



**HAL**  
open science

# ANALYSE DU METABOLOME URINAIRE HUMAIN PAR CHROMATOGRAPHIE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE A HAUTE RESOLUTION

Aurélie Roux

► **To cite this version:**

Aurélie Roux. ANALYSE DU METABOLOME URINAIRE HUMAIN PAR CHROMATOGRAPHIE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE A HAUTE RESOLUTION. Biochimie, Biologie Moléculaire. Université Pierre et Marie Curie - Paris VI, 2011. Français. NNT: . tel-00641529v2

**HAL Id: tel-00641529**

**<https://theses.hal.science/tel-00641529v2>**

Submitted on 21 Nov 2011

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

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

**UNIVERSITE PIERRE ET MARIE CURIE**  
**PARIS 6**  
**Ecole doctorale Iviv**  
**Année 2011**

Spécialité  
**CHIMIE ANALYTIQUE**

Présentée par  
**Aurélie ROUX**

Pour obtenir le grade de  
**DOCTEUR de l'UNIVERSITÉ PARIS 6**

---

**ANALYSE DU METABOLOME URINAIRE HUMAIN PAR CHROMATOGRAPHIE LIQUIDE  
COUPLEE A LA SPECTROMETRIE DE MASSE A HAUTE RESOLUTION**

---

**Thèse dirigée par les Dr. Christophe JUNOT et Pr. Jean Claude TABET**

**Soutenue le 21 Octobre 2011**

**Devant un jury composé de :**

**Dr. Christophe JUNOT**

**Pr. Jean Claude TABET**

**Pr. Gérard HOPFGARTNER**

**Pr. Pierre CHAMINADE**

**Pr. Germain TRUGNAN**

**Pr. Olivier LAPREVOTE**



# REMERCIEMENTS

Ce travail de thèse a été réalisé au CEA de Saclay, au sein du Service de Pharmacologie et d'Immunologie, dans le laboratoire d'Etude du Métabolisme des Médicaments. Je souhaite donc remercier en premier lieu le Dr. Christophe CREMINON, chef du service, et les Dr. Eric EZAN et Dr. Christophe JUNOT, successivement chef(s) de laboratoire, de m'avoir accueillie au sein du service et de m'avoir offert l'opportunité d'y effectuer ma thèse.

Je remercie sincèrement le Dr. Christophe JUNOT de m'avoir encadrée et guidée tout au long de ce travail, de s'être toujours montré à l'écoute et très disponible. Je tiens à remercier également le Pr. Jean-Claude TABET pour son aide précieuse et ses conseils avisés.

Je remercie le CEA et la DGA pour le financement de ma thèse.

Ce travail étant l'aboutissement de nombreuses collaborations, j'adresse mes remerciements à toutes les personnes qui y ont participé et plus particulièrement à Etienne THEVENOT, Cyrille PETAT, Arnaud MARTEL et François SEGUIN.

Je tiens à exprimer ma profonde gratitude aux membres du jury pour avoir accepté de juger ce travail de thèse.

Et enfin un grand merci toute l'équipe du SPI pour leur accueil chaleureux et avec qui j'ai passé de très bons moments.

Une thèse c'est un peu comme un parcours du combattant. Il faut être motivé et être bien entouré. Dans ce contexte, j'aimerais remercier toutes les personnes qui ont rendu cette expérience possible, qui m'ont soutenue tout au long de ces 3 ans.

Tout d'abord merci aux membres de l'équipe métabolome : Merci à toi Christophe pour la confiance et la liberté que tu m'as accordée, pour toutes les discussions très enrichissantes que nous avons eu ensemble et pour tout ce que tu m'as appris. Merci de m'avoir encadrée pendant ces 3 années. Je remercie également les anciens qui ont continué leur chemin, Jean-François pour ces précieux conseils et pour m'avoir initiée à l'utilisation du langage R, et Ying pour sa gentillesse et sa disponibilité. Merci également aux nouveaux, à Samia pour toutes nos conversations sur tout et n'importe quoi et des fous rires qui en découlent. Merci à Benoit pour ton humour et ta bonne humeur, et pour tous tes conseils sur la vie aux USA. Je remercie également toutes les personnes de passage dans l'équipe mais avec qui il a toujours été très agréable de travailler : aux twins et leur « clacos », Kinsley, Maria, Salma, Raoul, Dhouha, Kevin et Natali.

Je souhaite également remercier toutes les personnes du SPI, de Bertin Pharma et de Profilomic avec qui j'ai partagé déjeuners et conversations en tout genre. Merci notamment à François mon colocataire de bureau pour m'avoir supportée<sup>^</sup>. Merci à mes co-thésards, Cédric pour les nettoyages de hotte et les remises à niveau de la bulle, et Armelle sans qui j'aurais été STP (sans téléphone portable) il y a un an. Merci à mes collègues de jogging Céline et Céline, Mimine, Aurélie, Agnès, Alexandra... Merci aussi à Marie-Françoise, Pauline, Sophie, Sandrine, Laurie, Nicolas, Alexandre, Jérôme, François, Geoffrey, Stephanie, Denis, Jérôme, Victor... Un grand merci aussi à Christophe Creminon de m'avoir fait confiance pour les HNO et les manips un peu bizarres. J'ai vraiment adoré passer ces 3 ans avec vous, l'ambiance du labo et les personnes qui y travaillent vont me manquer c'est certain !

Je tiens ensuite à remercier ma famille, mes parents et ma sœur pour leur soutien sans faille. Merci d'être toujours là quand j'en ai besoin, et même pour les choses les plus futiles. Merci de m'avoir soutenue pendant mes études, de m'avoir fait totalement confiance quand je vous ai annoncé qu'après 6 ans d'études j'allais me lancer dans un doctorat de 3 ans, merci d'avoir tout fait pour que je réussisse, et si j'en suis arrivée là aujourd'hui c'est grâce à vous et je vous en serai toujours reconnaissante. Je vous aime.

Je remercie également du fond du cœur mes amis qui sont toujours là pour moi, avec qui je peux tout partager et qui ont su me redonner le moral quand ça n'allait pas. Merci Caroline Boulette pour tes gaffes et pour toutes les crises de rire que l'on a pu avoir ensemble. Et n'oublie pas, quand on a

soif il faut boire beaucoup de Quezac^^ Merci aussi pour les soirées « Déco » ou comment refaire la disposition de ton appart ! Merci Anaël Rebelle la princesse du groupe qui n'a jamais la langue dans sa poche. Merci pour ton franc-parler, ton déhanché endiablé et les échanges fringues ! Merci Marine Puceron pour l'inspiration que tu me prodigues et aussi pour les nombreuses soirées entre filles. Merci pour ton organisation quand on part en virée, merci pour les voyages à bord de la kiki mobile. Merci Anaëlle Coccinelle d'être toujours partante pour les trucs les plus improbables, de ne jamais t'énerver et d'avoir toujours de bons ragots à partager. Merci Lucie Bardet pour ton langage bien à toi, pour lequel il faut un décodeur certes, mais un jour on écrira un dictionnaire exprès^^. Merci pour ta blonditude c'est pour ça qu'on t'aime. Merci Clément Ourghanlian pour toute ta tendresse, j'en suis sûre plus tard on aura une grand maison avec cuisine intégrée et salle de bain toute équipée^^. Merci aussi à Noëlle et Delphine, car même si on ne se voit plus beaucoup on a quand même passé de très bons moments ensemble. Merci encore à vous, vous allez terriblement me manquer !

Merci aussi à tous les gens de la CEPPV, cette belle association dans laquelle j'ai passé presque 10 ans de ma vie, et aussi à toutes les personnes que j'ai connues à la faculté de pharmacie de Paris Descartes. Merci aux anciens pour le savoir qu'ils m'ont prodigué, et merci aux jeunes qui perpétuent la tradition et qui j'espère prendront autant de plaisir que j'en ai eu à être aux services des étudiants : Alexandra, Thaïs , Nico, Arnaud, Adrien, Dan, Antoine, PAX, Inès, Louise, Emilia, Julien Che', Samy, Marine, Thibaut, Jacques, Blanchot, Charles-Emmanuel, Julien Co', Roro, Léo, Pierrick, Hortense, Steph Pouet, Kelly... et tout ceux que je vais oublier...

Et enfin merci à l'ANEPF et aux 23 autres facultés de pharmacie de France pour tous ces bons moments partagés lors des AG, congrès mais aussi WE et sorties en tout genre. Merci Amiens pour les soirées au Club et pour les traversées nocturnes de votre belle ville (Merci à feu Maurice le canard). A Angers, petite fac en numéris clausus, mais grande fac de cœur. Merci Bezac pour l'accueil sous 15cm de neige, pas merci la SNCF. Merci Bordal pour le RPE cette année, c'était énorme ! Merci Caen pour tout, rallye, gala, embuscade et surtout pour votre bonne ambiance. Merci Clermont pour la verveine et les jeux à la con. Merci Dijon et Charles Trenet pour la mer ! Merci Grenoble pour vos chorégraphies toujours aussi difficiles à apprendre. Merci Lille de nous accueillir tous les ans à la braderie, merci pour la tournée des bars. Merci Limoges pour les cours de peinture sur corps. Vive Lyon pour ses combis Guantanamo. Merci Marseille de m'avoir entraînée un certain nombre de fois dans vos traquenards. Merci Montpellier pour les cours de corrida. Merci Nantes pour votre bonne humeur. Merci aux blondes de Nancy. Merci PXI d'être des gros boulets mais toujours partants pour faire la fête. Merci Poitiers d'avoir enfin organiser une AG. Merci les Pokémons de Rennes. Merci Reims pour votre champagne et votre bitume. Merci Rouen pour les bons moments passés au crit, à la dame de cœur et à la vodka Tabasco. Merci Strasbourg Illkirch-Graffenstaden. Merci Tours pour le RPE cette année également, vous avez été géniaux. Enfin merci Toulouse ;-)

Et vive la Pharmacie !



# TABLE DES MATIÈRES

LISTE DES FIGURES .....	XIII
LISTE DES TABLEAUX .....	XVII
LISTE DES ABRÉVIATIONS .....	XIX
INTRODUCTION GÉNÉRALE.....	2
<b>1 GENERALITES .....</b>	<b>4</b>
1.1. METABOLITES ET METABOLOME .....	5
1.2. BIOMARQUEURS ET METABOLOME.....	7
<b>2. L'APPROCHE METABOLOMIQUE.....</b>	<b>9</b>
2.1. LES DIFFERENTS TYPE D'APPROCHES.....	9
2.2. DEROULEMENT D'UNE ANALYSE METABOLOMIQUE .....	10
<b>3. PREPARATION DES ECHANTILLONS.....</b>	<b>11</b>
<b>4. METHODES D'ANALYSE.....</b>	<b>13</b>
4.1. LA RESONANCE MAGNETIQUE NUCLEAIRE.....	13
4.2. LES METHODES BASEES SUR L'UTILISATION DE LA SPECTROMETRIE DE MASSE .....	14
4.2.1. <i>Méthodes de spectrométrie de masse par introduction directe.....</i>	<i>19</i>
4.2.2. <i>Méthodes de spectrométrie de masse couplées à des méthodes séparatives .....</i>	<i>20</i>
4.3. APPROCHES ANALYTIQUES MULTIDIMENSIONNELLES .....	24
<b>5. TRAITEMENT DES DONNEES .....</b>	<b>24</b>
5.1. RESONANCE MAGNETIQUE NUCLEAIRE .....	25
5.2. CHROMATOGRAPHIE EN PHASE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE .....	25
<b>6. ANALYSE STATISTIQUE.....</b>	<b>31</b>
6.1. ANALYSE STATISTIQUE DESCRIPTIVE : ANALYSE EN COMPOSANTES PRINCIPALES (ACP).....	34
6.2. ANALYSE STATISTIQUE EXPLICATIVE : REGRESSION PLS.....	34
6.2.1. <i>L'analyse discriminante PLS.....</i>	<i>35</i>
6.2.2. <i>OPLS (Orthogonal Partial Least Square) [179].....</i>	<i>35</i>
<b>7. IDENTIFICATION DES VARIABLES D'INTERET .....</b>	<b>37</b>
7.1. GESTION DE LA REDONDANCE DU SIGNAL EN ESI .....	37
7.2. L'IDENTIFICATION EN SPECTROMETRIE DE MASSE .....	38
7.3. BASES DE DONNEES .....	40
<b>8. APPLICATIONS.....</b>	<b>42</b>
8.1. COMPREHENSION DES SYSTEMES VIVANTS .....	42

8.2.	APPLICATIONS EN BIOLOGIE CLINIQUE .....	44
8.3.	APPLICATIONS EN TOXICOLOGIE.....	46
<b>PRESENTATION DU TRAVAIL DE THESE.....</b>		<b>50</b>
<b>1</b>	<b>CONSTRUCTION D'UNE BASE DE DONNEES SPECTRALE POUR L'ANNOTATION DE JEUX DE DONNEES BIOLOGIQUES .....</b>	<b>52</b>
<b>2</b>	<b>EVALUATION DE L'IMPACT DE LA VARIABILITE BIOLOGIQUE SUR LE METABOLOME URINAIRE .....</b>	<b>53</b>
<b>3</b>	<b>CONDITIONS DE PRELEVEMENT ET DE STOCKAGE A COURT TERME DES URINES ET LEUR IMPACT SUR LE METABOLOME URINAIRE .....</b>	<b>53</b>
<b>CHAPITRE 1 : CONSTRUCTION D'UNE BASE DE DONNEES SPECTRALE POUR L'ANNOTATION DE JEUX DE DONNEES BIOLOGIQUES .....</b>		<b>54</b>
<b>1</b>	<b>INTRODUCTION .....</b>	<b>56</b>
<b>2</b>	<b>MATERIEL ET METHODES .....</b>	<b>61</b>
2.1	REACTIFS ET PRODUITS CHIMIQUES .....	61
2.2	ECHANTILLONS BIOLOGIQUES .....	61
2.3	PREPARATION DES ECHANTILLONS .....	62
2.3.1	<i>COMPOSES DE REFERENCE.....</i>	<i>62</i>
2.3.2	<i>ECHANTILLONS BIOLOGIQUES.....</i>	<i>62</i>
2.4	ACQUISITION DES DONNEES .....	62
2.4.1	<i>Appareillage.....</i>	<i>62</i>
2.4.2	<i>Analyses par spectrométrie de masse .....</i>	<i>63</i>
2.4.3	<i>Analyses par LC/MS.....</i>	<i>63</i>
2.4.4	<i>Séquences d'analyse.....</i>	<i>64</i>
2.5	TRAITEMENT DES DONNEES BIOLOGIQUES .....	65
2.5.1	<i>Conversion des fichiers de données en format universel.....</i>	<i>65</i>
2.5.2	<i>Déconvolution par XCMS.....</i>	<i>65</i>
2.6	ANALYSE STATISTIQUE .....	65
2.7	IDENTIFICATION.....	65
2.7.1	<i>Analyse des corrélations des abondances ioniques dans le domaine spectral .....</i>	<i>65</i>
2.7.2	<i>Analyse des corrélations des abondances ioniques inter-échantillons .....</i>	<i>66</i>
2.7.3	<i>Annotation par les bases de données publiques .....</i>	<i>66</i>
2.7.4	<i>Annotation par la base de données spectrale .....</i>	<i>66</i>
2.7.5	<i>Analyse structurale par expériences de fragmentations séquentielles.....</i>	<i>67</i>
<b>3</b>	<b>RESULTATS .....</b>	<b>68</b>
3.1	IDENTIFICATIONS A PARTIR DE COMPOSES DE REFERENCE .....	68
3.1.1	<i>Interprétation des spectres de masse des composés de référence .....</i>	<i>68</i>
3.1.2	<i>Annotation d'un jeu de données biologiques.....</i>	<i>69</i>
3.1.3	<i>Bilan des métabolites identifiés dans les urines humaines.....</i>	<i>70</i>
3.2	IDENTIFICATIONS A PARTIR D'UN JEU DE DONNEES BIOLOGIQUES .....	71
3.2.1	<i>Analyses statistiques.....</i>	<i>72</i>
3.2.2	<i>Regroupement des variables d'intérêt.....</i>	<i>75</i>

3.2.3	<i>Annotation des variables d'intérêt</i> .....	79
3.2.4	<i>Expériences complémentaires : MS<sup>n</sup></i> .....	80
3.2.5	<i>Interprétation des spectres de masse des composés putativement identifiés</i> .....	82
3.2.6	<i>Bilan des métabolites identifiés dans les urines humaines.</i> .....	83
<b>4.</b>	<b>CONCLUSION ET DISCUSSION</b> .....	<b>83</b>
<b>CHAPITRE 2 : EVALUATION DE L'IMPACT DES FACTEURS PHYSIOLOGIQUES SUR LE METABOLOME URINAIRE</b> .....		<b>88</b>
<b>1</b>	<b>INTRODUCTION</b> .....	<b>90</b>
<b>2</b>	<b>MATERIEL ET METHODES</b> .....	<b>94</b>
2.1	REACTIFS ET PRODUITS CHIMIQUES.....	94
2.2	ECHANTILLONS BIOLOGIQUES.....	94
2.3	PREPARATION DES ECHANTILLONS.....	94
2.4	ACQUISITION DES DONNEES.....	94
2.4.1	<i>Appareillage</i> .....	94
2.4.2	<i>Analyses par LC/MS</i> .....	94
2.4.3	<i>Séquences d'analyse</i> .....	94
2.5	TRAITEMENT DES DONNEES BIOLOGIQUES.....	94
2.6	ANALYSES STATISTIQUES.....	95
2.7	IDENTIFICATIONS.....	95
<b>3</b>	<b>RESULTATS</b> .....	<b>95</b>
3.1	ETUDE PRELIMINAIRE DE LA POPULATION.....	95
3.1.1	<i>Test d'indépendance entre les variables Genre et IMC</i> .....	95
3.1.2	<i>Test d'indépendance entre les variables Age et IMC</i> .....	96
3.1.3	<i>Test d'indépendance entre les variables Genre et Age</i> .....	97
3.1.4	<i>Choix des tests statistiques</i> .....	97
3.2	ANALYSES STATISTIQUES SUR LES DONNEES NON NORMALISEES.....	99
3.3	NORMALISATION.....	101
3.4	ANALYSES STATISTIQUES SUR LES DONNEES NORMALISEES.....	103
3.5	INTERPRETATIONS BIOLOGIQUES.....	108
3.5.1	<i>Variations liées à l'effet du genre</i> .....	110
3.5.2	<i>Variations liées à l'effet de l'âge</i> .....	112
3.5.3	<i>Variations liées à l'effet de l'IMC</i> .....	114
<b>4</b>	<b>CONCLUSION ET DISCUSSION</b> .....	<b>114</b>
<b>CHAPITRE 3 : CONDITIONS DE PRELEVEMENT ET DE STOCKAGE A COURT TERME DES URINES ET LEUR IMPACT SUR LE METABOLOME URINAIRE</b> .....		<b>119</b>
<b>1</b>	<b>INTRODUCTION</b> .....	<b>121</b>
<b>2</b>	<b>MATERIEL ET METHODES</b> .....	<b>123</b>
2.1	REACTIFS ET PRODUITS CHIMIQUES.....	123
2.2	ECHANTILLONS BIOLOGIQUES.....	123

2.3	PREPARATION DES ECHANTILLONS .....	124
2.4	ACQUISITION DES DONNEES .....	125
2.4.1	<i>Appareillage</i> .....	125
2.4.2	<i>Analyses par LC/MS</i> .....	125
2.4.3	<i>Analyses par RMN</i> .....	125
2.5	TRAITEMENT DES DONNEES BIOLOGIQUES .....	125
2.6	ANALYSES STATISTIQUES.....	126
2.7	IDENTIFICATIONS .....	126
<b>3</b>	<b>RESULTATS .....</b>	<b>126</b>
3.1	EVOLUTION DE LA CROISSANCE BACTERIENNE .....	126
3.2	ANALYSES STATISTIQUES MULTIVARIEES.....	127
3.2.1	<i>Impact de la contamination bactérienne sur les empreintes métaboliques</i> .....	130
3.2.2	<i>Impact de l'instabilité chimique sur les empreintes métaboliques</i> .....	133
3.3	VARIATIONS DES CONCENTRATIONS EN METABOLITES .....	134
3.3.1	<i>Métabolites impactés par la croissance bactérienne</i> .....	138
3.3.2	<i>Métabolites instables en solution</i> .....	140
<b>4</b>	<b>CONCLUSION ET DISCUSSION .....</b>	<b>141</b>
	<b>DISCUSSION GENERALE ET PERSPECTIVES .....</b>	<b>145</b>
	<b>ANNEXES.....</b>	<b>153</b>
	<b>ANNEXE 1 - SPECTROMETRIE DE MASSE : RAPPELS THEORIQUES .....</b>	<b>155</b>
<b>1.</b>	<b>LES DIFFERENTES SOURCES D'IONISATION.....</b>	<b>155</b>
1.1.	LES SOURCES BASEES SUR L'IONISATION EN PHASE GAZEUSE .....	155
1.2.	LES SOURCES BASEES SUR LA DESORPTION .....	157
<b>2.</b>	<b>LES DIFFERENTS ANALYSEURS.....</b>	<b>161</b>
2.1.	LES ANALYSEURS A BASSE RESOLUTION.....	161
2.2.	LES ANALYSEURS A HAUTE RESOLUTION .....	164
2.3.	LES DIFFERENTS MODES D'ACQUISITIONS.....	169
	<b>ANNEXE 2 – XCMS.....</b>	<b>173</b>
<b>1.</b>	<b>FONCTIONNEMENT D'XCMS [150].....</b>	<b>173</b>
1.1.	FILTRATION ET DETECTION DES PICS .....	173
1.2.	APPARIEMENT DES PICS ENTRE LES ECHANTILLONS .....	175
1.3.	ALIGNEMENT DES TEMPS DE RETENTION .....	176
1.4.	INTEGRATION DES PICS ET RECHERCHE DES VALEURS MANQUANTES .....	177
1.5.	ANALYSES STATISTIQUES DES RESULTATS ET VISUALISATION DES EIBPC .....	178
<b>2.</b>	<b>PARAMETRES XCMS.....</b>	<b>178</b>
	<b>ANNEXE 3 – LISTE DES METABOLITES IDENTIFIES DANS LES URINES HUMAINES.....</b>	<b>181</b>

<b>ANNEXE 4 – PUBLICATIONS, POSTERS ET COMMUNICATIONS ORALES .....</b>	<b>225</b>
PUBLICATIONS RELATIVES A LA THESE .....	225
PUBLICATIONS DANS LE CADRE DE COLLABORATIONS .....	225
POSTERS.....	225
COMMUNICATIONS ORALES.....	226
<b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>227</b>



# LISTE DES FIGURES

Figure 1 : Représentation schématique des technologies « Omiques ».....	5
Figure 2 : L'analyse métabolomique, une approche multidisciplinaire [42].....	10
Figure 3 : Schéma global d'un spectromètre de masse présentant les sources et analyseurs disponibles pour l'analyse de biomolécules. ....	15
Figure 4 : Types d'analyseurs utilisés pour les études métabolomiques par MS entre 2001 et 2010..	18
Figure 5 : Types de couplages utilisés pour les études métabolomiques par MS entre 2001 et 2010.	18
Figure 6 : Principe de l'électrophorèse capillaire de zone.....	23
Figure 7 : Représentation schématique en trois dimensions d'un chromatogramme obtenu par LC/MS .....	26
Figure 8 : Processus de traitement des données.....	26
Figure 9 : Analyses statistiques multivariées [42]. ....	31
Figure 10 : Exemple de matrice de données métabolomiques LC/MS.....	32
Figure 11 : Présentation des données lors des analyses multivariées. (A) carte factorielle du poids des individus et (B) carte factorielle du poids des variables. ....	35
Figure 12 : Présentation des données en OPLS.....	36
Figure 13 : Principe de la LC/ESI-MS et origine de la redondance du signal [42]. ....	38
Figure 14 : Processus d'indentification des molécules inconnues en métabolomique. ....	39
Figure 15 : Les principaux domaines d'application de la métabolomique.....	42
Figure 16 : Biosynthèse de l'acide ophtalmique dans les hépatocytes, et corrélation avec la concentration en glutathion hépatique [208]. ....	43
Figure 17 : Implication de la voie de synthèse du glutathion (schéma de gauche) dans la métabolisation du paracétamol (schéma de droite). ....	48
Figure 18 : Nombres de spectres disponibles dans les différentes bases de données spectrale entre 1969 et 2009 [246].....	57
Figure 19 : Comparaison de l'ionisation par impact électronique et l'électrospray. ....	58
Figure 20 : Annotation d'un jeu de données biologiques et mise en évidence de la redondance du signal en ESI-MS.....	59
Figure 21 : Construction d'une base de données spectrale ESI-MS .....	60
Figure 22 : Exemple de composition d'une séquence d'analyse d'urines humaines (HU) en mode d'ionisation positif. ....	64
Figure 23 : Liste des pics obtenus pour l'acide pantothénique en mode d'ionisation positif. ....	68
Figure 24 : Annotation de l'acide pantothénique à partir du composé de référence dans un jeu de données biologiques d'urines humaines. ....	69
Figure 25 : Mise en évidence du manque de résolution en C18.....	70
Figure 26 : Différentes étapes pour l'identification de variables inconnues à partir d'un jeu de données biologiques.....	71

Figure 27 : Score plot de l'ACP (échantillons biologiques + QC1 + QC2) en mode d'ionisation positif. Visualisation des données selon la nature des échantillons. ....	73
Figure 28 : Score plot de l'ACP (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction de l'ordre de passage.....	73
Figure 29 : Score plot de l'ACP (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction du genre. ....	74
Figure 30 : Score plot de la PLS-DA (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction du genre. ....	74
Figure 31 : Loading plot de la PLS-DA (échantillons biologiques) en mode d'ionisation positif. Visualisation des variables en fonction du genre .....	75
Figure 32 : Corrélations dans le domaine chromatographique : superposition des profils d'élution d'ions issus de molécules éluant au même temps de rétention .....	76
Figure 33 : Principe de la corrélation dans le domaine chromatographique [260].....	77
Figure 34 : Principe de la corrélation des abondances ioniques inter-échantillons [260].....	78
Figure 35 : Regroupement des signaux au temps de rétention 9.40 min en mode d'ionisation négatif. .....	79
Figure 36 : Annotation par les bases de données publiques (HMDB, KEGG et METLIN) des ions entre 9.33 et 9.67 minutes en mode d'ionisation négatif. ....	80
Figure 37 : Confirmation de l'identification du L-Urobilinogène par MS/MS .....	81
Figure 38 : Schéma de fragmentation du L-Urobilinogène. ....	81
Figure 39 : Annotation du L-urobilinogène à partir du spectre de masse .....	82
Figure 40 : Récapitulatif des métabolites présents dans la chimiothèque, la base de données spectrale, décrits ou non dans HMDB comme étant présents dans les urines et retrouvés ou non dans notre étude.....	83
Figure 41 : Différentes familles chimiques des métabolites identifiés.....	84
Figure 42 : Différences de structure entre banques et bases de données .....	91
Figure 43 : Evolution du nombre de banques et bases de données biologiques répertoriées par le NAR entre 1999 et 2011, classées par thématiques.....	92
Figure 44 : Test d'indépendance entre les variables Genre et IMC.....	96
Figure 45 : Test d'indépendance entre les variables Age et IMC .....	97
Figure 46 : Test d'indépendance entre les variables Age et Genre .....	97
Figure 47 : Evolution des abondances de l'Acide 4-méthyl-2-oxovalérique dans les deux modes d'ionisation, .....	102
Figure 48 : Evolution des abondances de plusieurs métabolites en mode d'ionisation positif, avant et après normalisation. ....	103
Figure 49 : Métabolites discriminants avant normalisation (rouge) et après normalisation (bleu). Les métabolites communs avant et après normalisation sont représentés en vert. ....	104
Figure 50 : Variations des abondances des hormones stéroïdiennes et des métabolites du métabolisme énergétique entre homme et femme (après normalisation).....	111
Figure 51 : Variations des abondances des métabolites du métabolisme des acides aminés et du tryptophane entre homme et femme (après normalisation) .....	112

Figure 52 : Variations des abondances des hormones stéroïdiennes et des métabolites du métabolisme des acides aminés avec l'âge (après normalisation).....	113
Figure 53 : Métabolisme de la caféine et variations des abondances de ces métabolites avec l'âge (après normalisation).....	114
Figure 54 : Préparation des échantillons pour les expérience de stabilité urinaire.....	124
Figure 55 : Evolution de la contamination bactérienne au cours des 3 expériences.....	127
Figure 56 : Score plot de l'ACP des données de stabilité (Expérience 3) en mode d'ionisation positif. ....	128
Figure 57 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. ....	129
Figure 58 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. Modèle restreint aux métabolites identifiés.....	129
Figure 59 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. Modèle restreint aux métabolites identifiés.....	130
Figure 60 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. ....	131
Figure 61 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif. Modèle restreint aux métabolites identifiés. ....	132
Figure 62 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif.....	132
Figure 63 Scpre plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif. Modèle restreint aux métabolites identifiés. ....	133
Figure 64 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif.....	134
Figure 65 : Métabolites retrouvés lors des études de stabilités des expériences 2 et 3.....	134
Figure 66 : Métabolites détectés en RMN connus pour varier avec la contamination bactérienne. ...	139
Figure 67 : Evolution de certains métabolites avec la contamination bactérienne pour l'expérience de Septembre 2010. ....	140
Figure 68 : Evolution de certains métabolites instables en solution pour l'expérience de Septembre 2010.....	141
Figure 69 : Schéma du principe de l'ionisation MALDI .....	158
Figure 70 : Schéma d'une source électrospray.....	160
Figure 71 : Schéma d'un analyseur quadripolaire [336].....	161
Figure 72 : Représentation schématique d'un triple quadripôle [337].....	162
Figure 73 : Représentation schématique d'un piège à ions 3D [339] .....	163
Figure 74 : Principe de fonctionnement d'un analyseur à temps de vol [337] .....	165
Figure 75 : Principe de fonctionnement du FT-ICR.....	167
Figure 76 : Piège électrostatique ou «Orbitrap» d'après la description de A. Makarov [369].....	169
Figure 77 : Stratégie de traitement des données mise en œuvre par XCMS. ....	173
Figure 78 : Illustration de la méthode de détection et d'intégration des pics par l'algorithme MatchedFilter de XCMS [150]. ....	174

Figure 79 : Exemple de regroupement des pic dans l'intervalle de m/z 337,975 à 338,225 (largeur 0,25m/z) [150].....	175
Figure 80 : Profil de déviation du temps de rétention [150].....	176
Figure 81 : EIBPC de deux groupes avant et après alignement des temps de rétention [150]. .....	177
Figure 82 : Exemple de superposition de chromatogrammes (gauche) et de boxplot (droite) générés par XCMS. ....	178

# LISTE DES TABLEAUX

Tableau 1 : Récapitulatif des performances des différents analyseurs [103] .....	17
Tableau 2 : Principaux logiciels utilisés pour le traitement des données métabolomiques. ....	28
Tableau 3 : Principales bases de données utilisées en analyse métabolomique [193].....	40
Tableau 4 : Classification des 227 individus en fonction de l'âge, du genre et de l'indice de masse corporel.....	61
Tableau 5 : Composition de la solution d'étalons internes.....	62
Tableau 6 : Méthodes chromatographiques utilisées pour l'analyse des échantillons. ....	64
Tableau 7 : Paramètres CAMERA utilisés .....	66
Tableau 8 : Paramètres utilisés lors des analyses MS/MS mode d'acquisition dépendant.....	67
Tableau 9 : Types de tests statistiques disponibles et critères de choix .....	98
Tableau 10 : Métabolites statistiquement significatifs pour au moins un des facteurs étudiés avant normalisation. ....	99
Tableau 11 : Métabolites statistiquement significatifs pour au moins un des facteurs étudiés après normalisation. ....	105
Tableau 12 : Métabolites statistiquement significatifs pour au moins un des trois facteurs physiologiques étudiés, hors métabolites marqués*. ....	108
Tableau 13 : Nombres de métabolites discriminants pour chaque facteurs en fonction de leur famille chimique. ....	110
Tableau 14 : Différentes études déjà publiées sur la conservation des urines à court, moyen et long terme.....	122
Tableau 15 : Différents échantillons et conditions utilisées pour les 3 expériences de stabilité urinaire. ....	124
Tableau 16 : Liste des métabolites discriminants dans au moins une des conditions des expériences 2 et 3.....	136
Tableau 17 : Métabolites connus pour varier avec la contamination bactérienne des urines [301;303] et variations observées lors des l'étude de stabilité. ....	138
Tableau 18 : Paramètres XCMS utilisés .....	179
Tableau 19 : Tableau récapitulatif des métabolites retrouvés dans les urines lors de l'analyse de la cohorte.....	182



# LISTE DES ABRÉVIATIONS

- ACN** : Acétonitrile
- AcB** : Acide borique
- ACP** : Analyse en Composantes Principales
- ADN** : Acide désoxyribonucléique
- AGC** (*Automatic Gain Control*) : Limite de commande automatique du gain
- ANOVA** (*ANalysis Of Variance*) : Analyse de variance
- APCI** (*Atmospheric Pressure Chemical Ionisation*) : Ionisation Chimique à Pression Atmosphérique
- APPI** (*Atmospheric Pressure Photo Ionisation*) : Photoionisation à Pression Atmosphérique
- ARN** : Acide ribonucléique
- ARNm** : Acide ribonucléique messager
- Asp** : Acide aspartique
- C<sub>18</sub>** (*Octadecyl*) : Chromatographie en phase liquide basée sur des interactions hydrophobes
- CAMERA** (*Collection of Algorithms for MEtabolite pRofile Annotation*) : Logiciel de traitement de données LC/MS
- CCM** : Chromatographie sur Couche Mince
- CE** (*Capillary Electrophoresis*) : Electrophorèse Capillaire
- CE/MS** : Electrophorèse Capillaire couplée à la spectrométrie de masse
- CEA** : Commissariat à l'Énergie Atomique
- ChEBi** (*Chemical Entities of Biological Interest*) : Base de données chimique orientée vers les composés chimiques de petite taille.
- CID** (*Collision Induced Dissociation*) : Dissociations induites par collision
- CV** : Coefficient de Variation
- CYP1A2** : Cytochrome P450 1A2
- D<sub>2</sub>O** : Eau deutérée
- Da** : Dalton
- DDBJ** (*DNA Data Bank of Japan*) : Base de données d'ADN Japonaise.
- DHEA** : Déhydroépiandrostérone
- DIOS** (*Desorption-Ionization mass spectrometry On porous Silicon*) : Désorption-ionisation sur silicium poreux
- EIBPC** (*Extracted Ion Base-Peak Chromatogram*) : Ion extrait à partir du chromatogramme
- EMBL** (*European Molecular Biology Laboratory*) : Base de données de génomique de l'institut européen de bioinformatique (EBI).
- ESI** (*Electrospray Source Ionization*) : Source d'ionisation électrospray
- FAB** (*Fast Atom Bombardment*) : Ionisation par bombardement d'ions ou d'atomes rapides
- FDR** (*False Discovery Rate*) : Probabilité que sur l'ensemble des tests significatifs (tests multiples) un certain pourcentage ne le soit en fait pas.
- FT-ICR** (*Fourier Transform Ion Cyclotron Resonance*) : Résonance cyclotronique des ions à transformée de Fourier
- FWHM** (*Full Width at High Maximum*) : Largeur du pic à mi-hauteur
- GC** (*Gas Chromatography*) : Chromatographie en phase gazeuse
- GC/MS** : Chromatographie en phase gazeuse couplée à la spectrométrie de masse
- Glu** : Gluamine
- H<sub>2</sub>O** : Eau
- HCOOH** : Acide formique
- HILIC** (*Hydrophilic Interaction Liquid Chromatography*) : Chromatographie en phase liquide basée sur des interactions hydrophiles
- HMDB** (*Human Metabolome DataBase*) : Base de données du métabolome humain

**HPLC** (*High Pressure Liquid Chromatography*): chromatographie liquide haute pression

**Ile** : Isoleucine

**IMC** : Indice de masse corporelle

**IR** : Infrarouge

**KEGG** (*Kyoto Encyclopedia of Genes and Genomes*) : Ensemble de bases de données relatives aux génomes, aux voies métaboliques et aux composés biochimiques.

**LC** (*Liquid Chromatography*) : Chromatographie en phase liquide

**LC/MS** : Chromatographie en phase liquide couplée à la spectrométrie de masse

**LCR** : Liquide céphalorachidien

**Leu** : Leucine

**LMCO** (*Low Mass Cut-Off*): limite de détection dans les basses masses

**LOQ** (*Limit Of Quantification*): Limite de quantification

**LOWESS** (*LOcally WEighted Scatterplot Smoothing*) : Méthode de regression locale

**LTDQ** (*linear trap quadrupole*): Triple quadripôle linéaire

**MALDI** (*Matrix Assisted Laser Desorption-Ionization*): Désorption-ionisation laser assistée par matrice

**MeOH** : Méthanol

**MetaCyc** : Base de données contenant un grand nombre d'informations sur le métabolisme et les enzymes de nombreux êtres vivants.

**METLIN** : Base de donnée métabolomique et MS/MS.

**MRM** (*Multiple Reaction Monitoring*) : détection de multiples réactions sélectionnées

**MS** (*Mass spectrometry*) : Spectrométrie de masse

**MS/MS** : Spectrométrie de masse en tandem

**MS<sup>n</sup>** : Fragmentation séquentielle d'ordre n

**m/z** : rapport masse sur charge

**NA** : Non applicable

**NIST** (*National Institute of standards and technology*) : Institut national des normes et de la technologie américain

**ns** : Non significatif

**NR** : Non retrouvé

**OPLS** (*Orthogonal Projection on Latent Structures*): Projection orthogonale sur structures latentes

**PBS** (*Phosphate buffered saline*) : Tampon phosphate salin

**PCR** (*Polymerase Chain Reaction*) : Réaction en Chaîne par Polymérase

**PFPP** (*PentaFluoroPhenylPropyl*) : Chromatographie en phase liquide basée sur des interactions mixtes, hydrophiles et hydrophobes.

**PIR** (*Protein Information Resource*) : Banque de données bio-informatique de séquences protéiques et d'outils d'analyse protéomique.

**PLS** (*Partial Least Square*) : Régression des moindres carrés partielle

**PLS-DA** (*Partial Least Square- Discriminant Analysis*) : Régression des moindres carrés partielle par analyse discriminante

**PubChem** : banque de données de molécules chimiques américaine gérée par le *National Center for Biotechnology Information* (NCBI), branche de la Bibliothèque nationale de médecine des États-Unis sous l'autorité de la *National Institutes of Health* (NIH).

**QC** (*Quality Control*) : Contrôle qualité

**Q-TOF** (*Quadrupole - Time Of Flight*) : Analyseur quadripolaire couplé à un analyseur à temps de vol

**RMN** : Résonance Magnétique Nucléaire

**RT** (*Room Temperature*): Température ambiante

**SIM ou SIR** (*Single Ion Monitoring ou Recording*) : Mode d'acquisition sélectif d'un ou plusieurs ions

**SIMS** (*secondary Ions Mass Spectrometry*) :  
Méthode d'ionisation par bombardement avec  
un faisceau d'ions primaires (généralement  
 $O^{2+}$ ,  $O^-$ ,  $Xe^+$  ou  $Ar^+$ ...) la surface d'un  
échantillon.

**S/N** (*Signal/Noise*) : rapport signal/bruit

**SRM** (*Selective Reaction Monitoring*) : Mode  
d'acquisition sélectif en mode MS/MS

**Th** (*Thomson*) : Unité de mesure relative à la  
spectrométrie de masse et correspondant à  
une unité de masse sur charge.

**TM ou TMA** : Triméthyle amine

**TMAO** : N-oxyde de triméthyle amine

**TOF** (*Time Of Flight*) : Analyseur à temps de  
vol

**TSP** (*TrimethylSilyl Propionate*) : Acide  
trimethylsilyl propanoïque (aussi appelé TMSP)

**tr** : temps de rétention

**UHPLC** (*UltraHigh Pressure Liquid  
Chromatography*) : chromatographie liquide à  
ultra haute pression

**Val** : Valine

**VIP** (*Variable Importance in the Projection*) :  
Score de significativité rendu par le logiciel  
SIMCA pour les analyses statistiques  
multivariées supervisées.

**XCMS** (*(X) of Chromatography Mass  
Spectrometry*) : Logiciel de traitement de  
données LC/MS



# INTRODUCTION GÉNÉRALE



# INTRODUCTION GÉNÉRALE

## 1 GENERALITES

Les avancées technologiques réalisées en biologie moléculaire ont permis d'aboutir dans le début des années 2000 au séquençage complet des génomes d'une vingtaine d'organismes [1] bactériens (*Escherichia coli*), de levures, de végétaux (riz, *Arabidopsis thaliana*) ou de l'espèce humaine. Un des enjeux scientifiques majeurs est maintenant de relier les gènes identifiés aux fonctions physiologiques d'une cellule ou d'un organisme grâce à des techniques d'analyse fonctionnelle.

L'exploration fonctionnelle du génome est actuellement réalisée à plusieurs niveaux :

- celui des **ARN** messagers avec l'étude du transcriptome
- celui des **protéines** exprimées à partir de ces ARN messagers avec l'étude du protéome
- celui des **métabolites** avec l'étude du métabolome

Le concept de métabolome fait référence à l'ensemble des métabolites contenus dans un système biologique donné : organisme, type cellulaire ou fluide biologique tel que l'urine ou le plasma [2]. Les métabolites sont des composés impliqués dans les processus métaboliques : substrats, produits ou cofacteurs de réactions enzymatiques ou simplement chimiques [3]. Le terme métabolite inclut par conséquent toutes les molécules de faibles masses moléculaires (en général <1500Da) telles que les acides organiques, les sucres, les acides gras, les métabolites conjugués, les acides aminés mais aussi certains peptides (comme le glutathion), les vitamines, les stéroïdes, les xénobiotiques et autres molécules exogènes [4].

Le terme métabolome a été conçu en référence au génome et aux autres approches omiques. Il est apparu pour la première fois dans une publication d'Oliver en 1998 [5]. L'étude du métabolome (ou métabolomique) s'inscrit au même titre que les études du transcriptome et du protéome dans un contexte post-génomique (Figure 1). Comparée aux études du transcriptome et du protéome, l'étude du métabolome possède des avantages complémentaires. Tout d'abord le métabolome est théoriquement plus facile à cerner dans sa globalité [6], il existe approximativement un facteur 10 entre le nombre de métabolites et le nombre de gènes (ex., 584 [4] et 6600 ([www.yeastgenome.org](http://www.yeastgenome.org)), respectivement chez *S.cerevisiae*). De plus, le métabolome, en tant que maillon final de la cascade des « omiques », représente l'ultime réponse d'un organisme à une altération génétique, une pathologie, une exposition à un toxique ou tout autre facteur susceptible de perturber son fonctionnement [7]. Il est donc le reflet le plus proche de la fonctionnalité et du phénotype d'une cellule

ou d'un organisme. Il est, tout comme le protéome, dépendant du contexte biologique et/ou environnemental. C'est-à-dire que les taux de protéines ou de métabolites sont modifiés en fonction de l'état physiologique, développemental, ou pathologique d'une cellule, d'un tissu, d'un organe ou d'un organisme [8]. En outre, il a été démontré par la théorie du contrôle métabolique [9-11], que des différences de concentration en métabolites intermédiaires doivent pouvoir être détectées par un effet d'amplification, alors que les modifications de concentrations en transcrits ou en protéines sont souvent très faibles ou indétectables [8].

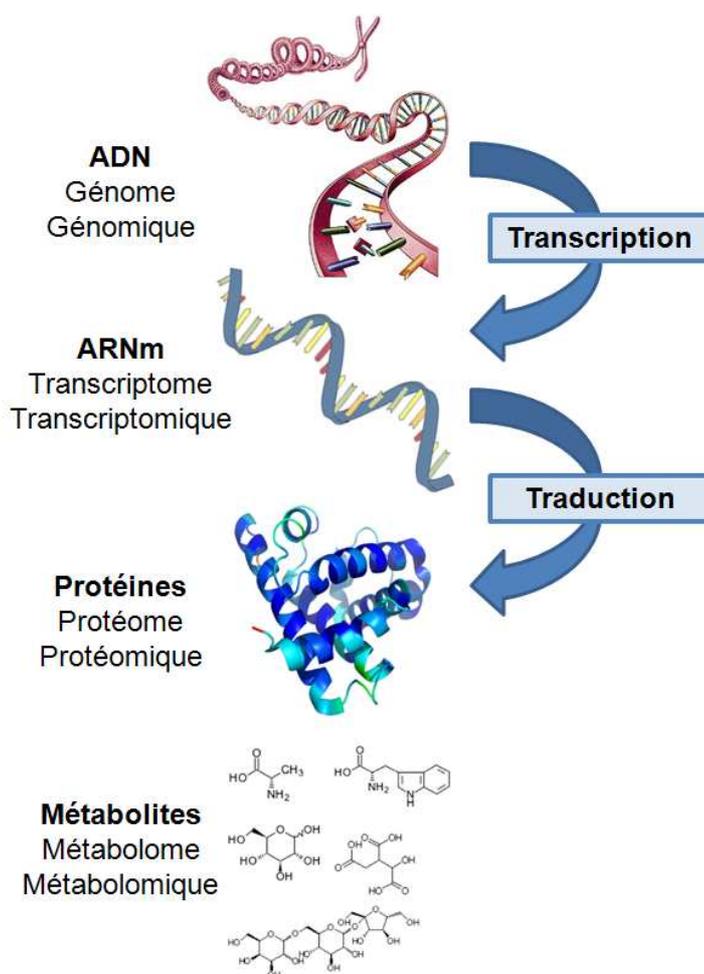


Figure 1 : Représentation schématique des technologies « Omiques ».

## 1.1. Métabolites et Métabolome

Pour certains scientifiques, la notion de métabolite inclut toutes les substances organiques d'origine naturelle provenant du métabolisme d'un organisme vivant et qui ne relèvent pas directement de l'expression des gènes. En fait, deux types de métabolites peuvent être distingués en fonction de leur origine : métabolites endogènes et exogènes.

Les métabolites endogènes peuvent être classés en métabolites primaires et secondaires. Les premiers sont ubiquitaires (règnes bactérien, végétal et animal) et sont directement impliqués dans

des processus essentiels à la vie tels que la croissance, le développement et la reproduction. C'est par exemple le cas pour les acides aminés ou des intermédiaires de la glycolyse. Les métabolites secondaires sont propres à chaque espèce, ont une distribution restreinte et sont synthétisés en vue d'une fonction biologique particulière. Citons par exemple les alcaloïdes chez les plantes ou des hormones chez les mammifères [12].

Les métabolites exogènes à un organisme sont les métabolites qui ne sont pas synthétisés directement par ce dernier. Cela concerne notamment les xénobiotiques qui représentent les composés chimiques synthétisés par l'Homme et leurs produits de biotransformation ou de métabolisation, résultant de réactions de phase I (modification de la molécule initiale visant à introduire un groupe fonctionnel) et/ou de conversion enzymatique de phase II (conjugaison) [13], mais aussi toute molécule étrangère à un organisme (par exemple la caféine issue du café n'est pas un xénobiotique mais un métabolite secondaire exogène pour l'Homme). Dans ce contexte particulier, Holmes et al. [3] ont proposé le concept de xénométabolome qui est une description du profil métabolique de xénobiotiques d'un individu exposé à des polluants environnementaux, médicaments, ou à des molécules exogènes provenant des aliments/composants alimentaires tels que les composés phytochimiques [14]. Crockford et al. [15] ont démontré le potentiel de cette approche en identifiant les métabolites de médicaments comme le paracétamol ou le disopyramide par hétérospectroscopie sur des empreintes urinaires acquises avec 600 MHz RMN  $^1\text{H}$  et UPLC/TOF-MSE obtenues à partir de plus de 80 patients. Ce concept étend l'approche développée dans le début des années quatre-vingt dix dans le domaine de l'épidémiologie moléculaire [16;17], grâce aux progrès techniques en chimie analytique.

Appliquée aux études épidémiologiques, l'analyse du xénométabolome pourrait améliorer la caractérisation des cohortes, notamment en permettant de valider les réponses aux questionnaires (observance à un traitement médicamenteux par exemple), mais également en documentant des expositions environnementales ou professionnelles à des produits chimiques. Ainsi, l'analyse du métabolome renseigne sur l'ensemble du métabolisme et des interactions d'un individu avec son environnement, alors que celle de l'expression des gènes (analyse transcriptomique) renseigne plutôt sur les possibilités d'adaptations fonctionnelles d'un organisme face à l'environnement. Les informations ainsi récoltées par l'analyse des métabolites dans les fluides biologiques sur le métabotype (phénotype métabolique) d'un individu ou d'une population peuvent être appliquées à la médecine personnalisée ou la santé publique [18]. L'analyse métabolomique permet donc d'accéder à un grand nombre de métabolites d'un organisme, qu'ils soient directement issus des gènes et protéines, ou qu'ils résultent des interactions de l'organisme avec son environnement. Ainsi on peut considérer l'analyse du métabolome comme une approche ouvrant des perspectives prometteuses dans des domaines aussi variés que la nutrition [19], le diagnostic de pathologie ou de toxicité (caractérisation de xénobiotiques [20], la mise en évidence de métabolites disrupteurs endocriniens [21]...), ou encore la médecine personnalisée et l'épidémiologie (phénotypage de cohorte [22], biomarqueur d'exposition).

## 1.2. Biomarqueurs et Métabolome

La relation entre la composition d'un fluide biologique et la présence d'une pathologie ou d'un toxique a été explorée depuis l'antiquité. On a par exemple longtemps diagnostiqué la présence d'un diabète par la saveur sucrée des urines (Susruta et Charaka, 2000 ans avant J.C., Papyrus de Thèbes 1550 ans avant J.C., Pharmaceutice rationalis T ;Willis 1674). En fait, chaque individu (quelle que soit l'espèce vivante) possède un état d'équilibre biologique appelé homéostasie. Initialement pressenti par Claude Bernard [23], le concept d'homéostasie fait référence à la capacité que peut avoir un organisme vivant à conserver son équilibre de fonctionnement en dépit des contraintes qui lui sont extérieures. Le terme fut formalisé par Walter Bradford Cannon qui parlait de l'homéostasie comme «la sagesse du corps». Il la définissait comme la stabilisation des états qui permettent les processus biologiques de la vie, telle qu'elle ressort de cette phrase extraite de son ouvrage «The Wisdom of the Body» : «Les êtres vivants supérieurs constituent un système ouvert présentant de nombreuses relations avec l'environnement. Les modifications de l'environnement déclenchent des réactions dans le système ou l'affectent directement (...) [24]». Les interactions avec l'environnement (exposition à des médicaments, produits chimiques...) ou l'apparition d'une maladie sont susceptibles de perturber cette homéostasie à différents niveaux de l'organisation biologique, y compris celui du métabolome. Les concentrations de métabolites endogènes peuvent être modifiées et des xénométabolites peuvent apparaître. Alors que ces derniers sont évidemment des marqueurs de l'exposition (biomarqueurs d'exposition), des signatures spécifiques de la maladie ou de l'exposition (souvent appelé profil métabolomique) ont pu être mises en évidence par l'analyse détaillée des métabolites endogènes.

La métabolomique en tant qu'approche globale peut alors permettre une meilleure compréhension des changements liés à une maladie ou à un effet pharmacologique [25], en mettant en évidence des interrelations métaboliques qui n'auraient pas pu être détectées avec des approches biochimiques ciblées traditionnelles. La notion de profil métabolique a été introduite par Williams en 1949 [26;27] qui a utilisé la chromatographie sur papier pour comparer les urines de 200000 sujets, parmi lesquels figuraient des alcooliques, des schizophrènes et des patients des hôpitaux psychiatriques. Il a démontré que certaines caractéristiques du profil métabolique pouvaient être associées à chacun de ces groupes. Griffiths et Wang [28] ont rapporté que les origines de la métabolomique remontent aux années 60 et 70 avec les travaux des Horning. Ils ont publié plusieurs articles sur la détermination de profils métaboliques dans l'urine par chromatographie en phase gazeuse couplée entre autre à la spectrométrie de masse [29;30]. Dans le même temps, Robinson et Pauling ont effectué une analyse quantitative de la vapeur d'urine et d'haleine par chromatographie en phase gazeuse pour le suivi de l'état de santé et le diagnostic de pathologies [31].

Les marqueurs biologiques ou biomarqueurs sont des indicateurs internes mesurables à la suite de modifications ou d'altérations cellulaires ou moléculaires qui peuvent apparaître dans un organisme pendant ou après une maladie ou suite à l'exposition à un toxique [32;33]. Cette définition, qui est utilisée en toxicologie environnementale et professionnelle, est plus large que celle de l'Institut

national de la santé américain (NIH) qui définit un biomarqueur comme «une caractéristique biochimique ou non (comme par exemple la mesure de la tensions artérielle) qui est objectivement mesurée et évaluée comme un indicateur de processus biologiques normaux, processus pathogènes, ou de procédés pharmacologiques pour une intervention thérapeutique» [34]. On considère comme biomarqueurs des composés ou un ensemble de composés (profil métabolomique) spécifiques, facilement mesurables et obtenus de manière, si possible, non invasive [35].

Ils peuvent être divisés en plusieurs catégories qui incluent les biomarqueurs d'exposition, les biomarqueurs d'effet et les biomarqueurs de susceptibilité. Un biomarqueur d'exposition renseigne quant à l'existence et l'importance de l'exposition à un toxique. La biosurveillance de l'exposition est utilisée depuis longtemps en milieu professionnel par exemple, pour la détermination des concentrations de Plomb [36] ou de métabolites du benzène [37] dans le sang ou l'urine.

Les biomarqueurs d'effet indiquent que l'exposition a donné lieu à une interaction entre la substance toxique et une cible biologique. Les substances mutagènes et/ou cancérigènes qui possèdent une ou plusieurs fonctions électrophiles se lient par exemple aux macromolécules biologiques comme les protéines, l'ADN ou les lipides. Les réactions chimiques impliquant l'hémoglobine sont souvent utilisées dans la biosurveillance en raison de la longue durée de vie des globules rouges (120 jours) et de sa facilité d'obtention. Citons également le stress oxydatif qui perturbe l'homéostasie de la cellule et conduit à la production de substances spécifiques comme la 8-hydroxy-2'-désoxyguanosine ou à des perturbations au niveau de la voie de biosynthèse ou de dégradation du glutathion [36].

Les biomarqueurs de susceptibilité décrivent les différences inter-individuelles en réponse à des substances toxiques, des facteurs génétiques ou non (âge, maladie du foie, des reins, alimentation, suppléments alimentaires...). Les polymorphismes des enzymes de détoxification et/ou d'activation ont été identifiés comme des facteurs clés dans ces variations entre individus. Haufroid et al. [38] ont ainsi démontré la relation entre l'excrétion urinaire d'acide phénylhydroxyéthylmercapturique (un métabolite issu d'une réaction entre le styrène et le glutathion) et le polymorphisme génétique de la glutathion S-transférase M1. Dans le même esprit, une approche dénommée «pharmacométabolomique», a été proposée afin d'étudier la réponse aux médicaments [39;40]. Dans ce contexte particulier, la métabolomique est utilisée comme un outil de génomique fonctionnelle.

Malgré les avancées réalisées dans le domaine analytique, qui ont grandement facilité la mise en évidence de biomarqueurs d'exposition ou d'effet, les biomarqueurs de susceptibilité restent beaucoup plus délicats à mettre en évidence. La variabilité inter-individuelle de la réponse d'un organisme à un facteur extérieur est multifactorielle. Notre connaissance très limitée des voies métaboliques ne nous permet pas toujours d'appréhender la complexité des réseaux métaboliques impliqués. Il est donc nécessaire d'obtenir une vision la plus intégrée possible qui passe notamment par la prise en compte des interactions individu-environnement pour la mise en évidence desquelles l'analyse métabolomique constitue une approche de choix.

## 2. L'APPROCHE METABOLOMIQUE

L'analyse du métabolome repose sur l'obtention d'empreintes métaboliques les plus exhaustives possibles (approche globale), ou de profils métaboliques plus spécifiques (approche ciblée), obtenus à l'aide de différentes méthodes analytiques qui seront détaillées ci-après. L'analyse différentielle de ces empreintes issues de différents groupes d'échantillons vise à caractériser les répercussions de la modification d'un facteur externe, et à mettre en évidence la manière dont un système biologique y répond.

### 2.1. Les différents type d'approches

Il existe différents types d'approches selon l'étape à laquelle l'identification des métabolites est réalisée. L'approche globale, ou non ciblée, est utilisée pour l'analyse du métabolome dans son ensemble, c'est-à-dire l'analyse complète de tous les métabolites mesurables dans un échantillon (y compris des signaux provenant d'analytes non encore caractérisés) que l'on cherchera à identifier par la suite. A l'inverse, l'approche ciblée se focalise sur un nombre limité de métabolites connus et prédéfinis avant l'analyse en elle-même afin de déterminer avec précision leurs concentrations (relatives ou absolues).

- **L'analyse métabolomique globale**

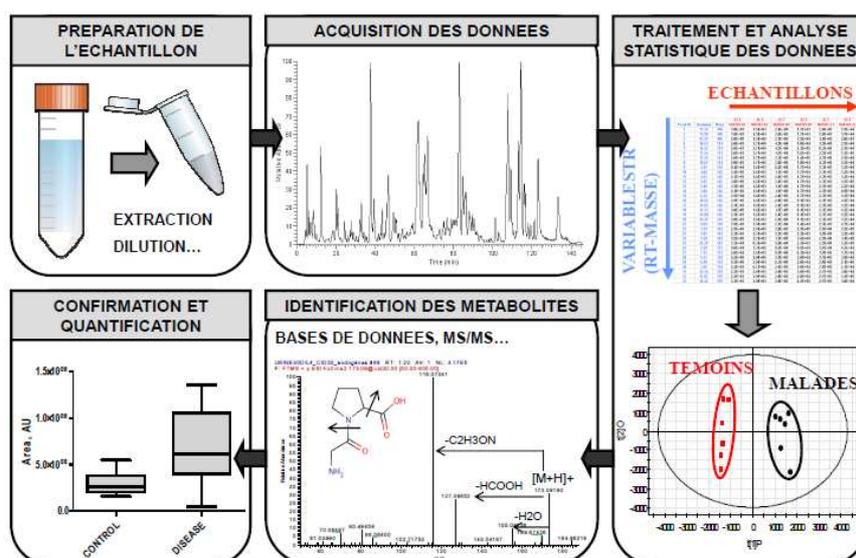
Elle a pour ambition la comparaison d'empreintes métaboliques obtenues avec différentes techniques (chromatographie en phase liquide ou gazeuse couplées à la spectrométrie de masse, infra-rouge ou encore résonance magnétique nucléaire), l'identification et parfois la quantification non biaisée des métabolites présents dans un échantillon biologique prélevé dans des conditions données. Cette approche vise à donner une vue d'ensemble de l'état métabolique d'un organisme, afin de repérer les réseaux de régulation, et/ou à construire un modèle statistique prédictif pour la classification ultérieure d'échantillons inconnus. Elle cherche ainsi à comparer des empreintes de métabolites dont les concentrations sont modifiées en réponse à une pathologie ou à une exposition à un toxique et à donner une signification biologique aux variations observées. Deux termes, métabolomique et métabonomique, sont employés de façon interchangeable selon les équipes pour décrire de telles approches. Le premier est historiquement employé pour des études concernant des systèmes cellulaires simples et surtout dans le cadre d'études d'organismes végétaux, et le second pour des études de systèmes biologiques intégrés, tels que des analyses d'urines ou de tissus, dans le cadre d'analyses toxicologiques. Un consensus est apparu récemment et tend à soutenir que la « métabolomique » fait référence à l'analyse de profils métaboliques exhaustifs tandis que la « métabonomique » s'intéresse à des changements métaboliques multiples (mais pas forcément exhaustifs) en réponse à un stress biologique (maladie, exposition à un agent toxique, ...) qui permettraient de comprendre le mécanisme de réponse de l'organisme au phénomène biologique étudié [41].

- **L'analyse métabolomique ciblée**

Elle a pour objet de mesurer spécifiquement un nombre réduit de métabolite comme les substrats ou les produits d'une réaction enzymatique par exemple. Elle peut aussi se focaliser sur des composés appartenant à une voie métabolique ou à une famille chimique donnée, on parle alors de profilage métabolique. Ce type d'approche a pour but une meilleure compréhension de la fonction ou de la régulation d'une voie métabolique et/ou de ses liens avec les autres voies métaboliques. Elle fait le plus souvent appel à des méthodes d'extraction sélectives de métabolites, ou plus généralement à une étape de traitement de l'échantillon importante, permettant d'optimiser la détection des analytes au sein d'une matrice biologique complexe. Les analyses effectuées dans ce cadre sont souvent quantitatives, précises et sensibles et s'apparentent à des dosages classiques en milieu biologique.

## 2.2. Déroulement d'une analyse métabolomique

L'analyse du métabolome est une approche multidisciplinaire mettant en jeu la chimie analytique, le traitement du signal et l'analyse statistique. Une expérience d'analyse métabolomique débute par un protocole expérimental approprié afin de veiller à ce que les données soient par la suite biologiquement interprétables. L'analyse commence par le traitement des échantillons biologiques avant l'injection dans le système analytique. Les profils métaboliques obtenus (données brutes) sont ensuite traités par des logiciels spécialisés dans la détection automatique des signaux, avant d'être analysés avec des outils statistiques. Enfin, l'identification des signaux discriminants est réalisée en combinant l'analyse de spectres, la consultation de bases de données et autres outils informatiques et mathématiques. Toutes ces étapes sont présentées dans la Figure 2 et seront détaillées dans les chapitres suivants.



### 3. PREPARATION DES ECHANTILLONS

Avant même de procéder au recueil et au traitement des échantillons, il convient de réfléchir à la démarche expérimentale à mettre en place. L'un des problèmes en métabolomique est l'existence de facteurs confondants qui masquent le phénomène biologique que l'on souhaite étudier. Ces facteurs confondants peuvent être d'origine analytique ou biologique et leur impact doit être contrôlé, autant que faire se peut, par un plan expérimental approprié.

De nombreux facteurs confondants présumés d'origine biologique ont déjà été répertoriés dans la littérature : l'âge [43], le genre [44], les effets chronobiologiques [45;46], l'impact des espèces animales et des souches [47;48], l'influence des facteurs environnementaux tels que l'alimentation, le mode de vie [49] ou encore l'impact de la microflore intestinale [50-52]. Certains de ces facteurs (l'alimentation ou le mode de vie, par exemple) varient d'un sujet à l'autre et ne peuvent pas être facilement contrôlés. Malheureusement dans la plupart des cas la seule option possible est de garder à l'esprit que ces facteurs existent et peuvent engendrer des variations des profils métaboliques. D'autres facteurs comme l'âge et le genre doivent être équilibrés dans les différents groupes afin de limiter leur impact sur les analyses statistiques. Dans ce contexte, il est particulièrement important de bien caractériser les profils métaboliques de fluides biologiques de sujets sains afin d'évaluer au préalable l'impact de ces facteurs physiologiques sur les concentrations en métabolites [53;54].

Une autre question importante est la normalisation des résultats. Cela est particulièrement important pour les échantillons urinaires. En effet, contrairement à la plupart des fluides biologiques ou des échantillons de tissus pour lequel la concentration de métabolites est clairement liée au volume ou la quantité prélevée, les concentrations urinaires de métabolites sont très fluctuantes en raison des variations du volume urinaire et de clairance rénale d'un prélèvement à l'autre, ce qui rend la normalisation des données nécessaire. Elle peut être réalisée en pondérant les abondances des signaux dans chaque échantillon par le volume urinaire, la concentration de créatinine ou l'osmolalité (sous réserve que ces paramètres ne soient pas eux-mêmes impactés par la pathologie ou le toxique étudié comme pour la toxicité glomérulaire rénale qui affecte la clairance à la créatinine, par exemple), ou encore par le signal total enregistré à partir des spectres [55], selon le type d'échantillon (spot urinaire ou collecte de 24h) et les informations disponibles (créatinine ou volume urinaire, par exemple).

A côté de ces facteurs confondants biologiques, les facteurs analytiques doivent également être pris en considération. Citons par exemple les variations d'abondance des signaux liés à l'encrassage progressif de la source électrospray au cours du déroulement de l'expérience. Cela conduit à une discrimination claire entre les échantillons analysés au début et à la fin d'une expérience, qui peut parfois gêner la visualisation de l'effet biologique d'intérêt. Pour limiter l'impact de ce type de biais, les échantillons peuvent être répartis aléatoirement au sein de la séquence d'analyse. On peut également faire appel à des algorithmes qui permettent de compenser ces effets, comme cela a déjà été publié dans le domaine de l'analyse lipidomique [56].

La qualité de la préparation des échantillons est un également un élément clé. Ce traitement de l'échantillon peut avoir plusieurs objectifs comme l'élimination d'une partie de la matrice biologique complexe (sérum, plasma, urine,...) afin de rendre l'échantillon compatible avec la technique analytique. Pour les analytes peu concentrés, il peut inclure une étape de pré-concentration. L'extraction des métabolites dépend fortement du type de milieu biologique (extraits cellulaires ou fluides biologiques), mais aussi du type de composés que l'on cherche préférentiellement à étudier comme dans les approches ciblées (composés polaires ou apolaire comme les lipides avec la méthode de Folch par exemple [57]). En revanche, en cas d'approche métabolomique globale, le traitement de l'échantillon se doit alors d'être le plus simple possible afin de limiter les sources de variabilité externes à l'échantillon et surtout la perte d'informations. Il est cependant nécessaire de retirer les éventuels sels et macromolécules qui pourraient détériorer la séparation chromatographique et/ou l'ionisation des petites molécules d'intérêt. Pour cela, quatre stratégies peuvent être envisagées : l'extraction sur phase solide (a), l'extraction liquide/liquide (L/L) (b), la précipitation par ajout de solvants organiques miscibles à l'eau (c), et les méthodes de filtration comme l'ultracentrifugation ou la dialyse (d).

Plusieurs études rapportent l'analyse d'échantillons bruts, sans aucun traitement, comme c'est par exemple le cas pour les urines [58;59] qui sont souvent simplement diluées avec de l'eau avant injection [60-62]. Les méthodes (a) et (b) permettent d'obtenir des mélanges plus «propres» mais la perte de certaines classes de molécules peut être observée [63]. Elles sont généralement utilisées pour les approches ciblées ou semi-ciblées. La méthode (c) est en général la plus utilisée car elle permet la précipitation de la plupart des macromolécules alors que la majorité des molécules d'intérêt restent en solution. Elle est notamment utilisée de façon spécifique pour la précipitation des protéines lors de l'analyse de fluides biologiques riches en protéines tels que le plasma ou le liquide céphalo-rachidien [64;65]. En effet une précipitation lente des protéines permet de limiter la quantité de métabolites piégés ou liés par ces dernières [66]. Pour les extraits cellulaires, le même genre de procédure peut être appliqué : après avoir été centrifugé pour séparer les cellules du surnageant, le culot cellulaire est remis en suspension dans un mélange de solvants aqueux et organiques, puis traité par ultrasons ou agité mécaniquement pour lyser les membranes cellulaires [67]. Quant aux tissus, ils doivent d'abord être rapidement prélevés et congelés en les plongeant dans l'azote liquide, par exemple. S'en suit une homogénéisation à froid dans des solvants organiques comme le méthanol. Un protocole d'extraction dérivé de la méthode Folch peut également être utilisé soit pour nettoyer la fraction polaire des lipides insolubles, soit pour analyser les fractions polaires et apolaires [68;69]. Le principal problème de ce type d'extraction concerne les métabolites de polarité intermédiaire (phospholipides par exemple) qui sont susceptibles de se distribuer entre les deux phases et pour lesquels on ne pourra pas avoir de quantification (absolue ou relative) pertinente. Selon le type de solvant organique utilisé pour l'extraction des métabolites, les échantillons sont dilués dans la phase mobile ou centrifugés, évaporés à sec et finalement remis en suspension dans un solvant compatible avec l'injection dans le système analytique.

## 4. METHODES D'ANALYSE

Même pour les organismes les mieux connus, la taille exacte de leur métabolome reste indéfinie. Certains auteurs ont néanmoins avancé des chiffres et proposent par exemples 2000 métabolites majeurs chez l'homme [70], et jusqu'à 200000 métabolites pour le règne végétal [71]. Les métabolomes sont des systèmes complexes et dynamiques. Ainsi, la distribution de nombreux métabolites est fortement soumise à des variations temporelles et spatiales avec par exemple l'existence de fluctuations circadiennes chez les mammifères. De plus, chez les mammifères, une variabilité biologique, induite par l'environnement et notamment par l'alimentation, est présente et vient encore complexifier l'analyse [72]. Enfin, les métabolites diffèrent aussi bien au niveau de leur taille, de la nature de leurs groupements fonctionnels, de leur volatilité, de leur polarité que de leurs concentrations physiologiques. A ce jour, il n'existe donc pas de méthode analytique universelle, capable de s'intéresser à l'ensemble de ces molécules. Chaque méthode analytique présente ses avantages et ses limitations. Le choix de la méthode analytique employée pour l'acquisition des empreintes métaboliques dépend donc des résultats attendus et d'éventuelles voies métaboliques d'intérêt et tient lieu de compromis entre rapidité, sélectivité et sensibilité.

De nombreuses méthodes ont été appliquées à des études métabolomiques : la spectroscopie Infrarouge à transformée de Fourier [5], l'absorption UV-visible et la fluorescence [73], la chromatographie sur couche mince (CCM) [2], la coulométrie [74], l'électrochimie [75], la résonance magnétique nucléaire (RMN), la spectrométrie de masse par pyrolyse [76], l'électrophorèse capillaire couplée à la Spectrométrie de Masse (CE/MS) [77;78], la chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC/MS) ou encore la chromatographie en phase liquide couplée à la spectrométrie de masse (LC/MS) [79;80] et la spectrométrie de masse à très haute résolution en injection directe avec des analyseurs à Résonance Cyclotronique des Ions à Transformée de Fourier (FT-ICR) [63;81]. Parmi toutes ces méthodes, les plus utilisées à l'heure actuelle sont la RMN, la GC/MS et la LC/MS, qui en plus de permettre des approches chimométriques, donnent accès à des informations sur les structures chimiques des métabolites, conduisant à leur possible identification afin de les replacer dans un contexte biologique.

### 4.1. La résonance magnétique nucléaire

La RMN est l'une des premières méthodes d'analyse à avoir été appliquée avec succès à l'étude du métabolome [82-85]. Elle bénéficie ainsi d'une vingtaine d'année de recul dans le domaine de la toxicologie animale. Cette méthode est basée sur l'existence de transitions entre les deux états énergétiques des noyaux atomiques ayant un moment magnétique non nul ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  ou  $^{31}\text{P}$  par exemple) qui existent lorsqu'ils sont placés dans un champ magnétique intense. La RMN présente un certain nombre d'avantages : elle est non destructrice, rapide, simple à mettre en œuvre, robuste, elle génère des spectres riches en informations structurales directes et dont les signaux subissent peu de décalage au cours de l'ensemble de l'analyse. De plus elle ne nécessite pas de présélection des conditions employées ce qui lui procure un avantage par rapport aux techniques chromatographiques

pour lesquelles les métabolites observés dépendent des conditions expérimentales (colonnes et conditions d'élution). Cependant, elle reste une méthode d'analyse peu sensible qui nécessite une grande quantité d'échantillon par analyse (souvent au-delà du milligramme). Si la présence universelle d'hydrogène dans les biomolécules fait de la RMN du proton ( $^1\text{H}$  RMN) une méthode de choix, la plupart des métabolites possédant chacun un grand nombre de signaux, l'attribution des signaux peut s'avérer délicate. En effet, les spectres de RMN obtenus lors de l'analyse de biofluides contiennent les résonances de centaines de métabolites, dont un certain nombre se recouvrent formant des multiplets indiscernables. Une deuxième dimension de RMN ( $^{13}\text{C}$  par exemple) ou des méthodes chromatographiques peuvent alors être couplées à la RMN  $^1\text{H}$ . Une méthode en deux dimensions utilisant la RMN  $^1\text{H}$ - $^{13}\text{C}$  HMBC (de l'anglais : *Heteronuclear Multiple Bonding Connectivity*) a ainsi été envisagée par Dumas et al [76]. La RMN 2D génère pour un même composé plusieurs signaux qui sont très corrélés. Un modèle utilisant la formation de dendrogrammes a été mis en place. Il oppose deux types de corrélations. Les fortes corrélations positives correspondent à des relations structurales entre les signaux, elles sont donc caractéristiques de signaux appartenant à un même composé. Les corrélations positives ou négatives moins marquées peuvent expliquer un lien physiologique entre ces signaux faiblement corrélés [76;86]. Enfin la caractérisation de molécules inconnues est délicate avec cette seule technique puisqu'elle ne donne aucune information sur la masse moléculaire des composés. Les récentes avancées technologiques ont permis d'améliorer sa sensibilité grâce au développement de la rotation à l'angle magique pour les échantillons solides [87] et des cryosondes [88;89]. Il faut ajouter que la RMN est la seule technique qui permet d'analyser des milliers d'échantillons provenant de cohortes de plusieurs milliers d'individus [49;90].

## 4.2. Les méthodes basées sur l'utilisation de la spectrométrie de masse

La spectrométrie de masse est une technique analytique mature qui s'est fortement implantée dans le domaine biomédical. Elle consiste à mesurer le rapport masse/charge ( $m/z$ ) d'espèces élémentaires ou moléculaires. De nombreuses évolutions techniques y ont régulièrement été apportées. A titre d'exemple, le développement des sources MALDI et ESI a permis l'essor de cette technique dans le domaine de la biologie. En métabolomique, le développement et les améliorations apportées aux analyseurs dont les temps de vol (*Time Of Flight*, TOF), FT-ICR et l'Orbitrap<sup>®</sup> ont donné accès à des informations plus précises et plus spécifiques.

La spectrométrie de masse offre plusieurs avantages pour les études métabolomiques. Elle est sensible et permet la détection d'espèces à des concentrations de l'ordre du  $\mu\text{M}$ , donnant ainsi accès à bon nombre de classes de métabolites détectables à leurs concentrations physiologiques. De plus, elle permet d'identifier les métabolites par l'obtention de leur masse moléculaire, avec une précision dépendante du type d'analyseur, et par l'étude de leurs spectres de fragmentations qui fournissent des informations structurales et qui peuvent être comparés avec celui d'un composé de référence. Il est ainsi possible de réaliser des identifications *de novo* de composés qui ne sont pas décrits dans les bases de données et d'atteindre une plus grande certitude dans l'identification des composés déjà décrits. Actuellement, de nombreux métabolites mis en évidence lors d'expériences métabolomiques

n'ont pas encore été identifiés et cela constitue une limitation majeure à l'expansion de cette approche dans la mesure où l'identification de nouveaux composés et la caractérisation reproductible de ceux déjà décrits au sein de jeux de données complexes est un pré requis à l'interprétation biologique des signatures métaboliques.

Fondamentalement, un spectromètre de masse est composé des éléments suivants (Figure 3) : un système d'introduction de la substance à analyser, une source pour ioniser les analytes, un ou plusieurs analyseurs pour séparer les différents ions générés dans la source, un détecteur pour dénombrer les ions et enfin, un système de traitement des données permettant la visualisation et l'enregistrement des spectres de masse et le contrôle les paramètres d'acquisition.

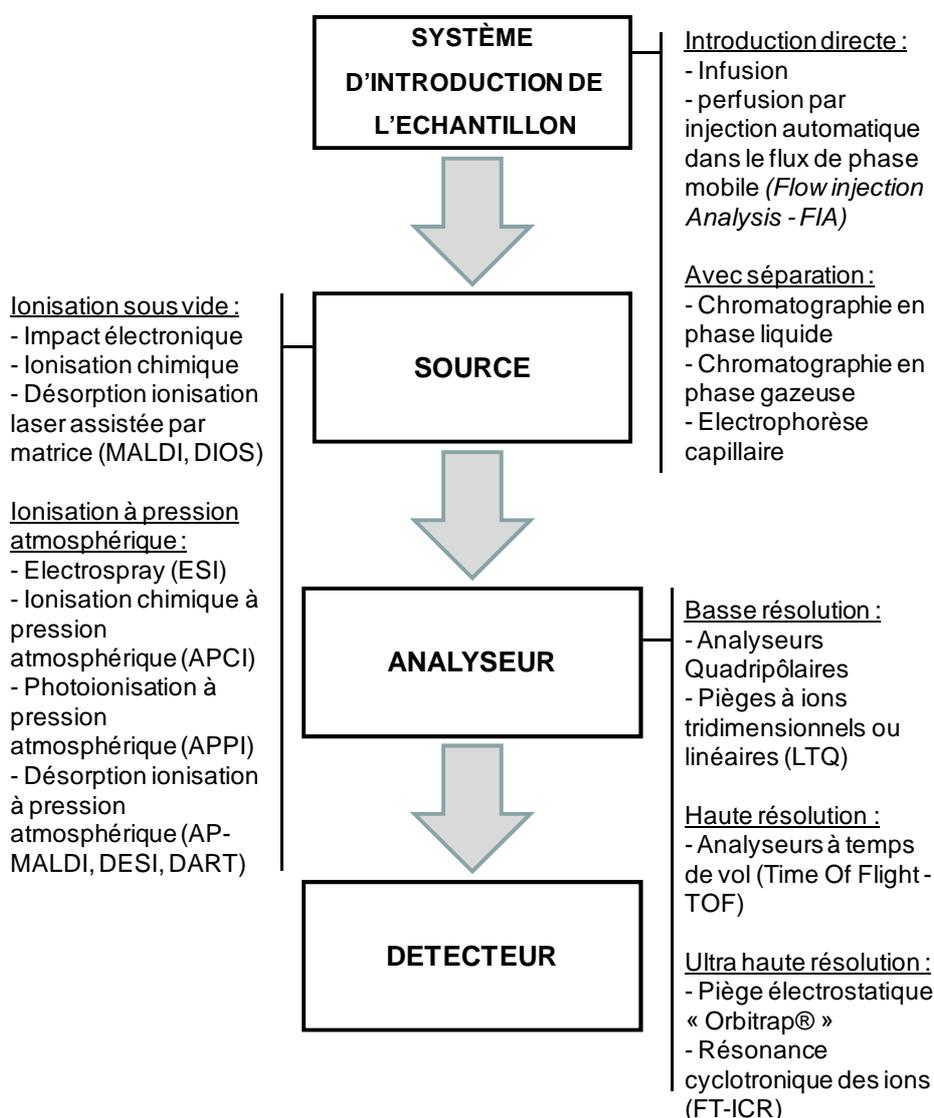


Figure 3 : Schéma global d'un spectromètre de masse présentant les sources et analyseurs disponibles pour l'analyse de biomolécules.

Les sources utilisées en spectrométrie de masse sont basées soit sur l'ionisation en phase gazeuse, soit sur la désorption-ionisation d'analytes. (Voir Annexe 1 - Spectrométrie de masse : rappels théoriques)

L'ionisation en phase gazeuse peut être réalisée sous vide (sources à impact électronique et à ionisation chimique utilisées en GC/MS) ou à pression atmosphérique comme par exemple l'APCI [91] (*Atmospheric Pressure Chemical Ionization*) ou l'APPI [92] (*Atmospheric Pressure Photoionization*). En APCI l'échantillon est ionisé à pression atmosphérique par application d'une décharge de corona. L'avantage de cette source est qu'elle peut être facilement couplée avec des techniques séparatives comme la LC. L'APCI est particulièrement bien adaptée aux molécules peu polaires ayant une masse moléculaire inférieure à 1500 Da. En APPI, l'ionisation résulte de l'interaction entre l'analyte et les photons émis par une lampe à décharge. Cette technique est adaptée à l'analyse de composés apolaires comme les stéroïdes [93], les pesticides [94;95] ou les hydrocarbures aromatiques [96].

La désorption-ionisation elle aussi peut être réalisée sous vide ou à pression atmosphérique. Les sources MALDI [97] (*Matrix Assisted Laser Desorption Ionization*) et DIOS [98] (*Desorption ionization on Porous Silicon*) opèrent à partir de supports solides, le DIOS permet d'accéder aux petites molécules, alors que le MALDI est réservé à l'étude des biomolécules de masse supérieure à 1000 Da. Les sources Electrospray [99;100] (*Electrospray Ionization*) et DESI (*Desorption Ionization Electrospray Ionization*) opèrent sur des nébulisats [101]. L'électrospray résulte de l'application d'un champ électrique de plusieurs milliers de volts sur un nébulisat, c'est l'une des sources les plus utilisées en LC/MS. Elle est par ailleurs adaptée au couplage avec l'électrophorèse capillaire [102]. Elle permet d'analyser sous forme quasiment intacte les biomolécules polaires à moyennement polaires qu'elles soient thermolabiles ou non.

Il existe de nombreux analyseurs répondant à des usages divers. Leurs performances sont évaluées au travers de leur résolution, leur exactitude en masse et de la gamme dynamique de masse dans laquelle ils répondent (voir Annexe 1 - Spectrométrie de masse : rappels théoriques).

- La résolution d'un analyseur est sa capacité à séparer des ions de rapport m/z proches mais différents. La résolution est initialement définie comme le rapport  $m/\Delta m$ , où  $\Delta m$  est la plus petite différence entre deux pics résolus. On lui préfère aujourd'hui la définition de FWHM (*Full Width at Half-Maximum peak height*) ou  $\Delta m$  correspond à la largeur du pic à mi-hauteur.
- L'exactitude en masse est la différence entre la masse exacte et la masse mesurée ; elle s'exprime en parties par million (ppm)

$$\text{Erreur de mesure de masse (ppm)} = \frac{M_{\text{exp}} - M_{\text{th}}}{M_{\text{th}}} \times 10^6$$

où  $M_{\text{exp}}$  représente la masse mesurée expérimentalement et  $M_{\text{th}}$  la masse monoisotopique théorique calculée de la molécule analysée.

- La gamme dynamique en masse correspond au domaine spectral (en  $m/z$ ) dans lequel l'instrument répond (*mass range*).

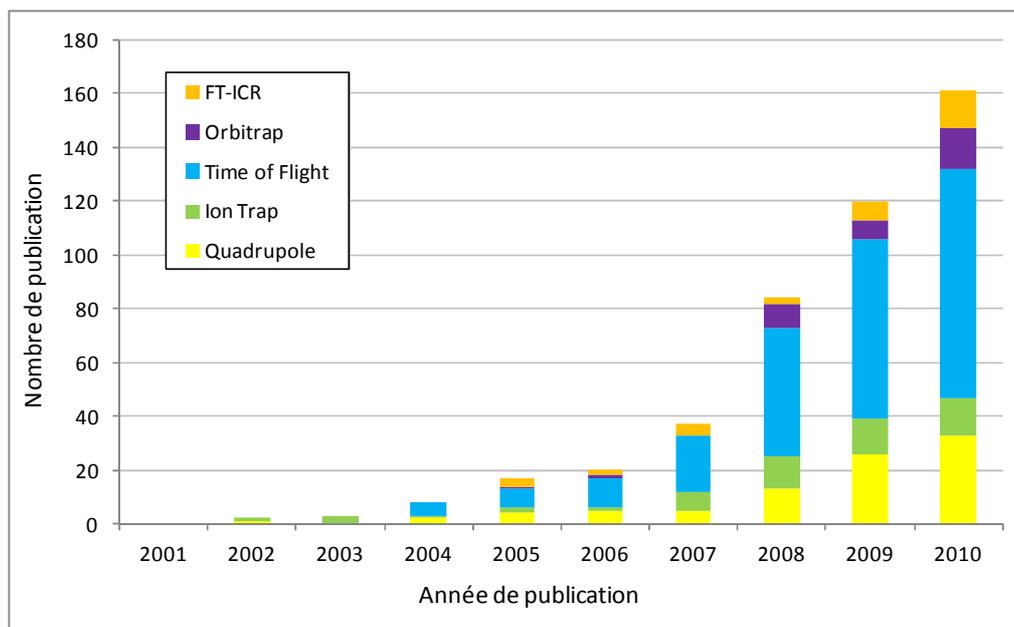
Le Tableau 1 présente un récapitulatif des performances des analyseurs disponibles sur le marché et utilisé en métabolomique.

Tableau 1 : Récapitulatif des performances des différents analyseurs [103]

<b>Analyseur</b>	<b>Résolution FWHM (<math>m/z</math> 400)</b>	<b>Exactitude en masse (ppm)</b>
Quadripôle	3000	200
Piège à ions (2D/3D)	5000	200
Temps de vol	8000-60000	2-10 <sup>a</sup>
Orbitrap®	Jusqu'à 100 000	<3
Résonance cyclotronique (FT-ICR)	Jusqu'à 10 000 000 <sup>b</sup>	<1

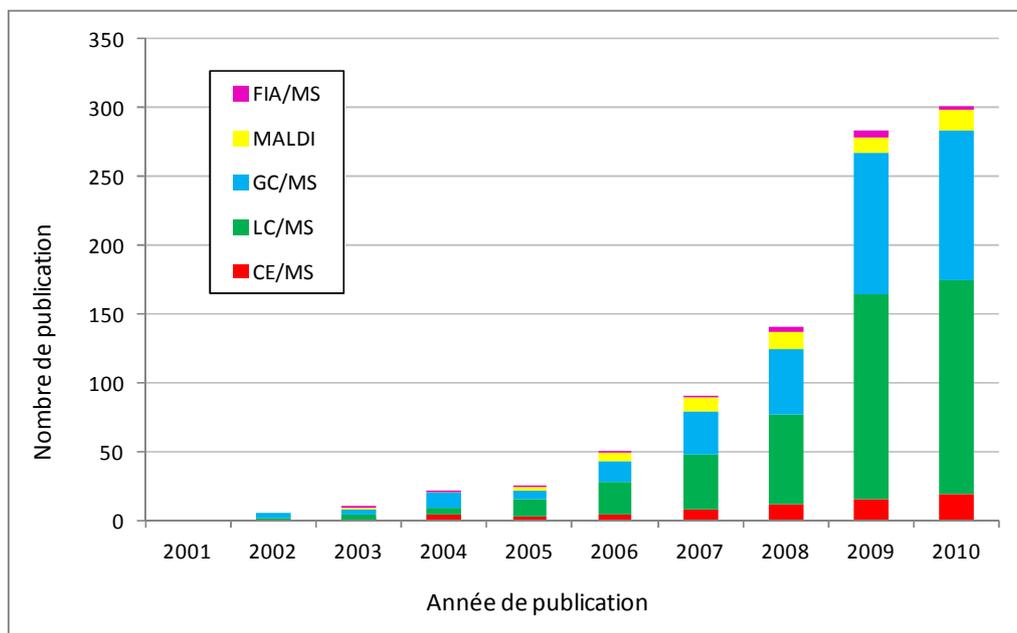
<sup>a</sup> En utilisant un calibrant interne ; <sup>b</sup> Selon la puissance du champ magnétique

Les premiers analyseurs à avoir été utilisés pour les analyses métabolomiques sont les analyseurs quadripolaires et les pièges à ions (Figure 4). Les quadripôles sont des analyseurs à faisceaux d'ions qui peuvent être montés en série (séparés par des cellules de collisions) comme dans le cas du triple quadripôle (TQ). Les TQ permettent différents modes d'acquisition en  $MS^2$  parmi lesquels les modes SRM/MRM (*Selected Reaction Monitoring / Multiple Reaction Monitoring*) qui sont particulièrement bien adaptés au dosage spécifique de molécules ou de familles chimique (approches ciblées). Les analyseurs à piégeage d'ions eux peuvent stocker les ions dans des pièges, ce qui permet de réaliser des fragmentations consécutives et séquentielles ( $MS^n$ ). La possibilité de réaliser ces expériences de  $MS^n$  avec les pièges ioniques fait des ces analyseurs des outils particulièrement adaptés à l'analyse structurale des métabolites. Le principal inconvénient de ces analyseurs et leur faible résolution, qui limite la séparation des ions de masse proches (isobares), et leur faible exactitude ce qui limite la détermination de formules brutes pour identification. Comme cela est montré sur la Figure 4, seuls ces deux type d'analyseurs étaient utilisés aux débuts de la métabolomique. Cependant, les analyseurs à temps de vol sont très rapidement devenus des analyseurs de choix en métabolomique de part leur meilleure exactitude en masse, et l'amélioration de leur résolution, notamment par couplage avec des quadripôles (Q-TOF). Ils sont d'ailleurs encore largement utilisés, mais le développement de la très haute résolution (> 50000, FWHM à  $m/z$  400), au début des années 2000, avec les analyseurs à transformée de Fourier (FT-ICR et Orbitrap®) a révolutionné la métabolomique, permettant de séparer facilement des composés isobares et d'accéder aux masses précises des ions pour déterminer leur formule brute. Les analyseurs de type Orbitrap® permettent d'accéder à la très haute résolution sans utiliser de champ magnétique comme pour le FT-ICR. Ces deux analyseurs sont la plupart du temps commercialisés couplés à des pièges ioniques afin de réaliser des expériences de fragmentation séquentielles de type  $MS^n$ . Ces analyseurs à haute résolution restent quand même moins répandus de part leur coût relativement élevé.



**Figure 4 : Types d'analyseurs utilisés pour les études métabolomiques par MS entre 2001 et 2010.**

Graphique représentant le nombre de publications sur ces 10 dernières années concernant l'analyse du métabolome par spectrométrie de masse. Pour chaque année les publications ont été classées par couleur en fonction du type d'analyseur utilisé. Les critères de recherche dans Pubmed étaient (metabolom\* OR metabonom\*), le type d'analyseur (Orbitrap, ICR, time of flight, ion trap et quadrupole) et l'année de publication (2001 à 2010).



**Figure 5 : Types de couplages utilisés pour les études métabolomiques par MS entre 2001 et 2010.**

Graphique représentant le nombre de publications sur ces 10 dernières années concernant l'analyse du métabolome par spectrométrie de masse. Pour chaque année les publications ont été classées par couleur en fonction du système d'introduction de l'échantillon utilisé. Les critères de recherche dans Pubmed étaient (metabolom\* OR metabonom\*), le système d'introduction utilisé (direct infusion, gas chromatography, liquid chromatography, capillary electrophoresis) et l'année de publication (2001 à 2010).

Il existe également plusieurs systèmes d'introduction de l'échantillon avant analyse dans le spectromètre de masse. Les plus utilisés sont les méthodes comme la GC/MS ou la LC/MS (Figure 5) car elles permettent d'apporter une dimension séparative supplémentaire à l'analyse. D'autres systèmes comme l'introduction directe par FIA (de l'anglais, *Flow Injection Analysis* pour injection directe dans le flux) ou le MALDI sont moins utilisés mais globalement ce sont la nature de l'échantillon, sa richesse potentielle en métabolites et le type de métabolites recherchés qui vont orienter le choix du système d'introduction.

#### 4.2.1. Méthodes de spectrométrie de masse par introduction directe

L'échantillon est introduit directement dans le spectromètre de masse sous forme gazeuse, liquide (infusion directe) ou solide (canne d'introduction directe, dépôt sur plaque MALDI, ...). L'utilisation de l'introduction directe permet des analyses rapides, à haut débit, d'échantillons bruts ou d'extraits. Malgré sa rapidité d'analyse et la garantie qu'elle apporte de ne pas perdre de métabolite durant les étapes de traitement de l'échantillon, l'utilisation de l'introduction directe présente des limitations. Lors des analyses en électrospray par exemple, l'effet matrice (suppression d'ionisation par exemple) doit en effet être pris en compte. Il est dû à la grande quantité d'ions analysés simultanément. Cet effet matrice est considéré comme équivalent pour les échantillons de même nature mais il empêche toute quantification absolue en l'absence d'étalons internes.

L'introduction directe a été utilisée [104;105] pour la classification et l'identification de microorganismes ou l'analyse d'organismes mutants dans le cadre d'études de génomique fonctionnelle. Elle est le plus souvent subordonnée à l'utilisation de spectromètres de masse à haute ou ultra-haute résolution. En effet, un grand nombre d'ions étant analysés simultanément, il est nécessaire de pouvoir accéder à des masses précises et à une bonne résolution afin de pouvoir discriminer et identifier les molécules isobares. Plusieurs études métabolomiques utilisant l'introduction directe ont été rapportées en couplage avec des FT-ICR et plus récemment avec l'Orbitrap® [106]. Il existe également un système d'infusion par nano-électrospray, ESI Chip® (Advion BioSciences Inc), qui permet d'améliorer la sensibilité en introduction directe [107;108] et de limiter le carry-over notamment pour l'analyse des lipides [109].

L'introduction d'échantillons solides est également possible en utilisant des méthodes de désorption-ionisation sur surface comme le MALDI, principalement utilisée pour les biomolécules de haut poids moléculaire comme les lipides [102;110;111] et les protéines. On peut également coupler ces méthodes d'introduction directe avec une cellule à mobilité ionique afin d'apporter une dimension séparative tout en conservant la rapidité d'analyse [102;112;113]. Dans cette cellule à mobilité ionique (*Ion Mobility-MS*), les molécules ionisées sont soumises à un champ électrique dans un courant de gaz. Les ions se déplacent alors suivant le champ électrique à une vitesse qui dépendra de leur interaction avec le gaz, de leur masse, leur taille et leur forme. On peut ainsi séparer des molécules de même masse ayant une forme différente.

## **4.2.2. Méthodes de spectrométrie de masse couplées à des méthodes séparatives**

### **4.2.2.1. La chromatographie en phase gazeuse couplée à la spectrométrie de masse**

Développée dans les années 1960, la chromatographie en phase gazeuse couplée à la spectrométrie de masse (à ionisation par impact électronique) offre une très bonne résolution chromatographique, une détection spécifique des analytes, et la possibilité de les quantifier, le tout avec une bonne sensibilité. Bien que l'usage de colonnes capillaires permette une bonne capacité de pic, dans le cas d'échantillons biologiques complexes, le pouvoir séparatif de la colonne est souvent insuffisant. Il a donc été développé des méthode de GC-2D (GCxGC), où chaque pic issu d'une première colonne, généralement apolaire, est ensuite transféré sur une seconde colonne polaire pour subir une seconde dimension de séparation [114].

L'énergie transmise aux analytes en ionisation électronique étant très importante, elle conduit à une fragmentation très reproductible, qui a permis de constituer des bibliothèques de spectres universelles. L'existence de ces bases de données regroupant les spectres de plus de 100 000 composés (celle du *National Institute of Standards and Technology* (NIST) par exemple), ainsi que l'obtention directe d'informations structurales facilitent l'identification des métabolites. Enfin, l'existence des indices de rétention (indices de Kovats [115]), spécifiques à la GC participe également à l'efficacité d'identification de cette technique. Ainsi, plus de 300 composés ont pu être caractérisés à partir d'une empreinte métabolomique d'extraits de feuilles d'*Arabidopsis thaliana* [116], grâce à une approche par GC/MS.

Cependant, l'analyse par GC/MS requiert que les analytes soient thermostables et volatiles. Ainsi, il est souvent nécessaire d'effectuer une étape de modification chimique (dérivation) pour améliorer la volatilité des composés polaires et peu volatiles. Ce traitement de l'échantillon est assez contraignant et peut conduire à des réactions de dégradation, en même temps qu'il peut nuire à l'aspect quantitatif, l'équilibre de la réaction pouvant être modifié dans certains échantillons. D'autre part, le couplage GC/MS repose souvent sur l'utilisation de sources à impact électronique qui permettent rarement d'obtenir la masse du composé intact mais seulement d'observer les ions issus de sa fragmentation. L'absence de l'espèce moléculaire rend donc délicate toute identification de composés inconnus ou non décrits dans les bases de données. Toutes ces raisons ont conduit au développement des techniques séparatives en phase liquide pour les approches métabolomiques.

