

# UNIVERSITÉ PARIS-SUD 11

## ECOLE DOCTORALE :

Chimie de Paris Sud (ED 470)

*PÔLE : CHIMIE*

## DISCIPLINE : CHIMIE

ANNÉE 2010 - 2011

SÉRIE DOCTORAT N° 1120

## THÈSE DE DOCTORAT

soutenue le 29/09/2011

par

**Erwan WERNER**

# ANALYSE DU METABOLOME PAR CHROMATOGRAPHIE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE :

## Application à la recherche de biomarqueurs indirects d'induction enzymatique

**Directeur de thèse :** Pr Pierre CHAMINADE

**Encadrant CEA :** Dr Christophe JUNOT

### Composition du jury :

*Rapporteurs :* Dr Laurent DEBRAUWER (INRA Toulouse)  
Dr Alain PARIS (INRA Paris – Agro Paris Tech)

*Examineurs :* Pr Fathi MOUSSA (Université Paris XI)  
Pr Jean-Claude TABET (Université Paris VI)  
Dr Vincent CROIXMARIE (Technologie Servier, Orléans)



## **Résumé**

Issue d'un partenariat de recherche entre le CEA et les laboratoires Servier, cette thèse avait pour objectif d'évaluer l'approche métabolomique par chromatographie liquide couplée à la spectrométrie de masse (LC-MS) pour l'identification de marqueurs indirects de l'induction dans les espèces de toxicologie. Le travail de thèse a débuté par l'optimisation de la méthode d'acquisition des empreintes métaboliques tant sur le plan analytique que dans le domaine du traitement des données brutes. Un outil reposant sur les matrices d'autocorrélation a alors été développé afin de s'affranchir d'une partie de la redondance du signal obtenu par spectrométrie de masse. Dans un troisième temps, les indices de Kendrick couplés à la sélectivité méthylène ont été appliqués à l'étude de composés biologiques en spectrométrie de masse haute résolution afin de proposer une méthode alternative de visualisation des données offrant une aide à l'identification des variables. Enfin, dans une dernière partie, les efforts se sont portés sur l'identification des composés endogènes modifiés au cours du protocole d'induction.

**Mots-Clés** : Métabolomique – Chromatographie liquide couplée à la spectrométrie de masse – Outils de visualisation – Biomarqueurs – Identification – Induction enzymatique.

## **ABSTRACT**

This work is the result of a research partnership between the CEA and Les laboratories Servier. It deals with the characterization of biomarkers of metabolic enzyme induction in rat biofluids using MS-based metabolomics. The first part of this work included methodological developments regarding the acquisition and the processing of metabolic fingerprints. A tool based on autocorrelation matrices was then implemented to reduce the redundancy of data generated with mass spectrometry and subsequently accelerate the isolation of discriminating variables. The next step consisted in the evaluation of the combined use of Kendrick mass defects and methylene selectivity as an alternative visualization tool for large data set, which would rely on compound chemical structures. Finally, the last part of the work was dedicated to the identification of discriminating signals raised up by a metabolomic global approach from rat biofluids collected before and after an induction assay.

**Keywords**: Metabolomics – Liquid Chromatography / Mass Spectrometry – Visualization tool – Biomarkers – Identification – Enzyme induction.



## Remerciements

En premier lieu, je remercie les docteurs Laurent Debrauwer et Alain Paris ainsi que l'ensemble des membres du jury pour avoir accepté de juger ce travail.

Les travaux présentés dans ce mémoire de thèse ont été réalisés au CEA de Saclay, au sein du Service de Pharmacologie et d'Immunologie, dans le Laboratoire d'Etude du Métabolisme des Médicaments. Je remercie donc les docteurs Jacques GRASSI et Christophe CREMINON, chefs du Service successifs, et le Dr. Eric EZAN, chef du laboratoire, pour m'avoir accueilli au sein de leur service et laboratoire, et pour m'y avoir offert d'excellentes conditions de travail.

Je tiens particulièrement à remercier le Dr. Christophe JUNOT ainsi que les Pr. Pierre CHAMINADE et Jean-Claude TABEL pour m'avoir encadré et dirigé durant ce travail de thèse. Merci pour vos conseils, votre soutien indéfectible et votre bonne humeur.

Ce travail étant réalisé en collaboration, je remercie sincèrement les Laboratoires Servier et plus encore l'entité Biopharmacie et son centre de pharmacocinétique et métabolisme Technologie Servier, dirigé par Bernard WALTHER, pour leur soutien financier, leur très bon accueil et la qualité de nos échanges au cours de cette collaboration. J'adresse un grand merci à l'ensemble du service pré-développement biopharmaceutique, dirigé par Claire Boursier-Neyret, et particulièrement à Thierry, Vincent, Françoise et Nathalie.

Un grand merci à tous les membres du SPI, jeunes et moins jeunes (ils se reconnaîtront) pour leur aide, leur soutien et leur bonne humeur tout au long de ces trois années partagées.

Enfin, merci à ma famille ainsi qu'à mes amis pour leurs encouragements dans les moments de doutes.



# Table des matières

<b>Résumés .....</b>	<b>3</b>
<b>Remerciements .....</b>	<b>5</b>
<b>Table des matières.....</b>	<b>7</b>
<b>Liste des Figures .....</b>	<b>11</b>
<b>Liste des Tableaux.....</b>	<b>15</b>
<b>Liste des Abréviations.....</b>	<b>17</b>
<b>CHAPITRE I : INTRODUCTION GÉNÉRALE.....</b>	<b>19</b>
<b>CHAPITRE II : ELEMENTS BIBLIOGRAPHIQUES .....</b>	<b>25</b>
<b>1 METABOLISATION DES XENOBIOTIQUES.....</b>	<b>27</b>
1.1 Introduction.....	27
1.2 Les enzymes de phase I .....	28
1.3 Les réactions de phase II .....	36
1.4 Les transporteurs de phase III .....	40
1.5 Induction enzymatique .....	43
1.6 Facteurs influençant le métabolisme.....	46
1.7 Les outils employés dans les études de métabolisme.....	47
<b>2 LE METABOLOME.....</b>	<b>52</b>
2.1 Définition.....	53
2.2 Les différentes approches métabolomiques .....	54
2.3 Applications .....	55
<b>3 METHODES D'ANALYSES DU METABOLOME .....</b>	<b>58</b>
3.1 La RMN .....	59
3.2 Les méthodes basées sur l'utilisation de la spectrométrie de masse.....	60
3.3 Approches analytiques multidimensionnelles .....	82
<b>4 PRÉTRAITEMENT DES DONNEES .....</b>	<b>83</b>
4.1 RMN .....	83

4.2	LC/MS.....	84
<b>5</b>	<b><u>TRAITEMENT DES DONNEES.....</u></b>	<b>88</b>
<b>6</b>	<b><u>IDENTIFICATION DES VARIABLES D'INTERET .....</u></b>	<b>92</b>
<b><u>CHAPITRE III : TRAVAUX PERSONNELS .....</u></b>		<b>97</b>
<b><u>Liste des communications.....</u></b>		<b>101</b>
<b>1</b>	<b><u>EXPERIMENTATION ANIMALE .....</u></b>	<b>103</b>
1.1	Introduction.....	103
1.2	Modèle animal.....	103
1.3	Caractérisation du modèle : Induction.....	106
<b>2</b>	<b><u>DÉVELOPPEMENTS MÉTHODOLOGIQUES .....</u></b>	<b>109</b>
2.1	Introduction.....	109
2.2	Optimisation analytique .....	110
2.3	Optimisation de l'extraction des variables.....	116
2.4	Analyses statistiques multivariées .....	121
2.5	Identification des variables d'intérêt .....	126
2.6	Conclusion.....	127
<b>3</b>	<b><u>DÉVELOPPEMENTS STATISTIQUES .....</u></b>	<b>129</b>
3.1	Introduction.....	129
3.2	Matériel et méthode.....	130
3.3	Résultats et Discussion .....	133
3.4	Conclusion.....	146
<b>4</b>	<b><u>OUTILS DE VISUALISATION : UNE AIDE A L'IDENTIFICATION ?.....</u></b>	<b>147</b>
4.1	Introduction.....	147
4.2	Matériel et méthode.....	148
4.3	Résultats et Discussion .....	150
4.4	Conclusion.....	160
<b>5</b>	<b><u>IDENTIFICATION DES VARIABLES D'INTERET .....</u></b>	<b>163</b>

5.1	Introduction.....	163
5.2	Matériel et méthode.....	165
5.3	Résultats.....	165
5.4	Intégration des résultats dans le contexte biologique .....	183
5.5	Conclusion.....	184
<b><u>CHAPITRE IV : CONCLUSION GENERALE .....</u></b>		<b>187</b>
<b><u>RÉFÉRENCES BIBLIOGRAPHIQUES .....</u></b>		<b>195</b>
<b><u>ANNEXES.....</u></b>		<b>213</b>



## Liste des Figures

Figure 1: Etapes du développement d'un médicament et place des études de métabolisme.....	22
Figure 2: Schéma général du processus de détoxification .....	27
Figure 3: Structure schématique d'un cytochrome P450.....	29
Figure 4: Cycle catalytique d'un cytochrome P450.....	30
Figure 5: Contribution des cytochromes P450 humains à la phase I du métabolisme des médicaments actuellement mis sur le marché <sup>13</sup> .....	32
Figure 6: Devenir d'un xénobiotique au cours du processus de métabolisation .....	34
Figure 7: Glucuronoconjugaison.....	36
Figure 8: Conjugaison au glutathion et synthèse d'acide mercapturique.....	40
Figure 9: Structure schématique des transporteurs de type ABC <sup>52</sup> .....	41
Figure 10: Transports hépatique et rénal des xénobiotiques .....	42
Figure 11: Structure d'un récepteur nucléaire.....	44
Figure 12: Implication des récepteurs PXR et CAR dans la détoxification des xénobiotiques .....	45
Figure 13 : Schéma du principe de l'ionisation MALDI.....	63
Figure 14 : Schéma d'une source électrospray.....	65
Figure 15 : Schéma d'un analyseur quadripolaire <sup>190</sup> .....	68
Figure 16 : Représentation schématique d'un triple quadripôle <sup>191</sup> .....	68
Figure 17 : Piège à ions 3D (Image issue de la documentation Thermo-Electron) .....	69
Figure 18 : Principe de fonctionnement d'un analyseur à temps de vol <sup>191</sup> .....	72
Figure 19 : Principe de fonctionnement du FT-ICR.....	73
Figure 20: Piège électrostatique ou « Orbitrap® » d'après la description de A. Makarov <sup>221</sup> .....	75
Figure 21: Courbe et équation de van Deemter.....	81
Figure 22: Représentation schématique de l'architecture 3D d'un chromatogramme obtenu par LC-MS .....	84
Figure 23 : Représentation des données lors des analyses ACP sous forme de carte factorielle des individus <sup>243</sup> .	89
Figure 24 : Représentation des données lors des analyses multivariées sous forme de carte factorielle des individus (A) et de carte des poids factoriels (B) .....	90
Figure 25: Présentation des données en OPLS.....	92
Figure 26 : Ratio d'expression des ARN <sub>m</sub> des principaux cytochromes P450 obtenus par RT-PCR.....	107
Figure 27: Déroulement d'une analyse métabolomique .....	110
Figure 28 : Exemple de profil métabolique standard .....	115
Figure 29 : Comparaison UPLC et HPLC.....	115
Figure 30 : Interface utilisateur du logiciel MetAlign® .....	117
Figure 31 : Comparaison des chromatogrammes obtenus avant et après traitement par le logiciel MetAlign® ..	119
Figure 32 : Cartes factorielles des individus obtenues par ACP. ....	123
Figure 33 : Cartes des poids factoriels obtenus en ACP. ....	124
Figure 34 : Analyse OPLS des données HPLC-piège-ionique en mode négatif - milieu neutre méthanol. ....	125
Figure 35 : Attribution des variables discriminantes en phase neutre-méthanol de l'effet traitement aux métabolites du PB.....	127
Figure 36 : Profils métaboliques urinaires de rat typiques obtenus en UPLC-TOF.....	134

Figure 37 : Présentation générale des matrices d'autocorrélation. ....	138
Figure 38 : Spectre de masse en ESI-Orbitrap® de deux pics à 24.02 min (A) et 24.52 min (B) contenant un ion de masse nominale $m/z$ 263. ....	140
Figure 39 : Spectre de masse Orbitrap® du dérivé mercapturique du phénobarbital. ....	141
Figure 40 : Données de spectrométrie de masse ayant servi à la caractérisation du signal d'intérêt $m/z$ 626. ....	143
Figure 41 : Proposition de schéma retraçant les voies métaboliques du Phénobarbital. ....	145
Figure 42 : Spectre de masse du PB en mode négatif. ....	145
Figure 43 : Comparaison des représentations en échelle IUPAC et de Kendrick de classes issues des acides carboxyliques aliphatiques. ....	152
Figure 44 : Extrait d'une représentation de Kendrick d'un profil métabolique urinaire de rat. ....	153
Figure 45 : Stratégie d'approche de données de métabolomiques pour l'isolation de série d'homologues en approches semi-ciblées ou pour la construction de bases de données chimiques. ....	154
Figure 46 : Spectres de fragmentation CID obtenus à partir des précurseurs $m/z$ 201.1134 à 22.06 et 27.08 min avec un pouvoir de résolution de 7500, une énergie de collision de 35 u.a. et la fonction wide band activation activée. ....	156
Figure 47 : Comparaison des chromatogrammes extraits pour les rapports $m/z$ des acides carboxyliques dans une urine de rat et pour leur standards commerciaux. ....	157
Figure 48 : Temps de rétention des acides dicarboxyliques en fonction de leur nombre d'unité méthylène. ....	158
Figure 49 : Matérialisation des séries d'homologues pour les composés isobares isolés à partir de leur défaut de masse de Kendrick. ....	159
Figure 50 : Stratégie d'identification d'un biomarqueur potentiel mis en évidence par approche métabolomique à l'aide d'expériences complémentaires en LC/MS et d'interrogation des bases de données. ....	164
Figure 51 : Comparaison des profils urinaires avant (rouge) et après (noir) traitement au PB après suppression des métabolites exogènes. ....	166
Figure 52 : Exemples de courbes de tendance obtenues pour les signaux les plus discriminants en analyse OPLS. ....	166
Figure 53 : Analyse en composante principale des groupes contrôle (NaCl) et PB. ....	167
Figure 54 : Valeurs individuelles pour des métabolites discriminant les individus avant et après traitement au PB et comparaison aux valeurs obtenues pour les individus du groupe NaCl. ....	168
Figure 55 : Analyse combinée des groupes 3-MC et BNF en analyses multivariées. ....	171
Figure 56 : Valeurs individuelles des 4 variables isolées en mode négatif d'après une évolution commune sous traitement 3-MC et BNF. ....	172
Figure 57 : Valeurs individuelles de la variable $m/z$ 366 à 9.28 pour l'ensemble des groupes analysés ayant pour véhicule l'huile de maïs. ....	173
Figure 58 : Détermination d'une composition élémentaire sur un système HPLC-Orbitrap®. ....	175
Figure 59 : Identification de l'acide pantothénique dans l'urine de rat. ....	177
Figure 60 : Etude de la variable d'intérêt $m/z$ 201.1133. ....	178
Figure 61 : Identification de l'acide sébacique dans l'urine de rat. ....	178
Figure 62 : (A) Structure proposée par la base CAS pour le composé inconnu à $m/z$ 201 ( $C_{10}H_{18}O_4$ ) et son mécanisme proposé de fragmentation avec une perte initiale de $H_2CO_2$ (B1) ou $CO_2$ (B2). ....	179

Figure 63 : Proposition de structure pour le composé d'intérêt $C_{10}H_{18}O_4$ correspondant à l'ion $m/z$ 201 élué à 22.06 minutes (A) à partir d'une proposition de mécanisme de fragmentation explicitant la formation de l'ensemble des ions détectés.....	180
Figure 64 : Valeurs individuelles de l'acide hippurique pour l'ensemble des profils en faveur de l'existence de sous-population d'animaux présentant des phénotypes de microbiote différents.....	184



## Liste des Tableaux

Tableau 1: Familles de cytochrome P450 et principaux substrats et fonctions .....	31
Tableau 2 : Récapitulatif des performances des différents analyseurs .....	67
Tableau 3: Principales bases de données utilisées en analyse métabolomique <sup>294</sup> .....	94
Tableau 4 : Composés inducteurs sélectionnés et leur profil d'induction associé.....	104
Tableau 5 : Groupes d'animaux et conditions de traitement.....	104
Tableau 6 : Mesure des activités enzymatiques des cytochromes P450.....	106
Tableau 7: Présentation du gradient utilisé pour l'éluion des métabolites sur le système HPLC-LCQ .....	111
Tableau 8 : Présentation du gradient utilisé pour l'éluion des métabolites sur le système UPLC-TOF.....	112
Tableau 9 : Impact de la durée du gradient sur la répétabilité des aires et temps de rétention des étalons. ....	114
Tableau 10 : Paramètres optimisés d'extraction des signaux avec le logiciel MetAlign®.....	120
Tableau 11 : Nombre de variables extraites selon le logiciel et la méthode analytique employés.....	121
Tableau 12 : Nombres de variables discriminantes issues de l'analyse différentielle des empreintes métaboliques. .....	126
Tableau 13 : Paramètres des modèles OPLS pour les différentes conditions analytiques.....	135
Tableau 14 : Nombres de variables discriminantes identifiées par l'analyse différentielle des empreintes.....	136
Tableau 15 : Résumé de l'impact des matrices d'autocorrélation sur le nombre de variables à identifier.....	139
Tableau 16 : Panorama des métabolites détectés en fonction des conditions d'analyse. ....	144
Tableau 17 : Calcul des masses nominales et indices de Kendrick pour les acides carboxyliques aliphatiques .	148
Tableau 18 : Rapports m/z ayant un indice de Kendrick de 0.105 et résultats des requêtes dans les bases HMDB et KEGG, obtenus à partir des formules brutes leur correspondant. ....	155
Tableau 19 : Régressions linéaires du temps de rétention des acides dicarboxyliques en fonction de leur nombre d'unité méthylène pour trois urines de rat. ....	159
Tableau 20 : Exemples de familles chimiques de composés biologiques isolées grâce aux indices de Kendrick. .....	160
Tableau 21 : Approche en statistiques classiques pour la sélection de variables communément modifiées après un traitement au 3-MC ou à la BNF.....	169
Tableau 22 : Résultats de l'approche SUS plot pour la sélection de variables communes, régulées de façon similaire avec les deux traitements.....	172
Tableau 23 : Composés identifiés au cours de l'analyse du groupe PB.....	182
Tableau 24 : Composés identifiés au cours de l'analyse des groupe 3-MC et BNF.....	183



## Liste des Abréviations

**ACN** : Acétonitrile

**AcONH<sub>4</sub>** : Acétate d'ammonium

**ACP** : Analyse en Composantes Principales

**ADN**: Acide désoxyribonucléique

**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 messenger

**ATP**: Adénosine tri-phosphate

**CE** (*Capillary electrophoresis*) : Electrophorèse Capillaire

**CE/MS** : Electrophorèse Capillaire couplée à la spectrométrie de masse

**CEA** : Commissariat à l'Energie Atomique

**CID** (*Collision Induced Dissociation*) : Dissociations induites par collision

**CV** : Coefficient de Variation

**CYP** : Cytochrome P450

**Da** : Dalton

**DIOS** (*Desorption-Ionization mass spectrometry On porous Silicon*) : Désorption-ionisation sur silicium poreuse

**EI** (*Electron Ionisation*) : Ionisation par impact d'électron

**ESI** (*Electrospray Source Ionization*) : Source électrospray

**FAB** (*Fast Atom Bombardment*) : Ionisation par bombardement d'ions ou d'atomes rapides

**FT-ICR** : Résonance cyclotronique des ions à transformée de Fourier

**GC** (*Gas Chromatography*) : Chromatographie en phase gazeuse

**GC/MS** : Chromatographie en phase gazeuse couplée à la spectrométrie de masse

**GS** : Glutathione synthase

**GSH** : glutathion réduit

**HCOOH** : Acide formique

**HILIC** (*Hydrophilic interaction liquid chromatography*) : Chromatographie en phase liquide basée sur des interactions hydrophiles

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

**IR** : Infrarouge

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

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

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

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

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

**MeOH** : Méthanol

**SRM** (*Single Reaction Monitoring*) : détection d'une réaction sélectionnée

**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

**NADH** (*nicotinamide adenine dinucleotide reduced*) : Adenosine 5'-(trihydrogen diphosphate)

**NADPH** (*Nicotinamide adenine dinucleotide phosphate – reduced*)

**NIST** (*National Institute of standards and technology*)

**OPLS** (*Orthogonal Projection on Latent Structures*)

**PCA** (*Perchloric acid*) : Acide perchlorique

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

**PEG** : PolyEthyleneGlycol

**PLS** (*Partial Least Square*)

**PLS-DA** (*Partial Least Square- Discriminant Analysis*)

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

**RMN** : Résonance Magnétique Nucléaire

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

**SIMS** (*secondary Ions Mass Spectrometry*)

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

**SRM** (*selective Reaction Monitoring*): Mode d'acquisition sélectif

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

**Tr** : temps de rétention

**UHPLC** (*Ultra-High Pressure Liquid Chromatography*) : chromatographie liquide à ultra haute pression

**UV** : Ultra-violet ou « Unit Variance » en analyses multivariées : centrage et réduction des données à la variance unitaire.

**VIP** : *Variable Importance in the Projection*

CHAPITRE I :

INTRODUCTION GÉNÉRALE



Le développement d'un médicament regroupe plusieurs étapes (Figure 1). Il commence par l'identification de la cible thérapeutique qui est classiquement un récepteur ou une enzyme avec lesquels le médicament devra interagir afin de modifier, corriger ou restaurer une ou des fonctions cellulaires. Une fois la cible validée, un test d'évaluation (cellulaire ou moléculaire) est mis au point pour sélectionner les composés les plus prometteurs en termes d'activité (« hits » : composés actifs sur des modèles *in vitro*, puis « leads » : composés actifs sur des modèles animaux). Cette étape, appelée le "criblage", a considérablement évolué ces 20 dernières années avec la mise en place des méthodes de criblage à très haut débit (HTS), de la chimie parallèle puis de la chimie combinatoire, et les approches *in silico*. Des tests sont ensuite pratiqués *in vitro* puis chez l'animal (pharmacologiques, pharmacocinétiques (PK), pharmacodynamiques, toxicologiques,...) afin d'étudier le comportement de la molécule *in vivo* et d'extrapoler l'innocuité relative du candidat médicament à l'homme. Cette étape correspond à la phase de développement et des études précliniques qui permettent la sélection d'un candidat médicament. Lorsque tous les résultats obtenus précédemment sont positifs, les tests cliniques chez l'homme peuvent débuter et comportent trois phases avant la mise sur le marché.

Les études de pharmacocinétique ciblent de façon classique quatre domaines qui sont l'absorption, la distribution, la métabolisation et l'excrétion (ADME). La partie métabolisme s'intéresse à l'identification des métabolites que ce soit *in vitro* ou *in vivo*, l'enzymologie des enzymes du métabolisme, ainsi que les mécanismes réactionnels et les voies métaboliques impliquées dans ces étapes. Ces études ont pour objectif d'étudier le devenir du médicament dans l'organisme, afin d'évaluer les risques toxicologiques liés à l'apparition de métabolites toxiques ou réactifs, ainsi que les risques d'interactions médicamenteuses.

Dans les années 90, la pharmacocinétique et le métabolisme d'un candidat médicament étaient la cause principale (40%) d'arrêt des programmes de R & D pharmaceutique. Avec l'introduction de tests *in vitro* dédiés à ces paramètres, dès les phases précoces de développement, cette origine de rejet était tombée à 10% en 2002<sup>1</sup>. Aujourd'hui, les données des études de métabolisme permettent de sélectionner les structures chimiques (hits) puis les « leads » démontrant des profils favorables en termes de métabolisation, de pharmacocinétique, et de sécurité, tandis que par la suite, les données issues des études précliniques facilitent la construction des études cliniques qui fourniront des données de sécurité liées à l'exposition de l'homme à la molécule en développement. Plus qu'un schéma séquentiel, comme présenté dans la Figure 1, ces étapes sont aujourd'hui fortement imbriquées.

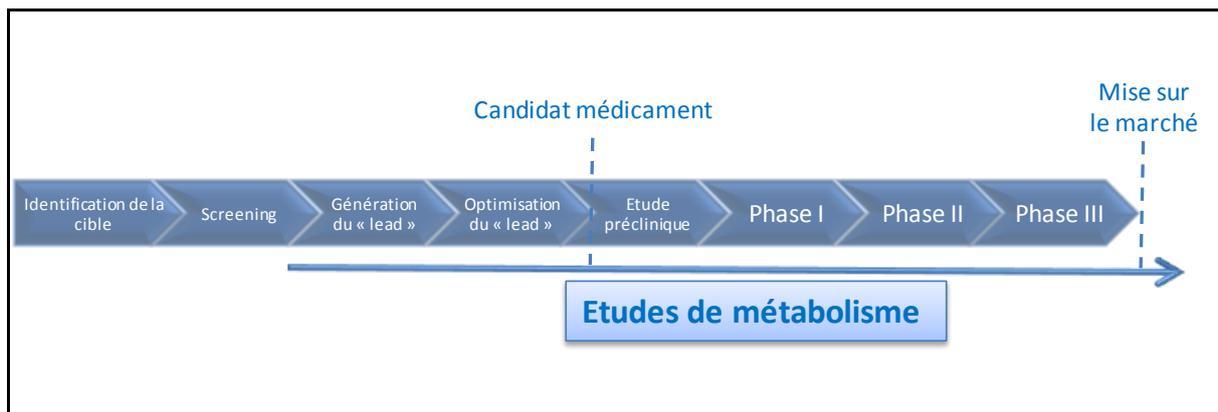


Figure 1: Etapes du développement d'un médicament et place des études de métabolisme.

L'étude du devenir métabolique d'une nouvelle molécule, éventuel candidat médicament, est devenue une étape majeure dans la recherche et le développement pharmaceutique (Figure 1). Le nombre de molécules à tester a considérablement augmenté avec les nouvelles méthodes de screening et les études de pharmacocinétiques ont dû s'adapter. Elles interviennent maintenant à un stade précoce, dès la phase de recherche alors qu'elles étaient jusqu'alors réalisées durant la phase de développement. Les coûts de développement étant par ailleurs intimement liés à l'avancement du programme de développement, et les contraintes réglementaires ne cessant de s'accroître, l'ensemble de ces facteurs conduit à la recherche de techniques de plus en plus efficaces, spécifiques et rapides.

Dans ce contexte, le CEA et Technologie Servier ont initié un partenariat de recherche à l'origine de ce travail de thèse. Ce partenariat avait pour objectif la caractérisation de biomarqueurs indirects d'induction enzymatique chez le rat, par administration à cette espèce de divers xénobiotiques, inducteurs spécifiques des différentes familles d'enzymes à cytochrome P450, impliquées dans la métabolisation des xénobiotiques. Reposant sur l'analyse du métabolome par chromatographie liquide couplée à la spectrométrie de masse, ce projet devait également proposer de nouveaux développements méthodologiques, afin d'évaluer l'impact des récentes avancées en matière d'instrumentation analytique, telles que l'UHPLC et la technologie Orbitrap<sup>®</sup>, sur les approches métabolomiques.

Après un rappel sur les voies biologiques impliquées dans le métabolisme des xénobiotiques, les facteurs pouvant les influencer et les méthodes employées à leur étude, nous brosserons un état des lieux sur les approches métabolomiques. A cette fin, nous effectuerons une revue des différentes approches existantes et des méthodes analytiques employées à leur étude. Nous nous focaliserons sur les méthodes employant la chromatographie et la spectrométrie de masse et l'ensemble des étapes que comprennent les approches globales (traitement et interprétation des données).

Dans un second temps, nous aborderons le travail de thèse à proprement parler qui comprendra quatre parties. La première sera dédiée à la comparaison des performances entre les méthodes développées au sein des deux groupes pour les approches globales, à leur optimisation et

au choix de la technique qui sera employée pour l'analyse des échantillons de l'étude. La seconde partie présentera l'application d'un outil statistique d'aide au traitement du signal obtenu par CL-MS afin d'en limiter la redondance. Dans un troisième temps, nous nous intéresserons à la mise en œuvre d'un outil de visualisation associé à une modélisation de la rétention en chromatographie liquide pour l'aide à l'identification des métabolites d'intérêt. Enfin, dans la dernière partie nous verrons l'approche retenue pour l'identification des métabolites d'intérêt, nous ferons un bilan des composés qui ont pu être identifiés au cours de ce travail et nous discuterons de leur pertinence biologique.



CHAPITRE II :  
ELEMENTS BIBLIOGRAPHIQUES



# 1 METABOLISATION DES XENOBIOTIQUES.

## 1.1 Introduction.

L'environnement nous expose quotidiennement, de façon directe ou indirecte à des substances chimiques d'origines diverses, rassemblées sous le terme de xénobiotiques. Ces composés comprennent des produits naturels d'origine végétale ou animale, des médicaments et des polluants de l'environnement tels que des dérivés de combustibles domestiques et industriels, des solvants, des additifs alimentaires, des pesticides, etc....

Généralement hydrophobes, ces molécules ont naturellement tendance à s'accumuler dans les phases lipidiques des tissus biologiques, et entraîneraient les organismes vers une mort certaine, si ces derniers ne s'étaient dotés, au cours de l'évolution, de systèmes enzymatiques permettant leur élimination. Le processus de détoxication (Figure 2) comprend trois phases regroupant différentes enzymes de biotransformation (phases I et II), ayant pour objectif d'augmenter l'hydrophilie des xénobiotiques, ainsi que des transporteurs d'efflux (phase III) qui facilitent l'excrétion du xénobiotique par voie urinaire ou biliaire.

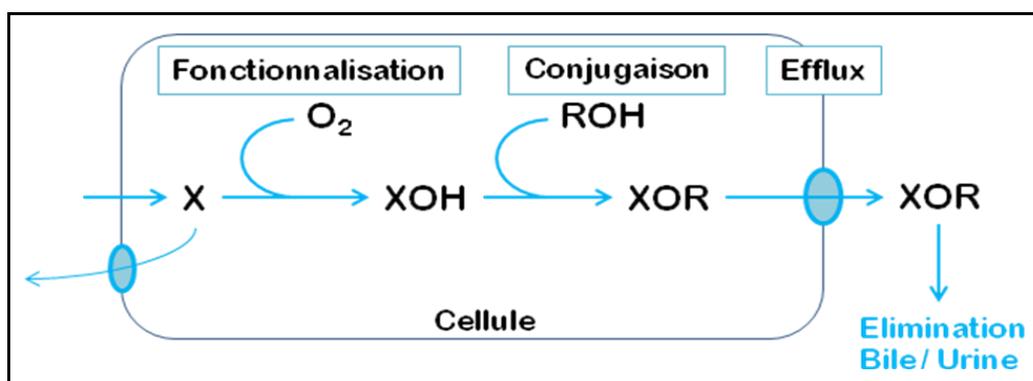


Figure 2: Schéma général du processus de détoxication

### 1.1.1 La phase I : Fonctionnalisation

Les réactions de phase I, dites de fonctionnalisation, incluent des réactions d'hydroxylation ( $R-H \rightarrow R-OH$ , R pouvant être aliphatique ou aromatique), de N-oxydation ( $R-NH-R' \rightarrow R-NOH-R'$ ), de S-oxydation ( $R-S-R' \rightarrow R-SO-R'$ ) où l'oxydation est évidente car il y a eu addition d'un atome d'oxygène, et des réactions de N- et O-déalkylation, où la fixation d'un atome d'oxygène n'a été qu'une étape intermédiaire et ne transparaît pas dans le produit final. Ces réactions sont majoritairement des réactions d'oxydation, mais elles incluent également des réactions d'hydrolyse, de réduction et de déhalogénéation. Elles sont essentiellement catalysées par les mono-oxygénases à cytochromes P450 mais font aussi intervenir les peroxydases et la quinone réductase. En outre, d'autres processus

hydrolytiques catalysés par des estérases, des amidases, des imidases, des époxyde-hydratases, etc., participent à l'augmentation de l'hydrophilie des xénobiotiques.<sup>2</sup>

### **1.1.2 La phase II : Conjugaison.**

Les réactions de phase II, dites de conjugaison, permettent l'ajout d'un radical hydrophile et sont réalisées soit directement sur le xénobiotique inchangé, soit sur ses métabolites "fonctionnalisés", issus de la phase I. Ces réactions de conjugaison augmentent l'hydrophilie de la molécule, facilitant son transport et son élimination par les voies rénale et/ou biliaire. Les radicaux hydrophiles conjugués sont de petites molécules endogènes polaires telles que le glutathion, l'acide glucuronique, le sulfate, les groupements méthyle et acétyle, et certains acides aminés comme la glycine ou la taurine. Ces réactions de conjugaison sont catalysées par différentes transférases, spécifiques de l'élément conjugué, parmi lesquelles se trouvent entre autres, les glutathion-S-transférases (GST), les UDP-glucuronosyl transférases (UGT), les sulfotransférases (SULT), ...

### **1.1.3 La phase III et ses transporteurs d'efflux.**

Si les xénobiotiques pénètrent facilement dans la cellule grâce à leur caractère lipophile, ils peuvent cependant en être expulsés grâce à des protéines d'efflux telles que la P-gp (P-glycoprotéine). Ils doivent également traverser des membranes cellulaires pour permettre leur excrétion après métabolisation. Ce rôle est assuré par les transporteurs de phase III. Ce sont essentiellement des glycoprotéines membranaires permettant le transport actif des xénobiotiques et des conjugués de la phase II hors de la cellule<sup>3</sup>. Elles participent à l'excrétion de ces molécules hors de l'organisme par les voies biliaire<sup>4</sup> et urinaire.

## **1.2 Les enzymes de phase I**

### **1.2.1 Les mono-oxygénases à cytochrome P450.**

#### **1.2.1.1 Localisation et distribution.**

Les mono-oxygénases à cytochromes P450 (CYP) sont des systèmes multienzymatiques, localisés essentiellement dans le réticulum endoplasmique et les membranes mitochondriales, dont le site actif est orienté du côté du cytoplasme<sup>5</sup>. Elles sont retrouvées essentiellement au niveau hépatique, mais également en quantité non négligeable dans les reins (zone corticale), les poumons, l'intestin, la peau, le cerveau, la vessie, les glandes surrénales...<sup>6, 7</sup> Leur expression peut être constitutive ou inductible. Le profil d'expression des isoenzymes de CYP varie entre les espèces, d'un individu à l'autre, sous l'effet de facteurs environnementaux et/ou génétiques<sup>8</sup>. Il existe aussi une distribution différente des cytochromes P450 entre les organes pour une même espèce<sup>9</sup>.

#### **1.2.1.2 Structure.**

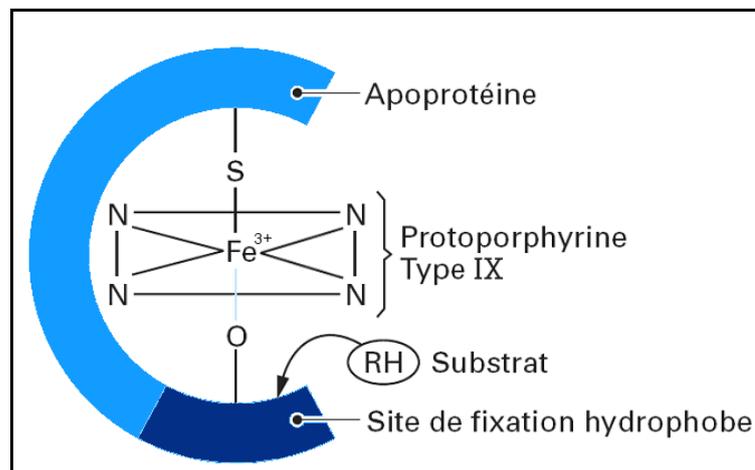
Les CYP, pour fonctionner, nécessitent d'autres protéines, chaque cycle catalytique comportant deux étapes successives de réduction monoélectronique. Dans la membrane du réticulum on trouve

deux protéines rédox qui présentent une activité cytochrome P450 réductase, une flavoprotéine: la NADPH-cytochrome P450 oxydoréductase, et un autre cytochrome: le cytochrome b5. Les différents éléments qui intègrent le système multienzymatique sont:

- Un donneur d'électrons : NADPH (nicotinamide adénine dinucléotide phosphate), ou éventuellement NADH (nicotinamide adénine dinucléotide)
- Une chaîne de transfert d'électrons comprenant :

**- le cytochrome P450 :**

C'est une hémoprotéine (Figure 3). Elle comprend une protéine monomérique d'environ 500 résidus (apoprotéine) qui ménage une cavité hydrophobe au niveau du site catalytique et un groupement prosthétique (groupement de nature non protéique d'une hétéroprotéine impliqué dans l'activité de cette dernière): une protoporphyrine IX, dont le métal est coordonné de façon permanente à la chaîne polypeptidique par la fonction thiolate d'une cystéine et de façon intermittente avec un acide aminé de la portion hydrophobe de l'apoprotéine qui permet la fixation du substrat.



**Figure 3: Structure schématique d'un cytochrome P450.** La liaison matérialisée en bleu ciel est rompue lors de la fixation du substrat sur la zone hydrophobe de l'apoprotéine.

**- la NADPH-cytochrome P450 oxydoréductase :**

C'est une flavoprotéine (protéine dont le groupement prosthétique est une flavine). Elle possède une partie hydrophobe qui la lie à la membrane du réticulum endoplasmique. Elle fait partie intégrante du système mono-oxygénase, mais dans certaines conditions, elle peut elle-même catalyser la réduction de certains xénobiotiques au niveau hépatique.

- le cytochrome b5 :

Il contribue au transfert d'électron à partir des cofacteurs NADPH et NADH. C'est une protéine héminique.

- la NADH-cytochrome b5 réductase :

C'est une flavoprotéine. Cette enzyme ne semble pas être indispensable au fonctionnement de tous les cytochromes P450<sup>10</sup>.

- la phosphatidylcholine :

Ce phospholipide agit comme transporteur d'électrons du cofacteur (NADPH) au complexe cytochrome P450-substrat et facilite l'interaction d'un substrat liposoluble à l'hémoprotéine.

1.2.1.3 Cycle catalytique de fonctionnement d'un Cytochrome P450.

Il comporte plusieurs étapes résumées dans la Figure 4. Lors de la fixation du substrat sur la zone hydrophobe de l'apoprotéine, la liaison entre un acide aminé de cette dernière et le fer de l'hème est rompue. Le fer qui reste sous forme ferrique n'est alors plus relié à l'apoprotéine que par la fonction thiolate d'une cystéine. Arrive ensuite un premier électron, transféré au CYP par la réductase, qui fait passer le fer à l'état ferreux. Le métal est alors capable de fixer l'oxygène moléculaire pour former un oxy-complexe. Un second électron est transféré au système par la réductase ou le cytochrome b5 pour conduire à un complexe oxénoïde. Enfin, tandis qu'un atome d'oxygène réagit avec les protons pour donner une molécule d'eau, le second est transféré au composé à métaboliser qui est ensuite libéré en même temps que le CYP, qui est revenu à son état initial.

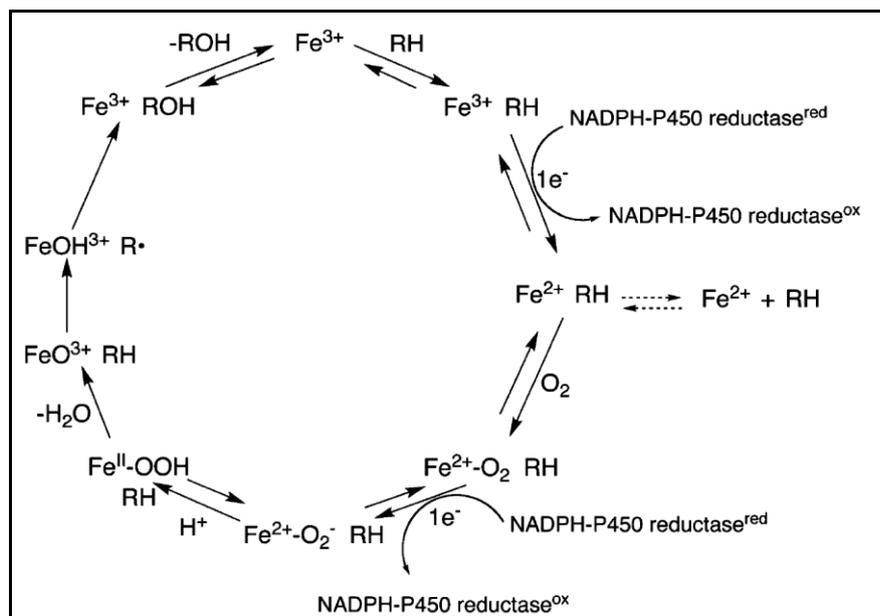


Figure 4: Cycle catalytique d'un cytochrome P450

#### 1.2.1.4 *Multiplicité et nomenclature des différentes formes de CYP450.*

De nombreux cytochromes P450 sont identifiés et caractérisés, et plus de 8500 gènes ont été décrits chez les organismes vivants (<http://drnelson.utmem.edu/CytochromeP450.html>). Chez le rat, on dénombre 89 gènes et 69 pseudogènes <sup>9</sup> (<http://drnelson.utmem.edu/rat.master.table.html>). Chez l'homme, on dénombre actuellement 18 familles de cytochromes P450, 44 sous-familles correspondant à 57 gènes et 58 pseudogènes <sup>11, 12</sup>.

Les CYPs, ayant un pourcentage d'homologie de leur séquence en acides aminés supérieur ou égal à 50%, sont regroupés en familles. Les CYPs, d'une même famille et ayant un pourcentage d'homologie de leur séquence supérieur à 59%, sont regroupés en sous familles, elles-mêmes subdivisées en isoformes. Le préfixe CYP signifie cytochrome P450 ; le premier chiffre arabe suivant renvoie à la famille, la lettre intermédiaire à la sous-famille et le dernier chiffre arabe à l'isoforme.

#### 1.2.1.5 *Rôles biologiques*

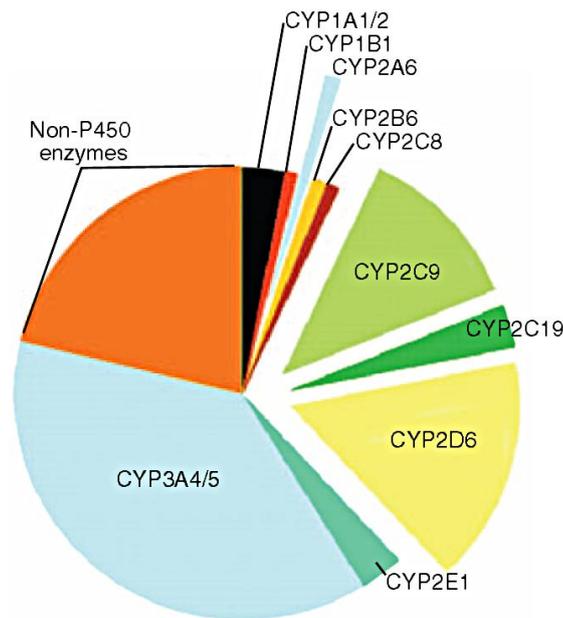
Les CYPs métabolisent des substrats endogènes comme les stéroïdes, les dérivés de l'acide arachidonique, les acides biliaires, certaines vitamines, ainsi que de très nombreux xénobiotiques. Le Tableau 1 présente les différentes familles de cytochromes P450 et leurs principaux rôles connus chez l'homme.

**Tableau 1: Familles de cytochrome P450 et principaux substrats et fonctions**

<b>Famille</b>	<b>Substrats</b>
CYP1	Xénobiotiques, Acide arachidonique, eicosanoïdes
CYP2	Xénobiotiques, Acide arachidonique, eicosanoïdes
CYP3	Xénobiotiques, Acide arachidonique, eicosanoïdes
CYP4	Acide arachidonique, eicosanoïdes, Acides gras
CYP5	Thromboxane A2 synthase
CYP7	Cholestérol, synthèse des acides biliaires
CYP8	Synthèse des prostaglandines et acides biliaires
CYP11	Stéroïdogénèse
CYP17	Stéroïde-17 $\alpha$ -hydrolase et 17/20 liase
CYP19	Aromatase (œstrogènes)
CYP20	Inconnu
CYP21	Stéroïde-21-hydroxylase
CYP24	Vitamine D <sub>3</sub> 24-hydroxylase
CYP26	Hydroxylation de l'acide rétinoïque
CYP27	Biosynthèse des acides biliaires, hydroxylation vitamine D <sub>3</sub>
CYP39	24-hydroxycholestérol-7 $\alpha$ -hydroxylase
CYP46	Cholestérol 24-hydroxylase
CYP51	Lanostérol 14 $\alpha$ -desméthylase

- **Implication dans la métabolisation des xénobiotiques.**

En présence de substrat exogène, les CYPs catalysent un grand nombre de biotransformations (décrites au 1.1.1). Celles-ci se distinguent principalement par l'activité biologique éventuelle du produit obtenu. Dans la plupart des cas, ce dernier en est fort heureusement dépourvu, et sa formation permet de faciliter l'élimination du composé initial par l'ajout de groupements polaires. Une étape de conjugaison vient éventuellement parachever ce mécanisme de détoxification, initié par les CYPs. Les CYPs, majoritairement impliqués dans la métabolisation des xénobiotiques au niveau du foie, font partie des familles 1, 2 et 3 (Figure 5).



**Figure 5: Contribution des cytochromes P450 humains à la phase I du métabolisme des médicaments actuellement mis sur le marché <sup>13</sup>.**

Le produit de biotransformation, appelé métabolite, est le plus souvent un composé biologiquement inactif dont l'excrétion rénale ou biliaire est favorisée. En ce sens, la biotransformation est un processus de détoxification. Toutefois, le métabolisme des xénobiotiques ne conduit pas toujours à l'inactivation de ces composés. Dans certains cas, le produit oxydé ou réduit peut manifester une très forte réactivité chimique qui se manifeste par des attaques électrophiles sur les macromolécules nucléophiles environnantes (protéines, acides nucléiques). Ainsi la formation d'adduits sur l'ADN peut entraîner l'apparition de mutations ponctuelles et initier des processus de cancérogenèse. La formation d'adduits aux protéines peut entraîner une nécrose tissulaire et des phénomènes d'allergies par immuno-allergisation <sup>14</sup>. Le schéma général présentant les devenir potentiels d'un xénobiotique est présenté sur la Figure 6.

### 1.2.2 Les autres enzymes de phase I.

- **Carboxyestérases**

Ce sont des enzymes microsomales ou cytosoliques qui sont retrouvées dans le foie, les reins, les testicules, les poumons, les globules rouges. Elles hydrolysent les esters, les amides; les thioesters, les esters d'acides phosphoriques et les anhydrides d'acide. Elles réalisent également des réactions de trans-estérification.

- **Peptidases**

Dans le sang et les tissus, on trouve des amino- et des carboxypeptidases qui clivent les peptides ou les protéines.

- **Oxydation des alcools et des aldéhydes**

Les alcools, les aldéhydes et les cétones sont oxydés ou réduits par des enzymes, incluant les alcools déshydrogénases (ADH), les aldéhydes déshydrogénases (ALDH).

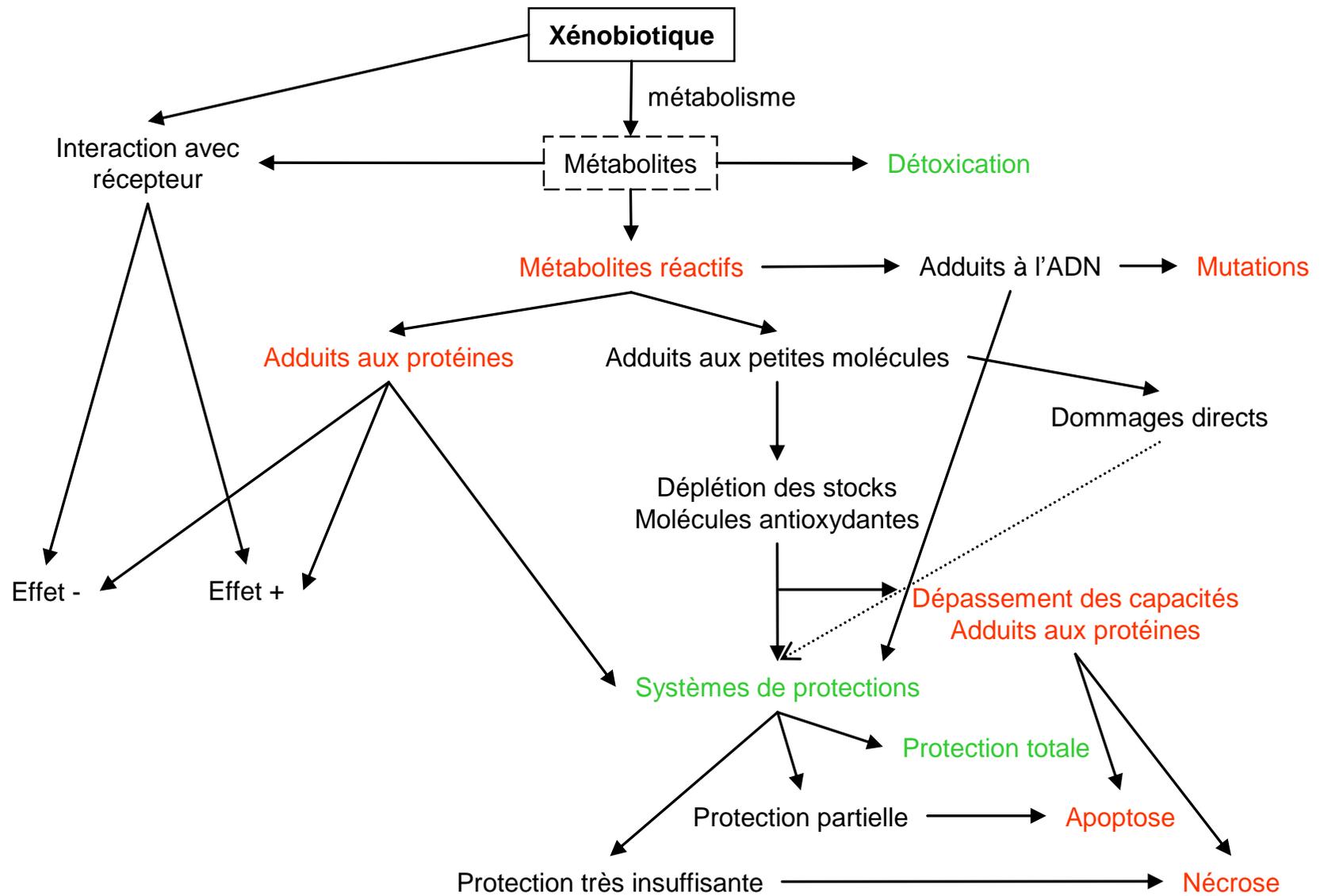


Figure 6: Devenir d'un xénobiotique au cours du processus de métabolisation

○ Les ADH sont des enzymes cytosoliques et mitochondriales, de structure dimérique, présentes dans plusieurs tissus: foie, rein, poumon, muqueuse gastrique. Il existe de nombreuses variations génétiques, trois gènes codant pour des sous-unités  $\alpha$ ,  $\beta$  1, 2, 3,  $\gamma$  1, 2. Elles utilisent le NAD (nicotinamide adénine dinucléotide) comme cofacteur.

○ Les aldéhydes sont en général toxiques. Leur oxydation conduit à l'acide correspondant, par l'intermédiaire des aldéhydes déshydrogénases à NAD. Il existe actuellement dix gènes connus codant pour les ALDH. Les principales enzymes, ALDH1 et ALDH3, sont cytosoliques, ALDH2 est mitochondriale et polymorphe. Elle est responsable du phénomène de « flush » observé chez les populations du sud-est asiatique, qui possèdent une forme peu active de cette enzyme conduisant, suite à l'absorption d'alcool, à l'accumulation d'aldéhyde qui ne peut alors pas être métabolisé en acide.

- **Peroxydases**

Cette classe comprend la prostaglandine H synthétase (PGHS) aussi appelée cyclooxygénase (COX) et les 5-, 12- et 15- lipoxygénases (LIPOX).

Il existe deux isoformes de la COX :

- la COX 1 (ou PGHS-1) qui est exprimée constitutivement dans la plupart des tissus de mammifères où elle participe au maintien de l'homéostasie cellulaire<sup>15</sup>.
- la COX 2 (ou PGHS-2) qui n'est exprimée constitutivement que dans certains tissus, comme le cerveau, la prostate, la moelle épinière<sup>16</sup>, les reins<sup>17</sup> et quelques tissus pathologiques comme les cellules cancéreuses du colon<sup>18</sup>. Cette isoforme est inductible.

Les lipoxygénases sont impliquées dans trois voies distinctes, à savoir : la transduction du signal (via la cascade de l'acide arachidonique), les réactions de peroxydation et la mobilisation de lipides<sup>19</sup>.

Certains xénobiotiques sont oxydés *in vitro* par la cyclooxygénase en présence d'acide arachidonique, et de peroxyde d'hydrogène ou de peroxydes lipidiques. La COX sert d'enzyme d'activation métabolique en complément ou en remplacement des mono-oxygénases à cytochromes P450 dans des réactions de bioactivation *in vivo*, surtout dans les tissus où l'activité de ces CYPs à fonction mixte est faible<sup>20</sup>. De nombreux composés chimiques peuvent être oxydés soit directement par la peroxydase de la COX, soit par des radicaux peroxydes qu'elle forme au cours de réactions avec d'autres composés. Son implication dans l'oxydation, par des radicaux peroxydes, d'hydrocarbures aromatiques polycycliques a ainsi été rapportée<sup>21</sup>.

La bioactivation de xénobiotiques liée à l'activité peroxydase de la COX reste le cas le plus courant. Plusieurs amines aromatiques ou phénols tels que le paracétamol <sup>22</sup>, le diethylstilbestrol <sup>23</sup> et l'hydroquinone (métabolite du benzène) <sup>24</sup> sont oxydés *in vitro* de la sorte.

Les LIPOX peuvent utiliser le glutathion réduit (GSH) comme substrat pour leur activité oxydase, générant la formation de xénobiotiques conjugués au glutathion<sup>25</sup>.

- **Flavines monooxygénases**

Les flavines monooxygénases (FMO) du foie, des reins et du poumon se trouvent au niveau des microsomes, fonctionnent en présence de NADPH et d'oxygène, et sont responsables de réactions d'oxydation qui peuvent aussi être catalysées par les CYPs. Les FMO catalysent l'oxydation des amines primaires, en N-oxydes, et des amines secondaires en hydroxylamine. Elles oxydent aussi les composés renfermant du soufre, tels que les thiols. Ces réactions de détoxification impliquent un intermédiaire peroxyde. Il existe cinq formes de FMO (FMO 1 à 5).

### 1.3 Les Réactions de phase II

- **Glucuroconjugaison** <sup>26</sup>

Principale réaction de conjugaison chez l'homme, elle conjugue le métabolite formé en phase I avec l'acide glucuronique, à partir de l'acide UDP-glucuronique (UDP : uridine-diphosphate) (Figure 7). Les atomes sur lesquels s'effectue la conjugaison sont l'oxygène d'un hydroxyle ou d'un carbonyle du substrat, ou parfois le soufre ou l'azote. Les substrats peuvent être des xénobiotiques ou des substances endogènes (stéroïdes, bilirubine).

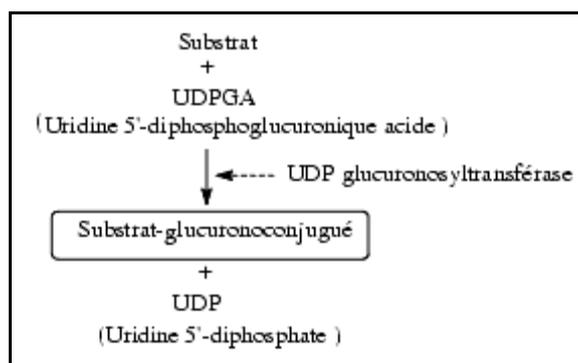


Figure 7: Glucuroconjugaison

Les UDP-glucuronosyl-transférases (UGT) sont une famille d'enzymes microsomales, hépatiques et rénales principalement. La conjugaison avec l'acide glucuronique implique une attaque nucléophile sur l'atome de carbone C<sub>1</sub> de l'acide glucuronique, de sorte que les glucuronides obtenus sont généralement de configuration bêta.

Il existe une multitude d'isoformes de spécificité variable. La famille UGT1 comporte les enzymes qui conjuguent la bilirubine, les quinones, les phénols, notamment ceux provenant du métabolisme des hydrocarbures aromatiques polycycliques<sup>27</sup>. La famille UGT2A comprend au moins un gène spécifique de la muqueuse nasale. La famille UGT2B comprend les gènes inductibles par le phénobarbital, la dioxine et d'autres hydrocarbures aromatiques polycycliques, les proliférateurs des peroxyosomes, les stéroïdes contraceptifs.

Ces enzymes présentent une grande variabilité d'activité. Kaniwa et al. ont étudié les différences ethniques de polymorphisme pour l'UGT1A1 au sein de populations africaines, caucasiennes et japonaises. A partir de 150 individus de chaque population, ils ont pu mettre en évidence dans la population afro-américaine un nouveau variant faiblement actif. De plus, les différences dans les profils d'haplotype suggèrent des variations de toxicité des molécules prises en charge par l'UGT1A1 selon les populations<sup>28</sup>. En outre, les UGTs sont peu matures chez le nouveau-né.

- **Sulfoconjugaison**

La formation d'esters sulfates est une voie importante de conjugaison. La sulfoconjugaison utilise les mêmes substrats, porteurs d'un groupement hydroxyle ou amine, que la glucuronoconjugaison.

Le donneur de sulfate est le 3'-phosphoadénosine-5'-phosphosulfate (PAP), formé à partir du sulfate minéral et de l'ATP<sup>29</sup>.

Les sulfotransférases sont une famille d'enzymes présentes dans la fraction soluble de différents tissus, dont le foie, l'intestin, et le rein. Dans l'espèce humaine, on dénombre 5 sulfotransférases cytosoliques. Si ces dernières ont une spécificité de substrat large, elles possèdent néanmoins des ligands préférentiels. Parmi les 5 sulfotransférases humaines, on trouve ainsi une première plus spécifique des estrogènes, une autre conjuguant les hydroxystéroïdes et trois phénolsulfotransférases<sup>30</sup>. L'équilibre entre glucurono- et sulfoconjugaison se fait en fonction de la disponibilité en sulfate, de la concentration du toxique, et de l'organe (le cerveau, par exemple, possède une forte activité sulfotransférase).

Elles sont impliquées dans l'inactivation de composés exogènes aussi bien qu'endogènes tels que les stéroïdes, les hormones thyroïdiennes, et les neurotransmetteurs<sup>31</sup>. Les composés sulfatés peuvent être éliminés de la cellule ou être stockés afin de pouvoir être réactivés sous l'action de sulfatases et utilisés *a posteriori*<sup>32</sup>.

- **Méthylation**

Elle s'effectue sur les groupements, amino, hydroxyle ou thiol par l'intermédiaire de méthyltransférases, demandant de la S-adénylméthionine qui est le groupement donneur de méthyle.

Les enzymes appartiennent majoritairement au cytosol et quelques-unes au réticulum endoplasmique. Une des méthyltransférases les plus connues est la catéchol-O-méthyl-transférase (COMT) qui est responsable de l'inactivation des catécholamines.

La méthylation est une voie de faible importance pour les xénobiotiques bien qu'elle joue un rôle fondamental dans la métabolisation de certains anticancéreux comme les thiopurines qui sont méthylées par une thiopurine S-méthyltransférase (TMST) <sup>33</sup>.

- **Acétylation**

L'acétylation est la voie de conjugaison principale pour des composés aminés, amines aromatiques ou aliphatiques ou des hydrazines. Le cofacteur de cette réaction est l'acétylcoenzyme A. Il nécessite l'ATP pour sa synthèse <sup>34</sup>.

Les N-acétyltransférases (NAT1 et NAT2), responsables de ces réactions sont des enzymes du cytosol de la cellule hépatique, de la muqueuse gastro-intestinale et des leucocytes. Leur masse moléculaire est de l'ordre de 37 000. La NAT1 se trouve dans tous les tissus et la NAT2 dans le foie et le tube digestif.

L'acétylation chez l'homme présente des variations individuelles importantes, d'origine génétique. Les NATs font ainsi partie des exemples les plus anciens et les mieux connus de polymorphismes génétiques.

Le polymorphisme de NAT2 (une quinzaine de variants alléliques) se traduit par l'expression de phénotypes « acétyleurs lents » et « acétyleurs rapides ». De très grandes variations interethniques dans la fréquence des acétyleurs lents sont décrites <sup>35, 36</sup>. Ce polymorphisme est à l'origine de nombreux effets indésirables ou toxiques lors de l'administration de médicaments. En effet, l'acétylation peut survenir après qu'un groupement amine a été oxydé par un CYP, conduisant au transfert de l'acétyle sur un oxygène (O-acétylation) <sup>37, 38</sup>. Si la N-acétylation est un mécanisme de détoxication, l'O-acétylation génèrent des intermédiaires nitrénium réactifs qui peuvent formés des adduits mutagènes à l'ADN <sup>39</sup>.

- **Conjugaison aux acides aminés**

Au contraire des réactions précédentes, le substrat doit être activé avec le coenzyme A. L'acide aminé mis en jeu le plus souvent est la glycine, dans une moindre mesure la glutamine et la taurine. Les acides carboxyliques sont des substrats de cette conjugaison. L'équilibre avec la glucuronoconjugaison des acides carboxyliques se fait en fonction de la masse moléculaire et de l'encombrement de la molécule. Cette conjugaison s'effectue dans les fractions mitochondriales et cytoplasmiques du foie et du rein.

- **Conjugaison au glutathion**

Cette réaction ne demande activation ni du substrat, ni du donneur, car le caractère nucléophile du glutathion suffit à le conjuguer avec des substrats électrophiles. Les réactions sont catalysées par des glutathion-S-transférases, famille d'enzymes de spécificité plus ou moins restreinte et chevauchante. Les enzymes se trouvent dans le cytoplasme, quelques-unes sont localisées au niveau de la membrane microsomale <sup>40</sup>. Les substrats sont électrophiles, ce sont des hydrocarbures aromatiques halogénés, des époxydes, des composés nitro-aromatiques et des alcènes. Certains de ces composés sont éliminés à l'état de conjugués à la N-acétylcystéine, également nommés « acides mercapturiques » (Figure 8). Ces conjugués sont le résultat d'une conjugaison initiale au glutathion suivie d'un clivage métabolique des résidus glutamyl (par la  $\gamma$ -glutamyltransférase,  $\gamma$ -GT) et glyciny (par une peptidase) puis d'une acétylation du reste cystéinyl par l'acétyltransférase (Figure 8). Il en existe sept classes : alpha, mu, pi, sigma, théta, kappa et zéta<sup>41</sup>. Les GST sont inductibles, par exemple, le phénobarbital et le 3-méthylcholanthrène augmentent l'expression des GST de classe alpha.

Les GSTs sont aussi un moyen de protection de l'organisme contre les dérivés du stress oxydant de par leur activité peroxydase qui transforment les peroxydes en composés moins réactifs<sup>42</sup>.

Plusieurs GST présentent un polymorphisme génétique chez l'homme.

La conjugaison au glutathion est initialement un processus détoxifiant, mais dans certains cas elle induit une augmentation de la toxicité des xénobiotiques, comme certains procarcinogènes qui sont activés par cette voie <sup>43</sup>. Quatre mécanismes peuvent en être à l'origine <sup>44</sup>. Ils entraînent la formation de :

- dérivés conjugués au glutathion d'alcènes halogénés, de thiocyanates et de nitrosoguanidine,
- dérivés d'alcane dihalogénés qui forment des ions episulfonium électrophiles,
- dérivés d'alcènes halogénés qui sont dégradés en métabolites toxiques par la  $\beta$ -lyase dans le rein,
- dérivés de quinones, quinone-imines ou isothiocyanates qui sont dégradés en métabolites toxiques par la  $\gamma$ -glutamyltranspeptidase dans le rein.

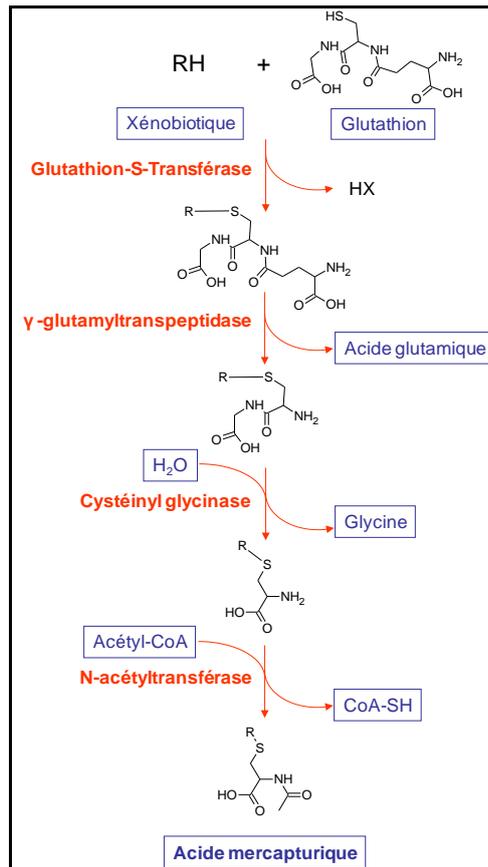
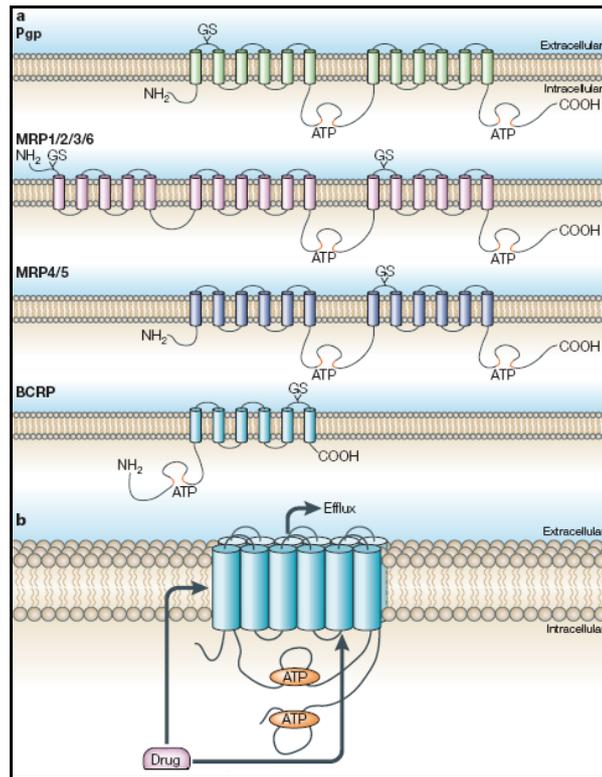


Figure 8: Conjugaison au glutathion et synthèse d'acide mercapturique

## 1.4 Les transporteurs de phase III

Si les xénobiotiques pénètrent facilement dans la cellule grâce à leur caractère lipophile, leur élimination nécessite l'intervention de pompes d'efflux. Ils doivent également traverser des membranes cellulaires pour permettre leur excrétion après métabolisation. Cette étape est assurée par les transporteurs de phase III, parmi lesquels les transporteurs ABC (« *ATP Binding Cassette* ») jouent un rôle prépondérant.

Cette superfamille est l'une des plus larges et comprend les protéines qui comptent parmi les plus exprimées. Chez l'homme, 49 gènes ABC sont classés en 7 familles (A à G)<sup>45-47</sup>. La plupart des transporteurs ABC requiert la liaison d'une molécule d'ATP (Adénosine Tri Phosphate) et son hydrolyse pour produire l'énergie nécessaire à la translocation du substrat à travers la membrane cellulaire. Les séquences en acides aminés de ces transporteurs ABC comportent des parties conservées comme le site de liaison à l'ATP incluant une lysine, un site d'interaction à Mg<sup>2+</sup> incluant un acide aspartique, et un motif spécifique à cette superfamille qui est une séquence conservée de 6 acides aminés situés entre les 2 sites précédemment cités<sup>48-50</sup>. Ce dernier motif aurait peut-être un rôle dans la reconnaissance, la liaison et l'hydrolyse de l'ATP<sup>51</sup>.



**Figure 9: Structure schématique des transporteurs de type ABC<sup>52</sup>**

**a/** modèles bidimensionnels des structures de la P-gp, des MRP et de la BCRP. Le modèle met en évidence les segments transmembranaires, les sites de glycosylation (GS), et les sites de liaison à l'ATP (Adénosine Tri-Phosphate). **b/** En réalité, les domaines transmembranaires sont rapprochés pour former une molécule plus compacte, la prise en charge du xénobiotique pouvant se faire à partir du cytosol ou de la membrane lipidique comme matérialisé par les flèches.

Les transporteurs ABC comprennent plusieurs familles de transporteurs (Figure 9), dont notamment la famille MDR (MultiDrug Resistance) ou ABCB (*ATP Binding Cassette sub-family B*) qui est la famille de la P-glycoprotéine (P-gp), la famille des MRP (*Multidrug Resistance Protein*) ou ABCC (*ATP Binding Cassette sub-family C*) et la BCRP (*Breast Cancer Resistance Protein*) ou ABCG2 (*ATP Binding Cassette sub-family G member 2*).

Ils sont, d'une manière générale, impliqués dans le transport des stérols et des acides gras <sup>53, 54</sup> ; mais peuvent avoir des rôles plus spécifiques selon le type cellulaire, comme la P-glycoprotéine qui est impliquée dans le transport des cytokines (IL-2, IL-4 et INF) chez les lymphocytes T périphériques normaux <sup>53</sup>.

Ces transporteurs sont aussi largement étudiés du fait de leur implication dans la biodisponibilité des médicaments, puisqu'ils sont responsables de l'efflux de ces derniers des tissus <sup>52, 55</sup>, ils protègent donc la cellule et les tissus contre les xénobiotiques <sup>56</sup>. La P-gp et la BCRP effluent préférentiellement de larges molécules hydrophobes chargées positivement tandis que les transporteurs de la famille MRP préfèrent les molécules hydrophobes non chargées et les composés anioniques solubles dans l'eau <sup>56</sup>.

Ils sont retrouvés au niveau des intestins, du foie, des reins, de la barrière hémato-encéphalique et de la barrière hémato-testiculaire. Ils sont ainsi présents dans les tissus ayant une fonction excrétoire. Les transporteurs ABC, prenant en charge l'excrétion des xénobiotiques, c'est-à-dire principalement la P-gp, la MRP2 et la BCRP, sont localisés à la face luminaire des cellules bordant les canalicules biliaires, les néphrons et la lumière intestinale. Les MRPs 1, 3, 4, et 6 sont principalement retrouvées sur les membranes baso-latérales où elles permettent la sécrétion des composés organiques dans le sang.

Parmi les autres familles de protéines responsables du transport des molécules chimiques, il faut citer les transporteurs SLC (« *SoLute Carrier* ») qui comprennent 43 familles et 320 membres qui ne seront pas détaillés ici. Ils sont impliqués dans l'absorption intestinale des xénobiotiques depuis la lumière intestinale vers le cytoplasme de l'entérocyte mais ils participent aussi à l'élimination de ces substances en permettant leur passage depuis le sang vers les hépatocytes ou les cellules rénales du tubule proximal. Ils sont donc situés sur les membranes baso-latérales et les faces cellulaires lumineuses. Ils sont multispécifiques comme en témoignent les membres d'une famille : « les transporteurs polypeptidiques des anions organiques ou Oatp (de l'anglais *Organic anion transporting polypeptide* » qui peuvent prendre en charge à la fois des acides organiques faibles, des composés neutres et même quelques espèces cationiques.

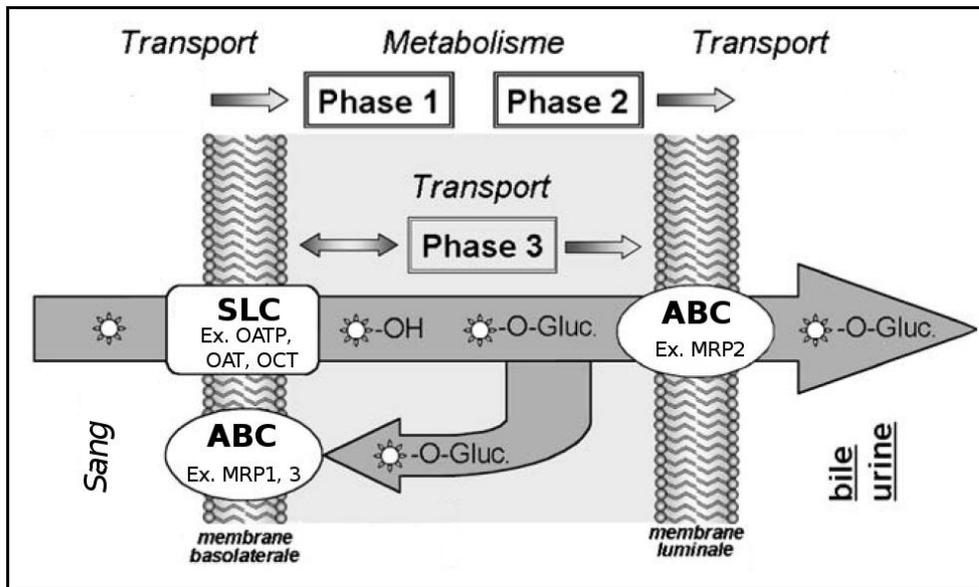


Figure 10: Transports hépatique et rénal des xénobiotiques

Traduit et adapté de Petzinger et Geyer.<sup>57</sup>

Le rôle des transporteurs dans la biodisponibilité des xénobiotiques et de certaines molécules endogènes n'a cessé de prendre de l'importance ces 15 dernières années. Il est aujourd'hui clair qu'ils sont responsables à la fois de l'absorption et de l'élimination des médicaments et autres composés chimiques (Figure 10) dans de nombreux tissus et qu'ils pourraient être les facteurs déterminants des caractéristiques pharmacocinétiques d'une molécule chimique. Ils ont ainsi permis d'expliquer, entre autre, les différences inter-espèces d'excrétion biliaire.

## 1.5 Induction enzymatique

### 1.5.1 Définition

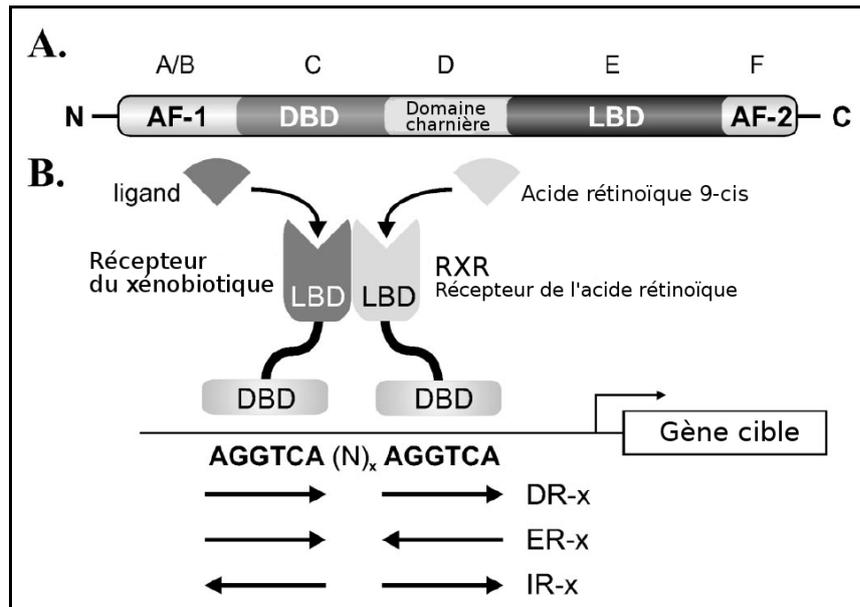
L'induction enzymatique correspond à l'augmentation de l'activité des enzymes du métabolisme. A l'inverse de l'inhibition enzymatique qui est une réponse immédiate impliquant un mécanisme de liaison compétitive, ou la formation de substrat suicide qui se fixe sur l'apoprotéine ou l'hème et conduit à l'inactivation du CYP, ou enfin à la formation de complexe réversibles ou non sur le fer, l'induction est un processus de régulation relativement lent qui requiert l'activation de la transcription du gène codant pour le CYP induit ou la stabilisation de son ARNm ou de la protéine elle-même.

### 1.5.2 Mécanismes

#### 1.5.2.1 *Activation de la transcription du gène codant pour l'enzyme*

La voie principale d'induction des enzymes du métabolisme utilise un mécanisme d'activation de la transcription de leurs gènes faisant intervenir des récepteurs nucléaires. Les récepteurs nucléaires sont des facteurs de transcription activés par la fixation d'un ligand. Au sein de ces récepteurs, certains ont pour ligand les stéroïdes, les hormones thyroïdiennes, les métabolites actifs des vitamines A et D et d'autres sont appelés récepteurs orphelins en l'absence de substrats physiologiques identifiés <sup>58</sup>.

Les récepteurs nucléaires possèdent une structure commune qui comprend un domaine central de liaison à l'ADN et un domaine C-terminal de liaison au ligand qui possède des sous-domaines permettant la dimérisation du récepteur, sa translocation nucléaire et la fixation de co-régulateurs <sup>59</sup> (Figure 11).



**Figure 11: Structure d'un récepteur nucléaire**

Traduit de Handschin <sup>60</sup>

**A.** Structure d'un récepteur nucléaire, comprenant 4 domaines : i) une région hypervariable N-terminale qui pour certains récepteurs nucléaires comprend une région activatrice (AF-1), ii) un domaine de liaison à l'ADN (DBD) constitué de 2 motifs en doigt de zinc, iii) un domaine charnière flexible, et iv) le domaine de liaison au ligand (LBD) qui inclut une région activatrice (AF-2).

**B.** Liaison du récepteur à l'ADN. Le récepteur du xénobiotique forme un hétérodimère avec RXR (récepteur de l'acide rétinoïque) afin d'obtenir une répétition de l'hexamère AGGTCA –qui permet la fixation à l'ADN– avec un espacement variable. L'arrangement des 2 hexamères peut être de 3 sortes : i) répétition directe (DR), ii) en miroir vers l'intérieur(ER) ou iii) en miroir vers l'extérieur (IR).

#### 1.5.2.1.1 Les récepteurs PXR et CAR

Les récepteurs nucléaires PXR (*Pregnane X Receptor*, NR112) et CAR (*Constitutive Androstane Receptor*, NR113) font partie de la sous famille des récepteurs nucléaires aux stéroïdes de type 2. Cette sous famille partage les caractéristiques générales des autres récepteurs nucléaires et tous ses membres possèdent comme récepteur d'hétérodimérisation commun le récepteur à l'acide rétinoïque (RXR). PXR et CAR sont activés par les xénobiotiques et jouent le rôle de régulateurs clefs des phases I à III du métabolisme impliquées dans la détoxification et l'élimination des stéroïdes, des acides biliaires et des xénobiotiques. En termes de spectre d'activité ces récepteurs sont responsables de l'induction au niveau de la phase I de plusieurs CYPs, au niveau de la phase II d'enzymes telles que les UGTs, les GSTs, et les SULTs <sup>61</sup>, et au niveau de la phase III de plusieurs transporteurs de la famille ABC dont MRP2 et MDR1 (P-gp) <sup>62, 63</sup> (Figure 12).

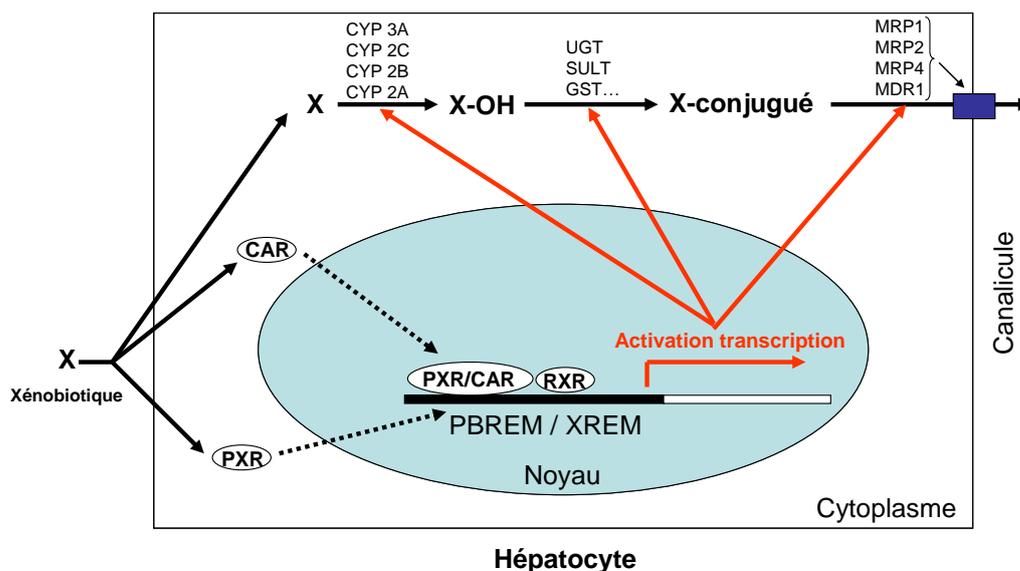


Figure 12: Implication des récepteurs PXR et CAR dans la détoxification des xénobiotiques

Traduit et adapté de Kretschmer et Baldwin<sup>64</sup>

PXR est activé par les prégnanes, la progestérone, et les glucocorticoïdes tandis que l'activité de CAR est régulée par les métabolites du noyau androstane, les œstrogènes, et la progestérone. Ces deux récepteurs possèdent donc une spécificité de ligand large et chevauchante formant un système plus ou moins redondant de protection de l'organisme. A l'inverse des autres récepteurs nucléaires, CAR possède une activité constitutive en l'absence de ligand que ce soit chez l'homme ou la souris<sup>65, 66</sup>.

Au niveau des CYPs, PXR et CAR sont responsables de l'induction de plusieurs membres des familles 1 à 3 qui sont impliquées dans le métabolisme des xénobiotiques. Le CYP3A et le CYP2B étant utilisés comme marqueurs respectifs de l'activité de PXR et CAR.

PXR et CAR régulent également les acteurs des phases II et III du métabolisme de façon chevauchante mais distincte<sup>64</sup>. Ainsi, CAR et PXR sont tous deux responsables de l'induction des GST<sub>α1</sub> et μ1, de l'UGT 1A1, de SULT 2A1 et des transporteurs MRP2 et MRP3. Cependant, PXR seul régule l'expression des transporteurs de phase III MDR1 (P-gp) et Oatp2 et CAR régule spécifiquement l'expression de l' amino N-sulfotransférase, de la 3'-phosphoadénosine-5'-phosphosulfate synthase (PAPS) qui synthétise le donneur de sulfate (PAP), et du transporteur de phase III MRP1<sup>67</sup>. Malgré ces spécificités de régulation, ces récepteurs fournissent à l'organisme deux mécanismes, en partie redondants, pour la reconnaissance et l'élimination des molécules toxiques *via* une activation concomitante par certains ligands communs et un profil d'activation, partiellement chevauchant.

### 1.5.2.1.2 Le récepteur AhR

Le récepteur Ah est une protéine de type hélice-boucle-hélice qui n'appartient pas à la même famille que PXR et CAR. Le récepteur Ah, de localisation cytosolique est activé par fixation d'un ligand qui peut être un hydrocarbure polycyclique aromatique ou des médicaments comme l'oméprazole. Le récepteur activé est ensuite transloqué dans le compartiment nucléaire où il se lie avec le facteur de translocation nucléaire ARNT (« AhR nuclear translocator ») pour former un hétérodimère. Ce dernier va pouvoir se lier aux séquences promotrices de l'ADN (éléments de réponse aux xénobiotiques (XRE)), en amont des gènes codants pour les CYP1, et les autres gènes régulés via AhR, afin de stimuler la transcription de ces gènes cibles. L'ensemble de ce processus est conservé dans de nombreux types cellulaires et à travers les différentes espèces. Il participe à l'induction des CYP1 sous l'effet des hydrocarbures polycycliques aromatiques dont de nombreux carcinogènes environnementaux retrouvés dans les fumées de combustion et de cigarette. Le mécanisme de l'induction des CYP1 et le rôle du couple AhR-ARNT sont relativement bien compris et ont été largement rapportés.

### 1.5.2.2 Stabilisation de l'ARNm de l'enzyme ou de l'enzyme elle-même

Ces voies d'activation sont rencontrées pour certains cytochromes dont l'un des mieux décrits est le CYP2E1<sup>68</sup>. Ce dernier voit ainsi son ARN<sub>m</sub> stabilisé grâce à la fixation réversible d'une protéine qui reste inconnue à ce jour. Le mécanisme d'induction principal de ce cytochrome fait intervenir une stabilisation de la protéine elle-même. En effet, en l'absence de ligand, le CYP2E présente deux cinétiques de dégradation, laissant supposer la présence de deux formes de l'enzyme, bien qu'elles ne soient pas expérimentalement différenciables. La première révèle une demi-vie de 32 heures tandis que celle de la seconde est de 7 heures. En présence de ligand, seule la cinétique à demi-vie longue est retrouvée.

## 1.6 Facteurs influençant le métabolisme.

La susceptibilité de différentes personnes aux interactions médicamenteuses peut dépendre d'une grande variété de facteurs, notamment :

**Génétique** : De nombreuses enzymes qui métabolisent les médicaments, par exemple, CYP2D6, CYP2C19, CYP2A6, CYP2C9 et la N-acétyltransférase, font l'objet de polymorphisme génétique à un point tel que la capacité de métaboliser les substrats de ces enzymes varie considérablement entre les personnes<sup>69, 70</sup>. Certains polymorphismes génétiques sont plus fréquents dans certains groupes ethniques. Par exemple, de 5 à 10 % de la population caucasienne métabolise mal les substrats du CYP2D6, contre seulement de 1 à 2 % de la population d'origine asiatique. À l'inverse, l'incidence des personnes qui métabolisent mal les substrats de l'enzyme CYP2C19 varie de 18 à 22% chez les asiatiques mais seulement de 2 à 6% chez les caucasiens<sup>69-71</sup>.

**Age** : L'amplitude et les conséquences cliniques des interactions médicamenteuses de nature pharmacocinétique peuvent dépendre de l'âge. Chez le nouveau-né, le métabolisme hépatique et l'excrétion rénale des médicaments sont réduits, car les fonctions hépatique et rénale ne sont pas encore tout à fait matures. La susceptibilité des personnes âgées à l'égard des interactions médicamenteuses peut être attribuable à des modifications de l'absorption, du métabolisme hépatique, de la clairance rénale ou du volume de distribution qui sont liées à l'âge <sup>72</sup>.

**Sexe** : Les études effectuées sur certains médicaments ont laissé entrevoir des différences dans les paramètres pharmacocinétiques ou pharmacodynamiques entre les hommes et les femmes qui évoquent des différences liées au sexe quant à la biodisponibilité, au volume de distribution, à l'activité des enzymes qui métabolisent les médicaments, à la clairance rénale ou aux caractéristiques physiologiques. La normalisation des paramètres pharmacocinétiques en fonction du poids corporel est souhaitable lorsqu'il s'agit d'évaluer les différences apparentes entre hommes et femmes au cours d'études sur les interactions médicamenteuses <sup>73</sup>.

**État de santé** : L'effet de l'atteinte hépatique sur l'ampleur et les conséquences des interactions médicamenteuses peut être complexe et difficile à prédire. Les patients qui présentent une atteinte rénale peuvent également être plus vulnérables aux interactions métaboliques des médicaments, puisque la contribution de la composante excrétoire du processus d'élimination est moins importante.

**Consommation de tabac et d'alcool** : La fumée de tabac induit les CYP1A1 et CYP1A2, et peut-être même le CYP2E1 <sup>74</sup>. La consommation d'alcool sur une longue période provoque l'induction du CYP2E1 <sup>75</sup>.

**Alimentation** : Le jus de pamplemousse renferme des substances chimiques qui sont de puissants inhibiteurs du CYP3A4 dans la couche muqueuse de la paroi intestinale <sup>76</sup>. Les légumes crucifères (p. ex. chou de Bruxelles, chou, chou-fleur) et les hydrocarbures présents dans la viande grillée au charbon de bois peuvent induire le CYP1A2 <sup>75</sup>. Le calcium des produits laitiers peut se lier aux médicaments tels que la tétracycline et les fluoroquinolones par chélation <sup>77</sup>. Les inhibiteurs non sélectifs et irréversibles de la monoamine oxydase réduisent le métabolisme de la noradrénaline endogène et de la tyramine exogène provenant des aliments et des boissons qui ont subi une dégradation de leurs protéines par le vieillissement, la fermentation, le marinage, le fumage ou la contamination bactérienne (p. ex. bière, vin, certains fromages et saucisses). La biodisponibilité accrue de la tyramine associée à l'augmentation des réserves de tyramine endogène peuvent entraîner une potentialisation de l'effet sympathomimétique indirect de la tyramine et provoquer une crise d'hypertension.

## 1.7 Les outils employés dans les études de métabolisme

Les approches et les techniques dédiées à l'étude du métabolisme ne sont pas forcément les mêmes en recherche et en développement car les besoins et les objectifs diffèrent à ces deux étapes. En recherche, l'intention première est le criblage d'un grand nombre de molécules afin de sélectionner

des candidats idéaux pour la phase de développement qui suivra. En conséquence, les techniques utilisées en recherche doivent être compatibles avec un haut débit. Elles regroupent des approches *in vitro*, *in silico*, *in situ* ou *in vivo* qui présentent de fortes variations de pertinence physiologique, de temps de mise en œuvre et de coût. A l'inverse, les études réalisées en développement nécessitent l'analyse plus en profondeur d'un seul composé et sont des approches maîtrisées et validées, répondant aux exigences des « Bonnes Pratiques de Laboratoire » (BPL) pour les études de toxicocinétique.

### 1.7.1 Méthodes *ex vivo*

- ***In silico*** :

Les simulations *in silico* et les études de relation structure-activité sont des outils utilisés en développement précoce pour éliminer les chimio-types les plus problématiques. Ces techniques sont basées sur la modélisation de la fixation des molécules chimiques étudiées sur les récepteurs nucléaires dont les structures ont été obtenues par cristallographie aux rayons X <sup>78</sup>. Si ces modèles sont prometteurs, l'extrême flexibilité du domaine de reconnaissance de PXR limite leur applicabilité. En effet, les changements de conformation qui sont susceptibles de se produire après la fixation du ligand sont difficilement modélisables <sup>79</sup>.

En outre, il existe des logiciels de prédiction des métabolites potentiels issus d'une molécule, comme METEOR (Lhasa Ltd., Leeds, GB) ou Pallas (CompuDrug Inc., San-Francisco, E-U). Basés sur des données issues de la littérature et sur des réactions chimiques préenregistrées, ces logiciels servent surtout à prédire les métabolites de phase I et II les plus attendus plutôt que l'ensemble des métabolites susceptibles d'être formés <sup>80</sup>.

- ***In vitro*** :

- Enzymes recombinantes : Il s'agit du système le plus simple d'étude de la métabolisation d'un composé chimique. Les ADNc des CYP les plus courants ont été clonés et les enzymes humaines recombinantes ont été exprimées dans une variété de cellules qui présentent une faible activité intrinsèque cytochrome P450, notamment des bactéries, des levures, et des cellules lymphoblastoïdes humaines <sup>81</sup>. Les avantages de ces systèmes sont leur simplicité et la possibilité d'étudier chaque enzyme du métabolisme séparément. Leur principal inconvénient réside dans la différence d'activité enzymatique par quantité de protéine qui existe entre le système d'expression et un micrososome hépatique humain.

- Les microsomes hépatiques : Des microsomes (vésicules de réticulum endoplasmique) hépatiques peuvent facilement être isolés par ultracentrifugation à partir d'homogénats de foie, ce qui fait de cette technique un outil très répandu. De façon à limiter la variabilité interindividuelle, ces systèmes sont généralement utilisés en regroupant des microsomes provenant de plusieurs individus donneurs. Ce système présente l'avantage de pouvoir étudier à la fois certaines réactions de phases I (CYP, flavine monooxygénase, hydrolases, ...) et II (UGT), moyennant l'adjonction des substrats et cofacteurs nécessaires.

Leur principal défaut est le temps limité de réaction puisque l'activité enzymatique décroît après 2 heures <sup>80</sup>. Les microsomes ne sont généralement pas appropriés pour l'étude des séquences de réactions métaboliques (i.e., le couplage des réactions de phases I et II), car l'organisation naturelle des éléments intracellulaires est détruite. Les microsomes hépatiques humains sont couramment utilisés pour l'exploration des effets d'un nouveau médicament sur les CYPs et pour l'obtention de données préliminaires sur les interactions susceptibles de survenir entre plusieurs médicaments. La prévision de ces dernières à partir de systèmes acellulaires comme les microsomes peut s'avérer fortement biaisée si les différences importantes entre les concentrations plasmatiques et les concentrations à l'intérieur des hépatocytes, observées *in vivo*, ne sont pas prises en compte.

La fraction intracellulaire S9 (surnageant d'homogénat de foie résultant de la précipitation des noyaux et des mitochondries par centrifugation) est utile pour l'étude du métabolisme des médicaments faisant appel à la fois aux enzymes microsomales et cytosoliques <sup>82</sup>.

- Les hépatocytes frais ou cryo-préservés : Ces modèles sont très intéressants pour la prédiction de la clairance et l'étude des interactions médicamenteuses. En effet, ces derniers comportent l'ensemble des enzymes de phase I et II ainsi que tous les cofacteurs nécessaires. Ceci autorise des temps d'incubation beaucoup plus long (jusqu'à 120 heures pour des cultures primaires). En inhibant la voie métabolique principale de la molécule étudiée, ces systèmes donnent accès aux voies métaboliques secondaires capables de la prendre en charge. Ils autorisent ainsi l'étude des conséquences d'une interaction médicamenteuse qui inhiberait le métabolisme principal de la molécule d'intérêt. Ces systèmes permettent également d'étudier l'impact des transporteurs sur le métabolisme. A noter cependant que les concentrations d'exposition au xénobiotique doivent être proches de celles qui seront rencontrées *in vivo* pour que ces résultats soient pertinents. La limitation majeure de ce modèle tenait à la faible disponibilité des hépatocytes humains mais les récents progrès en cryo-préservation ont permis une augmentation des ressources en cellules qui peuvent maintenant être conservées tout en maintenant une activité comparable aux cellules fraîches (à 10-20% près) <sup>80</sup>.

- Les cellules Caco-2 : Elles dérivent de cellules cancéreuses humaines du côlon. Cultivées sur des membranes poreuses, elles se différencient spontanément en monocouches de cellules polarisées. Ces monocouches de cellules Caco-2 sont utiles pour l'étude des interactions médicamenteuses dues à l'inhibition de l'efflux cellulaire lié à la P-gp. Elles servent de modèle pour l'absorption et le métabolisme dans l'intestin humain. Elles ont pour inconvénients de sous-exprimer les enzymes du métabolisme mais de surexprimer certains transporteurs. De plus, leur activité enzymatique s'atténue avec le temps et elles sont sujettes à la dérive clonale. Il est donc très difficile de réaliser des corrélations *in vitro* / *in vivo*

- Les systèmes tissulaires : Les foies humains utilisés ont pour origine les banques d'organes et sont ceux qui n'ont pas trouvé de receveur. Leurs propriétés varient donc selon l'état de santé, l'âge, le génotype du donneur, et des facteurs environnementaux tels que son

alimentation et sa consommation de tabac ou d'alcool. Les coupes précises de foie<sup>83</sup>, comme les foies perfusés, présentent la totalité des enzymes hépatiques du métabolisme, les cofacteurs, l'organisation naturelle des enzymes fixées aux membranes, et une restitution plus conforme du gradient de concentration du médicament de part et d'autre de la membrane cellulaire. Ils possèdent en outre un autre avantage qui est la conservation de l'architecture cellulaire du tissu et des communications intercellulaires. L'utilisation de ces systèmes requiert une grande expertise technique, et leur faible disponibilité conduit à les réserver aux études mécanistiques réalisées en développement.

### 1.7.2 Méthodes *in vivo*

Les espèces animales fournissent des modèles utiles pour déterminer, d'une part, si un métabolite produit *in vitro* par les microsomes de foie humain possède des effets pharmacologiques ou toxiques *in vivo* et, d'autre part, quel est l'impact de ces effets sur l'action de la molécule initiale. La connaissance des propriétés pharmacodynamiques et toxicologiques des métabolites d'un médicament chez l'homme permet également de prévoir dans une certaine mesure l'impact clinique de ce dernier en termes d'interactions médicamenteuses.

Cependant, les différences inter- et intra-espèces concernant les enzymes du métabolisme restreignent la capacité d'extrapoler à l'humain les résultats expérimentaux obtenus chez l'animal. Bien que les lignées de cellules animales et les fractions microsomales de cellules de foie animal puissent servir à l'acquisition de données préliminaires, celles-ci doivent donc toujours être confirmées par des études effectuées sur des préparations obtenues à partir de tissus humains.

En revanche, la comparaison du métabolisme d'un médicament chez l'humain et chez l'animal permet la sélection rationnelle des modèles animaux utilisés dans les études de toxicologie. Ainsi, l'existence de variations importantes dans le devenir métabolique d'un médicament chez un animal et chez l'humain réduit la valeur prédictive des données toxicologiques obtenues chez cet animal.

Enfin, l'emploi de modèles animaux transgéniques pour l'expression d'enzymes du métabolisme en systèmes hétérologues ou pour l'inactivation génique de certaines d'entre elles, constitue un moyen d'étudier *in vivo* le rôle d'enzymes spécifiques participant au métabolisme des médicaments<sup>84</sup>,<sup>85</sup>. Malgré cela, en raison de la variabilité inter-espèce, les paramètres issus de ces études ne seront pas nécessairement le reflet de la situation chez l'homme<sup>85</sup>.

En *in vivo*, les études comportent des dosages rapprochés, dans l'ensemble des fluides biologiques (sang, urine, fèces), de façon à suivre les cinétiques de disparition de la molécule d'intérêt et d'apparition de ses métabolites.

### 1.7.3 Les paramètres étudiés

- *In vivo*

Les paramètres utilisés en pharmacocinétiques comprennent la surface sous la courbe de la concentration du médicament dans le sang, le plasma ou le sérum par rapport au temps (AUC), la concentration maximale ( $C_{max}$ ), le temps nécessaire pour atteindre la  $C_{max}$  ( $T_{max}$ ), la concentration minimale ( $C_{min}$ ), la clairance (CL), le volume de distribution ( $V_d$ ) et la demi-vie ( $t_{1/2}$ ). Enfin, la fraction de médicament libre dans le plasma peut servir à l'évaluation des interactions médicamenteuses mettant en jeu certains agents fortement liés aux protéines plasmatiques.

La réglementation impose de déterminer les paramètres pharmacocinétiques de la molécule mère et des principaux métabolites actifs ou toxiques.

### 1.7.4 Corrélation entre études *in vitro* et *in vivo*

Pour évaluer la corrélation entre les études d'interactions médicamenteuses effectuées *in vitro* et *in vivo*, il est nécessaire de prendre en compte plusieurs éléments :

- Les études d'interactions médicamenteuses *in vitro* doivent être effectuées à des concentrations similaires à celles qui sont obtenues *in vivo*.
- La pertinence clinique des résultats positifs obtenus *in vitro* doit toujours être confirmée *in vivo*.
- Les études d'interactions médicamenteuses réalisées *in vitro* ne permettent pas toujours de prévoir exactement les interactions *in vivo*, lorsque d'autres voies de métabolisme ou d'élimination jouent un rôle de premier plan dans la clairance du médicament *in vivo*.
- Si les études *in vitro* et *in vivo* produisent des résultats divergents, les résultats obtenus *in vivo* doivent avoir préséance, pourvu que les études *in vivo* aient été menées dans des conditions pertinentes sur le plan clinique.

Plusieurs facteurs confusionnels peuvent empêcher l'établissement de corrélations exactes entre les études menées *in vitro* et *in vivo*<sup>86</sup> :

- Les concentrations de médicament libre dans le plasma (*in vivo*) peuvent être inférieures aux concentrations du même médicament en biophase hépatique, puisque la captation de nombreux composés lipophiles par le foie n'est pas nécessairement limitée par la liaison aux protéines.
- Lorsqu'une interaction médicamenteuse met en jeu à la fois des éléments inductifs et inhibiteurs, l'effet prédominant sur la clairance du médicament peut être soit positif, soit négatif, et fonction du temps (p. ex. inhibition métabolique dans des études à dose unique et induction à l'équilibre).
- Les résultats obtenus *in vitro* sur les interactions médicamenteuses peuvent donner une sous-estimation des interactions survenant *in vivo* pour les agents qui inhibent le transporteur

intestinal glycoprotéine P et ainsi augmenter la biodisponibilité des médicaments co-administrés qui sont des substrats de ce mécanisme d'efflux.

- Des concentrations élevées de protéines microsomales peuvent entraîner une sous-estimation de l'inhibition, si l'inhibiteur vient à manquer à cause du métabolisme microsomal ou de liaisons non spécifiques aux protéines microsomales <sup>71</sup>.

### 1.7.5 Etude et Enjeux du phénomène d'induction enzymatique

Les cultures primaires d'hépatocytes humains et les lignées cellulaires d'hépatocytes (p. ex. lignée cellulaire d'hépatome humain HepG2 <sup>87</sup>) sont utiles pour l'étude de l'induction. Les artéfacts, ou phénomènes parasites, sont cependant fréquents, car les hépatocytes présentent habituellement une perte de l'expression du CYP qui augmente avec le temps, ainsi qu'une diminution de leur capacité d'induction. La valeur prédictive des résultats obtenus à partir des lignées d'hépatocytes est souvent limitée par les différences phénotypiques substantielles qui existent entre ces cellules et les tissus dont elles proviennent. L'induction enzymatique peut être mesurée par le dosage de l'activité des isoformes spécifiques, par immunodétection de la protéine isoforme ou par le dosage quantitatif de son ARNm. Le recours à des inducteurs connus à titre de témoins positifs permet de vérifier la sensibilité des systèmes.

Le phénomène d'induction conduit à une altération de la pharmacocinétique et/ou de la pharmacodynamie <sup>88</sup>, voire à des effets toxiques potentiels <sup>89</sup> d'un agent thérapeutique par l'accélération de l'élimination de la molécule mère et de la formation de ses métabolites. Si les outils actuels permettent de relativement bien maîtriser le métabolisme et les conséquences sur ce dernier d'inhibitions enzymatiques, l'étude de l'induction reste encore problématique. En effet, les multiples sources de variabilité (génétiques, environnementales,...) de l'activité des enzymes du métabolisme et de leur capacité à être induites, et surtout le fait qu'elles soient non extrapolables à l'homme nécessite de trouver de nouveaux outils d'étude de ce phénomène, parmi lesquels l'analyse du métabolome représente une des pistes les plus prometteuses.

## 2 LE METABOLOME

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 <sup>90</sup> 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

Ces études cherchent à répondre aux nombreuses interrogations qui subsistent à propos des fonctions et des mécanismes de régulation de l'expression des gènes.

## 2.1 Définition

Apparu à la fin des années 1990 <sup>91, 92</sup>, en analogie aux notions de génome, de transcriptome et de protéome, 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 <sup>92</sup>. Il est par définition caractéristique d'un état physiologique donné. Les métabolites sont des composés impliqués dans les processus métaboliques, qu'ils en soient les substrats, les produits ou qu'ils soient nécessaires à leur bon déroulement <sup>93</sup>. 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, les vitamines, les stéroïdes, etc. <sup>94</sup>

Son analyse repose sur l'obtention d'empreintes métaboliques, les plus exhaustives possibles, obtenues à l'aide de différentes méthodes analytiques qui seront détaillées ci-après (III : Méthodes analytiques). L'analyse différentielle de ces empreintes issues d'au moins deux groupes d'échantillons, représentatifs d'états physiologiques différents, vise à caractériser les répercussions de la modification d'un facteur externe, et ainsi à mettre en évidence la manière dont un système biologique y répond. L'analyse du métabolome peut alors permettre une meilleure compréhension de la biologie des systèmes <sup>95</sup>, et des changements liés à une maladie ou à un effet pharmacologique <sup>96</sup>, en mettant en évidence des interrelations métaboliques qui n'auraient pas pu être détectées avec des approches biochimiques ciblées traditionnelles.

L'étude du métabolome s'inscrit au même titre que les études du transcriptome et du protéome dans un contexte post-génomique. Comparé 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, il existe approximativement un facteur 10 entre le nombre de métabolites et le nombre de gènes (ex., 584 <sup>94</sup> 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 « omes », 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 <sup>97</sup>. 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, 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 <sup>98</sup>. En outre, il a été démontré par la théorie du contrôle métabolique <sup>99-101</sup> 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 très faibles ou indétectables <sup>98</sup>.

