



HAL
open science

Développement de l'approche métabolomique par couplage chromatographie liquide / spectrométrie de masse : application à la nutrition

Hélène Pereira

► **To cite this version:**

Hélène Pereira. Développement de l'approche métabolomique par couplage chromatographie liquide / spectrométrie de masse : application à la nutrition. Alimentation et Nutrition. Université Blaise Pascal - Clermont-Ferrand II, 2010. Français. NNT : 2010CLF22013 . tel-00719399

HAL Id: tel-00719399

<https://theses.hal.science/tel-00719399v1>

Submitted on 19 Jul 2012

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

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

Numéro DU : 2013

Année 2010

THESE DE L'UNIVERSITE DE CLERMONT-FERRAND

Délivrée par

L'UNIVERSITE BLAISE PASCAL

ECOLE DOCTORALE DE SCIENCES FONDAMENTALES (636)

DIPLOME DE DOCTORAT (arrêté du 7 août 2006)

soutenue publiquement

le lundi 01 février 2010 par Hélène PEREIRA

**DEVELOPPEMENT DE L'APPROCHE METABOLOMIQUE
par couplage chromatographie liquide / spectrométrie de masse :
Application à la nutrition**

Directeurs de thèse : Jean Louis Sébédio et Estelle Pujos-Guillot

JURY

Laurent DEBRAUWER, rapporteur

Bruno LE BIZEC, rapporteur

Claire LARTIGUE, membre du jury

Mohamed SARAHA, membre du jury

Estelle PUJOS-GUILLOT, membre du jury

Jean Louis SEBEDIO, membre du jury

William CHRISTIE, invité

REMERCIEMENT

Je remercie Estelle Pujos-Guillot et Jean Louis Sébédio pour leur accueil au sein de la plateforme d'exploration du métabolisme et de leurs critiques. Je remercie tout particulièrement Estelle pour m'avoir permis de participer à de nombreux congrès.

J'adresse mes remerciements aux membres du jury pour avoir accepté de relire ce manuscrit et d'avoir jugé objectivement ce travail. Je suis reconnaissante à Claire Lartigue d'avoir présidé ce jury de soutenance et d'avoir apporté les corrections nécessaires.

J'adresse mes sincères remerciements à l'ANR METAPROFILE qui m'a financé pendant ces trois ans ainsi qu'à ces collaborateurs qui m'ont permis de développer cette approche.

Je remercie sincèrement les membres de mon comité de thèse (Jean Philippe Antignac , Christophe Junot et Bruno Comburieux) qui m'ont apporté un regard externe et expert sur mes travaux. Cet échange scientifique m'a permis de faire le point et de prendre du recul.

Un grand merci à Christophe Junot pour m'avoir ouvert les portes du CEA et m'avoir permis d'utiliser l'Orbitrap.

Je remercie tout particulièrement Jean François Martin de sa présence et de son aide tout au long de ces trois années. Merci de m'avoir transmis une partie de ton savoir en statistique et de m'avoir éclairé sur ce sujet large et complexe qu'est la métabolomique. Merci pour tes corrections et pour toutes les discussions « métabolomique » et personnelles échangées.

Je remercie également Bernard Lyan, Corinne Pouyet, pour tous les nombreux échanges et questionnements scientifiques tout au long de ces trois années. Merci pour votre ouverture d'esprit et votre intérêt pour cette discipline qu'est la chimie analytique et la spectrométrie de masse.

Merci à mes collègues du QTOF, Charlotte Joly, Marie Laure Bayle et Bernard Lyan avec qui nous avons dû résoudre de nombreuses pannes.

Merci à toute l'équipe, Charlotte Joly, Bernard Lyan, Carole Mignet, Corinne Pouyet, Jean-François Martin, Anne de la Foye et Anne Diot pour les grandes et petites discussions que nous avons eu qui ont toujours été très agréables et enrichissantes. Merci de leur soutien et leur écoute.

Un énorme merci à mes collègues non titulaires de bureau : Marie Laure et Xianfu dont les échanges scientifiques et personnels ont été très intéressants et très enrichissants pour mener à bien ce projet. Ils m'ont permis de me sentir moins seule dans ce monde complexe de la métabolomique.

Certains se sont investis dans la relecture et la correction constructive du présent manuscrit. Parce qu'ils le méritent bien, je les cite donc et en profite pour grandement les remercier : Anne de la Foye, Marie Laure, Bernard, Aurélie, Julia, David, Delphine, Lucie, Bénédicte. Merci à Nina pour notre week-end correction.

Merci à Sylvie Mercier pour m'avoir aidé à imprimer mon manuscrit. Merci à Isabelle Papet pour avoir eu la gentillesse de répondre à mes questions biologiques et de m'avoir aidé à corriger mon oral de présentation.

Merci à mes collègues de l'INRA Avenir, Mylène, Aurélie, Christelle, Solange et Vincent pour avoir réussi à réaliser cette journée INRA avenir 2008 pleine de rebondissement. Merci à tous les membres de l'ASICT et aux non permanents.

Pour finir, je voulais remercier très sincèrement mes collègues et ami(e)s du théâtre : Aurélie, Marjorie, Clementine, Julia, François, Nico, David ; mes colocataires Lucile, Marie et Tom ; mes ami(e)s de toujours : Nina, Del, Béné, Lucie, Tony, Charles, Eve, David et ma famille qui m'ont toujours encouragés.

Enfin merci à tous ceux que j'ai rencontré et qui ont contribué à mon adaptation dans cette région pas si austère que cela : Ania, Guillaume, Aurélie, Elena, Esperanza, José, Julia, , Mathilde, Marco, Sylvain, Damien, Marie,

LISTE DES ABBREVIATIONS

ACP Analyse en Composante Principale

ADN: Acide DésoxyriboNucléique

ANCOVA: Analyse de la Covariance

ANOVA: Analyse de la variance

ANR: Agence Nationale de Recherche

ARN : Acide RiboNucléique

DG : DiacylGlycérol

HMDB: Human Metabolom Database

HPLC : High Performance Liquid Chromatography

KEGG: Kyoto Encyclopedia of Genes and Genomes

LC: Chromatographie Liquide (Liquid chromatography)

MS: Spectrométrie de masse (Mass Spectrometry)

PC: Phosphatidylcholine

PE: Phosphatidyléthanolamine

PLS-DA : Partial Least Square and Discrimination Analysis

QTOF: Quadripôle couplé à un Temps de vol (Quadripôle and Time of Flight)

RMN: Résonance Magnétique Nucléaire

UPLC: Ultra Performance Liquid Chromatography

TABLE DES MATIERES

Table des matières

TABLE DES MATIERES	7
Introduction générale.....	17
Références	20
Chapitre I.....	25
I. Contexte	25
1. Données épidémiologiques.....	25
2. Développement de l'obésité et ses complications	28
3. Prévention.....	29
4. Marqueurs biologiques de l'obésité.....	29
II. Approches métabolomiques.....	32
1. L'approche ciblée	33
2. L'approche non ciblée	33
III. Techniques analytiques	35
1. Système de Chromatographie Liquide	36
2. Spectrométrie de masse	38
IV. Traitement des données.....	44
1. Traitement du signal	44
2. Extraction et alignement des ions.....	45
3. Prétraitement statistique	47
4. Analyses statistiques.....	47
V. Identification des ions sélectionnés	49
1. Les bases de données de métabolites.....	49
2. Description du LTQ-(en anglais Linear Trap Quadripole)-Orbitrap.....	50
VI. Validation des biomarqueurs.....	52
Références	53
Chapitre II	61
I. Introduction.....	61
II. Matériels et méthodes	64
1. Produits chimiques	64
2. Prélèvement et préparation des échantillons biologiques.....	64

3. Echantillons utilisés pour la validation de la méthode analytique.....	65
4. Analyse par LC/MS des échantillons plasmatiques.	65
5. Traitement de données.....	66
6. Analyse statistique des données	67
III. Résultats et discussions	68
1. Optimisation des conditions UPLC-MS	68
2. Prélèvement et préparation des échantillons plasmatiques.....	70
3. Validation de la méthode développée.....	75
IV. Conclusion.....	83
Références	84
Chapitre III	93
I. Introduction.....	93
II. Matériels et Méthodes.....	96
1. Traitements des échantillons biologiques.....	96
2. Traitement des données	97
III. Résultats	106
1. Choix des paramètres des logiciels MarkerLynx™ et XCMS.	106
2. Etude des matrices de données optimisées.....	109
3. Choix des méthodes de prétraitements appliquées à l'OSC-PLS-DA.....	113
IV. Discussion	116
1. Choix du paramètre du signal.....	116
2. Optimisation des paramètres des logiciels.....	116
3. Evaluation de la qualité des matrices de données obtenues.	117
4. Prétraitement des données	117
V. Conclusion.....	119
Référence.....	120
Chapitre IV	125
I. Introduction.....	125
II. Matériels et Méthodes.....	126
1. Etude d'intervention chez l'Homme	126
2. Traitement des échantillons.....	126
3. Traitement des données	127
4. Stratégie d'identification	128
III. Résultats	130

1. Ions sélectionnés en mode positif et négatif.....	130
2. Identification des ions moléculaires	133
IV. Discussion	148
V. Conclusion.....	150
Références	151
Conclusions et perspectives.....	155
Références	159
Chapitre V	163
I. Enjeux scientifiques de la thèse.	163
1. Contexte.....	163
2. Projet METAPROFILE financé par l'Agence Nationale pour la Recherche.....	163
II. Le sujet de thèse.....	165
1. Choix de la thèse	165
2. Moyens humains mis en place en place pour réaliser ce projet.....	165
3. Le réseau développé	166
III. Déroulement du projet.....	168
1. Préparation et orientation du projet	168
2. Conduite du projet	168
3. Coût du projet (Annexe 1).....	170
IV. Principales compétences développées au cours de la thèse	171
1. Compétences Scientifiques.....	171
2. Connaissances méthodologiques	172
3. Qualités personnelles exprimées	173
4. Réseau professionnel développé.....	174
V. Conclusion.....	175
Lexique.....	176
1. Structures & programmes.....	176
2. Termes scientifiques et techniques	176
Annexe 1 Evaluation du coût de la thèse.....	177
Annexe 2 Valorisation Scientifique.....	183
ANNEXE I (CHAPITRE II)	185
ANNEXE 2 (CHAPITRE II).....	191
ANNEXE 3 (CHAPITRE III)	193

"Science sans conscience n'est que ruine de l'âme". Rabelais

INTRODUCTION GENERALE

Introduction générale

L'étude des macromolécules biologiques telles que l'ADN (Acide DésoxyriboNucléique) , l'ARN (Acide RiboNucléique) ou les protéines ont permis ces dernières années de grandes avancées dans l'étude des mécanismes qui régulent l'expression des gènes. (Verhoeckx *et al.*, 2004). Toutefois, même si ces études ont permis de révéler la complexité de l'interaction existante entre la génétique et les facteurs environnementaux, il est désormais admis que la compréhension de ces interactions est impossible par la transcriptomique et difficile par la protéomique. (Nicholson *et al.*, 2004). Afin d'appréhender les problèmes de biocomplexité, il est nécessaire de suivre des **marqueurs biologiques dynamiques**, tels que **les métabolites** qui sont impliqués dans de nombreux processus biologiques. Par conséquent, l'étude des métabolites dans les fluides biologiques (Bollard *et al.*, 2005) fournit des informations instantanées sur les modifications de l'organisme induites par des facteurs internes et externes tels que la modification de l'expression des gènes (Raansdonk *et al.*, 2001) ou des facteurs environnementaux. Cette approche est la métabolomique.

Les termes de « **métabonomique ou métabolomique** » ont été définis par J. Nicholson (1999) puis par O. Fiehn (2001) comme « **la mesure quantitative de la réponse multiparamétrique liée au temps d'un système multicellulaire à des stimuli physiopathologiques ou à des modifications génétiques** ». Cette approche nécessite des techniques analytiques performantes dites haut débit tels que la **Résonance Magnétique Nucléaire (RMN)** et la **spectrométrie de masse**, capables de détecter la variation de la réponse de milliers de métabolites présents dans les fluides et tissus biologiques. Du fait de l'importance des données générées par ces techniques, l'utilisation d'analyses **statistiques uni- et multivariées** (Nicholson *et al.*, 1999) est indispensable pour l'interprétation de ces données analytiques complexes et la détermination de marqueurs biologiques potentiels.

De nombreuses études ont montré l'apport de cette approche dans différents domaines tels que la pharmacologie (Clayton *et al.*, 2006), la toxicologie (Beckwith-Hall *et al.*, 2002), la botanique (Jenkins *et al.*, 2004). Les études en nutrition sont encore récentes mais différents programmes de recherche concernent l'identification de marqueurs précoces de déséquilibres métaboliques associés à l'apparition de pathologies mais également l'identification de marqueurs.

Aujourd'hui, l'un des défis majeurs en nutrition est de lutter contre l'augmentation rapide de l'obésité induisant des complications souvent mortelles. (German *et al.*, 2005). Cette augmentation s'explique en partie par un dérèglement du métabolisme lié à une alimentation de plus en plus riche en lipides. Toutefois les connaissances scientifiques actuelles ne permettent pas de préciser les mécanismes mis en jeu. Dans ce contexte, l'utilisation d'une approche métabolomique, prenant en compte la biocomplexité, est une alternative à l'approche classique qui consiste à étudier l'effet d'un régime ou bien d'un nutriment donné sur une fonction particulière ou un organe cible.

Ainsi, l'objectif du projet METAPROFILE financé par l'ANR (Agence Nationale pour la Recherche) vise à **déterminer des biomarqueurs précoces de la prise de poids, suite à une alimentation hyperlipidique**. Pour répondre à cet objectif, deux types d'expérimentations ont été menées : l'une chez l'homme et l'autre chez la souris. L'étude métabolomique a été réalisée sur les fluides biologiques (urine et plasma) prélevés au cours des interventions, et analysés par Résonance Magnétique Nucléaire (RMN) et spectrométrie de masse (MS en anglais *Mass Spectrometry*).

Dans ce cadre, il a été essentiel de développer des méthodes capables de couvrir les différentes classes de métabolites et leur large gamme de concentrations dans les matrices biologiques. Ces méthodes analytiques devaient être robustes et capables de détecter les faibles variations induites par les interventions nutritionnelles. Nous avons choisi de baser notre étude sur l'utilisation de la spectrométrie de masse, qui est une des techniques analytiques les plus sensibles et « universelle » (Dettmer *et al.*, 2007). De plus, cette technique peut être couplée en aval par une chromatographie liquide (LC en anglais *Liquid Chromatography*) pour augmenter le nombre de molécules détectées.

Aux vues de ces différents éléments, ce travail s'est attaché à **étudier et optimiser les différentes étapes d'une approche métabolomique par spectrométrie de masse** : collecte et préparation de l'échantillon, acquisition des profils métaboliques, traitement des données analytiques et identification des biomarqueurs potentiels.

La première partie de ce manuscrit est consacrée aux rappels bibliographiques sur la problématique de la lutte contre l'obésité, l'apport des nouvelles approches métabolomiques, ainsi que la description des techniques analytiques mises en œuvre .

Le deuxième chapitre est consacré au développement et la validation d'une méthode d'analyse de profils plasmatiques.

Le troisième s'est intéressé à l'optimisation de la chaîne de traitement des données afin d'être capable de mettre en évidence les variations biologiques induites par les régimes.

Enfin, dans une dernière partie, nous nous sommes attachés à illustrer l'étape d'identification et de caractérisation des métabolites, à partir des échantillons plasmatiques issus de l'expérimentation chez l'homme.

Une conclusion générale reprend les résultats importants de cette étude et présente les perspectives de travail.

Références

Beckwith-Hall, B. M., Holmes, E., Lindon, J. C., Gounarides, J., Vickers, A., Shapiro, M. and Nicholson, J. K.: 2002, 'NMR-Based Metabonomic Studies on the Biochemical Effects of Commonly Used Drug Carrier Vehicles in the Rat.' *Chemical Research in Toxicology* **15**, 1136-1141.

Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K. and Holmes, E.: 2005, 'NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition.' *NMR IN BIOMEDICINE* **18**, 143–162.

Clayton, T. A., Lindon, J. C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., Provost, J.-P., Le Net, J.-L. c., Baker, D., Walley, R. J., Everett, J. R. and Nicholson, J. K.: 2006, 'Pharmacometabonomic phenotyping and personalized drug treatment.' **440**, 1073-1077.

Dettmer, K., Pavel, A.A., and Hammock, B.D.: 2007, Mass spectrometry Based Metabolomics. *Mass Spectrom. Reviews* **26**, 51-78.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N. and Willmitzer, L.: 2000, 'Metabolite profiling for plant functional genomics.' *Nature* **18**, 1157-1161.

German, J. B., Watkins, S. M. and Fay, L.-B.: 2005, 'Metabolomics in Practice: Emerging Knowledge to Guide Future Dietetic Advice toward Individualized Health.' *J Am Diet Assoc.* **105**:, 1425-1432.

Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R. J., Hall, R., Kopka, J., Lane, G. A., Lange, B. M., Liu, J. R., Mendes, P., Nikolau, B. J., Oliver, S. G., Paton, N. W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L. W., Wang, T., Walsh, S., Wurtele, E. S. and Kell, D. B.: 2004, 'A proposed framework for the description of plant metabolomics experiments and their results.' *Nature Biotechnology* **12**, 1601-1606.

Nicholson, J., Holmes, E., Lindon, J. and Wilson, I.: 2004, 'The challenge of modelling mammalian biocomplexity.' *Nature Biotechnology* **22**, 1268-1274.

Nicholson, J., Lindon, J. and Holmes, E.: 1999, 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.' *Xenobiotica* **29**, 1181-9.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berden, J. A., Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V., Dam, K. v. and Oliver, S. G.: 2001, 'A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.' *Nature Biotechnology* **19**, 45-50.

Verhoeckx, K. C. M., Bijlsma, S., Jespersen, S., Ramaker, R., Verheij, E. R., Witkamp, R. F., van der Greef, J. and Rodenburg, R. J. T.: 2004, 'Characterization of anti-inflammatory compounds using transcriptomics, proteomics, and metabolomics in combination with multivariate data analysis.' *International Immunopharmacology* **4**, 1499-1514.

CHAPITRE I : CONTEXTE ET TECHNIQUES

Chapitre I

I. Contexte

Le surpoids et l'obésité représentent un problème de santé publique majeur aussi bien dans les pays industrialisés que dans les pays en voie de développement. En 1997, l'obésité a été reconnue comme une maladie par l'Organisation Mondiale de la Santé.

L'obésité est définie comme un « *état caractérisé par un excès de masse adipeuse répartie de façon généralisée dans les diverses zones grasses de l'organisme* ». (Dictionnaire de médecine Flammarion)

1. Données épidémiologiques

L'embonpoint chez la femme ou l'homme fut longtemps synonyme de fertilité et de richesse. Pourtant, dès l'antiquité, Hippocrate observa que « *l'Obésité n'était pas une maladie en soi mais l'embryon d'autres* ». D'après ses observations, les patients en surpoids étaient moins fertiles et mouraient prématurément.

A la fin du 19^{ème} siècle, le développement des statistiques et les observations médicales faites depuis des siècles, permirent d'associer l'obésité à de nombreuses maladies telles que la dépression, l'apnée du sommeil, le diabète ou les maladies coronariennes. (Haslam, 2007).

Aujourd'hui, **l'obésité est devenue la première maladie non infectieuse de l'Histoire**. Cette pathologie frappe aussi bien les pays industrialisés que les pays en voie de développement. L'Organisation Mondiale de la Santé place actuellement **sa prévention et sa prise en charge comme une priorité dans le domaine de la pathologie nutritionnelle**. (Obésité : prévention et prise en charge de l'épidémie mondiale Rapport d'une consultation de l'OMS 2003).

Les derniers chiffres de l'Organisation Mondiale de la Santé (2008) comptent dans le monde **400 millions d'obèses et 1,6 milliard de personnes en surpoids**. L'augmentation rapide de cette maladie alerte les scientifiques. En effet, en 20 ans, le nombre d'individus obèses a doublé dans le monde. (Figure I. 1).

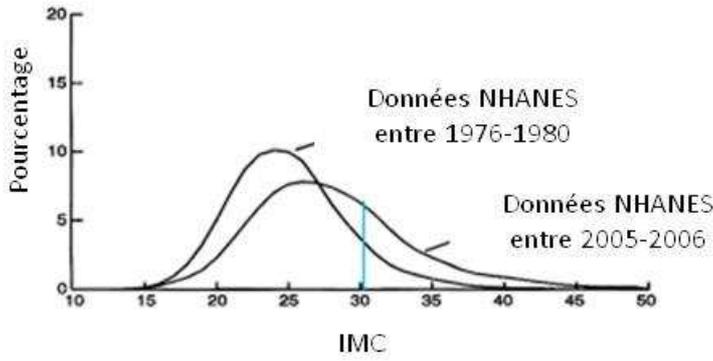


Figure. I.1 Répartition de l'indice de masse corporelle (IMC) des individus dans le monde, âge et sexe confondus. (NHANES National Health and Nutrition Examination Survey. (Haslam and James, 2005))

Les données de l'Organisation de Coopération et de Développement Economiques (OCDE), montrent que cette pathologie se développe surtout dans les pays anglo-saxons ou proches des Etats-Unis. (Figure I. 2). Aux Etats-Unis, 50% des américains sont en surpoids et 30% sont obèses. L'Europe compte près de 30% d'adultes en surpoids et 21% de personnes obèses.

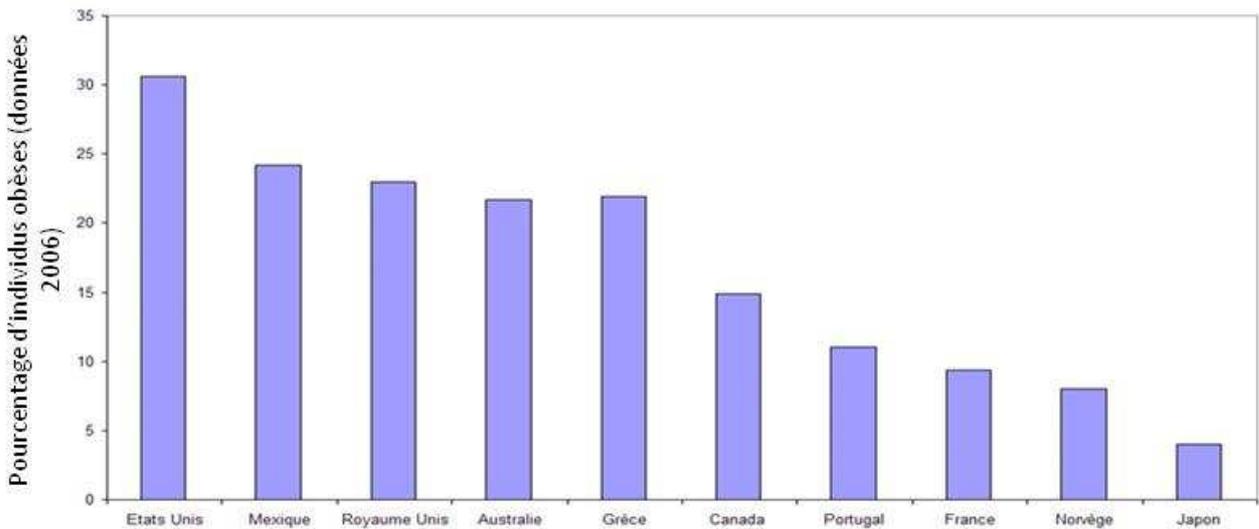


Figure. II.2 _ Répartition du pourcentage du nombre de personnes obèses en fonction des pays. (Economic Co-operation and Development (OCDE) Health Statistique 2000)

D'après les derniers chiffres de l'enquête épidémiologique ObEpi, 31,9% des français sont en surpoids et 14,5% sont obèses. La prévalence globale de l'obésité est significativement différente entre hommes (13,9%) et femmes (15,1%, $p < 0.01$). En 2009, de vastes disparités interrégionales demeurent : la prévalence de l'obésité varie du Sud au Nord

de 11,5% dans la région PACA à 20,5% dans le Nord-Pas de Calais et d'Est en Ouest: de 17,8% en Alsace à 12,2% en Bretagne.

Des études épidémiologiques ont montré l'impact significatif de l'obésité dans le diagnostic de maladies non transmissibles telles que le diabète, l'hypertension, l'hypercholestérolémie, les pathologies cardio-vasculaires, les cancers. La Figure I. 3 répertorie les maladies diagnostiquées chez les hommes et les femmes obèses en fonction de différentes régions du monde. D'après cette étude, les patients obèses d'Europe centrale et de l'Est développent presque systématiquement des complications (hypertension, diabète). Les conséquences de l'obésité sur la santé sont multiples allant des complications métaboliques jusqu'à une mortalité précoce. L'obésité représente ainsi le cinquième facteur de mortalité dans le monde. La définition de l'obésité est donc liée à une augmentation de la masse du tissu adipeux mais aussi à une augmentation du risque de mortalité ou de maladie.