### **4.2.2.2. La chromatographie en phase liquide couplée à la spectrométrie de masse**

Les couplages entre la chromatographie en phase liquide et la spectrométrie de masse sont apparus plus tardivement que ceux avec la chromatographie en phase gazeuse en raison des obstacles technologiques liés à leur réalisation, la principale limitation étant la nécessité d'évaporer la phase mobile chromatographique afin de faire passer les analytes en phase gazeuse avant leur

entrée dans le spectromètre de masse. Ces problèmes ont été résolus avec le développement de méthodes d'ionisation à pression atmosphérique telles que l'électrospray.

La chromatographie en phase liquide couplée à la spectrométrie de masse est aujourd'hui devenue une méthode de choix pour l'analyse du métabolome. En effet, elle permet de séparer des molécules isobares et souvent les isomères, de limiter la suppression d'ionisation et le bruit de fond, et contrairement à la GC/MS, elle présente l'avantage d'être applicable à l'analyse de composés non volatiles et/ou thermosensibles [117]. Par ailleurs, la LC/MS autorise, la plupart du temps, peu ou pas d'étape de traitement de l'échantillon. Enfin, contrairement aux études en GC/MS, les sources et les analyseurs utilisés en LC/MS permettent d'obtenir la masse moléculaire d'un métabolite et de réaliser ensuite des expériences complémentaires (MS/MS, MS<sup>n</sup>) afin de l'identifier. On distingue les systèmes chromatographique à haute (conventionnels) et à ultra-haute pression (UHPLC, de l'anglais «*ultra-high pressure liquid chromatography*»). Les systèmes basés sur la chromatographie capillaire et la nano-chromatographie sont également utilisés en analyse métabolomique.

- LC conventionnelle

Les méthodes les plus employées en métabolomique utilisent des colonnes narrow bore (diamètre interne de 2.1mm) à polarité de phases inversées, de type C<sub>8</sub> ou C<sub>18</sub>, avec un diamètre de particules de 3 à 5 µm. Ces phases sont peu adaptées aux métabolites très polaires qui ne sont pas retenus mais co-élués dans le temps mort. Quelques exemples d'analyses de tels composés par LC/MS avec l'utilisation de colonnes de type HILIC (*hydrophilic interaction liquid chromatography*), favorisant la rétention des composés hydrophiles mais de manière peu résolutive, sont décrits [118-121]. L'utilisation de la séparation par paire d'ion sur colonne à polarité de phases inversées a également été décrite pour la séparation des molécules polaires (avec l'utilisation de la tributylamine comme agent d'appariement, par exemple) et apolaires [122].

- LC capillaire et NanoLC

L'utilisation de colonnes capillaires et de nanocolonne a également été publiée. La diminution du diamètre intérieur (respectivement entre 200 et 320 µm et entre 10 et 100µm) et la petite taille des particules de remplissage de ces colonnes, entraînent une amélioration de l'efficacité de séparation chromatographique et une diminution du volume d'échantillon nécessaire. Les co-élutions des métabolites sont ainsi limités et les effets de suppression d'ionisation peuvent éventuellement être atténués. L'augmentation du rapport signal/bruit obtenue entraîne également une augmentation de la sensibilité. Les colonnes capillaires ont été utilisées lors d'études métabolomiques chez la plante *Arabidopsis thaliana* [123], et appliquées à l'étude d'urines de rats [80]. Dans cette dernière étude, l'utilisation d'une colonne capillaire entraîne l'augmentation du nombre de pics détectés ce qui permet d'améliorer la séparation entre les différents groupes d'échantillons étudiés, en d'autres termes, l'amélioration de la classification des échantillons en groupes. Les systèmes nanoLC restent encore peu utilisés en métabolomique [124;125], à cause probablement de l'étape pré-concentration,

nécessaire pour gagner en sensibilité quand on travaille avec de faibles volumes d'échantillons, mais qui exclut les métabolites polaires de l'analyse.

- UHPLC

Récemment, des analyses métabolomiques ont été effectuées à l'aide de systèmes UHPLC (*Ultra-High Pressure Liquid Chromatography*). Ces systèmes représentent une alternative aux colonnes capillaires avec l'utilisation de colonnes dont les diamètres de particules sont inférieurs à 2  $\mu\text{m}$  [126]. L'utilisation d'une granulométrie inférieure à 2,5  $\mu\text{m}$  (1.9  $\mu\text{m}$  chez Thermo et 1.7  $\mu\text{m}$  chez Waters) engendre un gain significatif en efficacité même pour des débits/vitesses linéaires plus élevés ce qui augmente la vitesse et la capacité de pics (nombre de pics résolus par unité de temps). Si la réduction de la granulométrie s'accompagne d'une amélioration des performances, elle se traduit également par une augmentation de la perte de charge, qui a nécessité des développements technologiques permettant de travailler à des pressions de travail de l'ordre de 10000 psi.

L'utilisation de l'UHPLC induit également une augmentation de la sensibilité [80], attribuable à l'augmentation de la vitesse linéaire qui conduit à une diminution de la largeur des pics, compensée par une augmentation de leur hauteur, ce qui améliore le rapport signal sur bruit. Ces systèmes permettent donc de réduire la durée des analyses en conservant ou en améliorant la résolution obtenue avec les colonnes « classiques » [80]. Les systèmes UHPLC présentent également une meilleure robustesse que les systèmes HPLC notamment au niveau des temps de rétention [80;127]. L'analyse de serum humain à l'aide d'un système UHPLC a permis l'augmentation de 20% du nombre de pics détectés [128] par rapport à l'HPLC.

A l'heure actuelle, il n'existe cependant que peu de phases disponibles commercialement (cinq à l'heure actuelle : C<sub>18</sub>, C<sub>8</sub>, Shield RP18, Phényle et HILIC). Enfin, un prétraitement de l'échantillon un peu plus important est parfois nécessaire pour éviter les problèmes de colmatage des colonnes. Des étapes de filtration et/ou de centrifugation sont notamment obligatoires pour les milieux les plus riches (plasma, extraits cellulaires).

#### **4.2.2.3. L'électrophorèse capillaire couplée à la spectrométrie de masse**

L'électrophorèse a été conçue pour séparer des espèces chimiques soumises à un champ électrique selon leur rapport taille sur charge et ce en fonction de leur vitesse de migration électrophorétique ( $\mu_{ep}$ ). L'électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) représente une alternative intéressante pour l'analyse des métabolites très polaires qui ne sont pas retenus par les systèmes chromatographiques à polarité de phase inversée [129;130]. L'électrophorèse capillaire est réalisée dans un capillaire en silice fondue, ce qui fait d'elle une méthode très résolutive, rapide et nécessitant une faible quantité d'échantillon. Contrairement à l'HPLC, l'électrophorèse capillaire n'implique pas de transfert de masse entre les phases et le profil d'écoulement du flux est plat (alors qu'il est parabolique pour un flux hydrodynamique) : il n'y a donc pas d'élargissement du pic comme en HPLC. L'efficacité de la séparation en électrophorèse capillaire

est généralement plus grande que celle d'autres méthodes de séparation telles que l'HPLC mais elle est moins robuste. De plus, de part le faible volume d'échantillon analysé elle est aussi moins sensible. C'est principalement l'électrophorèse capillaire de zone (CZE) qui est utilisée en CE/MS. En effet, les autres techniques nécessitent l'ajout d'additifs (tensioactifs par exemple) afin de modifier la sélectivité de la séparation, mais la plupart d'entre eux ne sont pas compatibles avec la spectrométrie de masse. Placés dans un tampon et sous l'effet d'un champ électrique constant, les métabolites migrent à différentes vitesses en fonction leur migration électrophorétique (dépend leur charge) et du flux électroosmotique (généralisé par la migration des analytes du tampon) (Figure 6).

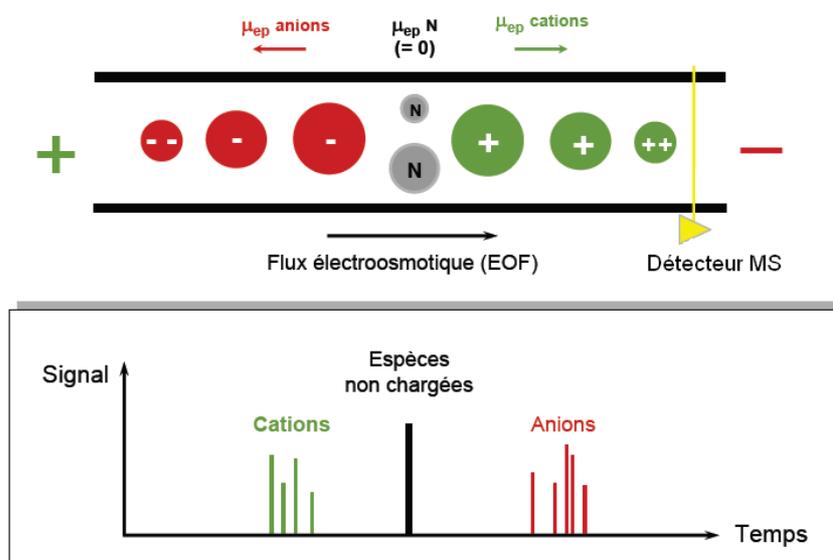


Figure 6 : Principe de l'électrophorèse capillaire de zone.

(image issue du site internet [http://nte-serveur.univ-lyon1.fr/afsep/fichier/225\\_18\\_1\\_2005.pdf](http://nte-serveur.univ-lyon1.fr/afsep/fichier/225_18_1_2005.pdf))

#### 4.2.2.4. La microfluidique

En raison de la diversité chimique des métabolites, de leurs concentrations généralement très faibles et de leur dynamique très rapide, leur quantification en milieu biologique est très complexe. La connaissance de ces conditions dynamiques représente un défi majeur et est le chaînon manquant pour une meilleure appréhension de la complexité des réseaux métaboliques. À cette fin, la microfluidique peut jouer un rôle important en raison de sa très bonne résolution aussi bien dans le domaine spatial que temporel et également de son faible temps d'analyse. De plus la faible quantité d'échantillon nécessaire et son traitement qui peut être intégré facilement à l'analyse en elle-même en font une technique d'avenir pour les analyses métaboliques. Le couplage de la microfluidique par CE ( $\mu$ CE) avec une source électrospray a été démontré en 2000 par Li. et al. [131]. Avec l'évolution rapide des techniques de miniaturisation, la microfluidique est maintenant facilement couplée à la spectrométrie de masse par le biais de source nano ESI. Elle est de plus en plus utilisée pour l'analyse des protéines et peptides, mais son utilisation pour l'analyse du métabolome est encore anecdotique [132-134].

### 4.3. Approches analytiques multidimensionnelles

Comme nous l'avons vu, le métabolome contient des composés appartenant à des classes chimiques très variées, et les analyser tous simultanément n'est pas envisageable avec une seule méthode d'analyse. En conséquence, des combinaisons de plusieurs techniques analytiques commencent à être envisagées afin de détecter le maximum de métabolites en une seule analyse. Ces couplages peuvent concerner la partie séparative, avec l'utilisation de différentes colonnes chromatographiques comme par exemple un couplage « colonne échangeuse d'ions/ C<sub>18</sub> », ou encore un couplage « C<sub>18</sub>/ HILIC » ou l'association de plusieurs méthodes séparatives comme l'électrophorèse capillaire, dédiée à l'analyse des composés polaires, et la LC à polarité de phase inversée pour les composés plus apolaires [135-137]. Plusieurs méthodes de détection peuvent également être couplées avec par exemple des couplages du type LC/RMN/MS et LC/RMN/MS/MS [138] qui se sont montrés très prometteurs mais qui engendrent une durée d'analyse conséquente. Ils permettent en effet une identification plus aisée des métabolites tout en conservant la sensibilité de la spectrométrie de masse, mais ils sont assez coûteux à cause de l'utilisation de phases mobiles deutérées. La complémentarité de ces deux techniques a déjà été mise en avant en tant qu'outil nécessaire à l'élucidation structurale de composés inconnus dans les recherches sur les produits naturels, les métabolites de médicaments et autres mélanges complexes [139;140]. Cependant, ces couplages nécessitent tout de même une RMN sensible (utilisation d'une cryosonde) et une chromatographie possédant une bonne résolution afin de pouvoir assigner de façon plus aisée les différents signaux. Une des difficultés majeures liées à l'emploi de plusieurs techniques reste la fusion des données [141].

## 5. TRAITEMENT DES DONNEES

Toutes les approches de génomique fonctionnelle reposent sur l'utilisation de l'informatique aussi bien au niveau de l'acquisition, du stockage, que de l'analyse d'une grande quantité de données. Si les techniques les plus matures comme la génomique, la transcriptomique et, dans une moindre mesure, la protéomique disposent aujourd'hui d'outils de traitement des données performants, de nombreuses limitations persistent dans le domaine de la métabolomique.

L'étape de traitement des données a pour principaux objectifs de sélectionner les signaux analytiquement pertinents présents dans les données brutes, d'aligner les empreintes des différents échantillons, d'éliminer le bruit de fond, et de présenter les données extraites sous un format matriciel, compatible avec leur injection dans les logiciels de traitement statistiques qui ne sont que rarement en mesure de lire les formats propriétaires utilisés par les appareils d'acquisition d'empreintes. Cette étape peut enfin comprendre une procédure de normalisation des données. Par la suite, ne seront développées que les méthodes employées dans le traitement des données issues de la RMN ou de la spectrométrie de masse.

## 5.1. Résonance magnétique nucléaire

Les spectres RMN d'échantillons biologiques (mesurés à 500 ou 600 MHz) sont complexes car ils contiennent plusieurs milliers de signaux correspondant aux déplacements chimiques des métabolites. La méthode de traitement des données la plus utilisée est fondée sur la division du signal en régions spectrales, également nommées fenêtres, qui correspondent à une bande spectrale dont la largeur typique est de 0,04 ppm [142], et sur laquelle l'ensemble du signal est moyenné. Cette méthode présente l'avantage de masquer les variations de position des pics et fournit une modélisation robuste. Elle est particulièrement adaptée aux approches chimiométriques. Cependant la perte de résolution qu'elle engendre peut être un désavantage car elle nécessite le réexamen de l'ensemble du spectre RMN afin d'identifier les métabolites responsables de la différenciation entre les groupes. De plus, l'utilisation de telles données peut conduire à des inexactitudes dans l'intensité mesurée des pics [143], à cause de l'intégration d'une quantité variable de bruit de fond suivant les échantillons.

Une approche alternative est l'analyse des données en résolution complète [86]. Elle implique la plupart du temps l'utilisation d'algorithmes automatiques d'alignement de pics afin de résoudre le problème des variations de déplacement chimique [144;145]. Ces variations peuvent affecter certaines molécules comme par exemple le citrate. Ces décalages entre les spectres peuvent être dus à une différence de pH entre les échantillons ou à d'autres effets environnementaux comme la concentration des ions métaux, les liaisons métabolites-protéines et les phénomènes d'échanges chimiques. L'analyse de l'ensemble du spectre est une avancée significative car elle permet l'interprétation directe des modèles chimiométriques. La résolution complète possède en effet une forte similarité en apparence avec les données brutes et cela aide à l'identification des métabolites.

## 5.2. Chromatographie en phase liquide couplée à la spectrométrie de masse

En LC/MS, le jeu de données initial est composé d'un ensemble de chromatogrammes (1 pour chaque échantillon). Pour chaque chromatogramme, représentant l'intensité du courant ionique total en fonction du temps de rétention, il existe en fait une troisième dimension qui correspond à la résolution en  $m/z$  des espèces ioniques détectées pour de chaque spectre (Figure 7).

Par rapport à la RMN, l'étape de traitement des données est rendue encore plus complexe par la présence d'une technique séparative qui amène une dimension supplémentaire, d'une importante quantité de bruit de fond, d'artefacts et de redondance d'information. Ainsi, la technique séparative introduit des décalages de pics temporels correspondant à des variations de temps de rétention. De plus, le bruit de fond peut être variable selon les différentes zones du chromatogramme et la forme des pics peut varier d'un pic à l'autre.

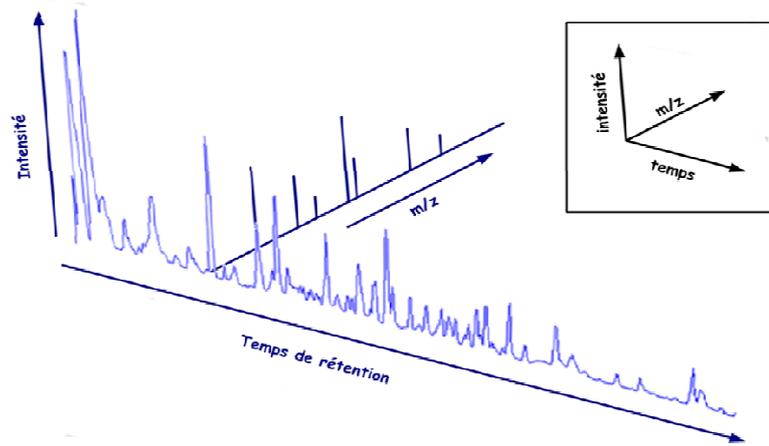


Figure 7 : Représentation schématique en trois dimensions d'un chromatogramme obtenu par LC/MS

Le traitement de ces données va donc consister à extraire les signaux analytiquement pertinents ( $m/z$ - $t_r$ -intensité) des données brutes pour chaque échantillon, puis à effectuer un réalignement dans les domaines temporel (ou chromatographique) et spectral et de les restituer sous un format qui soit adapté à leur analyse ultérieure. Ce processus de traitement des données a déjà été décrit dans certaines publications [146-152] et peut également comprendre une étape de normalisation des données (Figure 8). Il existe des logiciels spécifiques ayant pour but la mise en œuvre d'algorithmes permettant l'élimination du bruit de fond, la détection des pics et leur l'alignement. Nous allons donc nous intéresser à ces logiciels et à la normalisation des résultats mais avant nous allons voir comment les données brutes doivent être préparées préalablement à cette étape.

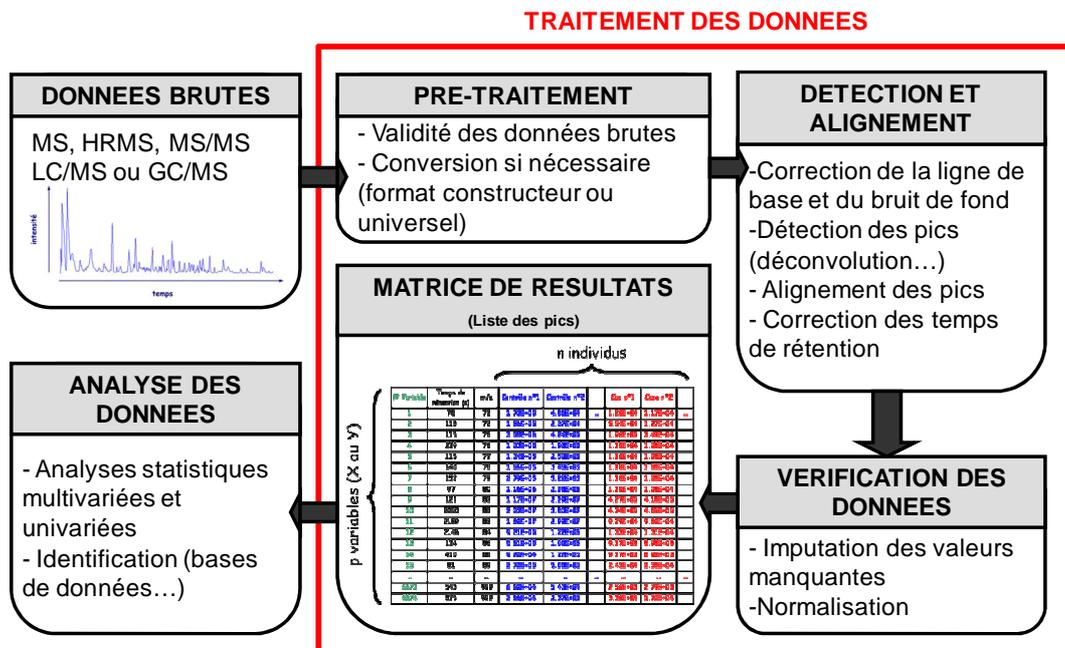


Figure 8 : Processus de traitement des données métaboliques

#### 4.2.1. Préparation des données

Avant de débiter cette étape, il est important de s'assurer de la qualité des données acquises. En effet, l'ionisation à pression atmosphérique étant sujette à de nombreuses variations et notamment au phénomène d'encrassement de la source ESI au cours du temps, il faut s'assurer que le signal enregistré ne varie pas significativement au cours de l'expérience car cela pourrait masquer les variations biologiques que l'on cherche à observer. La modification du signal MS peut par exemple être évaluée par le suivi des abondances des étalons internes au cours du temps.

Le type et le format des données brutes est également un point important à prendre en compte. En effet, si actuellement, suite aux nombreuses avancées technologiques, la plupart des logiciels sont capables de gérer des données haute résolution et/ou couplées à des méthodes séparatives, ce n'est pas le cas des données MS/MS. On distingue les logiciels commerciaux spécifiques d'un constructeur donné (MarkerLynx™ pour Waters, MarkerView™ pour AB SCIEX...) qui ne sont utilisables qu'avec des données acquises sur les appareils provenant du même fournisseur, des logiciels libres et gratuits (XCMS, MZmine). Les fournisseurs d'instruments utilisant des formats de données propriétaires, il est souvent nécessaire de convertir les données brutes en fichiers de formats universels afin de pouvoir utiliser ces logiciels gratuits. Deux types de formats universels ont été développés pour les données de spectrométrie de masse : le netCDF (Network Common Data Form) et le mzXML [153;154].

#### 4.2.2. Traitement des données [152]

Le traitement des données débute par une étape de soustraction du bruit de fond spectral. Les signaux sont ensuite extraits, alignés et restitués sous la forme d'une matrice contenant des variables, identifiées par un rapport  $m/z$  et un temps de rétention, et des numéros échantillons. Les valeurs contenues dans la matrice correspondent aux aires ou aux abondances de chaque variable pour chacun des échantillons. Différents logiciels de traitement de données sont actuellement disponibles. Certains d'entre eux sont des logiciels commerciaux, fournis ou non par les fabricants d'instruments, d'autres sont des logiciels à accès libre. Dans ce dernier cas, il est parfois possible d'accéder aux algorithmes de façon à les modifier ou à les améliorer. D'autre part, ces logiciels diffèrent aussi par les approches mises en jeu. Alors que la soustraction du bruit de fond repose souvent sur des algorithmes de filtrage classiquement utilisés en traitement du signal, de grandes différences sont observées au niveau des étapes d'extraction et d'alignement des signaux. A titre d'exemple, la détection de pics est effectuée dans deux dimensions (temps de rétention et rapports  $m/z$ ) de manière indépendante avec le logiciel Mzmine [155], alors que l'algorithme MatchedFilter de XCMS (XCMS pour « *various forms (X) of Chromatography Mass Spectrometry* ») extrait les ions à partir de fenêtres de rapports  $m/z$  [150].

XCMS n'est pas un logiciel à part entière mais un *package* du logiciel gratuit R (ou compilation d'outils qui ne se trouvent pas dans l'installation de base de R). Le langage R est un langage de programmation et un environnement mathématique utilisé pour le traitement de données et l'analyse statistique. L'avantage de R est qu'il dispose, dans sa version de base, de la plupart des

fonctionnalités utiles pour la statistique, mais ses possibilités deviennent énormes dès que l'on utilise des algorithmes chaînés (extensions ou *packages* comme XCMS ou CAMERA) écrits en R et mis librement à disposition par la communauté R. Cette communauté est très active, ce qui permet une évolution très rapide des fonctionnalités des algorithmes.

**Tableau 2 : Principaux logiciels utilisés pour le traitement des données métabolomiques.**

LOGICIEL	ACCES	DONNEES BRUTES	TRAITEMENT DES DONNEES	NORMALISATION	ANALYSE STATISTIQUE	IDENTIFICATION
XCMS (Scripps Research Institute)	Package gratuit du logiciel gratuit R "Open-access"	Formats universels mzXML, NetCDF (conversion préalable des données)	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Non	Analyses univariées (T-test et ANOVA)	Connexion à Metlin. Identification des adduits, fragments et isotopes (CAMERA R package).
Mzmine	Logiciel gratuit "Open-access" (Java)	Formats universels mzXML, NetCDF (conversion préalable des données) et Raw (format Thermo). plug-ins additionnels pour les autres formats constructeurs.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'utilisation de standards internes	Analyses multivariées (ACP, CDA, Sommon's plot)	Identification des adduits, fragments, isotopes et complexes. Connexion aux principales bases de données en ligne (KEGG, PubChem, HMDB...) et recherche possible dans une base de données personnalisée.
MarkerLynx™XS (Waters)	Logiciel constructeur payant (Apex Track™ algorithme breveté)	Données acquises sur des Spectromètres de masse Waters uniquement.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'intensité totale des pics ou par l'utilisation de standards internes.	Analyses multivariées (ACP). Intégration directe avec Umetrics Simca-P (logiciel payant) pour les statistiques multivariées type PLS.	Connexion aux principales bases de données ChemSpider et ChemFolder, et recherche possible dans une base de données personnalisée.
MetAlign (Plant Research International B.V)	Logiciel gratuit mais pas "Open-access"	Formats universels mzXML, NetCDF (conversion préalable des données)	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'utilisation de standards internes	Analyses univariées (T-test)	Non
Refiner MS (Genedata)	Logiciel payant (module de Genedata Expressionist®)	Données acquises sur des Spectromètres de masse ABI, Agilent, Bruker, Ciphergen, Thermo et Waters. Supporte aussi les formats universels mzXML et NetCDF.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Non	Module Genedata Analyst™ (T-test, ANOVA, PCA, PLS...)	Connexion à SEQUEST, Mascot, Mass Frontier et HMDB.
MarkerView™ (AB SCIEX)	Logiciel constructeur payant	Données acquises sur des Spectromètres de masse AB SCIEX uniquement.	Détection des pics par déconvolution, Alignement des pics.	Normalisation en fonction de la taille de l'échantillon ou du volume d'injection.	Analyses univariées (T-test) et multivariées (ACP)	Non
MathDamp (Wolfram Research, Inc.)	Package gratuit de Mathematica (Logiciel propriétaire payant)	Formats universels mzXML, NetCDF (conversion préalable des données)	Détection des pics directement à partir des données brutes, Alignement des pics.	Normalisation par l'intensité totale des pics ou par l'utilisation de standards internes.	Analyses univariées (T-test, Z-score et ANOVA)	Non
Sieve® (ThermoFisher Scientific)	Logiciel constructeur payant	Données acquises sur des Spectromètres de masse Thermo-Fisher uniquement.	Détection des pics directement à partir des données brutes, Alignement des pics.	Non	Analyses univariées (T-Test et ANOVA)	Connexion à ChemSpider et utilisation d'un outil de recherche SEQUEST.

Il existe également des logiciels qui réalisent une analyse comparative à partir des données brutes sans passer par une étape d'extraction des signaux. Citons également MathDamp [156] qui est

un package développé en langage Mathematica®, et qui est capable de travailler directement sur les données brutes, sans recours à une étape de détection et d'extraction des pics, ou encore Sieve®, un logiciel développé par Thermo-Fisher capable de travailler directement sur les données brutes des appareils du même constructeur, également sans étape de détection des pics (Tableau 2).

Cette liste n'est pas exhaustive. Elle cite les logiciels les plus utilisés à l'heure actuelle. De nouveaux logiciels apparaissent régulièrement et des améliorations sont continuellement apportées aux outils existants afin de répondre à la forte demande de logiciels de traitement de données qui soient performants et efficaces. Ces logiciels doivent également suivre les évolutions technologiques en matière de spectrométrie de masse. Citons par exemple le cas de la gestion des données acquises avec des spectromètres de masse à haute et très haute résolution couplées ou non à des méthodes séparatives qui a nécessité de développer des outils compatibles avec la gestion de jeux de données volumineux. Ils sont basés sur le même modèle formel de traitement [157] et de plus en plus ces logiciels intègrent des modules supplémentaires d'analyse des données comme les statistiques multivariées ou l'identification par recherche dans les bases de données (MZmine par exemple) (Tableau 2). Ainsi, des suites logicielles sont développées pour intégrer différents modules et réaliser une analyse métabolomique complète (de l'import des données brutes à la mise en évidence et l'identification des variables d'intérêt). C'est par exemple le cas de MetaboAnalyst [158], Bio-jETI [159;160] ou MeltDB [161], qui sont des plateformes regroupant plusieurs modules paramétrables, et qui utilisent comme module de traitement de données XCMS. A noter qu'il existe enfin beaucoup de logiciels propres aux différents laboratoires effectuant des analyses métabolomiques, qui développent leurs propres codes, principalement sous Matlab®.

Peu d'informations sont disponibles sur la validation des logiciels de détection de signal dans la littérature. Les raisons principales sont que la plupart des signaux présents dans les empreintes métaboliques restent non caractérisés et les résultats d'une procédure de détection automatique des métabolites peuvent être impactés à la fois par le type d'instrument et le type de matrice biologique. En conséquence, les utilisateurs doivent évaluer le logiciel avec leurs propres critères afin de sélectionner celui ou celle la plus adapté à leurs instruments et à leur matrice biologique. Tautenhahn et al. ont proposé une approche intéressante pour évaluer les logiciels d'extraction de signal. Elle est basée sur l'estimation de trois paramètres : le recouvrement, qui mesure la fraction des informations pertinentes extraites par l'algorithme, la précision, qui est le pourcentage des éléments pertinents par rapport aux faux positifs, et le temps d'analyse, qui est le temps nécessaire à l'algorithme pour obtenir une détection en fonction d'un ensemble de données [162]. En fait, de nombreux signaux artéfactuels sont présents dans les matrices de données résultant de l'extraction automatique des pics et de l'alignement du signal. Ils ont été évalués au nombre d'environ 400 pour 100 signaux pertinents [162]. Une façon d'aborder cette question consiste à effectuer des dilutions en série d'échantillons de contrôle qualité (CQ) et à sélectionner les ions dont les niveaux de concentration sont corrélés avec le facteur de dilution [67].

### 4.2.3. Normalisation [152]

L'objectif de l'étape de normalisation des données est de supprimer les biais expérimentaux concernant l'abondance des ions détectés entre les différents échantillons, tout en conservant les variations liées aux effets biologiques. La diversité chimique des métabolites et les variations interindividuelles entraînant par exemple des modifications des rendements d'extraction et d'ionisation, il est difficile de séparer les variations d'intérêt biologique des biais liés à l'instrumentation.

Les stratégies de normalisation des empreintes métaboliques peuvent être divisées en deux catégories :

- Celles fondées sur des modèles statistiques utilisés pour définir des facteurs de correction propres à chaque échantillon à partir du jeu de données complet [163], telles que la normalisation par l'écart-type [164], l'intensité moyenne globale [165] ou la méthode du maximum de vraisemblance [166].
- Celles reposant sur un ou de multiples composés de référence [167;168], standards internes ou externes, utilisés pour normaliser l'ensemble du chromatogramme (composé unique) ou certaines régions de rétention (normalisation de chaque zone en utilisant la référence qui y est éluée).

L'approche par modèles statistiques souffre de l'absence de composés dont la concentration est connue et qui puissent servir de référence pour les autres métabolites. En outre, l'utilisation d'un facteur correctif déterminé sur le signal total affecte la covariance du jeu de données et nécessite quelques précautions quant à l'interprétation des résultats des analyses multivariées, généralement menées par la suite [169]. L'utilisation d'une multitude de standards, adaptés à chaque signal en fonction du temps de rétention et d'un domaine de  $m/z$ , présente aussi des limitations dans la mesure où les temps de rétention et les rapports  $m/z$  des références peuvent ne pas être adaptés à certaines matrices ou certains métabolites. En effet, plusieurs classes chimiques de métabolites peuvent avoir des domaines d'élution chevauchants et il est alors déraisonnable de penser pouvoir appliquer le même facteur correctif à l'ensemble des espèces qui peuvent présenter des rendements d'ionisation très différents. De plus, l'introduction de standards peut influencer la réponse des composés présents dans l'échantillon. C'est ce qu'on appelle le phénomène de « cross-contribution ». En effet il existe des interférences entre standards et composés endogènes qui engendrent une perte d'information non négligeable [170]. Récemment, une approche fondée sur l'utilisation de standards internes multiples combinée avec la répétition des injections, baptisée NOMIS (de l'anglais « *Normalization using Optimal selection of Multiple Internal Standards* »), a été développée pour les analyses lipidomiques [171]. La présence d'injections répétées permet de déterminer les variations des métabolites qui sont corrélées à celles des étalons internes afin de pouvoir les éliminer, tandis que les variations de métabolites non corrélées à celles des étalons sont considérées comme spécifiques.

Il n'existe donc pas à l'heure actuelle de méthode permettant de s'assurer d'une normalisation fidèle des données. Les approches utilisant un marquage métabolique isotopique uniforme apportent une réponse, mais elles sont surtout applicables aux études de microbiologie et de biologie végétale [172;173]. Une étude récente a néanmoins utilisé un double marquage isotopique  $^{15}\text{N}$  et  $^{13}\text{C}$  pour faciliter l'assignation des formules brutes aux signaux détectés dans des extraits d'*Arabidopsis Thaliana* [174]. Signalons toutefois la possibilité de réaliser du marquage métabolique à l'azote 15 chez le rat [172], avec des applications dans le domaine de l'analyse protéomique, mais qui pourraient être aisément transposables à l'étude de métabolites azotés.

## 6. ANALYSE STATISTIQUE

De manière analogue aux études du transcriptome ou du protéome, l'analyse métabolomique repose sur l'analyse différentielle d'empreintes aboutissant à une expression semi-quantitative de résultats (facteurs d'induction ou de répression). La difficulté est donc de gérer et de comparer ces empreintes qui contiennent de nombreux signaux correspondant à plusieurs centaines, voir plusieurs milliers de métabolites. En pratique, il est difficile de comparer visuellement l'ensemble de ces données, d'autant plus quand le nombre d'échantillons est important. Les analyses statistiques multivariées sont utilisées afin d'étudier ou de décrire un ensemble de données [175]. Elles permettent en effet de synthétiser et de visualiser rapidement une grande quantité d'informations, ceci en projetant les données initiales dans un espace de dimensions réduites qui permet une visualisation aisée de l'ensemble des données.

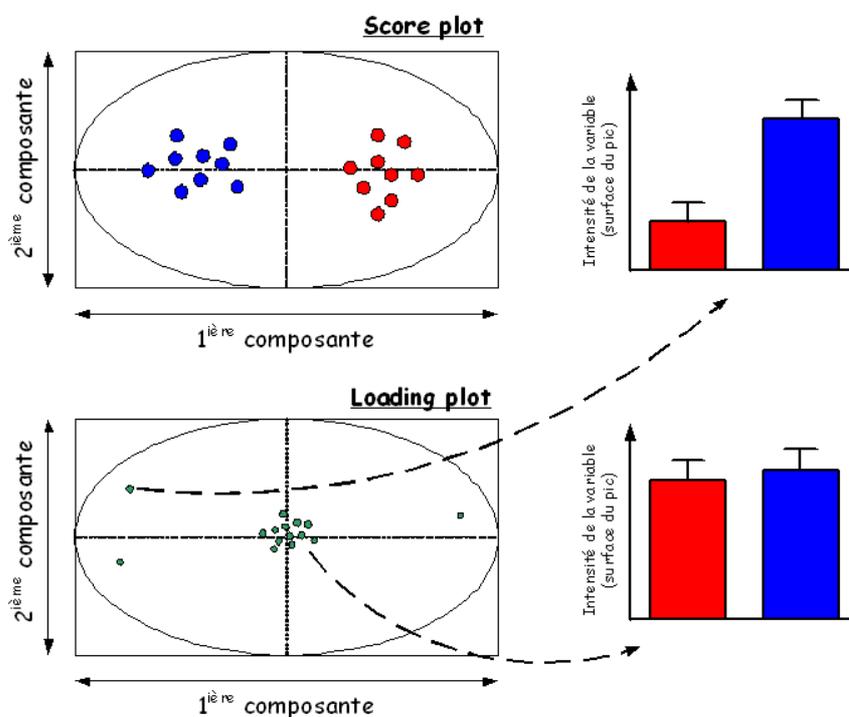


Figure 9 : Analyses statistiques multivariées [42].

Les statistiques multivariées permettent de représenter les données originelles (individus et variables) dans un espace de dimensions inférieures à l'espace d'origine, tout en limitant au maximum la perte d'information. La représentation des données dans des espaces de faibles dimensions facilite considérablement l'analyse. Pour cela, un changement de coordonnées est effectué afin de définir de nouveaux axes, appelés composantes principales, qui concentreront la plus grande partie de la variabilité des échantillons. Les composantes principales, extraites pour décrire la collection de données (variables expérimentales), sont classées en fonction de l'importance de la variabilité globale du jeu de données qu'elles expliquent. La première composante extrait la plus grande source de variance, la seconde composante, orthogonale à la première, extrait la plus grand source de variance de ce qui reste à expliquer, et ainsi de suite, si bien que les dernières composantes contiennent surtout le bruit aléatoire. En pratique, les deux ou trois premières composantes permettent souvent d'expliquer plus de 90 % de la variance totale. Les données sont présentées sous forme de cartes factorielles des individus ou « *score plots* », c'est-à-dire les coordonnées des échantillons dans le nouveau repère. Les échantillons appartenant à un même groupe sont proches les uns des autres. Les données peuvent également être étudiées sous forme de cartes factorielles du poids des variables ou « *loading plot* » qui représente la projection des variables ( $m/z$ ,  $tr$ ) sur les composantes. Les variables responsables de la séparation des deux groupes (Figure 9).

**n individus**

}-----}

# Variables	Temps de rétention (s)	m/z	n individus		Cas n°1		Cas n°2	
			Contrôle n°1	Contrôle n°2				
1	78	72	1.73E+05	4.86E+04	--	1.20E+04	1.17E+04	--
2	118	72	1.86E+05	2.37E+04		5.54E+04	1.27E+04	
3	119	76	3.39E+06	4.24E+03		1.96E+05	3.49E+04	
4	239	76	1.03E+05	1.93E+05		1.10E+04	1.08E+04	
5	115	77	1.34E+05	2.50E+05		1.10E+04	1.08E+04	
6	140	78	1.86E+05	3.48E+05		1.10E+04	1.08E+04	
7	182	79	2.79E+05	5.22E+03		1.10E+04	1.08E+04	
8	97	80	1.16E+06	2.18E+05		1.10E+04	1.08E+04	
9	121	83	1.17E+07	2.19E+07		4.27E+05	4.18E+05	
10	2080	83	2.03E+07	3.80E+07		4.74E+05	4.60E+05	
11	2189	83	1.60E+07	2.99E+07		9.79E+04	9.60E+04	
12	2146	84	9.21E+05	1.72E+05		1.93E+04	1.31E+04	
13	134	86	5.61E+05	1.05E+05		9.17E+03	8.98E+03	
14	410	86	9.29E+04	1.72E+05		9.17E+03	8.98E+03	
15	81	89	2.72E+05	5.08E+05		2.43E+04	2.38E+04	
--	--	--	--	--	--	--	--	--
6673	543	899	5.60E+04	5.43E+04		2.56E+03	2.79E+03	
6674	875	899	2.66E+04	2.37E+05		3.36E+04	3.70E+04	

}-----}

**p variables (X ou Y)**

Figure 10 : Exemple de matrice de données métabolomiques LC/MS.

Dans le cadre d'approches métabolomiques, les données initiales sont organisées sous forme de matrices issues de la détection automatique des ions. Chaque individu est décrit par un certain nombre de variables X, et parfois par des informations Y. Ainsi, les lignes des matrices sont constituées des variables X et parfois Y. Les colonnes qui constituent les matrices correspondent aux individus (échantillons) (Figure 10). Les données X, appelées variables correspondent par exemple aux rapports  $m/z$  pour les données acquises en spectrométrie de masse, ou des déplacements chimiques pour les données acquises en RMN. Dans le cas d'analyses explicatives, les variables Y,

appelées observations peuvent être la classe de l'échantillon, c'est-à-dire la dose ou le jour de traitement.

Ces données importées après la détection automatique des ions doivent également être normalisées. En effet les différences potentielles d'unités de mesures et d'échelle peuvent biaiser les résultats des analyses statistiques multivariées. Par exemple sans normalisation préalable, une variable de très forte intensité avec une faible variation entre les groupes aura une plus grande importance qu'une variable de faible intensité avec de fortes variations. Il convient donc de donner un poids équivalent à toutes les variables car elles peuvent montrer des différences d'abondance significatives. L'une des méthodes les plus utilisées est de centrer (soustraire la moyenne) et réduire (diviser par l'écart-type) les variables. Dans certains cas on préfère utiliser une normalisation de type Pareto (soustraire par la moyenne et diviser par le carré de l'écart type) afin de limiter la dilution du signal analytique (c'est-à-dire que l'on ne pas donne pas trop d'importance aux très faibles signaux du bruit de fond par exemple).

Une fois normalisées, les données peuvent être analysées. Il existe deux types d'analyses statistiques multivariées utilisées dans la littérature dans le cadre des études portant sur le métabolome:

- Les analyses statistiques **descriptives**, telle que l'analyse en composante principale (ACP), qui ne nécessitent pas d'information « *a priori* » sur la nature des échantillons. Elles permettent de décrire des données et de visualiser la répartition spontanée des échantillons.
- Les analyses statistiques **explicatives** qui visent à expliquer une ou plusieurs réponses (Y). Dans ce cas, des informations sont données sur la nature des échantillons. Parmi les méthodes explicatives, on trouve la régression « classique », dite des moindres carrés (ou « OLS » (*Ordinary Least Squares*)), ainsi que la régression PLS. Cette dernière présente deux avantages par rapport à la régression dite « classique ». D'une part, elle est insensible aux multicollinéarités, c'est-à-dire à la présence de variables très corrélées. Ceci est particulièrement intéressant en RMN où plusieurs déplacements chimiques sont associés à un même métabolite et sont donc fortement corrélés, ainsi qu'en spectrométrie de masse, où les rapports  $m/z$  correspondant aux fragments générés dans la source (in source CID) et aux ions adduits possèdent une forte corrélation entre eux et avec l'ion précurseur. D'autre part, elle est plus robuste vis-à-vis des données manquantes. Or les données extraites des chromatogrammes comportent la plupart du temps un certain nombre de données manquantes car tous les rapports  $m/z$  ou tous les déplacements chimiques ne sont pas nécessairement représentés dans chaque échantillon.

A ce stade il convient de distinguer les études exploratoires, qui tendent à mettre en évidence de nouveaux biomarqueurs potentiels dont la pertinence biologique doit être prouvée, et les études prédictives ayant pour but de catégoriser des individus inconnus dans un groupe définis (malade, sain...) et pour lesquelles une étape de validation est obligatoire. En effet ces mêmes méthodes peuvent également être utilisées dans un but prédictif. Un jeu de données initial est alors utilisé pour

construire un modèle explicatif, ce modèle est ensuite appliqué à de nouvelles données afin de classer les nouveaux échantillons dans les différents groupes.

### **6.1. Analyse statistique descriptive : Analyse en composantes principales (ACP)**

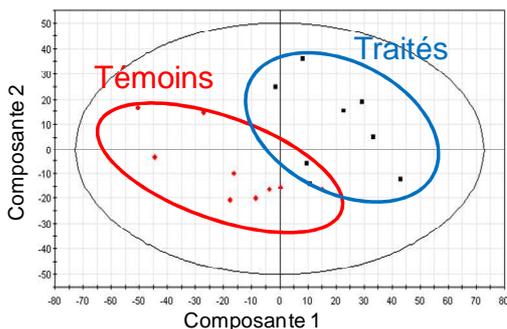
L'ACP [176,177] est la plus simple et la plus connue des techniques d'analyse de données multivariées. C'est une méthode de réduction des variables. Elle a donc pour objectif de décrire un ensemble de données ( $n$  individus,  $p$  variables) par de nouvelles variables en nombre restreint. C'est une méthode factorielle car la réduction des variables ne se fait pas par une simple sélection de certaines d'entre elles mais par la construction de nouvelles variables synthétiques obtenues à partir des variables initiales par combinaison linéaire. Cette méthode de classification est dite non supervisée car elle ne tient pas compte de l'appartenance des échantillons à des groupes préexistants (prise en compte uniquement des variables  $X$  et pas des  $Y$ ).

L'ACP est souvent la première méthode employée pour l'analyse des données afin d'étudier leur organisation et observer si les variations liées à l'appartenance à un groupe font partie des causes majeures de variance ou si cette dernière est plutôt reliée à un effet expérimental. Elle est suivie d'une analyse dite explicative permettant l'interprétation des données plus en profondeur afin de renforcer la séparation des groupes et de déterminer quelles sont les variables initiales du jeu de données qui contribuent à l'obtention de cette séparation.

### **6.2. Analyse statistique explicative : Régression PLS**

La régression PLS (Partial Least Square ou Projection to Latent Structure) fournit un modèle mathématique qui permet d'expliquer une ou plusieurs variables  $Y$  en fonction d'une matrice de variables expérimentales  $X$ . Le fonctionnement des techniques PLS ne sera pas décrit ici en détails. Ce type de modélisation n'utilise pas directement les variables explicatives, mais calcule d'abord les axes PLS sur lesquels sera ensuite réalisée la régression. Ces axes sont des combinaisons linéaires des variables initiales calculées de manière à décrire les plus grandes variations présentes dans le jeu de données et à maximiser la covariance entre les entrées ( $X$ ) et la sortie (le ou les  $Y$ ) du modèle. Ces axes sont tous orthogonaux entre eux. En régression PLS, toutes les variables importantes sont conservées et les variables sans importance sont soit exclues, soit participent au modèle, mais avec un faible poids. Les observations  $Y$  sont ainsi décrites en fonction des variables  $X$ . Pour ce qui est de la visualisation des données, les données sont présentées, tout comme pour l'ACP, sous forme de « score plot » ou carte factorielle du poids des individus, c'est-à-dire la projection des échantillons dans le nouveau repère généré par la PLS, ou sous forme de « *loading plot* » ou carte factorielle du poids des variables qui représente la projection des variables initiales sur les nouvelles composantes (Figure 11).

A. Score plot : carte factorielle du poids des individus



B. Loading plot : carte factorielle du poids des variables

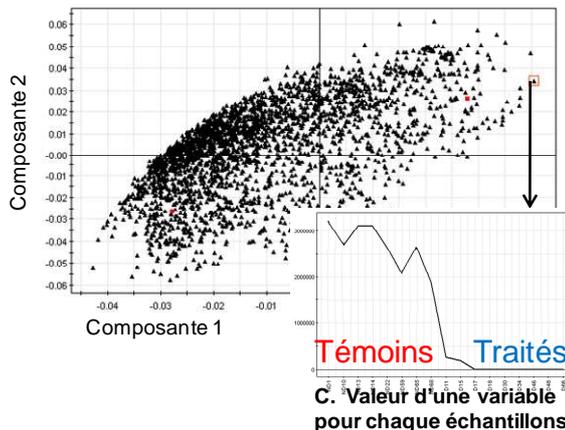


Figure 11 : Présentation des données lors des analyses multivariées. (A) carte factorielle du poids des individus et (B) carte factorielle du poids des variables.

(A) La carte factorielle du poids des individus («*score plot*» en anglais) permet de visualiser la répartition des échantillons dans un espace à 2 dimensions. Les axes du graphique, encore appelés composantes principales ou variables latentes, correspondent à des combinaisons linéaires des variables d'origine (signaux issus de la LC/MS). Ces combinaisons sont calculées de façon à visualiser la séparation des deux groupes sur les premières composantes. (B) La carte factorielle du poids des variables («*loading plot*», en anglais) met en évidence les principales variables responsables du clustering précédemment observé. (C) on peut visualiser la valeur de chacune des variable pour chaque échantillons.

### 6.2.1. L'analyse discriminante PLS

L'analyse discriminante PLS (PLSDA, de l'anglais «*partial least square-discriminant analysis*») est une variante de la régression PLS qui permet de construire un modèle qui maximise la séparation entre les classes auxquelles appartiennent les échantillons. Elle possède les avantages de la PLS : elle peut en effet gérer un nombre important de variables X, gérer les multicollinéarités et les données manquantes [178]. La différence entre les deux méthodes repose sur la nature des variables Y, pour une PLS classique les Y sont des variables quantitatives, pour le PLS-DA les variables Y sont qualitatives ou catégorielles.

La PLS, aussi bien que la PLS-DA permettent la construction d'un modèle explicatif. Ce modèle permet ainsi d'isoler les métabolites dont l'intensité est caractéristique d'un état biologique donné et qui contribuent à la séparation des différents groupes que forment les échantillons. Ce modèle construit à partir d'un jeu de données initiales peut également être utilisé pour expliquer un nouveau jeu de données. On cherche ainsi à prédire la classe d'appartenance des individus du nouveau jeu de données.

### 6.2.2. OPLS (Orthogonal Partial Least Square) [179]

L'OPLS est une récente modification de la PLS. Elle consiste à diviser la variation systématique de X en deux parties, dont la première est linéairement liée à Y et la seconde est indépendante (orthogonale) à Y. Ce fractionnement des données de la matrice des X facilite l'interprétation et

l'application du modèle à de nouveaux échantillons. Le modèle OPLS est composé de 2 sources de variation modélisées, l'une prédictive de Y et l'autre indépendante de Y, chacune correspondant à une composante. Seule la variation reliée à Y est utilisée pour modéliser Y. L'OPLS peut, comme la PLS-DA, être utilisée à des fins de discrimination (OPLS-DA) [180]. La Figure 12 illustre les avantages de l'OPLS vis-à-vis de la PLS-DA. Les variations interclasse et intra-classe sont séparées sur les deux composantes en OPLS-DA alors qu'elles ne l'étaient pas en PLS (Figure 11), ce qui facilite beaucoup l'interprétation du modèle OPLS-DA. Cloarec et al. [181] ont utilisé une combinaison de la covariance des variables (*loadings*) avec leur corrélation vis à vis de la séparation des classes afin de visualiser et résumer sur un seul graphique l'intervalle de variation de la variable et sa robustesse. Pour cela ils ont représenté la covariance de chaque variable et ont ajouté un codage couleur dépendant de la valeur de sa corrélation avec la séparation interclasse.

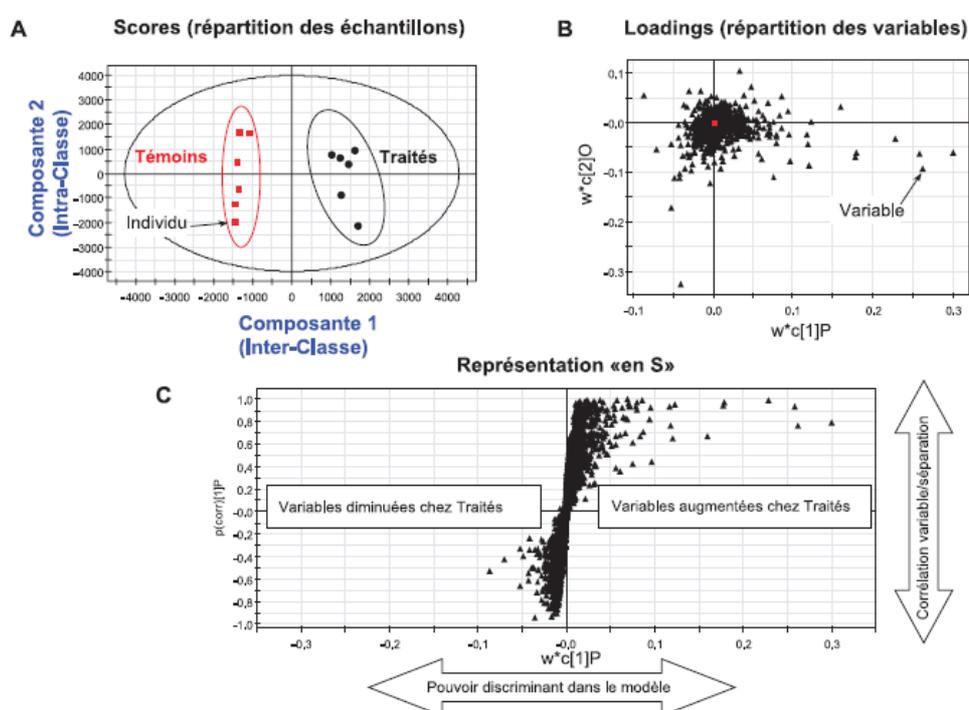


Figure 12 : Présentation des données en OPLS

Voir Figure 11 pour (A) et (B). (C) La représentation en S (S-plot en anglais) met sélectivement en évidence les variables discriminantes de la matrice X liées au facteur à étudier Y. En l'abscisse, on retrouve la covariance de la variable initiale, plus elle est éloignée du centre plus son intensité subit une forte variation entre les deux groupes. En ordonnée, on observe la corrélation entre l'intensité de la variable et le groupe de l'échantillon, ainsi, plus on s'éloigne du centre, plus l'intensité de la variable est homogène au sein de chaque groupe.

Cependant les variables d'intérêt mises en évidence par les analyses statistiques multivariées ne sont identifiées que par la valeur de leur déplacement chimique en RMN ou par le couple rapport masse/charge-temps de rétention dans le cas de la LC/MS. Il faut donc faire le lien entre ces valeurs et des métabolites afin de pouvoir les replacer dans le contexte biologique et évaluer leur pertinence. L'identification des composés d'intérêt reste un défi à relever pour la création de banques de données nécessaires à l'expansion des approches métabolomiques.

## 7. IDENTIFICATION DES VARIABLES D'INTERET

Une fois les variables d'intérêt isolées, il faut les identifier afin de leur donner une signification biologique. L'identification des signaux discriminants est probablement la tâche la plus complexe et la plus coûteuse en temps des analyses métabolomiques. Le grand nombre de métabolites et leur grande diversité chimique implique le plus souvent l'utilisation conjointe de plusieurs techniques analytiques, telles que la RMN, la MS, l'UV et l'IR [182;183]. En effet il est difficile voire impossible d'identifier un composé inconnu avec une seule technique et une identification formelle passe toujours par la comparaison des paramètres physico-chimiques du composé à identifier avec ceux d'un composé chimique de référence.

La spectrométrie de masse à pression atmosphérique est rapidement apparue comme un outil performant pour la métabolomique. Couplée à la chromatographie liquide, elle offre une bonne sensibilité, une importante gamme dynamique, une bonne adaptabilité aux différentes matrices biologiques et aux composés étudiés, tout en permettant, grâce aux techniques d'ionisation douce, d'avoir accès à la masse moléculaire des molécules intactes au sein de milieux complexes. Avec l'introduction des spectromètres de masse à haute et très haute résolution, l'obtention de mesures de masses précises, donnant accès à la formule brute des signaux, a considérablement accéléré l'identification des composés. L'utilisation conjointe de pièges ioniques qui sont utilisés pour la réalisation d'expériences de fragmentations séquentielles permet d'obtenir des informations structurales complémentaires, nécessaires à l'identification des métabolites d'intérêts [121;184]. Malheureusement, la spectrométrie de masse à pression atmosphérique présente une forte variabilité dans les profils de fragmentation générés sur différents appareils [185], ce qui a limité la construction de banques de données spectrales universelles à l'instar de celles obtenues en ionisation électronique ou par RMN [186]. En approche LC/MS, l'identification des signaux débute classiquement par une analyse du spectre de masse, afin de s'assurer que le signal correspond bien à une espèce moléculaire et non à un pic du massif isotopique, un ion adduit ou un ion fragment produit lors du processus l'ionisation.

### 7.1. Gestion de la redondance du signal en ESI

L'ionisation à pression atmosphérique est une méthode douce qui peut générer plusieurs types d'ions pour une même molécule (Figure 13). Il existe des outils informatiques qui permettent de détecter les pics issus d'un même métabolite. Ils regroupent les signaux en fonction des différences de masse spécifiques correspondant aux isotopes, adduits, fragments et en fonction des corrélations entre les intensités de paires d'ions, soit entre plusieurs spectres dans un même échantillon (corrélation dans le domaine spectral), soit entre tous les échantillons où les signaux sont observés [187-189]. La corrélation dans le domaine spectral repose sur le fait que des ions produits au moment de l'ionisation à partir d'une même molécule auront le même profil d'élution chromatographique. De la même manière leurs variations d'intensité entre différents échantillons seront toujours proportionnelles. Ainsi en comparant la forme des pics des différents ions à un même temps de

rétenion et les abondances dans les différents échantillons, on peut grouper les signaux entre eux et distinguer parmi eux l'espèce moléculaire protonée ou déprotonée (aussi appelé ion pseudomoléculaire).

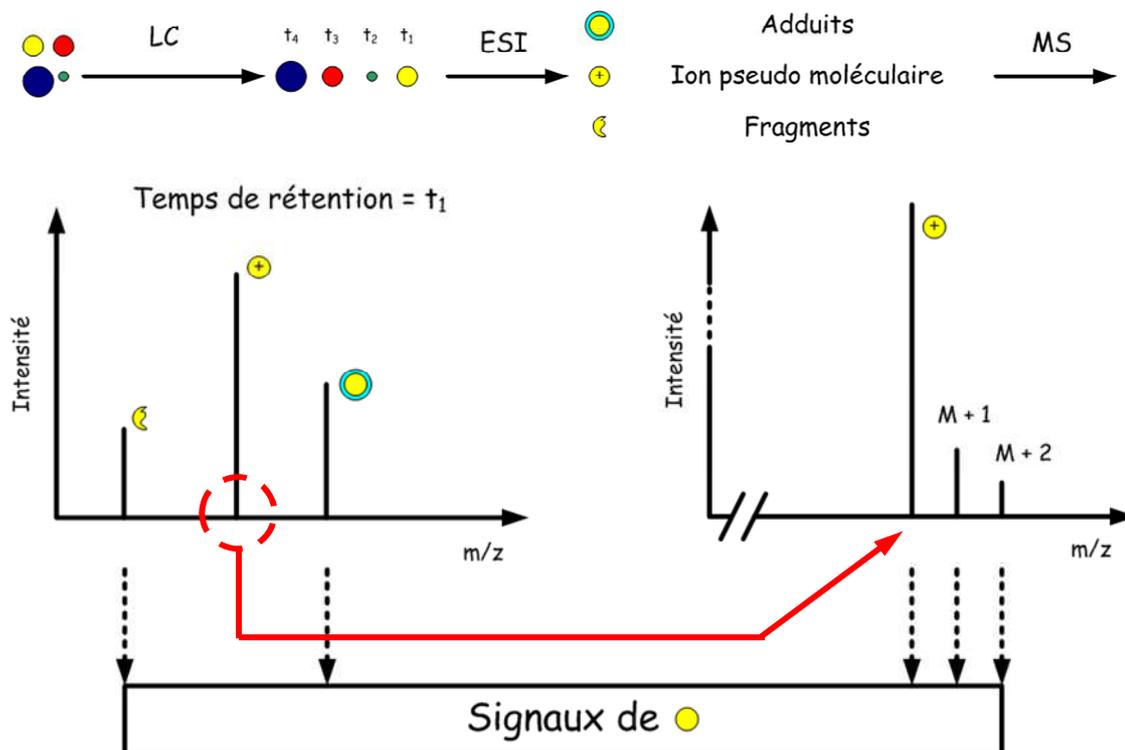


Figure 13 : Principe de la LC/ESI-MS et origine de la redondance du signal [42].

Quatre molécules représentées par des boules jaune, orange, bleue et verte sur la figure sont séparées par LC. Au temps de rétention  $t=1$ , la molécule jaune entre dans la source électrospray. Lors du processus d'ionisation plusieurs ions sont formés (adduits, fragments et espèce moléculaire). On retrouve ces différents ions sur le spectre de masse correspondant au temps de rétention  $t=1$ . De plus chaque ion possède un massif isotopique propre. Il apparaît clairement qu'il existe beaucoup de signaux différents pour une seule et même molécule. C'est de qu'on appelle la redondance du signal.

## 7.2. L'identification en spectrométrie de masse

Une ou plusieurs formules élémentaires peuvent alors être générée(s) si l'on travaille en haute résolution, ce qui fournit un premier élément pour effectuer une interrogation des bases de données existantes. L'analyse du massif isotopique d'un ion permet de réduire le nombre de formules possibles : l'analyse qualitative permet de déceler la présence ou non d'un atome par observation de son isotope (par exemple  $^{34}\text{S}$ ) et l'analyse quantitative (l'abondance relative d'un isotope) permet de remonter au nombre d'atome présent connaissant l'abondance naturelle de cet isotope. Cependant la précision de cette abondance relative varie en fonction de l'ionisation et de l'abondance de l'ion observé [190]. L'acquisition de spectres de fragmentation permet à ce stade de discriminer les réponses obtenues précédemment sur la base d'ions produits ou de pertes de neutre, caractéristiques de groupements chimiques. Certains composés peuvent encore être rejetés sur la base de leur temps de rétention et d'expériences complémentaires comme les échanges proton/deutérium (Figure 14).

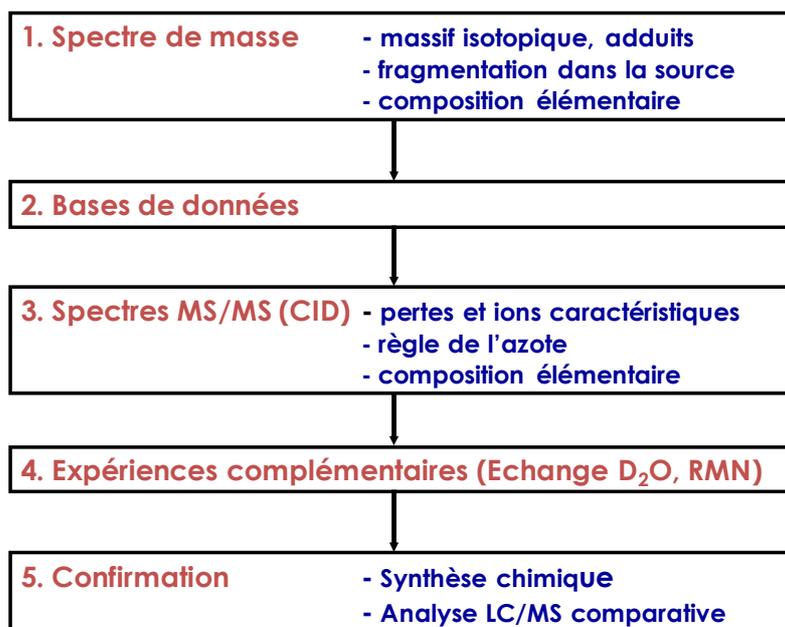


Figure 14 : Processus d'identification des molécules inconnues en métabolomique.

Au final, l'identification formelle est obtenue par injection de composés chimiques de référence (standards) sur le même système analytique. Le composé inconnu doit alors présenter le même temps de rétention, la même masse exacte et le même profil de fragmentation que le standard. En effet, dans le cadre de la « *Metabolomics Standards Initiative* » [191;192], Sumner et al. ont défini quatre niveaux d'identification en fonction des informations disponibles sur le métabolite à identifier :

- Métabolite identifié : un minimum de deux paramètres physico-chimiques indépendants identiques à ceux du standard dans les mêmes conditions analytiques. En LC/MS cela peut correspondre à la masse précise, au temps de rétention ou encore le spectre de MS<sup>n</sup>.
- Métabolite putativement annoté : en cas d'indisponibilité du standard, l'identification peut être basée sur les propriétés physico-chimiques et/ou les similarités spectrales (ex : MS/MS) avec les informations des bases de données publiques ou privées.
- Métabolite putativement caractérisés : l'identification peut être basée sur les propriétés physicochimiques d'une classe de composés et/ou les similarités spectrales (ex : MS/MS) avec des composés de la même classe chimique.
- Métabolite inconnu : bien que non identifiés, ces métabolites peuvent être différenciés à partir de données spectrales qui peuvent permettre une quantification relative.

L'analyse en spectrométrie de masse est suffisante pour identifier des métabolites quand ceux-ci sont bien décrits dans les bases de données commercialement disponibles et distinguables des autres isomères existants par une séparation chromatographique suffisante et des expériences de MS<sup>n</sup> adéquates. Dans ce cas, une simple comparaison du temps de rétention et du spectre de CID avec le standard permet de confirmer l'identification. Malheureusement, dans un grand nombre de cas, les composés d'intérêt ne figurent pas dans les bases de données et l'identification par l'utilisation de la spectrométrie de masse seule devient une tâche extrêmement complexe d'où la nécessité d'y

associer d'autres techniques analytiques pour parvenir à une formule développée. Il est ainsi le plus souvent nécessaire de procéder à de longues étapes de purification et de concentration pour permettre l'analyse par RMN de composés initialement détectés par spectrométrie de masse, mais très fréquemment, seule les données de spectrométrie de masse sont disponibles et la caractérisation des métabolites passe par une interprétation méticuleuse des spectres de masse et de MS<sup>n</sup>.

### 7.3. Bases de données

Le travail d'identification repose donc en partie sur la richesse des bases de données dont il existe plusieurs types (Tableau 3) qui vont être brièvement commentés.

Tableau 3 : Principales bases de données utilisées en analyse métabolomique [193]

Database		Thematic	Conception/URL
BiGG	●	Human	University of California (USA) ( <a href="http://www.biggs.ucsd.edu/">www.biggs.ucsd.edu/</a> )
BioCyc (HumanCyc, MetaCyc)	●	Biochemical pathways	SRI International (USA) ( <a href="http://www.biocyc.org/">www.biocyc.org/</a> )
ChEBI	●	General	European Bioinformatics Inst. (UK)/European Molecular Biology Lab ( <a href="http://www.ebi.ac.uk/chebi/">www.ebi.ac.uk/chebi/</a> )
ChemFinder	●	General	Cambridge Soft (USA) ( <a href="http://www.chemfinder.CambridgeSoft.com">www.chemfinder.CambridgeSoft.com</a> )
CHEMnetBASE (Dict. Nat. Prod.)	○	General	Chapman & Hall/CRC ( <a href="http://www.chemnetbase.com/">www.chemnetbase.com/</a> )
CSLS	●	General	CADD Lab. Med. Chem NCI, NIH (USA) ( <a href="http://129.43.27.140/cgi-bin/lookup/search">http://129.43.27.140/cgi-bin/lookup/search</a> )
Enhanced NCI Database Browser	●	General	CADD Lab. Med. Chem NCI, NIH (USA), U of Erlangen-Nuremberg (Germ.) ( <a href="http://129.43.27.140/ncidb2/-">http://129.43.27.140/ncidb2/-</a> )
Fiehn library	●	General	Fiehn Laboratory Univ California Davis: Genome center ( <a href="http://fiehnlab.ucdavis.edu/Metabolite-Library-2007">http://fiehnlab.ucdavis.edu/Metabolite-Library-2007</a> )
Golm	●	* Plant	Max Planck Institute for Molecular Plant Physiology (Germany) ( <a href="http://www.csbdb.mpimp-golm.mpg.de">www.csbdb.mpimp-golm.mpg.de</a> )
HMDB	●	* Human metabolites	Department of Computing Science, University of Alberta (Canada) ( <a href="http://www.hmdb.ca/extrindex.htm">www.hmdb.ca/extrindex.htm</a> )
KEGG ligand database	●	General	Kyoto University Bioinformatics center (Japan) ( <a href="http://www.genome.jp/kegg/ligand.html">www.genome.jp/kegg/ligand.html</a> )
KNAPSAcK	●	Natural products	RIKEN Plant Science Center (Japan) ( <a href="http://kanaya.naist.jp/KNAPSAcK/KNAPSAcK.php">http://kanaya.naist.jp/KNAPSAcK/KNAPSAcK.php</a> )
LipidMaps	●	Lipidomics	LIPID MAPS Bioinformatics Core (USA) ( <a href="http://www.lipidmaps.org/data/index.html">www.lipidmaps.org/data/index.html</a> )
LipidBank	●	Lipidomics	Japanese Conference on the Biochemistry of Lipids (Japan) ( <a href="http://www.lipidbank.jp/">www.lipidbank.jp/</a> )
Madison Metabolomics Consortium Database	●	General	National Magnetic Resonance Facility, University of Wisconsin-Madison ( <a href="http://mmcd.nmrfam.wisc.edu/">http://mmcd.nmrfam.wisc.edu/</a> )
MassBank	●	* General	Keio university, university of Tokyo, Kyoto university, RIKEN plant Science center (Japan) and others ( <a href="http://www.massbank.jp">www.massbank.jp</a> )
Merck Index	○	General	Merck publishing
Metlin	●	* Human metabolites	Scripps Center for Mass Spectrometry ( <a href="http://www.metlin.scripps.edu">www.metlin.scripps.edu</a> )
MoTo	●	* Metabolome database for tomato	Wageningen University ( <a href="http://www.appliedbioinformatics.wur.nl">www.appliedbioinformatics.wur.nl</a> )
MSlib	●	* Drugs, metabolites	University of Alberta ( <a href="http://www.ualberta.ca/~gjoness/mslib.htm">http://www.ualberta.ca/~gjoness/mslib.htm</a> )
NIST	⊙	* General	National institute for standard and technology (USA) ( <a href="http://www.nist.gov/srd/nist1a.htm">www.nist.gov/srd/nist1a.htm</a> )
PubChem	●	General	National Center for Biotechnology Information (USA) ( <a href="http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pccompound">www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pccompound</a> )
SDBS	●	* General	National Institute of Advanced Industrial Science and Technology ( <a href="http://www.riodb01.ibase.aist.go.jp/sdbs/">www.riodb01.ibase.aist.go.jp/sdbs/</a> )
SciFinder	○	General	American Chemical Society (USA)

(●) Free access, (⊙) partially free access, (○) licenced and (\*) spectral database.

Les bases de données peuvent être classées en 4 catégories :

- Les bases de données spectrales : Elles ont pour vocation de regrouper les spectres de fragmentation de standards afin de faciliter l'identification de composés inconnus sur la base d'une homologie de profil de fragmentation. Elles contiennent essentiellement des données de GC/MS par ionisation électronique. Citons par exemple *the Golm metabolome database* [194] spécialisée dans les

métabolites d'origine végétale, ou la base de données spectrale du *National Institute of Standards and Technology* (NIST11) qui comporte les spectres de fragmentation de plus de 200000 composés.

- Les bases de données métabolomiques : D'apparition récente, ces bases de données ont été constituées à partir d'expériences métabolomiques. Elles ont pour avantage de ne comporter que des composés pertinents en termes de biologie puisqu'elles ont été construites à partir d'analyses de divers milieux biologiques. En plus des données spectrales contenues par certaines d'entre elles, elles indiquent le milieu et l'espèce dans lequel le composé a été retrouvé. Elles peuvent être généralistes comme HMDB (*Human Metabolome Database*) [195] ou ciblées sur un organisme (ex. *Atomic Reconstruction of Metabolism* (ARM) qui référence plus de 2700 réactions du métabolisme d'*E.coli*), ou encore spécifiques d'une classe chimique de composés comme *LipidBank* [196] ([www.metabolom.jp](http://www.metabolom.jp)) qui regroupe des informations d'activité biologique, physico-chimiques et spectrales sur plus de 7000 lipides.

- Les bases de données métaboliques ou biochimiques : Elles sont parmi les plus utilisées actuellement car elles répertorient un grand nombre de composés biologiques. La plus célèbre est certainement *The Kyoto Encyclopedia of Genes and Genomes* (KEGG). Cette dernière permet de relier les génomes à la biologie des systèmes en accueillant à la fois un répertoire des gènes et des protéines associées (*KEGG genes*), un répertoire de métabolites endogènes et exogènes (*KEGG ligand*) et un répertoire de réactions et cycles métaboliques liant ces différents intervenants (*KEGG pathway*).

- Les bases de données chimiques généralistes : Elles comprennent un très grand nombre de molécules qui ne sont pas toutes d'intérêt biologique puisqu'elles comprennent de nombreux composés d'origine synthétique. Parmi ces bases généralistes, la plus riche est sans doute celle de la société américaine de chimie (ACS), nommée Scifinder (CAS : *Chemical Abstract Service*) qui recense des millions de composés.

Cette classification est cependant peu adéquate dans la mesure où la plupart des bases de données regroupent plusieurs de ces fonctions. Ainsi, de plus en plus de bases de données métabolomiques comportent également des données spectrales de fragmentation comme HMDB, Metlin [197], et LipidBank par exemple.

Dans le futur, les bases métabolomiques devraient connaître un véritable essor. Devenant de plus en plus riches, elles permettent néanmoins d'associer aux métabolites de précieuses métadonnées comme le milieu biologique et l'espèce où ils ont été retrouvés, et d'éliminer les composés chimiques synthétiques peu pertinents qui sont présents dans les bases généralistes. Par ailleurs, dans l'objectif de faciliter l'identification des métabolites, ces bases s'enrichissent toutes de données spectrales. Malheureusement, comme nous l'avons vu, la variabilité des profils de fragmentation obtenus par API-MS devrait imposer l'enregistrement de données obtenues sur différents types d'appareils. Cette initiative n'a cependant été prise que récemment dans les bases de données métabolomiques japonaises telles que *MassBank* [198] et *LipidBank*, où l'on trouve des données issues de triple quadripôles, de pièges ioniques et de Qq-TOF pour les différents composés. En attendant la généralisation de cette approche multi-instrument, beaucoup de laboratoire ont

recours à la construction de bases de données internes construites uniquement à partir de leurs instruments propres. Ces bases de données internes, très consommatrices en temps et en moyens, constituent un savoir-faire et une valeur ajoutée qui peuvent être valorisées commercialement, au détriment des approches collaboratives qui permettraient d'obtenir plus rapidement des bases beaucoup plus complètes, indispensables à la mise en œuvre efficace des approches métabolomiques.

## 8. APPLICATIONS

Comme indiqué en Figure 15, Les approches métabolomiques trouvent des applications dans différents domaines tels que la toxicologie [199], le phénotypage d'organismes génétiquement modifiés [200], la microbiologie [201], la biologie végétale, le marquage de pathologie, la sécurité alimentaire [202], la nutrition [203;204] et à un niveau plus fondamental la biologie des systèmes [205]. Nous aborderons ici principalement les applications réalisées dans le domaine de la santé (pharmacologie, biologie clinique, toxicologie...) et n'aborderons pas les thèmes relatifs aux microorganismes, aux plantes ou à la nutrition.



Figure 15 : Les principaux domaines d'application de la métabolomique.

### 8.1. Compréhension des systèmes vivants

Malgré les progrès réalisés en génétique et en biologie moléculaire, nos connaissances en matière de fonctions des gènes restent limitées, et les résultats issus des analyses métabolomiques sont largement exploitées dans le cadre d'études de génomique fonctionnelle en complément des

approches transcriptomiques et protéomiques. L'approche dite « FANCY » [8] constitue un travail pionnier en matière d'analyse métabolomique appliquée à la génomique fonctionnelle. Cette approche est basée sur l'idée que des modifications génétiques phénotypiquement silencieuses peuvent aboutir à des répercussions détectables au niveau du métabolome. Les auteurs de la publication décrivant cette approche suggèrent qu'il pourrait être possible d'attribuer une fonction à un gène en comparant les profils métaboliques de souches modifiées pour des gènes connus et inconnus.

Dans le même esprit, les approches métabolomiques ont été appliquées au phénotypage d'animaux mutants ou transgéniques et à l'étude des conséquences des modifications génétiques présentes chez ces animaux [85;206]. La mesure simultanée d'un grand nombre de métabolites peut aussi permettre la mise en évidence de nouvelles voies métaboliques. Quelques études ont ainsi relié des métabolites grâce à l'emploi de cartes de réseaux métaboliques et ont dès lors démontré l'intérêt des études métabolomiques pour visualiser la réponse de cellules ou d'organismes exposés à des agents toxiques [207]. Une publication traitant de la toxicité du paracétamol chez la souris illustre ces propos [208]. En utilisant l'électrophorèse capillaire (EC) couplée à un spectromètre de masse équipé d'un analyseur à temps de vol, Soga et coll. ont détecté 1859 signaux dans des extraits hépatiques de souris. En analysant les modifications métaboliques induites par le paracétamol, cette équipe a mis en évidence une corrélation entre l'augmentation des concentrations plasmatiques d'acide ophtalmique et la déplétion en glutathion hépatique (Figure 16). Ces auteurs ont donc proposé l'utilisation de cet acide en tant que biomarqueur de stress oxydant.

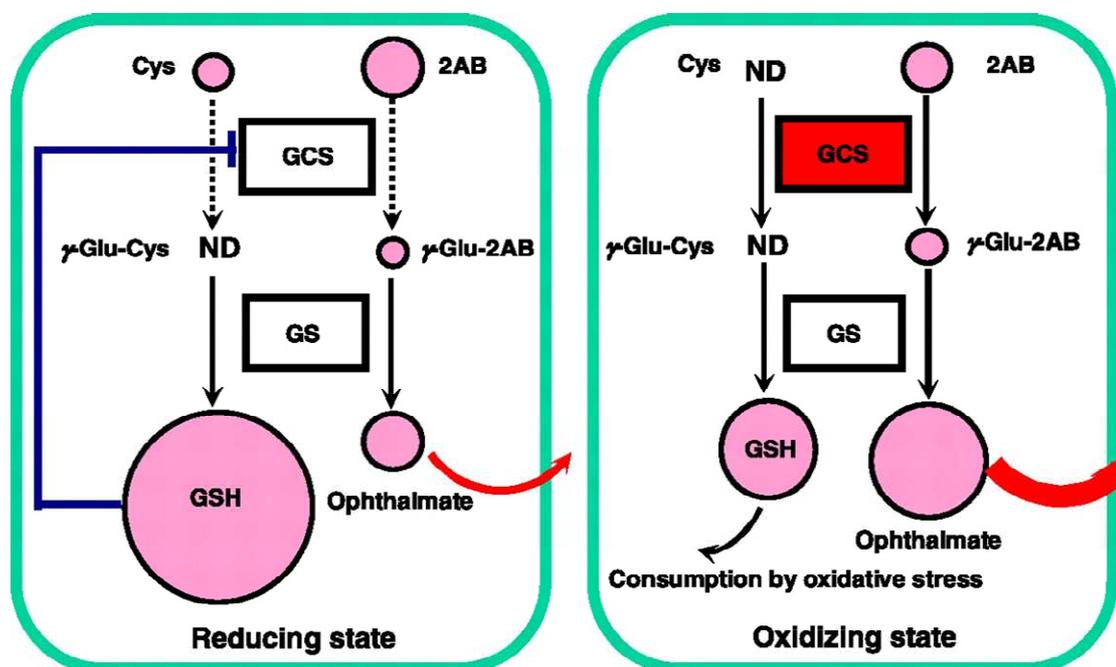


Figure 16 : Biosynthèse de l'acide ophtalmique dans les hépatocytes, et corrélation avec la concentration en glutathion hépatique [208].

En absence de stress oxydant (schéma de gauche), la  $\gamma$ -Glutamylcystéine Synthétase (GCS) est inhibée par de fortes concentrations en Glutathion hépatique (GSH). C'est cette enzyme (GCS) qui synthétise l'acide ophtalmique à partir de l'acide 2-aminobutyrique (2AB). En cas de stress oxydant (comme l'intoxication au paracétamol, voir le schéma de droite) le GSH et

est consommé et la GCS activée. L'acide ophtalmique est alors synthétisé en grande quantité est transporté hors de la cellule par des transporteurs de la membrane hépatocellulaire (MRP). La taille des cercles pour chaque métabolite est proportionnelle à leur concentration mesurée dans le foie. ND: non détecté.

Les approches métabolomiques ont également été utilisées pour assigner aux enzymes de nouvelles activités et de nouveaux substrats [209]. Pour cela, un mélange complexe de métabolites (mélange de nutriments utilisés pour la culture de microorganismes) est incubé avec des protéines purifiées et le mélange de la réaction est analysé par CE/MS. Saghatelian et al. ont réalisé le même genre d'expériences mais *in vivo* en inhibant l'enzyme d'intérêt et en analysant les extraits de tissus (foies et reins) par LC/MS [210]. Les variations de concentrations de métabolites observées en comparant les conditions dans lesquelles l'enzyme est inhibée ou non, ainsi que l'identification des métabolites dont les taux ont été modifiés permettent de préciser la nature de l'activité enzymatique et de connaître les substrats et les produits de la réaction.

## 8.2. Applications en biologie clinique

En plus d'être une nouvelle composante de l'exploration globale des systèmes biologiques, l'analyse métabolomique, initialement réservée aux chimistes réalisant du développement méthodologique, est de plus en plus utilisée par les médecins car elle offre des perspectives dans les domaines biomédicaux et pharmaceutiques. Parmi eux, la recherche de biomarqueurs de différentes pathologies pour leur diagnostic et leur suivi a fait l'objet de revues bibliographiques [211-213]. Les données les plus anciennes concernent la recherche d'anomalies innées du métabolisme par dosage des acides organiques urinaires en GC/MS [214]. D'autres pathologies comme la myopathie de Duchenne [215] ou le cancer [216;217] ont été explorées, principalement en RMN. Shaham et al. [218]. ont démontré la pertinence de l'utilisation de créatine plasmatique pour le biomarquage des pathologies de la chaîne respiratoire, plutôt que le lactate ou l'alanine, marqueurs actuellement connus mais moins sensibles. Ils ont confirmé l'augmentation des concentrations de créatine, observée sur cultures cellulaire, dans le plasma de patients et ont apporté une explication biologique à cette observation. La recherche de biomarqueurs peut également être mises en œuvre dans des cadres aussi divers que des études portant sur les effets de stress environnementaux [219] ou d'un traitement anabolisant dans le cadre de stratégies de contrôle de dopage [220].

L'analyse métabolomique peut également contribuer au développement de modèles animaux de pathologies humaines afin d'améliorer leur compréhension, d'identifier des biomarqueurs pour leur diagnostic biologique et de caractériser de nouvelles cibles thérapeutiques [206]. Wu et coll. ont ainsi utilisé le profilage métabolique par spectrométrie de masse pour développer un modèle murin d'étude d'une maladie métabolique humaine d'origine génétique portant sur une anomalie du métabolisme des acides aminés à chaîne ramifiée : la leucinose. Ils ont pour cela administré aux animaux un agent mutagène, la N-éthyl-N-nitroso-urée. Cette molécule provoque des mutations aléatoires dans le génome des souris. Grâce à l'analyse des urines de ces souris par spectrométrie de masse, les auteurs de cette étude ont pu sélectionner des animaux présentant un phénotype similaire à celui des patients atteints de leucinose.

Les études métabolomiques peuvent également permettre d'améliorer notre compréhension des pathologies. Pour cela, la complémentarité des différentes méthodes d'études post-génomiques commence à être exploitée. Une telle démarche a récemment été employée pour l'étude de la schizophrénie [221] à partir de tissus cérébraux humains en combinant les approches transcriptomiques, protéomiques et métabolomiques. Les auteurs de cette étude ont ainsi mis en évidence, grâce à ces trois approches post-génomiques, des preuves convergentes de perturbations dans un certain nombre de voies métaboliques mettant en évidence une augmentation du stress oxydatif dans la schizophrénie. Elles peuvent aussi permettre de caractériser des pathologies en regroupant des patients sur la base de leur phénotype métabolique (encore appelé « métabotype » [18]). On peut utiliser les variations observées entre ces métabotypes, et les relier à des données biologiques ou phénotypiques (maladie, obésité, pression artérielle) afin de renforcer la pertinence biologique des signatures métaboliques obtenues. Afin d'identifier de nouvelles anomalies métaboliques, Mochel et al. ont analysé le liquide céphalorachidien (LCR) de patients présentant des troubles neurodégénératifs complexes d'étiologie inconnue. Ils ont pu ainsi mettre en évidence une élévation d'acide sialique chez 5 patients présentant tous des symptômes communs dont une ataxie liée à une atrophie cérébelleuse. Ils ont ainsi caractérisé un syndrome n'avait jamais été décrit [222].

L'analyse métabolomique devient maintenant une technique génératrice d'hypothèses quant à l'implication de certaines voies métaboliques dans des pathologies telles que le cancer [223] ou la drépanocytose [224]. Sreekumar et al. ont ainsi montré l'intérêt de la sarcosine pour diagnostiquer des cancers agressifs de la prostate. La pertinence de la sarcosine a été renforcée par la mise en évidence de l'augmentation de l'activité méthyltransférase dans les tissus prostatiques. La compréhension du mécanisme de progression de cette pathologie ouvre des pistes pour la thérapeutique [223]. Des travaux mettant en jeu une démarche similaire ont été réalisés dans le domaine de la drépanocytose, une maladie génétique très fréquente et invalidante. L'analyse du métabolome des globules rouges par chromatographie liquide couplée à la spectrométrie de masse a permis d'obtenir une signature métabolique caractéristique des patients drépanocytaires. La modification de plusieurs voies métaboliques chez les globules rouges des malades (glycolyse, métabolisme du glutathion, du monoxyde d'azote et de l'arginine, renouvellement membranaire) a pu ainsi être mis en évidence [225]. Le métabolisme du NO revêt un intérêt particulier dans la mesure où il pourrait être lié aux complications vaso-occlusives observées chez ces patients. Dans une autre étude, l'analyse métabolomique de plasma de souris transgéniques constituant un modèle pour l'étude de la drépanocytose a révélé des concentrations d'adénosine plus élevées que la normale chez ces animaux. Cette augmentation de concentration a également été retrouvée dans les échantillons plasmatiques de patients atteints de drépanocytose. En utilisant des approches biochimiques et pharmacologiques, les auteurs de l'étude sont arrivés à montrer que la liaison de l'adénosine au récepteur  $A_{2B}$  est susceptible d'induire une augmentation des concentrations de 2,3-bisphosphoglycérate, qui provoque une déformation ou falciformation des érythrocytes, à l'origine de crises vaso-occlusives [224].

L'analyse métabolomique constitue une technique de choix pour appréhender les interactions entre le fonctionnement d'un organisme et son environnement, dont certaines seraient impliquées dans le développement de pathologies. Ainsi, l'utilisation de la métabolomique pour établir des relations entre flore intestinale et pathogenèse est une thématique en plein essor [52;226;227]. La flore intestinale est essentielle à la digestion des aliments, et notamment à l'absorption et au stockage de l'énergie par l'organisme. Elle est donc associée à homéostasie énergétique et il existe de nombreuses interactions entre l'activité enzymatique de la flore et le phénotype métabolique d'un individu. Cette symbiose entre la flore intestinale et l'individu hôte a déjà été mise en évidence par analyse métabolomique [52;228] ainsi que son implication dans certaines pathologies telles que l'obésité [227;229] ou le diabète [226]. Enfin, de récents travaux suggèrent l'existence d'entérotypes. Il s'agit de signatures bactériennes intestinales qui caractérisent un groupe d'individus, indépendamment de l'origine géographique des individus, et qui sont dus à des différences de composition en espèces bactériennes [230]. Ces entérotypes pourront peut-être expliquer des variations physiologiques interindividuelles et la prédisposition à certaines pathologies.

### **8.3. Applications en toxicologie**

La toxicologie vise à étudier les effets néfastes des produits chimiques (xénobiotiques) sur les organismes vivants. La toxicité d'une substance dépend de sa capacité à perturber les fonctions biologiques à différents niveaux d'organisation biologique (cellules, tissus ou organes). Elle est liée à l'importance et à la durée de l'exposition, au degré d'absorption du toxique par l'organisme, à sa distribution, sa biotransformation et son élimination ou accumulation. Les biomarqueurs sont utiles pour prédire un événement toxique avant la survenue de signes cliniques (biomarqueurs d'effet précoce), pour évaluer la gravité d'une intoxication (biomarqueurs d'effet), et également pour surveiller les patients exposés (biomarqueurs d'exposition). La survenue d'effets indésirables peut avoir des origines multifactorielles, notamment dues à des interactions individu-environnement, qui sont difficiles à évaluer en utilisant des approches conventionnelles. La métabolomique, qui permet de suivre les perturbations homéostatiques et les interactions hôte-environnement, est d'un intérêt particulier dans ce contexte. Les études pionnières basées sur l'utilisation de la RMN ont déjà été publiées [85;231-234]. Signalons enfin l'existence de consortiums tels que le COMET («*Consortium On MEtabonomic Technology*») coordonné par l'équipe de Jeremy Nicholson à l'Imperial College de Londres, qui ont commencé à développer des modèles basés sur l'analyse métabolomique pour la classification des toxiques [233].

L'analyse métabolomique permet de contribuer à une meilleure compréhension des mécanismes de toxicité. Williams et al. ont ainsi montré l'intérêt d'utiliser la métabonomique pour élucider le mécanisme de toxicité rénale de la D-sérine chez le rat [59;235]. Les analyses LC/MS ont conduit à l'identification de plusieurs métabolites (hydroxypyruvate, glycérate, acylcarnitine...) suggérant l'implication d'une enzyme, la D-amino-acide oxydase peroxysomale (D-AAO), qui produit le glycérate et l'hydroxypyruvate à partir de la D-sérine. Les auteurs ont émis l'hypothèse que l'hydroxypyruvate génère du peroxyde d'hydrogène, ce qui induit un stress oxydatif péroxisomal. Il a été démontré dans

une autre étude que la co-administration de D-sérine et le benzoate de sodium, un puissant inhibiteur compétitif de la D-amino-acide oxydase, prévient les lésions rénales, confirmant ainsi l'implication de cette enzyme dans le mécanisme de toxicité [236]. De nouveaux aperçus sur les mécanismes de toxicité seront probablement obtenus par le couplage des outils classiques histologiques et biochimiques avec d'autres approches omiques et d'imagerie. Dans ce contexte, un effort de recherche collaborative en matière de toxicologie moléculaire a été lancé par le Centre National pour la recherche toxicologique de la FDA et BG Medicine Inc. Soutenu par sept sociétés pharmaceutiques et trois fournisseurs de technologie, il vise à étudier la toxicité hépatique induite par les médicaments. Les résultats préliminaires sont présentés dans une publication dans laquelle les performances des plateformes analytiques vis-à-vis de deux premiers composés hépatotoxiques (entacapone et tolcapone) sont présentés et discutés [237].

De nombreuses études de recherche de biomarqueurs de toxicité sont réalisées en utilisant différentes plateformes analytiques, telles que <sup>1</sup>H-RMN, GC/MS et LC/MS, afin de maximiser la couverture de détection des métabolites. La plupart de ces études tentent d'aborder la question de la toxicité d'organe et visent à trouver des changements de concentration en métabolites liés à la substance toxique, survenant avant la détection clinique ou histopathologique, et susceptibles d'être plus spécifiques que des biomarqueurs conventionnels tels que les mesures d'activités alanine aminotransférase (ALAT), aspartate aminotransférase (ASAT) ou le dosage de la bilirubine pour l'hépatotoxicité, ou encore le dosage de l'azote uréique du sang pour une néphrotoxicité. Dans le cadre de ces études, le paracétamol (N-acétyl-p-aminophénol, APAP, ou encore acétaminophène) est fréquemment utilisé comme un médicament modèle d'hépatotoxicité. En étudiant la toxicité aiguë et chronique du paracétamol sur des rats Sprague-Dawley [238], Sun et al. ont mis en évidence une diminution des concentrations urinaires de 1-méthylnicotinate. Liée à la voie de biosynthèse du glutathion, cette molécule est produite lors de la conversion de la S-adénosylméthionine (SAM) en S-adénosylhomocystéine (SAH). La diminution de concentrations observées est liée à une déplétion en S-adénosylméthionine, qui est d'ailleurs détectée par les auteurs, et qui est liée à une activation de la voie de synthèse du glutathion, nécessaire à l'élimination du métabolite toxique du paracétamol : la N-acétyl-p-benzoquinone imine ou NAPQI (Figure 17). Ces résultats ont été confirmés par une autre étude qui portait sur la régulation de la voie trans-sulfuration dans des conditions de toxicité hépatique, et également des données d'expression génique [239].

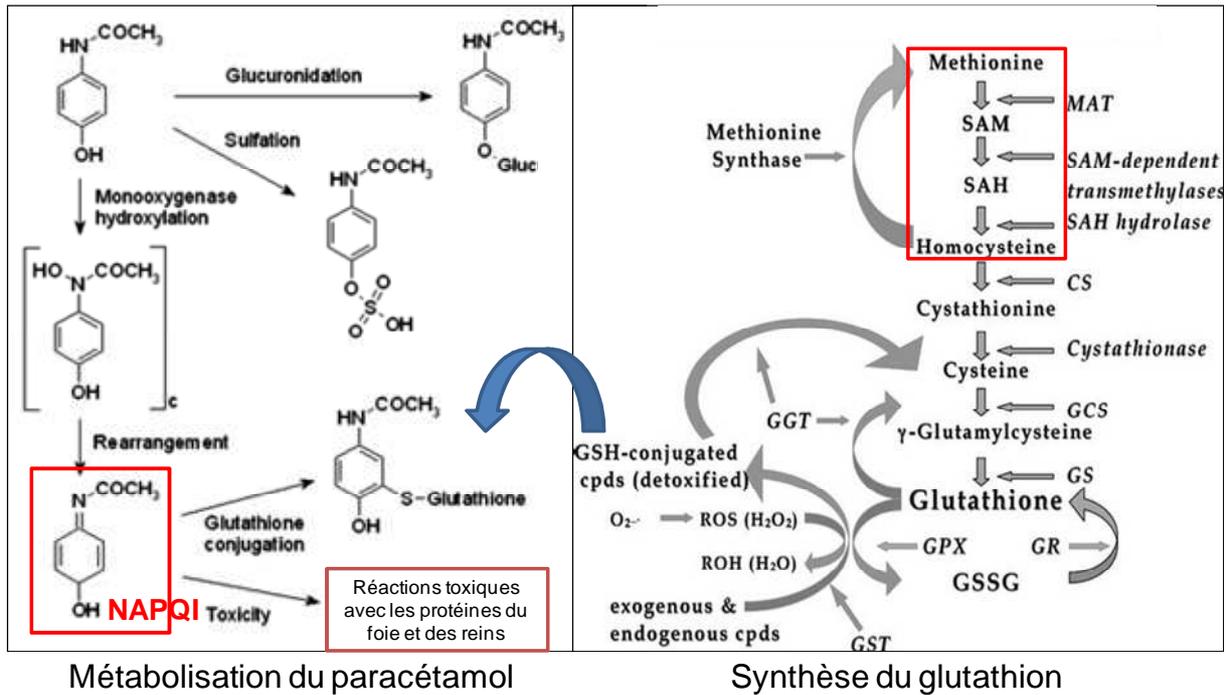


Figure 17 : Implication de la voie de synthèse du glutathion (schéma de gauche) dans la métabolisation du paracétamol (schéma de droite).