## 2.2 Les différentes approches métabolomiques

On peut distinguer deux grands niveaux d'analyses complémentaires :

- Les analyses dites **ciblées**. Elles s'avèrent utiles pour les analyses de métabolites endogènes faiblement concentrés, instables ou encore difficile à doser. Elles peuvent également être utilisées pour confirmer des hypothèses posées sur la base de l'interprétation de résultats d'analyses transcriptomiques et/ou protéomiques.
- Les analyses dites **globales**. Elles ont pour but l'analyse simultanée du plus grand nombre de métabolites possible.

Les approches métabolomiques sont conventionnellement classées en quatre niveaux d'études. Bien que les définitions de chaque niveau varient selon les auteurs <sup>90, 102, 103</sup>, celles qui sont données ci-dessous apparaissent les plus consensuelles :

- **L'analyse ciblée** (*target analysis*)

Elle a pour objet de mesurer quantitativement et spécifiquement un ou un nombre réduit de métabolite comme les substrats ou les produits d'une réaction enzymatique dont on veut évaluer l'activité *in vivo* ou *in vitro* suite à une altération de l'activité enzymatique (une mutation génétique, par exemple). Elle fait le plus souvent appel à des méthodes d'extraction ou plus généralement à une étape de traitement de l'échantillon importante, permettant la séparation de l'analyte d'avec la matrice. Les analyses effectuées dans ce cadre sont quantitatives, précises et sensibles et s'apparentent à des dosages en milieu biologique classique.

- **Le profilage métabolique** (*metabolite/metabolic profiling*)

Il s'agit ici d'analyser tous les composés appartenant à une voie métabolique ou à une famille chimique donnée. 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. Des méthodes spécifiques d'extraction sont parfois requises.

- **L’empreinte métabolique** (*metabolic fingerprinting*)

C’est une approche chimiométrique, qui fait appel aux analyses statistiques multivariées pour la construction de modèles prédictifs. Elle permet une approche simplifiée par rapport à l’approche métabolomique complète puisqu’elle se limite à la comparaison de profils d’élution (GC ou LC), de spectres IR ou profils RMN. Généralement, peu ou aucun traitement de l’échantillon n’est utilisé. Cette approche permet de classer très rapidement et à haut débit des échantillons inconnus selon des classes prédéfinies. Elle n’a pour objectif que de sélectionner un nombre suffisant de signaux qui puisse permettre la construction d’un modèle prédictif pour la classification ultérieure d’échantillons inconnus. Elle cherche ainsi à comparer des profils ou des empreintes de métabolites qui sont modifiés en réponse à une pathologie ou à une exposition à un toxique, mais l’identification et la quantification absolue des métabolites modifiés ne sont pas recherchées.

- **L’approche métabolomique ou métabonomique** (*metabolomic / metabonomic approach*)

Elle a pour ambition l’identification et la quantification non biaisée de tous les métabolites présents dans un échantillon biologique prélevé dans des conditions données. Cette approche vise à donner une vue d’ensemble de l’état des métabolismes primaires et secondaires d’un organisme, afin de permettre l’établissement de corrélations entre les différentes voies métaboliques, ou de repérer les réseaux de régulation. Elle cherche à donner une signification biologique aux variations observées. Les deux termes métabolomiques et métabonomiques 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. Récemment, un consensus tend à apparaître et 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 à la phénomène biologique étudié <sup>104</sup>.

## 2.3 Applications

Les approches métabolomiques trouvent des applications dans différents domaines tels que la toxicologie, le phénotypage, le marquage de pathologie, la nutrition <sup>105, 106</sup> et à un niveau plus fondamental la compréhension globale des systèmes vivants. Les publications actuelles traitent aussi bien de développements technologiques que d’applications biologiques, médicales et pharmaceutiques. La connaissance actuelle de la fonction des gènes reste limitée et les études métabolomiques sont largement exploitées dans le cadre de la génomique fonctionnelle en compléments des approches transcriptomiques et protéomiques. L’application de données métabolomiques dans la cadre de la génomique fonctionnelle a ainsi été initiée par l’approche dite

« FANCY »<sup>98</sup>. Cette approche est basée sur l'idée que des altérations génétiques mènent souvent à des modifications significatives des taux de métabolites même dans le cas de mutations « silencieuses » qui ne produisent pas de modifications phénotypiques en termes fonctionnels ou métaboliques. Ces approches « FANCY » visent ainsi à 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<sup>103, 107</sup>.

La mesure simultanée d'un grand nombre de métabolites peut aussi permettre de découvrir de nouvelles voies métaboliques. Quelques études ont ainsi mis en relation des métabolites grâce à l'emploi de diagrammes métaboliques et ont dès lors démontré l'intérêt des études métabolomiques pour cartographier les réponses de cellules ou d'organismes exposés à des agents toxiques<sup>108</sup>. Une récente publication traitant de la toxicité du paracétamol chez la souris illustre ces propos<sup>109</sup>. 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. Ces auteurs ont donc proposé l'utilisation de cet acide en tant que biomarqueur de stress oxydant.

Les approches métabolomiques ont également été utilisées pour assigner aux enzymes de nouvelles activités et spécificités de substrat<sup>110</sup>. 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 électrophorèse capillaire couplée à la spectrométrie de masse. La modification de la concentration des métabolites suggère la présence d'une activité enzymatique alors que l'identification des métabolites dont les taux ont été modifiés permet de connaître les substrats et les produits de la réaction.

En plus d'être une nouvelle composante dans l'exploration des systèmes biologiques, ce domaine de recherche offre des perspectives dans les domaines biomédicaux et pharmaceutiques. Parmi eux, la recherche de biomarqueurs de différentes pathologies pour leur diagnostic biologique a fait l'objet de revues bibliographiques<sup>111</sup>. Les données les plus anciennes concernent la recherche d'anomalies innées du métabolisme par dosage des acides organiques urinaires en GC/MS<sup>112</sup>. D'autres pathologies comme la myopathie de Duchenne<sup>113</sup> ou le cancer<sup>114, 115</sup> ont été explorées, principalement en RMN. Avec une démarche analogue, la possibilité de dépister des maladies coronariennes sur la base de spectres de RMN obtenues à partir d'échantillons plasmatiques de patients a été démontrée<sup>116</sup>. Cette étude est d'autant plus intéressante qu'il n'existe pas de méthode de diagnostic des affections coronariennes non invasives. La recherche de biomarqueurs peut également être mise en œuvre dans des cadres aussi divers que des études portant sur les effets de

stress environnementaux <sup>117</sup> ou d'un traitement anabolisant dans le cadre de stratégies de contrôle de dopage <sup>118</sup>.

Les études métabolomiques visent encore à une meilleure compréhension des pathologies dont l'étiologie est inconnue ou mal caractérisée. 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 <sup>119</sup> à partir de tissus cérébraux humains en combinant les approches transcriptomiques, protéomiques et métabolomiques. Les auteurs de cette étude ont ainsi mis en évidence des modifications du métabolisme énergétique cérébral et du fonctionnement mitochondrial, ainsi que l'existence de mécanismes de réponse à la présence d'un stress oxydant.

D'autre part, l'analyse métabolomique est maintenant appliquée au phénotypage d'animaux transgéniques. L'objectif est de caractériser biochimiquement des modèles animaux qui miment des pathologies humaines afin d'améliorer leur compréhension, d'identifier des biomarqueurs pour leur diagnostic biologique et de caractériser de nouvelles cibles thérapeutiques <sup>107</sup>. 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 ont également trouvé des applications dans l'industrie pharmaceutique pour la prédiction de la **toxicité** de molécules en développement ou le biomarqueage d'une efficacité thérapeutique de médicament en cours de développement. Elles sont particulièrement intéressantes en toxicologie car elles permettent d'obtenir des informations sur l'intégrité de l'organisme dans le temps suite à l'exposition à un médicament <sup>104, 120, 121</sup>. En effet, grâce à l'utilisation de biomarqueurs urinaires, les données peuvent être obtenues de façon répétée à partir d'un même animal, le sacrifice de ce dernier pouvant être évité. Ces approches permettent donc de diminuer le nombre d'animaux par étude <sup>122</sup> ainsi que la quantité de composé à tester devant être synthétisée.

Elles peuvent également contribuer à résoudre un problème mécanistique lors du développement d'un médicament. A titre d'exemple, Mortishire-Smith et coll. soulignent l'apport de l'analyse métabonomique à la caractérisation du mécanisme de toxicité d'un médicament en développement <sup>123</sup>. La molécule «MrkA» est un antagoniste sélectif de récepteurs des chimiokines qui pourrait être prescrite dans le cadre de la prise en charge des patients atteints du SIDA. Les études préliminaires de toxicologie ont montré que cette molécule présente une toxicité hépatique. Les approches classiques telles que l'anatomo-pathologie n'ont pas pu préciser le mécanisme de cette toxicité. La question était donc de savoir si celle-ci était attribuable au mode d'action pharmacologique donc à la classe de ce nouveau type de molécules, ou bien au composé lui-même. Des études menées chez des rats Sprague-Dawley ont permis de mettre en évidence une augmentation de la concentration des

acides subérique, pimélique et sébacique traduisant une altération du métabolisme des acides gras et plus particulièrement d'une de leur voie de dégradation : la  $\beta$ -oxydation. Ces résultats, obtenus à partir de plusieurs molécules, suggèrent que la toxicité provient du composé lui-même, et non pas de la classe thérapeutique. Enfin, les études métabolomiques ont été appliquées à l'étude du devenir des médicaments dans l'organisme. Plusieurs études rapportent ainsi l'utilisation de méthodes de LC/MS pour la caractérisation des métabolites de xénobiotiques <sup>124-128</sup>.

Des études exploratoires ont enfin été menées afin d'évaluer la possibilité d'adapter les traitements médicamenteux et l'alimentation <sup>129-133</sup> de chaque individu selon son statut métabolique, déterminé par approche métabolomique. Clayton et coll. ont ainsi publié une étude où ils ont été capables de prédire l'importance d'une partie du métabolisme urinaire du paracétamol chez le rat, à partir des profils métaboliques des animaux, avant même leur exposition au paracétamol. Ceci a pu être étendu à la prédiction de l'importance de l'atteinte hépatique des rats en cas de surdosage en paracétamol <sup>134</sup>. Cette étude a ouvert la voie à une nouvelle discipline, la pharmacométabolomique, qui tentera de compléter les résultats issus de la pharmacogénomique qui ne donnent qu'un reflet partiel du statut métabolique d'un individu, dans la mesure où la pharmacogénomique ne prend pas en compte les facteurs liés 1- à la régulation de la transcription des gènes des enzymes du métabolisme, 2- à la fonctionnalité des enzymes du métabolisme et enfin 3- à l'environnement.

### 3 METHODES D'ANALYSES DU METABOLOME

Pour les organismes même 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 <sup>135</sup>, et jusqu'à 200000 métabolites pour le règne végétal <sup>136</sup>. 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 <sup>137</sup>. Enfin, les composants diffèrent aussi bien au niveau de leur taille, de la nature de leurs groupements fonctionnels, de leur volatilité, de leur polarité que de leur concentration physiologique. 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 dans des études métabolomiques : la Spectroscopie Infrarouge à Transformée de Fourier <sup>91</sup>, l'absorption UV-visible et la Fluorescence <sup>138</sup>, la Chromatographie sur Couche Mince <sup>92</sup>, la coulométrie <sup>139</sup>, l'électrochimie <sup>140</sup>, la Résonance Magnétique Nucléaire (RMN), la spectrométrie de masse par pyrolyse <sup>141</sup>, l'Electrophorèse Capillaire

couplée à la Spectrométrie de Masse (CE/MS)<sup>142, 143</sup>, 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)<sup>144, 145</sup> et la spectrométrie de masse à très haute résolution seule avec des analyseurs à Résonance Cyclotronique des Ions à Transformée de Fourier (FT-ICR)<sup>146, 147</sup>. 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 chimiométriques, donnent accès à des informations sur les structures chimiques des métabolites, conduisant à leur possible identification afin de les replacer dans le contexte biologique.

### 3.1 La RMN

La RMN est l'une des premières méthodes d'analyse à avoir été appliquée avec succès à l'étude du métabolome<sup>103, 148-150</sup>. Elle bénéficie ainsi d'une vingtaine d'années 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 destructive, 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 dans lesquelles les métabolites observés dépendent des colonnes et des 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 recouvre 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  $^1\text{H}$ . L'utilisation d'une deuxième dimension a été envisagée par Dumas et al<sup>141</sup>. Une méthode utilisant la RMN  $^1\text{H}$ - $^{13}\text{C}$  HMBC (de l'anglais : *Heteronuclear Multiple Bonding Connectivity*) a ainsi été envisagée. 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<sup>141, 151</sup>. 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<sup>152</sup> et des cryosondes<sup>153, 154</sup>.

## 3.2 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, tandis qu'en métabolomique, le développement et les améliorations apportées aux analyseurs dont les temps de vol (*Time Of Flight*, TOF), FT-ICR et l'Orbitrap® ont donné accès à des informations plus précises et plus spécifiques.

Fondamentalement, un spectromètre de masse est composé des éléments suivants : 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 permettant de contrôler les paramètres d'acquisition.

La spectrométrie de masse offre plusieurs avantages pour les études métabolomiques. Elle est sensible et permet la détection d'espèces au  $\mu\text{M}$  au cours de l'acquisition des empreintes, donnant accès à bon nombre de classes de métabolites, détectables à leur concentration physiologique. De plus, elle autorise l'identification des métabolites *via* l'obtention de leur masse moléculaire, avec une précision dépendant du type d'analyseur, et *via* l'étude de leur spectre de fragmentation qui peut être comparé avec celui d'un composé de référence et qui fournit des informations structurales. Ces propriétés lui permettent donc d'effectuer 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 de métabolomique 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 larges jeux de données est un prérequis au développement des approches métabolomiques.

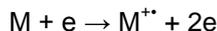
### 3.2.1 Les différentes sources d'ionisation.

#### 3.2.1.1 Les sources basées sur l'ionisation en phase gazeuse

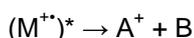
- **SOUS VIDE**
  - **L'impact électronique**

Cette source inventée par Dempster et perfectionnée par Bleakney<sup>155</sup> et par Nier<sup>156</sup> est parmi les plus couramment utilisées en spectrométrie de masse en couplage avec la chromatographie 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 gazeuses de l'échantillon analysé introduit dans la source. Lorsque l'énergie cinétique ( $E_c$ ) des électrons est de

l'ordre de 10 à 100 eV, les premiers ions formés résultent de l'arrachement d'un électron à la molécule M :



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 8 à 10 eV. En utilisant des électrons d'énergie cinétique supérieure à cette valeur, on communique aux ions moléculaires une énergie interne, de nature essentiellement vibrationnelle, située entre 0 et  $[E_c - EI]$ . Cet excédent d'énergie interne peut provoquer la dissociation des ions moléculaires et par conséquent la formation de fragments  $A^+$  et B :



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'ion moléculaire n'est pas toujours observé. Dans les conditions habituelles d'une source d'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 fragmentation robustes qui ont permis la constitution de banques de spectres universelles, facilitant ainsi l'identification ultérieure des composés.

#### ○ L'ionisation chimique

L'ionisation chimique <sup>157</sup> 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, ammoniaque,...) par ionisation électronique. L'ion ainsi produit entrera essentiellement en collision avec d'autres molécules du gaz réactif, formant par une série de réactions consécutives un plasma d'ionisation. Des 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 la molécule d'intérêt qui sera détectée sous forme protonée ou déprotonée. Cette source est utilisée en couplage avec la chromatographie gazeuse.

#### • A PRESSION ATMOSPHERIQUE

Ces sources ionisent l'échantillon à pression atmosphérique puis transfèrent les ions produits vers le spectromètre de masse. 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 obtenu à pression atmosphérique. 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*)<sup>158</sup>, l'APCI et l'APPI. Ces deux dernières techniques sont détaillées ci-dessous.

- **L'APCI (ionisation chimique à pression atmosphérique)**<sup>159</sup>

Cette technique, employée en couplage LC-MS, fait appel à des réactions ions-molécule 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, les molécules de solvant et d'analytes, sous forme gazeuse, sont ionisées à l'aide d'une décharge corona. 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 de molécules protonées 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<sup>146</sup>. 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<sup>160</sup>.

- **L'APPI (photoionisation à pression atmosphérique)**<sup>161</sup>

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 en captant les photons avant de les transférer à l'analyte. Leur utilisation améliore la formation d'ions

négatifs <sup>162</sup>. 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 <sup>163</sup>. Ces deux espèces sont l'ion moléculaire  $M^{\bullet+}$  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 <sup>164</sup>, les pesticides <sup>165, 166</sup>, des hydrocarbures aromatiques polycycliques <sup>167</sup>. Cette source est utilisée en couplage LC-MS.

### 3.2.1.2 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 <sup>168</sup>. 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 13). 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.

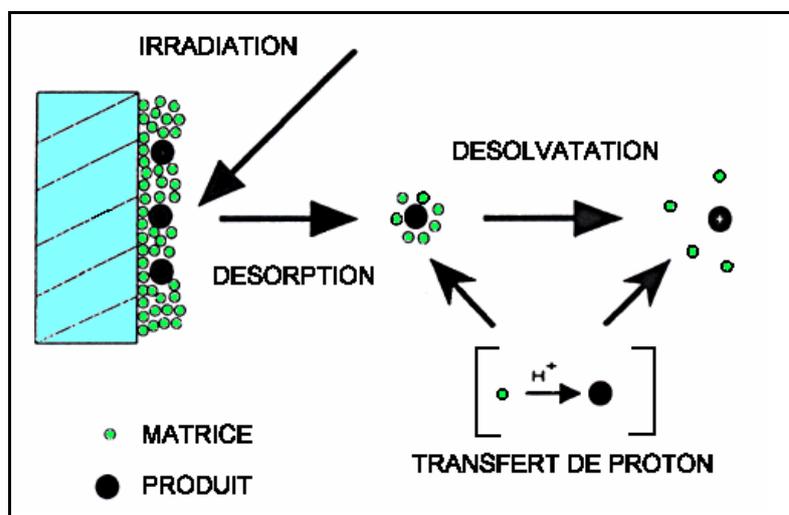


Figure 13 : Schéma du principe de l'ionisation MALDI

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 <sup>169</sup>. 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 <sup>170, 171</sup>. Des travaux effectués sur des extraits de bactéries ou de tissus <sup>172</sup> 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. Plus récemment, une approche métabolomique sur des extraits de tissu cardiaque de souris a été développée en MALDI/TOF/TOF <sup>173</sup>. Dans cette étude, 90 métabolites ont pu être identifiés à partir de 285 signaux correspondants à des ions négatifs.

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) <sup>174</sup>. 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<sup>n</sup>).

- **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*) <sup>179</sup> 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* <sup>180</sup>.

- **Le DIOS (désorption- ionisation laser sur silicium poreux)**

La désorption/ionisation sur silicium poreux (DIOS), a récemment été développée par l'équipe de Siuzdak <sup>181</sup>. 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 <sup>182</sup>. Le DIOS a également été utilisé dans les analyses protéomiques pour la caractérisation fonctionnelle et l'identification de petites protéines <sup>183</sup>. 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 <sup>184</sup>.

- **A PRESSION ATMOSPHERIQUE**

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

Cette technique, développée par J.B. Fenn au milieu des années 1980 <sup>175-178</sup>, 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. 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 d'agrégats chargés. Les agrégats chargés ainsi produits sont accélérés pour être désolvatés au cône alors que les molécules non chargées sont éliminées (Figure 14). Cette technique d'ionisation douce, permettant une faible fragmentation de l'ion 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.

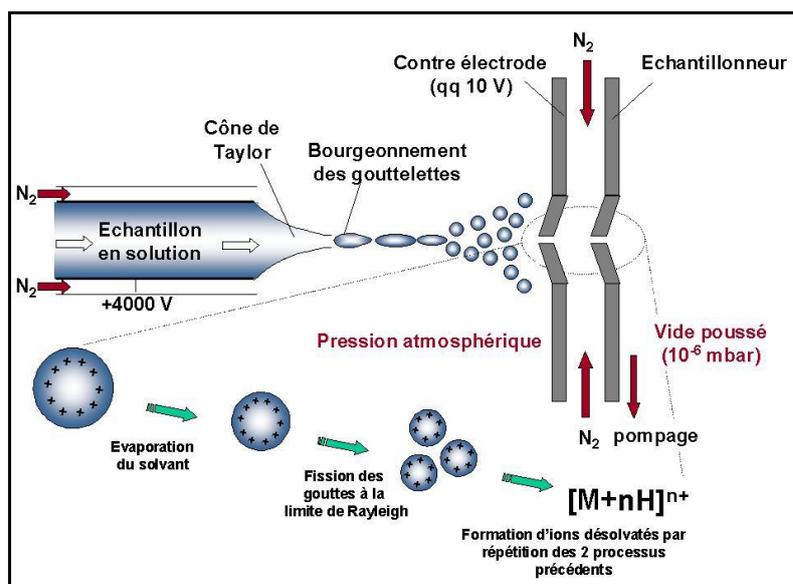


Figure 14 : Schéma d'une source électrospray

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

### ○ Le DESI (Désorption- Electrospray 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*)<sup>185</sup>. 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 permettent 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<sup>186, 187</sup>. Elle a été utilisée dans le cadre d'études métaboliques portant sur des urines de souris ayant ou non développées une tumeur<sup>179</sup> et sur le métabolisme central d'*E. coli* à partir d'extraits cellulaires et de prélèvements du milieu de culture<sup>188</sup>. Elle présente en outre une tolérance plus importante aux sels par rapport aux techniques d'ionisation comme l'ESI ou l'APCI<sup>179</sup> et permet donc l'analyse de milieux complexes sans traitement de l'échantillon. Enfin, elle offre la possibilité de faire de l'imagerie tissulaire sur de petites molécules comme les lipides, les médicaments et divers métabolites, même si cette application reste à l'heure actuelle à un stade de développement précoce.

#### 3.2.2 Les différents analyseurs

Tout comme il existe une grande variété de sources, il existe de nombreux analyseurs qui sont classés en deux catégories :

- **les analyseurs à faisceaux d'ions** (« *in space* » *instruments*): comme les quadripôles et les analyseurs à temps de vol (*time of flight*, TOF).
- **les analyseurs à piégeage d'ions** (« *in time* » *instruments*): comprenant les pièges ioniques à basse résolution comme la trappe de Paul (3D) ou les pièges ioniques linéaires (2D) et les analyseurs à ultra haute résolution, avec l'analyseur par résonance cyclotronique d'ions à transformée de Fourier (FT-ICR) et l'Orbitrap®.

Leurs performances sont évaluées grâce à plusieurs paramètres dont les définitions sont reportées ci-après pour les plus intéressants en terme d'analyse métabolique.

**La vitesse de balayage** qui est le temps que va mettre l'analyseur à établir un spectre pour une gamme de  $m/z$  donnée.

**La résolution**, qui représente la capacité de l'analyseur à séparer deux composés de  $m/z$  proches, et qui est donnée par  $R = m/\Delta m$ , où  $m$  correspond au rapport  $m/z$  mesuré et  $\Delta m$  à la largeur

du pic à x % de sa hauteur maximale. De plus en plus x=50%, et on parle de résolution FWHM de l'anglais *Full Width at Half-Maximum peak height*.

**La précision** qui renseigne sur la faculté de l'analyseur à mesurer une masse juste de la molécule analysée (la plus proche possible de la masse théorique). Cette précision est le plus souvent exprimée sous forme d'erreur de mesure de masse *en partie 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.

Le Tableau 2 présente un récapitulatif des performances des analyseurs disponibles sur le marché et utilisé en métabolomique.

Tableau 2 : Récapitulatif des performances des différents analyseurs

<b>Analyseur</b>	<b>Résolution (<math>m/\Delta m</math>, FWHM)</b>	<b>Exactitude en masse (ppm)</b>
Quadripôle	3000	200
Piège à ions (2D/3D)	5000	200
Temps de vol	8000-20 000	2-10 <sup>a</sup>
Orbitrap®	Jusqu'à 100 000	<3
Résonance cyclotronique	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

### 3.2.2.1 A basse résolution

- Les quadripôles

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 <sup>189</sup> ayant idéalement une section hyperbolique ou cylindrique auxquelles on applique des potentiels déphasés de 180° 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 15)

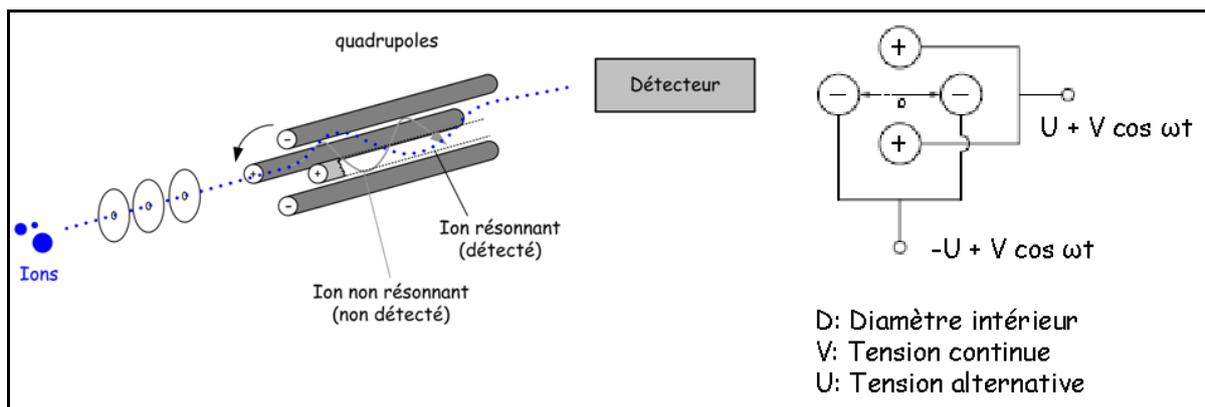


Figure 15 : Schéma d'un analyseur quadripolaire <sup>190</sup>

Certains appareils possèdent plusieurs analyseurs en série, séparés par une cellule de collision. La Figure 16 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.

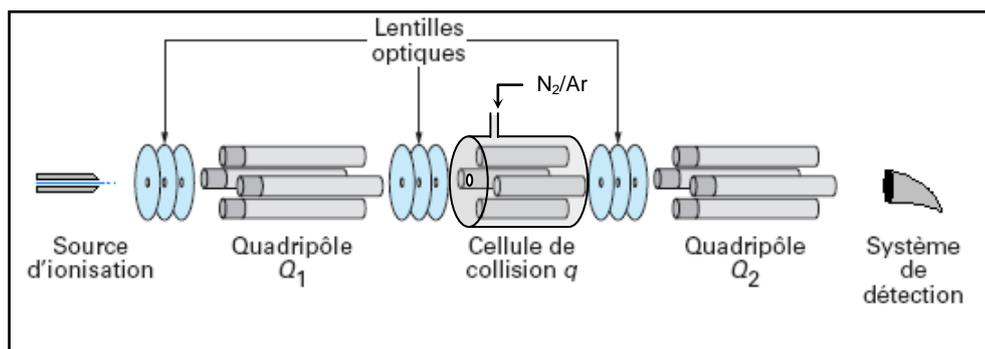


Figure 16 : Représentation schématique d'un triple quadripôle <sup>191</sup>

- **Les pièges à ions**

1.1. Les pièges à ions tridimensionnels <sup>192</sup>

Ils sont constitués d'une électrode circulaire en forme de diabolo, couverte de deux calottes sphériques (Figure 17). 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.

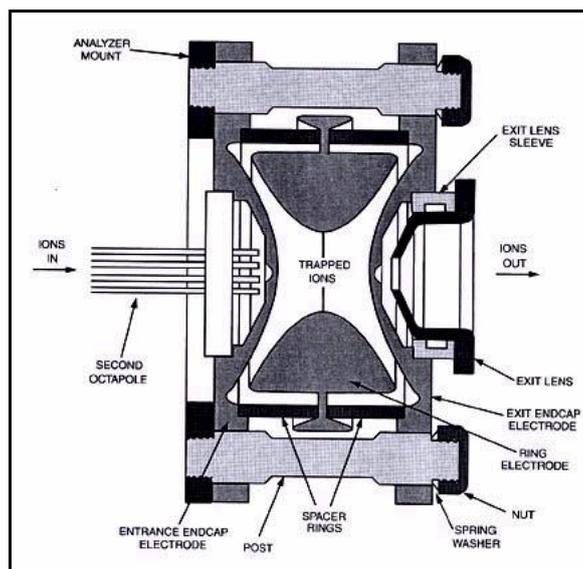


Figure 17 : Piège à ions 3D (Image issue de la documentation Thermo-Electron)

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<sup>193, 194</sup>. 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.

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 1/3 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<sup>195</sup>, 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.

## 1.2. Les pièges linéaires

Les pièges linéaires (pièges 2D)<sup>196</sup> 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<sup>197, 198</sup>. 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 sur son système à éjection radiale<sup>199, 200</sup>. 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 contrepartie, la fragmentation séquentielle à l'ordre 3<sup>198</sup>.

### 3.2.2.2 *Les analyseurs à haute résolution permettant des mesures de $m/z$ précis*

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®.

- **Analyseurs à temps de vol**

Le principe de l'analyseur à temps de vol (Figure 18), désigné TOF pour *Time of flight*, a été décrit en 1946 par Stephens<sup>201</sup>. Wiley et McLaren ont élaboré le premier TOF dans un but commercial en 1955<sup>202</sup>. 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.

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<sup>203</sup> 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.

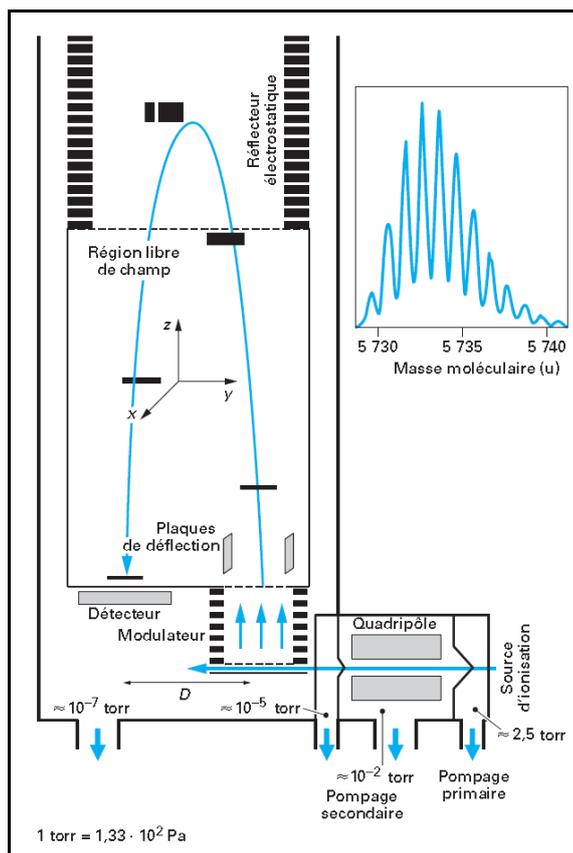


Figure 18 : Principe de fonctionnement d'un analyseur à temps de vol<sup>191</sup>

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<sup>204, 205</sup>. 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<sup>206-208</sup> sur les derniers appareils 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 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 10s. 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<sup>209, 210</sup>. 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 <sup>211</sup> 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 du 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 19).

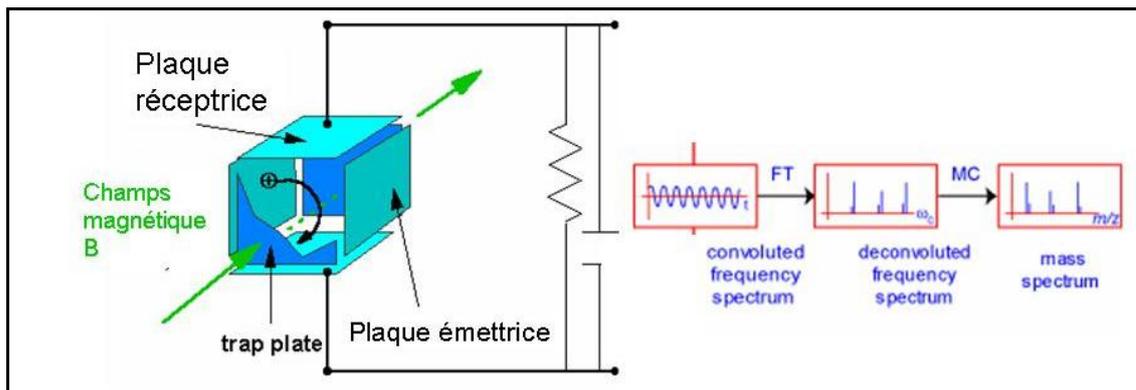


Figure 19 : Principe de fonctionnement du FT-ICR

Les spectromètres de masse par FT-ICR présentent un certain nombre d'avantages <sup>147, 212</sup> :

- 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 <sup>213</sup>.
- 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 <sup>214</sup>. 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'ions 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 <sup>146</sup>. Cependant, l'utilisation de la FT-ICR dans le cadre d'études du métabolome a fait l'objet d'une revue bibliographique <sup>147</sup> qui évalue les potentialités de ces spectromètres de masse pour les approches métabolomiques et quelques études ont été publiées depuis <sup>215-220</sup>. 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.

- L'Orbitrap®

D'introduction récente, l'Orbitrap® 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<sup>221</sup>. Brièvement, l'Orbitrap® 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 20). 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.

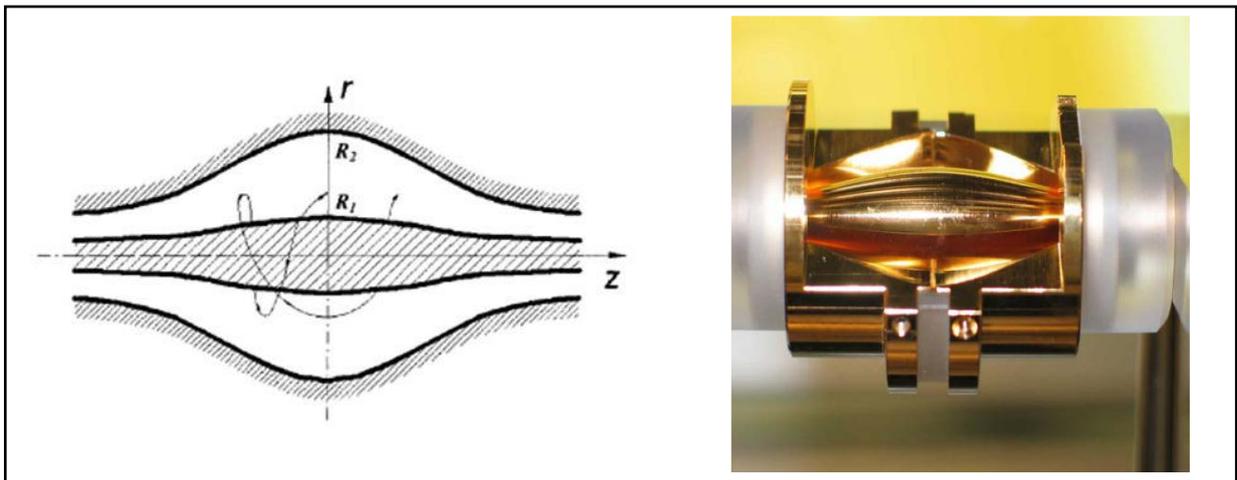


Figure 20: Piège électrostatique ou « Orbitrap® » d'après la description de A. Makarov<sup>221</sup>

L'Orbitrap® a été couplée par la société Thermo Electron avec un piège ionique linéaire (LTQ) pour former un spectromètre de masse hybride, le LTQ Orbitrap®. Le LTQ est capable d'acquérir ses 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 Da). Cet analyseur permet d'atteindre une précision en masse inférieure à 2 ppm et une résolution de 100000.<sup>222-224</sup>

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. Ce dernier a déjà été utilisé pour des approches de screening

de métabolites sur des incubats de microsomes hépatiques humains <sup>127</sup>, pour l'identification de métabolites de phénobarbital dans des urines de rat, mis en évidence par approche métabolomique globale <sup>128</sup>, pour l'acquisition d'empreintes métaboliques de bactéries <sup>225</sup> et de sera de mammifère en couplage avec l'UHPLC<sup>226</sup>, et a permis l'identification de biomarqueurs de prééclampsie <sup>227</sup>.

### 3.2.2.3 Les modes d'acquisitions des ions produits dans la source d'ions

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 successives (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*) <sup>228</sup> et ETD (*Electron transfert Dissociation*) et la technique IRMPD (*infra red multi photon dissociation*) <sup>229</sup>. Ces différentes techniques diffèrent dans les fragments qu'elles génèrent <sup>230</sup>. 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 <sup>231</sup>. 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 (cf. Figure 15) 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 troisième 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 <sup>232</sup>, de néphrotoxicité induite par la gentamicine <sup>233</sup>, ainsi que d'acides mercapturiques dans des urines humaines après une utilisation d'acétaminophène <sup>234</sup>. Le balayage des ions précurseurs a également été utilisé pour l'étude des sulfoconjugués <sup>232</sup>.

### 3.2.3 Préparation et méthodes d'introduction de l'échantillon

#### 3.2.3.1 Le traitement de l'échantillon

La qualité de la préparation des échantillons est un élément clé dans toute méthode analytique. 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. Les expériences de métabolomique peuvent être précédées d'une préparation de l'échantillon afin de cibler certains métabolites ou de mener des expériences quantitatives. En revanche, l'approche métabolomique globale cible tous les métabolites, le traitement de l'échantillon se doit alors d'être le plus simple et le plus universel possible afin de limiter les sources de variabilité

externes à l'échantillon. Il est cependant nécessaire de retirer les éventuels sels et macromolécules qui pourraient interférer avec la détection 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). 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 sont retenues. Elle est notamment utilisée de façon spécifique pour la précipitation des protéines lors des analyses de plasma, en effet une précipitation lente des protéines permet de limiter la quantité de métabolites entraînés par ces dernières lors de leur précipitation<sup>235</sup>. Les méthodes a et b permettent d'obtenir des mélanges plus « propres » mais la perte de certaines classes de petites molécules est observée<sup>147</sup>, elles sont en générale utilisées pour les approches ciblées ou semi-ciblées. Plusieurs études rapportent l'analyse d'échantillons bruts, sans aucun traitement, comme pour des urines<sup>124, 236</sup> ou des extraits cellulaires<sup>220, 237</sup>.

### 3.2.3.2 Introduction directe (infusion)

L'échantillon est introduit directement dans le spectromètre de masse. L'utilisation de l'introduction directe permet des analyses rapides, à haut débit, d'échantillons bruts ou d'extraits. Les sources électrospray et APCI sont généralement utilisées. 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, l'effet matrice (suppression ou augmentation d'ionisation) 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. L'introduction directe a été utilisée<sup>237, 238</sup>, notamment pour les classifications et identifications de microorganismes ou pour l'analyse de 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 isobares. Plusieurs études métabolomiques utilisant l'introduction directe ont été rapportées en couplage avec des FT-ICR et plus récemment avec l'Orbitrap<sup>®239</sup>. Dans le cas des infusions sans prétraitement, les sources DESI sont plus robustes que les sources ESI ou APCI, elles présentent également moins de problèmes avec les sels<sup>179</sup>. L'utilisation de sources nano-ESI peut également contribuer à réduire l'effet matrice et à limiter l'encrassement de la source<sup>240</sup>.

### 3.2.3.3 Les méthodes séparatives

#### 3.2.3.3.1 La GC/MS

Développée dans les années 1960, la chromatographie en phase gazeuse couplée à la spectrométrie de masse (IE) 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'un première colonne, généralement apolaire, est ensuite transféré sur une seconde colonne polaire pour subir une seconde dimension de séparation<sup>241</sup>.

L'énergie transmise aux analytes avec de l'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<sup>242</sup>), spécifiques à la GC participe également à la capacité 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*<sup>243</sup>, 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) afin de rendre volatiles les 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'ion moléculaire rend 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.

#### 3.2.3.3.2 La LC/MS

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 les 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<sup>244</sup>. 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.

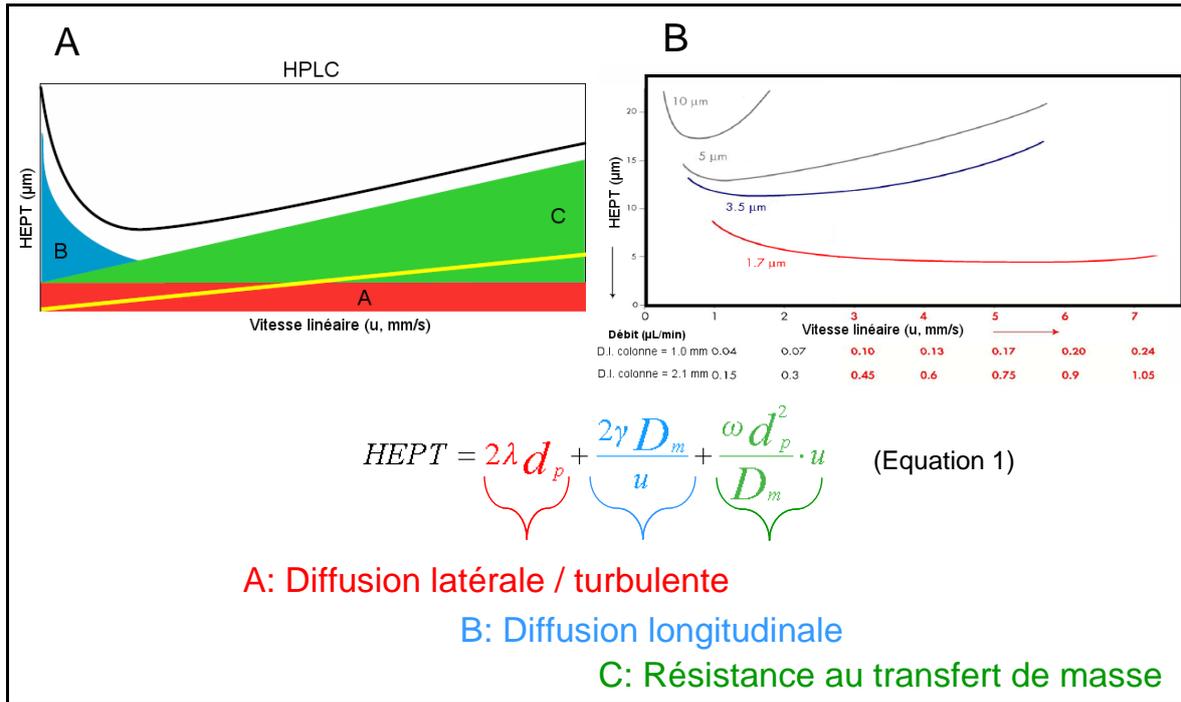
- LC conventionnelle

Les méthodes les plus employées en métabolomique utilisent des colonnes narrow bore (diamètre interne de 2.1mm) à polarité de phase inverse, du 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 volume mort. Quelques exemples d'analyse de tels composés par LC/MS avec l'utilisation de colonnes de type HILIC (*hydrophilic interaction liquid chromatography*), favorisant la rétention des composés hydrophiles, sont décrits<sup>194, 245-247</sup>.

La RP-HPLC conventionnelle est souvent insuffisante pour réaliser des séparations de milieu biologique complexes. L'utilisation de colonnes capillaires et de monolithes<sup>248</sup> pour améliorer la résolution a été rapportée. En effet, le diamètre intérieur (entre 200 et 320 µm) et la taille des particules de remplissage de ces colonnes diminuant, cela entraîne une amélioration de la résolution chromatographique. Les co-élutions des métabolites sont ainsi réduites 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. Ces colonnes ont été utilisées lors d'études métabolomiques chez la plante *Arabidopsis thaliana*<sup>248</sup>, ainsi que dans la classification d'urines de rats<sup>145</sup>. 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.

- 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 µm<sup>249</sup>. Les principes fondamentaux de la chromatographie peuvent être expliqués par la courbe de van Deemter<sup>250</sup> (Figure 21). La courbe de Van Deemter décrit la relation existant entre la vitesse linéaire (débit) et la hauteur équivalente à un plateau théorique ou HEPT, qui est directement reliable à l'efficacité de la colonne.



**Figure 21: Courbe et équation de van Deemter**

**A** : Courbe de van Deemter représentant la contribution des différentes formes de dispersion limitant l'efficacité d'une colonne en HPLC ; **B** : Influence du diamètre de particule sur l'efficacité de la colonne en fonction de la vitesse linéaire de la phase mobile ; **Equation 1** : Equation de van Deemter avec  $\lambda$  et  $\gamma$  : constantes liées à la géométrie des particules et à la qualité du remplissage (facteurs de tortuosité et d'obstruction) ;  $d_p$  : diamètre des particules ;  $D_m$  : coefficient de diffusion de la phase mobile ;  $u$  : vitesse linéaire de la phase mobile ;  $\omega$  : constante.

La taille des particules représente une des variables de cette équation (Equation 1). La courbe de van Deemter met en évidence que l'utilisation d'une granulométrie inférieure à 2,5  $\mu\text{m}$  engendre un gain significatif en efficacité qui est maintenu à des débits / vitesses linéaires plus élevés. Grâce aux granulométries plus fines, la vitesse et la capacité de pics (nombre de pics résolus par unité de temps), peuvent atteindre des performances inégalées. C'est sur cette approche que repose le développement de la technologie UHPLC. 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 selon la loi de Darcy (Equation 2), qui a nécessité des développements technologiques sur le matériel afin de permettre des pressions de travail de l'ordre de 10000 psi.

$$\Delta P = \Phi \cdot \frac{\eta \cdot L \cdot u}{d_p^2}$$

**Equation 2: Loi de Darcy**

$\Delta P$  : perte de charge ;  $\Phi$  : facteur de résistance à l'écoulement qui dépend de la forme des particules et de la qualité du remplissage de la phase stationnaire ;  $\eta$  : viscosité de la phase mobile ;  $L$  : longueur de la colonne ;  $u$  : vitesse linéaire de la phase mobile et  $d_p$  : diamètre des particules

L'utilisation de l'UHPLC induit également une augmentation de la sensibilité <sup>145</sup>, 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 » <sup>145</sup>. Les systèmes UHPLC présentent également une meilleure robustesse que les systèmes HPLC notamment au niveau des temps de rétention <sup>145, 251</sup>. L'analyse de *sera* humains à l'aide d'un système UPLC a permis l'augmentation de 20% du nombre de pics détectés <sup>252</sup> 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 de colonnes. Des filtrations et/ou des centrifugations sont ainsi obligatoires notamment pour les milieux les plus riches (plasma, extraits cellulaires).

### 3.3 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 avec l'électrophorèse capillaire dédiée à l'analyse des composés polaires et la LC à polarité de phase inversée pour les composés plus apolaires. Plusieurs méthodes de détection peuvent également être couplées avec par exemple des couplages du type LC/RMN/MS et LC/RMN/MS/MS <sup>253</sup> qui se sont montrés très prometteurs. Ils permettront en effet une identification plus aisée des métabolites tout en conservant la sensibilité de la spectrométrie de masse mais sont assez coûteux à cause de l'utilisation de phases mobiles deutériées. 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, l'analyse des métabolites de médicaments et l'analyse d'autres mélanges complexes <sup>254, 255</sup>. 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.

## 4 PRÉTRAITEMENT DES DONNEES

Toutes les approches de génomique fonctionnelle reposent sur 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 ce domaine pour la métabolomique.

L'étape de prétraitement des données est une étape critique mais nécessaire afin de présenter les empreintes métaboliques sous une forme compatible avec la réalisation des approches statistiques ultérieures. Jusqu'il y a peu, les méthodes de prétraitement des données étaient rarement décrites dans la littérature, mais l'apparition de « normes minimales de restitution des analyses métabolomiques »<sup>256-258</sup> et les nombreux travaux menés sur ce thème ont contribué à améliorer la situation.

L'étape de prétraitement des données a pour principaux objectifs d'extraire les signaux obtenus grâce aux différentes techniques analytiques, d'aligner les empreintes des différents échantillons, d'éliminer le bruit de fond, et de présenter les données extraites sous un format 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 pour le stockage des données. Cette étape peut enfin comprendre une procédure de normalisation des données.

Par la suite, ne seront développées que les méthodes employées dans le prétraitement des données issues de la RMN ou de la spectrométrie de masse.

### 4.1 RMN

Les spectres de RMN d'échantillons biologiques (mesurés à 500 ou 600MHz) sont complexes car ils contiennent plusieurs milliers de signaux correspondant aux déplacements chimiques des métabolites. La méthode de prétraitement 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<sup>259</sup>, 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<sup>260</sup>, correspondant à l'intégration d'une quantité variable de bruit de fond suivant les échantillons.

Une approche alternative est l'analyse des données en **résolution complète**<sup>151</sup>. Elle implique ou non l'utilisation d'algorithmes automatiques d'alignement de pics afin de résoudre le problème des variations de déplacement chimique<sup>261, 262</sup>. Ces variations peuvent affecter certaines molécules comme 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.

## 4.2 LC/MS

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 lors de chaque spectre. (Figure 22)

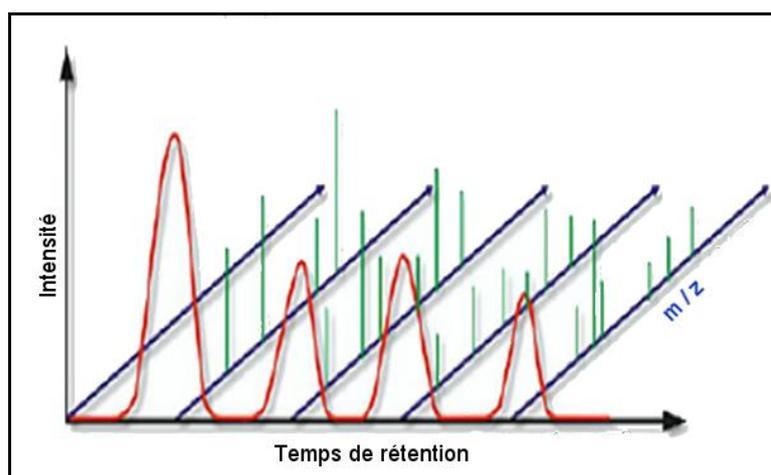


Figure 22: Représentation schématique de l'architecture 3D d'un chromatogramme obtenu par LC-MS

Par rapport à la RMN, l'étape de prétraitement des données est rendue encore plus complexe par la présence d'une technique séparative couplée, d'une importante quantité de bruit de fond, d'artefacts et de redondance dans les données. 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 se modifier d'un pic à l'autre. Le prétraitement de ces données va donc consister à extraire les pics (couple  $m/z$ - $t_r$ -intensité) dans chaque échantillon, puis à effectuer un réalignement dans les domaines chromatographique et spectral. Cette étape de prétraitement des données est décrite dans certaines publications<sup>121, 263-268</sup>.

**Fractionnement du chromatogramme** : Cette méthode permet de s'affranchir du problème lié à l'alignement des données. Il s'agit de diviser le chromatogramme en fractions temporelles et de sommer toutes les abondances correspondant à un  $m/z$  donné. Les analyses statistiques multivariées permettent de sélectionner les fenêtres les plus intéressantes et un retour aux chromatogrammes originaux permet de confirmer les variations<sup>121</sup>. La taille des fenêtres a pour but d'isoler un composé par fenêtre, Plumb et al ont par exemple divisé leur chromatogramme en zones de 100 scans, ce qui correspond à 1 minute. Ce fractionnement peut présenter un certain nombre de désavantages. En effet, les coupes entre les fenêtres sont choisies de façon arbitraire et le signal d'un analyte peut être divisé entre deux fenêtres. De plus, plusieurs composés indépendants peuvent contribuer au signal d'une seule fenêtre ce qui peut entraîner une difficulté pour discerner les variations de concentrations d'un composé lors d'études métabolomiques.

- **Utilisation de logiciels**: ils ont pour but la mise en œuvre d'algorithmes permettant la détection des pics, leur l'alignement, et l'élimination du bruit de fond. Certains seulement sont capables de déconvoluer les signaux.

Les logiciels capables de déconvolutions sont basés sur la théorie selon laquelle le spectre de masse d'un composé pur reste constant sur l'ensemble du pic chromatographique de ce composé. En conséquence, ils utilisent les SIR (*Single Ion Recording*) comme modèle pour la forme des pics. C'est pourquoi les pics qui coéluent mais qui ont des spectres de masse différents, des sommets séparés de moins d'une seconde et des composés présents à l'état de trace, peuvent être détectés. Parmi ces logiciels, on peut citer Markerlynx (Waters) qui permet la détection des pics et l'alignement spectral des données à l'aide de standards internes prédéfinis. Cependant, ce logiciel ne peut être utilisé qu'avec des spectromètres de masse Waters Micromass, ce qui rend son utilisation impossible pour des données provenant d'autres instruments, et n'est en outre pas capable d'effectuer un réalignement dans le domaine chromatographique. XCMS (Scripps Research Institute)<sup>266</sup> utilise un alignement linéaire ou non linéaire des temps de rétention et permet la détection des pics et la suppression du bruit de fond. Il divise les données en fenêtres de 0,1  $m/z$  qui sont utilisées pour générer les listes des ions extraits qui sont ensuite traitées individuellement. Il nécessite cependant des moyens informatiques importants (ordinateurs puissants) avec une capacité de mémoire vive importante puisqu'il ne travaille que dans la RAM. MS-resolver (Pattern recognition Systems) décompose les signaux enregistrés en contributions pour chaque composant. MetAlign<sup>TM</sup> (Plant Research International B.V)<sup>267</sup> inclue également une sélection des pics et la possibilité d'un alignement non linéaire et itératif. Cependant, il n'est basé sur aucun algorithme publié, à l'exception du manuel d'utilisation. De plus, il n'est pas possible de modifier ou d'étudier les codes utilisés, ces derniers étant protégés (brevetés). Sa principale limitation actuelle, en attendant la mise sur le marché d'une version haute résolution, est son incapacité à exploiter les masses précises, les variables générées correspondant à des couples  $tr-m/z$  où le rapport  $m/z$  est un nombre entier. Markerview (applied biosystems/MDS Sciex) a été utilisé dans le cadre de profiling ciblé de glucuronides dans des urines humaines<sup>269</sup>. MZ mine<sup>265</sup> permet de détecter les pics et d'aligner les chromatogrammes mais il ne réalise pas de reconstruction spectrale. Un outil nommé MET-IDEA (*Metabolomics Ion based Data*

Extraction Algorithm) a également été développé <sup>263</sup>. Il permet d'extraire les données relatives à l'abondance des ions associés aux différents métabolites séparés ou coéluant de données de LC ou GC/MS.

Il existe également des logiciels permettant l'analyse de données brutes ou de listes de variables, générées par un programme tiers, sans déconvolution comme par exemple MSFACTS (*Metabolomics Spectral Formatting, Alignment and Conversion Tools*). Il permet l'alignement des données chromatographiques et extrait l'information des fichiers bruts chromatographiques de type ASCII <sup>264</sup>. Citons également MathDamp <sup>270</sup> qui est un package développé pour Mathematica<sup>®</sup>, capable de travailler directement sur les données brutes, sans recours à une étape de détection et d'extraction des pics, ou Sieve<sup>®</sup>, 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.

Cette liste n'est pas exhaustive, elle cite les logiciels les plus utilisés à l'heure actuelle. Cette étape étant cruciale dans les études métabolomiques, de nouveaux logiciels apparaissent régulièrement et des améliorations sont apportées continuellement à ceux existants, afin de répondre à la forte demande de logiciels plus performants pour le traitement des données. 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<sup>®</sup>.

Certaines études utilisent plusieurs méthodes de traitement des données qu'elles comparent par la suite. Cette comparaison porte sur les données prétraitées et leur capacité à classer les différents échantillons <sup>245, 269</sup>. Or la plupart des logiciels utilisés génèrent des données permettant de classer les échantillons mais peu ou pas d'études comparent l'ensemble des variables générées. La mise en œuvre de méthodes de traitement des données matures étant un prérequis à l'expansion de la métabolomique, il existe un besoin croissant de jeux de données de référence, permettant la comparaison objective des différents logiciels <sup>268</sup>. Les efforts actuels de standardisation des analyses métabolomiques, tel que la « *Metabolomic Standards Initiative* », soutenue par la société de métabolomique, ont pour objectif de proposer de tels jeux de données, afin de mettre fin à l'absence actuelle de méthode de comparaison reconnue<sup>271</sup>.

- **Normalisation** <sup>268</sup> : L'objectif de l'étape de normalisation des données est de supprimer les biais instrumentaux 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<sup>272</sup>, telles que la normalisation par l'écart-type<sup>240</sup> ou l'intensité moyenne globale<sup>273</sup> et la méthode du maximum de vraisemblance<sup>274</sup>.
- Celles reposant sur un ou de multiples composés de référence<sup>275, 276</sup>, 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 éluee).

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<sup>277</sup>. 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. 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<sup>278</sup>. 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 métabolite 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 isotopique uniforme pourraient être une réponse, mais si elles sont applicables à l'étude de « petits » métabolomes bactériens<sup>279, 280</sup> par exemple, la diversité des structures chimiques rencontrées dans le règne végétal ou chez les organismes pluricellulaires, associée à un nombre important de composés qui sont encore non identifiés, empêchent leur mise en œuvre pour des études d'organismes complexes. Une étude récente a néanmoins utilisé un double marquage isotopique <sup>15</sup>N et <sup>13</sup>C pour faciliter l'assignation des formules brutes aux signaux détectés dans des extraits d'*Arabidopsis Thaliana*<sup>281</sup>.

## 5 TRAITEMENT DES DONNEES

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. 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.

Il existe deux types d'analyse statistique multivariée utilisés 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 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.

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.

Dans le cadre d'approches métabolomiques, les données sont organisées sous forme de matrices. Chaque individu est décrit par un certain nombre de variables X, et parfois par des informations Y. Ainsi, les colonnes des matrices sont constituées des variables X et parfois Y dans le

cas d'analyses explicatives. Les lignes qui constituent les matrices correspondent aux individus (échantillons). 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.

L'avantage des trois méthodes citées précédemment est qu'elles permettent une visualisation rapide et puissante des données dans un espace à deux ou trois dimensions. Ce sont donc ces trois types d'analyses que nous allons décrire plus en détails.

### 5.1.1 Analyse en composantes principales (ACP)

L'ACP <sup>282,283</sup> permet de représenter les données originelles (individus et variables) dans un espace de dimensions inférieures à l'espace originel, tout en limitant au maximum la perte d'information. La représentation des données dans des espaces de faible dimension en 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 « carte factorielle des individus » ou « *score plots* », c'est-à-dire les coordonnées des échantillons dans le nouveau repère généré par l'ACP. A titre d'exemple, les échantillons présentés Figure 23 sont représentés dans un espace à trois dimensions autour des 3 premières composantes principales.

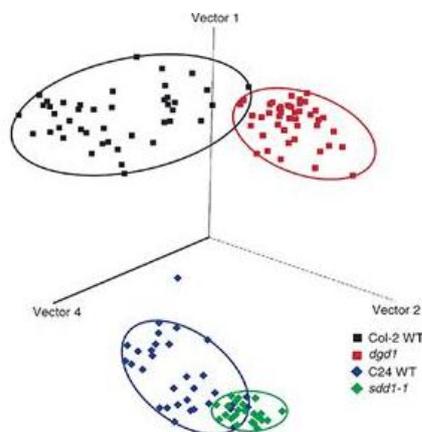


Figure 23 : Représentation des données lors des analyses ACP sous forme de carte factorielle des individus<sup>243</sup>

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.

## 5.1.2 Analyses explicatives et descriptives

### 5.1.2.1 Régression PLS

La PLS (Partial Least Square ou Projection to Latent Structure) fut développée vers la fin des années 60 par Herman Wold <sup>284</sup> et évolua jusqu'à sa forme actuelle en 1982. La PLS fournit un modèle mathématique qui permet d'expliquer un ou plusieurs 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 «carte factorielle des individus», c'est-à-dire les coordonnées des échantillons dans le nouveau repère généré par la PLS. On peut ainsi voir sur la Figure 24 la répartition des échantillons en deux groupes distincts. Les données peuvent également être étudiées sous forme de carte des poids factoriels ou « *loading plot* » qui décrivent les relations entre les variables dans le système des composantes principales (Figure 24).

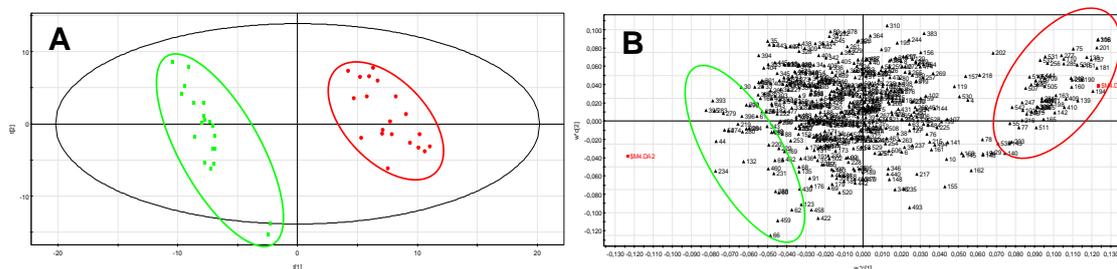


Figure 24 : Représentation des données lors des analyses multivariées sous forme de carte factorielle des individus (A) et de carte des poids factoriels (B)

### 5.1.2.2 *PLS-DA (Partial Least Square-Discriminant Analysis)*

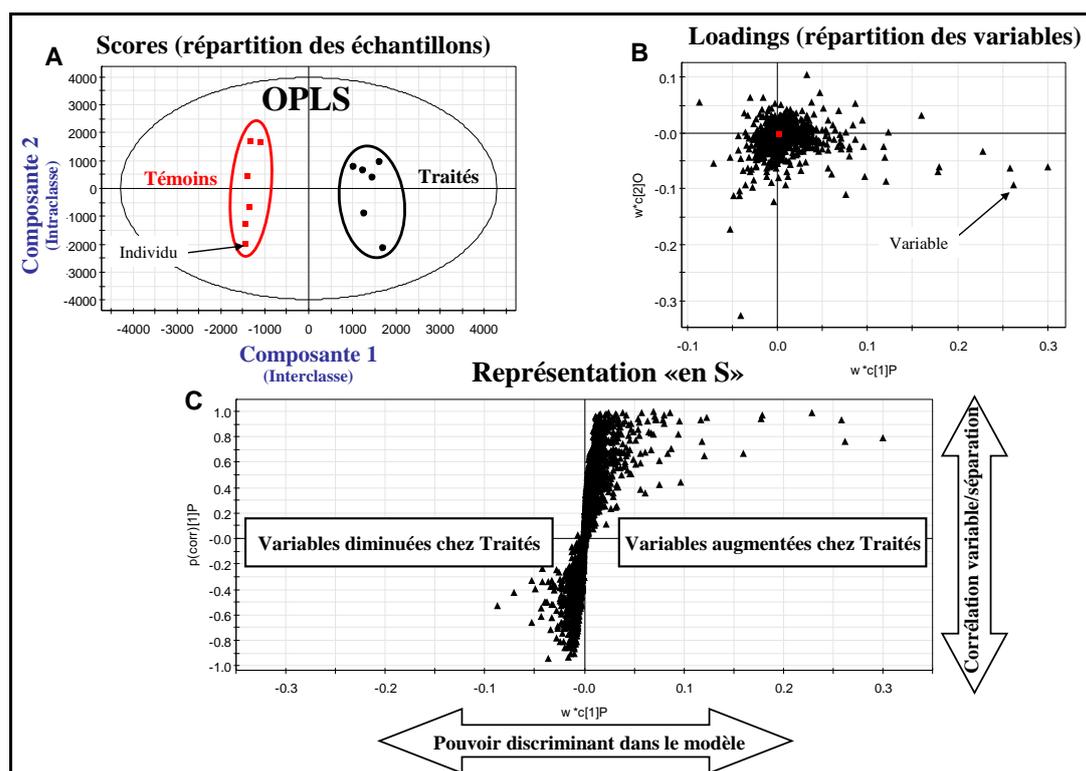
L'objectif de la PLS-DA ou analyse discriminante PLS est de construire un modèle qui permet de maximiser la séparation entre les classes auxquelles appartiennent les échantillons. Elle possède les avantages de la PLS : elle peut en effet gérer un nombre important de variables X, gérer les multicollinéarités et les données manquantes <sup>285</sup>. C'est une méthode proche de la PLS mais où la variable Y est catégorielle et non quantitative.

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 formation 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.

### 5.1.2.3 *OPLS (Orthogonal Partial Least Square)*<sup>286</sup>

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 reliée 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) <sup>287</sup>. La Figure 25 illustre les avantages de l'OPLS vis-à-vis de la PLS-DA. Les variations interclasse et intra-classe sont séparées en OPLS-DA alors qu'elles ne l'étaient pas en PLS-DA, ce qui facilite beaucoup l'interprétation du modèle OPLS-DA. Cloarec et al.<sup>288</sup> 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.

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.



**Figure 25: Présentation des données en OPLS**

(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 la première composante. (B) La carte factorielle du poids des variables («loading plot», en anglais) met en évidence les principales variables responsables du clustering précédemment observé. (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.

## 6 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. Elle requiert le plus souvent l'utilisation conjointe de plusieurs techniques analytiques, telles que la RMN, la MS, l'UV et l'IR <sup>289, 290</sup>.

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 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<sup>193, 194</sup>. Malheureusement, la spectrométrie de masse à pression atmosphérique présente une forte variabilité dans les profils de fragmentation générés sur différents appareils<sup>291</sup>, ce qui a limité la construction de banques de données spectrales universelles à l'instar de celles obtenues en ionisation électronique ou par RMN<sup>292</sup>.

En approche LC/MS, l'identification des signaux débute classiquement par la détermination de leur formule brute qui constitue un moyen simple, efficace et automatisable pour effectuer une interrogation des bases de données existantes. L'acquisition de spectres de fragmentation permet de discriminer les réponses obtenues précédemment sur la base d'ions produits ou de pertes de neutre, caractéristiques de groupement chimique. Certains composés peuvent encore être rejetés sur la base de leur temps de rétention et d'expériences complémentaires comme les échanges proton/deutérium. Au final, l'identification formelle est obtenue par injection de standards sur le même système analytique. Les standards doivent alors présenter le même temps de rétention, la même masse exacte et le même profil de fragmentation que le composé inconnu. En l'absence de réponse issue des bases de données, l'utilisation de la spectrométrie de masse seule pour l'identification des composés devient une tâche extrêmement complexe et il est très souvent nécessaire d'y associer d'autres techniques analytiques pour parvenir à une formule développée. Interviennent alors d'autres éléments de difficulté que sont les différences de sensibilité entre les méthodes. 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.

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

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 sont principalement originaires de la GC/MS par ionisation électronique, avec par exemple *the Golm metabolom database*<sup>293</sup> spécialisée dans les métabolites d'origine végétale, ou la bases de données spectrales du *National Institute of Standards and Technology* dont la base de données (NIST05) comporte les spectres de fragmentation de plus de 5000 composés.

Tableau 3: Principales bases de données utilisées en analyse métabolomique <sup>294</sup>

Database		Thematic	Conception/URL
BiGG	●	Human	University of California (USA) ( <a href="http://www.biigg.ucsd.edu/">www.biigg.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.csdb.mpg.de">www.csdb.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.nmr.fam.wisc.edu/">http://mmcd.nmr.fam.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/~gjoncs/mslib.htm">http://www.ualberta.ca/~gjoncs/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 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. Outre des données spectrales, elles indiquent le milieu et l'espèce dans lequel le composé a été retrouvé. Elles peuvent être généraliste comme HMDB (*Human Metabolom Database*) <sup>295</sup> ou ciblée sur un organisme (ex. *Atomic Reconstruction of Metabolism* (ARM) qui référence plus de 2700 réaction du métabolisme d'E.coli), ou encore spécifique d'une classe chimique de composés comme *LipidBank* <sup>296</sup> ([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 Génomes* (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<sup>297</sup>, 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 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*<sup>298</sup> et *LipidBank*, où l'on trouve des données issues de TQ, de pièges ioniques et de Qq-TOF pour les différents composés. En attendant la généralisation de cette approche multi-instrument, beaucoup de laboratoires ont recours à la construction de bases de données internes construites uniquement à partir de leurs instruments propres.

Cette base de données interne, très consommatrice de temps et de moyens, constitue un savoir-faire et une valeur ajoutée qui peuvent être valorisées commercialement, mais tendent également à ralentir les approches collaboratives en accès libre qui permettraient d'obtenir plus rapidement des bases plus complètes, indispensables à la mise en œuvre efficace des approches métabolomiques.



CHAPITRE III :  
TRAVAUX PERSONNELS



Après la mise en œuvre d'un protocole d'expérimentation animale consistant à exposer des groupes de 20 rats à différentes molécules inductrices, plus ou moins spécifiques des familles de cytochrome impliquées dans la métabolisation des xénobiotiques, nous avons tout d'abord conduit une comparaison des méthodes d'acquisition des profils métaboliques d'urines de rat par LC-MS sur les plateformes analytiques du CEA (HPLC-piège ionique) et de technologie Servier (UPLC-TOF), afin de sélectionner la méthode qui serait retenue pour l'analyse de l'ensemble des échantillons. Cette étape, réalisée sur le groupe de rats exposés au phénobarbital, outre l'aspect comparatif, a également permis d'effectuer des optimisations méthodologiques tant dans le domaine analytique (chromatographie et spectrométrie de masse), que dans celui du prétraitement et de l'analyse des données.

Dans un second temps, la nécessité d'identifier les signaux discriminants, mis en évidence par comparaison des urines de rat avant et après exposition au PB, nous a mené à développer de nouveaux outils d'organisation des signaux pour s'affranchir de la redondance des données caractéristique en spectrométrie de masse. L'application de ces outils et l'identification des signaux ont permis de caractériser de nombreux métabolites du phénobarbital, mettant en lumière la détection de biomarqueurs d'exposition au PB plutôt que celle de biomarqueurs d'induction enzymatique.

Dans une troisième étape, nous nous sommes intéressés à la mise en œuvre d'un outil de visualisation alternatif des données permettant de regrouper les signaux sur la base de leur composition chimique grâce à la spectrométrie de masse ultra-haute résolution. Nous avons ainsi appliqué les indices de Kendrick, développés pour l'analyse des hydrocarbures, aux urines pour l'organisation des signaux selon leur composition élémentaire. La diversité chimique du métabolome nous a conduit à y associer les modèles de rétention reposant sur la sélectivité méthylène afin d'améliorer leur apport en terme d'aide à l'élucidation structurale.

Enfin, la dernière partie de ce travail a consisté à caractériser les signaux de composés présumés endogènes, potentiels biomarqueurs d'induction enzymatique. Différentes approches statistiques ainsi qu'une stratégie d'identification combinant des expériences complémentaires en spectrométrie de masse ont ainsi été mises en œuvre.



## Liste des communications

### Articles (Annexe I)

Ducruix C, Vailhen D, Werner E, Fievet JB, Bourguignon J, Tabet JC et al. **Metabolomic investigation of the response of the model plant *Arabidopsis thaliana* to cadmium exposure: Evaluation of data pretreatment methods for further statistical analyses.** *Chemometrics and Intelligent Laboratory Systems* 2008; 91(1):67-77.

Werner E, Croixmarie V, Umbdenstock T, Ezan E, Chaminade P, Tabet JC et al. **Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution.** *Anal Chem* 2008; 80(13):4918-4932.

Werner E, Heillier JF, Ducruix C, Ezan E, Junot C, Tabet JC. **Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends.** *J Chromatogr B* 2008; 871(2):143-163.

### Communications Orales (sélection)

Erwan Werner<sup>1</sup>, Geoffrey Madalinski<sup>1</sup>, Thierry Umbdenstock<sup>2</sup>, Eric Ezan<sup>1</sup>, Jean-Claude Tabet<sup>3</sup>, Christophe Junot<sup>1</sup>. **Couplage LC-MS et analyse métabonomique : apport des nouvelles technologies de chromatographie et de spectrométrie de masse au profilage métabolique et à l'identification des métabolites.** 7<sup>ème</sup> congrès francophone de l'AfSep sur les sciences séparatives et les couplages, 20-22 mars 2007, Grenoble.

Erwan Werner, Geoffrey Madalinsky, Vincent Croixmarie, Thierry Umbdenstock, Eric Ezan, Jean-Claude Tabet, Christophe Junot. **Organisation et identification des signaux issus de l'analyse métabonomique par spectrométrie de masse : Apport de l'ultra-haute résolution et des outils mathématiques.** 24<sup>èmes</sup> Journées Françaises de Spectrométrie de Masse – Congrès de la société Française d'Electrophorèse et d'Analyse Protéomique (SMAP 2007), 17-20 juin 2007, Pau.

Erwan Werner, Eric Ezan, Pierre Chaminade, Jean-Claude Tabet, Christophe Junot. **Organisation et identification des signaux issus d'approches métabonomiques par spectrométrie de masse: Utilisation combinée des indices de Kendrick et de la sélectivité méthylène sur phase C<sub>18</sub>.** 3<sup>èmes</sup> Journées du réseau de métabolomique et fluxomique, 7-8 février 2008, Bordeaux.

### Affiches

Vincent Croixmarie, Thierry Umbdenstock, Françoise Dartiguelongue, Erwan Werner, Claire Boursier-Neyret, Bernard Walther. **A Metabolomics Approach to Assess in vivo Hepatic Induction in the Rat.** *Mebomeeting 3, Imperial college*, 18-19 Décembre 2006, Londres (GB).

Erwan Werner, Geoffrey Madalinski, Thierry Umbdenstock, Eric Ezan, Jean-Claude Tabet, Christophe Junot. **Combined use of a metabonomic approach by UPLC-TOFMS and ultra high resolution mass spectrometry for the characterization of Drug metabolites.** 55<sup>ème</sup> Congrès de la société américaine de spectrométrie de masse (ASMS), 2-7 juin 2007, Indianapolis (USA).

Geoffrey Madalinski, Erwan Werner, Denis Lesage, Sandra Alves, Jean Labarre, Eric Ezan, Jean-Claude Tabet, Christophe Junot (2007). **Evaluation of electrospray LTQ/Orbitrap mass spectrometry as a new tool for metabolite profiling.** 55<sup>ème</sup> Congrès de la société américaine de spectrométrie de masse (ASMS), 2-7 juin 2007, Indianapolis (USA).



# **1 EXPERIMENTATION ANIMALE**

## **1.1 Introduction.**

Les approches métabolomiques globales étant des analyses sans « a priori » du plus grand nombre de petites molécules présentes dans les fluides biologiques, tout facteur environnemental est susceptible d'avoir des répercussions sur les empreintes métaboliques générées. Il est donc nécessaire que l'environnement des individus puisse être contrôlé pendant la phase d'expérimentation, afin d'éviter que des différences de comportement, alimentaire par exemple, viennent impacter l'étude et introduire des facteurs confondants par rapport à l'effet qui veut être étudié.

En outre, l'étude du pouvoir inducteur des enzymes hépatiques du métabolisme d'une molécule, s'il s'agit d'un candidat médicament, étant une étape importante de son plan de développement, il a rapidement semblé sensé de mener notre approche sur une espèce de toxicologie. L'espèce la plus répandue et la plus utilisée dans les études de toxicologie n'est autre que le rat qui a donc été sélectionné. Ce choix présente plusieurs avantages. D'abord, les souches de rats utilisées en toxicologie sont très conservées de sorte que l'on peut comparer les résultats générés au cours des différentes études d'un même plan de développement mais aussi à travers les plans de développement de différentes molécules. Il en résulte que ce modèle est parfaitement caractérisé et que des données historiques permettent de s'assurer de la réussite des protocoles d'induction par exemple. Ensuite, les conditions d'élevage et d'expérimentation sont normalisées, ce qui limite l'influence des facteurs environnementaux qui pourraient venir obscurcir l'étude de l'effet d'intérêt. Et enfin, ces animaux de petite taille peuvent être accueillis en nombre, à un coût raisonnable, ce qui autorise à bénéficier d'un nombre d'individus relativement important. Les études de métabolomiques reposant sur l'analyse différentielle statistique d'empreintes métaboliques, le nombre d'individu conditionne directement leur réussite en déterminant la puissance statistique. La recherche de biomarqueurs indirects d'induction enzymatique devant mettre en évidence des modifications de concentration de métabolites endogènes, liées à l'augmentation de l'activité des enzymes hépatiques du métabolisme, une bonne puissance statistique sera nécessaire afin d'isoler l'impact de l'effet étudié des variations interindividuelles observables sur les métabolites endogènes.

## **1.2 Modèle animal**

Toutes les expériences ont été effectuées en accord avec le Décret sur l'Expérimentation Animale (Décret 87-848, 19 octobre 1987) sur des rats Wistar mâles (SPF Glx/BRL/Han) (Charles River, Domaine des Oncins, Saint-Germain-sur-l'Arbresle, France), âgés de 10 semaines et pesant entre 300 et 325 g au démarrage de l'étude.

### 1.2.1 Choix des composé inducteurs

Une fois l'espèce modèle arrêtée, les molécules employées pour les protocoles d'induction se sont imposées d'elles-mêmes, ces dernières étant d'usage classique chez notre partenaire et plus généralement en toxicologie animale. Le Tableau 4 présente l'ensemble des molécules utilisée ainsi que leur profil d'induction sur les enzymes hépatique du métabolisme des phases I et II.

Tableau 4 : Composés inducteurs sélectionnés et leur profil d'induction associé.

<u>Composé</u>	<u>Profil d'induction</u>
<b>PB</b> Phénobarbital	I: 2A1, 2B1 et 2B2, 2C, et 3A1 et 3A2 (mâle) et 2D1 (inhibition) II: UDP-glucuronyltransférase, Glutathion-S-Transférase
<b>INZ</b> Isoniazide	I: 2E1, par stabilisation de la protéine
<b>3-MC</b> 3-methylcholanthrène	I: 1A1 (+transcription via AhR) et 1A2, 1B1 (dépendante ou non de AhR)
<b>PCN</b> pregnenolone-16-carbonitrile	I: 3A1, 3A2, 3A9 et 3A18 II: UDP-glucuronyltransférase, Glutathion-S-Transférase
<b>BNF</b> beta-naphtoflavone	I: 1A (via AhR et autre voie Nrf2) II: cibles Nrf2: NAD-(P)H: quinone oxido-réductase, Glutathion S-transférase (sous-unité A), et hème oxygénase-1

### 1.2.2 Schéma expérimental

Afin de respecter la solubilité des composés chimiques utilisés comme agent inducteur, il a été nécessaire d'introduire deux groupes témoins correspondant aux véhicules d'administration, le chlorure de sodium pour les composés polaires et l'huile de maïs pour les composés lipophiles. Le Tableau 5 présente les différents groupes d'animaux et les termes de leur protocole d'induction.

Tableau 5 : Groupes d'animaux et conditions de traitement

Traitement	Témoin NaCl	PHENOBARBITAL	INZONIAZIDE
Groupe Animal	GROUPE 1	GROUPE 2	GROUPE 3
Nombre d'animaux	20 mâles	20 mâles	20 mâles
Souche	Wistar	Wistar	Wistar
Animaux n°	1 to 20	21 to 40	41 to 60
Dose en mg/kg/jour	0	80	200
Voie d'administration	Injection intra-péritonéale		
Volume (mL/kg)	2		
Véhicule	Chlorure de sodium 0.9% pour injection		
Fréquence - Durée	Une fois par jour pendant 4 jours		

Traitement	Témoïn huile	PCN	3MC	BNF
Groupe Animal	GROUPE 11	GROUPE 12	GROUPE 13	GROUPE 14
Nombre d'animaux	20 mâles	20 mâles	20 mâles	20 mâles
Souche	Wistar	Wistar	Wistar	Wistar
Animaux n°	61 to 80	81 to 100	101 to 120	121 to 140
Dose en mg/kg/jour	0	80	80	80
Voie d'administration	Injection intra-péritonéale			
Volume (mL/kg)	2			
Véhicule	Huile de maïs			
Fréquence - Durée	Une fois par jour pendant 4 jours			

### 1.2.3 Collecte des échantillons

Dans le cadre de cette expérimentation animale, il a été procédé au prélèvement de plusieurs fluides biologiques (urine, sérum) et de tissu (foie). Ces différents prélèvements ont servi à la réalisation des approches métabolomiques ainsi qu'à la caractérisation du modèle grâce à la détermination de paramètres biochimiques classiques et à la mesure des rapports des activités et des rapports d'expression des ARNm des cytochromes P450 les plus couramment impliqués dans le métabolisme des xénobiotiques, avant et après le protocole d'induction.

	Echantillons	Aliquots
Pré-dose (J-4 ou J-3)	Sérum (≈ 0.6 mL)	2 x 0.3mL
	Urine	500µl pour biochimie 2 x 1mL pour MBX
J+5	Sérum (≈ 2.5 mL)	5 x 0.5mL
	Urine	500µl pour biochimie 2 x 1mL pour MBX
	Foie	2 x 0.5g 1 x 6g

Sérum : Recueil sang (≈ 1.2mL en pré-dose et ≈ 6mL à J+5) dans des tubes contenant un séparateur de sérum et centrifugés rapidement après coagulation.

Urine : Collecte sur 16 heures en conditions réfrigérées

Foie : Collecte du foie entier puis séparation en 3 parties (2 x 0.5 g à partir du lobe gauche et le reliquat), immédiatement congelées par immersion dans l'azote liquide.

Tous les échantillons (sérum, urine et foie) ont été stockés à minimum -70 °C jusqu'à l'analyse

### 1.3 Caractérisation du modèle : Induction

La caractérisation du niveau d'induction induit par le protocole de traitement avec les différentes substances inductrices de référence a été effectuée grâce à la détermination du rapport des activités et de l'expression des ARNm des principaux CYPs participant au métabolisme des xénobiotiques avant et après traitement.

Les mesures de rapport d'activité (Tableau 6) sont obtenues par incubation de microsomes, préparés à partir du foie des animaux, avec des substrats spécifiques de certaines isoenzymes. Le principe repose sur le dosage de ces substrats et surtout de leurs métabolites, issus d'une réaction de biotransformation engendrée par une seule isoenzyme. Ce dosage est basé sur un couplage LC-MS/MS qui permet l'utilisation de substrats non radiomarqués. A noter que pour la caractérisation de l'activité du CYP2B, le dosage de la résorufine produite à partir de la pentoxyrésorufine est effectué par mesure de fluorescence avec  $\lambda_{\text{ex}} = 530\text{nm}$  et  $\lambda_{\text{em}} = 590\text{nm}$ .

**Tableau 6 : Mesure des activités enzymatiques des cytochromes P450 (ratio par rapport au control)**

	PB	INZ	PCN	3MC	BNF
Quantité de protéines	1.6	0.44	1.0	1.08	0.91
CYP1A (phénacétine)	na	0.93	na	4.3	3.6
CYP2B (pentoxyrésorufine, PROD)	56	1.0	1.5	2.1	1.7
CYP2B (16 $\beta$ -testostérone)	10	0.5	0	0	0
CYP2E (Chlorzoxazone)	na	2.7	na	na	na
CYP3A (6 $\beta$ -testostérone)	2.4	1.0	5.6	1.2	0.84
CYP3A (2 $\beta$ -testostérone)	4.4	0.33	7.0	2.0	0

1A : phénacétine dé-éthylase ; 2B : pentoxyrésorufine dé-éthylase et testostérone 16- $\beta$  hydroxylase ; 2E: chlorzoxazone 6-hydroxylation ; 3A : testostérone 2 et 6- $\beta$  hydroxylase.

La quantification de l'expression des ARNm des différentes isoformes a quant à elle été conduite par « reverse transcriptase – polymerase chain reaction » RT-PCR, à partir de prélèvements spécifiques réalisés dans le lobe gauche du foie des animaux (Figure 26).

L'ensemble de ces résultats confirme la réussite du protocole d'induction puisque la mesure des activités et les profils d'expression des ARNm montrent une augmentation de ces paramètres en accord avec le profil d'induction des composés de référence employés (Tableau 4). On notera toutefois la très grande disparité de l'effet inducteur des différents composés. Ainsi, les inducteurs les plus puissants en termes de quantité d'ARNm surexprimée sont le 3-MC et la BNF qui induisent une surexpression du CYP1A1 d'un facteur 1400 et du CYP1B1 d'un facteur 930 et 430, respectivement. A l'inverse, selon ces mêmes termes, ce sont le PB et l'INZ qui sont les inducteurs les moins puissants. Pour le dernier, l'absence d'augmentation de l'expression de son ARNm peut s'expliquer par le fait que le mécanisme d'induction n'est pas lié à une activation de l'expression du gène via un récepteur nucléaire et une séquence promotrice, mais participe de la stabilisation directe de la protéine par le composé.

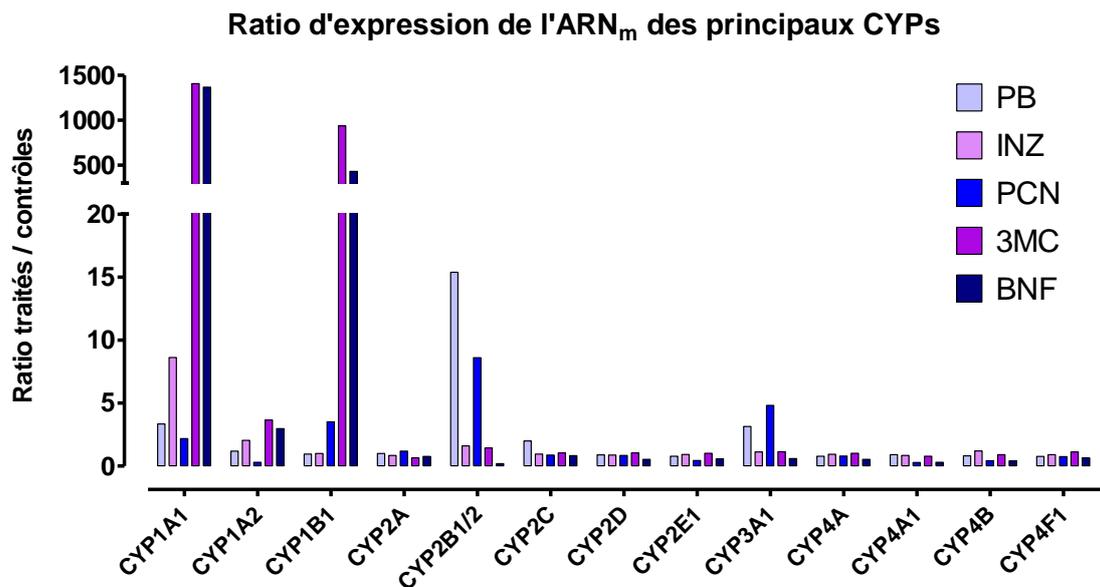


Figure 26 : Ratio d'expression des ARN<sub>m</sub> des principaux cytochromes P450 obtenus par RT-PCR.

Les résultats sont exprimés selon le calcul :  $Q_{\text{traités}} / Q_{\text{contrôles}}$

Plus étonnant, la mesure d'activité des CYPs révèle que l'augmentation d'activité de la protéine n'est pas du tout reliée à l'augmentation de l'expression de son ARN<sub>m</sub>. L'activité du CYP1A1 après traitement au 3-MC ou à la BNF n'est multipliée que par un facteur 4 malgré un ratio d'expression des ARN<sub>m</sub> de l'ordre de 1500, tandis que le PB induit une activité du CYP2B 56 fois supérieure pour un ratio d'expression de son ARN<sub>m</sub> d'un simple facteur 15.

Dès ce stade, on observe une différence de dynamique entre la transcriptomique et le phénotype qui explique pourquoi il est souvent si difficile de corréliser des résultats de génomique, de transcriptomique et de métabolomique. On comprend aussi immédiatement l'intérêt porté à la métabolomique qui est l'approche « omique » la plus proche du phénotype.



## 2 DÉVELOPPEMENTS MÉTHODOLOGIQUES

### 2.1 Introduction

Comme nous l'avons vu précédemment, les analyses métabolomiques globales ont pour objectif de mettre en évidence des variations biologiques correspondant à des modifications du statut physiologique des individus. Elles font appel à une association de méthodes analytiques pour l'acquisition des empreintes métaboliques, d'outils de traitement du signal pour l'extraction des signaux analytiques, et de méthodes statistiques afin de mener l'analyse différentielle des empreintes métaboliques en regard du statut physiologique des individus. La réussite de ce type d'approche repose donc à la fois sur la détection la plus exhaustive possible des petites molécules présentes dans les fluides biologiques analysés, sur la qualité de l'extraction des signaux ainsi générés nécessaire à l'obtention de jeux de données, ainsi que sur l'efficacité du traitement statistique appliqué à ce jeu de données pour la mise en évidence des signaux discriminants lors de l'analyse différentielle.

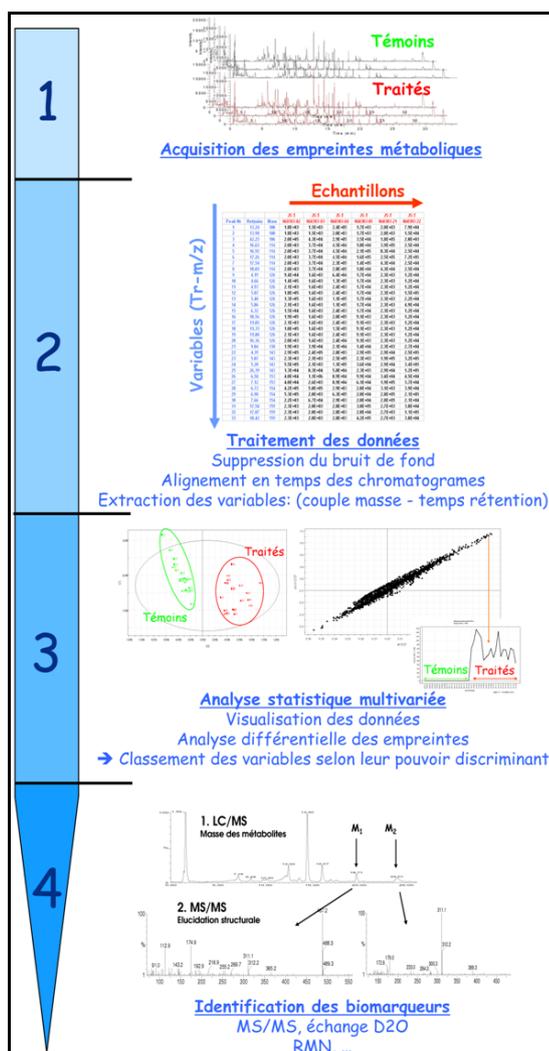


Figure 27: Déroulement d'une analyse métabolomique

La première partie de ce travail a donc logiquement consisté en une étape d'optimisation et de comparaison des méthodes analytiques et de traitement des empreintes métaboliques, développées au sein des deux groupes participant au projet, afin de sélectionner l'approche la plus susceptible de fournir des résultats pertinents en termes analytique et statistique avant d'envisager leur pertinence biologique. A cette fin, la méthode CEA (HPLC-Piège ionique / MetAlign<sup>®</sup>) et la méthode Servier (UPLC-TOF / Markerlynx<sup>®</sup>) ont été appliquées à l'analyse de sera de rats avant et après traitement par le phénobarbital.

## 2.2 Optimisation analytique

### 2.2.1 Matériel et méthode

#### 2.2.1.1 Réactifs

Le méthanol (qualité Lichrosolv) provient de la société Merck (Briare-le-Canal, France). L'acétate d'ammonium et l'acide formique (qualité réactifs pour analyse), la colchicine, la prednisone, l'acide déhydrocholique, sont fournis par la société Sigma (Saint-Quentin Fallavier, France).

#### 2.2.1.2 Echantillons d'urine de rats

Vingt rats ont été traités avec des injections intrapéritonéales de phénobarbital. Après une période d'acclimatation de 10 jours, les urines ont été collectées 3 jours avant le début des injections de PB et au 5<sup>ème</sup> jour de traitement. Elles ont ensuite été stockées à -80°C jusqu'à analyse.

#### 2.2.1.3 Appareillage

- **HPLC- Piège ionique**

Les analyses ont été réalisées sur un spectromètre de masse de type LCQ Duo provenant de la société Thermo-Fisher (Les Ulis, France). Ce spectromètre de masse comprend une source électrospray et un analyseur à piégeage d'ions (trappe 3D). Les séparations chromatographiques ont été réalisées sur une chaîne HPLC Surveyor<sup>®</sup> provenant également de chez Thermo-Fisher (Les Ulis, France). Les spectres enregistrés ont été traités à l'aide du logiciel Xcalibur<sup>®</sup> (version 1.4) provenant de la société Thermo-Fisher (Les Ulis, France). Le traitement des données a été réalisé au moyen du logiciel MetAlign<sup>®</sup>. Enfin, les analyses statistiques multivariées ont été menées à l'aide du logiciel SIMCA P11 (Umetrics, Suède).<sup>7</sup>

- **UPLC-TOF**

Les analyses ont été réalisées sur un spectromètre de masse de type LCT premier provenant de la société Waters (Saint-Quentin-en-Yvelines, France). Ce spectromètre de masse comprend une source électrospray équipée du système *lockspray*<sup>®</sup> et d'un analyseur à temps de vol. Les séparations

chromatographiques ont été réalisées sur une chaîne UPLC Acquity<sup>®</sup> provenant également de la société Waters. Les spectres de masses ont été acquis et traités à l'aide du logiciel MassLynx<sup>®</sup> (version 4.0, Waters). Le traitement des données a été réalisé au moyen du logiciel MarkerLynx<sup>®</sup> (Waters). Enfin, les analyses statistiques multivariées ont été réalisées à l'aide du logiciel SIMCA P11 (Umetrics, Suède).

#### 2.2.1.4 Méthode

- **HPLC-Piège ionique**

Les séparations chromatographiques ont été effectuées à 30°C en chromatographie liquide à polarité de phase inversée sur une colonne XTerra MS C<sub>18</sub> (2.1 x 150 mm, 5 µm) obtenue auprès de la société Waters (Saint-Quentin-en-Yvelines, France). L'élution des métabolites s'est faite en milieu neutre selon le gradient présenté dans le Tableau 7 :

Tableau 7: Présentation du gradient utilisé pour l'élution des métabolites urinaires sur le système HPLC-LCQ

Temps (minutes)	A (%)	B (%)
0	100	0
10	100	0
115	0	100
125	0	100
130	100	0
140	100	0

avec : A → H<sub>2</sub>O, AcONH<sub>4</sub> 10mM  
B → MeOH, AcONH<sub>4</sub> 10mM

Le débit de phase mobile, fixé à 0.2 mL/min, est maintenu constant sur l'ensemble de l'analyse. Avant d'être injectés dans le système chromatographique, les échantillons d'urine ont été surchargés avec une solution eau :méthanol (50 :50) contenant de la colchicine, de la prednisone et de l'acide déhydrocholique, à 3, 10 et 20 µg/mL, respectivement. Le volume d'échantillon injecté était de 20 µL.

La détection des métabolites a été effectuée simultanément dans les modes d'ionisation positifs et négatifs.

Les acquisitions ont été effectuées en mode balayage sur une gamme de rapports m/z allant de 75 à 1000 avec une résolution unitaire  $\Delta(m/z) = 1$ .

- **UPLC-TOF**

Les séparations chromatographiques ont été effectuées à 40°C en chromatographie liquide à polarité de phase inversée sur une colonne Acquity BEH C<sub>18</sub> (2.1 x 100 mm, 1.7 µm) obtenue auprès

de la société Waters (Saint-Quentin-en-Yvelines, France). L'élution des métabolites s'est faite en milieu acide selon le gradient présenté dans le Tableau 8, avec un débit de 0.4 mL/min.

**Tableau 8 : Présentation du gradient utilisé pour l'élution des métabolites urinaires sur le système UPLC-TOF**

Temps (minutes)	A (%)	B (%)
0	100	0
5	100	0
15	80	20
25	20	80
30	5	95
32	5	95
32.1	100	0
35	100	0

avec : A → H<sub>2</sub>O, Acide formique 0.1%  
B → ACN, Acide formique 0.1%

La détection des métabolites a été effectuée successivement dans les modes d'ionisation positifs et négatifs, au cours de 2 injections séparées.

Les acquisitions ont été effectuées en mode balayage sur une gamme de rapports m/z allant de 100 à 1000 et un pouvoir de résolution de l'ordre de 10000 (m/Δm).

## 2.2.2 Résultats

### 2.2.2.1 Optimisations méthode HPLC-piège ionique

- Optimisation des paramètres du spectromètre de masse :

Les paramètres du spectromètre de masse ont été optimisés à partir d'une solution méthanolique des composés utilisés comme références (colchicine, prednisone, et acide déhydrocholique) à 3, 10 et 20 µg/mL respectivement. L'optimisation de la température du capillaire et des débits de gaz a été effectuée au débit de la chromatographie grâce à la mise en place d'un système 3 voies.

Après optimisation, les paramètres les plus courants étaient en mode positif :

- Température du capillaire : 225°C
- Potentiel du cône: 4.5 kV
- Temps de remplissage maximum de la trappe : 50 à 100 ms.
- Débit du gaz principal : 80 (unités arbitraires)
- Débit du gaz auxiliaire : 20 (unités arbitraires)

- Optimisation de la durée du gradient.

La durée initiale du gradient étant de 140 min, et compte-tenu du nombre d'échantillons à analyser, une tentative de réduction de ce dernier à 60 min a été entreprise afin de limiter le temps d'analyse nécessaire à l'obtention des empreintes métaboliques. Cette réduction du temps d'analyse avait pour objectifs de limiter l'encrassement de la source avec le temps ainsi que d'améliorer la robustesse de la méthode en termes de répétabilité des temps de rétention (limitation des décalages de  $T_r$  au sein d'un batch analytique) afin de faciliter l'extraction des signaux analytiques (Cf. 3 Optimisation de l'extraction des variables). La maximisation de la répétabilité de la méthode est nécessaire afin d'éviter l'introduction d'un biais analytique lié à l'ordre d'injection des échantillons qui pourrait alors jouer le rôle de facteur confondant lors de l'analyse différentielle des empreintes métaboliques.

Dans un premier temps la réduction de la durée du gradient a été évaluée par l'injection répétée ( $n=3$ ) d'une urine témoin surchargée avec la solution de référence (colchicine, prednisone, et acide déhydrocholique). Cette première expérience a montré une conservation des coefficients de variations calculés à la fois pour les aires des références et leur temps de rétention.

Cependant, en conditions réelles d'analyse sur l'ensemble des urines, la réduction du temps d'analyse a montré l'introduction d'une importante variabilité sur l'aire des étalons (Tableau 9). Cette variabilité peut être attribuée au phénomène de suppression d'ionisation qui découle de l'augmentation du nombre de coélutions par suite de la réduction du temps de séparation. Avec des coefficients de variation (CV) sur l'aire des étalons compris entre 50 et 72 % pour l'ensemble de la séquence, le gradient 60 min est abandonné au profit de la conservation du gradient 140 min (CV sur l'aire des étalons compris entre 20 et 23 %) dans la mesure où le premier introduit une variabilité susceptible de perturber l'analyse différentielle des empreintes au moyen d'approches statistiques multivariées. A cette occasion, il est défini un critère d'acceptation analytique d'une séquence d'injection qui doit présenter un CV sur l'aire des étalons inférieur à 30 % pour être retenue.

## CHAPITRE III : Travaux personnels

**Tableau 9 : Impact de la durée du gradient sur la répétabilité des aires et temps de rétention des étalons.**

		Gradient 60 min						Mode négatif	
		colchicine		prednisone		ac dehydrocholique		ac dehydrocholique	
		Tr	aire	Tr	aire	Tr	aire	Tr	aire
run complet	Moyenne	30.47	1.38E+08	31.87	4.66E+07	30.21	6.47E+07	30.21	1.42E+07
	Ecart-type	0.14	7.10E+07	0.13	2.82E+07	0.16	3.64E+07	0.16	1.02E+07
	CV (%)	<b>0.44</b>	<b>51.57</b>	<b>0.42</b>	<b>60.69</b>	<b>0.54</b>	<b>56.27</b>	<b>0.54</b>	<b>72.12</b>
PD	Moyenne	30.47	1.79E+08	31.89	5.97E+07	30.22	7.89E+07	30.22	1.69E+07
	Ecart-type	0.14	2.81E+07	0.13	1.98E+07	0.14	2.03E+07	0.12	5.61E+06
	CV (%)	<b>0.46</b>	<b>15.67</b>	<b>0.40</b>	<b>33.15</b>	<b>0.45</b>	<b>25.70</b>	<b>0.40</b>	<b>33.23</b>
PT	Moyenne	30.48	1.10E+08	31.89	3.16E+07	30.19	4.02E+07	30.22	9.39E+06
	Ecart-type	0.12	3.34E+07	0.12	9.72E+06	0.18	1.88E+07	0.17	4.77E+06
	CV (%)	<b>0.39</b>	<b>30.49</b>	<b>0.37</b>	<b>30.71</b>	<b>0.61</b>	<b>46.68</b>	<b>0.55</b>	<b>50.79</b>
CQ	Moyenne	30.46	1.67E+08	31.84	6.30E+07	30.16	8.93E+07	30.14	2.17E+07
	Ecart-type	0.12	7.95E+07	0.15	3.21E+07	0.16	4.06E+07	0.19	1.31E+07
	CV (%)	<b>0.39</b>	<b>47.69</b>	<b>0.48</b>	<b>50.86</b>	<b>0.53</b>	<b>45.45</b>	<b>0.61</b>	<b>60.19</b>

		Gradient 140 min						Mode négatif	
		colchicine		prednisone		ac dehydrocholique		ac dehydrocholique	
		Tr	aire	Tr	aire	Tr	aire	Tr	aire
run complet	Moyenne	59.05	2.21E+08	62.88	7.46E+07	61.77	1.10E+08	61.76	8.33E+06
	Ecart-type	0.09	4.49E+07	0.10	1.69E+07	0.18	2.55E+07	0.16	1.74E+06
	CV (%)	<b>0.15</b>	<b>20.29</b>	<b>0.17</b>	<b>22.62</b>	<b>0.29</b>	<b>23.19</b>	<b>0.26</b>	<b>20.93</b>
PD	Moyenne	59.06	2.22E+08	62.89	7.38E+07	61.77	1.07E+08	61.76	7.84E+06
	Ecart-type	0.08	5.06E+07	0.10	1.86E+07	0.16	2.74E+07	0.13	1.65E+06
	CV (%)	<b>0.14</b>	<b>22.81</b>	<b>0.16</b>	<b>25.17</b>	<b>0.27</b>	<b>25.62</b>	<b>0.21</b>	<b>21.08</b>
PT	Moyenne	59.15	1.57E+08	62.85	7.88E+07	61.77	1.14E+08	61.76	8.92E+06
	Ecart-type	0.09	3.99E+07	0.10	1.44E+07	0.19	2.07E+07	0.16	1.64E+06
	CV (%)	<b>0.16</b>	<b>25.40</b>	<b>0.16</b>	<b>18.24</b>	<b>0.31</b>	<b>18.11</b>	<b>0.25</b>	<b>18.41</b>
CQ	Moyenne	59.00	2.42E+08	62.93	6.90E+07	61.78	1.19E+08	61.74	9.01E+06
	Ecart-type	0.09	5.29E+07	0.12	1.84E+07	0.22	3.05E+07	0.22	1.58E+06
	CV (%)	<b>0.15</b>	<b>21.86</b>	<b>0.19</b>	<b>26.72</b>	<b>0.36</b>	<b>25.66</b>	<b>0.36</b>	<b>17.54</b>

PD: prédose, PT: post traitement, CQ: Urine de référence servant de contrôle qualité

### 2.2.2.2 Comparaison des profils HPLC / UPLC

La comparaison des profils obtenus avec les deux méthodes est assez complexe et sujette à de nombreuses critiques dans la mesure où très peu de paramètres sont conservés. En effet, les différences technologiques (UPLC vs HPLC) et (TOF vs piège ionique), ainsi que l'emploi de phase mobiles différentes (milieu acide - acétonitrile vs milieu neutre - méthanol) ne permet pas une comparaison objective des performances des deux approches. En outre, la comparaison des deux méthodes qui avait arbitrairement été basée sur leur capacité à détecter un nombre maximal de métabolites nécessite la mise en œuvre complète de l'analyse métabolomique et inclut donc également l'étape de traitement du signal nécessaire à l'extraction des signaux détectés. En conséquence, le résultat de cette comparaison ne sera discuté qu'en fin de chapitre car de nombreux paramètres, ne dépendant pas uniquement de l'aspect analytique, sont à prendre en compte.

Néanmoins, afin de comparer de façon plus réaliste les deux méthodes, les échantillons ont été réinjectés en milieu neutre-méthanol sur le système UPLC-TOF (Figure 28). Comparativement à l'HPLC, l'UPLC apporte premièrement un gain de résolution chromatographique qui permet d'obtenir une meilleure qualité de séparation en un temps moindre. Ainsi le chromatogramme extrait correspondant aux dérivés N-glucuronidés du phénobarbital (Figure 29) montre que les deux isomères sont mieux résolus en UPLC qu'en HPLC, avec un retour complet à la ligne de base.