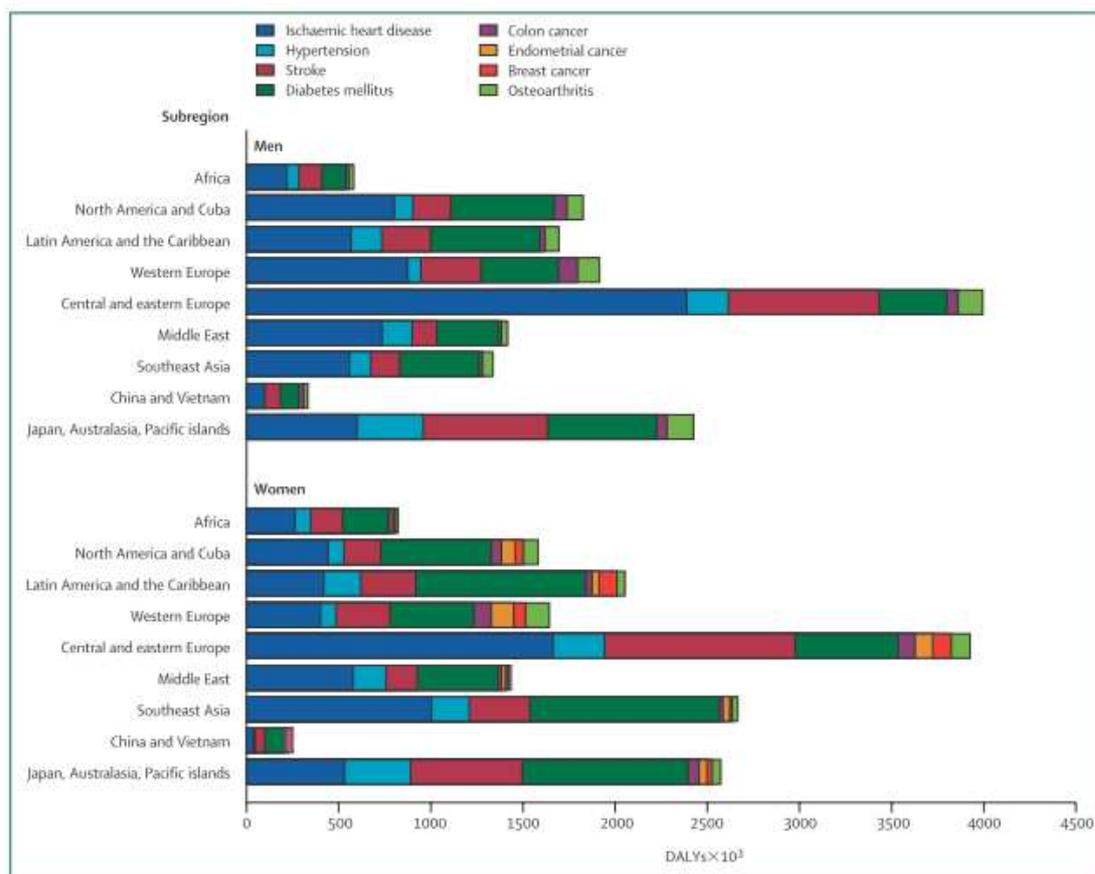


Figure. I.3 _ Répartition des maladies connues chez l'homme et la femme obèse dans les différentes régions du monde. (*James et colleagues*. DALYs (Disability Adjusted Life Years). (Haslam and James, 2005)

2. Développement de l'obésité et ses complications

L'obésité résulte d'une interaction complexe entre des facteurs génétiques, comportementaux et environnementaux. On admet actuellement qu'environ 30 % des cas d'obésité sont de nature génétique. Toutefois, le développement de l'obésité et de ses complications est dû en partie à des modifications métaboliques induites par des changements d'habitudes alimentaires (fast food et plats pré-cuisinés) plus riches en lipides.

Chez l'individu non malade, des mécanismes physiologiques permettent d'équilibrer les apports et les dépenses énergétiques totaux afin de maintenir la stabilité du poids corporel à long terme, c'est l'**homéostasie**. Chez l'individu obèse, les mécanismes de régulation du poids sont déséquilibrés suite à l'accumulation excessive de lipides stockés et libérés par les cellules adipeuses. On distingue deux phases dans le développement de l'obésité : **la phase dynamique** où la prise de poids survient et **la phase statique** où l'excès de poids est entretenu. (Figure I. 4)

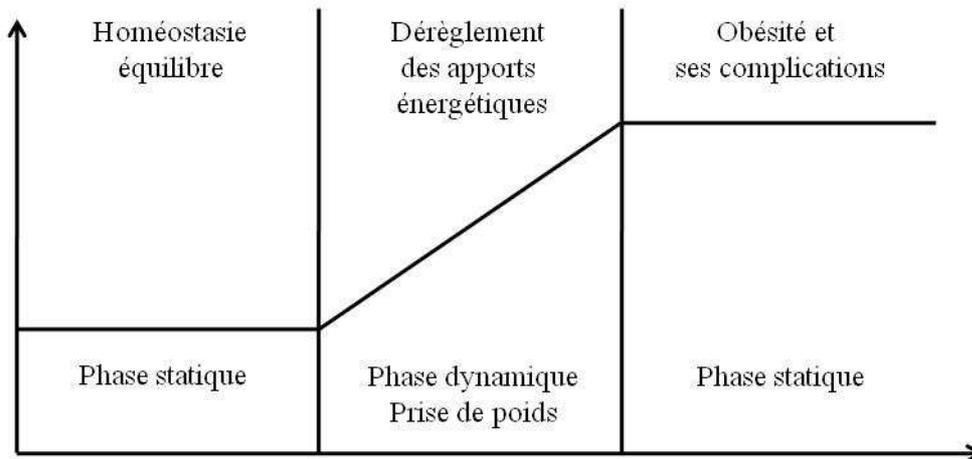


Figure. I.4 _ Représentation des différentes phases expliquant le développement de l'obésité et de ses complications.

Les études menées sur l'obésité ont permis de comprendre l'influence de l'accumulation des lipides sur le développement des pathologies associées à l'obésité telles que le diabète de type II, les maladies cardiovasculaires ou l'hypertension. L'identification de marqueurs de l'inflammation chez les individus obèses a montré que l'obésité était caractérisée par un état inflammatoire. L'apparition de l'état inflammatoire est liée à d'autres dérèglements comme la résistance à l'insuline, le stress oxydant et d'autres maladies comme

l'hypertension. (Fulop *et al.*, 2006). Bien que ces études aient décrit la complexité de cette pathologie, les connaissances scientifiques actuelles ne permettent pas de comprendre les interactions qui sont à l'origine du développement de l'obésité. Par conséquent, le seul moyen de lutter contre cette maladie est la prévention nutritionnelle.

3. Prévention

De nombreux programmes se sont mis en place à travers le monde. Leur objectif est d'informer la population des risques de l'obésité et de formuler des recommandations nutritionnelles permettant de changer les habitudes alimentaires désorganisées. Les stratégies de prévention ciblent les adultes sur leurs lieux de travail ainsi que les enfants dans les structures scolaires en les sensibilisant à l'éducation sportive, l'éducation alimentaire et les formations en technique du comportement. (Dehghan *et al.*, 2005). Malheureusement, **ces recommandations sont peu appliquées et restent générales**. C'est pourquoi, il est urgent de mieux comprendre les **multiples facteurs régissant la prise de poids** afin d'arriver à une prévention plus efficace (Hill *et al.*, 2008). Au vue de la complexité des interactions, l'utilisation d'une approche globale permettrait d'identifier des marqueurs biologiques dynamiques (métabolites) témoins de ces modifications.

4. Marqueurs biologiques de l'obésité

En médecine, les marqueurs biologiques sont des indicateurs d'un état pathologique particulier. Ils furent longtemps des indicateurs physiologiques comme la pression sanguine ou le rythme cardiaque. De nos jours, ce terme est synonyme de marqueur biologique moléculaire.

4.1. Les marqueurs cliniques de l'obésité

Un indicateur universel a été mis en place : **Indice de Masse Corporelle (IMC)**. Celui-ci est calculé en divisant le poids de la personne par le carré de sa taille. (exprimé en kg/m^2). Un seuil a été établi afin d'utiliser simplement cet indicateur.

- **IMC entre 25 et 30**, l'individu est en **surpoids**.
- **IMC supérieur à 30**, l'individu est considéré **obèse**.
- **IMC supérieur à 40**, on parle d'**obésité morbide** avec risque de complications.

L'IMC ne permet pas de distinguer le poids associé à la masse musculaire à celui associé à la masse grasseuse, c'est pourquoi d'autres mesures permettent de déclarer cliniquement un patient obèse :

- ✓ Les mesures anthropométriques sont le poids, la circonférence de tour de taille.
- ✓ La détermination de la répartition anatomique de la masse grasse (sous-cutanée ou abdominale) par IRM: Imagerie par Résonance Magnétique. La localisation de la masse grasse au niveau de l'abdomen est un facteur de risque.
- ✓ La concentration anormale en lipides (triglycérides, cholestérol) et glucose à jeûn.
- ✓ Le développement d'une hypertension artérielle.

4.2 Les marqueurs biologiques moléculaires de l'obésité.

L'étude du développement des pathologies liées à l'obésité est à l'origine de l'identification de marqueurs biologiques chez des patients obèses. Des protéines et des métabolites, marqueurs de l'inflammation, du stress-oxydant ou du métabolisme énergétique, ont été mis en évidence chez ces individus. (Wellen *et al.*, 2006 and Petersen *et al.*, 2006). Ces marqueurs biologiques sont des indicateurs spécifiques de certains états :

- L'inflammation sera caractérisé par une augmentation de la concentration plasmatique d'interleukine-6 (IL-6), « tumour necrosis factor- α » (TNF α) (Rexford *et al.*, 2008) et de lysophosphocholines (LPC) .
- Le stress oxydant sera déterminé par une augmentation de la vitamine E, des isoprostanes (Laight *et al.*, 1999) et une diminution des phosphocholines (PC), fort anti-oxydants (Pietilanien *et al.*, 2007).
- Le déséquilibre énergétique sera décrit par une diminution du ratio ATP/ADP (Serkova *et al.*, 2006) et une augmentation de la leucine, valine et isoleucine (Pietilanien *et al.*, 2008).

Tous ces marqueurs biologiques ont permis de comprendre les voies métaboliques perturbées à un stade d'obésité avancé. (deuxième phase statique où l'excès de poids est entretenu, Figure I. 4). A ce stade, la chirurgie ou le traitement médicamenteux sont

préconisés mais leur efficacité est réduite. Pour lutter efficacement contre cette maladie il est nécessaire de déterminer des marqueurs biologiques précoces témoins du dérèglement des apports énergétiques (phase dynamique de prise de poids).

L'étude des métabolismes suscite un intérêt majeur puisqu'elle reflète l'intégration de l'ensemble des phénomènes post-génomiques. L'émergence de l'outil métabolomique dans le champ de la nutrition offre la possibilité d'aborder les problèmes de biocomplexité en décrivant de façon exhaustive les métabolites générés ou modifiés dans certaines conditions nutritionnelles. Cette approche devrait ainsi permettre la détermination de biomarqueurs précoces de dysfonctionnement métaboliques.

II. Approches métabolomiques

Les avancées des techniques de Spectrométrie de Masse (MS) (Dettmer *et al.*, 2007) et de Résonance Magnétique Nucléaire (RMN) (Lenz *et al.*, 2007; Lindon *et al.*, 2008) ont permis l'analyse de profils métaboliques globaux. Toutefois, ces technologies génèrent des données volumineuses et complexes qui nécessitent l'utilisation d'outils chimiométriques et bioinformatiques (Idle *et al.*, 2007; Jonsson *et al.*, 2004; Moco *et al.*, 2007).

Une analyse métabolomique peut être décomposée en trois volets principaux : le traitement de l'échantillon, le traitement des données et l'étape d'identification. (Figure I. 5)

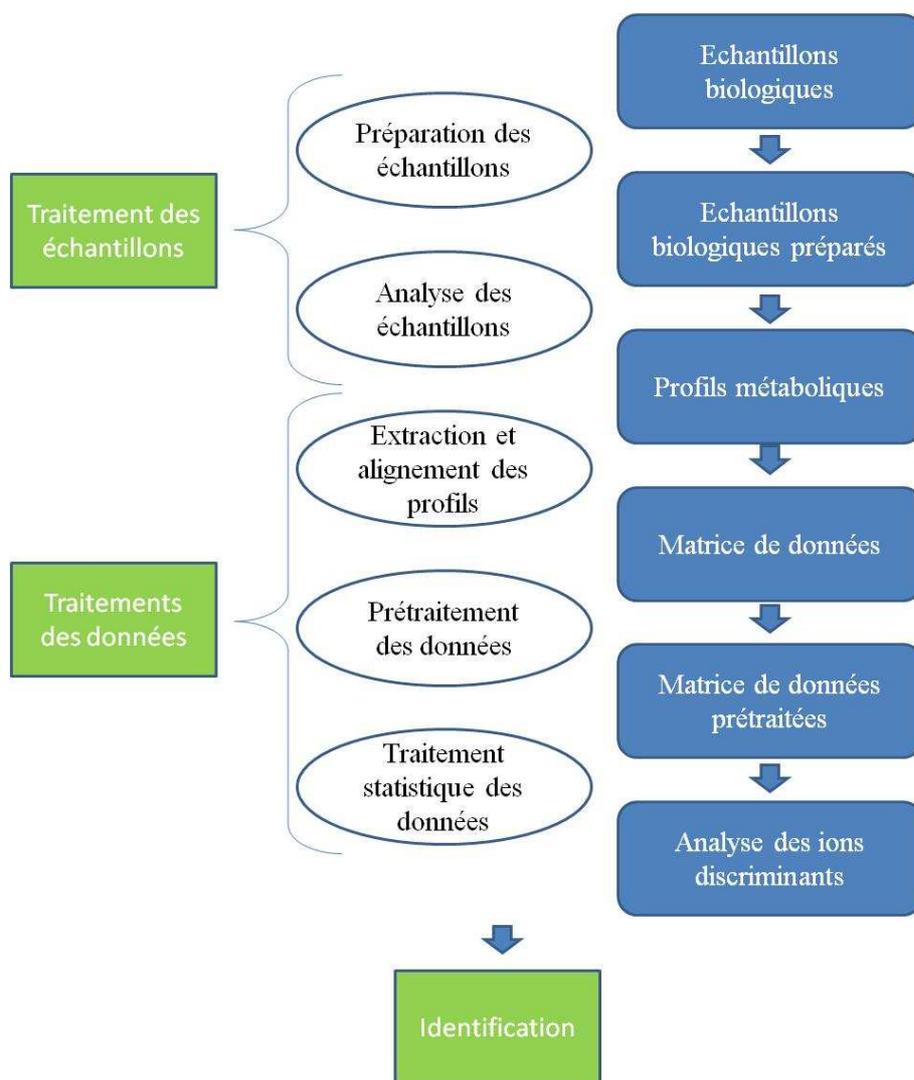


Figure. I.5 _ Schématisation des étapes de la métabolomique.

Deux approches métabolomiques ont été développées en parallèle : l'une dite « ciblée » : s'intéressant à une voie métabolique particulière et l'autre dite « non ciblée » : définissant une empreinte caractéristique de l'ensemble des métabolites.

1. L'approche ciblée

L'approche ciblée utilisée depuis de nombreuses années, s'intéresse à des familles particulières de métabolites, telles que les glucides ou les lipides par exemple. Des méthodes **analytiques robustes et sensibles** ont été développées pour caractériser et quantifier chaque famille de métabolites par chromatographie liquide (LC en anglais *Liquid Chromatography*), gazeuse (GC en anglais *Gaz Chromatography*) ou électrophorèse capillaire (CE en anglais *Capillary Electrophoresis*) couplées à la spectrométrie de masse (Houjou *et al.*, 2005, Watson *et al.*, 2006). Les données métabolomiques issues de différents instruments (Buscher *et al.*, 2009) sont intégrées afin de créer des profils caractéristiques des modifications métaboliques liées à une pathologie (le diabète, l'obésité, etc...) ou à une intervention nutritionnelle. L'utilisation de cette approche intégrative nécessite une connaissance au préalable des familles de métabolites recherchées ce qui constitue une limite. Afin de pallier cette contrainte, une stratégie non ciblée a été développée.

2. L'approche non ciblée

La métabolomique (Fiehn *et al.*, 2000, Nicholson *et al.*, 1999) non ciblée est l'analyse, sans *a priori*, de la modification de l'ensemble des petites molécules contenues dans des fluides biologiques (sang, urine, salive,...) induite par un stimulus tel qu'une intervention nutritionnelle, des médicaments ou une perturbation génétique. Cette approche récente s'est révélée très prometteuse notamment pour comprendre le métabolisme des nutriments, leurs effets biologiques et le rôle qu'ils peuvent jouer dans la prévention nutritionnelle de pathologies. L'objectif de cette approche est d'accéder à l'ensemble des métabolites présents dans les échantillons biologiques. Les techniques utilisées génèrent un grand nombre de données pour des temps d'analyse relativement courts. Des outils statistiques et bioinformatiques spécifiques permettent ensuite de comparer les profils métaboliques complets et d'en extraire l'information biologique. Cependant, deux facteurs sont limitants : d'une part la diversité physico-chimique et la large gamme de concentration des métabolites analysés, et d'autre part le volume et la complexité de données obtenues. C'est pourquoi, des

développements, aussi bien pour l'étape de traitement de l'échantillon que pour celle de traitement des données, sont nécessaires.

Toutefois, le choix de l'utilisation d'une de ces approches dépendra essentiellement de la question biologique posée et de la sensibilité requise.

Dans le cadre de notre étude, l'objectif était d'évaluer l'impact d'un régime hyperlipidique sur le métabolisme, afin de déterminer la ou les modifications métaboliques responsables de la prise de poids. C'est pourquoi, une approche non ciblée a été développée et optimisée pour analyser des fluides biologiques (urine et plasma).

En nutrition, les études métaboliques ont permis de déterminer des marqueurs biologiques endogènes expliquant les modifications métaboliques induites. (German *et al.*, 2002; Gibney *et al.*, 2005; Rezy *et al.*, 2006 ; Watkins *et al.*, 2001). **Toutefois, les doses nutritionnelles étudiées engendrent une réponse métabolique plus faible qu'en pharmacologie. Les techniques analytiques sensibles sont donc préférentiellement utilisées pour ces études.** La sensibilité et la résolution du couplage Chromatographie Liquide (LC en anglais *Liquid Chromatography*) avec un spectromètre de masse, font de cette technique un avantage pour l'approche métabolique dans le domaine de la nutrition.

III. Techniques analytiques

Bien que la Résonance Magnétique Nucléaire (RMN) fût longtemps la technique de choix en métabolomique, la sensibilité de la Spectrométrie de Masse (MS en anglais *Mass Spectrometry*) incite de plus en plus d'études à l'utiliser. Contrairement à la RMN, cet instrument peut être utilisé en couplage, ce qui permet d'augmenter le nombre d'ions détectés. De nombreux systèmes de séparation sont utilisés en métabolomique : la chromatographie gazeuse (GC) (Fiehn *et al.*, 2000), la chromatographie liquide (LC) et l'électrophorèse capillaire (CE) (Soga *et al.*, 2003). Le choix du couplage utilisé dépendra des composés à étudier. Par exemple, la GC permet d'analyser essentiellement des composés de faibles masses moléculaires et apolaires (volatils) alors que la LC/MS est une technique complémentaire qui couplée à l'*electrospray* permet l'ionisation de molécules polaires et de masses moléculaires comprises entre 10^1 et 10^6 . (Figure I. 6)

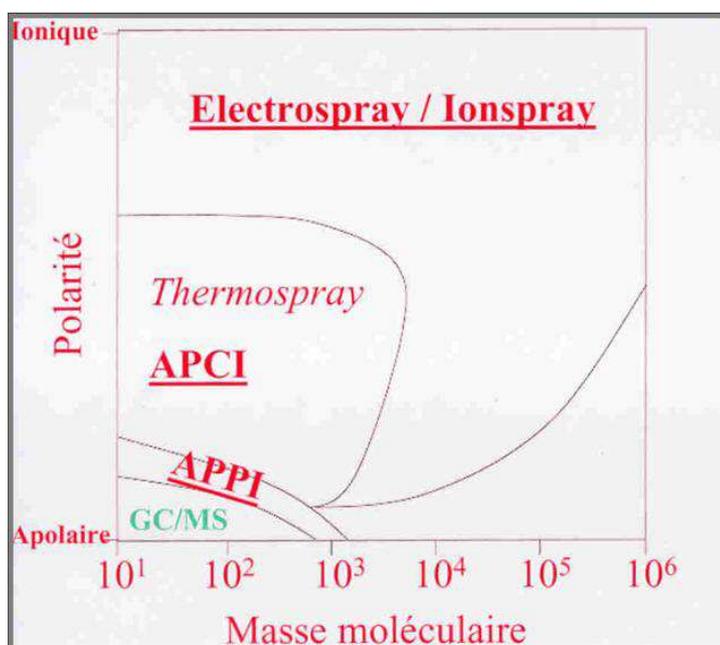


Figure. I.6 _ Choix de la technique de couplage en fonction de la nature des composés étudiés (polarité et masse moléculaire).

De plus, le **développement de la chromatographie en phase inverse a permis d'améliorer la séparation des composés polaires et d'augmenter le nombre de métabolites détectés dans les fluides biologiques** (Idborg *et al.*, 2005).

Par conséquent, le couplage entre la chromatographie liquide (phase inverse) (Issaq *et al.*, 2008) et la spectrométrie de masse, joue un rôle essentiel dans le développement de la métabolomique non ciblée.

1. Système de Chromatographie Liquide

1.1 Principe de l'HPLC (en anglais High Performance Liquid Chromatography).

La chromatographie est une technique analytique permettant la séparation de composés en fonction de leurs propriétés physico-chimiques. Cette technique est basée sur les différences d'affinités entre la substance à analyser et deux phases, l'une stationnaire (la colonne), l'autre mobile (l'éluant). L'échantillon traverse la colonne poussé par la phase mobile.

La détection des composés séparés est représentée par **un chromatogramme caractérisé par un ensemble de pics gaussiens**. Deux paramètres permettent de décrire un pic chromatographique : **le temps de rétention (t_r) et la largeur du pic à mi-hauteur ($w_{1/2}$)** (Figure I. 7). De ces grandeurs découlent d'autres paramètres caractéristiques:

- **le facteur de rétention (k)** qui détermine la capacité de la colonne à retenir les composés,

$k = (t_r - t_m) / t_m$ (t_m est le temps mort : il correspond au temps de rétention de substances non retenues par la phase stationnaire)

- **la résolution (R)** qui caractérise la séparation des pics chromatographiques.

$R = 2 * (t_{r2} - t_{r1}) / (I_1 + I_2)$. (I est la largeur du pic à sa base)

- **le nombre de plateaux théoriques (N)** qui caractérise le nombre d'équilibres successifs qu'a subit un composant du mélange dans la colonne.

$N = 5,54 * (t_r - t_m)^2 / (w_{1/2})$.

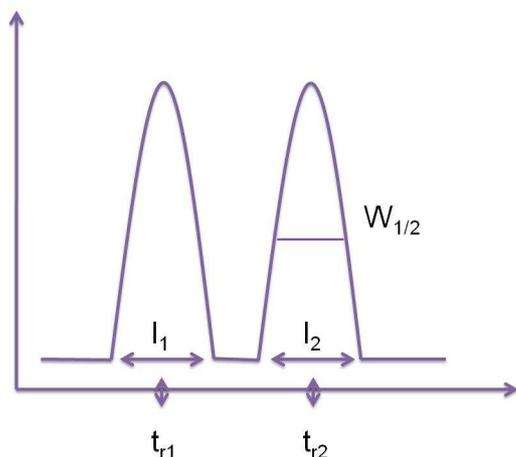


Figure. I.7 _ Représentation des paramètres chromatographiques.

Ces critères permettent de définir l'efficacité d'une colonne. Certains paramètres physiques tels que **la taille des particules de la phase stationnaire** auront une influence sur cette efficacité.

1.2 La phase stationnaire

Plusieurs types de phase stationnaire existent en LC: **les phases à polarité inverse (apolaires et hydrophobes) et les phases à polarité normale (polaires et hydrophiles)**. Chacune de ces phases est utilisée en fonction de la nature des composés à séparer. Dans notre étude, la colonne en phase inverse est plus adaptée aux composés ionisés par *electrospray*.

La phase inverse est composée de particules **de silice (groupe silanols Si-OH) sur lesquelles sont greffées des chaînes alkyles à 8 ou 18 atomes de carbones**. Pour modifier les caractéristiques d'une colonne (résolution), les fournisseurs peuvent modifier le taux de greffage des chaînes carbonées ou greffer d'autres fonctions chimiques (Si-O-CH₃ ou Si-O-C(CH₃)₃) sur les groupements silanols restants. Ainsi chaque colonne en phase inverse a des caractéristiques spécifiques.

1.3 La phase mobile

Deux modes de pompage des phases mobiles existent : le mode isocratique et le mode gradient. (Wiczling *et al.*, 2008). Lorsque la composition de la phase mobile change au cours de l'analyse, on parle du mode "gradient", en opposition au mode "isocratique", où la

composition de la phase mobile reste la même. Dans notre étude, le mode gradient était plus adapté pour la séparation de nombreux composés. Deux solvants sont en général utilisés, l'un aqueux et l'autre organique (méthanol, acétonitrile). Pour améliorer la séparation et l'ionisation des composés, la phase mobile aqueuse peut être tamponnée (sels ou réactifs ioniques ou tampons). (Kaliszan *et al.*, 2004). Toutefois, une forte concentration en sels peut perturber l'ionisation dans la source.

1.4 Particularité de l'UPLC

Cette technique fait appel aux mêmes principes que l'HPLC. Toutefois, les **phases stationnaires composées de particules inférieures à 2 μm permettent d'améliorer la résolution chromatographique, la vitesse d'analyse et la sensibilité.** (de Villiers *et al.*, 2006). Ainsi, en gardant les mêmes performances chromatographiques qu'en HPLC, l'augmentation du débit permettra d'améliorer la résolution ou de diminuer le temps d'analyse.

L'étude de matrices biologiques complexes augmente les phénomènes de co-élution, c'est-à-dire que les composés de propriétés physico-chimiques similaires peuvent éluer au même temps de rétention. Ce phénomène provoque alors des compétitions d'ionisation dans la source du spectromètre de masse. Il a été clairement démontré (Chambers *et al.*, 2007) que l'utilisation de l'UPLC améliorerait la séparation des pics chromatographiques, minimisant ainsi le phénomène de suppression d'ionisation dans la source. Le couplage UPLC-MS est donc optimal pour notre approche.

2. Spectrométrie de masse

2.1 Principe

La spectrométrie de masse permet de transformer des molécules de leur état naturel en ions à l'état gazeux. L'ion est mesuré en fonction de son rapport masse sur charge électrique (m/z), où m est l'unité de masse atomique et z le nombre de charges portées par l'ion. L'unité de mesure de m/z est le Thomson (Th).

Un spectromètre de masse est composé d'**une source d'ions** où se fait l'ionisation, d'**un analyseur** qui trie les ions en fonction de leur rapport m/z , d'**un détecteur** qui comptabilise

l'abondance des ions **et d'un ordinateur** qui permet de traiter et visualiser le signal. (Figure I. 8). Cette technique permet : **d'identifier des composés inconnus**, de **détecter la composition isotopique d'éléments**, de **déterminer la structure d'une molécule à l'aide de ses fragments** et de **quantifier les composés analysés**.