(Images issues des sites internet [http://en.wikipedia.org/wiki/File:Paracetamol\\_metabolism.png](http://en.wikipedia.org/wiki/File:Paracetamol_metabolism.png) et <http://herkules oulu.fi/isbn9514270703/html/x305.html>)

Outre la découverte de biomarqueurs, d'autres études métabolomiques rapportent la construction de modèles pour la prédiction et la classification des toxicités. Ce type de modèles est particulièrement intéressant pour l'industrie pharmaceutique. Ainsi, une étude réalisée chez le rat avec quatre hépatotoxiques différents (le R-naphtyle isothiocyanate ou ANIT, le tétrachlorure de carbone CCl<sub>4</sub>, l'APAP, et le diclofénac) [240] a permis de détecter des changements de concentration en métabolites caractéristiques de chaque hépatotoxine, et un modèle mathématique a été construit permettant de prévoir l'hépatotoxicité avec une prédictibilité supérieure à 95%. Une telle approche a également été réalisée par Boudonk et al. [241] avec 3 médicaments (gentamicine, cisplatine et tobramycine) induisant une néphrotoxicité tubulaire proximale. Une augmentation des concentrations en acides aminés et polyamines a été observée dans les urines et une diminution des concentrations en nucléosides puriques et pyrimidiques a été détectée dans les tissus rénaux avant la lésion rénale observable par histologie conventionnelle et les outils de la chimie clinique. Les métabolites urinaires présentant d'importants changements avec les 3 médicaments, tels que les acides aminés à chaîne ramifiée, l'hippurate et de glucose ont ensuite été sélectionnés pour construire un modèle de prédiction de l'apparition précoce de la néphrotoxicité.



# PRESENTATION DU TRAVAIL DE THESE



# PRESENTATION DU TRAVAIL DE THESE

La métabolomique ouvre de nouvelles perspectives pour la découverte de biomarqueurs dans le domaine de la nutrition et de la santé, et également dans le développement de la biologie des systèmes. Les premières années d'existence de la métabolomique ont été consacrées aux développements méthodologiques pour obtenir des données biologiquement exploitables. Les étapes de traitement des données et d'identification étaient alors particulièrement limitantes, ce qui n'est plus de cas de nos jours grâce aux évolutions en matière de traitement de l'information ainsi que l'avènement de la haute résolution. Cependant les profils métaboliques d'extraits cellulaires, de tissus ou de fluides biologiques obtenus par spectrométrie de masse peuvent contenir quelques centaines à plusieurs milliers de signaux dont la majeure partie demeure jusqu'à présent inconnue, ou du moins inexploitée car non caractérisée par les systèmes analytiques utilisés, ce qui limite l'obtention de données biologiquement pertinentes. Dans ce contexte, l'un des objectifs de cette thèse est le développement d'une base de données spectrale par ESI-LC/MS pour l'annotation de données issues d'analyses métabolomiques par spectrométrie de masse à haute résolution et son application à l'analyse des échantillons d'urine d'une cohorte de volontaires. D'autre part, un des points importants en métabolomique est le contrôle des conditions expérimentales et leur influence sur les empreintes métaboliques. Il faut pouvoir identifier les métabolites mais également pouvoir distinguer ceux discriminants pour des conditions particulières (maladie, exposition à un toxique...) de ceux fluctuants avec les nombreuses sources de variabilité environnementale, biologique ou analytique. Le second objectif de ce travail de thèse a donc été d'étudier les sources de variabilité biologique et leur influence sur les concentrations des métabolites identifiés. Enfin le dernier objectif de ce travail de thèse a été d'étudier l'impact de la variabilité due aux conditions de prélèvement et de conservation des échantillons d'urine.

## **1 CONSTRUCTION D'UNE BASE DE DONNEES SPECTRALE POUR L'ANNOTATION DE JEUX DE DONNEES BIOLOGIQUES**

Dans la première partie de ce travail de thèse, nous décrivons la construction d'une base de données spectrale pour l'identification de métabolites dans les échantillons biologiques. Ce travail a été réalisé sur des urines humaines issues d'une cohorte de 227 individus. En effet l'urine est un liquide organique composé des déchets de l'organisme. Elle est sécrétée en fin de chaîne de détoxification et d'élimination de l'organisme, et son recueil est non invasif. L'urine est donc un fluide très utilisé dans les études toxicologiques et cliniques, et la détermination des métabolites que l'on peut y trouver est de première importance. Deux approches complémentaires ont été utilisées : (i) l'obtention de spectres de masse et de MS/MS à partir de composés de référence, et (ii) l'identification de métabolites souvent non commercialement disponibles directement à partir des données biologiques. En effet, l'identification formelle d'un métabolite ne peut être obtenue que par comparaison des spectres de masse ou de CID et des temps de rétention chromatographiques du

composé à identifier avec ceux de composés de références. Or, le nombre de molécules commercialement disponibles est limité, et une approche complémentaire d'identification putative, basée sur l'interprétation de spectres de masse et de MS/MS et sur des données bibliographiques, est indispensable à la caractérisation d'un maximum de métabolites.

## **2 EVALUATION DE L'IMPACT DE LA VARIABILITE BIOLOGIQUE SUR LE METABOLOME URINAIRE**

Certains paramètres biologiques (âge, IMC, genre) peuvent avoir une influence sur les concentrations en métabolites urinaires [54] ou plasmatiques [53]. Ces facteurs pouvant être confondants lors d'études cliniques ou toxicologiques, il est pertinent d'évaluer cette variabilité biologique. A partir des métadonnées biologiques des 227 urines analysées pour l'élaboration de la base de données spectrale, des analyses statistiques ont été réalisées afin de déterminer quels métabolites sont corrélés à ces paramètres. A terme, l'objectif est de commencer la construction d'une banque de données relationnelle de métabolite dans différents biofluides.

## **3 CONDITIONS DE PRELEVEMENT ET DE STOCKAGE A COURT TERME DES URINES ET LEUR IMPACT SUR LE METABOLOME URINAIRE**

L'urine est un fluide biologique complexe dont les conditions de prélèvement et de stockage peuvent modifier la composition (contamination bactérienne, dégradation des métabolites...). Or il existe peu d'études sur le sujet et leurs résultats sont assez contradictoires. La dernière partie de ce travail de thèse a été de déterminer l'impact des conditions de prélèvement des urines, notamment sur les métabolites identifiés grâce à la base spectrale développée en première partie de thèse, afin d'uniformiser les conditions de prélèvement, de limiter la contamination bactérienne des échantillons et de ne pas biaiser les résultats biologiques potentiels avec une source de variabilité expérimentale non maîtrisée.

CHAPITRE 1 : CONSTRUCTION D'UNE  
BASE DE DONNEES SPECTRALE POUR  
L'ANNOTATION DE JEUX DE DONNEES  
BIOLOGIQUES



# CHAPITRE 1 : CONSTRUCTION D'UNE BASE DE

## DONNEES SPECTRALE POUR L'ANNOTATION DE

### JEUX DE DONNEES BIOLOGIQUES

## 1 INTRODUCTION

L'initiative de création de la première base de données de spectres de masse date des années 1950. L'institut pétrolier américain (*American Petroleum Institute, API*) a commencé à collecter des spectres par ionisation électronique (seule méthode d'ionisation disponible à l'époque) de mélange d'hydrocarbures dans un but de quantification plus que d'identification [242]. Parallèlement d'autres initiatives se sont développées, pour des molécules autres que les hydrocarbures [243]. Dans les années suivantes, les besoins des laboratoires de spectrométrie de masse en matière d'identification n'ont cessé de croître, ce qui a conduit à la création du comité de collecte de spectre de masse de référence (FWM) par l'organisation nationale de spectrométrie de masse et l'ASTM (*American Society for Testing and Materials*). L'objectif de ce comité était de permettre la centralisation des données de spectrométrie de masse. De nombreux laboratoires ont participé à l'initiative et en 1963 la base de données comportait plus de 4000 spectres de masse. En 1969, la base de données répertorie plus de 6800 spectres et est commercialisée pour la première fois par Wiley-Interscience [244]. Quelques années après Wiley, en 1978, c'est le bureau national des standards (*National Institute of Standards and Technology NIST*) qui distribue sa propre base de données contenant 2600 spectres (à cette même date celle de Wiley en comprends 41000). Après quelques années, la commercialisation des appareils de chromatographie en phase gazeuse couplés à la spectrométrie de masse a permis une forte expansion du nombre de spectres disponibles, et le développement d'algorithmes de comparaison et d'interprétation de spectres ont favorisé l'essor de ces bases de données (Figure 18). Ainsi en 1989 la base de données Wiley comporte 140000 spectres (correspondant à 107000 composés) et celle du NIST 54000 spectres. C'est cette année là qu'est commercialisée pour la première fois une base de données mixte, comprenant à la fois les spectres de la base Wiley et ceux du NIST. En 2009 cette base mixte contient presque 800000 spectres (correspondant à 667000 composés). D'autres bases de données de spectres basées sur l'ionisation à impact électronique ont également vu le jour dans le domaine de la métabolomique, combinant des spectres MS et des temps ou indices de rétention de chromatographie en phase gazeuse [245].

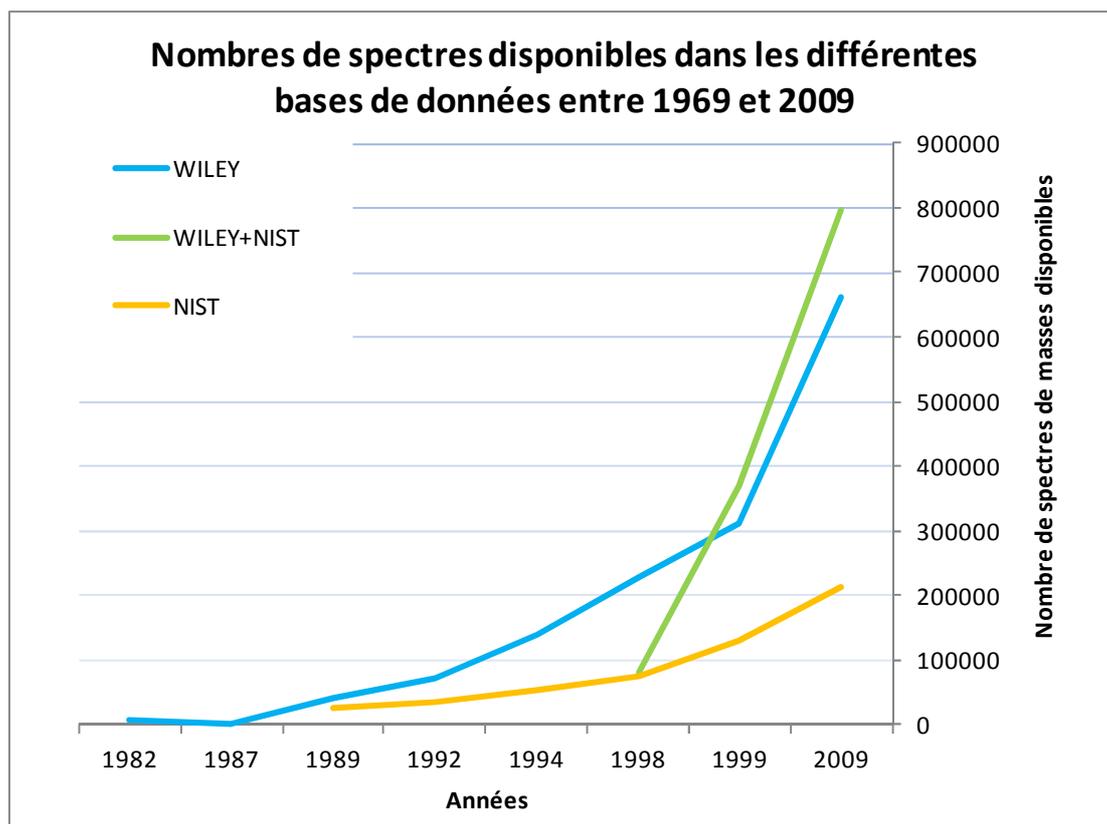


Figure 18 : Nombres de spectres disponibles dans les différentes bases de données spectrale entre 1969 et 2009 [246].

Ces bases de données s'avèrent indispensables pour les laboratoires utilisant la spectrométrie de masse à source d'ionisation par impact électronique (EI). Malheureusement, les informations contenues dans ces bases ne sont pas applicables aux données provenant de spectromètres de masse à sources d'ionisation à pression atmosphérique, telles que l'électrospray (ESI). En effet, EI et ESI sont deux méthodes d'ionisation radicalement différentes, et il en va de même pour les spectres de masse qui en résultent. Contrairement à l'ionisation par EI pour laquelle les spectres de masse sont riches en informations structurales, les spectres de masse obtenus en ESI permettent principalement de détecter l'espèce moléculaire et n'apportent dans la plupart des cas peu ou pas d'informations structurales (Figure 19). En effet, les ions produits en EI sont des ions radicaux (à nombres impairs d'électrons) qui sont très instables et se décomposent facilement à cause de la très forte énergie interne qui leur est transmise lors du processus d'ionisation. De nombreux fragments porteurs d'informations structurales sont donc présents dans le spectre de masse, et il est souvent difficile de détecter l'espèce moléculaire porteuse de l'information de la masse moléculaire du composé à analyser, comme en témoigne le spectre obtenu en EI pour l'acide cholique dans la figure 19. A l'inverse, les ions produits en ESI sont des espèces moléculaires (à nombre pair d'électrons) qui sont stables, et se fragmentent peu de manière spontanée. Il est donc nécessaire de faire de la MS/MS voire de la MS<sup>n</sup> afin d'obtenir des informations structurales. C'est la raison pour laquelle la majeure partie des travaux de banque de données en ESI sont réalisés en mode MS/MS.

Dès le début des années 2000, les premières bases de données d'ESI-MS/MS commencent à voir le jour. Elles sont initialement orientées vers la toxicologie [247;248] mais se développent aussi en

pharmacologie [249], en physiologie végétale [250], et également en métabolomique. A titre d'exemple, MassBank [198;251] est une base de données publique de MS, dédiée aux petites molécules, et qui regroupe des spectres de masse d'ESI-MS/MS mais aussi d'ESI-MS, et d'EI-MS. Cependant, le partage et l'utilisation des données est rendu délicat compte tenu du manque de robustesse et de reproductibilité des sources d'ionisation à pression atmosphérique [252]. Il existe tout de même de récentes initiatives qui tendent à faciliter le partage de ces données en ESI en essayant de standardiser les conditions d'acquisition des spectres [253] ou en développant des algorithmes pour surmonter le manque de reproductibilité de l'ESI [254;255].

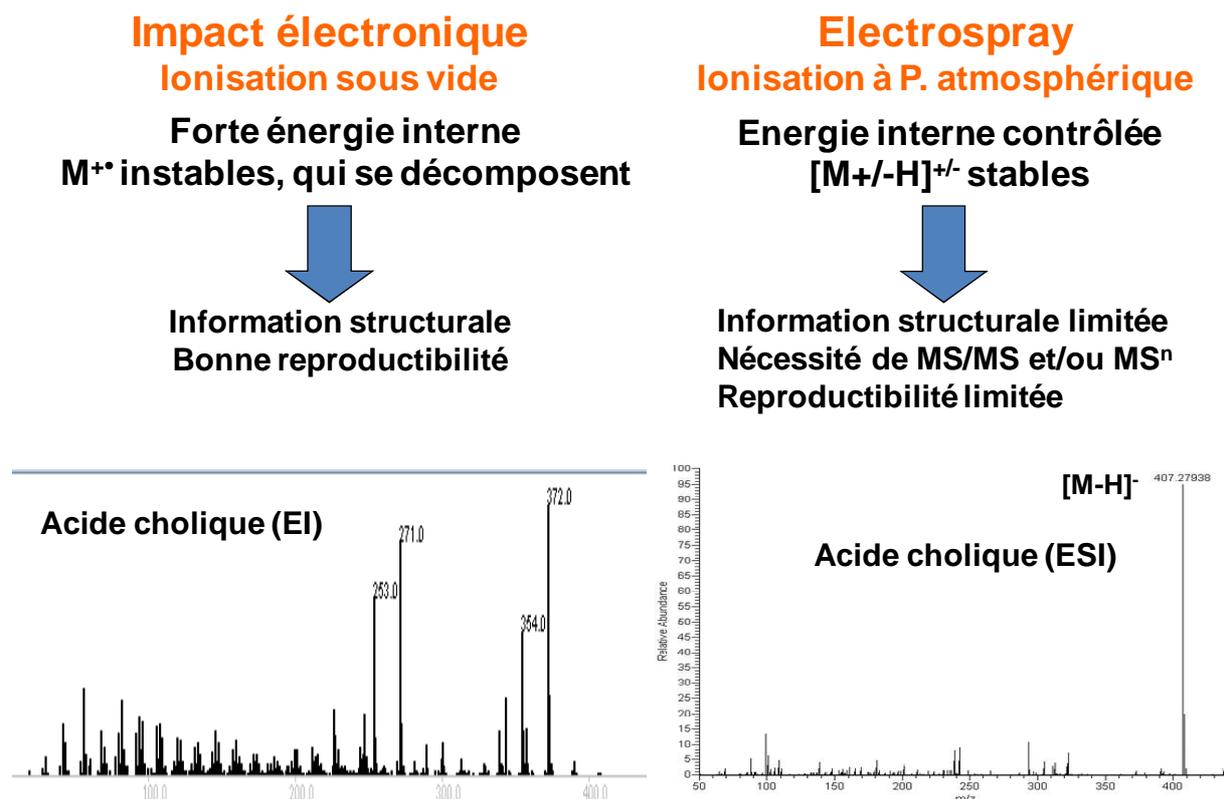


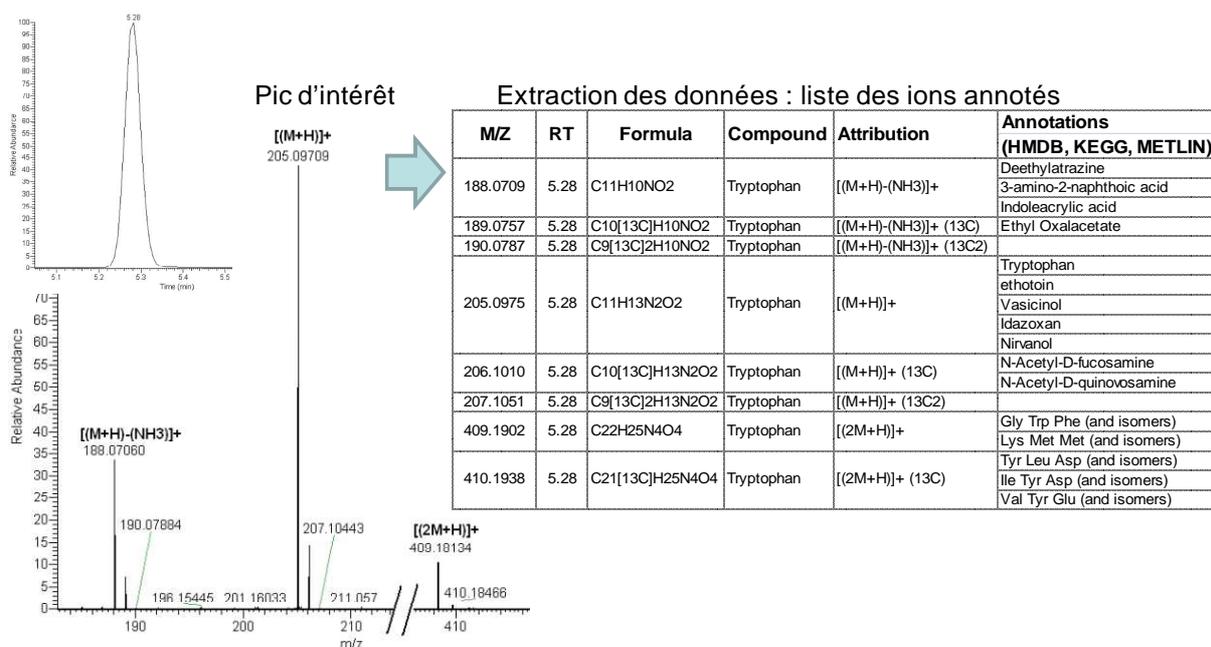
Figure 19 : Comparaison de l'ionisation par impact électronique et l'électrospray.

Cependant, l'analyse des données provenant des logiciels de détection automatique des signaux montre que les propos précédents concernant l'ionisation à pression atmosphérique doivent être tempérés. En effet, L'ESI est certes un processus d'ionisation doux qui génère des ions principalement moléculaires, permettant ainsi de remonter jusqu'à la formule brute de la molécule, mais également différents types d'ions qui peuvent parfois aider à la caractérisation de molécules inconnues :

- des ions isotopes (<sup>13</sup>C, <sup>15</sup>N, <sup>18</sup>O, <sup>34</sup>S, <sup>37</sup>Cl...) en fonction de la présence de certains atomes dans la molécule et de leur abondance naturelle.
- des ions adduits avec des cations dans le mode d'ions positifs (par exemple, Na<sup>+</sup>, K<sup>+</sup>, NH<sub>4</sub><sup>+</sup>) ou des anions dans le mode d'ions négatifs (HCOO<sup>-</sup>, CH<sub>3</sub>COO<sup>-</sup>, Cl<sup>-</sup>).

- des ions produits formés par la fragmentation prompte « en source » de l'ion précurseur (perte de H<sub>2</sub>O, NH<sub>3</sub>, CO<sub>2</sub>, SO<sub>3</sub>...).
- des homo ([2M+H]<sup>+</sup>, [2M-H]<sup>-</sup>).et hétéro dimères (adduits neutres [M+CH<sub>3</sub>CN+H]<sup>+</sup>).

Ces différents types d'ions peuvent être retrouvés dans les matrices de données générées par les logiciels de détection automatique, comme le montre la figure 20. Cette redondance du signal peut être dans un premier temps problématique car elle complique beaucoup le processus de traitement des données [103]. Elle nécessite en effet de connaître tous les ions produits par un métabolite afin d'éviter de se focaliser sur l'identification d'un ion qui en fait s'avère être l'adduit ou la perte d'une autre molécule (Figure 20). Elle peut cependant parfois permettre d'identifier directement des métabolites quand le spectre de masse contient suffisamment d'ions fragments. Il apparait ainsi pertinent de construire des bases de données ESI-MS pour faciliter l'identification des métabolites comme cela a été proposé dès 2008, dans le but de réaliser une annotation systématique et automatique du plus grand nombre possible de signaux afin d'obtenir une vision la plus exhaustive possible du métabolome [103;256-258].



**Figure 20 : Annotation d'un jeu de données biologiques et mise en évidence de la redondance du signal en ESI-MS.**

A droite de la figure est présentée la liste des ions extraits par un logiciel de détection automatique à partir d'une analyse métabolomique par LC/MS d'un échantillon biologique. Les annotations putatives des signaux sont fournies dans la dernière colonne. Elles ont été obtenues en comparant les masses mesurées expérimentalement avec les masses théoriques des métabolites contenus dans des bases de données publiques telles que HMDB, KEGG et Metlin. Ces annotations indiquent la présence présumée de nombreux composés tels que l'acide indoleacrylique, l'oxaloacétate d'éthyle ou le tryptophane. Une autre annotation par notre base de données spectrale confirme la présence de tryptophane, mais montre que les autres annotations par les bases de données publiques sont erronées car les ions en question correspondent à des ions du tryptophane générés par la source électrospray pendant le processus de désolvatation. Cet exemple illustre la complexité des données API-MS et démontre qu'une inspection approfondie des spectres de masse est nécessaire avant l'interprétation biologique.

Deux approches complémentaires ont été conçues pour construire notre base de données spectrale. La première consiste en l'analyse des spectres de masse de composés de référence de notre chimiothèque (environ 1200 produits chimiques). En effet selon la « *Metabolomics Standards Initiative* » [191], l'identification formelle d'un métabolite dans une matrice biologique nécessite la comparaison d'au moins deux paramètres physico-chimiques (un temps de rétention chromatographique, masse précise et un spectre MS ou de CID) du métabolite à identifier à ceux d'un composé de référence analysé dans les mêmes conditions expérimentales..

Cependant un grand nombre de métabolites ne sont pas commercialement disponibles et sont difficilement synthétisables. Le processus d'identification est donc limité même si certains signaux du spectre de masse peuvent être putativement identifiés sur la base de leur masse précise, leur similarité de temps de rétention avec une molécule proche et leur spectre de fragmentation. C'est pourquoi nous avons développé une deuxième approche basée sur la caractérisation de nouveaux métabolites directement à partir de données expérimentales (Figure 21). Suite à la détection automatique des ions à partir des données brutes, les signaux d'intérêt sont mis en évidence par analyses statistiques multivariées. Les informations redondantes (c'est-à-dire isotopes, fragments produits en source et adduit générés au cours du processus de désolvatation en plus des espèces moléculaires) sont regroupées à l'aide d'outils mathématiques et informatiques. Enfin, les ions d'intérêt sont annotés grâce à des requêtes automatiques dans les bases de données publiques sur la base des masses précises des espèces moléculaires et caractérisés par des expériences MS<sup>n</sup>.

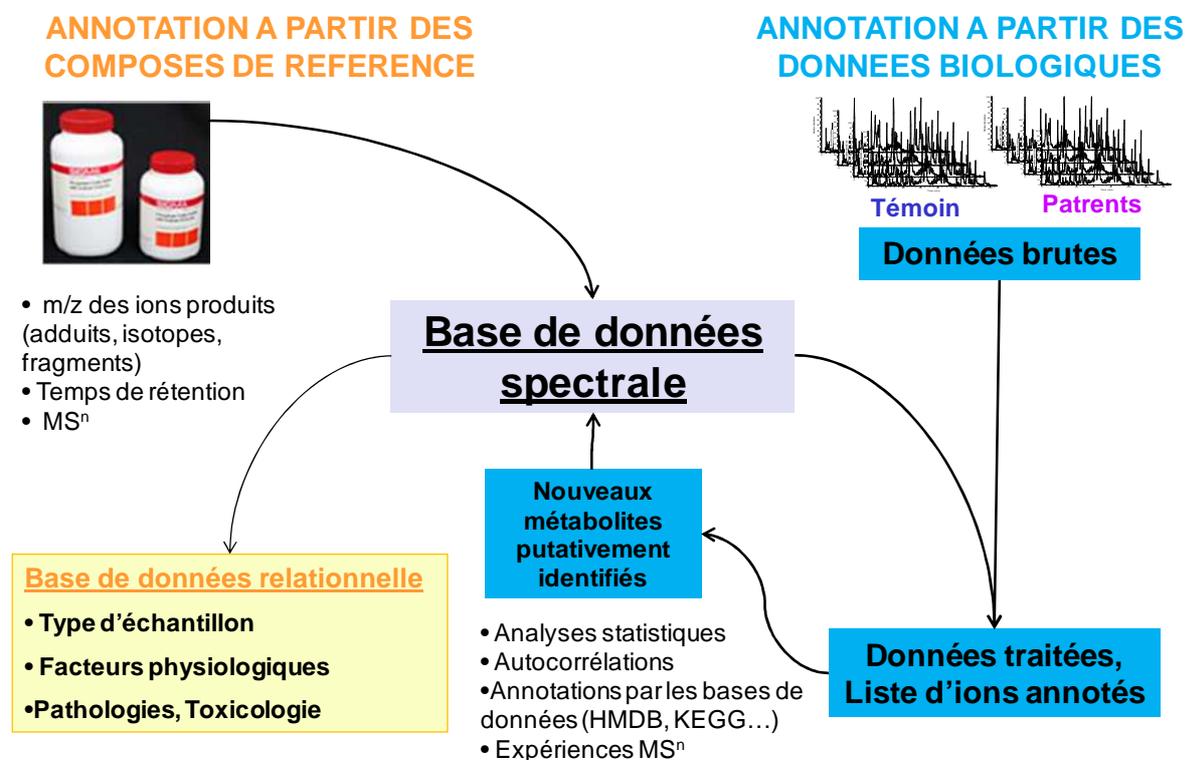


Figure 21 : Stratégie de construction d'une base de données spectrale ESI-MS

Les informations issues de ces deux systèmes sont importées sous forme de liste de pics dans la base de données spectrale. L'annotation automatique d'un jeu de données est réalisée par comparaison des valeurs de m/z et de temps de rétention du jeu de données avec celles de la base de données spectrale.

## 2 MATERIEL ET METHODES

### 2.1 REACTIFS ET PRODUITS CHIMIQUES

Les étalons internes et les composés de référence de notre chimiothèque ont été fournis par la société Sigma (Saint Quentin Fallavier, France), à l'exception des lipides qui ont été fournis par les sociétés Avanti Polar Lipids (Alabaster, USA) et Steraloids (Newport, USA). Les mélanges de standards utilisés pour la calibration de l'instrument MS (Calmix-positif et Calmix-négatif,) ont été fournis par Thermo Fisher Scientific (Les Ulis, France). L'acétonitrile (ACN) provient de la société SDS (Peypin, France) et l'acide formique de Merck (Briare-le-canal, France). L'eau a été déionisée et filtrée par un système de purification Millipore Milli-Q.

### 2.2 ECHANTILLONS BIOLOGIQUES

Les échantillons d'urine ont été recueillis par la médecine du travail du CEA lors de l'examen médical annuel. Des volontaires ont rempli un formulaire de consentement afin que le surplus de leurs prélèvements d'urine puissent être analysés dans notre laboratoire. Les échantillons récoltés sur plusieurs jours ont été anonymisés puis conservés à -80° C jusqu'à l'analyse. Les données physiologiques (âge, taille, poids, genre) ont été enregistrées pour chaque individu mais l'alimentation, la prise de médicaments et le mode de vie (exercice, tabac, alcool ...) n'ont pas été évalués. Les échantillons urinaires ont été recueillis auprès de 227 personnes, 121 hommes et 106 femmes, entre 18 et 81 ans. Les sujets peuvent être divisés en trois groupes d'âge : 18-25 ans ; 36-45 ans et de plus de 46 ans. La description détaillée de la cohorte est présentée dans le Tableau 4. Aucun critère d'exclusion ou de normalisation (alimentation, apport hydrique, exercice ...) n'a été appliqué afin d'obtenir un échantillon le plus représentatif possible de la population

Tableau 4 : Classification des 227 individus en fonction de l'âge, du genre et de l'indice de masse corporel.

Age (années)	18-35			36-45			>46			TOTAL AGE		
Genre*	F	H	Total	F	H	Total	F	H	Total	F	H	Total
<b>Indice de Masse Corporelle (IMC**)</b>												
Maigreur (<18.5)	3	0	3	2	0	2	0	2	2	5	2	7
Normal (18.5-24.9)	29	34	63	23	22	45	24	18	42	76	74	150
Surpoids (25-29.9)	2	11	13	8	9	17	8	22	30	18	42	60
Obésité (>30)	0	0	0	2	1	3	5	2	7	7	3	10
<b>TOTAL IMC</b>	<b>34</b>	<b>45</b>	<b>79</b>	<b>35</b>	<b>32</b>	<b>67</b>	<b>37</b>	<b>44</b>	<b>81</b>	<b>106</b>	<b>121</b>	<b>227</b>

\* F : Femme; H : Homme / \*\* IMC = poids (kg) / taille (m<sup>2</sup>)

## 2.3 PREPARATION DES ECHANTILLONS

### 2.3.1 COMPOSES DE REFERENCE

Les composés de références ont été pesés (1 mg) et dissous dans un solvant approprié pour obtenir une solution mère à 1 mg/ml. Une solution d'injection a ensuite été réalisée par dilution au 200<sup>ème</sup> : 5 µl de la solution mère ont été dilués dans 995 µl d'un mélange d'eau (H<sub>2</sub>O) et d'acétonitrile (ACN) (1/1, v/v) contenant 0,1% d'acide formique (HCOOH). Ces solutions d'injection ont été injectées en mode d'introduction directe dans le spectromètre de masse afin d'analyser les ions produits par la molécule. Des mélanges de solutions de composés de référence, contenant 10 à 15 composés, ont été préparés à partir de chaque solution mère et dilués dans le même mélange ACN/H<sub>2</sub>O (1/1, v/v) contenant 0.1% de HCOOH de façon à obtenir une concentration finale de 5 µg/ml par composé. Ces mélanges ont été injectés sur les différents systèmes chromatographiques afin de déterminer les temps de rétention des métabolites.

### 2.3.2 ECHANTILLONS BIOLOGIQUES

Les urines ont été centrifugées (5 minutes à 3000t/min) puis diluées dans de l'eau déminéralisée (dilution au 1/5 = 20 µl urine pour 80µl d'eau). Une partie de chacune des urines individuelles a été prélevée pour réaliser un pool d'urine qui servira de contrôle qualité (QC). Deux types d'échantillon de contrôle qualité sont préparés : un pool seul appelé QC1, et un pool auquel ont été ajouté des étalons internes (EI) appelé QC2. La solution mère d'EI utilisée est un mélange de 12 molécules non retrouvées dans les fluides biologiques en situation normale (molécules marquées, xénobiotiques). Les concentrations des EI, dans le mélange initial et dans les QC2 après dilution au 1/20 sont reportées dans le Tableau 5.

Tableau 5 : Composition de la solution d'étalons internes.

Nom du composé	Formule brute	Mw	EI(+)	EI(-)	Concentration de la solution mère	Concentration dans les QC (dilution 1/20)
Ethylmalonic acid	C5 H8 O4	132.04169	NI	131.03498	300 µg/mL	15 µg/mL
Aspartate <sup>15</sup> N	C4 H7 <sup>15</sup> N1 O4	134.03457	135.04182	133.02727	200 µg/mL	10 µg/mL
Arabinose <sup>13</sup> C	<sup>13</sup> C1 <sup>12</sup> C4 H10 O5	151.05618	174.04540 (Na+)	150.04890	300 µg/mL	15 µg/mL
Metformin	C4 H11 N5	129.10000	130.10872	NI	100 µg/mL	5 µg/mL
Glucose <sup>13</sup> C	<sup>13</sup> C1 <sup>12</sup> C5 H12 O6	181.06674	204.05596 (Na+)	180.05947	200 µg/mL	10 µg/mL
2-Aminoanthracene	C4 H11 N	193.09000	194.09643	192.08187	100 µg/mL	5 µg/mL
Amiloride	C6 H8 Cl N7 O	229.04846	230.05516	228.04061	100 µg/mL	5 µg/mL
Imipramine	C19 H24 N2	280.19452	281.20123	NI	50 µg/mL	2.5 µg/mL
Atropine sulfate	C17 H23 N O3	289.17000	290.17507	288.16052	100 µg/mL	5 µg/mL
Ampicilline	C16 H19 N3 O4 S	349.11020	350.11690	348.10235	100 µg/mL	5 µg/mL
Prednisone	C21 H26 O5	358.17830	359.18530	403.17482 (+HCOOH)	100 µg/mL	5 µg/mL
Colchicine	C22 H25 N O6	399.16876	400.17546	NI	100 µg/mL	5 µg/mL

NI = Non ionisé

## 2.4 ACQUISITION DES DONNEES

### 2.4.1 Appareillage

Les analyses ont été réalisées sur un spectromètre de masse de type LTQ-Orbitrap® Discovery de chez Thermo Fisher Scientific (Les Ulis, France). Ce spectromètre de masse est équipé une source électrospray et de deux analyseurs en série : un piège ionique linéaire (*linear trap quadrupole* LTQ) et un piège électrostatique à transformée de fourrier (Orbitrap). Les séparations chromatographiques ont été réalisées sur une chaîne UHPLC Accela couplée à ce spectromètre de masse et provenant également de chez Thermo Fisher Scientific (Les Ulis, France). Le logiciel d'interfaçage est XCalibur (version 2.1.) de Thermo Fisher Scientific (Les Ulis, France). Le spectromètre de masse a été étalonné avant chaque analyse en utilisant une solution d'étalonnage (Calimix) fournie par le fabricant (calibration externe).

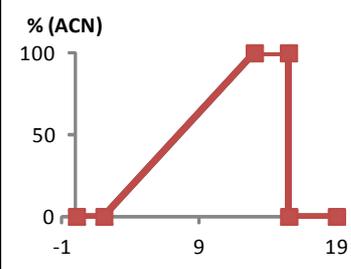
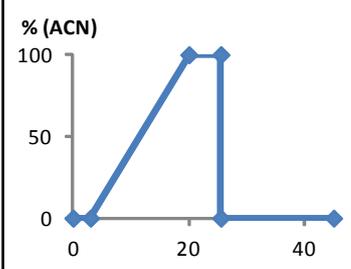
### 2.4.2 Analyses par spectrométrie de masse

Un volume de 50 µl d'échantillon a été injecté en mode FIA (de l'anglais « *flow injection analysis* » : injection directe dans le flux chromatographique) en utilisant une phase mobile constituée d'un mélange ACN/H<sub>2</sub>O (1/1 v/v) + 0.1% HCOOH à un débit de 100 µl/min. Les spectres de masse ont été enregistrés en mode continu et les spectres de MS/MS ont été acquis de manière concomitante (« *dependant scan* » : MS/MS de type CID sur l'ion le plus intense à 20 et 30% d'énergie de collision) entre 50 Th et 1000 Th avec un AGC (*Automatic gain control*) de  $5 \cdot 10^5$ . La résolution de l'appareil a été fixée à 30000 ( $m/\Delta m$ , FWHM à  $m/z$  400). En mode positif la tension électrique de la source électrospray a été fixée à 5 kV, celle du capillaire de transfert à 8 kV et le *tube lens offset* à 65 V. Les débits de gaz principaux et auxiliaire (azote) ont été optimisés à 35 et 8 unités arbitraire (*arbitrary units a.u.*) et la température a été fixée à 275°C. En mode négatif la tension électrique de la source électrospray a été fixée à -3.5 kV, celle du capillaire de transfert à -20 kV et le *tube lens offset* à -70 V. Les débits de gaz principaux et auxiliaire (azote) ont été optimisés à 45 et 15 unités arbitraire (*arbitrary units a.u.*) et la température a été fixée à 275°C.

### 2.4.3 Analyses par LC/MS

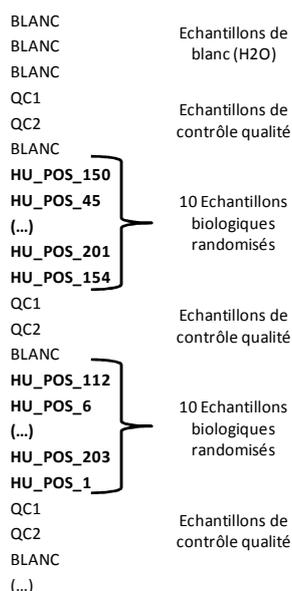
Deux type de séparations chromatographiques ont été utilisées : la première basée sur l'utilisation d'une colonne de type C<sub>18</sub> Hypersil Gold (2.1×150 mm, 1.9 µm), et la seconde impliquant une colonne de type PentaFluoroPhénylPropyl (PFPP) (Tableau 6). Un volume de 10 µl de chaque échantillon a été injecté sur le système chromatographique et a été élué dans les conditions de gradient définis dans le Tableau 6, avec des phases mobiles composées d'eau et d'acétonitrile contenant chacun 0,1% d'acide formique. Les acquisitions ont été effectuées dans les deux modes d'ionisation (négatif et positif) sur une gamme de rapports  $m/z$  allant de 75 Th à 1000 Th avec une résolution fixée à 30000 ( $m/\Delta m$ , FWHM à  $m/z$  400). Les conditions de sources sont les mêmes que pour l'analyse en FIA-MS. (voir paragraphe précédent)

**Tableau 6 : Méthodes chromatographiques utilisées pour l'analyse des échantillons.**

	Phase inverse	Phase mixte
Type de colonne	C <sub>18</sub>	PFPP
Dimensions	150 mm x 2.1 mm	250 mm x 2.1 mm
Diamètre des particules	1.9 µm	5 µm
références	hypersil GOLD	Discovery ® HS F5
Type de chromatographie	UHPLC	HPLC
Température	25°C	25°C
Débit de solvant	500 µl/min	200 µl/min
Durée d'analyse	19 minutes	45 minutes
Gradient		

### 2.4.4 Séquences d'analyse

Les échantillons biologiques ont été analysés de manière aléatoire dans le système LC/MS. Pour cela l'ordre des échantillons a été randomisé grâce au logiciel Excel. La séquence d'analyse commence toujours par des blancs (en général 3) afin de stabiliser le système d'analyse, et une première série d'échantillons de contrôle qualité. Les échantillons biologiques sont ensuite analysés 10 par 10, avec entre chaque, une série d'échantillons de contrôle qualité (Figure 22).



**Figure 22 : Exemple de composition d'une séquence d'analyse d'urines humaines (HU) en mode d'ionisation positif.**

Certains échantillons (ceux du début de séquence) ont été ré-analysés en fin de séquence. En mode d'ionisation positif, tous les échantillons ont été passés à la suite, ce qui représente 375 analyses au total, soit 5.5 jours d'analyse. En mode d'ionisation négatif les échantillons ont été passés en deux temps, avec un arrêt de l'appareil et un nettoyage de la source ESI entre les deux séquences d'analyses. La première séquence de 230 analyses a duré 3.2 jours, l'autre de 180 analyses a duré 2.5 jours.

## **2.5 TRAITEMENT DES DONNEES BIOLOGIQUES**

### **2.5.1 Conversion des fichiers de données en format universel**

Les données brutes (en format de fichier constructeur .raw) ont été converties en format universel netCDF avec le programme Xconvert disponible dans Xcalibur (version 2.1.) de Thermo Fisher Scientific (Les Ulis, France).

### **2.5.2 Déconvolution par XCMS**

Le traitement des données a été réalisé avec XCMS (Bioconductor version 2.8). C'est un *package* du logiciel R (version utilisée 2.11.1) qui effectue d'abord un filtrage des données pour éliminer le bruit de fond, puis une détection des pics par déconvolution suivie d'un alignement des pics détectés (voir Annexe 2). Les paramètres utilisés ainsi que leur signification, sont répertoriés en Annexe 2.

## **2.6 ANALYSE STATISTIQUE**

Les analyses statistiques multivariées (ACP, PLS...) ont été réalisées avec le logiciel SIMCA P11 (Umea, Suède).

## **2.7 IDENTIFICATION**

### **2.7.1 Analyse des corrélations des abondances ioniques dans le domaine spectral**

Le logiciel CAMERA (Bioconductor version 2.8) est un *package* gratuit du logiciel R (version utilisée 2.11.1). Il calcule la corrélation des abondances des ions d'un même pic chromatographique entre les différents spectres qui le composent. Il permet de regrouper les ions issus d'un même métabolite et de les annoter (isotopes, adduits et pertes les plus courants). Les paramètres utilisés pour ce travail ainsi que leur signification, sont répertoriés dans le Tableau 7.

Tableau 7 : Paramètres CAMERA utilisés

	Paramètre	Définition et choix du paramètre	UHPLC C18	HPLC PFPP
CAMERA	<b>sigma</b>	Multiplicateur appliqué à l'écart-type (SD) d'un pic pour déterminer sa largeur à mi-hauteur (en temps normal FWHM= SD*2.35)	6	6
	<b>perfwhm</b>	Pourcentage de la Largeur di pic à mi-hauteur = fenêtre de temps de rétention utilisée pour le calcul de la corrélation [RT_med+/- (FWHM*perfwhm)]	0.6	1
	<b>cor_eic_th</b>	Coefficient de corrélation minimum pour confirmer l'appartenance à un même groupe (métabolite.	0.6	0.6
	<b>maxcharge</b>	Charge maximum d'un ion	3	3
	<b>maxiso</b>	Nombre d'isotope maximum par ion	3	3
	<b>ppm</b>	Précision en ppm pour l'annotation des pertes et adduits	10	10
	<b>polarity</b>	Mode d'ionisation du jeu de donnée : positif ou négatif	positive	positive

### 2.7.2 Analyse des corrélations des abondances ioniques inter-échantillons

Un outil en langage R a été développé au laboratoire afin de pouvoir calculer la corrélation des abondances ioniques inter-échantillons. Il calcule un coefficient de corrélation à partir de l'abondance des ions au sein d'un même pic chromatographique pour les différents échantillons. Il permet de regrouper les ions issus d'un même métabolite et de les annoter (isotopes, adduits et pertes les plus courantes). Les paramètres utilisés sont la précision sur le temps de rétention (+/- 3 secondes), le mode d'ionisation utilisé pour l'acquisition du jeu de données et la précision en ppm pour l'annotation différentielle (1000 ppm).

### 2.7.3 Annotation par les bases de données publiques

Un outil en langage R a été développé au laboratoire afin de pouvoir réaliser une annotation de la liste de pics à partir des principales bases de données publiques : KEGG, HMDB et METLIN. Cet outil réalise une comparaison entre les masses des ions du jeu de données et celles contenues dans ces bases de données. Les paramètres utilisés sont le mode d'ionisation utilisé pour l'acquisition du jeu de données, la précision et le décalage en ppm pour l'annotation (en fonction de la précision du jeu de donnée à annoter, dans notre cas décalage 0 ppm et précision 5 ppm).

### 2.7.4 Annotation par la base de données spectrale

Un outil en informatique a été développé au laboratoire afin de pouvoir réaliser une annotation de la liste de pics à partir des informations stockées dans la base de données spectrale. Cet outil réalise une comparaison entre les masses des ions du jeu de données et celles contenues la base spectrale ainsi qu'une comparaison avec les temps de rétention enregistrés pour chaque méthode chromatographique. Les paramètres utilisés sont le mode d'ionisation utilisé pour l'acquisition du jeu de données, la précision en ppm pour l'annotation (en fonction de la précision du jeu de donnée à annoter, dans notre cas 5 ppm) et le décalage en temps de rétention (en fonction du celui du jeu de donnée) modélisé par une fonction exponentielle de type  $RT = a \pm RT^b$  (RT : temps de rétention chromatographique, a et b : constantes. Dans notre cas a = 2 et b = 0.5).

### 2.7.5 Analyse structurale par expériences de fragmentations séquentielles

Les analyses MS/MS et MS<sup>n</sup> ont été réalisées en mode d'acquisition concomittant lors de l'analyse LC/MS ou FIA/MS. Ce type d'acquisition de données permet de réaliser séquentiellement des analyses MS et CID, soit sur l'ion le plus intense (pour l'analyse de composés de référence en FIA/MS) soit sur une liste d'ions pré-sélectionnés (pour les analyses d'échantillons biologiques en LC/MS). L'utilisation d'une liste d'ions sélectionnés permet, via l'utilisation de l'option « *Global mass lists* » (Tableau 8), de sélectionner un ion, à un temps de rétention donné, et de fixer une énergie de collision pour l'acquisition des données de CID. On peut ainsi réaliser plusieurs expériences de fragmentation en une seule analyse, à partir du moment où les temps de rétention ne se chevauchent pas.

Tableau 8 : Paramètres utilisés lors des analyses MS/MS mode d'acquisition dépendant

GLOBAL	
<b>Global</b>	Range (m/z): 50.00-1000.00 for both selecting MS et MS <sup>n</sup> data dependent masses
<b>Mass Widths</b>	By mass, Low: 0.50, High: 0.50
<b>Scan Widths</b>	By mass, Low: 0.50, High: 0.50
<b>Dynamic Exclusion</b>	Not enabled
<b>Mass tags</b>	Not enabled
<b>Isotopic data dependence</b>	Not enabled
<b>Analog</b>	Not enabled
<b>Neutral loss</b>	Neutral loss mass sort mode: by intensity, Neutral loss mass width: By mass, Low: 0.50, High: 0.50
<b>Product</b>	Product loss mass sort mode: by intensity, Product loss mass width: By mass, Low: 0.50, High: 0.50
SEGMENT	
<b>Current segment</b>	Enable preview mode for FTMS master scans
<b>Chromatography</b>	Not enabled
<b>Parent Mass List</b>	Use global mass lists (uniquement pour des ions présélectionnés en LC/MS) sinon NA
<b>Reject Mass List</b>	Use global mass lists
<b>Charge state</b>	Enable charge state screening
<b>Neutral Loss</b>	NA
<b>Product Mass List</b>	NA
<b>Add/Sub</b>	Not enabled
SCAN EVENT	
<b>Current Scan Event</b>	Minimum signal threshold (counts): 500, Mass determined from scan event: 1, Nth most intense from list: 1 (pour des ions présélectionnés en LC/MS) Nth most intense ion: 1 (pour la sélection automatique de l'ion le plus intense en FIA/MS)
<b>Activation</b>	Activation type: CID, Default charge state: 1, Isolation width (m/z): 1.0, Normalized collision energy: 20 ou 30, activation Q: 0.250, Activation time (ms): 30
<b>FTHCD</b>	NA
<b>Mass Tags</b>	Neither partner

## 3 RESULTATS

### 3.1 IDENTIFICATIONS A PARTIR DE COMPOSES DE REFERENCE

Comme mentionné précédemment, l'ionisation électrospray génère une redondance du signal conséquente qu'il faut pourvoir rapidement mettre en évidence lors de l'analyse d'un jeu de données biologiques. L'objectif de l'analyse de composés de référence est de déterminer, à partir de l'analyse MS et MS/MS d'un composé, l'ensemble des ions produits pour une même molécule dans un système analytique donné. Une fois identifiés, ces ions sont importés dans la base de données spectrale pour pouvoir annoter les ions présents dans les listes d'ions issues de données biologiques.

#### 3.1.1 Interprétation des spectres de masse des composés de référence

Les composés de référence sont soumis à plusieurs types d'analyses. D'abord en FIA/MS afin d'annoter les ions produits en MS (espèce moléculaire, isotopes, adduits et fragments), puis en MS/MS pour aider à l'interprétation des ions fragments observés en MS. Enfin les composés sont analysés en LC/MS pour déterminer les temps de rétention dans les différents systèmes chromatographiques utilisés. A partir de la formule brute de la molécule analysée, les masses théoriques des ions sont calculées et comparées à celles extraites à partir du spectre de masse. Ces ions théoriques sont obtenus à partir d'une liste d'ions déjà observés dans notre système chromatographique : isotopes ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^{18}\text{O}$ ...) adduits ( $\text{HCOOH}$ ,  $\text{Na}$ ,  $\text{K}$ ...) pertes ( $\text{H}_2\text{O}$ ,  $\text{NH}_3$ ...) Une comparaison entre les  $m/z$  théoriques et expérimentaux est réalisée pour filtrer la liste des ions obtenus. Ce filtrage de la liste d'ions issus des spectres de masse permet d'obtenir, pour chaque molécule, une liste d'ions spécifiques. A ces informations sont ajoutées celles concernant la rétention dans divers systèmes chromatographiques (Figure 23).

Acide Pantothénique ESI(+)								Temps de rétention							
m/z (FIA)	Abundance	Relative	m/z (theoretical)	$\delta$ (ppm)	RDBE	Elemental Formula	Generated Ions	UHPLC (C18)	UHPLC (C8)	HPLC (PFPP)					
72.04431	7.7E+04	0.82	72.04439	-1.11	1.5	C3 H6 O N	[(M+H)-(C6H10O3)-(H2O)]	4.77	1.80	11.17					
90.05482	3.6E+05	3.78	90.05496	-1.55	0.5	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	4.77	1.80	11.17					
116.03400	2.7E+04	0.29	116.03422	-0.22	2.5	C4 H6 O3 N	[(M+H)-(C5H12O2)]+	4.77	1.80	11.17					
<b>Fragments et leurs isotopes</b>								626	-2.85	2.5	C7 H12 O2 N	[(M+H)-(C2H6O3)]+	4.77	1.80	11.17
								247	-0.30	1.5	C8 H16 O3 N	[(M+H)-(HCOOH)]+	4.77	1.80	11.17
								382	0.26	3.5	C9 H14 O3 N	[(M+H)-2(H2O)]+	4.77	1.80	11.17
								738	-0.46	2.5	C9 H16 O4 N	[(M+H)-(H2O)]+	4.77	1.80	11.17
<b>Pseudomoléculaire et ses isotopes</b>								795	0.18	1.5	C9 H18 N O5	[(M+H)]+	4.77	1.80	11.17
								131	0.20	1.5	C8 13C H18 N O5	[(M+H)]+ (13C)	4.77	1.80	11.17
								219	-2.40	1.5	C9 H18 N O4 18O	[(M+H)]+ (18O)	4.77	1.80	11.17
								555	2.20	1.5	C8 13C H18 N O4 18O	[(M+H)]+ (13C+18O)	4.77	1.80	11.17
238.12923	1.76E+03	0.02	238.12852	3.00	0.5	C9 H20 N O6	[(M+H)+(H2O)]+	4.77	1.80	11.17					
242.09992	6.61E+06	68.15	242.09990	0.10	1.5	C9 H17 N Na O5	[(M+Na)]+	4.77	1.80	11.17					
243.10331	4.44E+05	4.58	243.10325	0.25	1.5	C8 13C H17 N Na O5	[(M+Na)]+ (13C)	4.77	1.80	11.17					
244.10363	1.79E+04	0.18	244.10414	-2.09	1.5	C9 H17 N Na O4 18O	[(M+Na)]+ (18O)	4.77	1.80	11.17					
258.07386	1.29E+05	1.33	258.07383	0.10	1.5	C9 H17 K N O5	[(M+K)]+	4.77	1.80	11.17					
259.07731	1.49E+03	0.02	259.07719	0.47	1.5	C8 13C H17 K N O5	[(M+K)]+ (13C)	4.77	1.80	11.17					
264.08200	1.68E+05	1.74	264.08184	0.61	1.5	C9 H16 N Na2 O5	[(M-H+2Na)]+	4.77	1.80	11.17					
265.08498	3.08E+03	0.03	265.08520	-0.81	1.5	C8 13C H16 N Na2 O5	[(M-H+2Na)]+ (13C)	4.77	1.80	11.17					
310.08654	1.10E+03	0.01	310.08732	-2.52	1.5	C10 H18 N Na2 O7	[(M+Na)+(HCOONa)]+	4.77	1.80	11.17					
392	0.99	2.5	392	1.22	2.5	C18 H34 N2 Na O10	[(2M+Na)]+	4.77	1.80	11.17					
728	0.26	2.5	728	0.99	2.5	C17 13C H34 N2 Na	[(2M+Na)]+ (13C)	4.77	1.80	11.17					
728	0.26	2.5	728	0.26	2.5	C16 13C2 H34 N2 Na	[(2M+Na)]+ (13C2)	4.77	1.80	11.17					
251	0.10	2.5	251	0.10	2.5	C18 H33 N2 Na2 O11	[(2M-H+2Na)]+	4.77	1.80	11.17					

Figure 23 : Liste des ions obtenus pour l'acide pantothénique en mode d'ionisation positif.

### 3.1.2 Annotation d'un jeu de données biologiques

La liste des ions obtenus à partir des données biologiques est comparée aux ions présents dans la base de données spectrale. Ainsi, quand le temps de rétention et la masse précise correspondent, la molécule correspondante est annotée. Par exemple, l'acide pantothénique a été annoté dans les urines humaines (Figure 24). Une comparaison à partir de la liste des ions (et temps de rétention) obtenus avec le composé de référence en FIA (Figure 24, tableau blanc), a permis de retrouver, après filtrage des données, un certain nombre d'ions attribuables à l'acide pantothénique dans les urines humaines (Figure 24, tableau jaune). On retrouve l'espèce moléculaire et ses isotopes  $^{13}\text{C}$  et  $^{18}\text{O}$ , les adduits Na et K, une perte d' $\text{H}_2\text{O}$  et une autre perte spécifique, le tout au même temps de rétention que le composé de référence.

#### Acide Pantothénique ESI(+)

m/z (FIA)	Abundance	Relative	m/z (theoretical)	$\delta$ (ppm)	RDBE	Elemental Formula	Generated Ions	RETENTION TIMES (MIN)		
								UHPLC (C18)	UHPLC (C8)	HPLC (PFPP)
72.04431	7.7E+04	0.82	72.04439	-1.11	1.5	C3 H6 O N	[(M+H)-(C6H10O3)-(H2O)]+	4.77	1.80	11.17
90.05482	3.6E+05	3.78	90.05496	-1.55	0.5	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	4.77	1.80	11.17
116.03400	2.7E+04	0.29	116.03422	-0.22	2.5	C4 H6 O3 N	[(M+H)-(C5H12O2)]+	4.77	1.80	11.17
142.08585	1.0E+03	0.01	142.08626	-2.85	2.5	C7 H12 O2 N	[(M+H)-(C2H6O3)]+	4.77	1.80	11.17
174.11217	1.6E+04	0.17	174.11247	-0.30	1.5	C8 H16 O3 N	[(M+H)-(HCOOH)]+	4.77	1.80	11.17
184.09687	2.4E+03	0.03	184.09682	0.26	3.5	C9 H14 O3 N	[(M+H)-2(H2O)]+	4.77	1.80	11.17
202.10729	1.5E+04	0.16	202.10738	-0.46	2.5	C9 H16 O4 N	[(M+H)-(H2O)]+	4.77	1.80	11.17
<b>220.11799</b>	<b>9.69E+06</b>	<b>100</b>	<b>220.11795</b>	<b>0.18</b>	<b>1.5</b>	<b>C9 H18 N O5</b>	<b>[(M+H)]+</b>	4.77	1.80	11.17
221.12135	6.84E+05	7.06	221.12131	0.20	1.5	C8 13C H18 N O5	[(M+H)]+ (13C)	4.77	1.80	11.17
222.12166	4.59E+04	0.47	222.12219	-2.40	1.5	C9 H18 N O4 18O	[(M+H)]+ (18O)	4.77	1.80	11.17
223.12604	2.26E+03	0.02	223.12555	2.20	1.5	C8 13C H18 N O4 18O	[(M+H)]+ (13C+18O)	4.77	1.80	11.17
238.12923	1.76E+03	0.02	238.12852	3.00	0.5	C9 H20 N O6	[(M+H)+(H2O)]+	4.77	1.80	11.17
242.09992	6.61E+06	68.15	242.09990	0.10	1.5	C9 H17 N Na O5	[(M+Na)]+	4.77	1.80	11.17
243.10331	4.44E+05	4.58	243.10325	0.25	1.5	C8 13C H17 N Na O5	[(M+Na)]+ (13C)	4.77	1.80	11.17
244.10363	1.79E+04	0.18	244.10414	-2.09	1.5	C9 H17 N Na O4 18O	[(M+Na)]+ (18O)	4.77	1.80	11.17
258.07386	1.29E+05	1.33	258.07383	0.10	1.5	C9 H17 K N O5	[(M+K)]+	4.77	1.80	11.17
259.07731	1.49E+03	0.02	259.07719	0.47	1.5	C8 13C H17 K N O5	[(M+K)]+ (13C)	4.77	1.80	11.17
264.08200	1.68E+05	1.74	264.08184	0.61	1.5	C9 H16 N Na2 O5	[(M-H+2Na)]+	4.77	1.80	11.17
265.08498	3.08E+03	0.03	265.08520	-0.81	1.5	C8 13C H16 N Na2 O5	[(M-H+2Na)]+ (13C)	4.77	1.80	11.17
310.08654	1.10E+03	0.01	310.08732	-2.52	1.5	C10 H18 N Na2 O7	[(M+Na)+(HCOONa)]+	4.77	1.80	11.17
461.21113	6.45E+05	6.65	461.21057	1.22	2.5	C18 H34 N2 Na O10	[(2M+Na)]+	4.77	1.80	11.17
462.21438	1.08E+05	1.11	462.21392	0.99	2.5	C17 13C H34 N2 Na	[(2M+Na)]+ (13C)	4.77	1.80	11.17
463.21740	9.76E+03	0.1	463.21728	0.26	2.5	C16 13C2 H34 N2 N	[(2M+Na)]+ (13C2)	4.77	1.80	11.17
483.19256	4.21E+04	0.43	483.19251	0.10	2.5	C18 H33 N2 Na2 O1	[(2M-H+2Na)]+	4.77	1.80	11.17

m/z (experimental)	Abundance	Relative	m/z (theoretical)	$\delta$ (ppm)	Elemental Formula	Generated Ions	UHPLC (C18)	RT (min) (experimental)
90.0549	8.1E+06	1.65	90.0550	-0.50	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	4.77	4.74
202.1076	1.8E+06	0.36	202.1074	1.21	C9 H16 O4 N	[(M+H)-(H2O)]+	4.77	4.74
<b>220.1175</b>	<b>4.9E+08</b>	<b>100</b>	<b>220.1180</b>	<b>-1.85</b>	<b>C9 H18 N O5</b>	<b>[(M+H)]+</b>	<b>4.77</b>	<b>4.74</b>
221.1217	2.4E+07	4.92	221.1213	1.70	C8 13C H18 N O5	[(M+H)]+ (13C)	4.77	4.74
222.1221	1.5E+06	0.31	222.1222	-0.48	C9 H18 N O4 18O	[(M+H)]+ (18O)	4.77	4.74
242.1006	2.1E+06	0.43	242.0999	3.06	C9 H17 N Na O5	[(M+Na)]+	4.77	4.74
258.0753	4.9E+05	0.10	258.0738	5.74	C9 H17 K N O5	[(M+K)]+	4.77	4.74
259.0753	3.6E+04	0.01	259.0772	-7.47	C8 13C H17 K N O5	[(M+K)]+ (13C)	4.77	4.74

Urines Humaines

Figure 24 : Annotation de l'acide pantothénique à partir du composé de référence dans un jeu de données biologiques d'urines humaines.

L'annotation de ces listes des ions issus d'expériences biologiques implique une comparaison des temps de rétention avec ceux obtenus pour les composés de référence. Les échantillons biologiques ont été analysés avec une colonne de type C<sub>18</sub>, et certains d'entre eux avec une colonne de type PFPP (PentaFluoroPhenylPropyl, phase mixte). Nous avons pu ainsi observer que pour certaines molécules non retenues avec le système C<sub>18</sub>, plusieurs pics chromatographiques pouvaient être observés avec la colonne de type PFPP. Par exemple, la N-acétylputrescine est détectée dans les urines à un temps de rétention de 1.1 minutes avec une colonne de type C<sub>18</sub>, alors que deux pics

chromatographiques sont observés avec une colonne de type PFPP. Le premier pic correspond bien à la N-acétylputrescine dans la mesure où le temps de rétention correspond à celui du composé de référence, mais le deuxième pic correspond à un (ou peut-être plusieurs ?) isomères (Figure 25). Ce résultat souligne la complexité du métabolome et démontre qu'il peut être délicat d'identifier un métabolite avec un seul système chromatographique. De plus, l'existence de plusieurs molécules isomères non résolues chromatographiquement rend délicates les interprétations statistiques ultérieures. En effet, les variations significatives observées ne pourront pas être attribuées à l'une ou l'autre des molécules, mais à la somme des contributions de chacune des molécules. Inversement, une significativité biologique peut aussi être masquée si les contributions de chacun des métabolites sont opposées (c'est-à-dire que la concentration de l'un augmente tandis que l'autre diminue). Ce manque de résolution concerne 38 pics chromatographiques en C<sub>18</sub>, qui correspondent à 83 pics chromatographiques (et donc à au moins 83 métabolites) avec une colonne de type en PFPP. Sur les 38 pics obtenus avec la colonne C<sub>18</sub>, 37 sont élués dans le temps mort. Etant donné qu'une seule méthode chromatographique ne suffit pas pour séparer l'ensemble des métabolites présents, il est donc judicieux d'utiliser deux méthodes chromatographiques complémentaires pour identifier avec certitude un maximum de métabolites.

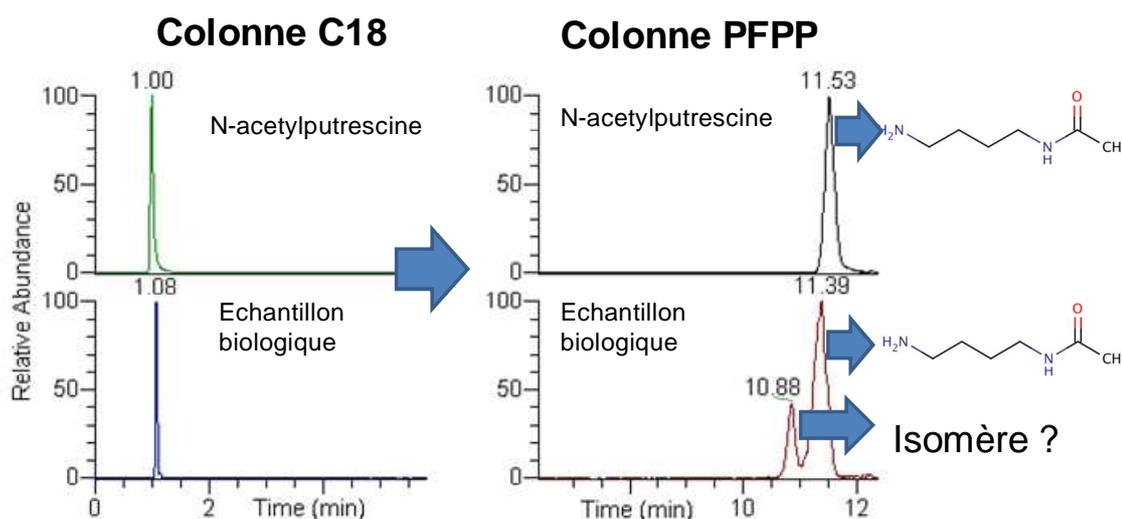


Figure 25 : Mise en évidence du manque de résolution en C18

### 3.1.3 Bilan des métabolites identifiés dans les urines humaines.

L'annotation du jeu de données biologique d'urines humaines à partir de la base de données spectrale de composés de référence a permis d'identifier 163 métabolites (voir annexe 3). L'utilisation de deux systèmes chromatographiques différents a permis de confirmer ces identifications. Il est intéressant de noter que sur ces 163 métabolites, 20 ne sont pas référencés dans les urines selon HMDB.

## 3.2 IDENTIFICATIONS A PARTIR D'UN JEU DE DONNEES BIOLOGIQUES

Bien que la comparaison à des composés de référence soit la méthode d'identification la plus rigoureuse, l'indisponibilité commerciale d'un grand nombre de métabolite nous a conduit à réaliser l'identification de certaines molécules directement à partir des données biologiques. L'identification de métabolites à partir de données biologiques peut être réalisée sur la base d'une mesure précise de la masse d'un ion qui permet d'obtenir sa formule brute, d'expériences de  $MS^n$ , voire même de la similitude de temps de rétention avec une molécule de structure proche (et ce même lorsque les composés de référence ne sont pas disponibles). Les ions correspondant à ces métabolites putativement identifiés peuvent enrichir la base de données spectrale, en complément des composés de référence.

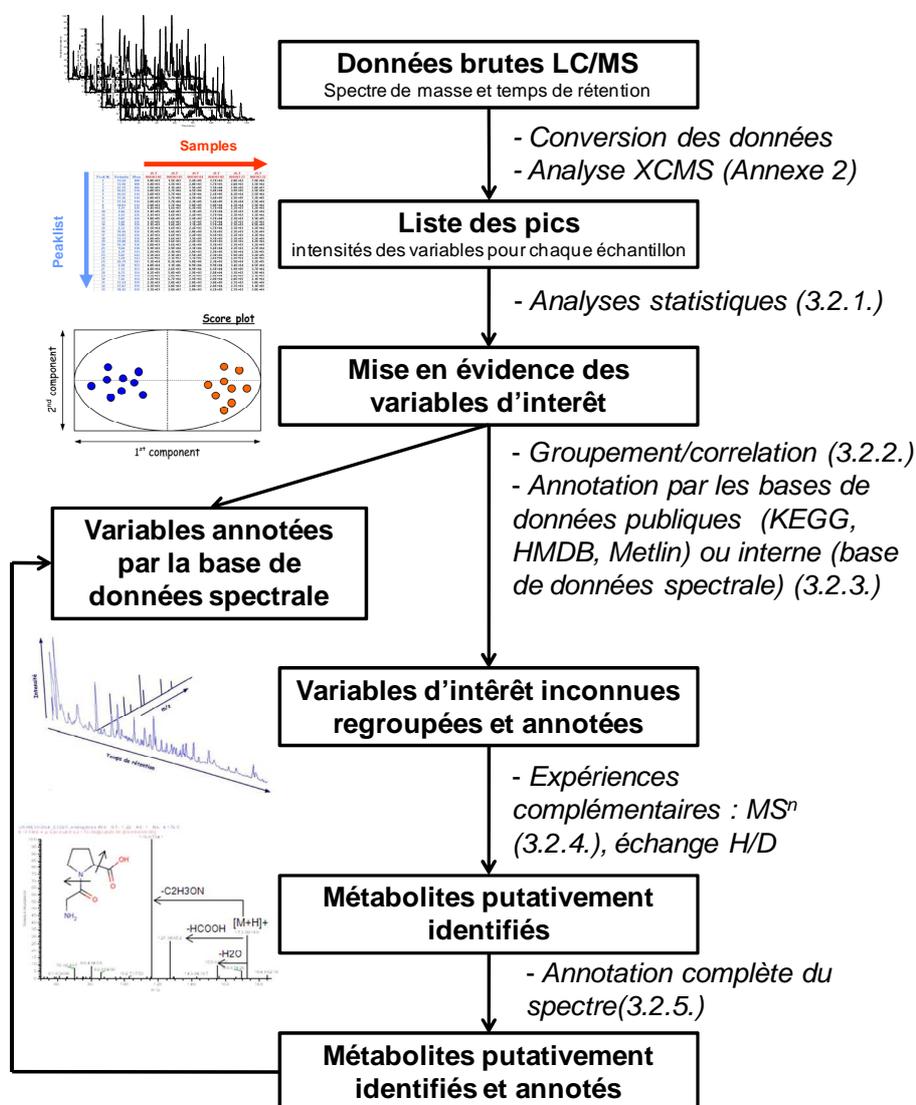


Figure 26 : Différentes étapes pour l'identification de variables inconnues à partir de données biologiques

(le numéro des chapitres correspondants est donné entre parenthèse à chaque étape).

Le processus d'identification à partir de données biologiques commence par l'analyse des échantillons et le traitement des données brutes afin d'obtenir la liste des pics (ou des variables, correspondant à des couples  $m/z$ , temps de rétention) et de leur abondances dans chaque échantillons biologiques (Figure 26). Dans le cas de notre cohorte issue du suivi médical de salariés du CEA-Saclay, les facteurs physiologiques (âge, genre et IMC) vont être utilisés pour structurer le jeu de données grâce aux analyses statistiques et ainsi mettre en évidence les variables d'intérêt. Les étapes suivantes d'identification se focalisent sur ces variables d'intérêt inconnues (qui ne sont pas déjà annotées par la base de données spectrale). Certaines d'entre elles qui sont issues d'un même métabolite, vont pouvoir être regroupées en utilisant les corrélations des abondances ioniques dans le domaine spectral et entre les différents échantillons. Elles seront ensuite annotées par les bases de données publiques, et les hypothèses d'identifications seront confirmées ou infirmées par des expériences de fragmentation. Enfin, une fois le métabolite putativement identifié, le spectre sera annoté en intégralité à partir des données brutes et non pas de la liste des pics obtenue avec XCMS, et la liste des ions obtenus sera importée dans la base de données spectrale. En effet certains ions peu intenses ne sont pas extraits par XCMS mais sont bien présents sur le spectre initial.

### 3.2.1 Analyses statistiques

Les données ont été dans un premier temps analysées à l'aide d'une l'analyse en composantes principales (ACP) afin de visualiser la répartition spontanée des échantillons dans un espace à deux dimensions. Dans un second temps, des analyses statistiques supervisées telles que la régression PLS ou la PLS-DA (analyse discriminante PLS, de l'anglais «*Partial Least Square -Discriminant Analysis*») ont été appliquées au jeu de données afin d'affiner les résultats et de déterminer quelles variables contribuent à la différenciation entre les groupes pour chaque facteurs (genre, âge, IMC).

La carte factorielle du poids des individus, encore appelée «*score plot*» et issue de l'ACP est représentée en Figure 27. Elle regroupe à la fois les chromatogrammes obtenus à partir des échantillons urinaires, mais également des échantillons de contrôle qualité, QC1 et QC2 (voir la section matériel et méthode). L'analyse de la carte factorielle montre que les échantillons QC1 et les échantillons biologiques se superposent, alors que les échantillons QC2 sont clairement individualisés, en raison de l'ajout d'étalons internes (Figure 27). Si, dans un deuxième temps, nous ne nous focalisons que sur les échantillons biologiques, nous remarquons alors que le jeu de données s'organise spontanément selon l'ordre de passage et non pas en fonction des facteurs physiologiques (Figure 28). Ceci signifie que la variabilité analytique de notre jeu de données biologiques est très importante, et que pour mettre en évidence plus facilement des variations biologiques, il va falloir utiliser des analyses supervisées telles que la PLS. En effet, si on garde la même répartition des échantillons (ou individus) et que l'on visualise les hommes (bleu) par rapport aux femmes (rouge) (Figure 29), la séparation entre les deux groupes n'est pas bonne, contrairement à celle que l'on peut obtenir par PLS (Figure 30).

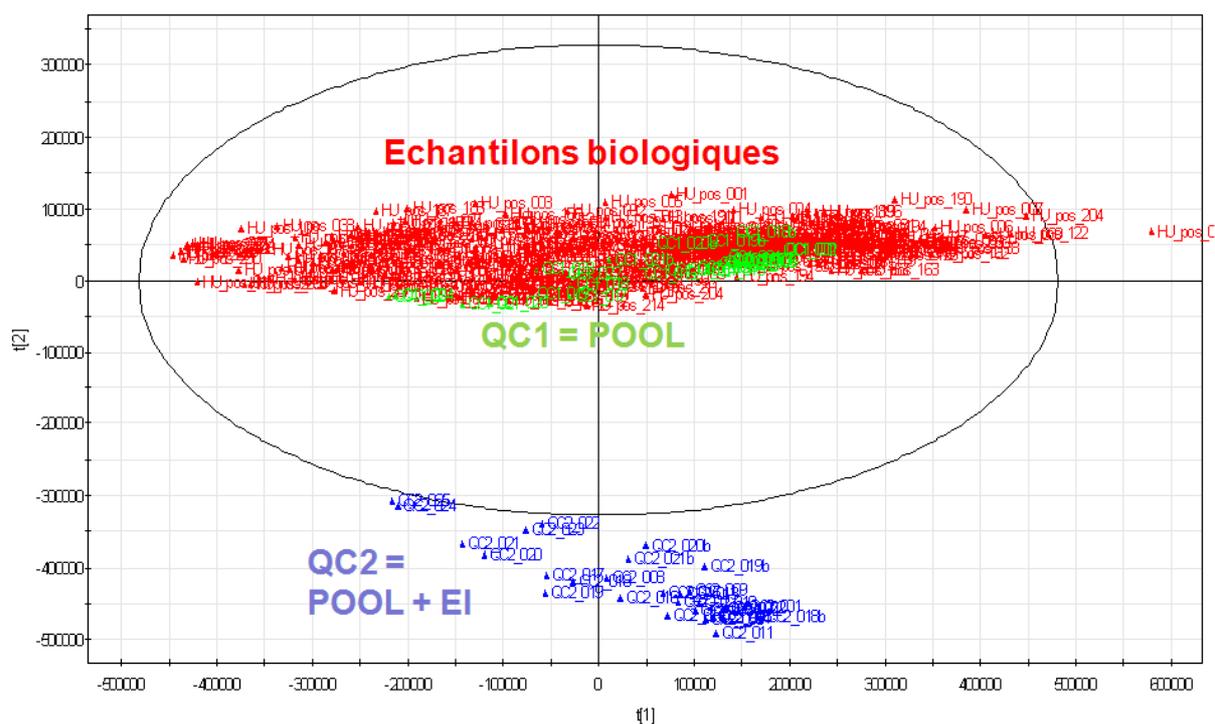


Figure 27 : Score plot de l'ACP (échantillons biologiques + QC1 + QC2) en mode d'ionisation positif. Visualisation des données selon la nature des échantillons.

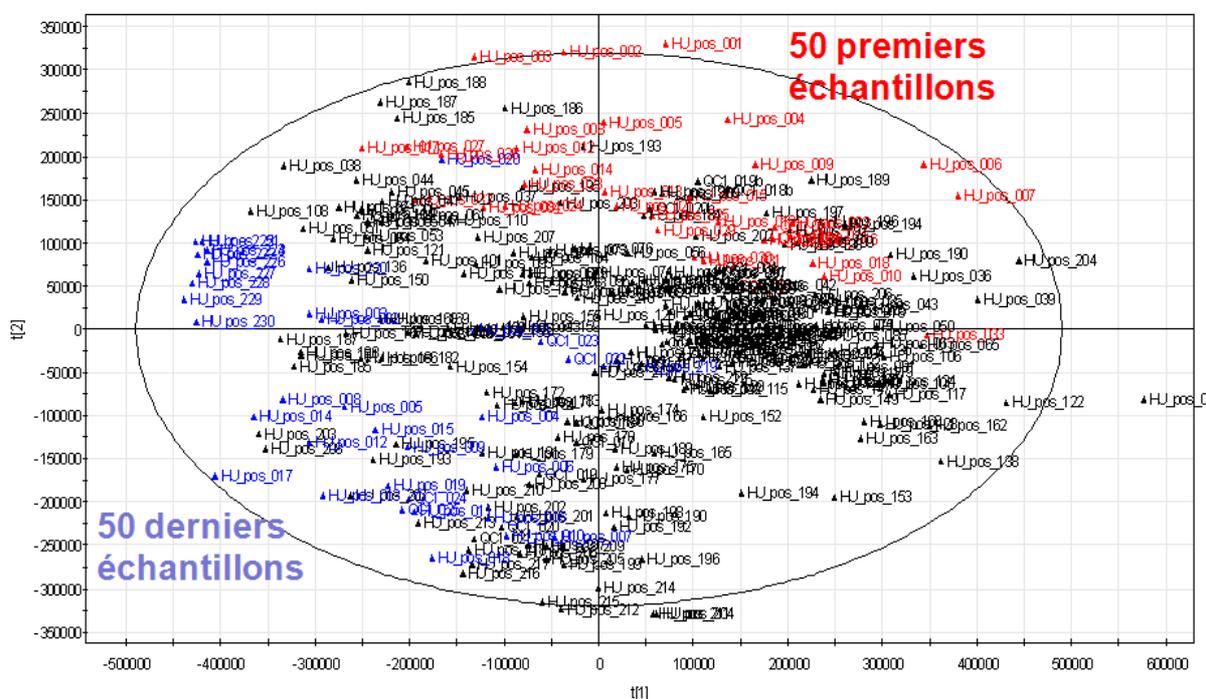


Figure 28 : Score plot de l'ACP (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction de l'ordre de passage.

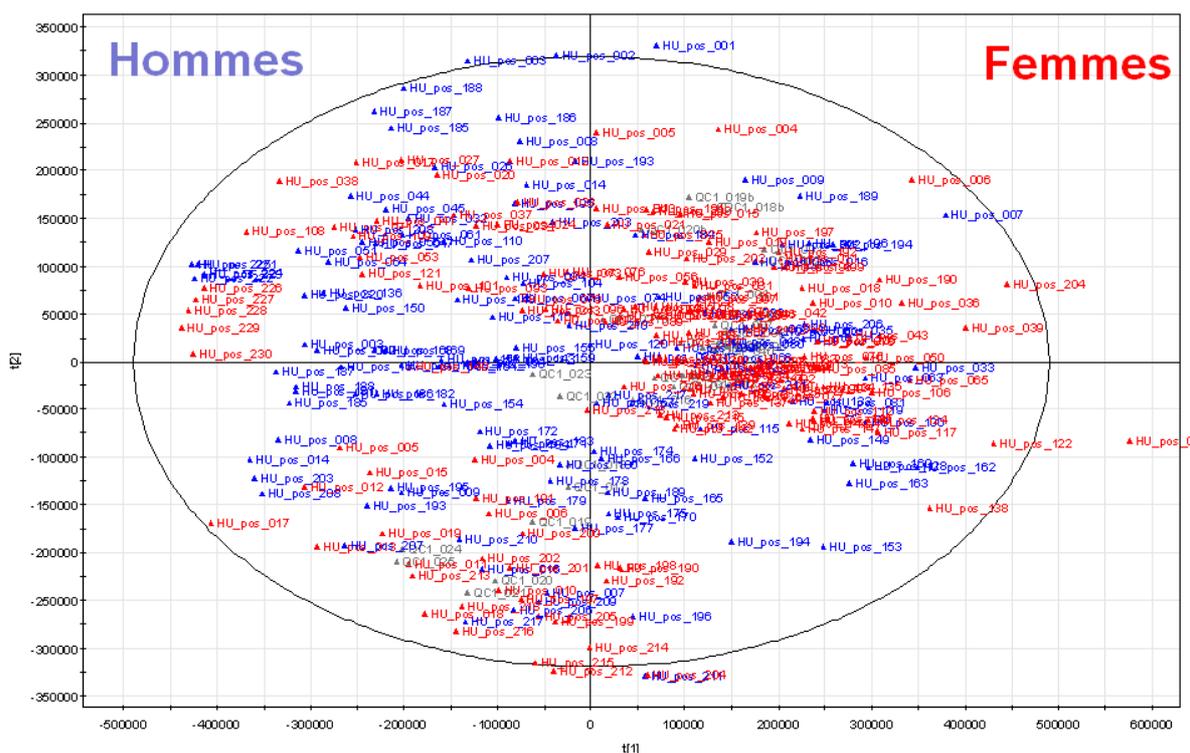


Figure 29 : Score plot de l'ACP (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction du genre.

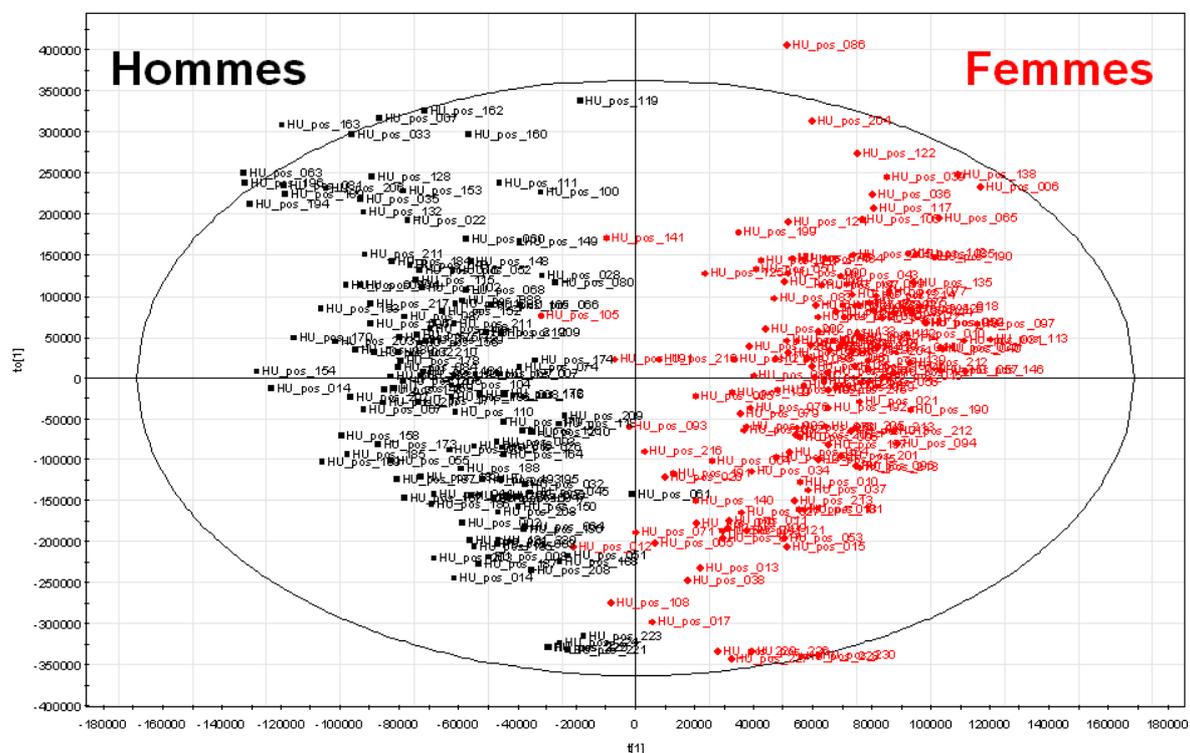


Figure 30 : Score plot de la PLS-DA (échantillons biologiques) en mode d'ionisation positif. Visualisation des données en fonction du genre.

C'est à partir de ce dernier type de modèle statistique explicatif que l'on va extraire les variables discriminantes pour un facteur donné. La carte factorielle du poids des variables, encore appelée «*Loading plot*», montre l'importance des variables pour chacun des facteurs. Par exemple, les métabolites les plus discriminants entre hommes et femmes sont les plus excentrés (Figure 31) comme c'est par exemple le cas pour la créatine, la L-leucine ou le tryptophane (identifiés par la base de données spectrale). Ce sont les signaux discriminants mais non encore caractérisés que nous allons maintenant chercher à identifier.

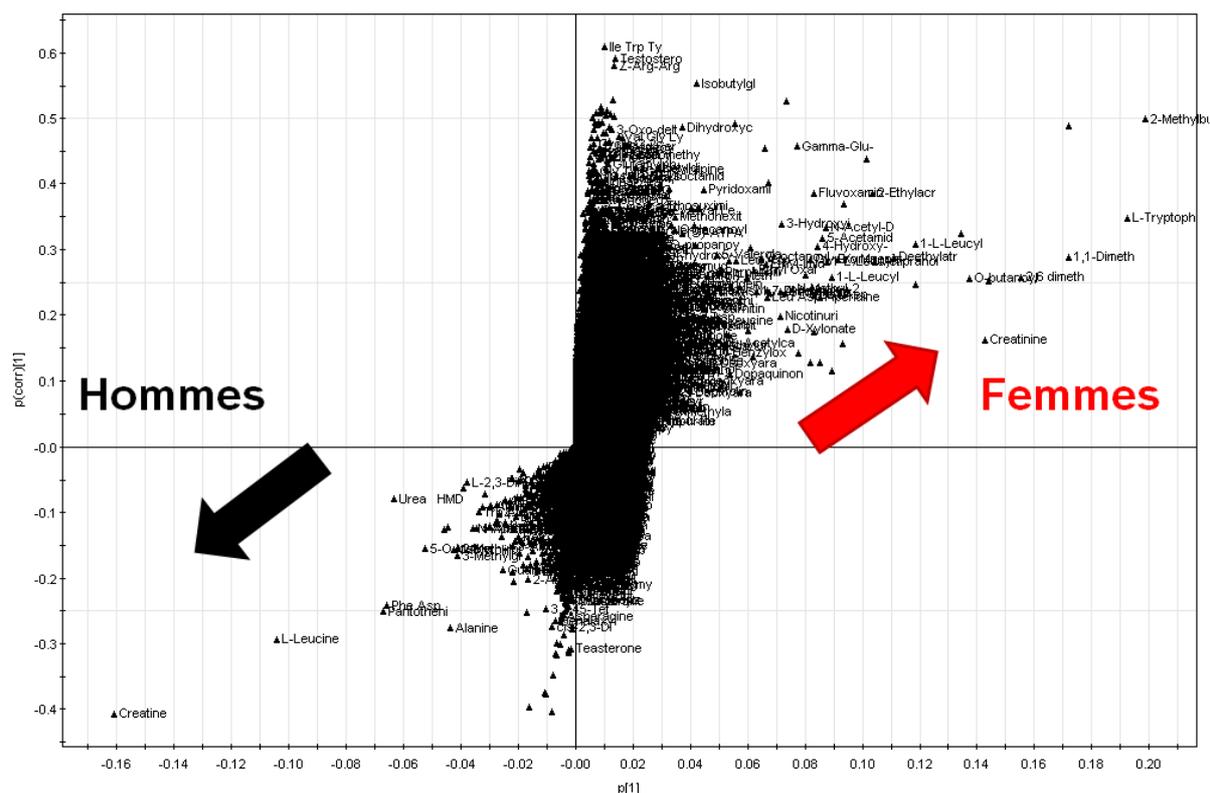


Figure 31 : Loading plot de la PLS-DA (échantillons biologiques) en mode d'ionisation positif. Visualisation des variables en fonction du genre

### 3.2.2 Regroupement des variables d'intérêt

L'objectif de cette étape est de regrouper les signaux correspondant à des ions issus d'un même métabolite. Certains de ces ions sont aisément identifiables par annotation différentielle lorsqu'on analyse un composé seul en solution. Il peut par contre s'avérer délicat de différencier une espèce moléculaire d'un adduit ou d'un isotope au sein d'un jeu de données biologiques complexe. Afin d'éviter les identifications erronées (Figure 20), il convient donc de regrouper les signaux issus d'un même métabolite, puis d'isoler l'espèce moléculaire pour ensuite d'identifier le métabolite d'origine.

Les signaux provenant d'un même métabolite ont le même profil chromatographique. En effet les ions sont produits à partir d'un métabolite qui a un profil d'éluion propre, ont le même profil chromatographique, à savoir celui du métabolite. Par exemple sur la Figure 32, les métabolites jaune et rouge co-éluent, mais en superposant les profils des ions issus de ces deux métabolites (ou en calculant le coefficient de corrélation qui reflète la similarité des profils), on arrive à distinguer les deux profils d'éluion et donc à savoir de quel métabolite est issu chaque ion. De plus les signaux provenant d'un même métabolite ont les mêmes évolutions en d'abondances relatives d'un échantillon à l'autre : si l'abondance de l'espèce moléculaire augmente, celles de ses isotopes augmentent aussi. Ainsi, les ratios entre les abondances des différents ions détectés sont théoriquement identiques d'un échantillon à l'autre. Cela signifie que les coefficients de corrélation entre les abondances de paires d'ions provenant d'une même molécule devraient théoriquement être égaux à 1, aussi bien pour l'ensemble des spectres qui constituent un pic chromatographique au sein d'un même échantillon (même profil chromatographique) qu'entre différents échantillons pour lesquels les ions sont observés. En raison de l'effet matrice et de la suppression d'ionisation, la valeur théorique de 1 est rarement atteinte et la corrélation entre les signaux est supposée être supérieure 0,8.

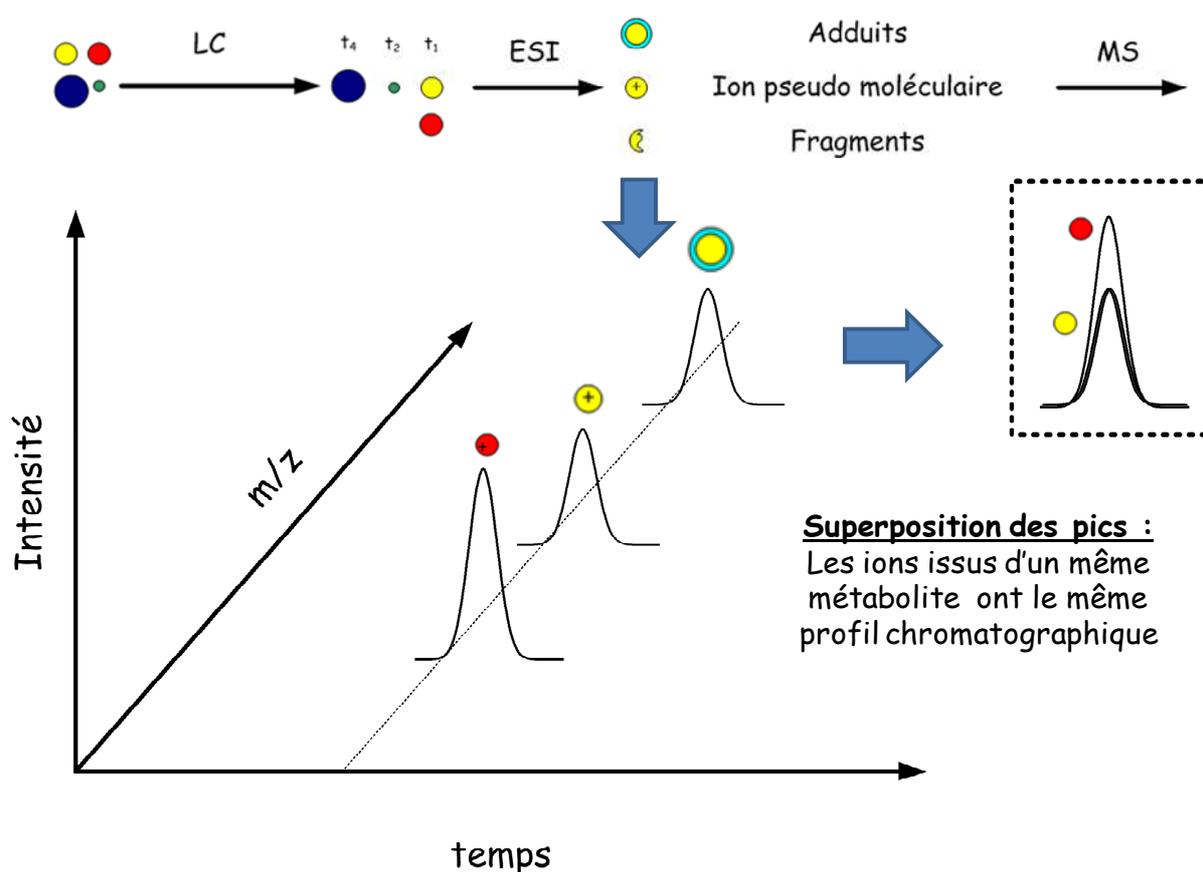


Figure 32 : Corrélation dans le domaine chromatographique : superposition des profils d'éluion d'ions issus de molécules éluant au même temps de rétention

- Corrélation dans le domaine de chromatographique : utilisation de CAMERA [259]. CAMERA (*Collection of Algorithms for MEtabolite pRofile Annotation*) est un *package* R pour l'annotation des données LC/MS qui interagit directement avec les données obtenues par XCMS (en particulier les

EIBPC voir annexe 2). Il utilise un algorithme qui regroupe les ions provenant d'une même molécule en comparant la forme des pics à un temps de rétention donné. Les adduits, isotopes et fragments d'un même métabolite présentent en effet le même profil d'élution (Figure 33). CAMERA utilise la régression linéaire pour estimer le degré d'homologie entre chaque spectre d'un pic chromatographique donné. Notons que pour les signaux présentant une faible intensité, et qui sont fortement impactés par le bruit de fond, les chromatogrammes sont souvent plat et/ou irréguliers, ce qui entraîne une faible corrélation. CAMERA réalise ensuite une annotation en fonction des différences de masse entre les différents ions pour permettre d'attribuer plus facilement les ions (Figure 33)

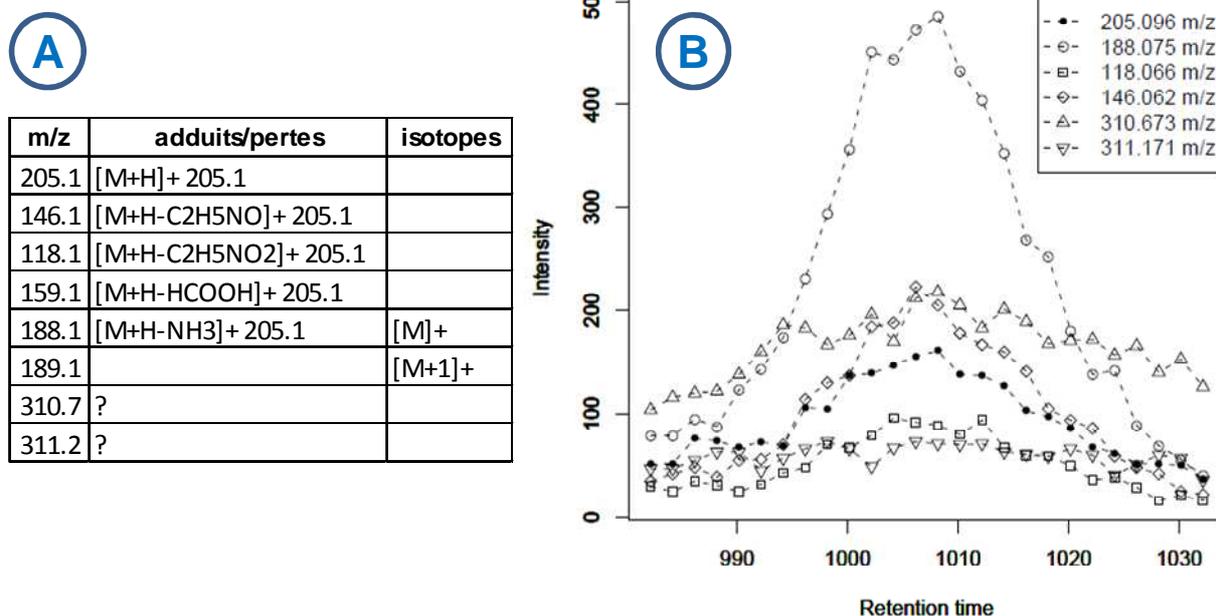


Figure 33 : Principe de la corrélation dans le domaine chromatographique [260]

Les EIBPC (« *Extracted Ion Base-Peak Chromatogram* » voir Annexe 2.) des ions co-éluant à même temps de rétention sont comparés (B) afin de déterminer lesquels appartiennent à la même molécule. Sur la base de cette comparaison de forme de pic CAMERA peut distinguer deux groupes d'ions, ceux liés au tryptophane (205.1, 146.1, 118.1, 159.1, 188.1 et 189.1) et ceux qui ne le sont pas (310.7 et 311.2) ; et ainsi réaliser une annotation différentielle pour aider à l'attribution des différents ions (A).