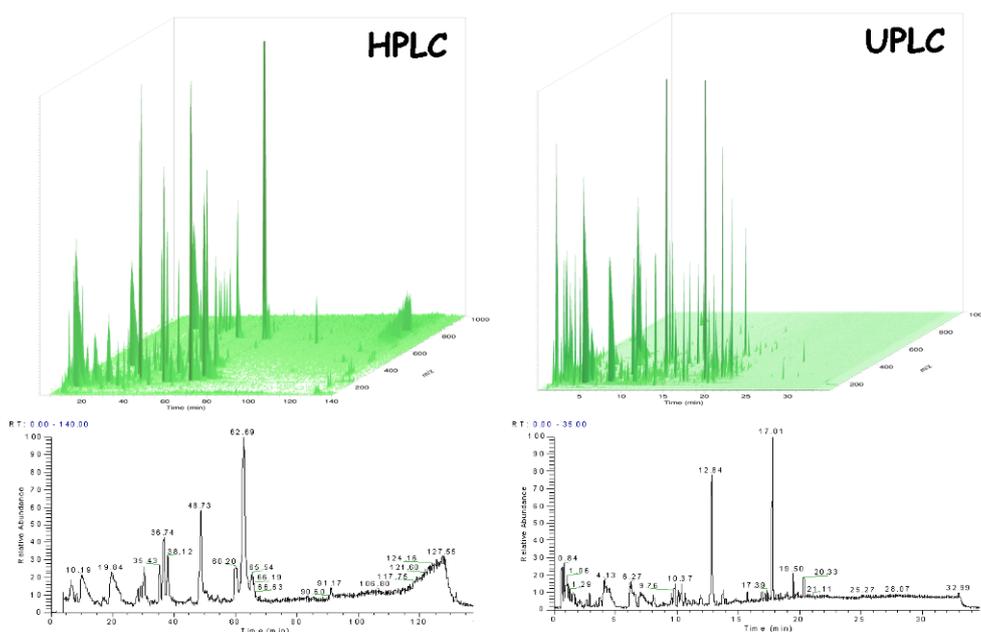


Figure 28 : Exemple de profil métabolique standard

Courant ionique total obtenu avec les deux méthodes (HPLC-piège ionique et UPLC-TOF) en mode négatif à partir de la même urine de rat et avec la même phase mobile (milieu neutre-méthanol).

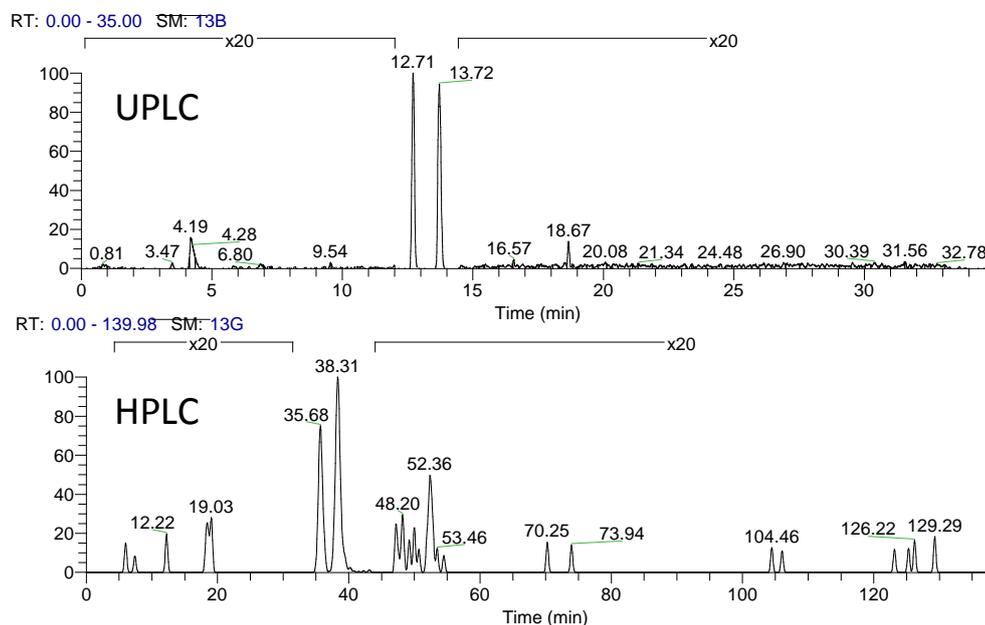


Figure 29 : Comparaison UPLC et HPLC.

Chromatogramme extrait ( $m/z$  407 - 407.5) obtenu avec les deux méthodes (HPLC-piège ionique et UPLC-TOF) en mode négatif à partir de la même urine de rat en conditions neutre-méthanol.

Deuxièmement, cette amélioration de la résolution et cette rapidité d'analyse s'accompagnent d'une réduction du temps d'éluion (finesse de pic) qui est à l'origine d'une amélioration des rapports

signal/bruit se traduisant par une amélioration de la sensibilité. L'amplification d'un facteur 20 du signal d'un facteur 20 de part et d'autre des pics correspondant aux glucuronides illustre la différence d'intensité des pics par rapport à la ligne de base. On notera également la présence d'au moins 2 pics à 4.19, et 18.67 min qui se détachent nettement de la ligne de base en UPLC alors qu'ils sont peu ou pas visibles en HPLC. Cet aspect peut être amplifié dans la mesure où la résolution du TOF participe également à l'amélioration du rapport signal/bruit lorsqu'on utilise une fenêtre d'extraction de rapport m/z réduite. Enfin, l'UPLC permet d'obtenir une excellente conservation des temps de rétention (Figure 31 B), laquelle autorise à s'affranchir de l'étape de réaligement des chromatogrammes lors du traitement du signal.

## 2.3 Optimisation de l'extraction des variables

### Optimisation MetAlign®

Le logiciel MetAlign® a été développé par A. Lommen du Plant Research Institute (Wageningen, Pays-Bas) pour le traitement des données de spectrométrie de masse au format non propriétaire NetCDF. Il est donc compatible avec l'ensemble des fournisseurs proposant la conversion des données sous ce format et présente entre autres avantages d'avoir été l'un des premiers logiciels disponibles et d'être gratuit.

Ce logiciel permet de :

- Détecter les signaux correspondants à des pics après lissage digital
- Calculer et réduire le bruit de fond par correction de la ligne de base en fonction du temps
- Aligner les chromatogrammes de façon itérative en se basant sur un nombre croissant de signaux communs dont le rapport signal/bruit initialement élevé s'amointrit au fur et à mesure que leur nombre augmente.
- Générer une matrice de données au format Excel® où pour chaque variable détectée (couple m/z-temps de rétention) est associée son amplitude maximale dans chaque échantillon
- Filtrer la matrice de données en fonction du ratio d'intensité moyenne des variables et/ou de leur rapport signal/bruit dans deux sous-groupes de chromatogrammes.

Les paramètres de détection des pics et de soustraction du bruit de fond doivent être optimisés en fonction de la technique analytique et des équipements employés. Cette optimisation a été effectuée deux fois afin de pouvoir traiter les données issues des deux méthodes.

L'interface utilisateur (Figure 30) est divisée en 3 parties (PART A, B, C) correspondant aux différentes fonctionnalités du logiciel :

PART A : Elle est dédiée à la correction de la ligne de base et à la suppression du bruit de fond. Elle repose sur l'élimination des signaux ne répondant pas aux critères de la définition d'un pic chromatographique selon l'utilisateur. Le chromatogramme est défini comme une suite adjacente de

spectres, l'utilisateur renseigne le premier et le dernier spectre à prendre en compte, l'amplitude maximale du signal détecté, la valeur du rapport signal/bruit à atteindre pour classer une variable comme appartenant à un pic et la largeur à mi-hauteur d'un pic chromatographique, exprimée en nombre de scan.

The screenshot displays the MetAlign software interface, organized into several sections:

- 1. Program configuration**: Includes buttons for selecting input data sets (2A, 2B, 3A, 3B) and a 'Clear' button.
- BASELINE AND NOISE ELIMINATION PARAMETERS**:
  - 4. Retention Begin (Scan nr): 150
  - 5. Retention End (Scan nr): 1800
  - 6. Maximum Amplitude: 1000000000
  - 7. Peak Slope Factor (x Noise): 1
  - 8. Peak Threshold Factor (x Noise): 2
  - 9. Average Peak Width at Half Height (Scans): 10
  - 10. Keep Peak Shape (no alignment):
  - 11. Run Baseline Correction: [Button]
- PART B: SCALING AND ALIGNING DATA SETS**:
  - 12. SCALING OPTIONS**:
    - No Scaling
    - Autoscaling on Total Signal
    - Scale on Marker Peak
    - From 1st Data Set, Group 1:
      - Nominal Mass: 100
      - Scan Nr.: 200
  - 13. INITIAL PEAK SEARCH CRITERIA**:
 

Scan Nr.	Max. Shift
Begin of 1st Region: 75	40
End of 1st Region: 1800	40
Begin of 2nd Region: 0	0
End of 2nd Region: 0	0
  - 14. TUNING ALIGNMENT OPTIONS AND CRITERIA**:
    - No Pre-align Processing (Rough)
    - Pre-align Processing (Iterative)
    - Calculation Criteria for Chromatography Shift Profiles:
      - 15. Maximum Shift per 100 Scans: 40
      - Mass Peak Selection:
 

1st Iteration	Last Iteration
16. Min. Factor (x Noise): 5	2
17. Min. Nr. of Masses: 10	5
  - SELECT MIN. NR. PER PEAK SET**:
    - 18. Group 1: 0 max.=19
    - or
    - 19. Group 2: 1 max.=1
  - 20. Run Scaling and Alignment: [Button]
  - 21. Detailed Ascii Output, Excel Compatible Output, Differential Retention Display: [Button]
- PART C: PEAK SELECTION AND EXPORT TO MS SOFTWARE FORMAT FOR VISUALISATION**:
  - PEAK SELECTION CRITERIA**:
    - 22. Significance Percentage: 99.00000
    - 23. Minimum Ratio between Means: 20
    - 24. Minimum S/N Ratio: 10
    - 25. Either in Gr. 1 or Gr. 2: >= 2 Masses/Compound
  - 26. FILTER ON CONDITION**:
    - Group 2 > Group 1
    - Group 1 > Group 2
  - 27. Run Peak Selection: [Button]
  - 28. Detailed Ascii Output, Excel Compatible Output, Masslynx Include List Output: [Button]
  - TOTAL PROCESSING**:
    - 29. Total Processing: [Button]
    - 30. Save and Exit: [Button]

Figure 30 : Interface utilisateur du logiciel MetAlign®.

PART B : A partir des variables sélectionnées précédemment, il s'agit de normaliser ou non l'intensité des variables au sein des différents échantillons en fonction soit d'un étalon soit de l'intensité totale du chromatogramme, ainsi que d'effectuer un réalignement en temps des variables afin de compenser les variations de temps de rétention dans les différents échantillons. L'utilisateur doit définir la méthode de normalisation, l'amplitude de correction en temps qu'il autorise ainsi que les critères de sélection des variables qui seront utilisées pour le réalignement itératif (nombre de variables par pic et leur rapport signal sur bruit, lors de la première et dernière itération). De plus, il existe un filtre pour la sélection des variables où l'utilisateur peut définir le nombre minimal d'échantillons où la variable doit apparaître afin d'être retenue. Enfin, il est possible de générer une matrice de données au format Excel® où pour chaque variable détectée (couple m/z-temps de rétention) est associé son amplitude maximale dans chaque échantillon.

PART C : Elle procède au tri des variables et à leur comparaison lorsque deux sous-populations d'échantillons ont été définies dans la partie A. Le logiciel calcule la moyenne de l'intensité de chaque variable pour chacun des sous-groupes, il conduit un test de Student pour chacune des variables, et calcule le rapport des moyennes des sous-groupes. Il est ainsi possible de sélectionner les variables en fonction de leur sens de variation dans les 2 groupes, du rapport de leur moyenne, et de leur rapport signal sur bruit. En pratique, cette partie n'a été que très peu utilisée dans la mesure où on lui a préféré les statistiques multivariées. En outre, la réalisation de ces calculs reste possible à partir de la matrice complète de données.

L'optimisation des parties A et B a pour objectif de produire un jeu de données de la meilleure qualité possible. Il s'agit d'extraire le maximum de variables afin de ne pas perdre d'information, tout en évitant la sélection d'éléments du bruit de fond dont le caractère aléatoire pourrait venir interférer avec l'analyse différentielle des empreintes. Il est par ailleurs important de ne pas échouer dans le réaligement des variables car dans le cas contraire, le traitement des données peut générer plusieurs variables correspondant au même signal issu d'un seul métabolite. Ceci est la conséquence de la réduction de la dimensionnalité des données de trois (temps de rétention,  $m/z$ , intensité) à deux dimensions (couples  $Tr$ - $m/z$ , intensité), nécessaire à l'application des statistiques multivariées comme la PCA ou la PLS. Ainsi, en cas de non alignement on pourra créer deux variables  $m/z$  201 à 2.05 min et  $m/z$  201 à 2.32 min alors qu'il n'existe qu'une seule variable à  $m/z$  201 à 2.05 min mais qu'il s'est produit un brusque changement de temps de rétention au milieu de la séquence, qui n'a pu être compensé par le réaligement. De nombreux phénomènes et les paramètres choisis peuvent être à l'origine de la perte ou de la génération artificielle de variables. Dans la partie A de la Figure 31, le retraitement de deux urines de rat traités par le PB montre pour la première la génération entre 62 et 67 minutes de 4 variables avec  $m/z = 231$  tandis que pour la seconde une seule variable à 63.11 min qui ne correspond à aucune des 4 précédentes est présente.

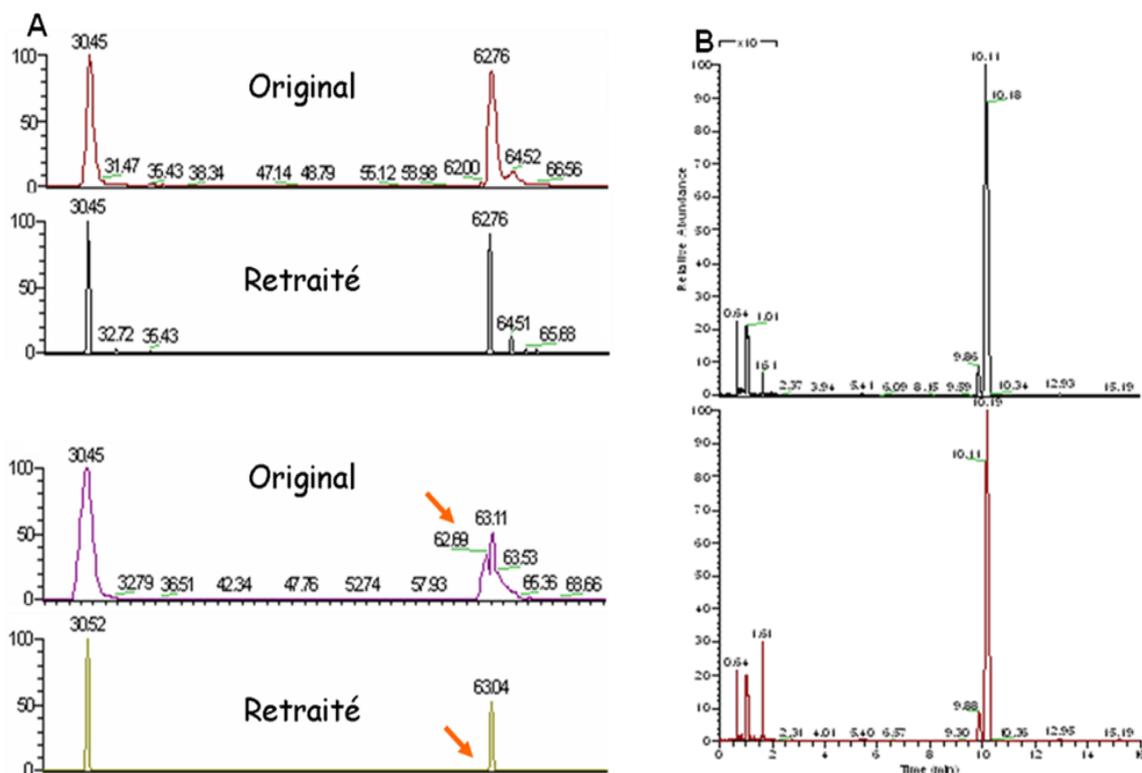


Figure 31 : Comparaison des chromatogrammes obtenus avant et après traitement par le logiciel MetAlign®

A : HPLC-piège ionique, B : UPLC-TOF. Chromatogrammes extraits ( $m/z$  231 = PB) obtenus à partir de deux urines de rat traité au PB et analysée en mode négatif.

La Figure 31 illustre également, pour la partie B, l'intérêt de l'UPLC-TOF qui en l'absence de tout retraitement, présente pour les deux urines des pics fins, symétriques, avec un très bon rapport signal/bruit et une parfaite analogie de temps de rétention.

L'une des principales limitations de l'utilisation du logiciel MetAlign® avec les données issues du système UPLC-TOF réside dans sa capacité à ne traiter que des masses nominales. Ainsi, il y a perte de la mesure de masse précise autorisée par le TOF lors du traitement des données. Une nouvelle version du logiciel, capable de traiter les masses précises, non disponible au moment du traitement des données, a levé cette limitation.

Après optimisation individuelle des paramètres par comparaison visuelle des chromatogrammes obtenus avant et après retraitement, les paramètres suivants ont été retenus comme paramètres génériques (Tableau 10). Cependant, en fonction des conditions analytiques (polarité d'ionisation, nature de la matrice, ...) ils peuvent être adaptés pour permettre le réalignement et augmenter ou limiter le nombre de signaux détectés.

Tableau 10 : Paramètres optimisés d'extraction des signaux avec le logiciel MetAlign®.

		HPLC	UPLC	
<b>Partie A</b>	Scan initial	100	50	
	Scan final	3700	4800	
	Amplitude	1000000000	100000	
	Facteur de pente	1	1	
	Seuil de détection de pic (x bruit)	2-3	5	
	Largeur de pic à mi-hauteur (scan)	15	10	
<b>Partie B</b>	Normalisation	Signal total	Signal total	
	Alignement	Itératif	Itératif	
	Correction d'alignement (scan/100)	40	25	
	S/N minimal	Itération 1	5	15
		Itération n	2	10
	Nombre minimal de m/z	Itération 1	4	4
Itération n		3	3	

Afin de pallier la perte de la mesure de masse précise avec le logiciel MetAlign®, les données obtenues avec le système UPLC-TOF ont également été traitées avec le logiciel MarkerLynx® (Waters). Ce dernier possède également de nombreux paramètres qui ne seront pas détaillés ici. En termes de différences majeures avec MetAlign®, ce dernier n'est pas capable de réaligner les chromatogrammes, il permet seulement de regrouper les rapports m/z égaux (tolérance déterminée en fonction de l'appareil) au sein d'une fenêtre de temps définie par l'utilisateur sous la même variable. Ce point est une limitation le plus souvent mineure grâce à la répétabilité des temps de rétention avec le système UPLC et à la précision de mesure en masse du TOF. En revanche, le logiciel fonctionne comme une relative « boîte noire » et l'optimisation des paramètres est peu intuitive. De plus, l'absence d'outil de visualisation des données avant et après traitement rend cette optimisation délicate. MarkerLynx® possède une fonction supplémentaire grâce à l'exploitation de la mesure de masse précise, puisqu'il permet la suppression des isotopes de la table des résultats. Cette étape consiste en la détection d'un delta de masse de 1.0034 (correspondant à la différence entre les masses des isotopes 12 et 13 du carbone) entre deux signaux présents au même temps de rétention. Lorsque de tels signaux sont détectés, le logiciel élimine de la matrice de données les variables correspondant au cluster d'ordre 1 du carbone, supprimant ainsi une partie de la redondance du signal observée en spectrométrie de masse (plusieurs signaux pour un même composé).

Le nombre de variables extraites avec chacune des méthodes (CEA ou Servier) est reporté dans le Tableau 11.

Tableau 11 : Nombre de variables extraites selon le logiciel et la méthode analytique employés.

	Nombre de variables extraites			
	MetAlign <sup>®</sup>		MarkerLynx <sup>®</sup>	
	Mode Positif	Mode Négatif	Mode Positif	Mode Négatif
HPLC-piège ionique (phase neutre)	2548	546	NA	NA
HPLC-piège ionique (phase acide)	1734	623	NA	NA
UPLC-TOF (phase acide)	5203	ND	3427 *	3600 *
UPLC-TOF (phase neutre)	-	2151	-	2464 *

\* après suppression des isotopes

NA: non applicable; ND: non déterminé

Le nombre important de variables associé aux différences techniques et analytiques (nature des phases mobiles) ne permettant pas d'évaluer les méthodes grâce à la détermination de l'équivalence des variables à cette étape du traitement, les échantillons ont été réinjectés en milieu neutre sur le système UPLC-TOF (acquisition en mode négatif uniquement) et en phase acide sur le système HPLC-piège ionique. Le mode négatif avec une phase mobile neutre-méthanol semble être assez favorable à la détection du PB et de ses métabolites d'après les résultats obtenus en HPLC.

La comparaison du nombre de variables extraites avec le logiciel MetAlign<sup>®</sup> pour le mode négatif en condition neutre-méthanol semble montrer une très nette supériorité de l'UPLC sur l'HPLC avec respectivement 2151 contre 546 variables extraites.

Cependant, de nombreuses variables détectées en HPLC-piège ionique ne sont pas retrouvées parmi celles issues du système UPLC-TOF. De plus, ce dernier semble plus favorable à la formation d'adduits avec le solvant organique et les ions de la phase mobile. La grande différence de variables détectées en UPLC-TOF pourrait donc en partie s'expliquer par la génération du signal plus redondant avec ce système. La comparaison des méthodes s'effectuera donc par leur capacité à mettre en évidence le plus grand nombre de variables discriminantes par analyse différentielle des empreintes métaboliques issues de l'analyse des urines avant et après exposition au PB.

## 2.4 Analyses statistiques multivariées

A partir des matrices de données générées précédemment, l'analyse différentielle des profils métaboliques urinaires avant et après exposition des rats au PB a été conduite à l'aide d'analyses statistiques multivariées.

En première approche, une ACP est réalisée afin de visualiser l'organisation des échantillons dans un espace à deux dimensions. Cette analyse permet d'évaluer les sources principales de variance des échantillons. Ainsi plus les échantillons seront séparés sur des composantes de faible rang, plus le phénomène à l'origine de la séparation sera explicatif d'une part importante de la variance globale du pool d'échantillons analysés. L'absence de séparation des échantillons en ACP

selon l'effet étudié montre que ce dernier n'est responsable que de façon négligeable de l'ensemble des variations entre les échantillons.

Dans un deuxième temps, est généralement appliquée une méthode descriptive comme la PLS-DA (plusieurs groupes d'échantillons à comparer) ou l'OPLS (deux groupes d'échantillons). Ces méthodes permettent de lier un effet aux variables initiales et de les classer en fonction leur importance relative (poids) dans l'explication de ce dernier. Par rapport à la PLS-DA, l'OPLS présente l'avantage de séparer les variables explicatives (matrice des X) en deux sous-groupes : le premier comportant les variables corrélées à l'effet étudié (Y) et le second comprenant les variables orthogonales à l'effet Y étudié. Il s'ensuit que la séparation des groupes selon l'effet Y est toujours placée sur la composante 1 (explicative de la variance inter-classe) tandis que les facteurs explicatifs de la variance des échantillons orthogonaux à Y sont placés sur la composante 2 (explication de la variance intra-classe).

En dehors d'une étape de normalisation des données, peu de paramètres peuvent être optimisés à ce niveau. Le succès de l'outil dépend avant tout de la qualité de la matrice de données (nombre de données manquantes, représentation du bruit par rapport aux signaux explicatifs, ...) ainsi que du plan expérimental. Ce dernier, longtemps négligé, est un élément crucial de la réussite de l'approche car il conditionne la présence ou non d'une grande partie des facteurs confondants pouvant être à l'origine du masquage de l'effet étudié. En effet, il est nécessaire que sa construction permette de s'affranchir de tous les facteurs non maîtrisés afin d'être en mesure d'isoler l'effet étudié lors de la comparaison des échantillons.

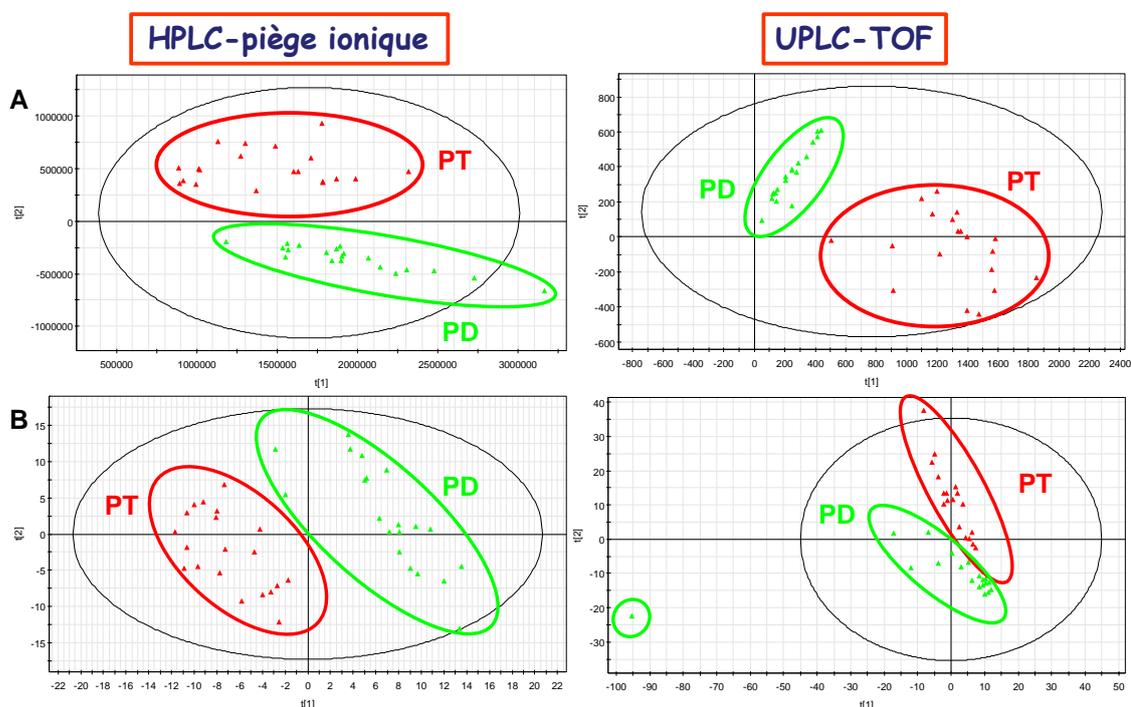
Différentes normalisations peuvent être appliquées :

- Aucune normalisation, on travaille avec l'intensité des variables telle qu'elle est introduite dans la matrice de données.
- Le centrage – réduction (UV pour Unit Variance) : chaque variable est centrée-réduite de sorte que la covariance ne soit plus un élément discriminant dans le model.
- Le centrage seul
- La normalisation Pareto : Méthode intermédiaire entre l'absence de normalisation et le centrage-réduction, la variable est centrée mais le facteur de pondération est la racine carrée de son écart-type. Cette normalisation présente l'avantage de lisser le poids des covariances des variables mais sans le supprimer. Elle est donc intéressante lorsqu'il existe une forte disparité dans les intensités des variables, ce qui est le cas pour les données issues de la spectrométrie de masse.

Dans la pratique, toutes les méthodes de normalisation peuvent être employées et sont testées de façon empirique, le choix final étant guidé par la validation ou non du model mathématique généré, de sa capacité à expliquer la variance globale des échantillons, et surtout de sa capacité prédictive.

### 2.4.1 Analyse en composantes principales (ACP)

L'ACP montre une très nette séparation des individus suivant qu'ils ont été ou non exposés au PB. La Figure 32 présente la carte factorielle des individus sur la première et la deuxième composante sans normalisation des données (A) et après centrage-réduction (B). En l'absence de normalisation, l'effet traitement explique la majorité de la variance du jeu de données issu du système UPLC-TOF (séparation des individus selon la première composante) tandis qu'il est la seconde cause explicative de la variance du jeu de données obtenu par HPLC-piège ionique (séparation des individus selon la deuxième composante). En revanche, après centrage et réduction des données, l'effet traitement permet de séparer les échantillons sur la première composante en HPLC-piège ionique tandis qu'il est réparti sur les deux composantes en UPLC-TOF. On remarque par ailleurs que la normalisation fait ressortir un échantillon très différent parmi les urines pré-doses en UPLC-TOF.



**Figure 32 : Cartes factorielles des individus obtenues par ACP.**

Modèles obtenus à partir des résultats générés en mode négatif par HPLC-piège ionique et UPLC-TOF en milieu neutre-méthanol à partir de 20 urines de rats collectées avant et après traitement au PB. PD : pré-dose ; PT : post-traitement. Les modèles ont été construits sans normalisation des données (A) et après centrage-réduction des variables (B).

Les cartes des poids factoriels (Figure 33) correspondant à ces modèles, qui représentent le poids relatif de chaque variable initiale sur les composantes 1 et 2, montrent que la majorité de la variance des échantillons en HPLC-piège ionique sans normalisation est liée à quelques variables dont l'intensité est extrêmement importante. Le fait de centrer et réduire les variables supprimant la notion de covariance, l'effet traitement qui se traduit par des variables présentant une moindre amplitude de variation de signal mais une meilleure corrélation entre la classe (pré-dose / traité) et les valeurs mesurées pour chaque sous groupe se trouve alors renforcé.

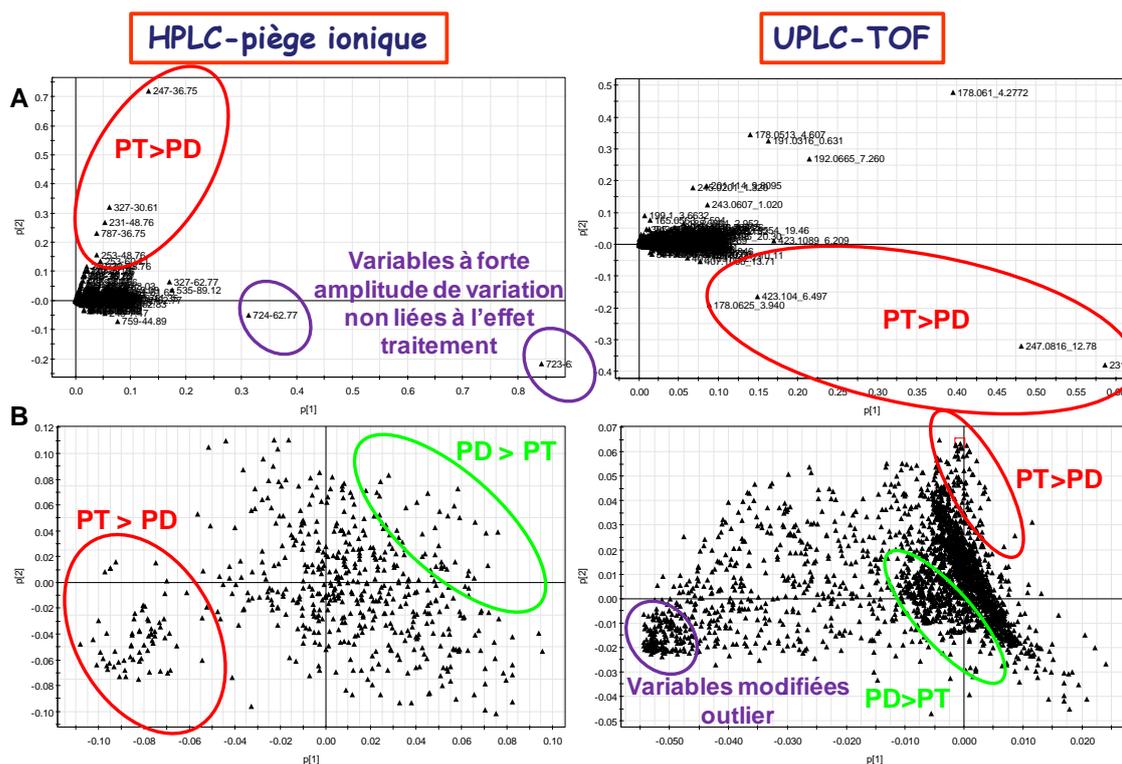


Figure 33 : Cartes des poids factoriels obtenus en ACP.

Modèles obtenus à partir des résultats générés en mode négatif par HPLC-piège ionique et UPLC-TOF en milieu neutre-méthanol à partir de 20 urines de rat collectées avant et après traitement au PB. PD : pré-dose ; PT : post-traitement. Les modèles ont été construits sans normalisation des données (A) et après centrage-réduction des variables (B).

## 2.4.2 Analyse OPLS (Orthogonal Partial Least Square)

L'ACP ayant permis de s'assurer que les individus sont naturellement structurés en classes d'après leur profil urinaire selon l'effet traitement, l'analyse des variables responsables de cette séparation a ensuite été approfondie par l'analyse OPLS en choisissant l'effet traitement comme variable catégorielle Y.

Les résultats obtenus pour l'HPLC-piège ionique sont présentés dans la Figure 34 à titre d'exemple. A partir de la carte factorielle des individus (Figure 34A), on observe comme attendu la séparation selon l'effet traitement (Y) sur la première composante, explicative de la variance inter-classe, la seconde permettant de modéliser la variance intra-classe. La Figure 34B présente les résultats explicatif et prédictif du modèle. Le Q2, égal à 0.98, montre que 98 % de la variance, liée à Y (effet traitement), est expliquée par le modèle tandis que le R2, égal à 0.96, reflète la capacité du modèle à prédire la classe sur Y d'un nouvel échantillon à partir des variables originelles sélectionnées. La Figure 34C, quant à elle, correspond à la validation Anova du modèle, où la p-value ( $1.3 \times 10^{-16}$ ) signe la très grande significativité statistique du modèle. La carte des poids factoriels est représentée sous forme de « S-plot » qui présente l'avantage de visualiser les variables selon l'amplitude de leur variation en abscisse ( $W^*c$ ) et selon la corrélation de ses valeurs avec la classe en ordonnées ( $p(\text{corr})$ ) (Figure 34D). Ainsi plus une variable originelle s'éloigne de l'axe des ordonnées

plus son amplitude de variation est importante, dans le cas présent plus elle s'éloigne vers la droite plus elle est augmentée chez les individus traités (PT), et plus elle s'éloigne vers la gauche, plus elle est élevée chez les pré-doses et donc diminuée chez les traités. Ensuite, plus une variable s'éloigne de l'axe des abscisses plus ses valeurs sont corrélées avec la classe, de sorte que les variables dont les valeurs sont peu chevauchantes entre les classe présentera un  $p(\text{corr})$  élevée. La variable, augmentée chez les traités, possédant la plus grande amplitude de variation entre les pré-doses et les traités et dont les valeurs pour les différents individus au sein d'une même classe sont peu dispersées se retrouvera donc dans le coin supérieur droit. Ici, il s'agit de  $m/z$  247, élué à 36.75 min. Les valeurs individuelles des échantillons pour cette variable sont représentées dans la Figure 34E et confirment l'intérêt du S-plot pour la sélection des variables les plus pertinentes.

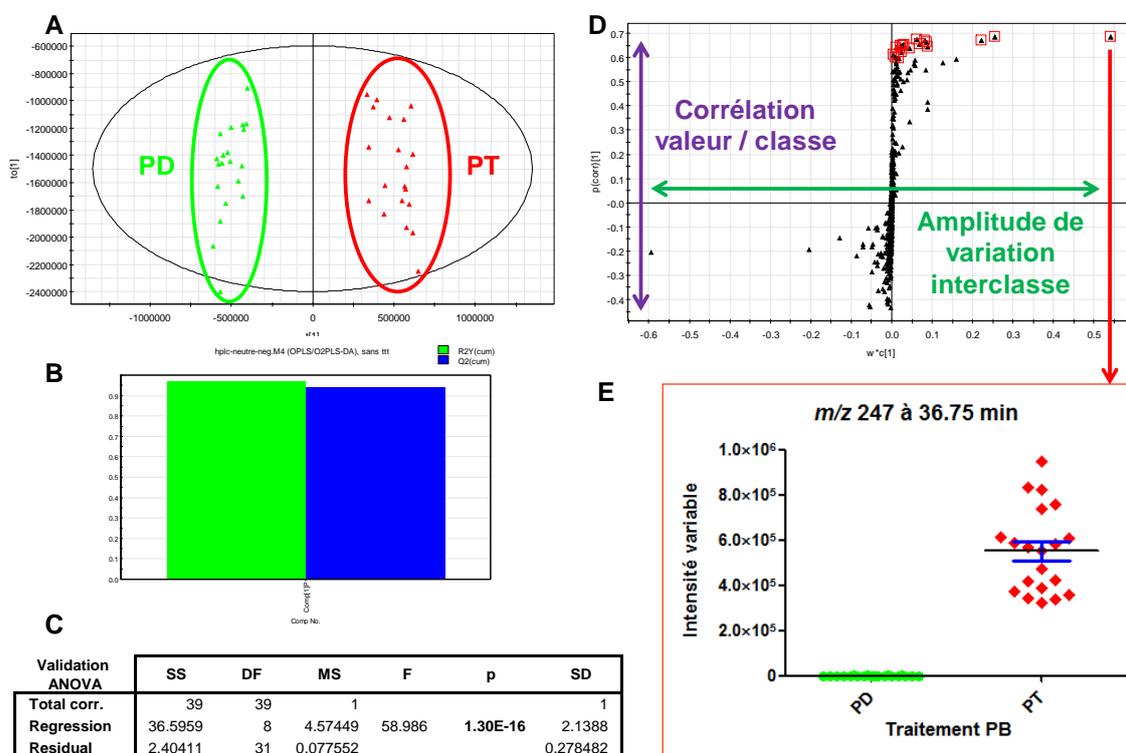


Figure 34 : Analyse OPLS des données HPLC-piège-ionique en mode négatif - milieu neutre méthanol.

A : Carte factorielle des individus sans normalisation des données, PD : pré-dose ; PT : post-traitement. B : Paramètres du modèle avec  $R^2=0.98$  et  $Q^2=0.96$ . C : Table de validation ANOVA du modèle. D : Représentation de la carte des poids factoriels sous forme de S-plot. E : Représentation des valeurs individuelles pour la variable  $m/z$  247 à 36 ;75 min qui est la variable la plus discriminante.

Les modèles explicatifs comme la PLS ou l'OPLS, à l'inverse des descriptifs, permettent par ailleurs de classer les variables originelles selon leur importance dans la construction du modèle. SIMCA<sup>®</sup> renvoie ainsi une valeur pour chaque variable, dénommée *variable importance on projection* (VIP), qui est supérieure à 1 pour les variables dont la contribution au modèle est significative.

Par expérience, et pour se focaliser sur les variables les plus contributives au modèle, n'ont été retenues que les variables avec un  $VIP > 1.1$ . En mode négatif neutre, il a ainsi été isolé 138 variables significatives en HPLC-piège ionique et 410 en UPLC-TOF.

L'analyse des valeurs individuelles des variables, les plus contributives au modèle, a montré que leur importance était liée au fait qu'elles soient absentes chez les pré-doses et n'apparaissent que dans les urines des individus traités (Tableau 12). La séparation des échantillons selon l'effet traitement peut correspondre à plusieurs phénomènes : 1- marquage d'exposition au composé administré avec la présence du composé inchangé et de ses métabolites chez les individus traités seulement, 2- la modification de métabolites endogènes pouvant traduire un effet pharmacologique ou toxicologique du composé administré et enfin 3- la modification de métabolites endogènes correspondant à l'induction des enzymes hépatiques du métabolisme après administration du composés inducteur, effet que nous cherchons à isoler.

Tableau 12 : Nombres de variables discriminantes issues de l'analyse différentielle des empreintes métaboliques.

Système analytique	Logiciel d'extraction	Variables extraites	Variables discriminantes VIP > 1.1		
			Total	Présentes PD et PT	Absentes PD
HPLC-piège ionique	MetAlign <sup>®</sup>	546	138	85	53
UPLC-TOF	MarkerLynx <sup>®</sup>	2464	410	368	42

La séparation étant principalement due à l'exposition au composé puisque les variables les plus discriminantes sont absentes des empreintes pré-doses, il faut d'abord chercher à s'affranchir de ces signaux avant de pouvoir accéder aux autres phénomènes. Les variables discriminantes (VIP>1.1) ont donc été triées selon un filtre qui consistait à attribuer au métabolisme du PB tous les signaux absents chez les pré-doses. Ce filtre n'est pas parfait car certains métabolites endogènes, présents à faible concentration ou possédant des facteurs de réponse faibles en spectrométrie de masse, peuvent produire des signaux inférieurs à la limite de détection en pré-dose, puis voir leur concentration augmentée par l'effet pharmacologique, par exemple, et être par conséquent classés comme reliés à l'exposition au PB.

## 2.5 Identification des variables d'intérêt

Une fois les variables d'intérêt isolées, il faut les identifier afin de leur donner une signification biologique. Cette identification repose en spectrométrie de masse couplée à la LC sur l'obtention de la formule brute du composé à l'origine du signal détecté. Elle s'appuie en fonction des appareils sur la mesure de masse précise et/ou l'obtention de spectres de fragmentation.

Dans notre cas, le TOF nous donne accès à la masse précise tandis que le piège ionique permet de mener des études structurales par l'acquisition de spectres CID. Il est donc nécessaire de réaliser un « matching » des données entre les deux appareils. Or nous avons vu au cours des étapes précédentes que cette opération, au-delà du fait qu'elle soit très consommatrice en temps, était pratiquement irréalisable en raison des différences de sensibilité, de nature de phase stationnaire chromatographique, de géométrie de source,... qui conduisent à la détection de signaux qui ne sont

que partiellement chevauchant pour les mêmes métabolites sur les deux systèmes. En l'absence d'une donnée commune très discriminante pour s'assurer de la détection du même composé sur les deux systèmes, leur utilisation complémentaire pour l'identification ne peut réalistement pas être envisagée.

Si a posteriori (cf chapitre II), les résultats obtenus avec les deux systèmes sont très comparables en terme de couvertures de métabolites exogènes détectés (Figure 35), la répétabilité des injections, la meilleure sensibilité et l'obtention d'une mesure de masse précise nous ont conduit à choisir le système UPLC-TOF pour la suite des expériences. En effet, en ce qui concerne les signaux non reliés au métabolisme du PB, il semble exister une nette supériorité de ce système.

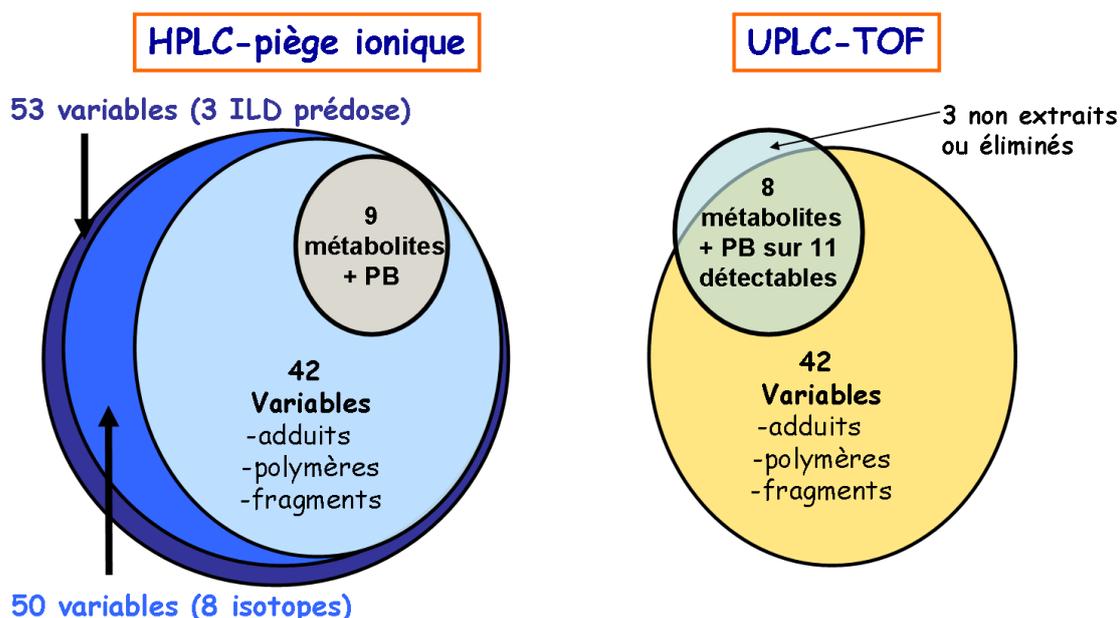


Figure 35 : Attribution des variables discriminantes en phase neutre-méthanol de l'effet traitement aux métabolites du PB.

ILD : inférieur à la limite de détection. Le système UPLC-TOF avait détecté l'ensemble des métabolites mais 3 ont été supprimés lors du traitement des données et recherchés a posteriori.

Par ailleurs, il est décidé, dans l'avenir, d'avoir recours à un système HPLC-Orbitrap® dont le détecteur haute résolution permettra un « matching » des variables entre les systèmes sur la base de la mesure de masse précise tout en donnant accès à l'acquisition de spectres de fragmentation grâce à la trappe linéaire présente en amont du système Orbitrap®.

## 2.6 Conclusion

Cette première partie du travail a permis de se familiariser avec l'ensemble des outils nécessaires à la réalisation des approches métabolomiques globales. Elle a aussi fait resurgir les limites technologiques actuelles qui constituent un frein au développement de telles approches comme la complexité d'obtenir une empreintes exhaustive de l'ensemble des métabolites présents dans un fluide biologique, la nécessité de développer des outils d'extraction de données qui soient à la fois capable de conserver l'ensemble de l'information analytique et de revenir à une présentation

analytique (possibilité de générer un chromatogramme) des données retraitées afin d'évaluer réellement l'impact du traitement, la difficulté de comparer des données issues d'appareils différents, et surtout la nécessité de développer des outils de regroupement des variables d'intérêt afin de limiter le nombre de signaux devant être identifiés tant cette étape est longue et parfois peu couronnée de succès, s'agissant de métabolites endogènes. Elle a également permis de sélectionner le système analytique qui permettrait de poursuivre ce travail et mis en lumière l'intérêt de l'utilisation de différentes phases mobiles (pH et nature du solvant organique) afin d'améliorer la couverture de détection de la multitude d'entités chimiques présentes dans un fluide biologique en chromatographie liquide couplée à la spectrométrie de masse, aspect qui sera discuté dans le chapitre suivant.

Une des difficultés, apparue lors de cette première partie, est la présence d'une multitude de signaux générés pour un seul métabolite. Cette multiplicité est liée à la méthode de détection et représente une forme de redondance qui complexifie l'exploitation des données. Il a donc été décidé de tenter de regrouper l'ensemble des signaux issus d'un même métabolite à l'aide d'outils statistiques à l'instar de ce qui a récemment été pratiqué en RMN pour le regroupement des déplacements chimiques issus de la même molécule. C'est ce travail et son application que nous aborderons dans la partie suivante.

## 3 DÉVELOPPEMENTS STATISTIQUES

### 3.1 Introduction.

Au cours des 20 dernières années, les avancées technologiques dans le domaine de la spectrométrie de masse ont conduit à une amélioration importante de la sensibilité et de la résolution des spectromètres de masse. L'introduction des spectromètres de masse haute résolution comme les système à temps de vol (TOF) ou à transformée de Fourier (FTICR, Orbitrap®), l'arrivée récente de l'UPLC qui permet une amélioration de la résolution chromatographique, de la capacité de pic, la sensibilité et de la répétabilité des analyses ont été des apports incontestables pour la métabolomique<sup>145, 299</sup>. Cependant, ceci a conduit à la détection dans les profils métaboliques d'un nombre de signaux en constante augmentation, qui se traduit par la nécessité de trier et d'identifier un nombre croissant de variables discriminantes suite à l'analyse différentielle des empreintes métaboliques à l'aide des statistiques multivariées. Les TOF sont couramment couplé à l'UHPLC pour l'acquisition des empreintes métaboliques grâce à un temps de cycle d'acquisition court (cycle propre de 0.1 à 0.3 s), et une mesure de masse précise à 5 ppm. Les Q-TOF ne permettent pas de mener des études de fragmentation avec de meilleures résolution et précision en masse, de plus la sélection du précurseur au moyen d'un quadripôle est assez peu résolutive, ce qui réduit la capacité d'identification obtenue avec ces systèmes<sup>300</sup>. A l'inverse, les systèmes FTICR ou Orbitrap® autorisent de meilleures résolution et précision de mesure en masse (<3 ppm) et l'acquisition de spectres de fragmentation avec détermination des filiations d'ions, mais leur temps de cycle plus long à ultra-haute résolution, de l'ordre de la seconde, rend leur couplage avec l'UHPLC difficile. Pour l'acquisition des empreintes et l'identification des variables d'intérêt, le recours aux deux types de systèmes semble donc inévitable.

En outre, en spectrométrie de masse couplée à la chromatographie liquide, un métabolite n'est pas à l'origine d'un seul signal (rapport  $m/z$ ) mais d'un ensemble de signaux correspondant à la formation au cours de la phase de désolvatation du processus d'ionisation de différentes espèces comme des adduits avec les cations métalliques présents ( $\text{Na}^+$ ,  $\text{K}^+$ , ...) ou des molécules de solvant ionisées, et à des ions produits issus de la fragmentation spontanée du métabolite dans la source (« in-source CID »). Bien qu'informative, nous y reviendrons, cette redondance du signal ralenti le processus d'identification des métabolites en complexifiant les données. Or l'étape d'identification des signaux discriminant a toujours été une étape limitante à l'expansion de la métabolomique.

Dans ce contexte, cette partie du travail a consisté à développer une stratégie de suppression de la redondance des données, impliquant un outil mathématique de traitement des données basé sur les matrices d'autocorrélation, et d'identification des signaux précédemment isolés grâce à l'utilisation d'un système couplant HPLC et spectrométrie de masse ultra-haute résolution.

A partir d'une approche métabolomique globale sur des urines de rats collectées avant et après traitement par le phénobarbital et analysées par UPLC-TOF, le regroupement des signaux redondants a été basé sur l'exploitation mathématique de la corrélation existant entre les intensités des ions issus

du même métabolite à travers l'ensemble du set d'échantillons (matrices de corrélation) afin de tenter d'isoler un signal unique pour chaque molécule détectée. L'isolement d'un nombre réduit de signaux à identifier et leur caractérisation en MS<sup>2</sup> et MS<sup>3</sup> sur un système HPLC-Orbitrap<sup>®</sup> a permis l'identification de 11 métabolites du PB dont certains n'avaient jamais été décrits.

## 3.2 Matériel et méthode

### 3.2.1 Réactifs

Le méthanol (qualité Lichrosolv) et l'acide formique proviennent de la société Merck (Briare-le-Canal, France). L'acétonitrile, l'acétate d'ammonium et les standards utilisés pour l'optimisation de l'Orbitrap<sup>®</sup> (glucuronide de testostérone, aldostérone, sulfate de DHEA, et acide adipique) ont été achetés auprès de la société Sigma (Saint-Quentin Fallavier, France). L'eau ultra-pure a été obtenue par déminéralisation couplée à un système de purification Milli-Q de chez Millipore. Les standards de calibration pour le système LTQ-Orbitrap<sup>®</sup> (Calmix-positive et Calmix-negative, pour respectivement les modes d'ionisation positif et négatif) ont été fournis par la société ThermoFisher Scientific (Les Ulis, France). Enfin, la solution de standards (Metabonomics MS system test mix) utilisée pour la validation analytique de l'acquisition des empreintes urinaires en UPLC-TOF avait pour origine la société Waters (Saint-Quentin en Yvelines, France).

### 3.2.2 Echantillons d'urine de rats

Urines de rats collectées avant et après administration de phénobarbital en intra-péritonéal selon le protocole expérimental précédemment décrit (Cf. Introduction expérimentale 1.2.2 et 0).

### 3.2.3 UPLC-TOF / acquisition des empreintes

Les urines ont été centrifugées à 10000 tpm pendant 5 min afin d'éliminer les particules et les protéines puis diluées au tiers avec de l'eau ultra-pure avant injection. Les urines des animaux avant et après traitement ont été analysées après randomisation au sein d'une même séquence au moyen d'un système UPLC-TOF. Les échantillons urinaires (10 µL par injection) ont été chromatographiés à l'aide d'une colonne ACQUITY BEH 1.7 µm C<sub>18</sub> 100 x 2.1 mm (Waters) sur un système ACQUITY UPLC en mode gradient. Deux conditions d'élution ont été utilisées, la première en milieu acétonitrile-acide avec une phase A et une phase B contenant respectivement de l'eau avec 0.1 % d'acide formique et de l'acétonitrile avec 0.1 % d'acide formique, et la seconde en milieu méthanol-neutre avec une phase A et une phase B contenant respectivement de l'eau avec 10 mM d'acétate d'ammonium et du méthanol avec 10 mM d'acétate d'ammonium. Dans les deux cas, le débit de phase mobile a été maintenu constant à 0.4 mL/min pendant les 35 minutes de durée du gradient dont le déroulement a été reporté dans le Tableau 2 du chapitre I.

L'acquisition des empreintes s'est faite sur un spectromètre Waters LCT premier en mode positif et négatif avec le mode W (double réflexion des ions) activé. Les ions protoné (*m/z* 556.2771) et déprotoné (*m/z* 554.2614) de la leucine encéphaline apportée à une concentration de 1.5 µg/mL dans

une solution eau : méthanol (50 :50) ont été utilisés comme référence de calibration en cours d'acquisition (système Lockspray<sup>®</sup>), en mode positif et négatif respectivement. La gamme d'acquisition en masse était comprise entre  $m/z$  100 et  $m/z$  1000 avec le mode DRE (Dynamic range enhancement) activé. Ce mode permet de dévier une partie du faisceau d'ion avant le détecteur du spectromètre afin d'éviter sa saturation. Les chromatogrammes et spectres de masses obtenus ont été traités à l'aide des logiciels MassLynx<sup>®</sup> 4.1 et MarkerLynx<sup>®</sup> (Waters). Une solution de références (Test mix) a été injectée tous les dix échantillons pour s'assurer de la stabilité du système analytique.

### 3.2.4 Traitement des données et analyses statistiques multivariées

Les données brutes ont été traitées à l'aide du logiciel MarkerLynx<sup>®</sup> (Waters). Elles ont été débruitées à l'aide d'un algorithme de suppression du bruit de fond opérant dans les dimensions chromatographique et spectrale, avant d'extraire les signaux correspondant à des pics. La dimensionnalité des données a été réduite à deux dimensions en agrégeant les informations  $m/z$  et temps de rétention (Tr) pour créer des variables définies par un couple  $m/z$ -Tr. L'ensemble des informations relatives aux intensités de ces variables a ensuite été combiné au sein d'une matrice regroupant toutes les variables extraites ainsi que leur intensité dans chaque échantillon. En cas de non détection d'une variable dans un échantillon ou pour le cas où son intensité serait inférieure à la limite fixée par l'utilisateur, le logiciel lui assigne une valeur dans la matrice, de sorte qu'il n'y ait pas de données manquantes. Enfin, le logiciel effectue une étape de normalisation des données pour chaque échantillon d'après la somme totale des intensités des variables détectées dans cet échantillon. Dans la mesure où il n'a pu être mis en évidence une différence significative entre les volumes urinaires avant et après traitement (test de Student,  $\alpha=0.05$ ), il n' a pas été procédé à une normalisation sur la concentration de créatine des échantillons.

La matrice de données résultante a ensuite été analysée avec le logiciel SIMCA P (version 11, Umetrics, Suède) au moyen d'outils de statistiques multivariées afin de réaliser l'analyse différentielle des empreintes avant et après traitement. Avant d'être analysés par ACP, PLS-DA, et OPLS, les intensités de la matrice ont d'abord été soit centrées-réduites soit normalisée selon la méthode Pareto précédemment décrite. Les modèles ont été validés par validation croisée et les modèles explicatifs testés avec un test de permutation ( $k=20$ ).

### 3.2.5 Matrices d'autocorrélation

Cette approche qui utilise la corrélation de l'intensité des variable à travers l'ensemble des échantillons est basée sur le calcul d'une matrice de corrélation C obtenue selon l'équation suivante<sup>288</sup> :

$$C = \frac{1}{n-1} X_1 \times' X_2$$

où pour des données de spectrométrie de masse  $X_1$  et  $X_2$  représentent des matrices de dimension  $n \times v_1$  et  $n \times v_2$ , avec  $n$  correspondant au nombre d'échantillons et  $V_1$  et  $V_2$  étant le nombre de variables (couple  $m/z$ -Tr) définissant ces échantillons dans chaque condition expérimentale.  ${}^tX_2$  est la transposée de  $X_2$ , et  $C$  est donc une matrice de dimension  $X_1 \times X_2$  dont les valeurs sont les coefficients de corrélation entre les deux variables de  $X_1$  et  $X_2$ . Pour le calcul d'une matrice d'autocorrélation  $X_1$  est égale à  $X_2$  et dans ce cas la matrice renvoie la corrélation de l'intensité des variables entre elles au travers de l'ensemble des échantillons. Nous avons émis l'hypothèse que les ions issus d'un même métabolite devaient présenter un temps de rétention identique et une variation d'intensité corrélée lors d'une modification de la concentration du métabolite source, en réponse à une perturbation, dans la mesure où les conditions analytiques étaient garanties constantes et que le coefficient de corrélation théorique de leur intensité à travers l'ensemble des échantillons devait théoriquement être égal à 1. En pratique, au sein d'échantillons complexes comme des fluides biologiques,  $r$  est toujours inférieur à 1 en raison du bruit spectral ou des coélutions qui sont à l'origine d'un effet matrice ou d'une suppression d'ionisation propre à chaque échantillon, les valeurs des coefficients de corrélation entre les intensités des différents ions (adduits, fragments intra source) issus d'un même métabolite restent néanmoins très élevées. Les matrices d'autocorrélations ont été calculées à partir de données centrées réduites à l'aide du logiciel Matlab® (version 6.5).

### 3.2.6 HPLC-LTQ-Orbitrap®

La séparation chromatographique des échantillons a été effectuée sur un système HP1100 (Agilent, Massy, France), équipé d'une colonne XTerra MS C<sub>18</sub> 5- $\mu$ m, 2.1 x 150 mm (Waters, Saint Quentin en Yvelines, France), munie d'un filtre en ligne (Interchim, Montluçon). Les phases mobiles employées étaient les mêmes que pour l'UPLC-TOF (neutre-méthanol et acide-acétonitrile). Le schéma de gradient était lui différent avec une phase initiale isocratique de 5 min en 100% phase A puis un gradient linéaire de 0 à 100% de phase B sur 40 minutes avec un débit maintenu à 0.2 mL/min. Après un retour à 100% de A en 1 min, la colonne était laissée en rééquilibrage pendant 10 minutes.

Les empreintes métaboliques ont été acquises, successivement en mode positif et négatif, sur une gamme de masse allant de  $m/z$  100 à  $m/z$  1000 avec un pouvoir de résolution fixé à 60000 ( $M/\Delta M$ , FWHM à  $m/z$  400). Les paramètres de source ont été optimisés à l'aide d'une solution contenant du glucuronide de testostérone, de l'aldostérone, du sulfate de DHEA, et de l'acide adipique, tous à 1  $\mu$ g/mL et infusés à 5  $\mu$ L/min. La calibration en masse a été réalisée en positif et en négatif au moyen des solutions commerciales fournies par le fabricant.

Les spectres de fragmentation CID ont été obtenus en utilisant la fonction « dependant scan » : le cycle d'acquisition comprenait un spectre de masse à un pouvoir de résolution de 15000 et deux spectres de fragmentation ( $MS^2$  et  $MS^3$ ) obtenus en dépendance à un pouvoir de résolution de 7500. L'ion le plus intense, détecté au cours du spectre initial, déclenchait l'acquisition de son spectre de fragmentation ( $MS^2$ ) dont l'ion le plus intense (en dehors du  $m/z$  du parent qui était exclu) entraînait à

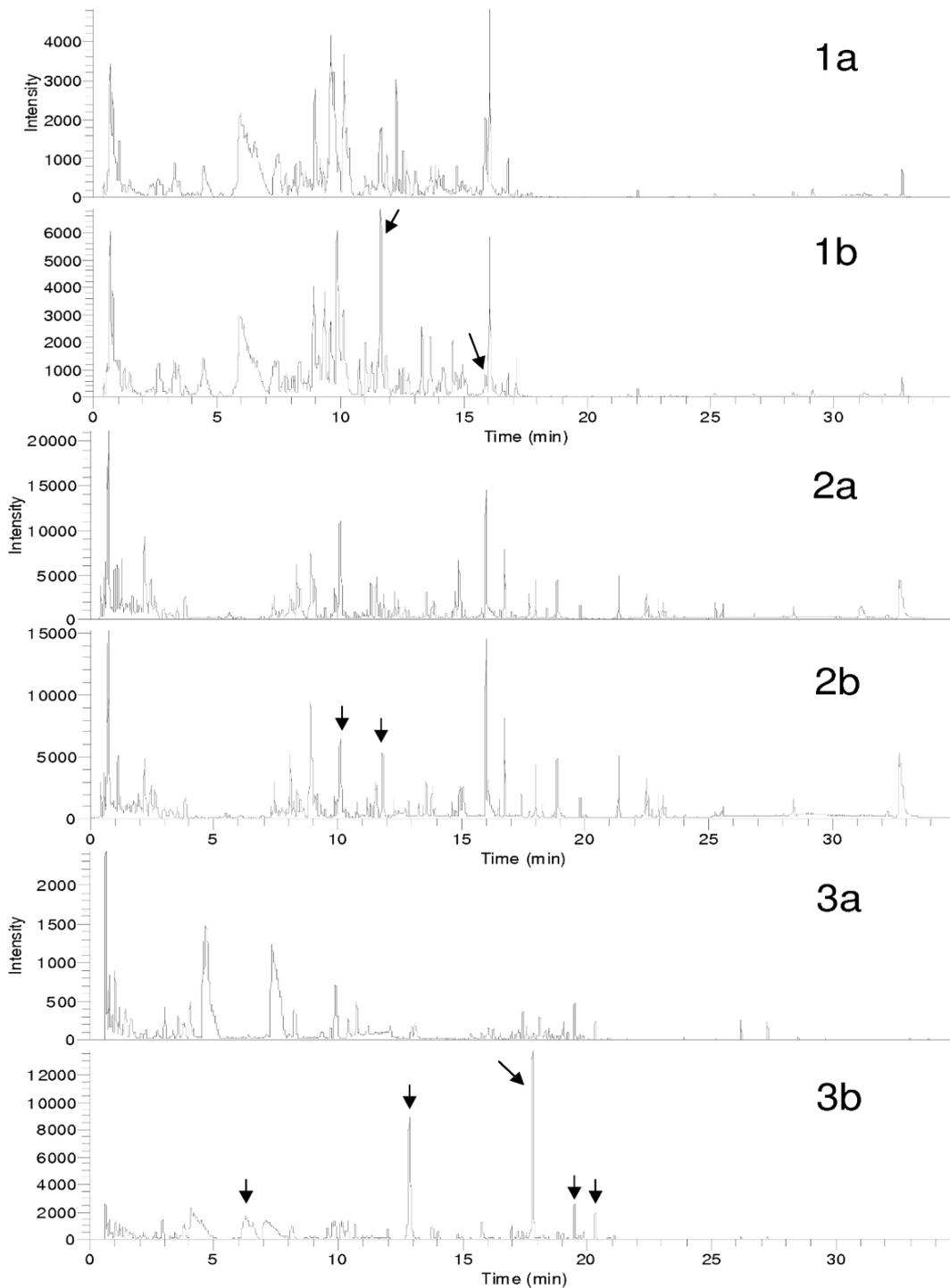
son tour l'acquisition de son propre spectre de fragmentation ( $MS^3$ ). La fenêtre d'isolation du parent était de  $\pm 0.5$  par rapport au  $m/z$  du parent, l'énergie normalisée de collision fixée à 35% et le temps d'activation réglé à 30 ms. Enfin, la détection à trois reprises successives d'un ion entraînait son placement en liste d'exclusion temporaire afin d'analyser des ions de moindre intensité. Toutes les données acquise avec ce système ont été analysées à l'aide du module Qualbrowser<sup>®</sup> du logiciel Xcalibur<sup>®</sup> (Thermo-Fisher, Les Ulis, France) qui a également permis la génération des formules brutes à partir des mesures de masse précise.

### 3.3 Résultats et Discussion

**Acquisition des empreintes.** La précision de mesure en masse du TOF a été garantie par l'utilisation du système lockspray tandis que la stabilité globale du système a été vérifiée par l'injection tous les 10 échantillons d'une solution de composés de référence, compatible avec toutes les conditions d'élution et de polarité d'ionisation. Ces injections ont permis de valider les performances chromatographiques (répétabilité des temps de rétention des références), et de détection en spectrométrie de masse (précision et répétabilité des mesures de masses précises et répétabilité des intensités des références). Les critères de validation analytique ont été les suivants : mesure de la masse précise des références avec une erreur inférieure à 5 ppm et coefficient de variation des Tr et des intensités, respectivement inférieurs à 5 et 25 %.

Les profils métaboliques obtenus dans les différentes conditions d'élution sont présentés dans la Figure 36. Si peu de variations sont visibles à l'œil nu suite au traitement par le PB en conditions acide-acétonitrile que ce soit en mode positif ou en mode négatif, les conditions neutre-méthanol font quant à elles apparaître de nombreuses modifications dans l'allure du chromatogramme. Ce constat suggère que l'utilisation de différentes compositions de phases mobiles est responsable d'importantes variations de réponse en ionisation électrospray. D'un point de vue chromatographique, l'utilisation de différents solvants organiques permet d'exploiter des mécanismes de rétention complémentaire en phase inverse. En effet, le méthanol affecte principalement les mécanismes d'interaction soluté / phase impliquant des liaisons hydrogènes tandis que l'acétonitrile affecte de façon prépondérante les interactions de type dipôle-dipôle. L'emploi de ces deux solvants organiques peut donc permettre de séparer des composés qui seraient coélus dans l'une ou l'autre condition. Enfin, la modification du pH de la phase mobile peut être responsable de la modification de l'état d'ionisation des métabolites ionisables pendant la migration, conduisant à des changements drastiques de rétention. Ce point est particulièrement intéressant pour les composés élués dans le volume mort dans l'une des conditions à cause de leur état ionisé puisqu'ils pourraient être retenus dans l'autre condition. Ainsi 27 % des variables extraites dans les conditions neutre-méthanol contre seulement 7% des variables extraites en condition acide-acétonitrile se trouvaient dans le volume mort. Ces changements de rétention ont également un impact en termes de détection par spectrométrie de masse puisque les composés coélus sont plus susceptibles de subir le phénomène de suppression d'ionisation. Bien que le design expérimental ne permette pas de différencier l'impact de la modification du pH de celui du changement de solvant organique, cette expérience montre néanmoins que l'emploi de conditions

chromatographiques différentes pour l'analyse des mêmes échantillons conditionne en partie l'étape d'extraction des variables et se révèle un moyen de mise en œuvre facile d'étendre la couverture des composés détectés dans un fluide biologique avec un système analytique unique.



**Figure 36 : Profils métaboliques urinaires de rat typiques obtenus en UPLC-TOF.**

Profils obtenus à partir de l'injection d'urines de rat avant (a) et après (b) un traitement de 4 jours au PB. L'acquisition a été menée en conditions éluantes acide-acétonitrile en mode négatif (1) et positif (2) et en conditions neutre-méthanol en mode négatif (3). Les flèches identifient les pics modifiés par le traitement et visibles à l'œil nu.

**Extraction des données et analyse différentielle des empreintes.** L'extraction des données analytiques a été menée avec le logiciel MarkerLynx<sup>®</sup> et a conduit à l'extraction de 2458, 8524 et 3627 variables pour respectivement les conditions neutre-méthanol en mode négatif et les conditions acide-acétonitrile en mode négatif et positif. Le nombre élevé de variables extraites dans le mode négatif en conditions acide-acétonitrile peut s'expliquer par un bruit de fond plus faible qu'en mode positif, ce qui a permis d'abaisser la valeur seuil d'extraction de variable. Comme nous l'avons mentionné au chapitre précédent, l'UPLC se révèle d'une aide précieuse pour l'extraction des variables puisqu'elle produit des pics fins (de l'ordre de 10 s à la base) et par conséquent un bon rapport signal/bruit pour toutes les conditions testées, et qu'elle offre une excellente répétabilité des temps de rétention avec par exemple un écart-type de 1.63 s pour l'acide hippurique (40 échantillons) et 0.55 s pour le PB (20 échantillons).

La matrice de données générée par MarkerLynx<sup>®</sup> a ensuite été soumise à l'analyse statistique multivariée après normalisation des variables (centrage puis réduction : variance unitaire ou Pareto). Les résultats obtenus en PCA et en OPLS ayant déjà été discutés dans le chapitre précédent ne seront pas repris en détail ici. L'analyse PCA a montré une séparation nette des individus selon un effet traitement suggérant que l'administration du PB était la source principale de variance des empreintes. Des méthodes supervisées comme la PLS-DA et l'OPLS, avec l'effet traitement en variable réponse (Y) ont ensuite été employées pour faciliter l'extraction des variables discriminantes entre les pré-doses et les traités. Pour toutes les conditions expérimentales les modèles OPLS construits présentaient des R2Y et des Q2 supérieurs à 0.9 et 0.85 (Tableau 13).

**Tableau 13 : Paramètres des modèles OPLS pour les différentes conditions analytiques**

Conditions	Normalisation	R2X <sup>a</sup>	R2Y <sup>b</sup>	Q2 <sup>c</sup>
<b>Acide positif</b>	Aucune	0.929	0.902	0.855
	Pareto <sup>d</sup>	0.410	0.933	0.917
	UV <sup>e</sup>	0.192	0.963	0.882
<b>Acide négatif</b>	Aucune	0.907	0.917	0.889
	Pareto <sup>d</sup>	0.303	0.965	0.928
	UV <sup>e</sup>	0.160	0.971	0.865
<b>Neutre négatif</b>	Aucune	0.783	0.861	0.834
	Pareto <sup>d</sup>	0.430	0.922	0.895
	UV <sup>e</sup>	0.125	0.955	0.663

<sup>a</sup> R2X correspond au pourcentage de variance des variables originelle expliquée par le modèle

<sup>b</sup> R2Y correspond au pourcentage de variance de la variable réponse (Y) expliquée par le modèle

<sup>c</sup> Q2 représente la proportion de variance des variables qui peut être prédite par le modèle

<sup>d</sup> Pareto : centrage des données et réduction de Pareto

<sup>e</sup> UV : centrage des données et réduction à la variance unitaire

Les variables discriminantes de ces modèles ont ensuite été extraites selon le critère précédemment défini  $VIP > 1.1$  et ceci afin de supprimer les variables présentant une contribution significative au modèle mais possédant 20 valeurs nulles dans un groupe et 16 dans l'autre donc de faible robustesse.

Le nombre de variables identifiées de la sorte est reporté pour chaque condition expérimentale dans le Tableau 14.

**Tableau 14 : Nombres de variables discriminantes identifiées par l'analyse différentielle des empreintes.**

Conditions UPLC/ToF <sup>a</sup>	Variables extraites (MarkerLynx <sup>®</sup> )	Variables discriminantes		
		Total	(Vip > 1.1)	
			Présentes PD et PT	Absentes PD
I	2458	670	628	42
II	8254	407	336	71
III	3627	1433	1362	71

I : conditions neutre-méthanol mode négatif, II : conditions acide-acétonitrile mode négatif, III : conditions acide-acétonitrile mode positif. PD : pré-dose, PT : post traitement.

Les variables liées au métabolisme du PB doivent être absentes des pré-dose et présentes chez les traités. Ce peut aussi être le cas de variables correspondant à des composés endogènes dont la faible concentration ne permet la détection chez les pré-doses mais qui se voient augmentés par le traitement au PB et deviennent détectables chez les traités. Dans ce cas, l'identification des composés est nécessaire afin de faire la part des métabolites et des composés endogènes.

A ce stade, il faut entreprendre l'identification des signaux discriminants et c'est l'objectif de la suite de ce travail avec le recours au système HPLC-Orbitrap<sup>®</sup>. Cependant, le nombre de signaux à identifier même en ce qui concerne les signaux potentiellement reliés au métabolisme du PB reste relativement important avec un total pour toutes les conditions de 184 (42 + 71 + 71) signaux extraits par l'analyse différentielle des empreintes métaboliques. De plus, parmi les 113 (42 + 71) variables mises en évidence en mode négatif, seules 19 sont communes aux deux conditions chromatographiques. De même, parmi les 142 signaux extraits en conditions acide-acétonitrile, seuls 10 sont communs (même temps de rétention, même masse monoisotopique calculée). Nous avons donc mis en place un outil statistique reposant sur les matrices d'autocorrélation afin de regrouper les signaux issu du même métabolite et ainsi limiter le nombre de variables à identifier.

**Matrices d'autocorrélation.** Elles ont initialement été développées pour le traitement des données de RMN <sup>288</sup>, lesquelles présentent souvent une forte redondance des données. L'objectif est alors de regrouper les déplacements chimiques issus de la même molécule en exploitant le fait que les rapports des changements d'intensité des signaux issus d'un même métabolite à la suite d'une perturbation doivent être identiques. Par analogie, nous avons ici appliqué les matrices d'autocorrélation afin de structurer le jeu de données en regroupant les différents signaux générés en spectrométrie de masse mais provenant du même métabolite. En effet, en ionisation électrospray, au cours du processus de désolvatation, un composé peut être à l'origine de plusieurs signaux correspondant à la formation d'adduits avec les cations du milieu comme les ions sodium, potassium et ammonium ou avec une molécule de solvant ionisée, ainsi qu'à la formation d'ions fragments formés par la dissociation spontanée de la molécule parent dans la source.

Bien que les adduits soient prévisibles et relativement facilement identifiables avec des mesures de masse précise, il n'en est pas de même pour les fragments générés spontanément dans la source, surtout en cas de coélution. La Figure 37A est la représentation graphique de la matrice d'autocorrélation obtenue pour les 71 variables discriminantes, attribuées au métabolisme du PB sur la base de leur absence chez les pré-doses, et issues de l'analyse différentielle des empreintes urinaires analysées en mode positif avec des conditions d'élution acide-acétonitrile. Chacune des variables, numérotées de 1 à 71 en fonction de leur temps de rétention, est retrouvée en abscisses et en ordonnées. Cet agencement des variables sur les axes conduit à la création d'un axe de symétrie correspondant à la diagonale de la matrice. Chaque point de cette diagonale correspond au coefficient de corrélation d'une variable avec elle-même, d'où leur valeur commune égale à 1. Cette représentation permet de visualiser la redondance globale des données en faisant apparaître des amas de points colorés de part et d'autre de la diagonale correspondant à des coefficients de corrélation élevés (variables 10 à 20, 40 à 50 et 50 à 60, par exemple). L'agencement en amas est une fois de plus lié à la classification par ordre d'élution des variables, les signaux issus de la même molécule possédant logiquement un temps de rétention identique. La Figure 37B présente 6 signaux coélus à 13.27 min en UPLC-TOF, auxquels correspond l'extrait de la matrice d'autocorrélation présenté dans la Figure 37C. Les coefficients de corrélation entre les intensités de ses 6 ions à travers les 20 échantillons traités ont été utilisés afin de déterminer s'ils étaient liés entre eux donc étaient issus de la même molécule parent. Le tracé de l'intensité de M1, M2, et M3 en fonction de celle de M2 pour chaque urine de traité (Figure 37D) confirme que ces ions proviennent de la même molécule parent comme l'avait laissé supposer la valeur des coefficients de corrélation de la matrice d'autocorrélation. L'analyse de ces ions conduit à l'attribution des espèces suivantes : M2 ( $m/z$  409) est la molécule protonée ( $[M+H]^+$ ), M3 ( $m/z$  391) correspond à la perte d'une molécule d'eau de M2, M4 ( $m/z$  431) est l'adduit sodium de la molécule parent ( $[M+Na]^+$ ) et enfin M1 ( $m/z$  839) est le dimère sodé de la molécule parent ( $[2M+Na]^+$ ). Pour M5 et M6, il est difficile de les relier à la molécule parent, et la Figure 37 E montre que leur intensité n'ont aucun lien avec celle de M2. En revanche, la Figure 37F prouve que ces deux ions sont liés entre eux et pourraient donc provenir d'un autre métabolite coélué. Au final, appliquées au Tr 13.27 min, les matrices d'autocorrélation on montré que les 6 signaux coélus ne correspondaient qu'à deux molécules différentes avec 4 signaux (M1, M2, M3, et

M4) issus d'une première molécule parent et 2 autres (M5 et M6) d'une seconde. De plus, comme mentionné dans la Figure 37B, M5 et M6 n'ont pu être détectés en HPLC-Orbitrap®. Cette approche permet donc de savoir que si deux signaux ont été perdus, seule une molécule n'a pas été détectée.

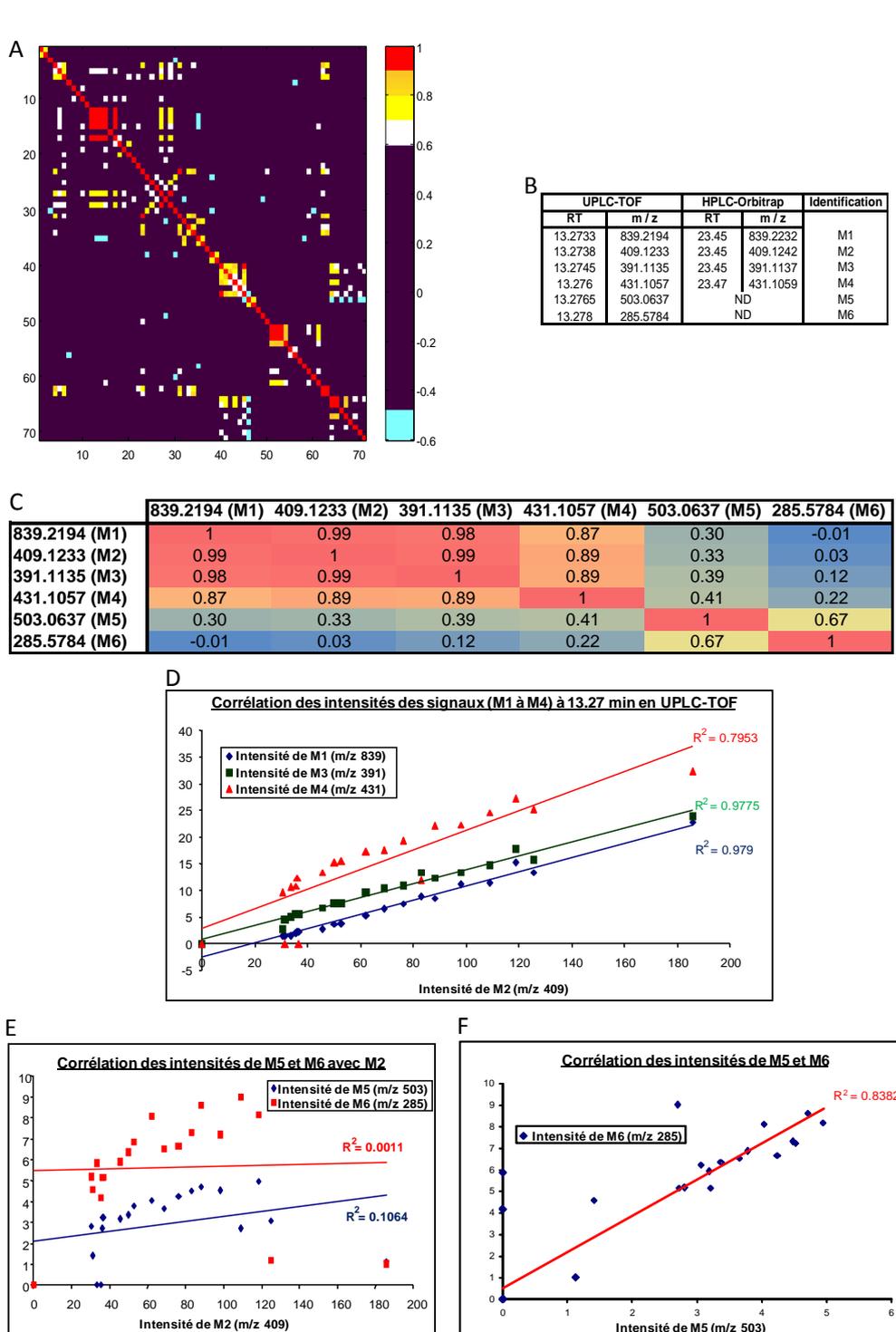


Figure 37 : Présentation générale des matrices d'autocorrélation.

(A) : Représentation graphique en échelle de couleur de la matrice d'autocorrélation obtenue pour les variables discriminantes des conditions acides-acétonitrile en mode positif. (B) Variables extraites en UPLC-TOF et correspondance avec les variables détectées en HPLC-Orbitrap®. (C) Extrait de la matrice d'autocorrélation pour les ions de TR = 13.27 min. (D) : Représentation de l'intensité des ions M1, M3 et M4 en fonction de celle de M2 pour les 20 échantillons traités. (E) : Représentation de l'intensité des ions M5 et M6 en fonction de celle de M2 pour les 20 échantillons traités. (F) : Représentation de l'intensité de l'ions M6 en fonction de celle de M5 pour les 20 échantillons traités.

Les matrices d'autocorrélation ont donc montré qu'elles étaient capables de regrouper les signaux issus d'une même molécule qu'il s'agisse d'adduits, de dimères ou d'ions fragments. Appliquées aux données expérimentales complètes soit aux 184 variables attribuées au métabolisme du PB toutes conditions confondues, les matrices d'autocorrélation ont permis de réduire ce nombre à 88 signaux ou groupes de signaux, soit l'élimination d'environ 50 % des signaux initialement à identifier, parmi lesquels 76 ont été retrouvés sur le système HPLC-Orbitrap<sup>®</sup>, ce qui correspond à une perte de signaux de 15 à 30 % lors du passage du système UPLC-TOF à l'HPLC-Orbitrap<sup>®</sup>. Les résultats détaillés pour chaque condition sont présentés dans le Tableau 15.

**Tableau 15 : Résumé de l'impact des matrices d'autocorrélation sur le nombre de variables à identifier.**

Conditions UPLC/TOF <sup>a</sup>	Variables discriminantes <sup>b</sup>	Variables ou groupes de variables <sup>c</sup>	Variables ou groupes de variables à identifier <sup>d</sup>	Métabolites identifiés
I	42	19	16	11, dont PB
II	71	34	28	11, dont PB
III	71	35	23	9

<sup>a</sup> I : conditions neutre-méthanol mode négatif, II : conditions acide-acétonitrile mode négatif, III : conditions acide-acétonitrile mode positif. PD : pré-dose, PT : post traitement.; <sup>b</sup> Vip > 1.1, Absentes en pré-dose; <sup>c</sup> Regroupement selon les matrices d'autocorrélation; <sup>d</sup> après suppression des variables non retrouvées sur le système HPLC-LTQ-Orbitrap<sup>®</sup>.

A partir de ces 76 signaux détectés avec l'HPLC-Orbitrap<sup>®</sup>, il a pu être identifié de 9 à 11 métabolites du PB par condition expérimentale. Ainsi, ces résultats montrent que bien qu'une partie de la redondance du signal soit mise en évidence à l'aide des matrices d'autocorrélation, certains adduits et ions fragments présentent une mauvaise corrélation avec l'ion protoné ou déprotoné de la molécule parent et persistent dans les variables à identifier. Au final, 33 signaux attribuables à des ions issus du PB ou de ses métabolites ont dû être investigués au cours d'expériences de fragmentation à l'aide du système HPLC-Orbitrap<sup>®</sup>.

**Identification des métabolites du PB sur le système HPLC-Orbitrap<sup>®</sup>.** Afin de lisser les variations interindividuelles et d'économiser du temps d'analyse, les 20 échantillons de pré-dose d'un côté et les 20 échantillons traités de l'autre ont été regroupés pour l'étape de caractérisation des signaux d'intérêt en HPLC-Orbitrap<sup>®</sup>.

Pour l'identification, les signaux d'intérêt mis en évidence jusqu'ici par le traitement des données UPLC-TOF ont été matchés avec les signaux obtenus en HPLC-Orbitrap<sup>®</sup> à partir des profils moyens à un pouvoir de résolution de 60 000. Les systèmes chromatographiques étant différents, la correspondance a été établie sur la base de la masse précise mesurée avec les deux systèmes (tolérance de 20 ppm), sur l'absence des signaux en pré-dose et sur des ions coélus lorsqu'il subsistait un doute. Malgré les différences de technologie (HPLC vs UPLC), de pente de gradient et de colonne, l'ordre d'élution des composés a presque toujours été conservé.

Une fois la correspondance établie, les compositions élémentaires correspondant aux signaux d'intérêt ont été générées à partir des données Orbitrap®, à l'aide du module dédié du logiciel Xcalibur®, avec une tolérance sur la mesure de masse de 5 ppm. Pour l'ensemble des conditions et des métabolites, cette erreur est toujours restée inférieure à 1.5 ppm. Les expériences menées avec l'Orbitrap® n'ont pas seulement confirmé les résultats obtenus avec l'UPLC-TOF, mais ont permis de lever certaines ambiguïtés sur des formules générées avec une précision de mesure moindre. Cependant, la précision de mesure en masse n'est pas le seul apport du détecteur Orbitrap®, le gain de résolution participe également pour beaucoup à l'attribution d'une formule élémentaire puisqu'il donne accès à une analyse fine du massif isotopique<sup>301</sup>. La Figure 38 illustre l'intérêt de l'analyse du massif isotopique avec les spectres de masse de deux pics contenant des ions de masse nominale  $m/z$  263 présentant une différence de  $m/z$  qui n'est que de 0.018 et des temps de rétention similaires.

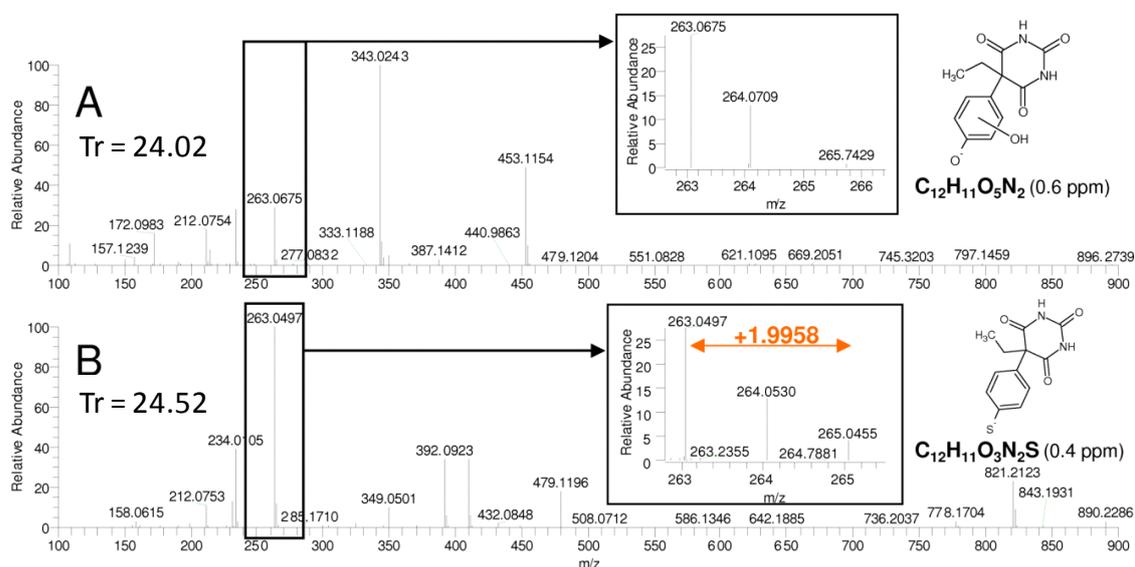


Figure 38 : Spectre de masse en ESI-Orbitrap® de deux pics à 24.02 min (A) et 24.52 min (B) contenant un ion de masse nominale  $m/z$  263.

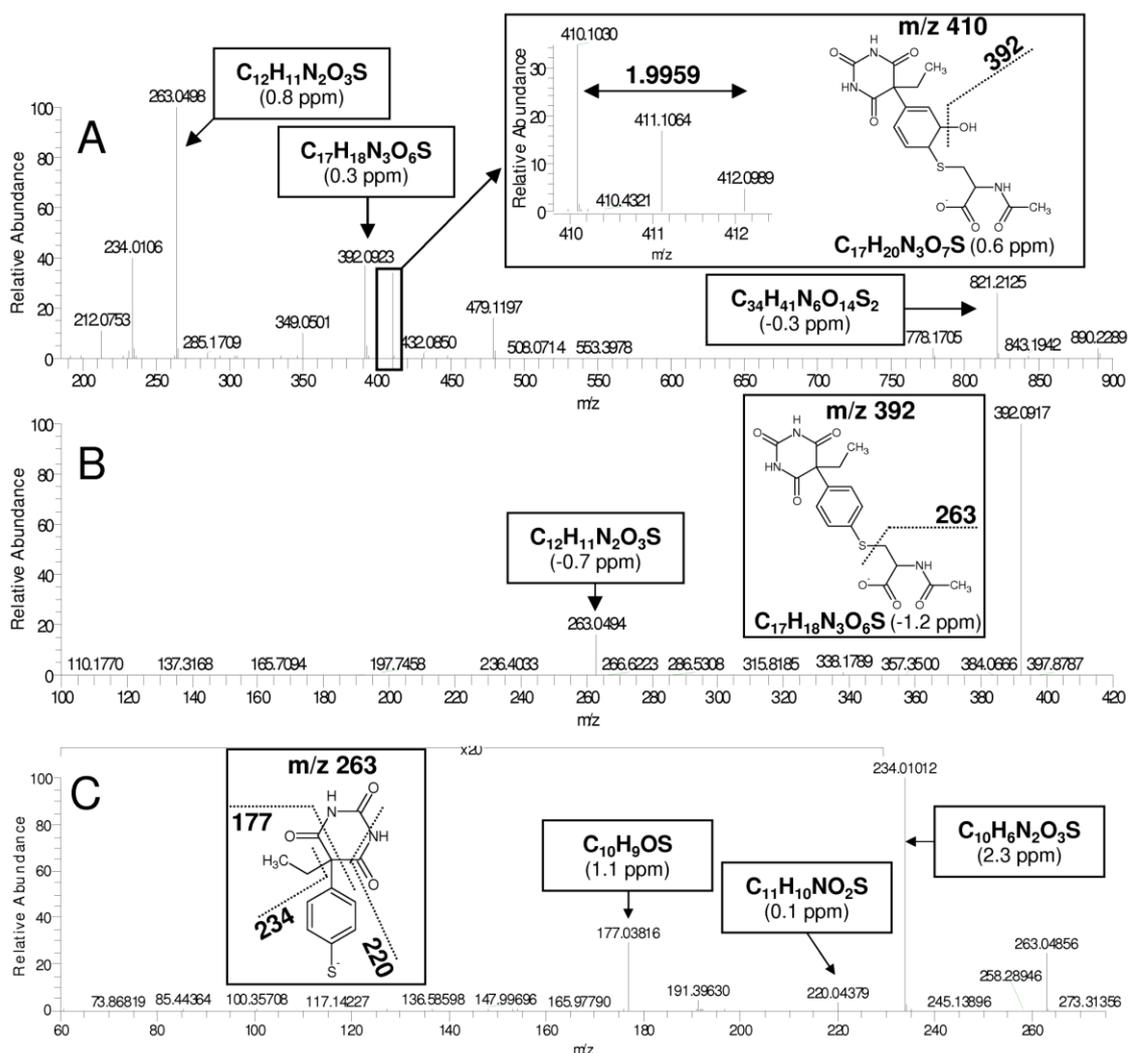
Spectre de masse enregistré en mode négatif avec un pouvoir de résolution de 60000. Les encarts présentent le massif isotopique détaillé de chacun des ions.

L'un contient un atome de soufre comme le montre son massif isotopique avec un pic d'abondance relative 4.5 % décalé de 1.9958 (théoriquement ce décalage devrait être de 1.9959) par rapport au pic monoisotopique, l'autre non. Il est ainsi possible de restreindre ou d'imposer les éléments pouvant être utilisés par le logiciel pour générer les formules brutes et ainsi limiter le nombre de solutions possibles. Il s'agit ici en fait de deux ions fragments provenant pour le premier de la fragmentation intra-source du conjugué sulfate du di-OH-PB, et pour le second de la fragmentation spontanée du dérivé mercapturique du PB.

Les expériences de fragmentation ( $MS^n$ ) peuvent aussi être utilisées pour confirmer ou infirmer une formule élémentaire. En effet, les spectres de fragmentation fournissent une seconde mesure de

masse précise du parent mais en plus permettent de déterminer la composition élémentaire des fragments, autorisant ainsi l'élimination des formules brutes incompatibles pour le parent.

En ce qui concerne la caractérisation de métabolites issus d'un xénobiotique, les expériences de fragmentation peuvent aussi mettre à jour des ions produits ou des pertes de molécules neutres communs entre le xénobiotique et son métabolite facilitant la filiation métabolique. Elles peuvent en outre faire apparaître des ions produits ou des pertes de neutres spécifiques de biotransformation. La Figure 39 illustre les expériences ayant permis la caractérisation d'un dérivé mercapturique du phénobarbital.



**Figure 39 : Spectre de masse Orbitrap® du dérivé mercapturique du phénobarbital.**

(A) Spectre obtenu en mode négatif avec un pouvoir de résolution de 60 000. Le massif isotopique du signal d'intérêt  $m/z$  410 est présenté en encart. (B) Spectre de fragmentation CID de  $m/z$  410 avec un pouvoir de résolution de 30 000. (C) Spectre de fragmentation séquentielle (MS<sup>3</sup>) de  $m/z$  410 →  $m/z$  392 déclenché automatiquement sur l'ion le plus intense de la MS<sup>2</sup>, parent exclu.

L'analyse du massif isotopique du dérivé mercapturique du PB sur le spectre de masse révèle que le composé contient un atome de soufre. Le spectre de MS<sup>2</sup> de  $m/z$  410 permet d'éliminer la partie de la molécule ajoutée par biotransformation (groupement N-Acétyle-cystéine). Enfin le spectre

de  $MS^3$  signe la présence du PB au sein de la molécule avec la présence de deux ions produits  $m/z$  234 et  $m/z$  177 qui correspondent à des pertes de neutres typiques du PB substitué en para (29u  $C_2H_5$  and 43u  $NCHO$ ). A noter que les matrices d'autocorrélation avaient mis en évidence un lien entre les intensités des ions  $m/z$  821,  $m/z$  410, et  $m/z$  392 correspondant respectivement aux espèces  $(2M+H)^+$ ,  $(M+H)^+$ , et  $(M+H-H_2O)^+$ . L'ion  $m/z$  234 quant à lui n'a pas été classé comme significatif au cours des analyses multivariées, il ne faisait donc pas partie des variables utilisées pour le calcul de la matrice d'autocorrélation.

Un second exemple est présenté dans la Figure 40 avec la caractérisation du signal d'intérêt  $m/z$  626. Dans celui-ci, l'analyse du massif isotopique exclu la présence d'atome avec une distribution isotopique à la masse monoisotopique +2 comme le potassium, le chlore, le soufre... Le spectre de fragmentation CID ( $MS^2$ ) permet d'éliminer une nouvelle fois la partie de la molécule ajoutée par biotransformation, ici, un probable N-Acétyl-hexapyranosyl, retrouvé sous forme d'ion fragment à  $m/z$  221. La fragmentation séquentielle  $m/z$  626  $\rightarrow$   $m/z$  405 permet quant à elle de faire la liaison avec le PB en produisant un ion  $m/z$  247 correspondant à l'OH-PB. Un mécanisme de fragmentation proposé pour ce composé est présenté dans la Figure 39D. Ce composé pourrait correspondre à un conjugué O-glucurono-N-Acétyl-hexapyranosylamine du PB, mais la validation de cette structure requiert une confirmation avec des méthodes d'analyse complémentaire comme la RMN. Cet exemple illustre donc bien la limitation des expériences de fragmentation CID et plus largement de la l'utilisation de la spectrométrie de masse seule pour la caractérisation d'une molécule.

Au final, cette approche a permis de détecter le PB, excrété sous forme inchangée, de même que 14 de ses métabolites qui sont présentés dans le Tableau 16 avec leurs conditions de détection. Il a ainsi été possible de proposer un schéma métabolique du PB (Figure 41). Le métabolisme du PB a été étudié *in vitro*, dans des hépatocytes de rats <sup>302, 303</sup>, et *in vivo* dans les biofluides de rongeurs <sup>302, 304, 305</sup>, conduisant à la mise en évidence du p-OH-PB et du m-OH-PB, ainsi que le PB-N-glucoside (chez la souris <sup>306</sup>), les PB-N-glucuronides (chez la souris <sup>307</sup>), et les dérivés O-glucuronidés de l'OH-PB. Un dérivé 2-phényl- $\gamma$ -butyrolactone du PB a aussi été décrit dans la littérature à partir d'incubations sur hépatocytes <sup>303</sup>. Ce composé a été détecté chez l'homme à la suite d'une intoxication au PB <sup>308</sup>, mais n'a pas été mis en évidence lors de cette approche chez le rat. A l'inverse, tous les autres métabolites identifiés par cette approche métabolomique n'avait jamais été identifiés chez le rat mais à partir d'échantillons biologiques humains <sup>309-313</sup>, à l'exception du conjugué O-glucurono-N-acétyl-hexapyranosylamine du PB qui est décrit ici pour la première fois.

Même si la caractérisation d'une molécule n'est que très rarement possible grâce à l'utilisation de la spectrométrie de masse seule, l'association de la mesure de masse précise, de la haute résolution et de la possibilité de mener des expériences séquentielles de fragmentation fournit des informations structurales de premier ordre qui peuvent suffire à prédire un schéma métabolique en association avec une approche métabolomique. Cette étude a également montré l'intérêt d'acquisitions multiples avec des conditions chromatographiques différentes, puisque certains métabolites n'ont été détectés qu'avec les conditions neutre-méthanol.

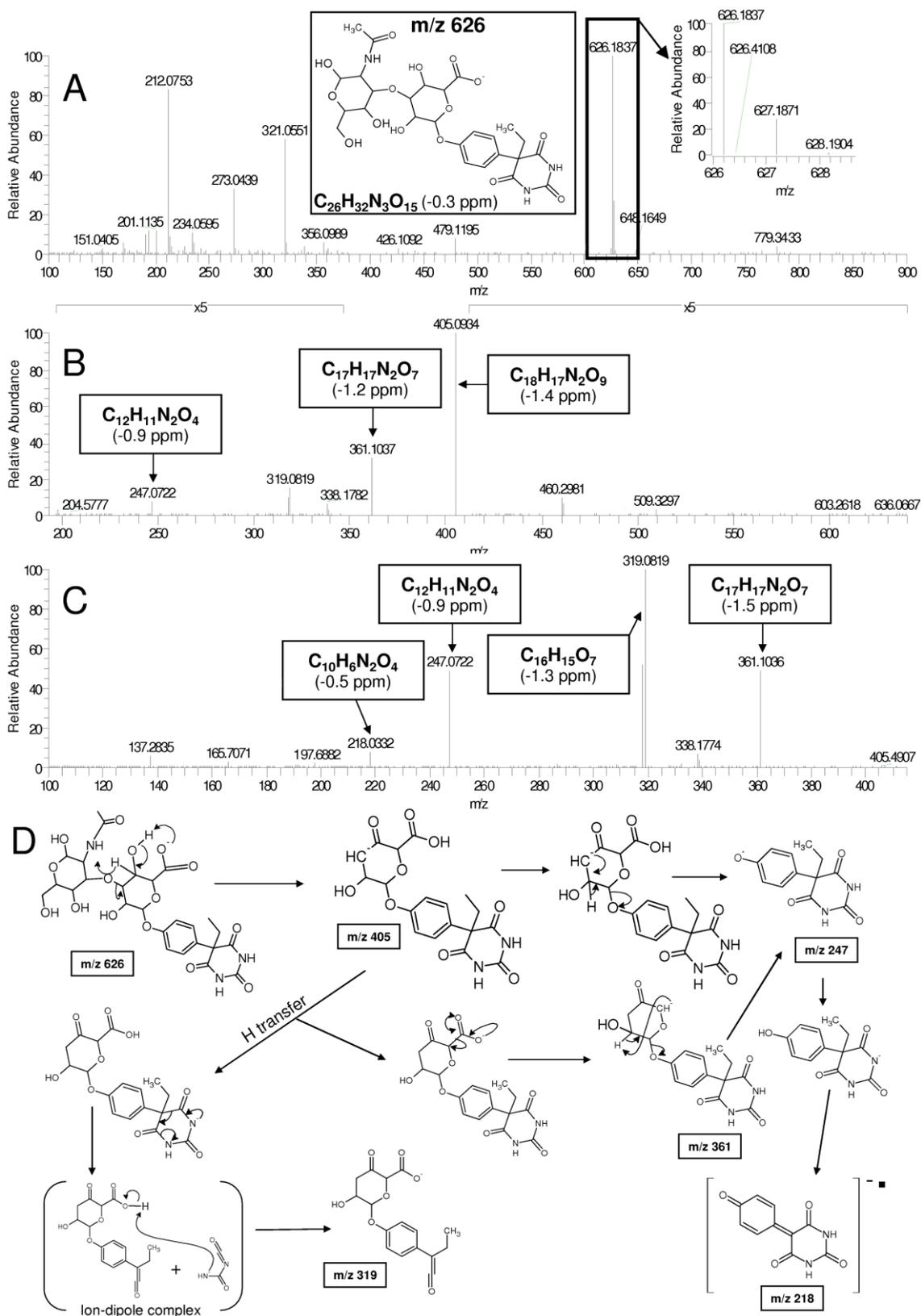


Figure 40 : Données de spectrométrie de masse ayant servi à la caractérisation du signal d'intérêt  $m/z$  626.

(A) : Spectre de masse Orbitrap<sup>®</sup> au temps de rétention de  $m/z$  626 obtenu avec un pouvoir de résolution de 60000 en mode négatif, l'encart offre une vue détaillée du massif isotopique. (B) Spectre de fragmentation CID de  $m/z$  626 avec un pouvoir de résolution de 30 000. (C) Spectre MS<sup>3</sup> séquentiel de  $m/z$  626 →  $m/z$  405 déclenché automatiquement sur l'ion le plus intense de la MS<sup>2</sup>, parent exclu.

Tableau 16 : Panorama des métabolites détectés en fonction des conditions d'analyse.

Composition	Structures envisagées	Conditions de détection <sup>a</sup>
C <sub>10</sub> H <sub>8</sub> O <sub>3</sub> N <sub>2</sub>	PB dééthylé (perte de la chaîne latérale)	I
C <sub>12</sub> H <sub>12</sub> O <sub>3</sub> N <sub>2</sub>	PB	I, II
C <sub>12</sub> H <sub>12</sub> O <sub>4</sub> N <sub>2</sub>	p-Hydroxy-PB	I, II, III
C <sub>12</sub> H <sub>12</sub> O <sub>5</sub> N <sub>2</sub>	Dihydroxy-PB	I
C <sub>12</sub> H <sub>14</sub> O <sub>5</sub> N <sub>2</sub>	Dihydrodiol-PB	I*, II
C <sub>12</sub> H <sub>12</sub> O <sub>7</sub> N <sub>2</sub> S	p-Hydroxy-PB-sulfate	I, II
C <sub>12</sub> H <sub>12</sub> O <sub>8</sub> N <sub>2</sub> S	Dihydroxy-PB-sulfate	I*, II
C <sub>17</sub> H <sub>22</sub> O <sub>8</sub> N <sub>2</sub>	PB-N-glucuronide avec ouverture du noyau barbital (perte de CO)	I, II, III
C <sub>17</sub> H <sub>20</sub> O <sub>9</sub> N <sub>2</sub>	Hydroxy-PB-O-glucuronide avec ouverture du noyau barbital	II*, III
C <sub>17</sub> H <sub>21</sub> O <sub>7</sub> N <sub>3</sub> S	Dérivé mercapturique du PB	I, II, III
C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	PB-N-glucuronide (énantiomère 1)	I, II, III
C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	PB-N-glucuronide (énantiomère 2)	I, II, III
C <sub>18</sub> H <sub>20</sub> O <sub>10</sub> N <sub>2</sub>	p-Hydroxy-PB-O-glucuronide	I, II, III
C <sub>19</sub> H <sub>22</sub> O <sub>11</sub> N <sub>2</sub>	O-Methyl-PB-O-glucuronide	I, II, III
C <sub>26</sub> H <sub>33</sub> O <sub>15</sub> N <sub>3</sub>	PB-O-glucuronide-N-acétyl-hexopyranosylamine	I*, III

<sup>a</sup> I : conditions neutre-méthanol mode négatif, II : conditions acide-acétonitrile mode négatif, III : conditions acide-acétonitrile mode positif.

En fait certains d'entre eux étaient bien détectables dans les conditions acide-acétonitrile mais n'ont pas été extraits par MarkerLynx<sup>®</sup> malgré une optimisation minutieuse des paramètres de ce logiciel. Ceci peut s'expliquer par leur faible ionisation dans ces conditions.