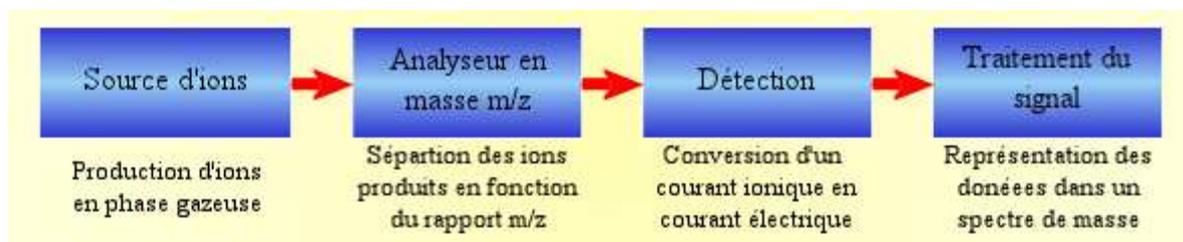


Figure. I.8 _ Schéma des différents éléments d'un spectromètre de masse.

Cet instrument sera caractérisé par certaines performances analytiques tels que :

- ✓ **La résolution** est la capacité d'un analyseur à distinguer entre eux deux signaux voisins. En considérant le pouvoir séparatif, on peut définir la résolution comme le rapport : $R=M/\Delta M$, où M est la masse et ΔM est la différence minimale entre le pic considéré et son plus proche voisin ou la largeur à mi-hauteur du pic étudié. (Figure I. 9).
- ✓ **La précision en masse** est la justesse avec laquelle une masse sera mesurée sur un spectromètre de masse. Cette valeur est définie en ppm.

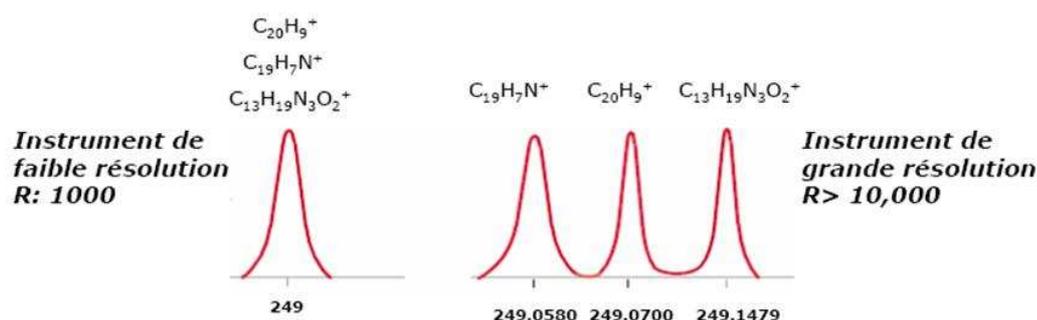


Figure. I.9 _ Avantage d'un instrument de grande résolution sur la séparation des pics mesurés.

2.2 Spectrométrie de masse et métabolomique.

Dans la mesure où la séparation se fait par chromatographie liquide, la source utilisée généralement est l'*electrospray*. Cette combinaison est optimale pour l'analyse des fluides biologiques. Dans le cadre de notre étude, deux analyseurs ont été utilisés : l'un permettant une analyse à haut débit (quadripôle couplé à un temps de vol QToF) et l'autre à haute résolution (LTQ-Orbitrap). La technique « haut débit » permet d'obtenir des profils métaboliques caractéristiques des individus. La comparaison de ces profils par analyses statistiques met en évidence les ions sensibles aux modifications biologiques entre les individus. L'accès à la masse exacte des ions recherchés est réalisé à l'aide d'un spectromètre de masse à haute résolution. Ces deux techniques sont présentées ci-dessous.

2.2.1 Source d'électronébulisation (ou electrospray)

L'*electrospray* (ESI) est une technique qui permet de faire passer des ions d'une phase liquide à une phase gazeuse, en utilisant le principe d'ionisation/désolvatation . Ce principe permet de ne pas dégrader les composés thermosensibles et d'ioniser des molécules de polarités et masses moléculaires (10^1 à 10^3) variées.

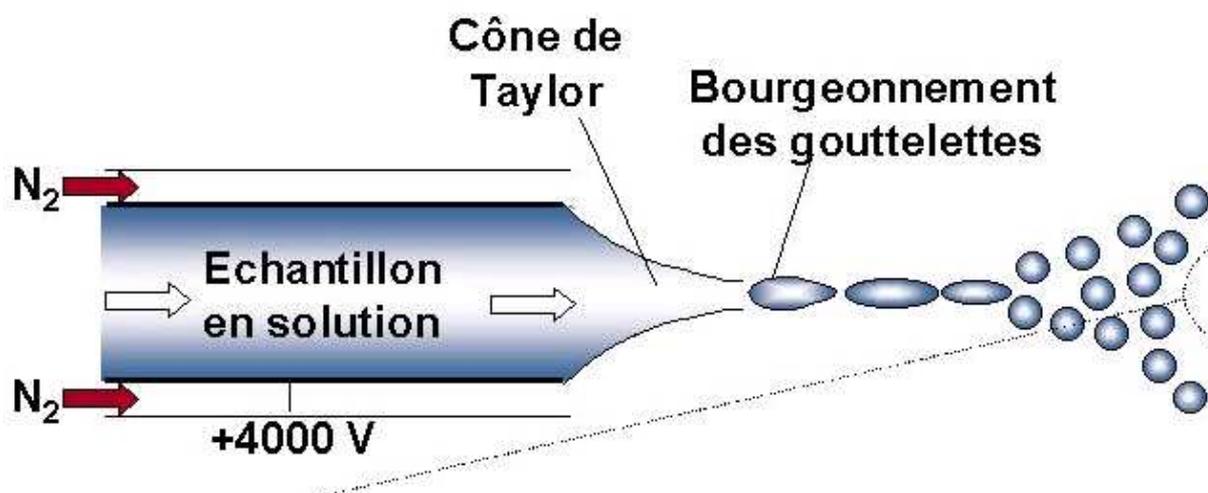


Figure. I.10 _Schéma du principe d'ionisation par l'*electrospray*.

Le processus est simplement décrit en 2 étapes (Figure I. 10):

- ✓ **Dispersion d'un liquide en gouttelettes multichargées.** L'échantillon en solution est introduit dans un capillaire porté à haut potentiel électrique. La détente de la solution

et le champ électrique à la sortie du capillaire provoquent la formation d'un nuage de gouttelettes chargées. De plus, un flux de gaz inerte est ajouté le long du capillaire pour faciliter la formation d'un fin brouillard de gouttelettes chargées, appelé cône de Taylor.

- ✓ **Fissions successives des gouttelettes.** Les gouttelettes vont se rétrécir par évaporation du solvant, assistée par un flux d'azote séchant à contre courant du brouillard, jusqu'à former un ion stable.

Deux paramètres ont une influence sur l'ionisation : **le débit et la concentration des ions.** Lorsque de nombreux ions arrivent simultanément dans le capillaire, le courant sature et ionise qu'une partie des ions en compétition, c'est le phénomène de **suppression d'ionisation**. L'utilisation de techniques séparatives en amont diminue en partie ce phénomène. Les ions formés sous pression atmosphérique sont alors transférés vers l'analyseur où règne un vide poussé.

2.2.2 Analyse haut débit: Quadripôle couplé au Temps de Vol (QTof)

Cet instrument est composé d'une série « hexapôle, quadripôle (Q1), hexapôle (cellule de collision Q2) » en amont de l'analyseur en temps de vol à injection orthogonal des ions muni d'un réflectron. (Figure I. 11).

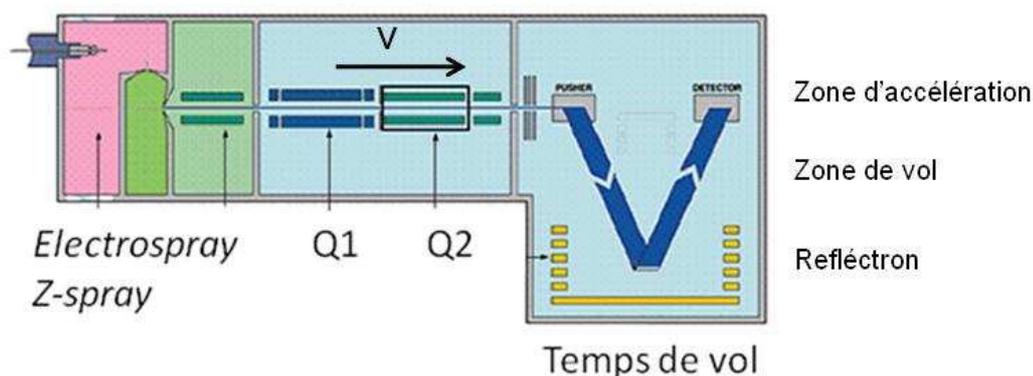


Figure. I.11 _ Schéma Waters de l'electrospray couplé à l'analyseur en tandem quadripôle/temps de vol.

Les ions formés dans la source sont transmis par l'hexapôle vers le quadripôle (Q1). **L'hexapôle et le quadripôle sont des analyseurs, constitués respectivement de six ou**

quatre barres parallèles associées électriquement deux à deux et entre lesquelles on applique un potentiel alternatif ($V \cos\omega t$) et un potentiel continu (U). Cet appareil transmet les ions en faisant varier simultanément ces tensions. Un balayage maintenant un rapport U/V constant permet de **séparer les ions en fonction de leurs rapports m/z.** En **mode MS (scanning)**, le quadripôle Q1 et l'hexapôle Q2 fonctionnent comme des guides d'ions. Les ions focalisés atteignent l'analyseur à temps de vol avec une distribution cinétique proche du niveau thermique. Cette étape permet une meilleure sensibilité et résolution de l'instrument. **En mode MS/MS**, le quadripôle Q1 devient un filtre d'ions capable de sélectionner les ions précurseurs. Ces ions sont alors accélérés puis fragmentés dans la cellule de collision Q2. Les ions obtenus après fragmentation sont transmis dans le temps de vol.

L'ion sélectionné est transmis dans un analyseur à temps de vol (TOF). Cet analyseur se compose d'une zone d'accélération où est appliquée la tension accélératrice, et d'une zone de vol où l'ion se déplace librement. En mode réflectron, un miroir électrostatique impose un champ électrique de direction opposée à celle du champ accélérateur initial. Cette technique permet à tous les ions de même rapport m/z de se focaliser sur un même plan. De plus, le réflectron permet d'allonger la distance de vol sans pour autant augmenter la taille de l'analyseur : les ions mettent plus de temps pour atteindre le détecteur, et réduisent aussi leur dispersion en temps, la résolution s'en trouve donc grandement améliorée.

Dans le cas d'un analyseur à temps de vol, le détecteur utilisé est le « **détecteur à microcanaux** ». (*MicroChannel Plates MCP*). La multiplication d'électrons est assurée en enduisant chaque canal d'un matériel semi-conducteur émettant des électrons secondaires. L'effet d'avalanche permet de multiplier le nombre d'électrons par 10^5 .

Avantages et limites du LC-QToF pour une utilisation en métabolomique.

L'avantage de cet appareil est de pouvoir analyser des profils métaboliques globaux. Pour cette technique, deux modes sont utilisés : le mode MS pour l'obtention des profils métaboliques et le mode MS/MS pour l'obtention de fragmentations. Les bonnes performances en termes de sensibilité et de résolution d'un temps de vol à accélération orthogonale et la possibilité de fragmenter les composés font de l'analyseur QToF un outil avantageux pour la métabolomique. La principale limitation de cet appareil est liée la faiblesse de la gamme de masse intrinsèque au mode de transmission des ions dans le

quadripôle qui diminue le gain de sensibilité du ToF. L'inconvénient du temps de vol est sa sensibilité à la température. Cet analyseur est constitué d'un matériau dont la longueur est fonction de la température. Cette variation induit un décalage des masses mesurées. Pour pallier ce problème, un étalon externe de leucine enképhaline est injecté en parallèle de l'analyse par un système de lockspray à double source. Le décalage en masse sera recalculé dans l'étape de traitement des données.

La combinaison LC-ESI permet d'analyser de nombreux composés, bien que cette technique possède une variabilité intrinsèque importante. Des contrôles qualités (QC) composés de standards injectés à des intervalles réguliers permettent de visualiser les variations de temps de rétention, masses et intensités. L'utilisation de l'UPLC améliore nettement la séparation et la reproductibilité des temps de rétention.

Les nombreux signaux obtenus par LC-QToF doivent être analysés par des techniques de traitement de données, qui sont discutées dans le chapitre III.

IV. Traitement des données

En métabolomique, le traitement de données est constitué de plusieurs étapes qui permettent de passer des profils métaboliques acquis par le spectromètre de masse (signal) à la sélection des ions significatifs de l'effet biologique étudié. L'acquisition des profils métaboliques par spectrométrie de masse se fait en mode continuum, c'est-à-dire que le signal ne subit aucun traitement (signal brut).

1. Traitement du signal

Les spectres de masse détectés en mode continuum doivent être centroïdés. Pour améliorer cette étape, deux traitements peuvent être réalisés : l'élimination du bruit de fond et le lissage du spectre de masse. (Figure I. 12 et 13). Comme nous pouvons le voir dans la Figure I. 12, une mauvaise soustraction du bruit de fond, augmente le nombre de pics à traiter.

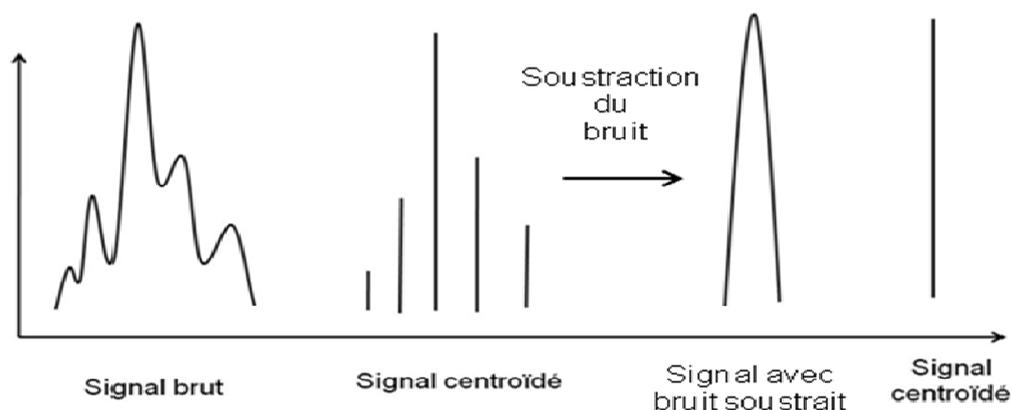


Figure. I.12 _ Modification du signal brut après soustraction du bruit de fond.

Dans notre étude, l'étape de lissage n'a pas été réalisée : en effet, lorsque plusieurs ions sont mal séparés, cette étape aura tendance à les combiner, induisant une erreur sur la masse mesurée. (Figure I. 13).

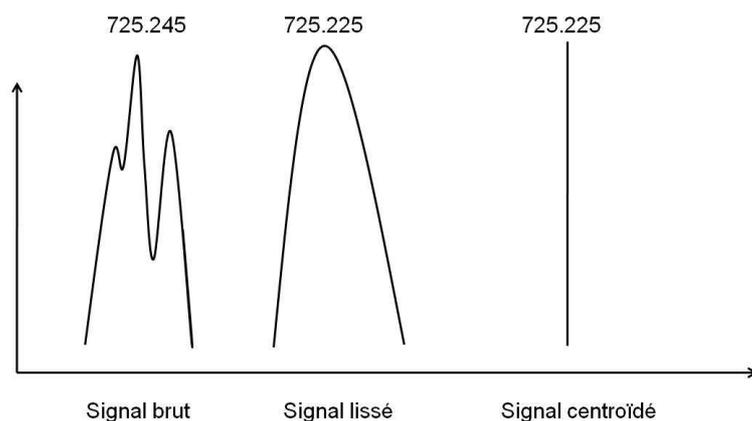


Figure. I.13 _ Modification du signal brut par le lissage.

Le signal centrôidé est ensuite corrigé en masse grâce à la valeur de masse de l'étalon externe, de Leucine enképhaline, injectée en parallèle dans la source par la technique du lock spray. La différence entre la masse mesurée et la masse théorique de leucine Enképhaline permet de recalculer toutes les masses détectées dans l'échantillon biologique à un temps donné.

2. Extraction et alignement des ions.

La deuxième étape consiste à transformer les profils métaboliques traités en une matrice de données exploitables par les outils d'analyses statistiques. Pour ce faire, des logiciels vont extraire dans chaque profil métabolique des ions et les aligner. (Figure I. 14). La matrice de données est composée en colonnes des mesures réalisées, et en lignes des échantillons sur lesquels ces variables ont été mesurées.

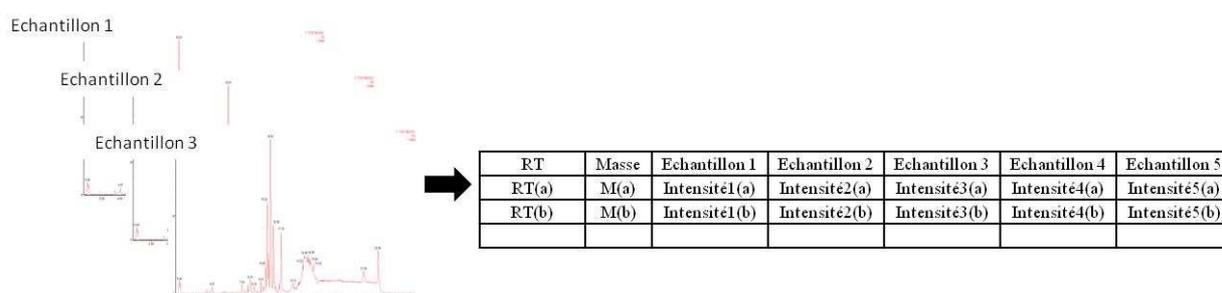


Figure. I.14_ Passage des profils métaboliques à une matrice de données : étape d'extraction et d'alignement des données.

De nombreux logiciels d'extraction et d'alignement des données LC-MS existent : MarkerLynxTM (Waters), XCMS (Smith *et al.*, 2006), Metalign (Tukunov *et al.*, 2005) et Mzmine (Katajamaa *et al.*, 2005).

Logiciels	Avantages	Inconvénients
MarkerLynx™	<ul style="list-style-type: none"> - Utilisation de fichier .raw. - Paramètres explicites. - Visualisation claire des données. - Possibilités d'aligner des centaines d'échantillons. 	<ul style="list-style-type: none"> - Payant. - L'utilisateur ne contrôle aucune étape. - Ajout d'une valeur nulle dans la matrice pour un ion non trouvé.
XCMS	<ul style="list-style-type: none"> - Gratuit - Chaque étape est contrôlée par l'utilisateur. - Correction du temps de rétention. - Possibilité d'aligner des centaines d'échantillons - Calcul statistique pour chaque ion et visualisation des données. - Ajout d'une valeur minimale (#0) dans la matrice pour un ion non trouvé. 	<ul style="list-style-type: none"> - Utilisation de fichiers NetCDF. - Paramètres complexes. - Interface non conviviale, codage sous R.
Mzmine	<ul style="list-style-type: none"> - Gratuit - Chaque étape est contrôlée par l'utilisateur. - Interface conviviale, paramètres explicites. - Possibilités de pré-traiter les données. - Ajout d'une valeur minimale (#0) dans la matrice pour un ion non trouvé. - Nombreuses méthodes de visualisation des données. 	<ul style="list-style-type: none"> - Utilisation de fichiers NetCDF. - Difficulté à aligner de nombreux échantillons.
Metalign	<ul style="list-style-type: none"> - Gratuit - Calcul local du bruit de fond - Correction de la ligne de base - Correction de la saturation - Normalisation des données 	<ul style="list-style-type: none"> - Utilisation de fichiers NetCDF. - Manque de convivialité. Paramètres complexes. - Difficulté à aligner des centaines d'échantillons.

Tableau. I.1 _ Avantages et Inconvénients des logiciels d'extraction et d'alignement testés.

En fonction des données étudiées, les paramètres et les méthodes les plus adaptés doivent être déterminés. Les avantages et les inconvénients de chacun de ces logiciels sont traités ci-dessus. En raison du nombre important des échantillons que nous devons analyser, **XCMS et MarkerLynx ont été utilisés et comparés dans le chapitre III de ce manuscrit.**

3. Prétraitement statistique

En métabolomique, à chaque étape de l'analyse, des sources de variations peuvent s'introduire et fausser l'interprétation biologique. Il est donc nécessaire d'avoir des critères d'évaluation pour chaque étape. L'objectif du prétraitement de données est de minimiser ces variations afin de pouvoir extraire l'information biologique. (Chapitre III).

4. Analyses statistiques

L'objectif des différentes analyses statistiques est de synthétiser et structurer l'information contenue dans les données mesurées par spectrométrie de masse. En métabolomique nutritionnelle, cette étape permet de sélectionner les ions pour lesquels on observe un effet significatif des facteurs nutritionnels étudiés. Les ions pertinents sélectionnés par les techniques univariées comme l'analyse de la variance, permettent de définir un modèle de discrimination par des techniques multivariées d'analyses discriminantes.

En métabolomique, la mesure de l'intensité des ions extraits des échantillons biologiques sont des variables aléatoires quantitatives et continues. Ces variables suivent en général une distribution normale ou log-normale. Certains paramètres de valeur centrale (médiane, moyenne) et de dispersion (variance) vont être définis et calculés afin de synthétiser l'information. En raison du grand nombre de données obtenues, plus de 1000 ions en général, la statistique multidimensionnelle permettra de représenter des informations biochimiques sous-jacentes en les simplifiant par des représentations graphiques. Trois types de techniques statistiques ont été appliquées sur ces données :

- ✓ Analyses univariées : Ces techniques traitent chaque ion indépendamment les uns des autres. Le principe consiste à décomposer la variabilité totale en variabilités dues aux facteurs étudiés (régime, phenotype, cinétique) et en

variabilité résiduelle. Un test de Fisher permet de tester la significativité des différentes sources de variabilité par rapport à la variabilité résiduelle. On obtient grâce à ces techniques une liste exhaustive d'ions sur lesquels les facteurs biologiques étudiés ont un effet significatif. Un ion sera significatif pour un facteur de variation si la valeur de probabilité calculée est inférieure à 0,05. Suivant le protocole biologique, on peut utiliser un test de Student, l'analyse de la variance (ANOVA) ou de la covariance (ANCOVA).

- ✓ Analyses multivariées non supervisées : Ces techniques prennent en compte l'ensemble des ions détectés. Elles ne nécessitent pas de connaissance préalable sur l'appartenance des individus à un groupe (phénotype) et permet de décrire visuellement les principales sources de variabilité dans les données. L'Analyse en Composantes Principales (ACP) permet de définir un sous espace de dimension inférieure au nombre d'ions. Dans ce nouvel espace défini par les composantes principales, on représente les ions et les individus ce qui permet de visualiser les relations entre ions et une éventuelle structuration dans les individus. La classification ascendante hiérarchique (CAH) est un autre technique multivariée basée elle sur une matrice de distance qui permet une représentation sous forme d'arbre montrant les similarités et dissimilarités entre individus (non utilisée dans notre étude).
- ✓ Analyses multivariées supervisées : Ces techniques permettent d'expliquer et de prédire l'impact d'une intervention nutritionnelle sur des groupes d'individus. Comme pour l'ACP ou la CAH, ces techniques utilisent l'ensemble des ions mesurés pour établir un modèle qui prédit l'appartenance à des groupes pré-définis correspondant aux facteurs étudiés (régimes, phénotypes ou cinétique). Du fait du nombre de variables (des milliers d'ions) largement supérieur au nombre d'individus (des dizaines à des centaines d'échantillons) nous avons utilisé la partial least square discriminant analysis (PLS-DA).

V. Identification des ions sélectionnés

L'identification d'un métabolite nécessite souvent plusieurs outils analytiques complémentaires. Après analyse des échantillons biologiques par LC-ESI-QToF, plusieurs centaines à plusieurs milliers de signaux sont révélées par les analyses statistiques. Ces signaux peuvent être des ions moléculaires, des fragments, des adduits ou des isotopes. Des techniques statistiques de corrélation permettent de supprimer les redondances d'information. Une fois que les ions moléculaires sont isolés, les échantillons sont ré-analysés par des techniques haute résolution telle que l'Orbitrap (Figure I. 15). La masse exacte obtenue est alors recherchée dans les bases de données. Si cette masse correspond à un métabolite connu alors une identification putative est possible mais si ce n'est pas le cas, une étape de caractérisation, permettant de générer une formule brute chimique à partir de la masse mesurée est envisagée.

1. Les bases de données de métabolites

- Human Metabolom Data Base (HMDB) (<http://www.hmdb.ca/>) est une base de données de métabolites présents dans des fluides biologiques humains. Elle contient différents outils pour aider à l'identification : des spectres de masse ou de RMN, la concentration, la localisation, le rôle des métabolites dans l'organisme. (Whishart *et al.*, 2007)
- Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg/) est une base de données biologiques qui contient une liste exhaustive de métabolites présents dans différentes espèces (plantes, rat, homme). Pour chaque espèce des voies métaboliques sont décrites.
- Nature Lipidomics Gateway (www.lipidmaps.org/) est une base de données qui contient des métabolites lipidiques.
- METLIN (http://metlin.scripps.edu/metabo_search.php) regroupe les métabolites contenus dans KEGG, Lipidmaps et HMDB. (Smith *et al.*, 2005)

Afin de valider les propositions obtenues dans les bases, des données spectrales complémentaires (profils isotopiques et spectres de fragmentations MSⁿ) sont nécessaires. L'utilisation de l'outil LTQ-Orbitrap est alors d'un intérêt majeur.

2. Description du LTQ-(en anglais Linear Trap Quadrupole)-Orbitrap

Le **LTQ Orbitrap** est un **instrument hybride constitué de deux analyseurs**: un trappe linéaire et un Orbitrap.