- Corrélation des abondances ioniques entre les différents échantillons. Le rapport des intensités entre deux ions issus d'une même molécule est théoriquement constant d'un échantillon sur l'autre (si les conditions d'analyse sont identiques). La corrélation entre les intensités d'ions pris deux à deux et détectés à un même temps de rétention dans différents chromatogrammes, est calculée pour confirmer le regroupement (Figure 34).

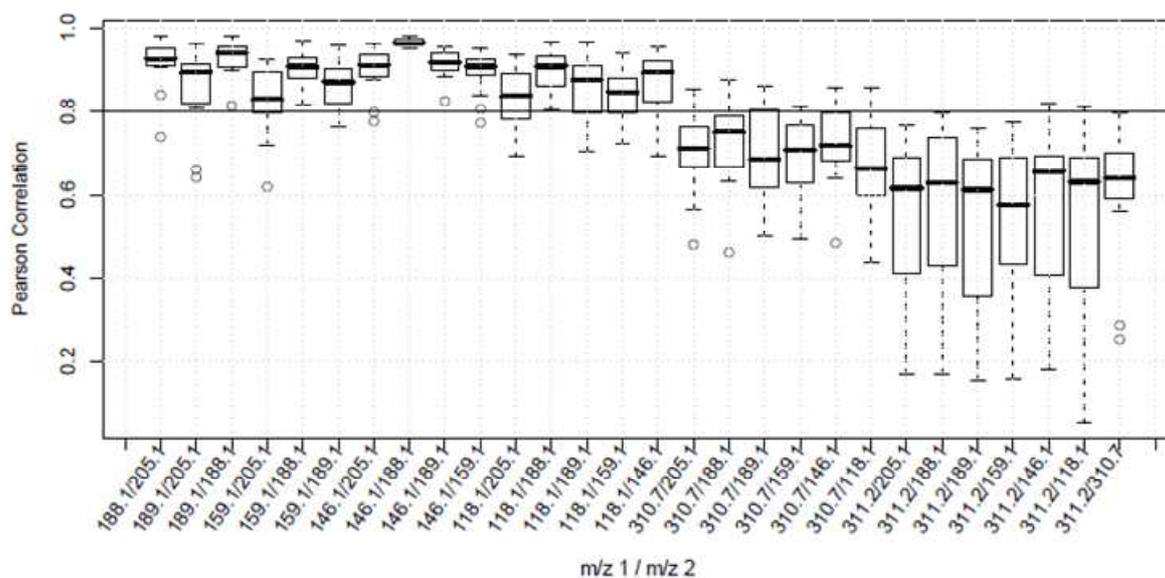


Figure 34 : Principe de la corrélation des abondances ioniques inter-échantillons [260]

Coefficients de corrélation de Pearson calculés pour les différentes paires d'ions présents au même temps de rétention. On constate que ces coefficients sont inférieurs à 0.8 quand ils concernent soit l'ion 310.7, soit l'ion 311.2, ce qui indique que ces deux ions ne sont pas reliés aux autres ions avec lesquels ils co-éluent, ils ne sont pas issus du même métabolite, en l'occurrence le tryptophane.

Le calcul de ces deux types de corrélation permet de regrouper les variables biologiques. Par exemple sur la Figure 35, 3 ions correspondant a priori à un même métabolite élué à 9.40 minutes ont été placés dans le même groupe (394) par CAMERA. L'annotation différentielle indique la présence potentielle de l'espèce moléculaire, de son isotope +1 ( $^{13}\text{C}$ ) et d'un adduit chlore. Les intensités observées sur le spectre de masse sont cohérentes avec cette annotation. Les valeurs des coefficients de corrélation ( $>0.8$ ) inter-échantillons confirment le fait que ces trois ions proviennent de la même molécule, avec comme espèce moléculaire l'ion 4679 à  $m/z = 595.3463$ . Certains ions supplémentaires peuvent également être observés sur le spectre de masse comme l'isotope  $^{18}\text{O}$  de l'espèce moléculaire, et les isotopes  $^{13}\text{C}$  et  $^{37}\text{Cl}$  de l'adduits chlore, mais ne sont pas retrouvés dans la liste de pics extraits par XCMS (Figure 35). Ces ions étaient vraisemblablement trop peu intenses pour être extraits. Mais cela soulève un point important : pour la détermination finale de tous les ions produits pour cette molécule (une fois qu'elle sera identifiée), il faudra obligatoirement faire l'annotation définitive à partir des données brutes (voir 3.2.5.) et non pas uniquement à partir de la liste de pics obtenus par XCMS qui ne prend pas en compte les ions de faible intensité.

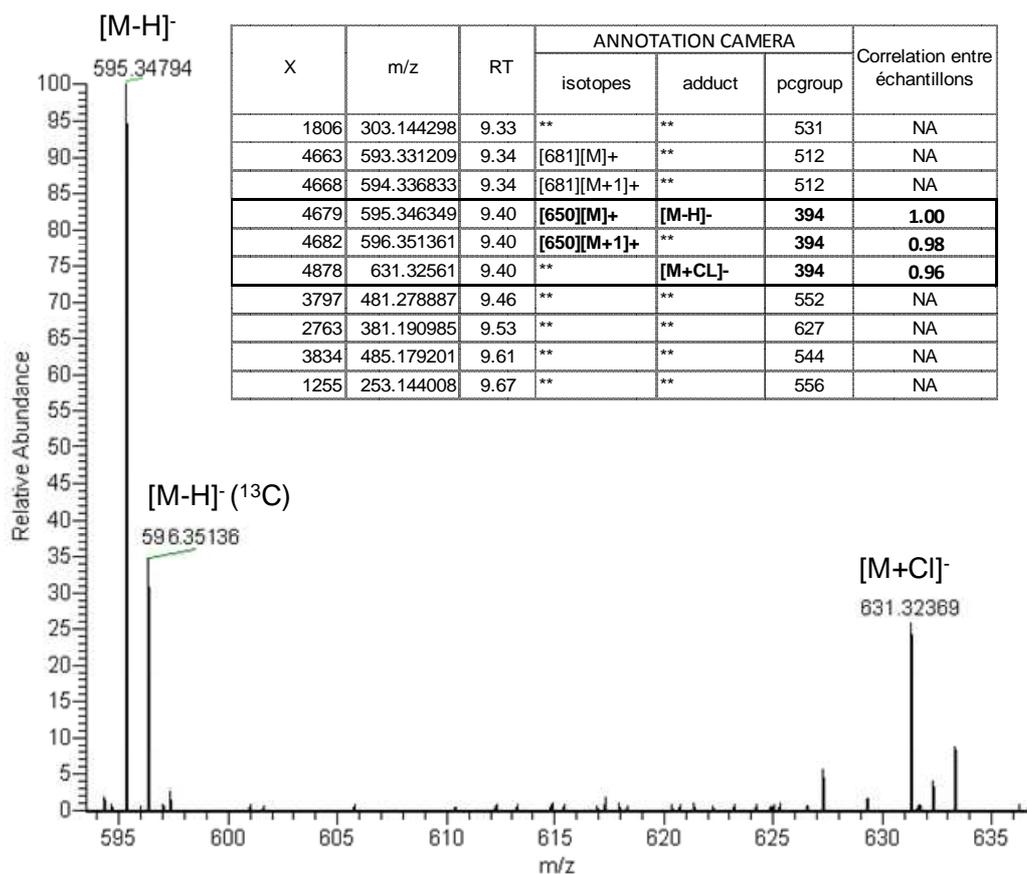


Figure 35 : Regroupement des signaux au temps de rétention 9.40 min en mode d'ionisation négatif.

### 3.2.3 Annotation des variables d'intérêt

Comme il l'a été mentionné précédemment, l'ionisation par ESI génère de nombreux ions, dont essentiellement des espèces moléculaires (protonées  $[M+H]^+$  ou déprotonées  $[M-H]^-$  selon la polarité) ce qui facilite l'annotation par les bases de données. En effet, dans les bases de données publiques (HMDB, KEGG, METLIN...) les molécules sont indexées par masses moléculaires. La masse moléculaire d'un métabolite peut être obtenue expérimentalement en ajoutant ou en soustrayant la masse d'un proton ( $= 1,00728$ ), suivant la polarité, à l'espèce moléculaire détectée. Connaissant la polarité de jeu de données, il est ainsi facile de faire une comparaison des masses expérimentales avec celles des bases de données publiques, avec une précision choisie en fonction de celle du spectromètre de masse utilisé. Dans notre cas la haute résolution et précision atteintes par le LTQ-Orbitrap nous permettent de restreindre le nombre de propositions.

HMDB, KEGG et Metlin ont été utilisées pour annoter les données biologiques. On ne retiendra que les propositions correspondant à une espèce supposée moléculaire. Si nous prenons l'exemple de l'ion N<sup>o</sup>4679 (Figure 36), deux composés sont compatibles : la C-Curarine et le L-Urobilinogène, et c'est la MS<sup>n</sup> qui peut confirmer ou non l'identification. Il se peut également que ces trois bases de

données ne génèrent aucune proposition pour certaines espèces moléculaires. Dans ce cas, il est possible de réaliser une annotation avec d'autres bases plus spécifiques comme Lipid Maps pour les lipides, ou plus généralistes comme PubChem.

X	m/z	RT	ANNOTATION CAMERA			Correlation entre échantillons	Annotation par les bases de données publiques
			isotopes	adduct	pcgroup		
1806	303.144298	9.33	**	**	531	NA	**
4663	593.331209	9.34	[681][M]+	**	512	NA	L-Urobilin
4668	594.336833	9.34	[681][M+1]+	**	512	NA	**
4679	595.346349	9.40	[650][M]+	[M-H]-	394	1.00	C-Curarine / L-Urobilinogen
4682	596.351361	9.40	[650][M+1]+	**	394	0.98	**
4878	631.32561	9.40	**	[M+CL]-	394	0.96	**
3797	481.278887	9.46	**	**	552	NA	GPCho(10:0/4:0) / GPCho(12:0/2:0)
2763	381.190985	9.53	**	**	627	NA	**
3834	485.179201	9.61	**	**	544	NA	Rutaevin / Nafenopinglycuronide
1255	253.144008	9.67	**	**	556	NA	**

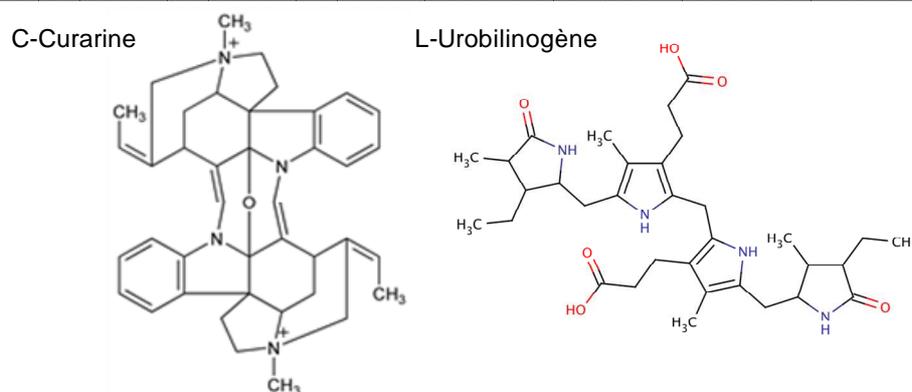


Figure 36 : Annotation par les bases de données publiques (HMDB, KEGG et Metlin) des ions entre 9.33 et 9.67 minutes en mode d'ionisation négatif.

### 3.2.4 Expériences complémentaires : MS<sup>n</sup>

Les expériences de masse en tandem sont essentielles, en particulier en l'absence de composé de référence disponible. La MS<sup>n</sup> permet de confirmer l'annotation par les bases de données publiques, ou tout du moins, en cas de propositions multiples, d'écarter les structures non cohérentes avec le schéma de fragmentation. Par exemple la C-Curarine et le L-Urobilinogène ont été proposés par les bases de données publiques pour l'ion N°4679 (Figure 36), La MS<sup>2</sup> par CID à 20% a été réalisée (Figure 37) et la fragmentation est cohérente avec le L-Urobilinogène mais pas avec la C-Curarine (Figure 38). Ce métabolite a donc été putativement identifié comme étant du L-urobilinogène.

X	m/z	RT	ANNOTATION CAMERA			Correlation entre échantillons	Annotation par les bases de données publiques	Identification
			isotopes	adduct	pcgroup			
1806	303.144298	9.33	**	**	531	NA	**	**
4663	593.331209	9.34	[681][M]+	**	512	NA	L-Urobilin	**
4668	594.336833	9.34	[681][M+1]+	**	512	NA	**	**
4679	595.346349	9.40	[650][M]+	[M-H]-	394	1.00	C-Curarine / L-Urobilinogen	L-Urobilinogen (M-H)-
4682	596.351361	9.40	[650][M+1]+	**	394	0.98	**	L-Urobilinogen (M-H)- (13C)
4878	631.32561	9.40	**	[M+CL]-	394	0.96	**	L-Urobilinogen (M+Cl)-
3797	481.278887	9.46	**	**	552	NA	GPCho(10:0/4:0) / GPCho(12:0/2:0)	**
2763	381.190985	9.53	**	**	627	NA	**	**
3834	485.179201	9.61	**	**	544	NA	Rutaevin / Nafenopinglucuronide	**
1255	253.144008	9.67	**	**	556	NA	**	**

MSN1\_20\_NEG #1915 RT: 9.38 AV: 1 NL: 9.17E4

F: FTMS - p ESI d Full ms2 595.35@cid20.00 [150.00-610.00]

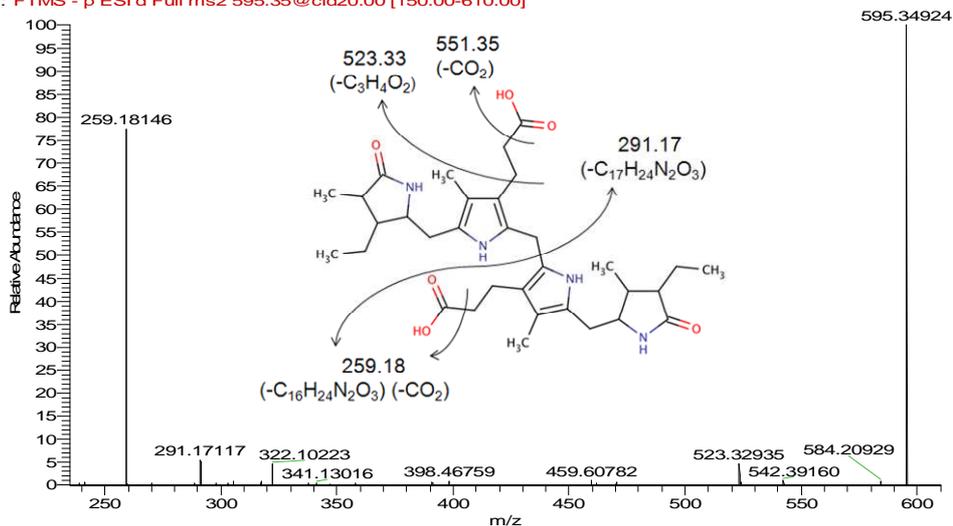


Figure 37 : Confirmation de l'identification du L-Urobilinogène par MS/MS

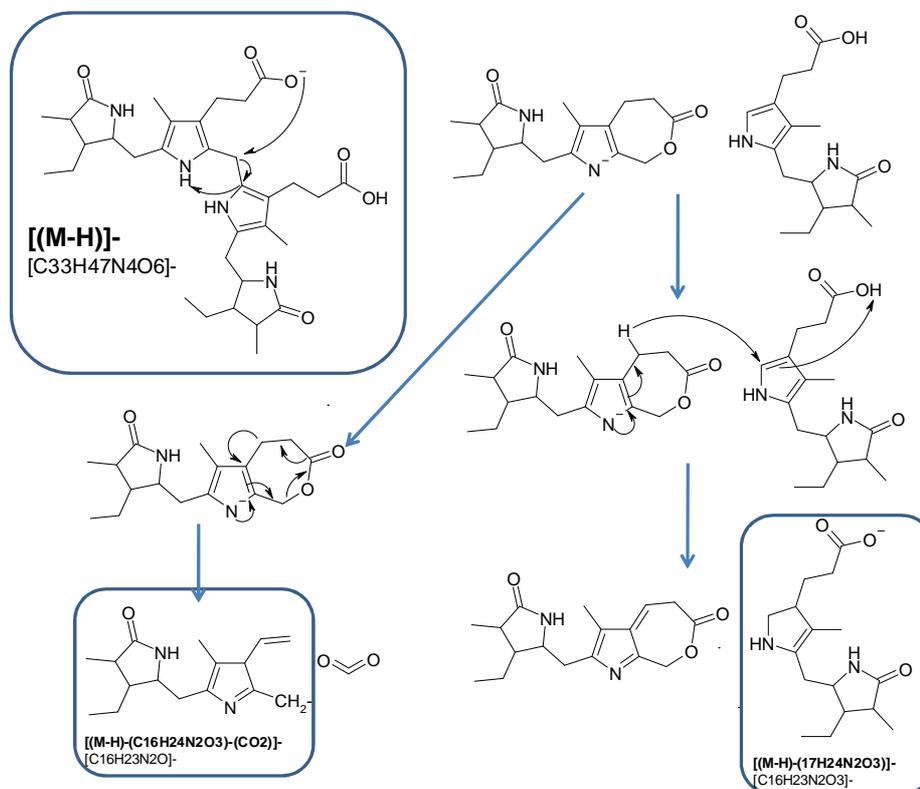


Figure 38 : Schéma de fragmentation du L-Urobilinogène.

### 3.2.5 Interprétation des spectres de masse des composés putativement identifiés

Une fois une molécule putativement identifiée, il est important de réaliser une annotation complète des ions produits pour cette molécule directement à partir des données brutes. En effet, maintenant que la formule brute de la molécule est connue, il est très simple de calculer les masses des ions théoriques (isotopes, adduits, pertes) de la même manière que pour les composés de référence, et de filtrer la liste des ions issus du spectre de masse de l'échantillon où cet ion est le plus intense. Cela permet d'annoter tous les ions produits pour une molécule, y compris ceux non extraits par XCMS (Figure 35). La liste totale des ions retrouvés pour le L-Urobilinogène sur le spectre de masse est présentée Figure 39. On y retrouve les ions initialement extraits par XCMS (les plus intenses en rouge) mais aussi d'autres ions de plus faible intensité (en vert), non extraits par XCMS, mais issus du L-Urobilinogène. Il est donc important d'annoter un maximum d'ion pour une molécule donnée. En effet, dans un contexte biologique particulier, cette molécule peut être présente à une très forte concentration dans le milieu biologique étudié, et les ions d'intensité plus faible peuvent alors être automatiquement détectés par XCMS. C'est cette liste d'ions, avec le(s) temps de rétention observés, qui sera importée dans la base de données spectrale en complément des composés de référence.

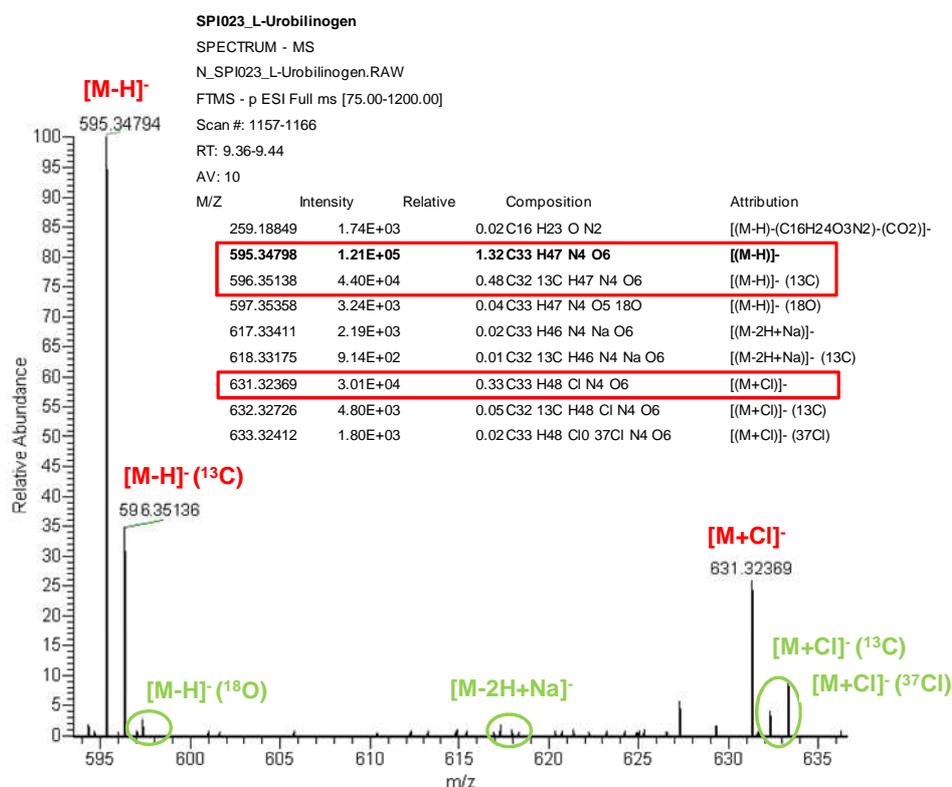


Figure 39 : Annotation du L-urobilinogène à partir du spectre de masse

Les ions présentés en rouge sont les ions initialement annotés, extraits par XCMS. Les ions présentés en vert sont les ions non extraits par XCMS et annotés directement à partir du spectre de masse.

### 3.2.6 Bilan des métabolites identifiés dans les urines humaines.

L'identification de nouveaux métabolites directement à partir du jeu de données biologiques d'urines humaines a permis de caractériser 81 métabolites supplémentaires (voir annexe 3). Ces nouveaux métabolites, dont les composés de référence ne sont pas disponibles commercialement, ont été importés dans la base de données spectrale afin de la compléter. Il est intéressant de noter que sur ces 81 métabolites, 58 n'ont pas encore été décrits dans les urines humaines selon HMDB.

## 4. CONCLUSION ET DISCUSSION

Nos travaux ont abouti à la caractérisation de 383 métabolites. Deux cent quarante quatre métabolites ont été identifiés (par comparaison de leurs paramètres physico-chimiques à ceux d'un composé de référence) ou putativement identifiés (sur la base de l'interprétation de spectres de masse et de CID), et 139 métabolites ont été uniquement caractérisés sur la base de leur masse précise (absence de spectres de MS/MS disponibles ou absence de propositions d'identification dans les bases de données). Ces 383 métabolites correspondent à environ 1000 ions dans chacun des modes d'ionisation.

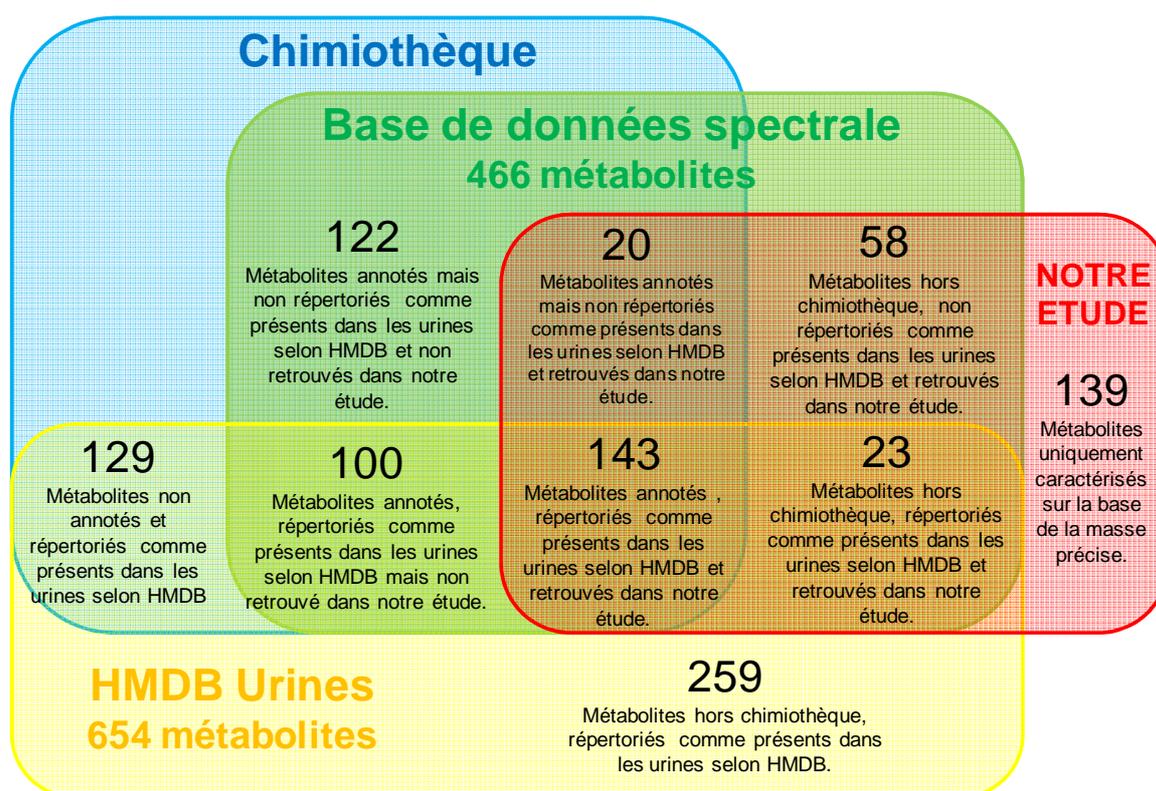


Figure 40 : Récapitulatif des métabolites présents dans la chimiothèque, annotés dans la base de données spectrale, décrits ou non dans HMDB comme étant présents dans les urines et retrouvés ou non dans notre étude.

Parmi les 244 métabolites identifiés, 78 ne sont pas répertoriés par HMDB comme étant présents dans les urines humaines. Bien que cette base de données soit mise à jour régulièrement, une

recherche parallèle sur PubMed a nous a finalement indiqué que 33 d'entre eux sont en fait détectables dans les urines humaines. Il s'agit principalement d'acylcarnitines (29 métabolites sur 33) qui ont été caractérisées à partir d'urines de 6 volontaires sains [261]. Il est aussi intéressant de noter que parmi les métabolites détectables dans les urines selon HMDB, 100 d'entre eux étaient déjà annotés à partir de composés de référence existant dans notre chimiothèque, mais ils n'ont pas été retrouvés dans les échantillons de notre cohorte (Figure 40). L'absence de ces métabolites peut être attribuée à un manque de sensibilité de notre technique de LC/MS. Elle peut également être mise sur le compte de la forte variabilité inter-individuelle de composition biochimique des urines. Enfin, nous disposons dans notre chimiothèque de 129 métabolites dont la présence urinaire est répertoriée par HMDB, mais qui n'ont pas encore été annotés. Ces métabolites n'ont donc pas été pris en compte dans notre étude il est donc pour l'instant impossible de statuer sur leur présence ou non dans les échantillons de notre cohorte.

Les 244 métabolites identifiés se répartissent dans les différentes classes chimiques connues de métabolites (Figure 41). Ils correspondent principalement à des acides organiques (et dérivés), à des acides aminés (et dérivés) et à des acylcarnitines. Notons qu'une publication récente a fait état de la détection de 385 acylcarnitines dans les urines humaines, alors que nous n'en avons trouvé que 29 [261]. Cette différence est probablement due au traitement spécifique de l'échantillon mis en œuvre : extraction sur phase solide, suivi d'une concentration d'un facteur 10 de l'éluat avant injection dans le système chromatographique. D'autre part, la méthode séparative de type UHPLC est plus résolutive que la notre, permettant peut-être de séparer un plus grand nombre d'isomères. Enfin, les expériences de MS<sup>2</sup> en mode de perte de neutre constante (59 u), de balayage d'ion parent (*m/z* 85) et surtout de SRM qui ont été utilisées pour détecter les acylcarnitines apportent peut-être un gain de sensibilité par rapport à un spectre de masse enregistré sur un LTQ-Orbitrap discovery.

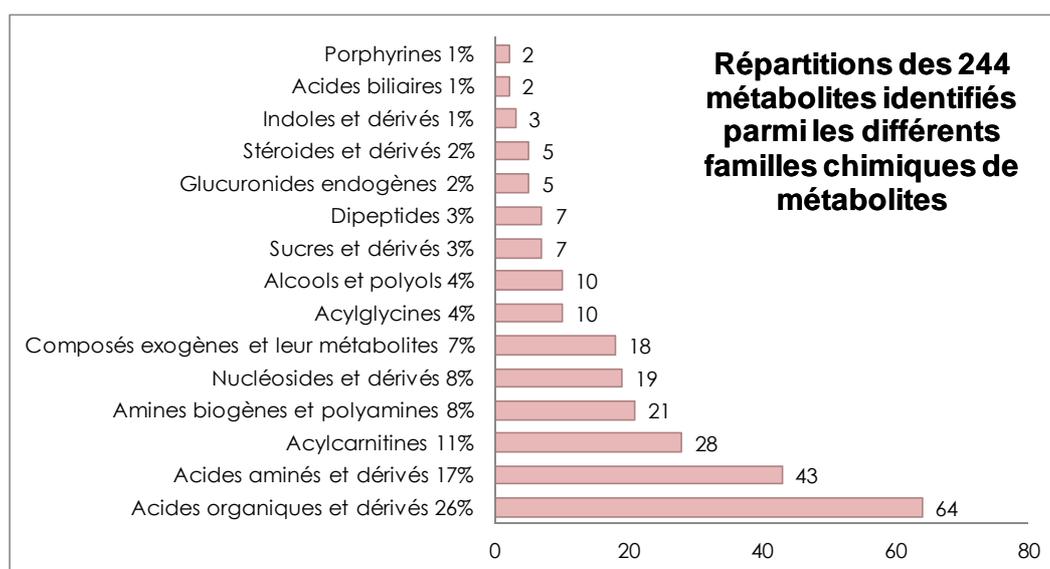


Figure 41 : Différentes familles chimiques des métabolites identifiés.

Notons que la fouille de données a été orientée de manière à faire ressortir les métabolites communs au plus grand nombre d'échantillons. Ainsi, les métabolites identifiés sont principalement des métabolites endogènes (primaires et secondaires), et dans une moindre mesure des xénobiotiques (7%) parmi les plus courants tels que les métabolites de la caféine, issus du tabac ou encore des médicaments très utilisés comme le paracétamol. L'étude plus approfondie des xénobiotiques passera par la mise en œuvre de méthodes de fouilles de données compatibles avec l'existence d'une forte variabilité inter-individuelle. Citons par exemple les techniques d'analyse en composantes indépendantes qui permettent de mettre en évidence de nouveaux métabolites à partir d'échantillons individuels. Enfin, parmi les métabolites les moins présents, on retrouve des familles de métabolites normalement peu ou pas excrétées dans les urines : les sucres (filtrés mais totalement réabsorbés) et les acides biliaires (éliminés principalement dans les urines uniquement en cas d'un obstacle sur les voies biliaires).

Durant cette étude, s'est posé le problème des critères à utiliser pour identifier des métabolites dans les fluides biologiques. Il est classiquement admis au sein de la communauté des chercheurs en métabolomique qu'un métabolite est formellement identifié quand il présente au moins deux paramètres physico-chimiques identiques à celui d'un composé de référence synthétisé chimiquement [191;192]. Nous avons ainsi pu identifier 163 métabolites dans les échantillons urinaires de notre cohorte sur la base de ces critères. Cette approche n'est malheureusement pas adaptée à la plupart des situations, en raison de l'indisponibilité commerciale d'un grand nombre de composés. Dans ce cas précis, le groupe de travail de la société de métabolomique en charge des critères d'identification des métabolites propose une identification dite putative basée sur des propriétés physico-chimiques et/ou des spectres similaires à ceux de composés répertoriés dans des bases de données publiques ou commerciales [191;192]. Nous avons pu identifier «putativement» 81 métabolites, principalement sur la base de l'interprétation de spectres de MS<sup>2</sup>, dans les échantillons d'urine de notre cohorte d'individus.

Les données de cette étude métabolomique ont été structurées afin de regrouper des ions correspondants à un même métabolite. L'espèce moléculaire porteuse de l'information de la masse moléculaire a été ensuite sélectionnée, et validée grâce à des expériences MS/MS pour confirmer ou infirmer les structures proposées par les bases de données publiques. En effet, même si l'ESI-MS génère plusieurs ions pour une même molécule, ces derniers sont structurellement peu informatifs, à l'exception des ions fragments générés dans la source et des isotopes tels que le <sup>34</sup>S. Ces types d'ions sont généralement peu observés, c'est-à-dire dans environ 20% des métabolites putativement identifiés dans le cadre de cette étude. Il s'est avéré que les informations fournies par la LC/MS (polarité, isotopes, fragments en source) et les spectres de MS/MS (fragments) sont la plupart du temps suffisantes pour identifier de manière putative un métabolite. Mais dans certains cas, il est difficile de faire la différence entre des isomères qui présentent le même profil de fragmentation et une polarité très proche, comme c'est par exemple le cas pour les dipeptides contenant de la leucine ou de l'isoleucine (Asp-Leu et Asp-Ile, par exemple) : les spectres MS/MS sont identiques et présentent des pertes de type H<sub>2</sub>O, NH<sub>3</sub>, HCOOH et C<sub>4</sub>H<sub>4</sub>O<sub>3</sub>N. La seule différence se situe au niveau du temps

de rétention : l'un des deux isomères est plus retenu, et donc plus apolaire que l'autre (5.00 min vs. 5.19 min). Sachant que c'est la leucine qui est plus apolaire que l'isoleucine (LogP= -1.52 vs. -1.70. HMDB), l'Asp-Ile peut être annoté à 5.00 min et l'Asp-Leu à 5.19 min. Dans ce cas précis, il y a 2 isomères possibles et deux pics chromatographiques. Par conséquent, l'attribution peut être faite sur la base d'autres informations que celle provenant de la spectrométrie de masse. Malheureusement ce n'est pas toujours possible : le pic de  $\gamma$ -Glu-Leu/Ile à 5.20 min peut correspondre à quatre isomères (d'après les spectres de MS/MS) : Glu-Leu, Glu-Ile,  $\gamma$ -Glu-Leu ou  $\gamma$ -Glu-Ile, mais il est impossible de faire la différence sur la base du temps de rétention car dans ce cas précis il n'y en a qu'un seul. La seule alternative ici serait la synthèse chimique de ces différents isomères (non commercialement disponibles) dans le cadre de collaborations avec des équipes spécialisées dans ce domaine.

La présence de nombreux isomères dans les fluides biologiques pose également le problème de la résolution chromatographique : en utilisant 2 méthodes chromatographiques différentes, il a été mis en évidence la co-élution (dans le volume mort) de nombreux isomères lorsqu'ils sont chromatographiés sur une colonne de type C<sub>18</sub>. Ce phénomène a pour conséquence une sous-estimation du nombre de métabolites en C<sub>18</sub> par rapport à l'élution sur colonne PFPP. Inversement, certaines molécules détectées avec la colonne C<sub>18</sub> ne sont pas éluées sur la colonne PFPP (bases organiques très apolaires). Le choix du type de séparation est donc extrêmement important et dépendant du type de métabolites que l'on s'attend à retrouver. Par exemple, pour des extraits cellulaires composés principalement de petites molécules polaires, il faudra plutôt choisir une phase stationnaire adaptée à ce type de molécules comme par exemple une phase normale (silice) ou HILIC. Les urines sont composées à la fois de petites molécules polaires (dont les isomères co-éluent dans le volume mort en C<sub>18</sub>), et de molécules plus lipophiles comme les carnitines (qui ne sont d'ailleurs pas toutes éluées avec la colonne PFPP). Il faudrait donc, afin d'être le plus exhaustif possible, utiliser plusieurs méthodes séparatives différentes, mais cela implique d'augmenter considérablement le nombre d'analyses.

Se pose enfin plusieurs problèmes relatifs à l'exploitation des données rassemblées lors de cette étude : comment et sous quelle forme partager ces informations ? Concernant les données spectrales, notre base de données d'ESI-MS est une base conçue pour annoter des jeux de données et elle ne tient pas compte des intensités relatives des ions, ce qui pose moins de problèmes de normalisation que les bases de données MS/MS. Le partage sous forme de liste d'ions est donc envisageable et devrait être utilisable pour annoter des listes d'ions provenant d'autres types d'appareils. A terme, l'idée est de pouvoir partager cette liste d'ions et de l'enrichir grâce à des collaborations avec des laboratoires travaillant sur d'autres systèmes analytiques. Concernant les données biologiques relatives à la présence de certains métabolites dans les urines, le partage devrait se faire via la base de données HMDB. Pour les métabolites formellement identifiés par comparaison au standard, le partage de l'information ne pose aucun problème, mais pour les métabolites putativement identifiés, il dépend du niveau de certitude de l'identification. Pour les métabolites sur lesquels il n'y a aucune ambiguïté (Asp-Leu et Asp-Ile par exemple), il n'y a aucune objection à publier dans HMDB, mais pour ceux pour lesquels il est impossible, pour le moment, de conclure

formellement ( $\delta$ -Glu-Leu/Ile par exemple), il n'est pas rigoureux de partager l'information via HMDB, même si, d'un point de vue biologique, cette information est pertinente. Une solution pourrait être de fournir, dans le cadre de publications, des listes d'ions correspondant à ces métabolites putativement identifiés. Ces listes pourraient être utilisées par d'autres laboratoires pour annoter leurs propres jeux de données.

Tirant parti des identifications obtenues lors de la construction de cette base de données spectrale, l'influence de facteurs physiologiques tels que le genre, l'âge et l'IMC sur les concentrations urinaires des métabolites peut maintenant être étudiée grâce à des outils de statistiques et en vue d'initier la construction d'une base de données relationnelle de métabolites dans les biofluides.

CHAPITRE 2 : EVALUATION DE L'IMPACT  
DES FACTEURS PHYSIOLOGIQUES SUR  
LE METABOLOME URINAIRE



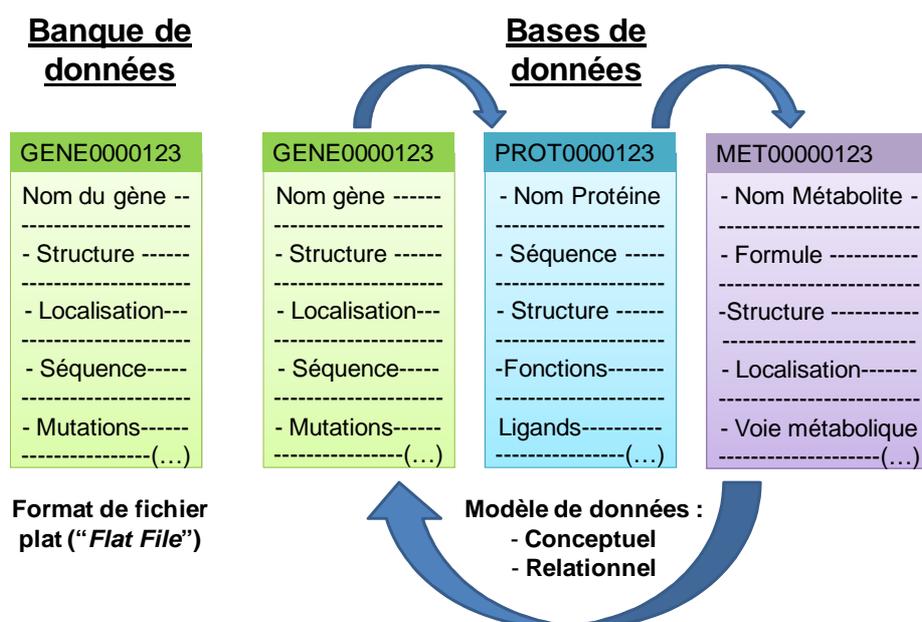
# CHAPITRE 2 : EVALUATION DE L'IMPACT DES

## FACTEURS PHYSIOLOGIQUES SUR LE

### METABOLOME URINAIRE

## 1 INTRODUCTION

Les banques de données biologiques sont des bibliothèques d'informations relatives aux sciences de la vie, recueillies à partir d'expériences scientifiques, de technologies d'expérimentation à haut débit, du traitement de ces données mais aussi de la littérature. La première banque de données voit le jour en 1965 et répertorie des données sur les protéines : *Atlas of Protein Sequences and Structures* [262]. Les données de cet atlas (en version papier) ont ensuite servies de base pour la création de la première banque de données informatique, la PIR [263] (*Protein Information Ressource*), en 1984 (<http://pir.georgetown.edu/>). Dans les années 80, les initiatives de banques de données se multiplient avec la création de GenBank™ [264], EMBL *Nucleotide Sequence Database* [265], DDBJ [266] (*DNA Data Bank of Japan*) pour l'étude du génome, et de Swiss-Prot [267] pour les protéines. Depuis, le nombre de banques disponibles n'a cessé d'augmenter. Sont également apparues les bases de données conceptuelles et relationnelles, plus complexes, et qui compilent des informations dans plusieurs domaines différents (Figure 42).



**Figure 42 : Différences de structure entre banques et bases de données**

La structure d'une banque de données est simple : elle compile des informations stockées dans des fichiers plats, sans relation entre eux. On peut alors accéder à une grande quantité de données sur un certain type d'entité (gène, protéine, métabolite...). En génomique par exemple, les informations répertoriées dans les banques de données sont la fonction du gène, sa structure, sa localisation (à la fois cellulaires et chromosomiques), les effets cliniques des mutations ainsi que les similitudes avec d'autres séquences. La structure d'une base de données est plus complexe : les fichiers sont reliés entre eux par des références croisées, on peut facilement naviguer d'une entité à l'autre et faire des recherches thématiques plus spécifiques grâce aux relations entre les différentes données, d'où le terme souvent utilisé de base de données relationnelle. Par exemple UniProt, base de données de protéines, répertorie à la fois les informations sur les protéines, mais aussi celles sur la séquence génétique (ADN et ARN) à l'origine de la protéine, sur sa structure et ses variations phylogéniques, par le biais de références croisées avec d'autres banques de données (EMBL, GenBank, GeneTree [268]...). Elles contiennent donc des données concernant divers domaines comme la génomique, la protéomique, la métabolomique, la transcriptomique, la phylogénétique.... Les banques et/ou bases de données biologiques peuvent être très générales ou très spécialisées (dédiées à une thématique et/ou un organisme). Il existe par exemple, en génomique, des bases de données spécifiques à certaines espèces : Colibase [269] est une base de données dédiée à *E. coli*, FlyBase [270] pour la drosophile, et WormBase [271] pour les nématodes *Caenorhabditis elegans* et *Caenorhabditis briggsae*.

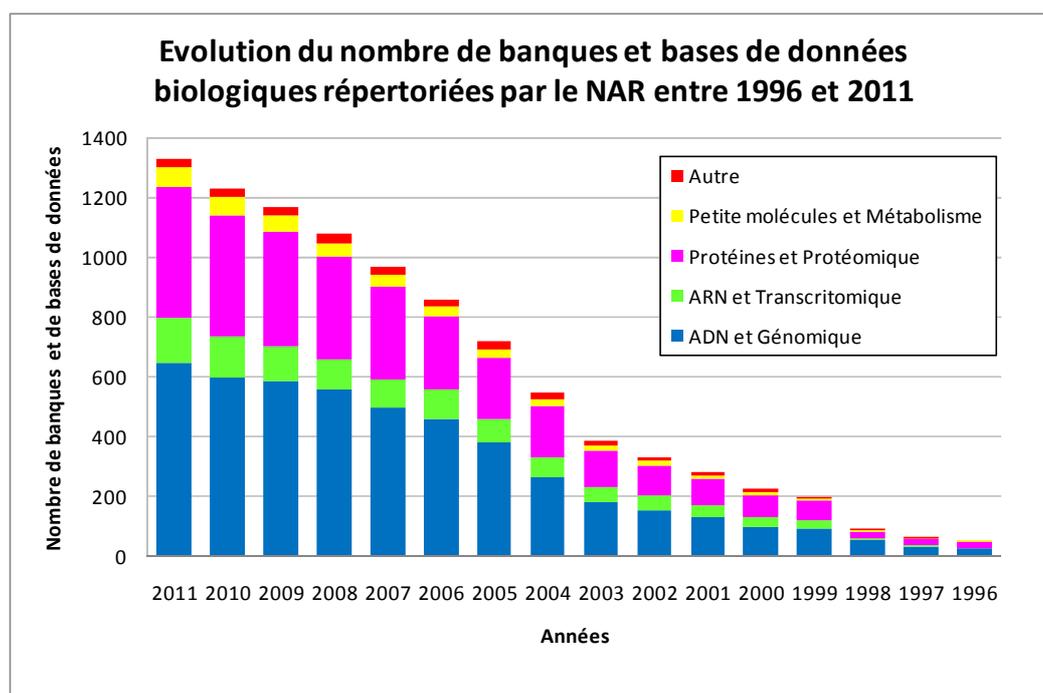


Figure 43 : Evolution du nombre de banques et bases de données biologiques répertoriées par le NAR entre 1999 et 2011, classées par thématiques.

(source <http://www.oxfordjournals.org/nar/database/a/>).

Depuis 1996, Le journal *Nucleic Acids Research* répertorie annuellement les bases et banques de données biologiques existantes au sein de la *Molecular Biology Database Collection* [272]. A la fin des années 90, il n'existait que peu de bases et banques de données biologiques (environ 50) et principalement dans les domaines de la génomique et de la protéomique (Figure 43). Dès le début des années 2000 les bases de données se diversifient et apparaissent les bases de données relatives à l'ARN et aux petites molécules. Depuis le nombre de bases de données disponibles n'a cessé d'augmenter, mais toujours avec la prédominance des bases relatives à l'ADN et aux protéines. La dernière mise à jour date de janvier 2011 et compile 1330 bases de données biologiques.

Ces banques et bases de données biologiques sont des outils de première importance qui aident les scientifiques à comprendre et à expliquer de nombreux phénomènes biologiques tels que la structure des biomolécules et leurs interactions entre elles, le métabolisme de divers organismes et la compréhension de l'évolution des espèces. Ces connaissances sont indispensables dans la lutte contre les maladies comme le cancer (physiopathologie, diagnostic...) ou pour l'élaboration de nouveaux médicaments. Si les bases de données sont très nombreuses dans certains domaines comme la génomique et la protéomique (Figure 43) grâce au recul existant pour ces disciplines, elles sont beaucoup moins développées pour ce qui est de l'étude des petites molécules et des voies métaboliques. Les bases de données structurales pour les petites molécules (ChEBI [273] - *Chemical Entities of Biological Interest* ou ChemBank [274]) peuvent être utilisées en métabolomique à des fins

d'identification, et les bases de données de voies métaboliques (KEGG [275] ou MetaCyc [276]) pour étudier les interactions entre les différents métabolites dans un organisme donné. Cependant les bases de données biologiques dédiées à la métabolomique sont peu nombreuses (HMDB [195], Metlin [197]), et les informations concernant les variations physiologiques ou pathologiques des concentrations de métabolites ne sont pas toujours disponibles ou du moins difficilement accessibles. De 1982 à 1986, des tables biologiques étaient éditées (*Geigy scientific tables* [277]) dans lesquelles ces informations étaient accessibles, mais depuis seule la base de données HMDB (*Human Metabolome Database*) répertorie ce type de données. Pour chaque métabolite enregistré, on trouve des renseignements concernant sa structure, sa famille chimique, ses origines biologiques, ses propriétés spectrales, mais aussi sur les concentrations normales ou anormales (pathologie, exposition à un toxique) dans différents fluides biologiques. Ces données quantitatives sont issues à la fois de la littérature, mais aussi d'études métabolomiques.

Cependant, peu d'études métabolomiques concernant les variations physiologiques et leur impact sur les concentrations des métabolites ont été publiées, la plupart préférant se focaliser sur la découverte de biomarqueurs spécifiques de pathologies ou résultant de l'exposition à un toxique. La seule étude actuellement publiée concerne l'analyse du métabolome plasmatique par LC/MS et GC/MS et les variations de concentrations de métabolites observées en fonction du genre, de l'âge et de l'origine ethnique [53]. Elle a été réalisée sur une cohorte de 269 individus. D'un point de vue analytique, l'utilisation de deux méthodes d'analyse a permis d'identifier 300 métabolites dans le plasma (60% par LC/MS et 40% par GC/MS). Des analyses statistiques réalisées sur les métabolites formellement identifiés (sur la base de la comparaison des temps de rétention chromatographiques et des spectres de MS/MS des composés à identifier à ceux de composés de référence) et présents dans au moins 2/3 des échantillons, ont montré que les concentrations d'une centaine d'entre eux sont impactées par au moins un des facteurs biologiques étudiés. Selon cette étude, l'âge des individus est le facteur qui a le plus de répercussion sur les concentrations plasmatiques des métabolites. Une influence de l'indice de masse corporelle a également été constatée sur les concentrations de certains métabolites déjà sensibles aux influences de l'âge, du genre ou de l'origine ethnique.

L'analyse des grandes cohortes (plus de quelques centaines d'échantillons) est principalement réalisée par RMN [233] en raison des excellentes performances de cette technique en termes de répétabilité et de reproductibilité, contrairement aux approches de type LC/MS dans leur état actuel de développement. Cependant, ces dernières, plus sensibles et permettant l'analyse plus poussée des espèces lipidiques, sont de plus en plus utilisées en métabolomique, mais posent encore quelques problèmes méthodologiques : un important effet matrice, qui peut induire un biais analytique conséquent au cours du temps (encrassement de la source ESI quand le nombre d'échantillon est élevé), et un manque de reproductibilité qui limite la comparaison d'expériences effectuées à des jours différents. L'étude de grandes cohortes comme celle citée précédemment [53] représente donc un challenge à la fois sur le plan analytique (pour s'affranchir de la dérive analytique) et sur le plan de l'analyse statistique (choisir les bons tests en fonction de la population et des facteurs étudiés). Cette

partie du travail de thèse a pour objectif d'évaluer, à partir de la population de 227 individus, l'impact des facteurs physiologiques étudiés (genre, âge et IMC) sur les concentrations urinaires des métabolites identifiés grâce à la base de données spectrale (chapitre 1) afin de commencer à construire une banque de données relationnelle de métabolites dans les biofluides humains.

## **2 MATERIEL ET METHODES**

### **2.1 REACTIFS ET PRODUITS CHIMIQUES**

Les réactifs et produits chimiques sont ceux présentés dans le Chapitre 1 (page 61).

### **2.2 ECHANTILLONS BIOLOGIQUES**

Les échantillons biologiques sont ceux présentés dans le Chapitre 1 (page 61).

### **2.3 PREPARATION DES ECHANTILLONS**

Les échantillons biologiques ont été préparés comme décrit dans le Chapitre 1 (page 62).

### **2.4 ACQUISITION DES DONNEES**

#### **2.4.1 Appareillage**

Les analyses par spectrométrie de masse ont été réalisées sur le matériel décrit dans le Chapitre 1 (page 62).

#### **2.4.2 Analyses par LC/MS**

Les analyses par LC/MS ont été réalisées comme décrit dans le chapitre 1 (page 63), mais un seul type de séparation chromatographique ( $C_{18}$ ) a été utilisé pour les analyses statistiques.

#### **2.4.3 Séquences d'analyse**

Les séquences d'analyse ont été présentées dans le chapitre 1 (page 64). Pour rappel, en mode d'ionisation positif, tous les échantillons ont été analysés au cours d'une même séquence d'analyse. En mode d'ionisation négatif les échantillons ont été passés en deux temps, avec un arrêt de l'appareil pour un nettoyage de la source ESI entre les deux séquences d'analyses.

### **2.5 TRAITEMENT DES DONNEES BIOLOGIQUES**

Le traitement des données a été réalisé comme décrit dans le Chapitre 1 (page 65). Les corrections par normalisation ont été réalisées avec le logiciel R.

## 2.6 ANALYSES STATISTIQUES

Les tests d'indépendance (tests du  $\chi^2$  de Pearson) ont été réalisés avec le logiciel GraphPad Prism (La Jolla, USA), et les analyses statistiques univariées (ANOVA et coefficients de corrélation de Spearman) ont été réalisées avec le logiciel R.

## 2.7 IDENTIFICATIONS

Les métabolites analysés sont ceux qui ont été annotés par la base de données spectrale, avec les paramètres décrits dans le Chapitre 1 (page 66).

# 3 RESULTATS

## 3.1 ETUDE PRELIMINAIRE DE LA POPULATION

Il est important de bien connaître la population étudiée, et notamment la dépendance ou l'indépendance des paramètres physiologiques. En effet dans certains cas il peut y avoir un lien entre deux facteurs physiologiques, et si on observe une variation des métabolites avec ces deux facteurs, il sera difficile de savoir lequel des deux est responsable de cette variation. Le test statistique utilisé, le  $\chi^2$  de Pearson, s'intéresse à la différence entre une proportion observée (ou valeur empirique) et la valeur attendue s'il y avait indépendance des facteurs (ou valeur théorique). Dans notre cas nous avons 3 facteurs : l'âge, le genre et l'IMC (l'âge et l'IMC, initialement variables quantitatives, ont été transformées en variables qualitatives pour pouvoir réaliser ce test). Des tests d'indépendance ont été réalisés pour chaque facteur deux à deux (Age vs. Genre, Genre vs. IMC et IMC vs. Age).

### 3.1.1 Test d'indépendance entre les variables Genre et IMC

Un premier test d'indépendance a été réalisé entre le genre et l'IMC. Celui-ci a d'abord été réalisé sur l'ensemble des classes des deux facteurs (Figure 44 gauche) et est ressorti significatif. Cependant la règle de Cochran (aucun effectif plus petit que 1 et moins de 20% d'entre eux inférieurs à 5) n'est pas respectée à cause des effectifs de très faible et très fort IMC (« Maigreur » et « Obésité »). Le test a donc été réalisé une seconde fois uniquement sur les classes « Normal » et « Surpoids » qui elles respectent la règle de Cochran (Figure 44 droite). Celui-ci est toujours significatif, ce qui implique que le genre et l'IMC sont des facteurs dépendants : la proportion de personnes en surpoids est plus élevée chez les Hommes, les proportions d'obèses et de maigres sont elles plus élevées chez les Femmes. Il y a donc bien un lien entre le genre et l'IMC d'un individu.

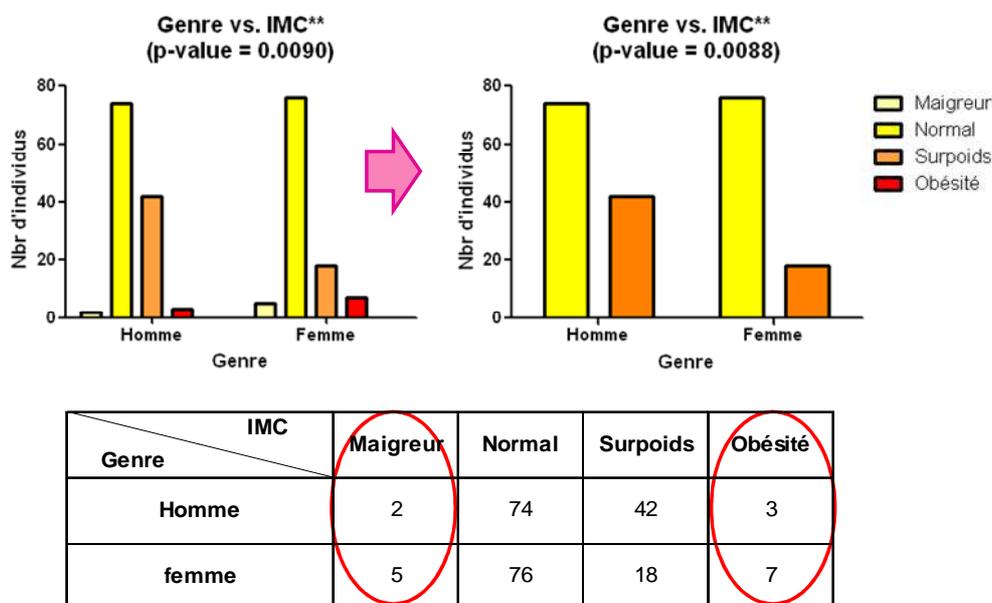


Figure 44 : Test d'indépendance entre les variables Genre et IMC

### 3.1.2 Test d'indépendance entre les variables Age et IMC

Un premier test d'indépendance a été réalisé entre l'âge et l'IMC. Celui-ci a d'abord été réalisé sur l'ensemble des classes des deux facteurs (Figure 45 gauche) et est ressorti significatif. Cependant la règle de Cochran (aucun effectif plus petit que 1 et moins de 20% d'entre eux inférieurs à 5) n'est pas respectée à cause des effectifs de très faible et très fort IMC (« Maigreur » et « Obésité »). Le test a donc été réalisé une seconde fois uniquement sur les classes « Normal » et « Surpoids » qui elles respectent la règle de Cochran (Figure 45 droite). Celui-ci est toujours significatif, ce qui implique que l'âge et l'IMC sont des facteurs dépendants : les proportions de personnes en surpoids et obèses augmentent avec l'âge, alors que la proportion de personnes ayant un IMC normal diminue. Il y a donc bien un lien entre l'âge et l'IMC d'un individu.

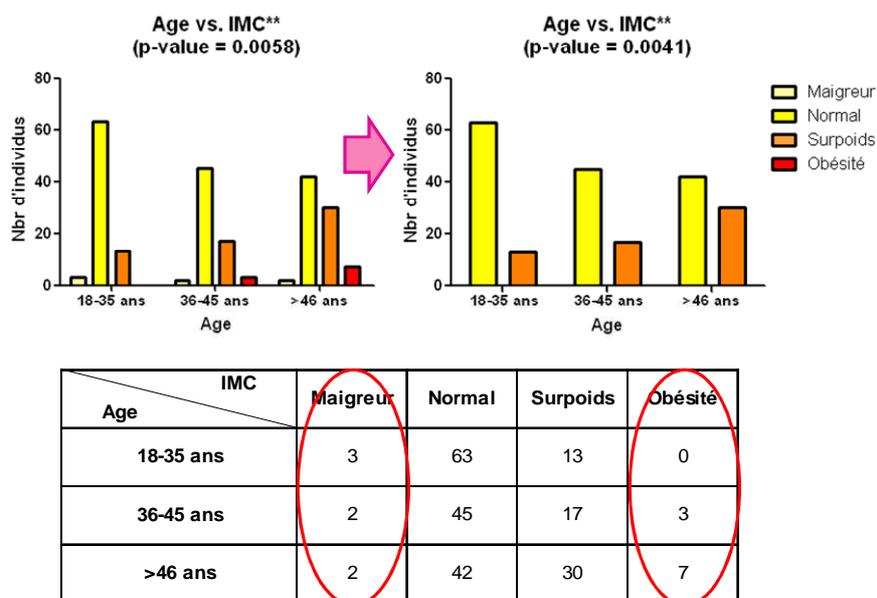


Figure 45 : Test d'indépendance entre les variables Age et IMC

### 3.1.3 Test d'indépendance entre les variables Genre et Age

Un premier test d'indépendance a été réalisé entre l'âge et le genre. Celui-ci a été réalisé sur l'ensemble des classes des deux facteurs (Figure 46) et n'est pas ressorti significatif. Comme la règle de Cochran est respectée, on peut conclure que l'âge et le genre sont des facteurs indépendants et donc qu'il n'y a aucun lien entre l'âge et le genre d'un individu.

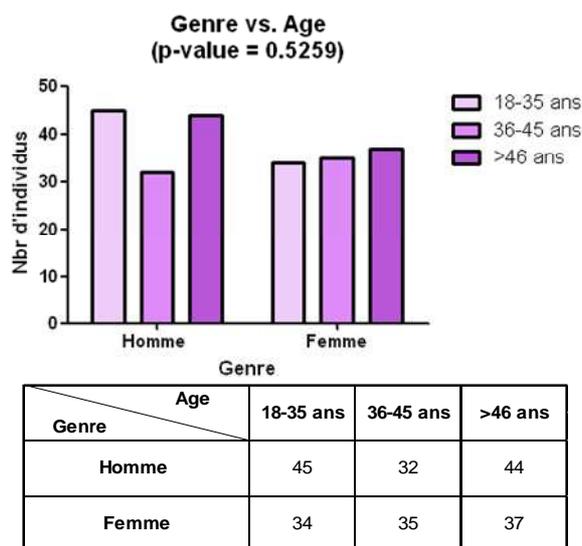


Figure 46 : Test d'indépendance entre les variables Age et Genre

### 3.1.4 Choix des tests statistiques

L'objectif est ici de réaliser des tests statistiques pertinents afin d'évaluer les variations de concentrations relatives des métabolites en fonction des facteurs physiologiques. De nombreux types de tests statistiques sont disponibles (Tableau 9). Le choix du type de test à réaliser dépend de la population et des variables étudiées. Pour des variables qualitatives, le type de test dépend du nombre de classes de la variable : test de comparaison de moyenne pour une variable à deux classes, au-delà de deux classes, une ANOVA ou analyse de variance. Pour les variables quantitatives, les tests disponibles sont des régressions ou tests de corrélation. Dans tous les cas, les tests peuvent être soit paramétriques quand la population initiale a une distribution normale, soit non paramétriques. Dans notre cas, la population d'origine étant inconnue (ou plutôt impossible à définir car elle correspond à l'ensemble des individus travaillant au CEA) on ne peut pas présager de sa distribution ni des valeurs de variances des variables au sein de celle-ci, les tests choisis seront des tests non paramétriques (Tableau 9). Les facteurs âge et genre étant étroitement liés à l'IMC, nous avons réaliser des tests indépendants pour chacun des facteurs.

- Le variable «genre» étant catégorielle (ou qualitative) à deux classes, le test effectué sera un test de comparaison des médianes de Mann-Whitney-Wilcoxon (Tableau 9).
- Les variables «âge» et «IMC» étant des variables quantitatives, les tests effectués seront des tests de corrélation de Spearman (Tableau 9).

**Tableau 9 : Types de tests statistiques disponibles et critères de choix**

	VARIABLES QUALITATIVES						VARIABLES QUANTITATIVES	
	k = 2 classes			k > 2classes				
Nom du Test ( <sup>P</sup> ) = paramétrique ( <sup>NP</sup> ) = non paramétrique	Student ( <sup>P</sup> )	Student avec correction de Welch ( <sup>P</sup> )	Mann-Whitney-Wilcoxon ( <sup>NP</sup> )	ANOVA ( <sup>P</sup> )	ANOVA avec correction de Welch ( <sup>P</sup> )	Kruskal-Wallis ( <sup>NP</sup> )	Pearson ( <sup>P</sup> )	Spearman ( <sup>NP</sup> )
Détails du test	Test de comparaison de 2 moyennes	Test de comparaison de 2 moyennes	Test de comparaison de 2 médianes	Test de comparaison de plus de 2 moyennes	Test de comparaison de plus de 2 moyennes	Tes de comparaison de plus de 2 médianes	Test de corrélation linéaire	Test de rang
Choix des individus dans la population	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire
Valeurs des variables	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Pas de valeurs manquantes**	Pas de valeurs manquantes**
Relation entre les variables							linéaire	monotone
Distribution des variables dans la population d'origine	Normale	Normale	NA	Normale	Normale	NA	Normale	NA
Variances des variables	identiques entre elles et à la population	NA	NA	identiques entre elles et à la population	NA	NA		

NA : Non Applicable

\*Si un individu appartient à la classe A il ne peut pas appartenir aussi à la classe B

\*\* Chaque individu doit posséder une valeur pour chacune des variables

Pour chacun des tests, une correction des tests multiples a été appliquée afin d'obtenir une «*q-value*». La «*q-value*» est une valeur corrigée de la «*p-value*» qui est calculée lors de tests multiples [278;279]. La «*p-value*» correspond à la probabilité que la différence définie comme significative pour un test soit en fait non significative (faux positif). La «*q-value*» elle, correspond au FDR (de l'anglais

«*False Discovery Rate*»), c'est-à dire la probabilité que sur l'ensemble des tests aux résultats significatifs (quand on réalise des tests multiples) un certain pourcentage ne le soit en fait pas.

### 3.2 ANALYSES STATISTIQUES SUR LES DONNEES NON NORMALISEES

Ces analyses statistiques, se sont focalisées sur les 244 métabolites déjà identifiés dans les urines et répertoriés dans la base de données spectrale (voir Chapitre 1). Parmi ces 244 métabolites, 31 ne sont pas statistiquement interprétables à cause de la mauvaise résolution chromatographique (Figure 25), il reste donc 213 métabolites statistiquement interprétables. Parmi ces 213 métabolites, ne seront conservés pour l'analyse statistique que ceux retrouvés dans les échantillons de QC. En effet, la présence d'un métabolite dans ces échantillons est indispensable pour pouvoir évaluer le biais analytique au cours du temps, biais analytique qui peut avoir un impact sur les résultats statistiques. 115 métabolites ont été retrouvés dans les QC en mode d'ionisation positif, 115 également en mode d'ionisation négatif, dont 32 métabolites communs aux deux modes, ce qui fait au total 198 métabolites identifiés et statistiquement interprétables retrouvés dans les QC. Il y a donc 15 métabolites sur les 213 qui ne sont pas retrouvés dans les QC et pour lesquels on ne pourra pas évaluer le biais analytique.

Les tests statistiques choisis au chapitre 3.1.4. ont été réalisés pour chacun des 198 métabolites. N'ont été conservés dans le Tableau 10 que les métabolites statistiquement significatifs (« *p-value* » < 0.05) pour au moins un des 3 facteurs. Les valeurs de « *q-value* » sont reportées dans le Tableau 10. La dérive analytique a également été évaluée pour chacune des variables en calculant le coefficient de corrélation (Spearman) de l'abondance dans les échantillons de QC en fonction de l'ordre de passage. Comme ces QC correspondent à un seul et même échantillon, une variation de l'intensité au fur et à mesure de l'ordre de passage indique un biais analytique lié à l'encrassement de la source ESI. Les résultats des tests statistiques sont, en cas de corrélation avec l'ordre de passage, à interpréter avec précaution. Comme on peut le voir sur le Tableau 10, cela concerne la majorité des métabolites.

**Tableau 10 : Métabolites statistiquement significatifs pour au moins un des facteurs étudiés avant normalisation.**

(Les métabolites sont classés par significativité en fonction du genre puis de l'âge)

Molécule	Mode d'ionisation	Avant normalisation			Corrélation avec l'ordre de passage - q.value Spearman
		q.value T-test Genre	q.value Spearman Age	q.value Spearman IMC	
Créatine	ESI(+)	<0.0001 (F>H)	ns	ns	ns
Adipoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003
Asp-Leu	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003
	ESI(-)	<0.0001 (H>F)	ns	ns	ns
(δ)Glu-Leu/Ile	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003
	ESI(-)	<0.0001 (H>F)	ns	ns	ns
Tiglylcarnitine / 2-Ethylacrylylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0002
Valérylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0001
Hydroxyoctanoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003

Octanoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	<b>0.0003</b>
Décénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	<b>0.0003</b>
Décénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	<b>0.0002</b>
Diméthylnonenoylcarnitine / Undécénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	<b>0.0003</b>
Décanoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	<b>0.0003</b>
Acide p-Anisique	ESI(-)	<0.0001 (F>H)	ns	ns	ns
Glucuronide de testostérone	ESI(-)	<0.0001 (H>F)	ns	ns	<b>0.0167</b>
Sulfate de déhydroépiandrostérone	ESI(-)	<b>0.0001 (H&gt;F)</b>	<0.0001 ↓	ns	<b>0.0045</b>
Octanoylcarnitine	ESI(+)	<b>0.0001 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
Nicotinamide	ESI(+)	<b>0.0001 (H&gt;F)</b>	ns	ns	ns
Asp-Ile	ESI(+)	<b>0.0002 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
	ESI(-)	<b>0.0004 (H&gt;F)</b>	ns	ns	
Nonénoylcarnitine	ESI(+)	<b>0.0006 (H&gt;F)</b>	ns	ns	<b>0.0062</b>
Acide 2-isopropylmalique	ESI(-)	<b>0.0006 (H&gt;F)</b>	ns	ns	ns
Décadiénylcarnitine	ESI(+)	<b>0.0007 (H&gt;F)</b>	ns	ns	<b>0.0002</b>
Indoleacétyle glutamine	ESI(+)	<b>0.0007 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
Acide 4-Hydroxymandélique	ESI(-)	<b>0.0010 (H&gt;F)</b>	ns	ns	<b>0.0099</b>
2,6 diméthylheptanoyl carnitine	ESI(+)	<b>0.0012 (H&gt;F)</b>	ns	ns	ns
Octénoylcarnitine	ESI(+)	<b>0.0013 (H&gt;F)</b>	ns	<b>0.0274 ↑</b>	ns
Acide Kynurénique	ESI(+)	<b>0.0014 (H&gt;F)</b>	ns	ns	<b>0.0059</b>
	ESI(-)	<b>0.0044 (H&gt;F)</b>	ns	ns	ns
Acide 3-Méthyl-2-oxovalérique	ESI(-)	<b>0.0016 (H&gt;F)</b>	ns	ns	ns
N2-Méthylguanosine	ESI(+)	<b>0.0019 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
Phénylalanine	ESI(+)	<b>0.0040 (H&gt;F)</b>	ns	ns	<b>0.0290</b>
Glucuronide de déhydroépiandrostérone	ESI(-)	<b>0.0042 (H&gt;F)</b>	<0.0001 ↓	ns	ns
Hepténoylcarnitine	ESI(+)	<b>0.0047 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
Acide Subérique	ESI(-)	<b>0.0056 (H&gt;F)</b>	ns	ns	<b>0.0354</b>
N4-Acétilycytidine	ESI(+)	<b>0.0060 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
	ESI(-)	ns	<b>0.0270 ↓</b>	ns	ns
Acide 4-Méthyl-2-oxovalérique (kétoleucine)	ESI(-)	<b>0.0070 (H&gt;F)</b>	ns	ns	
Norleucine	ESI(+)	<b>0.0097 (H&gt;F)</b>	ns	ns	ns
Butyrylcarnitine / Isobutyrylcarnitine	ESI(+)	<b>0.0098 (H&gt;F)</b>	ns	ns	<b>0.0001</b>
Xanthosine	ESI(-)	<b>0.0101 (H&gt;F)</b>	ns	ns	ns
2-méthoxy-N-[3-oxo-3-(pyrrolidin-1- yl)propyl]acétamide (et isomères)	ESI(+)	<b>0.0108 (H&gt;F)</b>	ns	ns	ns
Glucuronide d'acide 3-Indole carboxylique	ESI(-)	<b>0.0115 (H&gt;F)</b>	ns	ns	ns
Porphobilinogène	ESI(+)	<b>0.0125 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
	ESI(-)	<b>0.0159 (H&gt;F)</b>	ns	ns	ns
Diméthylguanosine	ESI(+)	<b>0.0168 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
	ESI(-)	<b>0.0138 (H&gt;F)</b>	ns	ns	ns
Glu-Val	ESI(+)	<b>0.0370 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
	ESI(-)	<b>0.0003 (H&gt;F)</b>	ns	ns	ns
Méthylinosine	ESI(-)	<b>0.0244 (H&gt;F)</b>	ns	ns	ns
Acide Heptylmalonique	ESI(-)	<b>0.0440 (H&gt;F)</b>	ns	ns	ns
3-Méthylglutarylcarnitine	ESI(+)	<b>0.0470 (H&gt;F)</b>	ns	ns	<b>0.0003</b>
Acide Quinique	ESI(-)	ns	<0.0001 ↓	ns	<b>0.0268</b>
Acide 1-Méthylurique	ESI(+)	ns	<0.0001 ↑	ns	<b>0.0027</b>
	ESI(-)	ns	<b>0.0386 ↑</b>	ns	ns
Acide 1,3-Diméthylurique	ESI(+)	ns	<0.0001 ↑	<b>0.0021 ↑</b>	<b>0.0002</b>
	ESI(-)	ns	<b>0.0024 ↑</b>	<b>0.0299 ↑</b>	ns
Acide 1,7-Diméthylurique	ESI(+)	ns	<0.0001 ↑	<b>0.0165 ↑</b>	<b>0.0001</b>

	ESI(-)	ns	<b>0.0033</b> ↑	ns	ns
Paraxanthine	ESI(+)	ns	<b>&lt;0.0001</b> ↑	ns	<b>0.0003</b>
Caféine	ESI(+)	ns	<b>&lt;0.0001</b> ↑	ns	<b>0.0003</b>
N-Acétylphénylalanine	ESI(+)	ns	<b>0.0002</b> ↑	ns	ns
4-Nitrophénol	ESI(+)	ns	<b>0.0011</b> ↑	ns	<b>0.0003</b>
1-Méthylxanthine	ESI(-)	ns	<b>0.0128</b> ↑	ns	ns
Proline	ESI(+)	ns	<b>0.0309</b> ↑	ns	ns
Acide Fumarique	ESI(-)	ns	<b>0.0309</b> ↑	ns	ns
Acide $\alpha$ -hydroxyhippurique	ESI(+)	ns	<b>0.0320</b> ↑	ns	<b>0.0011</b>
Acide Pyrroledicarboxylique	ESI(-)	ns	<b>0.0433</b> ↑	ns	ns

ns = non significatif

58 métabolites ressortent discriminants pour au moins un des 3 facteurs physiologiques étudiés, donc 11 métabolites communs aux deux modes d'ionisation. La plus grande source de variabilité est la différence Hommes vs. Femmes (45 métabolites sur 58), puis l'âge (16 métabolites sur 58). Les variations en fonction de l'IMC ne concernent que 3 métabolites (acide 1,3-diméthylurique, acide 1,7-diméthylurique et octenoylcarnitine), et ces métabolites sont déjà discriminants avec l'âge ou le genre. Il n'y a donc pas de variations spécifiques à l'IMC. Dans tous les cas ces résultats ne sont pas totalement fiables à cause du biais analytique évalué par le coefficient de corrélation avec l'ordre de passage. Une étape de normalisation est donc nécessaire afin de s'affranchir de cette dérive analytique susceptible de biaiser les résultats des tests statistiques.

### 3.3 NORMALISATION

L'un des plus gros problèmes rencontrés lors de l'analyse LC/MS de grandes quantités d'échantillons est la perte de signal liée à l'encrassement de la source ESI. Bien que la randomisation des échantillons biologiques tout au long de la séquence permette de répartir aléatoirement ce biais analytique, si ce dernier est trop important, il peut potentiellement masquer ou augmenter de manière fictive la variabilité biologique que l'on cherche à étudier. Cette variabilité analytique peut facilement être mise en évidence en visualisant l'évolution des intensités des différents métabolites dans les échantillons de QC au fur et à mesure de l'ordre de passage (Figure 47). La dérive analytique (assimilable à la pente de la droite) est nettement observable en mode positif (Figure 47 gauche). Comme dans ce mode d'ionisation tous les échantillons ont été analysés à la suite, cette dérive est monotone (va toujours dans le même sens). Par contre en mode négatif (Figure 47 droite), les échantillons ont été analysés en deux temps, et le biais analytique comprend à la fois la diminution de l'intensité au cours du temps, mais également une grosse différence d'intensité initiale entre les deux séquences.

Ce biais analytique est variable en fonction de la molécule étudiée. Il convient donc d'effectuer une correction des données individuellement pour chaque métabolite. Contrairement à ce qui est représenté Figure 47, le biais analytique n'est pas réellement linéaire. Pour pouvoir normaliser correctement les données il faut dans un premier temps modéliser de manière plus précise cette dérive analytique. Pour chaque métabolite, une régression polynomiale locale de degré 2 (LOWESS

ou *locally weighted scatterplot smoothing* [280]) a été réalisée à partir des intensités des QC. Les abondances réelles sont divisées par la valeur correspondante sur la courbe de régression obtenue afin d'obtenir les abondances normalisées [281]. Cette technique de normalisation par régression locale est très utilisée pour normaliser les données de puces à ADN [282;283].

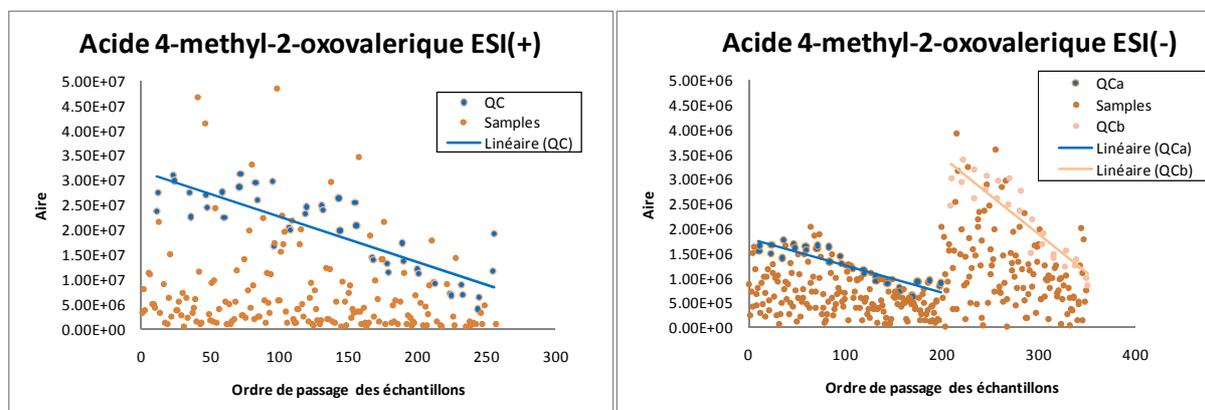


Figure 47 : Evolution des abondances de l'Acide 4-méthyl-2-oxovalérique dans les deux modes d'ionisation,

La régression locale est une méthode de modélisation qui s'appuie sur les méthodes classique de régression, telles que la régression linéaire ou non linéaire (polynomiale par exemple). Elle a été conçue pour répondre aux situations dans lesquelles les méthodes classiques ne fonctionnent pas bien ou ne peuvent pas être appliquées efficacement. La méthode LOWESS combine la simplicité de la régression linéaire avec la souplesse de la régression non linéaire. Le principe est de réaliser une régression locale, c'est-à-dire une pour chaque point du nuage de point, afin de tracer la courbe qui ajuste au mieux le nuage de points. Ainsi, pour chaque point, un polynôme de degré faible (0, 1 ou 2) est calculé afin d'ajuster ce point avec les points environnants grâce à une méthode moindre carrés (avec une pondération d'autant plus faible que le point est éloigné du point de référence). Selon le degré du polynôme, on obtient soit une moyenne mobile pondérée (degré 0), soit régression linéaire locale (degré 1, LOESS) soit une régression quadratique locale (degré 2, LOWESS). Cette méthode est simple dans le sens où il n'est pas nécessaire de spécifier une seule fonction pour tout le modèle et elle s'adapte donc à de très nombreuses formes de nuage de points. Cependant, une seule fonction doit être utilisée pour décrire le modèle. Dans le cas contraire, il est impossible d'obtenir une formule analytique pour décrire le nuage de point. De plus cette méthode demande autant de calculs que de points présents dans le nuage de points, et donc elle requiert donc une capacité de calcul importante.

Appliquée à notre jeu de données, cette méthode a permis de s'affranchir du biais analytique. Sur la Figure 48, les graphiques du haut représentent les données avant normalisation (QC en bleu et échantillons biologiques en orange). La courbe en bleu représente la courbe de régression locale obtenue avec les QC. C'est en utilisant les valeurs de cette courbe que les abondances des ions dans les échantillons (biologiques + QC) ont été normalisées. La courbe alors obtenue après normalisation a la forme d'une droite parallèle à l'axe des abscisses, ce qui signifie que la dérive analytique

(assimilable à la pente de la droite) n'impacte plus les abondances des métabolites du le jeu de données après normalisation. Ne sont alors conservés que les métabolites pour lesquels l'abondance est comprise dans l'intervalle de  $\pm 20\%$  de la moyenne après normalisation (lignes en pointillés rouge sur la Figure 48) pour au moins 2/3 des échantillons de QC. En effet si après cette étape de normalisation la variation de la réponse dans les échantillons de QC est encore trop importante, c'est qu'il existe une autre source de variabilité (ni biologique ni liée à la dérive analytique) qui risque de biaiser les résultats. Cette source de variabilité, si elle n'est pas identifiable, est impossible à évaluer et donc à corriger. Les métabolites dans cette situation (comme la norleucine Figure 48) sont marqués (\*) afin d'éviter une interprétation erronée des résultats statistiques.

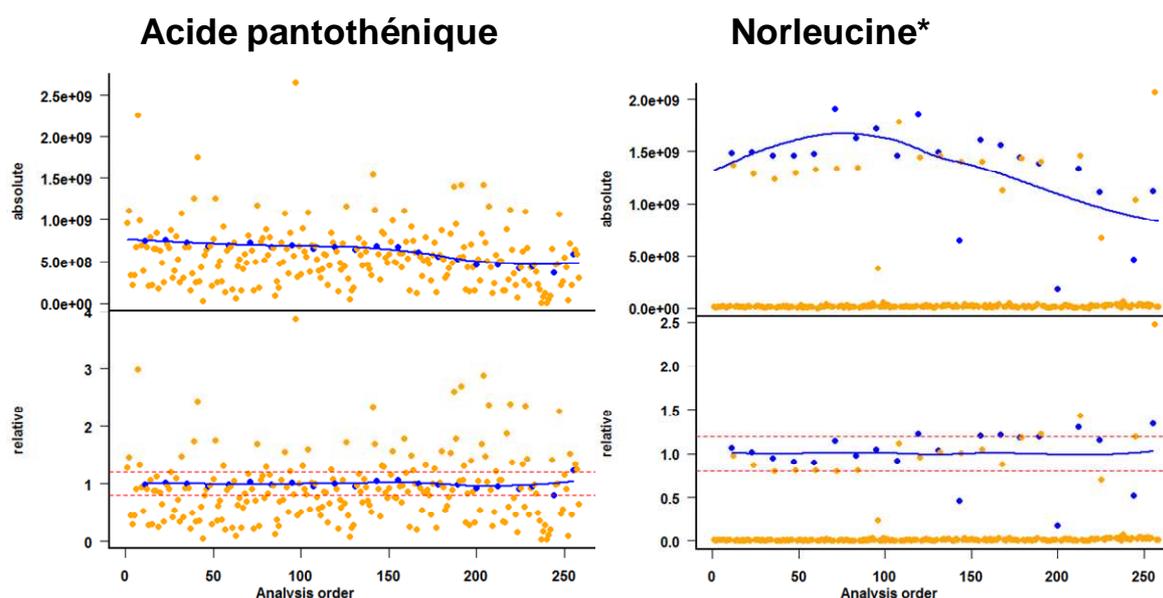


Figure 48 : Evolution des abondances absolues et relatives de plusieurs métabolites en mode d'ionisation positif, avant et après normalisation.

### 3.4 ANALYSES STATISTIQUES SUR LES DONNEES NORMALISEES

Le premier constat est qu'après normalisation, les coefficients de corrélation en fonction de l'ordre de passage sont tous non significatifs (Tableau 11). L'étape de normalisation a donc permis de s'affranchir de la dérive analytique. Cette étape a également corrigé la significativité des métabolites par rapport aux trois facteurs étudiés. Initialement 58 métabolites étaient discriminants, après normalisation il n'en reste que 51. Certains métabolites présentent des « *q-values* » moins significatives, comme la norleucine ou la butyrylcarnitine, alors que d'autres présentent au contraire des « *q-values* » plus significatives, comme l'acide pantothénique ou l'acide aspartique. Le nombre de métabolites discriminants pour chaque facteur avant et après normalisation est représenté Figure 49.

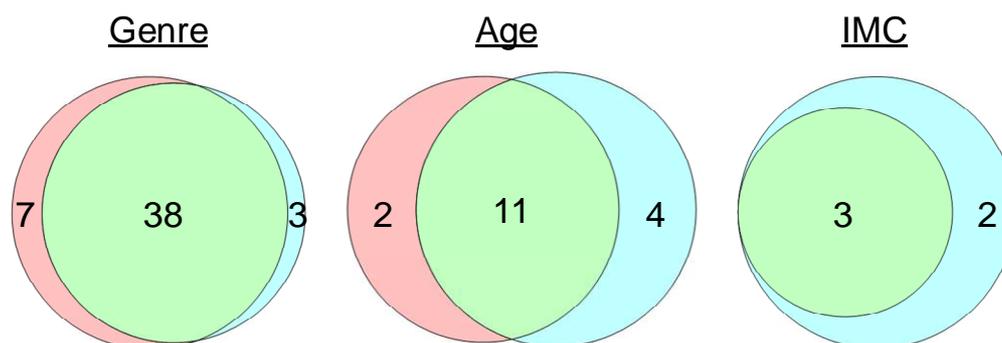


Figure 49 : Métabolites discriminants avant normalisation (rouge) et après normalisation (bleu). Les métabolites communs avant et après normalisation sont représentés en vert.

La perte de significativité est liée à la suppression de la dérive analytique, dont l'effet se superposait à la variabilité biologique. Une fois cette variabilité analytique supprimée, la variabilité résiduelle n'est plus assez importante pour être significative. Dans le cas de la méthylguanosine, de la proline et de la norleucine, la perte de significativité est probablement liée à la forte variabilité de la réponse analytique même après normalisation (ces métabolites sont marqués\* car la variabilité est supérieure à 20% pour plus du tiers des QC). La perte de significativité dépend de deux critères : l'importance de la significativité biologique (*q-value* très faible) et l'impact de la dérive analytique (« *q-value* » très faible). Ainsi, pour la N4-acétylcystéine, la butyrylcarnitine, le porphobilinogène, la diméthylguanosine et la 3-méthylglutarylcarnitine, la significativité biologique était importante (« *q-value* » entre 0.01 et 0.04), mais l'impact de la dérive analytique l'était encore plus (« *q-value* » < 0.001), ce qui a abouti à la perte de la significativité biologique. Par contre, en ce qui concerne les acides fumarique, heptylmalonique et  $\alpha$ -hydroxyhippurique, leur significativité biologique était peu importante, et même sans corrélation avec l'ordre de passage (pour les acides fumarique et heptylmalonique notamment), la suppression de la dérive analytique a aussi supprimé cette significativité biologique. Dans tous les cas, les métabolites très discriminants (« *q-value* » < 0.001) le restent même après normalisation, même si la corrélation avec la dérive analytique est très importante.

A l'inverse, lorsqu'il y a gain de significativité lié à la suppression de la dérive analytique, c'est que l'effet de ce biais analytique masquait la variabilité biologique. Une fois cette variabilité analytique supprimée, la variabilité résultante est assez importante pour être significative. Là encore le gain de significativité dépend de l'importance de la significativité biologique (ou plutôt dans ce cas la non significativité) et de l'impact de la dérive analytique (« *q-value* » très faible). Plusieurs situations sont observées : (i) l'impact de la dérive analytique est très important, et le gain de significativité après normalisation sera lui aussi très important (par exemple pour l'acide pantothénique), (ii) la corrélation avec le biais analytique est faible, et le gain de significativité l'est aussi (acide quinique, acide aspartique), et enfin (iii) bien que la corrélation avec la dérive analytique soit non significative, mais après la normalisation il résulte un gain de significativité plus ou moins important (acide 1-méthylurique, acide 5-hydroxyindolacétique, acide acétylaspartique).

**Tableau 11 : Métabolites statistiquement significatifs pour au moins un des facteurs étudiés après normalisation.**

(Les métabolites sont classés par significativité en fonction du genre puis de l'âge)

Molécule	Mode d'ionisation	Avant normalisation			Corrélation avec l'ordre de passage - q.value Spearman	Après Normalisation LOWESS			Corrélation avec l'ordre de passage - q.value Spearman
		q.value T-test Genre	q.value Spearman Age	q.value Spearman IMC		q.value T-test Genre	q.value Spearman Age	q.value Spearman IMC	
Créatine*	ESI(+)	<0.0001 (F>H)	ns	ns	ns	0.0001 (F>H)	ns	ns	ns
Adipoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	0.0001 (H>F)	ns	ns	ns
Asp-Leu	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	0.0003 (H>F)	ns	ns	ns
	ESI(-)	<0.0001 (H>F)	ns	ns	ns	0.0008 (H>F)	ns	ns	ns
(δ)Glu-Leu/Ile	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	<0.0001 (H>F)	ns	ns	ns
	ESI(-)	<0.0001 (H>F)	ns	ns	ns	<0.0001 (H>F)	ns	ns	ns
Tiglylcarnitine / 2-Ethylacrylylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0002	0.0044 (H>F)	ns	ns	ns
Valérylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0001	<0.0001 (H>F)	ns	ns	ns
Hydroxyoctanoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	<0.0001 (H>F)	ns	ns	ns
Octanoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	0.0003 (H>F)	ns	ns	ns
Décénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	<0.0001 (H>F)	ns	ns	ns
Deécénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0002	0.0001 (H>F)	ns	ns	ns
Diméthylnonenoylcarnitine / Undécénoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	<0.0001 (H>F)	ns	ns	ns
Décánoylcarnitine	ESI(+)	<0.0001 (H>F)	ns	ns	0.0003	<0.0001 (H>F)	ns	ns	ns
Acide p-Anisique	ESI(-)	<0.0001 (F>H)	ns	ns	ns	<0.0001 (F>H)	ns	ns	ns
Glucuronide de testostérone	ESI(-)	<0.0001 (H>F)	ns	ns	0.0167	<0.0001 (H>F)	0.0169 ↑	ns	ns
Sulfate de déhydroépiandrostérone	ESI(-)	0.0001 (H>F)	<0.0001 ↓	ns	0.0045	0.0002 (H>F)	<0.0001 ↓	ns	ns
Octanoylcarnitine	ESI(+)	0.0001 (H>F)	ns	ns	0.0003	0.0009 (H>F)	ns	ns	ns
Nicotinamide*	ESI(+)	0.0001 (H>F)	ns	ns	ns	0.0005 (H>F)	ns	ns	ns
Asp-Ile	ESI(+)	0.0002 (H>F)	ns	ns	0.0003	0.0034 (H>F)	ns	ns	ns
	ESI(-)	0.0004 (H>F)	ns	ns		0.0003 (H>F)	ns	ns	ns
Nonénoylcarnitine	ESI(+)	0.0006 (H>F)	ns	ns	0.0062	0.0016 (H>F)	ns	ns	ns
Acide 2-isopropylmalique	ESI(-)	0.0006 (H>F)	ns	ns	ns	0.0017 (H>F)	ns	ns	ns
Décadiénylcarnitine	ESI(+)	0.0007 (H>F)	ns	ns	0.0002	0.0037 (H>F)	ns	ns	ns
Indoleacétyle glutamine	ESI(+)	0.0007 (H>F)	ns	ns	0.0003	0.0194 (H>F)	ns	ns	ns
Acide 4-Hydroxymandélique	ESI(-)	0.0010 (H>F)	ns	ns	0.0099	0.0016 (H>F)	ns	ns	ns

2,6 diméthylheptanoylcarnitine	ESI(+)	<b>0.0012 (H&gt;F)</b>	ns	ns	ns	<b>0.0042 (H&gt;F)</b>	ns	ns	ns
Octénoylcarnitine	ESI(+)	<b>0.0013 (H&gt;F)</b>	ns	<b>0.0274 ↑</b>	ns	<b>0.0053 (H&gt;F)</b>	ns	<b>0.0341 ↑</b>	ns
Acide Kynurénique	ESI(+)	<b>0.0014 (H&gt;F)</b>	ns	ns	<b>0.0059</b>	<b>0.0159 (H&gt;F)</b>	ns	ns	ns
	ESI(-)	<b>0.0044 (H&gt;F)</b>	ns	ns	ns	<b>0.0080 (H&gt;F)</b>	ns	ns	ns
Acide 3-Méthyl-2-oxovalérique	ESI(-)	<b>0.0016 (H&gt;F)</b>	ns	ns	ns	<b>0.0031 (H&gt;F)</b>	ns	ns	ns
N2-Méthylguanosine*	ESI(+)	<b>0.0019 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
Phénylalanine	ESI(+)	<b>0.0040 (H&gt;F)</b>	ns	ns	<b>0.0290</b>	<b>0.0250 (H&gt;F)</b>	ns	ns	ns
Glucuronide de déhydroépiandrostérone	ESI(-)	<b>0.0042 (H&gt;F)</b>	<b>&lt;0.0001 ↓</b>	ns	ns	<b>0.0044 (H&gt;F)</b>	<b>&lt;0.0001 ↓</b>	ns	ns
Hepténoylcarnitine	ESI(+)	<b>0.0047 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	<b>0.0212 (H&gt;F)</b>	ns	ns	ns
Acide Subérique	ESI(-)	<b>0.0056 (H&gt;F)</b>	ns	ns	<b>0.0354</b>	<b>0.0313 (H&gt;F)</b>	ns	ns	ns
N4-Acétylcytidine	ESI(+)	<b>0.0060 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
	ESI(-)	ns	<b>0.0270 ↓</b>	ns	ns	ns	<b>0.0022 ↓</b>	ns	ns
Acide 4-Méthyl-2-oxovalérique (kétoleucine)	ESI(-)	<b>0.0070 (H&gt;F)</b>	ns	ns		<b>0.0195 (H&gt;F)</b>	ns	ns	ns
Norleucine*	ESI(+)	<b>0.0097 (H&gt;F)</b>	ns	ns	ns	ns	ns	ns	ns
Butyrylcarnitine / Isobutyrylcarnitine	ESI(+)	<b>0.0098 (H&gt;F)</b>	ns	ns	<b>0.0001</b>	ns	ns	ns	ns
Xanthosine	ESI(-)	<b>0.0101 (H&gt;F)</b>	ns	ns	ns	<b>0.0155 (H&gt;F)</b>	ns	ns	ns
2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide (et isomères)*	ESI(+)	<b>0.0108 (H&gt;F)</b>	ns	ns	ns	<b>0.0076 (H&gt;F)</b>	ns	ns	ns
Glucuronide d'acide 3-Indole carboxylique	ESI(-)	<b>0.0115 (H&gt;F)</b>	ns	ns	ns	<b>0.0037 (H&gt;F)</b>	ns	ns	ns
Porphobilinogène	ESI(+)	<b>0.0125 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
	ESI(-)	<b>0.0159 (H&gt;F)</b>	ns	ns	ns	ns	ns	ns	ns
Diméthylguanosine	ESI(+)	<b>0.0168 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
	ESI(-)	<b>0.0138 (H&gt;F)</b>	ns	ns	ns	<b>0.0195 (H&gt;F)</b>	<b>0.0104 ↓</b>	ns	ns
Glu-Val	ESI(+)	<b>0.0370 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
	ESI(-)	<b>0.0003 (H&gt;F)</b>	ns	ns	ns	<b>0.0002 (H&gt;F)</b>	ns	ns	ns
Méthylinosine	ESI(-)	<b>0.0244 (H&gt;F)</b>	ns	ns	ns	<b>0.0439 (H&gt;F)</b>	ns	ns	ns
Acide Heptylmalonique	ESI(-)	<b>0.0440 (H&gt;F)</b>	ns	ns	ns	ns	ns	ns	ns
3-Méthylglutaryl carnitine	ESI(+)	<b>0.0470 (H&gt;F)</b>	ns	ns	<b>0.0003</b>	ns	ns	ns	ns
Acide Quinique	ESI(-)	ns	<b>&lt;0.0001 ↓</b>	ns	<b>0.0268</b>	ns	<b>&lt;0.0001 ↓</b>	<b>0.0417 ↑</b>	ns
Acide 1-Méthylurique	ESI(+)	ns	<b>&lt;0.0001 ↑</b>	ns	<b>0.0027</b>	ns	<b>&lt;0.0001 ↑</b>	ns	ns
	ESI(-)	ns	<b>0.0386 ↑</b>	ns	ns	ns	<b>0.0334 ↑</b>	<b>0.0436 ↑</b>	ns

Acide 1,3-Diméthylurique	ESI(+)	ns	<0.0001 ↑	0.0021 ↑	0.0002	ns	0.0001 ↑	0.0049 ↑	ns
	ESI(-)	ns	0.0024 ↑	0.0299 ↑	ns	ns	0.0070 ↑	0.0018 ↑	ns
Acide 1,7-Diméthylurique	ESI(+)	ns	<0.0001 ↑	0.0165 ↑	0.0001	ns	<0.0001 ↑	0.0301 ↑	ns
	ESI(-)	ns	0.0033 ↑	ns	ns	ns	0.0034 ↑	0.0068 ↑	ns
Paraxanthine	ESI(+)	ns	<0.0001 ↑	ns	0.0003	ns	<0.0001 ↑	ns	ns
Caféine	ESI(+)	ns	<0.0001 ↑	ns	0.0003	ns	<0.0001 ↑	ns	ns
N-Acétylephénylalanine	ESI(+)	ns	0.0002 ↑	ns	ns	ns	0.0019 ↑	ns	ns
4-Nitrophenol	ESI(+)	ns	0.0011 ↑	ns	0.0003	ns	0.0223 ↑	ns	ns
1-Méthylxanthine	ESI(-)	ns	0.0128 ↑	ns	ns	ns	0.0123 ↑	ns	ns
Proline*	ESI(+)	ns	0.0309 ↑	ns	ns	0.0204 (H>F)	ns	ns	ns
Acide Fumarique	ESI(-)	ns	0.0309 ↑	ns	ns	ns	ns	ns	ns
Acide α-hydroxyhippurique	ESI(+)	ns	0.0320 ↑	ns	0.0011	ns	ns	ns	ns
Acide Pyrroledicarboxylique	ESI(-)	ns	0.0433 ↑	ns	ns	ns	0.0384 ↑	ns	ns
Acide 5-Hydroxyindoleacétique	ESI(-)	ns	ns	ns	ns	ns	0.0287 ↓	ns	ns
Acide Aspartique	ESI(-)	ns	ns	ns	0.0054	0.0220 (H>F)	0.0066 ↓	ns	ns
Acide Pantothénique	ESI(+)	ns	ns	ns	0.0003	0.0151 (F>H)	ns	ns	ns
Acide N-Acetylaspartique	ESI(-)	ns	ns	ns	ns	ns	0.0002 ↓	ns	ns
(hydroxyméthyl)pyrrolidine carboxylate de méthyle /(hydroxy)piperidine carboxylate de méthyle	ESI(-)	ns	ns	ns	ns	0.0007 (H>F)	ns	ns	ns

\*métabolites pour lesquels l'abondance est comprise dans l'intervalle de +/-20% de la moyenne pour moins de 2/3 des échantillons de QC.

ns = non significatif

### 3.5 INTERPRÉTATIONS BIOLOGIQUES

Nous ne porterons une interprétation biologique que sur les métabolites dont les variations sont significatives après l'étape de normalisation. Les variations des métabolites marqués d'une étoile (créatine, proline, norleucine, 2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide, N2-méthylguanosine et nicotinamide) ne seront pas interprétées à cause de leur forte variabilité même après normalisation (Tableau 12). Sont également notées, dans le Tableau 12, les informations biologiques relatives à chaque métabolite : présence dans les biofluides et variations (non pathologiques) répertoriées dans la littérature (HMDB). Certaines familles de métabolites sont plus sensibles que d'autres aux variations physiologiques étudiées (Tableau 13), comme par exemple les acylcarnitines qui sont très impactées par le genre, ou certains composés exogènes très sensibles aux variations de l'âge. Les concentrations des dérivés de stéroïdes sont, elles, impactées à la fois par le genre et par l'âge. Au total, sur 47 métabolites discriminants et interprétables, 36 le sont pour le genre, 18 pour l'âge et 5 pour l'IMC.