D'autres approches couramment employées<sup>314</sup> dans les études de métabolisme comme les études en perte de neutres ou de recherche de précurseurs, basées sur des fragmentations et l'utilisation de filtres de défaut de masse ne sont pas aussi exhaustives et présentent d'autres défauts. En effet, le profil de fragmentation du PB (Figure 42) diffère de celui de ses métabolites. Ceci peut s'expliquer par le fait que de nombreux métabolites sont issus du p-OH-PB qui est le métabolite de phase I majoritaire. Or la substitution en para du PB est responsable d'importantes modifications dans le profil de fragmentation puisqu'elle conduit à une fragmentation plus importante du noyau barbital et à la perte de la chaîne latérale éthyle, qui sont les pertes communes de la plupart des métabolites. Il est intéressant de noter que le dérivé mercapturique produit un premier fragment où le soufre est en position para, ce qui conduit au même type de fragmentation, que précédemment décrit, à l'étape suivante de fragmentation. En conséquence, une approche par perte de neutre basée sur la fragmentation du PB aurait certainement mis en évidence bien peu de métabolites. En revanche, une analyse des pertes de neutres basée sur les biotransformations de phase 2 aurait été plus efficace, même si le conjugué O-glucurono-N-acétyl-hexopyranosylamine du PB ne produit aucun fragment lié à ces biotransformations.

## CHAPITRE III : Travaux personnels

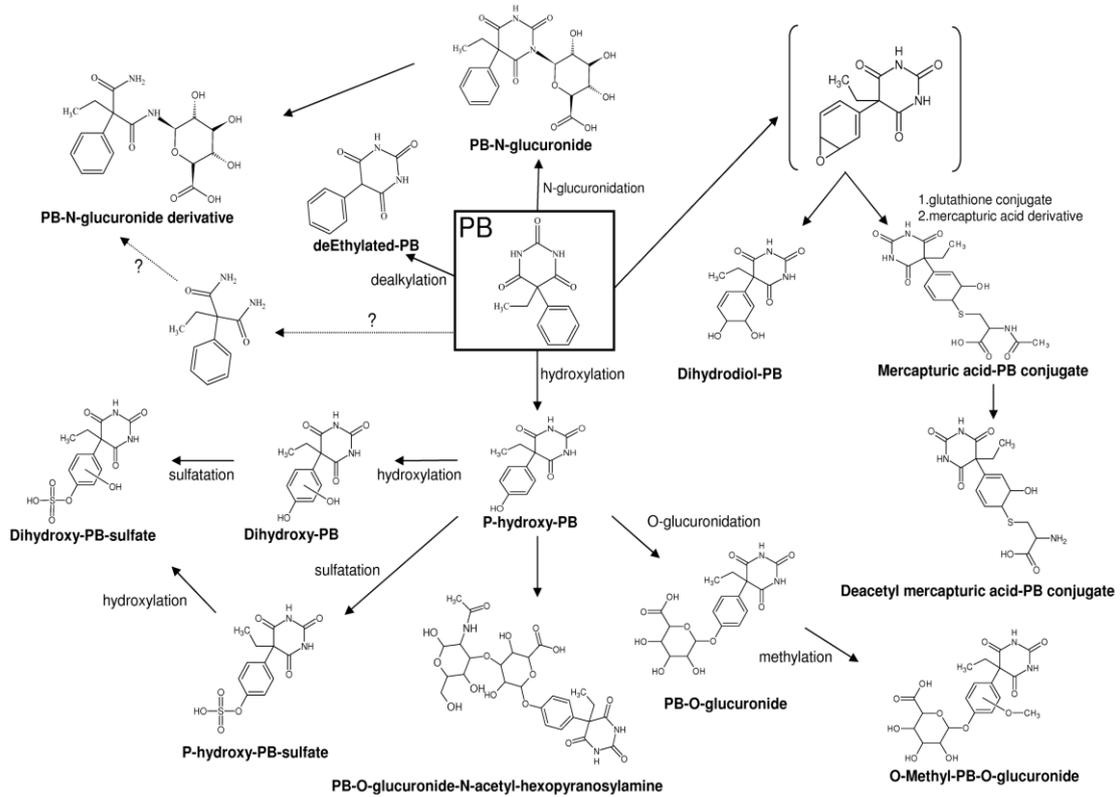


Figure 41 : Proposition de schéma retraçant les voies métaboliques du Phénobarbital.

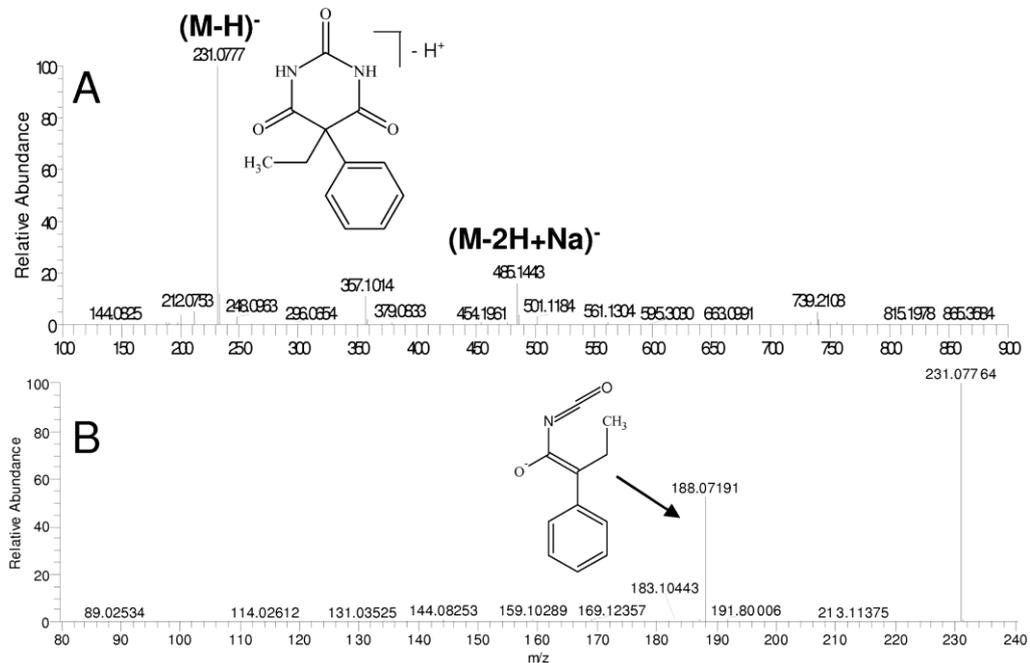


Figure 42 : Spectre de masse du PB en mode négatif.

(A) : Spectre de masse du PB obtenu en mode négatif sur un système Orbitrap<sup>®</sup> en mode négatif avec un pouvoir de résolution de 60000. (B) Spectre de fragmentation CID du PB obtenu avec un pouvoir de résolution de 30000 et montrant la fragmentation différente de ce xénobiotique en comparaison de ses métabolites avec une absence de pertes de neutres de 29 u et 43 u correspondant respectivement à la perte de la chaîne latérale éthyle et à l'ouverture du noyau barbital qui sont observées pour ses métabolites para-substitués.

Pour ce qui concerne les approches utilisant des filtres de défaut de masse dont le filtre est appliqué avec une tolérance couramment comprise entre 20 et 50 mDa<sup>315</sup>, elles sont plutôt dédiées à la mise en évidence de biotransformations mineures comme celles de phase I comprenant des oxydations, des réductions et le gain ou la perte de petites structures aliphatiques. Leur application au métabolisme du PB avec un filtre de 50 mDa n'aurait pas permis de détecter le conjugué O-glucuronon-N-acétyl-hexapyranosylamine, le dérivé PB-N-glucuronide ayant perdu CO par ouverture du noyau barbital et le di-OH PB-sulfate. Néanmoins, l'utilisation de logiciels permettant la prédiction de la masse des métabolites après les biotransformations classiques aurait très certainement rendu la détection de ses composés possible mais aurait également conduit à la sélection de nombreux composés endogènes.

Enfin, l'approche métabolomique du métabolisme possède elle aussi des limitations. D'abord parce qu'elle reste complexe et requiert donc un temps de mise en œuvre certain afin d'effectuer toutes les opérations d'analyse statistique, d'acquisition et d'interprétation des spectres CID, et ensuite parce qu'elle ne donne accès qu'à une information semi-quantitative. Ce dernier point est acceptable tant qu'il s'agit de l'identification des métabolites mais nécessite des expériences complémentaires quantitatives afin d'évaluer l'importance relative des différentes voies métaboliques. Pour finir, les variables de faible intensité ne permettent pas toujours d'obtenir un massif isotopique bien défini et de réaliser des expériences de fragmentation. Il est alors nécessaire de rajouter une étape de traitement de l'échantillon visant à le concentrer afin d'effectuer l'acquisition des spectre CID ou de pouvoir utiliser le massif isotopique pour l'attribution de la composition élémentaire des composés de plus de 400 Da.

### 3.4 Conclusion

Cette partie du travail a permis d'illustrer une nouvelle stratégie pour l'analyse des données de métabolomique obtenues par spectrométrie de masse et particulièrement pour son application à l'identification de métabolites dans les études de métabolisme. Premièrement, elle a mis en évidence l'apport de plusieurs conditions chromatographiques qui sont bénéfiques pour la séparation chromatographique mais aussi pour la détection en spectrométrie de masse et donc intéressantes pour augmenter la diversité chimique couverte par une technique analytique unique. Deuxièmement, les matrices d'autocorrélation ont montré qu'elles permettaient de supprimer plus de 50 % de la redondance des données MS afin de cibler l'identification sur les variables discriminantes les plus pertinentes. Troisièmement, le spectromètre LTQ-Orbitrap<sup>®</sup> s'est révélé un outil particulièrement adapté et performant pour l'identification des variables d'intérêts en offrant à la fois une mesure de masse précise, une excellente résolution permettant l'analyse des massifs isotopiques et la capacité de mener des expériences de fragmentation séquentielles à haute résolution elles aussi. Cependant, si la solution idéale semble être le couplage UPLC-Orbitrap<sup>®</sup>, le temps de cycle, supérieur à la seconde de ce dernier, limiterait très certainement l'acquisition des spectres CID en mode dépendance, nécessaire à l'obtention de la filiation des ions et si utile à l'identification.

## 4 OUTILS DE VISUALISATION : Une Aide à l'identification ?

### 4.1 Introduction.

Les analyses multivariées sont l'outil de visualisation le plus couramment employé pour l'analyse différentielle des empreintes métaboliques, générées dans le cadre d'approches métabolomiques globales. En revanche, dans le cadre d'approches semi-ciblées sur une ou plusieurs familles chimiques d'intérêt, menées a posteriori sur des données d'approche globale, ou pour la construction ou l'enrichissement de bases de données chimiques, il serait parfois plus intéressant de pouvoir organiser les signaux en fonction de leur structure chimique plutôt que par leur aspect discriminant vis-à-vis de sous-populations d'échantillons correspondant à des individus ayant subi une perturbation physiologique.

Avec l'expansion de la spectrométrie de masse ultra-haute résolution, dans laquelle nous classerons les spectromètres de masse à transformée de Fourier tels que les FTICR et l'Orbitrap<sup>®</sup>, il est aujourd'hui courant d'avoir accès à des données de spectrométrie de masse comportant une mesure de masse précise avec une erreur inférieure au ppm avec un FTICR ou de l'ordre de 1 à 3 ppm avec un détecteur Orbitrap<sup>®</sup>. En outre, ces spectromètres possèdent également un pouvoir de résolution très important (jusqu'à 100000 pour l'Orbitrap<sup>®</sup> <sup>222</sup> et au-delà du million pour un FTICR en fonction de l'intensité du champ magnétique de l'analyseur <sup>212</sup>) qui réduit considérablement les erreurs de mesures liées à la non séparation de deux rapports  $m/z$  proches au cours de l'acquisition du spectre. Ces performances permettent le plus souvent de déterminer les compositions élémentaires des molécules de bas poids moléculaires (< 500 Da) sur la base de la mesure de la masse précise, ce qui démontre qu'elles contiennent une information qualifiable de structurale. Il est dès lors évident de chercher à utiliser cette information structurale cachée afin d'organiser ses données en fonction d'éléments structuraux de discrimination des composés détectés.

En 1963, Edward Kendrick a proposé une échelle de masse reposant sur la masse du groupement méthylène (14 u par définition) en alternative à l'échelle de masse IUPAC qui est basée sur la masse de l'atome de carbone <sup>12</sup>C (12 u par définition). L'échelle de Kendrick convertit donc la masse du groupement méthylène (CH<sub>2</sub>) de 14.01565 u (selon l'échelle IUPAC) à 14 u (par définition), de sorte que les séries homologues, (c.-à-d., des composés chimiques présentant la même composition en hétéroatomes et le même degré d'insaturation et ne différant que par le nombre d'unité CH<sub>2</sub>) possèdent le même défaut de masse de Kendrick <sup>316</sup>. Le Tableau 17 illustre cette propriété avec la série des acides carboxyliques qui représente un intérêt biologique en tant que composante de la famille des acides gras. L'avantage majeur de cette technique de visualisation réside dans sa capacité à extrapoler à des composés de masse très élevée, ne permettant pas l'assignation d'une composition élémentaire unique à partir de la seule mesure de la masse, aussi précise soit elle, une classe chimique et une formule, déterminée à partir d'un composé de masse inférieure et ne différant que par son nombre d'unité CH<sub>2</sub> <sup>317</sup>.

L'analyse des indices de Kendrick a constitué une révolution pour l'interprétation des spectres de masse dans certains domaines comme la pétroliométrie. Elle a par ailleurs été récemment appliquée à dans une étude de lipidométrie cherchant à évaluer les modifications des concentrations de lipides au sein des tumeurs à la suite d'un traitement par adénovirus délivrant la protéine p53<sup>318</sup> et pour la caractérisation du contenu lipidique du lait de vache<sup>319</sup>.

Tableau 17 : Calcul des masses nominales et indices de Kendrick pour les acides carboxyliques aliphatiques

Formule chimique semi-développée	Masse IUPAC	Masse de Kendrick	Masse nominale de Kendrick	Défaut de masse de Kendrick
H <sub>3</sub> C-COOH	60.02113	59.9541099	60	0.0459
H <sub>3</sub> C-CH <sub>2</sub> -COOH	74.03678	73.9541099	74	0.0459
H <sub>3</sub> C-(CH <sub>2</sub> ) <sub>2</sub> -COOH	88.05243	87.9541099	88	0.0459
H <sub>3</sub> C-(CH <sub>2</sub> ) <sub>3</sub> -COOH	102.06808	101.9541099	102	0.0459

Cette partie du travail a consisté à évaluer la faisabilité de l'utilisation des indices de Kendrick au traitement des données de métabolomique obtenues par spectrométrie de masse ultra-haute résolution à l'aide d'un système Orbitrap<sup>®</sup>. Cette étude de faisabilité s'est poursuivie par l'application de la représentation de Kendrick à des urines de rats afin d'évaluer la pertinence d'une telle approche dans l'organisation de composés chimiques d'origine biologique.

## 4.2 Matériel et méthode

### 4.2.1 Réactifs

L'acétonitrile, l'acétate d'ammonium et les standards, utilisés pour l'optimisation de l'Orbitrap<sup>®</sup> (glucuronide de testostérone, aldostérone, sulfate de DHEA, et acide adipique) d'une part et pour les identifications d'autre part, proviennent de la société Sigma-Aldrich (Saint-Quentin Fallavier, France). L'eau ultra-pure a été obtenue par déminéralisation couplée à un système de purification Milli-Q de chez Millipore. Les standards de calibration pour le système LTQ-Orbitrap<sup>®</sup> (Calmix-positive et Calmix-negative, pour respectivement les modes d'ionisation positif et négatif) ont été fournis par la société ThermoFisher Scientific (Les Ulis, France).

### 4.2.2 HPLC-LTQ-Orbitrap<sup>®</sup>

La séparation chromatographique des échantillons a été effectuée sur un système Surveyor (ThermoFisher Scientific, Les Ulis, France), équipé d'une colonne XTerra MS C<sub>18</sub> 5- $\mu$ m, 2.1 x 150 mm (Waters, Saint Quentin en Yvelines, France), munie d'un filtre en ligne (Interchim, Montluçon). Les phases mobiles employées étaient les mêmes que pour l'UPLC-TOF (neutre-méthanol et acide-acétonitrile). Le schéma de gradient était lui différent avec un pallier initial isocratique de 5 min en

100% phase A puis un gradient linéaire de 0 à 100% de phase B sur 40 minutes avec un débit maintenu à 0.2 mL/min. Après un retour à 100% de A en 1 min, la colonne était laissée en rééquilibrage pendant 10 minutes.

Les empreintes métaboliques ont été acquises, successivement en mode positif et négatif, sur une gamme de masse allant de  $m/z$  100 à  $m/z$  1000 avec un pouvoir de résolution fixé à 60000 ( $M/\Delta M$ , FWHM à  $m/z$  400). Les paramètres de source ont été optimisés à l'aide d'une solution contenant du glucuronide de testostérone, de l'aldostérone, du sulfate de DHEA, et de l'acide adipique, tous à 1  $\mu\text{g/mL}$  et infusés à 5  $\mu\text{L/min}$ . La calibration en masse a été réalisée en positif et en négatif au moyen des solutions commerciales fournies par le fabricant. En mode négatif, les paramètres de source étaient les suivants : gaz de nébulisation, 50 (unité arbitraire UA) ; gaz auxiliaire, 10 UA ; température de capillaire, 275°C ; tension de source, 3500 V ; et tension de tube de transfert, -100V. En mode positif, les paramètres étaient identiques à l'exception du gaz auxiliaire (5 UA) et de la tension de source (5000 V).

Les spectres de fragmentation CID ont été obtenus en utilisant la fonction « dependant scan » : le cycle d'acquisition comprenait un spectre de masse à un pouvoir de résolution de 15000 et deux spectres de fragmentation ( $\text{MS}^2$  et  $\text{MS}^3$ ) obtenus en dépendance à un pouvoir de résolution de 7500. L'ion le plus intense, détecté au cours du spectre initial, déclenchait l'acquisition de son spectre de fragmentation ( $\text{MS}^2$ ) dont l'ion le plus intense (en dehors du  $m/z$  du parent qui était exclu) entraînait à son tour l'acquisition de son propre spectre de fragmentation ( $\text{MS}^3$ ). La fenêtre d'isolation du parent était de  $\pm 0.5$  par rapport au  $m/z$  du parent, l'énergie normalisée de collision fixée à 35% et le temps d'activation réglé à 30 ms. Enfin, la détection à trois reprises successives d'un ion entraînait son placement en liste d'exclusion temporaire afin d'analyser des ions de moindre intensité. Des expériences alternatives ont également été entreprises pour l'analyse des différentes classes chimiques. Pour ces dernières, le cycle d'acquisition comprenait un spectre de masse à un pouvoir de résolution de 15000 et un spectre de fragmentation ( $\text{MS}^2$ ) avec un pouvoir de résolution de 7500, obtenu par dépendant scan avec la fonction wide band activation activée. Le spectre de fragmentation était déclenché suite à la détection d'un ion appartenant à une liste de masse regroupant les rapports  $m/z$  de tous les membres de la classe chimique d'intérêt. Toutes les données acquises avec ce système ont été analysées à l'aide du module Qualbrowser<sup>®</sup> du logiciel Xcalibur<sup>®</sup> (Thermo-Fisher, Les Ulis, France) qui a également permis la génération des formules brutes à partir des mesures de masse précise.

### 4.2.3 Préparation des échantillons

Les urines utilisées provenaient du groupe contrôle (NaCl) de l'étude d'induction précédemment décrite. Avant injection, les urines ont uniquement été centrifugées à 10 000 tpm. Pour l'identification des familles des solutions de référence à 1  $\mu\text{g/mL}$ , préparées en solution méthanol : eau (50 :50, v/v) ont également été injectées sans autre traitement.

#### 4.2.4 Traitement des données

L'ensemble des spectres et chromatogrammes ont été traités à l'aide du module Qualbrowser<sup>®</sup> du logiciel Xcalibur<sup>®</sup>. Les compositions élémentaires ont été déterminées par étude visuelle du massif isotopique et au moyen du générateur de formule de Qualbrowser<sup>®</sup>.

Pour le calcul des indices de Kendrick, un spectre moyen de l'ensemble du chromatogramme a été généré avant d'être exporté au format Excel<sup>®</sup> (Suite Office 2003, Microsoft). Les masses paires et impaires ont ensuite été scindées en deux tables, puis pour chacune d'entre elles, ont été calculés les masses nominales et les défauts de masse de Kendrick. Enfin, les représentations graphiques ont été obtenues au moyen du logiciel Prism<sup>®</sup> (version 5.01, GraphPad software).

Les compositions élémentaires ont été utilisées pour interroger des banques de données (HMDB, KEGG et Metlin) de façon automatisée avec une macro, développée au laboratoire sous Excel<sup>®</sup>, afin d'évaluer le nombre de familles chimiques pouvant correspondre à un indice de Kendrick.

### 4.3 Résultats et Discussion

Dans un premier temps nous avons tenté d'isoler l'ensemble des acides dicarboxyliques à partir de la représentation de Kendrick d'un profil Orbitrap<sup>®</sup> d'urine de rats puisque nous avons mis en évidence un composé endogène modifié après le traitement au PB et l'avons identifié comme étant l'acide sébacique. L'acide sébacique ou décanedioïque présentait le double avantage d'appartenir à la famille des lipides pour lesquels les indices de Kendrick avaient déjà été mis en œuvre et d'avoir été modifié suite au traitement par le PB, rendant l'ensemble des acides dicarboxyliques intéressants, pour le cas où d'autres composés de cette série auraient également été impactés par le traitement.

L'acide sébacique possède une masse monoisotopique de 202.1205 qui correspond à une masse de Kendrick de 201.8950 qui se décompose en une masse nominale de Kendrick de 202 et un défaut de masse de Kendrick de 0.1050. A partir d'un profil urinaire de rat obtenu en HPLC-Orbitrap<sup>®</sup> en mode négatif, un spectre moyen est calculé, et l'ensemble des rapports m/z, au-delà d'un seuil d'intensité, en est extrait et exporté sous format Excel<sup>®</sup>. Les masses paires et impaires sont ensuite séparées, de sorte que les signaux correspondant au composé ne comportant que des carbone  $^{12}\text{C}_n$  et des composés  $^{13}\text{C}_1^{12}\text{C}_{n-1}$  d'une part et les composés à nombre impair ou pair d'azote d'autre part, ne soient pas regroupés. Ceci est possible grâce au mode d'ionisation puisque l'électrospray ne produit quasi exclusivement que des ions à nombre pair d'électron. A partir des signaux ainsi extraits, la masse moléculaire est calculée par ajout à la masse précise mesurée de la masse du proton (1.00728). Il s'agit ici d'une approximation puisque tous les rapports m/z mesurés sont considérés comme monochargés, mais dans la pratique, la plupart des composés présents dans les urines sont détectés sous cette forme. Néanmoins, une optimisation de l'approche pourrait inclure le développement d'un script de recherche du composé correspondant au  $^{13}\text{C}_1^{12}\text{C}_{n-1}$  dans la seconde table, afin de pouvoir calculer l'état de charge de l'ion considéré (m/z du  $^{13}\text{C}_1^{12}\text{C}_{n-1}$  décalé de

1.00335/z par rapport au m/z de l'ion du  $^{12}\text{C}_n$ ) et corroborer automatiquement la formule déterminée à partir de la table 1. Une fois la masse moléculaire calculée, la masse de Kendrick est obtenue en corrigeant la masse IUPAC du groupement méthylène de 14.01565 à 14 par définition d'après l'Équation 1 :

$$\text{Masse de Kendrick} = \frac{\text{Masse IUPAC} \times 14}{14,01565}$$

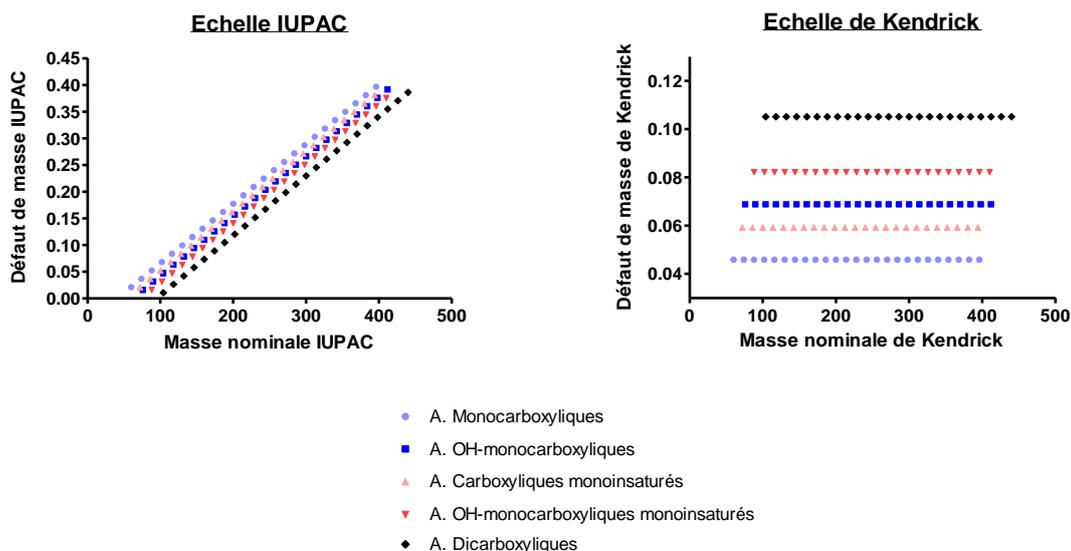
**Équation 1 : Calcul du changement d'échelle de masse du standard IUPAC vers Kendrick.**

D'où la masse nominale de Kendrick et le défaut de masse de Kendrick sont déduits d'après l'Équation 2 :

$$\text{Défaut de masse de Kendrick} = \text{Masse nominale de Kendrick} - \text{Masse exacte de Kendrick}$$

**Équation 2 : Calcul du défaut de masse de Kendrick.**

Il ne reste alors plus qu'à tracer les défauts de masse de Kendrick en fonction de la masse nominale. Il s'ensuit une organisation des signaux en fonction de la classe chimique des composés auxquels ils se rapportent, une classe chimique étant définie comme le regroupement de l'ensemble des compositions élémentaires ayant une composition identique en hétéroatomes (mêmes atomes et même quantité) et un même nombre d'insaturation. Dans la représentation de Kendrick, les composés d'une classe possèdent le même défaut de masse et se retrouvent donc alignés, chacun d'entre eux étant séparés du suivant par un intervalle fixe de 14u, correspondant à l'équivalent d'une elongation de la formule du précédent par une unité méthylène. A titre d'exemple, la Figure 43 présente la réalisation de représentations graphiques du défaut de masse en fonction de la masse nominale en échelle IUPAC et de Kendrick. Ces représentations correspondent à différentes « classes » issues des acides carboxyliques aliphatiques, modifiés par adjonction d'une insaturation ou d'un hétéroatome (un oxygène, équivalent à une hydroxylation).



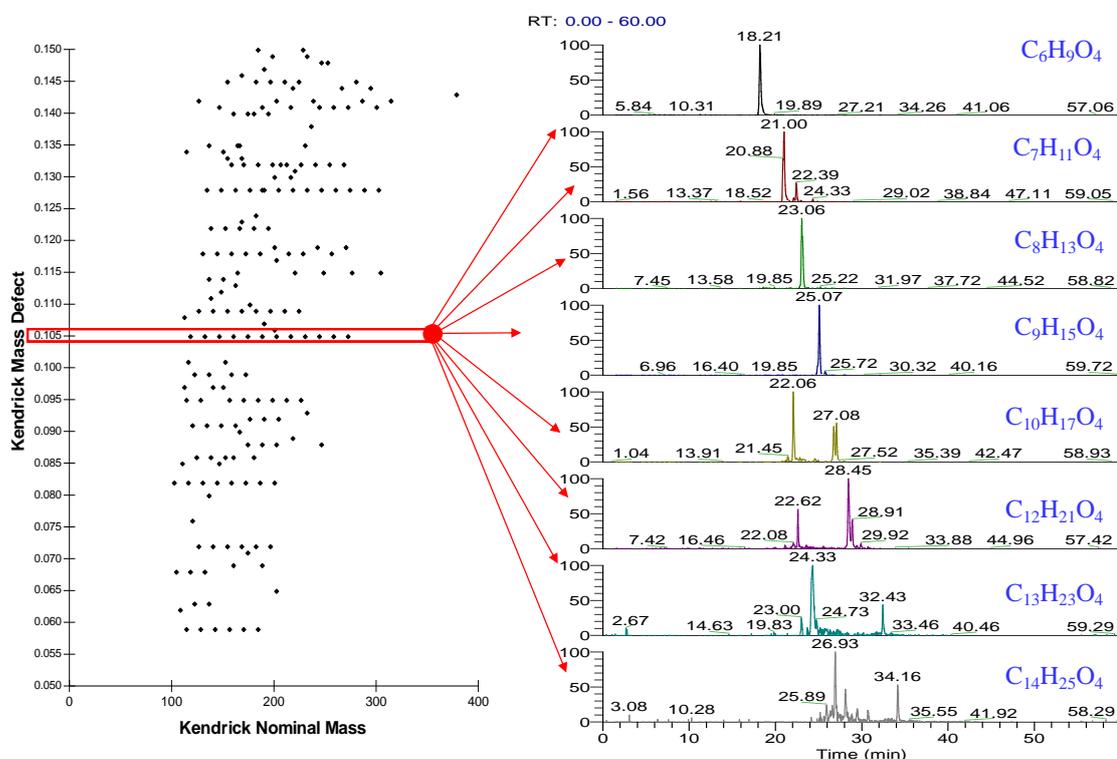
**Figure 43 : Comparaison des représentations en échelle IUPAC et de Kendrick de classes issues des acides carboxyliques aliphatiques.**

Graphiques obtenus en utilisant les masses théoriques pour l'échelle IUPAC, générées à partir des compositions élémentaires des acides carboxyliques aliphatiques, comprenant un nombre de carbone compris entre 2 et 25. Les masses de Kendrick ont été obtenues par le calcul à partir des précédentes.

On peut alors clairement visualiser l'intérêt de la représentation de Kendrick, où l'ensemble des composés de la même classe est aligné sur une seule ligne et séparé de 14u, l'introduction d'un nouvel hétéroatome ou d'une insaturation conduisant à un décalage du défaut de masse de Kendrick et l'ensemble des composés de cette nouvelle classe étant à nouveau alignés et séparés de 14u. On notera également que l'augmentation du nombre d'insaturation d'une unité entre deux classes ou l'insertion du même hétéroatome conduira toujours au même décalage en termes d'indice de Kendrick. Se pose alors le problème de la séparation des composés de masses proches et des isotopes. En effet, la diversité des molécules présentes dans un milieu biologique conduit à la nécessité de discriminer, par exemple, les composés  $^{14}N_y^{12}C_n H_x$  des composés  $^{14}N_{y-1}^{12}C_{n+1} H_{x+2}$ . Un calcul rapide montre que la différence de masse entre ces deux composés est de 0.0115 ( $12+2*1.00728-14.00307$ ), ce qui se traduit par un écart d'indice de Kendrick de 0.0115. De la même façon, il est nécessaire de discriminer les signaux correspondant aux massifs isotopiques des composés d'une classe chimique d'avec ceux d'une série voisine. On se retrouvera parfois dans la nécessité de différencier des composés  $^{13}C_2^{12}C_{n-2} H_x$  et des composés  $^{12}C_n H_{x+2}$ , comportant une insaturation de moins que la série d'homologues initialement considérée. Un calcul rapide montre que la différence de masse entre ces deux composés est de -0.00786 ( $2*1.00335 - 2*1.00728$ ) qui se traduit donc par un delta d'indice de Kendrick de 0.00785. Le recours à des indices de Kendrick à 3 décimales, dont le calcul est possible grâce à la mesure de masse précise obtenue au moyen du système Orbitrap<sup>®</sup>, permet de différencier ces composés et montre l'applicabilité théorique de cette approche à un milieu aussi complexe qu'un fluide biologique.

Appliquée à une empreinte métabolique d'urine de rat, acquise en mode négatif avec un système HPLC-Orbitrap<sup>®</sup>, la démarche ci-dessus a permis de mettre en évidence une ligne de composés,

correspondant à une classe chimique avec un défaut de masse de Kendrick égal à 0.105 (Figure 44). Ces composés possèdent donc la même formule brute que les acides dicarboxyliques aliphatiques. Cependant, comme le montre la partie droite de la Figure 44, à une formule brute donc une masse exacte peuvent correspondre plusieurs composés, d'où la présence de plusieurs pics pour certains chromatogrammes extraits des rapports m/z des acides carboxyliques aliphatiques. Il est en effet possible que soient résolus chromatographiquement des isomères de fonction ou de position. On atteint ici la limite des indices de Kendrick, qui correspond plus exactement à la limite de la détection en spectrométrie de masse, basée sur la mesure précise du rapport m/z d'un composé, et qui n'est donc pas capable de différencier des isobares. Dans cette représentation, les isomères de fonction possédant le même nombre d'hétéroatome et d'insaturation ne pourront donc pas être discriminés, il sera ainsi impossible de différencier une cétone avec une fonction alcool d'un acide carboxylique. De la même façon, des composés possédant une insaturation ou un groupement hydroxyl mais à des positions différentes ne pourront être résolus.



**Figure 44 : Extrait d'une représentation de Kendrick d'un profil métabolique urinaire de rat.**

Profil obtenu en mode négatif sur un système HPLC-Orbitrap<sup>®</sup> à un pouvoir de résolution de 60000. Seules les masses paires, déterminées par calcul en retour à partir des rapports m/z impairs détectés, sont représentées. Le défaut de masse de Kendrick correspondant aux acides carboxyliques aliphatiques (0.105) est mis en évidence (cadre). L'ensemble des chromatogrammes extraits correspondant aux m/z des ions monochargés des acides carboxyliques (tolérance  $\pm 5$  ppm) est présenté sur la droite.

Dans un contexte biologique où la diversité chimique est particulièrement importante, il faut donc avoir recours à l'acquisition de spectres de fragmentation pour confirmer la structure des composés regroupés à l'aide des indices de Kendrick, afin de s'assurer qu'il s'agit bien d'une série

d'homologues, et non juste de composés ayant la même formule brute à un groupement méthylène près.

Ceci nous amène à proposer une stratégie d'utilisation des indices de Kendrick, décrite dans la Figure 45, qui comporte une requête en base de données sur la base de la formule brute, déterminée à l'aide des indices de Kendrick, afin d'évaluer la diversité de composés pouvant correspondre à ces compositions élémentaires. Puis, dans un second temps, des expériences de fragmentation, dédiées à la confirmation de l'appartenance du composé à la classe chimique d'intérêt ou à la discrimination entre les différentes réponses renvoyées par les bases de données.

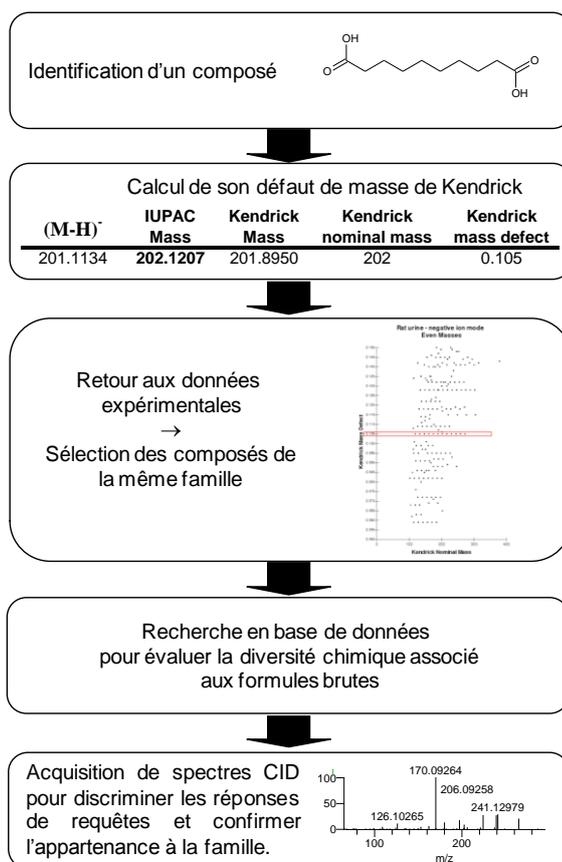


Figure 45 : Stratégie d'approche de données de métabolomiques pour l'isolation de série d'homologues en approches semi-ciblées ou pour la construction de bases de données chimiques.

L'application de cette stratégie pour la classe des acides dicarboxyliques dans les urines de rats permet d'isoler 12 composés. Le Tableau 18 résume les signaux détectés dans les urines de rats ayant un indice de Kendrick de 0.105, correspondant aux acides dicarboxyliques. Il présente également les résultats des requêtes sur les bases de données HMDB et KEGG, menées à partir la formule brute des composés, déterminée d'après la mesure expérimentale de leur masse précise. Comme l'avait laissé supposer la Figure 44, où l'on observait plusieurs pics chromatographiques pour un même rapport m/z, les résultats des requêtes en base de données montrent qu'à une formule brute, peuvent correspondre plusieurs composés biologiquement pertinents (Tableau 18). L'ultime

### CHAPITRE III : Travaux personnels

étape de la démarche proposée à la Figure 45 consiste à mener des expériences de fragmentation afin de tenter d'identifier les composés de la même série chimique.

**Tableau 18 : Rapports m/z ayant un indice de Kendrick de 0.105 et résultats des requêtes dans les bases HMDB et KEGG, obtenus à partir des formules brutes leur correspondant.**

Composé attendu	Masse exacte	Défaut de masse de K.	Masse expérimentale	Formule	Erreur (ppm)	KEGG	HMDB
Acide Butanedioïque	118.0266	0.105	118.0272	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	5	Succinic acid=butanedioic acid	Methylmalonic acid
						Methylmalonate	succinate
						Methyl oxalate	
						Pentanedioic acid=glutarate	Glutaric acid
						Acetolactate	Ethylmalonic acid
Acide Pentanedioïque	132.0423	0.105	132.0428	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>	4	Deoxyribonolactone	Methylsuccinic acid
						4-Hydroxy-2-oxopentanoate	Dimethylmalonic acid
						succinate semialdehyde	
						3-Hydroxy-3-methyl-2-oxobutanoic acid	
						dihydroxypentan-2,3-dione	
						2-Aceto-2-hydroxybutanoate	Methylglutaric acid
						Mevaldic acid	adipate
Acide Hexanedioïque	146.0579	0.105	146.0583	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	2.9	2-Dehydropantoate	2,2-Dimethylsuccinic acid
						3-Hydroxy-3-methyl-2-oxopentanoic acid	
						Adipate	
						2-Aceto-2-hydroxybutanoate	
						4-Hydroxy-2-oxohexanoic acid	
Acide Heptanedioïque	160.0736	0.105	160.074	C <sub>7</sub> H <sub>12</sub> O <sub>4</sub>	2.5	6-Carboxyhexanoate=pimelate	3,3-Dimethylglutarate
						3,4-Dihydroxycyclohexane-1-carboxylate	pimelate
Acide Octanedioïque	174.0892	0.105	174.0896	C <sub>8</sub> H <sub>14</sub> O <sub>4</sub>	2	Suberic acid	Suberic acid
						dimethyladipate	ethyladipate
Acide Nonanedioïque	188.1049	0.105	188.1051	C <sub>8</sub> H <sub>16</sub> O <sub>4</sub>	1.1	Nonanedioic acid	Nonanedioic acid
Acide Decanedioïque	202.1205	0.105	202.1207	C <sub>10</sub> H <sub>18</sub> O <sub>4</sub>	0.9	Decanedioic acid =sebacic	Sebacic acid
Acide Undecanedioïque	216.1362	0.105	216.1363	C <sub>11</sub> H <sub>20</sub> O <sub>4</sub>	0.7	Pas de réponse	Undecanedioic acid
Acide Dodecanedioïque	230.1518	0.105	230.1519	C <sub>12</sub> H <sub>22</sub> O <sub>4</sub>	0.3	Dodecanedioic acid	Dodecanedioic acid
						diisopropyladipate	
Acide Tridecanedioïque	244.1675	0.105	244.1676	C <sub>13</sub> H <sub>24</sub> O <sub>4</sub>	0.5	Pas de réponse	1,11-Undecanedicarboxylic acid
Acide Tétradecanedioïque	258.1831	0.105	258.1832	C <sub>14</sub> H <sub>26</sub> O <sub>4</sub>	0.2	Dibutyl adipate	Tétradecanedioic acid

A titre d'exemple, revenons à *m/z* 201.1134 correspondant au [M-H]<sup>-</sup> de l'acide sébacique ou décanedioïque. Sur la Figure 44, on voit qu'il existe 3 pics majeurs à 22.06, 26.5 et 27.08 min sur le chromatogramme extrait ( $\pm$  5ppm) de cet ion (C<sub>10</sub>H<sub>17</sub>O<sub>4</sub>). La Figure 46 montre des différences importantes dans les spectres de fragmentation de deux composés. A 22.06 min, le spectre de

fragmentation de l'ion  $m/z$  201.1134 fait apparaître deux ions à 157.1233 et 155.1078. Le premier correspond à une perte de  $\text{CO}_2$  tandis que le second, prédominant, vaut pour une perte de  $\text{H}_2\text{CO}_2$ . En mode négatif, ce motif signe une fonction acide substituée par un hydroxyl en  $\alpha$ . Cette substitution est à l'origine d'un mécanisme compétitif d'élimination de la fonction acide qui fait intervenir un intermédiaire cétone et conduit à la perte de  $\text{H}_2\text{CO}_2$ , au contraire des fonctions acides non substituées qui perdent directement  $\text{CO}_2$ . Les ions  $m/z$  139 et  $m/z$  121 correspondent respectivement à la perte d'une et deux molécules d'eau à partir du  $[\text{M}-\text{H}-\text{CO}_2]$ . Ces éléments éliminent donc la possibilité d'être en présence d'une molécule portant une deuxième fonction acide.

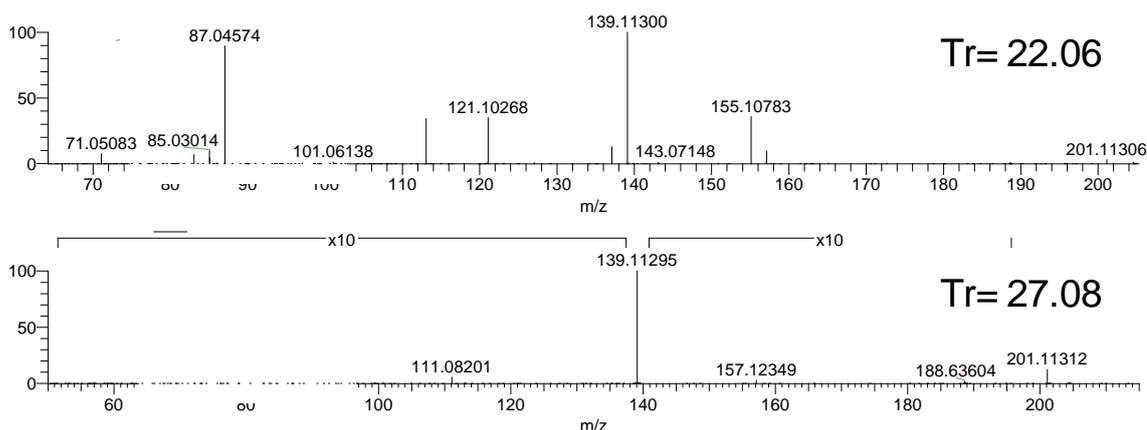


Figure 46 : Spectres de fragmentation CID obtenus à partir des précurseurs  $m/z$  201.1134 à 22.06 et 27.08 min avec un pouvoir de résolution de 7500, une énergie de collision de 35 u.a. et la fonction wide band activation activée.

A l'inverse, à 27.08 min, le spectre de fragmentation de l'ion  $m/z$  201.1134 fait apparaître un ion à 157.1235, équivalent à la perte d'un  $\text{CO}_2$ , caractéristique d'une fonction acide en mode négatif. L'ion majoritaire est l'ion  $m/z$  139.1130, témoin d'une perte d'eau supplémentaire. Enfin, l'ion  $m/z$  111.0820 correspond à une perte de  $\text{C}_2\text{H}_4$  à partir de  $m/z$  139.1130. L'ion majoritaire  $m/z$  139.1130 s'explique par la perte successive d'une molécule d'eau puis d'un  $\text{CO}_2$  qui correspond à la fragmentation classique des acides dicarboxyliques possédant plus de 6 atomes de carbone. En revanche, contrairement aux observations usuelles, l'ion  $m/z$  183, correspondant à la première perte d'eau sur l'une des fonctions acides, n'a pas été produit dans notre système, illustrant une fois encore la difficulté de construction de bibliothèques spectrales de fragmentation universelles en ESI-MS, tant la fragmentation peut être appareil dépendant.

Ce type d'approche a été mené pour l'ensemble des signaux correspondant aux ions des acides dicarboxyliques. Elle a permis de mettre en évidence un schéma de fragmentation conservé pour l'ensemble de la famille. En  $\text{MS}^2$  ces composés perdent une molécule d'eau et de  $\text{CO}_2$ , avec une intensité relative en faveur de la décarboxylation simple pour les acides à chaîne carbonée courte ( $<\text{C}6$ ) ou de la perte combinée d'une molécule d'eau et de gaz carbonique pour les acides à chaîne plus longue. En  $\text{MS}^3$ , on retrouve couramment une perte de 28, correspondant à l'élimination de  $\text{C}_2\text{H}_4$ . Il est à noter cependant que ce schéma a été observé pour plusieurs composés de même masse, résolus chromatographiquement, et qui pourraient correspondre à des isomères de position où les

deux fonctions acides ne seraient pas placées aux deux extrémités de la molécule. Ainsi, il n'est pas possible de discriminer les composés élués à 26.5 et 27.08 sur la base de leur spectre de fragmentation.

Afin de valider l'intérêt des indices de Kendrick et de l'approche proposée pour l'extraction des acides dicarboxyliques présents dans des urines de rats, l'ensemble des standards commerciaux correspondant aux composés ayant été mis en lumière dans les urines ont été achetés et injectés sur le système (Figure 47).

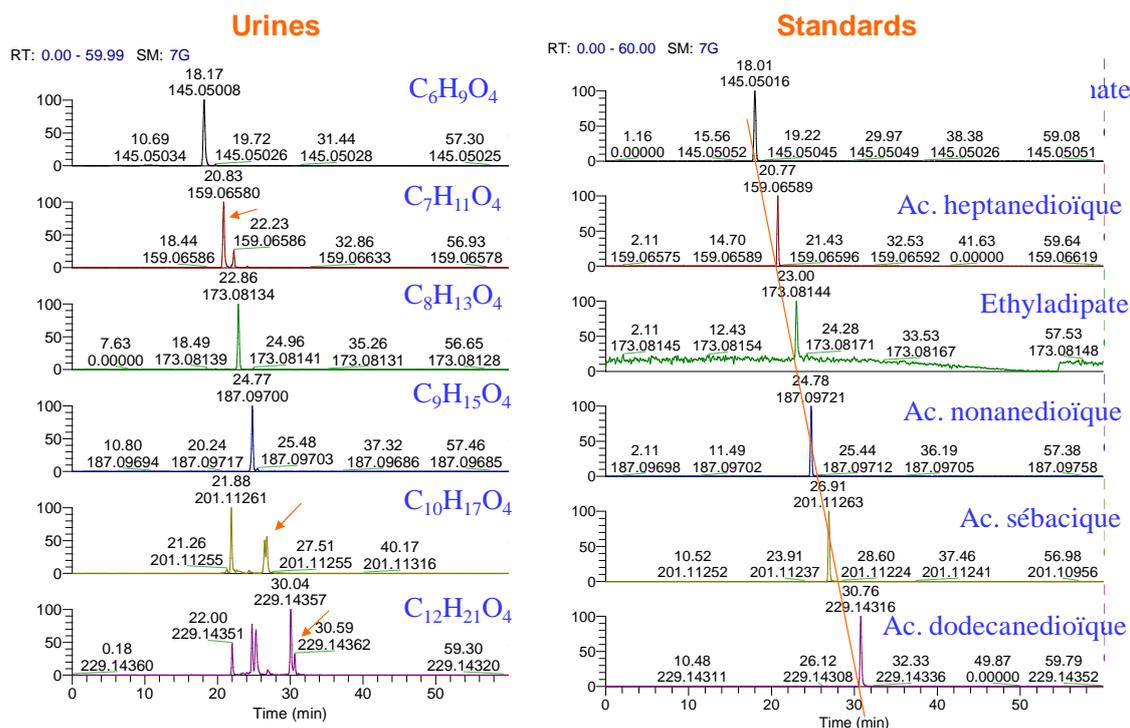


Figure 47 : Comparaison des chromatogrammes extraits pour les rapports m/z des acides carboxyliques dans une urine de rat et pour leurs standards commerciaux.

L'injection de composés de référence qui représente l'ultime étape d'identification d'un composé a permis de lever les ambiguïtés concernant les molécules présentant le même schéma de fragmentation. Par ailleurs, comme en témoigne la Figure 47 (partie droite), ces injections ont également révélé une relation qui semble linéaire entre la longueur de la chaîne carbonée du diacide et son temps de rétention.

Les travaux de Jandera et al.<sup>320</sup> ont montré que pour une série de composés homologues, le facteur de rétention d'un composé comprenant  $N_{CH_2}$  unité méthylène pouvait être modélisé par l'Équation 3 :

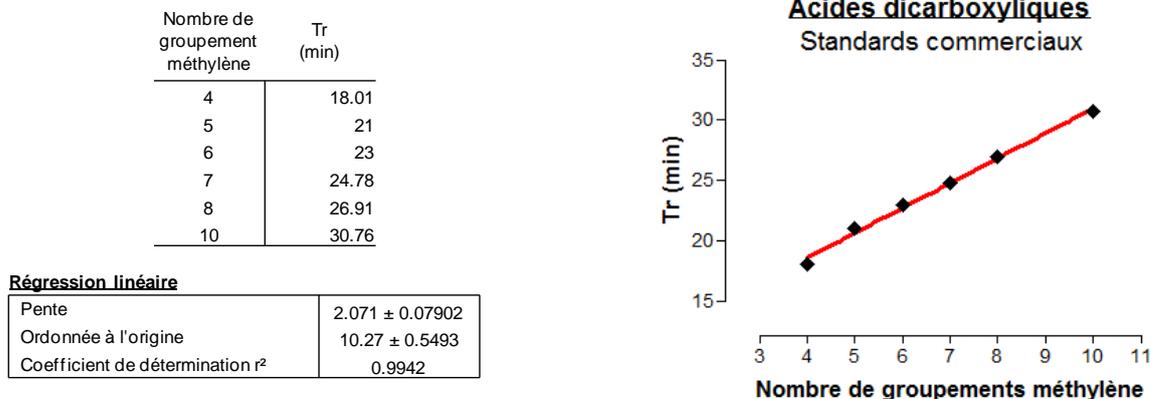
$$\log(k_{N_{CH_2}}) = \log(\beta) + \log(\alpha_{CH_2}) \cdot N_{CH_2}$$

Équation 3 : Relation entre le facteur de rétention et le nombre d'unité méthylène d'un membre d'une série homologue

où  $\beta$  et  $\alpha_{CH_2}$  (sélectivité méthylène) décrivent les contribution non spécifique et spécifique de la sélectivité méthylène à la rétention du composé.

A la différence de la notion de classe chimique, définie précédemment pour les indices de Kendrick, une série de composés homologues fait référence à des composés possédant une même structure de base (ex. l'acide acétique) et dont les autres membres ne diffèrent que par leur nombre d'unité méthylène (l'ensemble des acides carboxyliques aliphatique non ramifié). Cette dernière notion est donc plus restrictive. L'Équation 3, initialement établie en condition d'élution isocratique a, par la suite, été étendue aux gradients linéaires. Jandera et al. ont ainsi démontré que la diminution de  $\log \alpha_{CH_2}$  avec l'augmentation de la force éluante de la phase mobile (ou du pourcentage de solvant fort) répondait à un modèle quadratique et même linéaire lorsque la quantité de solvant fort incorporé restait limitée<sup>320</sup>. Il en découle qu'en condition de gradient linéaire, les composés appartenant à une série homologue seront élués à des intervalles plus ou moins réguliers en fonction de leur nombre d'unité méthylène. Leur rétention sera donc bien une combinaison d'un facteur non dépendant de la longueur de leur chaîne carbonée ( $\beta$ ), correspondant à leur structure de base commune, et d'un facteur dépendant de leur nombre d'unité méthylène ( $\alpha_{CH_2}$ ), la sélectivité méthylène.

L'utilisation de cette propriété doit pouvoir être mise à profit pour exploiter plus avant la représentation de Kendrick. En effet, si les indices de Kendrick permettent de classer les molécules en classes chimiques dont la structure de base est caractérisée de façon limitée par sa formule brute, l'exploitation de la sélectivité méthylène pourrait permettre de différencier les isomères de fonction et éventuellement de position.



**Figure 48 : Temps de rétention des acides dicarboxyliques en fonction de leur nombre d'unité méthylène.**

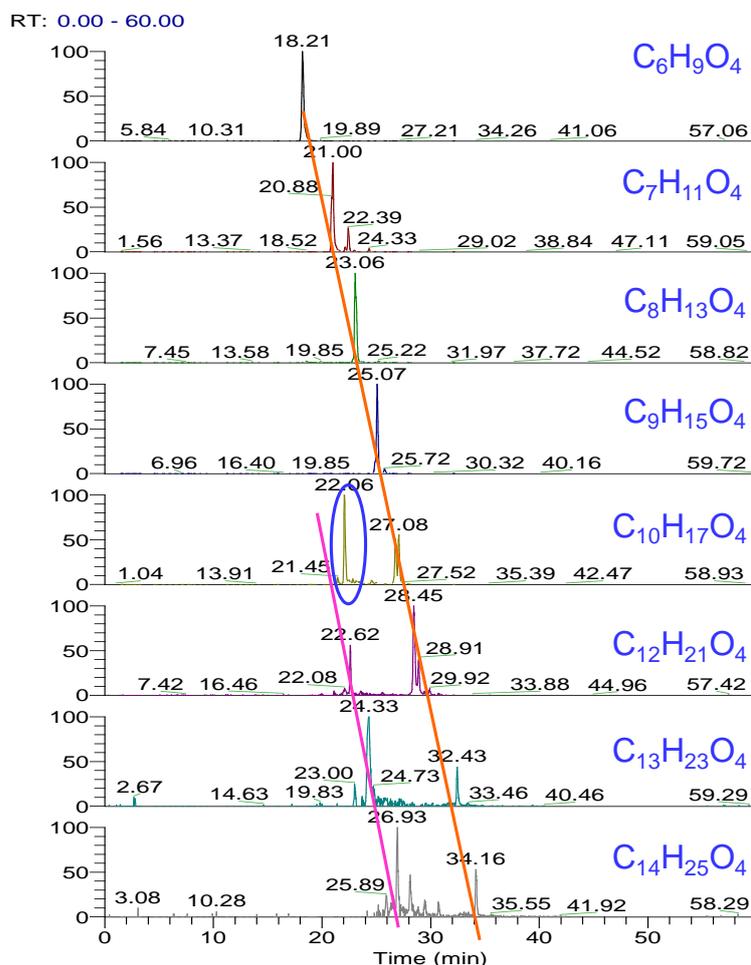
Temps de rétention obtenus en gradient linéaire sur un système HPLC-Orbitrap® par injection de standards commerciaux.

La première étape a consisté à injecter les standards commerciaux d'acide dicarboxyliques pour vérifier que leur rétention était bien linéairement reliée à leur nombre d'unité méthylène dans nos conditions chromatographiques (Figure 48). La linéarité étant bien respectée dans nos conditions d'élution avec des standards, les urines provenant de trois rats différents ont été analysées afin d'évaluer l'impact d'un milieu complexe sur la rétention de ces composés. Il n'est en effet pas rare que la rétention de composés soit modifiée par la matrice qui les contient, particulièrement dans le cas de fluides biologiques. L'injection des 3 urines différentes a montré que d'une part la linéarité de rétention

était assurée au sein de chaque échantillon mais qu'en plus les modèles linéaires obtenus pour ces trois différentes urines n'étaient pas significativement différents (Tableau 19).

**Tableau 19 : Régressions linéaires du temps de rétention des acides dicarboxyliques en fonction de leur nombre d'unité méthylène pour trois urines de rat.**

	Urine 1	Urine 2	Urine 3
<b>Pente</b>	2.07 ± 0.07	2.10 ± 0.06	2.02 ± 0.06
<b>Ordonnée à l'origine</b>	10.2 ± 0.47	9.86 ± 0.41	10.5 ± 0.42
<b>r<sup>2</sup></b>	0.996	0.997	0.996
<b>IC 95%</b>			
<b>Pente</b>	1.88 to 2.26	1.94 to 2.26	1.85 to 2.19
<b>Ordonnée à l'origine</b>	8.95 to 11.5	8.73 to 11.0	9.32 to 11.7
<b>Test d'égalité des pentes</b>			
<b>F</b>		0.456914	
<b>DDLn, DDLd</b>		2, 12	
<b>P value</b>		0.6438	
<b>Test d'égalité des ordonnées à l'origine</b>			
<b>F</b>		0.58822	
<b>DDLn, DDLd</b>		2, 14	
<b>P value</b>		0.5685	



**Figure 49 : Matérialisation des séries d'homologues pour les composés isobares isolés à partir de leur défaut de masse de Kendrick.**

La ligne orange matérialise les acides dicarboxyliques, le cercle bleu met en évidence un composé possédant une fonction acide associée à deux groupement hydroxyl et une double liaison, et la ligne rose témoigne d'une série d'homologues ayant un indice de Kendrick de 0.105.

En combinant la sélectivité méthylène à la représentation de Kendrick, on parvient à définir deux séries d'homologues et un composé différent parmi la classe chimique ayant un défaut de masse de Kendrick de 0.105 (Figure 49). Il est ainsi possible de séparer les acides dicarboxyliques d'une seconde série de composés ayant le même indice de Kendrick. On remarquera également le composé, élué à 22.06 min, n'appartenant à aucune de ces deux séries d'homologues et qui a été identifié comme un composé ne possédant qu'une fonction acide, mais deux fonctions hydroxyl et une double liaison (Cf. Chapitre IV).

L'association indice de Kendrick / sélectivité méthylène a confirmé son efficacité pour la mise en évidence des acides carboxyliques, des acides OH-carboxyliques monoinsaturés ou non, des acides dicarboxyliques monoinsaturés et des acides dicarboxyliques monoinsaturés hydroxylés.

Cette approche a été étendue à l'analyse d'autres classes chimiques. Le Tableau 20 présente diverses familles de métabolites qui ont pu être isolées dans des urines de rats grâce aux indices de Kendrick.

Tableau 20 : Exemples de familles chimiques de composés biologiques isolées grâce aux indices de Kendrick.

Défaut de masse de Kendrick	Formule brute	Hypothèses	Nombre (Urine)
0.128	$C_2O_5H-(CH_2)_n$	Acides OH-dicarboxyliques	12
0.122	$C_7H_6O_3-(CH_2)_n$	Acides OH-phényl carboxyliques	5
0.105	$C_2O_4H-(CH_2)_n$	Acides dicarboxyliques	11
0.087	$C_2H_5NO_2-(CH_2)_n$	Conjugués glycine Conjugués alanine	7
0.086	$C_6H_6O_2-(CH_2)_n$	Pyrocatechol	5
0.068-9	$CO_3H-(CH_2)_n$	Acides OH-carboxyliques	5
0.059	$C_3H_3O_2-(CH_2)_n$	Acides ène-carboxyliques	6
0.045	$CO_2H-(CH_2)_n$	Acides carboxyliques	5

La démarche est donc applicable à des composés biologiques autres que les lipides comme des conjugués aux acides aminés et des acides phénoliques. Sa restriction d'emploi concerne principalement les composés les plus polaires qui doivent être retenus sur la colonne pour pouvoir la mettre en œuvre.

#### 4.4 Conclusion

L'agencement d'un nombre important de composés en ligne ou classe lorsque l'ensemble des ions détectés dans une urine de rat sont représentés à l'aide des indices de Kendrick témoigne à lui seul de l'applicabilité de la méthode à un milieu biologique. Cette représentation permet d'exploiter la structure chimique des métabolites pour la représentation de larges jeux de données et apporte une

aide pour la détermination de formules brute d'un composé de masse élevée à partir de celle de composés de masse inférieure pour lesquels la composition élémentaire peut être obtenue sans ambiguïté sur des appareils haute résolution. La limitation majeure de l'utilisation de l'approche de Kendrick seule pour une caractérisation des métabolites endogènes au-delà de leur composition élémentaire est la difficulté d'obtenir des spectres de fragmentation pour ceux détectés avec une faible intensité. Il arrive par ailleurs que la fragmentation ne permette pas de différencier plusieurs composés élués à des temps différents. Dans ces conditions, le couplage avec la sélectivité méthylène se révèle d'une grande aide pour l'identification de métabolites endogènes inconnus. Cependant, l'utilisation de cette dernière possède elle aussi des limites. En effet, le nombre important de composés de structure proche conduit à de multiples pics chromatographiques pour une masse donnée même avec une restriction de sélection d'ion fixée à 5 ppm (tolérance de rapport m/z pour la construction des chromatogrammes extraits), c'est pourquoi il est nécessaire de posséder au moins deux standards d'une série d'homologues afin d'être sûr de sélectionner les composés selon la bonne équation reliant leur rétention à leur nombre d'unité méthylène. Enfin, la réussite de l'approche est fortement conditionnée par la précision de la mesure de masse et par la résolution de la méthode chromatographique qui doit impérativement reposer sur l'utilisation d'un gradient linéaire.

Cette approche combinée permet d'envisager une accélération de la construction de base de données pour des familles de composés dont seuls quelques représentants possèdent des standards commercialement disponibles. Elle pourra en outre faciliter l'identification ou l'extraction de familles de composés pour étudier l'impact d'une maladie ou d'un traitement sur l'ensemble de ses membres lorsque l'un d'eux a été mis en évidence par une approche plus classique reposant sur l'utilisation des statistiques multivariées.

Enfin, d'autres représentations, spécifiques à la spectrométrie de masse ultra-haute résolution pourraient venir enrichir le panel d'outils dédié à l'aide à l'identification en milieu biologique. Dans cette perspective, les indices de Van Krevelen <sup>321</sup>, qui reposent sur une représentation des composés en fonction de leur rapport O/C (nombre de carbone sur nombre d'oxygène) en x et H/C en y, sont également susceptibles d'être utilisés. Cette représentation présente l'intérêt d'organiser les composés en fonction de modifications structurales à ratios molaires constants. Ainsi, les métabolites correspondant à des phénomènes de méthylation / déméthylation s'agenceront sur une ligne ayant une ordonnée à l'origine toujours égale à 2 tandis que ceux reliés par une réaction d'hydratation / condensation s'agenceront sur une ligne de pente égale à 2 <sup>321</sup>. Cette représentation a d'ores et déjà été appliquée avec succès pour le regroupement et l'analyse de composés d'intérêt biologiques tels que les lipides, les celluloses, les protéines et les composés polycycliques aromatiques <sup>322-325</sup>.



## 5 IDENTIFICATION DES VARIABLES D'INTERET DISCUSSION DE LEUR PERTINENCE BIOLOGIQUE

### 5.1 Introduction

L'identification des signaux discriminants reste certainement l'étape la plus laborieuse et chronophage des approches métabolomiques. Elle débute classiquement par l'interrogation des bases de données existantes à partir de la masse mesurée, de la composition élémentaire ou d'un spectre de masse. De façon rigoureuse, cette tâche ne peut être accomplie qu'en ayant recours à une combinaison de méthodes d'analyses complémentaires, telles que la RMN, la MS, l'UV ou l'IR. Cependant, les différences de sensibilité entre les méthodes représentent un frein. Il est ainsi 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.

En couplage LC/MS, les sources d'ionisation douces utilisées permettent le plus souvent d'avoir accès à la molécule intacte, même au sein de mélanges complexes. L'essor de la spectrométrie de masse haute résolution avec les détecteurs à temps de vol et les détecteurs à transformée de Fourier a donné accès à la mesure de masse précise et facilite la détermination des compositions élémentaires. Par ailleurs, le recours à des pièges ioniques, par exemple, permet la réalisation d'expériences de fragmentation séquentielle qui peuvent fournir des données structurales complémentaires, nécessaires à l'identification. En revanche, la spectrométrie de masse à source d'ionisation à pression atmosphérique conduit à une forte variabilité de la fragmentation obtenue avec différents instruments, ce qui a empêché la construction de bases de données universelles, à l'instar de celles qui ont été construites en ionisation électronique ou en RMN. Pour l'ensemble de ces raisons, l'identification formelle des métabolites discriminants n'est que rarement atteinte et a pour conséquence que la majorité des signaux détectés par approche métabolomique reste non identifiée.

Notre stratégie d'identification des signaux discriminants reposant sur la technologie Orbitrap est présentée dans la Figure 50. Elle comprend une analyse du spectre de masse afin de vérifier que le signal est bien un ion monoisotopique et non un ion isotopologue ou un ion adduit. Elle se poursuit par la détermination de la composition élémentaire de l'ion d'intérêt afin de pouvoir utiliser cette information pour l'interrogation des bases de données. L'acquisition de spectres de fragmentation permet à ce stade de discriminer les réponses obtenues précédemment sur la base d'ions fragments ou de pertes de neutre, caractéristiques de groupement chimique. Certains composés peuvent encore être rejetés sur la base de leur temps de rétention ou d'expériences complémentaires comme les échanges proton/deutérium qui pourront également être utilisées pour la formulation d'hypothèses de structure. Au final, l'identification formelle d'un composé est obtenue par injection d'un standard sur le même système analytique. Le standard doit alors présenter le même temps de rétention, la même masse exacte et le même profil de fragmentation que le composé inconnu.

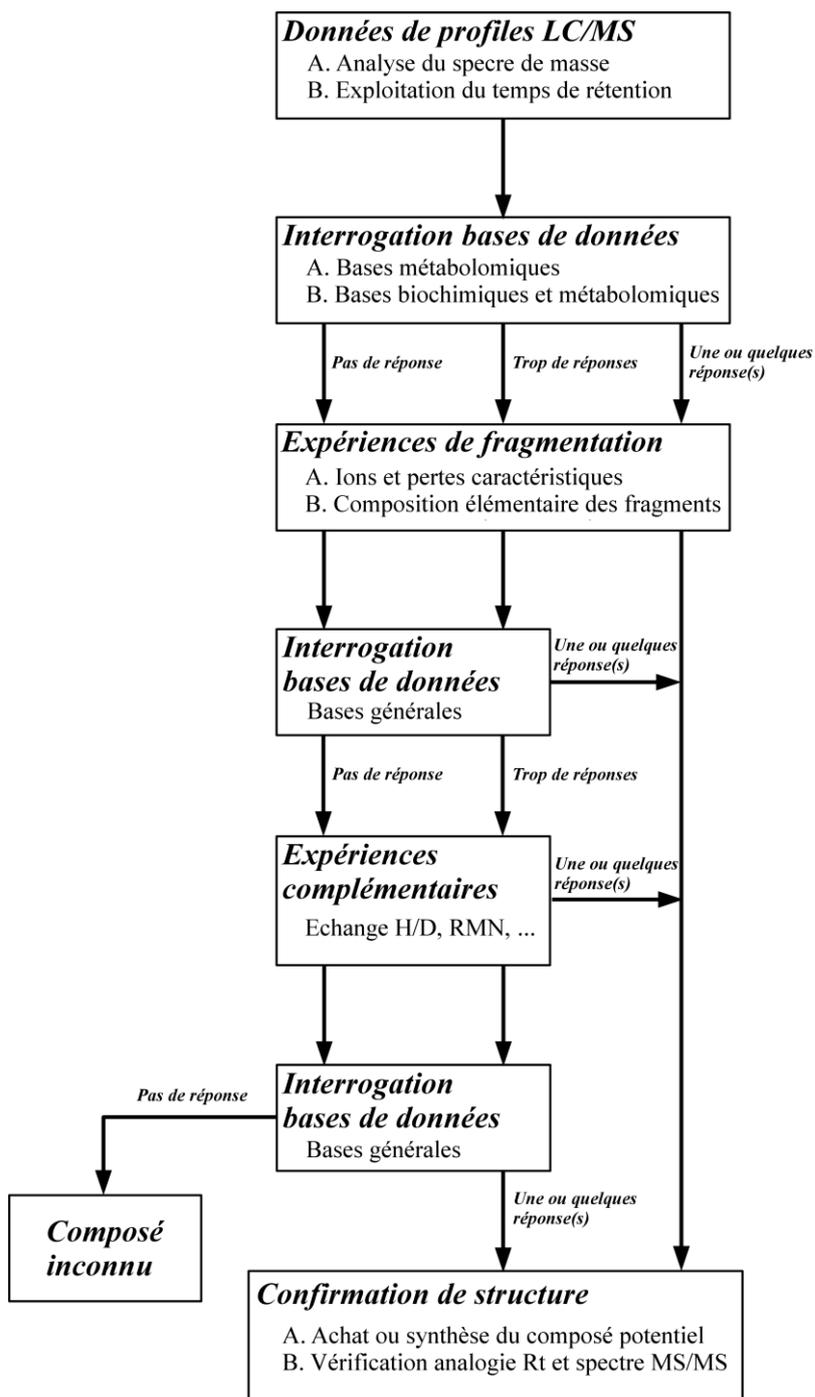


Figure 50 : Stratégie d'identification d'un biomarqueur potentiel mis en évidence par approche métabolomique à l'aide d'expériences complémentaires en LC/MS et d'interrogation des bases de données.

## 5.2 Matériel et méthode

**Expérimentation animale** : Cf. Introduction expérimentale

**Préparation des échantillons et réalisation des expériences d'échange H/D** : Pour aider à l'identification des composés endogènes, des expériences d'échange H/D ont été menées. A cette fin, des pools constitués à partir des urines de rats recueillies avant ou après traitement des animaux avec un composé inducteur ont été lyophilisées puis repris par le même volume de D<sub>2</sub>O. Après une seconde lyophilisation, les pools ont à nouveau été repris par D<sub>2</sub>O afin de s'assurer de l'absence de molécules d'eau résiduelles pour favoriser un échange le plus complet possible et éviter sa réversion. Au cours de ces expériences, l'eau de la phase mobile A a également été remplacée par du D<sub>2</sub>O. Ne disposant pas de méthanol deutéré, les expériences n'ont été conduites qu'avec l'acétonitrile comme solvant organique sur la base des conditions dites acides-acétonitrile, décrites précédemment.

**UPLC-TOF et HPLC-Orbitrap** : Cf. Chapitre II

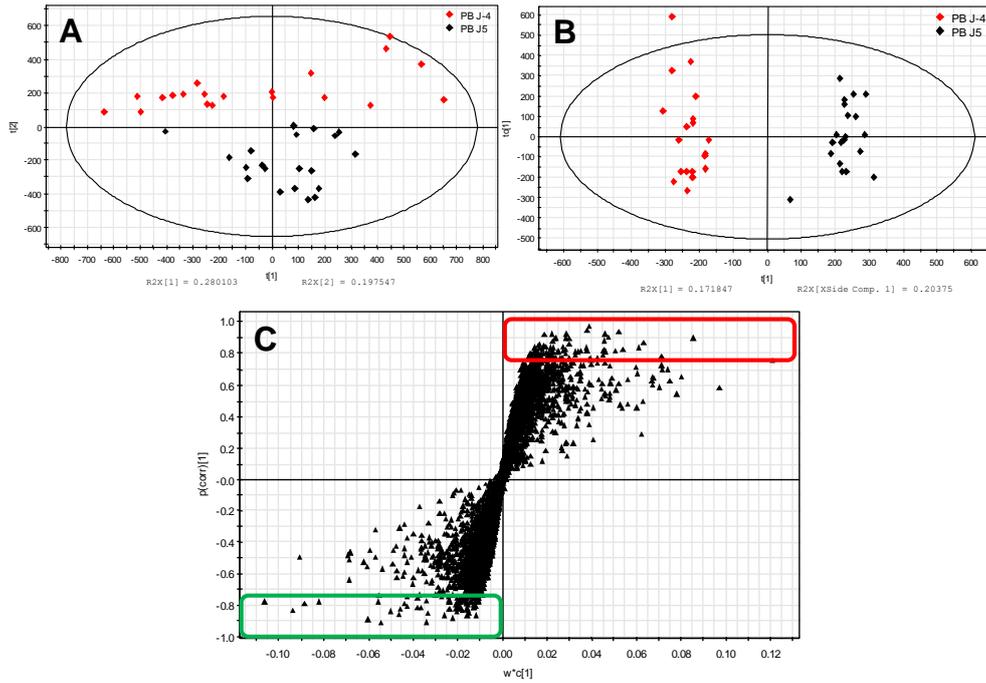
## 5.3 Résultats

### 5.3.1 Exploitation des profils métaboliques

#### 5.3.1.1 Comparaison PD/PT : Groupe PB

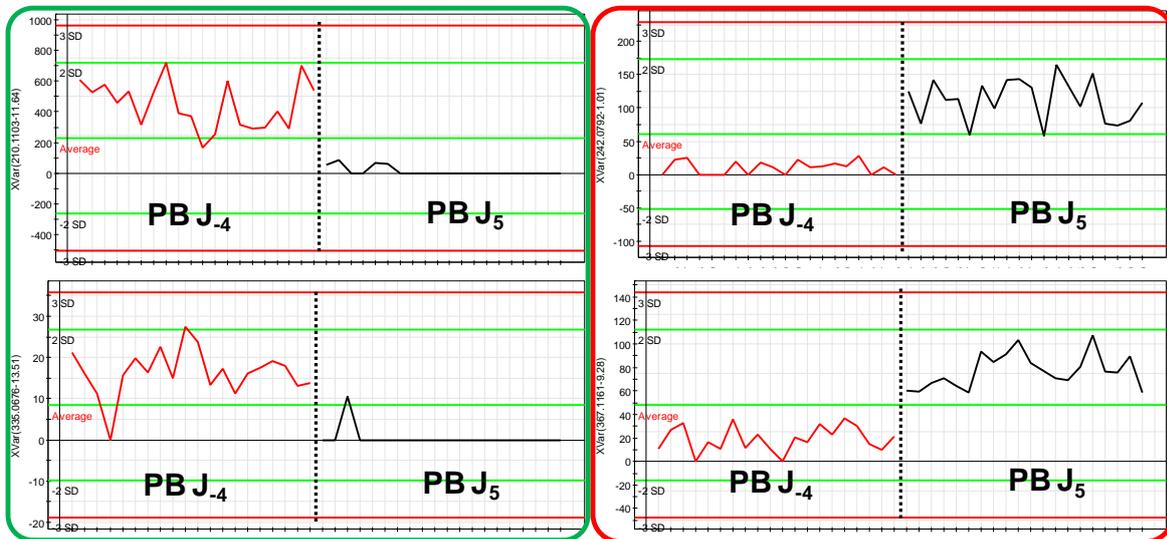
Comme nous l'avons vu au Chapitre II, l'utilisation d'un filtre « absent en prédose (PD) / présent après traitement (PT) » a permis d'isoler les variables associées au métabolisme du PB. Ces variables qui étaient responsables de la séparation des deux groupes en analyses multivariées masquaient les différences plus subtiles des profils urinaires correspondant à l'effet étudié. Après leur identification, elles ont donc été éliminées du jeu de données afin de ne pas contraindre les modèles statistiques (Figure 51). Une fois les métabolites du PB retiré du jeu de données, l'analyse en composante principale (Figure 51A) montre que la principale source de variance des profils n'est plus l'effet traitement puisque la séparation des profils avant et après traitement n'est obtenue que sur la seconde composante. La variabilité interindividuelle associée à la variabilité expérimentale (R<sup>2</sup>X composante 1 = 28 %) a donc un impact métabolique supérieur à l'effet traitement (R<sup>2</sup>X composante 2 = 20 %). Afin de cibler l'effet traitement, une analyse OPLS des données employant le traitement comme variable Y est réalisée et permet de mettre en évidence les variables les plus statistiquement significatives pour la discrimination des profils avant et après traitement (Figure 51C). A titre d'exemple, quelques courbes de tendance, correspondant aux signaux les plus discriminants, sont présentées dans la Figure 52.

### CHAPITRE III : Travaux personnels



**Figure 51 : Comparaison des profils urinaires avant (rouge) et après (noir) traitement au PB après suppression des métabolites exogènes.**

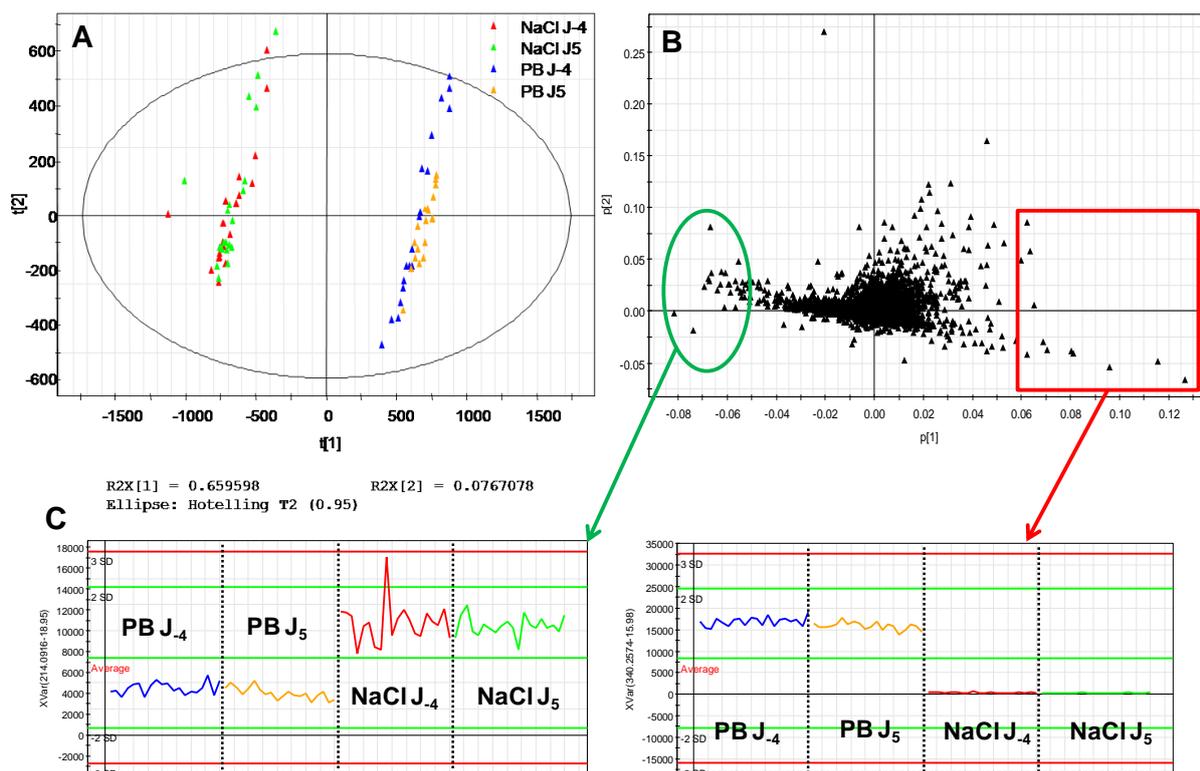
Modèles construits à partir des données obtenues en mode positif (normalisation Pareto). (A) Carte factorielle des individus en ACP, PBJ-4 (rouge) et PBJ5 (noir). (B) Carte factorielle des individus en OPLS, même code couleur. (C) Représentation en « S » pour la carte des poids factoriels du modèle OPLS. Les encadrés signalent les variables les plus discriminantes vis-à-vis de l'effet traitement qu'elles soient diminuées (vert) ou augmentées (rouge).



**Figure 52 : Exemples de courbes de tendance obtenues pour les signaux les plus discriminants en analyse OPLS.**

L'expérimentation animale comprenait également un groupe témoin qui recevait le véhicule (solution de NaCl à 0.9 %). Ce groupe avait pour objectif de suivre les éventuelles évolutions du profil urinaire des rats, indépendantes de l'administration des composés inducteurs, au cours de l'étude. En effet, les animaux utilisés dans l'étude, âgés de 8 semaines, sont toujours en cours de maturation et peuvent donc subir des variations de profil métabolique au cours du temps. De plus, les manipulations

et les injections nécessaires à l'administration des composés inducteurs sont vraisemblablement responsables de l'apparition d'un stress et d'une légère inflammation au point d'injection qui risquent de se traduire au niveau des profils urinaires. La démarche consistait donc à analyser simultanément les 20 individus du groupe contrôle et les 20 du groupe PB pour ne retenir que les modifications affectant uniquement les individus traités par le PB en incluant les profils des rats ayant reçu le véhicule pendant 5 jours aux profils obtenus avant toute manipulation lors des analyses supervisées. La Figure 53 présente l'analyse en composantes principale des 80 profils urinaires correspondants aux groupes NaCl et PB.



**Figure 53 : Analyse en composante principale des groupes contrôle (NaCl) et PB.**

(A) La carte factorielle des individus. (B) La carte factorielle du poids des variables. (C) Exemples de courbes de tendance de variables expliquant la séparation des groupes d'individus observés en (A) et présentant la valeur individuelle de la variable dans chaque échantillon. La représentation regroupe les valeurs en fonction du groupe et du temps de prélèvement.

Elle met en évidence que la majeure partie de la variance des profils urinaire (R2X (composante 1) = 66%) correspond à l'appartenance au groupe NaCl ou PB. L'exploration des signaux responsables de la séparation des individus selon leur groupe (Figure 53B) fait apparaître des variables dont la variabilité interindividuelle au sein d'un groupe est faible mais dont le valeur moyenne est très différente entre le groupe NaCl et le groupe PB (Figure 53C). Ce résultat est particulièrement gênant au regard de l'approche envisagée et de la reproductibilité de l'ensemble du processus d'analyse dans la mesure où les animaux prélevé à J-4 devraient être semblables à la variabilité individuelle près : qu'ils reçoivent par la suite le véhicule ou le PB n'ayant à ce stade normalement aucun impact. La reprise des conditions de l'expérimentation animale a révélé que tous les groupes n'avaient pu être accueillis simultanément en raison du nombre important d'animaux

requis pour l'ensemble de l'étude. De plus, même pour les groupes accueillis sur la même période, le nombre élevé d'animaux a entraîné un étalement des prélèvements pouvant conduire à un impact du rythme circadien sur les profils de différents groupes.

La difficulté tant dans l'approche des données que pour la future interprétation biologique réside dans ce que l'ensemble des variables ne se comporte pas de la même manière. Ainsi, nous avons vu précédemment des signaux dont la valeur moyenne était très différente en intergroupe mais homogène en intragroupe (Figure 53). L'exploration des variables discriminantes isolées à partir de la comparaison des individus du groupe PB avant et après traitement suite à la suppression des signaux des métabolites du PB, et leur analyse en regard des résultats obtenus pour le groupe NaCl met en lumière des variables dont les profils sont différents de ceux précédemment cités. Le fait de supprimer les contrôles a conduit à isoler des variables dont la variation est le fait d'un effet chronobiologique comme le 8-amino-7-oxo-nonanoate (Figure 54). Pour cette variable, les moyennes des valeurs pour les individus des 4 sous-groupes sont statistiquement différentes (au risque 5%) à l'exception des deux sous-groupes J5. En outre, on trouve des signaux pour lesquels les moyennes à J-4 ne sont pas statistiquement différentes entre le groupe PB et le groupe NaCl, et qui présentent une variation statistiquement significative uniquement pour les J5 du groupe PB ( $m/z$  366 à 9.28 min). Cependant, même pour ces variables, il est impossible de discriminer celles qui seraient de réels biomarqueurs d'induction plutôt que des biomarqueurs d'activité pharmacologique ou de toxicité. Enfin, les variations intergroupes n'étant pas comparables d'une variable à l'autre, aucun procédé de normalisation des valeurs mesurées dans les urines, en fonction de la diurèse ou de la créatinine, ne permet d'apporter une amélioration de la distribution des données. Pour qu'une telle approche puisse être appliquée, il faudrait être en mesure de prouver que l'intensité de la variable étudiée est bien corrélée avec sa concentration.

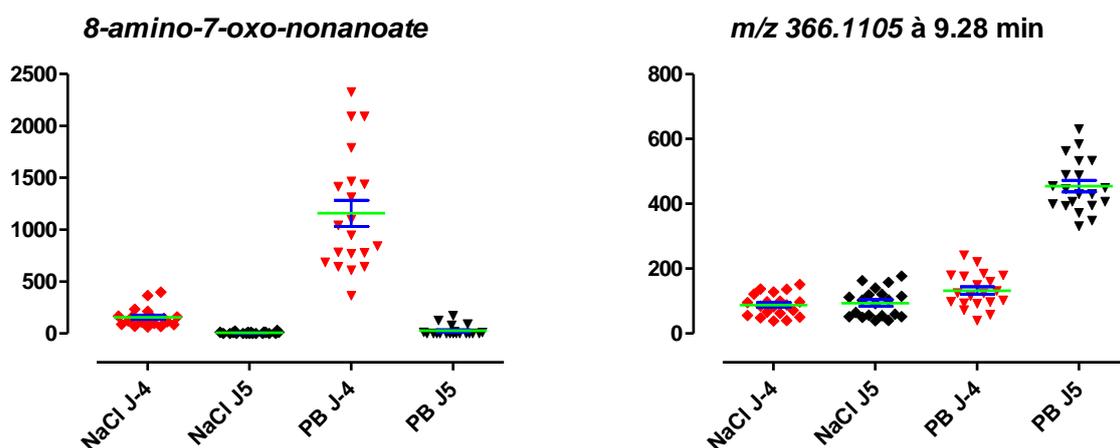


Figure 54 : Valeurs individuelles pour des métabolites discriminant les individus avant et après traitement au PB et comparaison aux valeurs obtenues pour les individus du groupe NaCl.

Moyenne (ligne verte)  $\pm$  écart-type de la moyenne (Barre d'erreur bleue)

### 5.3.1.2 Recherche de biomarqueurs communs : Groupes BNF/3MC

Le groupe contrôle (Huile de maïs) correspondant aux inducteurs 3-MC et BNF a souffert des mêmes problèmes d'hébergement des animaux et de temps de prélèvement que le groupe NaCl, empêchant à nouveau son exploitation en analyses multivariées. Cependant, les mesures d'expression des ARN<sub>m</sub> et des activités des CYPs réalisées (Cf Introduction expérimentale) montrent que les deux inducteurs 3-MC et BNF ont un profil d'induction semblable alors qu'ils sont administrés avec le même véhicule. Il a donc été décidé de tirer profit de cette similitude en extrayant les variables qui étaient modifiées avant et après traitement dans les deux groupes.

Une première approche, basée sur les statistiques classiques, a été appliquée aux profils obtenus avec le système UPLC-TOF. Pour chaque groupe (3-MC ou BNF), un test de Student a été réalisé sur chaque variable afin de sélectionner celles dont les moyennes différaient significativement avant et après traitement. En utilisant les informations de masse précise et de temps de rétention, les variables modifiées communes aux deux groupes ont pu être sélectionnées. Pour ce faire, les critères de sélection retenus étaient une différence de masse inférieure à 7 ppm et une différence de temps de rétention inférieure à 0.1 min. Les résultats de cette approche sont présentés dans le Tableau 21.

**Tableau 21 : Approche en statistiques classiques pour la sélection de variables communément modifiées après un traitement au 3-MC ou à la BNF.**

	Mode positif		Mode négatif	
	3-MC	BNF	3-MC	BNF
Variables extraites (MarkerLynx <sup>®</sup> )	2334	3580	3031	8211
Significativement différentes J-4 vs J5 *	398	606	457	616
Communes ( $\Delta M < 7$ ppm, $\Delta t_R < 0.1$ min)		81		55
Retrouvées sur le système Orbitrap		56		37
Exploitable (Formule brute générée)		46		32
Groupes de métabolites		26		23
Groupe avec expérience de fragmentation		21		19

\* Test-t :  $\alpha = 0.05$  pour le groupe 3-MC et  $\alpha = 0.01$  pour le groupe BNF

L'inconvénient majeur de l'approche en statistique classique (test de Student) vient de ce qu'au cours de la réalisation de ce test, il n'est pas tenu compte de l'intensité de la variable ni de sa variabilité intra-classe. En conséquence, de nombreuses variables sélectionnées appartiennent au bruit de fond ou présentent des variabilités intra-classe élevées qui conduisent au chevauchement des valeurs prises par la variable dans chacun des deux groupes comparés, de sorte qu'elles possèdent un pouvoir prédictif limité.

Par la suite, grâce à l'amélioration des capacités de traitement des logiciels d'extraction de variables (XCMS et RefinerMS), l'ensemble des groupes a été retraité simultanément de façon à ce que les profils de tous les individus soient caractérisés par des variables identiques au sein d'un même jeu de données et puissent être traités en analyses multivariées. En effet, une récente publication présente une extension de l'utilisation de la représentation en S (Décrite plus en détail au Chapitre I) pour la recherche de variables communes à deux effets ou traitements à partir de deux

analyse OPLS <sup>326</sup>. Dans cette nouvelle représentation, baptisée SUS plot (Shared and Unique Structure), la première OPLS oppose le groupe contrôle au groupe recevant le premier traitement et la seconde oppose le même groupe contrôle au groupe recevant le second traitement. Par construction, lorsque l'effet traitement est utilisé comme variable Y, la séparation des groupes selon Y est effectuée selon la première composante. Le SUS plot va consister à tracer la corrélation des variables avec la classe selon la composante 1 du second modèle contre la corrélation des variables avec la classe selon la composante 1 du premier modèle. Ceci revient à tracer pour chaque variable sa valeur sur l'axe des ordonnées du S plot du modèle 2 en fonction de valeur sur l'axe des ordonnées du S plot du modèle 1. Ainsi, une variable dont l'intensité présente une forte corrélation avec l'attribution de la classe et qui est augmentée dans les deux modèles se retrouvera dans le coin supérieur droit tandis qu'à l'inverse, si elle est diminuée et présente une forte corrélation avec l'attribution de la classe, elle sera placée dans le coin inférieur gauche. En outre, lorsqu'une variable est spécifique à la séparation dans un seul des deux modèles, sa corrélation avec l'attribution de la classe dans le second modèle est faible et elle se retrouve donc très proche de l'axe du modèle pour lequel elle est contributive. Ceci est particulièrement intéressant puisque typiquement les métabolites sont spécifiques de la molécule administrée et sont donc isolés des variables endogènes. L'extraction des potentiels biomarqueurs en est grandement facilitée.

La Figure 55 retrace la mise en œuvre de cette approche pour les profils urinaires en mode positif des groupes 3-MC et BNF avec :

- une première OPLS opposant les contrôles (composés des profils urinaires à j-4 pour les 2 groupes) aux profils urinaires J5 du groupe 3-MC (Figure 55A) et le S-plot associé (Figure 55C)
- Une seconde OPLS opposant les contrôles (comprenant les mêmes profils urinaires à j-4 que précédemment) aux profils urinaires J5 du groupe BNF cette fois (Figure 55B) et le S-plot associé (Figure 55D)
- Le SUS plot lui-même (Figure 55 E). L'axe des abscisses correspond à la corrélation des variables avec la classe pour l'effet traitement 3-MC tandis que l'axe des ordonnées correspond à la corrélation des variables avec la classe pour l'effet traitement BNF. Les rectangles rouges matérialisent les variables augmentées ou diminuées de façon commune parmi lesquelles peuvent se trouver de potentiels biomarqueurs d'induction. Les rectangles jaunes délimitent des espaces où les variables sont fortement corrélées avec exclusivement l'effet traitement 3-MC ; les métabolites du 3-MC sont susceptibles d'être regroupés dans le rectangle de droite qui correspond aux variables augmentées. Les rectangles bleus sont le pendant des précédents pour le traitement BNF. Enfin, les rectangles verts signalent des métabolites qui seraient fortement corrélés avec les deux effets traitement mais dont le sens de variation serait opposé selon le traitement.

### CHAPITRE III : Travaux personnels

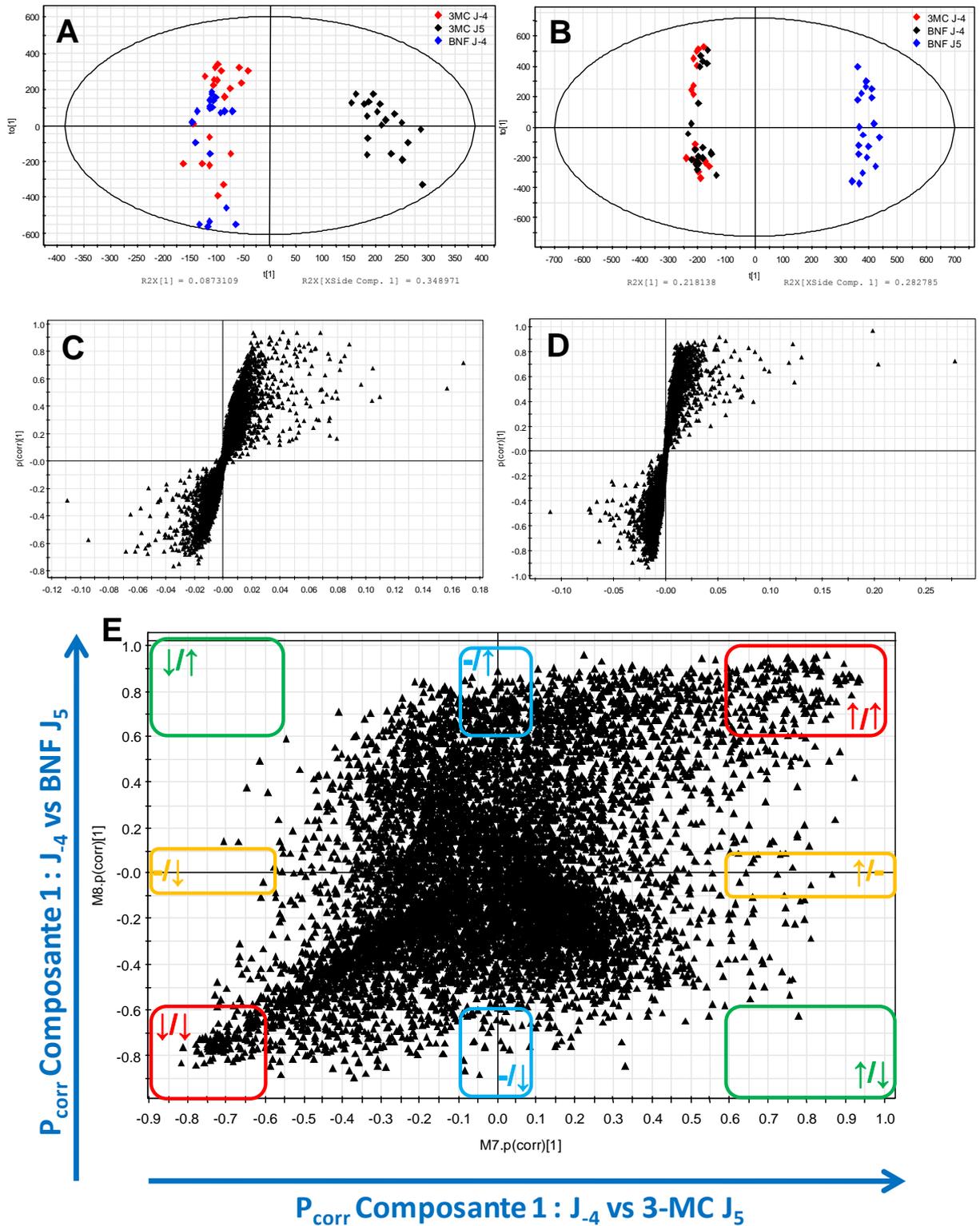


Figure 55 : Analyse combinée des groupes 3-MC et BNF en analyses multivariées.

(A) Carte factorielle des individus : modèle OPLS construit à partir de données normalisées Pareto et opposant les profils  $J_4$  des groupes 3-MC (rouge) et BNF (bleu) aux profils  $J_5$  du groupe 3-MC (noir). (B) Carte factorielle des individus : modèle OPLS construit à partir de données normalisées Pareto et opposant les profils  $J_4$  des groupes 3-MC (rouge) et BNF (noir) aux profils  $J_5$  du groupe BNF (bleu). (C) Carte des poids factoriels du modèle A. (D) Carte des poids factoriels du modèle B. (E) SUS plot construit à partir des modèles A et B. Il correspond au tracé de la corrélation à la composante principale du modèle B (ordonnée) contre celle du modèle A (abscisse).

En ne retenant que les variables corrélées à plus de 70 % avec les deux effets traitements, cette approche permet d'isoler les signaux d'intérêts potentiellement les plus pertinents (Tableau 22).

Tableau 22 : Résultats de l'approche SUS plot pour la sélection de variables communes, régulées de façon similaire avec les deux traitements.

Variables	Mode positif		Mode négatif	
	3-MC	BNF	3-MC	BNF
Extraites (RefinerMS®)	13119		4581	
Discriminantes (Modèle OPLS, IPcorr1 > 0.7)	137	418	28	156
Modifiées communes (IPcorr1 > 0.7) SUS plot	60		4	
Groupes de métabolites	39		4	

La pertinence, en termes d'effet commun, des variables sélectionnées est illustrée par la Figure 56 qui présente les intensités d'intégration individuelles des échantillons de l'ensemble de l'étude pour les 4 signaux isolés en mode négatif.

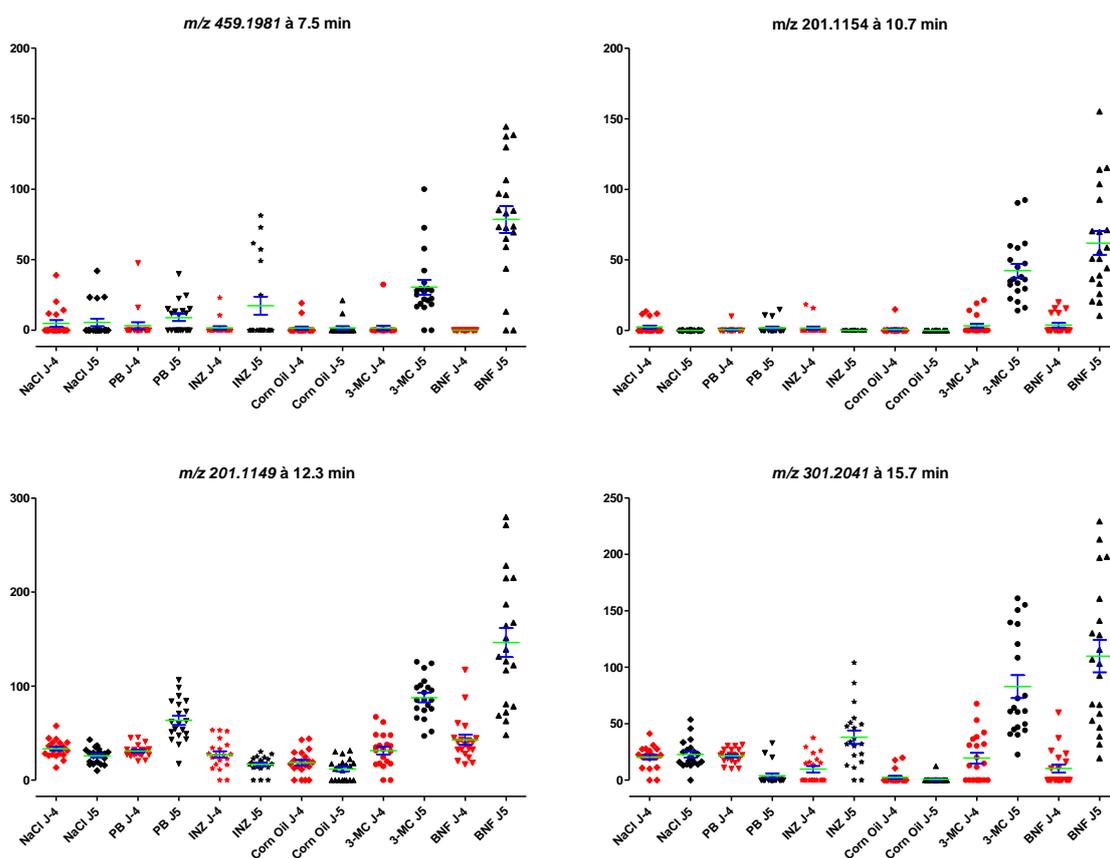


Figure 56 : Valeurs individuelles des 4 variables isolées en mode négatif d'après une évolution commune sous traitement 3-MC et BNF.

Si les signaux mis en évidence sont tous régulés de façon commune pour les groupes 3-MC et BNF, plusieurs d'entre eux sont également modifiés avec un autre traitement, isoniazide ( $m/z$  301 à

15.7 min) ou PB ( $m/z$  201 à 12.3 min) dont les profils d'induction sont très différents. Les résultats sont similaires en mode positif, où la variable  $m/z$  366 à 9.28 min, qui avait été mise en évidence lors de l'analyse du groupe PB (Figure 54), est à nouveau discriminante vis-à-vis d'un effet commun 3MC / BNF (Figure 57). Ceci démontre une fois encore l'incapacité de la structure de l'étude à isoler l'effet étudié. Plusieurs phénomènes peuvent être à l'origine de cette non spécificité : les profils d'induction des composés ne sont peut être pas aussi spécifiques qu'attendus, ou les différents inducteurs sont à l'origine d'un mécanisme pharmacologique ou toxique commun.

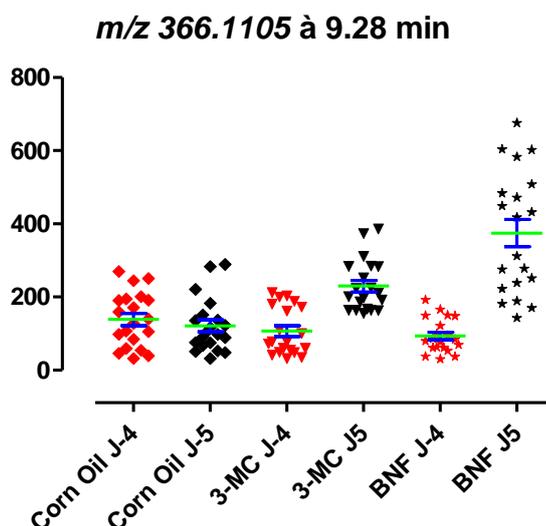


Figure 57 : Valeurs individuelles de la variable  $m/z$  366 à 9.28 pour l'ensemble des groupes analysés ayant pour véhicule l'huile de maïs.

Les signaux issus des dernières stratégies de traitement des données n'ont pas pu faire l'objet de campagne d'identification. Néanmoins, ils pourront être examinés avec intérêt au cours de la prochaine étude dont le design sera amélioré pour mieux contrôler les facteurs environnementaux et tenter de permettre d'isoler l'effet inducteur des autres facteurs confondants.

### 5.3.2 Identification

Une fois les variables discriminantes mises en évidence et regroupées dans la mesure du possible lorsqu'elles correspondent à un même métabolite, la première étape consiste à déterminer la composition élémentaire du composé d'intérêt en exploitant les données de mesure de masse précise afin de pouvoir interroger les bases de données.

#### 5.3.2.1 Détermination de la formule brute

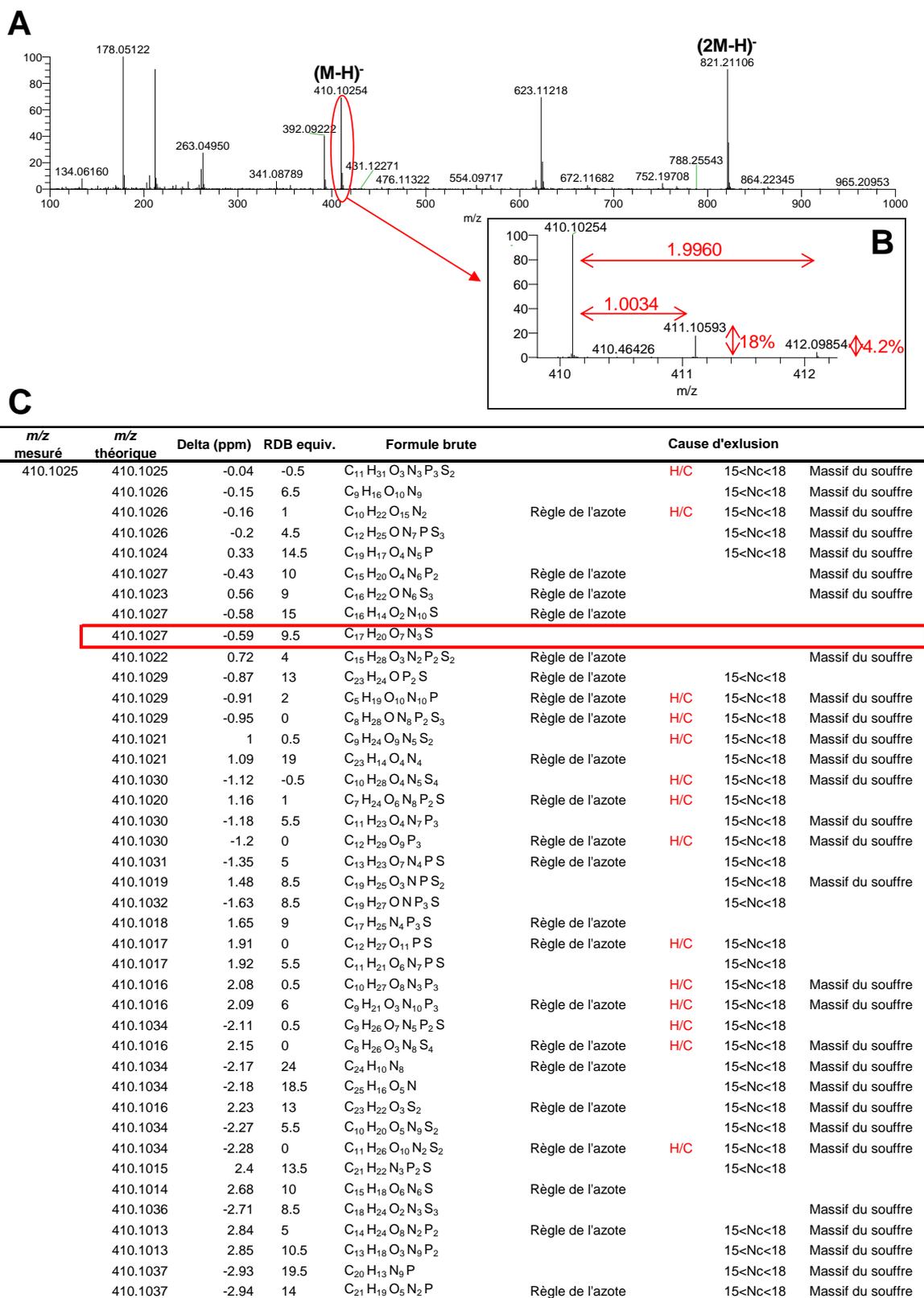
La Figure 58 présente un exemple de détermination de composition élémentaire grâce à un système HPLC-Orbitrap®. L'analyse du spectre de masse au temps de rétention du signal d'intérêt ( $m/z$  410.1025 en mode négatif) permet de s'assurer qu'il s'agit bien d'un ion monoisotopique correspondant à une molécule déprotonée et non d'un adduit ou d'un isotope. Elle révèle en outre la présence d'un dimère du composé à  $m/z$  821.21.