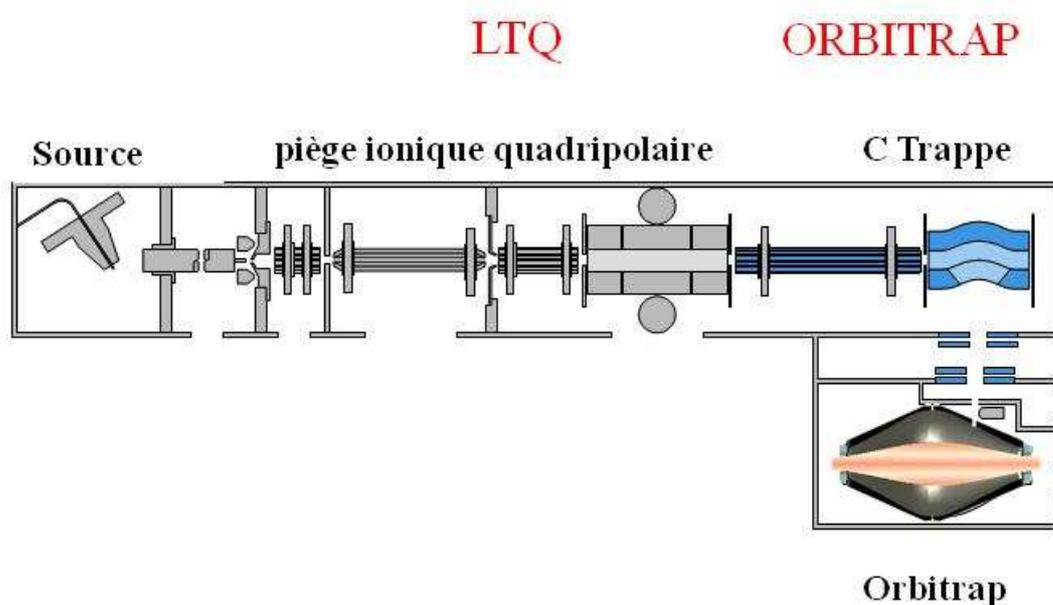


Figure. I.15. _ Schéma Thermo Fischer du piège à ions quadripolaire linéaire et l'Orbitrap.

Chacun de ces deux analyseurs dispose d'un système de détection des ions : deux dynodes et deux multiplicateurs d'électrons pour la trappe linéaire, détection par courant image puis **transformée de Fourier** pour l'Orbitrap. Lors d'une analyse, la trappe linéaire et l'Orbitrap peuvent être utilisés indépendamment l'un de l'autre ou simultanément : c'est la détection parallèle.

L'**Orbitrap** est un analyseur composé d'une électrode axiale placée au centre d'une électrode circulaire creuse. De par sa forme, l'électrode axiale génère un potentiel

électrostatique qui piège les ions. Ces ions sont alors animés d'un mouvement circulaire autour de l'électrode axiale et d'un mouvement oscillatoire perpendiculaire à leur mouvement de rotation induisant une **fréquence** ($f = \frac{1}{2\pi} (km/z)^{1/2}$ k est la courbure du champ). Le courant généré par ces oscillations et recueilli par deux électrodes est converti en un spectre de fréquence par une transformée de Fourier, puis en un spectre de masse.

La trappe linéaire du LTQ Orbitrap est l'analyseur idéal pour **générer des données MSⁿ**. Trois modes d'activation MS/MS et MSⁿ peuvent être utilisés:

- ✓ **mode CID** (en anglais *Collision Induced Dissociation*): dissociation induite par collision. L'ion précurseur est isolé puis fragmenté par augmentation de son énergie interne et collision avec des molécules gazeuses. Les fragments générés seront analysés dans l'Orbitrap.
- ✓ **mode PQD** (en anglais *Pulse Q Dissociation*): les ions sont activés en utilisant des valeurs de piégeage élevées.
- ✓ **mode HCD** (en anglais *High-energy Collisional Dissociation*): l'activation est non résonante et sans cut-off. La fragmentation est induite par CID dans la C-Trappe ou dans l'octapôle adjacent.

Les modes d'activation MS/MS permettent de cumuler des informations complémentaires dans le cas d'interprétations structurales difficiles. Dans notre étude nous n'avons utilisé que le mode CID.

Avantages et limites du LTQ-Orbitrap pour une utilisation en métabolomique

La haute résolution et la précision en masse de cet instrument permet d'identifier sans ambiguïté la présence de certains éléments (soufre, oxygène, azote) par lecture directe des signaux isotopiques spécifiques. En calibration externe, la précision de la mesure de masse est de 2ppm en moyenne alors qu'en calibration interne (Olsen *et al.*, 2005) les précisions sont systématiquement inférieures au ppm. Cette précision de la mesure de masse permet d'améliorer considérablement la qualité des recherches dans les bases de données ou la caractérisation.

La trappe ionique ayant un volume fini, une accumulation importante des ions peut provoquer le phénomène dit de *charge d'espace* qui a pour effet de « décalibrer » le piège. Un trop grand nombre d'ions crée un champ hexapolaire qui se superpose au champ quadripolaire du piège. Des systèmes comme l'AGC (automatic gain control) permettent d'éviter ce phénomène en réalisant une mesure avant l'analyse proprement dite.

VI. Validation des biomarqueurs

Une approche métabolomique nécessite plusieurs types de validation :

- 1- la validation des identifications réalisées : celle-ci est effectuée par l'analyse de composés standards purs.
- 2- la validation des biomarqueurs : afin de s'assurer de la spécificité et de la robustesse des biomarqueurs, des études sont généralement menées sur de nombreux sujets. Cette étape nécessite donc le développement de méthodes d'analyses ciblées quantitatives.

Références

Buscher, J. M., Czernik, D., Ewald, J. C., Sauer, U. and Zamboni, N. (2009), 'Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism.' *Analytical Chemistry* **81**, 2135-2143.

Chambers, E., Wagrowski-Diehl, D. M., Lu, Z. and Mazzeo, J. R. (2007), 'Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses.' *Journal of Chromatography B* **852**, 22–34.

de Villiers, A., Lestremau, F., Szucs, R., Gélébart, S., David, F. and Sandra, P. (2006), 'Evaluation of ultra performance liquid chromatography: Part I. Possibilities and limitations.' *Journal of Chromatography A* **1127**, 60-69.

Dehghan, M., Akhtar-Danesh, N. and Merchant, A. (2005), 'Childhood obesity, prevalence and prevention.' *Nutrition Journal* **4**, 24.

Dettmer, A., Hammock. (2007), 'Mass spectrometry Based Metabolomics.' *Mass spectrometry Reviews* **26**, 51-78.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N. and Willmitzer, L. (2000), 'Metabolite profiling for plant functional genomics.' *Nature* **18**, 1157-1161.

Fulop, T., Tessier, D. and Carpentier, A. (2006), 'The metabolic syndrome.' *Pathologie Biologie* **54**, 375-386.

Gibney, M. J., Walsh, M., Brennan, L., Roche, H. M., German, J. B. and van Ommen, B. (2005), 'Metabolomics in human nutrition : opportunities and challenges.' *American Journal of Clinical Nutrition* **82**, 497-503.

German, J. B., Roberts, M. A., Fay, L. and Watkins, S. M. (2002), 'Metabolomics and Individual Metabolic Assessment: The Next Great Challenge for Nutrition.' *J. Nutr.* **132**, 2486-2487.

Haslam, D. (2007), 'Obesity a medical history.' *Obesity Reviews* **8**, 31-36.

Haslam, D. W. and James, W. P. T. (2005), 'Obesity.' *The Lancet* **366**, 1197-1209.

Hill, J. O. (2009), 'Can a small-changes approach help address the obesity epidemic? A report of the Joint Task Force of the American Society for Nutrition, Institute of Food Technologists, and International Food Information Council.' *Am J Clin Nutr* **89**, 477-484.

Houjou, T., Kotoko, Y., Masayoshi, I., Takao, S. and Ryo, T. (2005), 'A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry.' *Rapid Communications in Mass Spectrometry* **19**, 654-666.

Idborg, H., Zamani, L., Edlund, P.-O., Schuppe-Koistinen, I. and Jacobsson, S. P. (2005), 'Metabolic fingerprinting of rat urine by LC/MS: Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry.' *Journal of Chromatography B* **828**, 9-13.

Idle, J. R. and Gonzalez, F. J. (2007), 'Metabolomics.' *Cell Metabolism* **6**, 348-351.

Issaq, H., Abbott, E. and Veenstra, T. (2008), 'Utility of separation science in metabolomic studies.' *Journal of separation science* **31**, 1936-1947.

Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M. and Moritz, T. (2004), 'A Strategy for Identifying Differences in Large Series of Metabolomic Samples Analyzed by GC/MS.' *Analytical Chemistry* **76**, 1738-1745.

Kaliszan, R., Wiczling, P. and Markuszewski, M. J. (2004), 'pH gradient high-performance liquid chromatography: theory and applications.' *Journal of Chromatography A* **1060**, 165-175.

Katajamaa, M. and Orešič, M. (2005), 'Processing methods for differential analysis of LC/MS profile data.' *BMC Bioinformatics* **6**, 179.

Laight, D. W., Desai, K. M., Gopaul, N. K., Änggård, E. E. and Carrier, M. J. (1999), 'F2-isoprostane evidence of oxidant stress in the insulin resistant, obese Zucker rat: effects of vitamin E.' *European Journal of Pharmacology* **377**, 89-92.

Lenz, W. (2007), 'Analytical Strategies in Metabonomics.' *Journal of Proteome Research* **6**, 443-458.

- Lindon, J. C. and Nicholson, J. K. (2008), 'Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery.' *TrAC Trends in Analytical Chemistry* **27**, 194-204.
- Moco, S., Bino, R. J., Vos, R. C. H. D. and Vervoort, J. (2007), 'Metabolomics technologies and metabolite identification.' *TrAC Trends in Analytical Chemistry* **26**, 855-866.
- Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. and Mann, M. (2005), 'Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap.' *Mol Cell Proteomics* **4**, 2010-2021.
- Petersen, K. F. and Shulman, G. I. (2006), 'Etiology of Insulin Resistance.' *The American Journal of Medicine* **119**, S10-S16.
- Pietilainen, K. H., Naukkarinen, J., Rissanen, A., Saharinen, J., Ellonen, P., Keranen, H., Suomalainen, A., Gotz, A., Suortti, T., Yki-Jarvinen, H., Oresic, M., Kaprio, J. and Peltonen, L. (2008), 'Global Transcript Profiles of Fat in Monozygotic Twins Discordant for BMI: Pathways behind Acquired Obesity.' *PLoS MEDICINE* **5**, 472-483.
- Pietilainen, K. H., Sysi-Aho, M., Rissanen, A., Seppanen-Laakso, T., Yki-Jarvinen, H., Kaprio, J., Oresic, M. and Waterfield, C. J. (2007), 'Acquired Obesity Is Associated with Changes in the Serum Lipidomic Profile Independent of Genetic Effects –A Monozygotic Twin Study.' *PLoS ONE* **5**, 472-483.
- Rexford, S. A. and Mitchell, A. L. (2008), 'Adipokines and the Peripheral and Neural Control of Energy Balance.' *Molecular Endocrinology* **22**, 1023–1031.
- Rezzi, S., Ramadan, Z., Fay, L. B. and Kochlar, S. (2006), 'Nutritional Metabonomics: Applications and Perspectives.' *Journal of Proteome Research* **Reviews** **6**, 513-525.
- Serkova, N., Jackman, M., Brown, J., Liu, T., Hirose, R., Roberts, J., Maher, J. and Niemann, C. (2006), 'Metabolic profiling of livers and blood from obese Zucker rats.' *J Hepatol* **44**, 956 - 962.

- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006), 'XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification.' *Anal. Chem.* **78**, 779-787.
- Smith, C., O'Maille, G., Want, E., Qin, C., Trauger, S., Brandon, T., Custodio, D., Abagyan, R. and Siuzdak, G. (2005), 'METLIN: a metabolite mass spectral database.' *Ther Drug Monit* **27**, 747 - 751.
- Soga, T., Baran, R., Suematsu, M., Ueno, Y., Ikeda, S., Sakurakawa, T., Kakazu, Y., Ishikawa, T., Robert, M., Nishioka, T. and Tomita, M. (2006), 'Differential Metabolomics Reveals Ophthalmic Acid as an Oxidative Stress Biomarker Indicating Hepatic Glutathione Consumption.' *The journal of Biological Chemistry* **281**, 16768–16776.
- Tikunov, Y., Lommen, A., Vos, C. H. R. d., Verhoeven, H. A., Bino, R. J., Hall, R. D. and Bovy, A. G. (2005), 'A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles.' *Plant Physiology* **139**, 1125–1137.
- Watkins, S. M., Hammock, B. D., Newman, J. W. and German, J. B. (2001), 'Individual metabolism should guide agriculture toward foods for improved health and nutrition.' *Am J Clin Nutr* **74**, 283-286.
- Watson, A. D. (2006), 'Lipidomics: A Global Approach to Lipid Analysis in Biological Systems.' *Journal of Lipids Research* **47**, 2101-2111.
- Wellen, K. E. and Hotamisligil, G. S. (2005), 'Inflammation, stress, and diabetes.' *J Clin Invest.* **115**, 1111–1119.
- Wiczling, P. and Kaliszan, R. (2008), 'Influence of pH on Retention in Linear Organic Modifier Gradient RP HPLC.' *Analytical Chemistry* **80**, 7855-7861.
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., MacInnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L.,

Marrie, T., Sykes, B. D., Vogel, H. J. and Querengesser, L. (2007), 'HMDB: the Human Metabolome Database.' *Nucl. Acids Res.* **35**, D521-526.

CHAPITRE II : TRAITEMENT ET ANALYSE DES ECHANTILLONS

Chapitre II

I. Introduction

En métabolomique non ciblée, la détermination de profils métaboliques de fluides biologiques par LC-MS nécessite l'optimisation de **méthodes analytiques robustes**. Ces méthodes doivent permettre **d'extraire et de détecter un maximum de métabolites** présents dans ces fluides, tout en limitant la variabilité analytique induite. Pour cela, il est nécessaire d'optimiser les étapes de préparation d'échantillons, de **séparation, d'ionisation et de détection des métabolites**.

Notre objectif a été de **développer et valider une méthode analytique non ciblée simple et robuste** sur la matrice plasma, dans le cadre d'une étude d'intervention nutritionnelle. En effet, en nutrition, le plasma est une matrice de choix car il est constitué de cellules sanguines qui témoignent des modifications nutritionnelles à travers leurs membranes lipidiques sensibles à l'alimentation de l'individu. (Skeaff *et al.*, 2008 et Gibney *et al.*, 2004).

Le plasma est une matrice complexe, constituée d'un grand nombre de métabolites ayant des propriétés physico-chimiques très diverses (masse, pKa, polarité...) et présents à des concentrations variables (Nicholson and Foxall, 1995). Il contient également d'importantes quantités de protéines, qui doivent être précipitées préalablement à l'analyse par LC/MS. Dans une approche métabolomique globale, la préparation d'échantillons doit être universelle et robuste (Bruce *et al.*, 2009; Want *et al.*, 2006), c'est-à-dire permettre une purification minimum tout en limitant la perte de métabolites lors de cette étape. Des études récentes (Boernsen *et al.*, 2005, Bruce *et al.*, 2009 and Want *et al.*, 2006) ont montré l'efficacité des différentes méthodes de déprotéinisation dans le sérum ou le plasma par addition de solvants organiques, de composés acides ou d'un traitement thermique. L'efficacité de ces méthodes a été évaluée en mesurant la quantité de protéines résiduelles et/ou le nombre de métabolites.

Cependant, l'utilisation de telles approches non ciblées en nutrition souffre aujourd'hui d'un manque d'identification de biomarqueurs spécifiques, notamment à cause du faible nombre de métabolites présents dans les bases de données de métabolites.

Cette étude propose donc **une méthode alternative** qui vise à développer et valider une approche globale tout en s'assurant de sa capacité à détecter des métabolites d'intérêt déjà décrits lors d'études métaboliques sur plasma humain. Pour cela nous avons effectué un suivi de différentes familles de métabolites déjà identifiés chez des sujets obèses, tels que la carnitine, la lysophosphatidylcholine, les acides gras libres, les acides aminés (Pietilainen *et al.*, 2007). Tout au long de cette étude (développement et validation), l'analyse des données a été effectuée à la fois sur les profils métaboliques complets, ainsi qu'à partir des réponses des métabolites d'intérêt sélectionnés.

Nous nous sommes focalisés dans un premier temps sur l'optimisation d'une méthode d'analyse par UPLC-MS. Ce développement a consisté principalement à trouver un compromis entre la solubilisation, la séparation et l'ionisation des métabolites plasmatiques. Dans un second temps, nous avons évalué l'influence relative des différentes étapes de collecte et de préparation de l'échantillon sur les profils plasmatiques obtenus. Cette étude a permis de traiter deux aspects essentiels: le choix de l'anticoagulant et la détermination d'une méthode de précipitation protéique permettant l'extraction d'un maximum de métabolites.

La méthode développée a ensuite été appliquée aux **suivis des métabolites (Tableau II. 1) et des standards**. Ces métabolites ont été sélectionnés pour leurs propriétés physico-chimiques variées, représentatives de l'ensemble des métabolites présents dans une analyse non ciblée. Ils devaient également être présents dans le plasma (Lawton *et al.*, 2008) et être détectés par LC-MS. (Piraud *et al.*, 2003; Houjou *et al.*, 2005; Guo *et al.*, 2006; Liebisch *et al.*, 2002; Liebisch *et al.*, 2004). Une fois ces deux étapes optimisées, 126 échantillons plasmatiques issus d'un protocole nutritionnel ont été analysés afin de valider notre approche métabolomique non ciblée.

	Nom	Masse Monoisotopique (g.mol ⁻¹)	[MH] ⁺	[MH] ⁻	UPLC RT (min)
Acides Aminés	Dimethyl-L-Arginine	202.142	203.150		0.40
	L-Arginine	174.112	175.118		1.08
	Leucine	131.095	132.102		0.45
	Phénylalanine	165.079	166.086		3.44
	Tryptophane	204.089	205.097		4.12
	Tyrosine	181.074	182.082		1.72
Amino ketones	Creatinine	113.058	114.067		0.46
Carnitines	Acetylcarnitine	203.115	204.124		0.98
	Betaine	117.078	118.087		0.43
	Carnitine	161.105	162.112		0.44
	Palmitoylcarnitine	399.334	400.345		11.13
Acides gras	acide valérique	102.068	103.075		3.46
	acide Stearique *	284.271		283.263	16.89
	acide Oleique *	282.255		281.248	14.32
	acide Palmitique *	256.240		255.232	14.13
	acide docosanoïque	340.334		339.326	12.37
	acide Arachidonique	304.240		303.232	13.49
	acide Octadécadiénoïque	280.240		279.232	13.61
Eicosanoïdes	acide Docohexanoïque	328.240	329.249		13.36
	Leukotriene E4	439.239	440.245		8.58
	Leukotriene B4	336.230	337.237		10.23
Phospholipides	LysoPC(16:0)**	495.332	496.340		10.54
	LysoPC(18:2)**	519.332	520.340		10.35
	LysoPC(18:1)**	521.348	522.356		10.88
	LysoPC(18:0)**	523.363	524.371		11.73
	LysoPC(14:0)**	467.303	468.310		9.58
	LysoPC(20:4)**	543.332	544.337		10.35
	PC (36:5)**	779.546	780.555		14.38
	PC (36:4)**	781.562	782.568		14.41
	PC (34:2)**	757.562	758.566		14.27
	LysoPE (18:1)**	479.301	480.299		10.63
	lysoPE (20:0)**	509.348		508.327	11.82
	Lyso PE (20:1)**	507.332		506.308	10.96
	Lyso PE(20:2)**	505.316		504.301	10.27
Porphyrones	Bilirubin	584.263	585.275		7.88
Prostanoides	6ketoprostaglandinF1a	370.235	371.227		6.75
	13.14-dihydroPGF1a	358.271	359.279		6.70
	6-ketoprostaglandinE1	368.219		367.212	7.41
Purine	acide urique	168.028	169.036		1.19
Stéroïdes	Cortisol	362.209	363.219		6.78
	Dehydrocholéstérol	384.339		383.335	15.20
Sphingolipides	CerP(d18:1/12:0)**	561.415	562.423		14.23

Tableau. II.1. Liste des ions sélectionnés dans le plasma. Le temps de rétention a été confirmé par l'analyse des standards des composés. * Ces ions sont des fragments de DG et/ou TG. **LysoPC Lysophosphatidylcholines. PC Phosphatidylcholines. LysoPE Lysophosphatidylethanolamines. CerP : Phosphoceramides

II. Matériels et méthodes

1. Produits chimiques

Le méthanol et l'acétonitrile (HPLC Gradient grade) ont été fournis par Sigma Aldrich, l'eau ultra pure (18,2 M Ω) par Millipore et l'acide formique par Fluka.

Une solution de Leucine Enképhaline (Sigma–Aldrich) à une concentration 0,5 ng/L (dans un mélange acétonitrile/eau, 50/50 v/v avec 0,1% d'acide formique) a été utilisée comme référence pour les mesures de masse. Une solution d'acide orthophosphorique à 0,1% (dans un mélange acétonitrile / eau, 50/50 v/v) a été préparée pour calibrer l'instrument avant chaque nouvelle série d'analyse.

Pour valider la méthode analytique, des solutions standards ont été préparées à une concentration de 10 μ g/mL (dans un mélange acétonitrile/eau, 50/50 v/v avec 0,1% d'acide formique). Ces solutions contiennent les composés suivants : la prednisone (Sigma), l'acide dehydrocholique (Sigma), le tryptophane (Sigma), la leucine (Sigma), la tyrosine (Sigma), la phenylalanine (Sigma), la méthionine (Sigma), la procaine (Fluka), la colchicine (Sigma), la 2-aminoanthracène (96%) (Sigma), la 4-aminoantipyrène (Riedel de Haen).

2. Prélèvement et préparation des échantillons biologiques

Pour tester l'effet des anticoagulants, 18 plasmas humains de volontaires sains ont été prélevés sur EDTA ou sur Héparine et conservés à -80°C. Pour chaque anticoagulant, trois méthodes de précipitation protéique lente ont été testées indépendamment. (Figure II. 1). Ces trois méthodes se distinguent par la nature et la quantité des solvants de précipitation. Ces solvants ont été conservés avant leur utilisation à 4°C.

Méthodes A et B : Après ajout de 2 ou 4 volumes de méthanol dans 200 μ L de plasma, les échantillons ont été conservés pendant 30 minutes à -20°C pour permettre une précipitation lente

Méthode C : Dans 200 μ L de plasma en constante agitation, 30 μ L d'acétonitrile ont été ajoutés toutes les 8 heures pendant 3 jours (méthode de Boernsen (2005)).

Après précipitation protéique lente (A, B et C), le précipité alors formé a été séparé du surnageant par centrifugation à 13000 tours par minute pendant 10 minutes. Enfin, une fois le surnageant repris et évaporé sous azote, l'extrait sec de plasma alors obtenu a été dissous dans différents milieux de reprise testés: (1) acétonitrile/eau (50/50, v/v), (2) acétonitrile/ eau

(50/50, v/v) avec 0.1 % d'acide formique et (3) acétonitrile/ eau (50/50, v/v) avec 10mM d'acétate d'ammonium.

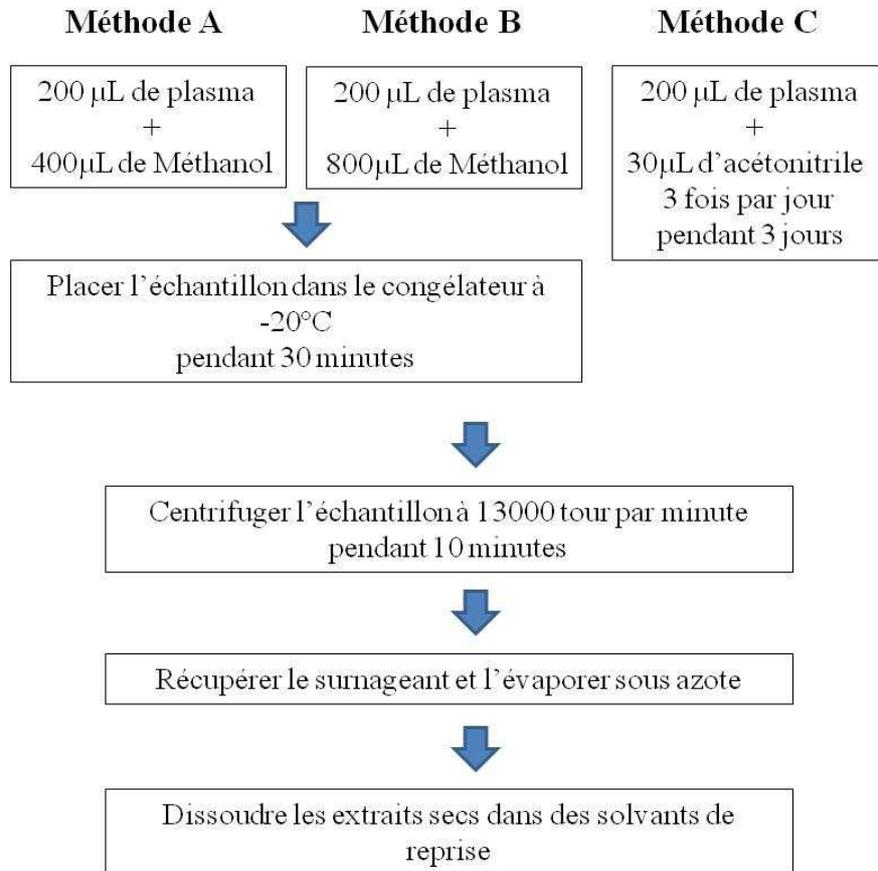


Figure. II.1. Schématisation des étapes de la préparation des échantillons plasmatiques.

3. Echantillons utilisés pour la validation de la méthode analytique.

126 échantillons de plasma prélevés sur héparine issus d'un protocole nutritionnel chez l'homme ont été utilisés. Pour suivre la réponse instrumentale, des contrôles qualités « poolés » (appelés en anglais *pool quality control* PQC) ont été utilisés. Les PQC sont préparés en mélangeant 20µL de chacun des 126 échantillons de plasma à analyser.