**Tableau 12 : Métabolites statistiquement significatifs pour au moins un des trois facteurs physiologiques étudiés, hors métabolites marqués\*.**

(Les métabolites sont classés par significativité en fonction du genre puis de l'âge)

Molécule	Voies métabolique concernée (KEGG)	Après Normalisation LOWESS			HMDB		
		q.value T-test Genre	q.value Spearman Age	q.value Spearman BMI	Urines	Sang	Autre
Diméthylnonenoylcarnitine / Undécénoylcarnitine	Catabolisme des lipides	<0.0001 (H>F)	ns	ns	OK [261]	OK	NA
Décénoylcarnitine	Catabolisme des lipides	<0.0001 (H>F)	ns	ns	OK [261]	OK	NA
(δ)Glu-Leu/Ile	Dégradation protéique	<0.0001 (H>F)	ns	ns	OK	NA	NA
		<0.0001 (H>F)	ns	ns			
Valérylcarnitine	Catabolisme des lipides	<0.0001 (H>F)	ns	ns	OK [261]	NA	LCR
Hydroxyoctanoylcarnitine	Catabolisme des lipides	<0.0001 (H>F)	ns	ns	OK [261]	NA	NA
Acide p-Anisique	Composé exogène (graines d'anis)	<0.0001 (F>H)	ns	ns	OK	OK	NA
Glucuronide de testostérone	Synthèse des hormones stéroïdienne	<0.0001 (H>F)	0.0169 ↓	ns	OK	NA	Sperme
Décénoylcarnitine	Catabolisme des lipides	<0.0001 (H>F)	ns	ns	OK [261]	OK	LCR
Décénoylcarnitine	Catabolisme des lipides	0.0001 (H>F)	ns	ns	OK [261]	OK	LCR
Sulfate de déhydroépiandrosterone	Synthèse des hormones stéroïdienne	0.0002 (H>F)	<0.0001 ↓	ns	OK	OK	NA
Glu-Val	Dégradation protéique	0.0002 (H>F)	ns	ns	OK	NA	NA
Asp-Leu	Dégradation protéique	0.0003 (H>F)	ns	ns	OK	NA	NA
		0.0008 (H>F)	ns	ns			
(hydroxyméthyl)pyrrolidine carboxylate de méthyle / (hydroxy)piperidine	NA	0.0007 (H>F)	ns	ns	NA	NA	NA

carboxylate de méthyle							
Asp-Ile	Dégradation protéique	<b>0.0034 (H&gt;F)</b>	ns	ns	OK	NA	NA
		<b>0.0003 (H&gt;F)</b>	ns	ns			
Octanoylcarnitine	Catabolisme des lipides	<b>0.0003 (H&gt;F)</b>	ns	ns	OK [261]	OK	NA
Octanoylcarnitine	Catabolisme des lipides	<b>0.0009 (H&gt;F)</b>	ns	ns	OK [261]	OK	NA
Nonenoylcarnitine	Catabolisme des lipides	<b>0.0016 (H&gt;F)</b>	ns	ns	OK [261]	NA	NA
Acide 4-Hydroxymandélique	Dégradation de l'aminobenzoate	<b>0.0016 (H&gt;F)</b>	ns	ns	OK	NA	NA
Acide 2-isopropylmalique	Métabolisme des acides aminés	<b>0.0017 (H&gt;F)</b>	ns	ns	OK	NA	NA
Acide 3-Méthyl-2-oxovalérique	Métabolisme des acides aminés	<b>0.0031 (H&gt;F)</b>	ns	ns	OK	OK	LCR (F>H)
2,6 diméthylheptanoylcarnitine	Catabolisme des lipides	<b>0.0042 (H&gt;F)</b>	ns	ns	OK [261]	NA	NA
Tiglylcarnitine / 2-Ethylacrylylcarnitine	Catabolisme des lipides	<b>0.0044 (H&gt;F)</b>	ns	ns	OK	OK	LCR
Glucuronide de déhydroépiandrosterone	Synthèse des hormones stéroïdienne	<b>0.0044 (H&gt;F)</b>	<b>&lt;0.0001 ↓</b>	ns	OK	NA	NA
Octénoylcarnitine	Catabolisme des lipides	<b>0.0053 (H&gt;F)</b>	ns	<b>0.0341 ↑</b>	OK [261]	NA	NA
Acide Kynurénique	Métabolisme du Tryptophane	<b>0.0159 (H&gt;F)</b>	ns	ns	OK	OK	LCR
		<b>0.0080 (H&gt;F)</b>	ns	ns			
Acide Pantothénique	Métabolisme des acides aminés	<b>0.0151 (F&gt;H)</b>	ns	ns	H>F et ↓ avec l'âge (0-13 ans)	↓ avec l'âge (0-13 ans)	LCR
Indoleacétyle glutamine	Métabolisme du Tryptophane	<b>0.0194 (H&gt;F)</b>	ns	ns	OK	NA	NA
Acide 4-Méthyl-2-oxovalérique (kétoleucine)	Métabolisme des acides aminés	<b>0.0195 (H&gt;F)</b>	ns	ns	OK	OK	LCR
Diméthylguanosine	Produit de dégradation des ARN de transfert.	<b>0.0195 (H&gt;F)</b>	<b>0.0104 ↓</b>	ns	OK	NA	NA
Hepténoylcarnitine	Catabolisme des lipides	<b>0.0212 (H&gt;F)</b>	ns	ns	OK [261]	NA	NA
Acide Aspartique	Métabolisme des acides aminés	<b>0.0220 (H&gt;F)</b>	<b>0.0066 ↓</b>	ns	↑ avec l'âge (0-1 an)	H>F	LCR (H>F), salive
Phénylalanine	Métabolisme des acides aminés	<b>0.0250 (H&gt;F)</b>	ns	ns	H>F et ↑ avec l'âge (0-18 ans)	H>F et ↓ avec l'âge (0-13 ans)	LCR, salive
Acide Subérique	Métabolisme des acides gras	<b>0.0313 (H&gt;F)</b>	ns	ns	↓ avec l'âge (0-13 ans)	OK	LCR
Méthylinosine	Produit de dégradation des ARN de transfert.	<b>0.0439 (H&gt;F)</b>	ns	ns	OK	OK	NA
Caféine	Métabolisme de la caféine	ns	<b>&lt;0.0001 ↑</b>	ns	OK	OK	NA
Paraxanthine	Métabolisme de la caféine	ns	<b>&lt;0.0001 ↑</b>	ns	OK	OK	NA
Acide Quinique	Composé exogène (alimentation)	ns	<b>&lt;0.0001 ↓</b>	<b>0.0417 ↑</b>	NA	OK	NA
Acide 1-Méthylurique	Métabolisme de la caféine	ns	<b>&lt;0.0001 ↑</b>	ns	OK	NA	NA
		ns	<b>0.0334 ↑</b>	<b>0.0436 ↑</b>			
Acide 1,3-Diméthylurique	Métabolisme de la théophylline	ns	<b>0.0001 ↑</b>	<b>0.0049 ↑</b>	OK	NA	NA
		ns	<b>0.0070 ↑</b>	<b>0.0018 ↑</b>			
Acide 1,7-Diméthylurique	Métabolisme de la caféine	ns	<b>&lt;0.0001 ↑</b>	<b>0.0301 ↑</b>	OK	NA	NA
		ns	<b>0.0034 ↑</b>	<b>0.0068 ↑</b>			
Acide N-Acétylaspartique	Métabolisme des acides aminés	ns	<b>0.0002 ↓</b>	ns	F>H et ↓ avec l'âge (0-	OK	LCR

					18 ans)		
N-Acétylphénylalanine	Métabolisme des acides aminés	ns	<b>0.0019</b> ↑	ns	OK	NA	NA
N4-Acétilycytidine	Produit de dégradation des ARN de transfert.	ns	<b>0.0022</b> ↓	ns	OK [284]	NA	NA
1-Méthylxanthine	Caffeine, Theophylline, Theobromine Metabolism	ns	<b>0.0123</b> ↑	ns	OK	NA	NA
4-Nitrophénol	Marqueur d'exposition aux pesticides.	ns	<b>0.0223</b> ↑	ns	OK	NA	NA
Acide 5-Hydroxyindoleacétique	Métabolisme du Tryptophane	ns	<b>0.0287</b> ↓	ns	↑ avec l'âge (0-13 ans)	OK	LCR
Acide Pyrroledicarboxylique	NA	ns	<b>0.0384</b> ↑	ns	NA	NA	NA

- ns = non significatif

**Tableau 13 : Nombres de métabolites discriminants pour chaque facteurs en fonction de leur famille chimique.**

(et % par rapport au nombre total de métabolites retrouvés pour chaque famille)

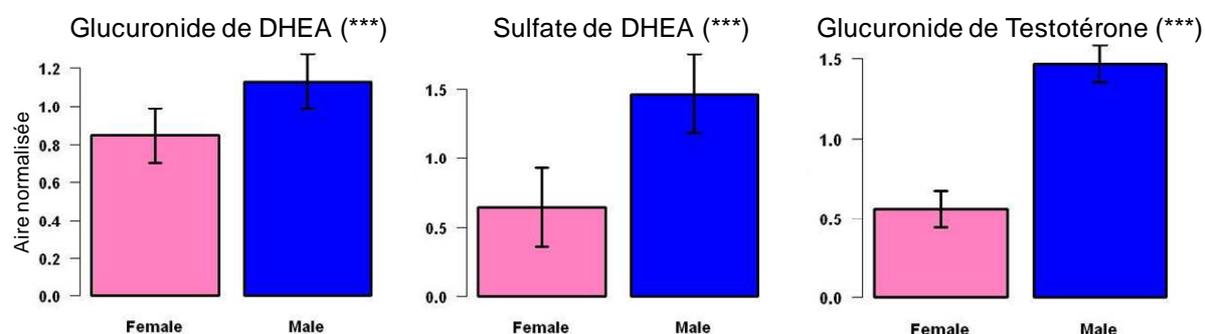
Famille de métabolites	TOTAL	Genre	Age	IMC
Acides organiques et dérivés	64	4 (6%)	1 (2%)	
Acides aminés et dérivés	43	6 (14%)	4 (9%)	
Acylcarnitines	28	13 (46%)		1 (4%)
Amines biogènes et polyamines	21			
Nucléosides et dérivés	19	2 (11%)	2 (11%)	
Composés exogènes et leurs métabolites	18	1 (6%)	8 (44%)	4 (22%)
Acylglycines	10			
Alcools et polyols	10	1 (10%)		
Sucres et dérivés	7			
Dipeptides	7	4 (57%)		
Glucuronides endogènes	5			
Stéroïdes et dérivés	5	3 (60%)	3 (60%)	
Indoles et dérivés	3			
Acides biliaires	2			
Porphyrines	2			
<b>TOTAL</b>	<b>244</b>	<b>34</b>	<b>18</b>	<b>5</b>

### 3.5.1 Variations liées à l'effet du genre

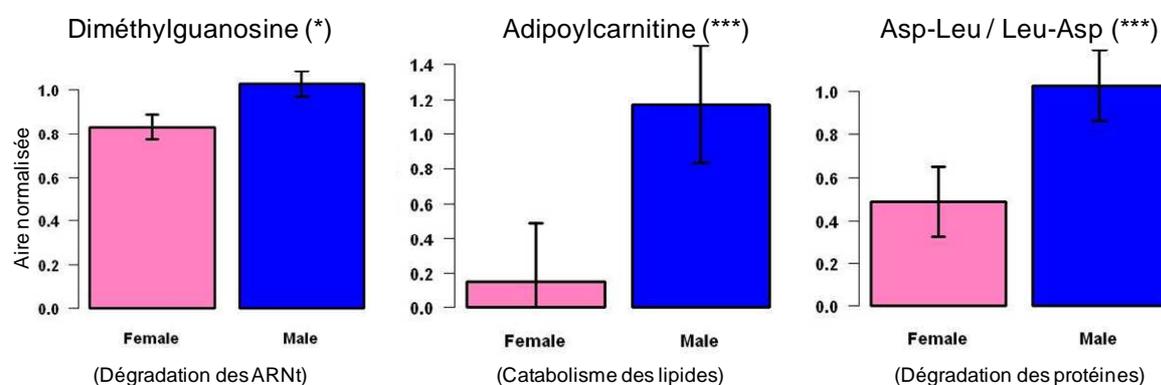
La source de variabilité la plus importante est la différence Homme/Femme. Certaines familles de métabolites sont plus impactées que d'autre par cette différence. Citons par exemple les hormones stéroïdiennes comme les dérivés de testostérone et de DHEA. La DHEA est connue comme étant en concentrations plus élevées chez l'homme [285], et cette différence est répertoriée dans HMDB. Il est donc également normal que ses dérivés, et notamment sa forme sulfatée qui est la forme circulante majoritaire, le soient aussi. Cependant pour le glucuronide et le sulfate de DHEA, aucune variation avec le genre n'est reportée dans HMDB. Il en va de même pour la testotérone [286] (Figure 50). Les métabolites du métabolisme énergétique sont eux aussi fortement impactés, que cela soit les métabolites du catabolisme des ARNt (dérivés nucléosidiques), des lipides (carnitines) ou des protéines (dipeptides). Ces résultats confirment que le catabolisme général est plus important chez

l'Homme que chez la Femme [287-289] (Figure 50). Là encore, pour ces métabolites, aucune variation entre Homme et Femme n'est répertoriée dans HMDB.

### **Hormones stéroïdiennes**



### **Métabolisme énergétique**



**Figure 50 : Variations des abondances des hormones stéroïdiennes et des métabolites du métabolisme énergétique entre homme et femme (après normalisation).**

Les barres d'erreur représentent l'intervalle de confiance de la moyenne, et les étoiles le niveau de significativité (\*\*\*)  $q$ -value < 0.0001, \*\*  $q$ -value entre 0.001 et 0.0001, \*  $q$ -value entre 0.05 et 0.001)

Le métabolisme des acides aminés, et particulièrement celui du tryptophane, sont aussi impactés par cette différence Homme/Femme. Les variations des métabolites du métabolisme des acides aminés vont aussi dans le sens d'une augmentation du métabolisme chez l'Homme par rapport à la Femme. Cependant, à l'exception de la Phénylalanine et de l'acide pantothénique, ces variations ne sont pas non plus répertoriées dans HMDB. Ce dernier est censé être plus élevé chez l'Homme que chez la Femme mais la variation observée dans cette étude est inverse (Figure 51). Le métabolisme du tryptophane et lui aussi plus important chez l'Homme, bien que cette différence ne soit pas reportée dans HMDB (ni pour le tryptophane, ni pour ces métabolites). Reste ensuite quelques métabolites du métabolisme des purines (xanthosine), des acides gras (acide subérique) et plus surprenant, des métabolites de l'anis (acide p-anisique) ou de la dégradation de l'acide aminobenzoïque (acide 4-hydroxymandélique). Pour ces métabolites là encore aucune variation liée au genre n'est disponible dans HMDB. D'autres n'ont pas de voie métabolique attribuée car ils ne sont pas décrits dans KEGG et non présents dans HMDB (acide heptylmalonique et (hydroxyméthyl)pyrrolidine carboxylate de méthyle / (hydroxy)piperidine carboxylate de méthyle).

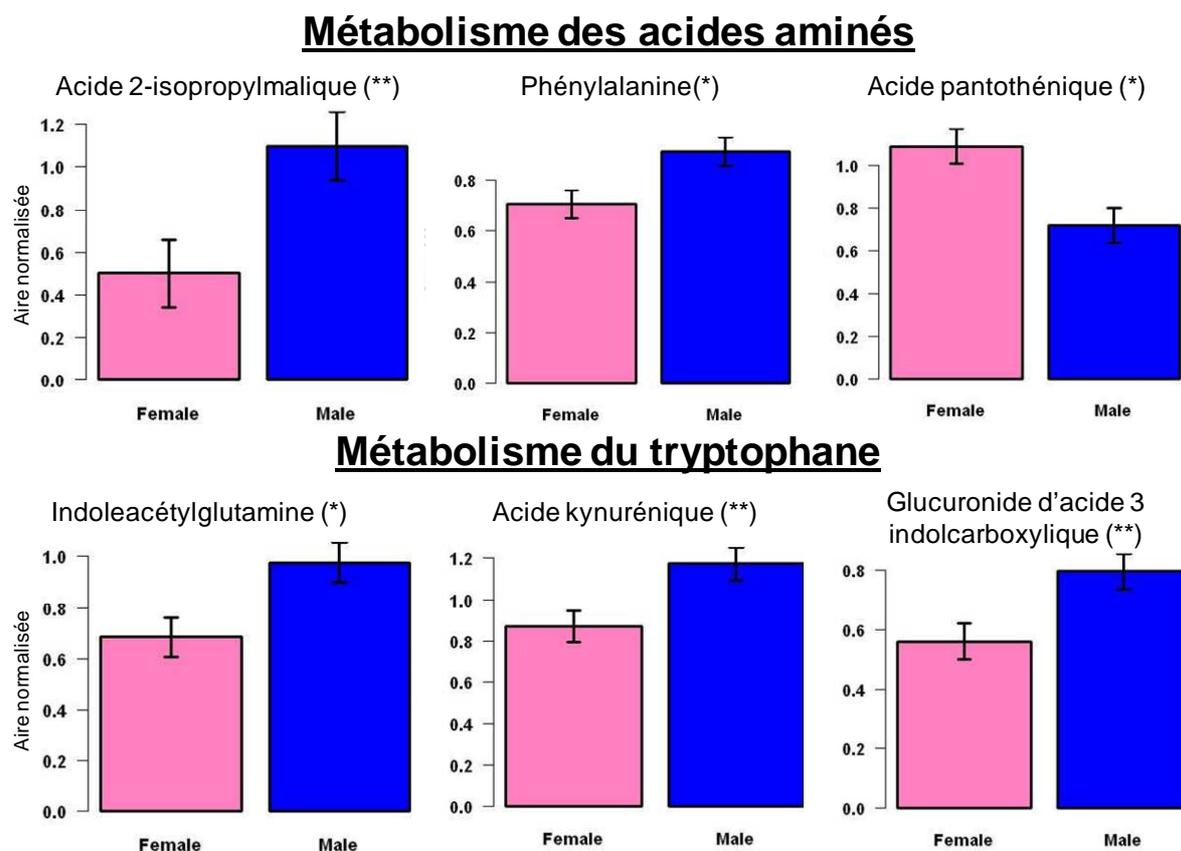


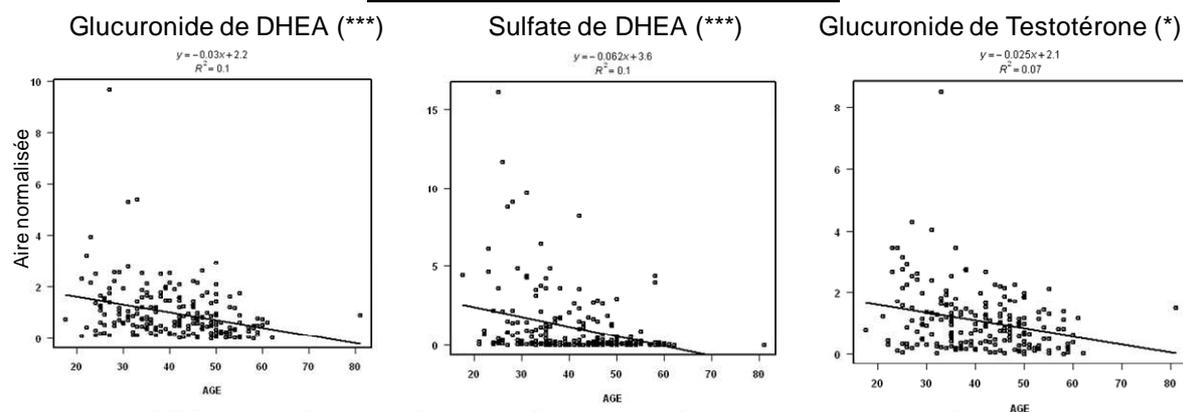
Figure 51 : Variations des abondances des métabolites du métabolisme des acides aminés et du tryptophane entre homme et femme (après normalisation)

Les barres d'erreur représente l'intervalle de confiance de la moyenne, et les étoiles le niveau de significativité (\*\*\*)  $q$ -value < 0.0001, \*\*  $q$ -value entre 0.001 et 0.0001, \*  $q$ -value entre 0.05 et 0.001)

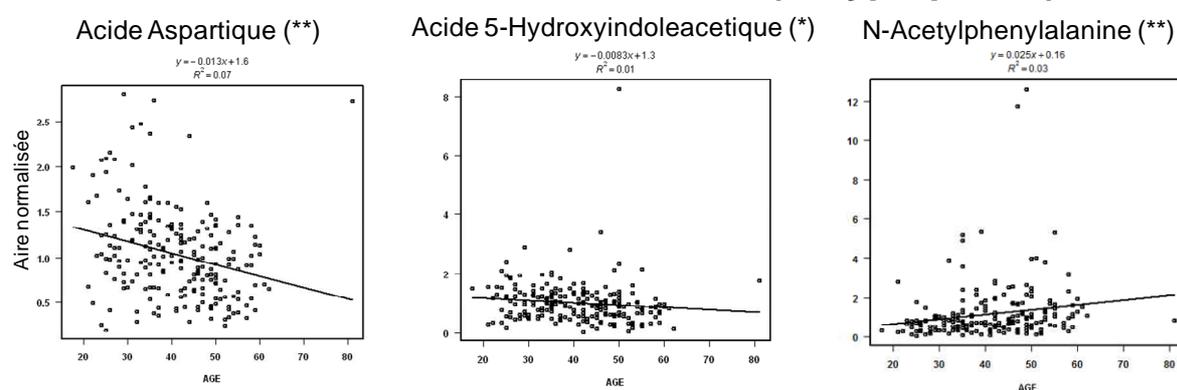
### 3.5.2 Variations liées à l'effet de l'âge

L'âge est la deuxième source de variabilité biologique après le genre. On retrouve notamment, dans les familles de métabolites impactées par l'âge, les hormones stéroïdiennes, qui étaient déjà impactés par le genre. L'âge et le genre étant des facteurs indépendants, les variations observées ici sont elle aussi indépendantes (Figure 52). Les concentrations d'hormones stéroïdiennes baissent avec l'âge et peuvent être expliquées par une diminution de leur production [290;291]. Le métabolisme des acides aminés et lui aussi impacté par l'âge (Figure 52) mais cela concerne des métabolites différents de ceux déjà discriminants entre hommes et femmes (à l'exception de l'acide aspartique). Les acides aminés et dérivés diminuent avec l'âge, à l'exception de la N-acétylphénylalanine. Il y a donc une diminution globale du métabolisme avec l'âge. D'autres métabolites endogènes diminuent avec l'âge comme la N4-acétylcytidine et la diméthylguanosine (qui sont des produits de dégradation des ARNt) et étaient déjà discriminants avec le genre. Enfin l'acide pyrroledicarboxylique diminue avec l'âge mais son rôle biologique reste inconnu.

## Hormones stéroïdiennes



## Métabolisme des acides aminés (+ tryptophane)



**Figure 52 : Variations des abondances des hormones stéroïdiennes et des métabolites du métabolisme des acides aminés avec l'âge (après normalisation).**

Les étoiles représentent le niveau de significativité (\*\*\*) q-value < 0.0001, \*\* q-value entre 0.001 et 0.0001, \* q-value entre 0.05 et 0.001)

Les autres métabolites dont les concentrations évoluent avec l'âge sont tous des métabolites d'origine exogène. La voie métabolique la plus touchée et celle de la métabolisation de la caféine. On retrouve les principaux métabolites de la caféine (paraxanthine, 1-méthylxanthine, acides 1,3-méthylurique, 1,7-méthylurique et 1-méthylurique), sauf la théophylline, mais celle-ci ayant des temps de rétention ( $C_{18}$  et PFPP) très proches de la paraxanthine, soit elle a subi un effet de suppression d'ionisation, vu la intensité importante du signal de la paraxanthine, soit les deux pics chromatographiques se superposent. Tous les métabolites retrouvés subissent une forte augmentation avec l'augmentation de l'âge. On sait que la demi-vie de la caféine augmente dans certaines conditions physiologiques ou médicales : grossesse [294], prise de contraceptifs [292], prise de médicaments [293]. Cette augmentation de la concentration en caféine avec l'âge indique une augmentation de sa demi-vie, qui peut être expliquée par une diminution de l'activité des cytochromes P450 (notamment le CYP1A2) avec l'âge [295;296]. La métabolisation de la caféine étant moins rapide, elle demeure plus longtemps dans l'organisme et il peut y avoir un effet cumulatif pour des prises régulières de café. D'autres composés exogènes sont également impactés par l'âge : diminution de l'acide quinique (issu de l'alimentation) et augmentation du 4-nitrophénol (connu comme marqueur d'exposition à certains xénobiotiques, dont les pesticides).

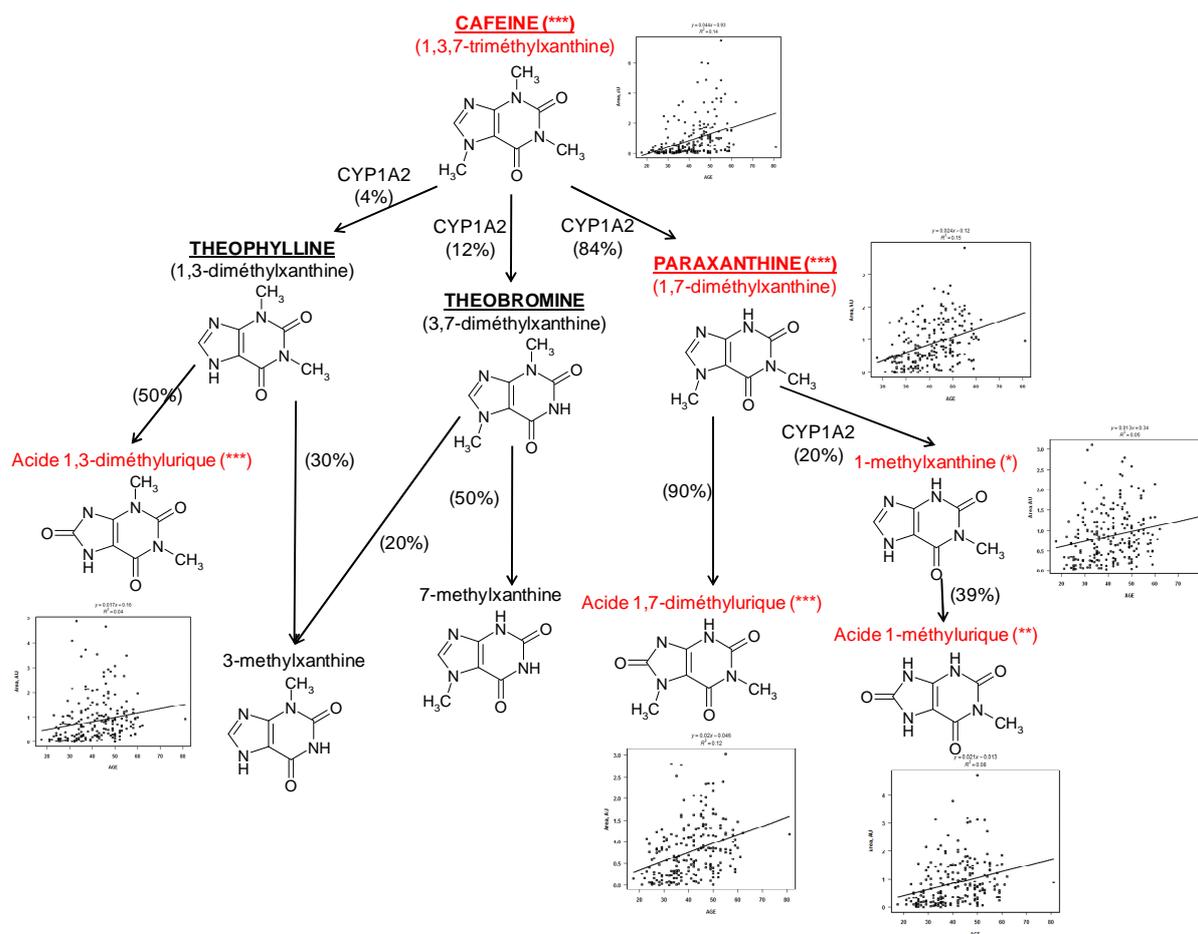


Figure 53 : Métabolisme de la caféine et variations des abondances de ces métabolites avec l'âge (après normalisation)

### 3.5.3 Variations liées à l'effet de l'IMC

Les variations avec l'IMC sont moins nombreuses et jamais spécifiques de ce paramètre. Les métabolites impactés par l'IMC le sont également soit par le genre soit par l'âge. Comme l'IMC est dépendant de ces deux facteurs, on ne peut pas dire si elles sont effectivement liées à l'IMC ou si elles sont artéfactuelles (les « *q-value* » étant quasiment toute plus faibles pour l'IMC que pour l'âge ou le genre). Les métabolites concernés sont l'acide 1,7-diméthylurique, l'acide 1,3-diméthylurique, l'acide 1-méthylurique, l'acide quinique et l'octenoylcarnitine.

## 4 CONCLUSION ET DISCUSSION

Depuis l'arrêt de la publication des tables Geigy [277], il n'existe plus de banque ou base de données spécifiquement dédiées aux concentrations des métabolites dans les fluides biologiques humains, et il n'existe que très peu d'études exhaustives dédiées à l'étude des variations physiologiques des molécules retrouvées dans ces milieux biologiques. Tirant avantage de la construction d'une base de données spectrales, nous avons étudié l'influence de facteurs tels que l'âge, le genre et l'IMC sur les concentrations urinaires de métabolites afin d'initier la construction

d'une base de données relationnelle de métabolites dans les biofluides humains. L'étude préliminaire de la population a permis de mettre en évidence les interactions entre certains facteurs étudiés (Genre/IMC et Age/IMC) mais également de choisir les tests statistiques appropriés. Ne connaissant pas à priori distribution sous-jacente des données, les tests appliqués sont non paramétriques. De plus il a été décidé de réaliser des tests indépendants pour étudier un par un chaque facteur physiologique en raison de la dépendance existantes entre certains facteurs. Les analyses statistiques ont été réalisées avant et après normalisation. Cette étape a permis de s'affranchir du biais analytique et de conforter les résultats biologiques.

Sur les 244 métabolites identifiés, 51 présentent des variations de concentrations significatives, dont 45 sont interprétables après normalisation, pour au moins un des trois facteurs physiologiques étudiés, ce qui représente un peu moins de 20% des métabolites identifiés. De plus, la base de données HMDB ne répertorie des données d'évolution de concentrations dans un fluide biologique avec l'âge ou le genre que pour 6 d'entre eux (Tableau 12). Ainsi, selon HMDB, les concentrations urinaires d'acide N-acétylaspartique, d'acide pantothénique et de phénylalanine varient significativement en fonction du genre, mais aussi de l'âge (entre 0 et 18 ans), et les concentrations des acides aspartique, subérique et 5-hydroxyindoleacétique ne varient significativement qu'avec l'âge (entre 0 et 18 ans, ce qui ne correspond pas à la tranche d'âge de notre étude). Finalement, étant donné que les variations de concentrations selon l'âge répertoriées par HMDB ne sont pas comparables aux nôtres et les données obtenus pour l'acide pathoténique étant en désaccord avec celles d'HMDB, seules les concentrations urinaires de phénylalanine présentent une évolution de concentrations avec le genre comparable à celles de notre étude. Notons cependant qu'il est possible d'obtenir indirectement des données sur certains métabolites tels que les dérivés de stéroïdes. En effet, ceux-ci (la testostérone et la DHEA, par exemple) sont détectés à l'état de conjugués, alors que les concentrations dans les biofluides, souvent obtenues par analyses GC/MS après hydrolyse de ces conjugués, sont répertoriées au niveau des aglycones correspondants dans HMDB. Finalement, cette confrontation des données que nous avons générées à celles qui sont répertoriées dans HMDB (une des principales banques de données de métabolomique) confirme que nous manquons de données physiologiques de concentrations de métabolites obtenues à partir de cohortes d'individus sains, ou tout venant qui pourraient permettre de faciliter la recherche de biomarqueurs de diagnostic et de suivi de pathologies.

La source majeure de variation de concentrations de métabolites est liée au genre, et elle concerne principalement des métabolites endogènes, tels que les hormones, les acides aminés, ou encore les dérivés lipidiques, confirmant ainsi les différences déjà observées dans le métabolisme général entre hommes et femmes [287-289]. La deuxième source de variabilité est liée à l'âge. Elle concerne à la fois des métabolites endogènes (diminution du métabolisme général avec l'âge) et exogènes (différences de métabolisation de ces composés avec l'âge). Enfin, aucune variation de concentration de métabolites spécifiquement liée à l'IMC n'a pu être rapportée, ce qui est étonnant car la masse corporelle est connue pour avoir un impact à la fois sur le métabolisme général et la métabolisation de composés exogène (augmentation du volume de distribution...). L'IMC étant dépendant du genre et de l'âge, l'absence de métabolites discriminants n'est pas incohérente car tout

métabolite discriminant avec le genre ou l'âge peut, à cause de cette dépendance, l'être aussi avec l'IMC mais de manière non spécifique. D'autre part, l'existence d'une interaction entre deux ou plusieurs facteurs vis-à-vis d'un métabolite donné est en soit une information biologique que nous prévoyons d'étudier à l'aide d'un modèle général d'ANOVA, mais tout en tenant compte du fait que certains facteurs sont déjà liés entre eux (IMC/âge et IMC/genre).

Nous avons également essayé de comparer nos résultats à ceux issue d'une publication traitant de l'impact de facteurs physiologiques sur les concentrations de métabolites plasmatiques [53], avec peu de succès. Tout d'abord les concentrations plasmatiques et urinaires de métabolites ne semblent pas influencées par les mêmes facteurs physiologiques. En effet, les variations de concentrations plasmatiques des métabolites sont principalement liées à l'âge, puis au genre, alors que les concentrations urinaires de métabolites détectés dans notre étude sont principalement influencées par le genre, puis dans un second temps par l'âge des individus. Dans les deux cas, l'IMC semble avoir un impact sur les taux de certains métabolites, mais de manière modérée, à cause de son lien étroit avec le genre et l'âge. Ensuite, bien qu'une grande partie des métabolites détectés dans le plasma [53] le soient aussi dans notre étude (acides aminés, acides organiques etc...), il y a peu de concordance entre les concentrations de métabolites impactées par le genre ou l'âge dans les deux études. Concernant les variations de concentrations liées au genre, seules deux sont communes aux deux études (phénylalanine et sulfate de DHEA), et seulement 3 métabolites (la caféine, la paraxanthine, et le sulfate de DHEA) présentent le même type d'évolution de concentrations liées à l'âge. Cette étude comparative décevante suggère (i) que les concentrations de métabolites ne sont pas forcément comparables d'un type de biofluide à un autre et/ou (ii) qu'il peut y avoir une grande disparité de résultats d'une cohorte à une autre.

Il a été nécessaire, pour mener à bien cette étude, de mettre au point une méthode de normalisation performante afin de s'affranchir de la dérive analytique souvent observée en LC/MS lors de l'étude d'un grand nombre d'échantillons. Le principe d'utiliser des échantillons urinaires de contrôle de qualité (ou QC) analysés tout au long de la séquence pour évaluer de biais analytique a été suggéré récemment [281]. Cette approche est en fait utilisée depuis longtemps dans les laboratoires de métabolomique, mais les données ne demeurent la plupart du temps interprétées qu'à un niveau chimométrique [297] : il s'agit par exemple de montrer que les QC sont regroupés sur une carte factorielle d'analyse en composantes principales. Grâce au travail d'identification de métabolite préalablement réalisé, nous avons maintenant la possibilité d'utiliser ces échantillons pour suivre l'évolution des concentrations de chaque métabolite identifié pendant le déroulement de l'expérience. Il est alors possible de normaliser les abondances de ces métabolites pour s'affranchir des biais analytiques.

La méthode de normalisation (LOWESS) qui a été récemment préconisée par Dunn et coll. [281] et dont nous nous sommes servis, est déjà utilisée pour traiter des jeux de données d'analyse transcriptomique, mais son application à un jeu de données métabolomique n'a pas encore été publiée. En utilisant des QC identiques d'une expérience à l'autre (en mode négatif), nous avons pu grâce à la méthode de normalisation, nous affranchir non seulement de la dérive analytique au sein

d'une expérience donnée, mais également des variations de performance de l'ionisation electrospray d'une expérience à une autre, ce qui ouvre des perspectives très intéressantes en matière de fusion de jeux de données. En effet, il devrait être possible de regrouper plusieurs expériences différentes au sein d'un même jeu de données en exprimant chacune des valeurs individuelles de concentration des métabolites par rapport aux valeurs trouvées dans les QC de l'expérience correspondante. Nous envisageons donc de constituer des échantillons de QC pour chaque matrice biologique étudiée au laboratoire. Ces échantillons de QC seront aliquotés puis conservés à  $-80^{\circ}\text{C}$ . La limitation principale de l'approche est la stabilité à long terme des analytes dans l'extrait biologique, qui est pour le moment inconnue. Pour l'étudier, il faudrait pouvoir s'affranchir du biais analytique inter-expérimental. L'utilisation de méthodes de quantification absolue à large échelle, grâce à l'utilisation de métabolites marqués, pourrait être une solution, qui est cependant compliquée et coûteuse à mettre en œuvre. Enfin, la méthode LOWESS ne modifie que peu les résultats des tests statistiques de cette étude, ce qui suggère que l'impact des facteurs physiologiques étudiés reste prépondérant par rapport au biais analytique, et cela même dans une population tout venant.

Enfin, là encore se pose le problème du partage des informations au niveau de bases de données telles que HMDB : Les variations physiologiques répertoriées dans HMDB le sont sous forme de résultats de dosages quantitatifs, or dans les études métabolomiques non ciblées, les résultats sont exprimés sous forme relative ( $H>F$ , par exemple) ou sous forme de rapport d'intensité ( $H/F$ , par exemple). De plus certains métabolites d'intérêt ne sont que putativement identifiés, et non pas rigoureusement identifiés. Il n'est donc pas possible, en l'état actuel des choses, de pouvoir y enregistrer les données obtenues dans cette étude, à moins de pouvoir réaliser des analyses quantitatives ciblées, et de pouvoir confirmer l'identification putative des métabolites d'intérêt.



CHAPITRE 3 : CONDITIONS DE  
PRELEVEMENT ET DE STOCKAGE A  
COURT TERME DES URINES ET LEUR  
IMPACT SUR LE METABOLOME URINAIRE



# CHAPITRE 3 : CONDITIONS DE PRELEVEMENT ET

# DE STOCKAGE A COURT TERME DES URINES ET

# LEUR IMPACT SUR LE METABOLOME URINAIRE

## **1 INTRODUCTION**

L'utilisation d'échantillons urinaires présente plusieurs intérêts en toxicologie : le recueil des urines est non invasif et, du fait du phénomène de concentration rénale, les taux de métabolites circulants sont amplifiés dans les urines. De plus, le milieu urinaire contient une grande diversité de composés différents tels que des acides aminés, acides organiques, lipides, sucres, hormones, peptides ou encore des produits finaux du métabolisme tels que les glucuro-conjugués et les sulfo-conjugués. Enfin, il est possible de réaliser des prélèvements ponctuels ou des recueils sur une période de temps déterminée pouvant aller jusqu'à 24h ou 48h. Cependant l'urine est un fluide biologique complexe dont les conditions de prélèvement et de stockage peuvent modifier la composition métabolique (contamination bactérienne, dégradation des métabolites...), en particulier quand les recueils réalisés dépassent plusieurs heures.

La stabilité des biomolécules dans les urines humaines a déjà été étudiée dans diverses conditions. Cependant, nous ne disposons que de quelques publications concernant les potentielles variations des métabolites (Tableau 14). Il a été ainsi démontré que les empreintes protéiques et métaboliques des urines peuvent être affectées par les conditions de prélèvement et de stockage des urines. La conservation à long et moyen terme des urines semble satisfaisante sur une période de 6 mois à conditions que la température soit inférieure ou égale à -20°C [298;299]. Par contre, la stabilité à court et moyen terme à des températures plus élevées est plus problématique, et ceci particulièrement pour les recueils d'urines sur plusieurs heures (prélèvement de 24h, cages à métabolisme pour les études animales, par exemple). Il en va de même pour la conservation d'échantillons traités dans l'injecteur automatique d'un système chromatographique (centrifugés, dilués...) avant analyse. Si la stabilité à court terme des échantillons traités est bien démontrée sur une période de 48h à 4°C (et 24H à température ambiante) [298-302], la stabilité des métabolites urinaires lors du prélèvement n'est pas clairement établie (une seule étude en protéomique par analyse MALDI-MS et électrophorèse [302]). C'est pourtant lors du recueil des échantillons d'urines que la contamination bactérienne, connue pour influer sur les concentrations de certains métabolites, est susceptible d'être la plus importante. En effet, l'urine bien qu'initialement stérile (sauf conditions pathologiques), peut être contaminée par le biais de l'urètre, qui peut contenir des germes commensaux alors emmenés par le flot urinaire. Les urines étant un bon milieu de culture, la

prolifération des bactéries y est très rapide. Celles-ci modifient alors ce milieu de culture en consommant et en produisant des métabolites.

**Tableau 14 : Différentes études déjà publiées sur la conservation des urines à court, moyen et long terme**

	Type d'échantillons	Prélèvement (pas de préparation d'échantillon)	Conservation à court terme (échantillon traités)	Conservation à moyen terme	Conservation à long terme	Méthodes
		0-48h	0-48h	2 jours à 4 semaines	>1 mois	
Gika J.Chrom. A 2008 [298]	Urines individuelles		Urines stables 48H, sans conservateur, à 4°C	Urines non stables, sans conservateur, 6 jours à 4°C	Urines stables 6 mois, sans conservateur, à -20 et -80°C	Métabolomique par LC/MS
Gika J. Prot. Res. 2007 [300]	Urines individuelles		Urines stables 20H, sans conservateur, à 4°C	Urines stables 1 mois, sans conservateur, à -20 et -80°C		Métabolomique par LC/MS
Thongboonkerd J. Prot. Res. 2007 [302]	Pool d'urines	Urines stables 24H et 16h, avec et sans conservateur, à 4°C (à RT = 12H)	Urines stables 48H, avec et sans conservateur, à 4°C (à RT = 24H)			Protéomique par MALDI et électrophorèse
Lauridsen Anal. Chem. 2007 [299]	Urines individuelles		Urines stables 48H, sans conservateur, à 4°C	Urines stables 1 semaine, avec conservateur, à 4°C	Urines stables 6 mois, avec ou sans conservateur, à -25 et -80°C (pas à 4°C)	Métabolomique par RMN
Saude Met. 2007 [303]	Urines individuelles			Urines stables 1 mois, sans conservateur, à -80°C (pas à 4°C ou RT)		Métabolomique par RMN
Maher Anal. Chem. 2007 [301]	Urines individuelles		Urines stables 24H, sans conservateur, à 4°C ou RT		Urines stables 3 mois, sans conservateurs, à -80, -40 et -20°C (pas à 4°C ou RT)	Métabolomique par RMN

Les différentes études publiées (Tableau 14) sont plutôt des études qualitatives basées sur des approches chimiométriques ou sur la comparaison visuelle de spectres. L'utilisation, dans la plupart des cas, d'urines individuelles limite la mise en évidence de possibles modifications des empreintes métaboliques, la variabilité inter-individuelle étant bien supérieure à celle générée par les problèmes de stabilité. Enfin, l'étude de l'impact de ces différentes conditions de conservation sur les concentrations en métabolites se limite à chaque fois aux métabolites déjà connus pour être directement liés à la prolifération bactérienne.

Dans ce contexte, le but de cette étude est de mettre en évidence les variations liées aux différentes conditions de prélèvement des urines, notamment celles liées à la contamination bactérienne des urines. Afin de limiter l'effet de la variabilité individuelle sur les résultats, nous avons travaillé sur des mélanges d'urines provenant de plusieurs individus. Ainsi lors de l'analyse statistique

multivariée des empreintes métaboliques, les seules variations observées seront liées à la variabilité analytique (de laquelle on ne peut pas s'affranchir) et la variabilité biologique, c'est-à-dire l'impact des conditions de stockage et du temps de prélèvement. Les urines ne sont pas traitées avant expérience afin de rester cohérent avec les conditions réelles de recueil des urines (pas de centrifugation ou de dilution). Concernant les conditions de conservation, deux facteurs principaux ont été testés : l'utilisation ou non de conservateurs et la température. Ces conditions ont été choisies pour être compatibles avec les conditions réelles de prélèvement des urines, comme par exemple la température qui peut être la température ambiante, ou une température de 0-4°C facilement obtenue en réalisant le prélèvement sur glace. Plusieurs expériences ont été réalisées (Janvier, Juillet et Septembre 2010, respectivement expériences 1, 2 et 3), dans des conditions de prélèvement différentes afin de tester la reproductibilité des résultats. Les prélèvements ont été effectués à intervalles réguliers (toutes les 4h pour la première expérience, toutes les 12/24h pour les deux autres) pendant 48h à 72h. Un recueil urinaire ne dépasse pas les 48h, mais pour être sûr d'observer des variations significatives [236;237;240], des prélèvements ont été effectués jusqu'à 72h. La présence de bactéries ainsi que leur croissance ont été suivies par turbidimétrie, c'est-à-dire à la mesure du degré de turbidité de la suspension dans le but d'effectuer une numération bactérienne totale. Cette mesure de turbidité peut être considérée comme une simple mesure d'absorbance et est donc effectuée avec un spectrophotomètre à 620nm. Les empreintes métaboliques quant à elles ont été obtenues par LC/MS (voir chapitre 1) et certains échantillons ont été analysés par RMN.

## **2 MATERIEL ET METHODES**

### **2.1 REACTIFS ET PRODUITS CHIMIQUES**

Les réactifs et produits chimiques sont les mêmes que ceux présentées dans le Chapitre 1 (page 61). Les conservateurs (acide borique et azide de sodium), le PBS (*Phosphate buffered saline*) et le TPS (Trimethylsilyl propionate) utilisés pour ces expériences ont été fournis par la société Sigma (Saint Quentin Fallavier, France).

### **2.2 ECHANTILLONS BIOLOGIQUES**

Les échantillons d'urine ont été recueillis parmi les travailleurs du laboratoire, de manière volontaire et anonyme. Seul le genre du donneur a été enregistré. Les échantillons ont été recueillis le jour même de l'expérimentation entre 9h30 et 10h30 (10h30=T0). Trois expériences ont été réalisées sur une période neuf mois en utilisant toujours le même protocole. Seules varient le nombre et le genre des donneurs, ainsi que les conditions testées (Tableau 15). Les temps de prélèvement pour analyses ont été réalisés à intervalles réguliers (toutes les 4h, ou 12h) entre 0h et 72h.

Tableau 15 : Différents échantillons et conditions utilisées pour les 3 expériences de stabilité urinaire.

Expérience	Constitution du mélange	Conditions testées		Temps de prélèvement (heures)										
		Conservateur	Température		0	4	8	12	16	20	24	36	48	72
			4°C	RT										
Janvier 2010 (Expérience 1)	5 Femmes - 5 Hommes	Acide borique AcB (2, 20 et 200 mM)	x	x (19°C)	x	x	x	x	x	x	x	x	x	x
		Azide de sodium (0.1, 1 et 10 mM)	x	x (19°C)										
		Sans conservateur	x	x (19°C)										
Juillet 2010 (Expérience 2)	2 Femmes - 3 Hommes	Sans conservateur	x	x (26°C)	x			x			x	x	x	x
Septembre 2010 (Expérience 3)	3 Femmes - 2 Hommes	Sans conservateur	x	x (23°C)	x			x			x		x	x
		Acide borique AcB (200 mM)		x (23°C)										

RT= Température ambiante

## 2.3 PREPARATION DES ECHANTILLONS

Une fois le mélange d'urines constitué, plusieurs aliquots ont été préparés, selon les conditions de l'expérience (avec ou sans conservateurs, concentrations différentes en conservateur) (Tableau 15). Ces aliquots ont ensuite été scindés en deux groupes. Les échantillons du premier groupe ont été conservés à température ambiante, (sur la paillasse du laboratoire, température maximale enregistrée voir Tableau 15), alors que ceux de l'autre groupe ont été placés à 4°C dans un réfrigérateur à température contrôlée.

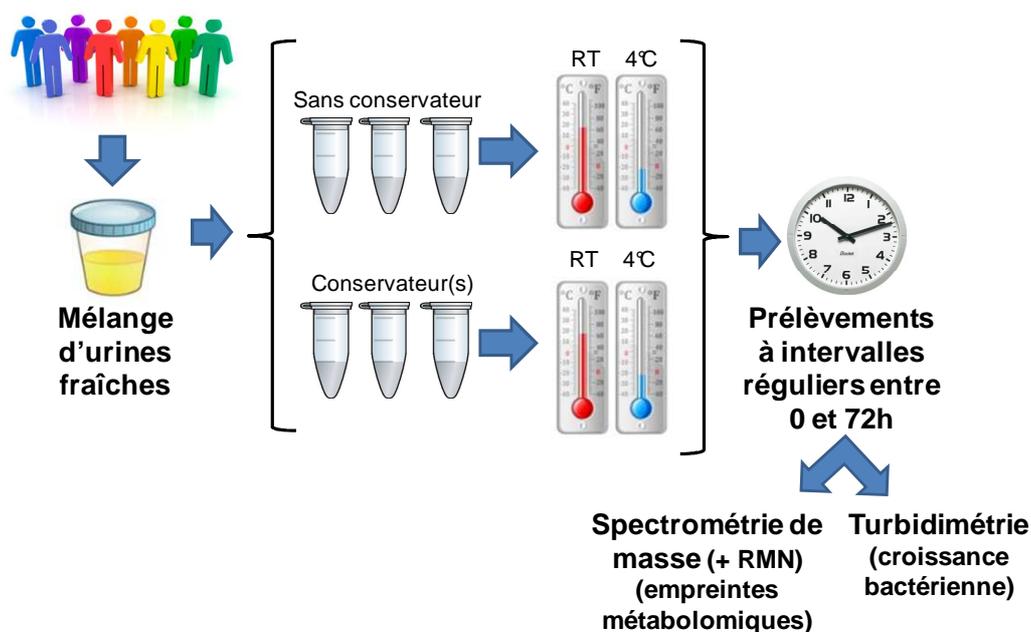


Figure 54 : Préparation des échantillons pour les expérience de stabilité urinaire.

Des échantillons ont été prélevés à intervalles réguliers (toutes les 4h, ou 12h) entre 0h et 72h. Une partie de ces prélèvements a été centrifugée (5 minutes à 3000t/min), diluée au cinquième dans de l'eau déminéralisée et congelée à -80°C pour analyse des empreintes ultérieure par spectrométrie de masse (et RMN). L'autre partie n'a pas été ni centrifugée, ni diluée et congelée directement à -80°C pour mesure ultérieure de la turbidité (suivi de la croissance bactérienne). Pour l'analyse par RMN, après décongélation, 400 µL d'urine diluée ont été ajoutés à 200 pl de PBS (pH=7,00) dans un mélange H<sub>2</sub>O/D<sub>2</sub>O (vol/vol) contenant 3mM d'acide 3-(triméthylsilyl) propionique-2,2,3,3 D<sub>4</sub>. Après 15 min, le pH a été ajusté à 7,00 ± 0,01, les échantillons ont ensuite été centrifugés 5 min à 12000 t/min puis analysés par RMN.

## **2.4 ACQUISITION DES DONNEES**

### **2.4.1 Appareillage**

Les analyses par spectrométrie de masse ont été réalisées sur le même matériel que décrit dans le Chapitre 1 (page 63). La mesure de la turbidité a été réalisée sur un lecteur de plaques Multiskan EX Thermo Fisher Scientific (Les Ulis, France) (Spectrophotomètre réglé à 620nm). Les spectres RMN <sup>1</sup>H ont été acquis avec un spectromètre Bruker Avance 500 MHz (à 500.13 MHz pour <sup>1</sup>H) et équipé d'une cryosonde de 5 mm thermostaté à 298 K.

### **2.4.2 Analyses par LC/MS**

Les analyses par LC/MS ont été réalisées comme précisé dans le chapitre1 (page 63), mais un seul type de séparation chromatographique (C<sub>18</sub>) a été utilisé

### **2.4.3 Analyses par RMN**

L'acquisition des spectres a été réalisée après une présaturation du signal de l'eau pendant de 2s et un temps de relaxation de 8s. Un total de 128 transients ont été accumulés sur une largeur spectrale de 5000 Hz (10 ppm), pour un temps d'acquisition total de 28 min et 24 s, et recueillis dans les points de données 32 k.

## **2.5 TRAITEMENT DES DONNEES BIOLOGIQUES**

Le traitement des données a été réalisé comme décrit dans le Chapitre 1 (page 65) pour les données LC/MS. Les spectres RMN ont été traités et analysés avec le logiciel MestReNova (version 6.2.0, Mesrelab, St Jacques de Compostelle, Espagne). Les spectres ont été obtenus après une transformation de Fourier sans apodisation. Après correction de phase et de la ligne de base, les zones des pics ont été mesurées par raccord de la conduite. Les concentrations des molécules ont été obtenues en utilisant comme référence le TSP.

## 2.6 ANALYSES STATISTIQUES

Les analyses statistiques multivariées ont été réalisées comme décrit dans le Chapitre 1 (page 65). Les analyses statistiques univariées (ANOVA et coefficients de corrélation de Spearman) ont été réalisées avec le logiciel R.

## 2.7 IDENTIFICATIONS

Les métabolites analysés sont ceux qui ont été annotés par la base de données spectrale, avec les mêmes paramètres que ceux décrits dans le Chapitre 1 (page 66).

# 3 RESULTATS

## 3.1 EVOLUTION DE LA CROISSANCE BACTERIENNE

La contamination bactérienne initiale des urines survient lors du recueil de ces dernières. Si l'urine vésicale est stérile en l'absence d'infection, il n'en est pas de même pour l'urètre, le méat urinaire et le périnée. Au niveau du méat urétral et du périnée, on retrouve une flore commensale cutanée (staphylocoques, corynébactéries) et une colonisation par les bactéries du tube digestif (*Escherichia coli*, *Klebsiella*, *Proteus*, entérocoques). La portion distale de l'urètre (les 2 derniers cm) est le siège d'une flore constituée de staphylocoques, corynébactéries, streptocoques, *Proteus*, entérocoques [304]. Il en résulte quelque difficulté à recueillir des urines sans contamination par cette flore commensale, le point important étant de limiter leur profération pendant et après le recueil.

La prolifération bactérienne a été suivie lors des 3 expériences. Les mesures d'absorbance ont été reportées pour les différentes conditions en fonction du temps (Figure 55). De grandes différences ont été observées aussi bien entre les différentes expériences, qu'entre les différentes conditions expérimentales au sein d'une même expérience (Figure 55A). Ainsi la croissance bactérienne est nulle à 4°C et ce quelque soit l'expérience (Figure 55C). Par contre, à température ambiante une croissance bactérienne plus ou moins importante est observée selon la date à laquelle a été réalisée l'expérience. A une température ambiante de 19°C environ, mesurée en Janvier (expérience 1), correspond une croissance bactérienne de très faible amplitude. En revanche, en juillet et en septembre (expériences 2 et 3), pour lesquels la température ambiante était supérieure à 20°C, la croissance bactérienne était beaucoup plus rapide et importante. L'ajout d'un conservateur a permis de supprimer cette croissance bactérienne à température ambiante (Figure 55B).

Si on se base uniquement sur l'évolution de la croissance bactérienne, il semble que l'ajout d'un conservateur ou le prélèvement sur glace soit suffisant pour limiter la contamination bactérienne des urines et donc pour également limiter son impact (qualitatif et quantitatif) sur les métabolites urinaires. De plus pour des températures ambiantes inférieures à 20°C, cette croissance bactérienne est très faible.

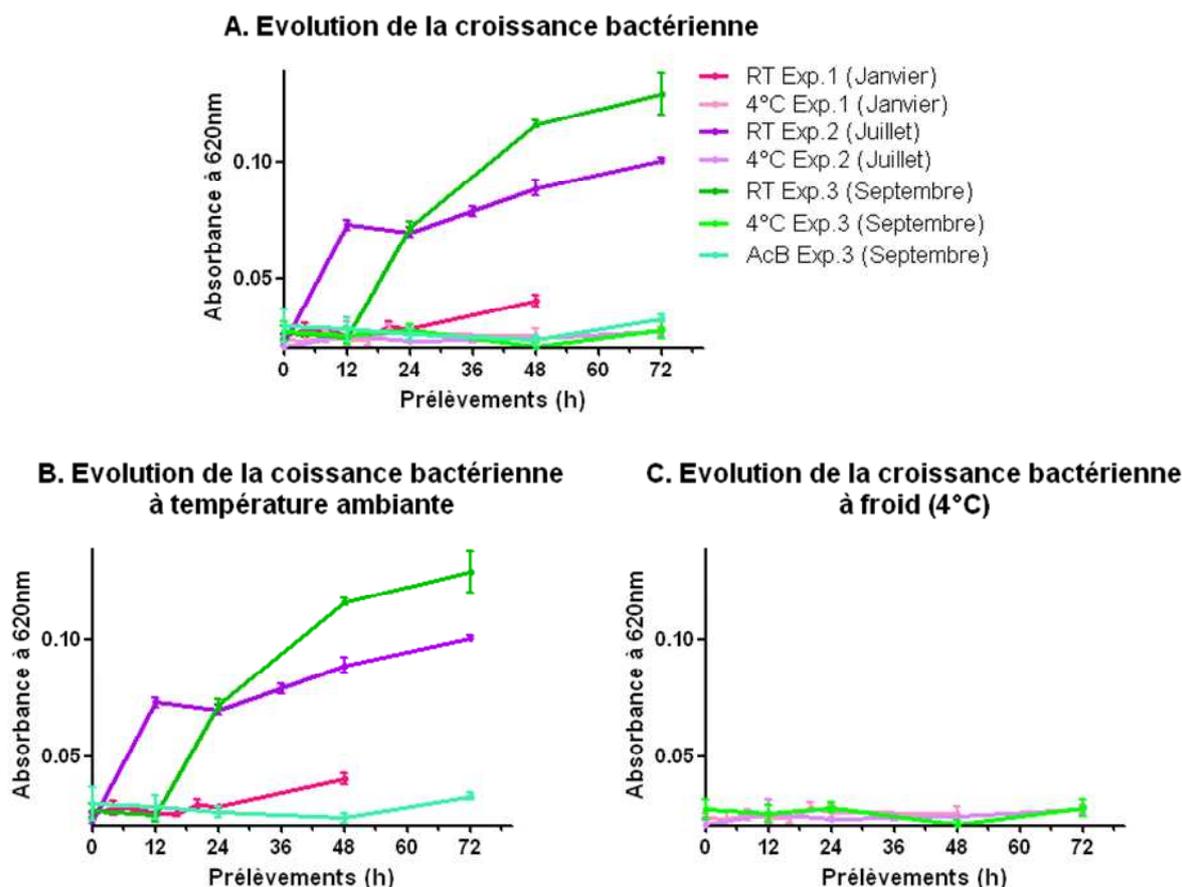


Figure 55 : Evolution de la contamination bactérienne au cours des 3 expériences

L'Absorbance à 620 nm en fonction du temps est reportée pour toutes les conditions réunies (haut), à température ambiante (bas gauche) ou à froid (bas droite). RT = température ambiante sans conservateurs, 4°C = 4°C sans conservateur et AcB = température ambiante avec conservateur (acide borique).

### 3.2 ANALYSES STATISTIQUES MULTIVARIABLES

Comme pour la cohorte décrite dans le chapitre 1, Les données de stabilité ont été dans un premier temps analysées en globalité à l'aide d'une analyse non supervisée ou ACP afin de visualiser la répartition «spontanée» des échantillons dans une espace à deux dimensions. Les résultats ont ensuite été affinés à l'aide d'analyses statistiques supervisées de type PLS.

La première ACP (Figure 56) regroupe les 3 conditions de conservations de l'expérience de septembre 2010 (Expérience 3) pour laquelle la prolifération bactérienne la plus importante a été observée : avec conservateur (AcB pour acide borique), sans conservateur à 4°C (4°C), et température ambiante (RT de l'anglais *Room Temperature*), ainsi que les échantillons de contrôle qualité. Ces derniers sont en fait assimilables aux échantillons en conditions initiales (T0) sans conservateur à 4°C et à température ambiante. C'est la raison pour laquelle ils se retrouvent regroupés au centre du graphique. Les échantillons avec conservateurs (AcB) sont clairement séparés des autres car l'ajout d'acide borique modifie drastiquement la composition des urines

(acidification du milieu, présence de l'acide borique lui-même). De plus, les échantillons sans conservateur à température ambiante dérivent au fur et à mesure du temps (les T72 étant totalement à droite du graphique) et donc au fur et à mesure que la croissance bactérienne augmente. On peut donc conclure que la première composante (axe des abscisses) permet de séparer les échantillons selon le niveau de prolifération bactérienne, alors que la deuxième composante (axe des ordonnées) permet de discriminer les échantillons selon la présence ou non de conservateur.

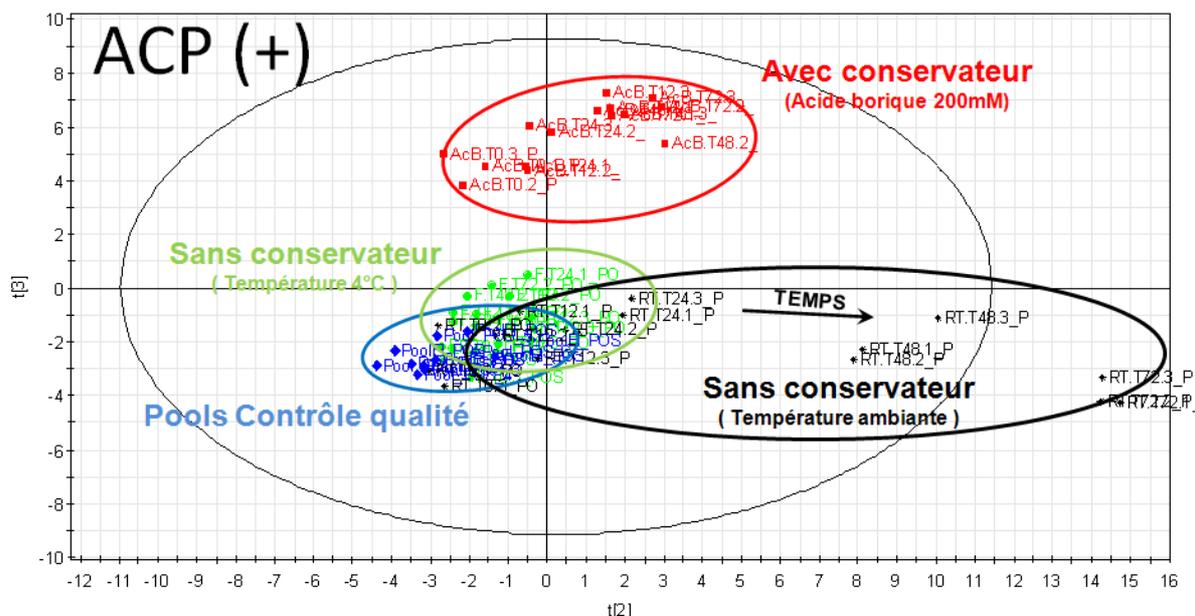


Figure 56 : Score plot de l'ACP des données de stabilité (Expérience 3) en mode d'ionisation positif.

Les résultats obtenus en PLS-DA avec les mêmes données, mais en renseignant cette fois l'appartenance de chaque échantillon à un groupe donné (Pools CQ, AcB, 4°C et RT) (Figure 57), donnent le même type de classification que l'ACP, avec peut-être une séparation des groupes un peu plus marquée. La première composante renseigne cette fois sur la présence ou l'absence de conservateur, alors que la deuxième composante permet de visualiser l'impact de la contamination bactérienne. Le peu de différences entre les deux modèles suggère que la source principale de variation observée est bien celle liée aux facteurs à étudier (stabilité, présence ou non de conservateurs). En effet le nombre d'échantillon étant restreint pour chaque expérience, la variabilité analytique est limitée. Les mêmes résultats ont été obtenus en mode négatif et pour l'expérience de juillet 2010 (Expérience 2).

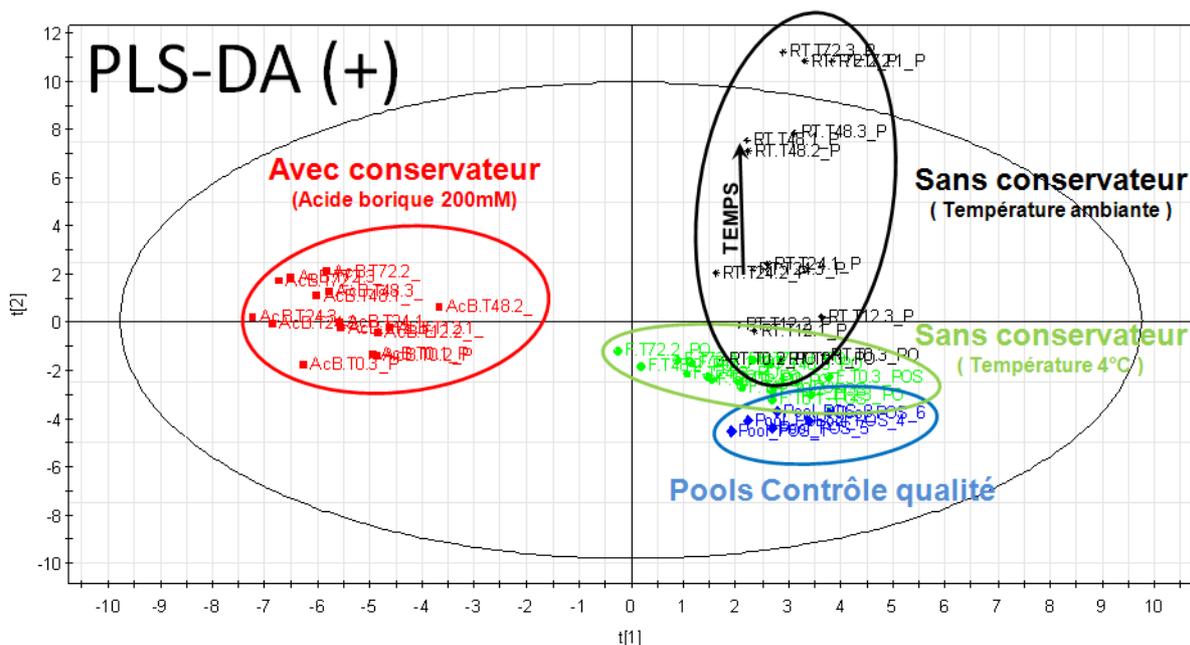


Figure 57 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif.

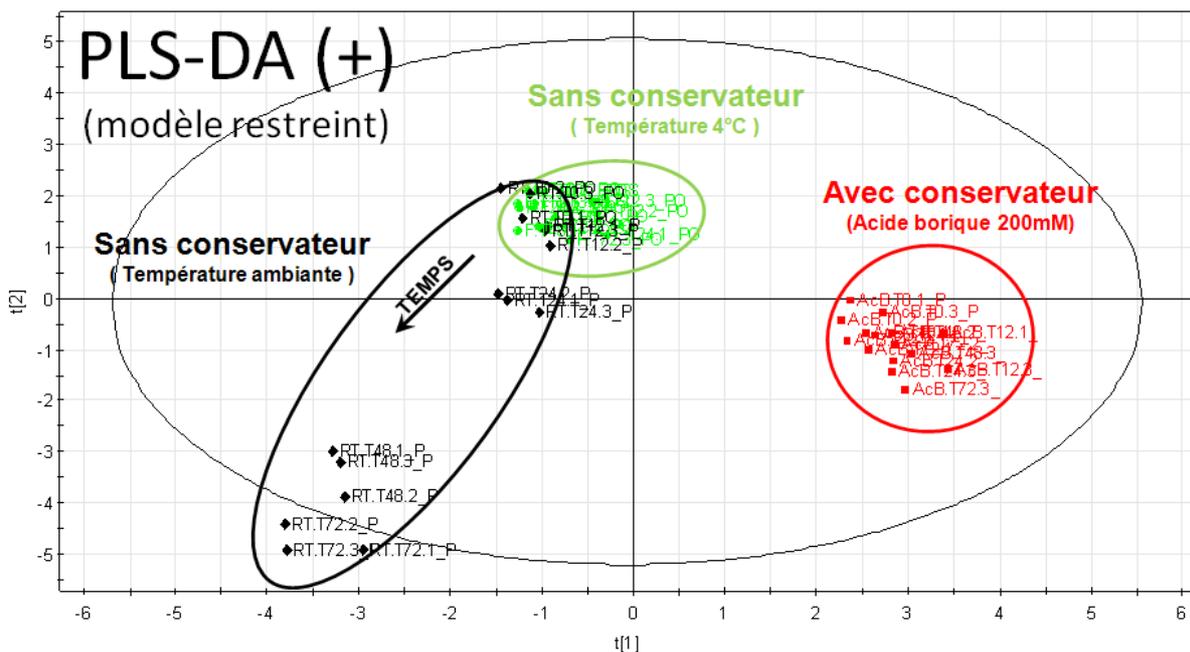


Figure 58 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. Modèle restreint aux métabolites identifiés.

Nous cherchons à connaître en priorité la stabilité des métabolites déjà identifiés dans les urines (Chapitre 1). Pour cela le modèle va être restreint uniquement aux variables correspondants à ces métabolites et à nos 3 conditions de conservation. Même avec ce modèle restreint (Figure 58) la disposition des groupes est similaire : le groupe avec conservateur distinct des groupes sans conservateur, Les groupes sans conservateurs qui se séparent au fur et à mesure du temps à cause de l'évolution de la prolifération bactérienne. En comparant les groupes entre eux, les variables

évoluant avec la contamination bactérienne et celles n'étant pas stables dans le temps indépendamment de cette contamination bactérienne (instabilité chimique) vont pouvoir être séparées. La significativité de ces variables sera ensuite vérifiée par des statistiques univariées (ANOVA et coefficients de corrélation de Spearman).

### 3.2.1 Impact de la contamination bactérienne sur les empreintes métaboliques

De part leur métabolisme propre, les bactéries potentiellement présentes dans les urines peuvent en modifier la composition en métabolites: Elles consomment et produisent des métabolites au cours du temps et des modifications vont apparaître sur les empreintes métaboliques. Ce sont ces variations que nous souhaitons mettre en évidence. Pour y parvenir, il faut donc comparer une condition expérimentale pour laquelle aucune croissance bactérienne n'est observée (4°C ou AcB) à une condition pour laquelle il y a croissance bactérienne (RT). Pour cela, nous utilisons comme précédemment, un modèle restreint aux variables déjà identifiées. En comparant les groupes sans conservateurs à 4°C et RT de l'expérience 3 (Figure 59), nous remarquons que les échantillons à T0 et T12 à RT, se confondent avec ceux à 4°C. A partir de T24 un début de séparation est observé, ce qui est cohérente avec le début de la croissance bactérienne (Figure 55). Les T48 et T72 sont encore plus éloignés ce qui montre en très bonne corrélation de ce modèle avec la croissance bactérienne observée en Figure 55. C'est la première composante qui explique les variations liées à la croissance bactérienne. Ont été étudiées en détails les variables qui contribuent majoritairement à cette composante (Figure 60).

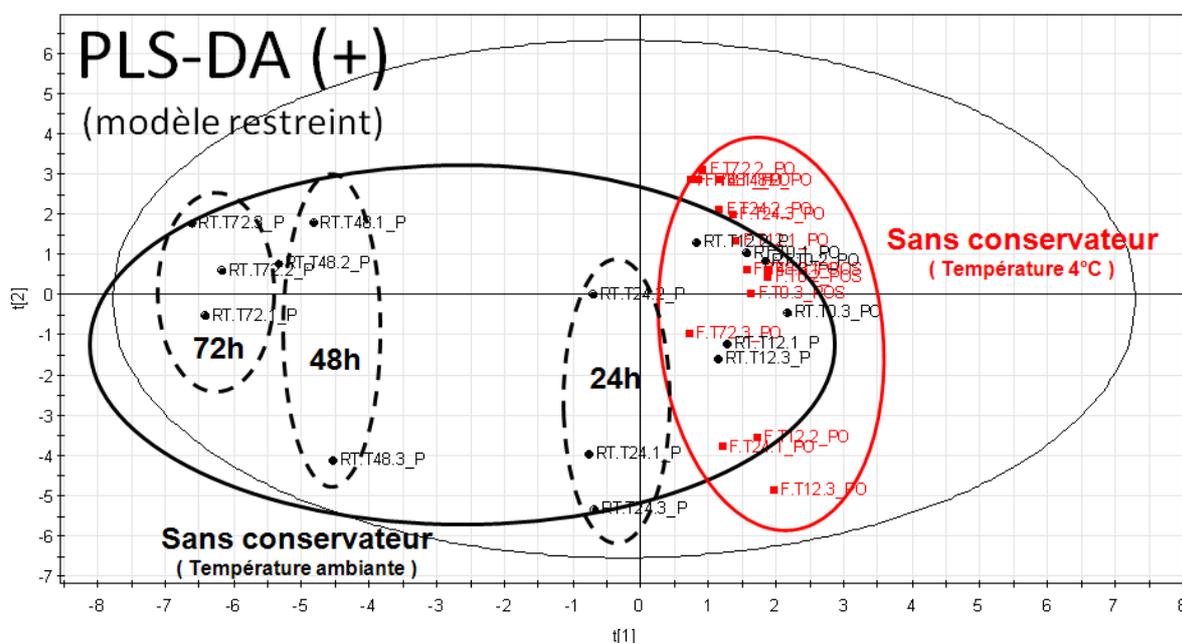


Figure 59 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif. Modèle restreint aux métabolites identifiés.

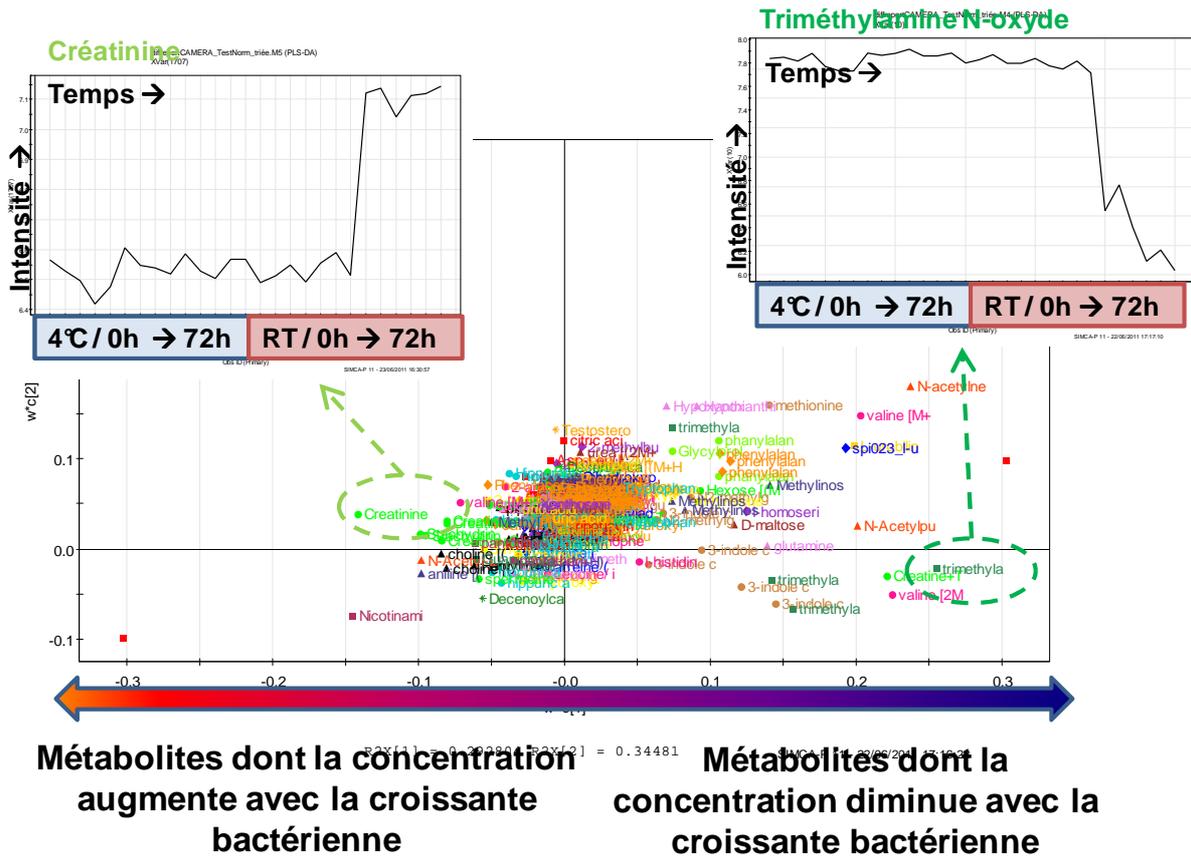


Figure 60 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation positif.

La carte factorielle des poids des variables sur la composante 1 permet de visualiser les variables (ou métabolites) impactés par la croissance bactérienne : à gauche les variables dont la concentration augmente avec la croissance bactérienne, et à droite celles dont la concentration diminue. Il est également possible de visualiser l'évolution des concentrations de chaque variable individuellement. La créatinine est l'une des variables dont la concentration augmente le plus car elle est l'une des plus excentrées à gauche du graphique. En visualisant son évolution nous observons qu'elle augmente drastiquement à T48 et T72. Inversement, le N-oxyde de triméthylamine, excentrée à droite du graphique, voit sa concentration diminuer significativement quand la croissance bactérienne augmente.

A partir de ce modèle, nous pouvons donc extraire une liste de variables les plus significatives, qui correspondent aux variables impactées par la contamination bactérienne observée en mode d'ionisation positif (la même démarche a été utilisée pour les variables du mode d'ionisation négatif). Par exemple, la concentration en acide férulique augmente avec la contamination bactérienne, alors que les concentrations en acide 3-méthyle-2-oxovalérique, elles, chutent à partir de 48h. Dans les deux modèles (obtenus pour les données acquises dans les modes d'ionisation positif et négatif), nous notons qu'il y a globalement plus de variables dont l'intensité diminue avec la contamination bactérienne, que de variables dont l'intensité augmente. Il y aurait donc vraisemblablement plus de consommation ou dégradation que de production de métabolites par les bactéries.

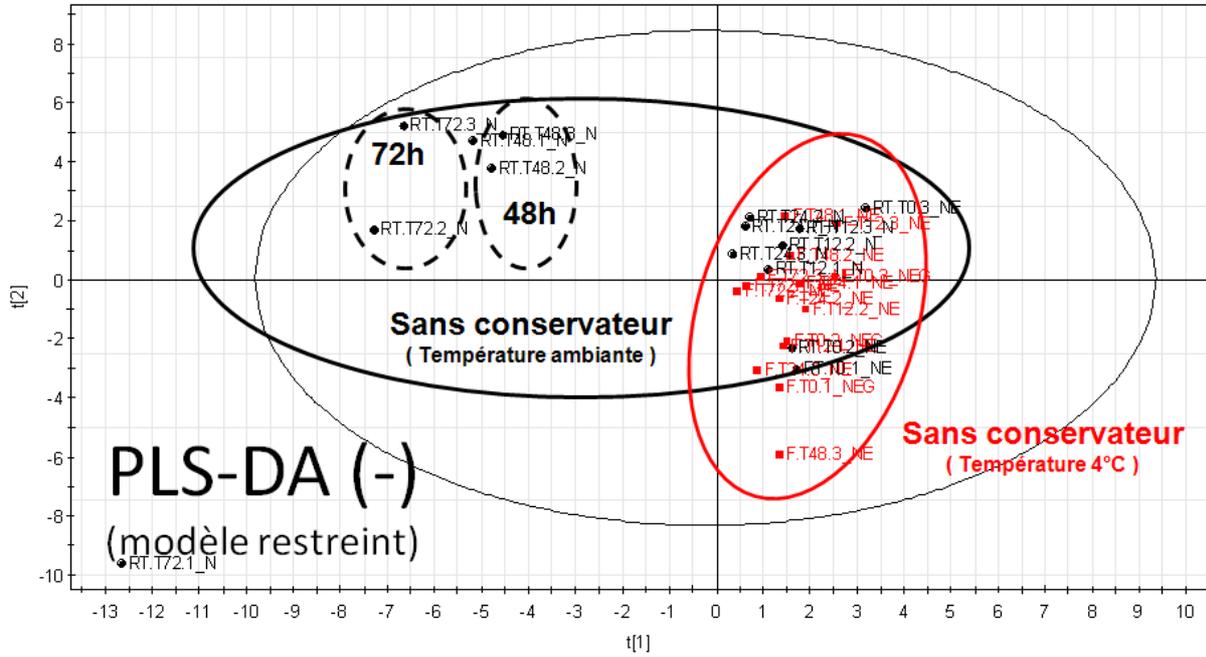


Figure 61 : Score plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif. Modèle restreint aux métabolites identifiés.

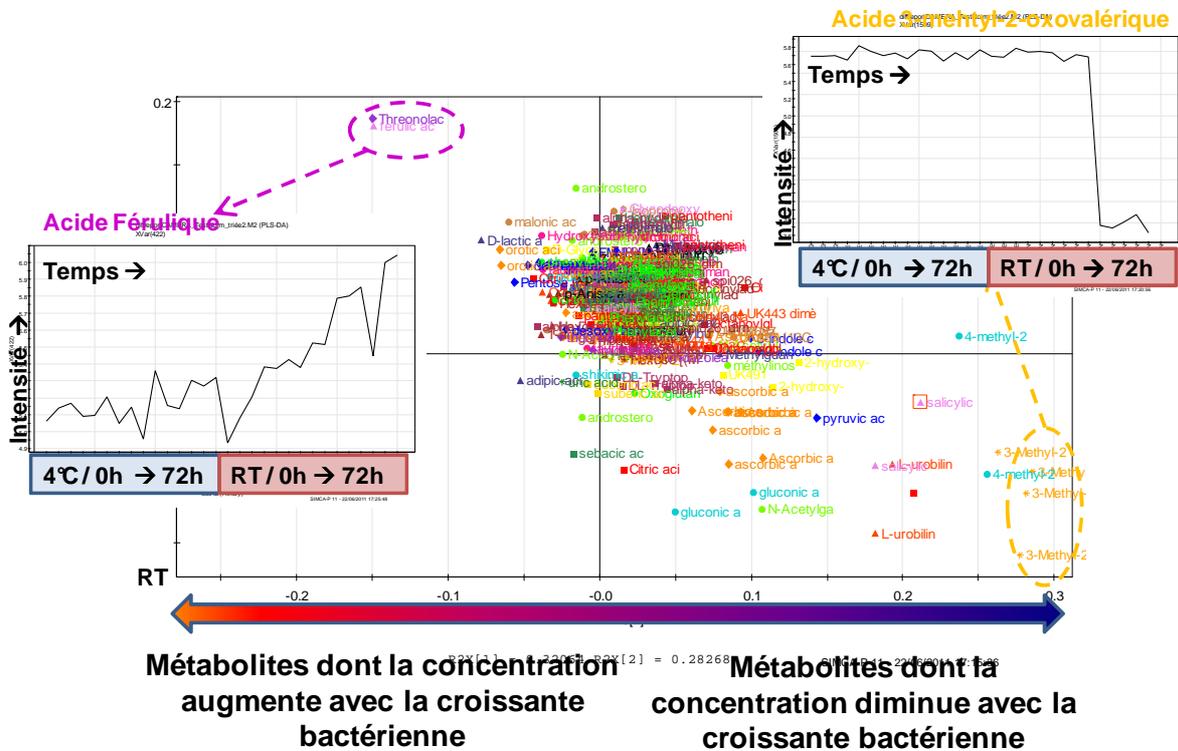


Figure 62 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif.

### 3.2.2 Impact de l'instabilité chimique sur les empreintes métaboliques

Comme nous avons pu le visualiser sur les Figure 59 et Figure 61, La source de variabilité la plus importante est celle liée à la contamination bactérienne. Une fois cette source de variabilité analysée, les échantillons incubés à température ambiante ont été retirés du modèle afin d'analyser les autres sources potentielles de variabilités (Figure 63). Une discrimination entre les échantillons incubés avec et sans conservateur (déjà mentionnée au paragraphe précédent) est observable sur la composante 1, mais également une discrimination importante au fur et à mesure du temps sur la deuxième composante. Ces échantillons n'étant pas impactés par la contamination bactérienne (voir paragraphe 3.1), il existe donc une autre source de variabilité qui modifie les empreintes métaboliques, et ce phénomène est amplifié à température ambiante. Il s'agit en fait d'une dégradation au cours du temps de certains métabolites, due à leur instabilité chimique en solution. Cette dégradation semble assez limitée à 4°C. En analysant la carte factorielle de la Figure 64, nous visualisons bien à droite du graphique les métabolites dont la concentration diminue avec le temps (dégradation) et à gauche ceux dont la concentration augmente (production de produits de dégradation). Cette discrimination en fonction du temps est beaucoup plus marquée en mode négatif, et n'est quasiment pas visible en mode positif, ce qui laisse penser que la plupart des métabolites présentant une faible stabilité chimique sont des métabolites ionisés en mode d'ionisation négatif (acides organiques par exemple).

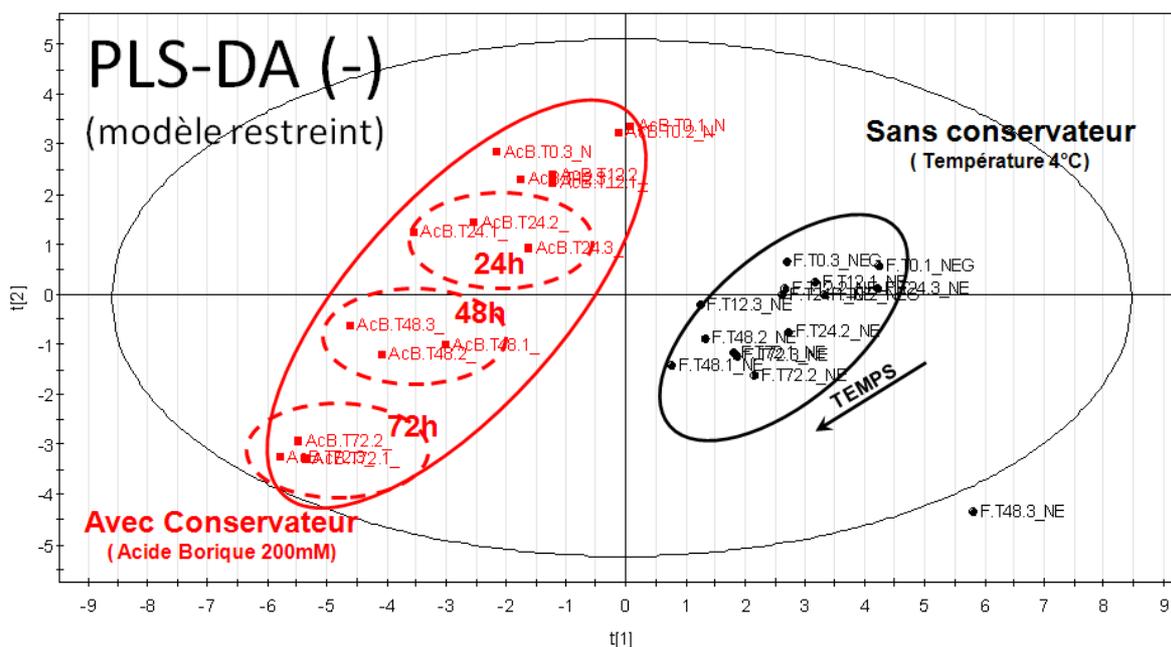


Figure 63 Scree plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif. Modèle restreint aux métabolites identifiés.

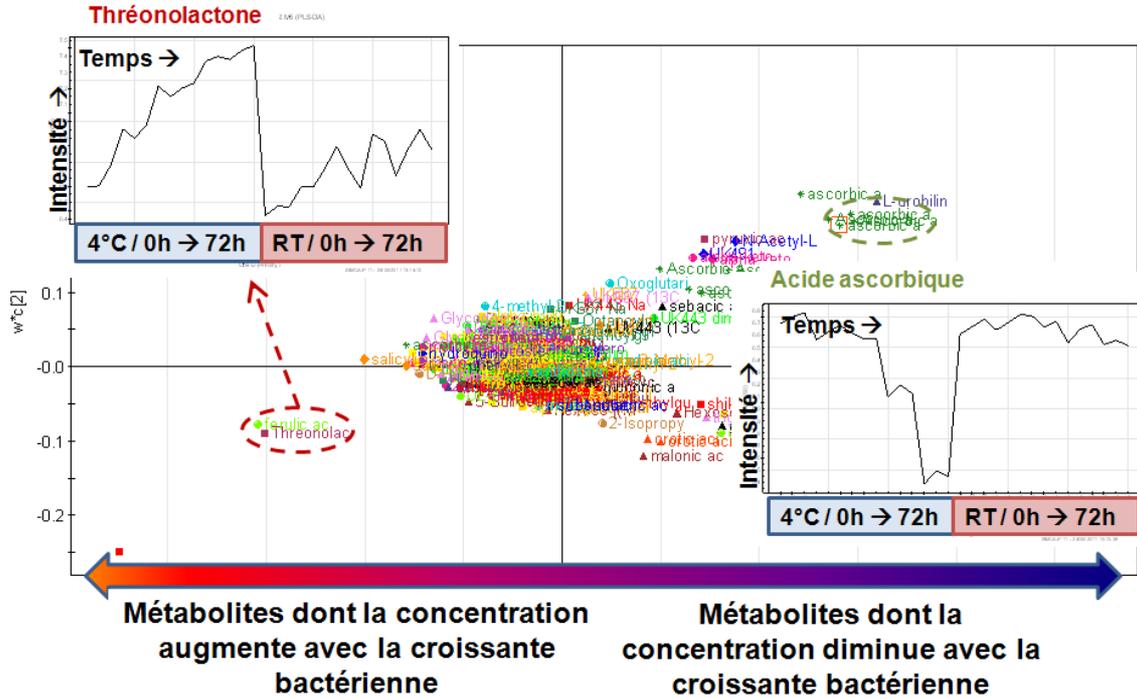


Figure 64 : Loading plot de la PLS-DA des données de stabilité (Expérience 3) en mode d'ionisation négatif.