L'analyse détaillée du massif isotopique (Figure 58 B) révèle :

1. La présence du premier cluster isotopique à M+1, ce qui signe le fait que l'ion soit monochargé,
2. Une abondance relative du pic à M+1.0034 correspondant à l'isotope  $^{12}\text{C}_{n-1} \ ^{13}\text{C}_1$  de 18 % qui indique un nombre de carbone proche de 16. En effet, comparativement à l'atome  $^{12}\text{C}$ , l'atome  $^{13}\text{C}$  présente un écart de masse théorique de 1.00335 et une abondance relative de 1.1 %.
3. Une abondance relative du pic M+1.9960 correspondant à l'isotope  $^{32}\text{S}_{n-1} \ ^{34}\text{S}_1$  de 4.2 % qui est la traduction de la présence d'un unique atome de soufre. Comparativement à l'atome  $^{32}\text{S}$ , l'atome  $^{34}\text{S}$  présente un écart de masse théorique de 1.9959 et une abondance relative de 4.4 %. Cette dernière partie de l'analyse illustre l'intérêt des spectromètres de masse à ultra-haute résolution qui permettent de résoudre en masse les différentes espèces du cluster M+2 et dont la mesure de masse précise permet l'assignement du pic sans ambiguïté au soufre tout en excluant un potassium ( $\Delta M+2 = 1.9981$  avec une abondance relative de 7 %.)

En l'absence de contrainte autre qu'une erreur de masse tolérée de 3 ppm et les seuls éléments C, H, N, O, P, et S autorisés, le générateur de formules du logiciel Xcalibur<sup>®</sup> propose 41 formules brutes à partir de l'ion  $m/z$  410.1025. Si les critères sont restreints à la présence d'un unique atome de soufre et à un nombre de carbone compris entre 15 et 18 (en supposant une erreur de mesure de l'abondance relative des pics du massif isotopique de 10 %), seules 3 formules sont proposées. Enfin, l'utilisation de la règle de l'azote permet de ne retenir qu'une seule composition élémentaire :  $\text{C}_{17}\text{H}_{20}\text{N}_3\text{O}_7\text{S}$ . La possibilité de générer le spectre théorique pour la formule restante en regard du spectre expérimental autorise une comparaison facile et soutient la réponse obtenue. Il existe néanmoins des limitations à l'utilisation du massif isotopique. En effet, dans les milieux biologiques, la précision de la mesure de l'abondance relative des ions d'un massif isotopique peut être affectée par des ions interférents ou par l'effet matrice. De plus, pour les ions de faible intensité, la précision de la mesure décroît, voire il est impossible de détecter leur massif qui est confondu avec le bruit chimique. Ainsi, sur notre appareil, l'exploitation des abondances relatives n'est réalisable que pour les ions monoisotopiques d'une intensité supérieure à  $5 \cdot 10^5$ .

### CHAPITRE III : Travaux personnels



**Figure 58 : Détermination d'une composition élémentaire sur un système HPLC-Orbitrap®.**

(A) Spectre de masse obtenu en électrospray négatif au temps de rétention du signal *m/z* 410.1025 devant être identifié. (B) Détail du massif isotopique de l'ion d'intérêt. (C) Tableau des formules brutes proposées par le générateur de formule du logiciel Xcalibur® à partir de la mesure de masse précise du rapport *m/z* de l'ion d'intérêt avec une tolérance d'erreur de mesure de masse de 3 ppm et justification de leur exclusion.

### 5.3.2.2 Exemples d'identification

D'après la démarche d'identification proposée en introduction, plusieurs situations de complexité différentes peuvent être rencontrées selon que la formule brute trouvée est référencée ou non dans les bases de données métaboliques. Nous présenterons donc ci-après 2 cas correspondant à chacune de ces deux situations :

- **Le composé d'intérêt est présent dans une base de données biochimique ou métabolomiques.**

Ce premier exemple traite d'un signal discriminant ( $m/z$  220.1177, détection électrospray positif) mis en évidence lors de l'analyse différentielle de profils urinaires de rats avant et après exposition au PB. La composition élémentaire déduite de l'analyse du spectre de masse est  $C_9H_{17}NO_5$ . Cette dernière est utilisée pour interroger la base de données métabolomique HMDB (Human Metabolom DataBase) qui renvoie pour unique réponse : l'acide pantothénique. Le spectre de fragmentation de l'ion à identifier (Figure 59 A') montre deux pertes successives de 18, conduisant aux ions  $m/z$  202 et  $m/z$  184, et correspondant à des pertes de molécules d'eau, ainsi qu'une perte de 28 (entre les ions  $m/z$  202 et  $m/z$  174) qui est attribuable à une perte de CO et non  $C_2H_4$  grâce à la mesure de masse précise obtenue avec l'Orbitrap®. Ces pertes suggèrent la présence d'un groupement hydroxyle et d'une fonction acide carboxylique qui sont classiquement éliminés sous forme d'eau pour le premier et sous forme d'une perte successive d'une molécule d'eau et d'un groupement carboxyle pour la seconde, en fragmentation CID positif. Par ailleurs, la base HMDB fournit un spectre MS/MS de l'acide pantothénique, obtenu en mode électrospray positif sur un triple quadripôle (Figure 59 D), qui présente d'importantes similitudes avec le composé à identifier.

L'achat et l'injection d'un standard d'acide pantothénique sur notre système (Figure 59 B) permet de comparer les temps de rétention du composé inconnu avec la référence et leur spectre de fragmentation sur le même appareillage. Si les spectres de fragmentation sont parfaitement concordant (Figure 59 A' et B'), on observe un léger décalage en ce qui concerne les temps de rétention qui sont respectivement de 16.69 min et 16.52 min pour le composé inconnu et la référence. L'injection d'une urine de rat surchargée avec la référence lève cette ambiguïté (Figure 59 C) et permet de conclure à l'identification formelle du composé d'intérêt.

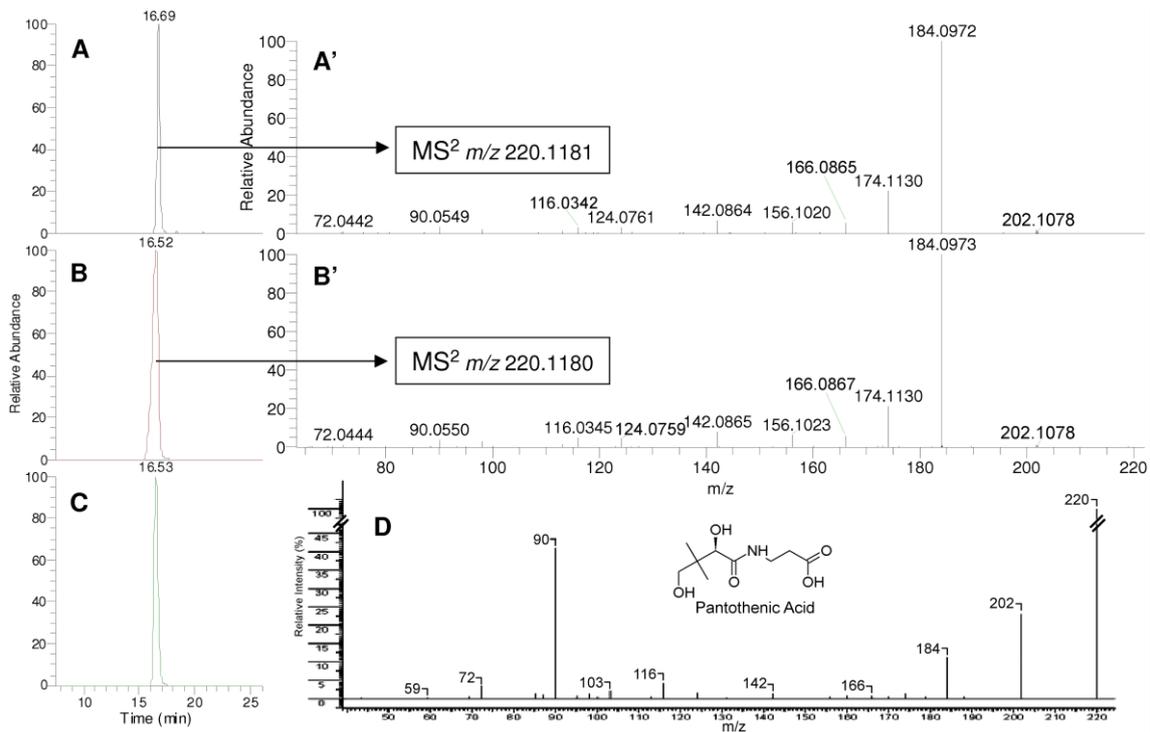


Figure 59 : Identification de l'acide pantothénique dans l'urine de rat. .

(A) Chromatogramme extrait de l'ion à identifier  $m/z$  220.1180 ( $\pm 2$  ppm) à partir d'un profil d'urine de rat obtenu sur un système HPLC-Orbitrap en électrospray positif. (A') Spectre de fragmentation CID de l'ion d'intérêt. (B) Chromatogramme extrait de l'ion  $m/z$  220.11795 ( $\pm 2$  ppm) correspondant à la masse théorique de l'acide pantothénique protoné et obtenu par injection d'un standard commercial. (B') Spectre de fragmentation CID du composé de référence. (C) Chromatogramme extrait de l'ion à identifier  $m/z$  220.1180 ( $\pm 2$  ppm) à partir d'un profil d'urine de rat surchargé avec une référence d'acide pantothénique. (D) Spectre de fragmentation de l'acide sébacique disponible sur HMDB (obtenu sur un TQ Quattro Waters®).

- **Le composé étudié est absent dans les bases de données biochimiques ou métabolomiques.**

Dans ce second cas de figure, les expériences de fragmentation ne permettent qu'une caractérisation partielle du potentiel biomarqueur. L'interprétation avancée du spectre de fragmentation ne conduit en effet qu'à une ou des structure(s) hypothétique(s) qui ne pourront être validées que par l'achat ou la synthèse des composés ainsi déterminés.

A titre d'exemple, nous présenterons l'ion  $m/z$  201.1133, détecté en LC-ESIMS mode négatif et mis en évidence au cours de l'analyse différentielle des empreintes urinaires de rats avant et après leur exposition au PB. Dès l'examen du chromatogramme extrait de l'ion avec une tolérance de masse de 2 ppm, le cas se révèle plus complexe puisque 3 pics donc 3 composés sont mis en évidence (Figure 60 A). L'analyse du massif isotopique des 3 composés a conduit à l'obtention de la même formule brute  $C_{10}H_{18}O_4$ .

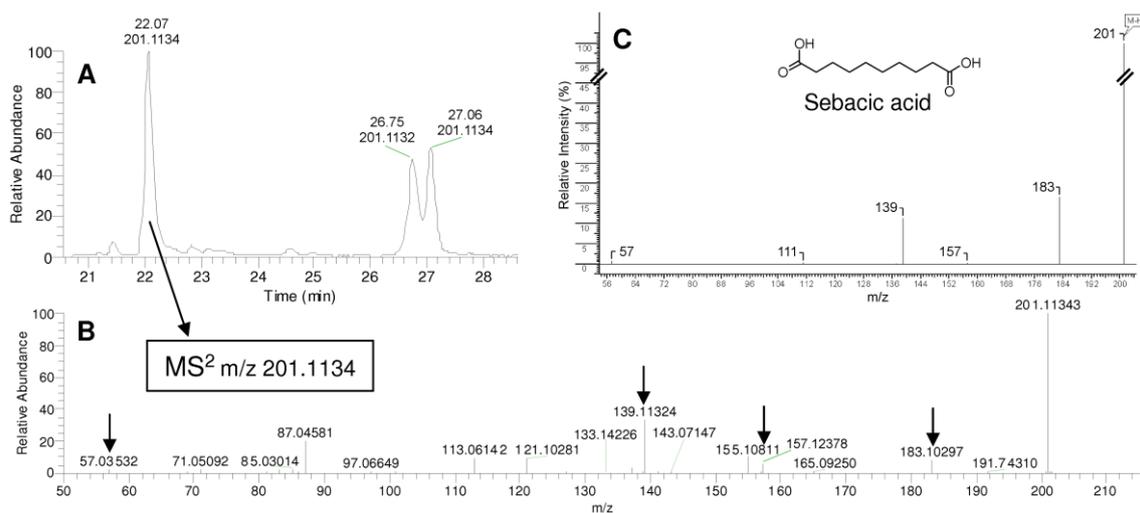


Figure 60 : Etude de la variable d'intérêt m/z 201.1133

(A) Chromatogramme extrait de l'ion m/z 201.1133 ( $\pm 2$  ppm) à partir d'un profil d'urine de rat obtenu sur un système HPLC-Orbitrap en électrospray négatif. (B) Spectre de fragmentation MS<sup>2</sup> de l'ion m/z 201.1134 à 22 min. Les flèches signalent les ions communs avec le spectre de fragmentation de l'acide sébacique présenté en (C) qui a été obtenu avec un TQ Quattro Waters® et est disponible dans la base HMDB.

Cependant, seuls les pics élués à 22 et 27 minutes ont vu leur intensité modifiée par le traitement au PB. L'interrogation des bases HMDB, Metlin et KEGG avec cette formule brute renvoie comme uniques réponses l'acide sébacique et le diméthyladipate. La base HMDB contient également le spectre de fragmentation CID de l'acide sébacique (Figure 60 C) qui présente de nombreuses similitudes avec les spectres de fragmentation obtenus pour les deux composés inconnus qui comprennent tous un ion majoritaire m/z 139 et d'autres ions fragments communs (m/z 183, 157 et 57). L'achat et l'injection de l'acide sébacique a permis d'identifier le composé élué à 27 minutes comme étant l'acide sébacique (Figure 61).

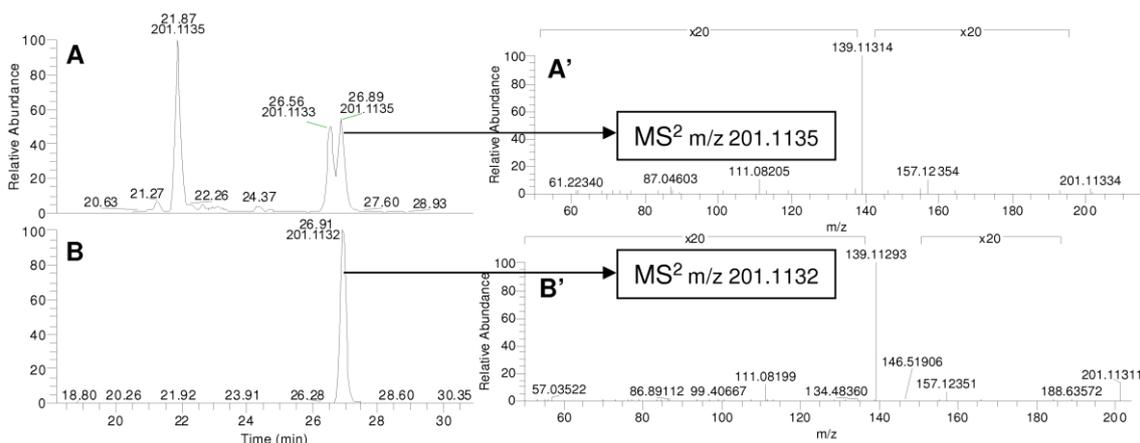


Figure 61 : Identification de l'acide sébacique dans l'urine de rat.

(A) Chromatogramme extrait de l'ion m/z 201.1133 ( $\pm 2$  ppm) à partir d'un profil d'urine de rat obtenu sur un système HPLC-Orbitrap en électrospray négatif. (B) Chromatogramme extrait de l'ion m/z 201.1133 ( $\pm 2$  ppm) à partir d'un profil d'urine de rat surchargé à 5  $\mu\text{g/mL}$  avec une référence d'acide sébacique obtenu sur le même système.

L'unique réponse restante pour notre composé inconnu est donc le diméthyladipate qui ne peut convenir au regard du spectre de fragmentation en l'absence de perte de groupement méthyle ou de méthanol. Il faut donc se tourner vers des bases de données chimiques généralistes comme la base CAS. Le principal inconvénient de ce type de base est leur richesse en molécules dont certaines sont sans pertinence biologique. Une requête sur la base CAS avec la formule  $C_{10}H_{18}O_4$  renvoie ainsi 2334 réponses. Il est alors obligatoire de conduire des expériences complémentaires afin de pouvoir restreindre les critères de recherche en imposant des fonctions qui auront été mises en évidence. Le spectre de fragmentation du composé à 22.06 min (Figure 60 B) montre une perte de 44, correspondant à  $CO_2$  et à l'origine de l'ion  $m/z$  157, et signe la présence d'une fonction carboxylique dont on peut supposer qu'elle est hydroxylée en  $\alpha$  pour expliquer en mode négatif la perte compétitive de 46 ( $(H_2CO_2)$ ). De plus, des expériences complémentaires d'échange de proton H/D ont confirmé la présence de 3 hydrogènes labiles sur la molécule d'intérêt qui orientent sa structure vers un acide carboxylique dihydroxylé comportant une insaturation plutôt que vers un acide dicarboxylique.

En interrogeant la base CAS et en restreignant les réponses avec ces informations, il ne subsiste plus qu'une réponse (Figure 62 A).

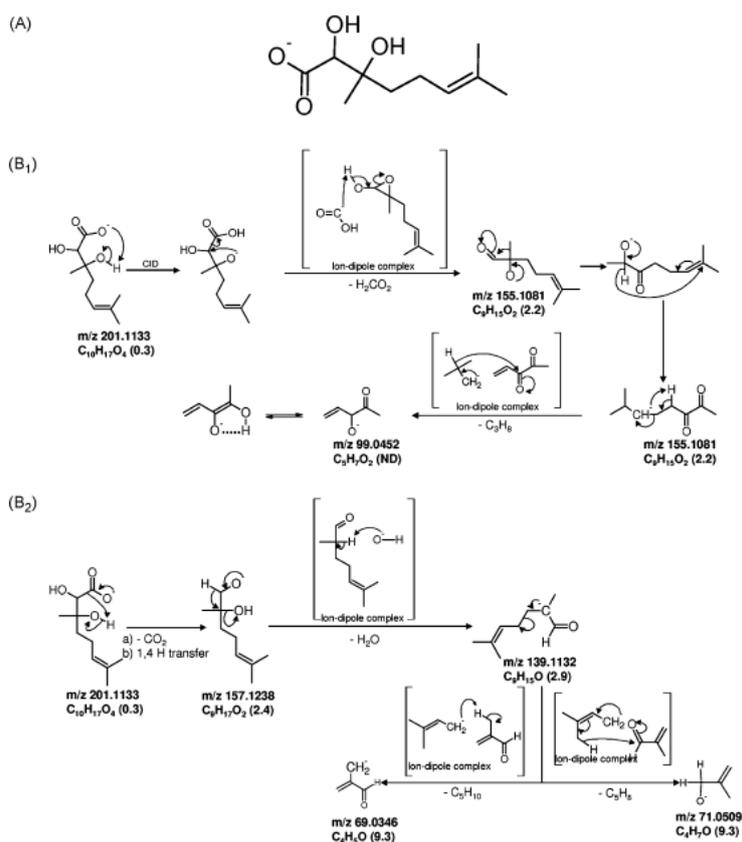
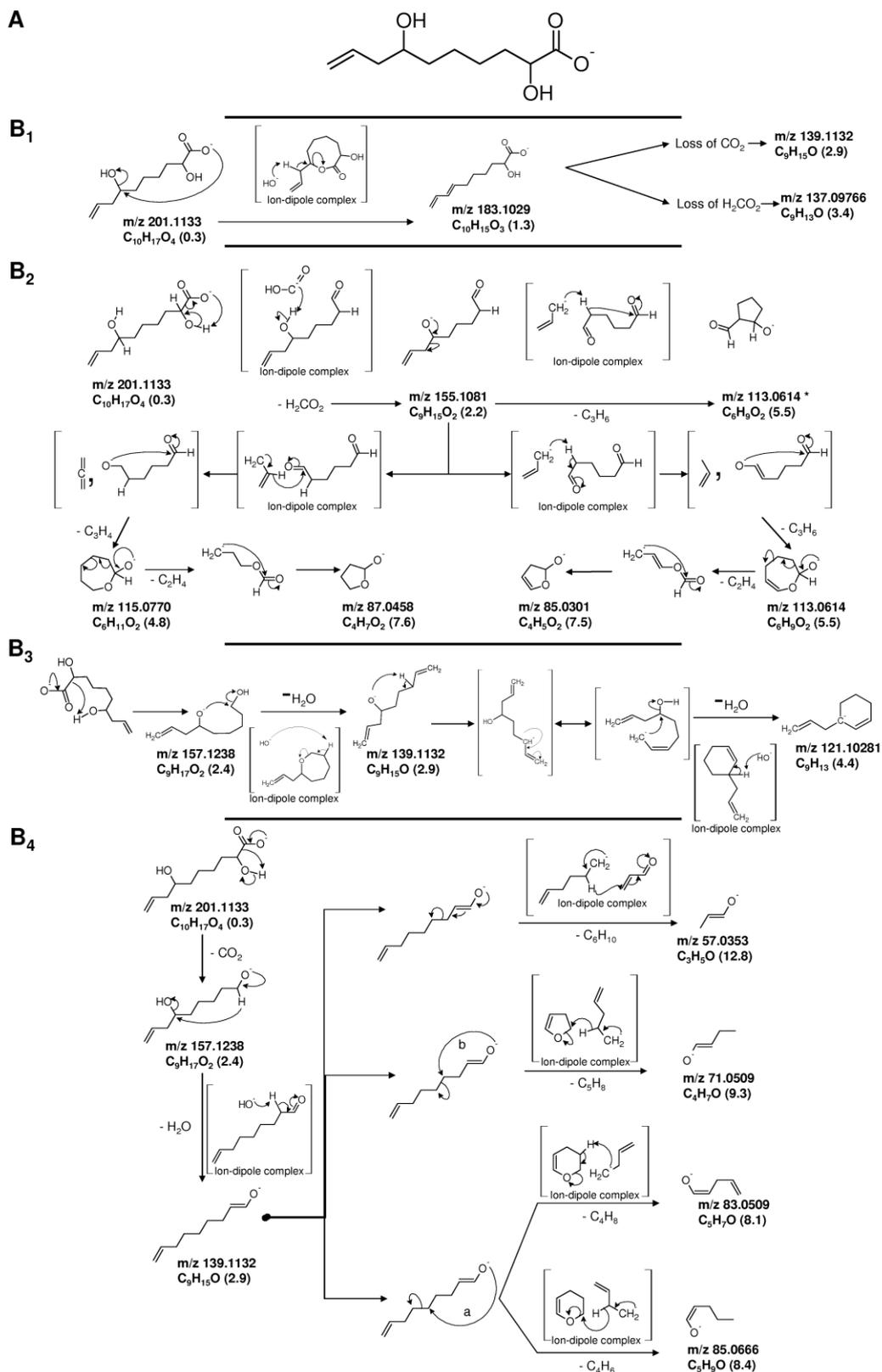


Figure 62 : (A) Structure proposée par la base CAS pour le composé inconnu à  $m/z$  201 ( $C_{10}H_{18}O_4$ ) et son mécanisme proposé de fragmentation avec une perte initiale de  $H_2CO_2$  (B1) ou  $CO_2$  (B2).

Pour chaque ion fragment sont présentés sa masse exacte, sa formule brute et l'erreur de masse correspondant à la masse précise mesurée en conditions expérimentales (ppm). ND = Non détecté

### CHAPITRE III : Travaux personnels



**Figure 63 : Proposition de structure pour le composé d'intérêt  $C_{10}H_{18}O_4$  correspondant à l'ion  $m/z$  201 élué à 22.06 minutes (A) à partir d'une proposition de mécanisme de fragmentation explicitant la formation de l'ensemble des ions détectés.**

Mécanisme de fragmentation débutant soit par la perte de  $H_2O$  (B1) soit par la perte de  $CO_2$  (B2). Pour chaque ion fragment sont présentées sa masse exacte, sa formule brute et l'erreur de masse correspondant à la masse précise mesurée en conditions expérimentales (ppm).

Malheureusement, avec ce composé il semble difficile d'expliquer l'ion  $m/z$  113 observé dans le spectre de fragmentation de la molécule inconnue. De plus, comme le suggère le mécanisme proposé dans la Figure 62, nous nous attendrions à la formation préférentielle d'ions  $m/z$  99 et 69 dont le premier n'est pas retrouvé en conditions expérimentales. En conséquence, une analyse minutieuse du spectre de fragmentation du composé inconnu a été entreprise afin de tenter de proposer une structure qui nous semble plus en adéquation avec son spectre de fragmentation. Notre tentative d'élucidation du mécanisme de fragmentation de la molécule ainsi que des ions fragments formés est présentée dans la Figure 63.

L'ion  $m/z$  113 peut être considéré comme un ion diagnostique de la position terminale de la double liaison de notre molécule. Il pourrait en effet correspondre à un clivage allylique de la molécule qui serait favorisé par la présence d'un groupement hydroxyle en position homoallylique. Ceci nous amène à proposer comme structure pour la molécule inconnue celle représentée en partie A de la Figure 63. Bien évidemment, cette structure ne pourra être confirmée qu'après la synthèse du composé proposé et son injection sur notre système ou le recours à des plateformes complémentaires comme la RMN.

### 5.3.2.3 Composés identifiés

Nous définirons 4 niveaux d'identification pour les différentes variables étudiées :

- Injection standard : pour les métabolites dont une référence commerciale ou synthétique a été injecté sur le système et donner lieu à la vérification du temps de rétention et du spectre de MS/MS.
- MS<sup>2</sup> concordante avec référence bibliographique : pour les métabolites dont l'interprétation de la fragmentation séquentielle a conduit à une hypothèse structurale pour laquelle un spectre de fragmentation a été trouvé dans la littérature ou les bases de données et nous a semblé similaire.
- Interprétation MS<sup>n</sup> : pour les composés où la proposition d'identification ne repose que sur notre interprétation des spectres de fragmentation séquentielle.
- Hypothèse à valider : pour les composés où une formule brute a été trouvée et a permis d'obtenir des hypothèses de structure mais dont l'analyse séquentielle de fragmentation se révèle non concluante ou n'a pas pu être réalisée.
- Le reste des composés est considéré comme inconnu.

Les résultats d'identification issus de l'analyse des groupes PB et 3-MC/BNF sont respectivement présentés dans les Tableaux ci-dessous.

## 5.3.2.3.1 Groupe PB

Tableau 23 : Composés identifiés au cours de l'analyse du groupe PB.

Métabolite	Niveau d'identification
Acide hippurique	Injection standard
N-acétyl leucine	Injection standard
Acide pantothénique	Injection standard
Acétylcarnitine	Injection standard
Ascorbate sulfate	Injection standard *
4-aminophénol sulfate	MS <sup>2</sup> concordante réf
Méthylimidazol acétate	MS <sup>2</sup> concordante réf
Acide férulique sulfate	MS <sup>2</sup> concordante réf
Guanidobutanoate	MS <sup>2</sup> concordante réf
Pipecolate	MS <sup>2</sup> concordante réf
Déoxycytidine	MS <sup>2</sup> concordante réf
Méthylhistidine	MS <sup>2</sup> concordante réf
Phénylacétylglycine	MS <sup>2</sup> concordante réf
Nicotinurate	MS <sup>2</sup> concordante réf
Acide indolelactique	MS <sup>2</sup> concordante réf
Diacetyl-ornithine	Interprétation MS <sup>n</sup>
Galactonate/gluconate/Gulonate *	Interprétation MS <sup>n</sup>
8-amino-7-oxo-nonanoate	Interprétation MS <sup>n</sup>
Pyrocatechol / quinol sulfate *	Interprétation MS <sup>n</sup>
Phénylglucuronide	Interprétation MS <sup>n</sup>
Acide 3-Hydroxysébacique	Interprétation MS <sup>n</sup>
Acide décènedioïque	Interprétation MS <sup>n</sup>
Méthyl-déoxycytidine	Interprétation MS <sup>n</sup>
3-hydroxydodecanedioïque	Interprétation MS <sup>n</sup>
Acétylpipecolate / diethyl maleamate	Hypothèse à valider
Coumarate sulfate	Hypothèse à valider
Dérivé d'hydroxy-oxo-glutarate	Hypothèse à valider
L-indospicine	Hypothèse à valider
N-méthyl-pyridone-carboxamide	Hypothèse à valider
Forme acétylée carnosine / anserine	Hypothèse à valider
Dihydroxyquinoline glucuronide	Hypothèse à valider
Acide Pyridylbutenoic/Méthyl-dioxy-indol-sulfate	Hypothèse à valider

\* non différenciables en MS

Parmi les composés toujours en cours de caractérisation, on trouve de nombreux conjugués sulfate ou acide glucuronique ainsi que des dérivés d'acides gras hydroxylés plus ou moins insaturés.

#### 5.3.2.3.2 Communs aux groupes traités 3-MC et BNF

Tableau 24 : Composés identifiés au cours de l'analyse des groupe 3-MC et BNF

Métabolite	Niveau d'identification
Acétyl-carnitine	Injection standard
Acide pantothénique	Injection standard
Nonanedioate	Injection standard
Acide sébacique	Injection standard
L-carnitine	MS <sup>2</sup> concordante avec réf.
Créatine	Interprétation MS <sup>n</sup>
3-OH-dodécènedioate	Interprétation MS <sup>n</sup>
N-méthyl-pyridone-carboxamide	Interprétation MS <sup>n</sup>
Acide urique	Interprétation MS <sup>n</sup>
Décènedioate	Interprétation MS <sup>n</sup>
Galactonate / gluconate / gulonate	Interprétation MS <sup>n</sup>
Vanillactate	Interprétation MS <sup>n</sup>
Atrolactate / phényllactate	Interprétation MS <sup>n</sup>

## 5.4 Intégration des résultats dans le contexte biologique

La variabilité non contrôlée des profils en partie due au design expérimental mais aussi aux progrès à accomplir pour améliorer la reproductibilité des empreintes d'un batch analytique à l'autre ne peut que nous amener à douter de la réelle pertinence biologique des variables identifiées. Cependant, parmi les composés identifiés, on retrouve l'acide hippurique, la phénylacétylglycine, divers acides chlorogéniques (dérivés de coumarine, et d'acide férulique), et un dérivé de quinoline qui sont tous reliés à l'activité métabolique de la microflore. Robosky et al. ont montré que des variations importantes de concentration urinaires en ces composés correspondaient à l'existence de colonies de rats ayant des microbiotes différents au sein même de Charles River <sup>327</sup>. Ceci pourrait expliquer certaines des variations observées entre les différents groupes d'animaux dans notre étude et particulièrement que les profils initiaux ne soient pas comparables dans tous les groupes (Figure 64).

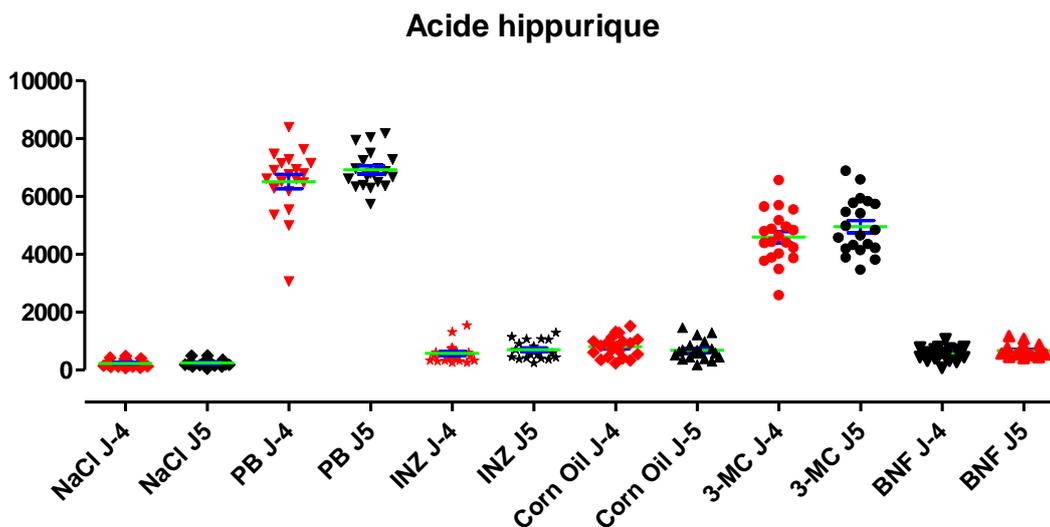


Figure 64 : Valeurs individuelles de l'acide hippurique pour l'ensemble des profils en faveur de l'existence de sous-populations d'animaux présentant des phénotypes de microbiote différents.

Par ailleurs, au cours des analyses du groupe PB, le sulfate d'acide ascorbique ainsi qu'un acide de formule ( $C_6H_{12}O_6$ ) pouvant être, entre autres, l'acide gulonique ont été mis en évidence. Une étude menée par les laboratoires Bristol Myers Squibb par approche métabolomique et transcriptomique a révélée que l'ascorbate et le gulonate étaient des marqueurs d'induction enzymatique dans les microsomes <sup>328</sup>. Enfin, les composés inducteurs utilisés présentent pour la plupart une toxicité hépatique pour laquelle a été remarqué une élévation des acides dicarboxyliques urinaires dont l'acide sébacique, l'acide subérique et, sous réserve de confirmation structurale formelle, l'acide pimélique <sup>123</sup>. Nos observations quant aux modifications de concentrations urinaires en acides dicarboxyliques dont l'acide sébacique sont donc certainement liées à la détection précoce d'une toxicité hépatique à la suite de l'administration de nos composés inducteurs.

## 5.5 Conclusion

Cette partie du travail a avant tout permis de mettre en évidence l'importance du design expérimental. Tous les animaux dont les profils sont destinés à être comparés par analyses multivariées doivent être évalués au niveau de leur profil initial afin de détecter d'éventuelles sous-populations (différence de microbiote par exemple) et doivent pouvoir être hébergés de façon simultanée. Il faut veiller à ce que les prélèvements soient réalisables dans un intervalle de temps réduit afin de minimiser l'impact du rythme circadien de certains métabolites. Par ailleurs, le design expérimental doit être construit de façon à permettre l'isolation de l'effet étudié des autres conséquences de l'administration d'un xénobiotique. Dans notre problématique d'étude de l'effet inducteur, l'administration d'une dose unique et le recueil d'échantillons à deux temps de prélèvement seulement ne permet pas d'éliminer les métabolites du composé administré et surtout n'autorise pas l'isolation de l'effet inducteur des potentiels effets pharmacologiques et toxiques des composés inducteurs. Enfin, même sur une durée courte d'une semaine, l'âge des rats semble avoir un impact sur leurs profils métaboliques et cette étude souligne donc l'importance de l'incorporation d'un groupe

contrôle en parallèle des groupes traités qui sera hébergé et prélevé dans des conditions analogues afin de suivre l'impact de l'ensemble des facteurs environnementaux sur les profils métaboliques.

L'approche SUS plot s'est révélée particulièrement intéressante puisqu'elle a permis d'isoler les métabolites propres à chacun des deux inducteurs simultanément analysés. En outre, la recherche de métabolites modifiés de façon similaire pour deux composés ayant des activités pharmacologiques différentes s'est révélée une approche efficace pour isoler l'effet inducteur.

La partie identification des composés endogènes a démontré l'intérêt de l'utilisation de spectromètres de masse donnant accès à l'ultra-haute résolution et à la possibilité de mener des expériences de fragmentation séquentielle. Elle a en outre mis en lumière la difficulté de changer d'appareillage entre l'acquisition des profils et l'étape d'identification. Elle a enfin montré l'extrême lourdeur de l'identification des composés endogènes inconnus et a permis d'inciter les deux groupes participants à initier la construction d'une base de données interne permettant d'alléger cette étape et de sauvegarder les acquis pour les études à venir.

Relativement décevante en termes de réponse apportée à l'objectif initial, cette étude restera au final un moyen de tirer des enseignements qui permettront d'améliorer la construction des futurs designs expérimentaux et les stratégies de traitement des données. La détection de métabolites discriminants reliés à l'ensemble des facteurs confondus dont l'effet inducteur témoigne cependant de l'intérêt de cette approche pour la découverte de biomarqueurs indirects d'induction enzymatique. Une nouvelle étude sera initiée mais comprendra cette fois-ci un nombre réduit de composés inducteurs donc un nombre réduits d'animaux, compatible avec les capacités d'accueil. Par ailleurs, plusieurs doses de composés inducteurs et plusieurs temps de prélèvement seront prévus de façon à tenter d'obtenir des empreintes qui permettront de séparer les effets toxiques du phénomène d'induction tout en apportant une information sur la cinétique des potentiels biomarqueurs. Enfin, l'utilisation de couples de composés inducteurs ayant le même profil d'induction comme le 3-MC et la BNF ou le PB et le TCPOBOP devrait également participer à cibler l'effet inducteur.



CHAPITRE IV :  
CONCLUSION GENERALE

La totalité du génome des espèces de toxicologie et de l'homme ayant été séquencée, ces dernières années ont vu l'émergence de nouvelles disciplines telles que la génomique fonctionnelle, la pharmacogénomique, ou encore la biologie des systèmes. La problématique s'est ainsi déplacée de la caractérisation des gènes vers leurs rôles fonctionnels et l'exploitation des différents génotypes lorsqu'ils induisent différents phénotypes pour la stratification des populations. Cette dernière vise à regrouper les individus / patients en fonctions de phénotypes qui en feront des sujets susceptibles ou non de développer une maladie, d'être de bons ou de mauvais répondeurs à un traitement, ou encore de présenter un risque accru d'interaction médicamenteuse ou de survenue d'effet toxique lors de la prise d'un médicament. Cette démarche est orientée vers la mise en place d'une médecine personnalisée où la thérapeutique s'adapte individuellement à chaque patient. Dans ce contexte, la métabolomique a connu et continue de présenter un fort potentiel attractif. En effet, puisqu'elle s'intéresse aux métabolites présents dans les différents fluides ou tissus biologiques et qu'ils sont les derniers maillons de l'expression effective des gènes, elle présente l'avantage d'explorer le statut génétique fonctionnel et la réponse globale d'un organisme à une quelconque perturbation tout en tenant compte des facteurs environnementaux, contrairement à la transcriptomique où la protéomique qui étudient des intermédiaires dans cette réponse et qui ont pour inconvénient majeur de ne pas discriminer les éléments fonctionnels de ceux qui ne le sont pas. De plus, Raamsdonk et al. ont montré chez la levure que la délétion de certains gènes ne produisait pas toujours l'effet escompté et par conséquent que l'étude du phénotype métabolique serait la plus susceptible de prédire l'impact d'une intervention thérapeutique <sup>98</sup>. Par analogie au génotype, le terme « métabotype » a ainsi été avancé pour définir le phénotype métabolique <sup>329</sup>. La caractérisation du métabotype d'individus a déjà pu être utilisée pour prédire leur susceptibilité à développer une intoxication au Paracétamol <sup>330</sup>. En revanche, la métabolomique reste une discipline jeune qui doit encore répondre à des améliorations techniques tant sur le plan analytique que dans les domaines de l'exploitation des données, de la caractérisation des métabolites et de la sauvegarde des connaissances.

L'objectif de ce travail de thèse était d'appliquer l'approche métabolomique à une problématique complexe de caractérisation de biomarqueurs indirects d'induction enzymatique dans une espèce de toxicologie. Au cours de cette étude, il s'agissait surtout de faire un état des lieux de la technique afin de mieux appréhender ses limitations actuelles et de tenter d'apporter des solutions aux problèmes rencontrés. L'objectif principal de cette thèse s'est donc naturellement orienté vers des développements méthodologiques plutôt que sur la découverte des biomarqueurs à proprement parler.

Dans un premier temps, le travail de thèse a consisté à comparer les méthodes analytiques employées au sein des deux entités partenaires dans ce projet de recherche. Bien que toutes deux basées sur la chromatographie liquide couplée à la spectrométrie de masse, ces deux méthodes reposaient sur des technologies et des partis pris très différents. Cette étape a permis de mettre en évidence l'impact important de la méthode analytique sur la partie traitement du signal pour l'obtention de jeux de données les plus informatifs et les moins biaisés possibles. Elle a confirmé la supériorité de l'UPLC comparativement à l'HPLC dans les approches employant comme technique séparative la

chromatographie liquide. Il y a deux raisons à cela : la première purement analytique étant le gain de résolution apporté par la technologie UPLC qui permet une amélioration de la séparation associée à une diminution du temps d'analyse. Ceci est à l'origine de deux conséquences favorables qui sont un gain en sensibilité et une réduction de la durée d'acquisition des profils métaboliques qui peut atteindre plusieurs jours voir plusieurs semaines en fonction du nombre d'échantillons. La seconde raison qui découle de la première est que l'UPLC, grâce à sa sensibilité améliorée, sa meilleure géométrie de pic et son excellente répétabilité en terme de temps de rétention favorise grandement l'efficacité des logiciels d'extraction des données en limitant la nécessité de réaligement des chromatogrammes ainsi qu'en améliorant le rapport signal/bruit. S'il est admis qu'aucune plateforme analytique ne permet une couverture exhaustive de la diversité chimique du métabolome, ce travail a en revanche illustré l'intérêt de recourir à plusieurs conditions d'élution chromatographiques (pH et solvant organique) afin d'améliorer la capacité de détection. Enfin, cette étape a permis d'opposer deux types de détecteur présentant chacun des caractéristiques intéressantes pour l'identification des composés détectés avec d'une part l'accès à la formule brute grâce à la mesure de masse précise du TOF et d'autre part la fragmentation séquentielle de la trappe ionique. L'idéal étant bien évidemment une combinaison des deux.

Dans un deuxième temps, nous avons tenté d'appliquer la méthode retenue (UPLC-TOF) à l'analyse d'échantillons biologiques en comparant des urines de rats avant et après exposition au phénobarbital. Si l'objectif premier de cette application était la mise en lumière de métabolites reliée à l'activité inductrice du PB, l'expérience a montré que les différences prédominantes entre les profils urinaires avant et après traitement étaient plus liées à l'exposition au composé exogène qu'à son pouvoir inducteur. Ce travail a été l'occasion de confirmer l'intérêt de l'approche métabolomique pour la caractérisation des métabolites d'un composé exogène <sup>124, 126</sup>. Cette stratégie se révèle très intéressante pour l'étude *in vivo* du métabolisme d'un médicament dans les phases précliniques précoces de son développement où la molécule radiomarquée n'est que très rarement disponible. En effet, contrairement aux approches couramment employées, cette stratégie ne fait appel ni à la modélisation a priori des biotransformations que la molécule étudiée est susceptible de subir, ni à l'utilisation des filtres de défaut de masse qui sont susceptibles d'extraire de nombreux composés endogènes et ainsi de complexifier l'identification des métabolites d'origine exogène. Au cours de cette étape, l'emploi de matrices de corrélation exploitant la corrélation qui existe au travers de l'ensemble des échantillons entre l'intensité des différents signaux, détectés en spectrométrie de masse à partir d'une même molécule parent, a permis de montrer qu'il était possible de s'affranchir d'environ 50% de la redondance du signal générée par ce type de détection. Le nombre très important de signaux discriminants, obtenus grâce aux analyses multivariées, associé à la complexité de leur identification, représente une étape limitante à l'application des approches métabolomiques de part le temps qu'il est nécessaire de lui consacrer. La possibilité de réduire le nombre de signaux discriminants en ciblant ceux correspondant à une molécule unique représente en conséquence une réelle avancée. Cette difficulté à identifier les signaux discriminants a également été l'occasion de prouver l'intérêt de l'utilisation de spectromètres de masse ultra-haute résolution avec l'emploi d'un système HPLC-Orbitrap<sup>®</sup> qui s'est révélé particulièrement performant. Ce détecteur permet non

seulement d'accéder à la mesure de masse précise, à un pouvoir de résolution élevé qui permet une exploitation avancée du massif isotopique mais encore il autorise l'accès à la fragmentation séquentielle dans sa partie trappe ionique qui ouvre la voie à l'exploitation de la filiation des ions dont on connaît l'intérêt dans le processus d'identification d'une molécule inconnue. Malheureusement, l'accès à l'ultra-haute résolution (spectromètre de masse à transformée de Fourier) se fait au prix d'un temps d'acquisition long, de l'ordre de la seconde, qui limite la possibilité de couplage de ces appareils avec l'UPLC.

Au cours d'une troisième partie, le travail a consisté à évaluer un outil de visualisation de larges jeux de données alternatif aux analyses multivariées. Il s'agissait ici d'étudier un moyen d'organiser les données reposant plus sur leurs caractéristiques structurales que sur leur pouvoir discriminant vis-à-vis de groupes d'échantillons afin de faciliter l'identification des variables et d'accélérer la construction de bases de données lorsque les données sont acquises avec des détecteurs autorisant la mesure de masses précises. A cette fin, nous avons démontré l'applicabilité de la représentation de Kendrick aux échantillons biologiques - plus particulièrement aux urines de rats - analysés sur un système HPLC-Orbitrap<sup>®</sup>. Ils se sont montrés efficaces pour le regroupement des signaux correspondant aux composés d'une même classe chimique (même composition élémentaire et même nombre d'insaturation au nombre d'unité méthylène près). Cependant, la diversité des molécules présentes dans les fluides biologiques, qui se traduit par la présence de nombreux isomères de fonction ou de position pour une même formule brute, a révélé les limites d'une telle approche, même associée à des expériences de fragmentation, destinées à discriminer les molécules par la mise en évidence de fragmentations spécifiques de certaines fonctions. Il a alors été proposé de les coupler avec l'exploitation de la sélectivité méthylène caractérisant les phases C<sub>18</sub> en chromatographie liquide à polarité de phase inverse. Cette association, utilisant la relation de linéarité qui existe entre le temps de rétention des composés d'une série d'homologues et leur nombre d'unité méthylène, a ainsi permis de regrouper les composés de structure proche possédant le même indice de Kendrick. Même si des expériences complémentaires et l'implémentation d'un script d'automatisation sont encore nécessaires, cette approche ouvre la voie à une accélération de l'identification des composés et de l'enrichissement des bases de données. Ce dernier point est particulièrement important dans le cadre d'approches métabolomiques reposant sur la chromatographie liquide couplée à la spectrométrie de masse où la fragmentation des composés est appareil dépendant et impose la construction d'une base de données collaborative pour chaque type d'appareils, voire la construction d'une base interne à chaque équipe.

Enfin, dans une quatrième partie, nous nous sommes intéressés à la problématique initiale, c'est-à-dire la mise en évidence et l'identification de biomarqueurs indirects d'induction enzymatique. Cette partie a notamment illustré l'importance de la maîtrise des conditions environnementales et du design expérimental, la complexité de l'identification de molécules endogènes, et la difficulté de replacer les composés mis en évidence dans des voies biochimiques biologiquement pertinentes avec l'effet étudié.

En conclusion, ce travail de thèse, s'il a modestement essayé de contribuer à résoudre certains freins empêchant l'exploitation du plein potentiel des approches métabolomiques a également révélé nombre d'étapes restant à franchir. Les principales limitations de l'approche sont encore :

- **L'absence d'une plateforme analytique permettant une couverture exhaustive de la diversité chimique du métabolome qui soit encore sensible, répétable, robuste et idéalement quantitative.** L'amélioration des capacités de traitement des données brutes a conduit à la multiplication des approches exploitant l'association de données issues de plusieurs plateformes analytiques pour l'analyse la plus large possible d'un même set d'échantillons. On observe aujourd'hui classiquement des approches reposant sur la combinaison de plateforme LC-ESIMS et GC-EIMS ou RMN par exemple. Il faut noter que l'EC-MS a connu un intérêt croissant <sup>331</sup> et des applications nombreuses <sup>332</sup> ces 5 dernières années pour l'analyse des molécules les plus polaires. La GCxGC-MS est également de plus en plus employée pour son exceptionnelle résolution et l'accès qu'elle offre aux bibliothèques spectrales universelles. En spectrométrie de masse couplée à la LC, les méthodes intègrent de plus en plus des contrôles de qualité permettant de garantir la stabilité du système <sup>333-335</sup>. Il a également été décrit l'utilisation de pools de l'ensemble des échantillons étudiés, injectés à différentes concentrations afin de s'assurer que les variations d'intensité de variables étaient bien corrélées avec sa concentration <sup>336</sup>. Le développement de méthodes employant des composés marqués avec des isotopes stables pourrait également permettre d'atteindre une mesure quantitative de certaines familles de métabolites dès l'acquisition des profils globaux. Enfin, plusieurs tentatives réussies de couplage de la LC à l'EIMS <sup>337</sup> ont été décrites et pourrait modifier dans un avenir proche l'étendue des techniques disponibles pour mener des approches métabolomiques par MS.

- **Le besoin d'améliorer l'extraction et le prétraitement des données.** Il s'agit d'un domaine en constante évolution pour lequel de nouveaux outils font régulièrement leur apparition. En complément de ceux précédemment cités dans la partie introductive, on citera les outils proposés aujourd'hui par l'ensemble des constructeurs de spectromètres de masse, mais qui présentent les inconvénients majeurs d'être, pour beaucoup, restreints au format propriétaire des données brutes et d'être des boîtes noires avec des algorithmes cachés. On trouve également une nouvelle famille de logiciels (nombreux et gratuits ou non) destinés à l'extraction automatique ciblée de métabolites à partir de bases de données comprenant les métabolites associés à leur masse exacte, leur temps de rétention, ou leur spectre de fragmentation et possédant des avantages et des inconvénients aussi divers que leur nombre. Enfin, certaines sociétés et communautés de chercheurs se sont spécialisées dans le développement de logiciels tiers, capables de traiter des données issues d'appareils multiples. Parmi ces derniers on citera pour la partie payante le logiciel RefinerMS<sup>®</sup> de la société GeneData qui est particulièrement abouti mais reste évolutif, et pour les gratuits la nouvelle version de MetAlign exploitant la haute résolution <sup>338</sup>, le logiciel MZmine V.2, le très évolutif XCMS dont la dernière évolution permet l'exploitation des données de fragmentations pour la recherche en base de données <sup>339</sup>, etc.... En tout état de cause le logiciel parfait n'est toujours pas disponible, même si les progrès accomplis en 5 ans sont réellement spectaculaires, l'attente étant toujours plus grande au fur et à mesure que les améliorations arrivent.

- **Le risque important d'avoir des protocoles expérimentaux qui ne sont pas adaptés aux approches métabolomiques.** Les progrès de la métabolomique et ses résultats prometteurs ont conduit à intégrer l'approche dans un nombre grandissant d'études précliniques et même cliniques. Le passage à l'homme a été à l'origine de l'introduction de nombreux facteurs confondants liés à l'impact de l'environnement des individus sur leur métabolome ainsi qu' à l'aspect multi-sites de beaucoup d'essais cliniques qui sont à l'origine de pratiques médicales et d'équipement souvent disparates. Dans ces conditions, il est nécessaire de définir des conditions standardisées de prélèvement, de stockage et de préparation des échantillons. Il est aussi important de construire le design expérimental en fonction de la question posée et on n'insistera jamais assez sur la nécessité de la présence de groupes contrôles indépendants qui permettent de suivre l'impact de facteurs tels que les effets chronobiologiques et ceux liés à l'environnement au cours de l'étude. Pour les études qui s'intéressent à des modifications fines des profils métaboliques, il est en outre recommandé d'utiliser des prélèvements multiples de façon à ce que chaque individu puisse se servir de propre témoin ou encore d'avoir recours à des designs en cross-over. Le choix des individus appartenant au groupe contrôle doit également être intimement lié à la question posée, en ce sens qu'il doit être composé d'individus homogènes avec le groupe recevant un traitement par exemple (co-médications, polyopathologies, ...) mais qu'il faille en plus que ces deux groupes soient ou non représentatifs de la population générale.

- **La difficulté d'identifier les métabolites discriminants.** Devant la multiplicité des signaux détectés lors d'une approche métabolomique globale, nulle doute que l'identification des potentiels biomarqueurs repose avant tout sur la capitalisation des efforts consentis au cours de chaque étude. Il est en effet difficilement concevable pour des contraintes de rentabilité, particulièrement dans le monde industriel, de prendre le temps d'identifier l'ensemble des signaux discriminants dans chaque étude, tant cette étape est consommatrice de temps et d'efforts. L'intérêt de la métabolomique n'est avéré que dans la mesure où les signaux discriminants sont identifiés, il faut donc ici souligner le rôle important qu'ont joué et doivent encore jouer les « assistants » afin d'étendre l'identification des molécules entrant dans la composition du métabolome. Comme nous l'avons vu à de multiples reprises, la métabolomique par approche LC-MS présente de nombreux avantages dont l'accès au composé intact grâce à des conditions d'ionisation douces, mais sa plus grande limitation est l'absence de fragmentation conservée. Il est important que soient développés des standards d'analyse, pour donner un sens à l'information de rétention, et de normalisation de l'énergie de collision appliquée, afin que des bases de données puissent être partagées, au moins pour les composés dont les standards ne sont pas disponibles. La démarche a été initiée en métabolomique pour la RMN où un système de stockage des données spectrales correspondant aux différents standards est accessible à tous, pour enrichissement et consultation, sous un format non propriétaire

- **La difficulté d'associer les modifications observées dans les profils métaboliques avec des hypothèses mécanistiques qui permettraient de valider la pertinence biologique de l'utilisation d'un métabolite comme biomarqueur.** Lors d'une étude, des métabolites discriminants entre les groupes analysés peuvent être mis en évidence. Cependant, il est souvent difficile de relier ces métabolites entre eux au sein des voies biochimiques et de trouver une liaison avec l'effet étudié. Ceci tient premièrement à notre trop faible connaissance des interrelations métaboliques mais aussi au fait qu'un métabolite peut être engagé dans plusieurs voies biochimiques. Or, dans ce cas, il est courant que la sensibilité des méthodes ne donne pas accès à d'autres composés qui permettraient de discriminer la voie impliquée. De plus, et ce paramètre est parfois oublié, toutes les voies biochimiques n'existent pas dans l'ensemble des organes et une variation au niveau de l'organe n'est pas forcément perceptible au niveau plasmatique. Les outils informatiques d'exploration de voies biochimiques tels que Ingenuity, ou MetaCore, ont initialement été développés pour la transcriptomique, et s'ils fournissent de premiers éléments, ils ne sont pas encore en mesure d'effectuer cette discrimination organe / biofluide. A ce niveau intervient aussi la notion de temps : quel est le temps nécessaire à la traduction plasmatique ou urinaire d'un processus localisé dans un organe ? Cette notion est également à placer au premier plan lorsque l'on cherche à corréler des données de métabolomique avec celles de protéomique ou de transcriptomique dont les échelles de temps sont bien souvent très différentes. Tous ces processus dynamiques sont pris en compte avec l'émergence des nouvelles approches comme la fluxomique et les projets de modélisation d'organe *in silico*, comme c'est le cas pour le cœur <sup>341</sup> par exemple, pour aboutir peut-être un jour à un modèle *in silico* d'organisme complet.

En définitive, la multiplicité des domaines impliqués (chimie analytique, traitement du signal, statistiques, mathématiques, élucidation structurales, interprétation biologique (biochimie, pharmacocinétique, pharmacologie) et informatique pour le stockage des données et la construction de bases de données fonctionnelles) ne peut qu'encourager la constitution d'équipes pluridisciplinaires dont seul le travail concerté, associé à la capitalisation des expériences, pourra aboutir à l'obtention d'une technique mature. Les progrès rapides dans l'ensemble de ces disciplines et la multiplications des efforts collaboratifs devraient permettre dans les années qui viennent de faire exploser le potentiel de la métabolomique dont les applications en biologie des systèmes appliquée au domaine médical sont d'ores et déjà toutes « imaginées » <sup>342</sup>.

En ce qui concerne l'étude de l'induction enzymatique, ce travail et les recherches bibliographiques qui l'ont accompagné ont mis en lumière la complexité des mécanismes biologiques mis en œuvre au cours de ce processus et la nécessité d'en améliorer notre compréhension afin d'être en mesure de trouver des design expérimentaux qui soient en mesure de répondre à la question posée. De nombreuses pistes sont envisageables pour une meilleure caractérisation du statut enzymatique et de son évolution pour les individus dont on compare les profils métaboliques, afin d'avoir des informations précises dont on puisse se servir pour tenter d'isoler des métabolites d'intérêt. A titre d'exemple, le développement de nouvelles « chips » de transcriptomique regroupant l'ensemble des CYPs, des transporteurs, des récepteurs nucléaires et des enzymes de conjugaison

## CHAPITRE IV : Conclusion générale

pourrait être un pas en avant même si la différence de dynamique entre le transcriptome et le métabolome ne garantit pas la réussite de l'établissement de corrélation entre ces deux approches. Par ailleurs, des publications récentes rapportent le développement de méthodes de quantification en spectrométrie de masse pour les cytochromes  $P_{450}$  ainsi que pour les transporteurs dont on pourrait suivre l'évolution sans toutefois avoir accès à leur activité. Enfin, ce n'est qu'en multipliant les approches et en obtenant une caractérisation la plus exhaustive possible des modèles employés dans les études qu'il pourra être possible de comprendre l'impact des processus les plus complexes.

# RÉFÉRENCES BIBLIOGRAPHIQUES

## Références bibliographiques

1. **Kola I. and J.Landis**, "Can the pharmaceutical industry reduce attrition rates?," *Nat.Rev.Drug Discov.* 3 (8): 711-715 (2004).
2. **Jakoby W.B.** 1994. Detoxication: Conjugation and hydrolysis. In *The Liver: Biology and Pathobiology*. Edited by I. M. Arias et al. New York, NY: Raven Press.
3. **Silverman J.A.**, "Multidrug-resistance transporters," *Pharm.Biotechnol.* 12: 353-386 (1999).
4. **Stieger B. and P.J.Meier**, "Bile acid and xenobiotic transporters in liver," *Curr.Opin.Cell Biol.* 10 (4): 462-467 (1998).
5. **Anandatheerthavarada H.K. et al.**, "Localization of multiple forms of inducible cytochromes P450 in rat liver mitochondria: immunological characteristics and patterns of xenobiotic substrate metabolism," *Arch.Biochem.Biophys.* 339 (1): 136-150 (1997).
6. **de Waziers I. et al.**, "Cytochrome P 450 isoenzymes, epoxide hydrolase and glutathione transferases in rat and human hepatic and extrahepatic tissues," *J.Pharmacol.Exp.Ther.* 253 (1): 387-394 (1990).
7. **Tredaniel J., G.Zalcman, and E.Douriez**, "[Genes and enzymes involved in the metabolism of carcinogens]," *Bull.Cancer* 82 Suppl 2: 77s-84s (1995).
8. **Hietanen E., K.HusgafvelPursiainen, and H.Vainio**, "Interaction between dose and susceptibility to environmental cancer: A short review," *Environmental Health Perspectives* 105: 749-754 (1997).
9. **Mugford C.A. and G.L.Kedderis**, "Sex-dependent metabolism of xenobiotics," *Drug Metab Rev.* 30 (3): 441-498 (1998).
10. **Shimada T., R.L.Mernaugh, and F.P.Guengerich**, "Interactions of mammalian cytochrome P450, NADPH-cytochrome P450 reductase, and cytochrome b(5) enzymes," *Arch.Biochem.Biophys.* 435 (1): 207-216 (2005).
11. **Nelson D.R. et al.**, "P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature," *Pharmacogenetics* 6 (1): 1-42 (1996).
12. **Nebert D.W. and D.W.Russell**, "Clinical importance of the cytochromes P450," *Lancet* 360 (9340): 1155-1162 (2002).
13. **Evans W.E. and M.V.Relling**, "Pharmacogenomics: translating functional genomics into rational therapeutics," *Science* 286 (5439): 487-491 (1999).
14. **Dansette P.M. et al.**, "Drug-induced immunotoxicity," *Eur.J.Drug Metab Pharmacokinet.* 23 (4): 443-451 (1998).
15. **Garavito R.M. and D.L.DeWitt**, "The cyclooxygenase isoforms: structural insights into the conversion of arachidonic acid to prostaglandins," *Biochim.Biophys.Acta* 1441 (2-3): 278-287 (1999).
16. **Simmons D.L. et al.** 1991. Multiple cyclooxygenases : cloning of a mitogen-inducible form. In *Prostaglandins, Leukotriens, Lipoxins and PAF*. Edited by J. M. Bradley. New-York: Plenum Press.
17. **Harris R.C. et al.**, "Cyclooxygenase-2 is associated with the macula densa of rat kidney and increases with salt restriction," *J.Clin.Invest* 94 (6): 2504-2510 (1994).
18. **Breder C.D., D.Dewitt, and R.P.Kraig**, "Characterization of inducible cyclooxygenase in rat brain," *J.Comp Neurol.* 355 (2): 296-315 (1995).
19. **Brash A.R.**, "Lipoxygenases: occurrence, functions, catalysis, and acquisition of substrate," *J.Biol.Chem.* 274 (34): 23679-23682 (1999).
20. **Degen G.H.**, "Prostaglandin-H synthase containing cell lines as tools for studying metabolism and toxicity of xenobiotics," *Toxicology* 82 (1-3): 243-256 (1993).
21. **Marnett L.J., A.Panthanickal, and G.A.Reed**, "Metabolic activation of 7,8-dihydroxy-7,8-dihydrobenzo[a]-pyrene during prostaglandin biosynthesis," *Drug Metab Rev.* 13 (2): 235-247 (1982).

## Références bibliographiques

22. **Potter D.W. and J.A.Hinson**, "The 1- and 2-electron oxidation of acetaminophen catalyzed by prostaglandin H synthase," *J.Biol.Chem.* 262 (3): 974-980 (1987).
23. **Degen G.H., T.E.Eling, and J.A.McLachlan**, "Oxidative metabolism of diethylstilbestrol by prostaglandin synthetase," *Cancer Res.* 42 (3): 919-923 (1982).
24. **Schlosser M.J., R.D.Shurina, and G.F.Kalf**, "Prostaglandin H synthase catalyzed oxidation of hydroquinone to a sulfhydryl-binding and DNA-damaging metabolite," *Chem.Res.Toxicol.* 3 (4): 333-339 (1990).
25. **Kulkarni A.P. and M.Sajan**, "Lipoxygenase-another pathway for glutathione conjugation of xenobiotics: A study with human term placental lipoxygenase and ethacrynic acid," *Arch.Biochem.Biophys.* 371 (2): 220-227 (1999).
26. **Tukey R.H. and C.P.Strassburg**, "Human UDP-glucuronosyltransferases: metabolism, expression, and disease," *Annu.Rev.Pharmacol.Toxicol.* 40: 581-616 (2000).
27. **Mulder G.J., M.W.H.Coughtrie, and B.Burchell**. 1990. Glucuronidation. In *Conjugation Reactions in Drug Metabolism*. Edited by G. J. Mulder. London: Taylor and Francis.
28. **Kaniwa N. et al.**, "Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686C> T (P229L) found in an African-American," *Drug Metab Dispos.* 33 (3): 458-465 (2005).
29. **Hilz H. and F.Lipmann**, "THE ENZYMATIC ACTIVATION OF SULFATE," *Proc.Natl.Acad.Sci.U.S A* 41 (11): 880-890 (1955).
30. **Negishi M. et al.**, "Structure and function of sulfotransferases," *Arch.Biochem.Biophys.* 390 (2): 149-157 (2001).
31. **Hobkirk R.**, "Steroid sulfation Current concepts," *Trends Endocrinol.Metab* 4 (2): 69-74 (1993).
32. **Strott C.A.**, "Steroid sulfotransferases," *Endocr.Rev.* 17 (6): 670-697 (1996).
33. **de Chaisemartin L. and M.A.Loriot**, "Pharmacogénétique des anticancéreux," *Pathol.Biol.(Paris)* 53 (2): 116-124 (2005).
34. **Weber W.W., G.N.Levy, and D.W.Hein**. 1990. Acetylation. In *Conjugation Reactions in Drug Metabolism*. Edited by G. J. Mulder. London: Taylor and Francis.
35. **Hein D.W. et al.**, "Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms," *Cancer Epidemiol.Biomarkers Prev.* 9 (1): 29-42 (2000).
36. **Garcia-Martin E.**, "Interethnic and intraethnic variability of NAT2 single nucleotide polymorphisms," *Curr.Drug Metab* 9 (6): 487-497 (2008).
37. **Hein D.W. et al.**, "Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases," *Carcinogenesis* 14 (8): 1633-1638 (1993).
38. **Hein D.W.**, "N-Acetyltransferase genetics and their role in predisposition to aromatic and heterocyclic amine-induced carcinogenesis," *Toxicol.Lett.* 112-113: 349-356 (2000).
39. **Hanna P.E.**, "N-acetyltransferases, O-acetyltransferases, and N,O-acetyltransferases: enzymology and bioactivation," *Adv.Pharmacol.* 27: 401-430 (1994).
40. **Otieno M.A. et al.**, "Immunolocalization of microsomal glutathione S-transferase in rat tissues," *Drug Metab Dispos.* 25 (1): 12-20 (1997).
41. **Miller M.S. et al.**, "Genetic polymorphisms in human drug metabolic enzymes," *Fundam.Appl.Toxicol.* 40 (1): 1-14 (1997).
42. **Ketterer B. et al.** 1990. GSTs and protection against oxidative stress. In *Glutathione-S-transferases and drug resistance*. Edited by J. D. Hayes, C. B. Pickett, and T. J. Mantle. London: Taylor and Francis.
43. **Seidegard J. and G.Ekstrom**, "The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics," *Environ.Health Perspect.* 105 Suppl 4: 791-799 (1997).

## Références bibliographiques

44. **Monks T.J. et al.**, "Glutathione conjugate mediated toxicities," *Toxicol.Appl.Pharmacol.* 106 (1): 1-19 (1990).
45. **Chang X.B.**, "A molecular understanding of ATP-dependent solute transport by multidrug resistance-associated protein MRP1," *Cancer Metastasis Rev.* 26 (1): 15-37 (2007).
46. **Dean M. and R.Allikmets**, "Complete characterization of the human ABC gene family," *J.Bioenerg.Biomembr.* 33 (6): 475-479 (2001).
47. **Klein I., B.Sarkadi, and A.Varadi**, "An inventory of the human ABC proteins," *Biochim.Biophys.Acta* 1461 (2): 237-262 (1999).
48. **Hung L.W. et al.**, "Crystal structure of the ATP-binding subunit of an ABC transporter," *Nature* 396 (6712): 703-707 (1998).
49. **Walker J.E. et al.**, "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold," *EMBO J.* 1 (8): 945-951 (1982).
50. **Sharom F.J. et al.**, "Insights into the structure and substrate interactions of the P-glycoprotein multidrug transporter from spectroscopic studies," *Biochim.Biophys.Acta* 1461 (2): 327-345 (1999).
51. **Leslie E.M., R.G.Deeley, and S.P.Cole**, "Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense," *Toxicol.Appl.Pharmacol.* 204 (3): 216-237 (2005).
52. **Loscher W. and H.Potschka**, "Drug resistance in brain diseases and the role of drug efflux transporters," *Nat.Rev.Neurosci.* 6 (8): 591-602 (2005).
53. **Borst P., N.Zelcer, and H.A.van**, "ABC transporters in lipid transport," *Biochim.Biophys.Acta* 1486 (1): 128-144 (2000).
54. **Schmitz G. and W.E.Kaminski**, "ABC transporters and cholesterol metabolism," *Front Biosci.* 6: D505-D514 (2001).
55. **Loscher W. and H.Potschka**, "Role of drug efflux transporters in the brain for drug disposition and treatment of brain diseases," *Prog.Neurobiol.* 76 (1): 22-76 (2005).
56. **de Lange E.C.**, "Potential role of ABC transporters as a detoxification system at the blood-CSF barrier," *Adv Drug Deliv.Rev.* 56 (12): 1793-1809 (2004).
57. **Petzinger E. and J.Geyer**, "Drug transporters in pharmacokinetics," *Naunyn Schmiedebergs Arch.Pharmacol.* 372 (6): 465-475 (2006).
58. **Whitfield G.K. et al.**, "Steroid hormone receptors: evolution, ligands, and molecular basis of biologic function," *J.Cell Biochem. Suppl* 32-33: 110-122 (1999).
59. **Bourguet W., P.Germain, and H.Gronemeyer**, "Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications," *Trends Pharmacol.Sci.* 21 (10): 381-388 (2000).
60. **Handschin C. and U.A.Meyer**, "Induction of drug metabolism: the role of nuclear receptors," *Pharmacol.Rev.* 55 (4): 649-673 (2003).
61. **Krishna D.R. and U.Klotz**, "Extrahepatic metabolism of drugs in humans," *Clin.Pharmacokinet.* 26 (2): 144-160 (1994).
62. **Borst P. et al.**, "A family of drug transporters: the multidrug resistance-associated proteins," *J.Natl.Cancer Inst.* 92 (16): 1295-1302 (2000).
63. **Litman T. et al.**, "From MDR to MXR: new understanding of multidrug resistance systems, their properties and clinical significance," *Cell Mol.Life Sci.* 58 (7): 931-959 (2001).
64. **Kretschmer X.C. and W.S.Baldwin**, "CAR and PXR: xenosensors of endocrine disrupters?," *Chem.Biol.Interact.* 155 (3): 111-128 (2005).

## Références bibliographiques

65. **Baes M. et al.**, "A new orphan member of the nuclear hormone receptor superfamily that interacts with a subset of retinoic acid response elements," *Mol.Cell Biol.* 14 (3): 1544-1552 (1994).
66. **Choi H.S. et al.**, "Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR," *J.Biol.Chem.* 272 (38): 23565-23571 (1997).
67. **Maglich J.M. et al.**, "Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification," *Mol.Pharmacol.* 62 (3): 638-646 (2002).
68. **Gonzalez F.J.**, "The 2006 Bernard B. Brodie Award Lecture. Cyp2e1," *Drug Metab Dispos.* 35 (1): 1-8 (2007).
69. **Ingelman-Sundberg M., M.Oscarson, and R.A.McLellan.** "Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment," *Trends Pharmacol.Sci.* 20 (8): 342-349 (1999).
70. **Daly A.K.** 1999. Pharmacogenetics. In *Handbook of Drug Metabolism*. Edited by T. F. Wolf. New York, NY: Marcel Dekker Inc.
71. **Lin J.H. and A.Y.Lu.** "Inhibition and induction of cytochrome P450 and the clinical implications," *Clin.Pharmacokinet.* 35 (5): 361-390 (1998).
72. **Hammerlein A., H.Derendorf, and D.T.Lowenthal.** "Pharmacokinetic and pharmacodynamic changes in the elderly. Clinical implications," *Clin.Pharmacokinet.* 35 (1): 49-64 (1998).
73. **Beierle I., B.Meibohm, and H.Derendorf.** "Gender differences in pharmacokinetics and pharmacodynamics," *Int.J.Clin.Pharmacol.Ther.* 37 (11): 529-547 (1999).
74. **Zevin S. and N.L.Benowitz.** "Drug interactions with tobacco smoking. An update," *Clin.Pharmacokinet.* 36 (6): 425-438 (1999).
75. **Johnson M.D., G.Newkirk, and J.R.White, Jr.**, "Clinically significant drug interactions," *Postgrad.Med.* 105 (2): 193-196 (1999).
76. **Ameer B. and R.A.Weintraub.** "Drug interactions with grapefruit juice," *Clin.Pharmacokinet.* 33 (2): 103-121 (1997).
77. **Fleisher D. et al.**, "Drug, meal and formulation interactions influencing drug absorption after oral administration. Clinical implications," *Clin.Pharmacokinet.* 36 (3): 233-254 (1999).
78. **Lemaire G. et al.**, "Discovery of a highly active ligand of human pregnane x receptor: a case study from pharmacophore modeling and virtual screening to "in vivo" biological activity," *Mol.Pharmacol.* 72 (3): 572-581 (2007).
79. **Handschin C., M.Podvinec, and U.A.Meyer.** "In silico approaches, and in vitro and in vivo experiments to predict induction of drug metabolism," *Drug News Perspect.* 16 (7): 423-434 (2003).
80. **Baranczewski P. et al.**, "Introduction to early in vitro identification of metabolites of new chemical entities in drug discovery and development," *Pharmacol.Rep.* 58 (3): 341-352 (2006).
81. **Crespi C.L.**, "Expression of cytochrome P450 cDNAs in human B lymphoblastoid cells: applications to toxicology and metabolite analysis," *Methods Enzymol.* 206: 123-129 (1991).
82. **Ekins S. et al.**, "Present and future in vitro approaches for drug metabolism," *J.Pharmacol.Toxicol.Methods* 44 (1): 313-324 (2000).
83. **Ferrero J.L. and K.Brendel.** "Liver slices as a model in drug metabolism," *Adv.Pharmacol.* 43: 131-169 (1997).
84. **McKinnon R.A. and D.W.Nebert.** "Cytochrome P450 knockout mice: new toxicological models," *Clin.Exp.Pharmacol.Physiol* 25 (10): 783-787 (1998).
85. **Friedberg T.** 1999. *In vivo* and *in vitro* recombinant DNA technology as a powerful tool in drug development. In *Handbook of Drug Metabolism*. Edited by T. F. Woolf. New York: Marcel Dekker Inc.
86. **Davit B. et al.**, "FDA evaluations using in vitro metabolism to predict and interpret in vivo metabolic drug-drug interactions: impact on labeling," *J.Clin.Pharmacol.* 39 (9): 899-910 (1999).

## Références bibliographiques

87. **Crespi C.L. and B.W.Penman**, "Use of cDNA-expressed human cytochrome P450 enzymes to study potential drug-drug interactions," *Adv.Pharmacol.* 43: 171-188 (1997).
88. **Dilger K., U.Hofmann, and U.Klotz**, "Enzyme induction in the elderly: effect of rifampin on the pharmacokinetics and pharmacodynamics of propafenone," *Clin.Pharmacol.Ther.* 67 (5): 512-520 (2000).
89. **Seeff L.B. et al.**, "Acetaminophen hepatotoxicity in alcoholics. A therapeutic misadventure," *Ann.Intern.Med.* 104 (3): 399-404 (1986).
90. **Sumner L.W., P.Mendes, and R.A.Dixon**, "Plant metabolomics: large-scale phytochemistry in the functional genomics era," *Phytochemistry* 62 (6): 817-836 (2003).
91. **Oliver S.G. et al.**, "Systematic functional analysis of the yeast genome," *Trends Biotechnol.* 16 (9): 373-378 (1998).
92. **Tweeddale H., L.Notley-McRobb, and T.Ferenci**, "Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis," *J Bacteriol.* 180 (19): 5109-5116 (1998).
93. **Dettmer K. and B.D.Hammock**, "Metabolomics--a new exciting field within the "omics" sciences," *Environ.Health Perspect.* 112 (7): A396-A397 (2004).
94. **Forster J. et al.**, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network," *Genome Res.* 13 (2): 244-253 (2003).
95. **Kell D.B.**, "Systems biology, metabolic modelling and metabolomics in drug discovery and development," *Drug Discov.Today* 11 (23-24): 1085-1092 (2006).
96. **van der Greef J., P.Stroobant, and R.van der Heijden**, "The role of analytical sciences in medical systems biology," *Curr.Opin.Chem.Biol.* 8 (5): 559-565 (2004).
97. **Keun H.C. et al.**, "Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles," *Chem.Res.Toxicol.* 17 (5): 579-587 (2004).
98. **Raamsdonk L.M. et al.**, "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations," *Nat Biotechnol.* 19 (1): 45-50 (2001).
99. **Kacser H. and J.A.Burns**, "The control of flux," *Symp.Soc.Exp.Biol.* 27: 65-104 (1973).
100. **Cascante M. et al.**, "Metabolic control analysis in drug discovery and disease," *Nat.Biotechnol.* 20 (3): 243-249 (2002).
101. **Kell D.B. and H.V.Westerhoff**, "Metabolic Control-Theory - Its Role in Microbiology and Biotechnology," *Fems Microbiology Reviews* 39 (4): 305-320 (1986).
102. **Fiehn O.**, "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks," *Comparative and Functional Genomics* 2 (3): 155-168 (2001).
103. **Nicholson J.K. et al.**, "Metabonomics: a platform for studying drug toxicity and gene function," *Nat.Rev.Drug Discov.* 1 (2): 153-161 (2002).
104. **Nicholson J.K., J.C.Lindon, and E.Holmes**, "Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica* 29 (11): 1181-1189 (1999).
105. **Whitfield P.D., A.J.German, and P.J.Noble**, "Metabolomics: an emerging post-genomic tool for nutrition," *Br.J.Nutr.* 92 (4): 549-555 (2004).
106. **German J.B., S.M.Watkins, and L.B.Fay**, "Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health," *J.Am.Diet.Assoc.* 105 (9): 1425-1432 (2005).
107. **Wu J.Y. et al.**, "ENU mutagenesis identifies mice with mitochondrial branched-chain aminotransferase deficiency resembling human maple syrup urine disease," *J Clin Invest* 113 (3): 434-440 (2004).

## Références bibliographiques

108. **Hirai M.Y. 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* 101 (27): 10205-10210 (2004).
109. **Soga T. et al.**, "Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption," *J Biol.Chem* 281 (24): 16768-16776 (2006).
110. **Saito N. et al.**, "Metabolomics approach for enzyme discovery," *J.Proteome Res.* 5 (8): 1979-1987 (2006).
111. **Lindon J.C. et al.**, "Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis," *Biomarkers* 9 (1): 1-31 (2004).
112. **Chace D.H.**, "Mass spectrometry in the clinical laboratory," *Chem.Rev.* 101 (2): 445-477 (2001).
113. **Griffin J.L. et al.**, "Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy," *FEBS Lett* 530 (1-3): 109-116 (2002).
114. **Griffin J.L. and J.P.Shockcor**, "Metabolic profiles of cancer cells," *Nat Rev Cancer* 4 (7): 551-561 (2004).
115. **Odunsi K. et al.**, "Detection of epithelial ovarian cancer using <sup>1</sup>H-NMR-based metabonomics," *Int.J.Cancer* 113 (5): 782-788 (2005).
116. **Brindle J.T. et al.**, "Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using <sup>1</sup>H-NMR-based metabonomics," *Nat Med* 8 (12): 1439-1444 (2002).
117. **Bailey N.J. et al.**, "Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via <sup>1</sup>H NMR spectroscopy and chemometrics," *Phytochemistry* 62 (6): 851-858 (2003).
118. **Dumas M.E. et al.**, "Selection of biomarkers by a multivariate statistical processing of composite metabonomic data sets using multiple factor analysis," *J.Proteome Res.* 4 (5): 1485-1492 (2005).
119. **Prabakaran S. et al.**, "Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress," *Mol.Psychiatry* 9 (7): 684-697, 643 (2004).
120. **Nicholls A.W. et al.**, "Metabonomic investigations into hydrazine toxicity in the rat," *Chem.Res.Toxicol.* 14 (8): 975-987 (2001).
121. **Plumb R.S. 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.* 16 (20): 1991-1996 (2002).
122. **Kroeger M.**, "How omics technologies can contribute to the '3R' principles by introducing new strategies in animal testing," *Trends Biotechnol.* 24 (8): 343-346 (2006).
123. **Mortishire-Smith R.J. et al.**, "Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity," *Chem.Res.Toxicol.* 17 (2): 165-173 (2004).
124. **Plumb R.S. 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.* 17 (23): 2632-2638 (2003).
125. **Castro-Perez J. et al.**, "Increasing throughput and information content for in vitro drug metabolism experiments using ultra-performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer," *Rapid Commun.Mass Spectrom.* 19 (6): 843-848 (2005).
126. **Giri S. et al.**, "The metabolomics of (+/-)-arecoline 1-oxide in the mouse and its formation by human flavin-containing monooxygenases," *Biochem.Pharmacol.* 73 (4): 561-573 (2007).
127. **Peterman S.M. et al.**, "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.* 17 (3): 363-375 (2006).

## Références bibliographiques

128. **Werner E. et al.**, "Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution," *Anal.Chem.* 80 (13): 4918-4932 (2008).
129. **German J.B., M.A.Roberts, and S.M.Watkins**, "Genomics and metabolomics as markers for the interaction of diet and health: lessons from lipids," *J.Nutr.* 133 (6 Suppl 1): 2078S-2083S (2003).
130. **German J.B., M.A.Roberts, and S.M.Watkins**, "Personal metabolomics as a next generation nutritional assessment," *J.Nutr.* 133 (12): 4260-4266 (2003).
131. **Rezzi S. et al.**, "Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals," *J.Proteome Res.* 6 (11): 4469-4477 (2007).
132. **Rezzi S. et al.**, "Nutritional metabonomics: applications and perspectives," *J.Proteome Res.* 6 (2): 513-525 (2007).
133. **Zeisel S.H. et al.**, "The nutritional phenotype in the age of metabolomics," *J.Nutr.* 135 (7): 1613-1616 (2005).
134. **Clayton T.A. et al.**, "Pharmaco-metabonomic phenotyping and personalized drug treatment," *Nature* 440 (7087): 1073-1077 (2006).
135. **Beecher C.W.W. 2003.** The human metabolome. In *Metabolic Profiling : Its Role In Biomarker Discovery And Gene Function Analysis*. Edited by G. G. Harrigan and R. Goodacre. Norwell: Kluwer Academic Publishers.
136. **Weckwerth W.**, "Metabolomics in systems biology," *Annu.Rev.Plant Biol.* 54: 669-689 (2003).
137. **Vigneau-Callahan K.E. et al.**, "Characterization of diet-dependent metabolic serotypes: analytical and biological variability issues in rats," *J.Nutr.* 131 (3): 924S-932S (2001).
138. **Fraser P.D. et al.**, "Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids," *Plant J* 24 (4): 551-558 (2000).
139. **Kristal B.S., K.E.Vigneau-Callahan, and W.R.Matson**, "Simultaneous analysis of the majority of low-molecular-weight, redox-active compounds from mitochondria," *Anal.Biochem.* 263 (1): 18-25 (1998).
140. **Gamache P.H. et al.**, "Metabolomic applications of electrochemistry/mass spectrometry," *J.Am.Soc.Mass Spectrom.* 15 (12): 1717-1726 (2004).
141. **Dumas M.E. et al.**, "Analyzing the physiological signature of anabolic steroids in cattle urine using pyrolysis/metastable atom bombardment mass spectrometry and pattern recognition," *Anal.Chem.* 74 (20): 5393-5404 (2002).
142. **Soga T. et al.**, "Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry," *Anal.Chem.* 74 (10): 2233-2239 (2002).
143. **Garcia-Perez I. et al.**, "Metabolic fingerprinting with capillary electrophoresis," *J.Chromatogr.A* (2008).
144. **Dettmer K., P.A.Aronov, and B.D.Hammock**, "Mass spectrometry-based metabolomics," *Mass Spect.Rev.* 26 (1): 51-78 (2007).
145. **Wilson I.D. et al.**, "HPLC-MS-based methods for the study of metabonomics," *J.Chromatogr.B* 817 (1): 67-76 (2005).
146. **Aharoni A. et al.**, "Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry," *OMICS.* 6 (3): 217-234 (2002).
147. **Brown S.C., G.Kruppa, and J.L.Dasseux**, "Metabolomics applications of FT-ICR mass spectrometry," *Mass Spect.Rev.* 24 (2): 223-231 (2005).
148. **Holmes E. 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* 44 (1-2): 245-255 (1998).

## Références bibliographiques

149. **Lenz E.M. et al.**, "A <sup>1</sup>H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects," *J Pharm.Biomed.Anal* 33 (5): 1103-1115 (2003).
150. **Nicholson J.K. and I.D.Wilson**, "High resolution proton magnetic resonance spectroscopy of biological fluids," *Progress in Nuclear Magnetic Resonance Spectroscopy* 21 (4-5): 449-501 (1989).
151. **Cloarec O. 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.* 77 (2): 517-526 (2005).
152. **Wang Y. 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.* 323 (1): 26-32 (2003).
153. **Keun H.C. et al.**, "Cryogenic probe <sup>13</sup>C NMR spectroscopy of urine for metabonomic studies," *Anal.Chem.* 74 (17): 4588-4593 (2002).
154. **Spraul M. et al.**, "Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe: application to the analysis of acetaminophen metabolites in urine," *Anal.Chem.* 75 (6): 1536-1541 (2003).
155. **Bleakney W.**, "A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor," *Phys.Rev.* 34 (1): 157 (1929).
156. **Nier A.O.C., E.P.Ney, and M.G.Ingham**, "Adjustable Gas Leak," *Review of Scientific Instruments* 18 (3): 191 (1947).
157. **Harrison A.G., T.Gaumann, and D.Stahl**, "Structure and Fragmentation of [C<sub>3</sub>H<sub>7</sub>O]<sup>+</sup> Ions Formed by Chemical Ionization," *Organic Mass Spectrometry* 18 (12): 517-524 (1983).
158. **Cody R.B., J.A.Laramee, and H.D.Durst**, "Versatile new ion source for the analysis of materials in open air under ambient Conditions," *Anal.Chem.* 77 (8): 2297-2302 (2005).
159. **Horning E.C. et al.**, "New picogram detection system based on a mass spectrometer with an external ionization source at atmospheric pressure," *Anal.Chem.* 45 (6): 936-943 (1973).
160. **Frycak P. et al.**, "Atmospheric pressure ionization mass spectrometry of purine and pyrimidine markers of inherited metabolic disorders," *J Mass Spectrom.* 37 (12): 1242-1248 (2002).
161. **Robb D.B., T.R.Covey, and A.P.Bruins**, "Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry," *Anal.Chem.* 72 (15): 3653-3659 (2000).
162. **Kauppila T.J. et al.**, "Negative ion-atmospheric pressure photoionization-mass spectrometry," *J Am Soc.Mass Spectrom.* 15 (2): 203-211 (2004).
163. **Kauppila T.J. et al.**, "Atmospheric pressure photoionization mass spectrometry. Ionization mechanism and the effect of solvent on the ionization of naphthalenes," *Anal.Chem.* 74 (21): 5470-5479 (2002).
164. **Lembcke J. et al.**, "Rapid quantification of free and esterified phytosterols in human serum using APPI-LC-MS/MS," *Journal Of Lipid Research* 46 (1): 21-26 (2005).
165. **Takino M., S.Daishima, and T.Nakahara**, "Determination of chloramphenicol residues in fish meats by liquid chromatography-atmospheric pressure photoionization mass spectrometry," *J.Chromatogr.A* 1011 (1-2): 67-75 (2003).
166. **Yoshioka N., Y.Akiyama, and K.Teranishi**, "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* 1022 (1-2): 145-150 (2004).
167. **Straube E.A., W.Dekant, and W.Volkel**, "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.* 15 (12): 1853-1862 (2004).
168. **Karas M. and F.Hillenkamp**, "Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons," *Anal.Chem.* 60: 2229-2301 (1988).

## Références bibliographiques

169. **Lavanant H. and C.Lange**, "Sodium-tolerant matrix for matrix-assisted laser desorption/ionization mass spectrometry and post-source decay of oligonucleotides," *Rapid Commun.Mass Spectrom.* 16 (20): 1928-1933 (2002).
170. **Mims D. and D.Hercules**, "Quantification of bile acids directly from urine by MALDI-TOF-MS," *Anal.Bioanal.Chem.* 375 (5): 609-616 (2003).
171. **Mims D. and D.Hercules**, "Quantification of bile acids directly from plasma by MALDI-TOF-MS," *Anal.Bioanal.Chem.* 378 (5): 1322-1326 (2004).
172. **Edwards J.L. and R.T.Kennedy**, "Metabolomic analysis of eukaryotic tissue and prokaryotes using negative mode MALDI time-of-flight mass spectrometry," *Anal.Chem.* 77 (7): 2201-2209 (2005).
173. **Sun G. et al.**, "Shotgun metabolomics approach for the analysis of negatively charged water-soluble cellular metabolites from mouse heart tissue," *Anal.Chem.* 79 (17): 6629-6640 (2007).
174. **Laiko V.V., M.A.Baldwin, and A.L.Burlingame**, "Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry," *Anal.Chem.* 72 (4): 652-657 (2000).
175. **Yamashita M. and J.B.Fenn**, "Electrospray Ion-Source - Another Variation on the Free-Jet Theme," *Journal of Physical Chemistry* 88 (20): 4451-4459 (1984).
176. **Whitehouse C.M. et al.**, "Electrospray Interface for Liquid Chromatographs and Mass Spectrometers," *Anal.Chem.* 57 (3): 675-679 (1985).
177. **Fenn J.B. et al.**, "Electrospray ionization for mass spectrometry of large biomolecules," *Science* 246 (4926): 64-71 (1989).
178. **Fenn J.B. et al.**, "Electrospray Ionization-Principles and Practice," *Mass Spectr.Rev.* 9 (1): 37-70 (1990).
179. **Chen H. et al.**, "Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation," *Rapid Commun.Mass Spectrom.* 20 (10): 1577-1584 (2006).
180. **Shen Z. et al.**, "Porous silicon as a versatile platform for laser desorption/ionization mass spectrometry," *Anal.Chem.* 73 (3): 612-619 (2001).
181. **Wei J., J.M.Buriak, and G.Siuzdak**, "Desorption-ionization mass spectrometry on porous silicon," *Nature* 399 (6733): 243-246 (1999).
182. **Huikko K. et al.**, "Feasibility of atmospheric pressure desorption/ionization on silicon mass spectrometry in analysis of drugs," *Rapid Commun.Mass Spectrom.* 17 (12): 1339-1343 (2003).
183. **Thomas J.J. et al.**, "Desorption/ionization on silicon (DIOS): a diverse mass spectrometry platform for protein characterization," *Proc.Natl.Acad.Sci.U.S.A* 98 (9): 4932-4937 (2001).
184. **Nordstrom A. et al.**, "Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics," *Anal.Chem.* 80 (2): 421-429 (2008).
185. **Cooks R.G. et al.**, "Detection Technologies. Ambient mass spectrometry," *Science* 311 (5767): 1566-1570 (2006).
186. **Chen H. et al.**, "Desorption electrospray ionization mass spectrometry for high-throughput analysis of pharmaceutical samples in the ambient environment," *Anal.Chem.* 77 (21): 6915-6927 (2005).
187. **Takats Z. et al.**, "Mass spectrometry sampling under ambient conditions with desorption electrospray ionization," *Science* 306 (5695): 471-473 (2004).
188. **Jackson A.U. et al.**, "Targeted metabolomic analysis of Escherichia coli by desorption electrospray ionization and extractive electrospray ionization mass spectrometry," *Anal.Biochem.* 375 (2): 272-281 (2008).
189. **Ferguson R.E., K.E.McCulloh, and H.M.Rosenstock**, "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* 42 (1): 100-106 (1965).

## Références bibliographiques

190. **Campana J.E.**, "Elementary theory of the quadrupole mass filter," *International Journal of Mass Spectrometry and Ion Physics* 33 (2): 101-117 (1980).
191. **Bouchoux G. and M.Sablier.** 2005. Spectrométrie de masse - Principes et appareillages. In *Techniques de l'ingénieur*. Paris: Editions T.I.
192. **March R.E. and R.J.Hughes.** 1989. *Quadrupole Storage Mass Spectrometry*. Wiley Interscience.
193. **Lafaye A. 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.* 17 (22): 2541-2549 (2003).
194. **Tolstikov V.V. and O.Fiehn**, "Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry," *Anal.Biochem.* 301 (2): 298-307 (2002).
195. **Parks J.H. and A.Szoke**, "Simulation of Collisional Relaxation of Trapped Ion Clouds in the Presence of Space-Charge Fields," *Journal of Chemical Physics* 103 (4): 1422-1439 (1995).
196. **Douglas D.J., A.J.Frank, and D.Mao**, "Linear ion traps in mass spectrometry," *Mass Spect.Rev.* 24 (1): 1-29 (2005).
197. **Hashimoto Y., H.Hasegawa, and I.Waki**, "Dual linear ion trap/orthogonal acceleration time-of-flight mass spectrometer with improved precursor ion selectivity," *Rapid Commun.Mass Spectrom.* 19 (11): 1485-1491 (2005).
198. **Schwartz J.C., M.W.Senko, and J.E.Syka**, "A two-dimensional quadrupole ion trap mass spectrometer," *J Am Soc.Mass Spectrom.* 13 (6): 659-669 (2002).
199. **Bantscheff M. et al.**, "Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer," *Mol.Cell Proteomics.* (2008).
200. **Griffin T.J. et al.**, "iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer," *J.Proteome Res.* 6 (11): 4200-4209 (2007).
201. **Stephens W.E.**, "A pulsed mass spectrometer with time dispersion," *Phys.Rev.* 69: 691 (1946).
202. **Wiley W.C. and I.H.McLaren**, "Time-of-Flight Mass Spectrometer with Improved Resolution," *Review of Scientific Instruments* 26 (12): 1150-1157 (1955).
203. **Mamyrin B.A. et al.**, "Mass-Reflectron, A New Nonmagnetic Time-Of-Flight High-Resolution Mass-Spectrometer," *Sov.Phys.JETP* 64 (1): 82-89 (1973).
204. **Roepenack-Lahaye E. et al.**, "Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry," *Plant Physiol* 134 (2): 548-559 (2004).
205. **Weckwerth W. et al.**, "Differential metabolic networks unravel the effects of silent plant phenotypes," *Proc.Natl.Acad.Sci.U.S.A* 101 (20): 7809-7814 (2004).
206. **Major H.J. et al.**, "A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition," *Rapid Commun.Mass Spectrom.* 20 (22): 3295-3302 (2006).
207. **Plumb R.S. 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.* 20 (19): 2800-2806 (2006).
208. **Williams R. et al.**, "A multi-analytical platform approach to the metabonomic analysis of plasma from normal and Zucker (fa/fa) obese rats," *Mol.Biosyst.* 2 (3-4): 174-183 (2006).
209. **Castro-Perez J., M.McCullagh, and A.Millar**, "720001170EN - The benefits of extended dynamic range for metabolite ID using UPLC/MS/MS," *Waters Technical Note 720001170EN* (2005).

## Références bibliographiques

210. **McCullagh M., J.Castro-Perez, and L.Calton**, "720001596EN - Q-ToF Premier dynamic range enhancement and accurate mass measurement performance," *Waters Technical Note 720001596EN* (2006).
211. **Marshall A.G., C.L.Hendrickson, and G.S.Jackson**, "Fourier transform ion cyclotron resonance mass spectrometry: a primer," *Mass Spectr.Rev.* 17 (1): 1-35 (1998).
212. **Marshall A.G.**, "Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development," *International Journal of Mass Spectrometry* 200 (1-3): 331-356 (2000).
213. **Hughey C.A., R.P.Rodgers, and A.G.Marshall**, "Resolution of 11,000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil," *Anal.Chem.* 74 (16): 4145-4149 (2002).
214. **Kruppa G. et al.**, "Multiple ion isolation applications in FT-ICR MS: exact-mass MS<sub>n</sub> internal calibration and purification/interrogation of protein-drug complexes," *Anal.Chem.* 74 (15): 3877-3886 (2002).
215. **Baidoo E.E. 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.* 80 (9): 3112-3122 (2008).
216. **Nakamura Y. et al.**, "Differential metabolomics unraveling light/dark regulation of metabolic activities in Arabidopsis cell culture," *Planta* 227 (1): 57-66 (2007).
217. **Ohta D., D.Shibata, and S.Kanaya**, "Metabolic profiling using Fourier-transform ion-cyclotron-resonance mass spectrometry," *Anal.Bioanal.Chem.* 389 (5): 1469-1475 (2007).
218. **Southam A.D. et al.**, "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.* 79 (12): 4595-4602 (2007).
219. **Stentiford G.D. et al.**, "Liver tumors in wild flatfish: a histopathological, proteomic, and metabolomic study," *OMICS.* 9 (3): 281-299 (2005).
220. **Takahashi H. et al.**, "Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry," *Anal.Bioanal.Chem.* 391 (8): 2769-2782 (2008).
221. **Makarov A.**, "Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis," *Anal.Chem.* 72 (6): 1156-1162 (2000).
222. **Hu Q. et al.**, "The Orbitrap: a new mass spectrometer," *J.Mass Spectrom.* 40 (4): 430-443 (2005).
223. **Yates J.R. et al.**, "Performance of a linear ion trap-Orbitrap hybrid for peptide analysis," *Anal.Chem.* 78 (2): 493-500 (2006).
224. **Makarov A. et al.**, "Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer," *Anal.Chem.* 78 (7): 2113-2120 (2006).
225. **Ding J. et al.**, "Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling," *Anal.Chem.* 79 (16): 6081-6093 (2007).
226. **Dunn W.B. 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.* 871 (2): 288-298 (2008).
227. **Kenny L.C. et al.**, "Detection and identification of novel metabolomic biomarkers in preeclampsia," *Reprod.Sci.* 15 (6): 591-597 (2008).
228. **Zubarev R.A., N.L.Kelleher, and F.W.McLafferty**, "Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process," *J.Am.Chem.Soc.* 120 (13): 3265-3266 (1998).
229. **Woodin R.L., D.S.Bomse, and J.L.Beauchamp**, "Multiphoton dissociation of molecules with low power continuous wave infrared laser radiation," *J.Am.Chem.Soc.* 100 (10): 3248-3250 (1978).

## Références bibliographiques

230. **Cooper H.J., K.Hakansson, and A.G.Marshall**, "The role of electron capture dissociation in biomolecular analysis," *Mass Spectr.Rev.* 24 (2): 201-222 (2005).
231. **Levsen K. 1978.** *Fundamental Aspects of Organic Mass Spectrometry*. Weinheim: Verlag chemie.
232. **Lafaye A. 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.* 39 (6): 655-664 (2004).
233. **Lenz E.M. 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* 10 (2-3): 173-187 (2005).
234. **Wagner S. 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.* 78 (4): 1296-1305 (2006).
235. **Boernsen K.O., S.Gatzek, and G.Imbert**, "Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma," *Anal.Chem.* 77 (22): 7255-7264 (2005).
236. **Williams R.E. et al.**, "D-Serine-induced nephrotoxicity: a HPLC-TOF/MS-based metabonomics approach," *Toxicology* 207 (2): 179-190 (2005).
237. **Smedsgaard J. and J.C.Frisvad**, "Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts," *Journal of Microbiological Methods* 25 (1): 5-17 (1996).
238. **Goodacre R. et al.**, "Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry," *Phytochemistry* 62 (6): 859-863 (2003).
239. **Madalinski G. et al.**, "Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis," *Anal.Chem.* 80 (9): 3291-3303 (2008).
240. **Scholz M. et al.**, "Metabolite fingerprinting: detecting biological features by independent component analysis," *Bioinformatics.* 20 (15): 2447-2454 (2004).
241. **Adahchour M. et al.**, "Twin comprehensive two-dimensional gas chromatographic system: concept and applications," *J.Chromatogr.A* 1086 (1-2): 128-134 (2005).
242. **Kovats E.S. 1965.** Gas Chromatographic Characterization of Organic Substances in the Retention Index System. In *Advances in Chromatography*. Edited by J. C. Giddins and R. A. Keller. New-York: M. Dekker Inc.
243. **Fiehn O. et al.**, "Metabolite profiling for plant functional genomics," *Nat.Biotechnol.* 18 (11): 1157-1161 (2000).
244. **Bowers L.D. and Sanaullah**, "Direct measurement of steroid sulfate and glucuronide conjugates with high-performance liquid chromatography-mass spectrometry," *J.Chromatogr.B* 687 (1): 61-68 (1996).
245. **Idborg H. et al.**, "Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry," *J.Chromatogr.B* 828 (1-2): 9-13 (2005).
246. **Lafaye A. et al.**, "Liquid chromatography-mass spectrometry and <sup>15</sup>N metabolic labeling for quantitative metabolic profiling," *Anal.Chem.* 77 (7): 2026-2033 (2005).
247. **Strege M.A.**, "Hydrophilic Interaction Chromatography-Electrospray Mass Spectrometry Analysis of Polar Compounds for Natural Product Drug Discovery," *Anal.Chem.* 70 (13): 2439-2445 (1998).
248. **Tolstikov V.V. et al.**, "Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics," *Anal.Chem.* 75 (23): 6737-6740 (2003).
249. **Plumb R. et al.**, "Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry," *Rapid Commun.Mass Spectrom.* 18 (19): 2331-2337 (2004).
250. **Van Deemter J.J., F.J.Zuiderweg, and A.Klinkenberg**, "Longitudinal Diffusion and Resistance to Mass Transfer As Causes of Nonideality in Chromatography," *Chemical Engineering Science* 5 (6): 271-289 (1956).

## Références bibliographiques

251. **Leandro C.C. et al.**, "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* 1103 (1): 94-101 (2006).
252. **Nordstrom A. et al.**, "Nonlinear Data Alignment for UPLC-MS and HPLC-MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum," *Anal.Chem.* 78 (10): 3289-3295 (2006).
253. **Bailey N.J. 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.* 48 (1): 42-46 (2000).
254. **Lindon J.C., J.K.Nicholson, and I.D.Wilson.** 2003. Biomedical and Pharmaceutical Applications of HPLC-NMR and HPLC-NMR-MS. In *On-Line LC-NMR and Related Techniques*. Edited by K. Albert.
255. **Shockcor J.P. 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.* 68 (24): 4431-4435 (1996).
256. **Bino R.J. et al.**, "Potential of metabolomics as a functional genomics tool," *Trends Plant Sci.* 9 (9): 418-425 (2004).
257. **Fiehn O. et al.**, "Minimum reporting standards for plant biology context information in metabolomic studies," *Metabolomics* 3 (3): 195-201 (2007).
258. **Goodacre R. et al.**, "Proposed minimum reporting standards for data analysis in metabolomics," *Metabolomics* 3 (3): 231-241 (2007).
259. **Holmes E. 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.* 220 (2): 284-296 (1994).
260. **Craig A. et al.**, "Scaling and normalization effects in NMR spectroscopic metabolomic data sets," *Anal.Chem.* 78 (7): 2262-2267 (2006).
261. **Forshed J. et al.**, "A comparison of methods for alignment of NMR peaks in the context of cluster analysis," *J Pharm.Biomed.Anal* 38 (5): 824-832 (2005).
262. **Stoyanova R. et al.**, "Sample classification based on Bayesian spectral decomposition of metabolomic NMR data sets," *Anal.Chem.* 76 (13): 3666-3674 (2004).
263. **Broeckling C.D. et al.**, "MET-IDEA: Data Extraction Tool for Mass Spectrometry-Based Metabolomics," *Anal.Chem.* 78 (13): 4334-4341 (2006).
264. **Duran A.L. et al.**, "Metabolomics spectral formatting, alignment and conversion tools (MSFACTs)," *Bioinformatics.* 19 (17): 2283-2293 (2003).
265. **Katajamaa M. and M.Oresic.**, "Processing methods for differential analysis of LC/MS profile data," *BMC Bioinformatics.* 6: 179 (2005).
266. **Smith C.A. et al.**, "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification," *Anal.Chem.* 78 (3): 779-787 (2006).
267. **Tikunov Y. et al.**, "A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles," *Plant Physiol* 139 (3): 1125-1137 (2005).
268. **Katajamaa M. and M.Oresic.**, "Data processing for mass spectrometry-based metabolomics," *J.Chromatogr.A* 1158 (1-2): 318-328 (2007).
269. **Lutz U., R.W.Lutz, and W.K.Lutz.**, "Metabolic Profiling of Glucuronides in Human Urine by LC-MS/MS and Partial Least-Squares Discriminant Analysis for Classification and Prediction of Gender," *Anal.Chem.* 78 (13): 4564-4571 (2006).