4. Analyse par LC/MS des échantillons plasmatiques.

La séparation a été effectuée grâce à un module Waters UPLC Acquity (Saint Quentin en Yvelines, France) et une colonne ACQUITY UPLC HSS T3 de 2,1 x 50 mm (Waters) contenant des particules de 1,8 µm, thermostatée à 30°C. Les échantillons ont été élués à un débit de 0,4 mL/minute en mode gradient linéaire (curve 6), 0 à 2 min : 100 % (A), 2 à 15 min : 0 à 100% (B), 15 à 22 min : 100% (B) et 22 à 26 min : 100% (A) pour la rééquilibration

de la colonne, A et B étant respectivement de l'eau et de l'acétonitrile contenant un milieu tampon. Deux pH de phases mobiles aqueuses ont été testés pour l'élution de solutions standards et d'échantillons de plasma: pH ~3 (avec 0,1% d'acide formique) et pH ~ 6 (avec 10 mM d'acétate d'ammonium). Le volume d'injection a été fixé à 6 μ L d'échantillon.

Le système UPLC a été couplé à un spectromètre de masse QToF micro (Chernushevich *et al.*, 2001) Waters (Saint Quentin en Yvelines, France), équipé d'une source *electrospray* utilisée en mode positif et négatif pour une gamme de masse fixée entre m/z 70 et 1000. La tension du capillaire a été fixée à 3 kV et la tension de cône a été optimisée à 30V, le temps de balayage et le dwell time ont été fixés à 1 et 0,1 seconde ; les températures de la source et désolvatation ont été fixées à 120 et 450 °C et les débits des gaz de désolvatation et nébulisation (azote) ont été fixés à 50 et 500 L / h, respectivement.

Le spectromètre de masse est calibré en utilisant une solution à 0,1% d'acide orthophosphorique, dans un mélange eau/acétonitrile (50/50 v/v). Pour effectuer la correction en masse, une solution de leucine enképhaline ($[M+H]^+$ m/z 556.2771 et $[M-H]^-$ m/z 554.2615) est infusée à un débit de 5 μ L/minute dans la source tout au long des analyses grâce à la technologie « lockspray » à une fréquence de 5s.

5. Traitement de données

Les données sont traitées grâce à des logiciels d'extraction et d'alignement de données LC-MS, qui permettent de comparer les profils plasmatiques. Deux logiciels ont été utilisés : Metalign (América *et al.*, 2006; De Vos *et al.*, 2007; Tikunov *et al.*, 2005), XCMS (Tikunov *et al.*, 2005 et Smith *et al.*, 2006) et MarkerLynx. Dans ce cas de Metalign et XCMS, les fichiers de données brutes « .raw » sont convertis au préalable au format « .cdf » à l'aide du logiciel Waters Databridge. Ces logiciels produisent des matrices de données contenant les temps de rétention, les masses exactes et les intensités des ions détectés.

Metalign a été utilisé sur les données obtenues au cours du développement analytique. Les paramètres Metalign ont été fixés comme suit: Seuil du rapport signal sur bruit de 3 et largeur du pic à mi-hauteur acceptable de 20 scans.

MarkerLynx et XCMS ont été utilisés sur les données obtenues au cours de l'étape de validation. Les paramètres de MarkerLynxTM ont été fixés comme suit : l'intervalle de masse et du temps de rétention acceptable pour aligner les pics présents dans les échantillons sont respectivement de 0,1 et 0,4, le seuil de bruit de fond au-delà duquel un ion est extrait est de 1

et le pourcentage de bruit de fond à éliminer dans le spectre chromatographique est de 30 %. Les paramètres XCMS ont été fixés comme suit: fwmh 10-12, step 0.01, steps 3, Sntreshold 10. (Ces paramètres sont expliqués et testés dans le chapitre III).

6. Analyse statistique des données

6.1. Données issues de l'optimisation analytique.

Chaque ion extrait et aligné a été testé pour sa significativité par un test de Student qui a comparé les 3 répétitions d'une méthode de précipitation protéique (A, B ou C) avec 2 répétitions d'échantillons dits « blancs » contenant du méthanol. Un ion est considéré comme significatif si la valeur de probabilité (p-value) est inférieure à 0,05 et si le rapport des intensités moyennes de l'ion contenu dans les 2 groupes (échantillons versus blancs) est supérieure à 10.

Des Analyses en Composante Principale (ACP) (analyse multivariée) ont également été effectuées en utilisant l'algorithme de NIPALS sur les données transformées en log avec normalisation Pareto, en utilisant le logiciel SIMCA-P v11 (Umetrics, Umeå, Suède).

6.2. Données issues de la validation de la méthode.

Pour la validation de méthode, une Analyse en Composante Principale (ACP) a été réalisée sur les données normalisées et centrées en utilisant le logiciel SIMCA-P v11.

III. Résultats et discussions

Considérant les études récentes publiées en métabolomique sur la matrice plasmatique, nos deux principaux objectifs ont été d'une part, de développer une méthode UPLC-MS permettant l'acquisition de profils métaboliques et d'autre part, d'évaluer l'influence relative des étapes de prélèvement et de préparation d'échantillons sur ces profils, afin de déterminer les conditions optimales.

1. Optimisation des conditions UPLC-MS

Pour cette étude, trois paramètres analytiques ont été testés : le milieu de reprise du résidu sec (ou solvant d'injection), le solvant d'élution et les paramètres de séparation chromatographique. Pour cette comparaison, les échantillons de plasma ont été préparés par la méthode A. Le choix de la méthode de préparation est discuté dans le deuxième paragraphe «Prélèvement et préparation des échantillons». **Les profils métaboliques complets ainsi que la réponse des métabolites d'intérêt sélectionnés (Tableau II. 1) ont été étudiés pour cette optimisation.**

1.1. Influence du solvant de reprise et du solvant d'élution

L'objectif a été de trouver le meilleur compromis entre une bonne solubilisation du résidu sec accompagnée d'une bonne ionisation et de bonnes performances chromatographiques.

L'analyse du nombre d'ions total détectés en mode positif et négatif dans les profils métaboliques obtenus ne montre pas d'influence significative du pH des phases mobiles. Cependant, en examinant la réponse des composés standards ciblés du tableau II. 1, nous avons constaté une meilleure séparation ainsi qu'une meilleure ionisation à pH~3, tant en mode positif que négatif.

Les profils métaboliques plasmatiques obtenus avec les phases mobiles contenant de l'acide formique ont été également de meilleure qualité, considérant la résolution des pics chromatographiques et la détection des composés polaires au début du chromatogramme. Ceci peut être expliqué par la présence d'acide formique qui améliore probablement la forme des pics chromatographiques détectés (Ma *et al.*, 2006) et par la diminution du phénomène de suppression d'ionisation en raison d'une meilleure rétention des composés polaires tels que les acides aminés. (Dimethyl-L-arginine, phénylalanine, tryptophane).

1.2. Influence du débit et du temps d'acquisition.

La technologie UPLC est une technique de séparation combinant des particules de diamètre inférieur à 2 μm et un système chromatographique fonctionnant jusqu'à une pression de 12000 psi, permettant **d'améliorer la résolution chromatographique, d'augmenter la vitesse d'analyse et la sensibilité**. Des études antérieures ont montré que l'UPLC était un outil adapté à la métabolomique car il permettait en particulier d'augmenter le rapport signal sur bruit nécessaire à une meilleure détection et intégration des pics. (Churchwell *et al.*, 2005; De Villiers *et al.*, 2006; Nordstrom *et al.*, 2006). De plus, l'amélioration de la reproductibilité des temps de rétention est déterminante pour l'alignement des données. Un meilleur alignement des données concède à une identification plus aisée des biomarqueurs. (Plumb *et al.*, 2006).

L'objectif de notre étude a été d'optimiser les conditions chromatographiques permettant de conserver la rétention des composés très polaires, tout en gardant l'élution de métabolites moins polaires tels que les phospholipides. Les critères utilisés pour déterminer la méthode optimale ont été les mêmes que précédemment : le nombre total d'ions détectés dans les profils métaboliques obtenus et le comportement chromatographique des métabolites sélectionnés (Tableau II. 1).

L'utilisation d'une colonne C18 greffé par une phase alkyle de densité de 1,6 $\mu\text{mol}/\text{m}^2$ a été choisie (Colonne HSS T3) afin de mieux retenir les composés les plus polaires.

Les meilleures performances de l'UPLC sont obtenues à des débits élevés (de l'ordre de 2 mL/min). Or, en raison de la faible rapidité de balayage du quadripôle et des limitations en pression, le débit optimal du solvant d'élution pour le couplage UPLC-QToF était de 0,4mL/minute. **Les conditions chromatographiques optimales d'analyse ont été obtenues en utilisant un gradient de 26 minutes.** Dans ces conditions les composés lipidiques ainsi que les acides aminés sont séparés et détectés.

Des échantillons de méthanol ont été analysés après six injections d'échantillons de plasma. Nous avons testé 3 temps de rééquilibration de la colonne : 4, 6 et 10 minutes. Aucune contamination croisée n'a été détectée dans les échantillons de méthanol. **Le temps de rééquilibration de la colonne a été fixé à 4 minutes.**

2. Prélèvement et préparation des échantillons plasmatiques.

Notre objectif principal était de déterminer, quelle étape (entre le prélèvement et la préparation d'échantillon) aurait le plus d'impact sur la variabilité des profils métaboliques obtenus, pour pouvoir déterminer dans un second temps les meilleures conditions pour extraire un maximum de métabolites tout en ayant une précipitation des protéines correcte. Dans notre étude trois méthodes de précipitation protéique et deux milieux de reprise ont été testés. (Voir matériels et méthodes)

Pour s'assurer de la bonne extraction de nos métabolites, l'étape de précipitation protéique devait être la plus lente possible comme la méthode lente de Boernsen (méthode C) qui a été réalisée en parallèle et comparée à nos méthodes.

2.1 Choix de l'Anticoagulant

Une analyse en Composante Principale (ACP) a été réalisée sur 36 échantillons plasmatiques (2 anticoagulants x 3 méthodes x 2 milieux de reprise x 3 réplicats) analysés en mode positif (Figure II. 2) et négatif (Figure II. 3). Les résultats ont montré que l'anticoagulant était le facteur le plus influant sur la composition des profils métaboliques. En effet la composante 1 de l'ACP (Figure II. 2) contribue à environ 35% de la variance totale observée alors que les autres facteurs (méthode de précipitation des protéines et milieu de reprise) représentés par la composante 2 contribuent à environ 10% de la variance.

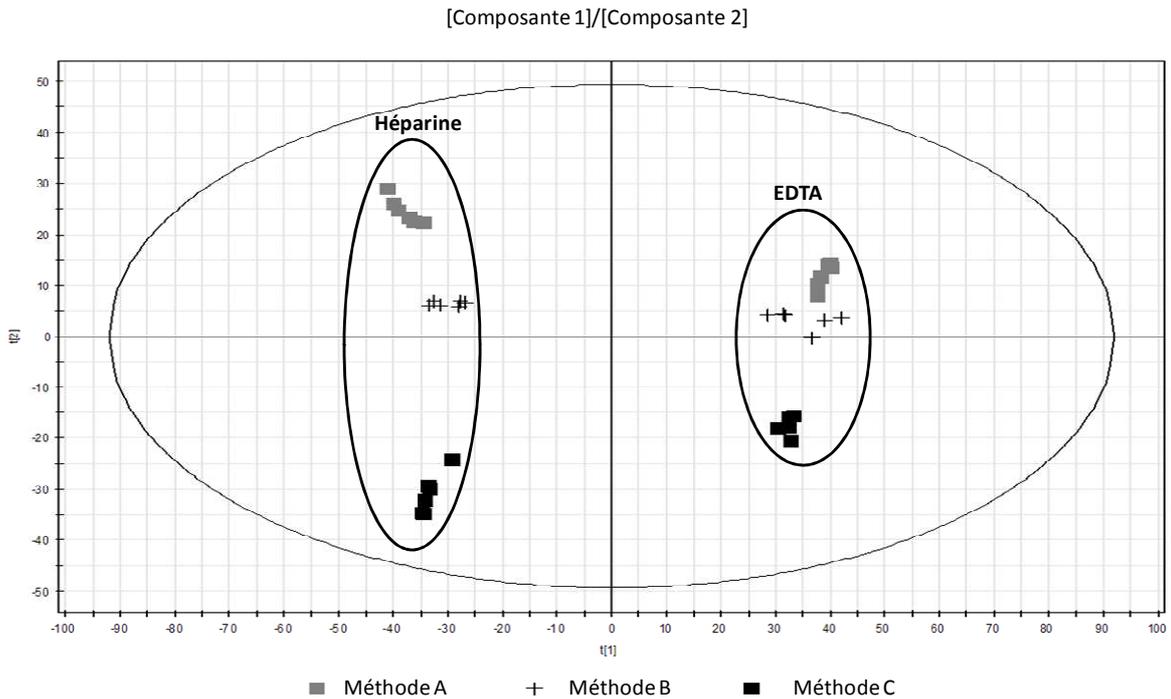


Figure. II.2. Analyse en Composante Principale (ACP) en mode positif des 36 échantillons plasmatiques (2 anticoagulants x 3 méthodes x 3 répétitions x 2 milieux de reprise). $R^2X[1]=0.352097$ et $R^2X[2]=0.101168$.

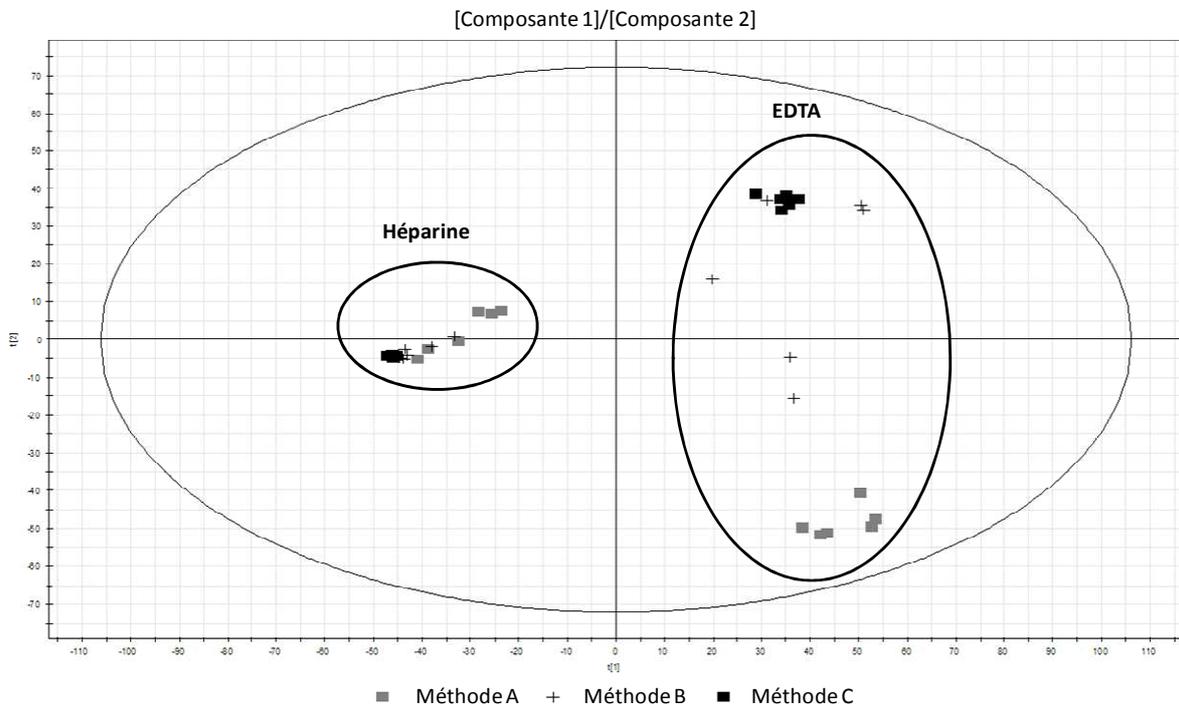


Figure. II.3. Analyse en Composante Principale (ACP) en mode négatif des 36 échantillons plasmatiques (2 anticoagulants x 3 méthodes x 3 répétitions x 2 milieux de reprise). $R^2X[1]=0.338548$ et $R^2X[2]=0.155618$.

De plus, les résultats montrent que les échantillons plasmatiques collectés sur héparine présentent un regroupement plus important en mode négatif (Figure II. 3), contrairement au mode positif où l'axe 2 de l'ACP discrimine clairement les différentes méthodes. L'observation des ions responsables de cette discrimination indique que ces métabolites (par exemple les acides aminés et les lysophosphatidylcholines) ont une meilleure réponse en mode positif, ce qui pourrait expliquer ce phénomène. L'ionisation est probablement améliorée dans les échantillons plasmatiques collectés sur héparine en raison de la présence de sels chélatés en présence d'EDTA. L'EDTA est en effet un puissant agent complexant (Sadagopan *et al.*, 2003) qui peut réagir et endommager la colonne LC. Un large pic chromatographique à 10 minutes, correspondant à l'EDTA (m/z 293), est obtenu après 30 injections. Ce pic provoque un phénomène de suppression d'ionisation dans toute cette zone du chromatogramme.

Par conséquent, **les échantillons de plasma seront collectés sur Héparine pour le reste de notre étude.**

2.2. Choix de la méthode de préparation

L'Analyse en Composante Principale (ACP) réalisée sur les 18 profils métaboliques de plasma collectés sur héparine montre que les 3 méthodes de précipitation donnent des profils métaboliques différents en mode positif (Figure II. 4) et mode négatif (Figure II. 5). En mode positif (Figure II. 4), la composante 1 de l'ACP représentant 24% de la variance totale observée permet de discriminer les méthodes, tandis que la composante 2 représentant 13% de la variance discrimine les milieux de reprise du résidu sec. En mode négatif (Figure II. 5), nous observons sur l'axe 1 une différence entre les échantillons préparés par la méthode A et repris dans la solution à pH~3, et les autres échantillons. L'ionisation des métabolites extraits par la méthode C ne semble pas être influencée par le milieu de reprise en mode positif puisque les deux milieux ne sont pas individualisés sur l'ACP alors qu'en mode négatif, cette discrimination est présente. Notons que l'observation précise des métabolites détectés par LC-MS montre que les métabolites observés dans un milieu acide pH~3 répondent mieux en terme d'intensité lorsqu'ils sont repris dans un milieu neutre. De plus, les temps de rétention et masses des pics en début du chromatogramme ont une meilleure reproductibilité. Par conséquent, nos molécules d'intérêts sont mieux solubilisées et ionisées dans le milieu de reprise eau-acétonitrile 50/50 v/v contenant 0,1% d'acide formique.

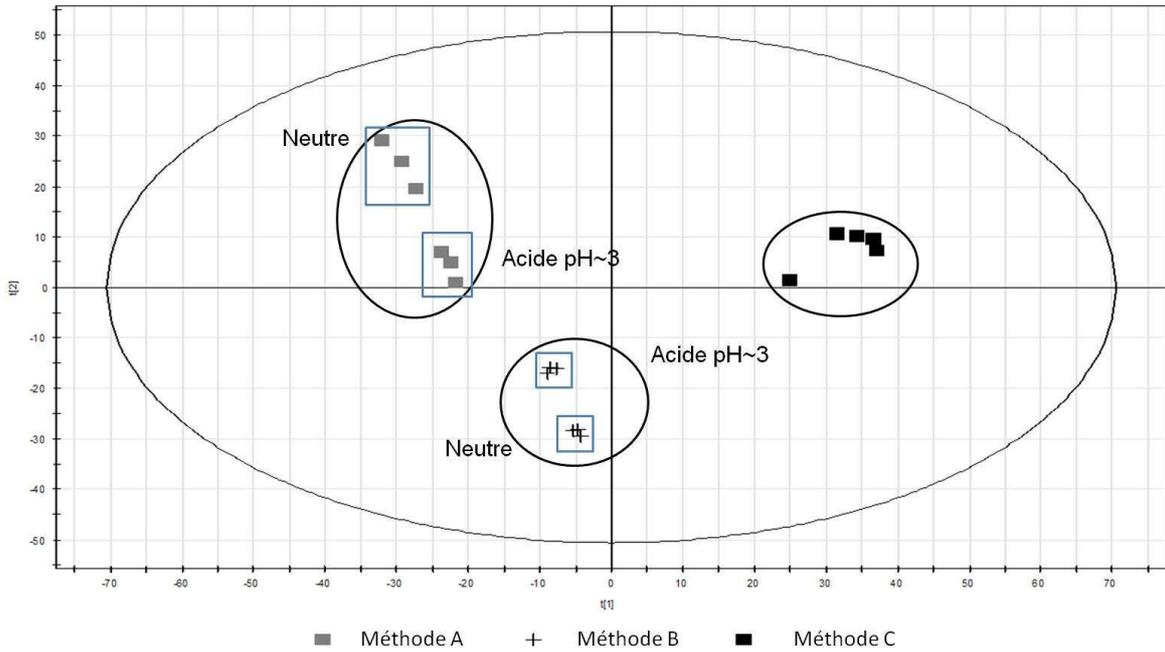


Figure. II.4. Analyse en Composante Principale (ACP) en mode positif des 18 échantillons plasmatiques (3 méthodes de précipitations x 3 répétitions x 2 milieux de reprise neutre (acétonitrile/eau (50/50) et acide pH ~3 (acétonitrile/eau (50/50) avec 0,1% d'acide formique)). $R2X[1]= 0.241958$ et $R2X[2]= 0.124957$.

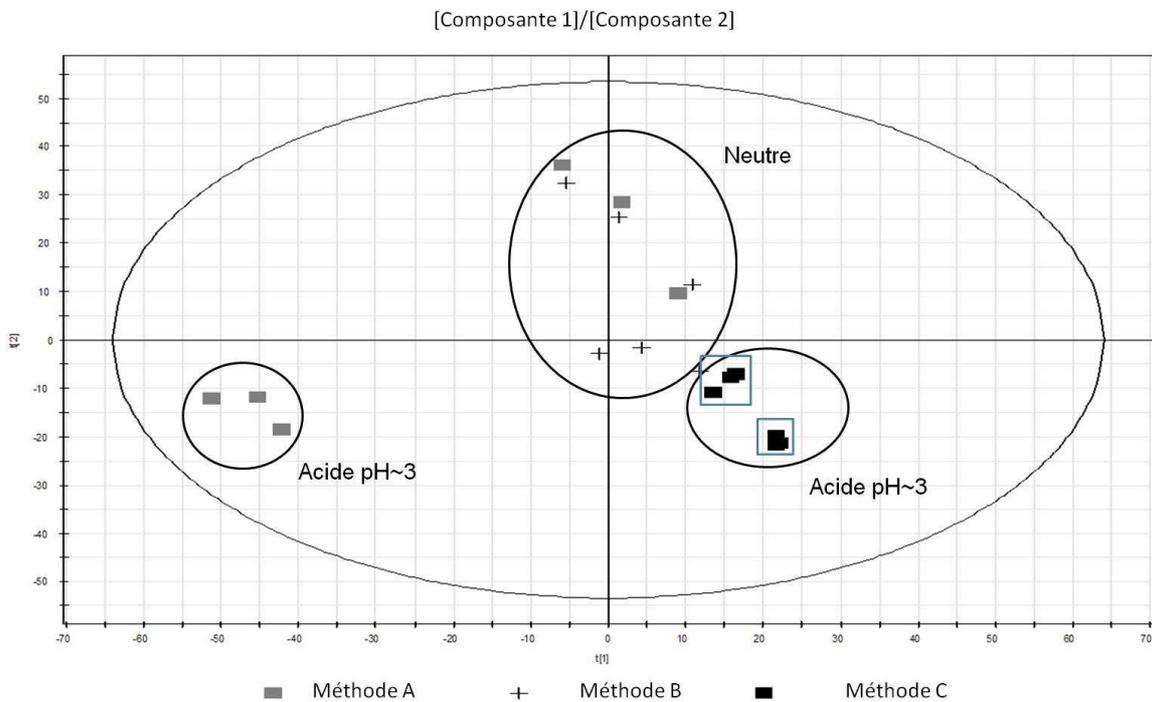
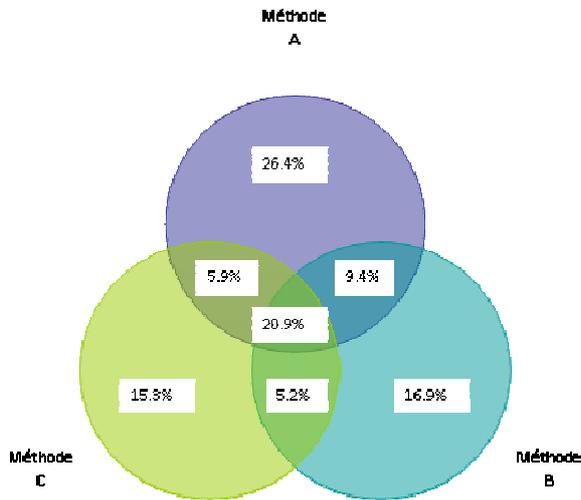
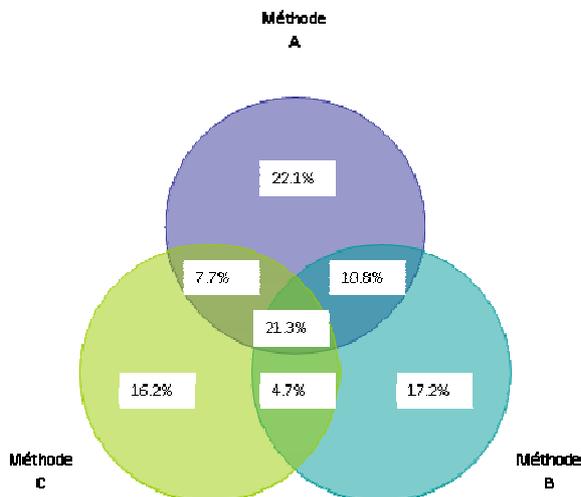


Figure. II.5. Analyse en Composante Principale (ACP) en mode négatif des 18 échantillons plasmatiques (3 méthodes de précipitations x 3 répétitions x 2 milieux de reprise neutre et acide pH ~3). $R2X[1]=0.195947$ et $R2X[2]=0.136438$.