### 3.3 VARIATIONS DES CONCENTRATIONS EN METABOLITES

Lors de ces expériences nous nous sommes focalisés sur les 244 métabolites déjà identifiés dans les urines présents dans notre base de données spectrale (voir Chapitre 1). Sur ces 244 métabolites, 125 ont été retrouvés en mode d'ionisation positif, et 119 en mode négatif pour l'expérience 2, et 109 et 126 en mode d'ionisation positif et négatif pour l'expérience 3 (Figure 65). Pour chaque des deux expériences, 49 métabolites communs sont retrouvés dans les deux modes.

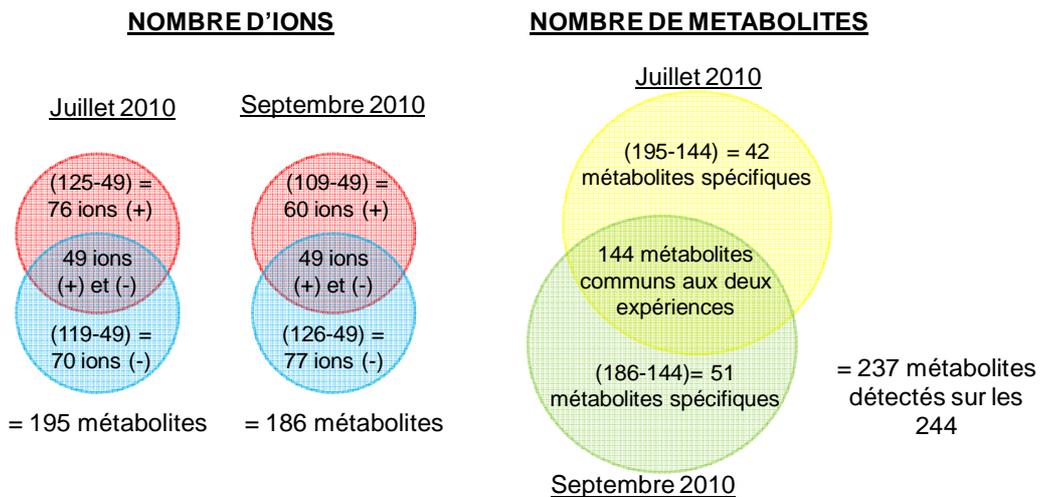


Figure 65 : Métabolites retrouvés lors des études de stabilités des expériences 2 et 3.

Au total on retrouve donc 195 métabolites et 186 pour les expériences 2 et 3, dont 144 sont communs aux deux expériences, ce qui fait un total de 237 métabolites détectés sur les 244 de la base de données spectrale.

Des analyses statistiques univariées en fonction du temps de prélèvement (ANOVA et coefficient de corrélation de Spearman, avec correction des tests multiples) ont été réalisées. Ces tests ont été menés indépendamment pour chaque condition (RT, 4°C, AcB) et pour chaque expérience (2 et 3). N'ont été conservées que les variables statistiquement interprétables (voir Chapitre 1 et Annexe 3) pour lesquelles une « *p-value* » est  $<0.05$  dans au moins une des conditions (RT, 4°C ou AcB). Les valeurs de « *q-value* » figurent dans le Tableau 16 ainsi que le premier temps significativement différent de T0. Pour rappel la « *q-value* » est une valeur corrigée de la « *p-value* » qui est calculée lors de tests multiples [278;279] (Chapitre 2 page 97). Une « *q-value* » significative ( $<0.05$ ) à RT uniquement indique un impact de la contamination bactérienne. Une « *q-value* » significative à RT et AcB (voire 4°C) indique une instabilité chimique.

Tableau 16 : Liste des métabolites discriminants dans au moins une des conditions des expériences 2 et 3.

Molécule	Ionisation	JUILLET 2010						SEPTEMBRE 2010								
		RT			4°C			RT			4°C			AcB		
		q.value SPEAR.	q.value ANOVA	Tps ≠ T0	q.value SPEAR.	q.value ANOVA	Tps ≠ T0	q.value SPEAR.	q.value ANOVA	Tps ≠ T0	q.value SPEAR.	q.value ANOVA	Tps ≠ T0	q.value SPEAR.	q.value ANOVA	Tps ≠ T0
N2-Méthylguanosine	ESI(+)	<b>0.0060</b>	<b>0.0358</b>	↓T12	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
	ESI(-)	<b>&lt;0.0001</b>	<b>0.0001</b>	↓T24	ns	ns	ns	ns	<b>0.0031</b>	↓T72	ns	ns	ns	ns	ns	ns
Acide 4-Méthyl-2-oxovalérique (ketoleucine)	ESI(-)	ns	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	ns	<b>&lt;0.0001</b>	↓T24	ns	ns	ns	ns	ns	ns
Oxyde de triméthylamine	ESI(+)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	<b>0.0005</b>	<b>&lt;0.0001</b>	↓T48	ns	ns	ns	ns	ns	ns
Acide 3-Méthyl-2-oxovalérique	ESI(-)	<b>0.0007</b>	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	<b>0.0098</b>	<b>&lt;0.0001</b>	↓T48	ns	ns	ns	ns	ns	ns
Acide ascorbique	ESI(-)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	<b>0.0005</b>	<b>0.0012</b>	↓T24	ns	ns	ns	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12
L-urobilinogène	ESI(+)	<b>&lt;0.0001</b>	<b>0.0004</b>	↓T24	ns	ns	ns	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12	<b>0.0001</b>	<b>0.0229</b>	↓T12	<b>0.0001</b>	<b>&lt;0.0001</b>	↓T12
	ESI(-)	<b>0.0001</b>	<b>0.0027</b>	↓T24	ns	ns	ns	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12	<b>0.0002</b>	ns	ns	<b>&lt;0.0001</b>	<b>0.0002</b>	↓T12
Acide pyruvique	ESI(-)	ns	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	<b>0.0034</b>	<b>&lt;0.0001</b>	↓T12	<b>0.0010</b>	<b>0.0049</b>	↓T48	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T12
Threonolactone	ESI(-)	ns	<b>0.0374</b>	↑T36	ns	ns	ns	<b>0.0019</b>	<b>0.0285</b>	↑T48	ns	ns	ns	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↑T24
N4-Acetylcytidine	ESI(+)	ns	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	<b>0.0004</b>	<b>&lt;0.0001</b>	↓T24	ns	ns	ns	ns	ns	ns
	ESI(-)	ns	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	NF	NF	NF	NF	NF	NF	NF	NF	NF
Glucuronide d'acide Ketoretinoïque	ESI(-)	ns	<b>0.0023</b>	↓T36	ns	ns	ns	NF	NF	NF	NF	NF	NF	NF	NF	NF
Glucuronide d'acide hydroxyretinoïque	ESI(-)	<b>&lt;0.0001</b>	<b>0.0005</b>	↓T36	ns	ns	ns	NF	NF	NF	NF	NF	NF	NF	NF	NF
Acide 2-Hydroxy 3-méthylbutyrique	ESI(-)	<b>&lt;0.0001</b>	<b>0.0002</b>	↓T12	ns	ns	ns	NF	NF	NF	NF	NF	NF	NF	NF	NF
Acide cholique	ESI(-)	<b>0.0035</b>	<b>0.0433</b>	↓T36	ns	ns	ns	NF	NF	NF	NF	NF	NF	NF	NF	NF
Acide α-ketoglutarique	ESI(-)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	↓T24	ns	ns	ns	ns	ns	ns	ns	ns	ns	<b>0.0035</b>	<b>0.0015</b>	↓T48
Diméthylguanosine	ESI(-)	<b>0.0007</b>	<b>0.0002</b>	↓T48	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Acide 2,5-dihydroxybenzoïque	ESI(-)	ns	<b>&lt;0.0001</b>	↓T12	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Acide phényllactique	ESI(-)	ns	<b>0.0027</b>	↑T72	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Acide homovanillique	ESI(-)	ns	<b>0.0092</b>	↑T24	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Acide adipique	ESI(-)	NF	NF	NF	NF	NF	NF	ns	<b>0.0009</b>	↑T72	ns	ns	ns	ns	ns	ns
Acide lactique	ESI(-)	NF	NF	NF	NF	NF	NF	<b>0.0098</b>	<b>0.0285</b>	↑T48	ns	ns	ns	ns	ns	ns
Tyrosine	ESI(+)	NF	NF	NF	NF	NF	NF	<b>0.0059</b>	<b>0.0200</b>	↓T48	ns	ns	ns	ns	ns	ns

N-Acetylphenylalanine	ESI(+)	NF	NF	NF	NF	NF	NF	0.0154	0.0200	↑T48	ns	ns	ns	ns	0.0043	↑T24
L-urobiline	ESI(-)	NF	NF	NF	NF	NF	NF	<0.0001	<0.0001	↓T12	0.0010	ns	ns	<0.0001	0.0005	↓T12
Acide malonique	ESI(-)	NF	NF	NF	NF	NF	NF	ns	0.0055	↑T12	0.0010	ns	ns	0.0019	0.0122	↑T48
Phénylalanine	ESI(+)	ns	ns	ns	ns	ns	ns	0.0098	0.0018	↓T24	ns	ns	ns	ns	ns	ns
Acide salicylique	ESI(-)	ns	ns	ns	ns	ns	ns	ns	0.0009	↑T48	ns	ns	ns	ns	ns	ns
Acide pipécolique	ESI(+)	ns	ns	ns	ns	ns	ns	ns	0.0368	↑T72	ns	ns	ns	ns	ns	ns
Créatine	ESI(+)	0.0455	ns	ns	ns	ns	ns	ns	0.0043	↑T72	ns	ns	ns	ns	ns	ns
Choline	ESI(+)	ns	ns	ns	ns	ns	ns	ns	0.0001	↑T48	ns	ns	ns	ns	ns	ns
Aniline (isomère)	ESI(+)	ns	ns	ns	ns	ns	ns	0.0059	<0.0001	↑T48	ns	ns	ns	ns	ns	ns
Glucuronide d'acide 3-Indole carboxylique	ESI(+)	ns	ns	ns	ns	ns	ns	ns	0.0007	↓T72	ns	ns	ns	ns	ns	ns
	ESI(-)	ns	ns	ns	ns	ns	ns	ns	<0.0001	↓T48	ns	ns	ns	ns	ns	ns
Acide citrique	ESI(+)	ns	ns	ns	ns	ns	ns	ns	0.0038	↓T48	ns	ns	ns	ns	ns	ns
Glutamine	ESI(+)	ns	ns	ns	ns	ns	ns	0.0154	0.0047	↓T48	ns	ns	ns	ns	ns	ns
Méthylinosine	ESI(+)	ns	ns	ns	ns	ns	ns	0.0002	0.0055	↓T48	ns	ns	ns	ns	ns	ns
	ESI(-)	ns	ns	ns	ns	ns	ns	ns	0.0013	↓T48	ns	ns	ns	ns	ns	ns
Octanoylglucuronide	ESI(-)	0.0248	ns	ns	ns	ns	ns	ns	0.0032	↓T72	ns	ns	ns	ns	ns	ns
Acide urique	ESI(+)	ns	ns	ns	ns	ns	ns	ns	0.0368	↑T72	ns	ns	ns	ns	ns	ns

- ns = non significatif, NF = *Not Found* (non retrouvé)

Notons qu'aucune variable n'est discriminante à 4°C pour l'expérience 2 et qu'une partie des variables discriminantes à RT l'est à la fois pour les expériences 2 et 3 (environ 1/3). Une autre partie des signaux discriminants n'est retrouvée spécifiquement que dans l'une des deux expériences, et enfin la dernière partie n'est discriminante que pour l'une ou l'autre. Certaines de ces variables impactées par la contamination bactérienne sont aussi discriminantes à 4°C et en présence de conservateur (AcB), ce qui indiqu'elles sont sensibles à la la présence de bactéries mais présentent aussi une instabilité chimique au cours du temps.

### 3.3.1 Métabolites impactés par la croissance bactérienne

Selon la bibliographie, certains métabolites sont déjà connus pour varier avec la contamination bactérienne (Tableau 17). Sur l'ensemble de ces métabolites, tous sont présents dans la base de données spectrale sauf 3 : l'acide acétique, la triméthylamine et l'acide formique. Même si ils étaient présents, l'acide formique et l'acide acétique sont de trop petites molécules pour être détectées avec notre méthode LC/MS. Sur les 11 métabolites restants, 3 ne sont pas retrouvés en LC/MS bien qu'ils soient bien présents dans la base de données spectrale : l'acide benzoïque, l'acide succinique et la glycine. L'alanine, quant à elle, n'est à priori pas présente car nous sommes plutôt en présence de l'un de ces isomères (voir Annexe 3). Sept de ces métabolites sont donc retrouvés dans nos expériences de stabilité, dont 3 sont non significatifs (ns) malgré la contamination bactérienne : l'urée, l'acide hippurique et la créatinine. Cette dernière est de toute façon non statistiquement interprétable à cause de la présence de plusieurs isomères qui sont co-élués dans nos conditions d'analyse chromatographique (Chapite 1 page 69).

**Tableau 17 : Métabolites connus pour varier avec la contamination bactérienne des urines [301;303] et variations observées lors des l'étude de stabilité.**

Metabolite	Variation attendue	Présent dans la base de données spectrale	Observation en MS + Mode d'ionisation	JUILLET 2010		SEPTEMBRE 2010		
				q.val ANOVA RT	q.val ANOVA 4°C	q.val ANOVA RT	q.val ANOVA 4°C	q.val ANOVA AcB
Acide benzoïque	↑	OUI	NF	NF	NF	NF	NF	NF
Créatine	↑	OUI	ESI(+)	ns	ns	<b>0.0042</b> (↑T72)	ns	ns
Créatinine*	↓	OUI	ESI(+)	ns	ns	ns	ns	ns
Acid lactique	↑	OUI	ESI(-)	ns	ns	<b>0.0285</b> (↑T48)	ns	ns
Acide citrique	↓	OUI	ESI(-)	ns	ns	<b>0.0037</b> (↓T48)	ns	ns
Acide hippurique	↓	OUI	ESI(-)	ns	ns	ns	ns	ns
Acide succinique	↑	OUI	NF	NF	NF	NF	NF	NF
Acide malonique	↑	OUI	ESI(-)	NF	NF	<b>0.0054</b> (↑T12)	ns	ns
Acide acétique	↑	NON	NF	NF	NF	NF	NF	NF
Acide formique	↑	NON	NF	NF	NF	NF	NF	NF
Triméthylamine	↑	NON	NF	NF	NF	NF	NF	NF
Glycine	↑	OUI	NF	NF	NF	NF	NF	NF
Alanine	↑	OUI	(autre isomère)	NF	NF	NF	NF	NF
Urée	↓	OUI	ESI(+)	ns	ns	ns	ns	ns

\* Non interprétable statistiquement car plusieurs isomères pour le même pic chromatographique  
NF = *Not Found* (non retrouvé)

Restent donc la créatine, l'acide lactique, l'acide citrique et l'acide malonique qui évoluent significativement, et de la même manière que dans la littérature, mais uniquement pour l'expérience 3. Comme mentionné au chapitre 3.1. (Figure 55), la contamination bactérienne était plus importante en septembre qu'en juillet. Ces variations en concentrations de métabolites ne sont peut-être significatives qu'au-delà d'un certain nombre de bactérie dans l'échantillon, nombre qui semble atteint en septembre mais pas en juillet. En effet ces variations sont significatives mais assez tardivement (sauf pour l'acide malonique que l'on ne retrouve pas en juillet), donc quand la contamination bactérienne est déjà importante. Certains de ces métabolites connus pour varier avec la contamination bactérienne, même si ils ne sont pas observés en LC/MS, sont détectables en RMN. Des analyses RMN complémentaires ont permis d'analyser les variations de ces métabolites (Figure 66). L'alanine, la glycine et la triméthylamine ne montrent aucune variation. Par contre l'acide acétique augmente significativement à RT, l'acide succinique montre une légère augmentation à RT en septembre et l'acide formique diminue à RT en septembre et juillet, mais aussi à 4°C en juillet. Cette évolution de l'acide formique n'est pas cohérente avec la bibliographie, d'autant plus qu'en juillet cette diminution concerne à la fois la condition RT et 4°C.

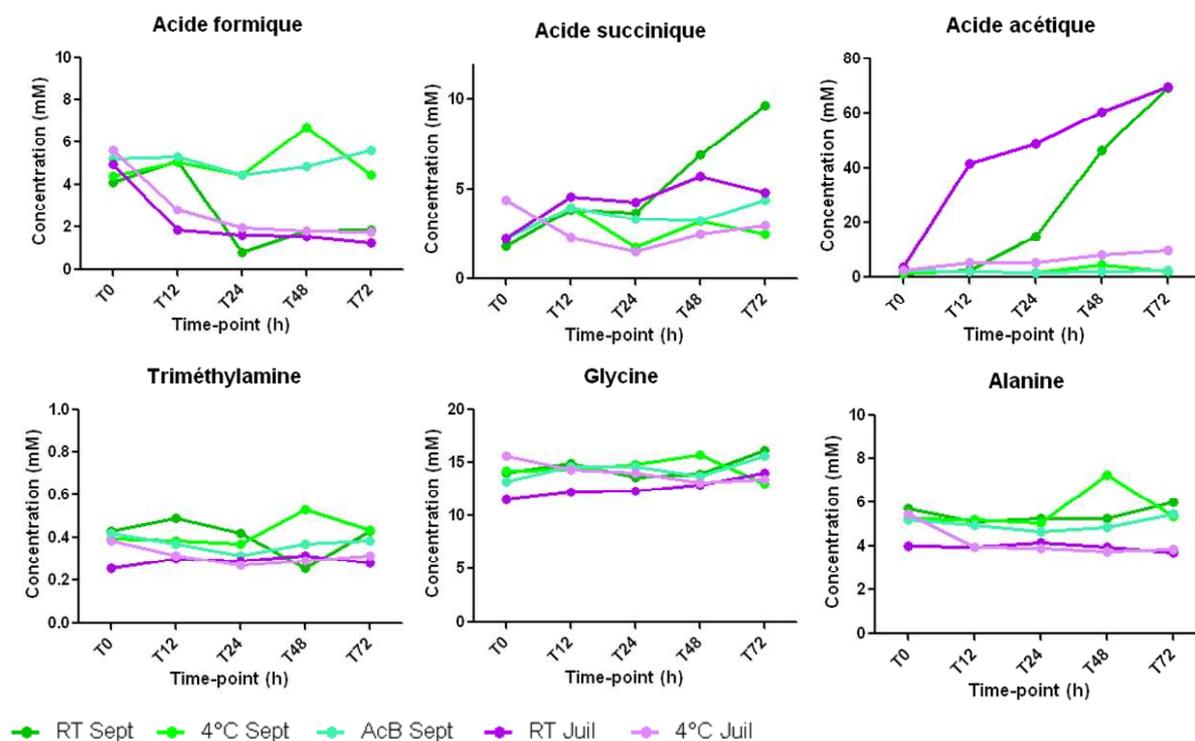


Figure 66 : Métabolites détectés en RMN connus pour varier avec la contamination bactérienne.

En plus de ces 4 métabolites du tableau 17, ce sont en tout 27 métabolites (sur les 237 observés) qui sont impactés par la contamination bactérienne (Tableau 16) à des niveaux plus ou moins important, comme par exemple l'oxyde de triméthylamine, la N4-acétylcytidine ou la méthylguanosine. Certaines des variations observées peuvent être facilement expliquées par le métabolisme de fermentation bactérien : augmentation des acides lactique, succinique, acétique. L'impact des bactéries sur certaines famille de molécules est également connu, comme par exemple la dégradation des

glucuronides [305] : on observe en effet la diminution de tous les glucuronides endogènes (Figure 67) détectés à l'exception des glucuronides de stéroïdes. Cependant certaines variations ne sont pas explicables : la triméthylamine (TM) est produite par les bactéries à partir de la choline, et par dégradation de l'oxyde de triméthylamine (TMAO). Si nous observons bien une diminution de la TMAO (dégradation), nous ne détectons en revanche aucune augmentation de la TM (Figure 67) et pas de diminution, voire même une augmentation de la choline (Figure 67). Enfin, nous ignorons encore le lien entre les variations de concentration de certaines molécules telles que la méthylguanosine, la méthylinosine, ou encore l'acide 3-méthyl-2-oxovalérique et la contamination bactérienne.

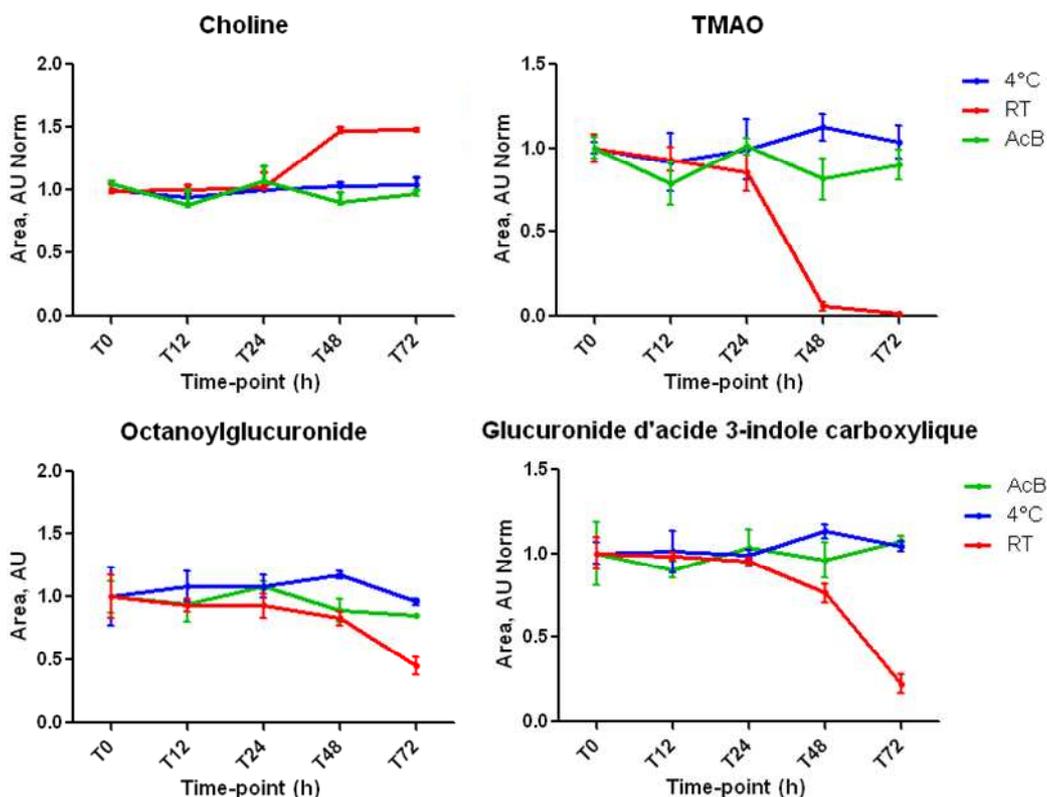


Figure 67 : Evolution de certains métabolites avec la contamination bactérienne pour l'expérience de Septembre 2010.

### 3.3.2 Métabolites instables en solution

Certains métabolites (8 sur les 237 observés), en plus de présenter des variations de concentrations à RT, en présentent aussi à 4°C et/ou AcB. Ces variations ne peuvent pas être expliquées uniquement par la contamination bactérienne car elle n'est pas présente dans ces conditions. Il s'agit donc d'un problème d'instabilité de ces métabolites en solution. Cela concerne l'acide ascorbique, la thréonolactone, L-urobilinogène et L-urobiline, l'acide pyruvique,  $\alpha$ -kétoglutarique, l'acide malonique et la N-Acétylphénylalanine. Pour tous ces métabolites, la variation va toujours dans le même sens (augmentation ou diminution) et, dans certains cas, elle est plus importante à RT qu'à AcB et enfin elle est minimale à 4°C, comme c'est par exemple le cas pour

l'acide ascorbique et son produit d'auto-oxydation la thréonolactone (Figure 68) : la dégradation par oxydation de l'acide ascorbique est limitée à 4°C, elle est plus importante pour AcB (conservateur mais à température ambiante) et atteint son maximum à RT. L'apparition de son produit de dégradation suit la même tendance. De la même manière, l'acide malonique, qui semble être un produit de dégradation des pyrimidines [306], augmente plus à RT qu'à AcB. Nous pouvons alors conclure que la contamination bactérienne et l'instabilité chimique sont toutes les deux responsables de ces variations. Pour les autres molécules, les variations sont similaires à RT et AcB, et limitées à 4°C. Dans ces cas précis la variation observée à RT n'est en rien due à la présence des bactéries, mais bien à une instabilité chimique dépendante de la température. Cette instabilité chimique est observée pour le L-urobilinogène et L-urobiline (Figure 68), l'acide pyruvique,  $\alpha$ -kétoglutarique et la N-Acétylphénylalanine.

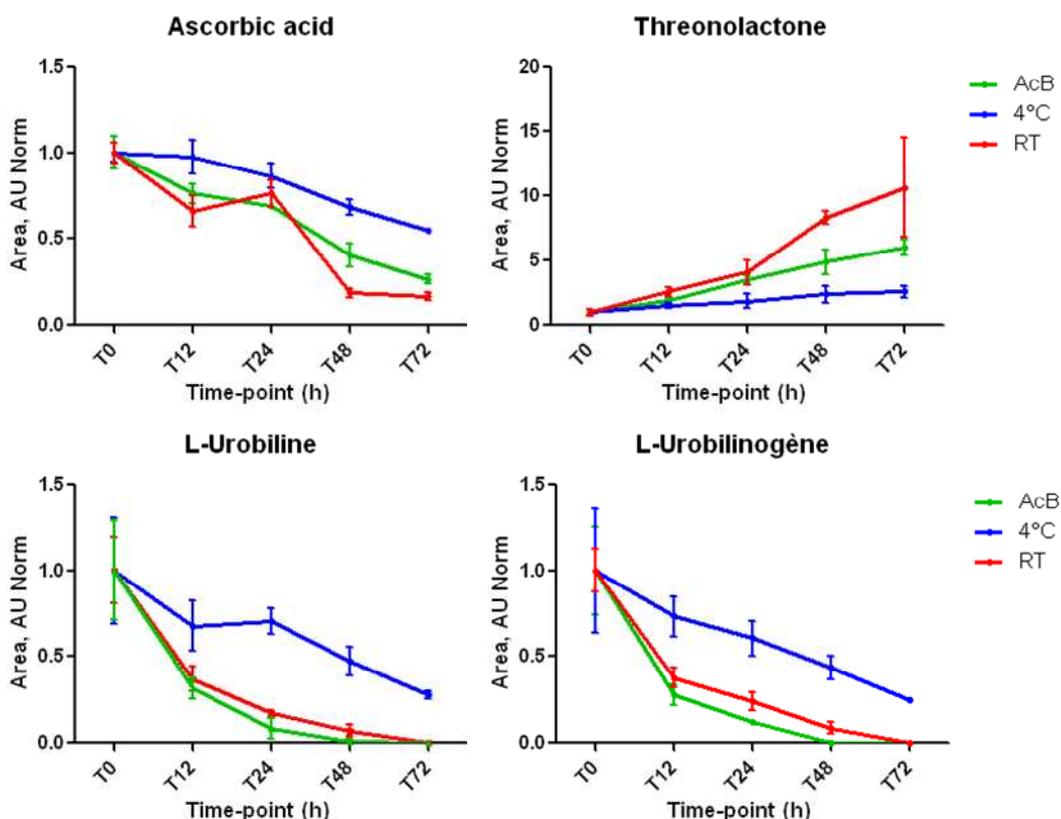


Figure 68 : Evolution de certains métabolites instables en solution pour l'expérience de Septembre 2010.

## 4 CONCLUSION ET DISCUSSION

Les conditions de prélèvement des biofluides peuvent avoir un impact qualitatif et quantitatif important sur le contenu des empreintes métaboliques. Nous l'avons démontré dans cette étude en analysant d'une part l'évolution et les conséquences de la contamination bactérienne dans des échantillons d'urines, mais aussi l'instabilité potentielle des métabolites en solution. Plusieurs facteurs susceptibles de modifier le contenu en métabolites des échantillons urinaires lors de leur prélèvement

ont été analysés : la température (RT ou 4°C) et la présence ou non de conservateurs. Leur impact a été évalué sur les métabolites identifiés grâce à la base de données spectrale. Trente six métabolites sur 237 détectés (environ 15%) montrent des variations de concentrations significatives en fonction des conditions de prélèvement et de conservation, qu'elles soient liées à la contamination bactérienne ou à leur instabilité en solution. La contamination bactérienne est bien présente mais est très sensible aux conditions expérimentales et en particulier à la température. Malgré cela, quand la contamination bactérienne est avérée, les résultats sont relativement robustes : ce sont les mêmes métabolites dont les concentrations sont modifiées, à des degrés différents en fonction de l'importance de cette contamination. A titre d'exemple, les concentrations de créatine et des acides lactique, citrique et malonique ne sont significativement impactés que dans l'expérience 3, celle pour laquelle le nombre de bactéries est le plus important. Il n'est pas non plus exclu que la nature des métabolites dont les concentrations ont été modifiées par la croissance bactérienne dans nos échantillons puisse varier en fonction du type de bactéries contaminantes, mais nous n'avons pas réalisé d'expériences d'identification bactérienne dans le cadre de cette étude. Outre les métabolites déjà connus pour être liés à la croissance bactérienne, on retrouve d'autres métabolites : vingt huit au total sur les trente six métabolites discriminants, comme ceux qui sont liés au métabolisme de la choline, ou encore des glucuroconjugués. Concernant l'instabilité en solution, elle est elle aussi très dépendante de la température mais concerne un nombre limité de métabolites : huit au total sur les trente six discriminants, comme l'acide ascorbique ou la L-urobiline. La visualisation des variables discriminantes du jeu de donnée en analyses statistiques multivariées (données non présentées) montrent qu'il y a d'autres signaux (10-20% du nombre total de signaux suivant l'expérience et le mode d'ionisation) dont les intensités varient avec la contamination bactérienne et l'instabilité chimique, mais qui n'ont pas encore été annotées ou identifiées.

Les conditions de prélèvement ont bien un impact sur la composition chimique des urines. Elles doivent donc être standardisées de manière à limiter les biais expérimentaux. D'une manière générale, les premiers effets du temps apparaissent après 24 heures. Mais ceci est bien évidemment dépendant de la température ambiante qui peut parfois être très élevée. Pour les prélèvements de moins 24h (24h inclus) il n'est pas, à priori, nécessaire de rajouter des conservateurs ou de faire le prélèvement à 4°C. L'ajout de conservateur se révèle efficace pour limiter l'impact de la contamination bactérienne, mais ne suffit pas pour limiter l'instabilité en solution, contrairement au prélèvement à 4°C (sur glace) qui empêche à la fois la prolifération bactérienne et la dégradation des métabolites instables en solution. C'est donc ce mode de prélèvement que nous préconisons.

Nos résultats sont en accord avec une étude réalisée sur des urines de rongeur en GC/MS [307] : le prélèvement en cage à métabolisme à 4°C limite les variations des métabolites dans les urines. D'autre part, l'ajout de conservateur modifie la composition chimique de l'échantillon, et, dans le cas de l'acide borique, l'acidification des échantillons a des répercussions sur les empreintes métaboliques enregistrées. Il est donc préférable d'éviter d'utiliser ces conservateurs. Si toutefois un doute subsiste sur les méthodes de prélèvement de certains échantillons analysés, il ne faudra pas prendre en compte dans le traitement des données, les métabolites que nous avons décrits comme étant instables ou impactés par la croissance bactérienne.

Cette étude des conditions de prélèvement a été réalisée en priorité à cause du manque flagrant de données dans la littérature et de l'impact potentiellement important de la contamination bactérienne sur les empreintes métaboliques. Cependant la stabilité à court terme des échantillons traités, les études de congélation/décongélation ainsi que la stabilité à long terme des urines, bien que déjà étudiées et publiées plusieurs fois (Tableau 14), doivent être revisitées. En effet les données publiées sont principalement basées sur des approches chimiométriques ou sur la comparaison visuelle de spectres, et utilisent des urines individuelles ce qui, à cause de la forte variabilité inter-individuelle, peut limiter la mise en évidence de la variabilité liée aux problèmes de stabilité [298;300]. Jusqu'à maintenant, les études de stabilité à long terme en LC/MS étaient difficiles à mettre en œuvre, tout simplement à cause du manque de reproductibilité analytique ce qui, quand des échantillons sont analysés à plusieurs mois d'intervalle, rend la comparaison des empreintes quasiment impossible. Grâce aux techniques de normalisation récemment décrites et à l'utilisation de QC, l'exploitation des données de stabilité à long terme sera plus simple et plus efficace. L'application de ces techniques de normalisation à l'étude de la stabilité d'autres biofluides comme le plasma ou le LCR est également envisageable. Cependant, chaque biofluide ayant des contraintes liées par exemple à leur composition biochimique (présence d'enzymes, par exemple), les études de stabilité doivent tenir compte des conditions réelles de prélèvement et de conservation de ces biofluides : le type d'anticoagulants pour le prélèvement du sang, le type de tube..., chacun de ces paramètres pouvant avoir un impact sur la composition chimique des échantillons biologiques.



DISCUSSION GENERALE ET  
PERSPECTIVES



## DISCUSSION GENERALE ET PERSPECTIVES

La métabolomique est une discipline relativement récente. Malgré les évolutions technologiques de ces dernières années réalisées dans les domaines de la chimie analytique, avec l'avènement de la spectrométrie de masse haute résolution, et du traitement des données, il reste encore de nombreux obstacles à surmonter. L'obtention de données biologiquement exploitables à partir des profils métaboliques issus d'analyses LC/MS reste limitée. En effet, ces profils peuvent contenir plusieurs centaines à quelques milliers de signaux dont la majeure partie demeure jusqu'à présent inconnue, ou du moins inexploitée car non caractérisée par les systèmes analytiques utilisés. De plus l'analyse et la manipulation de ces données est très coûteuse en temps et reste encore trop souvent éloignée de l'interprétation biologique.

Dans ce contexte, le premier objectif de ce travail de thèse a été de construire une base de données spectrale pour l'identification automatique des métabolites dans les fluides biologiques. Pour cela, deux approches complémentaires ont été développées. La première est basée sur l'analyse de composés de référence, et a permis de développer des outils d'interprétation spectrale et d'annotation automatique performants. Nous avons, grâce à ces outils, développé une base de données spectrale et avons eu à faire face à un problème majeur : l'indisponibilité commerciale d'un grand nombre de composés potentiellement détectables et biologiquement pertinents. Ce constat nous a amené à développer un deuxième type d'approche basé sur l'identification de métabolites directement à partir des données biologiques. Ces données étant très complexes (multitudes de signaux, bruit de fond important...), des outils de regroupement de variables et d'annotation différentielle ont été développés et optimisés pour structurer les jeux de données biologiques, c'est-à-dire regrouper les ions produits par un métabolite donné, afin de faciliter l'identification des métabolites. Avec les mesures de masses précises et la haute résolution de la spectrométrie de masse, complétées par des expériences complémentaires de  $MS^n$ , nous avons réussi à identifier 81 molécules non disponibles commercialement, en plus des métabolites pour lesquels nous disposons des composés de référence dans notre chimiothèque. Finalement, 244 métabolites ont été identifiés dans les échantillons d'urine de notre cohorte de salariés du CEA-Saclay. La présence dans les urines de 78 d'entre eux n'est pas répertoriée dans la base de données HMDB. Signalons cependant des publications, encore non prise en compte par HMDB, font état de la détection dans des urines humaines de 33 de ces métabolites, principalement des acylcarnitines.

La deuxième partie de ce travail de thèse a consisté à utiliser les identifications obtenues sur la cohorte de salariés du CEA pour étudier l'impact de 3 facteurs physiologiques (genre, âge et indice de masse corporelle) sur les concentrations urinaires des métabolites identifiés. L'interprétation des données sur une population aussi importante représente un challenge, notamment à cause de la dérive analytique générée lors des analyses LC/MS d'un grand nombre d'échantillons. Ainsi, notre première tâche a été d'évaluer ce biais susceptible de gêner l'étude de l'impact des facteurs physiologiques sur le contenu des empreintes métaboliques, afin de pouvoir ensuite nous en

affranchir. Une importance particulière a donc été apportée à la normalisation des données, mais également aux tests statistiques appliqués (choix des tests, correction des tests multiples). Ce travail a permis d'appliquer une méthodologie de normalisation pertinente (LOWESS) pour l'étude d'un grand nombre d'échantillons, mais aussi d'affiner le choix des tests statistiques en fonction des variables étudiées. Nous avons ainsi pu mettre en évidence une cinquantaine de métabolites impactés par le genre, l'âge ou l'IMC. La plus grande source de variabilité biologique est liée à l'effet du genre, avec des différences au niveau hormonal et au niveau du catabolisme (lipides, protéines...) et du métabolisme (acides aminés) général, puis l'âge avec principalement une diminution du métabolisme général et une diminution de l'efficacité de métabolisation des composés exogènes, notamment de la caféine. De manière surprenante, aucune variation n'a pu être attribuée spécifiquement à l'IMC. Cependant l'IMC est lié à l'âge et au genre, et cette dépendance peut interférer dans les analyses statistiques. Un biomarqueur pouvant naître d'une interaction entre deux facteurs, des tests de types ANOVA sont envisagés afin d'étudier les interactions entre facteurs, mais tout en tenant compte du fait que certains sont déjà liés entre eux (IMC/âge et IMC/genre). Enfin la connaissance de la variabilité métabolique inter-individuelle d'une population «tout venant» va permettre de faciliter l'interprétation des données de métabolomique d'un point de vue biologique. Il existe en effet un fossé entre l'identification d'un métabolite et la compréhension de son rôle d'un point de vue du métabolisme d'un organisme. Cette étude va donc nous permettre de poser les fondements d'une base de données relationnelle, qui fera à terme le lien entre un métabolite et un contexte biologique.

Enfin la dernière partie de mon travail de thèse s'est focalisée sur une étape en amont de l'analyse métabolomique mais qui a toute son importance dans la pertinence des résultats : les conditions de recueil des échantillons d'urine. La composition chimique des urines peut en effet être fortement influencée par la façon dont elles ont été prélevées. Différentes conditions de prélèvements ont été étudiées, et leur impact sur les empreintes métabolique a été évalué. Cette étude a permis d'analyser les conséquences de la prolifération bactérienne sur la composition en métabolite des urines, et également de mettre en évidence des phénomènes d'instabilité de certains métabolites en solution. Ces variations de concentration en métabolites sont assez reproductibles d'une expérience à l'autre, et semblent être proportionnelles au nombre de bactéries présentes dans l'échantillon. Il est probable que la nature des métabolites mis en évidence dans ce type d'étude dépende en partie du genre ou d'espèce des bactéries présentes dans la flore commensale. Il faudrait, pour le confirmer, réaliser le même type d'études dans d'autres laboratoires et être en mesure d'identifier les bactéries présentes dans d'autres échantillons urinaires.

La prolifération bactérienne est très sensible à la température, tout comme la stabilité chimique de certains métabolites. Ainsi, un prélèvement réalisés en présence de conservateurs ou à 4°C permet d'éviter la prolifération bactérienne. Cependant, seul le prélèvement à 4°C permet de limiter la dégradation des métabolites instables en solution. Les conditions optimales de recueil des urines ont pu être déterminées (prélèvement sur glace à 4°C) afin de limiter au maximum la variabilité liée aux méthodes de prélèvements.

En définitive, les résultats obtenus lors de ce travail de thèse démontrent l'intérêt des méthodologies mises en place pour l'identification de métabolites à partir de la base de données spectrale, et de l'utilisation de ces identifications pour étudier les variations de concentrations de ces métabolites aussi bien pour les études de grande envergure, que pour des études avec un nombre plus restreint d'échantillons. Cependant, les méthodes et outils développés dans le cadre de cette thèse présentent un certain nombre de limitations. Tout d'abord, une des principales limitations de la spectrométrie de masse basée sur l'ionisation à pression atmosphérique est le manque de reproductibilité d'un appareil à l'autre. Bien que la base de données ESI-MS soit une base d'annotation qui ne tienne pas compte des intensités relatives des ions, le degré d'homologie des spectres de masses produits par des systèmes analytiques différents reste à évaluer.

Il existe également des bases de données de MS/MS. Elles sont indispensables pour l'identification des métabolites et la construction de bases d'annotation en ESI-MS. Cependant celles-ci tiennent compte à la fois des ions produits et de leurs intensités relatives et il est délicat, voire impossible, de comparer des spectres MS/MS générés par des appareils différents, ou même par des laboratoires différents sur un même type d'appareil. La constitution d'une chimiothèque la plus complète possible coûte cher, et le partage de données MS/MS entre laboratoires est une bonne alternative pour favoriser l'enrichissement d'une telles bases de données. Il existe actuellement une initiative de standardisation des conditions d'acquisition des spectres MS/MS au sein du réseau français de métabolomique et de fluxomique (RFMF) [253]. Le but est d'évaluer la variabilité interlaboratoire et inter-instrumentale pour ensuite normaliser les conditions d'acquisition des spectres et faciliter le partage au sein de la communauté métabolomique.

Comme cela a déjà été évoqué, une des principales limitations des analyses métabolomiques par LC/MS est le manque de robustesse des méthodes d'ionisation à pression atmosphérique telles que l'ESI, qui rend délicate l'intégration d'expériences réalisées à des jours différents au sein d'un même jeu de données. L'utilisation de QC analysés à intervalles réguliers tout au long du déroulement de l'expérience, comme présenté dans ce travail de thèse, est une solution séduisante, qui reste néanmoins limitée aux métabolites majoritaires qui seront détectés ce type d'échantillons. Cette approche impose également de toujours utiliser un aliquot du même QC pour des analyses différentes, ce qui pose le problème de la disponibilité et de la stabilité à long terme des métabolites dans les extraits biologiques. Le développement de méthodes de quantification absolue à large échelle, grâce à l'utilisation de métabolites marqués, pourrait constituer une réponse à ce problème, mais aussi à celui de la comparaison de résultats obtenus avec des méthodes analytiques différentes. En effet il n'existe pas de méthode universelle d'analyse du métabolome, et pour avoir une image la plus exhaustive du métabolome urinaire, il faudrait utiliser des méthodes analytiques complémentaires telles que la RMN et la GC/MS, mais aussi développer des outils pour pouvoir comparer et mettre en commun les informations issues de ces différentes techniques. L'utilisation en routine des méthodes de quantification absolue reste actuellement limitée par le manque de disponibilité commerciale des métabolites marqués avec des isotopes stables. Collaborer avec des équipes de chimistes organiciens pour réaliser des synthèses chimiques est une alternative, et, bien qu'elle présente de

nombreuses limitations, elle est indispensable à moyen terme, aussi bien pour quantifier que pour confirmer l'identification de molécules non commercialement disponibles.

A côté des problèmes d'instrumentation et de chimie analytique, l'informatique reste un des principaux facteurs limitant l'utilisation des données d'analyse métabolomique : gestion de gros volumes de données, nécessité de développer des outils automatiques de traitement et d'analyse des données, et développement d'outils mathématiques et bioinformatiques pour visualiser les réseaux métaboliques et pour faciliter l'interprétation biologique.

Concernant le traitement des données, il existe maintenant des algorithmes permettant la détection automatique des ions, l'élimination du bruit de fond, l'alignement et le regroupement des signaux, ou encore l'annotation automatique à partir des bases de données. Un des enjeux actuels est de « chaîner » ces différents algorithmes afin de rendre le traitement des données plus efficace et convivial. Ainsi des outils intégrés d'analyse métabolomique, comme MZmine ou MetaboAnalyst, qui regroupent un certain nombre de ces modules et algorithmes développés ces trois dernières années, mais les résultats intermédiaires ne sont pas toujours accessibles ce qui limite la vérification des données entre chaque étape. De plus, la performance de chacun des modules est difficile à évaluer. Par exemple les algorithmes de détection des pics sont différents d'un logiciel à l'autre, et le seul critère actuellement utilisé pour évaluer leur efficacité est le nombre de pics détectés, parmi lesquels figurent un grand nombre d'ions provenant du bruit de fond. Il est donc important de pouvoir distinguer les pics analytiquement pertinents des autres, et un des moyens envisageable est d'annoter un maximum d'ions dans ces échantillons. Ce qui nous ramène à la nécessité de développer des bases spectrales ESI-MS d'annotation, aussi bien pour aider à l'interprétation biologique, que pour l'évaluation des performances des outils informatiques.

Se pose également le problème du partage des données. Si l'accès et l'échange de données brutes non traitées a été facilité ces dernières années grâce au développement de formats universels (mzXML, mzData, mzML, netCDF...) compatibles avec de nombreux logiciels commerciaux et mais aussi libres et gratuits, il n'en est pas de même pour les données traitées. Il n'existe pas encore de format reconnu et accepté par la communauté pour partager facilement les grandes quantités de données métabolomiques d'analyses LC/MS issues du traitement par des logiciels différents. Un certain nombre d'initiatives visant à uniformiser le stockage des données traitées de LC/MS ont déjà été publiées mais elles n'étaient jusqu'à maintenant que peu adaptées aux données de métabolomique. Citons par exemple le format FeatureXML [308], bien adapté pour les données LC/MS, mais plutôt destinées à des applications dans le domaine de la protéomique. Récemment un format plus complet, PeakML, développé spécifiquement pour les données métabolomiques, a été décrit par Schelteme et al. [309]. Ce format est compatible avec la plupart des logiciels existants (par conversion) et permet de stocker à la fois les informations sur l'acquisition (mode d'ionisation, séquence, données brutes...) et le traitement des données (mesure d'intensité, annotations et identifications...). De plus, le développement de modèles formels de déroulement d'analyse ainsi qu'une standardisation des méthodes d'analyse et de traitement de l'information au sein de la communauté devraient faciliter l'utilisation de ce type de données par les biologistes.

Enfin, l'objectif des analyses métabolomique est de produire des données biologiquement exploitables, notamment en révélant des interrelations métaboliques qui auraient été difficilement détectables avec des approches ciblées conventionnelles. Cela passe par le développement de nouvelles méthodes de visualisation des réseaux métaboliques. Un moyen classique pour étudier le métabolisme consiste à se concentrer sur de petits sous-ensembles de réactions biochimiques qui sont appelées voies métaboliques. La principale limite de cette approche est que les voies métaboliques sont considérées comme des processus distincts, représentés par des cartes métaboliques répertoriées dans des bases de données telles que KEGG ou MetaCyc. Cette vision statique et simpliste est erronée puisque la plupart des métabolites sont impliqués dans plusieurs voies. Dans une perspective de biologie des systèmes, le métabolisme est analysé au niveau global en intégrant le plus grand nombre possible de voies métaboliques en un seul réseau. En effet, grâce à l'annotation fonctionnelle de génomes complets et aux progrès de la protéomique et de la métabolomique, il est aujourd'hui envisageable d'étudier le réseau métabolique «théorique» d'un organisme dans sa globalité. L'avantage de cette approche est de permettre d'étudier les interactions entre les voies métaboliques, et de faciliter la découverte de voies alternatives. Dans ce contexte, la bioinformatique apportera là encore une contribution indispensable à l'analyse globale des réseaux métaboliques, de la reconstruction automatique de celui-ci sous forme de graphes, au développement de nouveaux algorithmes d'analyse de ces réseaux.



# ANNEXES



# ANNEXE 1 - SPECTROMETRIE DE MASSE :

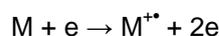
## RAPPELS THEORIQUES

### 1. LES DIFFERENTES SOURCES D'IONISATION

#### 1.1. Les sources basées sur l'ionisation en phase gazeuse

- SOUS VIDE
  - L'impact électronique

Cette source inventée par Dempster et perfectionnée par Bleakney [310] et Nier [311] est parmi les plus couramment utilisées en spectrométrie de masse couplée à la chromatographie en phase gazeuse. Elle est constituée d'un filament chauffé qui émet des électrons. Ceux-ci sont accélérés vers une anode à l'aide d'une différence de potentiel et entrent en interaction avec les molécules en phase gazeuse de l'échantillon analysé introduit dans la source. Lorsque l'énergie cinétique ( $E_c$ ) des électrons est de l'ordre de 70 eV, les premiers ions formés résultent de l'arrachement d'un électron à la molécule M. En moyenne, à 70 eV, on produit dans les conditions habituelles de pression, un ion pour mille molécules entrant dans la source.



L'ion  $M^{+\bullet}$  est appelé « ion moléculaire », on notera qu'il s'agit d'un cation-radical, c'est-à-dire d'une entité portant une charge positive et un électron célibataire. L'énergie d'ionisation (EI), d'une molécule organique se situe autour de 10 eV. Or, lors du processus d'ionisation, de 10 à 20 eV sont transférés aux molécules. Cet excédent d'énergie interne peut provoquer la dissociation des espèces moléculaires et par conséquent la formation de fragments  $A^+$  (cation) et  $B^\bullet$  (radical neutre) ou  $A^{+\bullet}$  (radical cation) et B (molécule neutre)



Des processus analogues peuvent se produire de manière compétitive ou consécutive et conduire ainsi à de nombreux ions fragments de rapports  $m/z$  inférieurs à la masse nominale de la molécule. Cette technique induit donc une fragmentation importante de telle sorte que l'espèce moléculaire n'est pas toujours observée. Dans les conditions habituelles d'une source à impact électronique, la formation d'ions négatifs est inefficace par rapport à la formation d'ions positifs.

Cette source est principalement employée en couplage avec la chromatographie en phase gazeuse. L'énergie transférée pour l'ionisation (70 eV), permet l'obtention de spectres de fragmentations robustes qui ont permis la constitution de banques de spectres universelles, facilitant ainsi l'identification des composés.

- **L'ionisation chimique**

L'ionisation chimique [312] consiste à produire des ions par collision de la molécule à ioniser avec des ions primaires présents dans la source. Il existe dans la source une petite enceinte d'environ 1 cm de côté où règne une pression de 60 Pa. Hors de celle-ci, la pression est de l'ordre de  $10^{-3}$  Pa. Lorsqu'un électron entre dans l'enceinte, il ionise préférentiellement le gaz réactif (méthane, ammoniac, ...) par ionisation électronique. L'ion ainsi produit entrera en collision avec d'autres molécules du gaz réactif, formant par une série de réactions consécutives un plasma d'ionisation. Les ions de la substance à analyser, tant positifs que négatifs, se formeront par réactions chimiques avec les ions de ce plasma, donnant lieu à des réactions de transfert de protons, d'abstraction d'hydrure, d'addition, de transfert de charges, etc. Il s'agit d'une méthode d'ionisation douce, conduisant à une faible fragmentation de l'espèce moléculaire. Cette source est également utilisée en couplage avec la chromatographie en phase gazeuse.

- **A PRESSION ATMOSPHERIQUE**

Ces sources ionisent l'échantillon à pression atmosphérique puis transfèrent les ions produits vers le spectromètre de masse. Les échantillons liquides sont directement introduits dans un nébuliseur. Sous l'effet d'un jet d'air ou d'azote, le liquide est transformé en fin brouillard. Un chauffage assure la désolvatation des composés. Ces derniers sont ensuite ionisés à pression atmosphérique. Une interface est nécessaire pour transférer les ions formés à pression atmosphérique vers le vide poussé de l'analyseur de masse. Une technique de pompage différentiel sur un ou plusieurs étages de pompage séparés entre eux par des écorceurs (ou skimmers) à l'aide de pompes de très grande capacité est alors utilisée. Le transfert des ions de la pression atmosphérique vers le vide du spectromètre induit nécessairement des pertes d'ions, liées à l'étroitesse des orifices de transfert qui doivent permettre de maintenir les différences de pression d'un compartiment à l'autre. Mais ces pertes sont compensées par le meilleur rendement d'ionisation. L'avantage de ces sources est la simplicité avec laquelle elles permettent le couplage direct au spectromètre de masse des différentes techniques de séparation en phase liquide (HPLC, CE). Les sources basées sur l'ionisation en phase gazeuse à pression atmosphérique sont le DART (*Direct Analysis in Real Time*) [313], l'APCI et l'APPI. Ces deux dernières techniques sont détaillées ci-dessous.

- **L'APCI (Ionisation chimique à pression atmosphérique) [314]**

Cette technique, employée en couplage LC/MS, fait appel à des réactions ions-molécules en phase gazeuse à pression atmosphérique. C'est donc une méthode analogue à l'ionisation chimique où les ions primaires sont produits par des décharges corona sur un aérosol de solvant contenant les analytes. La source APCI est principalement appliquée à l'analyse des composés polaires et ioniques avec une masse moléculaire inférieure à environ 1500 Da, même si elle est capable d'analyser des composés relativement non polaires. Elle utilise un nébuliseur pneumatique à air ou à azote pour produire un fin aérosol au sein d'une chambre chauffée (350 à 500 °C), le vaporisateur. Au sein de la source, une décharge corona ionise l'air ambiant et crée un plasma, i.e. un 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. En effet, les molécules de solvant, en large excès, sont ionisées préférentiellement et jouent le rôle d'ions réactifs. En ionisation en mode positif, le transfert de proton à partir des ions du solvant conduit à la formation d'ions protonés de l'analyte, suivant les affinités protoniques en phase gazeuse du solvant et de l'analyte. En ionisation en mode négatif, les ions de l'espèce moléculaire sont obtenus par transfert d'électron ou par abstraction de proton. La formation d'ions adduits peut se produire dans les deux modes. Cette technique d'ionisation a été utilisée dans le cadre d'analyses métabolomiques visant à identifier des métabolites dont la concentration varie avec les étapes de maturation d'un fruit [81]. Les échantillons sont également analysés en parallèle avec une source électrospray. Grâce à la complémentarité des deux sources, plus de 1,000 métabolites ont ainsi pu être détectés. L'APCI est en effet utilisée pour analyser des métabolites moins polaires que ceux analysés en électrospray ou pour améliorer la sensibilité [91].

- **L'APPI (photoionisation à pression atmosphérique) [92]**

L'échantillon en solution est vaporisé à l'aide d'un nébuliseur chauffant. Après vaporisation, l'analyte interagit avec des photons émis par une lampe à décharge. Ces photons induisent alors une série de réactions en phase gazeuse qui conduit à l'ionisation de l'analyte. L'ionisation directe de l'analyte se caractérise généralement par une faible efficacité due en partie à l'absorption des photons par les molécules de solvant. Dès lors, l'APPI utilisant un dopant a été décrite. Les dopants les plus utilisés sont le toluène et l'acétone car ils sont photoionisables et agissent comme intermédiaires entre les photons et l'analyte. Leur utilisation améliore la formation d'ions négatifs [315]. Les spectres de masse de type APPI obtenus en mode positif sont caractérisés par la formation de deux types d'ions de l'espèce moléculaire qui peuvent coexister [316]. Ces deux espèces sont l'ion moléculaire  $M^{*+}$  et l'ion moléculaire protoné  $[M+H]^+$ . Cette technique s'adresse principalement à des composés peu polaires connus pour être difficilement ionisables en ESI ou en APCI. Elle est notamment performante pour des composés comme des flavonoïdes, les stéroïdes [93], les pesticides [94;95], ou les hydrocarbures aromatiques polycycliques [96] et est utilisée en couplage LC/MS.

## 1.2. Les sources basées sur la désorption

- **SOUS VIDE**

- **Le MALDI**

Introduite en 1988 par Karas et Hillenkamp, la méthode de désorption-ionisation laser assistée par matrice (MALDI) a été développée afin d'analyser les molécules de haute masse moléculaire [97]. Elle consiste à mélanger la substance à analyser à une solution de petites molécules organiques, appelée matrice. L'irradiation de cette matrice, capable d'absorber à la même longueur d'onde que le laser, va provoquer l'accumulation d'une importante quantité d'énergie dans la phase condensée par excitation électronique des molécules de la matrice. Il en résulte alors la formation d'un panache où se

produit la désorption des ions. Des échanges de protons au sein des agrégats qui constituent le panache (matière photo-excitée) permettent l'ionisation de la substance analysée (Figure 69). Les matrices utilisées en MALDI sont des chromophores de faible enthalpie de sublimation. Les plus utilisées sont les dérivés des acides cinnamiques et benzoïques : l'acide  $\alpha$ -cyano-4-hydroxycinnamique (HCCA), l'acide sinapinique (SA), l'acide ferrulique, l'acide caféique et l'acide gentisique (DHB). Leur rôle est multiple. Elles permettent en phase condensée de favoriser la séparation des molécules à étudier en réduisant les forces intermoléculaires et d'absorber l'énergie du faisceau laser. De plus, elles favorisent l'ionisation en induisant des transferts de protons. Le choix de la matrice est donc important.

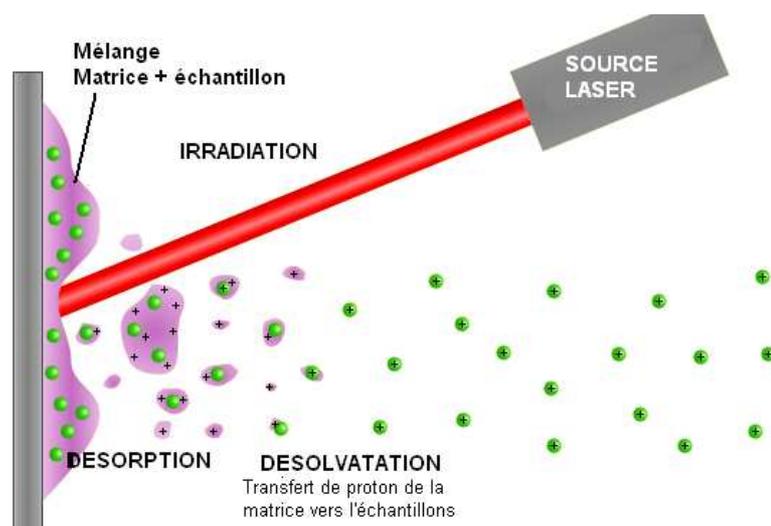


Figure 69 : Schéma du principe de l'ionisation MALDI

(image issue du site internet <http://fr.wikipedia.org/wiki/Fichier:Maldi.png>)

La méthode de désorption-ionisation laser assistée par matrice (MALDI) est à l'heure actuelle couramment utilisée dans les analyses protéomiques mais très peu d'applications pour l'analyse du métabolome ont été réalisées à ce jour en raison de l'importance des interférences de la matrice dans la détection des composés de faible masse moléculaire. De plus, cette méthode étant peu sensible en ionisation en mode négatif avec les matrices couramment utilisées, elle ne permet généralement pas la détection de molécules telles que les acides organiques ou les composés phosphorylés s'ionisant préférentiellement en mode négatif [317]. Cependant, de récentes avancées technologiques ont permis de rendre possible l'analyse en MALDI des composés de faible masse moléculaire ou s'ionisant préférentiellement en ionisation en mode négatif [318;319]. Des travaux effectués sur des extraits de bactéries ou de tissus [320] ont montré que cette méthode pouvait être utilisée pour la détection de métabolites dans des extraits biologiques complexes suggérant ainsi son utilité pour l'analyse du métabolome. Les approches MALDI sont notamment très utilisées pour l'analyse des protéines ou des lipides, que ce soit en analyse lipidomique [321] ou en imagerie [322;323].

Bien que les conditions sous vide soient plus adaptées pour créer et conserver les ions, une version à pression atmosphérique de cette méthode existe (AP-MALDI) [324]. Elle permet d'étudier des systèmes fragiles, ainsi que de coupler cette technique à des analyseurs tels que les pièges à ions et ainsi de faire des expériences de fragmentations séquentielles ( $MS^n$ ).

- **Désorption laser sur surfaces modifiées ou non**

La désorption/ionisation laser directe d'analytes déposés sans matrice a été étudiée sur toutes sortes de surfaces sans rencontrer de réel succès car la dégradation de l'échantillon est habituellement observée. Cependant des résultats encourageants ont été obtenus par la technique désignée sous l'acronyme SALDI (*Surface-Activated Laser Desorption Ionisation*) [327] qui utilise le graphite comme surface. Mais l'utilisation du silicium poreux comme nouveau support s'est avéré encore plus prometteuse et a conduit au développement d'une nouvelle technique dénommée DIOS pour *Desorption/Ionisation On Silicon* [328]

La désorption/ionisation sur silicium poreux (DIOS), a récemment été développée par l'équipe de Siuzdak [329]. Le DIOS ne nécessite pas de matrice et se présente comme une alternative intéressante pour l'étude de molécules de masse moléculaire inférieure à 3000 Da et de polarités variables. Ce sont donc essentiellement des petits systèmes qui sont étudiés, tels que les peptides, les acides gras, des stupéfiants [330]. Le DIOS a également été utilisé dans les analyses protéomiques pour la caractérisation fonctionnelle et l'identification de petites protéines [98]. L'échantillon est simplement déposé sur du silicium poreux, aucune matrice organique n'est ajoutée. Le silicium poreux permet de piéger la molécule analyte et du fait de sa forte absorption dans l'UV permet à l'échantillon d'être désorbé et ionisé sous irradiation laser UV. Le DIOS présente une limite de sensibilité équivalente à la technique du MALDI mais est en plus très tolérante vis-à-vis de la présence de sels ou de tampon. Il présente cependant une faible robustesse due aux problèmes de conservation des cibles. Les publications sont principalement d'ordre méthodologique et peu d'études sur des analyses en milieux biologiques ont été publiées. Une récente étude a montré l'intérêt d'utiliser plusieurs sources d'ionisation pour élargir le nombre de composés détectés dans les empreintes métaboliques. Cette étude a montré l'intérêt du DIOS qui a fourni environ 50% des ions détectés par un seul type d'ionisation [331].

- **A PRESSION ATMOSPHERIQUE**

- **Les sources ESI (*Electrospray Ionisation*)**

Cette technique, développée par J.B. Fenn au milieu des années 1980 [99;100;325;326], est utilisée en spectrométrie de masse pour produire des ions à partir de composés en solution. L'électrospray est produit par application à pression atmosphérique d'un fort champ électrique sur un liquide traversant un tube capillaire avec un faible débit. Le champ électrique est obtenu par application d'une différence de potentiel de 2 à 6 kV entre ce capillaire et la contre-électrode. Il provoque l'accumulation de charges à la surface du liquide, situé à l'extrémité du capillaire, qui va se rompre pour former des gouttelettes fortement chargées. L'évaporation du solvant contenu dans ces

gouttelettes va provoquer la réduction de leur taille jusqu'au moment où les forces coulombiennes vont approcher le niveau des forces de cohésions des gouttelettes et provoquer leur explosion.

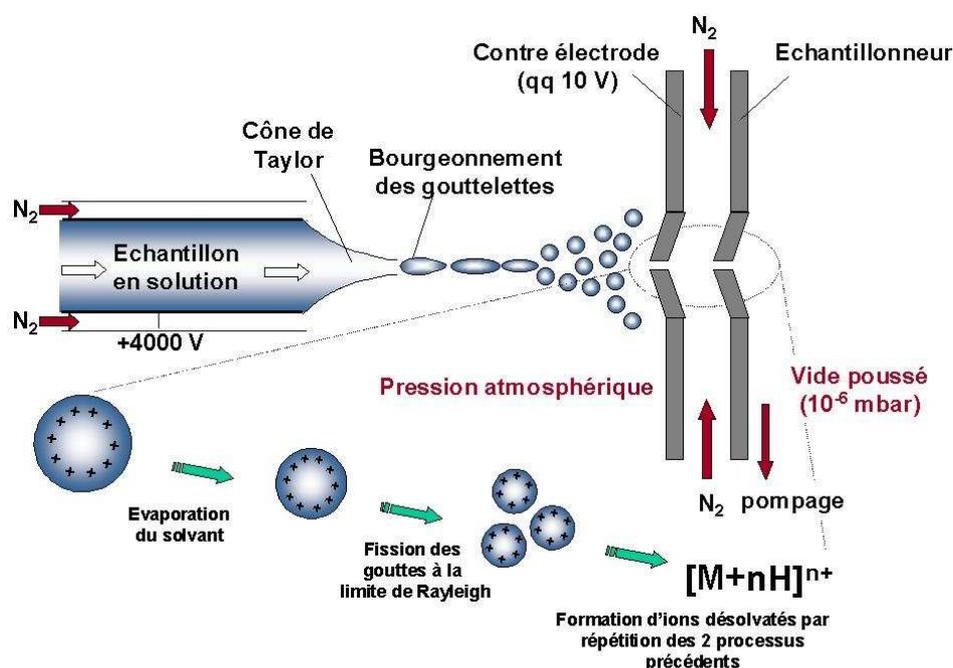


Figure 70 : Schéma d'une source électrospray

(image issue du site internet [www.pierre-de-lune.fr/cjism/fiche\\_esi.php](http://www.pierre-de-lune.fr/cjism/fiche_esi.php))

Une cascade de fissions conduit à des gouttelettes de plus en plus petites, jusqu'au moment où le champ électrique à leur surface devient suffisant pour provoquer la désorption des ions. Les ions ainsi produits sont accélérés vers l'analyseur alors que les molécules non chargées sont éliminées (Figure 70). Les états de charges sont plus nombreux qu'en MALDI. Cette technique d'ionisation douce, permettant une faible fragmentation de l'espèce moléculaire et donnant ainsi accès à la masse moléculaire des analytes, est l'une des techniques les plus utilisées pour les études métabolomiques en couplage LC/MS.

#### o Le DESI (Désorption- Electropray Ionisation)

La technique DESI implique l'utilisation d'un spray de gouttelettes chargées qui frappent la surface d'intérêt, conduisant à la désorption des petites molécules organiques et des biomolécules de larges masses. Les analytes sont ainsi ionisés et amenés dans le spectromètre de masse sous forme d'ions désolvatés. Elle peut être considérée comme une version à pression atmosphérique de la technique SIMS (*Secondary Ion Mass spectrometry*) [332]. La technique SIMS analyse en effet les ions secondaires émis quand une surface est irradiée par un faisceau énergétique d'ions primaires également appelés ions incidents. Les sources d'ions avec un faisceau d'ions primaires de faible courant sont appelées sources statiques parce qu'elles n'endommagent pas la surface de l'échantillon, par opposition aux sources dynamiques qui produisent l'érosion de la surface analysée

et donc permet une analyse en profondeur de cette surface. Cette technique d'ionisation s'applique surtout à des échantillons solides. Elle possède une haute spécificité et sensibilité, ne nécessite pas de séparation préalable et seul un traitement minimal de l'échantillon est nécessaire [101;333]. Elle a été utilisée dans le cadre d'études métabolomiques portant sur des urines de souris ayant ou non développées une tumeur [327] et sur le métabolisme central d'*E. coli* à partir d'extraits cellulaires et de prélèvements du milieu de culture [334]. Elle présente en outre une tolérance plus importante aux sels par rapport aux techniques d'ionisation comme l'ESI ou l'APCI [327] et permet donc l'analyse de milieux complexes sans traitement de l'échantillon.

## 2. LES DIFFERENTS ANALYSEURS

### 2.1. Les analyseurs à basse résolution

- Les filtres quadripolaires

Ils utilisent la stabilité des trajectoires dans un champ électrique oscillant pour séparer les ions selon leur rapport  $m/z$ . Ils sont constitués de quatre barres [335] ayant idéalement une section hyperbolique ou cylindrique auxquelles on applique des potentiels déphasés de  $180^\circ$  dont l'un est continu ( $U$ ) et l'autre alternatif de radiofréquence  $V \cos \omega t$ . L'application de ces potentiels crée un champ quadripolaire entre les électrodes. Aussi, par un jeu de lentilles électrostatiques, les ions sont extraits de la source, accélérés et focalisés selon l'axe  $z$  jusqu'à leur entrée dans l'analyseur quadripolaire. A un instant donné, seuls les ions dont le rapport  $m/z$  permet une trajectoire stable dans le filtre atteignent le détecteur. Les autres, dont la trajectoire n'est pas stable, sont éjectés avant d'atteindre l'extrémité de l'analyseur et ne sont pas détectés. (Figure 71)

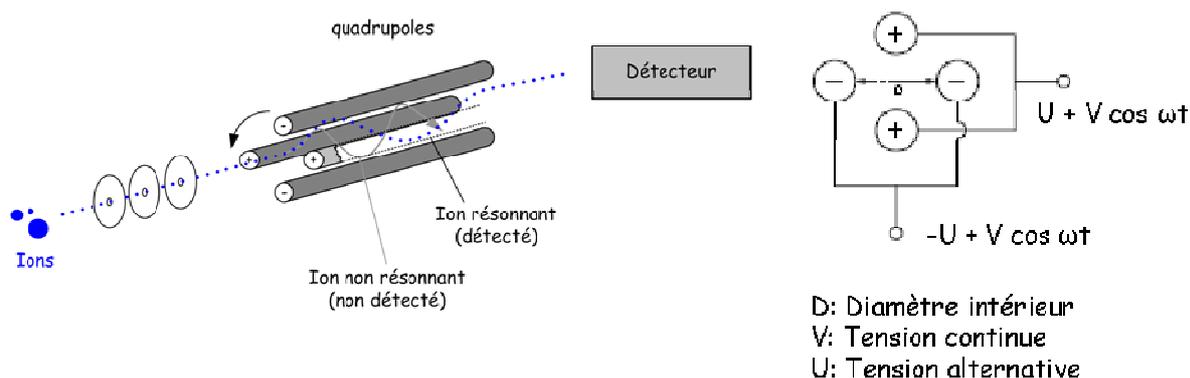


Figure 71 : Schéma d'un analyseur quadripolaire [336]

Certains appareils possèdent plusieurs analyseurs en série, séparés par une cellule de collision. La Figure 72 donne le schéma de principe d'un appareil à trois quadripôles : le triple-quadripôle (TQ). Dans le quadripôle central, on peut introduire un gaz de collision à une pression telle qu'un ion entrant dans le quadripôle subira en moyenne une collision avant de se dissocier. Dans un TQ, l'énergie de fragmentation délivrée est transmise à l'ensemble des ions (parents et fragments) contenus dans la

cellule de collision par application d'une différence de potentiel. Ainsi certains ions produits peuvent se fragmenter à nouveau pour produire des ions produits de 3<sup>ème</sup> génération qui peuvent eux-mêmes re-fragmenter,...

Les triples quadripôles permettent trois configurations de balayage : les balayages d'ions fragments, d'ions précurseurs et de perte de neutre constante. Ces dernières seront décrites dans les paragraphes suivants. La multiplicité des choix de configuration de balayage fait des triples quadripôles des outils particulièrement adaptés aux expériences quantitatives spécifiques.

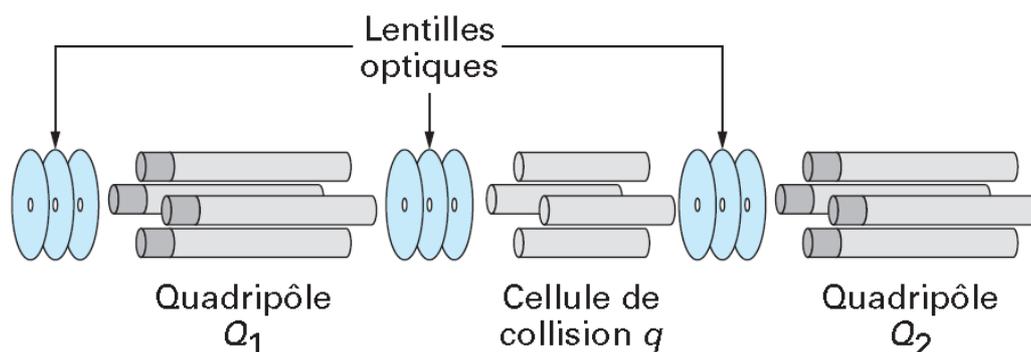


Figure 72 : Représentation schématique d'un triple quadripôle [337]

- **Les pièges à ions**
- Les pièges à ions tridimensionnels [338]

Ils sont constitués d'une électrode circulaire en forme de diabol, couverte de deux calottes sphériques (Figure 73). L'application de tensions alternatives sur l'électrode circulaire permet d'obtenir un « piège » à l'intérieur duquel les ions sont stockés. L'expulsion des ions et leur manipulation dans le piège sont effectuées par déstabilisation de leur trajectoire. Pour cela, l'amplitude de la tension alternative appliquée sur l'électrode circulaire est progressivement augmentée et une modulation axiale est appliquée sur les calottes, en parallèle et déphasée de 180°. On obtient une expulsion des ions selon les rapports  $m/z$  croissants de manière consécutive car on atteint, soit la frontière du diagramme de stabilité, soit une valeur d'instabilité due à des résonances non linéaires dans le diagramme de stabilité.

Les analyseurs à piégeage d'ions peuvent être utilisés pour produire des fragmentations séquentielles ( $MS^n$ ). En effet, en sélectionnant correctement les valeurs de tension appliquées de la modulation axiale mais également les fréquences appliquées sur les calottes et l'amplitude de la tension alternative, on peut ne conserver dans le piège que les ions ayant un rapport  $m/z$  donné. Au cours du temps, les ions sélectionnés se fragmentent dans le piège qui contient alors tous les ions produits. On fait ainsi des séquences  $MS/MS$  « temporelles » au lieu de les faire « dans l'espace » avec plusieurs analyseurs couplés. L'expulsion des ions produits à l'extérieur du piège se fait ensuite

comme exposé précédemment (par déstabilisation de leur trajectoire) et permet ainsi d'enregistrer un spectre de masse correspondant aux ions produits issus d'un ion précurseur de rapport  $m/z$  donné. La possibilité de faire des séquences  $MS^n$  en pièges à ions fait de ces analyseurs des outils particulièrement adaptés aux études structurales car ils permettent d'établir des filiations entre les différents ions produits générés lors des expériences de CID, facilitant ainsi leur identification. C'est un atout primordial dans l'étude de composés inconnus isolés lors d'études métabolomiques [121;184]. Dans le piège, la MS/MS permet une transmission exclusive de l'énergie de fragmentation aux ions précurseurs par application d'une fréquence de résonance qui leur est propre. Au fur et à mesure des collisions de basse énergie avec les molécules d'hélium présentes dans le piège, l'ion précurseur va progressivement gagner en énergie interne jusqu'à obtenir sa dissociation. Théoriquement les ions issus d'une telle fragmentation ne peuvent se fragmenter simultanément en ions produits de génération supérieure.

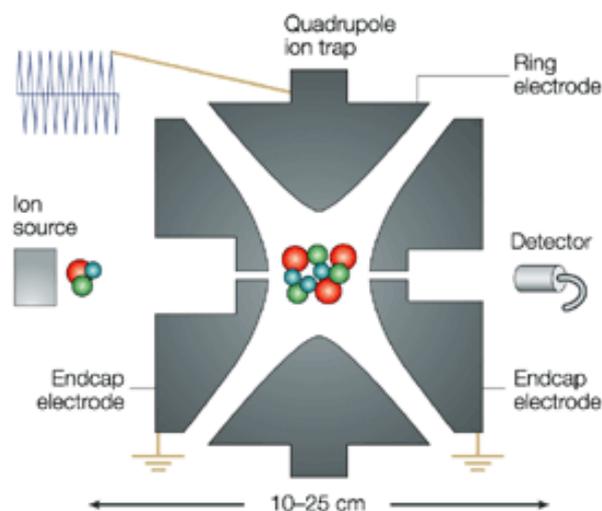


Figure 73 : Représentation schématique d'un piège à ions 3D [339]

Néanmoins, l'utilisation des pièges à ions présente une limitation importante due à l'existence du phénomène de LMCO (*Low Mass Cut-Off*) qui limite la détection des ions produits de bas rapports  $m/z$  en modes  $MS/MS$  et  $MS^n$ . Le seuil de détection des basses masses se situe à environ  $1/4$  de la masse de l'ion parent. Lors de l'étude de peptides, les informations relatives aux fragments  $y_1$  et aux ions immoniums peuvent être perdues. Cette limite est liée au fait que l'activation d'un ion, nécessaire à sa fragmentation nécessite sa mise en oscillation dans le puits de pseudo potentiel dans lequel il se trouve. Or pour avoir accès aux ions de faibles masses, la quantité d'énergie nécessaire est trop importante et entraîne l'éjection de l'ion précurseur du puits de pseudo potentiel. De plus, les pièges 3D possèdent un faible volume de piégeage et ont donc une capacité de stockage limitée. Une quantité trop importante d'ions présents simultanément dans le piège entraîne l'apparition du phénomène de charge d'espace qui se traduit par un décalage et un élargissement des fréquences

d'oscillations [340], aboutissant à une moindre efficacité de la sélection et de l'excitation d'ions de  $m/z$  donné et donc à une dégradation de la précision en masse, de la résolution, et de la gamme dynamique en masse (capacité à discriminer des ions d'abondances différentes dans un même spectre), les mouvements des ions de faible abondance étant particulièrement affectés.

- Les pièges linéaires

Les pièges linéaires (pièges 2D) [341] sont des analyseurs constitués de quatre barres comme pour un quadripôle qui serait fermé par des électrodes d'entrée et de sortie permettant de repousser les ions vers l'intérieur du quadripôle, en appliquant des tensions continues positives et négatives, pour piéger respectivement les ions positifs et négatifs. Les ions sont « refroidis » en énergie cinétique par collisions avec un gaz inerte à faible pression. Ils sont piégés dans un mouvement d'oscillation entre les électrodes terminales suivant l'axe  $z$ , et simultanément oscillent dans le plan  $xy$  par application de tensions RF de signe opposé et d'un déphasage de  $180^\circ$  sur les barres. Suivant les constructeurs, deux modes d'expulsion sont utilisés : éjection axiale suivant  $z$  (Q-trap<sup>®</sup>, Applied Biosystem) ou l'éjection radiale (LTQ, thermo-Fisher). Un grand avantage des pièges linéaires est d'accepter environ 10 fois plus d'ions que les pièges 3D avant d'avoir un effet de charge d'espace notable [342;343]. Ceci est dû à la fois au plus grand volume de piégeage disponible et au fait que la focalisation se fait autour de l'axe  $z$  plutôt qu'autour d'un point. Suivant la géométrie du piège, il n'y a plus de LMCO. En effet, les pièges à éjection axiale n'ont plus de LMCO alors que pour les pièges à éjection radiale, cette limite existe encore. Les pièges linéaires possèdent une vitesse de balayage supérieure à celle des pièges 3D améliorant ainsi la sensibilité et permettant la détection des composés minoritaires lors des analyses métabolomiques par exemple. Il faut noter que Thermo-Fisher a développé un nouveau mode de fragmentation pour son LTQ : le *Pulsed-Q-Dissociation* (PQD<sup>®</sup>) qui permet de supprimer le phénomène de LMCO son système à éjection radiale[344;345]. Ce mode implique une activation de l'ion précurseur avec un  $Q$  élevé, suivi d'un temps de latence permettant la fragmentation du précurseur puis d'un rapide changement vers un  $Q$  de faible intensité permettant de piéger l'ensemble des ions produits qui peuvent ainsi être détectés jusqu'à  $m/z$  50. L'utilisation d'un système à éjection axiale permet de conserver les différents modes d'acquisition propre au TQ mais limite, en contre partie, la fragmentation séquentielle à l'ordre 3 [343].

## 2.2. Les analyseurs à haute résolution

Ces analyseurs ont contribué à l'essor de la métabolomique par spectrométrie de masse en permettant la discrimination des espèces isobares nombreuses dans les fluides biologiques complexes comme le plasma ou les urines et en donnant accès à la formule brute des ions détectés. Les principaux types d'appareils actuellement commercialisés permettant d'accéder à la haute résolution et à la mesure de masses précises sont les appareils de types TOF, TOF-TOF ou Qq-TOF, les FT-ICR et plus récemment l'Orbitrap<sup>®</sup>.

- **Analyseurs à temps de vol**

Le principe de l'analyseur à temps de vol (Figure 74), désigné TOF pour Time of flight, a été décrit en 1946 par Stephens [346]. Wiley et McLaren ont élaboré le premier TOF dans un but commercial en 1955 [347]. Son principe de fonctionnement est le suivant : les ions sont accélérés vers le tube de vol par une différence de potentiel appliquée entre une électrode et la grille d'extraction. Les particules accélérées ayant toutes la même énergie cinétique passent ensuite dans le tube de vol, libre de champ, où elles sont séparées en fonction de leur vitesse acquise. Les vitesses sont inversement proportionnelles à la racine carrée de leurs rapports  $m/z$ . Ainsi, pour des ions de même énergie cinétique, les ions plus légers possèdent une grande vitesse et arrivent au détecteur plus rapidement que les ions plus lourds possédant une plus faible vitesse.

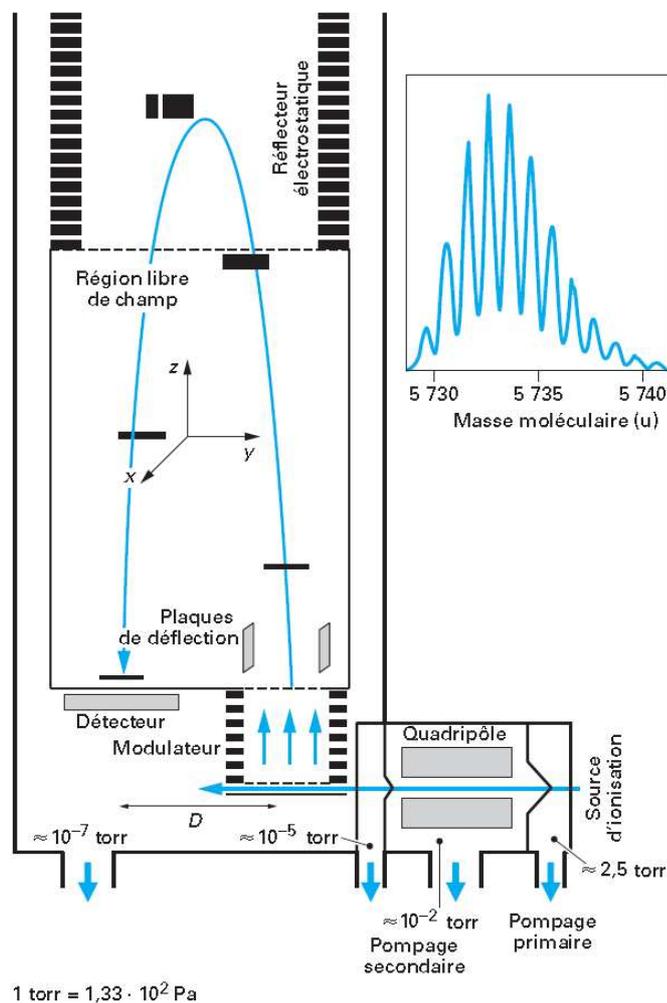


Figure 74 : Principe de fonctionnement d'un analyseur à temps de vol [337]

Afin d'améliorer la résolution en masse des analyseurs à temps de vol, des modifications techniques ont été mises au point : l'extraction retardée qui consiste à introduire un retard entre la formation des ions et l'extraction de ceux-ci de la source, et/ou l'utilisation d'un miroir ou réflecteur

électrostatique [348] qui permet la focalisation temporelle au niveau du détecteur des ions quittant la source avec le même rapport  $m/z$  mais présentant une dispersion en énergie cinétique.

Les analyseurs hybrides Qq-TOF (Quadripôle –Temps de vol) sont apparus en 1995. Ces analyseurs peuvent être considérés comme un triple quadripôle dans lequel le troisième quadripôle est remplacé par un analyseur à temps de vol (TOF). Ils ont vu le nombre de leurs applications se multiplier dans le cadre des études du métabolome ces dernières années [349;350]. En effet, ces analyseurs présentent de nombreux avantages tels que leur sensibilité qui est équivalente à celle des triples quadripôles mais également un important pouvoir de résolution en masse ( $m/\Delta m > 10\,000$  pour une masse de 1000 Da) et une précision en masse inférieure à 10 voire 5 ppm [351-353] lorsqu'un système de calibration interne comme le lockspray<sup>®</sup> est utilisé, facilitant ainsi les identifications préliminaires de composés inconnus. Ils peuvent être couplés à la GC ou à la LC. Leur principal atout est leur vitesse d'acquisition, par rapport aux analyseurs de type piège à ions, qui est de l'ordre de 300 ms et qui en fait des analyseurs particulièrement compatibles avec la dernière innovation en termes de technique séparative en phase liquide : la chromatographie liquide à très haute pression (UHPLC) dont les largeurs de pics à la base sont couramment inférieures à 10 s. Par ailleurs, ces instruments permettent l'utilisation des modes de balayage de parents ou de fils, présentant un intérêt particulier lors des analyses de familles chimiques particulières. Leur limitation majeure en métabolomique est leur faible gamme dynamique qui est classiquement de 2 à 3 ordres de magnitude, bien qu'un fabricant (Waters) ait développé un nouveau système baptisé DRE<sup>®</sup> pour *Dynamic Range Enhancement* qui permet d'atteindre 4 ordres de magnitude [354;355]. Leur possibilité de couplage avec l'UHPLC les place parmi les analyseurs les plus utilisés en profilage métabolomique.

- **Les analyseurs à ultra haute résolution.**

Une précision en masse et une résolution encore meilleures peuvent être obtenues grâce à deux types d'appareils de plus en plus utilisés en métabolomique. Pour ces deux types d'instrument, le mélange de métabolites ionisés est piégé sur une trajectoire orbitale. La fréquence d'oscillation des ions sur leur orbite dépend de leur rapport  $m/z$  et peut être mesurée, ce qui permet d'obtenir une précision en masse exceptionnelle.

- **Le FT-ICR**

La spectrométrie de masse par FT-ICR [356] repose sur le principe du cyclotron : le piégeage des ions produits par une source externe (MALDI ; électrospray...) s'effectue sur une trajectoire circulaire conséquence de l'application d'un champ magnétique intense, orienté perpendiculairement à leur direction de déplacement, et qui crée un puits de pseudo-potentiel. La fréquence résultante du mouvement de l'ion dépend du rapport  $m/z$  des ions et de l'intensité du champ magnétique. Un champ électrique annexe maintient les ions dans une direction parallèle au champ magnétique. Cette configuration leur impose un mouvement cyclotronique selon une fréquence inversement proportionnelle à la valeur de leur rapport  $m/z$ . Initialement, les ions sont proches du centre de la cellule ICR, du fait de leur faible énergie cinétique, l'application de radiofréquences d'excitation permet

d'élargir leur orbite et de les regrouper de façon homogène. L'élargissement de l'orbite dû à l'excitation les rapproche des plaques réceptrices qui enregistrent les courants induits par les paquets d'ions. Le signal détecté appelé « transient » correspond à la superposition des différents courants induits. Les différentes fréquences sont extraites du signal grâce à l'utilisation d'une transformée de Fourier et le spectre des fréquences résultantes est converti en un spectre de rapports  $m/z$  grâce à l'utilisation de la relation qui lie rapport  $m/z$  et fréquence (Figure 75).

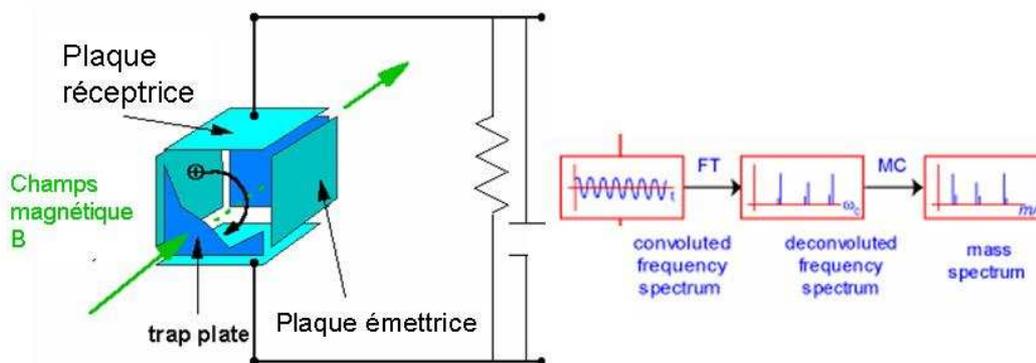


Figure 75 : Principe de fonctionnement du FT-ICR

Les spectromètres de masse par FT-ICR présentent un certain nombre d'avantages [63;357]

- L'obtention de mesures de masse avec une précision suffisante ( $< 1$  ppm) pour obtenir la composition élémentaire d'une molécule de faible masse moléculaire. En effet, pour une masse de 400 Da, une précision en masse de 1 ppm permet d'aboutir à une composition élémentaire unique alors qu'une précision en masse de 10 ppm (TOF) en génère plus d'une dizaine.
- Une résolution qui peut atteindre des valeurs supérieures à 1 000 000 ( $m/\Delta m$ , FWHM), ce qui permet l'analyse de mélange très complexe. A titre d'exemple, Marshall et al. ont pu montrer la résolution de plus de 10 000 espèces moléculaires distinctes dans un seul spectre de masse réalisé à partir d'huile de pétrole [358].
- La possibilité de mener des expériences de fragmentation  $MS^n$ , ce qui permet de compléter les mesures de masse précises et hautement résolues par des informations structurales. La précision en masse est conservée pour les ions fragments, il est donc possible d'accéder à la composition élémentaire des ions produits[359]. Lorsque la sélection du précurseur est effectuée dans la cellule ICR, l'ion précurseur peut être isolé à haute résolution ( $R= 10000$  à  $20000$ ,  $m/\Delta m$ , FWHM à  $m/z = 600$ ).
- La possibilité d'utiliser des modes de fragmentation différents de la CID et pouvant permettre d'obtenir des informations structurales complémentaires.

Cependant, lorsqu'aucune méthode de chromatographie n'est couplée avec ce spectromètre de masse, malgré le gain de temps, certaines limitations apparaissent. En effet, les isomères chimiques comme les hexoses ne sont pas différenciés car ils possèdent une masse moléculaire identique. De

plus la présence de fragments et d'ions adduits, difficilement différenciables d'espèces moléculaires, est un facteur limitant, ainsi que la difficulté de quantifier liée à la favorisation de la suppression d'ionisation en absence de technique séparative couplée. A l'heure actuelle, peu de publications décrivant l'utilisation de la FT-ICR pour des applications métabolomiques sont parues [81]. Cependant, l'utilisation de la FT-ICR dans le cadre d'études du métabolome a fait l'objet d'une revue bibliographique [63] qui évalue les potentialités de ces spectromètres de masse pour les approches métabolomiques et quelques études ont été publiées depuis [360-365]. Les analyseurs FT-ICR sont principalement utilisés en introduction directe, soulevant la principale limitation de leur utilisation en métabolomique, qui est leur temps de cycle d'acquisition. En effet, si les performances (précision en masse et résolution) sont exceptionnelles, elles sont atteintes au prix d'un temps d'acquisition de transient long (>1s) qui limite le couplage de cet analyseur avec l'UHPLC dont la résolution, la robustesse, et la rapidité d'analyse ont conduit à sa rapide expansion dans le domaine de l'acquisition des empreintes métaboliques. Enfin, les performances des FT-ICR sont directement reliées à l'intensité du champ magnétique utilisé (actuellement jusqu'à 15T), nécessitant l'emploi d'aimants supraconducteurs dont le prix de maintenance et d'entretien restreint le nombre d'utilisateurs. Malgré toutes ces limitations, de plus en plus d'études métabolomiques sont réalisées sur ce type d'appareil, couplé ou non à un système chromatographique [366-368].

- **L'Orbitrap<sup>®</sup>**

D'introduction récente, l'Orbitrap<sup>®</sup> est une alternative intéressante au FT-ICR pour la métabolomique puisqu'elle donne accès à l'ultra-haute résolution sans avoir recours à un champ magnétique intense. Son principe a été décrit par Alexander Makarov [369]. Brièvement, l'Orbitrap<sup>®</sup> est constituée de deux électrodes, la première est centrale et possède la forme d'un fuseau, la deuxième est extérieure et possède la forme d'un tonneau coupé en deux parties égales séparées par un espace étroit (Figure 76). Un champ électrique quadri-logarithmique est maintenu entre les électrodes. Les ions sont injectés tangentiellement par l'interstice séparant les deux moitiés de l'électrode externe. Les ions sont injectés à une énergie cinétique de quelques keV et se mettent à osciller en orbite de l'électrode centrale dans les directions axiales et radiales sous l'action concomitante des tensions qui génèrent le champ électrique, des forces centrifuges et de la géométrie du piège. Le vide y est poussé ( $10^{-10}$  torr). A chaque passage devant les électrodes externes, les ions génèrent des courants induits sur lesquels est appliquée une transformée de Fourier afin de déconvoluer les différentes fréquences d'oscillation axiale (sur l'axe z) des ions qui sont ensuite convertie en rapports  $m/z$  précis.

L'Orbitrap<sup>®</sup> a été couplée par la société Thermo Electron avec un piège ionique linéaire (LTQ) pour former un spectromètre de masse hybride, le LTQ Orbitrap<sup>®</sup>. Le LTQ est capable d'acquérir ces propres spectres et permet de faire de la  $MS^n$ . Les ions produits peuvent à leur tour être envoyés dans l'Orbitrap et analysés avec la même résolution que l'ion précurseur. Une différence importante vis-à-vis des FT-ICR est la sélection et la fragmentation des précurseurs au sein de la trappe linéaire qui

implique donc leur sélection à basse résolution (fenêtre minimale d'isolation de 0,3 Th). Cet analyseur permet d'atteindre une précision en masse inférieure à 2 ppm et une résolution de 100000 [370-372].

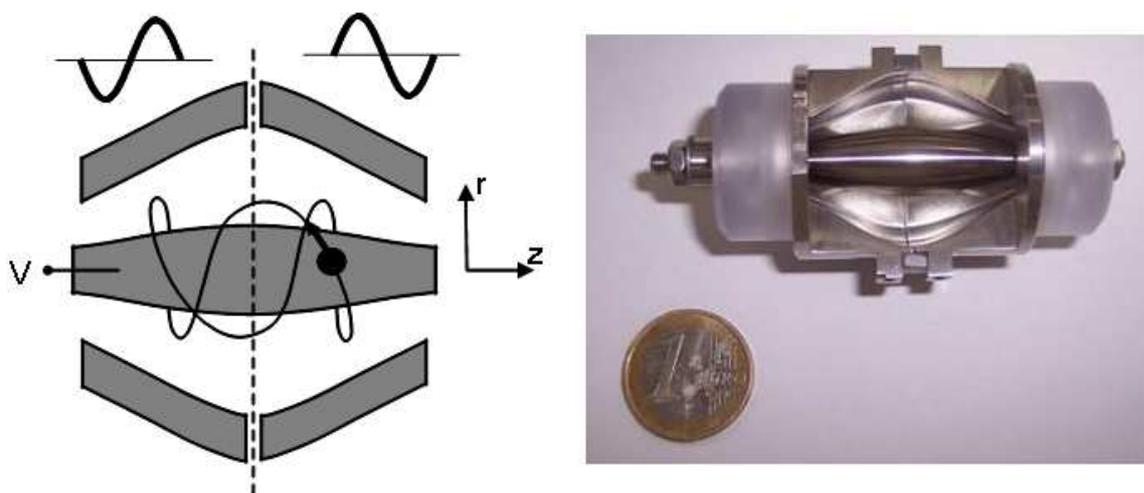


Figure 76 : Piège électrostatique ou «Orbitrap» d'après la description de A. Makarov [369]

(images issues des sites internet <http://aston.chem.purdue.edu/research/instrumentation/orbitrap/ion-motion-control> et <http://www-lpg.obs.ujf-grenoble.fr/Workshop-techniques-de-spectrometrie-de-masse>)

La simplicité d'utilisation, les bonnes performances et les possibilités d'étude structurale ont permis à l'Orbitrap® de s'imposer très rapidement sur le marché et de rencontrer un réel engouement dans le domaine de la métabolomique. Pour ce dernier, il a déjà été utilisé pour des approches de screening de métabolites sur des incubats de microsomes hépatiques humains [373], pour l'identification de métabolites de phénobarbital dans des urines de rat, mis en évidence par approche métabolomique globale [62], pour l'acquisition d'empreintes métaboliques de bactéries [374] et de sera de mammifère en couplage avec l'UHPLC [375], et a permis l'identification de biomarqueurs de prééclampsie [376].