## Références bibliographiques

270. **Baran R. et al.**, "MathDAMP: a package for differential analysis of metabolite profiles," *BMC Bioinformatics*. 7: 530 (2006).
271. **Vorst O. et al.**, "A non-directed approach to the differential analysis of multiple LC-MS-derived metabolic profiles," *Metabolomics* 1 (2): 169-180 (2005).
272. **Crawford L.R. and J.D.Morrison**, "Computer Methods in Analytical Mass Spectrometry - Identification of An Unknown Compound in A Catalog," *Anal.Chem.* 40 (10): 1464-& (1968).
273. **Wang W.X. et al.**, "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards," *Anal.Chem.* 75 (18): 4818-4826 (2003).
274. **Oresic M. et al.**, "Phenotype characterisation using integrated gene transcript, protein and metabolite profiling," *Appl.Bioinformatics*. 3 (4): 205-217 (2004).
275. **Hermansson M. et al.**, "Automated quantitative analysis of complex lipidomes by liquid chromatography/mass spectrometry," *Anal.Chem.* 77 (7): 2166-2175 (2005).
276. **Bijlsma S. et al.**, "Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation," *Anal.Chem.* 78 (2): 567-574 (2006).
277. **Aitchison J.** 2003. *The statistical Analysis of Compositional Data*. Caldwell, NJ: Blackburn Press.
278. **Sysi-Aho M. et al.**, "Normalization method for metabolomics data using optimal selection of multiple internal standards," *BMC Bioinformatics*. 8: 93 (2007).
279. **Wu L. 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.* 336 (2): 164-171 (2005).
280. **Birkemeyer C. et al.**, "Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling," *Trends in Biotechnology* 23 (1): 28-33 (2005).
281. **Hegeman A.D. et al.**, "Stable isotope assisted assignment of elemental compositions for metabolomics," *Anal.Chem.* 79 (18): 6912-6921 (2007).
282. **Pearson K.**, "On lines and planes of closest fit to systems of points in space," *Philos Mag* (2): 559-572 (1901).
283. **Hotelling H.**, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology* 24: 417-441 (1933).
284. **Wold S. and K.-G.Joreskog.** 1982. Soft modeling. The basic design and some extensions. In *Systems Under Indirect Observation Vols. I and II*.
285. **Wold S., M.Sjöström, and L.Eriksson.** 1998. *PLS in Chemistry, The Encyclopedia of Computational Chemistry*. Edited by P.V.R.Schleyer et al.
286. **Trygg J. and S.Wold**, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics* 16 (3): 119-128 (2002).
287. **Trygg J., E.Holmes, and T.Lundstedt**, "Chemometrics in metabonomics," *J.Proteome Res.* 6 (2): 469-479 (2007).
288. **Cloarec O. et al.**, "Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic <sup>1</sup>H NMR data sets," *Anal.Chem.* 77 (5): 1282-1289 (2005).
289. **Wolfender J.L., K.Ndjoko, and K.Hostettmann**, "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* 1000 (1-2): 437-455 (2003).
290. **Molyneux R.J. and P.Schieberle**, "Compound identification: a Journal of Agricultural and Food Chemistry perspective," *J.Agric.Food Chem.* 55 (12): 4625-4629 (2007).
291. **Bogusz M.J. et al.**, "Poor reproducibility of in-source collisional atmospheric pressure ionization mass spectra of toxicologically relevant drugs," *J.Chromatogr.A* 844 (1-2): 409-418 (1999).

## Références bibliographiques

292. **Cui Q. et al.**, "Metabolite identification via the Madison Metabolomics Consortium Database," *Nat.Biotechnol.* 26 (2): 162-164 (2008).
293. **Kopka J. et al.**, "GMD@CSB.DB: the Golm Metabolome Database," *Bioinformatics.* 21 (8): 1635-1638 (2005).
294. **Werner E. et al.**, "Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends," *J.Chromatogr.B* 871 (2): 143-163 (2008).
295. **Wishart D.S. et al.**, "HMDB: the Human Metabolome Database," *Nucleic Acids Res.* 35 (Database issue): D521-D526 (2007).
296. **Watanabe K., E.Yasugi, and M.Oshima**, "How to search the glycolipid data in \"LIPIDBANK for Web\", the newly developed lipid database in Japan," *Trends Glycosci.Glycotechnol.* 12 (65): 175-184 (2000).
297. **Smith C.A. et al.**, "METLIN: a metabolite mass spectral database," *Ther.Drug Monit.* 27 (6): 747-751 (2005).
298. **Horai, H., Suwa, K., Arita, M., Nihei, Y., and Nishioka, T.** Mass Spectral Database for Metabolome Analysis. 2007. Indianapolis, USA. 55th ASMS Conference on Mass Spectrometry and Allied Topics. Ref Type: Conference Proceeding
299. **Plumb R.S. et al.**, "A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice," *Analyst* 130 (6): 844-849 (2005).
300. **Sleno L., D.A.Volmer, and A.G.Marshall**, "Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power," *J.Am.Soc.Mass Spectrom.* 16 (2): 183-198 (2005).
301. **Stoll N., E.Schmidt, and K.Thurow**, "Isotope pattern evaluation for the reduction of elemental compositions assigned to high-resolution mass spectral data from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry," *J.Am.Soc.Mass Spectrom.* 17 (12): 1692-1699 (2006).
302. **Levin S.S. et al.**, "The Metabolism and Excretion of Enzyme-Inducing Doses of Phenobarbital by Rats with Bile Fistulas," *Xenobiotica* 16 (3): 213-224 (1986).
303. **Verite P. et al.**, "Phenobarbital metabolism by hepatocytes isolated from rat," *Drug Metabol.Drug Interact.* 13 (1): 41-55 (1996).
304. **Caldwell J. et al.**, "The metabolic fate of [14C]-phenobarbitone in the rat and the effect of chronic administration and dose size (proceedings)," *Br.J.Pharmacol.* 60 (2): 295P-296P (1977).
305. **Anderson G.D. and R.H.Levy**, "The effect of valproate on the metabolism of phenobarbital in the rat," *Pharm.Res.* 9 (12): 1622-1628 (1992).
306. **Soine W.H. et al.**, "Identification of phenobarbital N-glucosides as urinary metabolites of phenobarbital in mice," *J.Pharm.Sci.* 80 (2): 99-103 (1991).
307. **Neighbors S.M. and W.H.Soine**, "Identification of phenobarbital N-glucuronides as urinary metabolites of phenobarbital in mice," *Drug Metab Dispos.* 23 (5): 548-552 (1995).
308. **Andresen B.D. et al.**, "Synthesis and characterization of alpha-phenyl-gamma-butyrolactone, a metabolite of glutethimide, phenobarbital and primidone, in human urine," *Res.Commun.Chem.Pathol.Pharmacol.* 15 (1): 21-29 (1976).
309. **Harvey D.L., L.Glazener, and L.Stratton**, "Detection of a 5-(3,4-dihydroxy-1,5-cyclohexadien-1-yl)-metabolite of phenobarbital and mephobarbital in rat, guinea pig and human.," *Res.Comm.Chem.Path.Pharm.* 3: 557-565 (1972).
310. **Tang B.K., W.Kalow, and A.A.Grey**, "Metabolic fate of phenobarbital in man. N-Glucoside formation," *Drug Metab Dispos.* 7 (5): 315-318 (1979).
311. **Kallberg N. et al.**, "Quantitation of phenobarbital and its main metabolites in human urine," *Eur.J.Clin.Pharmacol.* 9 (2-3): 161-168 (1975).

## Références bibliographiques

312. **Soine W.H. et al.**, "Stereochemical characterization of the diastereomers of the phenobarbital N-beta-D-glucose conjugate excreted in human urine," *Pharm.Res.* 7 (4): 402-406 (1990).
313. **Bernus I. et al.**, "Urinary excretion of phenobarbitone and its metabolites in chronically treated patients," *Eur.J.Clin.Pharmacol.* 46 (5): 473-475 (1994).
314. **Kostiainen R. et al.**, "Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies," *Journal of Mass Spectrometry* 38 (4): 357-372 (2003).
315. **Zhu M. et al.**, "Detection and characterization of metabolites in biological matrices using mass defect filtering of liquid chromatography/high resolution mass spectrometry data," *Drug Metab Dispos.* 34 (10): 1722-1733 (2006).
316. **Kendrick E.**, "A Mass Scale Based on CH<sub>2</sub> = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds.," *Anal.Chem.* 35 (13): 2146-2154 (1963).
317. **Hughey C.A. et al.**, "Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra," *Anal.Chem.* 73 (19): 4676-4681 (2001).
318. **He H. et al.**, "Method for lipidomic analysis: p53 expression modulation of sulfatide, ganglioside, and phospholipid composition of U87 MG glioblastoma cells," *Anal.Chem.* 79 (22): 8423-8430 (2007).
319. **Lerno L.A., Jr., J.B.German, and C.B.Lebrilla.**, "Method for the identification of lipid classes based on referenced Kendrick mass analysis," *Anal.Chem.* 82 (10): 4236-4245 (2010).
320. **Jandera P. and M.Spacek.**, "Method for characterization of selectivity in reversed-phase liquid chromatography. III. Retention behaviour in gradient-elution chromatography: application to the chromatography of pesticide compounds," *J.Chromatogr.* 366: 107-126 (1986).
321. **van Krevelen D.W.**, "Graphical-statistical method for the study of structure and reaction processes of coal," *Fuel* 29: 269-284 (1950).
322. **Hedges J.I.**, "The Chemistry of Archaeological Wood," *Advances in Chemistry Series* (225): 111-140 (1990).
323. **Kim S., R.W.Kramer, and P.G.Hatcher.**, "Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram," *Analytical Chemistry* 75 (20): 5336-5344 (2003).
324. **Wu Z., R.P.Rodgers, and A.G.Marshall.**, "Two- and three-dimensional van krevelen diagrams: a graphical analysis complementary to the kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband fourier transform ion cyclotron resonance mass measurements," *Analytical Chemistry* 76 (9): 2511-2516 (2004).
325. **Sleighter R.L. and P.G.Hatcher.**, "The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter," *Journal of Mass Spectrometry* 42 (5): 559-574 (2007).
326. **Wiklund S. et al.**, "Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models," *Anal.Chem.* 80 (1): 115-122 (2008).
327. **Robosky L.C. et al.**, "Metabonomic identification of two distinct phenotypes in Sprague-Dawley (CrI:CD(SD)) rats," *Toxicol.Sci.* 87 (1): 277-284 (2005).
328. **Aranibar N. et al.**, "Modulation of ascorbic acid metabolism by cytochrome P450 induction revealed by metabonomics and transcriptional profiling," *Magn Reson.Chem.* 47 Suppl 1: S12-S19 (2009).
329. **Holmes E., I.D.Wilson, and J.K.Nicholson.**, "Metabolic phenotyping in health and disease," *Cell* 134 (5): 714-717 (2008).
330. **Nicholson J.K., E.Holmes, and I.D.Wilson.**, "Gut microorganisms, mammalian metabolism and personalized health care," *Nat.Rev.Microbiol.* 3 (5): 431-438 (2005).
331. **Ramautar R. et al.**, "Metabolic profiling of human urine by CE-MS using a positively charged capillary coating and comparison with UPLC-MS," *Mol.Biosyst.* 7 (1): 194-199 (2011).

## Références bibliographiques

332. **Ramautar R. et al.**, "CE-MS for metabolomics: Developments and applications in the period 2008-2010," *Electrophoresis* 32 (1): 52-65 (2011).
333. **van der Greef J. et al.**, "The art and practice of systems biology in medicine: mapping patterns of relationships," *J.Proteome Res.* 6 (4): 1540-1559 (2007).
334. **Guy P.A. et al.**, "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.* 871 (2): 253-260 (2008).
335. **Peters S., V.E.van, and H.G.Janssen**, "Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples," *Anal.Bioanal.Chem.* 394 (5): 1273-1281 (2009).
336. **Croixmarie V. et al.**, "Integrated comparison of drug-related and drug-induced ultra performance liquid chromatography/mass spectrometry metabonomic profiles using human hepatocyte cultures," *Anal.Chem.* 81 (15): 6061-6069 (2009).
337. **Palma P. et al.**, "Electron ionization in LC-MS: recent developments and applications of the direct-EI LC-MS interface," *Anal.Bioanal.Chem.* 399 (8): 2683-2693 (2011).
338. **Lommen A.**, "MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing," *Anal.Chem.* 81 (8): 3079-3086 (2009).
339. **Benton H.P. et al.**, "XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization," *Anal.Chem.* 80 (16): 6382-6389 (2008).
340. **Ulrich E.L. et al.**, "BioMagResBank," *Nucleic Acids Res.* 36 (Database issue): D402-D408 (2008).
341. **Noble D.**, "Systems biology and the heart," *Biosystems* 83 (2-3): 75-80 (2006).
342. **van der Greef J.**, "Systems biology, connectivity and the future of medicine," *Syst.Biol.(Stevenage.)* 152 (4): 174-178 (2005).

# ANNEXES



**Annexe 1 : Articles**

Ducruix C, Vailhen D, Werner E, Fievet JB, Bourguignon J, Tabet JC et al.

**Metabolomic investigation of the response of the model plant *Arabidopsis thaliana* to cadmium exposure: Evaluation of data pretreatment methods for further statistical analyses.**

*Chemometrics and Intelligent Laboratory Systems 2008; 91(1):67-77.*



# Metabolomic investigation of the response of the model plant *Arabidopsis thaliana* to cadmium exposure: Evaluation of data pretreatment methods for further statistical analyses

Céline Ducruix<sup>a</sup>, Dominique Vailhen<sup>b</sup>, Erwan Werner<sup>a</sup>, Julie B. Fievet<sup>c</sup>, Jacques Bourguignon<sup>c</sup>, Jean-Claude Tabet<sup>d</sup>, Eric Ezan<sup>a</sup>, Christophe Junot<sup>a,\*</sup>

<sup>a</sup> CEA, Service de Pharmacologie et d'Immunologie, DSV/DRM, 91191 Gif-sur-Yvette, France

<sup>b</sup> CEA, DAM/DASE/SRCE, BP 12, 91680 Bruyères-le-Chatel, France

<sup>c</sup> CEA, DSV/iRTSV/Laboratoire de Physiologie Cellulaire Végétale, UMR5168 CEA CNRS UJF INRA, 38054 Grenoble France

<sup>d</sup> Laboratoire de Chimie Structurale Organique et Biologique, CNRS UMR 7613, Université Pierre et Marie Curie, 75252 Paris, France

Received 31 December 2006; received in revised form 6 May 2007; accepted 1 August 2007  
Available online 17 August 2007

## Abstract

Understanding the metabolic answer and the adaptation of plants towards heavy metal exposure opens the way to future phytoremediation of polluted sites. For this reason, we studied the impact of cadmium, a toxic heavy metal, on the metabolome of the model plant *Arabidopsis thaliana*. The analytical methodology (liquid chromatography coupled with mass spectrometry) that was used within the framework of a designed experiment conducted on *A. thaliana* cells exposed to cadmium generated an important volume of data. Multivariate statistical analyses appeared relevant to compare the metabolic fingerprints in order to isolate and identify some discriminating metabolites.

Three types of data pretreatment, *i.e.*, reduction of dimensionality, bucketing and automatic processing by the MetAlign™ software were compared for efficiency in extracting the information. The pretreated data were then subjected to multivariate statistical analysis by principal component analysis (PCA) and partial least square regression (PLS). Finally, an OSC (Orthogonal Signal Correction)-PLS2 approach performed on kinetic and dose ranging studies allowed to visualize time- and cadmium dose-induced changes on the metabolism of *A. thaliana* cells.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Metabolomics; LC/MS; Data pretreatment; Chemometrics; Multivariate statistical analysis; PLS

## 1. Introduction

Metabolomics deals with the comprehensive analysis of metabolites present in a biological sample by the combined use of analytical methods and multivariate statistical analyses [1,2]. It has emerged at the end of the 90's as the third major path of functional genomics beside mRNA profiling (transcriptomics) and proteomics [3,4]. Metabolomic approaches seek to profile metabolites in a non targeted way, *i.e.* to reliably separate and detect as many metabolites as possible in a single analysis. The term “metabolite” refers to all small molecular weight com-

pounds including organic acids, sugars, amino acids, vitamins, as well as small peptides. These compounds are involved in metabolic processes, whether they are the products or being necessary to the mechanism.

The metabolome is characterized by a huge chemical diversity, making the detection of all metabolites by a single and universal method impossible. As a consequence, metabolome analysis has already been reported by using nuclear magnetic resonance (NMR) [5,6], Fourier transform infrared spectroscopy [7], pyrolysis/electron impact-mass spectrometry (pyrolysis/EI-MS) [8], gas chromatography/electron impact-mass spectrometry [9], electrospray mass spectrometry (ESI-MS) [10] and also ESI-MS coupled with liquid chromatography (LC/MS) [11–13]. We selected the latter technology because it provides not only chemometric information, but also useful information

\* Corresponding author.

E-mail address: [christophe.junot@cea.fr](mailto:christophe.junot@cea.fr) (C. Junot).

concerning the structure and identity of metabolites, such as the molecular mass of intact compounds.

Following their acquisition by a given analytical tool, the metabolic fingerprints are compared by using multivariate statistical analyses, with the aim to reduce the complexity of the data sets and to highlight the analytical information of biological relevance. In the field of metabolomics, a preliminary step often relies on the use of unsupervised analyses such as principal component analysis (PCA) [14,15]. This descriptive method allows to visualize the original data in a two or three dimensional space by reducing the dimensionality of complex data sets. Explicative analyses are then performed by using supervised tools such as PLS (projection to latent structures or partial least squares) regression [16], PLS-discriminant analysis (PLS-DA) [17], or more recently OPLS (Orthogonal Projection on Latent Structure) [18]. However, the study of metabolic fingerprints generated with LC/MS is complex, due to the high background noise, the presence of artefacts and also the occurrence of redundancy (*i.e.* different  $m/z$  ratios that are related to the same metabolite). Moreover, the background noise can vary with the retention time, the peak shape can be different from one peak to another and retention time of compounds may exhibit some variations from one chromatogram to another. Thus the hypothetical “direct analysis” of LC/MS data by a statistical software may result in serious nonsense.

To overcome these shortcomings, chemometricians have recently developed special techniques. Warping is a way known to deal with time shift and “align” different chromatograms. However and according to our knowledge, when this study was initiated, no efficient warping software was able to deal safely with few hundreds of peaks together with notable baseline drift. Furthermore, the raw output of a single sample is 2-D: variables (*i.e.* couple of retention time and  $m/z$  ratio, for LC/MS or GC/MS) are displayed in columns whereas their related intensities in the samples correspond to entries and the samples correspond to rows. Comparing samples would mean a 3-D data set, which methods such as Tucker or Parafac can obviously handle. However, as statistical packages combining PLS and 3-D methods were not available in our laboratory for the data to come, we focused on 2-D methods known to us and decided to solve the problem by an adapted preprocessing of raw data. A coarse way to get a 2-D data set is to collapse the time axis, thus ignoring the LC separation [19,20]. A more elaborate way is to truncate the time axis, and collapse the time axis within defined regular brackets, thus dealing with several pooled mass spectra, as if collected in a bucket over each time window. It implies dealing with as many 2-D data sets as created buckets [11]. Otherwise, an alternative solution relies on the use of softwares allowing peak detection and correction of retention time shifts [21–24]. However, the pretreatment of metabolomic data sets is not restricted to the extraction of relevant analytical information. It also deals with scaling procedures, aimed at adjusting the huge differences observed in the metabolite concentrations, and with normalization, which addresses the issue of the comparison of one sample to another. This issue has been addressed for NMR [25] and GC/MS [26] data sets, but LC/MS are not structured in the same way.

In this framework, the issue of the pretreatment of LC/MS data for further multivariate statistical analysis was addressed by

using a metabolomic study performed on *Arabidopsis thaliana* (*A. thaliana*) cells exposed to cadmium. Heavy metals such as lead (Pb), silver (Ag), tin (Sn) or cadmium (Cd) are major environmental pollutants, particularly in areas where the anthropogenic pressure is high. The presence of those metals in the atmosphere, the soil and the water, even at low concentration, can generate toxic effects on all organisms. Their bioaccumulation among the food chain represents a major public health problem [27]. Plants display a wide range of potential mechanisms at the cellular level that are involved in detoxification and thus tolerance to heavy metal stress [28]. Among them, chelation of metals by phytochelatins (Pcs) followed by their transport into the vacuole is the most important [29].

Phytochelatins are peptides which are induced by heavy metal exposure. They are characterized by the general structure  $(\gamma\text{-Glu-Cys})_n\text{-Gly}$ , but in some plant species, the C-terminal glycine can be replaced by serine, glutamine, glutamate or alanine, leading to *iso*-phytochelatins (*iso*-Pcs). Pcs are synthesized from glutathione by a constitutive enzyme which is a  $\gamma$ -EC dipeptidyl transpeptidase named phytochelatin synthase (PCS). Its molecular identity has only recently been determined by 3 groups who cloned and characterized the genes encoding PCS (*AtPCS1*) [30–32]. On the other hand, the biosynthesis of *iso*-Pcs remains unclear. Although their distribution is considered to differ from one specie to another, we previously showed by using a metabolomic based approach that *A. thaliana* cells are able to synthesize most Pc-related peptides (Pcs and *iso*-Pcs) described in the literature [33,34].

In this context, the aim of the present work was to use some already explored metabolomic data sets in order to evaluate the impact of the different methods of pretreatment of the data on the isolation of 14 metabolites of biological relevance (*i.e.* Pcs and *iso*-Pcs) from complex metabolic fingerprints by using multivariate statistical analyses. Various statistical tools such as PCA, PLS, PLS2 and OSC (Orthogonal Signal Correction)-PLS2 were then applied to the analysis of a structured experiment along with an experimental design combining dose and exposure time factors with the aim to visualize the metabolic changes induced by cadmium in *A. thaliana* cells.

## 2. Materials and methods

### 2.1. Reagents

MS culture medium (ref. M5519), Methanol, prednisone, colchicine, Tris and cadmium nitrate were from Sigma (Saint-Quentin Fallavier, France). Acetonitrile (ACN) was from SDS (Peypin, France). Formic acid was from Merck (Briare-le-canal, France). Water was deionized and filtered through a Millipore Milli-Q water purification system.

### 2.2. Sample preparation

*A. thaliana* cell culture and Cd treatment was described elsewhere [34]. Before each Cd treatment, 4-day-subcultures of *Arabidopsis* cell suspensions were carried out twice, and after the second subculture, cells were treated, from day 4, with



lens offset at 26 V. Injection time was 50 ms. The MS cycle time was 1 s and the positive to negative switch time was 400 ms. Mass spectra were recorded at unit mass resolution without in-source fragmentation from  $m/z$  100 to 1500.

Samples were randomly run. Experiments were judged suitable for further biological interpretation if the coefficients of variation (CVs) obtained from the areas of the chromatographic peaks of colchicines and prednisone, the two reference compounds present in the extraction buffer, were below 20% for all chromatographic injections.

## 2.4. Data pretreatment

### 2.4.1. Reduction of dimensionality and bucketing

The recorded mass spectra were averaged either over the complete chromatographic acquisition (reduction of dimensionality) or over the successive 5 min zones (bucketing) and were then exported to an Excel file. However, when no signal was recorded for a given  $m/z$  ratio, it was omitted in the Excel sheet, leading to problems of variable alignment in the resulting two-dimensional data sets (samples  $\times$  variables, *i.e.*  $m/z$  ratios). This has been corrected by using an in-house Visual Basic program running on Excel that made match all the  $m/z$  ratios from one sample to another before performing multivariate statistical analyses.

### 2.4.2. Software supported method

MetAlign™ is a software from the Plant Research International (<http://www.metalign.nl>, PRI, NL) dedicated to metabolomics [35]. It performs baseline correction of ion traces by estimating the local background noise as a function of the retention time, calculates and stores peak maximum amplitudes (*i.e.* peak height), and aligns the chromatograms of a data set by using peaks of high signal to noise ratio that are common to all chromatograms. An iterative fine alignment procedure is also available by including an increasing number of landmark peaks with lower signal to noise ratios.

Table 1  
Phytochelatin and *iso*-phytochelatin identified in *A. thaliana* cells [33]

Identity	Formula	$m/z$	Tr	
Pcs <sup>a</sup>	Pc <sub>2</sub>	$\gamma$ -(Glu-Cys) <sub>2</sub> -Gly	540	9.7
	Pc <sub>3</sub>	$\gamma$ -(Glu-Cys) <sub>3</sub> -Gly	772	21.2
	Pc <sub>4</sub>	$\gamma$ -(Glu-Cys) <sub>4</sub> -Gly	1004	23.3
	Pc <sub>5</sub>	$\gamma$ -(Glu-Cys) <sub>5</sub> -Gly	1236	25.0
	<i>Iso</i> -Pc <sup>b</sup> (Ala)	<i>Iso</i> -Pc <sub>2</sub> (Ala)	$\gamma$ -(Glu-Cys) <sub>2</sub> -Ala	554
<i>Iso</i> -Pc <sub>3</sub> (Ala)		$\gamma$ -(Glu-Cys) <sub>3</sub> -Ala	786	22.1
<i>Iso</i> -Pc <sub>4</sub> (Ala)		$\gamma$ -(Glu-Cys) <sub>4</sub> -Ala	1018	24
<i>Iso</i> -Pc (Ser)	<i>Iso</i> -Pc <sub>3</sub> (Ser)	$\gamma$ -(Glu-Cys) <sub>3</sub> -Ser	802	20.6
	<i>Iso</i> -Pc <sub>4</sub> (Ser)	$\gamma$ -(Glu-Cys) <sub>4</sub> -Ser	1034	22.9
<i>Iso</i> -Pc (Gln)	<i>Iso</i> -Pc <sub>3</sub> (Gln)	$\gamma$ -(Glu-Cys) <sub>3</sub> -Gln	843	20.8
	<i>Iso</i> -Pc <sub>4</sub> (Gln)	$\gamma$ -(Glu-Cys) <sub>4</sub> -Gln	1075	23.0
<i>Iso</i> -Pc (Glu)	<i>Iso</i> -Pc <sub>3</sub> (Glu)	$\gamma$ -(Glu-Cys) <sub>3</sub> -Glu	844	20.8
	DesGluPc	Cys- $\gamma$ -(Glu-Cys) <sub>2</sub> -Gly	643	17.0
	DesGluPc <sub>4</sub>	Cys- $\gamma$ -(Glu-Cys) <sub>3</sub> -Gly	875	21.6

<sup>a</sup> Pcs: phytochelatin.

<sup>b</sup> *iso*-Pcs: *iso*-phytochelatin.

Table 2

Intensities of phytochelatin and *iso*-phytochelatin MS signals as a function of cadmium concentrations

Variable ID	Cadmium ( $\mu$ M)			
	0	5	50	200
Pc <sub>2</sub>	4.6e5 $\pm$ 3.1e5 <sup>a</sup>	5.5e6 $\pm$ 1.2e6	1.4e7 $\pm$ 7.8e5	1.1e7 $\pm$ 1.2e6
Pc <sub>3</sub>	1.8e6 $\pm$ 1.6e6	1.2e7 $\pm$ 3.8e6	9.5e7 $\pm$ 1.5e7	1.1e8 $\pm$ 1.6e7
Pc <sub>4</sub>	1.2e6 $\pm$ 1.1e6	1.8e6 $\pm$ 7.2e5	3.4e7 $\pm$ 6.9e6	8.9e7 $\pm$ 1.2e7
Pc <sub>5</sub>	ND	ND	1.3e6 $\pm$ 5.1e5	1.2e7 $\pm$ 3.8e6
<i>Iso</i> -Pc <sub>2</sub> (Ala)	ND	ND	ND	3e5 $\pm$ 1e5
<i>Iso</i> -Pc <sub>3</sub> (Ala)	ND	ND	2.3e6 $\pm$ 5.3e5	9.7e6 $\pm$ 1.5e6
<i>Iso</i> -Pc <sub>4</sub> (Ala)	ND	ND	5.1e5 $\pm$ 8.7e4	5.6e6 $\pm$ 1.1e6
<i>Iso</i> -Pc <sub>3</sub> (Ser)	ND	ND	ND	1.8e6 $\pm$ 2.6e5
<i>Iso</i> -Pc <sub>4</sub> (Ser)	ND	ND	ND	1.4e6 $\pm$ 1.5e5
<i>Iso</i> -Pc <sub>3</sub> (Gln)	ND	ND	ND	1.5e6 $\pm$ 2.3e5
<i>Iso</i> -Pc <sub>4</sub> (Gln)	ND	ND	ND	5.5e5 $\pm$ 1.4e5
<i>Iso</i> -Pc <sub>3</sub> (Glu)	ND	ND	ND	4.3e5 $\pm$ 1.3e5
DesGluPc <sub>3</sub>	ND	3.8e5 $\pm$ 2.1e5	1.1e6 $\pm$ 8.8e5	7.3e5 $\pm$ 5e5
DesGluPc <sub>4</sub>	ND	4.2e5 $\pm$ 1.2e5	2.3e6 $\pm$ 1.7e6	4.4e6 $\pm$ 8.4e5

ND: not detected.

<sup>a</sup> Mean intensity  $\pm$  one standard deviation. Data are expressed in arbitrary units.

All HPLC-MS data were processed using the MetAlign software (Plant Research International) to yield a data matrix (Excel file) containing information such as retention time, mass and normalized peak intensity. Parameters for MetAlign were adjusted as follows: the retention's beginning and end were fixed at 80 and 1600 scan numbers, respectively, and the maximum signal amplitude at 10<sup>9</sup>. The peak slope factor and peak threshold were fixed at "1" and "2", respectively, and the average peak width at half height was set at 20 scans. Of note, we found that this latter parameter had a crucial impact on the final results since low values for these parameters (5 for example) induced some peak slicing, leading to artefactual variables. Alignment of data sets was performed by using the iterative procedure, with a maximum shift of 40 per 100 scans, without any scaling. For the mass peak selection, the minimum factor times noise was fixed at "5" and "2", for the first and the last iteration, respectively, and the minimum number of masses was set at "4" and "3", respectively.

## 2.5. Multivariate statistical analysis

Excel files obtained from pretreated data sets were organized as a matrix where each row corresponds to a sample, each column to a  $m/z$  ratio and each value in the data table the related intensity. They were subsequently exported to the SIMCA software (SIMCA P11, Umetrics, Sweden) for multivariate statistical analysis. PCA (Principal Component Analysis) was performed first, to visualise the data and test for a possible structure, regardless of the cadmium dose. Surprisingly, this descriptive (*i.e.* unsupervised) method revealed that our data were structured according to the cadmium dose. Supervised methods, such as the Partial Least Squares Regression (PLS) were then used to build a model between the various  $m/z$  or couples ( $m/z$ , retention times), considered as independent variables and the cadmium dose, regarded as the dependant variable. The optimal number of factors of the PLS models were selected by SIMCA (autofit procedure).

### 3. Results and discussion

#### 3.1. Description of the metabolomic data

The data set is derived from a previous combined proteomic and metabolomic study aimed at evaluating the early response of *A. thaliana* cells to cadmium exposure [33]. This study was performed on cell extracts by using liquid chromatography coupled to electrospray mass spectrometry (LC/MS). A typical chromatographic profile of a cell extract from the plant *A. thaliana* is shown in Fig. 1. Over a 70 min-time period, several tens of chromatographic peaks are detected. As shown in the inset of Fig. 1, several tens of  $m/z$  ratio can be counted under a single chromatographic peak. By using this procedure, we showed that *A. thaliana* cells are able to synthesize most Pc-related peptides (Pcs and *iso*-Pcs) described in the literature [33]. These peptides, whose structures have been assessed by tandem mass spectrometry, are shown in Table 1, with their corresponding  $m/z$  ratio and chromatographic retention times, ranging from  $m/z$  540 to 1075 and from 10 to 25 min in our analytical conditions, respectively.

The intensities of the Pc-derived peptides are displayed in Table 2. Some peptides such as Pc<sub>2</sub>, Pc<sub>3</sub> and Pc<sub>4</sub> are already detected in the control group, whereas others such as *iso*-Pc (Ser) and *iso*-Pc (Gln) only occur at the highest cadmium concentration (200  $\mu$ M). These metabolites are also present at different abundances. Indeed, assuming that they are ionized in the same way in the electrospray source, the canonical Pcs (*i.e.* ( $\gamma$ Glu-Cys)<sub>*n*</sub>-Gly) are the most abundant species, whereas *iso*-Pcs represent less than 10% of these kinds of peptides. Furthermore, the highest levels are observed at the highest cadmium dose, with increased concentrations as a function of cadmium dose for some peptides (Pc<sub>3</sub>, Pc<sub>4</sub> or desGluPc<sub>4</sub>, for example).

In brief, the 14 metabolites of interest listed in Table 1 have been isolated and formally identified from chromatograms containing few hundreds to several thousands of variables.

These metabolites are present at a wide range of concentration and exhibit different concentration patterns in response to cadmium exposure. We then took the opportunity to work on this data set in order to evaluate the impact of the pretreatment of the data and of different multivariate statistical tools on the classification of the samples and the isolation of discriminatory variables.

#### 3.2. Pretreatment of data

Three kinds of pretreatment methods were used in this study i) the reduction of dimensionality by averaging the intensity of each  $m/z$  ratio over the acquisition time, ii) bucketing by truncating the chromatograms into fractions of 5 min, iii) the use of the software MetAlign. The resulting data were subjected to PLS analysis and the classification of our variables of interest are displayed in Tables 3 (reduced dimensionality), 4 (bucketing) and 5 (MetAlign). Different scaling procedures have also been evaluated: the no scaling, the scaling to unit variance (*i.e.* UV-scaling: the variables are centered and divided by their standard deviation) and the pareto-scaling (the variables are centered and divided by the square root of their standard deviation).

The quality of the clustering (*i.e.* the visual discrimination between the different cadmium-exposed groups in the first principal plan), the rank of the variables (variable significance) and their associated contribution to the models (*i.e.* the VI: variable importance on the projections) are also shown in these tables, for comparison purposes. Of note, a VI value above 1 indicates a significant contribution to the PLS model. All the variables having a VI value above 1 have therefore been studied.

##### 3.2.1. Reduction of dimensionality

Reduction of dimensionality has been achieved by averaging the intensity of each  $m/z$  ratio from 4 to 70 min. Although it induces the loss of information concerning isobaric ions (*i.e.*

Table 3  
PLS models obtained by reducing the dimensionality of the data set

Variable ID	No normalization			Unit variance-scaling			Pareto-scaling		
	Clusters	Rank	VI <sup>a</sup>	Clusters	Rank	VI	Clusters	Rank	VI
Pc <sub>2</sub>		5	4.97		26	2.1		6	5.7
Pc <sub>3</sub>		1	17.7		7	2.4		1	12.2
Pc <sub>4</sub>		2	14.8		3	2.6		2	11.8
Pc <sub>5</sub>		9	3.4		5	2.5		7	5.3
desGluPc <sub>3</sub>		33	1.2		56	1.9		30	2.0
desGluPc <sub>4</sub>		ND			ND			ND	
<i>Iso</i> -Pc <sub>2</sub> (Ala)	0;5/50/200	ND		0;5/50/200	ND		0;5/50/200	ND	
<i>Iso</i> -Pc <sub>3</sub> (Ala)		11	2.0		4	2.5		9	3.9
<i>Iso</i> -Pc <sub>4</sub> (Ala)		21	1.4		14	2.2		13	3.1
<i>Iso</i> -Pc <sub>3</sub> (Ser)		ND			ND			ND	
<i>Iso</i> -Pc <sub>4</sub> (Ser)		ND			ND			ND	
<i>Iso</i> -Pc <sub>3</sub> (Gln)		ND			ND			ND	
<i>Iso</i> -Pc <sub>4</sub> (Gln)		ND			ND			ND	
<i>Iso</i> -Pc <sub>3</sub> (Glu)		ND			ND			ND	

ND: not detected.

<sup>a</sup> Variable importance.

Table 4  
PLS models obtained by using a bucketing procedure

Variable ID	Bracket [min]	No normalization			Unit variance-scaling			Pareto-scaling		
		Clusters	Rank	VI <sup>a</sup>	Clusters	Rank	VI	Clusters	Rank	VI
Pc <sub>2</sub>	[9–14]	0/5/50/200	1	33.8	0;5/50/200	25	2.4	0/5/50/200	1	19.1
Pc <sub>3</sub>	[19–24]	0/5/50/200	1	26.9	0/5/50/200	50	1.8	0/5/50/200	1	15.8
Pc <sub>4</sub>	[19–24]	0/5/50/200	2	19.8	0/5/50/200	4	1.9	0/5/50/200	2	14.9
Pc <sub>5</sub>	[24–29]	0;5/50/200	1	22.5	0;5/50/200	7	2.0	0;5/50/200	1	11.7
desGluPc <sub>3</sub>	[14–19]	0;5/50/200	2	12.8	0;5/50/200	135	1.6	0/5/50/200	1	6.3
desGluPc <sub>4</sub>	[19–24]	0/5/50/200	12	1.3	0/5/50/200	7	1.9	0/5/50/200	11	3.6
Iso-Pc <sub>2</sub> (Ala)	[14–19]	0;5/50/200	5	7.0	0;5/50/200	1	2.8	0;5/50/200	2	6.0
Iso-Pc <sub>3</sub> (Ala)	[19–24]	0/5/50/200	9	2.1	0/5/50/200	8	1.9	0/5/50/200	7	4.8
Iso-Pc <sub>4</sub> (Ala)	[19–24]	0/5/50/200	15	0.9	0/5/50/200	46	1.8	0/5/50/200	12	3.1
	[24–29]	0;5/50/200	5	3.9	0;5/50/200	31	1.8	0/5/50/200	5	4.4
Iso-Pc <sub>3</sub> (Ser)	[19–24]	0/5/50/200	37	0.5	0/5/50/200	35	1.8	0/5/50/200	31	2.1
Iso-Pc <sub>4</sub> (Ser)	[19–24]	0/5/50/200	57	0.3	0/5/50/200	10	1.9	0/5/50/200	41	1.9
Iso-Pc <sub>3</sub> (Gln)	[19–24]	0/5/50/200	45	0.4	0/5/50/200	5	1.9	0/5/50/200	35	2
Iso-Pc <sub>4</sub> (Gln)	[19–24]	0/5/50/200	85	0.2	0/5/50/200	58	1.7	0/5/50/200	61	1.4
Iso-Pc <sub>3</sub> (Glu)	[19–24]	0/5/50/200	42	0.4	0/5/50/200	9	1.9	0;5/50/200	34	2.0

<sup>a</sup> Variable importance.

ions having the same nominal mass), we chose it because it gave us the opportunity to suppress the problem of retention time shifts that may be observed from one chromatogram to the other. This procedure led to a matrix of 19 individuals (rows) containing 1402 variables, *i.e.* *m/z* ratios (columns). The corresponding results of PLS analyses are displayed in Table 3, together with the impact of scaling. The classification of control and cadmium-exposed samples was successful whatever the scaling methods used, leading to 3 clusters: the first one corresponds to the samples from control and 5  $\mu\text{M}$ -cadmium-treated cells, whereas the second and the third one to the samples from the 50  $\mu\text{M}$  and the 200  $\mu\text{M}$ -cadmium-treated cells, respectively.

However, only the most abundant metabolites, *i.e.* the canonical phytochelatins and few *iso*-Pcs such as *iso*-Pc<sub>3</sub> (Ala) and *iso*-Pc<sub>4</sub> (Ala) were detected. This loss of information may be attributed to the high levels of background noise observed at the end of the gradient, in the presence of the highest amount of acetonitrile. As a matter of fact, in these experimental conditions,

the intensity of the background noise is in the same range as that of the variables exhibiting the lowest concentrations. However, this issue could be addressed by using high resolution mass spectrometers and by summing the mass spectra instead of averaging them [20]. According to our experience in the field of LC coupled to ultra-high resolution mass spectrometry (LTQ-Orbitrap, ThermoFisher, Les Ulis France), such a procedure limits the loss of information, although it still occurs, especially in complex biofluids such as urines in which we observed many isomers (data not shown).

### 3.3. Bucketing procedure

Bucketing can be viewed as an improvement of the dimensionality reduction procedure. By truncating the chromatograms into zones of 5 min, we aimed at adopting a compromise solution between the time dedicated to the pretreatment and the loss of information caused by the presence of isobaric

Table 5  
PLS models obtained from LC/MS data processed with MetAlign

Variable ID	No normalization			Unit variance-scaling			Pareto-scaling		
	Clusters	Rank	VI <sup>a</sup>	Clusters	Rank	VI	Clusters	Rank	VI
Pc <sub>2</sub>		8	4.9		204	1.6		11	5.3
Pc <sub>3</sub>		1	36.9		56	2.3		1	20.4
Pc <sub>4</sub>		2	27.2		5	2.76		2	20.0
Pc <sub>5</sub>		9	4.3		19	2.68		7	7.6
desGluPc <sub>3</sub>		97	0.38		1126	0.87		166	1.1
desGluPc <sub>4</sub>		21	1.3		41	2.4		17	4.0
<i>iso</i> -Pc <sub>2</sub> (Ala)	0/5/50/200	684	0.1	0;5/50/200	39	2.4	0;5/50/200	194	1.02
<i>Iso</i> -Pc <sub>3</sub> (Ala)		10	3.2		3	2.8		8	6.8
<i>Iso</i> -Pc <sub>4</sub> (Ala)		15	2.05		8	2.75		14	5.3
<i>Iso</i> -Pc <sub>3</sub> (Ser)		53	0.6		4	2.79		25	2.9
<i>Iso</i> -Pc <sub>4</sub> (Ser)		60	0.5		6	2.76		34	2.6
<i>Iso</i> -Pc <sub>3</sub> (Gln)		62	0.5		11	2.74		33	2.6
<i>Iso</i> -Pc <sub>4</sub> (Gln)		244	0.19		27	2.58		91	1.5
<i>Iso</i> -Pc <sub>3</sub> (Glu)		414	0.13		36	2.49		143	1.24

<sup>a</sup> Variable importance.

compounds, and by high background noise. Some authors reported the use of smaller time intervals, *i.e.* 1 min [11]. Here, such a procedure would be of limited interest to correct for time shifts that may exceed 1 min.

For each zone, a PLS model has been built. The results obtained when using our bucketing procedures are displayed in Table 4. According to their retention time, the variables of interest were displayed in their respective time window. All the phytochelatin-derived peptides are now detected. One of them, the *iso*-Pc<sub>4</sub> (Ala) is present in two brackets since its retention time of 24 min is straddling the time intervals of [19–24 min] and of [24–29 min]. Interestingly, this metabolite does not exhibit a significant contribution to the PLS model obtained with unscaled variables on the [19–24 min] bracket since its variable importance (VI) on the projection is inferior to 1.0, whereas its weight is significant in the next zone of [24–29 min]. As a consequence, this variable has not been lost by the bucketing procedure.

The classification of control and cadmium-exposed samples was successful whatever the scaling methods used, leading to 3 or even 4 clusters: in some cases, the 5 μM cadmium-exposed group was well separated from the controls. The quality of the PLS models was evaluated by three parameters:  $R^2(X)$ , corre-

sponding to the proportion of the variance of the variables that is explained by the model,  $R^2(Y)$ , defining the proportion of the variance of the response variable (*i.e.* the cadmium concentration) explained by the model, and  $Q^2(\text{cum})$ , expressing the cumulative proportion of the variance of the variables that can be predicted by the model.

The [19–24 min] bracket consists in a matrix containing 19 observations and 1396 variables. The values of  $R^2(X)$  were 0.949, 0.336 and 0.446 without any scaling, with UV- and pareto-scaling, respectively. The values of  $R^2(Y)$  were 0.994, 0.995 and 0.993, with unscaled, UV- and pareto-scaled variables, respectively.  $Q^2(\text{cum})$  are 0.972, 0.817 and 0.914 with unscaled, UV- and pareto-scaled variables, respectively. The same range of values was obtained for  $R^2(X)$  and  $R^2(Y)$  with the other three brackets: [9–14 min], [14–19 min] and [24–29 min], which contain 1401, 1396 and 1396 variables, respectively. For the first two brackets the values of  $Q^2(\text{cum})$  were lower and ranged from 0.59 to 0.87.

The values of  $R^2(Y)$  indicate that most of the variance related to our response variable, *i.e.* cadmium exposure, is explained by the PLS models. Furthermore, the values of  $R^2(X)$  that have been obtained for UV-scaling are lower than those related to non scaled and pareto-scaled variables, suggesting that a weaker

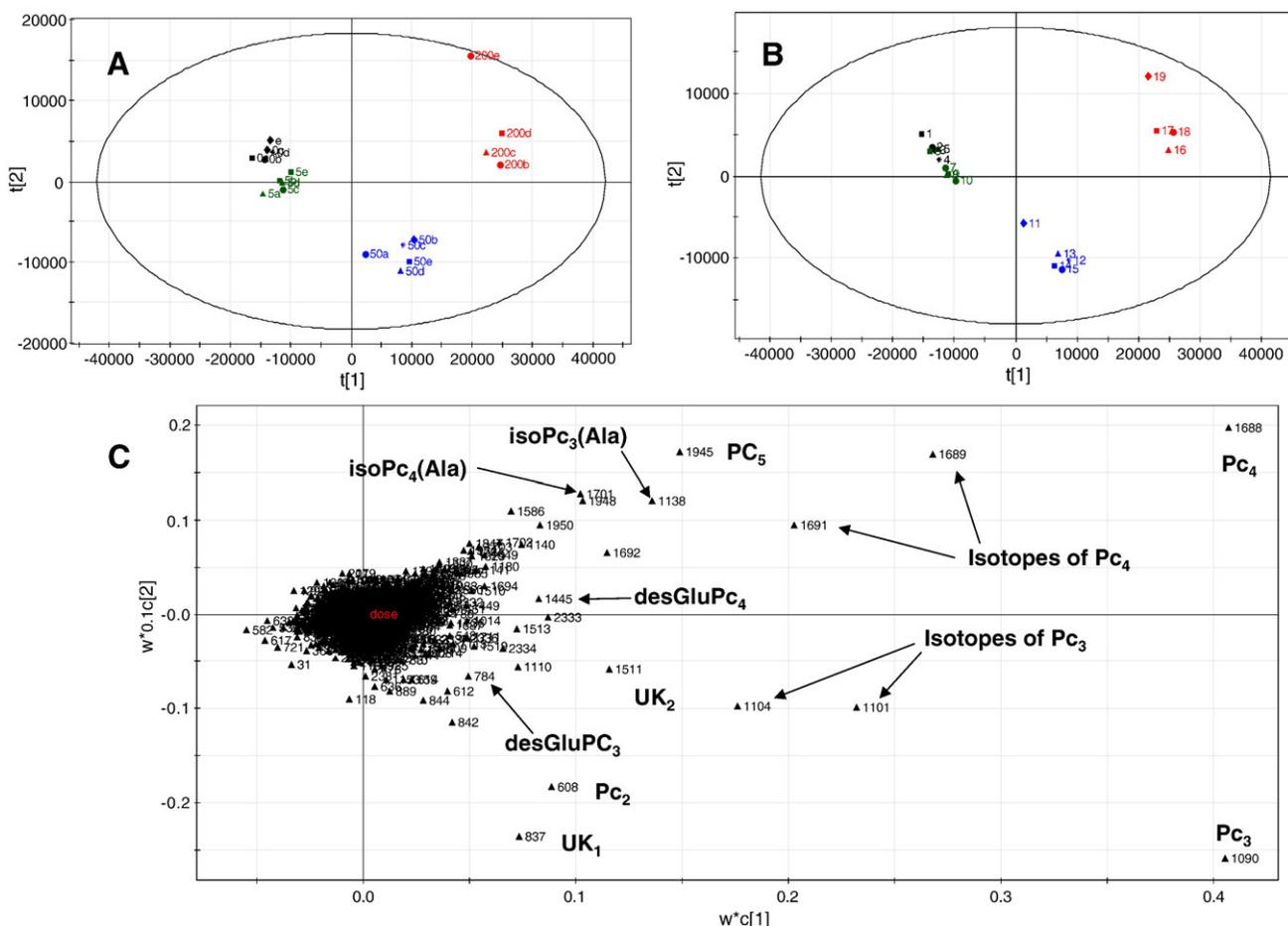


Fig. 2. Multivariate statistical analyses resulting from the dose ranging experiment. The treatment with MetAlign resulted in 2639 variables. A: PCA scores plot. The control and Cd-treated samples are clearly distinguished on the first component which account for 37% of the total variance. The second component, which account for 7% of the total variance is related to inter-individual differences. B: PLS score plot and C: PLS loadings plot.

proportion of variance related to the variables is explained by the PLS models in this scaling condition. This is reflected by the ranks observed with this scaling procedure for our metabolites of interest (Table 4). For example,  $Pc_2$ ,  $Pc_3$  and  $desGluPc_3$ , which are classified in the first or second position in their related brackets without any scaling or in pareto-scaled conditions, occurs at the 25th, 50th and 135th positions in the UV-scaled data sets, respectively. As a consequence, by giving an equal weight to all variables, UV-scaling induced a dilution of the analytical information of biological relevance.

However, UV-, but also pareto-scaling allowed to isolate variables present at low concentrations, such as *iso*-Pc(Ser) and *iso*-Pc(Gln) that would not have been taken into account in the unscaled data set. In conclusion for this data set, pareto-scaling appears to be a good compromise. Of note, in some cases (urine samples for example, data not shown), we obtained optimal results with UV-scaling by using the same analytical methodology, thus indicating that the choice of a scaling method is context dependent.

### 3.3.1. Use of the MetAlign software

MetAlign treated data are displayed as a matrix (Excel file) containing the variables (*i.e.* couples of  $m/z$  ratios and retention times) and their intensities for each individual. Treatment of the data set by MetAlign resulted in 2639 variables. Contrary to the data sets obtained after reduction of dimensionality or bucketing, the variables are now expressed as couples of  $m/z$  and retention times. As described in the last section about the bucketing procedure, the quality of the PLS models have been evaluated by

comparing the values of  $R^2(X)$ ,  $R^2(Y)$  and  $Q^2(\text{Cum})$  for the different kinds of scalings. The values of  $R^2(X)$  were of 0.973, 0.191 and 0.439 for unscaled, UV- and pareto-scaled variables, respectively. The values of  $R^2(Y)$  were of 0.981, 0.646 and 0.995 for unscaled, UV- and pareto-scaled variables, respectively.  $Q^2(\text{cum})$  were of 0.972, 0.171 and 0.953 for unscaled, UV- and pareto-scaled variables, respectively. These values are closed to that observed with the bucketing procedure: a weak proportion of variance related to the variables is explained by the PLS models in UV-scaling conditions.

The results of PLS analyses obtained with LC/MS data processed by MetAlign are displayed in Table 5. As previously observed with the other two methods of data pretreatment, the classification of control and cadmium-exposed samples was successful whatever the scaling method, leading to 3 or 4 clusters, each one being related to a given cadmium concentration.

Concerning the impact of the scaling on the classification of the variables of biological interest, the same conclusions as those stated for the bucketing procedure can be drawn: a weak proportion of variance related to the variables is explained by the PLS models in UV-scaling condition. However, this kind of scaling allowed the less abundant variables to be taken into account. Here again, pareto-scaling produces the most exploitable results since all the metabolites of interests exhibit significant contributions to the resulting PLS model.

The use of an automatic method for the pretreatment of LC/MS data presents the main advantage of time saving, since a single statistical analysis is performed instead of one for each bucket. Its use is also simple. On the other hand, some

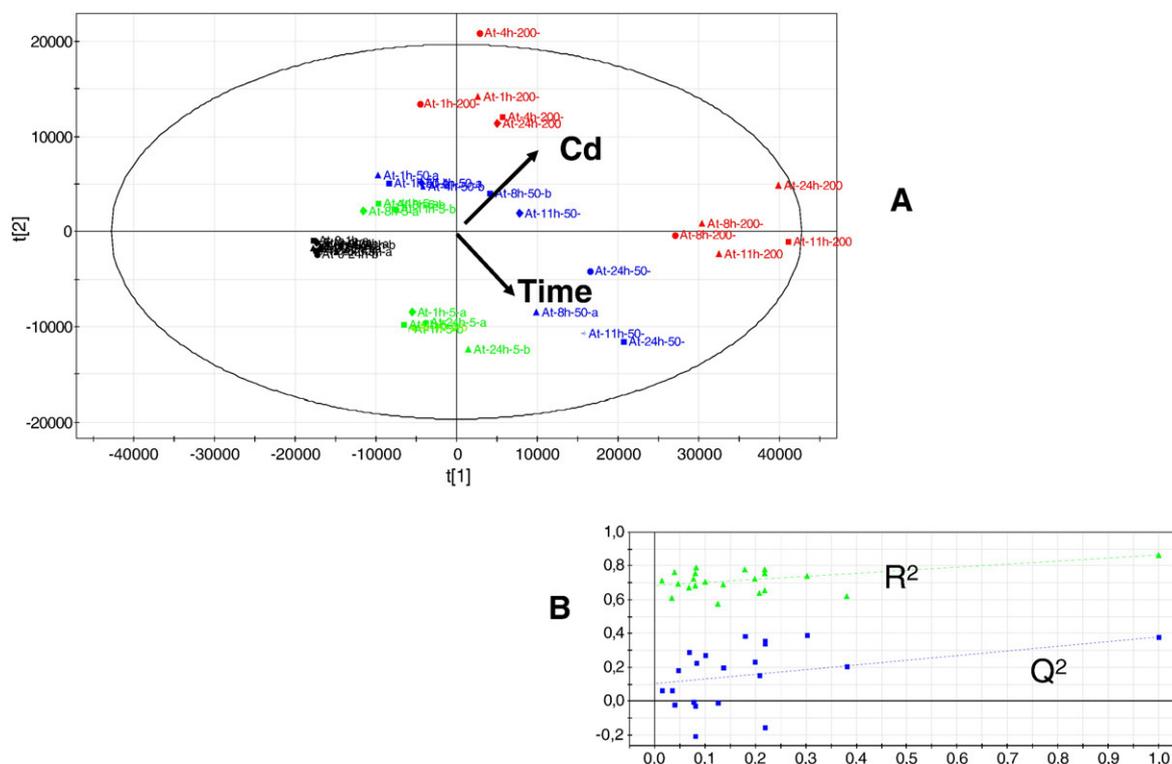


Fig. 3. PLS2 scores plot resulting from the analysis of the kinetic experiment A. and validation by permutation test B. The data set comes from a range dose (0, 5, 50 and 200  $\mu\text{M}$  Cd) and kinetic (1, 4, 8, 11 and 24 h) study. The treatment of the data set by MetAlign resulted in 3666 variables for 40 observations.  $R^2(X)=0.283$ ,  $R^2(Y)=0.801$  and  $Q^2(\text{Cum})=0.348$ . The model has been validated by a permutation test ( $n=20$ ).  $R^2$  intercept is at 0.682 and  $Q^2$  intercept is at 0.104.

chromatographic peaks exhibiting an asymmetric shape may be incidentally discarded or may lead to several artefactual variables instead of a single one. To address this issue, it is crucial to perform a test baseline correction by comparing some processed chromatograms to the original data [35]. No such instance was detected with the present data set.

To sum up, the reduction of dimensionality by collapsing the time axis clearly resulted in the loss of information, whereas the MetAlign program led to the same results as those obtained by using the bucketing procedure in terms of clustering and of classification of the variables.

#### 3.4. Multivariate statistical analysis of the response of *A. thaliana* cells to cadmium exposure from a dose ranging study

A metabolomic analysis was performed on extracts from *A. thaliana* cells that have been treated for 24 h with different cadmium concentrations (0–5–50–200  $\mu\text{M}$ , 4–5 independent experiments per condition). PCA was performed on the MetAlign pretreated data set and the variables were pareto-scaled, as described in Section 3. PCA provided an overview of the way the data set was structured. Fig. 2A illustrates the PCA

scores plot that was generated, where the ellipse marks the 95% Hotelling  $T^2$  control chart. No observation statistically lies outside this critical limit, indicating the absence of outliers. The scores plot shows that the first principal component has the ability to clearly distinguish the control from the treated groups (5, 50 and 200  $\mu\text{M}$  Cd). Most of the variation is then related to class difference. The first and second principal components account for 37% and 7% of the total variance, respectively. The scores plot illustrates the strong differences existing between the samples, already leading to a clear separation of the control from the treated groups without using supervised techniques.

The data were then subjected to supervised multivariate statistical analyses by using PLS regression. The scores plot obtained with the model is shown in Fig. 2B. Four clusters corresponding to the different groups are observed. The characteristics of this multivariate models were as follow:  $R^2(X)$ ,  $R^2(Y)$  and  $Q^2(\text{Cum})$  of 0.439, 0.995 and 0.953, respectively. Of note, the value of  $R^2(X)$  is quite low. As a matter of fact, this model has been built with 10% of the variables (*i.e.* those exhibiting a VIP score superior to 1). It would probably be more efficient for classification purposes to build a model on informative signals which could be selected

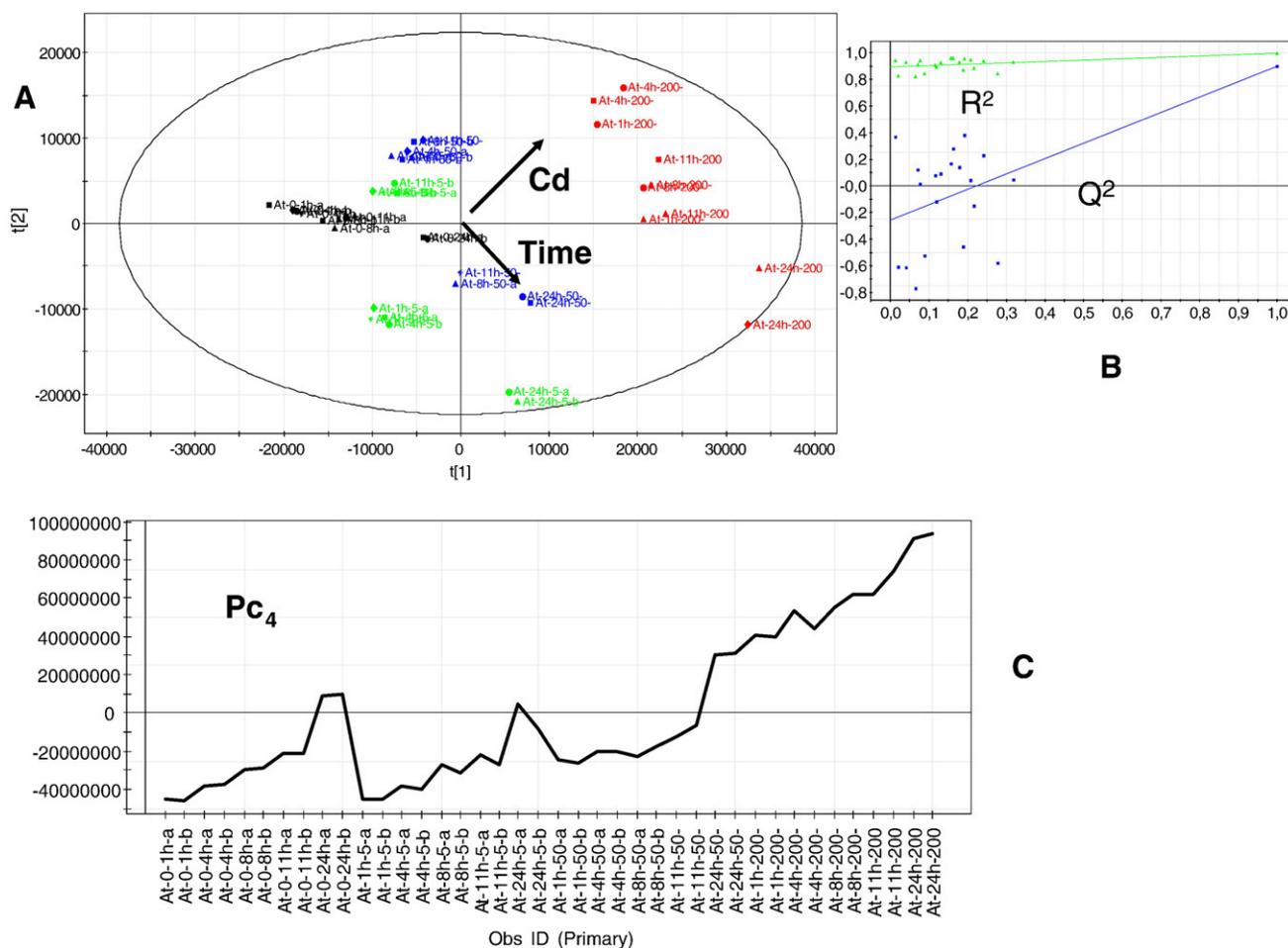


Fig. 4. OSC-PLS2 scores plot resulting from the analysis of the kinetic experiment A., validation by permutation test B., and typical variable line plot C. The data set comes from a range dose (0, 5, 50 and 200  $\mu\text{M}$  Cd) and kinetic (1, 4, 8, 11 and 24 h) study. The treatment of the data set by MetAlign resulted in 3666 variables for 40 observations.  $R^2(X)=0.473$ ,  $R^2(Y)=0.995$  and  $Q^2(\text{Cum})=0.897$ . The model has been validated by a permutation test ( $n=20$ ).  $R^2$  intercept is at 0.893 and  $Q^2$  intercept is at  $-0.258$ .

by an ANOVA filter as it has recently been described [20]. However, our objective was rather to isolate these variables than to develop a model for sample classification.

Furthermore, using PLS regression alone could be dangerous in the sense that it may hide a non monotonous relation between the cause and the effect (here between cadmium dose and expression of phytochelatins). In this context, the use of PLS-DA, in which the dependent variable ( $Y$ ) is not a continuous, but a categorical variable related to the cadmium dose could be useful to detect some non monotonous relation between dose and effect. However, comparative results between PLS and PLS-DA clearly showed that the latter tool added little extra value (data not shown).

Fig. 2C displays the corresponding loadings plot which enables to visualize the variables that contribute to the clustering observed in the scores plot: the canonical Pcs ( $Pc_2$ ,  $Pc_3$ ,  $Pc_4$  and  $Pc_5$ ), and also *iso*- $Pc(Ala)_3$ , *iso*- $Pc(Ala)_4$ , *desGluPc*<sub>3</sub> and *desGluPc*<sub>4</sub>. Other discriminating variables correspond to isotopes of the formers. Furthermore, some variables such as UK 1 and UK 2 are new *Pc*-related peptides whose determination of the formal identity is under progress in our laboratory by performing complementary mass spectrometric experiments, whereas other non annotated variables remain unknown. These results emphasize the interest of metabolomics for the characterization of new metabolites of biological or toxicological relevance.

### 3.5. Multivariate statistical analysis of the response of *A. thaliana* cells to cadmium exposure from a kinetic study

The multivariate statistical analysis of metabolic fingerprints allows visualizing time- and toxic-induced changes on the metabolism of biological organisms. Such an approach has been pioneered by NMR studies of the biochemical effects of hepatotoxins on biochemical composition of rat biofluids [36]. The 3 investigated toxins gave unique metabolic trajectories that were visualized on two-dimensional PCA maps.

We here report the global analysis of a data set consisting in a range of dose (0, 5, 50 and 200  $\mu$ M  $CdNO_3$ ) and kinetic (1, 4, 8, 11 and 24 h) studies acquired in duplicate by using PLS2-related multivariate methods, the two response variables being the time and the cadmium concentration. The treatment of the data set by MetAlign resulted in 3666 variables for 40 observations. The resulting PLS2 was characterized by  $R^2(X)$  and  $Q^2(Cum)$  values of 0.283 and 0.348, respectively, indicating that a weak proportion of the variance of the variables was explained by this model and suggesting a poorer prediction power. The related scores plot is displayed in Fig. 3A. The dose effect is clearly visualized whereas the time effect cannot be detected as easily.

This model was improved by using orthogonal signal correction (OSC) before PLS2. OSC is a data filtering method that minimizes the influence of background noise present in the mass spectra [37], thus leading to sharper separation between the variables affected by Cd dose and time, and those that were not. By this means, the proportion of the explained and predicted variances of the variables ( $R^2(X)$  of 0.473 and  $Q^2(Cum)$  of 0.897) were higher than that obtained with PLS2. Furthermore,

the OSC-PLS2 model was validated by performing permutations tests ( $n=20$ ) (Fig. 4B), which gave improved results when compared to those obtained in PLS2 (Fig. 3B). Of note, the variables of biological interest that have been pointed out by multivariate statistical analyses have been confirmed by at least 5 different and independent experiments.

The scores plot of the OSC-PLS2 model is displayed in the Fig. 4A. When compared to Fig. 3A, the effect of time is now observed on the first axis. The groups corresponding to the 24 hours-time points at the 5 and 50  $\mu$ M doses are clearly individualized on the right of the scores plot. This can be explained by the late and abundant production of *Pc*-related peptides in *A. thaliana* cells at these cadmium doses, as it is observed on the typical variable line plot displayed for  $Pc_4$  in the Fig. 4C. Finally, although the experiments presented in this article deal with a cell model, we also applied this metabolomic approach to the investigation of cadmium toxicity on *A. thaliana* leaf and root extracts, with analytical performances similar to those observed with cell extracts: the CVs of the chromatographic peaks our two reference compounds (*i.e.* colchicine and prednisone) were below 20%. The results were close to that observed on the cell model since many *iso*-phytochelatins were detected in root extracts.

## 4. Conclusion

It is no surprise that this work emphasizes the impact of the preparation of LC/MS data before multivariate statistical analysis for metabolomics. Among the three ways of pretreating the data, binning and automatic processing gave satisfactory and similar results in extracting the analytical information of biological interest, whereas reducing the dimensionality resulted in a significant loss of discriminating variables in our analytical conditions (*i.e.* low resolution mass spectrometry). The effect of data scaling was also evaluated and pareto-scaling was found to be the most appropriate for the studied data set. However, the nature of information contained in metabolic fingerprints is very different from one biological medium to another, especially in terms of chemical structures and of metabolites concentrations. This precludes the development of standardized procedures and makes the pretreatment of data depending on the biological context and on analytical methodology. The pretreated data were then subjected to multivariate statistical analysis by principal component analysis (PCA), partial least square (PLS) and partial least square discriminant analysis (PLS-DA) for classification. These two PLS methods provided almost similar results for the classification of samples and for the extraction of discriminating variables. An OSC-PLS2 approach was also performed on kinetic and dose ranging studies, allowing visualizing time- and cadmium dose-induced changes on the metabolism of *A. thaliana* cells.

## Acknowledgements

This work was funded by the “Programme CEA de Toxicologie Nucléaire” and by the “Programme inter-organismes

CEA-CNRS-INRA-INSERM de Toxicologie Nucléaire Environnementale”.

## References

- [1] O. Fiehn, *Compar. Funct. Genom.* 2 (2001) 155–168.
- [2] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, *Nat. Rev. Drug Discov.* 1 (2002) 153–161.
- [3] O. Fiehn, *Plant Mol. Biol.* 48 (2002) 155–171.
- [4] L.W. Sumner, P. Mendes, R.A. Dixon, *Phytochemistry* 62 (2003) 817–836.
- [5] J.K. Nicholson, J.C. Lindon, E. Holmes, *Xenobiotica* 29 (1999) 1181–1189.
- [6] J.K. Nicholson, I.D. Wilson, *Prog. Nucl. Magn. Reson. Spectrosc.* 21 (1989) 449–501.
- [7] H.E. Johnson, D. Broadhurst, D.B. Kell, M.K. Theodorou, R.J. Merry, G.W. Griffith, *Appl. Environ. Microbiol.* 70 (2004) 1583–1592.
- [8] M.E. Dumas, L. Debrauwer, L. Beyet, D. Lesage, F. Andre, A. Paris, J.C. Tabet, *Anal. Chem.* 74 (2002) 5393–5404.
- [9] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, *Nat. Biotechnol.* 18 (2000) 1157–1161.
- [10] R. Goodacre, E.V. York, J.K. Heald, I.M. Scott, *Phytochemistry* 62 (2003) 859–863.
- [11] R.S. Plumb, C.L. Stumpf, M.V. Gorenstein, J.M. Castro-Perez, G.J. Dear, M. Anthony, B.C. Sweatman, S.C. Connor, J.N. Haselden, *J. Rapid Commun. Mass Spectrom.* 16 (2002) 1991–1996.
- [12] H. Idborg-Bjorkman, P.O. Edlund, O.M. Kvalheim, I. Schuppe-Koistinen, S.P. Jacobsson, *Anal. Chem.* 75 (2003) 4784–4792.
- [13] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, J.C. Tabet, E. Ezan, *Rapid Commun. Mass Spectrom.* 17 (2003) 2541–2549.
- [14] K. Pearson, *Philos. Mag.* (1901) 559–572.
- [15] H. Hotelling, *J. Educ. Psychol.* 24 (1933) 417–441.
- [16] S. Wold, K.-G. Joreskog, *Soft Modeling. The Basic Design and Some Extensions*, 1982.
- [17] M. Sjöström, S. Wold, B. Söderström, in: E.S. Gelsema, L.N. Kanal (Eds.), *Pattern Recognition in Practice II*, Elsevier, Amsterdam, 1986, p. 486.
- [18] O. Cloarec, M.E. Dumas, J. Trygg, A. Craig, R.H. Barton, J.C. Lindon, J.K. Nicholson, E. Holmes, *Anal. Chem.* 77 (2005) 517–526.
- [19] H. Idborg, L. Zamani, P.O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, *J. Chromatogr., B, Analyt. Technol. Biomed. Life Sci.* 828 (2005) 14–20.
- [20] J. Boccard, E. Grata, A. Thiocone, J.-Y. Gaurvit, P. Lantéri, P.-A. Carrupt, J.-L. Wolfender, S. Rudaz, *Chemometr. Intell. Lab. Syst.* 86 (2007) 189–197.
- [21] V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, *Anal. Chem.* 75 (2003) 6737–6740.
- [22] M. Katajamaa, M. Oresic, *BMC Bioinformatics* 6 (2005) 179.
- [23] C.A. Smith, E.J. Want, G. O’maille, R. Abagyan, G. Siuzdak, *Anal. Chem.* 78 (2006) 779–787.
- [24] R.S. Plumb, C.L. Stumpf, J.H. Granger, J. Castro-Perez, J.N. Haselden, G.J. Dear, *Rapid Commun. Mass Spectrom.* 17 (2003) 2632–2638.
- [25] A. Craig, O. Cloarec, E. Holmes, J.K. Nicholson, J.C. Lindon, *Anal. Chem.* 78 (2006) 2262–2267.
- [26] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, *BMC Genomics* 7 (2006) 142.
- [27] L. Sanita di Topi, R. Gabbrielli, *Environ. Exp. Bot.* 4 (1999) 105–130.
- [28] J.L. Hall, *J. Exp. Bot.* 53 (2002) 1–11.
- [29] C.S. Cobbett, *Curr. Opin. Plant Biol.* 3 (2000) 211–216.
- [30] S. Clemens, E.J. Kim, D. Neumann, J.I. Schroeder, *EMBO J.* 18 (1999) 3325–3333.
- [31] S.B. Ha, A.P. Smith, R. Howden, W.M. Dietrich, S. Bugg, M.J. O’Connell, P.B. Goldsbrough, C.S. Cobbett, *Plant Cell* 11 (1999) 1153–1164.
- [32] O.K. Vatamaniuk, S. Mari, Y.P. Lu, P.A. Rea, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 7110–7115.
- [33] J.E. Sarry, L. Kuhn, C. Ducruix, A. Lafaye, C. Junot, V. Hugouvieux, A. Jourdain, O. Bastien, D. Vailhen, B. Amekraz, C. Moulin, E. Ezan, J. Garin, *J. Bourguignon, Proteomics* 6 (2006) 2180–2198.
- [34] C. Ducruix, C. Junot, J. Fiévet, F. Villiers, E. Ezan, J. Bourguignon, *Biochimie* 88 (2006) 1733–1742.
- [35] R.C.H. De Vos, S. Moco, A. Lommen, J.J.B. Keurentjes, R.J. Bino, R. Hall, *Nat. Protoc.* 2 (2007) 778–791.
- [36] B.M. Beckwith-Hall, J.K. Nicholson, A.W. Nicholls, P.J. Foxall, J.C. Lindon, S.C. Connor, M. Abdi, J. Connelly, E. Holmes, *Chem. Res. Toxicol.* 4 (1998) 260–272.
- [37] B.M. Beckwith-Hall, J.T. Brindle, R.H. Barton, M. Coen, E. Holmes, J.K. Nicholson, H. Antti, *Analyst* 127 (2002) 1283–1288.



Werner E, Croixmarie V, Umbdenstock T, Ezan E, Chaminade P, Tabet JC et Junot C.

**Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution.**

*Anal Chem* 2008; 80(13):4918-4932.



# Mass Spectrometry-Based Metabolomics: Accelerating the Characterization of Discriminating Signals by Combining Statistical Correlations and Ultrahigh Resolution

Erwan Werner,<sup>†</sup> Vincent Croixmarie,<sup>‡</sup> Thierry Umbdenstock,<sup>‡</sup> Eric Ezan,<sup>†</sup> Pierre Chaminade,<sup>§</sup> Jean-Claude Tabet,<sup>||</sup> and Christophe Junot<sup>\*,†</sup>

CEA, DSV/iBiTec-S/SPI/LEMM, Batiment 136, 91191 Gif-sur-Yvette Cedex, France, Technologie Servier, 25/27 rue Eugène Vignat, 45000 Orléans, France, Groupe de chimie analytique de paris sud—E.A. 4041, Faculté de pharmacie de Châtenay-Malabry, Université Paris XI, 5 rue J.B. Clément, 92296 Châtenay-Malabry cedex, France, Laboratoire de Chimie Structurale Organique et Biologique, UMR 7613-CNRS, Université Paris 6, 4 place Jussieu, 75252 Paris cedex 5, France

A strategy combining autocorrelation matrices and ultrahigh resolution mass spectrometry (MS) was developed to optimize the characterization of discriminating ions highlighted by metabolomics. As an example, urine samples from rats treated with phenobarbital (PB) were analyzed by ultrahigh-pressure chromatography with two different eluting conditions coupled to time-of-flight mass spectrometric detection in both the positive and negative electrospray ionization modes. Multivariate data analyses were performed to highlight discriminating variables from several thousand detected signals: a few hundred signals were found to be affected by PB, whereas a few tenths of them were linked to its metabolism. Autocorrelation matrices were then applied to eliminate adduct and fragment ions. Finally, the characterization of the ions of interest was performed with ultrahigh-resolution mass spectrometry and sequential MS<sup>n</sup> experiments, by using a LC-LTQ-Orbitrap system. The use of different eluting conditions was shown to drastically impact on the chromatographic retention and ionization of compounds, thus providing a way to obtain more exhaustive metabolic fingerprints, whereas autocorrelation matrices allowed one to focus the identification work on the most relevant ions. By using such an approach, 14 PB metabolites were characterized in rat urines, some of which have not been reported in the literature.

Over the past few years, “omics tools” (genomics, transcriptomics, proteomics, metabolomics) have been developed to clarify living organisms’ responses to a perturbation, of whatever source. The terms metabolomics refer to the untargeted quantitative or semiquantitative analysis of the metabolome (set of all metabolites present in a biological system) by differential analysis of metabolite

patterns in different experimental groups.<sup>1–3</sup> Once the metabolic fingerprints have been acquired, they are compared using multivariate statistical analysis<sup>4,5</sup> for discriminating analysis purposes. Multivariate analysis (MVA) visualizes sample distribution and highlights any signals that discriminate between experimental groups.

Almost every analytical technique can be employed for the acquisition of metabolic fingerprints, but nuclear magnetic resonance (NMR) and mass spectrometry (MS)-based techniques have become established as the major tools for metabolomics because of their reliability and selectivity, and also because they identify the metabolites of interest. Initial studies relied on NMR, but despite recent technological developments such as magic angle spinning<sup>6</sup> and cryogenic sample probes,<sup>7,8</sup> the sensitivity and dynamic range of NMR-based techniques are still not as good as those of MS-based methods. As a result, MS-based techniques rapidly emerged as powerful alternatives to NMR for metabolome analysis.

Because of the chemical and physical diversity of the compounds present in a complex biofluid such as urine or plasma, MS detection is classically coupled with a separation technique to maximize the number of metabolites detected. Gas chromatography with electron ionization (GC/EIMS) was first used for metabolic profiling. It provides high chromatographic resolution, good analyte selectivity, and quantification and identification of metabolites thanks to the existing comprehensive EI-mass spectra databases. However, the necessary thermal stability of the investigated compounds, the required derivatization of nonvolatile

- (1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–71.
- (2) Villas-Boas, S. G.; Rasmussen, S.; Lane, G. A. *Trends Biotechnol.* **2005**, *23*, 385–6.
- (3) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–9.
- (4) Holmes, E.; Antti, H. *Analyst* **2002**, *127*, 1549–57.
- (5) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–79.
- (6) Wang, Y.; Bollard, M. E.; Keun, H.; Antti, H.; Beckonert, O.; Ebbels, T. M.; Lindon, J. C.; Holmes, E.; Tang, H.; Nicholson, J. K. *Anal. Biochem.* **2003**, *323*, 26–32.
- (7) Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2002**, *74*, 4588–93.
- (8) Spraul, M.; Freund, A. S.; Nast, R. E.; Withers, R. S.; Maas, W. E.; Corcoran, O. *Anal. Chem.* **2003**, *75*, 1536–41.

\* To whom correspondence should be addressed. Phone: +33169084366. Fax: +33169085907. E-mail: christophe.junot@cea.fr.

<sup>†</sup> CEA, DSV/iBiTec-S/SPI/LEMM.

<sup>‡</sup> Technologie Servier.

<sup>§</sup> Université Paris XI.

<sup>||</sup> Université Paris 6.

compounds prior to analysis, and the frequent absence of molecular ions have made liquid chromatography (LC) coupled to atmospheric-pressure ionization mass spectrometry a more and more widely used tool for metabolomics. It classically relies on a separation on reversed-phase (RP) columns, followed by a detection based on electrospray ionization mass spectrometry (ESI-MS). LC/MS is able to separate and detect thousands of compounds with good reproducibility and sensitivity over a wide dynamic range. However, conventional RP-HPLC is often insufficient for a well-resolved separation and comprehensive analysis of complex biological samples. Recently, implementation of ultraperformance liquid chromatography (UPLC) improved chromatographic resolution, peak capacity, and even sensitivity, which is of great value in the context of metabolomics.<sup>9,10</sup>

Another milestone was reached in terms of technological development with the introduction of high-resolution mass spectrometers such as time-of-flight (TOF) and Fourier transform (FT) mass spectrometers. So far, UPLC-TOF/MS has been employed successfully in metabolomics for the detection and characterization of the metabolites of several drugs and natural compounds,<sup>11,12</sup> lipid analysis,<sup>13</sup> and gender or pathology classification.<sup>10</sup> It has spread thanks to good chromatographic separations, full-scan sensitivity, and accurate mass measurement. However, while such a technique is very effective for metabolic profile acquisitions because of a low acquisition cycle duration (from 0.1 to 0.3 s), its averaged resolution of ~5 ppm is a limitation for metabolite characterization.<sup>14</sup> An alternative way of achieving better mass accuracy to facilitate metabolite characterization would be to work with FTMS. Its use has been reported for phenotyping of plant extracts using direct infusion-Fourier transform ion cyclotron resonance mass spectrometry (FTICR)<sup>15</sup> and drug metabolism studies using LC-LTQ-Orbitrap.<sup>16</sup> However, accurate mass measurement is achieved at the expense of longer cycle duration, especially with FTICR, limiting the association of ultrahigh resolution with UPLC, whereas combination with HPLC has already been used for proteomics<sup>17</sup> and metabolite identification.<sup>16</sup>

Despite these technological advances, metabolomic approaches still need to overcome several problems. First, it is impossible to detect all metabolites by a single and universal analytical method. The signal extraction step (i.e., mass spectral background noise suppression and correction of retention time shifts occurring from one chromatogram to the other) must also be improved. It can

rely either on manual procedures<sup>18–20</sup> or on dedicated softwares.<sup>21</sup> Although the latter approach seems to be the most attractive, some metabolites are present in trace amounts or ionize poorly and so the resulting low-intensity signals are missed. Some artifactual signals can also be generated due to erratic background noise suppression or to insufficient correction from retention time shifts. Finally, characterization of discriminating signals is a very difficult and time-consuming step.

The present study focuses on this last point and deals with the different reasons that make the identification of discriminating signals a bottleneck of the metabolomic approach. The first one is a direct consequence of the improvements in sensitivity and mass accuracy, which allow the detection of an increased number of signals, leading to hundreds of compounds to be identified after MVA. Moreover, a metabolite does not produce a single  $m/z$  peak: adduct and product ions can be produced during the desolvation step following ionization/desorption aggregate formation. Although informative, such signal redundancy slows down the identification procedure. Lastly, profiling experiments are often performed with TOF mass spectrometers that can be coupled to UPLC but that are unsuitable for further CID experiments. Of course, Q-TOF instruments enable MS/MS experiments but do not allow sequential MS<sup>n</sup> experiments and, therefore, ion filiation determinations that are very useful for compound identification. As a result, another device may be required for MS/MS or MS<sup>n</sup> experiments to get the additional structural information needed for metabolite identification.

In this context, the present work was designed to provide a suitable strategy for highlighting and characterizing discriminating metabolites generated by LC/MS-based metabolomics. The proof of concept was performed on urine from phenobarbital (PB)-treated rats. PB is of interest for metabolomics because it exerts pleiotropic pharmacological and toxicological effects. It is also metabolized<sup>22,23</sup> and is known to be a cytochrome P450 inducer in rat.<sup>24,25</sup> UPLC/TOFMS experiments were performed using two different eluting conditions and the detection was performed in both the positive and negative ESI modes, with the aim of improving the recovery of metabolites. MVAs were then performed to highlight discriminating variables. Before identification of the ions of biological relevance, the signal redundancy due to the presence of adduct and product ions was removed by using a mathematical tool based on autocorrelation matrices. The characterization of the PB-related ions was finally achieved by performing MS<sup>2</sup> and MS<sup>3</sup> experiments on a LC-LTQ-Orbitrap mass spectrometer. Behind the linear ion trap (LTQ, Thermo-Fisher), the Orbitrap cell is based on a new type of mass analyzer designed

(9) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *817*, 67–76.

(10) Plumb, R. S.; Granger, J. H.; Stumpf, C. L.; Johnson, K. A.; Smith, B. W.; Gaultz, S.; Wilson, I. D.; Castro-Perez, J. *Analyst* **2005**, *130*, 844–9.

(11) Giri, S.; Idle, J. R.; Chen, C.; Zabriskie, T. M.; Krausz, K. W.; Gonzalez, F. J. *Chem. Res. Toxicol.* **2006**, *19*, 818–27.

(12) Chen, C.; Ma, X.; Malfatti, M. A.; Krausz, K. W.; Kimura, S.; Felton, J. S.; Idle, J. R.; Gonzalez, F. J. *Chem. Res. Toxicol.* **2007**, *20*, 531–42.

(13) Rainville, P. D.; Stumpf, C. L.; Shockor, J. P.; Plumb, R. S.; Nicholson, J. K. *J. Proteome Res.* **2007**, *6*, 552–8.

(14) Sleno, L.; Volmer, D. A.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 183–98.

(15) Aharoni, A.; Ried, V.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenow, D. B. *OMICS* **2002**, *6*, 217–34.

(16) Peterman, S. M.; Duczak, N.; Kalgutkar, A. S.; Lame, M. E.; Soglia, J. R. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 363–75.

(17) Meng, F.; Wiener, M. C.; Sachs, J. R.; Burns, C.; Verma, P.; Paweletz, C. P.; Mazur, M. T.; Deyanova, E. G.; Yates, N. A.; Hendrickson, R. C. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 226–33.

(18) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991–6.

(19) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *828*, 14–20.

(20) Ducruix, C.; Vailhen, D.; Werner, E.; Fievet, J. B.; Bourguignon, J.; Tabet, J. C.; Ezan, E.; Junot, C. *Chemom. Intell. Lab. Syst. In press.*

(21) Katajamaa, M.; Oresic, M. *J. Chromatogr., A* **2007**, *1158*, 318–28.

(22) Harvey, D. L.; Glazener, L.; Stratton, L. *Res. Commun. Chem. Pathol. Pharmacol.* **1972**, *3*, 557–65.

(23) Tang, B. K.; Kalow, W.; Grey, A. A. *Drug Metab. Dispos.* **1979**, *7*, 315–8.

(24) Handschin, C.; Meyer, U. A. *Pharmacol. Rev.* **2003**, *55*, 649–73.

(25) Pustynnyak, V. O.; Gulyaeva, L. F.; Lyakhovich, V. V. *Toxicology* **2005**, *216*, 147–53.

by Makarov and co-workers.<sup>26</sup> As an ingeniously modified Kingdon ion trap, it operates by radially trapping ions about a central spindle electrode and generates a transient ion signal related to the axial ion motion rather than to the radial detection used for FTICR. It achieves high mass resolution (up to 100 000, as expressed as full width at half-maximum peak height, i.e. fwhm), high mass accuracy (below 5 ppm with external calibration), a mass/charge range of 6000, and a dynamic range greater than 3 orders of magnitude. Moreover, it allows MS<sup>n</sup> experiments to be performed with a large space charge capacity due to the linear ion trap, which is the front part of this device.<sup>27–29</sup>

## EXPERIMENTAL SECTION

**Chemicals.** Acetonitrile, ammonium acetate, phenobarbital, and standards, used for tuning the Orbitrap mass spectrometer (i.e., testosterone glucuronide, aldosterone, DHEA sulfate, and adipic acid), were from Sigma (Saint Quentin Fallavier, France). Methanol and formic acid were from Merck (Briare-le-canal, France). Water was deionized and filtered through a Millipore Milli-Q water purification system. The standard mixtures used for the external calibration of the LTQ-Orbitrap instrument (Calmix-positive and Calmix-negative, for both the positive and negative ionization modes, respectively) were from ThermoFisher Scientific (Les Ulis, France). Metabonomics MS system test mix used as reference during UPLC-TOF fingerprint acquisition was purchased from Waters.

**Animals and Treatments.** Forty Wistar male rats (SPF Glx/BRL/Han) from Charles River Laboratories France (Domaine des Oncins, 69210 Saint-Germain-sur-l'Arbresle, France) were acclimated to housing conditions for a period of one week. The animals were 10 weeks old at the start of the dosing period. Their bodyweight ranged approximately from 300 to 325 g based on supplier data.

PB was formulated as a 40 mg/ml suspension in 0.9% w/v aqueous sodium chloride and was administered during 4 days (from day 1 to day 4) at 2 mL/kg to male Wistar rats ( $n = 20$ ), via intraperitoneal injection. Urine samples were collected twice (day -3 and day 5) over a 16-h period on a refrigerated surface using metabolism cages. They were then stored at -80 °C until analysis. All studies on animals complied with the *Décret sur l'Expérimentation Animale* (French law on rules for animal experimentation, Decree 87-848, October 19, 1987).

**Ultrapformance Liquid Chromatography Coupled to Electrospray Time-of-Flight Mass Spectrometry.** Urine samples were centrifuged at 10 000 rpm for 5 min to remove particles and proteins and then diluted with two volumes of water prior to sample injection. Predose and postdose urine samples were randomly analyzed together by UPLC/TOFMS. Urine samples (10  $\mu$ L/injection) were separated on a 100  $\times$  2.1 mm Acquity BEH 1.7- $\mu$ m C<sub>18</sub> column (Waters, Saint Quentin en Yvelines, France) using an Acquity UPLC system (Waters) with a gradient mobile phase. Two eluting conditions were used: a first gradient mobile

**Table 1. UPLC Gradient Conditions**

time (min)	A (%)	B (%)
0	100	0
5	100	0
15	80	20
25	20	80
30	5	95
32	5	95
32.1	100	0
35	100	0

phase (hereafter referred to as “acetonitrile-acidic eluting conditions”) comprising 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B) and a second one (hereafter referred to as “methanol-neutral eluting conditions”) comprising 10 mM ammonium acetate (A) and methanol containing 10 mM ammonium acetate (B). In both cases, a 0.5 mL/min flow rate was maintained in a 35-min run with relative proportions of A and B as mentioned in Table 1.

Mass spectrometry experiments were performed on a Waters LCT Premier operating in negative and positive ion modes with W-optics enabled. For the negative ion mode, the nebulizer gas flow was 100 L/h, the desolvation gas flow 700 L/h at 400 °C with the cone gas flow 50 L/h, and the source temperature 150 °C. The capillary voltage and the cone voltage were 2700 and 50 V, respectively. Deprotonated leucine-enkephalin was used as the lock mass ( $m/z$  554.2614) for accurate mass calibration and introduced at a concentration of 1.5  $\mu$ g/mL in 50% aqueous methanol using the LockSpray interface at 50  $\mu$ L/min. For the positive ion mode, parameters were similar, except for the capillary voltage and the lock mass, which were 3000 V and  $m/z$  556.2771 (corresponding to protonated leucine-enkephalin), respectively. Data were acquired in centroid mode from  $m/z$  100 to 1000 with dynamic range enhancement. Following data acquisition, UPLC-MS chromatograms and spectra were further processed by MassLynx 4.1 and MarkerLynx application softwares (Waters) on an IBM intelligistation Z Pro. The metabonomics MS system test mix (Waters) was injected at regular intervals (every 10 injections) to check for consistency of analytical results.

### Data Preprocessing and Multivariate Statistical Analysis.

Raw data were analyzed using the Micromass MarkerLynx Application Manager (version 1.0, Waters). The UPLC/TOFMS data were peak-detected and noise-reduced in both the LC and MS domains to exclude peaks related to noise. A list of the intensities of the peaks detected was then generated for each sample, using a combination of the retention time (RT) and the  $m/z$  ratio as an identifier. An arbitrary number was assigned to each of these RT- $m/z$  pairs in order of elution. The data were then combined into a single matrix by aligning peaks with the same mass-retention time pair together from each data file in the data set. In the case of either nondetected signals or detected signal intensities under the set cutoff value, MarkerLynx assigned a zero value to the variable in the matrix for the samples concerned. Once this was completed, the ion intensities for each detected peak were normalized, within each sample, to the sum of the peak intensities in that sample. Difference between urinary volumes before and after treatment was not statistically significant (paired  $t$  test,  $\alpha = 0.05$ ). As a consequence, there was no need for a normalization step against creatinine. The resulting data were

(26) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham, C. R. *J. Mass Spectrom.* **2005**, *40*, 430–43.

(27) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977–82.

(28) Hu, Q.; Makarov, A. A.; Cooks, R. G.; Noll, R. J. *J. Phys. Chem. A* **2006**, *110*, 2682–9.

(29) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113–20.

introduced into SIMCA P (version 11.0; Umetrics) for multivariate analyses (MVAs). Prior to MVAs by unsupervised and supervised methods, principal component analysis (PCA), partial least-squares or projection on latent structure-discriminant analysis (PLS-DA), and orthogonal partial least-squares or projection on latent structure (OPLS) in discriminant analysis mode, respectively, data exported from MarkerLynx were first mean-centered and scaled either to unit variance or to Pareto variance. Models were validated by using the cross-validation function of SIMCA P. PLS-DA models were tested with permutation tests ( $k = 20$ ).

**Autocorrelation Matrices.** Such an approach was originally developed for the processing of NMR data. As described by Cloarec et al.,<sup>30</sup> it is based upon the calculation of a correlation matrix  $C$  according to the following equation:

$$C = (1/(n-1))X_1 \times {}^tX_2 \quad (1)$$

where, for MS experiments,  $X_1$  and  $X_2$  denote the autoscaled experimental matrices of  $n \times v_1$  and  $n \times v_2$ , respectively;  $n$  is the number of samples and  $v_1$  and  $v_2$  are the number of variables (i.e., couple RT- $m/z$ ) in the chromatograms for each matrix.  ${}^tX_2$  is the transpose of the matrix  $X_2$ , and  $C$  is therefore a matrix of  $v_1 \times v_2$ , where each value is a correlation coefficient between two variables of the matrices  $X_1$  and  $X_2$ .  $X_1$  is equal to  $X_2$  for autocorrelation analysis. In this latter case, the matrix displays the correlation among the intensities of the variables across the whole sample. By assuming that the different ions from a single metabolite should exhibit the same retention time and the same change in concentration in response to a perturbation, and that instrumental conditions were kept identical from one sample to another, the relative intensities of such ions should theoretically be fully correlated (correlation coefficient  $r = 1$ ). In complex samples such as biofluids,  $r$  was always less than 1 because spectral noise or peak overlaps occur from other molecules, thus leading to a matrix effect or ion suppression. However, in practice, the correlation matrix from a set of mass spectra containing different amounts of the same molecules showed very high correlations between the variables corresponding to the various ions related to the same molecule (adduct ions, in-source CID product ions). Autocorrelation matrices were calculated from mean-centered and scaled to unit variance signal intensities, by using Matlab 6.

**Liquid Chromatography Coupled to Electrospray-LTQ-Orbitrap Mass Spectrometry.** An HP1100 series LC system (Agilent Technologies, Massy, France) was used for sample introduction and separation. The chromatographic separation was performed on a XTerra MS  $C_{18}$  5- $\mu$ m, 2.1  $\times$  150 mm column (Waters) equipped with an online prefilter (Interchim, Montluçon, France). The mobile phases were either (A) 100% water and (B) 100% methanol, both containing 10 mM ammonium acetate buffer pH 7.0 or (A) 100% water and (B) 100% acetonitrile, both containing 0.1% formic acid. After an isocratic step of 5 min at 100% phase A, a linear gradient from 0 to 100% B was run over the next 40 min with a mobile phase flow of 0.2 mL/min. Returning to 100% A over 1 min, the column was then allowed to equilibrate for 10

min leading to a total run time of 60 min. These will later be referred to as "methanol-neutral eluting conditions" and "acetonitrile-acidic eluting conditions", respectively.

ESI-MS metabolic profiles were acquired within the range  $m/z$  100–1000, successively in both positive and negative ion modes with a resolution set to 60 000 ( $M/\Delta M$  at fwhm) in centroid mode. For the negative ion mode, the sheath gas was 50 (arbitrary units), the auxiliary gas 10 (arbitrary units), and the capillary temperature 275 °C. The source voltage and the tube lens voltage were 3500 and -100 V, respectively. Lens parameters were optimized using a tune solution containing testosterone glucuronide, aldosterone, DHEA sulfate, and adipic acid at 1  $\mu$ g/mL in methanol/water (50:50), infused at 5  $\mu$ L/min. Gas flows were then optimized with the same compounds but with a 0.2 mL/min HPLC flow. For the positive ion detection mode, the sheath gas was 50 (arbitrary units), the auxiliary gas 5 (arbitrary units), and the capillary temperature 275 °C. The source voltage and the tube lens voltage were 5000 and +100 V, respectively. Orbitrap calibration was performed in the positive and negative ion modes by using calibration mixture from Thermo-Fisher.

CID spectra were acquired using data-dependent scanning function. The scan event cycle comprised a full scan mass spectrum at a resolution power of 15 000 and two data-dependent ( $MS^2$  and  $MS^3$ ) events acquired with a resolution set to 7500. The most intense ion detected during full scan underwent data-dependent scanning ( $MS^2$ ) whose most intense ion ( $MS^2$  parent ion excluded) triggered  $MS^3$ . Microscan count was set to unity, and a repeat count for dynamic exclusion was set to 3.  $MS^n$  acquisition parameters were an isolation width of 1 u, normalized collision energy of 35%, and activation time of 30 ms.

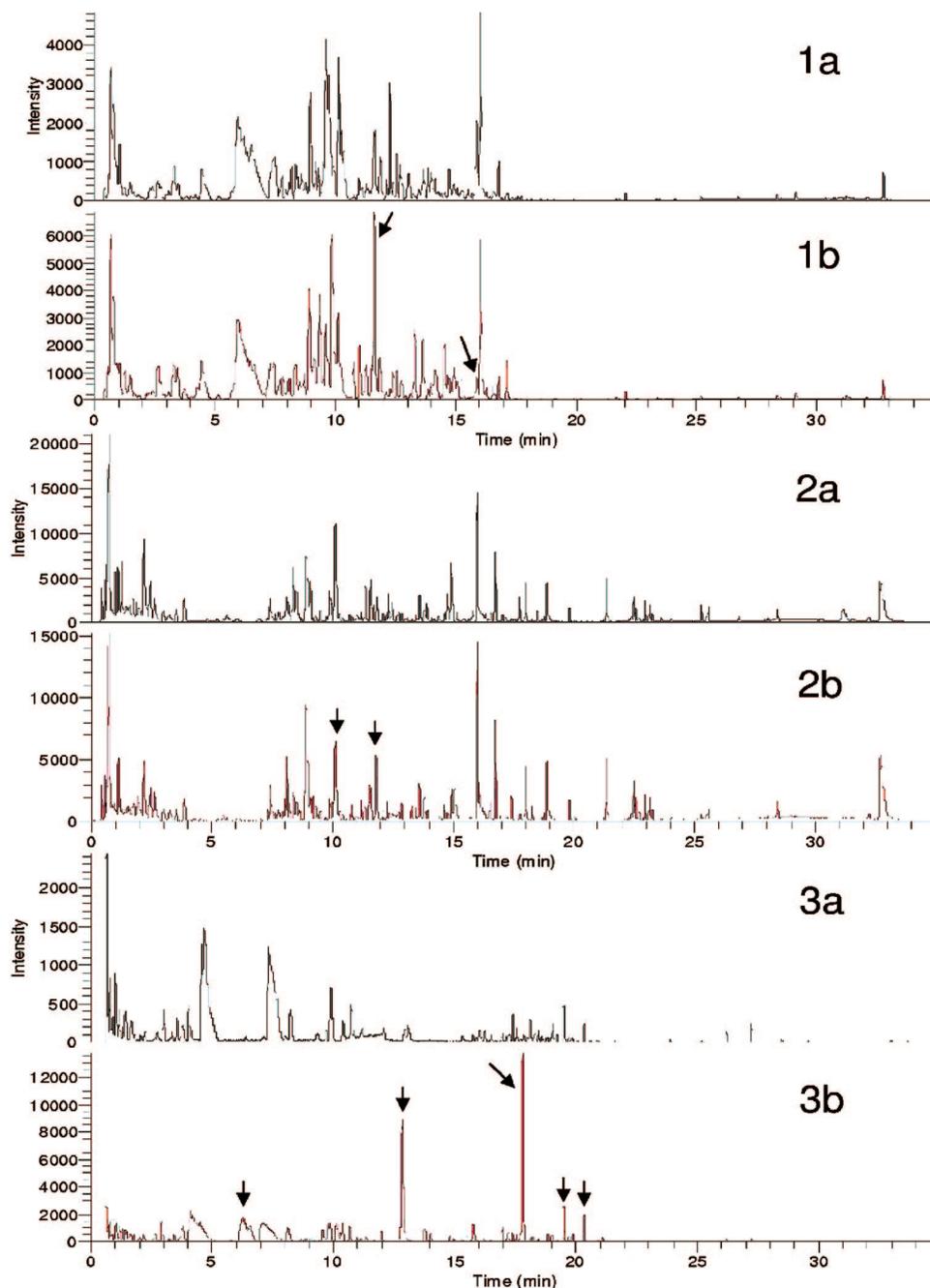
All data were processed using Qualbrowser (Thermo-Fisher), and its chemical formula generator was used to provide elemental compositions.

## RESULTS AND DISCUSSION

This study deals with an approach to analyzing UPLC/MS-based metabolomics data sets, from the analytical method to the characterization of discriminating variables, highlighted by multivariate analyses. The proof of concept was performed on urine samples from PB-treated rats.

**Data Acquisition.** The analysis of urine samples was carried out using a 35-min UPLC reversed-phase gradient coupled with ESI-TOFMS detection in both the positive and the negative electrospray modes. The mass accuracy of the mass spectrometer was ensured by the use of the lockspray interface. Moreover, the metabolomics MS system test mix (Waters) was injected every 10 samples, to check for consistency of analytical results. This mix was compatible with the different chromatographic elution conditions in terms of distribution of reference compounds along the gradient. It allowed to assess the performances of both the chromatographic column (consistency of retention times and peak widths of reference compounds) and the mass spectrometer during the experiments (consistency of mass accuracy and signal intensity of reference compounds). Experiments were judged suitable for further biological interpretation if the mass accuracy of the test mix standards was below 5 ppm and if the coefficients of variation obtained from the retention times and signal intensities of the chromatographic peaks of the standards were below 5 and 25%, respectively.

(30) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282–9.



**Figure 1.** Representative base peak ion reconstructed chromatogram resulting from the injection of urine samples from male Wistar rats obtained before (a) and after (b) a four-day treatment with PB. Acquisition was performed in negative (1) and positive (2) ion electrospray with a UPLC-TOFMS separation on a 2.1 mm × 150 mm, 1.7 μm C<sub>18</sub> bonded column using a water/acetonitrile (both with 0.1% formic acid) gradient over 30 min at 500 μL/min or using a water/methanol (both containing ammonium acetate 10 mM) gradient (same chromatographic conditions) in the negative ion detection mode (3). Arrows highlight peaks that visually underwent intensity modifications.

Two sets of eluting conditions were used. Typical UPLC/TOF base peak ion chromatograms of predose (a) and treated (b) rat urine are displayed in Figure 1 for each experimental condition. No major modifications of the metabolic fingerprints were observed from visual inspection of the reconstructed chromatograms obtained in acetonitrile-acidic eluting conditions for both positive and negative electrospray modes. This was, however, not the case for the chromatograms obtained with the methanol-neutral eluting conditions, for which dramatic modifications were observed in metabolic fingerprints between predose and treated animal urine samples. This suggests that different mobile-phase compositions lead to significant variations of ionization efficiency.

Furthermore, the use of different solvents allowed us to exploit complementary retention mechanisms on reversed-phase columns. Indeed, two different organic modifiers affecting preferentially hydrogen bonds (methanol) or dipole-dipole interactions (acetonitrile) may discriminate coeluting compounds in one or the other condition. Moreover, changing pH-eluting conditions modifies the ionization state of ionizable compounds during the chromatographic migration and, therefore, modifies their chromatographic retention behavior. In this way, compounds that were eluted in the dead volume under one set of conditions due to an ionized state may be retained on the column under the other set of conditions. Finally, regarding mass spectrometry, the use of

two organic modifiers is likely to modify the relative ionization efficiency of some compounds in the ESI source. Although such an experimental design did not allow us to discriminate between the effects related to either pH or solvent changes, different signals were highlighted from one condition to the other, showing that the use of different chromatographic conditions has a strong impact on the peak-picking step and is mandatory for maximizing information exhaustiveness when using a single chromatographic system.

Mass spectrometric detection was performed by using a TOF analyzer operated in the W-mode, which provided accurate mass measurements within 5 ppm, using internal calibration. This is of particular interest for the analysis of the chromatographic dead volume, which contains highly polar metabolites that are not retained on the reversed-phase column. Such metabolites represented 27, 7, and 18% of the signals that were detected in methanol-neutral eluting conditions with negative electrospray, and in acetonitrile-acidic eluting conditions with both the negative and positive modes, respectively. As a consequence, a significant part of the information that is lost with low-resolution measurements can be taken into account. However, matrix effects and ion suppressions are still more likely to occur in such circumstances and isomers remain indiscernible.

**Statistical Data Reduction and Analysis.** A multivariate analysis was conducted in order to visualize the data and to highlight signals allowing the discrimination of samples from treated and predose animals. Before performing statistical analysis, raw data were preprocessed in order to extract the analytically relevant variables after having removed background noise and isotopes. MarkerLynx extracted 2458 variables in the negative ion mode for methanol-neutral eluting conditions and 8524 or 3627 variables for acetonitrile-acidic eluting conditions for both the negative and positive ion modes, respectively. In the latter case, the high number of variables that were extracted in the negative ion mode could be attributed to lower signal and chemical noise intensities than those observed in the positive ion mode, thus allowing extraction of variables with a lower intensity cutoff limit. Although some of these extracted signals may be erratic, all of the discriminating variables that were investigated in the course of this study were found to be analytically relevant.