A ce stade, il est encore difficile de conclure sur le choix de la méthode de précipitation. Afin de déterminer la méthode la plus efficace en termes de nombres d'ions extraits, le pourcentage des ions spécifiques et communs aux 3 méthodes a été calculé (Figures II. 6 (a) et (b)).



(a)



(b)

Figure. II.6. Répartition des ions détectés en dans les échantillons plasmatiques préparés par chaque méthode et repris dans un milieu acide à pH~3. (a) mode positif et (b) mode négatif.

En mode positif et négatif, 31% des ions détectés sont communs aux méthodes A et B, 26% aux méthodes B et C et 26%-29% aux méthodes A et C. 21% des ions sont communs aux trois méthodes. Parmi les trois méthodes, la méthode A contient le plus d'ions spécifiques non détectés par les 2 autres méthodes (26,9% en mode positif et 22,1% en mode négatif). Ces données montrent que la précipitation protéique a une influence sur les profils métabolomiques obtenus. Ainsi pour pouvoir comparer les données entre différentes plateformes il est essentiel de trouver des méthodes de préparation standardisées. De plus, tous les métabolites de la liste (Tableau II. 1) sont détectés par la méthode de précipitation protéique A.

La répétabilité des trois méthodes a ensuite été étudiée en utilisant les profils métaboliques obtenus en mode positif et négatif. Pour cela, trois répétitions de méthode d'extraction et d'injection ont été effectuées. D'après la FDA (Food and Drug Administration Guidance, 2001), un coefficient de variation de 15% est acceptable pour l'étude quantitative d'un composé.

Les échantillons préparés par la méthode A, repris dans un solvant à pH~3 et analysés par LC-MS donnent 1108 ions en mode négatif et 814 ions en mode positif. Des valeurs similaires ont été obtenues pour les deux autres milieux de reprise. Les échantillons traités par les méthodes B et C apportent moins d'information. De plus, la méthode A démontre une plus grande répétabilité puisque 72 à 73 % des ions spécifiques à la méthode A et détectés en mode positif et négatif ont un coefficient de variation inférieur à 15%. (Tableau II. 2).

<i>Mode Positif</i>	<i>Méthode A</i>	<i>Méthode B</i>	<i>Méthode C</i>
CV<15%	72,2	68,3	44,8
<i>Mode Négatif</i>	<i>Méthode A</i>	<i>Méthode B</i>	<i>Méthode C</i>
CV<15%	73,1	71,1	47,4

Tableau. II.2. Pourcentage des ions spécifiques à chaque méthode dont le coefficient de variation est inférieur à 15%. Chaque méthode d'extraction a été répétée 3 fois.

Par conséquent, la méthode de précipitation protéique A est sélectionnée.

3. Validation de la méthode développée.

L'étape de validation permet de confirmer que la méthode analytique employée est appropriée pour les critères sélectionnés. Les résultats obtenus après **l'étape de validation permettent de confirmer l'exactitude des résultats.**

Dans le domaine de la métabolomique appliquée à la nutrition, plusieurs étapes sont source de variabilités telles que le prélèvement, la conservation et la préparation des échantillons. Il est donc très important de quantifier la variabilité analytique. Ainsi, cette validation est axée sur **l'étude de la robustesse** de l'étape de préparation des échantillons et sur la stabilité de la réponse instrumentale en termes de temps de rétention, de masse et d'intensité.

La validation de cette méthode s'est faite en plusieurs étapes : (i) étude de l'effet matrice, (ii) validation de la méthode d'extraction de métabolites et (iii) étude de la stabilité instrumentale en termes de temps de rétention, de masse et d'intensité.

3.1. Etude de l'effet matrice

Afin de caractériser les performances de la méthode analytique, les échantillons plasmatiques ont été dopés et analysés par LC-MS. Evaluer l'effet matrice est important lorsqu'une méthode analytique est développée. La coélution des composés de la matrice plasmatique dans la LC peut en effet induire une diminution voire une suppression du signal causé par la compétition de l'ionisation des composés élués dans la source. (Van de Steene *et al.*, 2008).

Six standards de xénobiotiques ont été utilisés afin d'étudier cet effet. Ces composés ont été sélectionnés pour leurs propriétés physico-chimiques (polarité, masse moléculaire et fonctions chimiques) similaires aux métabolites d'intérêts étudiés, ainsi que pour leur bonne solubilisation dans le plasma et leur bonne ionisation en *electrospray*. L'effet matrice a été étudié en utilisant différentes dilutions du plasma (5 fois, 10 fois, 15 fois, 20 fois) dans un mélange eau/acétonitrile (50/50, v/v) avec 0,1% d'acide formique. Pour chaque dilution, trois échantillons ont été préparés.

D'après la Figure II. 7, la réponse instrumentale (aire des pics) des six standards de xénobiotiques est constante quelque soit la dilution de la matrice plasmatique. Par conséquent, l'effet matrice est reproductible dans ces conditions analytiques.

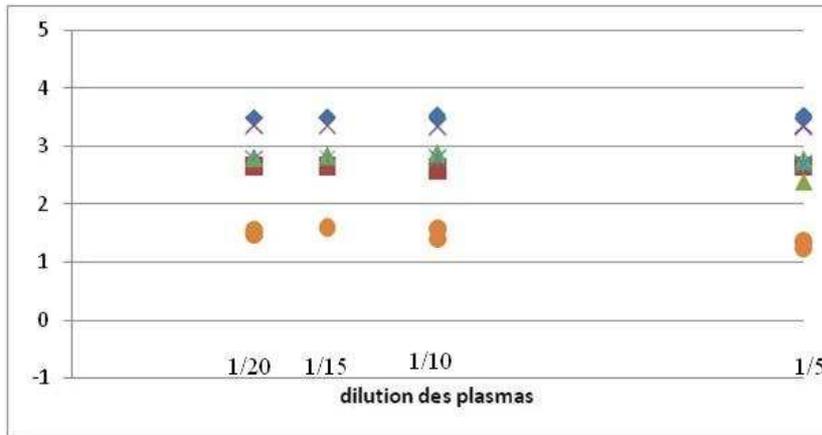


Figure. II.7. Logarithme de la valeur de l'aire des pics chromatographiques des standards de xénobiotiques dans des plasmas de dilution différente (1/20, 1/15, 1/10 et 1/5) : ◇Colchicine, x Aminopyridine, △ Aminoanthracène, *Procaine, □ Prednisone, ○ Acide Dehydrocholique.

3.2. Validation de l'étape de préparation des échantillons plasmatiques

Douze échantillons de plasma de 200µL ont été dopés avec six standards de xénobiotiques et six standards biologiques de concentration constante avant et après précipitation protéique. Pour chaque standard, un rendement d'extraction a été calculé à partir des valeurs de l'aire des pics chromatographiques avant et après précipitation protéique. (Tableau II. 3).

	<i>Composés</i>	<i>Temps de rétention (min)</i>	<i>de CV (%)</i>	<i>Rendement(%)</i>
Standards	Leucine	0,45	1	86±4
Biologiques	Méthionine	0,50	1	90±1
	Phénylalanine	3,45	$6 \cdot 10^{-2}$	72±8
	Tryptophane	4,12	$8 \cdot 10^{-1}$	80±7
	LysoPC(16:0)	10,54	$1 \cdot 10^{-3}$	100±2
	LysoPC(18:1)	10,88	$1 \cdot 10^{-3}$	102±1
	LysoPC(18:0)	11,73	$1 \cdot 10^{-3}$	100±2
Standards de xénobiotiques	2-aminoantipyrène	3,98	$8 \cdot 10^{-1}$	72±13
	4.aminoanthracène	7,89	$4 \cdot 10^{-1}$	81±10
	Acide	7,85	$4 \cdot 10^{-2}$	67±6
	Dehydrocholic Colchicine	6,35	$5 \cdot 10^{-1}$	68±4
	Prednisone	6,71	$4 \cdot 10^{-1}$	68±4
	Procaine	4,14	$8 \cdot 10^{-1}$	81±5

Tableau. II.3. Rendement de l'extraction des standards dopés dans le plasma après précipitation protéique lente. Pour chaque composé, le coefficient de variation a été calculé.

Les résultats montrent que pour les 12 standards, les rendements sont supérieurs à 67% avec des coefficients de variation inférieurs à 15%. Ces résultats confirment que l'étape d'extraction des métabolites est convenable pour les standards suivis.

3.3. Validation de la méthode.

Dans cette deuxième partie, 126 échantillons de plasma ont été préparés et analysés par UPLC-MS. Ces échantillons ont été répartis dans quatre séries d'analyses selon un plan d'expérience. Dans chaque série d'analyses, après trois injections de méthanol, la répétabilité de l'injection a été vérifiée, en injectant à trois reprises un échantillon de contrôle qualité composés de standards (SQC en anglais *standards quality control*). La stabilité du système analytique a été suivi en injectant des échantillons de contrôle qualité composés d'un mélange des plasmas à analyser. (PQC en anglais *pool quality control*). Ces échantillons ont été injectés 1 fois après les SQC puis 2 fois après injection de 10 échantillons plasmatiques. Au total 42 à 52 injections ont été effectuées pour chaque série d'analyses. La calibration de l'instrument a été faite au début de chaque série. La source a été nettoyée après 60 injections de plasma, comme le suggère Zelena *et al.* (2009).

Pour cette validation, nous avons examiné la variabilité des temps de rétention, de masses exactes, et d'intensités des ions présents dans les PQC. Pour les composés présents dans le tableau II. 1 et détectés dans les échantillons de PQC, une faible variabilité des temps de rétention a été observée (CV inférieurs à 1,4%). De même les coefficients de variation des masses exactes ont été inférieurs à 0,03% dans une série d'analyses et inférieurs à 0,3% pour l'ensemble de l'analyse. La stabilité de la masse exacte et du temps de rétention mesurés, est cruciale pour l'étape d'extraction et d'alignement des données discutés dans le chapitre III de ce manuscrit.

Dans une étude récente analysant du sérum par UPLC-MS (Zelena *et al.*, 2009), il a été montré que la stabilité du système n'était effective qu'après l'injection de dix échantillons de PQC. D'après la Figure II. 8, environ 50% des ions détectés dans deux échantillons de PQC successives ont des variations d'intensités inférieures à 20%.

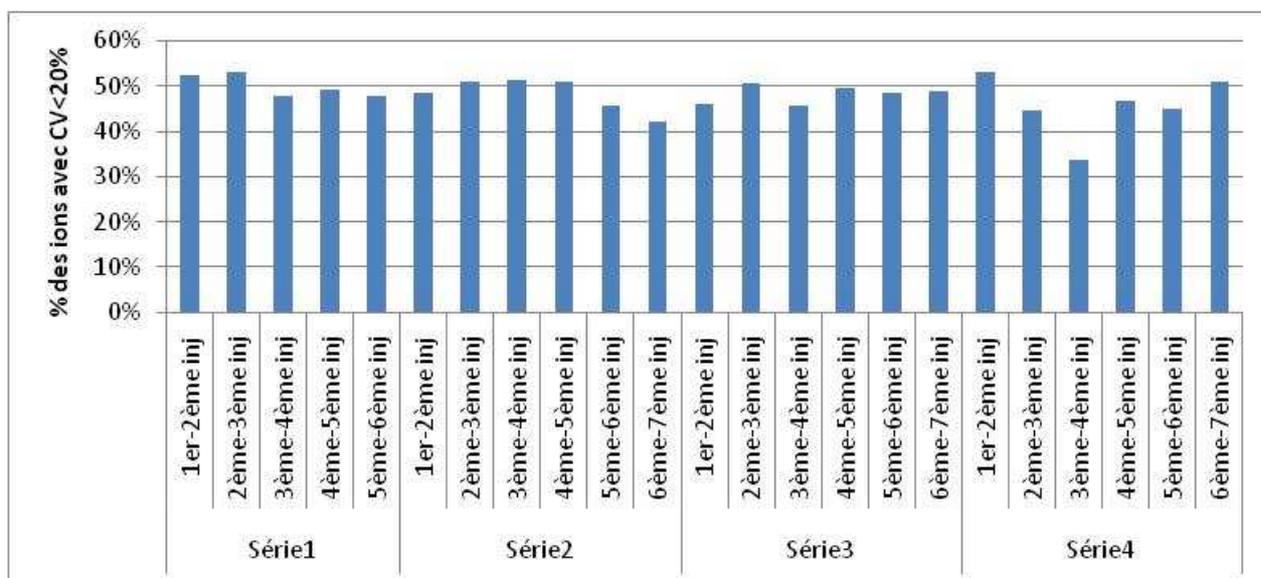


Figure. II.8. Pourcentage des ions dont les intensités varient de moins de 20% entre deux injections de contrôle qualité d'une même série d'analyse.

Notre temps d'acquisition de 22 minutes ainsi que notre temps de rééquilibrage de colonne de 4 minutes peuvent expliquer ces résultats. L'injection d'échantillons de SQC a permis de suivre en temps réel le système chromatographique par la vérification visuelle des pics chromatographiques des standards.

La variation des intensités des ions détectés dans les échantillons de PQC pendant quatre jours est représentée dans la Figure II. 9. Les ions de faibles intensités ($I < 100$ coups par seconde) présentent des coefficients de variation toujours supérieurs à 30%, tandis que pour les ions de fortes intensités ($I > 10000$ coups par seconde) ces coefficients de variation sont inférieurs à

15%. Ces résultats montrent que la variabilité est dépendante de l'intensité du signal. Ce phénomène a été observé dans des études précédentes (Gika *et al.*, 2008). Cette variabilité peut être attribuée à la réponse instrumentale, mais également à l'étape d'extraction des pics durant le traitement des données. Sur ce jeu de données, l'étape d'extraction des ions a été réalisée par le logiciel MarkerLynx. Ce logiciel a pour particularité d'extraire des ions de faible intensité ce qui peut expliquer les valeurs élevées de coefficients de variation. (Discuté dans le chapitre III).

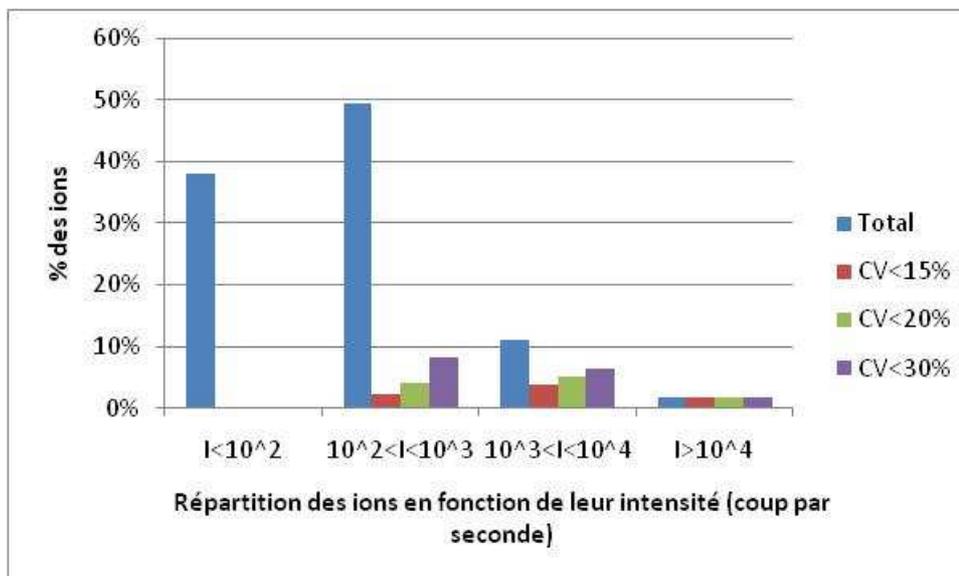


Figure. II.9 . Pourcentage des ions totaux détectés (ou avec un CV < 15%, 20% ou 30%) dans les échantillons plasmatiques de contrôle qualité (PQC) en fonction de la valeur moyenne de la hauteur du signal de l'ion.

Dans le cas d'études de nombreux composés une variabilité instrumentale de 20% semble être un critère approprié. (Zelena *et al.*, 2009). Le pourcentage des ions détectés dans les profils métaboliques dont le coefficient de variation remplit ce critère sont présentés Figure II. 10. Pour des ions de faibles intensités (I compris entre 100 et 1000 coups par seconde) moins de 40% des pics ont un coefficient de variation inférieur à 20% tandis que 100% des ions intenses (mais inférieurs à la saturation) remplissent ce critère. Ces résultats indiquent aussi que la variabilité intra-série d'analyses est inférieure à la variabilité inter-séries d'analyses.

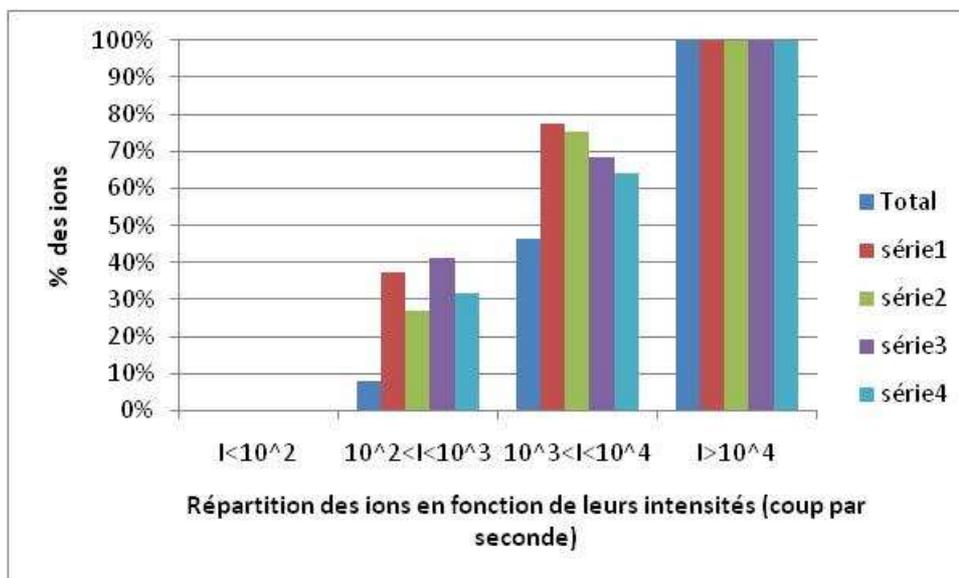


Figure. II.10. Pourcentage des ions avec un $CV < 20\%$ détectés dans les échantillons plasmatiques de contrôle qualité (PQC) et analysés dans les 4 séries d'analyses en fonction de la valeur moyenne de la hauteur d'intensité.

La Figure II. 11 représente une Analyse en Composante Principale des profils métaboliques des 126 échantillons de plasma et des 27 échantillons de PQC, en mode positif. Les données ont été normalisées par la valeur moyenne de l'intensité des ions dans une série d'analyse. Notons que les échantillons de PQC ne sont pas regroupés au centre de l'ellipse de l'ACP, comme observé par Michopoulos *et al.* (2009). Des effets matrices causés par la compétition d'ionisation des composés non retenus peuvent expliquer ce phénomène.

Cette figure illustre que la variabilité entre les échantillons biologiques est supérieure à la variabilité entre les échantillons de PQC (en noir). Par conséquent, la plus grande source de variabilité peut donc être attribuée à la variation biologique et non pas à la variation liée au système analytique. Ce résultat démontre la performance de la méthode analytique développée pour des études métabolomique en nutrition, impliquant des analyses différentielles du métabolome.

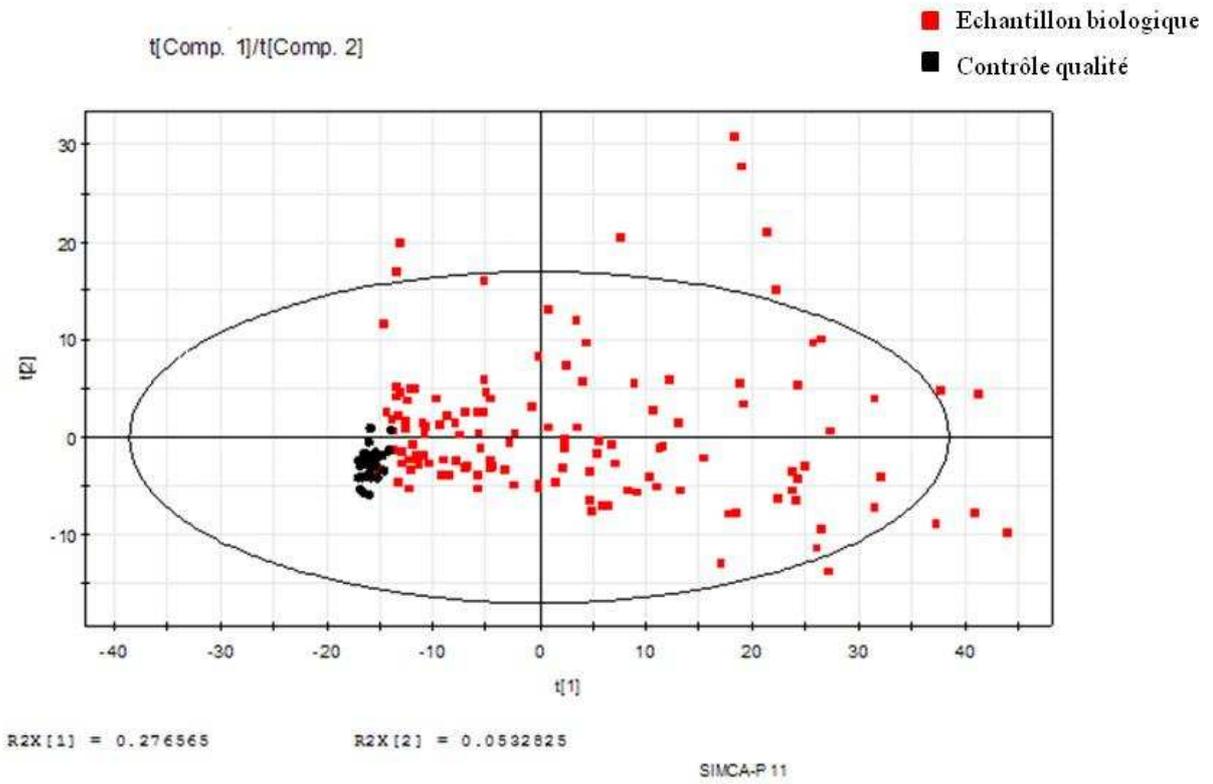


Figure. II.11. Analyse en Composante Principale des 126 échantillons biologique et 27 échantillons de contrôle qualité. Les données ont été normalisées et centrées.

IV. Conclusion

Dans cette étude, nous avons développé une méthode simple et robuste pour analyser par UPLC-MS les échantillons de plasma issus d'une étude métabolomique en nutrition. Cette méthode permet d'obtenir des profils métaboliques complets tout en gardant une détection optimale des métabolites d'intérêt, préalablement sélectionnés pour caractériser des dérégulations métaboliques. Cette méthode ne permet pas la quantification de métabolites mais elle est la première étape d'identification de ces derniers amenant à la découverte de biomarqueurs par analyse différentielle. Les biomarqueurs devront ensuite être validés et pourront être quantifiés par des méthodes ciblées.

D'après nos résultats, **l'héparine** est **l'anticoagulant** le plus approprié pour l'étude non ciblée du métabolome. De plus, la méthode de **précipitation protéique lente au méthanol avec un rapport méthanol/plasma (2/1)** est la plus reproductible. La validation de la méthode développée montre que l'intensité du signal qui dépend de la nature du métabolite et de sa concentration induit la plus grande source de variabilité. Pour contrôler cette variation, dans le cas de longues séries d'analyse (nombre d'échantillons supérieur à 100), il est conseillé de **répartir les échantillons dans des séries indépendantes pour la validation de marqueurs biologiques**.

Références

- America, A.H.P., Cordewener, J.H.G., van Geffen, M.H.A., Lommen, A., Vissers, J.P.C., Bino, R.J., Hall, R.D. (2006). Alignment and statistical difference analysis of complex peptide data sets generated by multidimensional LC-MS. *Proteomics*. **6**, 641-653.
- Benton, H.P., Wong, D.M., Trauger, S.A., Siuzdak, G. (2008). XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. *Anal. Chem.* **80**, 6382-6389.
- Boernsen, K.O., Gatzek, S., Imbert, G. (2005). Controlled Protein Precipitation in Combination with Chip-Based Nanospray Infusion Mass Spectrometry. An Approach for Metabolomics Profiling of Plasma. *Anal. Chem.* **77**, 7255-7264.
- Bruce, S.J., Jonsson, P., Antti, H., Cloarec, O., Trygg, J., Marklund, S.L., Moritz, T. (2008). Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis. *Anal. Biochem.* **372**, 237-249.
- Bruce, S.J., Tavazzi, I., Parisod, V., Rezzi, S., Kochhar, S., Guy, P.A. (2009). Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* **81**, 3285-3296.
- Chambers, E., Wagrowski-Diehl, D.M., Lu, Z., Mazzeo, J.R. (2007). Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses. *J. Chromatogr. B.* **852**, 22-34.
- Chernushevich, I.V., Loboda, A.V., Thomson, B.A. (2001). An introduction to quadrupole-time-of-flight mass Spectrometry. *J. Mass Spectrom.* **36**, 849-865.
- Churchwell, M.I., Twaddle, N.C., Meeker, L.R., Doerge, D.R. (2005). Improving LC-MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS. *J. Chromatogr. B.* **825**, 134-143.
- De Villiers, A., Lestremau, F., Szucs, R., Gélébart, S., David, F., Sandra, P., (2006). Evaluation of ultra performance liquid chromatography: Part I. Possibilities and limitations. *J. Chromatogr. A.* **1127**, 60-69.