### 2.3. Les différents modes d'acquisitions

On distingue deux principaux modes d'acquisition : le mode de spectrométrie de masse simple (MS) et le mode de spectrométrie de masse en tandem (MS/MS) ou à fragmentation consécutives (MS<sup>n</sup>).

- Pour le mode MS, les spectres peuvent être acquis en mode de balayage analytique (scan) ou en mode de détection d'ions sélectionnés (*selected ion monitoring*, en abrégé SIM). Ainsi, en balayage analytique, des spectres de masse sont acquis entre deux rapports  $m/z$  extrêmes et sont accumulés par répétition. Tous les ions formés dans la source sont détectés. Le mode SIM a quant à lui pour but de détecter des substances ciblées avec un maximum de sensibilité.

- Les modes d'acquisition MS/MS et MS<sup>n</sup> impliquent une activation des ions de laquelle va résulter leur fragmentation. On obtient alors le spectre des ions produits résultant de la fragmentation. Il existe différents modes d'activation: la collision par gaz (CID : *Collision Induced Dissociation*), la technique SORI (*Sustained Off-Resonance Irradiation*), les techniques ECD (*Electron Capture Dissociation*) [377] EDD (*Electron Detachment Dissociation*) et ETD (*Electron transfert Dissociation*) et la technique IRMPD (*infra red multi photon dissociation*) [378]. Ces différentes techniques diffèrent dans les fragments qu'elles génèrent [379]. Mis à part l'activation CID, les autres techniques ne sont disponibles que sur les spectromètres de masse FT-ICR et l'Orbitrap<sup>®</sup> s'est récemment doté de l'ETD. On obtient alors le spectre des ions produits résultant de la fragmentation.

La fragmentation CID reste la méthode d'activation la plus répandue avec les instruments actuels. Elle repose sur la survenue de collisions entre l'ion précurseur et des molécules d'un gaz neutre cible, qui conduisent à la conversion d'une partie de l'énergie translationnelle de l'ion en énergie interne [380]. Lorsque l'énergie interne de l'ion dépasse le seuil de dissociation, l'ion précurseur peut se fragmenter. Selon les types d'instrument, les énergies de collisions peuvent être de l'ordre du keV (collisions à haute énergie) ou comprises entre 1 et 200 eV (collisions à basse énergie). Les collisions à haute énergie ont l'avantage d'être très robustes vis-à-vis des variations des conditions expérimentales, conduisant à des spectres de fragmentation très reproductibles mais ne sont principalement possibles qu'avec des appareils de type TOF-TOF ou à secteurs magnétiques qui ne sont équipés que de sources MALDI ou de désorption par laser et qui ne sont donc pas couplables avec les techniques séparatives en phase liquide. Les instruments capables de fragmentation et employés en métabolomique avec un couplage LC sont donc limités aux collisions à basses énergies, peu reproductibles d'une expérience à l'autre et d'un instrument à l'autre, empêchant la constitution de bibliothèques spectrales universelles.

En ce qui concerne les analyseurs « dans l'espace », comme les TQ et les Qq-TOF, différents modes de balayage en MS/MS peuvent être utilisés : les modes de balayage analytique (scan) ou de SIM utilisés en MS mais également le mode SRM (*selected reaction Monitoring*) impliquant la détection de réactions de décompositions d'ions sélectionnées, spécifiques des substances à analyser. Le spectromètre est réglé de manière à ne transmettre que les ions produits issus d'une réaction de décomposition choisie dans la zone de réaction. Les potentiels U et V du premier quadripôle sont fixés sur un ion parent donné et ceux du troisième quadripôle sur un ion issu de la fragmentation de l'ion parent. Ce mode permet d'obtenir un gain de sensibilité et de sélectivité par rapport au mode SIM car on ne suit qu'une transition.

Les équipements permettant de faire de la spectrométrie de masse en mode MS/MS peuvent être exploités de plusieurs manières dont les trois principales sont :

- Le balayage d'ions fragments : un ion précurseur choisi est sélectionné dans le premier spectromètre pour subir des collisions dans la cellule de collision et se fragmente. Les produits de réactions sont analysés par le second spectromètre de masse.

- Le balayage des ions précurseurs : un ion est sélectionné dans le deuxième analyseur pendant que le premier balaye en rapports  $m/z$  de façon à détecter tous les ions parents capables, en se fragmentant, de produire l'ion sélectionné dans le troisième analyseur.

- Le balayage de perte neutre constante: il consiste à balayer en parallèle le premier et le dernier analyseur en imposant entre les deux un décalage en rapport  $m/z$  constant. On détecte ainsi tous les ions qui en se fragmentant perdent un neutre de masse constante (uniquement sur les TQ).

Ces différentes configurations sont particulièrement adaptées à la mise en œuvre de dosages spécifiques, ainsi que dans les approches ciblées. Le balayage de perte de neutre constante a par exemple été utilisé dans des études de toxicité induite par des métaux lourds sur les sulfoconjugués urinaires chez le rat [381], de néphrotoxicité induite par la gentamicine [382], ainsi que d'acides mercapturiques dans des urines humaines après une utilisation d'acétaminophène [383]. Le balayage des ions précurseurs a également été utilisé pour l'étude des sulfoconjugués [381].



## ANNEXE 2 – XCMS

### 1. FONCTIONNEMENT D'XCMS [150]

La stratégie de traitement des données mise en œuvre par XCMS combine plusieurs étapes (Figure 77). Cette stratégie est modulaire et permet la substitution ou l'ajout d'étapes de traitement supplémentaires (comme par exemple avec le package CAMERA).

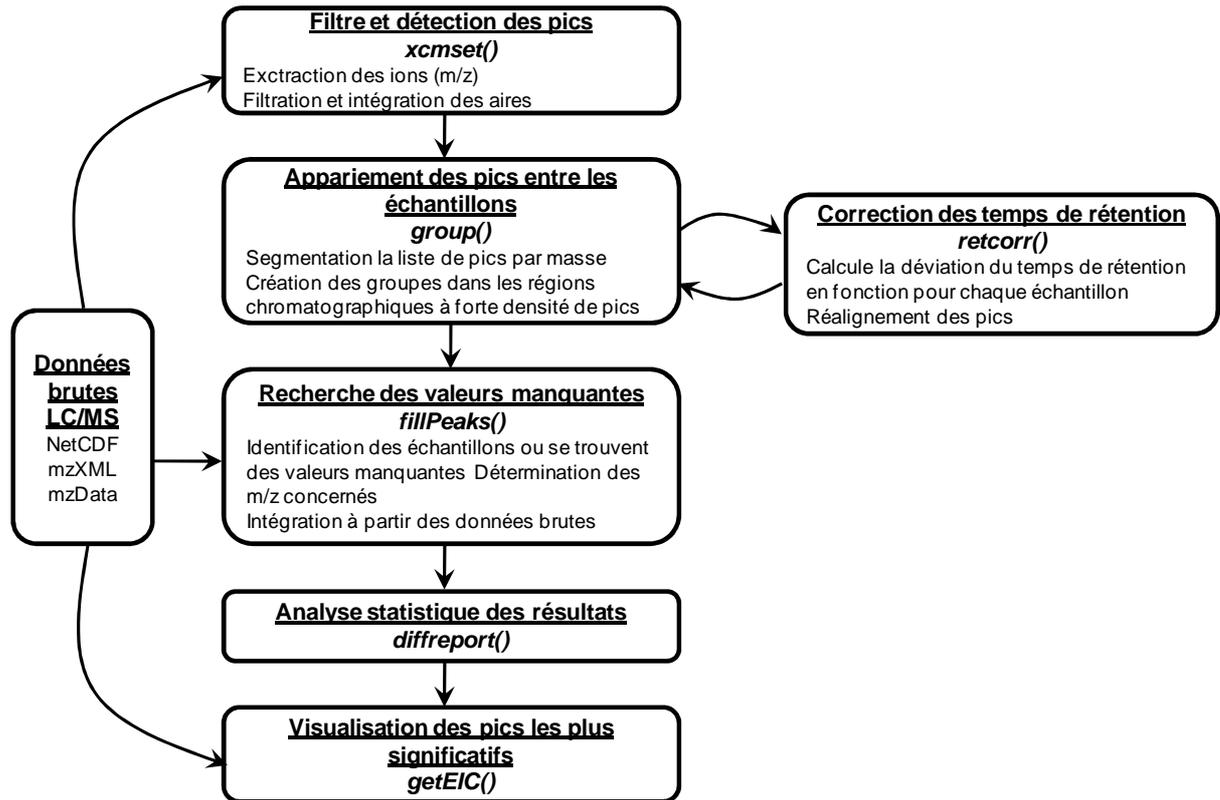


Figure 77 : Stratégie de traitement des données mise en œuvre par XCMS.

#### 1.1. Filtration et détection des pics

Il existe deux algorithmes de filtration et de détection des pics dans XCMS : MatchedFilter et CentWave. Nous allons détailler ici celui utilisé lors de ce travail de thèse à savoir MatchedFilter (qui est aussi celui par défaut).

L'algorithme de détection des pics MatchedFilter est basé sur la réduction du spectre de chaque échantillon en tranches de  $m/z$  (paramètre « *step* » par défaut 0,1 de largeur) puis l'exploitation de ces tranches individuelles dans le domaine chromatographique. Dans chaque tranche, le signal est déterminé en prenant le maximum d'intensité en fonction du temps. La représentation d'un ion extrait à partir du chromatogramme est appelée EIBPC pour « *Extracted Ion Base-Peak Chromatogram* ». Les EIBPC sont ensuite combinés (par défaut 2 à 2, correspond au paramètre « *steps* ») pour obtenir

un pic chromatographique (Figure 78A). Chaque tranche combinée est ensuite filtrée en utilisant la dérivée seconde de la fonction Gaussienne (La Gaussienne est le modèle pour la forme du pic chromatographique, avec comme paramètre de base la largeur à mi-hauteur ou « *fwhm* ») (Figure 78B). Après filtration, les pics sont sélectionnés en utilisant comme paramètre le rapport signal/bruit (« *snthresh* ») puis sont intégrés. (Figure 78C)

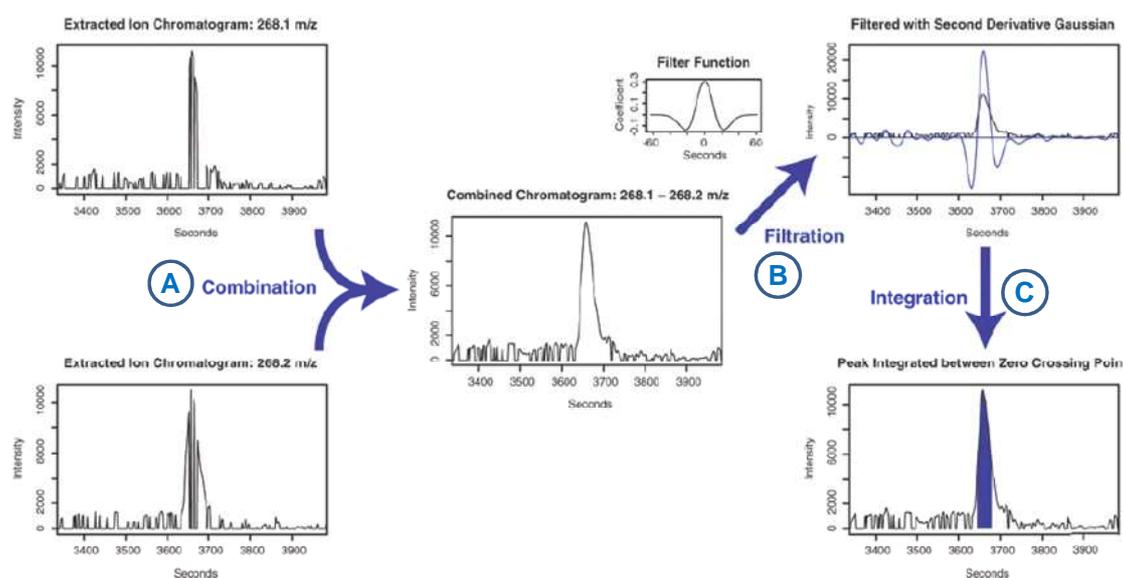


Figure 78 : Illustration de la méthode de détection et d'intégration des pics par l'algorithme MatchedFilter de XCMS [150].

En raison de la partie négative du pic générée par la filtration par la dérivée seconde, la suppression du bruit de fond, qui est réalisée sur données filtrées, peut être problématique. En effet le rapport réel du signal/bruit dans les données non filtrées peut être beaucoup plus faible que dans les données filtrées, il convient donc de surestimer le « *snthresh* », et également de limiter le nombre de pics détectés par EIBPC (« *max* ») afin d'éviter l'intégration du bruit de fond sous forme de pics chromatographique.

Un détail important est la relation entre la largeur des tranches en  $m/z$  (« *step* ») et la précision de l'appareillage. En effet sur des spectromètres basse résolution la largeur de 0.1th est beaucoup plus petite que la précision en masse et ainsi un même métabolite détecté (=pic chromatographique) débordera sur plus de 2 EIBPC, dans ce cas il faudra soit augmenter le « *step* » ou augmenter le nombre d'EIBPC à combiner pour créer un pic chromatographique (« *steps* »). Inversement avec les spectromètre de masse haute résolution ou largeur des tranches en  $m/z$  (« *step* ») est plus élevée que la précision en masse (ou dans le cas des données en centroïde), il faut soit diminuer le « *step* » soit passer à un « *steps* » =1.

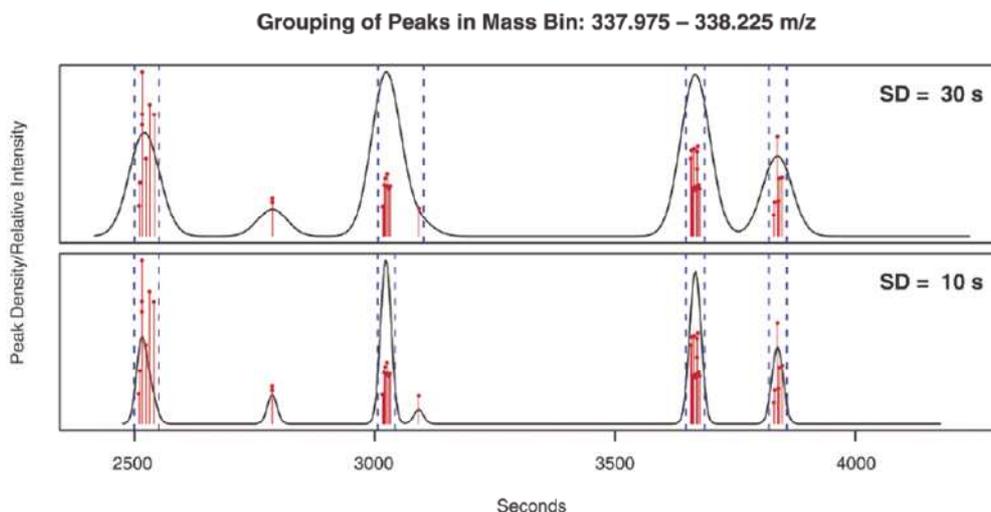
Le découpage en tranche de  $m/z$  du spectre et la combinaison d'EIBPC réduit nécessairement la précision des résultats. Pour éviter cette perte en précision, la masse de chaque pic est calculée à partir des données brutes : la masse en haute résolution est déterminée dans chaque spectre

contenant le pic, puis la masse globale est calculée comme une moyenne pondérée de toutes les masses en haute résolution. L'algorithme de détection de pic MatchedFilter gère ainsi les données en basse et haute résolution, et les données en profile ou centroïde.

## 1.2. Appariement des pics entre les échantillons

Après avoir détecté les pics dans des échantillons individuels, ces pics doivent ensuite être comparés entre les différents échantillons pour permettre le calcul des écarts temps de rétention et de comparaison de leurs intensités relatives. Il existe là encore plusieurs algorithmes d'appariement dans XCMS (group.density, group.mzclust, group.nearest). Nous allons détailler ici celui utilisé lors de ce travail de thèse à savoir group.density (qui est aussi celui par défaut).

L'algorithme group.density va appairer les pics en tenant compte des deux dimensions de la LC/MS : le temps de rétention et la masse. Parce que la précision en masse est souvent mieux connue et relativement meilleure que celle des temps de rétention, nous utilisons des intervalles fixes de  $m/z$  (« *mzwid* ») pour regrouper les pics dans le domaine de masse (Figure 79). Pour éviter de scinder en deux un pic chromatographique en raison du découpage arbitraire des intervalles de  $m/z$ , il y a chevauchement de moitié des intervalles (ex : 100,0 à 100,25, 100,125 à 100,375, etc.). Au cours de l'appariement, chaque pic est compté deux fois dans chaque intervalle qui se chevauche, mais une étape est intégrée pour supprimer ces doublons.



**Figure 79 : Exemple de regroupement des pic dans l'intervalle de  $m/z$  337,975 à 338,225 (largeur 0,25 $m/z$ ) [150].**

Chaque chromatogramme représente le profil chromatographique des pics de cet intervalle de masse, la différence entre les deux chromatogrammes et l'écart-type de la gaussienne obtenue lors du lissage. Plus cet écart-type est grand, plus les pics sont larges ce qui a un impact sur le regroupement dans le domaine chromatographique observé en pointillés bleus.

Après avoir initialement regroupé les pics dans le domaine spectral, il faut différencier les groupes avec des temps de rétention différents dans chaque intervalle en  $m/z$ . ne connaissant pas à priori les écarts sur les temps de rétention à ce stade, on ne peut pas appliquer des intervalles fixes

comme pour le domaine spectral. L'algorithme calcule donc la répartition globale des pics chromatographiques et identifie de manière dynamique les régions où de nombreux pics ont des temps de rétention similaires. En fonction de ce regroupement dans le domaine chromatographique, tous les pics de même temps de rétention sont placés dans un même groupe. Ce regroupement est impacté par la largeur des pics obtenus par lissage (paramètre « *bw* » à choisir en fonction du type de chromatographie) (Figure 79)

L'algorithme ne garde que les groupes dont les pics sont présents dans plus de la moitié des échantillons (« *minfrac* » = 0.5 par défaut). Si les échantillons proviennent de conditions différentes (plus de 2 groupes), les données brutes peuvent être divisées en plusieurs sous-groupes de telle manière à ce que cette fraction minimum (qui est modifiable) concerne chaque sous-groupe et non pas la totalité des échantillons. Dans ce cas, les pics sont éliminés si aucun des sous-groupes d'échantillons ne satisfait ce critère. On peut également définir le nombre d'échantillon minimum dans lequel doit être présent le pic pour le conserver (« *minsamp* »).

### 1.3. Alignement des temps de rétention

L'alignement temps de rétention permet de corriger simultanément les temps de rétention de tous les échantillons en une seule étape. Il existe ici encore plusieurs méthodes disponibles dans XCMS (*retcor.loess*, *retcor.linear*, *retcor.orbiwarp*) Nous allons détailler ici celui utilisé lors de ce travail de thèse à savoir *retcor.loess* (qui est aussi celui par défaut). L'algorithme utilisé à cette étape identifie d'abord des groupes dans lesquels très peu d'échantillons n'ont pas de pics attribués (« *missing* ») et/ou très peu d'échantillons ayant plus d'un pic attribué (« *extra* »). Ces pics ayant une forte probabilité d'être bien groupés, ils vont être temporairement utilisés comme pics de références. Pour chacun de ces groupes, l'algorithme calcule le temps de rétention moyen et l'écart de la moyenne dans les différents échantillons. Ainsi pour chaque échantillon on peut construire, à partir des valeurs de déviation par rapport à la moyenne des groupes de référence, une courbe de déviation (Figure 80).

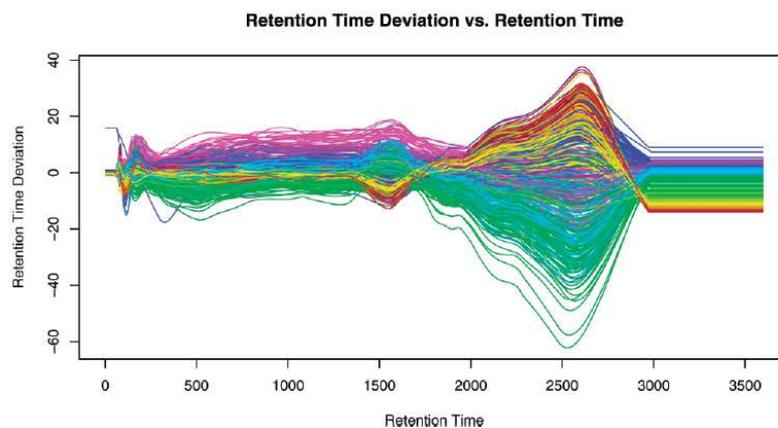


Figure 80 : Profil de déviation du temps de rétention [150].

Chaque ligne correspond à un échantillon, les couleurs correspondent à l'ordre d'analyse : rouges les premiers échantillons, violet les derniers. Une déviation positive indique que les composés sont élués après le temps de rétention moyen, une valeur négative que les composés sont élués avant le temps de rétention moyen.

Une méthode mathématique est alors utilisée (« *method* ») afin de modéliser les parties de la courbe (temps de rétention) ou il n'y a pas de groupe de référence et de corriger les temps de rétention pour chaque échantillon en fonction de cette courbe. Cette procédure d'alignement peut être répétée de façon itérative afin d'obtenir un alignement de plus en plus précis. Suite à cette étape d'alignement des temps de rétention, une nouvelle étape d'appariement des pics doit être réalisée, car si la dérive dans le domaine chromatographie est importante, certains pics peuvent alors être répartis dans deux groupes différents (ce qui équivaut à une détection de deux pics différents pour un seul et même pic réel) suite à la première étape d'appariement (GroupA ou GA). Cette erreur qui peut être corrigée en réalisant de nouveau un appariement (GroupB ou GB) (Figure 81)

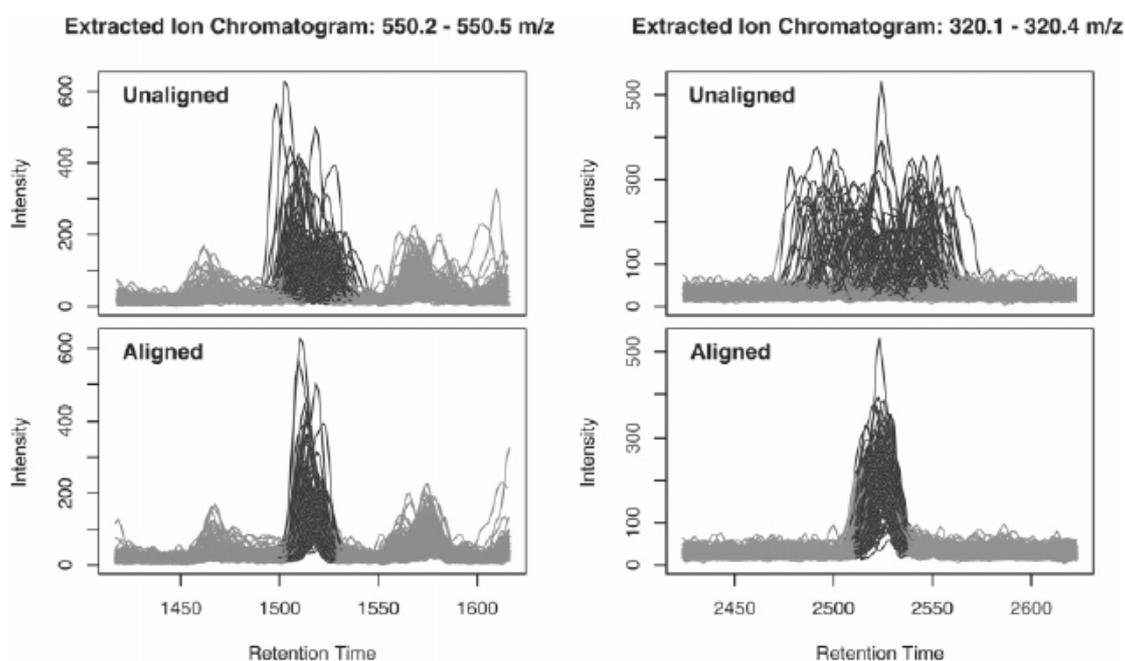


Figure 81 : EIBPC de deux groupes avant et après alignement des temps de rétention [150].

#### 1.4. Intégration des pics et recherche des valeurs manquantes

Même après un deuxième appariement des groupes, il y a toujours des pics qui ne sont pas retrouvés dans tous les échantillons. Cette méthode détermine les intensités de ces pics dans les échantillons en intégrant les  $m/z$  du pic concerné directement dans les données brutes. Selon le type de données brutes, il ya 2 différentes méthodes disponibles dans XCMS : pour les données LC/MS la méthode "*chrom*" intègre les données brutes dans le domaine chromatographique, alors que «*MSW*» est utilisé pour les listes de pics sans informations sur les temps de rétention comme pour les spectres de FIA. Il n'y a aucun paramètre modifiable pour ces méthodes.

## 1.5. Analyses statistiques des résultats et visualisation des EIBPC

Cette dernière étape gère la création d'un rapport statistique entre deux séries d'échantillons (« *class1* » et « *class2* »). Il effectue un T-test avec correction de Welch pour chaque pic et classe ces derniers en fonction de leur p-value (« *sortpval* »). Les EIBPC les plus significatifs (« *eicmax* ») sont extraits, les chromatogrammes issus des données brutes sont générés pour visualiser les différences entre les deux groupes, et un box plot représentant les intensités de ce EIBPC dans chaque groupe est également créé (moyenne et écart-type) (Figure 82).

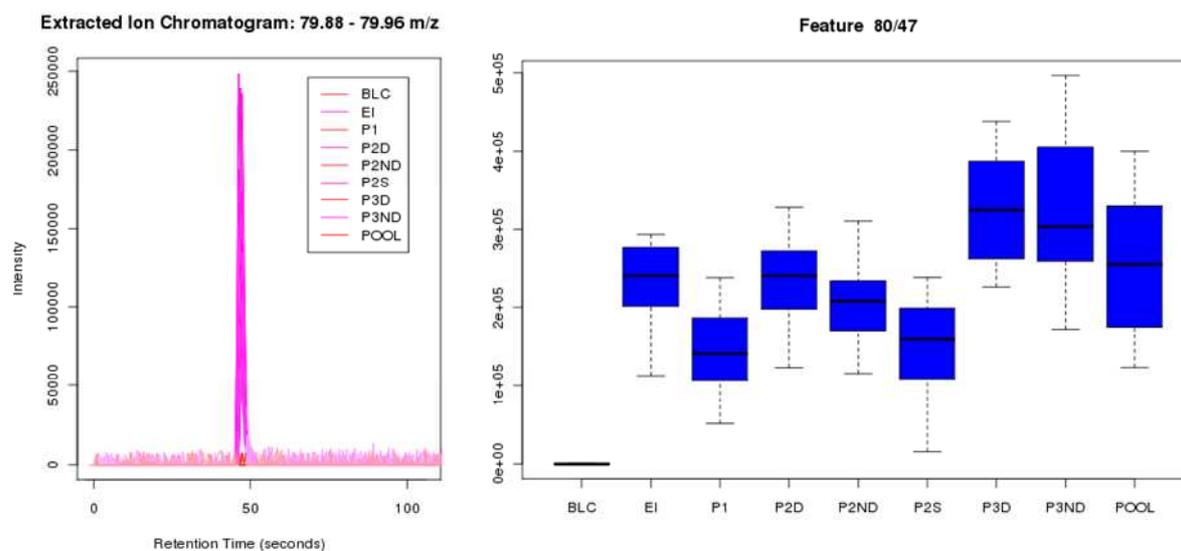


Figure 82 : Exemple de superposition de chromatogrammes (gauche) et de boxplot (droite) générés par XCMS.

## 2. PARAMETRES XCMS

Voici le récapitulatif des paramètres utilisés pour ce travail de thèse. Ils sont classés en fonction des différentes étapes du processus de traitement des données expliqué dans ce chapitre.

Tableau 18 : Paramètres XCMS utilisés

	Paramètre	Définition et choix du paramètre	Paramètres utilisés
findPeaks.matchedFilter-methods	<b>fwhm</b>	Largeur des pics chromatographiques à mi-hauteur (en secondes), dépendant de la méthode utilisée (HPLC ou UPLC)	4
	<b>step</b>	Largeur de la tranche m/z lors du découpage du chromatogramme dans le domaine spectral afin de créer les groupes de pics (" <i>Extracted Ion Base Peak Chromatograms</i> " ou EIBPC).	0.05
	<b>steps</b>	Nombre de EIBPC voisins combinés avant le filtre et la détection des pics.	2
	<b>mzdiff</b>	Différence minimale en m/z pour des pics dont les temps de retentions se superposent.	0.01
	<b>max</b>	Nombre maximum de pics chromatographiques par EIBPC	50
	<b>snthresh</b>	Rapport signal/bruit maximum (pour le tri après filtration)	3
group-methods	<b>minfrac_GA</b>	Minimum relatif (%) d'échantillon où apparaît un pic pour faire un groupe	0.5
	<b>minsamp_GA</b>	Minimum absolu d'échantillon où apparaît un pour faire un groupe	1
	<b>bw_GA</b>	Ecart-type ou demi-largeur à mi-hauteur de la gaussienne lissée, dépendant de la méthode utilisée (HPLC ou UPLC)	5
	<b>mzwid_GA</b>	Largeur de l'overlap de la tranche en m/z pour créer un groupe.	0.01
	<b>max_GA</b>	Nombre maximum de pics chromatographiques par tranche de m/z.	50
retcor-methods	<b>missing</b>	Nombre d'échantillon dont le pic peut-être absent pour garder un groupe	1
	<b>extra</b>	Nombre d'échantillon dont le pic peut-être présent en double pour garder un groupe.	1
	<b>method</b>	Méthode d'alignement du temps de rétention (" <i>loess</i> " pour l'alignement non linéaire ou " <i>linear</i> " pour l'alignement linéaire)	loess
	<b>span</b>	Degré de lissage pour la régression polynomiale locale	0.2
	<b>family</b>	Modèle pour la forme du pic chromatographique	gaussian
	<b>plottype</b>	Type de graphique obtenu	deviation
group-methods	<b>minfrac_GB</b>	Minimum relatif (%) d'échantillon où apparaît un pic pour faire un groupe	0.3
	<b>minsamp_GB</b>	Minimum absolu d'échantillon où apparaît un pour faire un groupe	1
	<b>bw_GB</b>	Ecart-type ou demi-largeur à mi-hauteur de la gaussienne lissée, dépendant de la méthode utilisée (HPLC ou UPLC)	5
	<b>mzwid_GB</b>	Largeur de l'overlap de la tranche en m/z pour créer un groupe.	0.01
	<b>max_GB</b>	Nombre maximum de pics chromatographiques par tranche de m/z.	50
diffreport-methods	<b>class1</b>	Classe d'échantillon 1	BLC
	<b>class2</b>	Classe d'échantillon 2	BIO
	<b>eicmax</b>	Nombre de EIBPC les plus significatifs à créer	50
	<b>eicwidth</b>	Largeur de fenêtre des EIBPC produits	200
	<b>sortpval</b>	Classement en fonction de la p-value (=T=true=oui ou =F=false=non)	T
	<b>metlin</b>	Incertitude sur la masse en vue d'une interrogation de metlin (0 ou F supprime la colonne)	0



## ANNEXE 3 – LISTE DES METABOLITES

### IDENTIFIES DANS LES URINES HUMAINES

#### LEGENDE :

<b>Statut</b>	a : m/z correspondant à structure
	b : MS <sup>n</sup> cohérente avec la structure
	c : tr identiques à ceux du standard (C18 et PFPP)
	d : MS <sup>n</sup> identique à celle du standard
<b>Présence dans les urines</b>	a : métabolite répertorié dans HMDB
	b : métabolite présent dans les urines selon HMDB
	c : métabolite présent dans les urines selon PubMed (recherche = nom métabolite + urine)

Tableau 19 : Tableau récapitulatif des métabolites retrouvés dans les urines lors de l'analyse de la cohorte.

Molécule	formule brute	Statut	Interprétation statistique	Informations sur l'identification	Famille chimique	Voie(s) métabolique(s) (KEGG)	Présence dans les urines	Masse précise	RT C18 (min)	RT HSF5 (min)
Urée	CH4N2O	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés / Cycle de l'urée	Oui <sup>a,b</sup>	60.0324	0.91	NR
Oxyde de Triméthylamine	C3H9NO	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme du méthane / Métabolisme microbien	Oui <sup>a,b</sup>	75.06841	0.93	8.75
Acide Pyruvique	C3H4O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Glycolyse / Glyconéogénèse / Cycle de Krebs	Oui <sup>a,b</sup>	88.0160	1.07/1.63	NR
Alanine isomère	C3H7NO2	Caractérisé <sup>a</sup>	Oui	Les tr en C <sub>18</sub> correspondent à la fois à l'Alanine, la Sarcosine et la Beta-alanine, mais les tr en PFPP ne correspondent à aucuns de ces isomères. Autres isomères possibles : Lactamide, Ester de méthyle Glycine...	NA	NA	NA	89.0477	0.94	4.06
Acide Oxalique	C2H2O4	Caractérisé <sup>a</sup>	Oui	NA	Acides organiques	Dégradation des chloroalcènes et chloroalcènes / Métabolisme du glyoxylate et des dicarboxylates / Métabolisme microbien	Oui <sup>a,b</sup>	89.9953	1.02	NR
Acide Lactique	C3H6O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Glyconéogénèse / Métabolisme du pyruvate	Oui <sup>a,b</sup>	90.0317	1.08	5.46
Aniline isomère	C6H7N	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Aniline ou à la 2-picoline. Autre isomère possible : 3-picoline, 4-picoline.	NA	NA	NA	93.05785	0.95	11.76
Acide Valérique isomère	C5H10O2	Caractérisé <sup>a</sup>	Oui	Autres isomères possibles : Acides Ethylméthylacétique, Isovalérique ou	NA	NA	NA	102.0681	5.14	NR

				Valérique.						
Diméthylglycine	C4H9NO2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme de la bétaine / Métabolisme des acides aminés	Oui <sup>a,b</sup>	103.0633	1.02	4.04
Acide 2-Aminobutyrique	C4H9NO2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	103.0633	1.02	5.38
Acide Aminobutyrique isomère	C4H9NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 2-Aminoisobutyrique, 3-Aminoisobutyrique ou 4-aminobutyrique. Autres isomères possibles : l'Acide 3-Aminobutyrique, N-Méthylalanine, N-Ethylglycine, N-Acétyléthanolamine...	NA	NA	NA	103.0633	1.02	7.80
Acide Aminobutyrique isomère	C4H9NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 2-Aminoisobutyrique, 3-Aminoisobutyrique ou 4-aminobutyrique. Autres isomères possibles : l'Acide 3-Aminobutyrique, N-Méthylalanine, N-Ethylglycine, N-Acétyléthanolamine...	NA	NA	NA	103.0633	1.02	9.08
Acide Aminobutyrique isomère	C4H9NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 2-Aminoisobutyrique, 3-Aminoisobutyrique ou 4-aminobutyrique. Autres isomères possibles : l'Acide 3-Aminobutyrique, N-Méthylalanine, N-Ethylglycine, N-Acétyléthanolamine...	NA	NA	NA	103.0633	1.02	12.20
Acide Malonique	C3H4O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme des acides aminés / Métabolisme des pyrimidines /	Oui <sup>a,b</sup>	104.0110	1.07	5.83

						Synthèse des acides gras				
Choline	C5H14NO+	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme de la bétaine / Métabolisme de la méthionine / Synthèse des phospholipides	Oui <sup>a,b</sup>	104.107	0.90	7.50
Sérine	C3H7NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Métabolisme des shingolipides / Métabolisme du soufre	Oui <sup>a,b</sup>	105.0426	0.88	3.60
Acide Glycérique	C3H6O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme des glycérolipides / Métabolisme des acides aminés	Oui <sup>a,b</sup>	106.0266	0.97	3.89
Aminophénol isomère	C6H7NO	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas au 3-Aminophénol ou au 2-Aminophénol.	NA	NA	NA	109.0528	1.05	12.95
Aminophénol isomère	C6H7NO	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas au 3-Aminophénol ou au 2-Aminophénol.	NA	NA	NA	109.0528	1.05	13.75
4-Aminophénol	C6H7NO	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Alcools et polyols / Phénols et dérivés	Métabolite minoritaire mais néphrotoxique de la phénacétine et du paracétamol.	Oui <sup>a,c</sup> [384]	109.0528	1.05	10.90
Cytosine	C4H5N3O	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des pyrimidines	Non <sup>a</sup>	111.0433	1.09	NR
Histamine	C5H9N3	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme de l'histidine	Oui <sup>a,b</sup>	111.0796	0.83	NR
Uracile	C4H4N2O2	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme de la beta-alanine / Métabolisme des pyrimidines	Oui <sup>a,b</sup>	112.0273	1.16	NR
Créatinine	C4H7N3O	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés	Oui <sup>a,b</sup>	113.0589	0.94	9.30

Créatinine isomère	C4H7N3O	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas créatinine.	NA	NA	NA	113.0589	0.94	8.30
5,6-Dihydrouracile	C4H6N2O2	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme de la beta-alanine / Métabolisme des pyrimidines / Synthèse du pantothénate et des CoA	Oui <sup>a,b</sup>	114.0429	1.10	5.90
Proline	C5H9NO2	Annoté <sup>a,c</sup>	Oui mais potentiellement d'autres ions sous le même pic chromatographique (fragments de Gly-Pro et Pro-Gly)	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	115.0633	1.07	4.74
Acide Fumarique	C4H4O4	Identifié <sup>a,c,d</sup>	Oui	NA	Acides organiques	Métabolisme des acides aminés / Cycle de l'urée / Cycle de Krebs	Oui <sup>a,b</sup>	116.0110	1.08	11.37
Acide Maléique	C4H4O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme du butanoate / Métabolisme du nicotinate et du nicotinamide / Métabolisme microbien	Oui <sup>a,b</sup>	116.0110	1.40	11.48
Acide Guanidoacétique	C3H7N3O2	Identifié <sup>a,b,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Cycle de l'urée	Oui <sup>a,b</sup>	117.0538	0.90	4.85
Acide Guanidoacétique isomère	C3H7N3O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas l'Acide Guanidoacétique.	NA	NA	NA	117.0538	0.90	5.95
Bétaine	C5H11NO2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	117.0790	0.91	4.25
Valine	C5H11NO2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Métabolisme du propanoate	Oui <sup>a,b</sup>	117.0790	0.91	7.00
Acide Méthylmalonique	C4H6O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme des acides aminés / Métabolisme du	Oui <sup>a,b</sup>	118.0266	2.31	10.88

						propanoate				
Acide Méthylmalonique isomère	C4H6O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au Diméthyl oxalate, Acide Succinique, Erythrono-1,4-lactone ou Acide Méthylmalonique. Autres isomères possibles : Acide Acétoxyacétique, Oxalate de monoéthyle...	NA	NA	NA	118.0266	1.71	8.37
Thréonolactone	C4H6O4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O et CO <sub>2</sub> .	Alcools et polyols	Produit d'auto-oxydation de l'acide ascorbique	Non <sup>a</sup>	118.0266	1.96	9.60
Acide 3-Hydroxy 3-méthylbutyrique	C5H10O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	NA	Oui <sup>a,b</sup>	118.0630	3.95	11.30
Acide 2-Hydroxy 2-méthylbutyrique	C5H10O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	NA	Oui <sup>a,b</sup>	118.0630	4.53	11.75
Acide 2-Hydroxy 3-méthylbutyrique	C5H10O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	NA	Oui <sup>a,b</sup>	118.0630	4.85	12.20
Thréonine	C4H9NO3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Dégradation de la thréonine et du 2-oxobutanoate	Oui <sup>a,b</sup>	119.0582	1.02	4.01
Thréonine isomère	C4H9NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr en C <sub>18</sub> correspondent à la fois à la thréonine et à l'homosérine, mais pas les tr en PFPP. Autre isomères possibles : Méthylsérine ou Acide Aminohydroxybutyrique .	NA	NA	NA	119.0582	1.02	5.02
Acide Benzoïque isomère	C7H6O2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent ni à l'acide Benzoïque ni au 4-Hydroxybenzaldéhyde.	NA	NA	NA	122.0368	5.38	NR
Nicotinamide	C6H6N2O	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme du nicotinate et du nicotinamide	Oui <sup>a,b</sup>	122.0480	1.10	11.02

Erithrol (+ isomères)	C4H10O4	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les polyols en C <sub>4</sub> (Erithrol, Threitol...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Alcools et polyols / Alcools de sucres	Interconversion des pentoses et du gluconate / Voie des pentoses phosphates	Oui <sup>a,b</sup>	122.0579	0.91	3.61
Acide Nicotinique	C6H5NO2	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme du nicotinate et du nicotinamide	Oui <sup>a,b</sup>	123.0320	1.10	5.47
Acide Nicotinique isomère	C6H5NO2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide Picolinique ni à l'Acide Nicotinique.	NA	NA	NA	123.0320	4.47	12.15
Alcool Hydroxybenzylique isomère	C7H8O2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au Guaiacol, Méthylcatechol, ou à l'alcool 4-Hydroxybenzylique.	NA	NA	NA	124.0524	6.20	14.91
Alcool 3-hydroxybenzylique	C7H8O2	Annoté <sup>a,c</sup>	Oui	NA	Alcools et polyols / Phénols et dérivés	Dégradation du toluène / Métabolisme microbien	Non	124.0524	5.13	12.62
Alcool 2-hydroxybenzylique ( Saligénine)	C7H8O2	Annoté <sup>a,c</sup>	Oui	NA	Alcools et polyols / Phénols et dérivés	produit de dégradation de la salycine, un anestésique local.	Non	124.0524	5.59	NR
Taurine	C2H7NO3S	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Synthèse d'acides biliaires / Métabolisme de la taurine et de l'hypotaurine	Oui <sup>a,b</sup>	125.0147	0.90	3.50
Taurine isomère	C2H7NO3S	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le spectre de masse montre la présence d'un isotope <sup>34</sup> S.	NA	NA	NA	125.0147	0.90	3.60
Méthylcytosine isomère	C5H7N3O	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à la 5-Méthylcytosine.	NA	NA	NA	125.0589	1.02	10.88
Méthylcytosine isomère	C5H7N3O	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à la 5-Méthylcytosine.	NA	NA	NA	125.0589	1.02	11.42
Méthylcytosine isomère	C5H7N3O	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à la 5-Méthylcytosine.	NA	NA	NA	125.0589	1.02	13.10

N3-Méthylhistamine	C6H11N3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme de l'histidine	Oui <sup>a,b</sup>	125.0953	0.85	11.71
N1-Méthylhistamine	C6H11N3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme de l'histidine	Oui <sup>a,b</sup>	125.0953	0.85	32.96
Méthylhistamine isomère	C6H11N3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à la N1, N3 ou la Nα Méthylhistamine. Autres isomères possibles : α, 4 ou 2-Méthylhistamine.	NA	NA	NA	125.0953	0.85	37.85
Nα-Méthylhistamine	C6H11N3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme de l'histidine	Non <sup>c</sup> [385]	125.0953	0.85	6.11
Acide Imidazoleacétique isomère	C5H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à l'Acide Imidazoleacétique. Autres isomères possibles : Thymine, Méthyluracile, Acide Méthylpyrazole carboxylique ou Imidazolecarboxylate de méthyle.	NA	NA	NA	126.0429	1.05	7.60
Acide Imidazoleacétique isomère	C5H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à l'Acide Imidazoleacétique. Autres isomères possibles : Thymine, Méthyluracile, Acide Méthylpyrazole carboxylique ou Imidazolecarboxylate de méthyle.	NA	NA	NA	126.0429	1.05	9.20
Acide Imidazoleacétique isomère	C5H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à l'Acide Imidazoleacétique. Autres isomères possibles : Thymine, Méthyluracile, Acide Méthylpyrazole	NA	NA	NA	126.0429	1.05	11.29

				carboxylique ou Imidazolecarboxylate de méthyle.						
Hydrouracile isomère	C5H8N2O2	-aractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	128.0586	5.84	4.95
Hydrouracile	C5H8N2O2	Caractérisé <sup>a</sup>	Oui	Les tr semblent cohérents par rapport à ceux de l'Uracile et la Dihydrouracile.	NA	Métabolisme des pyrimidines	Oui <sup>a,b</sup>	128.0586	0.96	6.75
Oxoproline	C5H7NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme du glutathion	Oui <sup>a,b</sup>	129.0426	1.10/1.60	5.50/8.00
Oxoproline isomère	C5H7NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Oxoproline.	NA	NA	NA	129.0426	5.80	13.05
Acide Pipécolique	C6H11NO2	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme de la lysine / Synthèse de métabolites secondaires	Oui <sup>a,b</sup>	129.0790	1.06	4.80
Acide 3-Méthyl-2-oxovalérique	C6H10O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme des acides aminés / Synthèse de métabolites secondaires	Oui <sup>a,b</sup>	130.0630	5.30	14.00
Acide 4-Méthyl-2-oxovalérique (kétoleucine)	C6H10O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Synthèse du pantothénate et des CoA / Métabolisme des acides aminés / Synthèse de métabolites secondaires	Oui <sup>a,b</sup>	130.0630	5.64	14.62
N-Acétyleputrescine	C6H14N2O	Identifié <sup>a,b,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés	Oui <sup>a,b</sup>	130.1106	1.05	11.40
N-Acétyleputrescine isomère	C6H14N2O	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à la N-Acétyleputrescine. Autres isomères possibles : L-Leucinamide, 2-Amino-N,N-diéthylacétamide (...)	Amines biogènes et Polyamines	NA	NA	130.1106	1.05	10.90
Hydroxyproline	C5H9NO3	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères	Les tr correspondent à la fois à la Cis-4-hydroxyproline et la	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	131.0582	1.05/1.60	3.91

			pour un même pic chromatographique	Trans-3-Hydroxyproline mais pas à la Trans-4-hydroxyproline.						
N-Acétylalanine	C5H9NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Synthèse de métabolites secondaires	Oui <sup>a,b</sup>	131.0582	2.02	9.63
Hydroxyproline isomère	C5H9NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Hydroxyproline (Cis-4, trans-4, Trans-3), la N-Acétylalanine ou l'Acide 5-Aminolevulinique.	NA	NA	NA	131.0582	1.05/1.60	6.00
Hydroxyproline isomère	C5H9NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Hydroxyproline (Cis-4, trans-4, Trans-3), la N-Acétylalanine ou l'Acide 5-Aminolevulinique.	NA	NA	NA	131.0582	1.05/1.60	7.30
Hydroxyproline isomère	C5H9NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Hydroxyproline (Cis-4, trans-4, Trans-3), la N-Acétylalanine ou l'Acide 5-Aminolevulinique.	NA	NA	NA	131.0582	1.05/1.60	8.20
Hydroxyproline isomère	C5H9NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Hydroxyproline (Cis-4, trans-4, Trans-3), la N-Acétylalanine ou l'Acide 5-Aminolevulinique.	NA	NA	NA	131.0582	1.05/1.60	8.97
Créatine	C4H9N3O2	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	131.0695	0.96	5.15
Leucine	C6H13NO2	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	131.0946	1.05	12.90
Norleucine	C6H13NO2	Identifié <sup>a,c,d</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	131.0946	2.42	14.20
Leucine isomère	C6H13NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Isoleucine, la β-Leucine ou l'Acide 6-Aminocaproïque.	NA	NA	NA	131.0946	1.05	12.30

Leucine isomère	C6H13NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Isoleucine, la β-Leucine ou l'Acide 6-Aminocaproïque.	NA	NA	NA	131.0946	1.05	13.50
Asparagine	C4H8N2O3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Métabolisme de l'azote	Oui <sup>a,b</sup>	132.0535	1.05	3.77
Acide 3-uréidopropionique	C4H8N2O3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Synthèse du pantothénate et des CoA / Métabolisme des acides aminés / Métabolisme des pyrimidines	Oui <sup>a,b</sup>	132.0535	1.05/1.14	5.43
Acide Aspartique	C4H7NO4	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Cycle de l'urée	Oui <sup>a,b</sup>	133.0375	0.95	3.90
Acide Malique	C4H6O5	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Cycle de Krebs / Métabolisme des pyrimidines	Oui <sup>a,b</sup>	134.0215	1.05	4.75
Déoxyribose	C5H10O4	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les Deoxypentose ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés	Voie des pentoses phosphates	Non <sup>a</sup>	134.0579	1.00	3.67
Déoxyribose isomère	C5H10O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au Déoxyribose. Autres isomères possibles : glycerol 1-acétate, Acide 2,3-Dihydroxyvalérique ou 2,3-Dihydroxy-isovalérique (...)	NA	NA	NA	134.0579	2.44	6.73
Acide Thréonique / Acide Erythronique	C4H8O5	Annoté <sup>a,c</sup>	Oui	L'Acide Erythronique et Thréonique sont deux isomères possibles.	Acides organiques	Métabolisme de l'ascorbate et l'aldarate	Oui <sup>a,b</sup>	136.0372	0.93	3.66
Hypoxanthine	C5H4N4O	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	136.0385	1.10/1.50	10.50
Hypoxanthine isomère	C5H4N4O	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Hypoxanthine.	NA	NA	NA	136.0385	4.58	6.65
1-Méthylnicotinamide	C7H9N2O+	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Amines biogènes et Polyamines	Métabolisme du nicotinate et du nicotinamide	Oui <sup>a,b</sup>	136.0637	0.91	9.20

Méthylnicotinamide isomère	C7H9N2O+	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à la 1-Méthylnicotinamide. Isomères possibles : N-méthylnicotinamide, 2-méthylnicotinamide, 6-Méthylnicotinamide.	NA	NA	NA	136.0637	0.91	8.00
Trigonelline	C7H7NO2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Composés exogènes, xénobiotiques et leur métabolites	Métabolisme du nicotinate et du nicotinamide	Oui <sup>a,b</sup>	137.0477	0.94	5.45
Trigonelline isomère	C7H7NO2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Aminobenzoïque, Pyridylacétique ou à la Salicylamide.	NA	NA	NA	137.0477	0.94	4.90
Acide Salicylique (Acide 2-hydroxybenzoïque)	C7H6O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Utilisé comme médicament et comme précurseur de l'acide acétylsalicylique, l'aspirine.	Non <sup>a</sup>	138.0317	7.35	17.00
Acide 4-Hydroxybenzoïque	C7H6O3	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Synthèse de l'ubiquinone	Oui <sup>a,b</sup>	138.0317	5.51	13.58
Acide Aminonicotinique isomère	C6H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Urocanique ou à l'oxyde de Nicotinamide. Plusieurs isomères possible d'acide Aminopyridine-carboxylique.	NA	NA	NA	138.0429	1.09	12.40
Acide Aminonicotinique isomère	C6H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Urocanique ou à l'oxyde de Nicotinamide. Plusieurs isomères possible d'acide Aminopyridine-carboxylique.	NA	NA	NA	138.0429	1.09	13.60
Acide Aminonicotinique isomère	C6H6N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Urocanique ou à l'oxyde de Nicotinamide. Plusieurs isomères possible	NA	NA	NA	138.0429	1.09	14.45

				d'acide Aminopyridine-carboxylique.						
4-Nitrophénol	C6H5NO3	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolite phénolique urinaire indiquant l'exposition aux pesticides.	Oui <sup>a,b</sup>	139.0269	7.50	17.09
Acide Hydroxynicotinique isomère	C6H5NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 3-Hydroxypicolinique ou 6-hydroxynicotinique. Plusieurs isomères possibles d'Acide Hydroxynicotinique, Hydroxyisonicotinique ou Hydroxypicolinique.	NA	NA	NA	139.0269	1.07/1.52	9.80
Acide Hydroxynicotinique isomère	C6H5NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 3-Hydroxypicolinique ou 6-hydroxynicotinique. Plusieurs isomères possibles d'Acide Hydroxynicotinique, Hydroxyisonicotinique ou Hydroxypicolinique.	NA	NA	NA	139.0269	1.07/1.52	10.80
Acide Hydroxynicotinique isomère	C6H5NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide 3-Hydroxypicolinique ou 6-hydroxynicotinique. Plusieurs isomères possibles d'Acide Hydroxynicotinique, Hydroxyisonicotinique ou Hydroxypicolinique.	NA	NA	NA	139.0269	1.07/1.52	11.43
Acide Hydroxynicotinique isomère	C6H5NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide 3-Hydroxypicolinique ou 6-hydroxynicotinique. Plusieurs isomères possibles d'Acide Hydroxynicotinique, Hydroxyisonicotinique ou Hydroxypicolinique.	NA	NA	NA	139.0269	4.48	13.21
Acide 1-Méthyl-4-imidazoleacétique	C6H8N2O2	Identifié <sup>a,b,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme de l'histidine	Oui <sup>a,b</sup>	140.0586	1.03	8.40

Acide Méthylimidazole-acétique isomère	C6H8N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Autres isomères possibles : Acide 1-Methyl-5-imidazoleacétique, Diméthyluracil, Acide Imidazolepropionique, 1-Méthylthymine (...)	NA	NA	NA	140.0586	1.03	10.98
Acide Méthylimidazole-acétique isomère	C6H8N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Autres isomères possibles : Acide 1-Methyl-5-imidazoleacétique, Diméthyluracil, Acide Imidazolepropionique, 1-Méthylthymine (...)	NA	NA	NA	140.0586	1.03	11.95
Acide Méthylimidazole-acétique isomère	C6H8N2O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Autres isomères possibles : Acide 1-Methyl-5-imidazoleacétique, Diméthyluracil, Acide Imidazolepropionique, 1-Méthylthymine (...)	NA	NA	NA	140.0586	1.03	13.33
5-hydroxyméthyluracile	C5H6N2O3	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	La 5-Hydroxyméthyluracile est un produit d'oxydation dérivé de la thymine ou de la 5-méthylcytosine.	Oui <sup>a,b</sup>	142.0378	1.05	4.90
Stachydrine (Proline bétaine)	C7H13NO2	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de HCOOH, C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> et C <sub>4</sub> H <sub>6</sub> O <sub>2</sub> .	Composés exogènes, xénobiotiques et leurs métabolites	Analogue de la glycine bétaine retrouvés dans les agrumes.	Oui <sup>a,b</sup>	143.0946	1.00	5.45
Acide 4-Acétamidobutanoïque isomère	C6H11NO3	Caractérisé <sup>a</sup>	Oui	Autres isomères possibles : Isobutyrylglycine, N-Butyrylglycine, Acide 4-Acétamidobutanoïque (...)	NA	NA	NA	145.0739	1.80	10.85
Acide 4-Acétamidobutanoïque isomère	C6H11NO3	Caractérisé <sup>a</sup>	Oui	Autres isomères possibles : Isobutyrylglycine, N-Butyrylglycine, Acide 4-Acétamidobutanoïque (...)	NA	NA	NA	145.0739	2.80	11.05
Acide 4-Acétamidobutanoïque isomère	C6H11NO3	Caractérisé <sup>a</sup>	Oui	Autres isomères possibles : Isobutyrylglycine, N-Butyrylglycine, Acide 4-Acétamidobutanoïque	NA	NA	NA	145.0739	3.80	11.40

				(...)						
Spermidine	C7H19N3	Identifié <sup>a,c,d</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés / Métabolisme du glutathion	Oui <sup>a,b</sup>	145.1579	0.81	NR
Acide α-Kétoglutarique	C5H6O5	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques	Cycle de Krebs / Métabolisme des acides aminés / Synthèse du pantothénate et des CoA	Oui <sup>a,b</sup>	146.0215	1.06	4.56
Acide Adipique	C6H10O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Dégradation du caprolactam / Métabolisme microbien	Oui <sup>a,b</sup>	146.0579	5.08	12.22
Acide Méthylglutarique	C6H10O4	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Les tr correspondent à la fois à l'Acide 2-Méthylglutarique et 3-Méthylglutarique.	Acides organiques	Cycle de Krebs / Métabolisme des acides aminés	Oui <sup>a,b</sup>	146.0579	5.14	12.39
Acide 2,2-Diméthylsuccinique	C6H10O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	NA	Oui <sup>a,b</sup>	146.0579	5.52	12.87
Acide Méthylglutarique isomère	C6H10O4	Caractérisé <sup>a</sup>	Yes	Les tr ne correspondent pas à l'Acide 2-Méthylglutarique, 3-Méthylglutarique ou Monométhylglutarique.	NA	NA	NA	146.0579	4.21	NR
Glutamine	C5H10N2O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Purine, Pyrimidine, Cycle de l'urée	Oui <sup>a,b</sup>	146.0691	0.88	3.92
Lysine	C6H14N2O2	Identifié <sup>a,c,d</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Synthèse des carnitines / Métabolisme de la biotine	Oui <sup>a,b</sup>	146.1055	0.81	6.15
O-Acétylsérine	C5H9NO4	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme de la méthionine et cystéine / Métabolisme du soufre / Métabolisme microbien	Non <sup>a</sup>	147.0532	1.05	4.05

O-Acétylsérine isomère	C5H9NO4	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide Glutamique, l'Acide N-Méthyl-D-aspartique (NMDA) ou la N-Acetylserine.	NA	NA	NA	147.0532	1.05	10.80
Acide Mévalonique isomère	C6H12O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide Mévalonique.	NA	NA	NA	148.0736	4.55	11.25
Acide Mévalonique isomère	C6H12O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide Mévalonique.	NA	NA	NA	148.0736	5.20	12.40
Acide (2-méthoxyéthoxy) propanoïque	C6H12O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acide Mévalonique. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH, C <sub>3</sub> H <sub>6</sub> O (+H <sub>2</sub> O), C <sub>3</sub> H <sub>8</sub> O, C <sub>3</sub> H <sub>4</sub> O <sub>2</sub> (+H <sub>2</sub> O). 2 Autres isomères possibles : Acide 2-(2-méthoxyéthoxy) propanoïque ou 3-(2-méthoxyéthoxy) propanoïque.	Acides organiques	NA	Non	148.0736	4.95	11.55
Acide (2-méthoxyéthoxy) propanoïque	C6H12O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acide Mévalonique. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH, C <sub>3</sub> H <sub>6</sub> O (+H <sub>2</sub> O), C <sub>3</sub> H <sub>8</sub> O, C <sub>3</sub> H <sub>4</sub> O <sub>2</sub> (+H <sub>2</sub> O). 2 Autres isomères possibles : Acide 2-(2-méthoxyéthoxy) propanoïque ou 3-(2-méthoxyéthoxy) propanoïque.	Acides organiques	NA	Non	148.0736	4.75	11.80
Méthionine	C5H11O2NS	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	149.0511	1.05/1.35	8.10
Méthionine isomère	C5H11O2NS	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à la Méthionine. Autres isomères possibles : Thiaïsoleucine, Ethylcystéine ou	NA	NA	NA	149.0511	1.05/1.35	11.13

				Pénicillamine.						
Méthyladenine isomère	C6H7N5	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à N6-Méthyladenine. Peut être un autre isomère de méthyladenine (1, 3, 7 ou 9)	NA	NA	NA	149.0701	1.09	12.60
Méthyladenine isomère	C6H7N5	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à N6-Méthyladenine. Peut être un autre isomère de méthyladenine (1, 3, 7 ou 9)	NA	NA	NA	149.0701	1.09	13.13
Pentose	C5H10O5	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les pentose (ribose, arabinose, xylose...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés / Monosaccharides	Interconversion des pentoses et du gluconate / Voie des pentoses phosphates	Oui <sup>a,b</sup>	150.0528	0.97	3.93
Guanine	C5H5N5O	Caractérisé <sup>a</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	151.0494	1.09	11.31
Paracétamol	C8H9NO2	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Médicament contre la douleur.	Oui <sup>a,b</sup>	151.0633	4.88	12.36
Paracétamol isomère	C8H9NO2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au paracétamol. Autres isomères possibles : 2-Phénylglycine, Acide 2-Ethylisonicotinique, Leukoaminochrome, Dopamine quinone (...)	NA	NA	NA	151.0633	5.70	13.50
Xanthine	C5H4N4O2	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	152.0334	1.14/1.73	NR
Acide 3-Hydroxyphényl-acétique	C8H8O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Métabolisme des acides aminés / Métabolisme microbien	Oui <sup>a,b</sup>	152.0473	6.00	13.85
Acide 2-Hydroxyphényl-acétique	C8H8O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Métabolisme des acides aminés / Métabolisme microbien	Oui <sup>a,b</sup>	152.0473	6.20	14.28

Acide p-Anisique	C8H8O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques	Acide phénolique présent dans les graines d'anis.	Oui <sup>a,b</sup>	152.0473	7.18	15.88
Acide 4-Hydroxyphényl-acétique	C8H8O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Métabolisme des acides aminés / Métabolisme microbien	Oui <sup>a,b</sup>	152.0473	5.66	13.42
Arabitol (+ isomères)	C5H12O5	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les polyols en C5 (Adonitol, Arabitol...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Alcools et polyols / Alcools de sucres	Interconversion des pentoses et du gluconate / Voie des pentoses phosphates	Oui <sup>a,b</sup>	152.0685	4.17	3.50
Acide Hydroxy-anthranilique isomère	C7H7NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide 3-Hydroxyanthranilique ou 3-Aminosalicilylique.	NA	NA	NA	153.0426	4.55	12.00
Acide Hydroxy anthranilique isomère	C7H7NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide 3-Hydroxyanthranilique ou 3-Aminosalicilylique.	NA	NA	NA	153.0426	4.70	13.50
Acide 3,5-dihydroxybenzoïque / Acide 3,4-dihydroxybenzoïque	C7H6O4	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Les tr correspondent à la fois à l'acide 3,5-dihydroxybenzoïque et à l'acide 3,4-dihydroxybenzoïque.	Acides organiques / Acides Phénoliques	Dégradation du benzoate / AminoDégradation du benzoate / Métabolisme microbien	Non <sup>a</sup>	154.0266	4.85	12.50
Acide 2,5-dihydroxybenzoïque (Acide Gentisique)	C7H6O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Dégradation du benzoate / AminoDégradation du benzoate / Métabolisme microbien	Oui <sup>a,b</sup>	154.0266	5.56	14.02
Acide 2,3-dihydroxybenzoïque (Acide 2-Pyrocatechique)	C7H6O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Dégradation du benzoate / AminoDégradation du benzoate / Métabolisme microbien	Oui <sup>a,b</sup>	154.0266	5.96	14.45
Acide dihydroxybenzoïque isomère	C7H6O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas ni à l'acide 3,4-Dihydroxybenzoïque ni à l'acide 3,5-Dihydroxybenzoïque. Isomères possibles : acides 2,4 ou 2,6-Dihydroxybenzoïque.	NA	NA	NA	154.0266	6.18	16.35

Acide Pyrrole-dicarboxylique	C6H5NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre deux pertes de CO <sub>2</sub> . Isomère possible : Acide 1H-pyrrole-2,5-dicarboxylique ou Acide 1H-pyrrole-3,4-dicarboxylique.	Acides organiques	NA	Non	155.0219	5.20	12.62
Histidine	C6H9N3O2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	155.0695	0.85	5.45
Histidine isomère	C6H9N3O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Histidine.	NA	NA	NA	155.0695	0.85	6.05
Acide Orotique	C5H4N2O4	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides organiques	Métabolisme des pyrimidines	Oui <sup>a,b</sup>	156.0171	1.04	5.42
Acide Orotique isomère	C5H4N2O4	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Acide orotique.	NA	NA	NA	156.0171	1.04	3.49
Acide Imidazolelactique isomère	C6H8N2O3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Imidazolelactique. La MS <sup>n</sup> montre une perte de HCOOH.	NA	NA	NA	156.0535	0.97	6.80
Acide Imidazolelactique isomère	C6H8N2O3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'acide Imidazolelactique. La MS <sup>n</sup> montre une perte de HCOOH.	NA	NA	NA	156.0535	0.97	8.90
Tiglylglycine	C7H11NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N (-CO). La distinction entre les deux isomères a été faite sur la base du logP donné par HMDB	Acyle Glycines / Dérivés Glycine aliphatiques	Métabolisme des acides gras	Oui <sup>a,b</sup>	157.0739	5.25	12.50
3-Methylcrotonyl-glycine	C7H11NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N (-CO). La distinction entre les deux isomères a été faite sur la base du logP donné par HMDB	Acyle Glycines / Dérivés Glycine aliphatiques	Métabolisme des acides gras	Oui <sup>a,b</sup>	157.0739	5.15	12.65

Valérylglycine isomère	C7H13NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>5</sub> H <sub>8</sub> O. Autres isomères possibles : 2-Méthylbutyrylglycine, Isovalérylglycine, Valérylglycine.	Acyle Glycines	NA	Oui <sup>a,b</sup>	159.0895	5.06	12.43
Valérylglycine isomère	C7H13NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>5</sub> H <sub>8</sub> O. Autres isomères possibles : 2-Méthylbutyrylglycine, Isovalérylglycine, Valérylglycine.	Acyle Glycines	NA	Oui <sup>a,b</sup>	159.0895	5.20	12.62
Carboxylate de méthyl (hydroxyméthyl)-pyrrolidine / Carboxylate de méthyl (hydroxy)pipéridine	C7H13NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> et H <sub>2</sub> O. 2 Autres isomères possibles : méthyl (hydroxyméthyl)-pyrrolidine-carboxylate ou méthyl (hydroxy)piperidine-carboxylate.	Acides organiques / Esters	NA	Non	159.0895	4.60	11.63
Acide 2-Oxadipique	C6H8O5	Caractérisé <sup>a</sup>	Oui	les tr C <sub>18</sub> et PFPP sont cohérent en comparaison de ceux de l'acide Adipique. La distinction entre les deux isomères a été faite sur la base du logP donné par HMDB.	Acides organiques	Métabolisme des acides aminés	Oui <sup>a,b</sup>	160.0372	4.65	11.75
Acide 3-Oxadipique	C6H8O5	Caractérisé <sup>a</sup>	Oui	les tr C <sub>18</sub> et PFPP sont cohérent en comparaison de ceux de l'acide Adipique. La distinction entre les deux isomères a été faite sur la base du logP donné par HMDB.	Acides organiques	Dégradation du benzoate / Métabolisme microbien	Oui <sup>a,b</sup>	160.0372	4.50	11.97
Acide Pimélique	C7H12O4	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides organiques	Métabolisme de la biotine	Oui <sup>a,b</sup>	160.0734	5.75	12.98
Acide 3-Méthyladipique	C7H12O4	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides organiques	Catabolisme de l'acide phytanique	Oui <sup>a,b</sup>	160.0734	5.75	13.07

Acide 3,3-Diméthylglutarique	C7H12O4	Identifié <sup>a,c,d</sup>	Oui	NA	Acides organiques	NA	Oui <sup>a,b</sup>	160.0734	5.85	13.14
Adipate de Monométhyle isomère	C7H12O4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	160.0734	4.38	NR
Acide 2,2-Diméthylglutarique	C7H12O4	Identifié <sup>a,c,d</sup>	Oui	NA	Acides organiques	NA	Non	160.0734	5.95	13.30
Adipate de Monométhyle	C7H12O4	Identifié <sup>a,c,d</sup>	Oui	NA	Acides organiques esters	NA	Non	160.0734	6.25	13.81
Tryptamine	C10H12N2	Identifié <sup>a,c,d</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés / synthèse des indolalkaloïdes	Oui <sup>a,b</sup>	160.1001	5.67	NR
Acide 2-Aminoadipique	C6H11NO4	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	161.0688	1.00	5.45
Carnitine	C7H15NO3	Identifié <sup>a,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acyles carnitines / Acyles carnitines à courte chaîne	Métabolisme des acides gras	Oui <sup>a,b</sup>	161.1052	0.91	8.34
Carnitine isomère	C7H15NO3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	le tr en PFPP ne correspond pas à la Carnitine.	NA	NA	NA	161.1052	0.91	10.09
Nicotine	C10H14N2	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la nicotine	Oui <sup>a,b</sup>	162.1157	1.10	11.06
Déoxyhexose	C6H12O5	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les Déoxyhexoses (Fucose, Rhamnose, Rhamnulose...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés	Glycolyse / Glyconéogénèse / Voie des pentoses phosphates / Métabolisme du galactose	Oui <sup>a,b</sup>	164.0685	1.02	3.70
Phénylalanine	C9H11NO2	Identifié <sup>a,c,d</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	165.0790	4.44	15.60
Acide Ribonique (+ isomères)	C5H10O6	Caractérisé <sup>a</sup>	Oui	Tous les Acides de sucre ne C5 ont les mêmes tr en C <sub>18</sub> et PFPP.	Acides organiques	Métabolisme de l'ascorbate et l'aldarate	Non <sup>a</sup>	166.0477	0.94	3.54

7-Méthylxanthine	C6H6N4O2	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine, théophylline et de la théobromine	Oui <sup>a,b</sup>	166.0491	3.96	11.13
1-Méthylxanthine	C6H6N4O2	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine, théophylline et de la théobromine	Oui <sup>a,b</sup>	166.0491	4.30	11.45
9-Méthylxanthine	C6H6N4O2	Caractérisé <sup>a</sup>	Oui	Le tr en C <sub>18</sub> correspond avec la 3-Méthylxanthine mais pas le tr en PFPP, probablement la 9-Méthylxanthine.	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine, théophylline et de la théobromine	Oui <sup>a,b</sup>	166.0491	4.24	11.23
Acide Phényllactique	C9H10O3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques	Métabolisme de la phénylalanine / Synthèse d'alcaloïdes dérivés du Tropane, de la piperidine et de la pyridine	Oui <sup>a,b</sup>	166.0629	6.47	14.53
Pyridoxal isomère	C8H9NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au Pyridoxal ou à la 4-Hydroxyphénylglycine. Isomère possible : isopyridoxal.	NA	NA	NA	167.0582	1.09	11.20
Acide Urique	C5H4N4O3	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines / Métabolisme microbien	Oui <sup>a,b</sup>	168.0283	1.11/1.40	5.44/7.65
Acide 4-Hydroxymandélique	C8H8O4	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques / Acides Phenoliques	AminoDégénération du benzoate / Métabolisme microbien	Oui <sup>a,b</sup>	168.0423	2.32	11.30
Acide 3,4-Dihydroxyphénylacétique	C8H8O4	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques / Acides Phenoliques	Métabolisme de la tyrosine / Métabolisme microbien	Oui <sup>a,b</sup>	168.0423	5.21	12.62
Acide Hydroxymandélique isomère	C8H8O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Homogentisique, Vanilique, 3-Hydroxymélicique ou 5-Méthoxysalicilyque. La MS <sup>n</sup> montre une perte de CO <sub>2</sub> .	NA	NA	NA	168.0423	5.60	13.21

Acide Méthoxysalicylique isomère	C8H8O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Homogentisique, Vanilique, 3-Hydroxymételicque ou 5-Méthoxysalicylique. La MS <sup>n</sup> montre une perte de CO <sub>2</sub> et de CH <sub>3</sub> .	NA	NA	NA	168.0423	6.15	14.90
Acide Méthoxysalicylique isomère	C8H8O4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Homogentisique, Vanilique, 3-Hydroxymételicque ou 5-Méthoxysalicylique. La MS <sup>n</sup> montre une perte de CO <sub>2</sub> et de CH <sub>3</sub> .	NA	NA	NA	168.0423	6.84	15.94
Pyridoxamine isomère	C8H12N2O2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à la Pyridoxamine. La MS <sup>n</sup> montre une perte de H <sub>2</sub> O et C <sub>3</sub> H <sub>9</sub> NO.	NA	NA	NA	168.0899	4.60	9.20
Pyridoxamine isomère	C8H12N2O2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à la Pyridoxamine. La MS <sup>n</sup> montre une perte de H <sub>2</sub> O, CO et C <sub>4</sub> H <sub>5</sub> NO <sub>2</sub> .	NA	NA	NA	168.0899	4.45	11.35
Méthylhistidine	C7H11N3O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en C <sub>18</sub> est compatible avec la méthylhistidine, mais plusieurs isomères sont possibles : 1-Méthylhistidine, 3-Méthylhistidine, alpha-Méthylhistidine ou Histidine méthyl ester.	NA	NA	Oui <sup>a,b</sup> pour la 1 et la 3.	169.0851	0.90	5.45
Méthylhistidine	C7H11N3O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en C <sub>18</sub> est compatible avec la méthylhistidine, mais plusieurs isomères sont possibles : 1-Méthylhistidine, 3-Méthylhistidine, alpha-Méthylhistidine ou Histidine méthyl ester.	NA	NA	Oui <sup>a,b</sup> pour la 1 et la 3.	169.0851	0.90	6.10

Méthylhistidine	C7H11N3O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en C1 <sub>8</sub> est compatible avec la méthylhistidine, mais plusieurs isomères sont possibles : 1-Méthylhistidine, 3-Méthylhistidine, alpha-Méthylhistidine ou Histidine méthyl ester.	NA	NA	Oui <sup>a,b</sup> pour la 1 et la 3.	169.0851	0.90	6.80
Dihydroxyphényl-éthylène glycole isomère	C8H10O4	Caractérisé <sup>a</sup>	Oui	La MS <sup>n</sup> montre une perte de H <sub>2</sub> O, 2(H <sub>2</sub> O), CH <sub>2</sub> O et C <sub>2</sub> H <sub>2</sub> O <sub>2</sub> .	NA	NA	NA	170.0579	5.32	12.34
Glycérol Phosphate	C3H9O6P	Caractérisé <sup>a</sup>	Oui	Les tr sont cohérents avec la structure. Isomères possibles : glycerol-2-phosphate ou glycérol-3-phosphate.	Glycerophosphates	Métabolisme des glycérolipides / Métabolisme des glycérophospholipides	Non <sup>a</sup>	172.0137	0.93	3.60
Prolylglycine	C7H12N2O3	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Dipeptides	Produit de dégradation protéolytique des protéines.	Oui <sup>a,b</sup>	172.0848	1.07	11.34
Glycylproline	C7H12N2O3	Putativement Identifié <sup>a,b</sup>	Non, plusieurs isomères pour un même pic chromatographique	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH, C <sub>2</sub> H <sub>3</sub> NO (+HCOOH)	Dipeptides	Produit de dégradation protéolytique des protéines. (produit final de dégradation du collagène)	Oui <sup>a,b</sup>	172.0848	1.07	12.24
N-Acétyleucine	C8H15NO3	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes et leurs métabolites	Médicament.	Non <sup>a</sup>	173.1051	6.15	13.14
Isovalérylalanine isomère	C8H15NO3	Caractérisé <sup>a</sup>	Oui	La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , C <sub>2</sub> H <sub>2</sub> O, C <sub>5</sub> H <sub>8</sub> O, C <sub>3</sub> H <sub>4</sub> O et H <sub>2</sub> O.	NA	NA	NA	173.1051	6.34	13.44
Hexanoylglycine	C8H15NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , H <sub>2</sub> O et C <sub>6</sub> H <sub>10</sub> O. Ce schéma de fragmentation peut correspondre à l'Hexanoylglycine.	Acyle Glycines / Dérivés Glycine aliphatiques	Métabolisme des acides gras	Oui <sup>a,b</sup>	173.1051	6.45	13.68
N-Acétylisoleucine	C8H15NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	NA	Oui <sup>c</sup> [386]	173.1051	6.02	13.30
Sulfate de Phényle	C6H6O4S	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de SO <sub>3</sub> , SO <sub>2</sub> et C <sub>6</sub> H <sub>5</sub> O. Ce schéma de fragmentation peut	Acides organiques	NA	Oui <sup>c</sup> [387]	173.9987	5.33	11.93

				correspondre au Phényl sulfate et à l'acide phénolsulfonique. Le tr de l'acide p-Phénolsulfonique ne correspond pas.						
Acide Shikimique isomère	C7H10O5	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide shikimique.	NA	NA	NA	174.0528	0.89	3.52
Acide Subérique	C8H14O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides carboxyliques à longue chaîne	Métabolisme des acides gras	Oui <sup>a,b</sup>	174.0892	6.42	13.83
Arginine	C6H14N4O2	Annoté <sup>a,c</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acides aminés et dérivés	Métabolisme des acides aminés / Cycle de l'urée	Oui <sup>a,b</sup>	174.1117	0.86	6.27
Arginine isomère	C6H14N4O2	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Les tr ne correspondent pas à l'Arginine.	NA	NA	NA	174.1117	0.86	6.77
Acide N-Acétylaspartique	C6H9NO5	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme de l'alanine, de l'aspartate et du glutamate	Oui <sup>a,b</sup>	175.0481	1.05	5.50/6.84
Acide Indoleacétique	C10H9NO2	Annoté <sup>a,c</sup>	Oui	NA	Indoles et dérivés	Métabolisme du tryptophane	Oui <sup>a,b</sup>	175.0633	7.34	15.85
Acide Indoleacétique isomère	C10H9NO2	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide Indoleacétique. La MS <sup>n</sup> montre une perte de C <sub>5</sub> H <sub>5</sub> ON.	NA	NA	NA	175.0633	4.78	13.81
Citrulline isomère	C6H13N3O3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à Citrulline.	NA	NA	NA	175.0957	1.07	9.21
Citrulline isomère	C6H13N3O3	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	Le tr en PFPP ne correspond pas à Citrulline.	NA	NA	NA	175.0957	1.07	11.30
Acide Ascorbique	C6H8O6	Identifié <sup>a,c,d</sup>	Oui	NA	Acides organiques	Métabolisme de l'ascorbate et l'aldarate / Métabolisme du glutathion	Oui <sup>a,b</sup>	176.0321	1.02	5.20

Acide 2-isopropylmalique	C7H12O5	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques	Métabolisme des acides aminés / Métabolisme du pyruvate	Oui <sup>a,b</sup>	176.0685	5.52	12.61
Cotinine	C10H12N2O	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la nicotine	Oui <sup>a,b</sup>	176.095	1.83	14.31
Sérotonine	C10H12N2O	Annoté <sup>a,c</sup>	Oui	NA	Indoles et dérivés	Métabolisme du tryptophane / Neurotransmetteur	Oui <sup>a,b</sup>	176.095	4.17	NR
N-Formylméthionine	C6H11NO3S	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	177.0460	5.11	12.39
N-Formylméthionine isomère	C6H11NO3S	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	177.0460	4.64	12.09
Acide Hippurique	C9H9NO3	Identifié <sup>a,c,d</sup>	Oui	NA	Acyle Glycines / dérivés glycine hétérocycliques	Métabolisme des acides aminés	Oui <sup>a,b</sup>	179.0582	5.82	13.50
5-Méthylthioribose	C6H12O4S	Caractérisé <sup>a</sup>	Oui	Le spectre de masse montre la présence de l'isotope <sup>34</sup> S.	Sucres et dérivés	Métabolisme des acides aminés	Non <sup>a</sup>	180.0456	1.09	5.12
Acide Nicotinurique isomère	C8H8N2O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Nicotinurique. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH (or CO <sub>2</sub> ), C <sub>2</sub> H <sub>3</sub> NO et C <sub>2</sub> H <sub>5</sub> NO.	NA	NA	NA	180.0535	5.31	12.96
Hexose	C6H12O6	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les hexoses (Glucose, Fructose, Galactose, Mannose...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés	Glycolyse / Glyconéogénèse / Voie des pentoses phosphates / Métabolisme du galactose	Oui <sup>a,b</sup>	180.0634	0.90	3.67
Théobromine	C7H8N4O2	Identifié <sup>a,b,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la théobromine	Oui <sup>a,b</sup>	180.0647	4.74	11.40
Paraxanthine / Théophylline	C7H8N4O2	Identifié <sup>a,b,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique (théophylline)	Les temps de rétention peuvent correspondre à la Paraxanthine et la Theophylline (mai s plus proches de la paraxanthine)	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine	Oui <sup>a,b</sup>	180.0647	5.08	12.16

Tyrosine	C9H11NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	181.0739	1.05	12.07
Tyrosine isomère	C9H11NO3	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à la Tyrosine ou l'Ortho-Tyrosine. La MS <sup>n</sup> montre des pertes de NH <sub>3</sub> , H <sub>2</sub> O et HCOOH. Isomères possibles : Beta ou Meta-Tyrosine.	Acides aminés et dérivés	NA	NA	181.0739	2.25	12.66
L-Thréo-3-Phénylsérine	C9H11NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> .	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	181.0739	6.02	13.50
Acide 1-Méthylurique	C6H6N4O3	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine	Oui <sup>a,b</sup>	182.044	4.10	11.15
Acide Isohomovanillique	C9H10O4	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines / Cathécholamines vanilliques	Métabolisme de la tyrosine	Non <sup>a</sup>	182.0579	6.23	13.97
Acide Hydroxyphényllactique	C9H10O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides Phénoliques	Métabolisme de la tyrosine	Oui <sup>a,b</sup>	182.0579	5.25	12.66
Acide Homovanillique	C9H10O4	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines / Cathécholamines vanilliques	Métabolisme de la tyrosine	Oui <sup>a,b</sup>	182.0579	6.10	13.35
Mannitol (+ isomères)	C6H14O6	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les polyols en C6 (Mannitol, sorbitol, Galactitol...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Alcools et polyols / alcools simples	Métabolisme du fructose et du mannose / Métabolisme du galactose	Oui <sup>a,b</sup>	182.0790	0.95	3.50
Acide 4-Pyridoxique	C8H9NO4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Pyridoxals et dérivés	Métabolisme de la vitamine B6	Oui <sup>a,b</sup>	183.0532	3.22	11.65
Acide 4-Pyridoxique isomère	C8H9NO4	Caractérisé <sup>a</sup>	Yes	Les tr ne correspondent pas à l'acide 4-Pyridoxique ou à la 3,5-Dihydroxyphénylglycine. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N.	NA	NA	NA	183.0532	5.63	13.20

Acide Tétrahydro-hippurique	C9H13NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH (ou CO <sub>2</sub> ) et C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N.	Acyle Glycines / dérivés glycine hétérocycliques	NA	Non	183.0895	6.33	13.80
N1-Acétyspermidine	C9H21N3O	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, NH <sub>3</sub> , C <sub>4</sub> H <sub>9</sub> N (+NH <sub>3</sub> ) et C <sub>5</sub> H <sub>12</sub> N <sub>2</sub> O. Les tr ne correspondent pas à la N8-Acétyspermidine.	Amines biogènes et Polyamines	NA	Oui <sup>a,b</sup>	187.1685	0.88	43.70
Acide Azélaïque	C9H16O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides carboxyliques à longue chaîne	NA	Oui <sup>a,b</sup>	188.1049	7.08	14.65
Acide Kynurénique	C10H7NO3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques	Métabolisme du tryptophane	Oui <sup>a,b</sup>	189.0426	5.45	14.56
Sulfate de Pyrocatechole (+ isomères)	C6H6O5S	Putativement Identifié <sup>a,b</sup>	Oui	La MS montre la présence d'un isotope <sup>34</sup> S. La MS <sup>n</sup> montre une perte de SO <sub>3</sub> . Isomères possibles : Sulfate de Pyrocatechol, Sulfate de Quinol ou Sulfocatechol.	Alcools et polyols	NA	Non	189.9936	4.88	11.47
Acide Hydroxysubérique isomère	C8H14O5	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	190.0841	4.90	11.72
Acide Hydroxysubérique isomère	C8H14O5	Caractérisé <sup>a</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O et CO <sub>2</sub> +H <sub>2</sub> O.	NA	NA	NA	190.0841	5.15	12.31
Acide Hydroxysubérique isomère	C8H14O5	Caractérisé <sup>a</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>7</sub> H <sub>7</sub> O.	NA	NA	NA	190.0841	6.30	13.65
Acide 6-(2-hydroxyéthoxy)-6-oxohexanoïque	C8H14O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CO <sub>2</sub> +H <sub>2</sub> O et C <sub>2</sub> H <sub>4</sub> O.	Acides organiques	NA	Non	190.0841	5.35	11.99
Acide 6-(carboxyméthoxy)-hexanoïque	C8H14O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CO <sub>2</sub> +H <sub>2</sub> O, C <sub>2</sub> H <sub>4</sub> O <sub>2</sub> , C <sub>2</sub> H <sub>4</sub> O <sub>3</sub> et C <sub>3</sub> H <sub>4</sub> O <sub>4</sub> .	Acides organiques	NA	Non	190.0841	5.65	12.88
Acide 5-Hydroxyindole-acétique	C10H9NO3	Identifié <sup>a,b,c,d</sup>	Oui	NA	Indoles et dérivés	Métabolisme du tryptophane	Oui <sup>a,b</sup>	191.0582	5.69	13.76

Acide Citrique	C6H8O7	Identifié <sup>a,b,c,d</sup>	Oui	NA	Acides organiques	Cycle de Krebs / Métabolisme du glyoxylate et des dicarboxylates	Oui <sup>a,b</sup>	192.0270	1.07/1.49	5.52
Acide Quinique	C7H12O6	Annoté <sup>a,c</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Acide organiques retrouvés dans les grains de café, les feuilles de tabac, les feuilles de carottes, les pommes, pêches, poires, prunes, et autres fruits et légumes.	Non <sup>a</sup>	192.0634	0.95	3.85
Hydroxycotinine isomère	C10H12N2O2	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>2</sub> H <sub>5</sub> NO. Autres isomères possibles : Hydroxycotinine, Oxoamide, Oxyde de Cotinine (...)	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la nicotine	Oui <sup>a,b</sup>	192.0899	1.11	12.11
Acide 2-Méthylhippurique	C10H11NO3	Annoté <sup>a,c</sup>	Oui	NA	Acyle Glycines / Conjugués Heterocyclique de glycine	Métabolisme des acides gras	Oui <sup>a,b</sup>	193.0739	6.07	13.90
Acide 4-Méthylhippurique / Acide 3-Méthylhippurique	C10H11NO3	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Les tr correspondent à la fois à l'Acide 4-Méthylhippurique et 3-Méthylhippurique.	Acyle Glycines / Conjugués Heterocyclique de glycine	Métabolites du xylène	Oui <sup>a,b</sup>	193.0739	6.70	14.50
Acide Méthylhippurique isomère	C10H11NO3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide 2-Méthylhippurique, 4-Méthylhippurique ou 3-Méthylhippurique. Autres isomères possibles : Phénylacétylglycine ou Hippurate de méthyle.	NA	NA	NA	193.0739	6.20	14.19
Acide Glucuronique (+ isomères)	C6H10O7	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Les tr correspondent à la fois à l'Acide Galacturonique et Glucuronique.	Sucres et dérivés	Interconversion des pentoses et du gluconate / Métabolisme de l'ascorbate et l'aldarate	Oui <sup>a,b</sup>	194.0427	0.96	3.50
Acide Isoféruleque	C10H10O4	Identifié <sup>a,b,c,d</sup>	Oui	Le tr en C <sub>18</sub> correspond à la fois à l'Acide Ferulique et Isoferulique, mais le tr	Acides organiques / Acides Phénoliques	Métabolisme de la caféine	Oui <sup>a,b</sup>	194.0578	6.65	14.79

				en PFPP correspond à l'Acide Isoféruilique.						
Dihydroxy-cinnamoate de méthyle	C10H10O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acide Féruilique ou Isoféruilique. La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , CH <sub>3</sub> et CO <sub>2</sub> +CH <sub>3</sub> .	Acides organiques / Esters	NA	Non	194.0578	5.70	12.63
Malonate de benzyle	C10H10O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acide Féruilique ou Isoféruilique. La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , CO <sub>2</sub> +CO et C <sub>3</sub> H <sub>2</sub> O <sub>3</sub> .	Acides organiques / Esters	NA	Non	194.0578	6.95	15.69
Phthalate d'éthyle	C10H10O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acide Féruilique ou Isoféruilique. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CO <sub>2</sub> , C <sub>2</sub> H <sub>4</sub> , C <sub>2</sub> H <sub>5</sub> O et C <sub>3</sub> H <sub>3</sub> O <sub>2</sub> (+CO <sub>2</sub> )	Composés exogènes, xénobiotiques et leurs métabolites	Excipient de cosmétiques et médicaments.	Oui <sup>a,b</sup>	194.0578	7.45	16.06
Pyridylacétylglycine	C9H10N2O3	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'acide aminohippurique. Autres isomères possibles : 2-Pyridylacétylglycine, 3-Pyridylacétylglycine.	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la nicotine	Non	194.0691	4.74	13.75
Caféine	C8H10N4O2	Identifié <sup>a,b,c,d</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine	Oui <sup>a,b</sup>	194.0804	5.56	12.84
Acide α-hydroxyhippurique	C9H9NO4	Identifié <sup>a,c,d</sup>	Oui	NA	Acides aminés et dérivés	utilisé dans la formulation de certains herbicides.	Oui <sup>a,b</sup>	195.0532	5.10	12.55
Acide Hydroxyhippurique isomère	C9H9NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Alpha-Hydroxyhippurique. La MS <sup>n</sup> montre une perte de C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> . Autres isomères possibles : Acides Salicylurique, N-acetyl-4-aminosalicylique, N-Acetyl-5-aminosalicylique ou la 4-Carboxyphénylglycine.	NA	NA	NA	195.0532	4.85	12.24

Acide Hydroxyhippurique isomère	C9H9NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Alpha-Hydroxyhippurique. La MS <sup>n</sup> montre une perte de C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> . Autres isomères possibles : Acides Salicylurique, N-acetyl-4-aminosalicylique, N-Acetyl-5-aminosalicylique ou la 4-Carboxyphénylglycine.	NA	NA	NA	195.0532	6.00	13.95
Acide Hydroxyhippurique isomère	C9H9NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'acide Alpha-Hydroxyhippurique. La MS <sup>n</sup> montre une perte de C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> . Autres isomères possibles : Acides Salicylurique, N-acetyl-4-aminosalicylique, N-Acetyl-5-aminosalicylique ou la 4-Carboxyphénylglycine.	NA	NA	NA	195.0532	6.39	14.78
Acide Gluconique (+ isomères)	C6H12O7	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les Acides de sucres en C6 (Gluconic acid, Galactonic acid...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Acides organiques	Voie des pentoses phosphates / Interconversion des pentoses et du gluconate / Métabolisme de l'ascorbate et l'aldarate	Oui <sup>a,b</sup>	196.0583	0.95	3.51
Acide 1,3-Diméthylurique	C7H8N4O3	Identifié <sup>a,c,d</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la théophylline	Oui <sup>a,b</sup>	196.0596	4.67	11.55
Acide 1,7-Diméthylurique	C7H8N4O3	Identifié <sup>a,c,d</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolisme de la caféine	Oui <sup>a,b</sup>	196.0596	4.96	11.84
Acide 3,7-Diméthylurique	C7H8N4O3	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme de la caféine, théophylline et de la théobromine	Oui <sup>a,b</sup>	196.0596	4.25	11.20
Acide Diméthylurique isomère	C7H8N4O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas aux acides (1,7) (1,3) ou (3,7)	NA	NA	NA	196.0596	4.81	12.23

				diméthylurique. Isomères possibles (1,9) (3,9) ou (7,9).						
Acide Diméthylurique isomère	C7H8N4O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas aux acides (1,7) (1,3) ou (3,7) diméthylurique. Isomères possibles (1,9) (3,9) ou (7,9).	NA	NA	NA	196.0596	5.09	12.55
Acide Diméthylurique isomère	C7H8N4O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas aux acides (1,7) (1,3) ou (3,7) diméthylurique. Isomères possibles (1,9) (3,9) ou (7,9).	NA	NA	NA	196.0596	6.01	13.95
Acide Sébacique	C10H18O4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides carboxyliques à longue chaîne	Métabolisme des acides gras	Oui <sup>a,b</sup>	202.1205	7.65	15.38
Acide Heptylmalonique	C10H18O4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CH <sub>2</sub> O <sub>2</sub> , CO <sub>2</sub> +2(H <sub>2</sub> O).	Acides organiques / Acides carboxyliques à longue chaîne	NA	Non	202.1205	6.05	13.15
Acide 2-éthylsubérique	C10H18O4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CO <sub>2</sub> +H <sub>2</sub> O, CO <sub>2</sub> +2(H <sub>2</sub> O), C <sub>6</sub> H <sub>10</sub> O <sub>2</sub> , C <sub>4</sub> H <sub>8</sub> O <sub>2</sub> .	Acides organiques / Acides carboxyliques à longue chaîne	NA	Non	202.1205	6.59	14.09
Spermine	C10H26N4	Annoté <sup>a,c</sup>	Oui	NA	Amines biogènes et Polyamines	Métabolisme des acides aminés / Métabolisme du glutathion	Oui <sup>a,b</sup>	202.2157	0.81	NR
Acide N2-Acétylamino-adipique	C8H13NO5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, CO <sub>2</sub> et C <sub>2</sub> H <sub>2</sub> O.	Acides aminés et dérivés	Métabolisme de la lysine	Non	203.0794	4.05	11.23
Acide (éthoxycarbonyl)-hydroxypyrrolidine-carboxylique / Acide (méthoxyacetyl)-hydroxypyrrolidine-carboxylique	C8H13NO5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , C <sub>2</sub> H <sub>6</sub> O et CO <sub>2</sub> +H <sub>2</sub> O.	Acides organiques	NA	Non	203.0794	4.55	11.72
Acide (éthoxycarbonyl)-hydroxypyrrolidine-carboxylique / Acide	C8H13NO5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , C <sub>2</sub> H <sub>6</sub> O et CO <sub>2</sub> +H <sub>2</sub> O.	Acides organiques	NA	Non	203.0794	5.10	12.30

(méthoxyacetyl)-hydroxypyrrolidine-carboxylique										
Acétylcarnitine	C9H17NO4	Identifié <sup>a,b,c,d</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	Acyles carnitines / Acyles carnitines à courte chaîne	Métabolisme des acides gras	Oui <sup>a,b</sup>	203.1158	1.1/1.50	15.95
Acétylcarnitine isomère	C9H17NO4	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	NA	NA	NA	203.1158	1.1/1.50	19.77
Tryptophane isomère	C11H12N2O2	Caractérisé <sup>a</sup>	Oui	Le tr en C <sub>18</sub> correspond au Tryptophane, mais pas le tr en PFPP.	NA	NA	NA	204.0899	5.28	30.47
Acide Xanthurénique	C10H7NO4	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides hétérocycliques	Métabolisme du tryptophane	Oui <sup>a,b</sup>	205.0375	5.18	14.99
Cinnamoylglycine	C11H11NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> .	Acyle Glycines	NA	Oui <sup>a,b</sup>	205.0739	7.03	15.05
N-Acétylphénylalanine	C11H13NO3	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme des acides aminés	Oui <sup>a,b</sup>	207.0895	6.50	14.48
Acétate de 2-acétamido-4-méthylphényle	C11H13NO3	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>8</sub> H <sub>6</sub> O, C <sub>2</sub> H <sub>2</sub> O et C <sub>2</sub> H <sub>5</sub> NO. Ce schéma de fragmentation peut correspondre à l'acétate de 2-acétamido-4-méthylphényle ou au 2-(4-acétylphénoxy)-N-méthylacetamide et méthyl 3-acétamido-4-méthylbenzoate.	Acides organiques / Esters	NA	Non	207.0895	6.72	14.75
Kynurénine isomère	C10H12N2O3	Caractérisé <sup>a</sup>	Oui	Le tr en C <sub>18</sub> correspond avec la kynurénine, mais pas le tr en PFPP. Autre isomère possible : Formyl-5-hydroxykynurénamine.	NA	NA	NA	208.0848	4.41	17.40
Acide Aminosalicylurique	C9H10N2O4	Caractérisé <sup>a</sup>	Oui	La MS <sup>n</sup> montre une perte de HCOOH.	NA	NA	Non	210.0641	4.48	11.40