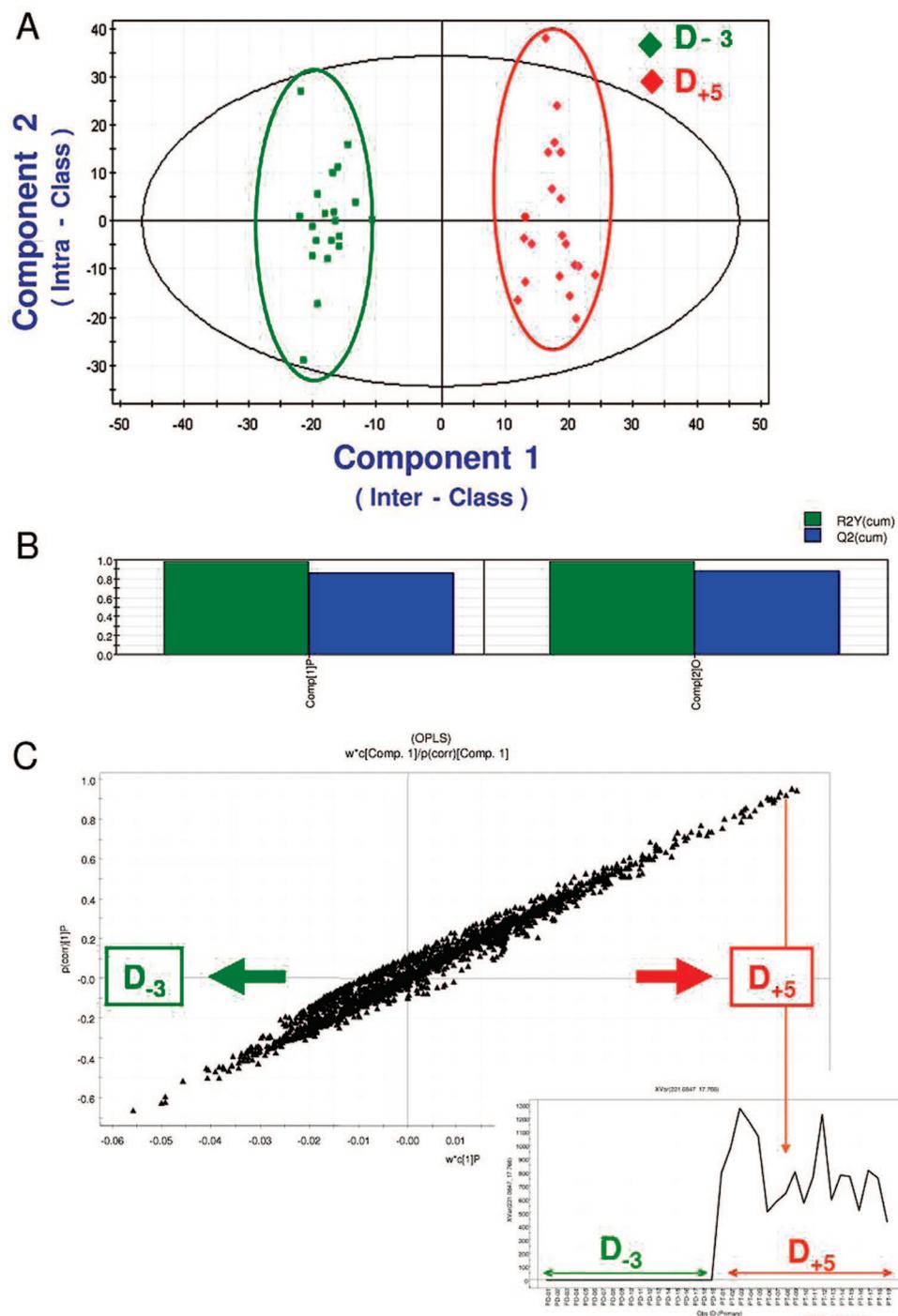
The main issues encountered during software-assisted signal treatment are first to ensure selection of all the relevant peaks without extracting peak spikes or background noise and then to avoid peak splitting. In this context, UPLC offered great advantages compared with HPLC. First, it reduced analysis time for the same or even better peak resolution, thus improving signal recoveries by limiting ion suppression in the latter case. It also improved signal-to-noise ratios due to narrower peak widths. These latter were of 9.5, 9.5, and 10 s for acidic positive, acidic negative, and neutral negative chromatographic conditions, respectively. As a consequence, UPLC allowed the detection of compounds occurring at lower levels and improved software-assisted peak detection and variable extraction compared with conventional LC/MS-based systems (data not shown). Moreover, UPLC provided good reproducibility and robustness of data generated. There were no, or at least minor, retention time shifts, and so peak realignment step was unnecessary (e.g., standard deviations on retention time

were 1.63 (40 samples) and 0.55 s (20 samples) for hippuric acid and phenobarbital, respectively).

Processed data were then submitted to multivariate analysis using the software SIMCA P11 (Umetrics, Umea, Sweden). Before and after normalization (mean-centering and scaling to “unit variance” or to “Pareto variance”), the preprocessed data were first submitted to PCA, an unsupervised projection method used to visualize the structure of the data set. Strong clustering in the PCA scores between predose and treated samples was observed with the first two components, suggesting that phenobarbital administration was the main source of variance in the samples (data not shown).

Supervised methods such as projection on latent structure-discriminant analysis (PLS-DA) and orthogonal-PLS (O-PLS) were then used to facilitate the isolation of the variables responsible for the discrimination between postdose and predose samples. Two-component OPLS models were constructed for each experimental condition and led in any case to R<sup>2</sup><sub>Y</sub> and Q<sup>2</sup> above 0.9 and 0.85, respectively (detailed results are provided in Table S1 as Supporting Information). As an example, Figure 2 displays the results obtained with the acetonitrile-acidic eluting conditions coupled to negative ion mode detection. SIMCA returned a variable importance on projection (VIP) score, which reflects the contribution of the variables to the model. A variable is considered as important for the model when its VIP is above 1.0. To focus on highly significant variables, only those with a VIP above 1.1 were considered. Removing variables with a VIP within the range 1.0–1.1 excluded variables with, for instance, 20 zero values for a group and 16 zero values for the other. Such variables are not very meaningful because of their poor robustness, even if they exhibit a statistical significance in the discrimination between the two groups. Moreover, as the identification step is already the most time- and work-consuming, it is worth limiting the number of variables to be characterized.

The numbers of variables identified by multivariate statistical analyses, in each set of conditions, are displayed in Table 2. It is possible to classify these variables according to their presence in both the predose and treated groups or just in the treated group. Those occurring in both groups could be related to susceptibility biomarkers, i.e., endogenous metabolites whose concentrations were altered due to PB treatment (pharmacological or toxicological effects, CYP induction). In contrast, the signals that are only present in the treated group could be attributed to either PB metabolites (i.e., biomarkers of exposure) or to endogenous signals that are not detected in the predose group and whose levels are increased following PB administration. In the latter case, further MS<sup>n</sup> experiments are required in order to assess the exogenous or endogenous origin of these signals. This is the aim of the following parts of this work, by using LC-ESI-LTQ-Orbitrap. However, a high number of variables had to be investigated, i.e., 42 signals for methanol-neutral eluting conditions, with the negative ion mode detection, and 71 signals for acetonitrile-acid eluting conditions, with both positive and negative ion mode detection. Furthermore, among the 113 (42 + 71) signals highlighted by MVA in the negative mode, only 19 of them were common to the two sets of conditions. In the same way, only 10 common variables (i.e., same retention times and back calculated molecular weights) were found among the 142 (71 + 71) signals



**Figure 2.** OPLS model based on metabolite profiles of urine samples from 20 male Wistar rats three days before ( $D_{-3}$ ) and five days after ( $D_{+5}$ ) a four-day treatment with daily ip injection of phenobarbital. Profiles were acquired by using UPLC-TOFMS in the negative ion mode with acetonitrile-acidic eluting conditions. (A) Scores plot. Data were mean-centered and scaled to unit variance. (B) Model overview: R2Y (cumulative fraction of the variation of the Y variable explained after the selected component) and Q2 (cumulative predicted fraction (cross-validation) of the variation of Y). (C) OPLS S-plot. Data points in the top right corner indicate increased ions responsible for variation between predose and treated samples. The inset displays the intensity of one of them in each sample. This ion is only detected in treated sample, as a result its high discriminating power.

highlighted by MVA for both the positive and the negative mode in acetonitrile-acidic conditions. A statistical tool based on autocorrelation matrices was therefore implemented to reduce the number of signals to be identified via the suppression of redundant signals.

**Autocorrelation Matrices.** This mathematical tool was originally developed for the processing of NMR data, for which signal redundancy is frequently observed. In this context, the aim is to

group the different chemical shifts originating from a single molecule. It is based on the fact that different signals from a single metabolite should exhibit the same change in concentration in response to a perturbation. By analogy, autocorrelation matrices were here used to structure the data set by trying to group signals related to the same “parent molecule”. With ESI-MS, in the course of the desolvation process, a single compound may produce several signals corresponding to the formation of adduct ions with

**Table 2. Number of Discriminating Signals Identified by Multivariate Statistical Analyses**

UPLC/TOF conditions <sup>a</sup>	extracted variables (MarkerLynx)	discriminating variables ( $V_{ip} > 1.1$ )		
		total	present at D <sub>-3</sub> and D <sub>5</sub>	absent at D <sub>-3</sub>
I	2458	670	628	42
II	8254	407	336	71
III	3627	1433	1362	71

<sup>a</sup> I, neutral-water–methanol eluting conditions, negative electrospray detection. II, acid–water–acetonitrile eluting conditions, negative electrospray detection. III, acid–water–acetonitrile eluting conditions, positive electrospray detection.

cations such as sodium, potassium, or ammonium or to product ions formed by the spontaneous in-source fragmentation of the parent molecule.

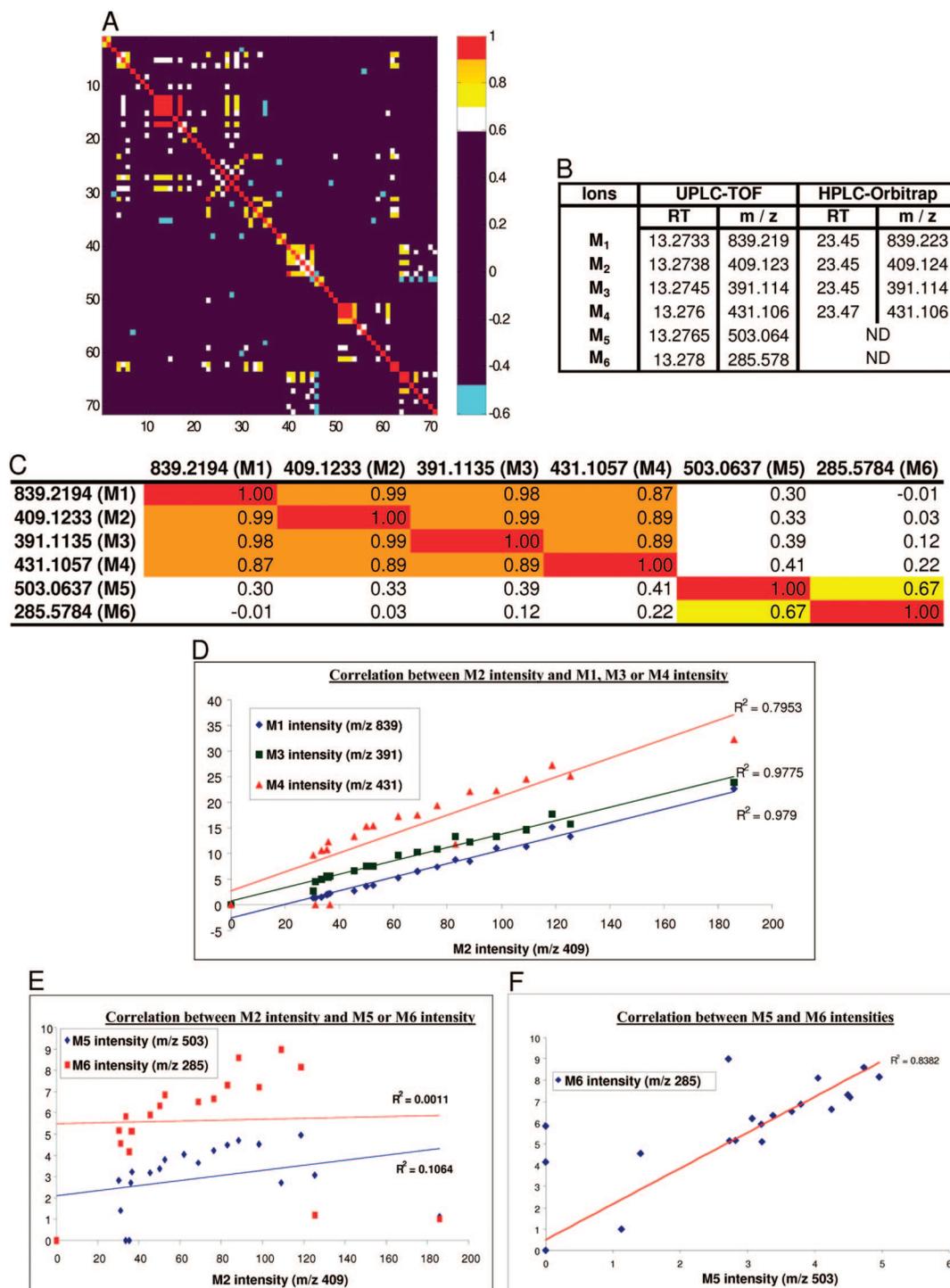
Although adduct ions can easily be evidenced by accurate mass measurement and use of scripts, it is far more difficult to link each molecular species with its own in-source fragment ions in the case of coeluting compounds. Figure 3A shows the color-scaled autocorrelation matrix calculated from the 71 signals sorted as discriminant by MVA in the acetonitrile-acidic eluting conditions and positive ion mode. On the  $X$  and  $Y$  axis, signals (i.e., variables = couples  $RT-m/z$ ) are numbered from 1 to 71 according to their retention time. As all variables are present on both axes in the same order, the matrix exhibits a symmetry line corresponding to its diagonal. Each point of this diagonal corresponds to the correlation of one variable with itself and therefore returns a correlation coefficient equal to 1. The use of autocorrelation matrices visualized the global redundancy of data (Figure 3A). This plot reveals the presence of aggregates of colored spots around the diagonal (e.g., between variables 10–20, 40–50, and 50–60). These spots correspond to high correlation coefficients between two ions because of similar behavior during PB treatment, and as they are grouped into aggregates around the diagonal, they exhibit similar retention times. They are therefore likely to represent a parent molecule with its related signals. As an example, Figure 3B displays six ions ( $M_1$ – $M_6$ ) coeluted at 13.27 min on the UPLC/TOF system for which we extracted the related part of the autocorrelation matrix (Figure 3C). The resulting correlation coefficients between the intensity of each pair of ions across all samples were used to determine whether or not there is a link between these coeluting signals. Plots of intensity of  $M_1$ ,  $M_3$ , and  $M_4$  as a function of  $M_2$  intensity for each sample (Figure 3D) confirmed that these four ions are related to the same parent molecule, as previously attested by correlation coefficients (Figure 3C).  $M_2$  ( $m/z$  409) can be attributed to the protonated molecule ( $MH^+$ ), whereas  $M_3$  ( $m/z$  391) corresponds to the loss of a molecule of water from  $M_2$ ,  $M_4$  ( $m/z$  431) to the sodium cationized molecule ( $[M + Na]^+$ ), and  $M_1$  ( $m/z$  839) to the sodium cationized dimer ( $[2M + Na]^+$ ). For  $M_5$  and  $M_6$ , it is difficult to establish a chemical link with either  $M_2$  or between these two ions. However, plotting their intensities as a function of the intensity of  $M_2$  ( $m/z$  409) (Figure 3E) indicates that there is no correlation between  $M_2$  and these two ions. Finally, plotting the intensity of  $M_5$  as a function of  $M_6$  intensity (Figure 3F) highlighted a link between these two ions. In this case, autocorrelation matrix calculation indicated that these six signals were due to only two parent molecules, the first one generating four signals ( $M_1$  to  $M_4$ ) and the second one generating two signals ( $M_5$  and  $M_6$ ). Furthermore, neither  $M_5$  nor  $M_6$  was detected using the HPLC-Orbitrap system

for signal identification (Figure 3B), indicating that part of the information provided by UPLC/TOFMS has been lost, but that only one compound was undetected despite the loss of two signals when switching from one instrument to the other. However, further structural information regarding  $M_5$  and  $M_6$  could not be acquired to confirm or validate the link that was highlighted by autocorrelation matrices.

Thus, autocorrelation matrices allowed the grouping of ions, adduct ions, and product ions generated by in-source CID. It was possible in this way to exclude some redundant information generated by MS detection so as to focus the further identification process only on protonated or deprotonated molecules. However, this redundant information, such as cationized molecules or dimers, may be helpful to confirm the relevance of elemental compositions as it allowed several accurate mass measurements on ions related to the same molecule. Finally, when applied to the experimental conditions, this statistical tool pointed out 88 signals to be investigated from the 184 signals that were obtained for all the detection conditions.

The identification process is summarized in Table 3. Autocorrelation matrices allowed us to remove 55% of the initial variables that were highlighted in the negative electrospray mode from methanol-neutral elution conditions. Most of the variables (i.e., 16 signals or group of signals) were detected by using the LC-LTQ-Orbitrap, leading to the characterization of PB and 10 of its metabolites by further  $MS^n$  experiments. For acetonitrile-acidic conditions, ~51% of the initial variables were removed by autocorrelation matrices. Of note, depending on the chromatographic conditions and ionization modes, 20–30% of the UPLC/TOF signals were not detected by using the LC-LTQ-Orbitrap system. Finally, around 20 and 30 variables or groups of variable were investigated, leading to the identification of 9 and 10 PB metabolites plus PB excreted unchanged by further sequential  $MS^n$  experiments, in the positive and negative ion modes, respectively. Taken together, these results indicate that although autocorrelation matrices allow removal of more than a half of the signals to be characterized, some of them, corresponding to product and adduct ions exhibiting poor correlation with protonated or deprotonated ions, still require investigation. Finally, 31 signals of interest for all the detection conditions, corresponding to either protonated or deprotonated PB-related ions, had to be investigated by further sequential  $MS^n$  experiments.

**Identification of PB Metabolites Using LC-ESI-LTQ-Orbitrap.** Signals highlighted by multivariate statistical analyses and by autocorrelation matrices were characterized using LC-ESI-LTQ-Orbitrap. To this end, the 20 predose samples and the 20 treated samples were pooled, thus saving time and averaging interindividual variations.



**Figure 3.** Overview of autocorrelation matrices. (A) Autocorrelation plot from discriminating PB-related signals detected using UPLC/TOF in the positive ion mode. (B) Extract of discriminating variables: RT–*m/z* pairs for UPLC/TOF and correspondence with HPLC/Orbitrap. (C) Autocorrelation matrix. (D) Plot of M<sub>1</sub>, M<sub>3</sub>, and M<sub>4</sub> ion intensities as a function of M<sub>2</sub> intensity across the 20 treated samples. (E) Plot of M<sub>5</sub> and M<sub>6</sub> ion intensities as a function of M<sub>2</sub> intensity across the 20 treated samples. (F) Plot of M<sub>6</sub> ion intensities as a function of M<sub>5</sub> intensity across the 20 treated samples.

The identification process was as follows. First, HPLC-Orbitrap and UPLC-TOF signals were matched. As the chromatographic systems were different, the matching was based on mass accuracy and also on the absence of PB-related signals in predose samples. Ions to be characterized were extracted from HPLC-Orbitrap chromatograms and compared with those occurring in UPLC-TOF within a mass range of 20 ppm.

Despite the presence of many variations in chromatographic conditions (gradient, column, UPLC vs HPLC), elution order was

almost always maintained. Accurate mass filtering of the full scan mass spectra often led to the extraction across the entire chromatogram of several ions with an identical *m/z* ratio, but comparison of predose and treated samples and coeluted ions ensured matching between the signals of the two methods.

Formulas were then generated using the Qualbrowser elemental composition calculator (ThermoFisher) for each UPLC-TOF variable detected with the HPLC-Orbitrap system with a mass accuracy of 5 ppm. Whatever the eluting conditions, none of the

**Table 3. Impact of Autocorrelation Matrices on the Number of Discriminating Variables To Be Identified**

UPLC/TOF conditions <sup>a</sup>	discriminating variables <sup>b</sup>	variables or groups of variables <sup>c</sup>	variables or groups of variables to identify <sup>d</sup>	metabolites identified
I	42	19	16	11, including PB
II	71	34	28	11, including PB
III	71	35	23	9

<sup>a</sup> I, neutral-water–methanol eluting conditions, negative electrospray detection. II, acid–water–acetonitrile eluting conditions, negative electrospray detection. III, acid–water–acetonitrile eluting conditions, positive electrospray detection. <sup>b</sup>  $V_{ip} > 1.1$ , absent at D<sub>-3</sub>. <sup>c</sup> Highlighted by autocorrelation matrices. <sup>d</sup> After suppression of nondetected signals on the HPLC/LTQ-Orbitrap system.

measured masses differed from theoretical masses by more than 1.5 ppm.

Experiments performed with LTQ-Orbitrap not only confirmed results obtained from UPLC/TOFMS experiments but also dispelled ambiguities about formulas previously generated with lower mass accuracy measurements. However, even a mass accuracy of 1 ppm does not always enable a unique assignment to a molecular formula.<sup>31</sup> But, analysis of isotopic distribution helped to determine the nature of elements to be selected for calculating elemental compositions.<sup>32</sup> As an example, Figure 4 displays two ions of nominal  $m/z$  263. These two ions exhibit a  $m/z$  difference of 0.018 and similar retention times (24.02 and 24.52 min on HPLC-Orbitrap system). One contains one sulfur atom, as attested by a 4.5% relative isotopic abundance at monoisotopic peak plus 1.9958 (theoretically 1.9959), whereas the other does not. It is therefore possible to limit the number of elements in the elemental composition of the compound to be identified and thus to decrease the number of formulas generated by the software. In fact, these two signals corresponded to fragment ions generated by in-source CID of the sulfoconjugate and the mercapturic acid conjugate of PB.

MS<sup>n</sup> experiments are also useful to confirm or to discriminate between relevant formulas. CID spectra provide a second accurate mass measurement of the parent ion and accurate mass determination of its product ions. Product ion formulas can be generated, thus reinforcing or invalidating initial hypotheses. MS<sup>2</sup> fragment ion or neutral loss formula determination may also highlight the presence of a particular functional group or signal a metabolite when a product ion or a neutral loss is common to the parent drug fragmentation pattern. These points are illustrated in Figures 5 and 6.

The negative ion CID spectra of the PB–mercapturic acid conjugate are shown in Figure 5. The isotope pattern analysis in full scan MS revealed the presence of a sulfur atom. MS<sup>2</sup> experiments removed biotransformation add-on (i.e., the *N*-acetylcysteine moiety). Finally, the MS<sup>3</sup> product spectrum confirmed the presence of PB with two ions  $m/z$  234 and 177 corresponding to typical neutral losses for para-substituted PB (29u C<sub>2</sub>H<sub>5</sub> and 43u NCHO). Of note, autocorrelation matrix

showed a link between ions at  $m/z$  821, 410, and 392, corresponding to (2M + H)<sup>+</sup>, (M + H)<sup>+</sup>, and (M + H – H<sub>2</sub>O)<sup>+</sup>, respectively. The ion at  $m/z$  234 that was observed on mass spectrum and that may be a product ion from the ion at  $m/z$  263, was not highlighted by MVA (i.e., not sorted as a discriminating signal). As a consequence, it was not included in the autocorrelation matrix which was built from MVA results.

Another example is given in Figure 6. In this case, the isotope pattern excluded the occurrence of elements with natural isotopes at the monoisotopic peak plus 2u, such as K, Cl, S, and P. MS<sup>2</sup> experiments removed the biotransformation add-on (*N*-acetylhexopyranosyl, 221u). Sequential MS<sup>3</sup> CID spectra demonstrated linkage of the compound with PB through a fragment ion corresponding to *p*-OH-PB ( $m/z$  247). A proposed mechanism of fragmentation is presented in Figure 6D. This metabolite could correspond to an *O*-glucuronide-*N*-acetylhexopyranosylamine conjugate of PB. However, this structure must be confirmed by further investigations relying on D<sub>2</sub>O exchange and NMR experiments. The latter example clearly highlights the limitations of conventional CID experiments as a stand-alone tool for the identification of discriminating variables identified by metabolomic approaches.

Finally, such an approach detected PB, excreted unchanged in urine, as well as 14 of its metabolites, as summarized in Table 4, and could be used to propose a metabolic pathway of PB in the rat (Scheme 1). The metabolism of phenobarbital has been investigated in vitro, in rat hepatocytes,<sup>34</sup> and also in vivo in rodent biofluids,<sup>36</sup> thus leading to the characterization of *p*- and *m*-hydroxyphenobarbital, together with the phenobarbital-*N*-glucoside (in mice),<sup>37</sup> the phenobarbital-*N*- (in mice),<sup>38</sup> and the hydroxyphenobarbital-*O*-glucuronides. A 2-phenyl- $\gamma$ -butyrolactone derivative of phenobarbital was also reported from hepatocyte incubations.<sup>34</sup> This metabolite was detected in human biofluids following PB intoxication<sup>39</sup> but was not found in our experimental conditions. According to our knowledge, all the other metabolites that were highlighted by our metabolomic approach have not been identified in rats, but rather from experiments with human materials,<sup>22,23,40–42</sup> except for the PB-*O*-glucuronide-*N*-acetylhexopyranosylamine metabolite, that is reported here for the first time.

Even if complete characterization can rarely be achieved using exclusively MS data, ultrahigh resolution combined with sequential MS<sup>n</sup> experiments provides great structural information that can be used to predict metabolic pathways. Of note, the fragment ions obtained from some CID experiments on PB and its related metabolites are displayed in the Table S2 (Supporting Information). Furthermore, this study showed the value of performing several acquisitions in different eluting and detection conditions. Some metabolites such as deethyl-PB, dihydroxy-PB, dihydrodiol-PB,

(31) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*, 234.

(32) Stoll, N.; Schmidt, E.; Thurow, K. J. *Am. Soc. Mass Spectrom.* **2006**, *17*, 1692–9.

(33) Levin, S. S.; Vars, H. M.; Schleyer, H.; Cooper, D. Y. *Xenobiotica* **1986**, *16*, 213–24.

(34) Verite, P.; Cave, C.; Menager, S.; Davy, J.; Andre, D.; Farnoux, C. C.; Lafont, O. *Drug Metabol. Drug Interact.* **1996**, *13*, 41–55.

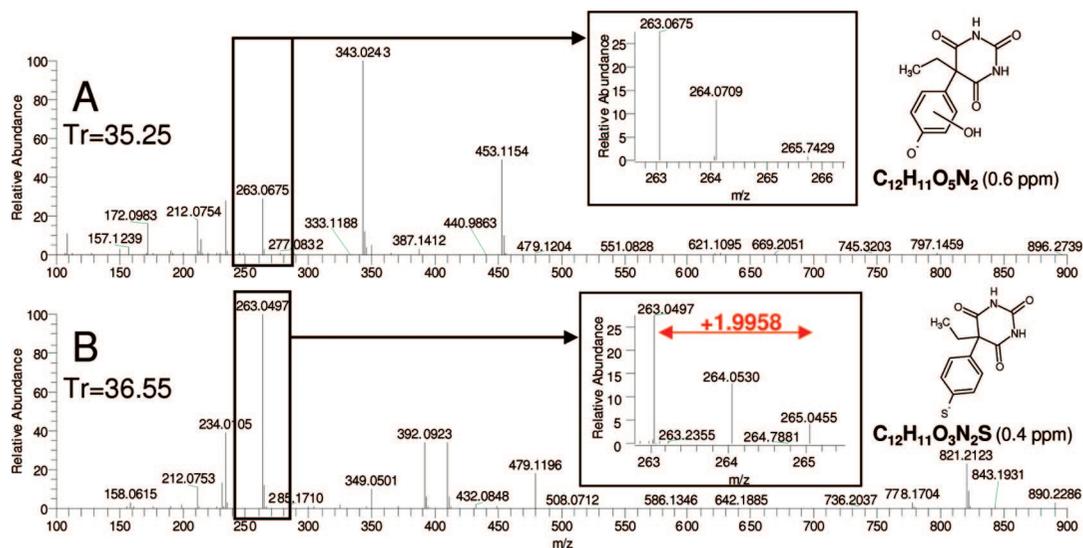
(35) Caldwell, J.; Croft, J. E.; Smith, R. L.; Snedden, W. *Br. J. Pharmacol.* **1977**, *60*, 295P–6P.

(36) Anderson, G. D.; Levy, R. H. *Pharm. Res.* **1992**, *9*, 1622–8.

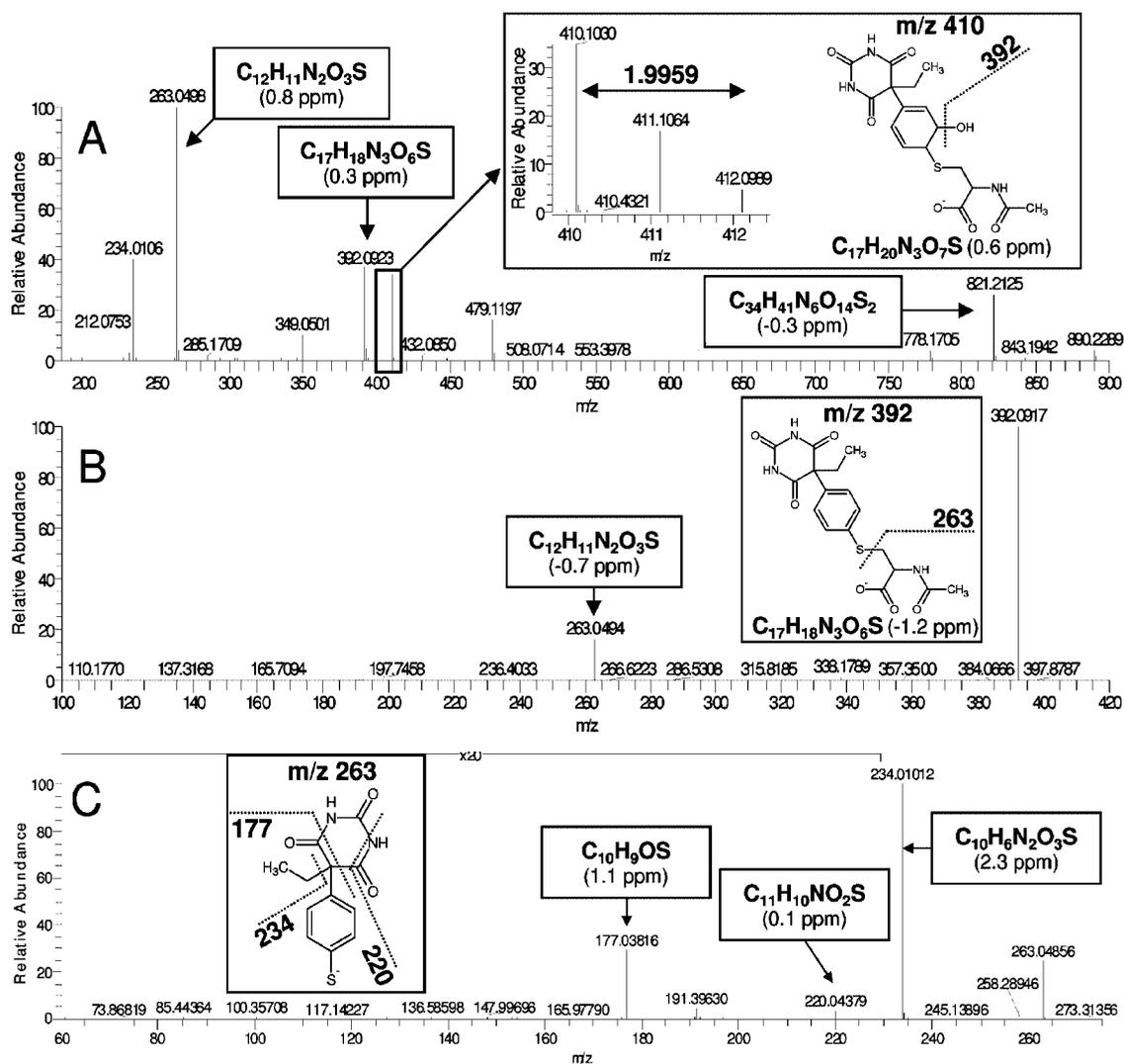
(37) Soine, W. H.; Soine, P. J.; England, T. M.; Ferkany, J. W.; Agriesti, B. E. *J. Pharm. Sci.* **1991**, *80*, 99–103.

(38) Neighbors, S. M.; Soine, W. H. *Drug Metab. Dispos.* **1995**, *23*, 548–52.

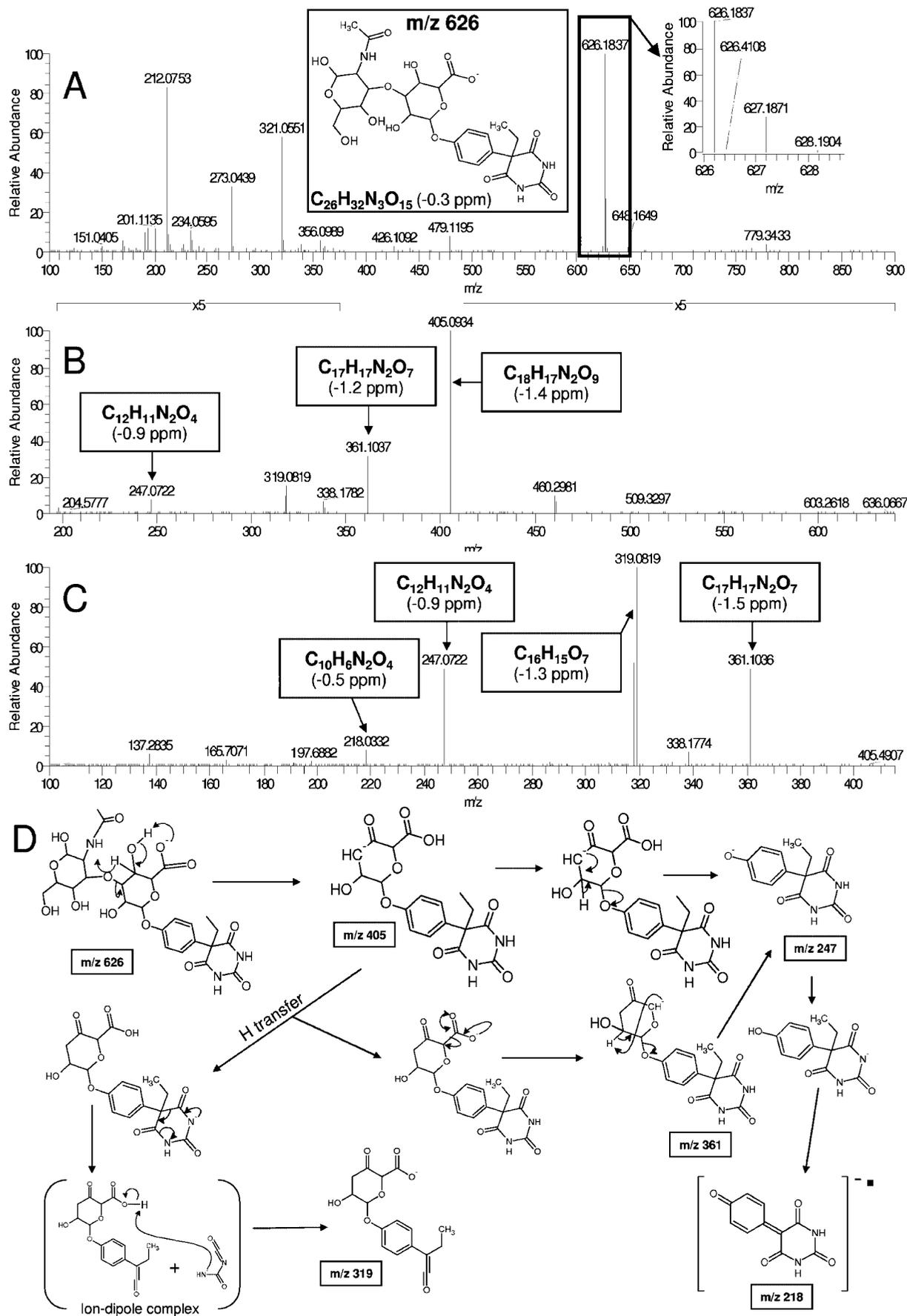
(39) Andersen, B. D.; Davis, F. T.; Templeton, J. L.; Hammer, R. H.; Panzik, H. L. *Res. Commun. Chem. Pathol. Pharmacol.* **1976**, *15*, 21–9.



**Figure 4.** Orbitrap electrospray mass spectra of two LC peaks containing  $m/z$  263: RT = 35.25 min (A) and RT = 36.55 min (B). The resolution was set to 60 000, and the mass spectra were recorded in the negative ion mode. Detailed isotopic distribution for each ion is provided as inset.



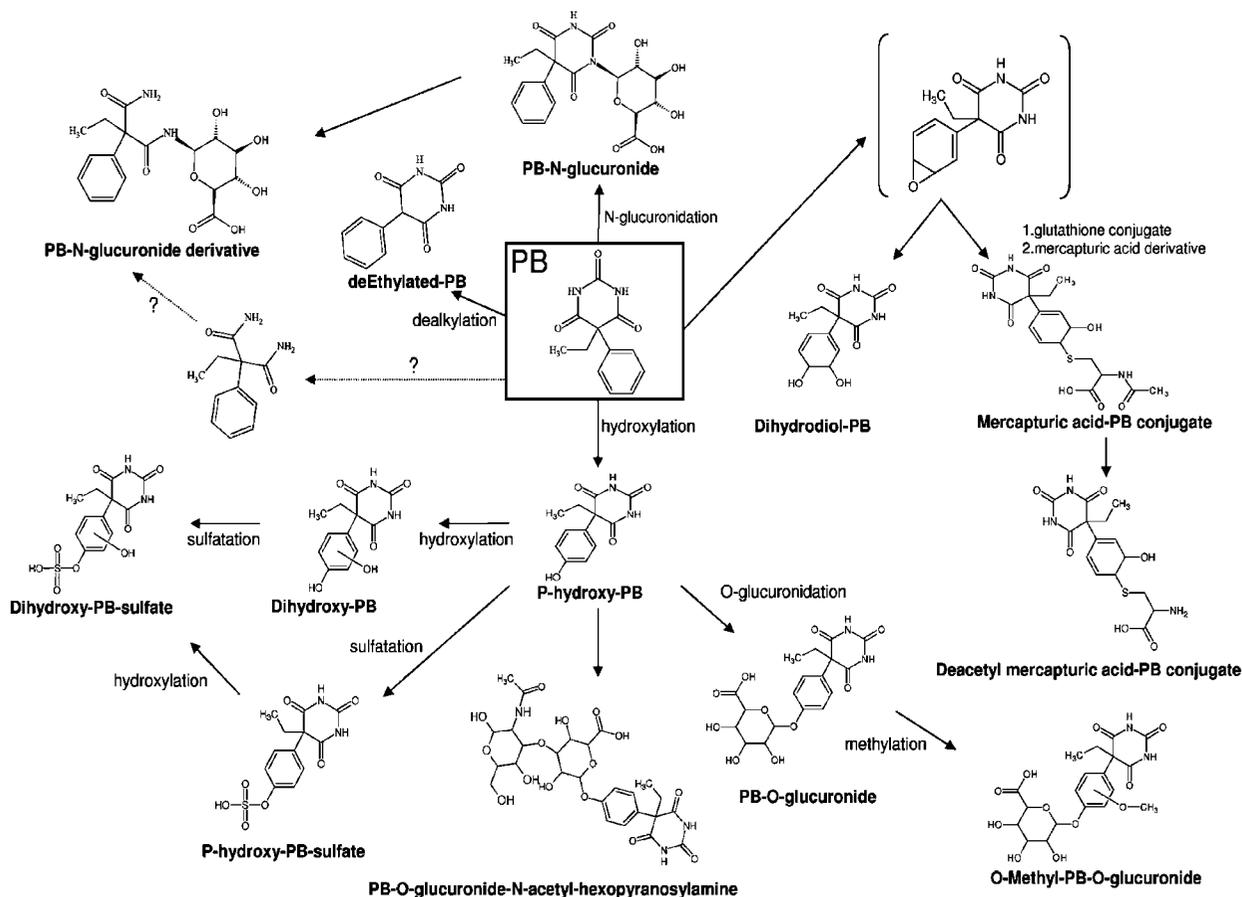
**Figure 5.** (A) Orbitrap electrospray mass spectrum of the LC peak containing  $m/z$  410 recorded in the negative ion mode. The resolution was set to 60 000, and detailed natural isotopic distribution is provided as inset. (B) CID product mass spectrum of the molecular ion  $m/z$  410 with a resolution set to 30 000. (C) Sequential MS<sup>3</sup> spectrum of the most intense ion produced in MS<sup>2</sup> spectrum of  $m/z$  410 ( $m/z$  392).



**Table 4. Overview of Detected Metabolites According to Detection Conditions**

composition	putative structure	detection conditions <sup>a</sup>
C <sub>10</sub> H <sub>8</sub> O <sub>3</sub> N <sub>2</sub>	PB without ethyl lateral chain	I
C <sub>12</sub> H <sub>12</sub> O <sub>3</sub> N <sub>2</sub>	PB	I, II
C <sub>12</sub> H <sub>12</sub> O <sub>4</sub> N <sub>2</sub>	<i>p</i> -hydroxy-PB	I, II, III
C <sub>12</sub> H <sub>12</sub> O <sub>5</sub> N <sub>2</sub>	dihydroxy-PB	I
C <sub>12</sub> H <sub>14</sub> O <sub>5</sub> N <sub>2</sub>	dihydrodiol-PB	I <sup>b</sup> , II
C <sub>12</sub> H <sub>12</sub> O <sub>7</sub> N <sub>2</sub> S	<i>p</i> -hydroxy-PB-sulfate	I, II
C <sub>12</sub> H <sub>12</sub> O <sub>8</sub> N <sub>2</sub> S	dihydroxy-PB-sulfate	I <sup>b</sup> , II
C <sub>17</sub> H <sub>22</sub> O <sub>8</sub> N <sub>2</sub>	PB- <i>N</i> -glucuronide with an opening of the barbital moiety (loss of CO)	I, II, III
C <sub>17</sub> H <sub>20</sub> O <sub>9</sub> N <sub>2</sub>	hydroxy-PB- <i>O</i> -glucuronide with an opening of the barbital moiety (loss of CO)	II <sup>b</sup> , III
C <sub>17</sub> H <sub>21</sub> O <sub>7</sub> N <sub>3</sub> S	PB-mercaptopuric acid conjugate	I, II, III
C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	PB- <i>N</i> -glucuronide enantiomer	I, II, III
C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	PB- <i>N</i> -glucuronide	I, II, III
C <sub>18</sub> H <sub>20</sub> O <sub>10</sub> N <sub>2</sub>	<i>p</i> -hydroxy-PB- <i>O</i> -glucuronide	I, II, III
C <sub>19</sub> H <sub>22</sub> O <sub>11</sub> N <sub>2</sub>	<i>O</i> -methyl-PB- <i>O</i> -glucuronide	I, II, III
C <sub>26</sub> H <sub>33</sub> O <sub>15</sub> N <sub>3</sub>	PB- <i>O</i> -glucuronide- <i>N</i> -acetylhexopyranosylamine	I <sup>b</sup> , III

<sup>a</sup> I, neutral-water-methanol eluting conditions, negative electrospray detection. II, acid-water-acetonitrile eluting conditions, negative electrospray detection. III, acid-water-acetonitrile eluting conditions, positive electrospray detection. <sup>b</sup> Detected under the mentioned condition but not extracted by MarkerLynx.

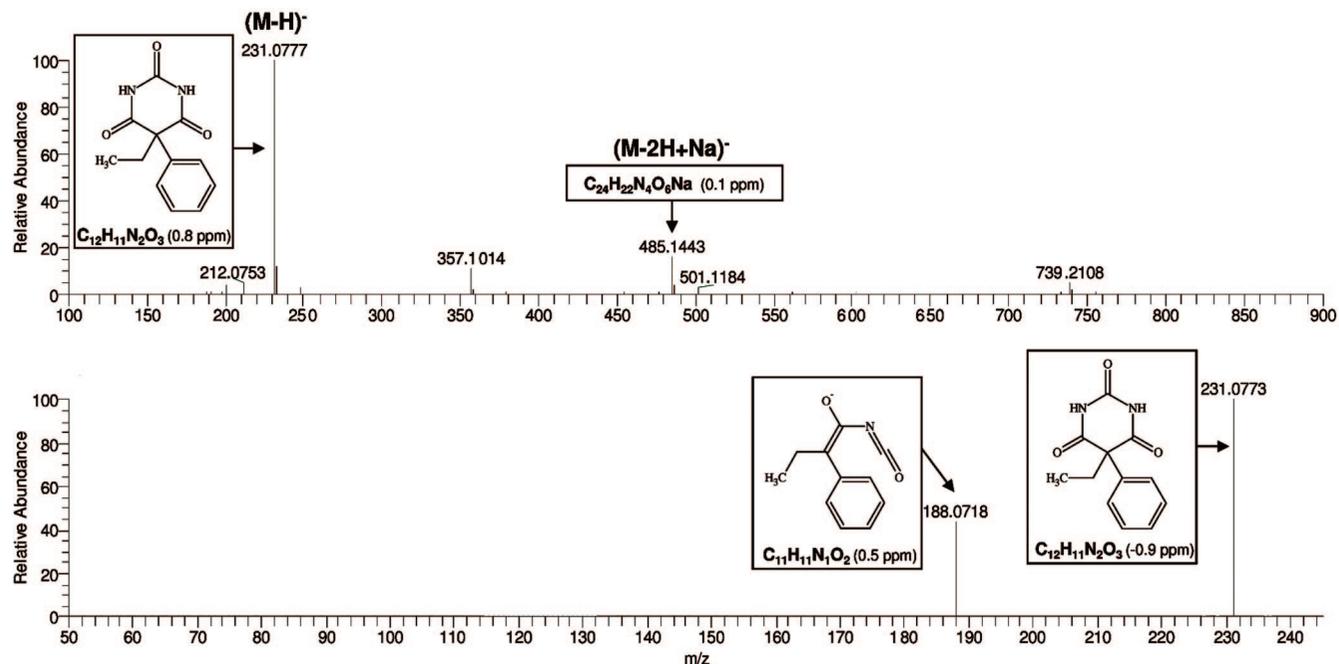
**Scheme 1. Proposed Metabolic Pathway of Phenobarbital from Metabolites Identified in Rat Urine Samples**

dihydroxy-PB-sulfate, a hydroxy-PB-*O*-glucuronide derivative and PB-*O*-glucuronide-*N*-acetyl-hexopyranosylamine were only observed in one of the three experimental conditions. Although these metabolites were detected in the other conditions, they were not extracted by MarkerLynx, despite careful optimizations of the parameters of this software, due to poor ionization efficiency in these conditions.

Other approaches such as neutral loss monitoring and the use of mass defect filters are also currently applied in metabolism

studies.<sup>43</sup> However, they have some drawbacks and are not as exhaustive as the metabolomic approach. The fragmentation

- (40) Kallberg, N.; Agurell, S.; Ericsson, O.; Bucht, E.; Jalling, B.; Boreus, L. O. *Eur. J. Clin. Pharmacol.* **1975**, *9*, 161–8.
- (41) Soine, W. H.; Soine, P. J.; Mongrain, S. E.; England, T. M. *Pharm. Res.* **1990**, *7*, 402–6.
- (42) Bernus, I.; Dickinson, R. G.; Hooper, W. D.; Eadie, M. J. *Eur. J. Clin. Pharmacol.* **1994**, *46*, 473–5.
- (43) Kostianen, R.; Kotiaho, T.; Kuuranne, T.; Auriola, S. *J. Mass Spectrom.* **2003**, *38*, 357–72.



**Figure 7.** (A) Negative Orbitrap electrospray mass spectrum of phenobarbital. The resolution was set to 60 000. (B) CID product spectrum of PB molecular ion (resolution 30 000) showing the particular fragmentation pattern of the drug compared with its metabolites: a neutral loss of 43u corresponding to the opening of the barbital moiety was observed, whereas no neutral loss of 29u corresponding to the release of the ethyl lateral chain was found.

pattern of PB, which is shown in Figure 7, differs from that of its metabolites. This can be explained by the fact that most PB metabolites are derived from *p*-OH-PB, which is the main phase I metabolite of PB. However, the hydroxyl group, which is located in the para position of the phenyl ring, was responsible for a drastic change in the fragmentation pattern since it led to a further fragmentation of the barbital moiety and to the loss of the ethyl lateral chain, which were the common losses retrieved for other metabolites (two neutral losses of 43u instead of one and a loss of 29u in negative ion mode, respectively). It is also interesting to note that the PB mercapturic derivative produced a first fragment with sulfur in the para position, which resulted in the same fragmentation pattern. As a consequence, most of the PB metabolites that have been identified by the metabolomic approach would not have been detected by a neutral loss procedure based on PB CID fragmentation. A neutral loss procedure characteristic of phase 2 biotransformations should be more efficient. However, even in this latter case, the putative PB-*O*-glucuronide-*N*-acetylhexopyranosylamine metabolite, which does not exhibit any neutral loss related to classical phase 2 biotransformations in its MS<sup>2</sup> spectrum, would not have been detected.

The same kinds of drawbacks may apply to the mass defect filter strategy. Such filters are generally applied within the range  $\pm 20$ –50 mDa.<sup>44</sup> They are therefore designed to identify metabolites undergoing only minor metabolic transformations. These transformations correspond to phase I metabolites obtained via oxidation, reduction, and loss or gain of small aliphatic groups. If applied to the present results, *m/z* 626 (neg) or 628 (pos), *m/z* 381 (neg) or 383 (pos), and *m/z* 343 (neg), corresponding to PB-*O*-glucuronide-*N*-acetylhexopyranosylamine, PB-*N*-glucuronide with

an opening of the barbital moiety (loss of CO) and dihydroxy-PB-sulfate, respectively, would have been discarded with a mass defect tolerance of 50 mDa. However, computational prediction of metabolite masses undergoing common biotransformations such as sulfation or glucuronidation may have helped to select these ions, but would also have led to the selection of many other endogenous compounds.

Nevertheless, metabolomics-based approaches to metabolism studies also have some drawbacks. They are complex and time-consuming since they require the development of several statistical procedures such as multivariate statistical analysis and autocorrelation matrices, followed by the acquisition and interpretation of CID spectra. Another limitation is the access to only relative quantitation. This is acceptable as far as metabolite identification is concerned. However, additional information about the abundances of the metabolites will require further experiments in order to achieve absolute quantitation. Moreover, low-intensity variables often result in the absence of CID spectra and reliable isotopic distribution, which are of value to constrain the elements entering molecular formulas. Further sample pretreatment, such as pre-concentration may therefore be required to obtain a signal intensity suitable for sequential MS<sup>*n*</sup> experiments or for well-defined isotopic pattern so as to identify compounds over 300 Da.

## CONCLUSION

Our data illustrate a new strategy for the analysis of MS metabolomic data, especially in the field of metabolite identification. First, the use of different chromatographic conditions can drastically impact on the chromatographic migration and ionization efficiency of compounds, thus providing a way to obtain more exhaustive information during metabolic fingerprint acquisition. Second, autocorrelation matrices combined with pattern recogni-

(44) Zhu, M.; Ma, L.; Zhang, D.; Ray, K.; Zhao, W.; Humphreys, W. G.; Skiles, G.; Sanders, M.; Zhang, H. *Drug Metab. Dispos.* **2006**, *34*, 1722–33.

tion deciphered more than 50% of the MS data redundancy to focus the identification process on the most relevant discriminating ions. Third, the LTQ-Orbitrap rapidly identifies ions of interest. These results also confirm that the metabolomic approach provides effective tools for in vivo drug metabolite screening.<sup>45</sup> However, two main limitations remain: the absence of absolute quantification with the MS-based untargeted metabolomic approach and the impossibility of a fully comprehensive analysis of the metabolome with one single analytical tool.

In our view, a promising strategy may be to combine UPLC with Orbitrap detection. In this way, UPLC would provide robustness, better sensitivity and better software-assisted peak extraction, whereas the Orbitrap would allow accurate mass measurement and multistep fragmentation for identification purposes. However, accurate mass measurements at ultrahigh resolution (over 30 000) with the Orbitrap require scan times of ~1 s, and typical peak widths achieved with UPLC in our experimental conditions are ~10 s. This could limit the performances of the UPLC/LTQ-Orbitrap coupling for data-dependent MS<sup>n</sup> acquisitions at ultrahigh resolution, which are very useful for ion filiation determination.

Compound identification remains a challenging task and thus highlights the crucial need for any reliable mathematical tool or complementary technique that simplifies it. Further NMR experiments are necessary to achieve comprehensive identification of compounds and to obtain quantitative results. However, NMR is less sensitive than LC/MS and often requires isolation and

(45) Plumb, R. S.; Stumpf, C. L.; Granger, J. H.; Castro-Perez, J.; Haselden, J. N.; Dear, G. J. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2632–8.

concentration of compounds prior to its implementation as a complementary technique to formally characterize signals initially detected by an LC/MS-based approach. Databases of compounds identified by LC/MS have to be enriched to take full advantage of the tremendous potential applications of metabolomics, as a tool for elucidation of interrelated and interregulated biological processes. In the latter application, autocorrelation may highlight correlation or anticorrelation between ions eluted at different retention times and thus could help piece together metabolic networks. Finally, once more comprehensive databases become available; the use of isotopic dilution should provide a way to achieve absolute quantitation as soon as fingerprints are acquired.

#### **ACKNOWLEDGMENT**

This work was supported by a research contract between CEA and Technologie Servier. We thank Arnaud Martel (GIPSI, CEA) for providing a very useful visual basic program run on Excel able to search for multiple databases and to sort and link results, and François Fenaille and Elisabeth Métreau-Grognon for their help with data handling and useful comments.

#### **SUPPORTING INFORMATION AVAILABLE**

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review January 14, 2008. Accepted April 16, 2008.

AC800094P



# Mass spectrometry-based metabolomics: combining statistical correlations and ultra-high resolution accelerates the characterization of discriminating signals.

*Erwan Werner<sup>§</sup>, Vincent Croixmarie<sup>&</sup>, Thierry Umbdenstock<sup>&</sup>, Eric Ezan<sup>§</sup>, Pierre Chaminade<sup>#</sup>, Jean-Claude Tabet<sup>¶</sup>, Christophe Junot<sup>§,\*</sup>.*

<sup>§</sup>CEA, DSV / iBiTec-S / SPI / LEMM, Batiment 136, 91191 Gif-sur-Yvette Cedex, FRANCE.

<sup>&</sup>Technologie Servier, 25/27 rue Eugène Vignat, 45000 Orléans, FRANCE

<sup>#</sup>Groupe de chimie analytique de paris sud – E.A. 4041, Faculté de pharmacie de Châtenay-Malabry, Université Paris XI, 5 rue J.B. Clément, 92296 Châtenay-Malabry cedex, FRANCE.

<sup>¶</sup>Laboratoire de Chimie Structurale Organique et Biologique, UMR 7613-CNRS, Université Paris 6, 4 place Jussieu, 75252 Paris cedex 5, FRANCE.

**Supporting information (SI)**

**Table S1:** Parameters of the statistical models.

<b>Conditions</b>	<b>Pre-processing</b>	<b>R2X<sup>a</sup></b>	<b>R2Y<sup>b</sup></b>	<b>Q2<sup>c</sup></b>
<b>Acidic positive</b>	Without	0.929	0.902	0.855
	Pareto <sup>d</sup>	0.410	0.933	0.917
	UV <sup>e</sup>	0.192	0.963	0.882
<b>Acidic negative</b>	Without	0.907	0.917	0.889
	Pareto <sup>d</sup>	0.303	0.965	0.928
	UV <sup>e</sup>	0.160	0.971	0.865
<b>Neutral negative</b>	Without	0.783	0.861	0.834
	Pareto <sup>d</sup>	0.430	0.922	0.895
	UV <sup>e</sup>	0.125	0.955	0.663

<sup>a</sup> R2X corresponds to the proportion of the variance of the variables that is explained by the model. <sup>b</sup> R2Y defines the proportion of the variance of the response variable that is explained by the model. <sup>c</sup> Q2 expresses the proportion of the variance of the variables that can be predicted by the model. <sup>d</sup> Data were first mean-centered and scaled to Pareto variance. <sup>e</sup> Data were first mean-centered and scaled to unit variance.

**Table S2.** Example of fragmentation observed for each PB metabolites

m/z	Elemental composition		Fragment ions				Putative structure	Conditions <sup>a</sup>
	Mass accuracy (ppm).		MS <sup>2</sup>		MS <sup>3</sup>			
203.0463	C <sub>10</sub> H <sub>7</sub> O <sub>3</sub> N <sub>2</sub>	(0.2)	160.0408 116.0512	C <sub>9</sub> H <sub>6</sub> O <sub>2</sub> N (2.3) C <sub>8</sub> H <sub>6</sub> N (5.6)	203 → 116 116.0511	C <sub>8</sub> H <sub>6</sub> N (4.2)	PB without ethyl lateral chain	I
231.0776	C <sub>12</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub>	(0.4)	188.07175 144.08217	C <sub>11</sub> H <sub>10</sub> O <sub>2</sub> N (0.3) C <sub>10</sub> H <sub>10</sub> N (2.0)	Not available		PB	I
247.0724	C <sub>12</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	(0.2)	218.03329 204.0667 176.07199	C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub> (-0.2) C <sub>11</sub> H <sub>10</sub> O <sub>3</sub> N (0.5) C <sub>10</sub> H <sub>10</sub> O <sub>2</sub> N (1.6)	247 → 204 176.0719	C <sub>10</sub> H <sub>10</sub> O <sub>2</sub> N (1.3) C <sub>10</sub> H <sub>9</sub> O <sub>2</sub> (1.7) C <sub>9</sub> H <sub>9</sub> O (3.5)	p-Hydroxy-PB	I
263.0675	C <sub>12</sub> H <sub>11</sub> O <sub>5</sub> N <sub>2</sub>	(0.6)	161.0611 160.0772 133.0663	C <sub>10</sub> H <sub>9</sub> O <sub>2</sub> (2.0) C <sub>10</sub> H <sub>10</sub> ON (2.3) C <sub>9</sub> H <sub>9</sub> O (3.3)	Not available		Dihydroxy-PB	I
265.0830	C <sub>12</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	(0.2)	234.0280 220.0615	C <sub>10</sub> H <sub>6</sub> O <sub>5</sub> N <sub>2</sub> (-0.9) C <sub>11</sub> H <sub>10</sub> O <sub>4</sub> N (-0.1)	265 → 204 → 186.05591 175.02777 147.03236	C <sub>11</sub> H <sub>6</sub> O <sub>2</sub> N (-0.8) C <sub>9</sub> H <sub>5</sub> O <sub>3</sub> N (1.6) C <sub>8</sub> H <sub>5</sub> O <sub>2</sub> N (-1.5)	Dihydrodiol-PB	I II
327.0295	C <sub>12</sub> H <sub>11</sub> O <sub>7</sub> N <sub>2</sub> S	(0.6)	283.0389 247.07221 241.0177 218.0333 204.0667 160.07703	C <sub>11</sub> H <sub>11</sub> O <sub>5</sub> N <sub>2</sub> S (-1.8) C <sub>12</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub> (-1.3) C <sub>10</sub> H <sub>9</sub> O <sub>5</sub> S (0.2) C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub> (-0.2) C <sub>11</sub> H <sub>10</sub> O <sub>3</sub> N (0.4) C <sub>10</sub> H <sub>10</sub> ON (1.5)	327 → 247 → 218.0333 204.06671 161.01102	C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub> (-0.0) C <sub>11</sub> H <sub>10</sub> O <sub>3</sub> N (0.5) C <sub>10</sub> H <sub>9</sub> O <sub>2</sub> (2.0)	p-hydroxy-PB-sulfate	I
343.0239	C <sub>12</sub> H <sub>11</sub> O <sub>8</sub> N <sub>2</sub> S	(-0.8)	263.06711 234.02821 220.06166	C <sub>12</sub> H <sub>11</sub> O <sub>5</sub> N <sub>2</sub> (-0.9) C <sub>10</sub> H <sub>6</sub> O <sub>5</sub> N <sub>2</sub> (0.0) C <sub>11</sub> H <sub>10</sub> O <sub>4</sub> N (0.6)	343 → 263 → 234.02765 220.06134 177.05560 155.04633	C <sub>10</sub> H <sub>6</sub> O <sub>5</sub> N <sub>2</sub> (-2.4) C <sub>11</sub> H <sub>10</sub> O <sub>4</sub> N (-0.9) C <sub>10</sub> H <sub>9</sub> O <sub>3</sub> (-0.7) C <sub>6</sub> H <sub>7</sub> O <sub>3</sub> N <sub>2</sub> (0.8)	Dihydroxy-PB-sulfate	II
381.1304	C <sub>17</sub> H <sub>21</sub> O <sub>8</sub> N <sub>2</sub>	(0.2)	337.1396 247.1084 192.0515 175.0251 174.0411 156.0305 130.0515 112.0411	C <sub>16</sub> H <sub>21</sub> O <sub>8</sub> N <sub>2</sub> (-2.7) C <sub>13</sub> H <sub>15</sub> O <sub>3</sub> N <sub>2</sub> (-1.7) C <sub>6</sub> H <sub>10</sub> O <sub>6</sub> N (0.6) C <sub>6</sub> H <sub>7</sub> O <sub>6</sub> (1.4) C <sub>6</sub> H <sub>6</sub> O <sub>5</sub> N (1.7) C <sub>6</sub> H <sub>6</sub> O <sub>4</sub> N (2.0) C <sub>5</sub> H <sub>6</sub> O <sub>3</sub> N (4.3) C <sub>5</sub> H <sub>6</sub> O <sub>2</sub> N (6.2)	381 → 192 → 174.0409 156.0304 133.0146 130.0513 112.0409 102.0566 100.0409 94.0305 89.0251 85.0300	C <sub>6</sub> H <sub>6</sub> O <sub>5</sub> N (0.6) C <sub>6</sub> H <sub>6</sub> O <sub>4</sub> N (0.8) C <sub>4</sub> H <sub>5</sub> O <sub>5</sub> (2.7) C <sub>5</sub> H <sub>5</sub> O <sub>3</sub> N (2.9) C <sub>5</sub> H <sub>6</sub> O <sub>2</sub> N (4.5) C <sub>4</sub> H <sub>6</sub> O <sub>2</sub> N (5.7) C <sub>4</sub> H <sub>6</sub> O <sub>2</sub> N (5.2) C <sub>4</sub> H <sub>5</sub> ON (6.7) C <sub>3</sub> H <sub>5</sub> O <sub>3</sub> (7.4) C <sub>4</sub> H <sub>5</sub> O <sub>2</sub> (5.8)	PB-N-glucuronide with an opening of the barbital moiety (loss of CO)	I
397.1242	C <sub>17</sub> H <sub>21</sub> O <sub>9</sub> N <sub>2</sub>	(0.1)	379.12027 336.10941 203.08203 160.07582 132.08087	loss H <sub>2</sub> O C <sub>16</sub> H <sub>18</sub> O <sub>7</sub> N (4.8) C <sub>11</sub> H <sub>11</sub> O <sub>2</sub> N <sub>2</sub> (2.6) C <sub>10</sub> H <sub>10</sub> ON (0.8) C <sub>9</sub> H <sub>10</sub> N (0.7)	397 → 160 → 132.08081	C <sub>9</sub> H <sub>10</sub> N (0.3)	p-Hydroxy-PB-O-glucuronide with an opening of the barbital moiety (loss of CO)	III
407.1095	C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	(-0.3)	364.1027 320.1133 231.0774 174.0411 114.02020	C <sub>17</sub> H <sub>18</sub> O <sub>8</sub> N (-1.5) C <sub>16</sub> H <sub>18</sub> O <sub>6</sub> N (-0.5) C <sub>12</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub> (-0.79) C <sub>6</sub> H <sub>8</sub> O <sub>5</sub> N (4.8) C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N (5.5)	407 → 364 → 320.1135 258.1134 174.0409 156.0304 114.0202	C <sub>16</sub> H <sub>18</sub> O <sub>6</sub> N (-0.2) C <sub>15</sub> H <sub>16</sub> O <sub>3</sub> N (1.5) C <sub>6</sub> H <sub>8</sub> O <sub>5</sub> N (0.9) C <sub>6</sub> H <sub>6</sub> O <sub>4</sub> N (1.1) C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N (4.9)	PB-N-glucuronide enantiomer	I
407.1096	C <sub>18</sub> H <sub>19</sub> O <sub>9</sub> N <sub>2</sub>	(-0.1)	364.1027 320.1132 231.0774 174.0411 114.0203	C <sub>17</sub> H <sub>18</sub> O <sub>8</sub> N (-2.9) C <sub>16</sub> H <sub>18</sub> O <sub>6</sub> N (-2.5) C <sub>12</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub> (-0.6) C <sub>6</sub> H <sub>8</sub> O <sub>5</sub> N (1.8) C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N (5.8)	407 → 364 → 320.1135 258.1133 218.0308 174.0410 156.0304 114.0202	C <sub>16</sub> H <sub>18</sub> O <sub>6</sub> N (-1.3) C <sub>15</sub> H <sub>16</sub> O <sub>3</sub> N (-1.2) C <sub>7</sub> H <sub>8</sub> O <sub>5</sub> N (0.6) C <sub>6</sub> H <sub>8</sub> O <sub>5</sub> N (0.9) C <sub>6</sub> H <sub>6</sub> O <sub>4</sub> N (1.3) C <sub>4</sub> H <sub>4</sub> O <sub>3</sub> N (4.7)	PB-N-glucuronide	I
410.1029	C <sub>17</sub> H <sub>20</sub> O <sub>7</sub> N <sub>3</sub> S	(0.5)	392.09161 263.04935	C <sub>17</sub> H <sub>18</sub> O <sub>6</sub> N <sub>3</sub> S (-1.4) C <sub>12</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub> S (-0.9)	410 → 392 → 263.04935	C <sub>12</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub> S (-0.9)	PB-mercapturic acic conjugate	I

423.1047	C <sub>18</sub> H <sub>19</sub> O <sub>10</sub> N <sub>2</sub>	(0.4)	405.0934	C <sub>18</sub> H <sub>17</sub> O <sub>9</sub> N <sub>2</sub>	(-1.3)	423 → 247 →	C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub>	(-0.25)	p-Hydroxy-PB-O-glucuronide	I
			380.0979	C <sub>17</sub> H <sub>16</sub> O <sub>9</sub> N	(-2)					
			379.1141	C <sub>17</sub> H <sub>19</sub> O <sub>7</sub> N <sub>2</sub>	(-1.6)					
			361.1039	C <sub>17</sub> H <sub>17</sub> O <sub>7</sub> N <sub>2</sub>	(-0.7)					
			336.1088	C <sub>16</sub> H <sub>18</sub> O <sub>7</sub> N	(-0.14)					
			247.0723	C <sub>12</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	(-0.5)					
			218.0332	C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub>	(-0.07)					
			204.0667	C <sub>11</sub> H <sub>10</sub> O <sub>3</sub> N	(0.3)					
			204.0666	C <sub>11</sub> H <sub>10</sub> O <sub>3</sub> N	(0.02)					
			175.0250	C <sub>6</sub> H <sub>7</sub> O <sub>6</sub>	(1.3)					
			157.0145	C <sub>6</sub> H <sub>5</sub> O <sub>5</sub>	(1.8)					
			129.0198	C <sub>5</sub> H <sub>5</sub> O <sub>4</sub>	(3.7)					
			113.0250	C <sub>5</sub> H <sub>5</sub> O <sub>3</sub>	(5.0)					
453.1152	C <sub>19</sub> H <sub>21</sub> O <sub>11</sub> N <sub>2</sub>	(0.2)	435.1027	C <sub>19</sub> H <sub>19</sub> O <sub>10</sub> N <sub>2</sub>	(-1.6)	453 → 277 →	C <sub>11</sub> H <sub>6</sub> O <sub>5</sub> N <sub>2</sub>	(-0.7)	O-methyl-PB-O-Glucuronide	I
			409.1237	C <sub>18</sub> H <sub>21</sub> O <sub>9</sub> N <sub>2</sub>	(-1.2)					
			366.1181	C <sub>17</sub> H <sub>20</sub> O <sub>8</sub> N	(-3.6)					
			277.0824	C <sub>13</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	(-2.0)					
			248.0436	C <sub>11</sub> H <sub>9</sub> O <sub>5</sub> N <sub>2</sub>	(-1.0)					
			175.0251	C <sub>6</sub> H <sub>7</sub> O <sub>6</sub>	(0.9)					
			113.0251	C <sub>5</sub> H <sub>5</sub> O <sub>3</sub>	(6.0)					
626.1837	C <sub>26</sub> H <sub>32</sub> O <sub>15</sub> N <sub>3</sub>	(-0.3)	405.0934	C <sub>18</sub> H <sub>17</sub> O <sub>9</sub> N <sub>2</sub>	(-1.5)	405 → 361 →	C <sub>17</sub> H <sub>17</sub> O <sub>7</sub> N <sub>2</sub>	(-1.4)	PB-O-glucuronide-N-acetyl-hexopyranosylamine	I
			361.1037	C <sub>17</sub> H <sub>17</sub> O <sub>7</sub> N <sub>2</sub>	(-1.2)					
			319.0821	C <sub>16</sub> H <sub>15</sub> O <sub>7</sub>	(-0.7)					
			247.0723	C <sub>12</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	(-0.6)					
			361.10350	C <sub>17</sub> H <sub>17</sub> O <sub>7</sub> N <sub>2</sub>	(-1.4)					
			319.08194	C <sub>16</sub> H <sub>15</sub> O <sub>7</sub>	(-1.2)					
			318.09799	C <sub>16</sub> H <sub>16</sub> O <sub>6</sub> N	(-1.0)					
247.07237	C <sub>12</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	(-0.2)								
218.0332	C <sub>10</sub> H <sub>6</sub> O <sub>4</sub> N <sub>2</sub>	(-0.5)								

<sup>a</sup> I: neutral-water-methanol eluting conditions, negative electrospray detection, II: acid-water-acetonitrile eluting conditions, negative electrospray detection, III: acid-water-acetonitrile eluting conditions, positive electrospray detection.



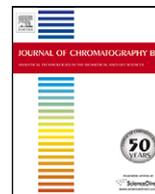


Werner E, Heilier JF, Ducruix C, Ezan E, Junot C, Tabet JC.

**Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends.**

*J Chromatogr B* 2008; 871(2):143-163.





## Review

# Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends<sup>☆</sup>

Erwan Werner<sup>a</sup>, Jean-François Heilier<sup>a,b</sup>, Céline Ducruix<sup>a,1</sup>, Eric Ezan<sup>a</sup>, Christophe Junot<sup>a,\*</sup>, Jean-Claude Tabet<sup>c,\*</sup>

<sup>a</sup> Commissariat à l'Energie Atomique, DSV-iBiTec-S-SPI, Laboratoire d'analyse du Métabolisme du Médicament, 91191 Gif-sur-Yvette, France

<sup>b</sup> Université catholique de Louvain, Faculty of Medicine, Industrial Toxicology and Occupational Medicine Unit, 1200 Brussels, Belgium

<sup>c</sup> Laboratoire de Chimie Structurale Organique et Biologique, CNRS UMR 7613, UPMC Univ. Paris 06, F-75005 Paris, France

## ARTICLE INFO

## Article history:

Received 25 February 2008

Accepted 1 July 2008

Available online 12 July 2008

## Keywords:

Metabolomics

Metabonomics

Identification

Mass spectrometry

## ABSTRACT

The metabolome is characterized by a large number of molecules exhibiting a high diversity of chemical structures and abundances, requiring complementary analytical platforms to reach its extensive coverage. Among them, atmospheric pressure ionization mass spectrometry (API-MS)-based technologies, and especially those using electrospray ionization are now very popular. In this context, this review deals with strengths, limitations and future trends in the identification of signals highlighted by API-MS-based metabolomics. It covers the identification process from the determination of the molecular mass and/or its elemental composition to the confirmation of structural hypotheses. Furthermore, some tools that were developed in order to address the MS signal redundancy and some approaches that could facilitate identification by improving the visualization and organization of complex data sets are also reported and discussed.

© 2008 Elsevier B.V. All rights reserved.

## Contents

1. Introduction .....	144
2. Mass spectrometry-based tools .....	145

**Abbreviations:** ACS, American Chemical Society; APCI, atmospheric pressure chemical ionization; API, atmospheric pressure ionization; APPI, atmospheric pressure photo ionization; ARM, atomic reconstruction of metabolism; ArMet, architecture for metabolomics; CAS, chemical abstract service; CE, capillary electrophoresis; ChEBI, chemical entities of biological interest; CID, collision-induced dissociation; CID-SORI, collision-induced dissociation-sustained off-resonance irradiation; CSLS, chemical structure lookup service; C-trap, curve linear trap (shaped like the letter "C"); DART, direct analysis in real time; DRE, dynamic range enhancement;  $E_{com}$ , energy at the centre of mass; EI, electron ionization (*a.k.a.* electronic impact); EID, electron-induced dissociation;  $E_{lab}$ , energy in the laboratory framework; ERMS, energy resolved mass spectrometry; ESI, electrospray ionization; eV, electron volt; FTICR, Fourier transform ion cyclotron resonance; FT/MS, Fourier transform mass spectrometry; FWHM, full width at half maximum (peak height); GC, gas chromatography; H/D exchange, hydrogen/deuterium exchange; HMDB, Human Metabolome DataBase; IR, infra-red; IRMPD, infra-red multiphoton dissociation; ITMS, ion trap mass spectrometer; KEGG, Kyoto Encyclopaedia Of Genes And Genomes; LC, liquid chromatography; LMSD, The Lipid Maps Structure Database; LTQ, linear ion trap; LTQ-Orbitrap<sup>®</sup>, linear ion trap-Orbitrap<sup>®</sup>; m/z, mass-to-charge ratio; MALDI, matrix-assisted laser desorption ionization; MeMo, metabolomic modelling project; METPR, metabolite enrichment by tagging and proteolytic release; MIAMET, minimum information about a metabolomics experiment;  $M_{iso}$ , monoisotopic mass; MoTo DB, the metabolome database for tomato; mRNA, messenger ribonucleic acid; MS, mass spectrometry; MS<sup>E</sup>, untargeted mixed-mode (low- and high-energy) tandem mass spectrometry; MS<sup>n</sup>, multistage (sequential) tandem mass spectrometry; MS/MS, tandem mass spectrometry; MVA, multivariate (data) analysis; MW, molecular weight; MySQL, My Structured Query Language; NCBI, National Center for Biotechnology Information; NCI, National Cancer Institute; NIST, National Institute of Standards and Technology; NMR, nuclear magnetic resonance; PHP, hypertext preprocessor; ppm, part per million; QqTOF, hybrid quadrupole time-of-flight mass spectrometer; Q TRAP<sup>®</sup>, hybrid quadrupole linear ion trap; RDBE, ring plus double bond equivalent; Rf-only, radiofrequency only; RIA, relative isotopic abundance measurement; SMB, supersonic molecular beam; SQL, Structured Query Language; TAIR, The Arabidopsis Information Resource; Th, Thompson; TOF, time of flight; TQ, triple quadrupole; u, atomic mass unit; UPLC, ultra performance liquid chromatography; UV, ultra-violet; UV-vis, ultra-violet-visible; XML, eXtensible Markup Language.

<sup>☆</sup> This paper is part of a special volume entitled "Hyphenated Techniques for Global Metabolite Profiling", guest edited by Georgios Theodoridis and Ian D. Wilson.

\* Corresponding authors.

E-mail addresses: [christophe.junot@cea.fr](mailto:christophe.junot@cea.fr) (C. Junot), [tabet@ccr.jussieu.fr](mailto:tabet@ccr.jussieu.fr) (J.-C. Tabet).

<sup>1</sup> Present address: The Cancer Research UK Centre for Cancer Therapeutics, The Institute of Cancer Research, 15 Cotswold Road, Belmont, Sutton, Surrey, SM2 5NG, United Kingdom.

2.1.	The analyzers: mass resolving power and accuracy .....	146
2.2.	The determination of elemental compositions .....	146
2.3.	The different kinds of MS/MS experiments .....	149
2.3.1.	Fragmentation spectra: various ion activation modes.....	149
2.3.2.	Collision-induced dissociation .....	149
2.3.3.	Other activation modes.....	150
2.3.4.	Mass analyzers and MS/MS experiments.....	150
2.4.	Complementary approaches.....	151
2.4.1.	H/D exchange.....	151
2.4.2.	Derivatization strategies.....	151
3.	Databases.....	151
3.1.	General chemical databases .....	152
3.2.	Biochemical and metabolic databases .....	152
3.3.	Metabolomic databases .....	153
3.4.	Spectral databases .....	153
4.	Endogenous metabolite identification: some case studies .....	153
4.1.	First situation: the hypothetic metabolite is described in biochemical or metabolomic databases.....	154
4.2.	Second situation: the hypothetical structure of the metabolite cannot be obviously deduced from databases.....	155
5.	How to address MS signal redundancy?.....	158
6.	Mathematical tools to improve analysis of MS data.....	158
6.1.	Kendrick plot.....	158
6.2.	van Krevelen plot.....	159
7.	Conclusion and perspectives .....	160
7.1.	Tools for the standardization of metabolomic data.....	160
7.2.	Analytical platforms .....	160
7.3.	Toward relational databases .....	160
7.4.	Statistical and mathematical tools.....	161
	Acknowledgments .....	161
	References.....	161

## 1. Introduction

Metabolomics deals with the comprehensive analysis of metabolites present in a biological sample by the combined use of analytical methods and multivariate statistical analyses (MVA) [1,2]. It emerged at the end of the 1990s as the third major path of functional genomics beside mRNA profiling (transcriptomics) and proteomics [3,4].

Metabolomics experiments start with the acquisition of metabolic fingerprints using various analytical platforms (RMN, GC/MS, LC/MS, FT/IR). When using MS-based methods, the resulting fingerprints are then pre-processed (*i.e.*, suppression of background chemical noise, variable alignment, peak picking) [5] before being eventually pre-treated (centering, scaling, etc.) [6], and finally processed. Processing generally relies on MVA, which facilitate data visualization through data dimensionality reduction and highlight relevant biological information [7,8]. Finally, identification of the discriminating signals is undertaken by combining mass spectrum analysis and database consultation.

Four particular issues should be addressed in the frame of metabolomics. First, the large diversity of metabolites in terms of chemical structures and concentrations prohibits the coverage of the whole metabolome with a single analytical method. Several analytical platforms are therefore used, relying either on nuclear magnetic resonance (NMR) [9], or on mass spectrometry with different ion sources and mass analyzers [10–19]. Each tool provides complementary but sometimes redundant information.

Secondly, processing the data prior to MVA may also be a source of errors. Artifactual signals could be generated due to erratic chemical background noise suppression or insufficient correction of retention time shifts when coupling with liquid chromatography (LC), and metabolites present in trace amounts or which poorly ionize may not be extracted due to low-intensity signals. In addition, the consistency of MVA results may be impacted not only by ana-

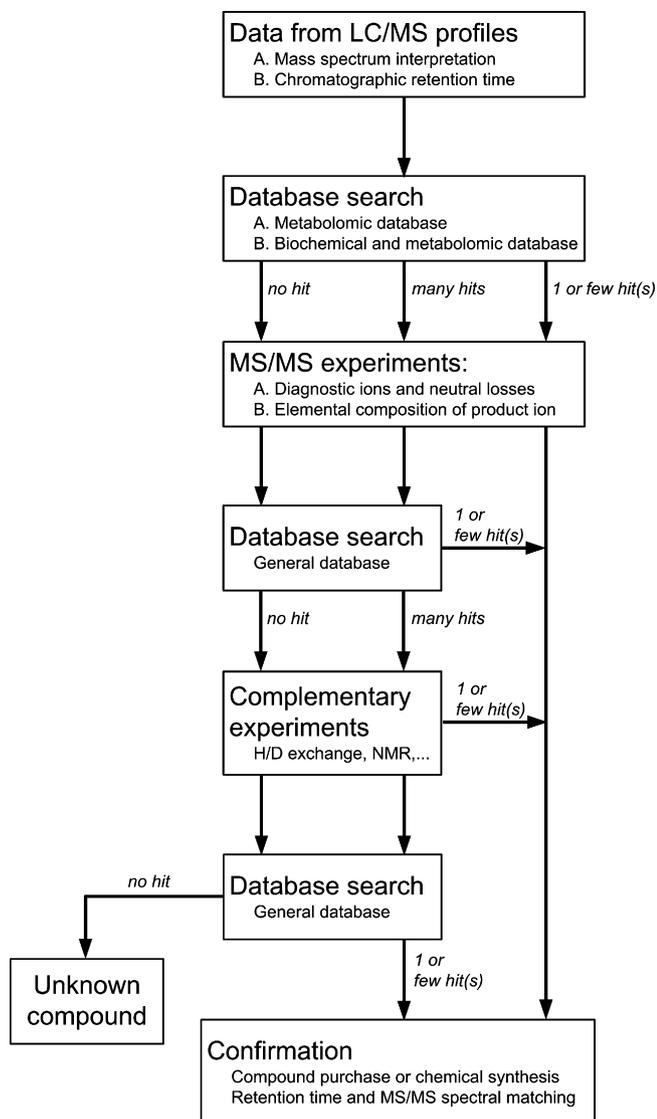
lytical and data pre-processing issues, but also by the scaling and normalization procedures [20–22].

Thirdly, another matter of concern is the high number of MS signals to be identified after MVA. A metabolite does not produce a single  $m/z$  peak since adduct and product ions can be generated during the desolvation step following the API process. Although informative, such a signal redundancy slows down the identification procedure by increasing the number of variables that have to be investigated and also by leading to unsuccessful database queries.

Fourthly, the identification of discriminating signals is perhaps the most laborious and time-consuming step of the metabolomics experiment. It is usually achieved by searching databases with the recorded metabolite molecular mass, the elemental composition, or the whole mass spectra. Rigorously, a panel of complementary spectroscopic methods such as NMR, UV, MS and IR detection is required to reach this goal [23,24].

Atmospheric pressure ionization mass spectrometry (API-MS)-based techniques have rapidly emerged as a popular and powerful tool for metabolomics. When combined with liquid chromatography, they exhibit a good sensitivity, high dynamic range and versatility but also provide soft ionization conditions giving access to the molecular mass of intact molecules from complex mixtures. Furthermore, the development of high and ultra-high resolution analyzers (TOF/MS and FT/MS, respectively) has yielded accurate mass measurements, whereas ion trap analyzers, among others, are used to perform  $MS^n$  experiments to get the additional structural information needed for metabolite identification [25,26]. Unfortunately, API-MS exhibits poor reproducibility and high inter-instrument variability in the generation of fragmentation patterns, and this has hampered the constitution of universal databases as done with electron ionization mass spectrometry [27] or with NMR [28].

For all these reasons, formal identification of metabolite structure is not easily achieved by using the API-MS experiments alone.



**Fig. 1.** Flowchart of a proposed strategy for the identification of potential biomarkers highlighted by MS-based metabolomics using complementary LC/API-MS(/MS) approaches and database queries.

As a result, most MS signals detected in metabolic fingerprints remain unidentified.

Recently, Chen et al. [29] reported an integrated approach for the identification of biomarkers detected in LC/MS fingerprints, based on LC/MS<sup>2</sup> experiments, micropreparative ultra-performance liquid chromatography (UPLC), Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS), gas chromatography (GC) retention time indices, database search and generation of stable-isotope-labeled standards. Based on the work of Chen et al. [29] and in our experience, a flow chart of the identification process of signals from metabolomics relying on complementary API-MS strategies is presented in Fig. 1. It starts with an interpretation of the mass spectra in order to ensure that the signal of interest corresponds to a monoisotopic ion and not to a natural isotopologue ion or an adduct ion. One or more relevant elemental compositions are deduced from accurate mass measurements if high or very high-resolution mass spectrometry is available, for further database queries. Collision-induced dissociation (CID) spectra are then acquired and interpreted in order to get information about the chemical structure. At this stage, chemical database queries may

be refined and the highlighted compounds, if any, are kept for further consideration or ruled out based on chromatographic retention time and CID mass spectra information. Complementary experiments (*i.e.*, other sequential MS<sup>n</sup> experiments or H/D exchanges) may be required to deduce a putative structure. Finally, formal identification is achieved when the metabolite to be characterized exhibits the same retention time and CID spectra as those of the reference molecule that has to be purchased or synthesized.

In this context, this review deals with advantages, limitations and future trends of API mass spectrometry for the identification of signals from LC/MS-based metabolomics. It aims to cover the identification process, from the determination of the molecular mass and/or elemental composition, to the confirmation of structural hypotheses. Furthermore, as the identification process cannot be reduced solely to the interpretation of mass and CID spectra, some complementary tools developed to address MS signal redundancy and some approaches that could facilitate identification by improving the visualization and organization of complex data sets will also be reported and commented upon.

## 2. Mass spectrometry-based tools

The coupling of liquid chromatography with mass spectrometry uses API sources to convert aqueous LC effluent containing analytes to gas-phase molecules. Ionization at atmospheric pressure differs from ionization under high vacuum conditions by the ability of the generated precursor ions to dissipate part of their acquired internal energy. In high-vacuum ionization conditions, the internal energy is not strictly controlled, ranging from 0 to 15 eV for electron ionization and being less than 3 eV for chemical ionization [30]. The situation is entirely different under API conditions for which the ion activation/cooling allows the production of more or less energetic charged molecular species [31]. By using gas phase ionization, *via* atmospheric pressure chemical ionization (APCI)[32], atmospheric pressure photo ionization (APPI) [33] or direct analysis in real time (DART) [34,35], it is possible to produce stable MH<sup>+</sup> (or [M+NH<sub>4</sub>]<sup>+</sup>) and [M-H]<sup>-</sup> (or [M-X]<sup>-</sup>, X being an halogen anion or other deprotonated organic and inorganic acid) and to a lesser extent, M<sup>2+</sup> or M<sup>3+</sup> (mainly in APPI). With electrospray (ESI) [36] and desorption electrospray ionisation [37–39], although ions are produced by desorption, the desolvation occurring at the skimmer produces the same ions as those observed under gas phase ionization, except that ESI is known to produce multiply charged ions. Finally, under API conditions, ion aggregates are desolvated at the skimmer where the internal ion energy can be modulated through the controlled nozzle-skimmer potential difference in order to either analyze intact species or obtain in source CID spectra [40,41].

Among the high diversity of analyzer technologies that can be implemented with LC/ESI-MS systems, the analyst has the choice between two possible approaches: “in-space” instruments, and “in-time” mass spectrometers [42,43]. Two kinds of “in-space” instruments can be distinguished: (i) without field scanning devices (*i.e.*, TOF/MS-MS), and (ii) electrostatic or magnetic field scanning instruments (*i.e.*, quadrupole and sector analyzers, respectively).

“In-time” mass spectrometers are based upon ion storage in ion traps, enabling sequential MS/MS experiments. Different kinds of ion traps are commercially available: some of them rely on the scanning of quadrupolar or nonlinear fields, which is the case of 3D (Paul’s trap) [44–46] and 2D (linear) ion traps [17,47], whereas others are based on ion trapping by using either an electromagnetic field (*i.e.*, the Penning trap used for ICR instruments) [48] or a quadro-logarithmic electrostatic field, as for the Kingdon cell (*i.e.*, the Orbitrap®) developed by Makarov and coworkers [49–52].

**Table 1**  
Typical values for mass resolving power and mass accuracy of commercially available analyzers

Analyzers	Mass resolving power <sup>a</sup>	Mass accuracy
Triple quadrupoles and ion traps	~500	~0.3 to 1 u
TOF devices	8000–20,000	1–5 ppm <sup>b</sup>
LTQ-Orbitrap	Up to 100,000	<3 ppm
FTICR	Up to 1,000,000 <sup>c</sup>	<1 ppm

<sup>a</sup>  $m/\Delta m$ , FWHM.

<sup>b</sup> When internal calibration or lock mass is used.

<sup>c</sup> Higher values can be achieved depending on the magnetic field intensity.

The Orbitrap<sup>®</sup> operates by radially trapping ions around a central spindle electrode and then generates a transient ion signal related to the axial ion motion rather than to the radial ion motion used with FTICR. Oscillation frequencies for stored ions with different  $m/z$  ratios are then obtained by using Fourier transform processes, resulting in an accurate reading of  $m/z$  ratios through the mass scale.

Multi-stage mass spectrometers (*i.e.*, tandem instruments) can be constructed by coupling the large variety of analyzers, leading to either homogenous tandem mass spectrometers such as TOF–TOF, triple quadrupole or multipoles, 2D and 3D (linear) ion traps, or hybrid instruments, such as QqTOF, Q TRAP<sup>®</sup>, and 2D trap coupled with ICR cell or Orbitrap<sup>®</sup>.

For identification purposes, the performance of the different kinds of analyzers will be discussed through their mass resolving power and mass accuracy (high and ultra-high resolution instruments), which are two major parameters involved in the determination of elemental compositions. In addition, their ability to provide structural information will be addressed through their capacity to perform two-, three- or multi-stage MS/MS (*i.e.*, MS<sup>*n*</sup>) experiments. The issue of spectral database building with API data will also be discussed.

### 2.1. The analyzers: mass resolving power and accuracy

The mass resolving power is the capacity of a mass spectrometer to separate ions of adjacent but different  $m/z$  ratios. It is defined as the ratio of the measured mass  $m$  to  $\Delta m$ , the full width of the peak at half its maximum height (*i.e.*,  $m/\Delta m$ , full width at half maximum peak height, FWHM). The mass accuracy is the difference between the exact mass of an ion and its measured mass and is commonly expressed as parts per million (ppm) [43,53,54]. Mass resolving powers and mass accuracies of commercially available analyzers discussed below, are displayed in Table 1.

The mass resolving power of low-resolution analyzers such as ion trap and quadrupole devices does not exceed 500 at  $m/z$  250 ( $m/\Delta m$ , FWHM) [55]. As a consequence, they permit the separation of adjacent peaks by 1 Th, but not that of isobaric ions at the same nominal  $m/z$  ratio. Higher resolution mass spectra can be obtained by reducing the scan rate (up to 5000 at  $m/z$  250 using the Ultra ZoomScan<sup>®</sup> mode on a 2D ion trap), but the acquisition window is then limited to the 100 Th scale.

High-resolution analyzers based on TOF technology (TOF and hybrid Q-TOF mass spectrometers) provide accurate mass measurements with errors below 5 ppm when internal calibration or lock mass is used, and achieve isobaric ion separation within the limits of their resolving power ranging from 8000 to 20,000 ( $m/\Delta m$ , FWHM). However, TOF analyzers suffer from a modest dynamic range that is typically around two to three orders of magnitude. This has been improved to four orders of magnitude by using a new technology named Dynamic Range Enhancement (DRE<sup>®</sup>) [56,57].

By using FT-ICR devices, elemental compositions of low molecular weight analytes (*i.e.*, below 1000 Da) are provided by very accurate mass measurements with sub-ppm errors in mass spec-

tra. However, FT-ICR devices are not popular because they utilize strong and maintenance-expensive magnetic fields. Accurate mass measurements are achieved at the expense of longer recording times (transient signal digitizing duration  $\geq 1$  s). This may limit their coupling with LC and especially ultra high pressure liquid chromatography (*i.e.*, UHPLC) [14]. The increase in superconducting magnetic fields up to 12 T allows better resolution for lower transient signal lifetimes.

More recently, a new technology was introduced with the release of the LTQ-Orbitrap<sup>®</sup> [52]. This new type of hybrid mass spectrometer consists of a linear ion trap coupled to an Orbitrap analyzer. The linear ion trap (LTQ) [58] is able to record its own full-scan mass spectra and sequential MS<sup>*n*</sup> activation spectra from low-resolution precursor ion selection (mass window  $\geq 0.3$  u). Orbitrap<sup>®</sup> devices achieve accurate mass measurements with errors below 3 ppm and resolving powers up to 100,000 ( $m/\Delta m$  at  $m/z$  1000, FWHM). Their performances in terms of  $m/z$  ratio stability and dynamic range of mass accuracy have been evaluated by serial injections of five reference compounds [59]. The Orbitrap<sup>®</sup> provided accurate mass measurements (error < 2 ppm) for over 24 h post-calibration, whereas the dynamic range of mass accuracy was over three orders of magnitude. Similar values of mass accuracy were found in biological matrices in other studies [60,61].

High and ultra-high resolution analyzers are becoming increasingly popular in the field of metabolomics because they provide accurate mass measurements, which are useful for the determination of elemental compositions of metabolites, together with MS/MS or sequential MS<sup>*n*</sup> experiments which provide useful additional structural information, especially when product ions are analyzed at high resolution. However, there is a lack of comparative studies about their analytical performances in biological matrices in terms of dynamic ranges of concentration and mass accuracy.

### 2.2. The determination of elemental compositions

When accurate mass measurements are available, the elemental composition determination of monoisotopic ions (mass noted as  $M_{\text{iso}}$ ) is the first step of metabolite characterization from the mass spectrum because it provides a simple, efficient and automatable way to search chemical and metabolic databases.

It is often believed that mass accuracy is the most important parameter for the determination of elemental compositions. Nevertheless, the number of possible elemental compositions increases exponentially with increasing ion mass, even with ultra-high resolution instruments (*e.g.*, 15 different molecular formulas for mass > 600 Th within a 1-ppm window) [62]. Therefore, restrictive criteria based on physico-chemical rules and spectral information, as found in mass spectrometry textbooks [43,53] (*i.e.*, nitrogen rules, valence considerations, isotopic patterns), are required in order to remove irrelevant proposals.

The most popular of these chemical rules is the nitrogen rule that states that odd nominal molecular mass compounds contain an odd number of nitrogen atoms [54,63]. It is important to keep in mind that ESI produces almost exclusively even electron ions (during the ionization step) and therefore facilitates the determination of the integer nominal mass of the compound. Instrument software that calculate the elemental compositions can automatically take care of the nitrogen rule, which is also important for the product ions since it helps to determine the number of nitrogen atoms lost or if the product ion is a radical (*i.e.*, odd electron). Nevertheless, such a rule becomes unreliable for masses above 500 u because the large number of elements contributing to the mass may result in a mass defect above 0.5 u that leads to overestimation of the integer mass by 1 u.

The number of ring plus double bond equivalents (RDBE) in a molecule consisting of  $C_aH_bN_cO_d$  can be evaluated by the formula:  $RDBE = a - (1/2)b + (1/2)c + 1$  [64]. Depending upon their valence, other elements should be counted as C, H or N. For example, in biochemical applications, halogen atoms that have a valence state of 1 should be counted as hydrogen atoms, whereas O, which has a valence of 2 has not to be included with C, H, and N since it has no impact on the number of H attached to a given C. RDBE for singly protonated ions should be above  $-0.5$  for an elemental composition to be theoretically possible. Furthermore, the RDBE formula distinguishes between odd- and even-electron ions, since RDBE values are integers in the first case and non-integers with a remainder of 0.5 in the second case [54]. The RDBE formula only takes account of the lowest valence state of elements (e.g., 3 and not 5 for phosphorus). Therefore, RDBE determinations of phosphorus-containing molecules may be of limited interest since more than one RDBE value should be obtained [65,66]. In addition, in a biological context, phosphorus is often present in a valence state of five (i.e., as phosphate), which is not considered by the RDBE formula. The situation is far more complicated with sulfur that can have three different valences (i.e., 6, 4 and 2 for sulfone, sulfoxide, and thiol or thioether, respectively). Moreover, if there are two or more sulfur atoms in a given molecule, they can be simultaneously present at different valence states. Advanced softwares calculate a range for RDBE but none to our knowledge uses every valence states of elements to return all possible results because it would require too much calculation time. This shows the limitation of the use of the RDBE formula which is not of great help for advanced structure elucidation when elements other than C, H, N, O are concerned.

Analysis of isotopic patterns is helpful in elemental composition determination. Kind and Fiehn [67] showed that interpretation of isotopic abundance patterns removes more than 95% of false candidate formulas for molecules above 500 u. Moreover, they concluded that instruments with 3 ppm mass accuracy and 2% relative error for isotopic abundance pattern outperformed those with less than 1 ppm accuracy that do not include isotope information in the calculation of molecular formulas [67].

Instruments providing ultra-high resolution (i.e., Orbitrap® and FT-ICR) are of special interest in this context. Thanks to their resolving power, such instruments allow resolution of isotope peaks within isotope clusters. So, the isotopic contribution of each element may be isolated so as to give evidence of atoms with particular isotopic pattern or determine atom amounts.

In addition, the charge state could be determined based upon the fact that the smallest mass difference arising from natural isotopic variations in a molecule is close to 1 u. Thus, the spacing between the  $^{12}C_n$  and  $^{12}C_{n-1}^{13}C_1$  (isotopologue ions) on the  $m/z$  scale is  $(1\text{ u})/z$ , which allows the charge to be directly elucidated by measuring the  $m/z$  variation between the isotopologue ( $^{12}C_n$ ,  $^{12}C_{n-1}^{13}C_1$ ,  $^{12}C_{n-2}^{13}C_2$ , and so on) ions [63,68].

A typical example of the determination of a chemical formula from a LTQ-Orbitrap® mass spectrum is provided in Fig. 2. The ESI mass spectrum has been extracted from an LC/MS chromatogram of rat urine in the course of a metabolomics experiment. The discriminating signal to be identified is an ion at  $m/z$  410.1025 in negative mode. The analysis of the mass spectrum before the calculation of possible formulas by the Qualbrowser/Xcalibur® post-processing software reduces the number of hypothetical structures. Analysis of the isotopic pattern reveals: (i) the presence of the first isotope cluster at  $(M_{iso} + 1)$ , indicating that the ion is singly charged, (ii) the peak at  $(M_{iso} + 1.0034)$  that exhibits a relative abundance of 18% indicates a number of carbon atoms close to 16 (compared with  $^{12}C$ ,  $^{13}C$  shows 1.1% relative abundance and a difference in mass of 1.00335 u), and (iii) the presence of a single sulfur-atom in the elemental composition (which simultaneously excludes the presence

of Cl, Br or K), thanks to the separation of the  $^{34}S$  (peak of 4.4% relative intensity at  $(M_{iso} + 1.9959)$ ) from the  $^{13}C_2$  isotopologue at the  $(M_{iso} + 2)$  isotope cluster.

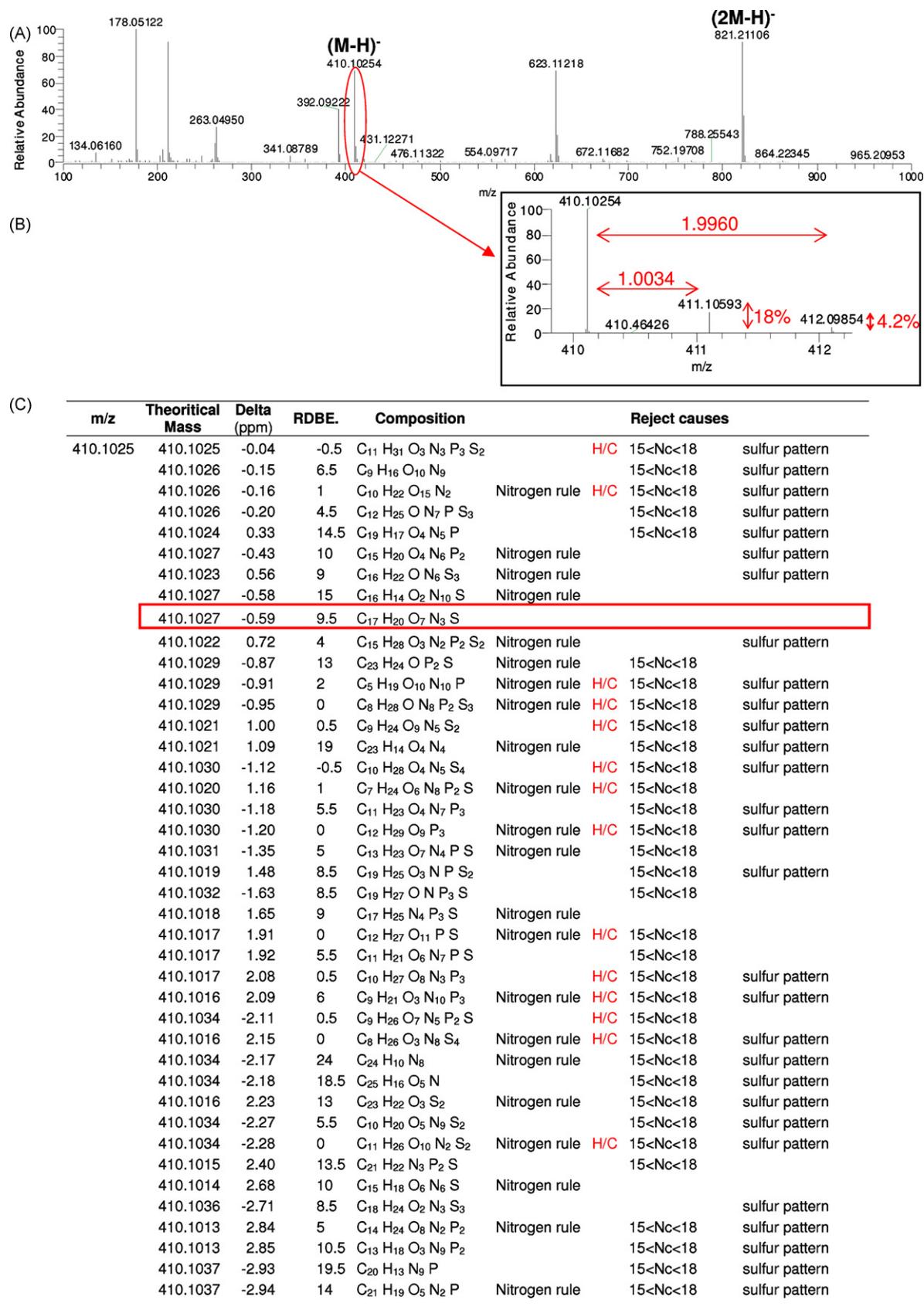
The Xcalibur® formula generator returned 41 possible elemental compositions within 3 ppm for the measured mass at  $m/z$  410.1025, as shown in Fig. 2. If element settings are restricted with (i) at least one sulfur atom and (ii) 15–18 carbon atoms (corresponding to a 10% relative error in isotopic ion abundance measurement), then only four formulas remain. In addition, according to the classical nitrogen rules previously mentioned, the ion of interest should have an odd number of nitrogen atoms. Finally, a single formula remains:  $C_{17}H_{20}N_3O_7S$ . Simulation of the proposed elemental compositions under the recorded mass spectrum supports this elemental composition.

However, there are some limitations in the use of isotopic patterns for the selection of relevant elemental compositions in biological media. Isotope ratio measurements can be affected by interference and saturation effects due both to the matrix and the experimental conditions (that increase spectral chemical background noise or limit sensitivity). In addition, there is very limited information available in the literature about the precision of relative isotopic ion abundance measurements (RIA) because it depends upon both the matrix and the experimental conditions, and also because many signals cannot be formally identified. As a matter of fact, although it has been stated that isotopic ratios can be measured within a 2% signal intensity precision by using TOF devices [69], Grange et al. [70] reported relative isotopic abundance measurement errors in a hydro-organic medium within 20% of their predicted values for 35 compounds whose ion  $m/z$  ratios exceeded 140 Th by using a single MS stage orthogonal accelerated TOF mass spectrometer [70]. Other studies reported excellent agreement between experimental and theoretical distributions for selected ions [60]. However, no statistics about a significant number of measures on replicates and also different ions were provided.

In this context, it could be relevant to use stable isotope labeling when available (e.g., for microorganism or plant metabolomics). The comparison of the monoisotopic masses from unlabeled and labeled biological extracts gives access to the number of both C and N atoms, decreasing the number of possible chemical formulas. A proof of concept has been performed on extracts from the model plant *Arabidopsis thaliana*: double isotopic labeling ( $^{13}C$  and  $^{15}N$ ) resulted in unique assignment of 87% of 5000 formulas versus 20% with a typical approach [71].

Another challenge of metabolomics is to deal with the huge number of masses to be investigated. It is then imperative to develop automatic procedures for the determination of relevant elemental formulas. In this context, pioneering work has already been performed in the analysis of petroleum oil and natural organic matter by using FTICR-MS [72]. Kujawinski and Behn [73] proposed an automated compound identification algorithm for the analysis of ultra-high resolution mass spectra of natural organic matter. Their approach looks for functional group relationships, such as  $CH_2$ ,  $CO_2$ ,  $NH$  or  $C_2H_2O$ , between all compounds within mass spectra. When applied to Suwannee River dissolved organic matter, such a procedure led to elemental composition assignments in 80–90% of the cases, whereas only 44% of assignments were obtained with bacterial dissolved organic matters. This example stresses the need for additional restrictive criteria when considering biological data in metabolomics.

Koch et al. also proposed some suitable techniques to be implemented in software for the analysis of ultra-high resolution mass spectra [62]. Formulas are generated without any restriction on hypothetical elements, but exclusion criteria, such as instrument error, the nitrogen rule, RDBE, and thresholds for molecular element ratios, are included. Two additional steps are added: (i) the



**Fig. 2.** The determination of relevant elemental compositions. (A) Full-scan ESI spectrum at the retention time of the MS signal to be identified. Acquisition was performed on a LC/LTQ-Orbitrap® system scanning within the range 75–1000Th at a mass resolving power of 60,000 ( $m/\Delta m$ , FWHM) in the negative-ion mode. The natural isotopic abundance distribution is given in an inset (B) and indicates an approximate number of carbon atoms equal to sixteen as well as the presence of a sulfur atom. (C) Table of the molecular formulas returned by the post-processing software based on the accurate mass measurement of  $m/z$  410 (assumed mass accuracy better than 3 ppm), and summary of their reasons for rejection.

first one consists in sorting molecular formulas into homologous series (e.g., molecules differing by  $\text{CH}_2$  units) and (ii) the second one deals with the exploitation of mass spectral information such as isotope ratios and predicted carbon numbers for intense peaks. A similar approach based on compound classification into homologous series followed by formula extrapolation from lower to higher masses has also been successfully applied to crude petroleum oil by using Kendrick plots [74]. This will be addressed further in this article.

Finally, Kind and Fiehn described an algorithm for filtering molecular formulas which is derived from seven heuristic rules: (i) restriction on the number of elements, (ii) Lewis and Senior rules, which filter elemental compositions on the basis of valence considerations, (iii) isotopic pattern, (iv) H/C ratios, which are comprised between 0.125 and 2 in most cases, (v) heteroatom ratios, i.e. (NOPS)/C ratios, (vi) element probabilities and (vii) the presence of trimethylsilylated compounds (for GC/MS applications) [66]. All the generated formulas, which comply with the rules, are also checked for presence in a customizable list of chemical databases. These rules have been implemented in an automated Excel<sup>®</sup> script [66]. The adduct ion removal must be done manually before entering any values, but an adduct ion mass calculator is provided.