- De Vos, C.H.R., Moco, S., Lommen, A., Keurentjes, J.J., Bino, R.J., Hall, R.D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protoc.* **2**, 778-791.
- Dettmer, K., Pavel, A.A., and Hammock, B.D. (2007). Mass spectrometry Based Metabolomics. *Mass Spectrom. Reviews* **26**, 51-78.
- Dumas, M.-E., Barton, R. H., Toye, A., Cloarec, O., Blancher, C., Rothwell, A., Fearnside, J., Tatoud, R., Blanc, V., Lindon, J. C., Mitchell, S. C., Holmes, E., McCarthy, M. I., Scott, J., Gauguier, D., Nicholson, J. K. (2006). Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *PNAS.* **103**, 12511-12516.
- Fardet, A., Llorach, R., Martin, J. F., Besson, C., Lyan, B., Pujos-Guillot, E. and Scalbert, A. (2008). A Liquid Chromatography-Quadrupole Time-of-Flight (LC-QTOF)-based Metabolomic Approach Reveals New Metabolic Effects of Catechin in Rats Fed High-Fat Diets. *J. Proteome Res.* **7**, 2388–2398.
- FDA Guidance for industry, Bioanalytical method Validation, Food and Drug Administration. **May 2001**. *Center for Drug Evaluation and research (CDER)*.
- Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nature.* **18**, 1157-1161.
- German, J.B., Roberts, M.A., Fay, L., Watkins, S.M. (2002). Metabolomics and Individual Metabolic Assessment: The Next Great Challenge for Nutrition. *J. Nutr.* **132**, 2486-2487.
- Gibney, M.J., Gibney, E.R. Symposium on ‘Implications for dietary guidelines of genetic influences on requirements’ Diet, genes and disease: implications for nutrition policy.(2004) *Proceedings of the Nutrition Society.* **63**, 491–500.
- Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, J.B., Van Ommen, B. (2005). Metabolomics in human nutrition : opportunities and challenges. *Am. J. Clin Nutr.* **82**, 497-503.
- Gika, H., Macpherson, E., Theodoridis, G., Wilson, I.D. (2008). Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B.* **871**, 299-305.

- Guo, Z., Yarasheski, K., Jensen, M.D. (2006). High-precision isotopic analysis of palmitoylcarnitine by liquid chromatography/electrospray ionization ion-trap tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 3361-3366.
- Guy, P.A., Tavazzi, I., Bruce, S.J., Ramadan, Z., Kochhar, S. (2008). Global metabolic profiling analysis on human urine by UPLC-TOFMS: Issues and method validation in nutritional metabolomics. *J. Chromatogr. B.* **871**, 253-260.
- Houjou, T., Kotoko, Y., Masayoshi, I., Takao, S., Ryo, T. (2005). A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **19**, 654-666.
- Idle, J.R. and Gonzalez, F.J. (2007). Metabolomics. *Cell Metabolism.* **6**, 348-351.
- Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, T., Moritz, M. (2004). A Strategy for Identifying Differences in Large Series of Metabolomic Samples Analyzed by GC/MS. *Anal. Chem.* **76**, 1738-1745.
- Koutsari, C. and Jensen, M. D. (2006). Thematic review series: Patient-Oriented Research. Free fatty acid metabolism in human obesity. *J. Lipid Res.* **47**, 1643-1650.
- Lawton, K.A., Berger, A., Mitchell, M., Milgram, K.E., Evans, A.M., Guo, L.N., Hanson, R.W., Kalhan, S.C., Ryals, J.A., Milburn, M. V. (2008). Analysis of the adult human plasma metabolome. *Pharmacogen.* **9**, 383-397.
- Lenz, W. (2007). Analytical Strategies in Metabonomics. *J Proteome Res.* **6**, 443-458.
- Liebisch, G., Drobnik, W., Lieser, B., Schmitz, G. (2002). High-Throughput Quantification of Lysophosphatidylcholine by Electrospray Ionization Tandem Mass Spectrometry. *Clin Chem.* **48**, 2217-2224.
- Liebisch, G., Lieser, B., Rathenber, J., Drobnik, W., Schmitz, G. (2004). High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim. Bioph. Act. - Molecular and Cell Biology of Lipids.* **1686**, 108-117.

- Lindon, J. C., Nicholson, J. K. (2008). Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC*. **27**, 194-204.
- Ma, L., Luo, H., Dai, J., Carr, P. W., (2006). Development of acid stable, hyper-crosslinked, silica-based reversed-phase liquid chromatography supports for the separation of organic bases. *J. Chromato.A*. **1114**, 21-28.
- Maher, A.D., Zirah, S.F.M., Holmes, E., Nicholson, J.K. (2007). Experimental and Analytical Variation in Human Urine in ¹H NMR Spectroscopy-Based Metabolic Phenotyping Studies. *Anal. Chem.* **79**, 5204-5211.
- Major, H., Williams, R., Wilson, A., Wilson, I. (2006). A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition. *Rapid Commun Mass Spectrom.* **20**, 3295–3302.
- Michopoulos, F., Lai, L., Gika, H., Theodoridis, G., Wilson, I. (2009). UPLC-MS-Based Analysis of Human Plasma for Metabonomics Using Solvent Precipitation or Solid Phase Extraction. *J. Proteome Res.* **8**, 2114-2121.
- Moco, S., Bino, R.J., Vos, R.C.H.D., Vervoort, J. (2007). Metabolomics technologies and metabolite identification. *TrAC*. **26**, 855-866.
- Nicholson, J.K., Foxall, P. (1995). 750 MHz ¹H and ¹H ¹³C NMR Spectroscopy of Human Blood Plasma. *Anal. Chem.* **67**, 793-811.
- Nicholson, J.K, Lindon, J.C, Holmes, E. (1999). Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. **29**, 1181-1189.
- Nordstrom, A., O'Maille, G., Qin, C., Siuzdak, G. (2006). Nonlinear Data Alignment for UPLC-MS and HPLC-MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum. *Anal. Chem.* **78**, 3289-3295.
- Pietilainen, K.H., Sysi-Aho, M., Rissanen, A., Seppanen-Laakso, T., Yki-Jarvinen, H., Kaprio, J., Oresic, M., Waterfield, C.J. (2007). Acquired Obesity Is Associated with Changes in the Serum Lipidomic Profile Independent of Genetic Effects –A Monozygotic Twin Study. *PLoS ONE*. **5**, 472-483.

- Piraud, M., Vianey-Saban, C., Petritis, K., Elfakir, C., Steghens, J.P., Morla, A., Bouchu, D. (2003). ESI-MS/MS analysis of underivatized amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. *Rapid Commun. Mass Spectrom.* **17**, 1297-1311.
- Plumb, R.S., Johnson, K.A., Rainville, P.D., Shockcor, J.P., Williams, R., Granger, J.H., Wilson, I.D. (2006). 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**, 2800-2806.
- Plumb, R.S., Johnson, K.A., Rainville, P.D., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., Nicholson, J.K. (2006). UPLC/MS; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **20**, 1989-1994.
- Rezzi, S., Ramadan, Z., Fay, L.B., Kochlar, S. (2006). Nutritional Metabonomics: Applications and Perspectives. *J. Proteome Res. Reviews* **6**, 513-525.
- Sadagopan, N.P., Li, W., Cook, J.A., Galvan, J.A., Weller, D.L., Fountain, S.T., Cohen, L.H. (2003). Investigation of EDTA anticoagulant in plasma to improve the throughput of liquid chromatography/tandem mass spectrometric assays. *Rapid Commun. Mass Spectrom.* **17**, 1065-1070.
- Sébedio, J.L., Martin, J.F., Pujos, E. (2008). Nutritional Metabolomics: What are the perspectives?. *OCL*. **15**, 341-345.
- Skeaff, C.M, Hodson, L, McKenzie, J.E. (2006). Dietary-induced changes in fatty acid composition of human plasma, platelet, and erythrocyte lipids follow a similar time course. *J. Nutr.* **136**, 565-569.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779-787.

- Tikunov, Y., Lommen, A., de Vos, C.H.R., Verhoeven, H.A., Bino, R.J., Hall, R.D., Bovy, A.G. (2005). A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. *Plant Physiology*. **139**, 1125–1137.
- Van De Steene, J. C. and Lambert, W. E. (2008). Comparaison of Matrix Effects in HPLC-MS/MS and UPLC-MS/MS Analysis of Nine Basic Pharmaceuticals in Surface Waters. *Am. Soc. Mass Spectrom.* **19**, 713-718.
- Want, E., O'Maille, G., Smith, C., Brandon, T., Uritboonthai, W., Qin, C., Trauger, S., Siuzdak, G. (2006). Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal Chem.* **78**, 743-752.
- Watkins, S.M., Hammock, B.D., Newman, J.W., German, J.B. (2001). Individual metabolism should guide agriculture toward foods for improved health and nutrition. *Am J Clin Nutr.* **74**, 283-286.
- Williams, R., Lenz, E., Wilson, A., Granger, J., Wilson, I., Major, H., Stumpf, C., Plumb, R. (2006). A multi-analytical platform approach to the metabonomic analysis of plasma from normal and Zucker (fa/fa) obese rats. *Mol Biosyst.* **2**, 174-183.
- Wilson, I.D., Plumb, R., Granger, J., Major, H., Williams, R., Lenz, E.M. (2005). HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr. B.* **817**, 67-76.
- Wong, M.C.Y., Lee, W.T.K., Wong, J.S.Y., Frost, G., Lodge, J. (2008). An approach towards method development for untargeted urinary metabolite profiling in metabonomic research using UPLC/QToF MS. *J. Chromatogr. B.* **871**, 341-348.
- Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D., Halsall, A., Wilson, I.D., Kell, D.B. (2009). Development of a Robust and Repeatable UPLC;MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.* **81**, 1357-1364.

CHAPITRE III : TRAITEMENT DE DONNEES

Chapitre III

I. Introduction

La métabolomique est une approche à haut débit qui engendre une **multitude de données complexes**. Pour extraire l'information biologique, ces données doivent être traitées avec les outils appropriés. (Scalbert *et al.*, 2008).

En nutrition, l'effet du régime sur le métabolisme est souvent de faible intensité et très variable entre les sujets. (German *et al.*, 2005). Pour que la puissance statistique des tests effectués soit suffisante, **le protocole biologique doit être réalisé sur un grand nombre d'individus** conduisant, en général, à l'analyse de **plusieurs centaines d'échantillons**. (Gibney *et al.*, 2005). De plus, les études nutritionnelles s'accompagnent souvent au cours du protocole d'un suivi des observations biologiques (points cinétiques) des sujets entraînant encore une augmentation du nombre d'échantillons.

Dans ce contexte, en métabolomique nutritionnelle, l'analyse de ces centaines d'échantillons nécessite une technique à haut débit, telle que le couplage entre la chromatographie liquide et la spectrométrie de masse (Fardet *et al.*, 2008 et Gika *et al.*, 2008). Pour traiter l'ensemble des données issues de l'analyse par LC-MS plusieurs étapes sont nécessaires :

- 1) Elimination des signaux du bruit de fond et correction de la masse des signaux détectés et sélectionnés.
- 2) Extraction et alignement des signaux sélectionnés dans chaque échantillon en fonction de leur temps de rétention, de leur masse exacte et de leur intensité.
- 3) Sélection des ions sensibles à l'effet du régime grâce à l'analyse statistique.

Dans l'étape d'extraction et d'alignement des données, correspondant à la deuxième étape, le choix des paramètres des logiciels peut engendrer des biais. Par conséquent, le choix de ces paramètres doit être considéré comme **l'étape critique dans le traitement des données métabolomiques** (Dunn *et al.*, 2008).

De plus, en LC-MS, des effets de variation de temps de rétention et de masse risquent de dupliquer les ions extraits. Des phénomènes de suppression d'ionisation dans la source demeurent et peuvent considérablement modifier les intensités du signal mesurées. Par conséquent, la comparaison d'un grand nombre d'échantillons amplifiera ces variations. Ainsi, malgré l'optimisation de l'extraction et de l'alignement des données, **une étape de prétraitement des données avant l'analyse statistique est souvent nécessaire.** (Idgord *et al.*, 2005). Les méthodes de prétraitement sont nombreuses : centrage, centrage-réduction, transformation en logarithme. Elles sont adaptées aux données et modèles statistiques utilisés. Dans les protocoles nutritionnels, le modèle statistique utilisé doit permettre de vérifier que les données analytiques obtenues sont modifiées par les phénotypes observés chez les individus soumis à un régime. Par conséquent, des **modèles supervisés de classification sont utilisés tel que la PLS-DA.** (en anglais *Partial Least Square- Discriminant Analysis*).

Notre objectif était de proposer une démarche de traitement des données obtenues par LC-MS appliquée à de longues séries d'analyses (plus de 100 échantillons). Deux étapes ont été cruciales dans le traitement des données : la détermination des paramètres optimaux des logiciels d'extraction et d'alignement des données et l'évaluation de l'effet des différentes méthodes de prétraitement sur le modèle de discrimination statistique PLS-DA. En effet, diverses études présentent les effets de différents types de prétraitement (Van den Berg, *et al.*, 2006) ou différents types de modèles statistiques (van Velzen *et al.*, 2008) mais, à notre connaissance aucune étude n'a comparé simultanément ces deux effets.

Pour réaliser cette étude l'ensemble des processus de traitement des données a été appliqué sur des données provenant de l'analyse d'échantillons d'urines prélevées sur cent souris soumises à un régime hyper-lipidique pendant 23 semaines dans le cadre du projet ANR Metaprofile. Par conséquent, 482 échantillons ont été répartis en 26 séries d'analyse et analysés par LC-MS.

En combinant suivant un plan factoriel toutes les options utilisées aux différentes étapes du traitement de données, on a obtenu 48 matrices de données qui ont été analysés par une analyse discriminante supervisée, la PLS-DA. (Figure III. 1).

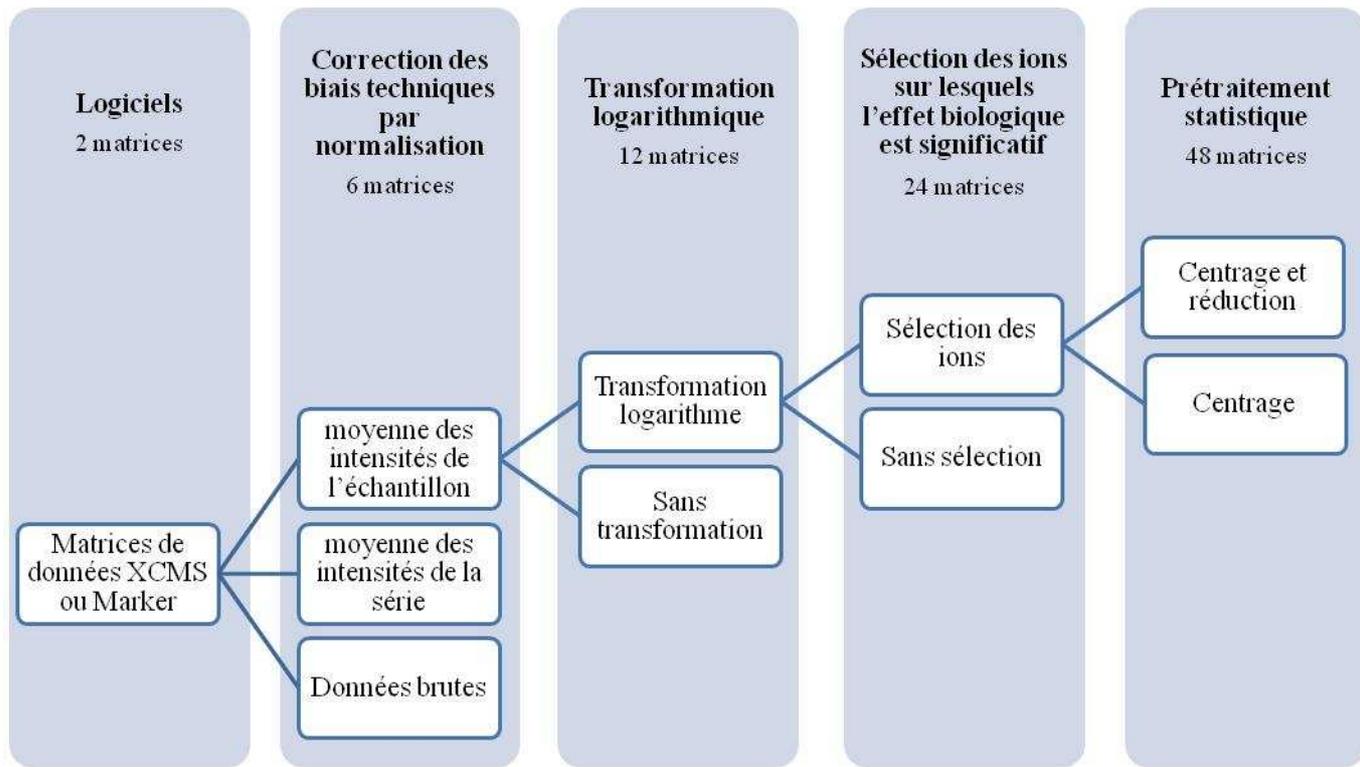


Figure III.1 _ Schématisation du plan factoriel représentant les étapes successives des prétraitements des données appliqués sur les données brutes obtenues pour les deux logiciels d'extraction et d'alignement : MarkerLynx et XCMS.

II. Matériels et Méthodes

1. Traitements des échantillons biologiques

1.1. Prélèvement des échantillons

Le protocole a été réalisé sur un modèle de souris C57B6/6J pouvant développer un phénotype obèse suite à un régime hyper-lipidique. Après acclimatation, 100 souris de 5 semaines ont été réparties dans différentes cages. 20 souris ont été logées par cage et nourries avec un régime hyper-lipidique contenant 72% de matière grasse (huile de maïs et saindoux), 28% de protéines et 1% de glucides pendant 23 semaines. Les prélèvements urinaires ont été recueillis après 6 heures à jeun lors de la 0, 4^{ème}, 8^{ème}, 14^{ème} et 23^{ème} semaine représentant la cinétique de l'étude nutritionnelle. Ces échantillons d'urine ont été conservés à -80 ° C. Il est à noter que le protocole expérimental a été approuvé par les institutions locales de protection des animaux

1.2. Préparation des échantillons

Dans notre étude le volume d'urine prélevé variait entre 0 et 150 µL. La variation de ce volume urinaire expliquée par la consommation d'eau de l'individu et/ou d'autres facteurs physiologiques, induit des variations de la concentration des métabolites endogènes. Il est donc nécessaire de normaliser les échantillons d'urine en fonction d'un dénominateur commun afin de minimiser la variation inter-individu engendrée par ces volumes. Il est courant de normaliser l'urine par rapport à un métabolite endogène, la créatinine. Pour réaliser la normalisation, tous les échantillons sont dilués de façon à obtenir la même concentration en créatinine (égale à la plus petite concentration). Avant toute dilution, il est important de vérifier la réponse de l'échantillon d'urine contenant la plus faible concentration en créatinine. En effet, si la valeur minimale en créatinine est trop faible, les profils métaboliques ne seront plus détectables, mais si la valeur minimale en créatinine est trop élevée, le détecteur sature. Dans le cadre de notre étude la concentration minimale en créatinine (112 mg/L) a permis de réaliser des dilutions de façon à obtenir des profils métabolomique optimaux. La concentration en créatinine a été mesurée par RMN.

1.3. Analyse des échantillons

482 échantillons d'urine préparés ont été analysés par un couplage de chromatographie liquide HPLC (Waters Alliance 2696) avec un spectromètre de masse (ESI / QToF Micro Waters). L'observation du poids des souris à la 23^{ème} semaine nous a permis de définir grâce à la médiane deux phénotypes: 50% de souris étaient obèses et 50% non obèses. Les échantillons ont été répartis dans 26 séries d'analyse de façon à équilibrer le facteur obésité. Au final, chaque série d'analyse contenait les échantillons de 4 souris de phénotypes différents et leurs cinétiques. (Les semaines de prélèvement).

2. Traitement des données

2.1. Choix des paramètres des logiciels d'extraction et alignement.

Les profils métaboliques ont été extraits et alignés par deux logiciels MarkerlynxTM et XCMS (Smith *et al.*, 2006). Il en résulte une **matrice de données** composée verticalement des mesures instrumentales (temps de rétention, masse, intensité) et horizontalement des échantillons sur lesquels ces variables ont été mesurées. Cette étape est cruciale car une mauvaise utilisation des paramètres des logiciels peut induire des biais. Par exemple :

- si le logiciel n'extrait pas un ion dans plusieurs échantillons alors que l'ion est détecté par l'instrument, cet ion peut devenir discriminant statistiquement.
- si le logiciel n'aligne pas correctement un ion présent dans tous les échantillons en raison du décalage du temps de rétention ou de masse, cet ion sera dédoublé dans la matrice de données. Ces deux ions auront des valeurs d'intensité mesurées complémentaires dans les échantillons. En effet, si l'ion A est présent dans l'échantillon 1, l'ion B en est absent. Par conséquent, pour chacun de ces ions les valeurs d'intensité mesurées dans chaque échantillon peuvent le rendre significatif par rapport au phénotype recherché.

Par conséquent, ce biais lié à l'étape d'extraction et d'alignement provoque des erreurs dans l'analyse statistique et l'interprétation biologique. C'est pourquoi, des critères de qualité doivent être définis pour les matrices de données obtenues.

2.1.1. Critères de qualité d'une matrice de données

Une matrice de données contient l'ensemble des ions extraits et alignés par le logiciel. Chaque ion est caractérisé par son temps de rétention, sa masse et la valeur de son intensité dans chaque échantillon. Pour permettre de caractériser une matrice de données, des critères ont été établis : (1) le nombre total d'ions extraits et alignés, (2) le nombre d'ions avec une valeur d'intensité moyenne supérieure à 5000, à 1000, à 500 et à 100, (3) le nombre d'ions avec une valeur d'intensité moyenne inférieure à 100, (4) le nombre d'ions présents dans plus de 95% des échantillons et (5) le nombre d'ions dédoublés. Par conséquent, les critères définis permettront d'évaluer l'effet des paramètres des logiciels sur les matrices de données. Dans la suite, nous nous intéresserons qu'aux paramètres dont l'influence sur l'extraction et l'alignement a été avérée dans des études préliminaires.

2.1.2. Les paramètres MarkerLynx™.

MarkerLynx™ est un logiciel d'extraction et d'alignement fourni par la société Waters avec le QToF Micro, qui utilise des fichiers bruts d'acquisition Waters dont l'extension du fichier est « .raw ». MarkerLynx™ fonctionne en deux étapes. Tout d'abord le logiciel extrait les pics chromatographiques définis par le temps de rétention dans chaque échantillon puis vérifie que la masse des pics existe dans au moins deux échantillons et les aligne. Des études préliminaires ont montré que cinq paramètres contrôlaient principalement l'extraction et l'alignement des données décrit ci-dessous. De plus, ces tests nous ont permis d'établir des valeurs optimales ou critiques pour chacun de ces paramètres. :

- le pourcentage de bruit de fond à éliminer dans le spectre chromatographique est défini automatiquement par un algorithme de MarkerLynx, (PPBN en anglais *Peak to Peak Baseline Noise*)
- les ions extraits doivent avoir une intensité supérieure ou égale à six fois l'intensité du bruit de fond du spectre de masse (NEL en anglais *Noise Elimination Level*)
- les ions extraits doivent avoir une largeur de pic de masse à 5% de la hauteur de 15 ou 30 secondes (PW5 en anglais *Peak Width at 5% Height*),
- l'intervalle de masse acceptable pour aligner les pics présents dans les échantillons est de 0.05 ou 0.1 Thomson (MW en anglais *Mass Window*)

- l'intervalle du temps de rétention acceptable pour aligner les pics présents dans les échantillons est de 0.4 ou 0.9 minutes. (RTW en anglais *Retention Time Window*)

Pour trois de ces cinq paramètres, deux valeurs critiques ont été testées. Pour pouvoir déterminer l'influence respective de ces trois paramètres, un plan factoriel décrit en Figure III. 2 a été réalisé.

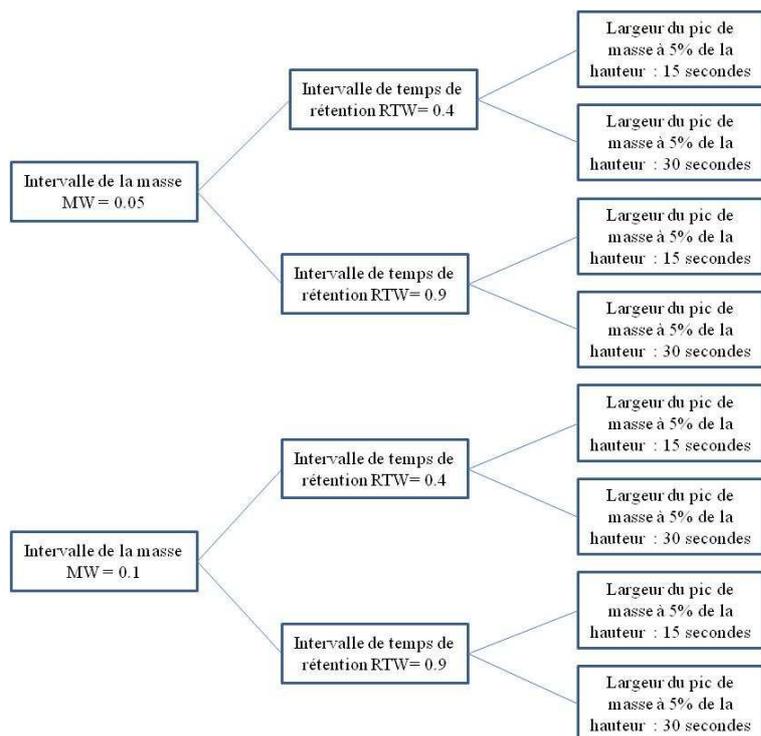


Figure. III.2 _ Plan factoriel appliqué aux paramètres Markerlynx sélectionnés et testés.

Les résultats de ce plan factoriel sont présentés dans la partie III-Résultats. Comme pour MarkerLynxTM, nous décrivons ci-dessous les paramètres de XCMS influençant l'extraction et l'alignement des données.

2.1.3. Les paramètres XCMS

XCMS est une bibliothèque du logiciel gratuit R (logiciel informatique). Les fichiers bruts d'acquisition issus du spectromètre de masse dont l'extension est « .raw » doivent être transformés en fichiers au format « NETcdf » d'extension « .cdf ». La conversion des fichiers se fait par le logiciel Databridge Waters.