Déoxyribose phosphate (+ isomères)	C5H11O7P	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les deoxypentoses phosphate ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés / Monosaccharide phosphates	Voie des pentoses phosphates / Métabolisme des pyrimidines	Non <sup>a</sup>	214.0242	0.96	3.67
2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide (+ isomères)	C10H18N2O3	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas la Dethiobiotine ou Val-Pro. La MS <sup>n</sup> montre une perte de C <sub>6</sub> H <sub>9</sub> NO <sub>3</sub> . Ce schéma de fragmentation peut correspondre à l'Acide 3-{{[2-(pyrrolidin-1-yl)éthyl]carbamoyl} propanoïque ou 2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide.	Amines biogènes et Polyamines	NA	Non	214.1317	1.05	14.22
2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide (+ isomères)	C10H18N2O3	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas la Dethiobiotine ou Val-Pro. La MS <sup>n</sup> montre une perte de C <sub>6</sub> H <sub>9</sub> NO <sub>3</sub> . Ce schéma de fragmentation peut correspondre à l'Acide 3-{{[2-(pyrrolidin-1-yl)éthyl]carbamoyl} propanoïque ou 2-méthoxy-N-[3-oxo-3-(pyrrolidin-1-yl)propyl]acétamide.	Amines biogènes et Polyamines	NA	Non	214.1317	1.72	16.66
Propionylcarnitine isomère	C10H19NO4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	217.1314	3.19	20.57
Propionylcarnitine isomère	C10H19NO4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	217.1314	4.65	34.26
Acide 5-Sulfosalicylique	C7H6O6S	Annoté <sup>a,c</sup>	Oui	La MS montre la présence d'un isotope <sup>34</sup> S et d'une perte de SO <sub>3</sub> . Les tr correspondent également au standard.	Composés exogènes, xénobiotiques et leurs métabolites	dérivé de l'acide salicylique.	Non <sup>a</sup>	217.9885	5.20	11.55
Acide Sulfosalicylique isomère	C7H6O6S	Caractérisé <sup>a</sup>	Oui	La MS montre la présence d'un isotope <sup>34</sup> S et d'une perte de SO <sub>3</sub> .	NA	NA	NA	217.9885	4.80	11.80

N-Acétylsérotonine isomère	C12H14N2O2	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	218.1055	5.74	13.15
N-Acétylsérotonine isomère	C12H14N2O2	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	218.1055	5.93	13.72
Pantothénic acid	C9H17NO5	Identifié <sup>a,c,d</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme de la beta-alanine	Oui <sup>a,b</sup>	219.1107	4.74	11.47
Hydroxytryptophane isomère	C11H12N2O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au 5-Hydroxytryptophane.	NA	NA	NA	220.0848	3.07	12.54
Hydroxytryptophane isomère	C11H12N2O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au 5-Hydroxytryptophane.	NA	NA	NA	220.0848	4.54	14.23
Hydroxytryptophane isomère	C11H12N2O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas au 5-Hydroxytryptophane.	NA	NA	NA	220.0848	5.67	15.53
Acétyltyrosine isomère	C11H13NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acétyltyrosine.	NA	NA	NA	223.0845	4.40	14.48
Porphobilinogène	C10H14N2O4	Annoté <sup>a,c</sup>	Oui	NA	Acides aminés et dérivés	Métabolisme de la porphyrine et de la chorophylle	Oui <sup>a,b</sup>	226.0954	4.71	11.51
Butyrylcarnitine / Isobutyryl-L-carnitine	C11H21NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>17</sub> NO <sub>2</sub> .	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	231.1471	4.95	16.80
Bioptérine isomère	C9H11N5O3	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à la 6-Bioptérine ou la Sépiaptérine. Autres isomères possibles : Orinaptérine, Primaptérine et Dysproptérine.	NA	NA	NA	237.0862	5.15	NR
Thymidine	C10H14N2O5	Identifié <sup>a,c,d</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des pyrimidines	Oui <sup>a,b</sup>	242.0903	1.04	11.19
Tiglylcarnitine / 2-Ethylacrylylcarnitine	C12H21NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>8</sub> H <sub>17</sub> NO <sub>2</sub> .	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	243.1471	5.53	32.20
Uridine isomère	C9H12N2O6	Caractérisé <sup>a</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic	Les doubles tr C <sub>18</sub> correspondent à l'uridine mais pas ceux en PFPP. Isomère	NA	NA	NA	244.0695	1.04/1.73	5.42/ 5.90

			chromatographique	possible : Pseudouridine.						
Valérylcarnitine	C12H23NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>8</sub> H <sub>19</sub> NO <sub>2</sub> . Ce schéma de fragmentation peut correspondre à la 2-Méthylbutyroylcarnitine, Valérylcarnitine et Isovalérylcarnitine. La distinction entre les différents isomères est basée sur la valeur de LogP donnée par HMDB.	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	245.1627	5.74	13.40
Isovalérylcarnitine	C12H23NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>8</sub> H <sub>19</sub> NO <sub>2</sub> . Ce schéma de fragmentation peut correspondre à la 2-Méthylbutyroylcarnitine, Valérylcarnitine et Isovalérylcarnitine. La distinction entre les différents isomères est basée sur la valeur de LogP donnée par HMDB.	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	245.1627	5.63	13.90
2-Méthylbutyroylcarnitine	C12H23NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>8</sub> H <sub>19</sub> NO <sub>2</sub> . Ce schéma de fragmentation peut correspondre à la 2-Méthylbutyroylcarnitine, Valérylcarnitine et Isovalérylcarnitine. La distinction entre les différents isomères est basée sur la valeur de LogP donnée par HMDB.	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	245.1627	5.55	13.99
N-Acétyltryptophane isomère	C13H14N2O3	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	246.1004	7.00	15.01
N-Acétyltryptophane isomère	C13H14N2O4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	246.1004	5.74	13.07

N-Acétyltryptophane isomère	C13H14N2O4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	246.1004	6.82	14.65
Asp-Ile	C10H18N2O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, NH <sub>3</sub> , HCOOH et C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N. La distinction a été faite sur la base du tr en C <sub>18</sub> et de l'hydrophobicité de la leucine par rapport à l'isoleucine.	Dipeptides	Produit de dégradation protéolytique des protéines.	Oui <sup>a,b</sup>	246.1216	5.00	12.30
Asp-Leu	C10H18N2O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, NH <sub>3</sub> , HCOOH et C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N. La distinction a été faite sur la base du tr en C <sub>18</sub> et de l'hydrophobicité de la leucine par rapport à l'isoleucine.	Dipeptides	Produit de dégradation protéolytique des protéines.	Non	246.1216	5.19	12.81
Glu-Val	C10H18N2O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, NH <sub>3</sub> , HCOOH et C <sub>3</sub> H <sub>7</sub> O <sub>3</sub> N.	Dipeptides	Produit de dégradation protéolytique des protéines.	Oui <sup>a,b</sup>	246.1216	4.41	14.58
Phe-Ser / Ser-Phe	C12H16N2O4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de CO <sub>2</sub> , 2(CO <sub>2</sub> ), CH <sub>3</sub> NO <sub>2</sub> et C <sub>2</sub> H <sub>3</sub> O <sub>4</sub> N. Les tr ne correspondent pas à Ala-Tyr ou Tyr-Ala.	Dipeptides	Produit de dégradation protéolytique des protéines.	Non	252.111	5.01	15.16
Dihydroneoptérine isomère	C9H13N5O4	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	255.0968	4.86	12.05
dérivé de carnitine	C12H17NO5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de NC <sub>3</sub> H <sub>9</sub> et C <sub>3</sub> H <sub>9</sub> N+C <sub>3</sub> H <sub>4</sub> O <sub>3</sub> .	Acyles carnitines	NA	Oui <sup>c</sup> [261]	255.1101	4.62	14.37
Hexanoylcarnitine	C13H25NO4	Annoté <sup>a,c</sup>	Oui	NA	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	259.1783	6.58	NR
Hexanoylcarnitine isomère	C13H25NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas l'Hexanoylcarnitine.	NA	NA	Oui <sup>c</sup> [261]	259.1783	6.23	NR
Hexanoylcarnitine isomère	C13H25NO4	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas l'Hexanoylcarnitine.	NA	NA	Oui <sup>c</sup> [261]	259.1783	6.35	NR

(δ)Glu-Leu/Ile	C11H20N2O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, 2(H <sub>2</sub> O), HCOOH et C <sub>5</sub> H <sub>6</sub> NO <sub>3</sub> .	Dipeptides	Produit de dégradation protéolytique des protéines.	Oui <sup>a,b</sup>	260.1372	5.20	14.23/15.07
Phénylacétyl-glutamine	C13H16N2O4	Putativement Identifié <sup>a,b</sup>	Oui	Les tr ne correspondent pas à l'Acetyl-N-formyl-5-methoxykynurenamine. La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, NH <sub>3</sub> , C <sub>5</sub> H <sub>7</sub> O <sub>3</sub> N et C <sub>8</sub> H <sub>9</sub> NO.	Acides aminés et dérivés	Métabolisme du Phénylacétate / Synthèse des acides aminés et des protéines	Oui <sup>a,b</sup>	264.111	5.84	13.06
Adénosine	C10H13N5O4	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	267.0967	1.11	NR
N-Acétylecarnosine isomère	C11H16N4O4	Caractérisé <sup>a</sup>	Non, plusieurs isomères pour un même pic chromatographique	NA	NA	NA	NA	268.1166	1.10	10.85
N-Acétylecarnosine	C11H16N4O4	Putativement Identifié <sup>a,b</sup>	Non, plusieurs isomères pour un même pic chromatographique	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O, HCOOH et C <sub>5</sub> H <sub>7</sub> O <sub>2</sub> N (+HCOOH).	Amines biogènes et Polyamines	Métabolisme de l'histidine / Métabolisme de la beta-alanine	Non <sup>a</sup>	268.1166	1.10	11.55
Hepténoylcarnitine	C14H25NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>3</sub> H <sub>9</sub> N.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	271.1784	6.52	18.29
Méthylinosine	C11H14N4O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> . Autres isomères possibles : N1/N7/O6 Méthylinosine	Nucléosides et dérivés	NA	Oui <sup>a,b</sup> pour la 1 pas la 7	282.0964	4.45	11.24
Méthylinosine	C11H14N4O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> . Autres isomères possibles : N1/N7/O6 Méthylinosine	Nucléosides et dérivés	NA	Oui <sup>a,b</sup> pour la 1 pas la 7	282.0964	4.26	11.47
Xanthosine	C10H12N4O6	Identifié <sup>a,b,c,d</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines / Métabolisme de la caféine	Oui <sup>a,b</sup>	284.0757	4.28	11.09
N4-Acétylecytidine	C11H15N3O6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> et C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> +C <sub>2</sub> H <sub>2</sub> O.	Nucléosides et dérivés	Produit de dégradation des ARN de transfert (ARNt).	Oui <sup>a,c</sup> [284]	285.0961	4.51	11.38
Octénoylcarnitine	C15H27NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>4</sub> H <sub>6</sub> O <sub>3</sub> et C <sub>9</sub> H <sub>23</sub> NO <sub>2</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	285.194	7.31	NR

Octanoylcarnitine	C15H29NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>15</sub> NO <sub>2</sub> et C <sub>9</sub> H <sub>25</sub> NO <sub>2</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	287.2097	7.50	NR
Octanoylcarnitine	C15H29NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>15</sub> NO <sub>2</sub> et C <sub>9</sub> H <sub>25</sub> NO <sub>2</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	287.2097	7.94	NR
Adipoylcarnitine	C13H23NO6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>13</sub> O <sub>2</sub> N (+H <sub>2</sub> O), C <sub>9</sub> H <sub>19</sub> NO <sub>3</sub> .	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	289.1525	5.30	21.06
Méthylglutaryl-carnitine	C13H23NO6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>13</sub> O <sub>2</sub> N (+H <sub>2</sub> O, +2H <sub>2</sub> O, +CO), C <sub>9</sub> H <sub>19</sub> NO <sub>3</sub> .	Acyles carnitines / Acyles carnitines à courte chaîne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	289.1525	4.70	31.48
Méthylglutaryl-carnitine	C13H23NO6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>6</sub> H <sub>14</sub> O <sub>4</sub> , C <sub>7</sub> H <sub>13</sub> O <sub>2</sub> N (+H <sub>2</sub> O, +2H <sub>2</sub> O, +CO), C <sub>9</sub> H <sub>19</sub> NO <sub>3</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	289.1525	4.83	33.11
N2-Méthylguanosine	C11H15N5O5	Identifié <sup>a,c,d</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	297.1073	4.30	11.61
Nonénoylcarnitine	C16H29NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N, C <sub>8</sub> H <sub>15</sub> NO <sub>4</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	299.2097	7.60	NR
Nonénoylcarnitine	C16H29NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N, C <sub>8</sub> H <sub>15</sub> NO <sub>4</sub> .	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	299.2097	7.90	NR
2,6-diméthylhéptanoyl carnitine	C16H31NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>9</sub> H <sub>17</sub> O (+H <sub>2</sub> O, +CO). Ce schéma de fragmentation peut correspondre à la 2,6 diméthylhéptanoyl carnitine et la Nonanoylcarnitine. La distinction entre les différents isomères est basée sur la valeur de LogP donnée par HMDB..	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	301.2253	8.06	NR

Nonanoylcarnitine	C16H31NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N, C <sub>9</sub> H <sub>17</sub> O (+H <sub>2</sub> O, +CO). Ce schéma de fragmentation peut correspondre à la 2,6 diméthylheptanoyl carnitine et la Nonanoylcarnitine. La distinction entre les différents isomères est basée sur la valeur de LogP donnée par HMDB..	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	301.2253	8.18	NR
Indoleacétyl glutamine	C15H17N3O4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de NH <sub>3</sub> , H <sub>2</sub> O, NH <sub>3</sub> +HCOOH, C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub> (+CO).	Acides aminés et dérivés	Métabolisme du tryptophane	Oui <sup>a,b</sup>	303.1219	6.05	13.40
Hydroxyoctanoyl-carnitine	C15H29O5N	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N (+H <sub>2</sub> O), C <sub>6</sub> H <sub>10</sub> O <sub>3</sub> et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N (+H <sub>2</sub> O).	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	303.2046	6.60	30.22
dUMP	C9H13N2O8P	Identifié <sup>a,c,d</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des pyrimidines	Non <sup>a</sup>	308.041	1.01	NR
Décadiénylecarnitine	C17H27NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	309.194	7.66	NR
Décadiénylecarnitine	C17H27NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	309.194	7.88	NR
Diméthylguanosine	C12H17N5O5	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> . 2 isomères possibles : N2,N2- et 1,7-Diméthylguanosine.	Nucléosides et dérivés	Métabolisme des purines / Produit de dégradation des ARN de transfert	Oui <sup>a,b</sup> pour la 2,2 pas la 1,7	311.123	4.75	12.10
Décénoylecarnitine	C17H31NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N (+H <sub>2</sub> O).	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	313.2259	8.38	18.29
Décénoylecarnitine	C17H31NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N (+H <sub>2</sub> O).	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	313.2259	8.17	19.12

Décanoylcarnitine	C17H33NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>10</sub> H <sub>10</sub> O.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	315.241	8.74	22.77
Décanoylcarnitine	C17H33NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>9</sub> N et C <sub>10</sub> H <sub>10</sub> O.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>a,c</sup> [261]	315.241	8.98	NR
Octanoyl-glucuronide	C14H24O8	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> (C <sub>8</sub> H <sub>16</sub> O <sub>2</sub> ) <sup>-</sup> , C <sub>8</sub> H <sub>16</sub> O <sub>2</sub> (C <sub>6</sub> H <sub>7</sub> O <sub>6</sub> ) <sup>-</sup> , C <sub>8</sub> H <sub>14</sub> O (C <sub>6</sub> H <sub>7</sub> O <sub>6</sub> +H <sub>2</sub> O) <sup>-</sup> et C <sub>9</sub> H <sub>18</sub> O <sub>5</sub> (C <sub>8</sub> H <sub>16</sub> O <sub>2</sub> +CO <sub>2</sub> +H <sub>2</sub> O) <sup>-</sup>	Glucuronides endogènes	NA	Oui <sup>a,c</sup> [388]	320.1468	8.06	15.60
Glucopyranosyl-moranoline	C12H23NO9	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> .	Composés exogènes, xénobiotiques et leurs métabolites	Médicament	Non	325.1373	4.55	17.33
Glucuronide de Paracétamol	C14H17NO8	Identifié <sup>a,c,d</sup>	Oui	NA	Composés exogènes, xénobiotiques et leurs métabolites	Métabolite de médicaments	Oui <sup>a,b</sup>	327.0954	4.15	11.11
Diméthylnonénoyl-carnitine / Undécénoyl-carnitine	C18H33NO4	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre une perte de C <sub>3</sub> H <sub>9</sub> N et C <sub>7</sub> H <sub>15</sub> O <sub>3</sub> N. 2 isomères possibles: diméthylnonénoylcarnitine et Undécénoylcarnitine.	Acyles carnitines / Acyles carnitines à chaîne moyenne	Catabolisme des lipides / Transport des acides gras / Production d'énergie	Oui <sup>c</sup> [261]	327.241	8.94	18.31
AMP Cyclique	C10H12N5O6P	Annoté <sup>a,c</sup>	Oui	NA	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	329.0525	1.01/3.60	10.86
Déoxycortico-stérone	C21H30O3	Annoté <sup>a,c</sup>	Oui	NA	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Non <sup>a</sup>	330.2195	9.62	NR
Glucuronide d'acide 3-Indole carboxylique	C15H15NO8	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>2</sup> montre une perte de C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> , et La MS <sup>3</sup> montre une perte de CO <sub>2</sub> et H <sub>2</sub> O.	Glucuronides endogènes	Métabolisme du tryptophane	Oui <sup>a,b</sup>	337.0798	6.20	13.49
Sucrose (+ isomères)	C12H22O11	Annoté <sup>a,c</sup>	Oui Mais potentiellement plusieurs isomères pour un même pic chromatographique	Tous les disaccharides (Sucrose, Maltose, Isomaltose, Lactulose...) ont les mêmes tr en C <sub>18</sub> et PFPP.	Sucres et dérivés / Disaccharides	Métabolisme du galactose / Métabolisme du sucrose et de l'amidon	Oui <sup>a,b</sup>	342.1162	0.97	3.84

Acide Chlorogénique	C16H18O9	Annoté <sup>a,c</sup>	Oui	NA	Acides organiques / Acides à longue chaîne	Synthèse de métabolites secondaires	Oui <sup>a,b</sup>	354.0951	5.49	NR
Dehydroépiandrosterone sulfate	C19H28O5S	Annoté <sup>a,c</sup>	Oui	NA	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	368.1657	8.35	NR
Sulfate de Testostérone	C19H28O5S	Caractérisé <sup>a</sup>	Oui	Autre isomère possible : Sulfate d'Épitéstostérone	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	368.1657	8.65	15.75
Cortol / beta-Cortol	C21H36O5	Caractérisé <sup>a</sup>	Oui	NA	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	368.2563	6.85	NR
Riboflavine (Vitamine B2)	C17H20N4O6	Annoté <sup>a,c</sup>	Oui	NA	Alcools et polyols / Ptérides et Dihydroptérides	Métabolisme de la Riboflavine	Oui <sup>a,b</sup>	376.1383	5.75	12.47
Succinyladénosine	C14H17N5O8	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> et C <sub>4</sub> H <sub>4</sub> O <sub>4</sub> .	Nucléosides et dérivés	Métabolisme des purines	Oui <sup>a,b</sup>	383.1077	4.77	11.62
Acide Cholique	C24H40O5	Annoté <sup>a,c</sup>	Oui	NA	Acides biliaires	Synthèse d'acides biliaires / Sécrétion biliaire	Oui <sup>a,b</sup>	408.2876	9.18	16.61
Phe-Tyr-Asp (+ isomères)	C22H25N3O7	Caractérisé <sup>a</sup>	Oui	NA	Tripeptides	Produit de dégradation protéolytique des protéines.	NA	443.1693	5.83	13.40
Phe-Tyr-Asp (+ isomères)	C22H25N3O7	Caractérisé <sup>a</sup>	Oui	NA	Tripeptides	Produit de dégradation protéolytique des protéines.	NA	443.1693	5.83	13.55
Glucuronide d'acide 3-(carboxyméthyl)-5-hydroxy-adamantylacétique (+ isomères)	C20H28O11	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> (+CO <sub>2</sub> , +2CO <sub>2</sub> ), C <sub>14</sub> H <sub>18</sub> O <sub>4</sub> et C <sub>14</sub> H <sub>20</sub> O <sub>5</sub> . C'est un glucuronide de C <sub>14</sub> H <sub>20</sub> O <sub>5</sub> avec 2 fonctions acides carboxylique (3-(carboxyméthyl)-5-hydroxy-adamantylacetic acid...)	Glucuronides endogènes	NA	Non	444.1625	8.59	16.07
Acide Chénodéoxyglycocholique	C26H43NO5	Caractérisé <sup>a</sup>	Oui	Les tr ne correspondent pas à l'Acide glycodéoxycholique, glyochénodéoxycholique ou	Acides biliaires	Synthèse d'acides biliaires / Sécrétion biliaire	Non <sup>a</sup>	449.3141	8.86	16.51

				Glycoursodéoxy- cholique.						
FMNH	C17H23N4O9P	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	Non <sup>a</sup>	458.1203	5.70	12.30
Glucuronide de Déhydroépiandrost érone	C25H36O8	Annoté <sup>a,c</sup>	Oui	NA	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	464.2410	8.30	15.72
Glucuronide de Testostérone isomère	C25H36O8	Caractérisé <sup>a</sup>	Oui	NA	Glucuronides endogènes	Synthèse d'hormones stéroïdes	NA	464.2410	7.50	14.60
Glucuronide de Testostérone	C25H36O8	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de H <sub>2</sub> O et C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> (+H <sub>2</sub> O).	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	464.2410	7.98	15.25
Acide Glycocholique	C26H43NO6	Annoté <sup>a,c</sup>	Oui	NA	Acides biliaires	Synthèse d'acides biliaires / Sécretion biliaire	Oui <sup>a,b</sup>	465.3091	8.42	15.40
Acide Glycocholique isomère	C26H43NO6	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	465.3091	7.12	14.12
Acide Glycocholique isomère	C26H43NO6	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	465.3091	7.50	14.50
Acide Glycocholique isomère	C26H43NO6	Caractérisé <sup>a</sup>	Oui	NA	NA	NA	NA	465.3091	7.75	14.75
Glucuronide d'Androstérone	C25H38O8	Annoté <sup>a,c</sup>	Oui	Autre isomère possible : Glucuronide d'Etiocholanolone	Stéroïdes et dérivés	Synthèse d'hormones stéroïdes	Oui <sup>a,b</sup>	466.2567	8.65	16.03
Glucuronide d'acide Kétorétinoïque	C26H34O9	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> , C <sub>6</sub> H <sub>10</sub> O <sub>7</sub> , C <sub>20</sub> H <sub>24</sub> O <sub>2</sub> et C <sub>20</sub> H <sub>26</sub> O <sub>3</sub> . C'est un glucuronide de C <sub>20</sub> H <sub>26</sub> O <sub>3</sub> (Estradiol acetate/ Gestadienol/ hydroxynoréthistérone/ 6-Kétonoréthindrone/ Acide 4-Kétoretinoïque)	Glucuronides endogènes	NA	Non	490.2203	8.85	16.78
Glucuronide d'acide Hydroxylétinoïque	C26H36O9	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>6</sub> H <sub>8</sub> O <sub>6</sub> , C <sub>6</sub> H <sub>10</sub> O <sub>7</sub> , C <sub>20</sub> H <sub>26</sub> O <sub>2</sub> et C <sub>20</sub> H <sub>28</sub> O <sub>3</sub> . C'est un glucuronide de C <sub>20</sub> H <sub>28</sub> O <sub>3</sub> (Cafestrol/	Glucuronides endogènes	NA	Non	492.2359	8.95	16.70

				Gestonorone/ Acides Hydroxyrétinoïques/ Epoxyrétinoïque/ hydroxyméthétienone... )						
Lysophosphatidylcholine LysoPC(16:0)	C24H50NO7P	Caractérisé <sup>a</sup>	Oui	NA	Lysophosphatidylcholines	Métabolisme des glycéro-phospholipides	Non <sup>a</sup>	495.3325	11.67	15.91
1-Linoléoylglycéro-phosphocholine GPCCho(18:2)	C26H50NO7P	Caractérisé <sup>a</sup>	Oui	NA	Lysophosphatidylcholines	Métabolisme des glycéro-phospholipides	Non <sup>a</sup>	519.3325	11.17	12.57
1-Oleoylglycéro-phosphocholine LysoPC(18:1)	C26H52NO7P	Caractérisé <sup>a</sup>	Oui	NA	Lysophosphatidylcholines	Métabolisme des glycéro-phospholipides	Non <sup>a</sup>	521.3481	12.00	16.75
L-urobiline	C33H46N4O6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>7</sub> H <sub>9</sub> NO, C <sub>8</sub> H <sub>9</sub> NO, C <sub>17</sub> H <sub>22</sub> N <sub>2</sub> O <sub>3</sub> (+H <sub>2</sub> O) et C <sub>17</sub> H <sub>24</sub> N <sub>2</sub> O <sub>3</sub> (+H <sub>2</sub> O).	Porphyrines	Dégradation de l'hème / Métabolisme de la porphyrine et de la chlorophylle	Non <sup>a</sup>	594.3417	9.34	17.40
L-urobilinogène	C33H48N4O6	Putativement Identifié <sup>a,b</sup>	Oui	La MS <sup>n</sup> montre des pertes de C <sub>3</sub> H <sub>4</sub> O <sub>2</sub> , C <sub>17</sub> H <sub>24</sub> O <sub>3</sub> N <sub>2</sub> et C <sub>17</sub> H <sub>24</sub> O <sub>5</sub> N <sub>2</sub> .	Porphyrines	Dégradation de l'hème / Métabolisme de la porphyrine et de la chlorophylle	Non <sup>a</sup>	596.3574	9.40	17.45

- NA = non applicable, NR = non retrouvé

## ANNEXE 4 – PUBLICATIONS, POSTERS ET

### COMMUNICATIONS ORALES

#### **PUBLICATIONS RELATIVES A LA THESE**

Roux A, Lison D, Junot C, Heilier JF. Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clin Biochem.* 2011 Jan;44(1):119-35.

Roux A, Xu Y, Heilier JF, Olivier MFO, Ezan E, Tabet JC, Junot C. Development of an electrospray-mass spectral database for annotating metabolomics datasets: application to the analysis of the adult human urinary metabolome (en cours)

Roux A, Xu Y, Thevenot E, Olivier MFO, Paris A, Ezan E, Junot C. Analysis of adult human urinary métabolome : impact of gender, age and BMI on metabolite concentrations (en cours)

Roux A, Thevenot E, Olivier MFO, Junot C. Evaluation of the impact of collection and short-term storage conditions on the metabolite content of urine samples as analyzed by liquid chromatography coupled to mass spectrometry (en cours)

#### **PUBLICATIONS DANS LE CADRE DE COLLABORATIONS**

Forquin MP, Hébert A, Roux A, Aubert J, Proux C, Heilier JF, Landaud S, Junot C, Bonnarme P, Martin-Verstraete I. Global Regulation of the Response to Sulfur Availability in the Cheese-Related Bacterium *Brevibacterium aurantiacum*. *Appl Environ Microbiol.* 2011 Feb;77(4):1449-59.

Hébert A, Forquin-Gomez MP, Roux A, Aubert J, Junot C, Loux V, Heilier JF, Bonnarme P, Beckerich JM, Landaud S. Exploration of sulfur metabolism in the yeast *Kluyveromyces lactis*. *Appl Microbiol Biotechnol.* 2011 Sep;91(5):1409-23.

#### **POSTERS**

Roux A, Xu Y, Madalinski G, Heilier JF, Ezan E, Tabet JC, Junot C. Urinary metabolome analysis by LC/MS: Use of preservatives and temperature during sample collection, and the impact on metabolomic fingerprints. 4èmes Journées Scientifiques du Réseau Français de Métabolomique et Fluxomique - mai 2010 - Marseille, France.

Roux A, Xu Y, Madalinski G, Heilier JF, Ezan E, Tabet JC, Junot C. Urinary metabolome analysis by LC/MS: evaluation of the impact of collection and short-term storage conditions on the metabolite

content. 27èmes Journées Françaises de Spectrométrie de Masse - septembre 2010 - Clermont-Ferrand, France.

### **COMMUNICATIONS ORALES**

Roux A, Xu Y, Heilier JF, Madalinski G, Ezan E, Tabet JC, Junot C. Development of an ESI-mass spectral database for LC/MS based metabolomics. 4èmes Journées Scientifiques du Réseau Français de Métabolisme et Fluxomique - mai 2010 - Marseille, France.

Roux A, Xu Y, Heilier JF, Olivier MFO, Ezan E, Tabet JC, Junot C. Development of an electrospray-mass spectral database for annotating metabolomics datasets: application to the analysis of the adult human urinary metabolome. 59th ASMS Conference on Mass Spectrometry and Allied Topics – June 2011 - Denver, Colorado, USA.

# RÉFÉRENCES BIBLIOGRAPHIQUES



## RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Sumner LW, Mendes P, Dixon RA. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 2003;62: 817-36.
- [2] Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol* 1998;180: 5109-16.
- [3] Dettmer K, Hammock BD. Metabolomics--a new exciting field within the "omics" sciences. *Environ Health Perspect* 2004;112: A396-A397.
- [4] Forster J, Famili I, Fu P, Palsson BO, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 2003;13: 244-53.
- [5] Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998;16: 373-8.
- [6] Nobeli I, Ponstingl H, Krissinel EB, Thornton JM. A structure-based anatomy of the *E. coli* metabolome. *Journal of Molecular Biology* 2003;334: 697-719.
- [7] Keun HC, Ebbels TM, Bollard ME, et al. Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. *Chem Res Toxicol* 2004;17: 579-87.
- [8] Raamsdonk LM, Teusink B, Broadhurst D, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 2001;19: 45-50.
- [9] Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol* 1973;27: 65-104.
- [10] Cascante M, Boros LG, Comin-Anduix B, et al. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol* 2002;20: 243-9.
- [11] Kell DB, Westerhoff HV. *Metabolic Control-Theory - Its Role in Microbiology and Biotechnology*. *Fems Microbiology Reviews* 1986;39: 305-20.
- [12] Herbert RB. *The biosynthesis of secondary metabolites*. 2nd ed.: Chapman and Hall, 1989.
- [13] Shargel L, Yu A. *Applied Biopharmaceutics and Pharmacokinetics*. 4th ed.: McGraw-Hill, 1999.
- [14] Wishart DS. Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology* 2008;19: 482-93.
- [15] Crockford DJ, Maher AD, Ahmadi KR, et al. 1H NMR and UPLC-MS(E) statistical heterospectroscopy: characterization of drug metabolites (xenometabolome) in epidemiological studies. *Anal Chem* 2008;80: 6835-44.
- [16] DeCaprio AP. Biomarkers: Coming of Age for Environmental Health and Risk Assessment. *Environ Sci Technol* 1997;31: 1837-48.
- [17] Wolfe DA. Insights on the utility of biomarkers for environmental impact assessment and monitoring. *Human and Ecological Risk Assessment* 1996;2: 245-50.
- [18] Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008;134: 714-7.
- [19] Fave G, Beckmann ME, Draper JH, Mathers JC. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes and Nutrition* 2009;4: 135-41.
- [20] Patterson AD, Gonzalez FJ, Idle JR. Xenobiotic Metabolism: A View through the Metabolometer. *Chemical Research in Toxicology* 2010;23: 851-60.
- [21] Southam AD, Lange A, Hines A, et al. Metabolomics Reveals Target and Off-Target Toxicities of a Model Organophosphate Pesticide to Roach (*Rutilus rutilus*): Implications for Biomonitoring. *Environmental Science & Technology* 2011;45: 3759-67.
- [22] Bictash M, Ebbels TM, Chan Q, et al. Opening up the "Black Box": Metabolic phenotyping and metabolome-wide association studies in epidemiology. *Journal of Clinical Epidemiology* 2010;63: 970-9.

- [23] Claude BERNARD, Introduction à l'étude de la médecine expérimentale. Rééd. Champs, Flammarion, 1865.
- [24] Walter B. Cannon, The Wisdom of the Body. Peter Smith Pub Inc, 1932.
- [25] van der Greef J, Stroobant P, van der Heijden R. The role of analytical sciences in medical systems biology. *Curr Opin Chem Biol* 2004;8: 559-65.
- [26] Williams RJ, Berry LJ, Beerstecher E. Individual Metabolic Patterns, Alcoholism - Genetotropic Diseases. *science* 1949;109: 441.
- [27] Williams RJ. *Biochemical Individuality*.: John Wiley & Sons, 1956.
- [28] Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem Soc Rev* 2009;38: 1882-96.
- [29] Horning EC, Horning MG. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin Chem* 1971;17: 802-9.
- [30] Horning EC, Horning MG. Human metabolic profiles obtained by GC [gas chromatography] and GC/MS [gas chromatography/mass spectrometry]. *J Chromatogr Sci* 1971;9: 129-40.
- [31] Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Natl Acad Sci U S A* 1971;68: 2374-6.
- [32] Bennett DA, Waters MD. Applying biomarker research. *Environ Health Perspect* 2000;108: 907-10.
- [33] Paustenbach DJ. The practice of exposure assessment. In: Hayes AW, Ed. *Principles and Methods of Toxicology*. London: Taylor and Francis; 2001.
- [34] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69: 89-95.
- [35] Timbrell JA. Biomarkers in toxicology. *Toxicology* 1998;129: 1-12.
- [36] Angerer J, Ewers U, Wilhelm M. Human biomonitoring: State of the art. *International Journal of Hygiene and Environmental Health* 2007;210: 201-28.
- [37] Pearce SJ, Schrenk HH, Yant WP, Microcolorimetric determination of benzene in blood and urine. 1936.
- [38] Haufroid V, Jakubowski M, Janasik B, et al. Interest of genotyping and phenotyping of drug-metabolizing enzymes for the interpretation of biological monitoring of exposure to styrene. *Pharmacogenetics* 2002;12: 691-702.
- [39] Clayton TA, Lindon JC, Cloarec O, et al. Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* 2006;440: 1073-7.
- [40] Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci U S A* 2009;106: 14728-33.
- [41] Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29: 1181-9.
- [42] Roux A, Lison D, Junot C, Heilier JF. Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clinical Biochemistry* 2011;44: 119-35.
- [43] Park SY, Kim YW, Kim JE, Kim JY. Age-associated changes in fat metabolism in the rat and its relation to sympathetic activity. *Life Sci* 2006;79: 2228-33.
- [44] Hodson MP, Dear GJ, Roberts AD, et al. A gender-specific discriminator in Sprague-Dawley rat urine: the deployment of a metabolic profiling strategy for biomarker discovery and identification. *Anal Biochem* 2007;362: 182-92.
- [45] Plumb R, Granger J, Stumpf C, et al. Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* 2003;128: 819-23.
- [46] Williams RE, Lenz EM, Evans JA, et al. A combined (1)H NMR and HPLC-MS-based metabonomic study of urine from obese (fa/fa) Zucker and normal Wistar-derived rats. *J Pharm Biomed Anal* 2005;38: 465-71.

- [47] Ebbels TM, Holmes E, Lindon JC, Nicholson JK. Evaluation of metabolic variation in normal rat strains from a statistical analysis of <sup>1</sup>H NMR spectra of urine. *J Pharm Biomed Anal* 2004;36: 823-33.
- [48] Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK. An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* 2000;484: 169-74.
- [49] Dumas ME, Maibaum EC, Teague C, et al. Assessment of analytical reproducibility of H-1 NMR spectroscopy based metabolomics for large-scale epidemiological research: the INTERMAP study. *Analytical Chemistry* 2006;78: 2199-208.
- [50] Gu H, Chen H, Pan Z, et al. Monitoring diet effects via biofluids and their implications for metabolomics studies. *Anal Chem* 2007;79: 89-97.
- [51] Phipps AN, Stewart J, Wright B, Wilson ID. Effect of diet on the urinary excretion of hippuric acid and other dietary-derived aromatics in rat. A complex interaction between diet, gut microflora and substrate specificity. *Xenobiotica* 1998;28: 527-37.
- [52] Wikoff WR, Anfora AT, Liu J, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* 2009;106: 3698-703.
- [53] Lawton KA, Berger A, Mitchell M, et al. Analysis of the adult human plasma metabolome. *Pharmacogenomics* 2008;9: 383-97.
- [54] Saude EJ, Adamko D, Rowe BH, Marrie T, Sykes BD. Variation of metabolites in normal human urine. *Metabolomics* 2007;3: 439-51.
- [55] Warrack BM, Hnatyshyn S, Ott KH, et al. Normalization strategies for metabonomic analysis of urine samples. *J Chromatogr B Analyt Technol Biomed Life Sci* 2009;877: 547-52.
- [56] Sysi-Aho M, Katajamaa M, Yetukuri L, Oresic M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics* 2007;8.
- [57] FOLCH J, LEES M, Stanley GHS. A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *Journal of Biological Chemistry* 1957;226: 497-509.
- [58] Plumb RS, Stumpf CL, Granger JH, et al. Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun Mass Spectrom* 2003;17: 2632-8.
- [59] Williams RE, Major H, Lock EA, Lenz EM, Wilson ID. D-Serine-induced nephrotoxicity: a HPLC-TOF/MS-based metabonomics approach. *Toxicology* 2005;207: 179-90.
- [60] Jankevics A, Liepinsh E, Liepinsh E, et al. Metabolomic studies of experimental diabetic urine samples by H-1 NMR spectroscopy and LC/MS method. *Chemometrics and Intelligent Laboratory Systems* 2009;97: 11-7.
- [61] Wang J, Reijmers T, Chen L, et al. Systems toxicology study of doxorubicin on rats using ultra performance liquid chromatography coupled with mass spectrometry based metabolomics. *Metabolomics* 2009;5: 407-18.
- [62] Werner E, Croixmarie V, Umbdenstock T, et al. Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal Chem* 2008;80: 4918-32.
- [63] Brown SC, Kruppa G, Dasseux JL. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrometry Reviews* 2005;24: 223-31.
- [64] Bruce SJ, Tavazzi I, Parisod V, et al. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal Chem* 2009;81: 3285-96.
- [65] Want EJ, O'maille G, Smith CA, et al. Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem* 2006;78: 743-52.
- [66] Boernsen KO, Gatzek S, Imbert G. Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal Chem* 2005;77: 7255-64.
- [67] Croixmarie V, Umbdenstock T, Cloarec O, et al. Integrated comparison of drug-related and drug-induced ultra performance liquid chromatography/mass spectrometry metabonomic profiles using human hepatocyte cultures. *Anal Chem* 2009;81: 6061-9.

- [68] Soga T, Baran R, Suematsu M, et al. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J Biol Chem* 2006;281: 16768-76.
- [69] Soga T, Igarashi K, Ito C, et al. Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Anal Chem* 2009;81: 6165-74.
- [70] Beecher CWW. The human metabolome. In: Harrigan GG, Goodacre R, Eds. *Metabolic Profiling : Its Role In Biomarker Discovery And Gene Function Analysis*. Norwell: Kluwer Academic Publishers; 2003, p. 311-20.
- [71] Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* 2003;54: 669-89.
- [72] Vigneau-Callahan KE, Shestopalov AI, Milbury PE, Matson WR, Kristal BS. Characterization of diet-dependent metabolic serotypes: analytical and biological variability issues in rats. *J Nutr* 2001;131: 924S-32S.
- [73] Fraser PD, Pinto ME, Holloway DE, Bramley PM. Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J* 2000;24: 551-8.
- [74] Kristal BS, Vigneau-Callahan KE, Matson WR. Simultaneous analysis of the majority of low-molecular-weight, redox-active compounds from mitochondria. *Anal Biochem* 1998;263: 18-25.
- [75] Gamache PH, Meyer DF, Granger MC, Acworth IN. Metabolomic applications of electrochemistry/mass spectrometry. *J Am Soc Mass Spectrom* 2004;15: 1717-26.
- [76] Dumas ME, Debrauwer L, Beyet L, et al. Analyzing the physiological signature of anabolic steroids in cattle urine using pyrolysis/metastable atom bombardment mass spectrometry and pattern recognition. *Anal Chem* 2002;74: 5393-404.
- [77] Soga T, Ueno Y, Naraoka H, et al. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal Chem* 2002;74: 2233-9.
- [78] Garcia-Perez I, Vallejo M, Garcia A, Legido-Quigley C, Barbas C. Metabolic fingerprinting with capillary electrophoresis. *J Chromatogr A* 2008.
- [79] Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007;26: 51-78.
- [80] Wilson ID, Plumb R, Granger J, et al. HPLC-MS-based methods for the study of metabonomics. *J Chromatogr B* 2005;817: 67-76.
- [81] Aharoni A, Ric d, V, Verhoeven HA, et al. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS* 2002;6: 217-34.
- [82] Holmes E, Nicholson JK, Nicholls AW, et al. The identification of novel biomarkers of renal toxicity using automatic data reduction techniques and PCA of proton NMR spectra of urine. *Chemometrics and Intelligent Laboratory Systems* 1998;44: 245-55.
- [83] Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF. A <sup>1</sup>H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. *J Pharm Biomed Anal* 2003;33: 1103-15.
- [84] Nicholson JK, Wilson ID. High resolution proton magnetic resonance spectroscopy of biological fluids. *Progress in Nuclear Magnetic Resonance Spectroscopy* 1989;21: 449-501.
- [85] Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 2002;1: 153-61.
- [86] Cloarec O, Dumas ME, Trygg J, et al. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in <sup>1</sup>H NMR spectroscopic metabonomic studies. *Anal Chem* 2005;77: 517-26.
- [87] Wang Y, Bollard ME, Keun H, et al. Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning <sup>1</sup>H nuclear magnetic resonance spectroscopy of liver tissues. *Anal Biochem* 2003;323: 26-32.
- [88] Keun HC, Beckonert O, Griffin JL, et al. Cryogenic probe <sup>13</sup>C NMR spectroscopy of urine for metabonomic studies. *Anal Chem* 2002;74: 4588-93.
- [89] Spraul M, Freund AS, Nast RE, et al. Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe: application to the analysis of acetaminophen metabolites in urine. *Anal Chem* 2003;75: 1536-41.

- [90] Holmes E, Loo RL, Stamler J, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008;453: 396-U50.
- [91] Frycak P, Huskova R, Adam T, Lemr K. Atmospheric pressure ionization mass spectrometry of purine and pyrimidine markers of inherited metabolic disorders. *J Mass Spectrom* 2002;37: 1242-8.
- [92] Robb DB, Covey TR, Bruins AP. Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Anal Chem* 2000;72: 3653-9.
- [93] Lembcke J, Ceglarek U, Fiedler GM, et al. Rapid quantification of free and esterified phytosterols in human serum using APPI-LC-MS/MS. *Journal Of Lipid Research* 2005;46: 21-6.
- [94] Takino M, Daishima S, Nakahara T. Determination of chloramphenicol residues in fish meats by liquid chromatography-atmospheric pressure photoionization mass spectrometry. *J Chromatogr A* 2003;1011: 67-75.
- [95] Yoshioka N, Akiyama Y, Teranishi K. Rapid simultaneous determination of o-phenylphenol, diphenyl, thiabendazole, imazalil and its major metabolite in citrus fruits by liquid chromatography-mass spectrometry using atmospheric pressure photoionization. *Journal of Chromatography A* 2004;1022: 145-50.
- [96] Straube EA, Dekant W, Volkel W. Comparison of electrospray ionization, atmospheric pressure chemical ionization, and atmospheric pressure photoionization for the analysis of dinitropyrene and aminonitropyrene LC-MS/MS. *J Am Soc Mass Spectrom* 2004;15: 1853-62.
- [97] Karas M, Hillenkamp F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal Chem* 1988;60: 2229-301.
- [98] Thomas JJ, Shen Z, Crowell JE, Finn MG, Siuzdak G. Desorption/ionization on silicon (DIOS): a diverse mass spectrometry platform for protein characterization. *Proc Natl Acad Sci U S A* 2001;98: 4932-7.
- [99] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *science* 1989;246: 64-71.
- [100] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray Ionization-Principles and Practice. *Mass Spectrometry Reviews* 1990;9: 37-70.
- [101] Takats Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *science* 2004;306: 471-3.
- [102] Kraly J, Fazal MA, Schoenherr RM, et al. Bioanalytical applications of capillary electrophoresis. *Analytical Chemistry* 2006;78: 4097-110.
- [103] Werner E, Heilier JF, Ducruix C, et al. Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008;871: 143-63.
- [104] Goodacre R, York EV, Heald JK, Scott IM. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. *Phytochemistry* 2003;62: 859-63.
- [105] Smedsgaard J, Frisvad JC. Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. *Journal of Microbiological Methods* 1996;25: 5-17.
- [106] Madalinski G, Godat E, Alves S, et al. Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal Chem* 2008;80: 3291-303.
- [107] Boernsen KO, Gatzek S, Imbert G. Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Analytical Chemistry* 2005;77: 7255-64.
- [108] Zamfir A, Vakhrushev S, Sterling A, et al. Fully automated chip-based mass spectrometry for complex carbohydrate system analysis. *Analytical Chemistry* 2004;76: 2046-54.
- [109] Schiopu C, Flangea C, Capitan F, et al. Determination of ganglioside composition and structure in human brain hemangioma by chip-based nanoelectrospray ionization tandem mass spectrometry. *Analytical and Bioanalytical Chemistry* 2009;395: 2465-77.
- [110] Herring KD, Oppenheimer SR, Caprioli RM. Direct tissue analysis by matrix-assisted laser desorption ionization mass spectrometry: Application to kidney biology. *Seminars in Nephrology* 2007;27: 597-608.
- [111] Ishida Y, Kitagawa K, Nakayama A, Ohtani H. On-probe sample pretreatment for direct analysis of lipids in gram-positive bacterial cells by matrix-assisted laser desorption ionization mass spectrometry. *Applied and Environmental Microbiology* 2005;71: 7539-41.

- [112] Dwivedi P, Puzon G, Tam M, et al. Metabolic profiling of *Escherichia coli* by ion mobility-mass spectrometry with MALDI ion source. *Journal of Mass Spectrometry* 2010;45: 1383-93.
- [113] Dwivedi P, Schultz AJ, Hill HH. Metabolic profiling of human blood by high-resolution ion mobility mass spectrometry (IM-MS). *International Journal of Mass Spectrometry* 2010;298: 78-90.
- [114] Adahchour M, Jover E, Beens J, Vreuls RJ, Brinkman UA. Twin comprehensive two-dimensional gas chromatographic system: concept and applications. *J Chromatogr A* 2005;1086: 128-34.
- [115] Kovats ES. Gas Chromatographic Characterization of Organic Substances in the Retention Index System. In: Giddins JC, Keller RA, Eds. *Advances in Chromatography*. New-York: M. Dekker Inc.; 1965, p. 229-47.
- [116] Fiehn O, Kopka J, Dormann P, et al. Metabolite profiling for plant functional genomics. *Nat Biotechnol* 2000;18: 1157-61.
- [117] Bowers LD, Sanaullah. Direct measurement of steroid sulfate and glucuronide conjugates with high-performance liquid chromatography-mass spectrometry. *J Chromatogr B* 1996;687: 61-8.
- [118] Idborg H, Zamani L, Edlund PO, Schuppe-Koistinen I, Jacobsson SP. Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J Chromatogr B* 2005;828: 9-13.
- [119] Lafaye A, Labarre J, Tabet JC, Ezan E, Junot C. Liquid chromatography-mass spectrometry and <sup>15</sup>N metabolic labeling for quantitative metabolic profiling. *Anal Chem* 2005;77: 2026-33.
- [120] Strege MA. Hydrophilic Interaction Chromatography-Electrospray Mass Spectrometry Analysis of Polar Compounds for Natural Product Drug Discovery. *Anal Chem* 1998;70: 2439-45.
- [121] Tolstikov VV, Fiehn O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal Biochem* 2002;301: 298-307.
- [122] Lu WY, Clasquin MF, Melamud E, et al. Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer. *Analytical Chemistry* 2010;82: 3212-21.
- [123] Tolstikov VV, Lommen A, Nakanishi K, Tanaka N, Fiehn O. Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal Chem* 2003;75: 6737-40.
- [124] Uehara T, Yokoi A, Aoshima K, et al. Quantitative Phosphorus Metabolomics Using Nanoflow Liquid Chromatography-Tandem Mass Spectrometry and Culture-Derived Comprehensive Global Internal Standards. *Analytical Chemistry* 2009;81: 3836-42.
- [125] Myint KT, Aoshima K, Tanaka S, Nakamura T, Oda Y. Quantitative Profiling of Polar Cationic Metabolites in Human Cerebrospinal Fluid by Reversed-Phase Nanoflow Liquid Chromatography/Mass Spectrometry. *Analytical Chemistry* 2009;81: 1121-9.
- [126] Plumb R, Castro-Perez J, Granger J, et al. Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2004;18: 2331-7.
- [127] Leandro CC, Hancock P, Fussell RJ, Keely BJ. Comparison of ultra-performance liquid chromatography and high-performance liquid chromatography for the determination of priority pesticides in baby foods by tandem quadrupole mass spectrometry. *Journal of Chromatography A* 2006;1103: 94-101.
- [128] Nordstrom A, O'maille G, Qin C, Siuzdak G. Nonlinear Data Alignment for UPLC-MS and HPLC-MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum. *Anal Chem* 2006;78: 3289-95.
- [129] Monton MRN, Soga T. Metabolome analysis by capillary electrophoresis-mass spectrometry. *Journal of Chromatography A* 2007;1168: 237-46.
- [130] Soga T, Ohashi Y, Ueno Y, et al. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research* 2003;2: 488-94.
- [131] Li JJ, Kelly JF, Chemushevich I, Harrison DJ, Thibault P. Separation and identification of peptides from gel-isolated membrane proteins using a microfabricated device for combined capillary electrophoresis/nanoelectrospray mass spectrometry. *Analytical Chemistry* 2000;72: 599-609.
- [132] Enders JR, Marasco CC, Kole A, et al. Towards monitoring real-time cellular response using an integrated microfluidics-matrix assisted laser desorption ionisation/nanoelectrospray ionisation-ion mobility-mass spectrometry platform. *let Systems Biology* 2010;4: 416-27.

- [133] Kumar A, Malik AK, Pico Y. Sample preparation methods for the determination of pesticides in foods using CE-UV/MS. *Electrophoresis* 2010;31: 2115-25.
- [134] Wurm M, Schopke B, Lutz D, Muller J, Zeng AP. Microtechnology meets systems biology: The small molecules of metabolome as next big targets. *Journal of Biotechnology* 2010;149: 33-51.
- [135] Fairchild JN, Horvath K, Gooding JR, Campagna SR, Guiochon G. Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. *Journal of Chromatography A* 2010;1217: 8161-6.
- [136] Guo XH, Lankmayr E. Multidimensional approaches in LC and MS for phospholipid bioanalysis. *Bioanalysis* 2010;2: 1109-23.
- [137] Tomas R, Kleparnik K, Foret F. Multidimensional liquid phase separations for mass spectrometry. *Journal of Separation Science* 2008;31: 1964-79.
- [138] Bailey NJ, Cooper P, Hadfield ST, et al. Application of directly coupled HPLC-NMR-MS/MS to the identification of metabolites of 5-trifluoromethylpyridone (2-hydroxy-5-trifluoromethylpyridine) in hydroponically grown plants. *J Agric Food Chem* 2000;48: 42-6.
- [139] Lindon JC, Nicholson JK, Wilson ID. Biomedical and Pharmaceutical Applications of HPLC-NMR and HPLC-NMR-MS. In: Albert K, Ed. *On-Line LC-NMR and Related Techniques*. 2003, p. 45-87.
- [140] Shockcor JP, Unger SE, Wilson ID, et al. Combined HPLC, NMR spectroscopy, and ion-trap mass spectrometry with application to the detection and characterization of xenobiotic and endogenous metabolites in human urine. *Anal Chem* 1996;68: 4431-5.
- [141] Moco S, Forshed J, De Vos RCH, Bino RJ, Vervoort J. Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography-mass spectrometry and nuclear magnetic resonance. *Metabolomics* 2008;4: 202-15.
- [142] Holmes E, Foxall PJ, Nicholson JK, et al. Automatic data reduction and pattern recognition methods for analysis of <sup>1</sup>H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal Biochem* 1994;220: 284-96.
- [143] Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabolomic data sets. *Anal Chem* 2006;78: 2262-7.
- [144] Forshed J, Torgrip RJ, Aberg KM, et al. A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *J Pharm Biomed Anal* 2005;38: 824-32.
- [145] Stoyanova R, Nicholson JK, Lindon JC, Brown TR. Sample classification based on Bayesian spectral decomposition of metabolomic NMR data sets. *Anal Chem* 2004;76: 3666-74.
- [146] Broeckling CD, Reddy IR, Duran AL, Zhao X, Sumner LW. MET-IDEA: Data Extraction Tool for Mass Spectrometry-Based Metabolomics. *Anal Chem* 2006;78: 4334-41.
- [147] Duran AL, Yang J, Wang L, Sumner LW. Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). *Bioinformatics* 2003;19: 2283-93.
- [148] Katajamaa M, Oresic M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* 2005;6: 179.
- [149] Plumb RS, Stumpf CL, Gorenstein MV, et al. Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. *Rapid Commun Mass Spectrom* 2002;16: 1991-6.
- [150] Smith CA, Want EJ, 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: 779-87.
- [151] Tikunov Y, Lommen A, de Vos CH, et al. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol* 2005;139: 1125-37.
- [152] Katajamaa M, Oresic M. Data processing for mass spectrometry-based metabolomics. *J Chromatogr A* 2007;1158: 318-28.
- [153] Pedrioli PG, Eng JK, Hubley R, et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 2004;22: 1459-66.

- [154] Rew RK, G.P.Davis. NetCDF: An Interface for Scientific Data Access. *IEEE computer graphics and applications* 1990;10: 76-82.
- [155] Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 2006;22: 634-6.
- [156] Baran R, Kochi H, Saito N, et al. MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* 2006;7: 530.
- [157] Spasic I, Dunn WB, Velarde G, et al. MeMo: a hybrid SQL/XML approach to metabolomic data management for functional genomics. *BMC Bioinformatics* 2006;7.
- [158] Xia JG, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* 2009;37: W652-W660.
- [159] Lamprecht AL, Margaria T, Steffen B. Bio-jETI: a framework for semantics-based service composition. *BMC Bioinformatics* 2009;10.
- [160] Margaria T, Kubczak C, Steffen B. Bio-jETI: a service integration, design, and provisioning platform for orchestrated bioinformatics processes. *BMC Bioinformatics* 2008;9.
- [161] Neuweger H, Albaum SP, Dondrup M, et al. MeltDB: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics* 2008;24: 2726-32.
- [162] Tautenhahn R, Bottcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 2008;9: 504.
- [163] Crawford LR, Morrison JD. Computer Methods in Analytical Mass Spectrometry - Identification of An Unknown Compound in A Catalog. *Anal Chem* 1968;40: 1464-&.
- [164] Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* 2004;20: 2447-54.
- [165] Wang WX, Zhou HH, Lin H, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* 2003;75: 4818-26.
- [166] Oresic M, Clish CB, Davidov EJ, et al. Phenotype characterisation using integrated gene transcript, protein and metabolite profiling. *Appl Bioinformatics* 2004;3: 205-17.
- [167] Hermansson M, Uphoff A, Kakela R, Somerharju P. Automated quantitative analysis of complex lipidomes by liquid chromatography/mass spectrometry. *Anal Chem* 2005;77: 2166-75.
- [168] Bijlsma S, Bobeldijk I, Verheij ER, et al. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal Chem* 2006;78: 567-74.
- [169] Aitchison J. *The statistical Analysis of Compositional Data*. Caldwell, NJ: Blackburn Press, 2003.
- [170] Redestig H, Fukushima A, Stenlund H, et al. Compensation for Systematic Cross-Contribution Improves Normalization of Mass Spectrometry Based Metabolomics Data. *Analytical Chemistry* 2009;81: 7974-80.
- [171] Sysi-Aho M, Katajamaa M, Yetukuri L, Oresic M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics* 2007;8: 93.
- [172] Wu L, Mashego MR, van Dam JC, et al. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly <sup>13</sup>C-labeled cell extracts as internal standards. *Anal Biochem* 2005;336: 164-71.
- [173] Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J. Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. *Trends in Biotechnology* 2005;23: 28-33.
- [174] Hegeman AD, Schulte CF, Cui Q, et al. Stable isotope assisted assignment of elemental compositions for metabolomics. *Anal Chem* 2007;79: 6912-21.
- [175] Wiklund S, Johansson E, Sjostrom L, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry* 2008;80: 115-22.
- [176] Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag* 1901; 559-72.
- [177] Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 1933;24: 417-41.

- [178] Wold S, Sjöström M, Eriksson L. PLS in Chemistry, The Encyclopedia of Computational Chemistry. 1998.
- [179] Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* 2002;16: 119-28.
- [180] Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. *J Proteome Res* 2007;6: 469-79.
- [181] Cloarec O, Dumas ME, Craig A, et al. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic <sup>1</sup>H NMR data sets. *Anal Chem* 2005;77: 1282-9.
- [182] Wolfender JL, Ndjoko K, Hostettmann K. Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites. *J Chromatogr A* 2003;1000: 437-55.
- [183] Molyneux RJ, Schieberle P. Compound identification: a Journal of Agricultural and Food Chemistry perspective. *J Agric Food Chem* 2007;55: 4625-9.
- [184] Lafaye A, Junot C, Ramounet-Le Gall B, et al. Metabolite profiling in rat urine by liquid chromatography/electrospray ion trap mass spectrometry. Application to the study of heavy metal toxicity. *Rapid Commun Mass Spectrom* 2003;17: 2541-9.
- [185] Bogusz MJ, Maier RD, Kruger KD, et al. Poor reproducibility of in-source collisional atmospheric pressure ionization mass spectra of toxicologically relevant drugs. *J Chromatogr A* 1999;844: 409-18.
- [186] Cui Q, Lewis IA, Hegeman AD, et al. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* 2008;26: 162-4.
- [187] Matsuda F, Yonekura-Sakakibara K, Niida R, et al. MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J* 2009;57: 555-77.
- [188] Werner E, Heilier JF, Ducruix C, et al. Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *J Chromatogr B* 2008;871: 143-63.
- [189] Werner E, Croixmarie V, Umbdenstock T, et al. Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal Chem* 2008;80: 4918-32.
- [190] Xu Y, Heilier JF, Madalinski G, et al. Evaluation of Accurate Mass and Relative Isotopic Abundance Measurements in the LTQ-Orbitrap Mass Spectrometer for Further Metabolomics Database Building. *Analytical Chemistry* 2010;82: 5490-501.
- [191] Sumner LW, Urbanczyk-Wochniak E, Broeckling CD. Metabolomics data analysis, visualization, and integration. *Methods Mol Biol* 2007;406: 409-36.
- [192] Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 2007;3: 211-21.
- [193] Werner E, Heilier JF, Ducruix C, et al. Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *J Chromatogr B* 2008;871: 143-63.
- [194] Kopka J, Schauer N, Krueger S, et al. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 2005;21: 1635-8.
- [195] Wishart DS, Tzur D, Knox C, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007;35: D521-D526.
- [196] Watanabe K, Yasugi E, Oshima M. How to search the glycolipid data in "LIPIDBANK for Web", the newly developed lipid database in Japan. *Trends Glycosci Glycotechnol* 2000;12: 175-84.
- [197] Smith CA, O'maille G, Want EJ, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005;27: 747-51.
- [198] Horai H, Suwa K, Arita M, Nihei Y, Nishioka T, Mass Spectral Database for Metabolome Analysis. Indianapolis, USA 2007.
- [199] Robertson DG, Watkins PB, Reilly MD. Metabolomics in Toxicology: Preclinical and Clinical Applications. *Toxicological Sciences* 2011;120: S146-S170.
- [200] Kusano M, Redestig H, Hirai T, et al. Covering Chemical Diversity of Genetically-Modified Tomatoes Using Metabolomics for Objective Substantial Equivalence Assessment. *Plos One* 2011;6.

- [201] Reaves ML, Rabinowitz JD. Metabolomics in systems microbiology. *Current Opinion in Biotechnology* 2011;22: 17-25.
- [202] Shepherd LVT, Fraser P, Stewart D. Metabolomics: a second-generation platform for crop and food analysis. *Bioanalysis* 2011;3: 1143-59.
- [203] Whitfield PD, German AJ, Noble PJ. Metabolomics: an emerging post-genomic tool for nutrition. *Br J Nutr* 2004;92: 549-55.
- [204] German JB, Watkins SM, Fay LB. Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health. *J Am Diet Assoc* 2005;105: 1425-32.
- [205] Takahashi H, Morioka R, Ito R, et al. Dynamics of Time-Lagged Gene-to-Metabolite Networks of *Escherichia coli* Elucidated by Integrative Omics Approach. *Omics-A Journal of Integrative Biology* 2011;15: 15-23.
- [206] Wu JY, Kao HJ, Li SC, et al. ENU mutagenesis identifies mice with mitochondrial branched-chain aminotransferase deficiency resembling human maple syrup urine disease. *J Clin Invest* 2004;113: 434-40.
- [207] Hirai MY, Yano M, Goodenowe DB, et al. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2004;101: 10205-10.
- [208] Soga T, Baran R, Suematsu M, et al. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J Biol Chem* 2006;281: 16768-76.
- [209] Saito N, Robert M, Kitamura S, et al. Metabolomics approach for enzyme discovery. *J Proteome Res* 2006;5: 1979-87.
- [210] Saghatelian A, McKinney MK, Bandell M, Patapoutian A, Cravatt BF. A FAAH-regulated class of N-acyl taurines that activates TRP ion channels. *Biochemistry* 2006;45: 9007-15.
- [211] Lindon JC, Holmes E, Bollard ME, Stanley EG, Nicholson JK. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* 2004;9: 1-31.
- [212] Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: A global biochemical approach to drug response and disease. *Annual Review of Pharmacology and Toxicology* 2008;48: 653-83.
- [213] Kaddurah-Daouk R. Metabolomics: A Global Biochemical Approach to the Study of Neuropsychiatric Disorders. *Biological Psychiatry* 2009;65: 79S.
- [214] Chace DH. Mass spectrometry in the clinical laboratory. *Chem Rev* 2001;101: 445-77.
- [215] Griffin JL, Sang E, Evens T, Davies K, Clarke K. Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy. *FEBS Lett* 2002;530: 109-16.
- [216] Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. *Nat Rev Cancer* 2004;4: 551-61.
- [217] Odunsi K, Wollman RM, Ambrosone CB, et al. Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *Int J Cancer* 2005;113: 782-8.
- [218] Shaham O, Slate NG, Goldberger O, et al. A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. *Proc Natl Acad Sci U S A* 2010;107: 1571-5.
- [219] Bailey NJ, Oven M, Holmes E, Nicholson JK, Zenk MH. Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via 1H NMR spectroscopy and chemometrics. *Phytochemistry* 2003;62: 851-8.
- [220] Dumas ME, Canlet C, Debrauwer L, Martin P, Paris A. Selection of biomarkers by a multivariate statistical processing of composite metabonomic data sets using multiple factor analysis. *J Proteome Res* 2005;4: 1485-92.
- [221] Prabakaran S, Swatton JE, Ryan MM, et al. Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol Psychiatry* 2004;9: 684-97, 643.
- [222] Mochel F, Sedel F, Vanderver A, et al. Cerebellar ataxia with elevated cerebrospinal free sialic acid (CAFSA). *Brain* 2009;132: 801-9.
- [223] Sreekumar A, Poisson LM, Rajendiran TM, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 2009;457: 910-4.
- [224] Zhang YJ, Dai YB, Wen JM, et al. Detrimental effects of adenosine signaling in sickle cell disease. *Nature Medicine* 2011;17: 79-U252.

- [225] Darghouth D, Koehl B, Madalinski G, et al. Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. *Blood* 2011;117: E57-E66.
- [226] Larsen N, Vogensen FK, van den Berg FWJ, et al. Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. *Plos One* 2010;5.
- [227] Turnbaugh PJ, Hamady M, Yatsunencko T, et al. A core gut microbiome in obese and lean twins. *Nature* 2009;457: 480-U7.
- [228] Backhed F, Crawford PA. Coordinated regulation of the metabolome and lipidome at the host-microbial interface. *Biochimica et Biophysica Acta-Molecular and Cell Biology of Lipids* 2010;1801: 240-5.
- [229] Backhed F, Ding H, Wang T, et al. The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101: 15718-23.
- [230] Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473: 174-80.
- [231] Beger RD, Sun J, Schnackenberg LK. Metabolomics approaches for discovering biomarkers of drug-induced hepatotoxicity and nephrotoxicity. *Toxicol Appl Pharmacol* 2010;243: 154-66.
- [232] Griffin JL. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 2003;7: 648-54.
- [233] Lindon JC, Nicholson JK, Holmes E, et al. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol Appl Pharmacol* 2003;187: 137-46.
- [234] Shockcor JP, Holmes E. Metabonomic applications in toxicity screening and disease diagnosis. *Curr Top Med Chem* 2002;2: 35-51.
- [235] Ganote CE, Peterson DR, Carone FA. The nature of D-serine-induced nephrotoxicity. *Am J Pathol* 1974;77: 269-82.
- [236] Williams RE, Lock EA. Sodium benzoate attenuates D-serine induced nephrotoxicity in the rat. *Toxicology* 2005;207: 35-48.
- [237] McBurney RN, Hines WM, Von Tungeln LS, et al. The liver toxicity biomarker study: phase I design and preliminary results. *Toxicol Pathol* 2009;37: 52-64.
- [238] Sun J, Schnackenberg LK, Holland RD, et al. Metabonomics evaluation of urine from rats given acute and chronic doses of acetaminophen using NMR and UPLC/MS. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008;871: 328-40.
- [239] Schnackenberg LK, Chen M, Sun J, et al. Evaluations of the trans-sulfuration pathway in multiple liver toxicity studies. *Toxicol Appl Pharmacol* 2009;235: 25-32.
- [240] La S, Yoo HH, Kim DH. Liquid chromatography-mass spectrometric analysis of urinary metabolites and their pattern recognition for the prediction of drug-induced hepatotoxicity. *Chem Res Toxicol* 2005;18: 1887-96.
- [241] Boudonck KJ, Mitchell MW, Nemet L, et al. Discovery of metabolomics biomarkers for early detection of nephrotoxicity. *Toxicol Pathol* 2009;37: 280-92.
- [242] National Bureau of Standards, American Petroleum Institute Catalog of Mass Spectral Data, Project 44. 2011.
- [243] Zeman PD. Punched Card Catalog of Mass Spectra Useful in Qualitative Analysis. *Analytical Chemistry* 1950;22: 920-2.
- [244] Adahchour M, Jover E, Beens J, Vreuls RJ, Brinkman UA. Twin comprehensive two-dimensional gas chromatographic system: concept and applications. *J Chromatogr A* 2005;1086: 128-34.
- [245] Kind T, Wohlgemuth G, Lee DY, et al. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Analytical Chemistry* 2009;81: 10038-48.
- [246] McLafferty FW, Stauffer DA, Loh SY, Wesdemiotis C. Unknown identification using reference mass spectra. Quality evaluation of databases. *Journal of the American Society for Mass Spectrometry* 1999;10: 1229-40.
- [247] Marquet P, Venisse N, Lacassie E, Lachatre G. In-source CID mass spectral libraries for the "general unknown" screening of drugs and toxicants. *Analisis* 2000;28: 925-934A.

- [248] Dresen S, Gergov M, Politi L, Halter C, Weinmann W. ESI-MS/MS library of 1,253 compounds for application in forensic and clinical toxicology. *Analytical and Bioanalytical Chemistry* 2009;395: 2521-6.
- [249] Dresen S, Ferreiros N, Gnann H, Zimmermann R, Weinmann W. Detection and identification of 700 drugs by multi-target screening with a 3200 Q TRAP(A (R)) LC-MS/MS system and library searching. *Analytical and Bioanalytical Chemistry* 2010;396: 2425-34.
- [250] Matsuda F, Yonekura-Sakakibara K, Niida R, et al. MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant Journal* 2009;57: 555-77.
- [251] Horai H, Arita M, Kanaya S, et al. MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* 2010;45: 703-14.
- [252] Jansen R, Lachatre G, Marquet P. LC-MS/MS systematic toxicological analysis: Comparison of MS/MS spectra obtained with different instruments and settings. *Clinical Biochemistry* 2005;38: 362-72.
- [253] Farid Ichou, Denis Lesage, Xavier Machuron-Mandard, Christophe Junot, Jean-Claude Tabet, Calibration method for creating inter-instrumental and inter-laboratory spectral databank. in: *ASMS2011 SMQApn1*, editor. 2011.
- [254] Oberacher H, Weinmann W, Dresen S. Quality evaluation of tandem mass spectral libraries. *Analytical and Bioanalytical Chemistry* 2011;400: 2641-8.
- [255] Oberacher H, Pavlic M, Libiseller K, et al. On the inter-instrument and inter-laboratory transferability of a tandem mass spectral reference library: 1. Results of an Austrian multicenter study. *Journal of Mass Spectrometry* 2009;44: 485-93.
- [256] Mohamed R, Varesio E, Ivosev G, et al. Comprehensive Analytical Strategy for Biomarker Identification based on Liquid Chromatography Coupled to Mass Spectrometry and New Candidate Confirmation Tools. *Analytical Chemistry* 2009;81: 7677-94.
- [257] Brown M, Dunn WB, Dobson P, et al. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* 2009;134: 1322-32.
- [258] Chen J, Zhao X, Fritsche J, et al. Practical approach for the identification and isomer elucidation of biomarkers detected in a metabonomic study for the discovery of individuals at risk for diabetes by integrating the chromatographic and mass spectrometric information. *Analytical Chemistry* 2008;80: 1280-9.
- [259] Kuhl C, Tautenhahn R, Neumann S. LC-MS Peak Annotation and Identification with CAMERA. *bioconductor fhrc org* 2011.
- [260] Tautenhahn R, Bottcher C, Neumann S. Annotation of LC/ESI-MS Mass Signals. *Proceedings of BIRD 2007 - 1st International Conference on Bioinformatics Research and Development* 2007;4414/2007: 371-80.
- [261] Zuniga A, Li LA. Ultra-high performance liquid chromatography tandem mass spectrometry for comprehensive analysis of urinary acylcarnitines. *Analytica Chimica Acta* 2011;689: 77-84.
- [262] Jones IG. Eck,Rv - Atlas of Protein Sequence and Structure 1966. *Nature* 1967;214: 854-&.
- [263] Barker WC, George DG, Mewes HW, Tsugita A. The Pir-International Protein-Sequence Database. *Nucleic Acids Research* 1992;20: 2023-6.
- [264] Rindone WP, Perry HM, Goad WB, Bilofsky HS, Carrico CK. Genbank Tm - the Genetic Sequence Data-Bank. *Dna-A Journal of Molecular & Cellular Biology* 1983;2: 173.
- [265] Stoesser G, Sterk P, Tuli MA, Stoehr PJ, Cameron GN. The EMBL Nucleotide Sequence Database. *Nucleic Acids Research* 1997;25: 7-13.
- [266] Shin'i T, Gojobori T. [DDBJ database and genetic information analysis softwares]. *Tanpakushitsu kakusan koso Protein, nucleic acid, enzyme* 1994;39.
- [267] Bairoch A, Boeckmann B. The Swiss-Prot Protein-Sequence Data-Bank. *Nucleic Acids Research* 1991;19: 2247-8.
- [268] Page RDM. GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics* 1998;14: 819-20.
- [269] Chaudhuri RR, Khan AM, Pallen MJ. ColiBASE: an online database for Escherichia coli, Shigella and Salmonella comparative genomics. *Nucleic Acids Research* 2004;32: D296-D299.
- [270] Ashburner M, Drysdale R. Flybase - the Drosophila Genetic Database. *Development* 1994;120: 2077-9.

- [271] Harris TW, Lee R, Schwarz E, et al. WormBase: a cross-species database for comparative genomics. *Nucleic Acids Research* 2003;31: 133-7.
- [272] Galperin MY, Cochrane GR. The 2011 *Nucleic Acids Research Database Issue* and the online *Molecular Biology Database Collection*. *Nucleic Acids Research* 2011;39: D1-D6.
- [273] Degtyarenko K, De Matos P, Ennis M, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research* 2008;36: D344-D350.
- [274] Seiler KP, George GA, Happ MP, et al. ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Research* 2008;36: D351-D359.
- [275] Ogata H, Goto S, Sato K, et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 1999;27: 29-34.
- [276] Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database. *Nucleic Acids Research* 2002;30: 59-61.
- [277] Patrick JM. Geigy Scientific Tables, 8Th Edition, Vol 1, 1981, Vol 2, 1982, Vol 3, 1984 - Ciba-Geigy. *Ergonomics* 1986;29: 173-4.
- [278] Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100: 9440-5.
- [279] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 1995;57: 289-300.
- [280] Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 1979;74: 829-36.
- [281] Dunn WB, Broadhurst D, Begley P, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* 2011;6: 1060-83.
- [282] Sirbu A, Ruskin HJ, Crane M. Cross-Platform Microarray Data Normalisation for Regulatory Network Inference. *Plos One* 2010;5.
- [283] Alston MJ, Seers J, Hinton JCD, Lucchini S. BABAR: an R package to simplify the normalisation of common reference design microarray-based transcriptomic datasets. *BMC Bioinformatics* 2010;11.
- [284] Bond A, Dudley E, Lemiere F, et al. Analysis of urinary nucleosides. V. Identification of urinary pyrimidine nucleosides by liquid chromatography/electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry* 2006;20: 137-50.
- [285] Shamim W, Yousufuddin M, Bakhai A, Coats AJS, Honour JW. Gender differences in the urinary excretion rates of cortisol and androgen metabolites. *Annals of Clinical Biochemistry* 2000;37: 770-4.
- [286] Rehman KS, Carr BR. Sex differences in adrenal androgens. *Seminars in Reproductive Medicine* 2004;22: 349-60.
- [287] Kiens B. Gender differences in lipid metabolism. *Chemistry and Physics of Lipids* 2008;154: S11.
- [288] Lamont LS, McCullough AJ, Kalhan SC. Gender differences in the regulation of amino acid metabolism. *Journal of Applied Physiology* 2003;95: 1259-65.
- [289] Tipton KD. Gender differences in protein metabolism. *Current Opinion in Clinical Nutrition and Metabolic Care* 2001;4: 493-8.
- [290] Petrov RV, Kuzina IN, Kilikovsky VV, Smirnova OV. Age-dependent changes in the concentration of active sex steroids, precursors, metabolites, and regulatory agents in blood serum of male subjects. *Russian Journal of Developmental Biology* 2009;40: 373-81.
- [291] Dharía S, Parker CR. Adrenal androgens and aging. *Seminars in Reproductive Medicine* 2004;22: 361-8.
- [292] Meyer FP, Canzler E, Giers H, Walther H. Time Course of Impaired Elimination of Caffeine by the Oral Depot Contraceptive Deposiston. *Zentralblatt für Gynäkologie* 1991;113: 297-302.
- [293] Bolton S, Null G. Caffeine Psychological Effects, Use and Abuse. *Journal of Orthomolecular Psychiatry* 1981;10: 202-11.

- [294] Ortweiler W, Simon HU, Splinter FK, et al. Elimination of Caffeine and Metamizol in Pregnancy and After Delivery As An In vivo Method for Characterization of Various Cytochrome-P-450 Dependent Biotransformation Reactions. *Biomedica Biochimica Acta* 1985;44: 1189-99.
- [295] Chung WG, Kang JH, Park CS, Cho MH, Cha YN. Effect of age and smoking on in vivo CYP1A2, flavin-containing monooxygenase, and xanthine oxidase activities in Koreans: Determination by caffeine metabolism. *Clinical Pharmacology & Therapeutics* 2000;67: 258-66.
- [296] Tang BK. Caffeine Metabolism As A Measure of Cytochrome-P450 Function. *Journal of Cellular Biochemistry* 1995; 187.
- [297] Guy PA, Tavazzi I, Bruce SJ, Ramadan Z, Kochhar S. Global metabolic profiling analysis on human urine by UPLC-TOFMS: issues and method validation in nutritional metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008;871: 253-60.
- [298] Gika HG, Theodoridis GA, Wilson ID. Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine - Sample stability under different handling and storage conditions for metabolomics studies. *Journal of Chromatography A* 2008;1189: 314-22.
- [299] Lauridsen M, Hansen SH, Jaroszewski JW, Cornett C. Human urine as test material in H-1 NMR-based metabolomics: Recommendations for sample preparation and storage. *Analytical Chemistry* 2007;79: 1181-6.
- [300] Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-day reproducibility of an HPLC-MS-Based method for metabolomic analysis: Application to human urine. *Journal of Proteome Research* 2007;6: 3291-303.
- [301] Maher AD, Zirah SFM, Holmes E, Nicholson JK. Experimental and analytical variation in human urine in H-1 NMR spectroscopy-based metabolic phenotyping studies. *Analytical Chemistry* 2007;79: 5204-11.
- [302] Thongboonkerd V, Saetun P. Bacterial overgrowth affects urinary proteome analysis: Recommendation for centrifugation, temperature, duration, and the use of preservatives during sample collection. *Journal of Proteome Research* 2007;6: 4173-81.
- [303] Saude EJ, Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* 2007;3: 19-27.
- [304] Maskell RM. The natural history of urinary tract infection in women. *Medical Hypotheses* 2010;74: 802-6.
- [305] Barber M, Brooksbank BWL, Haslewood GAD. Destruction of Urinary Glucuronide by Bacteria. *Nature* 1948;162: 701-2.
- [306] Vogels GD, Vanderdrift C. Degradation of Purines and Pyrimidines by Microorganisms. *Bacteriological Reviews* 1976;40: 403-68.
- [307] Bando K, Kawahara R, Kunimatsu T, et al. Influences of biofluid sample collection and handling procedures on GC-MS based metabolomic studies. *Journal of Bioscience and Bioengineering* 2010;110: 491-9.
- [308] Sturm M, Kohlbacher O. TOPPView: An Open-Source Viewer for Mass Spectrometry Data. *Journal of Proteome Research* 2009;8: 3760-3.
- [309] Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R. PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis. *Analytical Chemistry* 2011;83: 2786-93.
- [310] Bleakney W. A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor. *Phys Rev* 1929;34: 157.
- [311] Nier AOC, Ney EP, Ingham MG. Adjustable Gas Leak. *Review of Scientific Instruments* 1947;18: 191.
- [312] Harrison AG, Gaumann T, Stahl D. Structure and Fragmentation of [C<sub>3</sub>H<sub>7</sub>O]<sup>+</sup> Ions Formed by Chemical Ionization. *Organic Mass Spectrometry* 1983;18: 517-24.
- [313] Cody RB, Laramee JA, Durst HD. Versatile new ion source for the analysis of materials in open air under ambient Conditions. *Anal Chem* 2005;77: 2297-302.
- [314] Horning EC, Horning MG, Carroll DI, Dzidic I, Stillwell RN. New picogram detection system based on a mass spectrometer with an external ionization source at atmospheric pressure. *Anal Chem* 1973;45: 936-43.
- [315] Kauppila TJ, Kotiaho T, Kostianen R, Bruins AP. Negative ion-atmospheric pressure photoionization-mass spectrometry. *J Am Soc Mass Spectrom* 2004;15: 203-11.

- [316] Kauppila TJ, Kuuranne T, Meurer EC, et al. Atmospheric pressure photoionization mass spectrometry. Ionization mechanism and the effect of solvent on the ionization of naphthalenes. *Anal Chem* 2002;74: 5470-9.
- [317] Lavanant H, Lange C. Sodium-tolerant matrix for matrix-assisted laser desorption/ionization mass spectrometry and post-source decay of oligonucleotides. *Rapid Commun Mass Spectrom* 2002;16: 1928-33.
- [318] Mims D, Hercules D. Quantification of bile acids directly from urine by MALDI-TOF-MS. *Anal Bioanal Chem* 2003;375: 609-16.
- [319] Mims D, Hercules D. Quantification of bile acids directly from plasma by MALDI-TOF-MS. *Anal Bioanal Chem* 2004;378: 1322-6.
- [320] Edwards JL, Kennedy RT. Metabolomic analysis of eukaryotic tissue and prokaryotes using negative mode MALDI time-of-flight mass spectrometry. *Anal Chem* 2005;77: 2201-9.
- [321] Valdes-Gonzalez T, Goto-Inoue N, Hirano W, et al. New approach for glyco- and lipidomics - Molecular scanning of human brain gangliosides by TLC-Blot and MALDI-QIT-TOF MS. *Journal of Neurochemistry* 2011;116: 678-83.
- [322] Hart PJ, Francese S, Claude E, Woodroffe MN, Clench MR. MALDI-MS imaging of lipids in ex vivo human skin. *Analytical and Bioanalytical Chemistry* 2011;401: 115-25.
- [323] Hankin JA, Farias SE, Barkley R, et al. MALDI Mass Spectrometric Imaging of Lipids in Rat Brain Injury Models. *Journal of the American Society for Mass Spectrometry* 2011;22: 1014-21.
- [324] Laiko VV, Baldwin MA, Burlingame AL. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* 2000;72: 652-7.
- [325] Yamashita M, Fenn JB. Electrospray Ion-Source - Another Variation on the Free-Jet Theme. *Journal of Physical Chemistry* 1984;88: 4451-9.
- [326] Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal Chem* 1985;57: 675-9.
- [327] Chen H, Pan Z, Talaty N, Raftery D, Cooks RG. Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. *Rapid Commun Mass Spectrom* 2006;20: 1577-84.
- [328] Shen Z, Thomas JJ, Averbuj C, et al. Porous silicon as a versatile platform for laser desorption/ionization mass spectrometry. *Anal Chem* 2001;73: 612-9.
- [329] Wei J, Buriak JM, Siuzdak G. Desorption-ionization mass spectrometry on porous silicon. *Nature* 1999;399: 243-6.
- [330] Huikko K, Ostman P, Sauber C, et al. Feasibility of atmospheric pressure desorption/ionization on silicon mass spectrometry in analysis of drugs. *Rapid Commun Mass Spectrom* 2003;17: 1339-43.
- [331] Nordstrom A, Want E, Northen T, Lehtio J, Siuzdak G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal Chem* 2008;80: 421-9.
- [332] Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Detection Technologies. Ambient mass spectrometry. *science* 2006;311: 1566-70.
- [333] Chen H, Talaty NN, Takats Z, Cooks RG. Desorption electrospray ionization mass spectrometry for high-throughput analysis of pharmaceutical samples in the ambient environment. *Anal Chem* 2005;77: 6915-27.
- [334] Jackson AU, Werner SR, Talaty N, et al. Targeted metabolomic analysis of *Escherichia coli* by desorption electrospray ionization and extractive electrospray ionization mass spectrometry. *Anal Biochem* 2008;375: 272-81.
- [335] Ferguson RE, McCulloh KE, Rosenstock HM. Observation of the Products of Ionic Collision Processes and Ion Decomposition in a Linear, Pulsed Time-of-Flight Mass Spectrometer. *The Journal of Chemical Physics* 1965;42: 100-6.
- [336] Campana JE. Elementary theory of the quadrupole mass filter. *International Journal of Mass Spectrometry and Ion Physics* 1980;33: 101-17.
- [337] Bouchoux G, Sablier M. Spectrométrie de masse - Principes et appareillages. *Techniques de l'ingénieur*. Paris: Editions T.I.; 2005, p. **P 2 645-1-P 2 645-32**.
- [338] March RE, Hughes RJ. *Quadrupole Storage Mass Spectrometry*.: Wiley Interscience, 1989.

- [339] Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery* 2003;2: 140-50.
- [340] Parks JH, Szoke A. Simulation of Collisional Relaxation of Trapped Ion Clouds in the Presence of Space-Charge Fields. *Journal of Chemical Physics* 1995;103: 1422-39.
- [341] Douglas DJ, Frank AJ, Mao D. Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* 2005;24: 1-29.
- [342] Hashimoto Y, Hasegawa H, Waki I. Dual linear ion trap/orthogonal acceleration time-of-flight mass spectrometer with improved precursor ion selectivity. *Rapid Commun Mass Spectrom* 2005;19: 1485-91.
- [343] Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 2002;13: 659-69.
- [344] Bantscheff M, Boesche M, Eberhard D, et al. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics* 2008.
- [345] Griffin TJ, Xie H, Bandhakavi S, et al. iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J Proteome Res* 2007;6: 4200-9.
- [346] Stephens WE. A pulsed mass spectrometer with time dispersion. *Phys Rev* 1946;69: 691.
- [347] Wiley WC, McLaren IH. Time-of-Flight Mass Spectrometer with Improved Resolution. *Review of Scientific Instruments* 1955;26: 1150-7.
- [348] Mamyrin BA, Karataev VI, Shmikk DV, Zagulin VA. Mass-Reflectron A New Nonmagnetic Time-Of-Flight High-Resolution Mass-Spectrometer. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki* 1973;64: 82-9.
- [349] Roepenack-Lahaye E, Degenkolb T, Zerjeski M, et al. Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol* 2004;134: 548-59.
- [350] Weckwerth W, Loureiro ME, Wenzel K, Fiehn O. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci U S A* 2004;101: 7809-14.
- [351] Major HJ, Williams R, Wilson AJ, Wilson ID. A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition. *Rapid Commun Mass Spectrom* 2006;20: 3295-302.
- [352] Plumb RS, Johnson KA, Rainville P, et al. The detection of phenotypic differences in the metabolic plasma profile of three strains of Zucker rats at 20 weeks of age using ultra-performance liquid chromatography/orthogonal acceleration time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2006;20: 2800-6.
- [353] Williams R, Lenz EM, Wilson AJ, et al. A multi-analytical platform approach to the metabonomic analysis of plasma from normal and Zucker (fa/fa) obese rats. *Mol Biosyst* 2006;2: 174-83.
- [354] Castro-Perez J, McCullagh M, Millar A. 720001170EN - The benefits of extended dynamic range for metabolite ID using UPLC/MS/MS. *Waters Technical Note* 2005;720001170EN.
- [355] McCullagh M, Castro-Perez J, Calton L. 720001596EN - Q-ToF Premier dynamic range enhancement and accurate mass measurement performance. *Waters Technical Note* 2006;720001596EN.
- [356] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998;17: 1-35.
- [357] Marshall AG. Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. *International Journal of Mass Spectrometry* 2000;200: 331-56.
- [358] Hughey CA, Rodgers RP, Marshall AG. Resolution of 11,000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. *Anal Chem* 2002;74: 4145-9.
- [359] Kruppa G, Schnier PD, Tabei K, Van Orden S, Siegel MM. Multiple ion isolation applications in FT-ICR MS: exact-mass MSn internal calibration and purification/interrogation of protein-drug complexes. *Anal Chem* 2002;74: 3877-86.
- [360] Baidoo EE, Benke PI, Neuss C, et al. Capillary electrophoresis-fourier transform ion cyclotron resonance mass spectrometry for the identification of cationic metabolites via a pH-mediated stacking-transient isotachophoretic method. *Anal Chem* 2008;80: 3112-22.

- [361] Nakamura Y, Kimura A, Saga H, et al. Differential metabolomics unraveling light/dark regulation of metabolic activities in *Arabidopsis* cell culture. *Planta* 2007;227: 57-66.
- [362] Ohta D, Shibata D, Kanaya S. Metabolic profiling using Fourier-transform ion-cyclotron-resonance mass spectrometry. *Anal Bioanal Chem* 2007;389: 1469-75.
- [363] Southam AD, Payne TG, Cooper HJ, Arvanitis TN, Viant MR. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal Chem* 2007;79: 4595-602.
- [364] Stentiford GD, Viant MR, Ward DG, et al. Liver tumors in wild flatfish: a histopathological, proteomic, and metabolomic study. *OMICS* 2005;9: 281-99.
- [365] Takahashi H, Kai K, Shinbo Y, et al. Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry. *Anal Bioanal Chem* 2008;391: 2769-82.
- [366] Taylor NS, Weber RJM, Southam AD, et al. A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* 2009;5: 44-58.
- [367] Nakamura Y, Sakurai N, Iijima Y, et al. A metabolome analyzing system for high-throughput processing of data obtained from liquid chromatography-fourier transform-mass spectrometry (LC/FT-ICR-MS). *Plant and Cell Physiology* 2007;48: S68.
- [368] Nakamura Y, Shinbo Y, Asahi H, et al. Development of a high-throughput comprehensive analyzing system for metabolome analyses using FT-ICR/MS. *Plant and Cell Physiology* 2006;47: S246.
- [369] Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000;72: 1156-62.
- [370] Hu Q, Noll RJ, Li H, et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 2005;40: 430-43.
- [371] Yates JR, Cociorva D, Liao L, Zabrouskov V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal Chem* 2006;78: 493-500.
- [372] Makarov A, Denisov E, Kholomeev A, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 2006;78: 2113-20.
- [373] Peterman SM, Duczak N, Jr., Kalgutkar AS, Lame ME, Soglia JR. Application of a linear ion trap/orbitrap mass spectrometer in metabolite characterization studies: examination of the human liver microsomal metabolism of the non-tricyclic anti-depressant nefazodone using data-dependent accurate mass measurements. *J Am Soc Mass Spectrom* 2006;17: 363-75.
- [374] Ding J, Sorensen CM, Zhang Q, et al. Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling. *Anal Chem* 2007;79: 6081-93.
- [375] Dunn WB, Broadhurst D, Brown M, et al. Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system. *J Chromatogr B Analyt Technol Biomed Life Sci* 2008;871: 288-98.
- [376] Kenny LC, Broadhurst D, Brown M, et al. Detection and identification of novel metabolomic biomarkers in preeclampsia. *Reprod Sci* 2008;15: 591-7.
- [377] Zubarev RA, Kelleher NL, McLafferty FW. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J Am Chem Soc* 1998;120: 3265-6.
- [378] Woodin RL, Bomse DS, Beauchamp JL. Multiphoton dissociation of molecules with low power continuous wave infrared laser radiation. *J Am Chem Soc* 1978;100: 3248-50.
- [379] Cooper HJ, Hakansson K, Marshall AG. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrometry Reviews* 2005;24: 201-22.
- [380] Levsen K. *Fundamental Aspects of Organic Mass Spectrometry*. Weinheim: Verlag chemie, 1978.
- [381] Lafaye A, Junot C, Ramounet-Le Gall B, et al. Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats. *J Mass Spectrom* 2004;39: 655-64.
- [382] Lenz EM, Bright J, Knight R, et al. Metabonomics with <sup>1</sup>H-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. *Biomarkers* 2005;10: 173-87.

- [383] Wagner S, Scholz K, Donegan M, et al. Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. *Anal Chem* 2006;78: 1296-305.
- [384] VanBocxlaer JF, Clauwaert KM, Lambert WE, DeLeenheer AP. Quantitative colorimetric determination of urinary p-aminophenol with an automated analyzer. *Clinical Chemistry* 1997;43: 627-34.
- [385] Nishiwaki F, Kuroda K, Inoue Y, Endo G. Determination of histamine, 1-methylhistamine and N-methylhistamine by capillary electrophoresis with micelles. *Biomedical Chromatography* 2000;14: 184-7.
- [386] Jellum E, Horn L, Thoresen O, Kvittingen EA, Stokke O. Urinary-Excretion of N-Acetyl Amino-Acids in Patients with Some Inborn-Errors of Amino-Acid-Metabolism. *Scandinavian Journal of Clinical & Laboratory Investigation* 1986;46: 21-6.
- [387] Ogata M, Yamasaki Y, Kawai T. Significance of Urinary Phenyl Sulfate and Phenyl Glucuronide As Indexes of Exposure to Phenol. *International Archives of Occupational and Environmental Health* 1986;58: 197-202.
- [388] Duran M, Ketting D, Vanvossen R, et al. Octanoylglucuronide Excretion in Patients with A Defective Oxidation of Medium-Chain Fatty-Acids. *Clinica Chimica Acta* 1985;152: 253-60.



L'objectif de ce travail de thèse est de développer une base de données spectrale pour faciliter l'annotation et l'interprétation biologique des jeux de données d'analyse métabolomique obtenus en utilisant la chromatographie liquide couplée à la spectrométrie de masse. Deux approches ont été utilisées : (i) l'identification par comparaison aux spectres de masse de composés de références et (ii) l'identification directement à partir des données biologiques. Pour la première approche une chimiothèque de métabolite a été constituée et analysée. L'identification à partir de données biologiques a été réalisée sur une cohorte de volontaires de 227 individus travaillant au CEA. 244 métabolites ont ainsi été identifiés dans les urines humaines, donc 78 qui n'avaient jamais été décrits comme faisant parti du métabolome urinaire. 139 métabolites ont également été caractérisés sur la base de leur masse précise mais sans identification formelle. Ces 383 métabolites représentent environ 1000 ions dans chacun des modes d'ionisation. Les variations physiologiques au sein de la cohorte, en fonction de l'âge, du poids et du genre, de ces différents métabolites ont été étudiées afin de construire une base de données relationnelle. Enfin, le métabolome urinaire pouvant être affecté par les conditions de prélèvement des échantillons d'urines, nous avons réalisé des études de stabilité dans les conditions de prélèvement des métabolites précédemment caractérisés. Ces études nous ont permis de proposer des recommandations en termes de conditions de prélèvement et de stockage à court terme des urines et de mesurer l'impact de la contamination bactérienne sur les concentrations de différents métabolites urinaires.

The aim of this work was to develop tools to help metabolites identification and biological interpretation of metabolomics analysis by LC / MS. In one hand, a spectral database was developed to facilitate the annotation of biological datasets from these analyses. Two approaches were performed: identification by comparison with reference compounds spectra and direct identification from biological dataset. For the first approach a chemical database of metabolites was formed and analyzed. The identification from biological data was performed on a cohort of volunteers of 227 CEA workers. 244 metabolites were identified in human urine, of which 78 never described as part of the urinary metabolome. 139 metabolites were also characterized based on accurate masses but without formal identification. These 383 metabolites account for about 1000 ions in each ionization mode. Physiological changes within the cohort, depending on age, weight and sex, these metabolites were studied in order to build a relational database. This database aims to better understand data generated by metabolomics studies. Finally, as the urinary metabolome can be affected by sampling conditions, stability studies were performed. It allowed us to determine the best conditions for sampling and short-term storage of urine and to measure bacterial overgrowth impact on urinary metabolite content.

### **Mots-clés**

Métabolomique, Métabolome, urines humaines, Chromatographie liquide couplée à la spectrométrie de masse, Base de données, Variations physiologiques, stabilité des métabolites urinaires