### 2.3. The different kinds of MS/MS experiments

LC/MS based on various ionization and scanning techniques can be considered as the first step of the identification process since it provides information about the metabolite molecular mass and elemental composition, when high-resolution analyzers are used. Although direct introduction of samples into the mass spectrometer is possible, it may suffer from lack of sensitivity in complex biofluids such as plasma or urine due to matrix effects, and it does not address the issue of isomeric compounds. However, LC/MS experiments only provide information restricted to retention times and molecular weights and therefore must be complemented by further LC coupled to tandem mass spectrometry experiments to access structural information. The following section deals with the different types of ion activation that are relevant in the field of metabolomics, i.e., collision induced dissociations and infra-red multiphoton dissociation (IRMPD) [75], and also with several LC coupled to tandem mass spectrometry approaches that have been reported in the literature for metabolomics.

#### 2.3.1. Fragmentation spectra: various ion activation modes

Ion dissociation of metabolites is related to the internal energy as well as the electron state (i.e., open- or closed-shell species, which correspond to odd- and even-electron ions, respectively) of singly or multiply charged positive or negative ions. Odd-electron ions that are, for example, produced in electron ionization present high internal energy. Their gas phase stability is also lower than that of even-electron species, when they are formed under high-vacuum conditions in the ion source. This explains why many product ions are observed in EI mass spectra. Conversely, and in many cases (i.e., ESI, APPI or APCI), the API process generates stable molecular species yielding limited “in-source fragmentations”. Furthermore, these generated ions can be charged at various functionalized positions according to the experimental source conditions (solvent included), which results in an enhanced competitive dissociation of precursors. This yields a broad variety of fragmentations useful for ion structure elucidation.

There are two different modes of ion activation: (i) ergodic activation modes that deal with collisional to internal energy transfers, such as collision-induced dissociation (CID) and infra-red multiphoton dissociation (IRMPD) [76,77], and (ii) chemical activation

modes, such as electron capture dissociation [78], which involve electron transfers.

#### 2.3.2. Collision-induced dissociation

CID remains the most common ion activation technique used in current instruments. Collisions between a fast precursor ion and a neutral target gas lead to an increase in internal energy of the ion through the conversion of part of its translational energy into internal energy [79]. If the internal energy imparted is beyond the threshold for dissociation, precursor ion fragmentation may occur. Depending on instruments, collision energies can be in the keV range (high-energy collisions) or in the 1–200 eV range (low-energy collisions).

**2.3.2.1. High-energy collisions.** High energy collision conditions allow access to selected precursor ion excitation through electronic levels and indirectly through the vibrational levels of its electronic ground state [80,81]. The reproducibility of fragmentation patterns under variable collision conditions such as variations in high kinetic energy, target pressure, or initial internal energy is noteworthy and very useful for the building of product ion libraries from MS/MS experiments.

Such an excitation mode is mainly possible with TOF/TOF tandem instruments [82,83], although it has been used with multiselector instruments for a long time [84]. However, TOF-TOF instruments [85] are only combined with MALDI and other laser desorption sources which are not suitable for LC coupling. Recently, a shotgun metabolomic approach has been developed using MALDI-TOF/TOF MS from mouse heart tissues [86]. From 285 peaks corresponding to negatively charged ions, 90 metabolites were identified and discussed. Further developments, enabling TOF-TOF operation with spray ionization sources could be very promising.

Among the most specific processes, charge remote processes involve homolytic pathways where the charge is not involved in the fragmentation [87–90]. Such dissociation takes place from deprotonated fatty acids (unsaturated, branched and oxidized side chain) or their cationized forms in positive ion mode. It occurs statistically along the hydrocarbon chain. Furthermore, using such an approach allows the localization of alkyl, double bond, hydroxyl, or ketone groups. It was applied to various compounds, including phospho- and glycolipids, acylglycerols and steroids which are relevant compounds for metabolomic applications [91].

Fragmentations promoted by charge occur also under high-energy collision conditions. However, distribution of ion species and their respective abundances are very different from those observed with charge remote fragmentation processes, as shown by a recent study of March et al. [92].

**2.3.2.2. Low-energy collision.** Low collision energy experiments (i.e.,  $E_{\text{lab}}$  lower than 200 eV,  $E_{\text{lab}}$  representing the ion kinetic energy in the laboratory reference frame) are the most popular activation methods. In low collision energy experiments, the selected precursor ions are exclusively vibrationally excited. In high energy CID, often using helium as target, less than 10–30% of precursor ions undergo dissociation, which results in a low product ion recovery efficiency (ion beam sputtering), whereas low energy CID can yield a total disappearance of precursor ions that is accompanied by a better product ion recovery. However, while the CID spectra recorded at several keV are reproducible, independently of the instrument, this is not so for low-energy CID spectra, which strongly depend upon (i) the instrument characteristics (e.g., quadrupole versus ion trap or TOF mass spectrometers, for example), (ii) lens potentials, (iii) the collision energy, i.e.,  $E_{\text{lab}}$  values and (iv) the target gas pressure, a parameter that is difficult to control. A solution to this problem is to calibrate the collision energy, i.e., to work in

constant experimental conditions in which the relative abundances of both the precursor and product ions are maintained constant. This can be achieved by using reference compounds for tuning the system in order to achieve a given ratio, and then by plotting the abundance of precursor and product ions as a function of  $E_{\text{lab}}$  (Energy Resolved Mass Spectrometry, ERMS). Only under these conditions, is it possible to build MS/MS libraries without large variations in relative ion abundances.

However, two essential differences characterize collisional activation occurring from ion beams in the Rf-only quadrupole collision cell and from the stored ions in 3D (and 2D) ion trap cells: (i) in the quadrupole collision cell, collisional activation takes place as a *fast heating* [93] from ion beam collision experiments in contrast to storage experiments in which *slow heating* [93] occurs due to the large number of low-energy collisions with collision gas, and (ii) a selective excitation of precursor ions occurs in ion traps, whereas both precursor and product ions can be activated in triple quadrupole instruments. These differences suggest that the number of consecutive dissociations must be reduced in the ion trap compared with that occurring in the quadrupole collision cell [44]. Furthermore, the detection of product ions depends upon the dissociation rate constant of precursor ions which must be consistent with the time-window for ion observation (*i.e.*,  $10^{-5}$  s for triple quadrupoles and  $10^{-2}$  s for IT/MS).

All these findings show why the building of database for CID in MS/MS experiments is difficult without many drastic precautions when CID spectra are recorded with various tandem instruments (*i.e.*, TOF, quadrupole, 3D–2D ion traps or FTICR analyzers). Indeed, all the previous experimental parameters must be controlled to be able to compare CID spectra.

### 2.3.3. Other activation modes

The IRMPD mode was first introduced by McLafferty and coworkers [77] and Hunt et al. [94] in FTICR devices. It is now conventionally available on FTICR [95] and as prototype on 3D–[96,97] and 2D-ion [47] traps. Following precursor ion fragmentation, product ions are trapped together with the selected precursor ions and can be activated during the next IRMPD process. The stored product ions being submitted to IR radiation can dissociate in turn, yielding consecutive cleavages not necessarily observed by CID in Rf-only multipole collision cells or during CID-sustained off-resonance irradiation (CID-SORI) processes in ICR cells. IRMPD was used by Hakansson and coworkers [75] for phosphate-containing metabolite investigation and was compared with CID processes. CID was shown to be more efficient for low-mass compounds (<600 u), but IRMPD appeared to provide complementary information since it was preferred for higher mass compounds (>600 u).

Recently, following the work of Freiser and coworker [98], McLafferty and coworkers [77], and Beauchamp and coworkers [76], Zubarev and coworkers [99] and Hakansson's group [75] used EID (electron-induced dissociation) for metabolite analysis with an FTICR instrument and demonstrated that both CID and EID processes give complementary fragmentation patterns.

### 2.3.4. Mass analyzers and MS/MS experiments

Triple quadrupole (TQ) instruments are very popular in the field of metabolomics [100,101]. They are dedicated to low-energy dissociation processes. The collision cell can be a Rf-only quadrupole or multipole (hexapolar or octapolar fields), both of them being suitable for ion transmission. Compared with the high keV ion beam tandem (TOF–TOF) mass spectrometers, TQ-instruments have a lower scanning speed and a limited resolution, which does not allow accurate mass measurements.

The advantage of the triple stage instruments is the possibility of simple screening by using different scanning MS/MS modes: (i)

detection of the precursor ion of a selected product ion by maintaining the second mass filter on the  $m/z$  value of product species and by scanning the first mass filter which sequentially transmits all ions from the source to the collision cell (*i.e.*, precursor ion scans) and (ii) detection of the precursor ions able to release a common neutral after collision and excitation in the activation cell (*i.e.*, neutral loss scans). To this end, both the mass filters scan together with a mass shift corresponding to the released neutral. These different MS<sup>2</sup> scanning modes available with triple quadrupole instruments are relevant for targeted approaches by allowing the selective detection of the components of the same chemical class [102,103]. The neutral loss scanning combined with natural and stable-isotope labeled compounds allows high-throughput screening of metabolites based upon the MS signatures [104]. On the other hand, the constant precursor, product ion and neutral loss scanning modes were employed advantageously to identify metabolites in urine samples [105].

“In-time” analyzers, such as ion traps, can be used to perform low-mass resolution sequential MS<sup>n</sup> experiments, which are of interest for structural elucidation [26,106]. The relevance of using LC/ITMS for complex product authentication and identification was shown by Kite et al. for the metabolomic analysis of a hundred saponins in crude plant extracts [107]. However, these experiments are limited by the intrinsic low mass cut-off (*i.e.*, the  $m/z$  ratios of product ions must be higher than 1/3 to 1/5 of that of the precursor ions). Linear ion traps are the last generation of ion trap mass devices and have several advantages over 3D-ion traps, such as a larger ion storage capacity, a higher trapping efficiency, a reduced low mass cut-off thanks to *pulsed-Q-dissociation* (PQD), which is a novel fragmentation mechanism developed for the LTQ [108,109], and the potential to use resonant radial ion activation with complementary advantages for product ion scanning [110]. Of note, linear ion traps have already been used for metabolomic purposes [111].

Triple quadrupole and ion trap analyzers provide complementary information. For example, a triple quadrupole operated in the neutral loss and precursor ion scanning modes was used to characterize sulfoconjugates in rat urine. Further MS<sup>n</sup> experiments performed in an ion trap mass spectrometer, identified 4-ethylphenol sulfate and discriminated it from its two isomers (*i.e.*, 2- and 3-ethylphenol sulfates) by matching of chromatographic retention times, MS<sup>2</sup> and MS<sup>3</sup> spectra of the unknown to the synthesized compounds [102].

Such an approach is much more easily performed by using a hybrid Q TRAP<sup>®</sup> instrument [47]. Combining a triple quadrupole analyzer with linear ion trap technology in a single instrument retains the conventional triple quadrupole scan functions such as neutral loss, precursor and product ion scanning modes, and the selected product ion and multiple reaction monitoring modes, and provides access to sensitive MS scans over a large  $m/z$  ratio range as well as to MS<sup>3</sup> experiments [112]. Furthermore, “triple quadrupole-like” fragmentations without any low mass cut-off can be obtained because the precursor ion selection in the first quadrupole and the fragmentation in Rf-only cell are decoupled. Some applications have been published in the field of small molecule profiling [113], metabolomics [103] and metabolism studies [114,115].

A limitation of the previous instruments for identification purposes is their inability to provide mass measurements accurate enough to achieve elemental composition determination. Combining different analyzers such as quadrupole and TOF in QqTOF devices overcomes this limitation while retaining fragmentation capability. However, to maintain TOF properties, the quadrupole filter and the multipole collision cell are almost orthogonal to the TOF axis in order to limit the error due to the spatial ion beam distribution on the resolution of the TOF analyzer. The

use of a push–pull system transforms the continuous low kinetic ion beam coming from the ESI/quadrupole filter into a higher kinetic pulsed ion beam [116]. Such a device achieves high-resolution CID spectra of relevance for metabolite identification purposes [117].

QqTOF devices can be used for MS<sup>2</sup> experiments in the product ion mode and can be coupled to UPLC, as reported in a study of metabolomic maps for understanding the cell response to ionizing radiations [117]. A new approach called UPLC/MS<sup>E</sup> has been developed to improve the collection of information from MS/MS experiments. Acquisitions are performed at low and high collision energies without any precursor ion selection, thus requiring additional software to generate information on both precursor and product ions [118]. Otherwise, 3D ion traps combined with TOF [119] and even better 2D ion traps/TOF [120,121] provide some interesting characteristics, which include MS<sup>n</sup> scanning capability together with accurate mass measurement of product ions generated in the ion trap.

The use of TOF analyzer is also focused on accurate mass measurements for elemental composition determinations. However, greater accuracy is achieved using FT/MS instruments such as the LTQ-Orbitrap<sup>®</sup> which provides accurate mass measurements in multistage MS/MS experiments for metabolite profiling [122] and metabolism studies [59,60,123]. Interestingly, the C-trap, initially designed to store ions before their orthogonal ejection into the Orbitrap cell for high resolution analysis, can now be used as a collision chamber to enable triple quadrupole-like fragmentations, as recently shown for peptide sequencing [124]. This new CID approach is promising for metabolomic applications.

Finally, the best accuracy and mass resolving power are achieved with tandem QhFTICR instruments (h for hexapole) [125]. Furthermore, FTICR devices select precursor ions with high resolution and perform *in situ* (i.e., in the ICR cell) ion activation such as IRMPD. This cannot be achieved with the Orbitrap<sup>®</sup> since until now isolation and fragmentation of precursor ions occur in the ion trap (maximum resolution for precursor ion selection: 0.3 Th). The selection of precursor ions at high resolution can be achieved in the ICR cell by injecting some pulsed collision gas. This decreases the resolution power of the device, with typical values ranging from 10,000 to 20,000 at FWHM (for an ion at *m/z* 600, JC Tabet Personal Communication). The performances of FTICR for proteomic and metabolomic applications have been reported in a study coupling 20 kpsi reverse-phase liquid chromatography to a 11.4 T-FTICR mass spectrometer: more than 5000 metabolites were detected in a bacterial cell extracts. However, complementary LC/MS/MS experiments were performed in the ion trap that is in front of the ICR cell, suggesting that such experiments (MS/MS) remain challenging in the FTICR cell [126].

## 2.4. Complementary approaches

### 2.4.1. H/D exchange

Hydrogen/deuterium exchange in solution is a widely used strategy for elucidation of mass fragmentation mechanisms [127–130], especially in the field of metabolism studies. Determination of the number of exchangeable hydrogen atoms (i.e., attached to O, N, and S atoms) facilitates structural elucidation of metabolites. For example, from the formula C<sub>10</sub>H<sub>18</sub>O<sub>4</sub>, several compounds are possible: a dicarboxylic acid, a monocarboxylic acid with two OH groups and an unsaturation (a ring or a double-bond), a monocarboxylic acid with a keto function and a hydroxyl group, and a compound with four OH groups and two unsaturations. These compounds will be partially discriminated by H/D exchange experiments since they bear 2, 3, 2 and 4 exchangeable hydrogen atoms,

respectively. Furthermore, an increase in *m/z* ratio for a product ion during H/D exchange experiments will supply information on the position of a particular chemical function containing mobile protons in the charged molecule. Karlsson [131], one of the pioneers of H/D exchange application, introduced the use of deuterium oxide as a mobile phase for microcolumn liquid chromatography coupled to ESI/MS, which is now used extensively for structural elucidation of small molecules [132–135]. The application of deuterium oxide as the sheath liquid in CE/MS for structural elucidation purpose has also been reported [136]. In addition, gas phase H/D exchange experiments can be performed by introducing various deuterated agents (e.g., CH<sub>3</sub>OD, D<sub>2</sub>O, ND<sub>3</sub>) at the skimmer and in the collision cell. The extent and rate of H/D exchanges increase with the gas-phase basicity of the reagent, so it is possible to selectively label functional groups of the molecule. In addition, this sometimes permits the distinction of isomers [137]. However, the yield of deuteration may not be optimal as compared with other classic H/D experiments. A relatively poor yield may lead to erroneous conclusions.

### 2.4.2. Derivatization strategies

Derivatization can be used to achieve numerous and different objectives, such as to improve the ionization efficiency of a family of compounds, highlight and localize a chemical function (e.g., hydroxyl group, glucuronide), trap reactive metabolites, assist or induce ion dissociation, and modify polar compound retention. These latter points have been extensively reviewed [138,139] and although pharmaceutical and drug discovery issues were addressed, it is easily transposable to metabolomics.

A chemoselective tagging strategy termed “metabolite enrichment by tagging and proteolytic release” (METPR) has been introduced for enrichment and profiling of small molecules [140]. Briefly, metabolites are captured on a solid support by conjugation to resin-bound reactive groups that target different classes of metabolites. The trapped molecules are then released by a cleavage step promoted by a protease. This approach was used to profile highly polar compounds by derivatizing the metabolites with a hydrophobic *p*-Cl-phenylalanine residue, which improved their retention on reverse-phase columns. The chlorine of the tag also allowed discrimination of the tagged metabolites from background peaks [141].

## 3. Databases

A first objective of database inquiry is the annotation of MS signals. This requires queries by molecular mass or by elemental composition, depending on the instrument. Databases can as well be used for biological purposes. Indeed, it is essential to give biological sense to the acquired data, for example, to determine the biological functions of metabolites.

For these purposes, four kinds of tools can be distinguished: (i) general chemical databases (e.g., PubChem) that encompass synthetic and/or natural compounds, (ii) metabolic databases that deal with annotated metabolic pathways, (iii) metabolomics databases that originate from metabolomic research project in a specific field (e.g., LipidMaps in lipidomics) and (iv) mass spectral databases (e.g., NIST).

Actually, some databases combine several of these functions. It is important to keep in mind that all databases are works in progress, with some going through a rapid growth phase with constant additions and corrections being made. Therefore, additional functions could be found in the near future. Table 2, which is certainly not exhaustive, summarizes some of the most widely used databases that are addressed in the following sections.

**Table 2**  
Databases for the identification of metabolomic signals by mass spectrometry

Database		Thematic	Conception/URL	Reference
BiGG	●	Human	University of California (USA) ( <a href="http://www.biggs.ucsd.edu/">www.biggs.ucsd.edu/</a> )	[197]
BioCyc (HumanCyc, MetaCyc)	●	Biochemical pathways	SRI International (USA) ( <a href="http://www.biocyc.org/">www.biocyc.org/</a> )	[198,148]
ChEBI	●	General	European Bioinformatics Inst. (UK)/European Molecular Biology Lab ( <a href="http://www.ebi.ac.uk/chebi/">www.ebi.ac.uk/chebi/</a> )	[144]
ChemFinder	●	General	Cambridge Soft (USA) ( <a href="http://www.chemfinder.CambridgeSoft.com">www.chemfinder.CambridgeSoft.com</a> )	[143]
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> )	[142]
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> )	[153]
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> )	[149]
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> )	[199]
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> )	[151]
LipidMaps	●	Lipidomics	LIPID MAPS Bioinformatics Core (USA) ( <a href="http://www.lipidmaps.org/data/index.html">www.lipidmaps.org/data/index.html</a> )	[200]
LipidBank	●	Lipidomics	Japanese Conference on the Biochemistry of Lipids (Japan) ( <a href="http://www.lipidbank.jp/">www.lipidbank.jp/</a> )	[201]
Madison Metabolomics Consortium Database	●	General	National Magnetic Resonance Facility, University of Wisconsin-Madison ( <a href="http://mmcd.nmr.fam.wisc.edu/">http://mmcd.nmr.fam.wisc.edu/</a> )	[28]
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> )	[202]
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> )	[150]
MoTo	●	Metabolome database for tomato	Wageningen University ( <a href="http://www.appliedbioinformatics.wur.nl">www.appliedbioinformatics.wur.nl</a> )	[152]
MSlib	●	* Drugs, metabolites	University of Alberta ( <a href="http://www.ualberta.ca/~gjoncs/mslib.htm">http://www.ualberta.ca/~gjoncs/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> )	[203]
SciFinder	○	General	American Chemical Society (USA)	

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

### 3.1. General chemical databases

The American Chemical Society (ACS) provides several databases, including Scifinder (CAS: Chemical Abstract Service), that contain general information about millions of compounds. A subscription is required to access these databases. Searches can be performed by entering elemental compositions and substructures, which can also be used to refine queries.

Cactus (CADD group's Chemo informatics Tools and User Service) provides access to the Chemical Structure Lookup Service (CSLS) and Enhanced NCI Database Browser. CSLS allows searches of millions of indexed structures from 80 free and commercial chemical databases [142].

Chemfinder is a free chemistry database (CambridgeSoft Corporation, Cambridge, MA), but a subscription is required for its professional version ChemIndex [143]. The latter contains data on over 75,000 compounds including structures, names and synonyms.

PubChem is a database of chemical molecules managed by the National Center for Biotechnology Information (NCBI) which is part of the United States National Library of Medicine (NHL). The PubChem substance, the largest PubChem database, contains more than 38 millions of records describing chemical structures, molecular weight and formula. This database includes structures from chemical manufacturers (e.g., Sigma-Aldrich) and from licensed databases (DiscoveryGate Elsevier, Thomson Pharma). The queries

can be formulated as molecular weight (and MW range), molecular formula or using a chemical structure that may be sketched using the PubChem sketcher.

Chemical Entities of Biological Interest (ChEBI) is a freely available database of small chemical compounds. ChEBI combines information (on both synthetic and natural products) from the KEGG and IntEnz (the Integrated relational Enzyme database) databases. All entries have been reviewed in order to standardize biochemical terminology. In addition, ChEBI offers an ontological classification that gives comprehensive relationships between molecular entities [144].

The Merck Index is an encyclopedia of chemicals, drugs and biological compounds with over 10,000 monographs containing information such as CAS registry number, chemical formula, molecular weight on single substances. It is available by subscription.

CHEMnetBASE is an online and CD-ROM chemical database from Chapman & Hall/CRC that includes "The Handbook of Chemistry & Physics" and some other combined chemical dictionaries as "The dictionary of natural products" for example. The latter provides a comprehensive source of chemical data on natural products.

### 3.2. Biochemical and metabolic databases

These databases are mostly used to give some biological sense to metabolomic data. They are also a good starting point for the iden-

tification of signals from cell extracts, plant biology or microbiology experiments.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database of biological systems consisting of genetic building blocks of genes and proteins (KEGG GENES), chemical building blocks of both endogenous and exogenous metabolites (KEGG LIGAND, [145], molecular wiring diagrams of interaction and reaction networks (KEGG PATHWAY), and hierarchies and relationships of various biological objects (KEGG BRITE). It provides a reference knowledge base for linking genomes to biological systems [146].

Biocyc [147] is a collection of 371 pathways corresponding to various genomes. It describes the genome and pathways of a single organism. Two types of materials can be downloaded: data files and an executable program that combines the pathway software with the Biocyc database [148].

The Arabidopsis information resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available include the complete genome sequence, gene product information, metabolism, gene expression, DNA seed stocks, genome maps, genetic and physical markers, publications and information about the *Arabidopsis* research community. The downloadable program Aracyc 4.1 contains 238 pathways, 85% of which are experimentally confirmed. Only queries by name among the 1908 compounds are available.

### 3.3. Metabolomic databases

Those databases are the most recent ones and in the near future should be the most comprehensive ones. They are of value for the first steps of identification because they contain spectral data about biochemically relevant metabolites in specific biological contexts such as human biofluids or plant extracts, for example.

The Human Metabolome Database (HMDB) is a comprehensive, Web-accessible metabolomic database that brings together quantitative chemical, physical clinical and biological data about thousands of endogenous human metabolites [149]. It contains records for more than 2180 endogenous metabolites and provides literature derived data, mass spectra, NMR-spectra and validated metabolite concentrations. The HMDB offers as well a relational data extraction tool and both MS and NMR spectra searching tools.

Metlin is a metabolite database containing over 15,000 entries. It provides a data management system designed to help the researcher to identify metabolites by offering public access to its repository of current and comprehensive mass spectral metabolite data. Metabolite data provide a list of known metabolites, their mass, chemical formula and structure, each of them being linked to outside resources such as KEGG. Metlin also contains FT/MS data consisting in FT/MS spectra of chromatographically separated human serum fractions. It gives access to LCMS profiles from various tissues and biofluids, which are provided with the experimental conditions [150].

Atomic Reconstruction of Metabolism (ARM) is a project of the Arita, Nishioka and Kanaya groups and represents the metabolism of more than 2700 reactions in about 1700 *Escherichia coli* subclasses. The official website (<http://www.metabolome.jp>) provides software tools and databases such as MassBank, flavonoid viewer, LipidBank and KNApSACK. MassBank will be detailed in a further section. LipidBank contains biological activities, physico-chemical properties, literature information and spectral data on 7009 lipids. It is an open publicly free database of natural lipids including fatty acids, glycolipids, sphingolipids, steroids and various vitamins. KNApSACK is a comprehensive metabolites-species relationship database. It includes 20,000 metabolites and their relative species information. It can be searchable by name, molecular weight, molecular formula and mass spectra [151].

The metabolome database for tomato (MoTo DB) is an open access metabolite database for LC/MS dedicated to tomato fruit [152]. It is based on literature information combined with experimental data derived from LC/MS-based metabolomic experiments. The assignment of mass signals relies on the combination of accurate mass (Q-TOF), retention time, UV-vis information and MS/MS fragmentation data. The accurate mass, together with a mass accuracy setting, is the main search entry for the database. Mass accuracy can be set from 1 to 1000 ppm, thus enabling the matching of data from detectors generating mass with either low or high accuracy. Links with PubChem and MedLine databases are also available.

### 3.4. Spectral databases

Mass spectral databases were initially constituted with data from GC/MS experiments due to high reproducibility between instruments. This is for example the case for the Golm metabolome database [153]. The NIST05 database offers a fully evaluated collection of electron ionization mass spectra obtained at standardized electron energy which also contains a MS/MS library (5191 compounds) and retention index data (25728 compounds) (<http://www.nist.gov>). LC/MS/MS using the particle beam interface [154] was expected to give EI spectra and be useful for constructing libraries, but was not sufficiently sensitive [155].

Ionization and fragmentation in LC/MS instruments could not be standardized even with similar instruments [54,156]. This observation highlights the need for the constitution of spectral libraries taking into account the information provided by different kinds of instruments. Despite this limitation, databases containing API CID spectra such as HMDB, Metlin, MassBank from [metabolome.jp](http://www.metabolome.jp) ([www.metabolome.jp/](http://www.metabolome.jp/)), and LipidMaps are starting to be released.

The MS search function of HMDB allows access to 1200 MS/MS spectra acquired with a triple quadrupole at 3 different collision energies (i.e., low, medium and high) for around 400 compounds.

Metlin incorporates 282 MS/MS spectra acquired in ESI+ and 140 MS/MS spectra acquired in ESI-. By specifying a precursor mass range on the MS/MS search form, a research can easily reference the MS/MS profile of known metabolites in Metlin against the MS/MS profile of an unknown compound [150].

The MassBank of [metabolome.jp](http://www.metabolome.jp) provides high-resolution mass spectra of 1749 metabolites. Experimental conditions of separation are also described whenever a separation technique (LC, GC, CE) is used. LC/MS spectra are recorded on QqTOF, TQ and ion trap instruments. Ten thousand spectral data are searchable through a spectral browser and a peak search function [157,158].

The LipidMaps structure database (LMSD) comprises structures and annotations of biologically relevant lipids. Lipid standards include MS/MS values with links to fragmentation spectra, including structures of principal product ions, as well as links to commercial catalogs and literature references. This online tool predicts possible structures based on MS data for precursor ions, mass tolerance, ion mode and head groups. The “structure drawing” function is able to create a fatty acyl structure from information about carbon chain and functional groups.

## 4. Endogenous metabolite identification: some case studies

Some metabolite identifications reported in metabolomic studies are typically not novel as they consist in assigning a signal to a compound that has already been rigorously described in the literature. Thus, these metabolites are often identified based upon the retrieval of properties such as retention time, accurate mass, fragmentation pattern that are common to endogenous metabolites and authentic reference compounds (standards). In high-resolution

mass spectrometry-based techniques, the identification of a compound typically begins with a database query based either directly on its experimental accurate mass measurement or on its deduced elemental composition.

Database queries here return one or more hits. If fragmentation experiments are available, they can supply information about chemical groups that are present in the compound by highlighting neutral losses or product ions, which are characteristic of a functional group and can serve to discriminate between database hits. In addition, spectral information from the authentic compound may be accessible via the database or the literature and lead to a confident structure assignment. As no consensus on what constitutes valid metabolite identification has been reached, Sumner et al. have reported four different levels of identification according to the information provided for the compound characterization within the Metabolomics Standards Initiative [159]:

- (i) *Identified compounds*: A minimum of two independent and orthogonal types of data relative to an authentic compound analyzed under identical experimental conditions. In MS-based techniques this could include: retention time/index and mass spectrum, or accurate mass and tandem MS.
- (ii) *Putatively annotated compounds*: Without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries.
- (iii) *Putatively characterized compound classes*: Based upon characteristic physicochemical properties of a chemical class of

compounds, or by spectral similarity to known compounds of a chemical class.

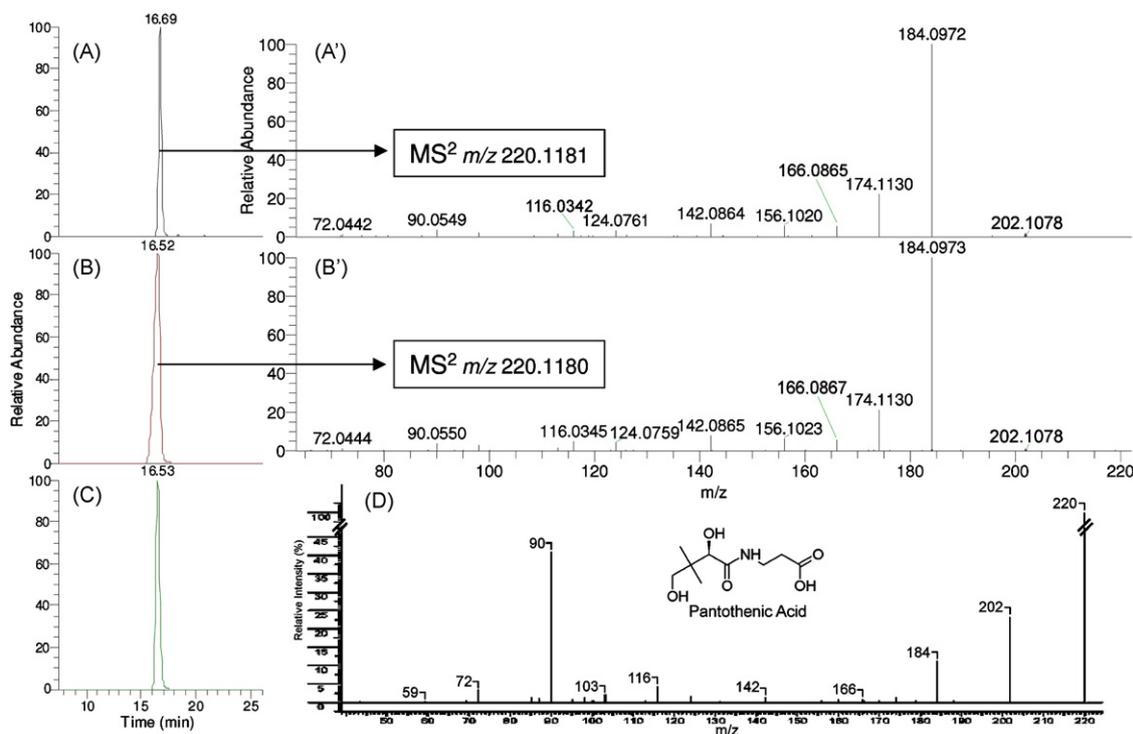
- (iv) *Unknown compounds*: Although unidentified or unclassified these metabolites can still be differentiated based upon spectral data, thus enabling relative quantification.

Two concrete cases are hereafter presented to illustrate the varying complexity of compound identification. These two examples are extracted from a metabolomic study and concern the identification of metabolites detected in rat urine by using an LC-LTQ-Orbitrap® system.

#### 4.1. First situation: the hypothetical metabolite is described in biochemical or metabolomic databases

This first example deals with the identification of a discriminating signal at  $m/z$  220.1180, recorded in positive electrospray ionization (Fig. 3A). Elemental composition determination led to the formula  $C_9H_{17}NO_5$  ( $-1.1$  ppm), which was used to query the human metabolome database. HMDB returned pantothenic acid as single hit.

From the CID spectrum, presented in Fig. 3A', it was possible to observe 2 losses of 18 (corresponding to water losses) and a loss of 28 (CO and not  $C_2H_4$ , as shown by accurate mass measurement), which indicated the probable presence of a hydroxyl and a carboxyl group. HMDB provided a CID spectrum of the authentic compound (pantothenic acid), acquired in positive ESI mode with a



**Fig. 3.** Identification of pantothenic acid in rat urine. (A) Extracted ion chromatogram of the ion to be identified at  $m/z$  220.1180 ( $\pm 2$  ppm). (A') CID spectrum of the ion of interest at  $m/z$  220.1180. (B) Extracted ion chromatogram from the authentic pantothenic acid (theoretical  $[M+H]^+$  at  $m/z$  220.11795 ( $\pm 2$  ppm)) and its CID mass spectrum (B'). (C) Extracted ion chromatogram of  $m/z$  220.1180 from rat urine spiked with pantothenic acid. (D) CID mass spectrum of the authentic pantothenic acid provided by HMDB (acquired on a TQ instrument (Quatro Waters®)). *Conditions:* The chromatographic separation was performed on an XTerra MS C<sub>18</sub> 5- $\mu$ m, 2.1 mm  $\times$  150 mm column (Waters, Saint Quentin en Yvelines, France). The mobile phases were A: 100% water and B: 100% acetonitrile, both containing 0.1% formic acid. After an isocratic step of 5 min at 100% phase A, a linear gradient from 0 to 100% of B was run over the next 40 min with a mobile phase flow of 0.2 ml/min. Returning to 100% A over 1 min, the column was then allowed to equilibrate for 10 min leading to a total run time of 60 min. ESI-MS metabolic profiles were acquired using a LC/LTQ-Orbitrap® system within the range 75–1000 Th, successively in both positive and negative ion modes with a resolution set to 60,000 ( $m/\Delta m$  at FWHM) in centroid mode. CID spectra were acquired using data-dependent scanning function. The scan event cycle comprised two data-dependent ( $MS^2$  and  $MS^3$ ) events acquired with a resolution set to 7500. Microscan count was set to unity and a repeat count for dynamic exclusion was set to 3.  $MS^2$  acquisition parameters were an isolation width of 1 Th, normalized collision energy of 35%, and an activation time of 30 ms.

triple quadrupole instrument (Fig. 3D). Following the observation of strong similarities between the CID spectra of our compound of interest and those provided by HMDB, we injected the authentic pantothenic acid into our system for formal identification by comparing retention times (Fig. 3A–C, corresponding to unknown compound in rat urine, reference compound in buffer, and reference compound in spiked urine, respectively), accurate mass measurements, isotopic abundance and fragmentation patterns (Fig. 3A' and B').

#### 4.2. Second situation: the hypothetical structure of the metabolite cannot be obviously deduced from databases

In this case, CID experiments may allow putative annotation or characterization of the metabolites. Careful interpretation of such CID spectra is required to generate structural hypotheses that need to be confirmed by further chemical syntheses.

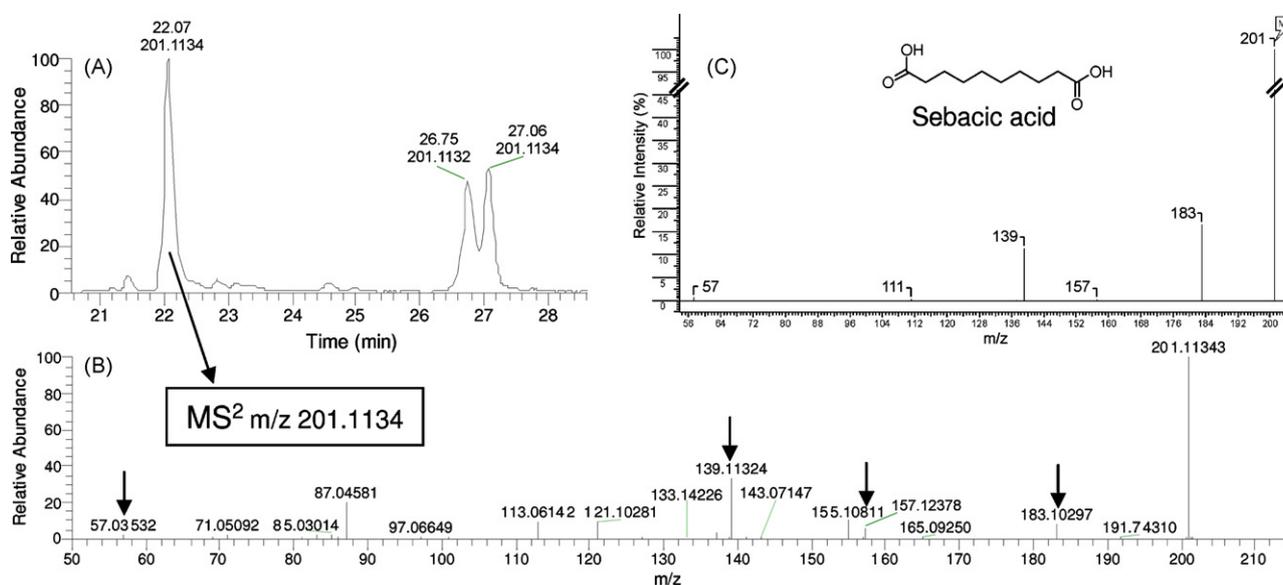
The second example deals with an ion at  $m/z$  201.1133 that was observed in LC/MS metabolic fingerprints from rat urine in negative electrospray ionization. This case is more complex because the extracted ion chromatogram of the ion of interest displayed three main peaks, even with a mass tolerance of 2 ppm (Fig. 4A). However, only the first eluted peak at 22 min was highlighted by MVA and therefore had to be identified. Isotope pattern analysis and elemental composition determination led to the same formula for the three peaks that is  $C_{10}H_{18}O_4$  (RDBE = 2).

A query based on the elemental composition in the human metabolome database returned sebaccic acid as a unique hit. The low-resolution CID spectrum of sebaccic acid provided in the database (Fig. 4C) presented similarities (principal product ion at  $m/z$  139, and several common product ions at  $m/z$  183, 157, and 57) with the one obtained for the compound of interest (Fig. 4B, arrows mark the peaks of the compound of interest similar to those of the reference molecule). The injection of the authentic compound from Sigma resulted in the chromatogram and CID spectrum presented in Fig. 5B. From Fig. 5 and contrary to what was expected, it is clear that the endogenous metabolite of interest is not sebaccic acid since it exhibits a different retention time.

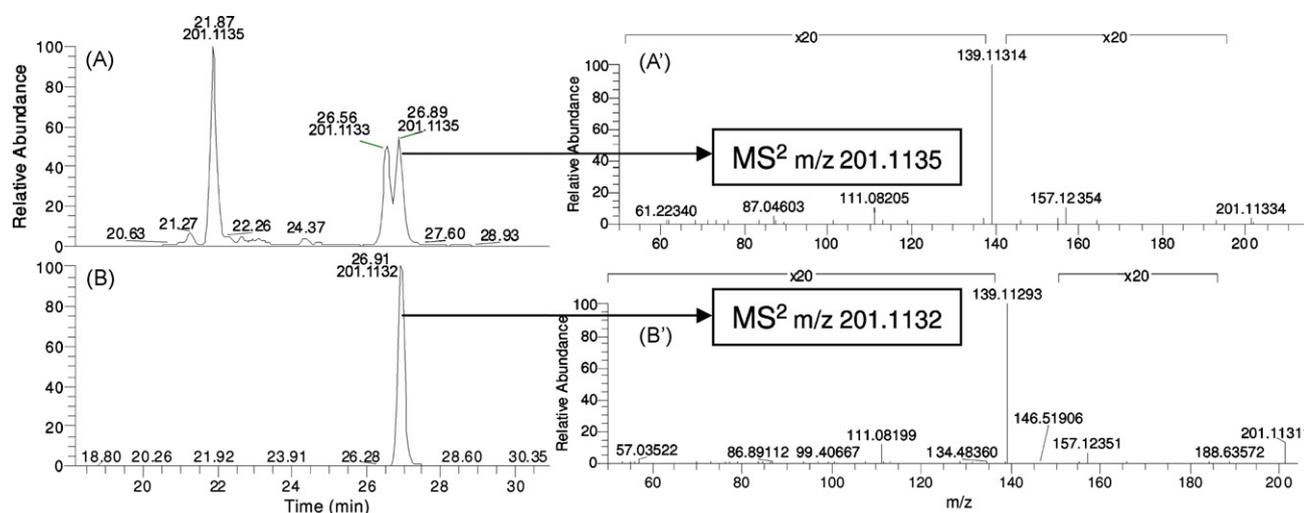
CID and spiking experiments showed that sebaccic acid corresponds to the third peak at 27 min. The three isomers presented in Fig. 4A can only be discriminated from their  $MS^3$  fragmentation patterns, revealing the limitation of the sole use of high resolution for complex biological sample analysis (data not shown). Furthermore, although the product ion at  $m/z$  183 had a high relative abundance in the CID spectrum of deprotonated sebaccic acid from HMDB (recorded by using a triple quadrupole instrument) and in that of the unknown compound eluting at 22 min, it was produced neither for the authentic compound nor for the endogenous sebaccic acid in our system (Fig. 5A and B). This highlights the difficulty of building multi-instrument databases with API-MS techniques.

As the sole hits from biologically oriented databases (HMDB, Metlin, KEGG) were sebaccic acid and diethyl adipate, which were inconsistent with CID spectra, we had to move on to wider databases. However, querying CAS with  $C_{10}H_{18}O_4$  returned 2334 hits. It was therefore necessary to conduct complementary experiments to restrict the search by using imposed chemical substructures. CID experiments (Fig. 4B) showed a neutral loss of 44 u ( $CO_2$ ), yielding a peak at  $m/z$  157 and so highlighted the presence of a carboxylic acid function that should be hydroxylated in the  $\alpha$  position to explain a competitive loss of 46 u ( $H_2CO_2$ ) in the negative ion mode [160]. Moreover H/D exchange experiments supported the previous result by revealing the presence of three exchangeable protons in the compound. In this way, it was possible to assume a dihydroxylated carboxylic acid with an additional unsaturation rather than a dicarboxylic acid.

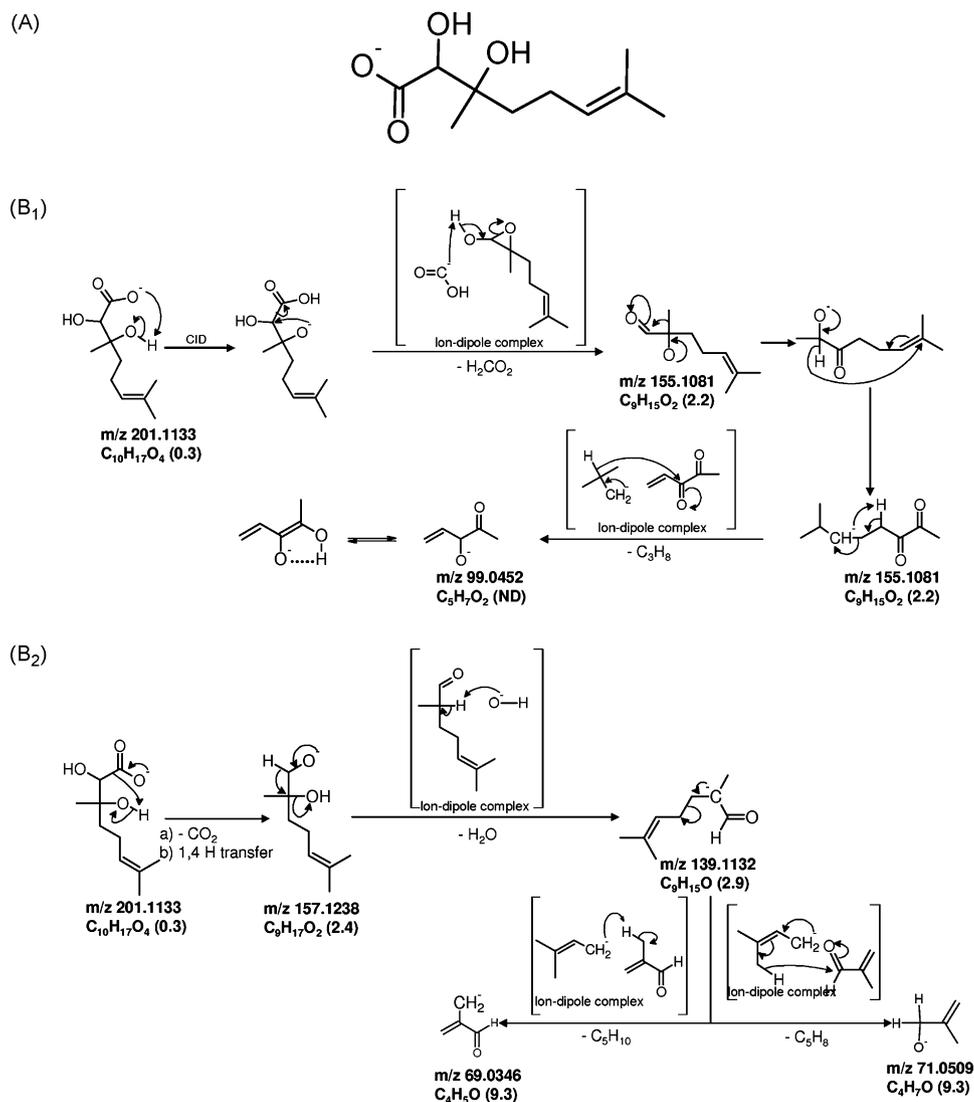
Using this information, the search in CAS returned 19 hits that were discriminated based upon H/D exchange experiments. This led to a single candidate since only one out of the 19 compounds comprised a carboxylic acid, two hydroxyl groups and a single unsaturation (Fig. 6A). Based on this formula, we tried to explain the product ions observed in the multi-stage MS/MS experiments by proposing a mechanism of fragmentation. Unfortunately, in our opinion, the compound found in CAS is unlikely to generate the observed fragmentation pattern, especially the ion at  $m/z$  113. Indeed, we would rather have expected product ions at  $m/z$  99 and 69 according to the putative proposed mechanisms shown



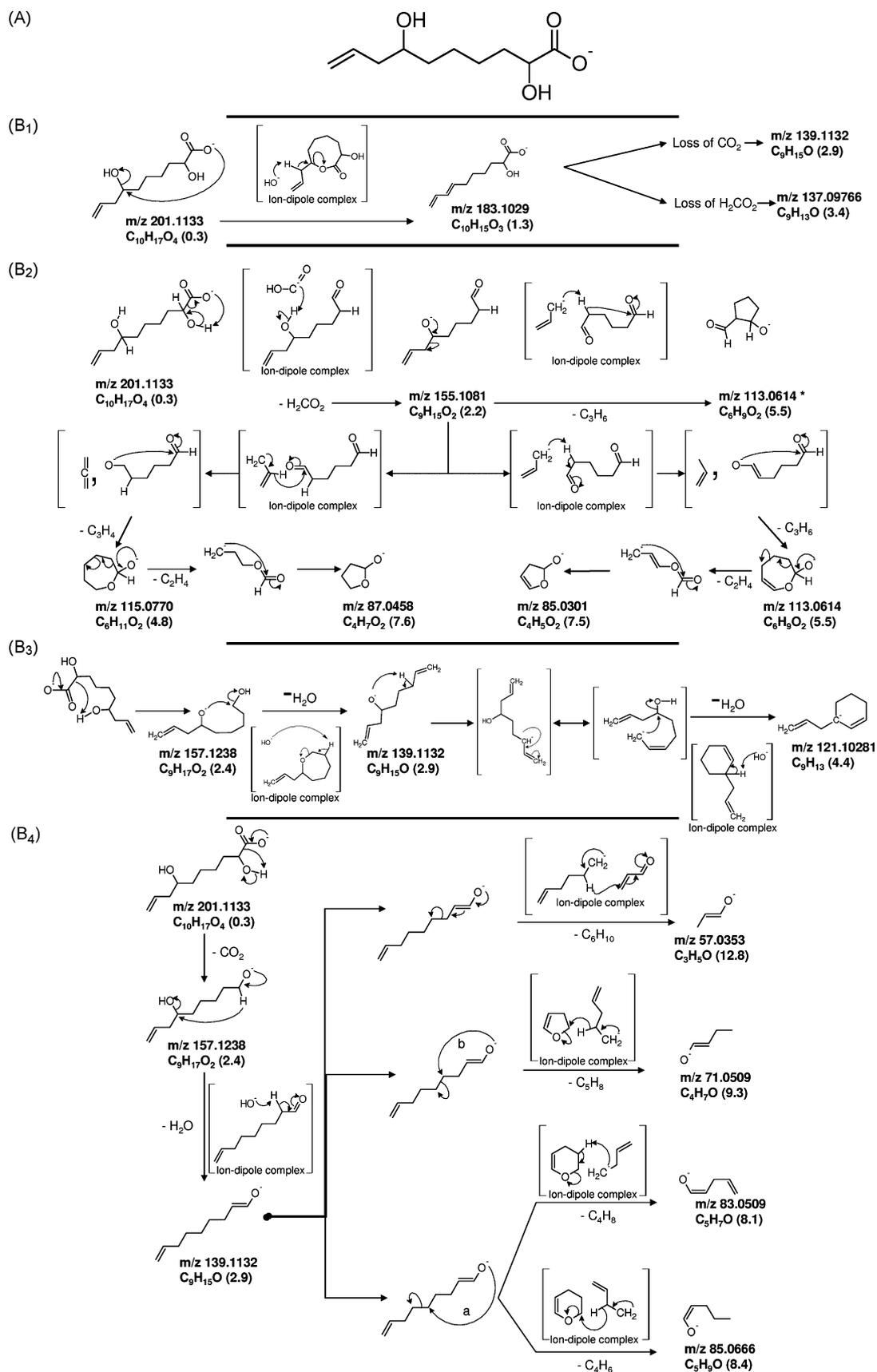
**Fig. 4.** Identification of an unknown compound in rat urine. (A) Extracted ion chromatogram at  $m/z$  201.1133 ( $\pm 2$  ppm) from a LC/MS profile of rat urine acquired on an LC/LTQ-Orbitrap<sup>®</sup> system in negative ion mode. (B)  $MS^2$  spectra of  $m/z$  201.1134 at 22 min (with wide band activation function enabled). Arrows mark peaks similar to those observed at low-resolution on the TQ CID spectra provided by HMDB for sebaccic acid. (C)  $MS/MS$  spectra of sebaccic acid provided by HMDB (acquired with a TQ instrument Quattro Waters<sup>®</sup>). Same experimental conditions as those reported for Fig. 3.



**Fig. 5.** Identification of sebacic acid in rat urine. (A) Extracted ion chromatogram at  $m/z$  201.1133 from an LC/MS profile of rat urine acquired on a LC/LTQ-Orbitrap® system in negative ion mode. (B) Extracted ion chromatogram at  $m/z$  201.1133 ( $\pm 2$  ppm) from an LC/MS profile of rat urine spiked with sebacic acid and acquired on an LC/LTQ-Orbitrap® system in negative-ion mode. Same experimental conditions as those reported for Fig. 3.



**Fig. 6.** (A) Structure proposed by CAS for the unknown metabolite at  $m/z$  201.1133 ( $C_{10}H_{18}O_4$ ) and putative predictive fragmentation mechanisms B1: starting with a neutral loss of  $H_2CO_2$  and B2: starting with a neutral loss of  $CO_2$ . Exact masses and masses error on formula determination (ppm) are indicated under each product ion (ND = not detected).



**Fig. 7.** (A) Proposed structure for the unknown metabolite at  $m/z$  201.1133 ( $C_{10}H_{18}O_4$ ) based on putative fragmentation mechanisms B1: starting with a neutral loss of  $H_2O$  yielding the ion at  $m/z$  183.1029, B2: starting with a loss of  $H_2CO_2$  yielding the ion at  $m/z$  155.1081 and then to ions at  $m/z$  113.0614, B3: starting with a loss of  $CO_2$  and leading to the fragment ion at  $m/z$  121.1028 and B4: starting with a loss of  $CO_2$  but leading to fragment ions at  $m/z$  85.0666, 83.0509, 71.0509, 57.0353. Exact masses and masses error on formula determination (ppm) are indicated under each product ion.

in Fig. 6B. However, such product ions were not retrieved in the experimental CID spectrum (Fig. 4A). As a consequence, we then undertook an interpretation of the CID spectrum of the endogenous compound of interest in order to propose a relevant structure that would fit with the observed product ions. A tentative structural elucidation of the product ions generated by the fragmentation of the metabolite of interest is shown in Fig. 7. The peak at  $m/z$  113 can be considered as a diagnostic ion of the terminal position of the unsaturation. Indeed, it could be generated by an allylic cleavage that would be enhanced by the presence of a hydroxyl group at the homoallylic site. Of course, this putative structure must be confirmed by NMR experiments and further chemical synthesis for formal identification.

## 5. How to address MS signal redundancy?

With ESI-MS, a single compound may produce several signals corresponding to the formation of (i) adduct ions with cations in the positive ion mode (e.g.,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{NH}_4^+$ ) or anions in the negative ion mode (e.g.,  $\text{Cl}^-$ ), (ii) product ions formed by the spontaneous “in-source” CID of the precursor ion, and (iii) homo and heterodimeric ions (e.g.,  $[\text{2M}+\text{H}]^+$ ,  $[\text{M}+\text{CH}_3\text{CN}+\text{H}]^+$ ,  $[\text{2M}-\text{H}]^-$ ). Such signal redundancy may be problematic for metabolomics because it increases time spent on identification and complicates database queries.

Using ESI ionization, it is not *a priori* obvious to determine whether the ion represents a quasi-molecular ion (i.e., a protonated or deprotonated molecule) or an adduct ion. In positive ESI experiments,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{NH}_4^+$  adduct ions can be observed in competition with  $\text{MH}^+$ . With ultra-high resolution instruments, the search for specific mass differences can help make this assignment. For example, looking for mass differences of 17.02655 ( $M_{\text{isoNH}_4} - M_{\text{isoH}}$ ), 21.98195 ( $M_{\text{isoNa}} - M_{\text{isoH}}$ ) and 37.95589 ( $M_{\text{isoK}} - M_{\text{isoH}}$ ) traces the presence of  $\text{NH}_4^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , adduct respectively. Product ions could be identified by the search for typical characteristic product ions and neutral losses.

A similar approach consists in scanning mass spectra to detect MS peaks differing by exact masses that correspond to known typical compositional changes occurring during the generation of adduct or product ions. At least two programs present this functionality: the ESI package and IntelliXtract<sup>®</sup>. The ESI package is a freely available package implemented in R [161]. It is able to annotate peaks such as isotope, adduct and product ions from a user-defined difference mass list that can be amended. IntelliXtract<sup>®</sup> (ACD Lab) is a commercial program compatible with most MS instrument raw data formats. The program performs component extraction and automatic ion assignment of molecular species (i.e., protonated/deprotonated ions, cationized or anionized molecules). Peaks are flagged on mass spectra in which different colors discriminate related peaks in the case of co-eluting compounds.

Another approach is based on the fact that the ratios between the intensities of quasi-molecular and adduct or fragment ions are theoretically fixed, at least within a limited range of concentration of the parent molecule. It is therefore possible to assume that correlation coefficients between the intensities of pair of related ions should theoretically be equal to 1 either across several spectra within a sample or across all samples where the signals are observed. In the first case, the intensity correlation calculation is performed in the chromatographic domain and highlights the theoretical linear dependence between the extracted ion chromatograms of the molecule and its related compounds (i.e., related ions should exhibit the same detection profile). In the second case, by assuming that the different ions from a single metabolite should exhibit the same retention time and the same change in response to a perturbation, and that instrumental conditions were kept iden-

tical from one sample to another, the relative intensities of such ions across all samples where they are observed should theoretically be fully correlated (correlation coefficient  $r = 1$ ). Due to matrix effects and ionization suppression the theoretical value of  $r$  is rarely achieved and relation between signals is assumed for  $r > 0.8$  [162]. Such an approach has been manually implemented by Chen et al. as the first step of the metabolite identification process [29]. Computational scripts relying on correlation calculations are also available in software such as the ESI package [161], openMS [163] or IntelliXtract<sup>®</sup>. Finally, autocorrelation matrices that were initially developed for NMR data processing [165] can also be applied to LC/MS data sets in order to highlight redundant information and also correlation or anti-correlation between ions eluted at different retention times with the aim of piecing together metabolic networks [166].

## 6. Mathematical tools to improve analysis of MS data

Multivariate statistical analyses constitute the most popular kind of visualization tool by emphasizing differences between samples. However, in some situations such as the characterization of unknown signals for database building or targeted approaches, it may be of value to sort the variables according to their chemical structures. In this context, ultra-high resolution mass spectrometry provides useful tools such as the Kendrick and van Krevelen representations for data set organization [63].

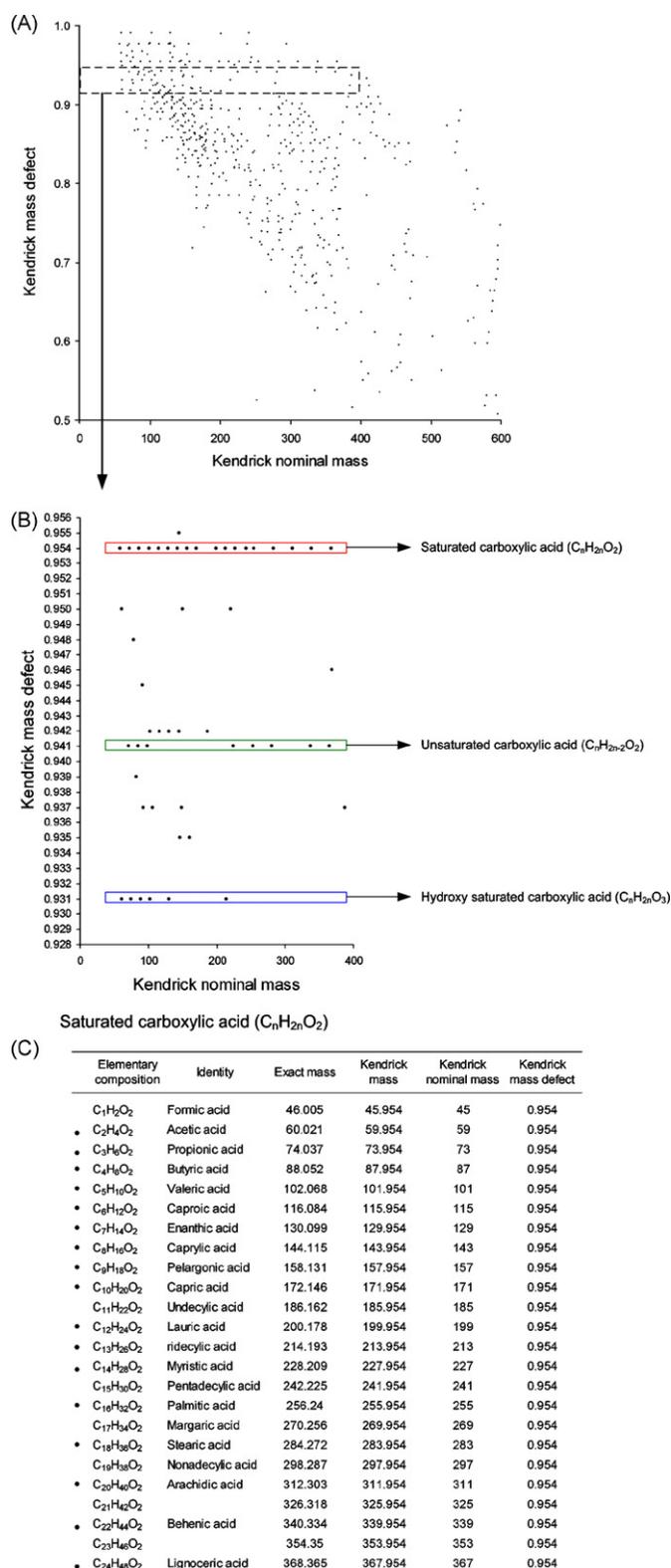
### 6.1. Kendrick plot

Kendrick proposed a mass scale based on the mass of  $\text{CH}_2$  as an alternative to the IUPAC mass scale which is based on the mass of  $^{12}\text{C}$  [167]. Such a scale converts the mass of  $\text{CH}_2$  from 14.01565 to 14 u. Thus, homologous series (namely, compounds with the same constitution of heteroatoms and number of rings plus double bonds, but different numbers of  $\text{CH}_2$  groups) will have an identical Kendrick mass defect [74,168]. The Kendrick representation plots the Kendrick mass defect as a function of the Kendrick nominal mass, thereby producing rows of data separated horizontally by numbers of alkyl groups and vertically according to degree of saturation and class.

The application of Kendrick plot improves the assignment of elemental formulas of high-molecular-weight compounds. As the elementary compositions of few low-molecular-weight compounds related to a given class have been identified, extension to higher masses with the same Kendrick mass defect (i.e., belonging to the same “chemical class”) allows for confident elemental composition assignment of ions [74].

The use of Kendrick mass defect analysis has recently been reported for lipidomics, in the course of a study aimed at evaluating changes in polar lipid levels in tumor cells in response to adenovirus therapy [169]. In this context, Kendrick plots provided a simple visual classification of different chemical classes of polar lipids such as phosphatidylinositols, gangliosides and sulfatides.

To illustrate the relevance of Kendrick indices in organizing biologically relevant compounds and their value for metabolomics, we provide here a Kendrick plot based on the exact masses of 7000 compounds extracted from the KEGG database. We limited Kendrick plot to compounds that exhibit even molecular weights within the range 50–1000 u and contain only C, H, N, O, P, and S atoms. As can be seen in Fig. 8A, it is possible to organize signals by sorting them into homologous series. Compounds of the same chemical class can be identified by calculating the Kendrick mass defect of one of them and then locating all the signals corresponding to this mass defect. This is illustrated by the isolation of carboxylic fatty acids (Fig. 8B and C). Furthermore, the Kendrick representa-



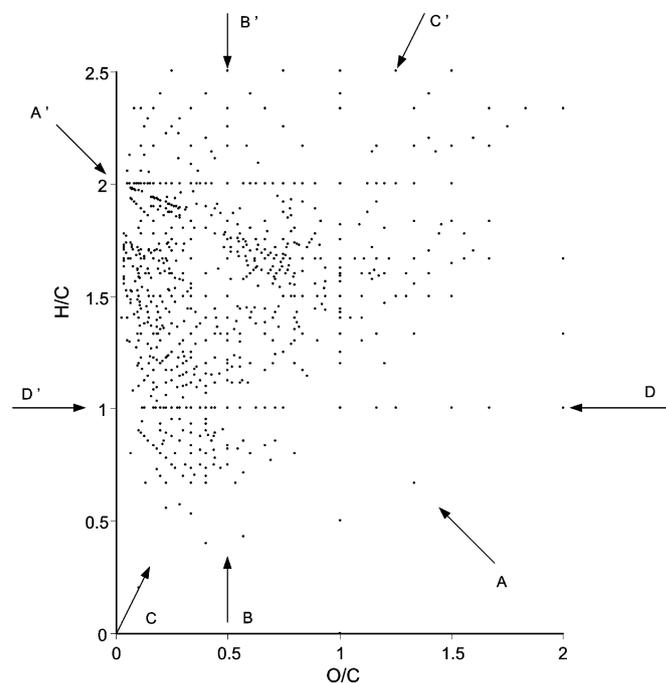
**Fig. 8.** (A) Kendrick plot based on the exact masses of 7000 compounds extracted from the KEGG database. The database query was constrained with the following criteria: (i) compound exhibiting even molecular weights within the range 50–1000 u and (ii) formulas containing C, H, N, O, P, and S atoms. (B) Zoom on the [0.928–0.956] Kendrick mass defect area showing the organization of biological compounds into homologous series that appear as ranks in the plot. (C) Table summarizing the carboxylic acids present in the database (marked by dots) and their associated IUPAC and Kendrick masses.

tion highlights some generic chemical modifications of compounds from homologous series. For example, hydroxy-carboxylic acid as well as mono-unsaturated carboxylic acids produce homologous series that can be distinguished from those related to saturated carboxylic acids, thus enabling rapid monitoring of biotransformation of a particular class of compounds (Fig. 8B). However, Kendrick plots do not succeed in discriminating isomers. As an example, the three compounds presented in Fig. 5 have the same Kendrick mass defect although they exhibit different chemical structures.

## 6.2. van Krevelen plot

Another way to organize data sets from accurate mass measurements is provided by the van Krevelen diagrams [170,171]. These diagrams are element-ratio plots, in which each dot represents the molar ratio of hydrogen to carbon (H/C) on the y-axis and molar ratio of oxygen to carbon (O/C) on the x-axis for one specific molecule. A prerequisite of such representations is to convert the accurate mass measurements of the data sets into elemental formulas. Then, the chemical classes exhibiting any characteristic H/C and/or O/C ratios cluster within specific regions of the diagram, as previously shown for biologically derived compounds, such as lipids, cellulose, lignins, proteins and condensed polyaromatic hydrocarbons [170–174].

In the van Krevelen plot, trends along the lines can be indicative of structural relationships among families of compounds linked by reactions that involve loss or gain of elements in a specific molar ratio. Lines from each reaction path have characteristic slopes or intercepts that can be easily demonstrated from mathematical calculations [170]. From these lines, a series of peaks, possibly products from various chemical reactions, can be visually identified. For example, a trend line representing methylation/demethylation reactions always intersects the ordinate at an H/C value of 2 and



**Fig. 9.** Van Krevelen diagram based on the molecular formulas of 10,000 compounds extracted from the KEGG database. The database query was constrained with the following criteria: (i) compound exhibiting molecular weights within the range 50–1000 u and (ii) formulas containing C, H, N, O, P, and S atoms. The flagged lines in the plot denote the following chemical reactions: (AA') methylation/demethylation, or alkyl chain elongation; (BB') hydrogenation/dehydrogenation; (CC') hydration/condensation; and (DD') oxidation/reduction.

hydration/condensation reactions induce changes along a trend line with a slope of 2 [170,171].

Fig. 9 displays a van Krevelen diagram based on the exact masses of 10,000 compounds extracted from the KEGG database according to the following criteria: (i) molecular masses ranging from 50 to 1000; (ii) formulas containing C, H, N, O, P, and S atoms. The flagged lines in the plot denote the following chemical reactions: (AA') methylation/demethylation, or alkyl chain elongation; (BB') hydrogenation/dehydrogenation; (CC') hydration/condensation; and (DD') oxidation/reduction. Although the use of van Krevelen representations is still limited by the lack of software able to automatically generate and test elemental compositions, this example emphasizes the relevance of such a tool for metabolomic purposes. Alone or combined with Kendrick plots [168], they could be used to separate the signals previously highlighted by multivariate statistical analyses into different chemical families, thus providing another way to visualize the analytical information of biological relevance.

## 7. Conclusion and perspectives

The metabolome is characterized by a huge diversity of chemical structures requiring complementary analytical platform to reach its extensive coverage. Among these, atmospheric pressure ionization based technologies, and especially electrospray ionization are now very popular because they are adapted to the analysis of biological molecules and also because they can be combined with liquid chromatography. Furthermore, they give access to the molecular mass of uncharacterized metabolites, which facilitates the identification work. The aim of this review was to determine what can be expected from atmospheric pressure ionization mass spectrometry for the identification of signals highlighted by metabolomics. Mass spectrometry experiments alone are sufficient when the investigated metabolites are well described in databases and commercially available. In this case, identification is achieved by matching the retention time and CID spectra of the compound of interest to those of the putatively related synthetic reference molecule. In more complex situations, *i.e.*, when the metabolite is not reported in biochemical or metabolomic databases, additional analytical tools such as NMR will be used. However, in this context, atmospheric pressure ionization mass spectrometry can be used as a first approach for compound characterization. We have shown that careful and precise interpretation of CID spectra, combined with additional experiments such as H/D exchange, is useful for providing new structural hypotheses and in refining queries of general chemical databases. Collaboration with organic chemists is then necessary to synthesize the putative chemical structures, if possible, for further LC and CID spectral matching with the signal of interest. NMR experiments may also be required to confirm the structure and to address the issue of isomers.

In terms of perspectives, the identification process should be improved by developments in various fields reflecting the multidisciplinary nature of the metabolomic-based approach: the development of tools for the standardization of metabolomic data, analytical platforms, relational databases, absolute quantification and statistical and mathematical tools. These different points are briefly addressed below.

### 7.1. Tools for the standardization of metabolomic data

Methodologies employed in metabolomics are diverse and evolving. Standardization of metabolomic data is required for the data to be re-used, shared and compared between laboratories.

Several aspects have already been addressed or pointed out by the standard Metabolic Reporting Structure Working group [175] and also by working groups of the metabolomic society, including analytical chemistry [159], statistical analysis and experimental metadata [176,177].

One major issue remains the difficulty of comparing and evaluating the information generated by existing API-mass spectrometry platforms. To this end, reference mixtures of selected metabolites at appropriate concentrations or even perhaps synthetic biofluids have to be defined.

Absolute quantification could facilitate the comparison of results between laboratories. However, accurate and absolute quantification in LC/ESI-MS is not straightforward. Interestingly, some studies now report quantification of amine-containing molecules in biological matrices by using derivatization with a stable isotopic tag, thus suggesting that large-scale quantification of metabolite is possible [178,179].

Furthermore, we lack a mass spectral repository, *i.e.*, a central open source mass spectral database of CID spectra acquired with various instruments in different laboratories. Information such as the biological matrix, instrument and analytical conditions should be precisely documented, as previously suggested [159,175].

### 7.2. Analytical platforms

LC-NMR has already been applied to structural elucidation of drug and plant metabolites [180,181]. LC/MS and NMR can be coupled either on-line or off-line in order to improve structure determination [182,183]. However and despite the introduction of microflow NMR allowing investigations on the microgram scale [184], NMR still lacks sensitivity and requires appropriate chromatographic purification when applied to complex biofluids.

Ion mobility mass spectrometers can be coupled to conventional mass spectrometers. They allow separation of isomers or isobars by providing a third separation dimension if liquid chromatography is also used [185]. Ion mobility mass spectrometry has been evaluated for metabolomic applications: over 1000 metabolite peaks were observed from two-dimensional mobility mass spectra of cell extracts from *E. coli* cultures, with 42 isobaric pairs identified [186].

While API-MS techniques have proven efficient in the determination of ion monoisotopic masses, they also lack reproducibility in terms of fragmentation pattern, which has until now limited the constitution of LC/MS/MS spectral databases. Thus, an interesting improvement would be the interfacing of liquid chromatography with electron ionization mass spectrometry. Such a system has recently been reported: it exhibited minimal matrix effects, was fully compatible with non volatile buffers and salts present in mobile phases and allowed comparisons with EI spectra databases [187]. Although it lacks sensitivity, LC/EI-MS could be used to complement CID spectra obtained with API-mass spectrometers. Doing so, metabolite characterization could be improved through access to large EI spectral libraries and complementary fragmentation information.

Another promising approach consists in soft thermal vaporization followed by supersonic expansion and direct sample compound ionization, while in a supersonic molecular beam (SMB). This technique produces EI mass spectra with enhanced molecular ions that are used to eliminate false search candidates [188,189].

### 7.3. Toward relational databases

There is a crucial need to develop relational database covering the entire metabolomic experiment and allowing comparison

of data sets and interpretation of experimental results by various teams. To this end, associated metadata, *i.e.* biological data (treatment or status) or experimental protocols (sample origin, tissue and experimental conditions) should be stipulated, as has already been proposed for proteomics [190] and transcriptomics (microarray) [191]. Similar steps have been taken in the metabolomics field with MIAMET that provides a checklist of information that should be described in metabolomic publications [192].

A further improvement has been achieved with ArMet (*i.e.*, Architecture for Metabolomics) [193]. Its objective is to provide a formal description of plant metabolomic data sets, including associated metadata, for comparison, mining, data storage and exchange. Finally, standardization of relational databases remains a challenge that includes the choice of spectral data formats (instrument manufacturer format or rather open access file formats such as netCDF, tabular data format and storage format (*e.g.*, XML)). Among recent initiatives, MeMo is a hybrid SQL/XML based on ArMet structure that has been proposed for metabolomic data management (*i.e.*, data storage, organization and annotation) [194].

#### 7.4. Statistical and mathematical tools

The performance of metabolome analysis could be improved by the development of multivariate statistical and chemometric methods that could link the information provided by complementary analytical tools on the same data set. Such approaches, based for example on correlation matrixes [195,196], could help to characterize unknown metabolites in analytical systems and could also be useful in gaining new insights into metabolic pathways.

#### Acknowledgments

EW receives financial support through a research contract funded by the CEA (Commissariat à l'Energie Atomique) and Technologie Servier. JFH is supported by a grant provided by the FSR (Fonds Spéciaux de Recherche, Université catholique de Louvain). The authors wish to thank the editor and anonymous referees for their useful comments and suggestions.

#### References

- [1] O. Fiehn, S. Kloska, T. Altmann, *Curr. Opin. Biotechnol.* 12 (2001) 82.
- [2] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, *Nat. Rev. Drug Discov.* 1 (2002) 153.
- [3] O. Fiehn, *Plant Mol. Biol.* 48 (2002) 155.
- [4] L.W. Sumner, P. Mendes, R.A. Dixon, *Phytochemistry* 62 (2003) 817.
- [5] M. Katajamaa, M. Oresic, *J. Chromatogr. A* 1158 (2007) 318.
- [6] S. Bijlsma, I. Bobeldijk, E.R. Verheij, R. Ramaker, S. Kochhar, I.A. Macdonald, O.B. van, A.K. Smilde, *Anal. Chem.* 78 (2006) 567.
- [7] E. Holmes, H. Antti, *Analyst* 127 (2002) 1549.
- [8] J. Trygg, E. Holmes, T. Lundstedt, *J. Proteome Res.* 6 (2007) 469.
- [9] E. Holmes, A.W. Nicholls, J.C. Lindon, S. Ramos, M. Spraul, P. Neidig, S.C. Connor, J. Connelly, S.J. Damment, J. Haselden, J.K. Nicholson, *NMR Biomed.* 11 (1998) 235.
- [10] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, *Anal. Chem.* 72 (2000) 3573.
- [11] A. Aharoni, C.H. Ric de Vos, H.A. Verhoeven, C.A. Maliepaard, G. Kruppa, R. Bino, D.B. Goodenow, *OMICS* 6 (2002) 217.
- [12] S.A. Trauger, E.P. Go, Z. Shen, J.V. Apon, B.J. Compton, E.S. Bouvier, M.G. Finn, C. Siuzdak, *Anal. Chem.* 76 (2004) 4484.
- [13] S. Vaidyanathan, D. Jones, T. Jenkins, D.B. Kell, R. Goodacre, *Abstracts of Papers of the 229th ACS National meeting*, 2005, p. BIOT 414.
- [14] I.D. Wilson, J.K. Nicholson, J. Castro-Perez, J.H. Granger, K.A. Johnson, B.W. Smith, R.S. Plumb, *J. Proteome Res.* 4 (2005) 591.
- [15] S. Vaidyanathan, S. Gaskell, R. Goodacre, *Rapid Commun. Mass Spectrom.* 20 (2006) 1192.
- [16] K. Dettmer, P.A. Aronov, B.D. Hammock, *Mass Spectrom. Rev.* 26 (2007) 51.
- [17] G. Huang, H. Chen, X. Zhang, R.G. Cooks, Z. Ouyang, *Anal. Chem.* 79 (2007) 8327.
- [18] L.W. Jia, C. Wang, S.M. Zhao, X. Lu, G.W. Xu, *J. Chromatogr. B* 860 (2007) 134.
- [19] S. Vaidyanathan, D. Jones, J. Ellis, T. Jenkins, C. Chong, M. Anderson, R. Goodacre, *Rapid Commun. Mass Spectrom.* 21 (2007) 2157.
- [20] A. Craig, O. Cloarec, E. Holmes, J.K. Nicholson, J.C. Lindon, *Anal. Chem.* 78 (2006) 2262.
- [21] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, *BMC Genomics* 7 (2006) 142.
- [22] C. Ducruix, D. Vailhen, E. Werner, J.B. Fievet, J. Bourignon, J.C. Tabet, E. Ezan, C. Junot, *Chemom. Intell. Lab. Syst.* 91 (2008) 67.
- [23] J.L. Wolfender, K. Ndjoko, K. Hostettmann, *J. Chromatogr. A* 1000 (2003) 437.
- [24] R.J. Molyneux, P. Schieberle, *J. Agric. Food Chem.* 55 (2007) 4625.
- [25] V.V. Tolstikov, O. Fiehn, *Anal. Biochem.* 301 (2002) 298.
- [26] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, J.C. Tabet, E. Ezan, *Rapid Commun. Mass Spectrom.* 17 (2003) 2541.
- [27] M.J. Bogusz, R.D. Maier, K.D. Kruger, K.S. Webb, J. Romeril, M.L. Miller, *J. Chromatogr. A* 844 (1999) 409.
- [28] Q. Cui, I.A. Lewis, A.D. Hegeman, M.E. Anderson, J. Li, C.F. Schulte, W.M. Westler, H.R. Eghbalnia, M.R. Sussman, J.L. Markley, *Nat. Biotechnol.* 26 (2008) 162.
- [29] J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, W. Wang, X. Lu, H.U. Haring, E.D. Schleicher, R. Lehmann, G. Xu, *Anal. Chem.* 80 (2008) 1280.
- [30] K. Vekey, *J. Mass Spectrom.* 31 (1996) 445.
- [31] V. Grill, J. Shen, C. Evans, R.G. Cooks, *Rev. Sci. Instrum.* 72 (2001) 3149.
- [32] E. Rosenberg, *J. Chromatogr. A* 1000 (2003) 841.
- [33] A. Raffaelli, A. Saba, *Mass Spectrom. Rev.* 22 (2003) 318.
- [34] R.B. Cody, J.A. Laramée, H.D. Durst, *Anal. Chem.* 77 (2005) 2297.
- [35] G. Morlock, Y. Ueda, *J. Chromatogr. A* 1143 (2007) 243.
- [36] C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, *Anal. Chem.* 57 (1985) 675.
- [37] Z. Takats, J.M. Wiseman, B. Gologan, R.G. Cooks, *Science* 306 (2004) 471.
- [38] H.W. Chen, Z.Z. Pan, N. Talaty, D. Raftery, R.G. Cooks, *Rapid Commun. Mass Spectrom.* 20 (2006) 1577.
- [39] Z.Z. Pan, H.W. Gu, N. Talaty, H.W. Chen, N. Shanaiah, B.E. Hainline, R.G. Cooks, D. Raftery, *Anal. Bioanal. Chem.* 387 (2007) 539.
- [40] B.B. Schneider, D.D.Y. Chen, *Anal. Chem.* 72 (2000) 791.
- [41] B.B. Schneider, D.J. Douglas, D.D.Y. Chen, *J. Am. Soc. Mass Spectrom.* 12 (2001) 772.
- [42] J.C. Schwartz, A.P. Wade, C.G. Enke, R.G. Cooks, *Anal. Chem.* 62 (1990) 1809.
- [43] E. De Hoffmann, V. Stroobant, *Mass Spectrometry: Principles and Applications*, 3rd ed., John Wiley and Sons Ltd., West Sussex, UK, 2007.
- [44] R.E. March, J.F.J. Todd, *Practical Aspects of Ion Trap Mass Spectrometry: Fundamentals of Ion Trap Mass Spectrometry*, vol. I, CRC-Press, Boca Raton, FL, 1995.
- [45] R.E. March, J.F.J. Todd, *Practical Aspects of Ion Trap Mass Spectrometry: Ion Trap Instrumentation*, vol. II, CRC-Press, Boca Raton, FL, 1995.
- [46] R.E. March, J.F.J. Todd, *Practical Aspects of Ion Trap Mass Spectrometry: Chemical, Environmental and Biomedical Applications*, vol. III, CRC-Press, Boca Raton, FL, 1995.
- [47] D.J. Douglas, A.J. Frank, D.M. Mao, *Mass Spectrom. Rev.* 24 (2005) 1.
- [48] R.M.A. Heeren, A.J. Kleinnijenhuis, L.A. McDonnell, T.H. Mize, *Anal. Bioanal. Chem.* 378 (2004) 1048.
- [49] M. Hardman, A.A. Makarov, *Anal. Chem.* 75 (2003) 1699.
- [50] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R.G. Cooks, *J. Mass Spectrom.* 40 (2005) 430.
- [51] A. Makarov, E. Denisov, O. Lange, S. Horning, *J. Am. Soc. Mass Spectrom.* 17 (2006) 977.
- [52] A. Makarov, *Anal. Chem.* 72 (2000) 1156.
- [53] F.W. McLafferty, *Interpretation of Mass Spectra*, 4th ed., Wiley University Science Book, New York, 1993.
- [54] A.W.T. Bristow, K.S. Webb, A.T. Lubben, J. Halket, *Rapid Commun. Mass Spectrom.* 18 (2004) 1447.
- [55] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, J.C. Tabet, *Abstracts of Papers of the 229th ACS National meeting*, 2005, p. ANYL-212.
- [56] J. Castro-Perez, M. McCullagh, A. Millar, *Waters Technical Note*, 720001170EN, 2005.
- [57] M. McCullagh, J. Castro-Perez, L. Calton, *Waters Technical Note*, 720001596EN, 2006.
- [58] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, *Anal. Chem.* 78 (2006) 2113.
- [59] H.K. Lim, J. Chen, C. Senseshauser, K. Cook, V. Subrahmanyam, *Rapid Commun. Mass Spectrom.* 21 (2007) 1821.
- [60] S.M. Peterman, N. Duczak Jr., A.S. Kalgutkar, M.E. Lame, J.R. Soglia, *J. Am. Soc. Mass Spectrom.* 17 (2006) 363.
- [61] Q. Ruan, S. Peterman, M.A. Szewc, L. Ma, D. Cui, W.G. Humphreys, M. Zhu, *J. Mass Spectrom.* 43 (2008) 251.
- [62] B.P. Koch, T. Dittmar, M. Witt, G. Kattner, *Anal. Chem.* 79 (2007) 1758.
- [63] J. Meija, *Anal. Bioanal. Chem.* 385 (2006) 486.
- [64] F.W. McLafferty, F. Turecek, *Interpretation of Mass Spectra*, 4th ed., Wiley University Science Book, New York, 1993.
- [65] A.W. Bristow, *Mass Spectrom. Rev.* 25 (2006) 99.
- [66] T. Kind, O. Fiehn, *BMC Bioinform.* 8 (2007) 105.
- [67] T. Kind, O. Fiehn, *BMC Bioinform.* 7 (2006) 234.
- [68] K.D. Henry, E.R. Williams, B.H. Wang, F.W. McLafferty, J. Shabanowitz, D.F. Hunt, *Proc. Natl. Acad. Sci. U.S.A.* 86 (1989) 9075.
- [69] S. Ojanpera, A. Pelander, M. Pelzing, I. Krebs, E. Vuori, I. Ojanpera, *Rapid Commun. Mass Spectrom.* 20 (2006) 1161.
- [70] A.H. Grange, M.C. Zumwalt, G.W. Sovocool, *Rapid Commun. Mass Spectrom.* 20 (2006) 89.

- [71] A.D. Hegeman, C.F. Schulte, Q. Cui, I.A. Lewis, E.L. Huttlin, H. Eghbalnia, A.C. Harms, E.L. Ulrich, J.L. Markley, M.R. Sussman, *Anal. Chem.* 79 (2007) 6912.
- [72] A.C. Stenson, A.G. Marshall, W.T. Cooper, *Anal. Chem.* 75 (2003) 1275.
- [73] E.B. Kujawinski, M.D. Behn, *Anal. Chem.* 78 (2006) 4363.
- [74] C.A. Hughey, C.L. Hendrickson, R.P. Rodgers, A.G. Marshall, K. Qian, *Anal. Chem.* 73 (2001) 4676.
- [75] H.J. Yoo, H. Liu, K. Hakansson, *Anal. Chem.* 79 (2007) 7858.
- [76] R.L. Woodin, D.S. Bomse, J.L. Beauchamp, *J. Am. Soc. Mass Spectrom.* 100 (1978) 3248.
- [77] D.P. Little, J.P. Speir, M.W. Senko, P.B. O'Connor, F.W. McLafferty, *Anal. Chem.* 66 (1994) 2809.
- [78] N.A. Kruger, R.A. Zubarev, D.M. Horn, F.W. McLafferty, *Int. J. Mass Spectrom.* 187 (1999) 787.
- [79] K. Levsen, *Fundamental Aspects of Organic Mass Spectrometry*, Verlag chemie, Weinheim, 1978.
- [80] R.G. Cooks, *Collision spectroscopy*, vol. xiv, Plenum Press, New York, 1978.
- [81] K.L. Busch, G.L. Glish, S.A. McLuckey, *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, VCH Publishers, New York, 1988.
- [82] E. Clayton, R.H. Bateman, *Rapid Commun. Mass Spectrom.* 6 (1992) 719.
- [83] U. Lewandrowski, A. Resemann, A. Sickmann, *Anal. Chem.* 77 (2005) 3274.
- [84] R.G. Cooks, J.H. Beynon, J.F. Litton, *Org. Mass Spectrom.* 10 (1975) 503.
- [85] R.J. Cotter, W. Griffith, C. Jelinek, *J. Chromatogr. B* 855 (2007) 2.
- [86] G. Sun, K. Yang, Z. Zhao, S. Guan, X. Han, R.W. Gross, *Anal. Chem.* 79 (2007) 6629.
- [87] N.J. Jensen, K.B. Tomer, M.L. Gross, *J. Am. Chem. Soc.* 107 (1985) 1863.
- [88] K.B. Tomer, F.W. Crow, M.L. Gross, *J. Am. Chem. Soc.* 105 (1983) 5487.
- [89] J. Adams, *Mass Spectrom. Rev.* 9 (1990) 141.
- [90] C.F. Cheng, M.L. Gross, *Mass Spectrom. Rev.* 19 (2000) 398.
- [91] R.C. Murphy, J. Fiedler, J. Hevko, *Chem. Rev.* 101 (2001) 479.
- [92] R.E. March, H.X. Li, O. Belgacem, D. Papanastasiou, *Int. J. Mass Spectrom.* 262 (2007) 51.
- [93] S.A. McLuckey, D.E. Goeringer, *J. Mass Spectrom.* 32 (1997) 461.
- [94] D.F. Hunt, J. Shabanowitz, J.R. Yates, *J. Chem. Soc. Chem. Commun.* 8 (1987) 548.
- [95] Y.O. Tsybin, M. Witt, G. Baykut, F. Kjeldsen, P. Hakansson, *Rapid Commun. Mass Spectrom.* 17 (2003) 1759.
- [96] Y. Hashimoto, H. Hasegawa, K. Yoshinari, I. Waki, *Anal. Chem.* 75 (2003) 420.
- [97] J.J. Wilson, J.S. Brodbelt, *Anal. Chem.* 79 (2007) 2067.
- [98] R.B. Cody, B.S. Freiser, *Anal. Chem.* 51 (1979) 547.
- [99] B.A. Budnik, K.F. Haselmann, Y.N. Elkin, V.I. Gorbach, R.A. Zubarev, *Anal. Chem.* 75 (2003) 5994.
- [100] H. Idborg-Bjorkman, P.O. Edlund, O.M. Kvalheim, I. Schuppe-Koistinen, S.P. Jacobsson, *Anal. Chem.* 75 (2003) 4784.
- [101] S.U. Bajad, W.Y. Lu, E.H. Kimball, J. Yuan, C. Peterson, J.D. Rabinowitz, *J. Chromatogr. A* 1125 (2006) 76.
- [102] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, E. Ezan, J.C. Tabet, *J. Mass Spectrom.* 39 (2004) 655.
- [103] S. Wagner, K. Scholz, M. Donegan, L. Burton, J. Wingate, W. Volkel, *Anal. Chem.* 78 (2006) 1296.
- [104] Z. Yan, G.W. Caldwell, *Anal. Chem.* 76 (2004) 6835.
- [105] W. Bicker, M. Lammerhofer, D. Genser, H. Kiss, W. Lindner, *Toxicol. Lett.* 159 (2005) 235.
- [106] L. Coulier, R. Bas, S. Jespersen, E. Verheij, M.J. van der Werf, T. Hankemeier, *Anal. Chem.* 78 (2006) 6573.
- [107] G.C. Kite, M.J. Howes, M.S. Simmonds, *Rapid Commun. Mass Spectrom.* 18 (2004) 2859.
- [108] M. Bantscheff, M. Boesche, D. Eberhard, T. Matthieson, G. Sweetman, B. Kuster, *Mol. Cell Proteomics* (2008).
- [109] T.J. Griffin, H. Xie, S. Bandhakavi, J. Popko, A. Mohan, J.V. Carlis, L. Higgins, *J. Proteome Res.* 6 (2007) 4200.
- [110] J.C. Schwartz, M.W. Senko, J.E. Syka, *J. Am. Soc. Mass Spectrom.* 13 (2002) 659.
- [111] A. Koulman, B.A. Tapper, K. Fraser, M. Cao, G.A. Lane, S. Rasmussen, *Rapid Commun. Mass Spectrom.* 21 (2007) 421.
- [112] J.W. Hager, J.C.Y. Le Blanc, *Rapid Commun. Mass Spectrom.* 17 (2003) 1056.
- [113] Y.A. Hammel, R. Mohamed, E. Gremaud, M.H. Lebreton, P.A. Guy, *J. Chromatogr. A* 1177 (2008) 58.
- [114] G. Hopfgartner, E. Varesio, V. Tschappat, C. Grivet, E. Bourgoigne, L.A. Leuthold, *J. Mass Spectrom.* 39 (2004) 845.
- [115] A.C. Li, M.A. Gohdes, W.Z. Shou, *Rapid Commun. Mass Spectrom.* 21 (2007) 1421.
- [116] A.N. Krutchinsky, I.V. Chernushevich, V.L. Spicer, W. Ens, K.G. Standing, *J. Am. Soc. Mass Spectrom.* 9 (1998) 569.
- [117] A.D. Patterson, H. Li, G.S. Eichler, K.W. Krausz, J.N. Weinstein, A.J. Fornace Jr., F.J. Gonzalez, J.R. Idle, *Anal. Chem.* 80 (2008) 665.
- [118] R.S. Plumb, K.A. Johnson, P. Rainville, B.W. Smith, I.D. Wilson, J.M. Castro-Perez, J.K. Nicholson, *Rapid Commun. Mass Spectrom.* 20 (2006) 1989.
- [119] K. Tanaka, T. Tamura, S. Fukuda, J. Batkhuu, C. Sanchir, K. Komatsu, *Phytochemistry* 69 (2008) 2081.
- [120] J.M. Campbell, B.A. Collings, D.J. Douglas, *Rapid Commun. Mass Spectrom.* 12 (1998) 1463.
- [121] T. Yokosuka, K. Yoshinari, K. Kobayashi, A. Ohtake, A. Hirabayashi, Y. Hashimoto, I. Waki, T. Takao, *Rapid Commun. Mass Spectrom.* 20 (2006) 2589.
- [122] J. Ding, C.M. Sorensen, Q. Zhang, H. Jiang, N. Jaitly, E.A. Livesay, Y. Shen, R.D. Smith, T.O. Metz, *Anal. Chem.* 79 (2007) 6081.
- [123] M. Thevis, A.A. Makarov, S. Horning, W. Schanzer, *Rapid Commun. Mass Spectrom.* 19 (2005) 3369.
- [124] J.V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, *Nat. Methods* 4 (2007) 709.
- [125] S.C. Brown, G. Kruppa, J.L. Dasseux, *Mass Spectrom. Rev.* 24 (2005) 223.
- [126] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith, *Anal. Chem.* 77 (2005) 3090.
- [127] M.M. Siegel, *Anal. Chem.* 60 (1988) 2090.
- [128] H.T. Chi, J.K. Baker, *Org. Mass Spectrom.* 28 (1993) 12.
- [129] M.D. Sierra, A. Furey, B. Hamilton, M. Lehane, K.J. James, *J. Mass Spectrom.* 38 (2003) 1178.
- [130] A.M. Kamel, B. Munson, *Eur. J. Mass Spectrom.* 10 (2004) 239.
- [131] K.E. Karlsson, *J. Chromatogr.* 647 (1993) 31.
- [132] D.Q. Liu, C.E.C.A. Hop, M.G. Beconi, A. Mao, S.H.L. Chiu, *Rapid Commun. Mass Spectrom.* 15 (2001) 1832.
- [133] A.M. Kamel, H.G. Fouda, P.R. Brown, B. Munson, *J. Am. Soc. Mass Spectrom.* 13 (2002) 543.
- [134] A.M. Kamel, K.S. Zandi, W.W. Massefski, *J. Pharm. Biomed. Anal.* 31 (2003) 1211.
- [135] B. Zhou, Z.Y. Zhang, *Methods* 42 (2007) 227.
- [136] M.E. Palmer, L.W. Tetler, I.D. Wilson, *Rapid Commun. Mass Spectrom.* 14 (2000) 808.
- [137] S. Campbell, M.T. Rodgers, E.M. Marzluff, J.L. Beauchamp, *J. Am. Chem. Soc.* 117 (1995) 12840.
- [138] D.Q. Liu, C.E.C.A. Hop, *J. Pharm. Biomed. Anal.* 37 (2005) 1.
- [139] C. Prakash, C.L. Shaffer, A. Nedderman, *Mass Spectrom. Rev.* 26 (2007) 340.
- [140] E.E. Carlson, B.F. Cravatt, *Nat. Methods* 4 (2007) 429.
- [141] E.E. Carlson, B.F. Cravatt, *J. Am. Chem. Soc.* 129 (2007) 15780.
- [142] M.C. Nicklaus, W.D. Ihlenfeldt, D. Filimonov, V.V. Poroikov, *Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26–30, 2001, CINF-091*.
- [143] J.S. Brecher, *Chimia* 52 (1998) 658.
- [144] K. Degtyarenko, M.P. de, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, M. Ashburner, *Nucleic Acids Res.* 36 (2008) D344–D350.
- [145] S. Goto, Y. Okuno, M. Hattori, T. Nishioka, M. Kanehisa, *Nucleic Acids Res.* 30 (2002) 402.
- [146] M. Kanehisa, S. Goto, M. Hattori, K.F. oki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, *Nucleic Acids Res.* 34 (2006) D354–D357.
- [147] M. Krummenacker, S. Paley, L. Mueller, T. Yan, P.D. Karp, *Bioinformatics* 21 (2005) 3454.
- [148] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, P.D. Karp, *Nucleic Acids Res.* 36 (2008) D623–D631.
- [149] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeronics, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Minciari, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. Macinnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, *Nucleic Acids Res.* 35 (2007) D521–D526.
- [150] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, *Ther. Drug Monit.* 27 (2005) 747.
- [151] Y. Shinbo, Y. Nakamura, M. Altaf-Ul-Amin, H. Asahi, K. Kurokawa, M. Arita, K. Saito, D. Ohta, D. Shibata, S. Kanaya, *Biotechnol. Agric. For.* 57 (2006) 165.
- [152] S. Moco, R.J. Bino, O. Vorst, H.A. Verhoeven, G.J. de, T.A. van Beek, J. Vervoort, C.H. de Vos, *Plant Physiol.* 141 (2006) 1205.
- [153] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A.R. Fernie, D. Steinhauser, *Bioinformatics* 21 (2005) 1635.
- [154] R.C. Willoughby, R.F. Browner, *Anal. Chem.* 56 (1984) 2626.
- [155] J.C. Wolff, P.N. Hawtin, S. Monte, M. Balogh, T. Jones, *Rapid Commun. Mass Spectrom.* 15 (2001) 265.
- [156] R. Jansen, G. Lachatre, P. Marquet, *Clin. Biochem.* 38 (2005) 362.
- [157] M. Arita, *Genome Res.* 13 (2003) 2455.
- [158] M. Arita, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 1543.
- [159] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reilly, J.J. Thaden, M.R. Viant, *Metabolomics* 3 (2007) 211.
- [160] M.L. Bandu, T. Grubbs, M. Kater, H. Desaire, *Int. J. Mass Spectrom.* 251 (2006) 40.
- [161] R. Tautenhahn, C. Bottcher, S. Neumann, in: S. Hochreiter, R. Wagner (Eds.), *Bioinformatics Research and Development*, Springer, Berlin, 2007, p. 371.
- [162] C.B. Clish, E. Davidov, M. Oresic, T.N. Plasterer, G. Lavine, T. Londo, M. Meys, P. Snell, W. Stochaj, A. Adourian, X. Zhang, N. Morel, E. Neumann, E. Verheij, J.T. Vogels, L.M. Haveskes, N. Afeyan, F. Regnier, J. van der Greef, S. Naylor, *OMICS* 8 (2004) 3.
- [163] E. Lange, C. Gropl, O. Schulz-Trieglaff, A. Leinenbach, C. Huber, K. Reinert, *Bioinformatics* 23 (2007) i273–i281.

- [165] O. Cloarec, M.E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J.C. Lindon, E. Holmes, J. Nicholson, *Anal. Chem.* 77 (2005) 1282.
- [166] E. Werner, V. Croixmarie, T. Umbdenstock, E. Ezan, P. Chaminade, J.C. Tabet, C. Junot, *Anal. Chem.* 80 (2008) 4918.
- [167] E. Kendrick, *Anal. Chem.* 35 (1963) 2146.
- [168] Z. Wu, R.P. Rodgers, A.G. Marshall, *Anal. Chem.* 76 (2004) 2511.
- [169] H. He, C.A. Conrad, C.L. Nilsson, Y. Ji, T.M. Schaub, A.G. Marshall, M.R. Emmett, *Anal. Chem.* 79 (2007) 8423.
- [170] D.W. van Krevelen, *Fuel* 29 (1950) 269.
- [171] S. Kim, R.W. Kramer, P.G. Hatcher, *Anal. Chem.* 75 (2003) 5336.
- [172] J.H. Reuter, E.M. Perdue, *Mitteilungen des Geologisch-Paläontologischen Instituts der Universität Hamburg* 56 (1984) 249.
- [173] J.I. Hedges, *Adv. Chem. Ser.* 225 (1990) 111.
- [174] R.L. Sleighter, P.G. Hatcher, *J. Mass Spectrom.* 42 (2007) 559.
- [175] J.C. Lindon, J.K. Nicholson, E. Holmes, H.C. Keun, A. Craig, J.T. Pearce, S.J. Bruce, N. Hardy, S.A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R.D. Beger, E.R. Verheij, A. Amberg, D. Baunsgaard, G.H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J.N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I.D. Wilson, M.D. Reily, D.G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, M. Spraul, *Nat. Biotechnol.* 23 (2005) 833.
- [176] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Calmani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjostrom, J. Trygg, F. Wulfer, *Metabolomics* 3 (2007) 231.
- [177] N.W. Hardy, C.F. Taylor, *Metabolomics* 3 (2007) 243.
- [178] M.R. Shortreed, S.M. Lamos, B.L. Frey, M.F. Phillips, M. Patel, P.J. Belshaw, L.M. Smith, *Anal. Chem.* 78 (2006) 6398.
- [179] K. Guo, C. Ji, L. Li, *Anal. Chem.* 79 (2007) 8631.
- [180] J.C. Lindon, J.K. Nicholson, I.D. Wilson, *J. Chromatogr. B* 748 (2000) 233.
- [181] J.L. Wolfender, K. Ndjoko, K. Hostettmann, *Phytochem. Anal.* 12 (2001) 2.
- [182] Z. Yang, *J. Pharm. Biomed. Anal.* 40 (2006) 516.
- [183] G. Glauser, D. Guillaume, E. Grata, J. Boccard, A. Thiocone, P.A. Carrupt, J.L. Veuthey, S. Rudaz, J.L. Wolfender, *J. Chromatogr. A* 1180 (2008) 90.
- [184] D.L. Olson, J.A. Norcross, M. O'Neil-Johnson, P.F. Molitor, D.J. Detlefsen, A.G. Wilson, T.L. Peck, *Anal. Chem.* 76 (2004) 2966.
- [185] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H. Hill Jr., *J. Mass Spectrom.* 43 (2008) 1.
- [186] P.P. Dwivedi, P. Wu, S.J. Klopsch, G.J. Puzon, L. Xun, H.H. Hill, *Metabolomics* 4 (2008) 63.
- [187] A. Cappiello, G. Famigliini, E. Pierini, P. Palma, H. Truffelli, *Anal. Chem.* 79 (2007) 5364.
- [188] A. Amirav, O. Granot, *J. Am. Soc. Mass Spectrom.* 11 (2000) 587.
- [189] O. Granot, A. Amirav, *Int. J. Mass Spectrom.* 244 (2005) 15.
- [190] S. Orchard, W. Zhu, R.K. Julian Jr., H. Hermjakob, R. Apweiler, *Proteomics* 3 (2003) 2065.
- [191] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, M. Vingron, *Nat. Genet.* 29 (2001) 365.
- [192] R.J. Bino, R.D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B.J. Nikolau, P. Mendes, U. Roessner-Tunali, M.H. Beale, R.N. Trethewey, B.M. Lange, E.S. Wurtele, L.W. Sumner, *Trends Plant Sci.* 9 (2004) 418.
- [193] H. Jenkins, N. Hardy, M. Beckmann, J. Draper, A.R. Smith, J. Taylor, O. Fiehn, R. Goodacre, R.J. Bino, R. Hall, J. Kopka, G.A. Lane, B.M. Lange, J.R. Liu, P. Mendes, B.J. Nikolau, S.G. Oliver, N.W. Paton, S. Rhee, U. Roessner-Tunali, K. Saito, J. Smedsgaard, L.W. Sumner, T. Wang, S. Walsh, E.S. Wurtele, D.B. Kell, *Nat. Biotechnol.* 22 (2004) 1601.
- [194] I. Spasic, W.B. Dunn, G. Velarde, A. Tseng, H. Jenkins, N. Hardy, S.G. Oliver, D.B. Kell, *BMC Bioinform.* 7 (2006) 281.
- [195] M.E. Dumas, C. Canlet, L. Debrauwer, P. Martin, A. Paris, J. Proteome Res. 4 (2005) 1485.
- [196] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zerah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, *Anal. Chem.* 78 (2006) 363.
- [197] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, B.O. Palsson, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 1777.
- [198] P. Romero, J. Wagg, M.L. Green, D. Kaiser, M. Krumpal, P.D. Karp, *Genome Biol.* 6 (2005) R2.
- [199] M. Kanehisa, S. Goto, *Nucleic Acids Res.* 28 (2000) 27.
- [200] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, *Nucleic Acids Res.* 35 (2007) W606–W612.
- [201] K. Watanabe, E. Yasugi, M. Oshima, *Trends Glycosci. Glycotechnol.* 12 (2000) 175.
- [202] H. Horai, K. Suwa, M. Arita, Y. Nihei, T. Nishioka, *Abstracts of Papers of the 55th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, USA, 2007.*
- [203] O. Yamamoto, K. Someno, N. Wasada, J. Hiraishi, K. Hayamizu, K. Tanabe, T. Tamura, M. Yanagisawa, *Anal. Sci.* 4 (1988) 233.





## **Résumé**

Issue d'un partenariat de recherche entre le CEA et les laboratoires Servier, cette thèse avait pour objectif d'évaluer l'approche métabolomique par chromatographie liquide couplée à la spectrométrie de masse (LC-MS) pour l'identification de marqueurs indirects de l'induction dans les espèces de toxicologie. Le travail de thèse a débuté par l'optimisation de la méthode d'acquisition des empreintes métaboliques tant sur le plan analytique que dans le domaine du traitement des données brutes. Un outil reposant sur les matrices d'autocorrélation a alors été développé afin de s'affranchir d'une partie de la redondance du signal obtenu par spectrométrie de masse. Dans un second temps, les indices de Kendrick couplés à la sélectivité méthylène ont été appliqués à l'étude de composés biologiques en spectrométrie de masse haute résolution afin de proposer une méthode alternative de visualisation des données offrant une aide à l'identification des variables. Enfin, dans une troisième partie, les efforts se sont portés sur l'identification des composés endogènes modifiés au cours du protocole d'induction.

## **Mots-clés**

Métabolomique – Chromatographie liquide couplée à la spectrométrie de masse – Traitement du signal – Outils de visualisation – Biomarqueurs – Identification – Induction enzymatique

## **Laboratoires**

Laboratoire de Chimie analytique – Faculté de Pharmacie  
Groupe de chimie analytique de Paris sud – EA 4041

Laboratoire d'Etude du Métabolisme des Médicaments (LEMM)  
Service de Pharmacologie et d'immuno-analyse (SPI)  
iBiTec-S / Direction des Sciences du Vivant (DSV)  
CEA Saclay

UNIVERSITÉ PARIS-SUD 11

UFR «FACULTÉ DE PHARMACIE DE CHATENAY-MALABRY »

5, rue Jean Baptiste Clément

92296 CHÂTENAY-MALABRY Cedex