Le logiciel XCMS fonctionne en quatre étapes caractérisées par des paramètres décrits ci dessous:

(1) **Extraction des pics** (appelé sous R la fonction `xcmsSet`): le logiciel permet le fractionnement du spectre de masse en différentes sections. Dans chacune de ces sections, l'extraction du pic sera effectuée si celui-ci est présent dans plusieurs balayages (scan) successifs. Afin de réaliser l'extraction des pics, certains paramètres doivent être fixés préalablement :

Fwhm représente la valeur minimale de la largeur à mi-hauteur d'un pic de masse pour qu'il soit extrait : 5-7 ou 12-14 secondes

Snthreshold représente la valeur seuil qui permet l'extraction d'un ion d'intensité supérieure à 5 fois le bruit de fond,

Step définit la largeur de la fenêtre de m/z pour l'extraction et la définition des pics chromatographiques de chaque ions (ion base peak chromatogram). Deux valeurs ont été testées : 0.05 ou 0.1.

(2) **Comparaison des pics extraits des différents échantillons et création de groupes** (appelé sous R la fonction `group`). C'est la phase d'alignement des pics à travers les échantillons. Un « group » est un ion qui est considéré comme identique dans un ensemble d'échantillons. Cet alignement est contrôlé par le paramètre `mzwid`.

Mzwid: qui représente la largeur acceptable entre les pics chromatographiques des ions extraits pour pouvoir être aligné. On a testé deux valeurs : 0.05 ou 0.25 seconde.

Minsamp représente le nombre d'échantillons dans lequel l'ion extrait doit être présent pour être aligné : 2 échantillons

Bw représente la largeur à mi-hauteur minimum du pic chromatographique extrait et aligné: 12-14 secondes.

(3) **Correction du temps de rétention** (appelé sous R la fonction `Retcor`). Dans l'ensemble des ions alignés, un certain nombre d'ions bien représentés dans la majorité des échantillons seront utilisés comme référence pour corriger la valeur du temps de rétention de l'ensemble des ions. Le logiciel recalcule le temps de rétention de tous les groupes. Cette fonction utilise

un paramètre nommé « missing » qui correspond au nombre maximum d'échantillons où le pic est manquant dans ces groupes de référence. Ce paramètre est fixé à 460.

(4) **Recherche des données manquantes** (appelé sous R la fonction Fillpeaks): Si un ion n'est pas extrait dans un échantillon, le logiciel retourne dans le chromatogramme et récupère la valeur d'intensité trouvée à l'endroit supposé du pic.

Comme précédemment, dans le but de déterminer l'influence de trois paramètres critiques, un plan factoriel a été réalisé. (Figure III. 3).

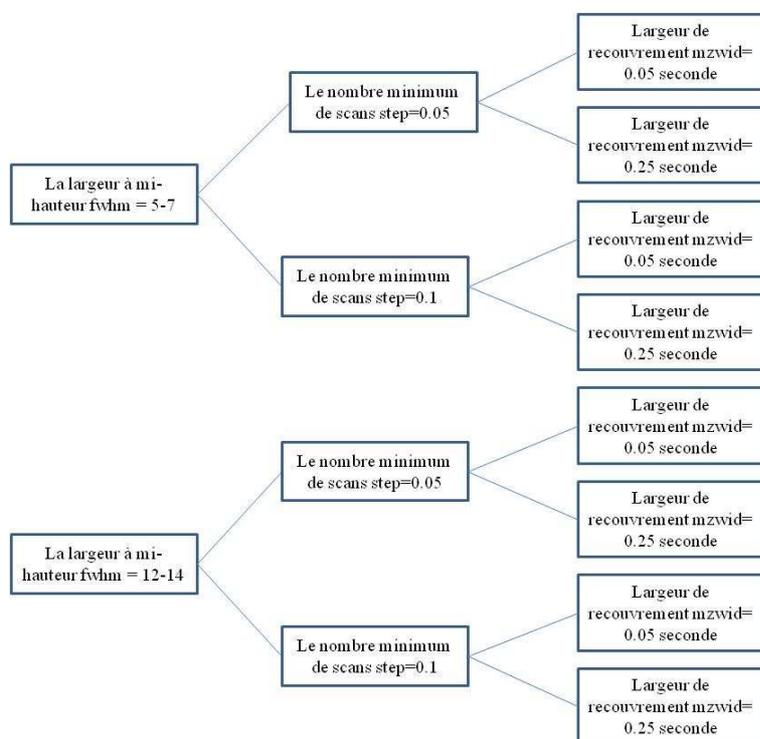


Figure. III.3 _ Plan factoriel appliqué aux paramètres XCMS sélectionnés et testés.

Les résultats du plan factoriel sont décrits dans la partie III.Résultats. Une fois cette étape réalisée, des méthodes de prétraitements seront appliquées aux données métabolomique.

2.2. Prétraitement des données

Cette étape de prétraitement des données permet de minimiser toutes les sources de variation introduites dans les données métabolomiques. Plusieurs phénomènes sont à l'origine de la complexité des données : la variation instrumentale des intensités d'un ion, l'écart des valeurs absolues entre les ions (en raison de la concentration et de leur capacité d'ionisation)

et la détection de polluants. Afin de corriger ces effets, et par conséquent d'améliorer le modèle de prédiction statistique (PLS-DA), les données d'intensité sont transformées par des méthodes de prétraitement : la normalisation, la transformation logarithme ou les prétraitements statistiques. Sélectionner préalablement des ions où l'effet du phénotype est significatif par une analyse de variance peut permettre d'améliorer le modèle de prédiction. (Smilde *et al.*, 2005). Toutes ces méthodes seront testées individuellement ou successivement.

Nous avons par conséquent appliqué ces méthodes de prétraitement sur chacune des matrices de données XCMS et MarkerLynxTM présentées ci dessus.

2.2.1. Correction des effets séries d'analyse et échantillons.

Dans la mesure où les échantillons biologiques ont été répartis de façon équilibrée dans les séries d'analyse, la variation de l'intensité d'un ion est supposée proportionnelle à l'intensité totale des ions dans un échantillon ou dans une série d'analyse.

Afin de corriger l'effet échantillon, les intensités brutes I de l'ion q dans l'échantillon p ont été transformées à l'aide de la formule suivante:

$$X_{pq} = I_{pq} / \sum_p I_q$$

Afin de corriger l'effet série d'analyse, l'intensité brute I de l'ion q détecté dans l'échantillon p injecté dans la série d'analyse k ont été transformées à l'aide de la formule suivante:

$$X_{pqk} = I_{pqk} / \sum_{q,k} I_{qk}$$

2.2.2. Transformation logarithme. (Snipen *et al.*, 2009)

La transformation des données en logarithme est généralement appliquée pour rendre symétrique la distribution des données d'intensité et corriger la variation statistique intrinsèque au modèle utilisé appelée hétéroscédasticité (Kvalheim et Brakstad, 1994). Cette étape est importante lors des analyses multivariées. En générale, la transformation logarithme est conseillée lorsque les données présentent des valeurs de variances hétérogènes. Pour obtenir des données transformées en logarithme la formule suivante est appliquée:

$$\tilde{y}_{ij} = \log_{10} (y_{ij})$$

2.2.3. Sélection des ions significatifs pour l'effet obésité.

Chaque matrice de données obtenue a été analysée par l'analyse de covariance (ANCOVA) sur mesures répétées en utilisant un modèle mixte. Ce modèle mixte permet de décomposer les différents effets susceptibles d'avoir une influence sur les données. Ainsi, nous pourrions déterminer la significativité d'un effet pour chaque ion. Un programme a été réalisé sur le logiciel SAS version 9.1 (SAS Institute Cary, NC) pour chaque ion.

$$x_{ijk} = x_0 + \alpha_i + \alpha_j + (\tilde{\alpha}\alpha)_{ij} + \alpha_k + (\tilde{\alpha}\alpha)_{ik} + \alpha_{s(i)} + \alpha_{ijsk}$$

- x_0 est la valeur du métabolite au départ;
- α_i est l'effet du phénotype;
- α_j est l'effet du régime ;
- $(\tilde{\alpha}\alpha)_{ij}$ est l'interaction entre le phénotype et le régime ;
- α_k effet de la série d'analyse ;
- $(\tilde{\alpha}\alpha)_{ik}$ est l'interaction entre le phénotype et la série d'analyse ;
- $\alpha_{s(i)}$ est l'effet aléatoire.
- α_{ijsk} est l'erreur aléatoire.

Tous les ions présentant un effet phénotype significatif ou une interaction phénotype-cinétique significatif ont été sélectionnés pour l'analyse par PLS-DA.

2.2.4. Prétraitement statistique des données

Les transformations liées au prétraitement statistique permettent de stabiliser les variances des données. De nombreuses méthodes ont été décrites par Vandenberg *et al.* (2006). Nous avons testés les méthodes qui avaient le plus d'influence sur le modèle statistique PLS-DA.

Centrage

Centrer les données permet d'ajuster et de comparer des métabolites dont les intensités moyennes sont très différentes. Ainsi, les différences entre les données seront plus visibles que les similitudes. Pour obtenir des données centrées \tilde{y}_{ij} , chaque intensité brute d'un ion y_{ij} est soustraite par la valeur moyenne de l'intensité \bar{y}_i de cet ion pour tous les échantillons à l'aide de la formule suivante: $\tilde{y}_{ij} = y_{ij} - \bar{y}_i$

Centrage et réduction (UV en anglais *Unit Variance*)

Centrer et réduire les données permet de comparer des données d'échelles variées. Cette technique est utilisée pour donner le même poids à des ions de moyenne et de variance différente. Pour obtenir des données centrées et réduites \tilde{y}_{ij} , chaque intensité brute d'un ion y_{ij} est soustraite par la valeur moyenne de l'intensité \bar{y}_i et divisé par l'écart type de cet ion S_i à l'aide de la formule suivante: $\tilde{y}_{ij} = (y_{ij} - \bar{y}_i) / S_i$

Afin de déterminer la meilleure combinaison de méthodes de prétraitement sur les données biologiques et ainsi déterminer le meilleur modèle de prédiction du phénotype obésité, toutes les matrices de données transformées sont traitées par PLS-DA, une analyse statistique discriminante supervisée.

2.3. Validation du modèle PLS-DA.

Les différentes options de traitement choisies ont été évaluées grâce à la qualité du modèle PLS-DA à discriminer le phénotype obésité. Pour améliorer le modèle de discrimination PLS-DA, une filtration (OSC) des données a été réalisée afin d'éliminer toutes les variations indépendantes de l'effet phénotype. (en anglais OSC *Orthogonal Signal Correction*).

Dans la mesure où le nombre de variables prédictives est largement supérieur au nombre d'individus, il est toujours possible de trouver des variables qui discriminent parfaitement les groupes. Il est donc important de valider les modèles de prédiction des données. (Westerhuis *et al.*, 2008) Il faut donc s'assurer que les composantes trouvées sont robustes aux changements des échantillons. Deux critères permettent d'évaluer la qualité de l'ajustement et la capacité de prédiction du modèle :

(1) la validation croisée, qui consiste à scinder les données en deux groupes. Le groupe d'apprentissage permet de calculer un modèle de discrimination tandis que le groupe de validation, qui ne participe pas au calcul du modèle, sert à le valider et à en estimer la qualité de prédiction. Deux paramètres sont utilisés pour estimer la qualité du modèle : R^2Y qui rend compte de la capacité explicative (jeux d'apprentissage) et le Q^2Cum qui rend compte de la qualité prédictive du modèle.

(2) le logiciel SIMCAP que nous avons utilisé propose un test de permutation, qui consiste à réaffecter aléatoirement les individus dans les lots de phénotypes afin de vérifier que le modèle obtenu n'est pas dû au hasard.

Plus la valeur de Q^2_{Cum} est proche de 1, plus le modèle est apte à prédire correctement les phénotypes étudiés. Toutefois, il est possible que les variables discriminent parfaitement les groupes par hasard. C'est pourquoi il est nécessaire de confirmer cette valeur de Q^2_{Cum} par un test de permutation qui consiste à réaffecter au hasard les échantillons dans les deux classes de phénotype. Le modèle est considéré comme valide si on observe une dégradation des indices R^2Y et Q^2 pour les échantillons permutés. Un modèle de validation correcte signifie que la valeur de R^2Y est comprise entre 0.2 et 0.3 et que la valeur de Q^2 est inférieure à 0. (Figure III. 4).

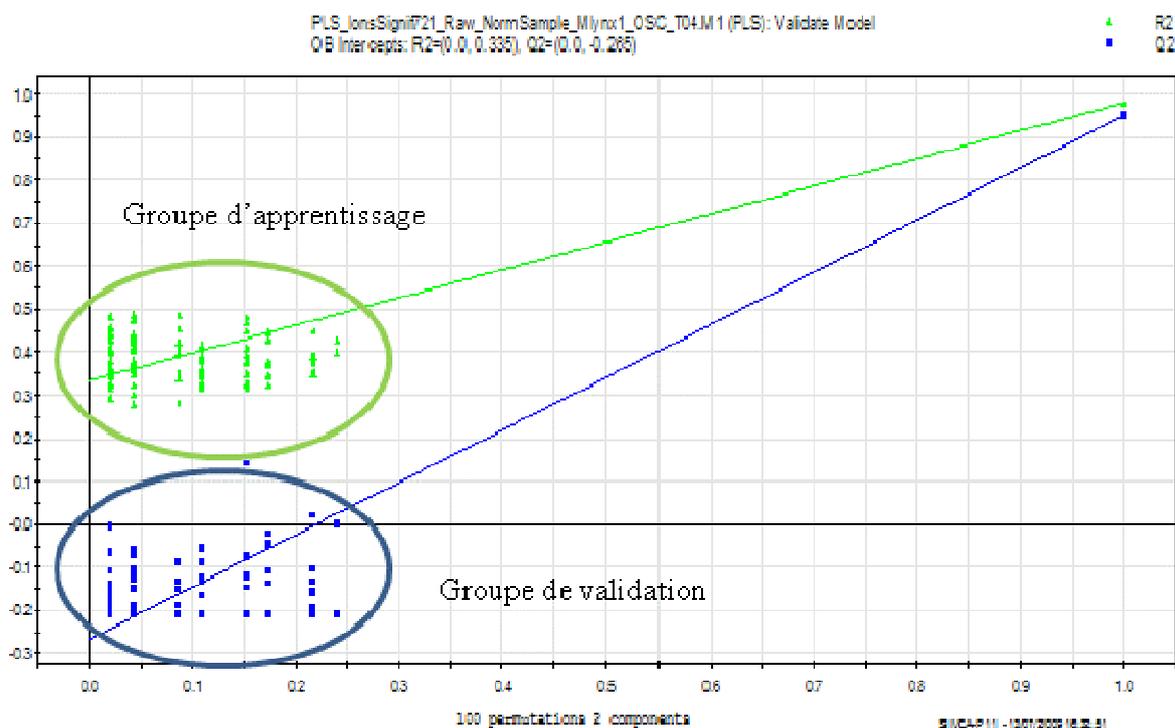


Figure III.4 Validation du modèle OSC-PLS-DA par SIMCA-P version 11. Pour les échantillons avec individus permutés (à gauche) on obtient $R^2Y= 0.335$ et $Q^2=-0.265$. Ces valeurs sont dégradées par rapport à l'échantillon normal correspondant à la valeur d'abscisse 1. ceci montre que le modèle n'est pas obtenu au hasard.

III. Résultats

1. Choix des paramètres des logiciels MarkerLynx™ et XCMS.

Pour déterminer l'influence des paramètres MarkerLynx™ et XCMS sur la qualité de la matrice de données, deux plans factoriels ont été réalisés. (« Matériel et Méthodes »). Un test d'analyse de variance (ANOVA en anglais *Analysis of Variance*) est appliqué pour mesurer l'influence des paramètres des logiciels sur la matrice de données, sur les valeurs des critères d'évaluation. (Tableau III. 1). Dans le Tableau III. 1, pour chaque critère observé, nous regardons si le paramètre MarkerLynx testé par ANOVA est significatif. Cette significativité est observée si la valeur de probabilité du test est inférieure à 0.05.

<i>Paramètres testés et critères observés</i>	<i>Total des ions détectés</i>	<i>Ions d'intensité moyenne supérieure à 5000</i>	<i>Ions présents dans plus de 95% des échantillons</i>	<i>Ions dédoublés</i>
PW5	0.310	0.116	0.039	0.752
RTW	0.0001	0.116	0.0001	0.0001
MW	0.0001	0.016	0.011	0.197

Tableau. III.1 _ Influence des paramètres MarkerLynx™ sur les critères d'évaluation de la matrice de données observés. L'ANOVA nous permet d'avoir une valeur de probabilité. Si cette valeur est inférieure à 0.05, le facteur testé est significatif. (**en gras**).

Par conséquent, la largeur du pic à 5% de la hauteur (PW5) est significative sur le nombre d'ions présents dans plus de 95% des échantillons. L'intervalle du temps de rétention (RTW) et l'intervalle de la masse (MW) sont significatifs pour trois critères. L'intervalle du temps de rétention (RTW) a une influence sur l'alignement des ions tandis que la largeur du pic à 5% de la hauteur (PW5) et l'intervalle de la masse exacte (MW) ont une influence sur l'extraction des ions.

Les problèmes d'alignement induisent une augmentation du nombre d'ions dédoublés et une diminution du nombre d'ions présents dans plus de 95% des échantillons. Ainsi, les valeurs des paramètres sélectionnés doivent permettre de minimiser le nombre d'ions dédoublés et de

maximiser le nombre d'ions présents dans plus de 95% des échantillons et le nombre d'ions avec une intensité moyenne supérieure à 1000. La Figure III. 5 représente le nombre d'ions calculés pour chaque combinaison de valeur des paramètres de MarkerLynx™.

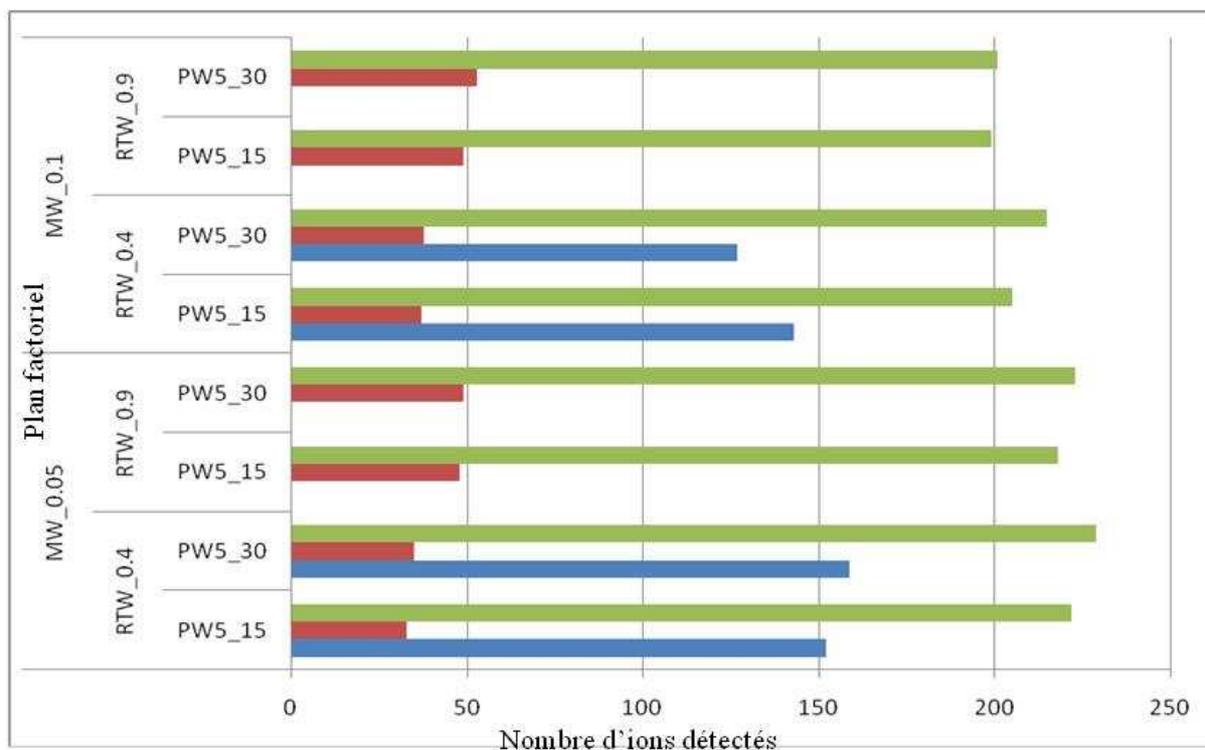


Figure. III.5 _ Plan factoriel appliqué aux paramètres MarkerLynx sélectionnés et testés. Cette figure représente en vert, le nombre d'ions d'intensité moyenne supérieure à 1000, en rouge, le nombre d'ions présents dans plus de 95% des échantillons et en bleu, le nombre d'ions dédoublés.

D'après la Figure III. 5, les paramètres optimaux permettent d'obtenir 0 ions dédoublés, 53 ions présents dans 95% des échantillons et 201 ions d'une intensité moyenne supérieure à 1000. Les valeurs des paramètres optimaux sont 30 secondes pour la largeur du pic de masse à 5% de la hauteur de, 0.1 Thomson d'intervalle de masse et 0.9 minutes d'un intervalle de temps de rétention.

Une démarche similaire a été effectuée pour les paramètres de XCMS. Le plan factoriel a été analysé par une analyse de variance (Tableau III. 2).

<i>Paramètres testés et critères observés</i>	<i>Total des ions détectés</i>	<i>Ions d'intensité moyenne supérieure à 5000</i>	<i>Ions présents dans plus de 95% des échantillons</i>	<i>Ions dédoublés</i>
fwhm	0.153	0.561	0.001	0.004
step	0.067	1	0.046	0.337
mzwid	0.189	0.034	0.843	0.401

Tableau. III.2 _ Influence des paramètres XCMS sur les critères d'évaluation de la matrice de données observés. L'ANOVA nous permet d'avoir une valeur de probabilité. Si cette valeur est inférieure à 0.05, le facteur testé est significatif. (**en gras**).

Le Tableau III. 2 montre que les paramètres du nombre de balayage (step) et de la largeur de recouvrement tolérée entre les pics chromatographiques des ions extraits dans chaque échantillon (mzwid) ont une influence sur l'extraction des pics alors que la largeur à mi-hauteur d'un pic de masse (fwhm) a une influence sur l'extraction et l'alignement des pics. Pour déterminer les meilleures valeurs des trois paramètres testés, il est nécessaire de comparer les valeurs absolues obtenues pour chaque critère. **D'après la Figure III. 6**, pour XCMS, la combinaison qui donne le moins d'ions dédoublés (0), le plus d'ions présents dans 95% des échantillons (136) et le plus d'ions d'intensités moyenne supérieure à 1000 coups par seconde (198) est une largeur à mi-hauteur (fwhm) de 12-14, un fenêtrage de m/z (step) fixé à 0.1 et une largeur de chevauchement (mzwide) de 0.25.

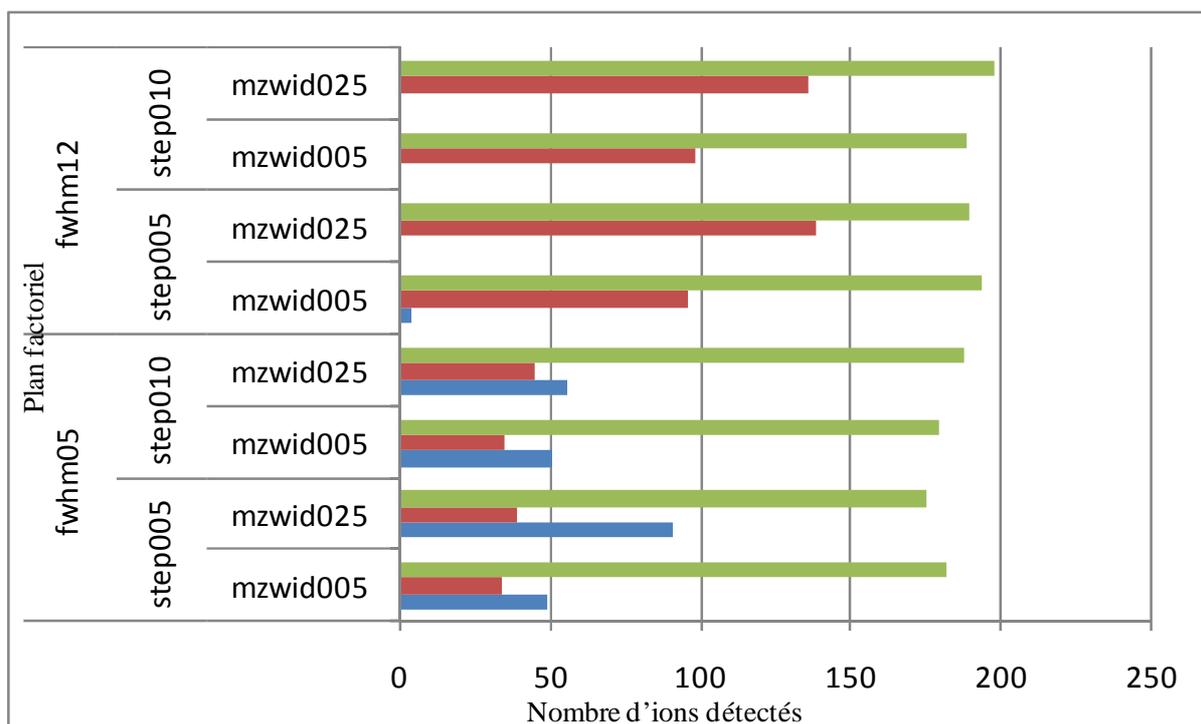


Figure. III.6 _ Plan factoriel appliqué aux paramètres XCMS sélectionnés et testés. Cette figure représente en vert, le nombre d'ions d'intensité moyenne supérieure à 1000, en rouge, le nombre d'ions présents dans plus de 95% des échantillons et en bleu, le nombre d'ions dédoublés.

Après optimisation des paramètres des logiciels d'extraction et d'alignement, il est nécessaire d'étudier les matrices de données.

2. Etude des matrices de données optimisées

Pour évaluer la qualité des matrices de données obtenues, des critères sont utilisés: les valeurs des intensités extraites pour un ion et la somme des intensités totales des ions dans chaque série d'analyse.

2.1. Comparaison des valeurs des intensités obtenues par MarkerlynxTM et XCMS

Pour comparer les valeurs d'intensité extraites par les deux logiciels, nous avons suivi un métabolite urinaire connu, l'acide hippurique dont l'intensité moyenne est de l'ordre de 500 et un ion m/z 304,29 non identifié d'intensité moyenne de plus de 1000 coups par seconde.

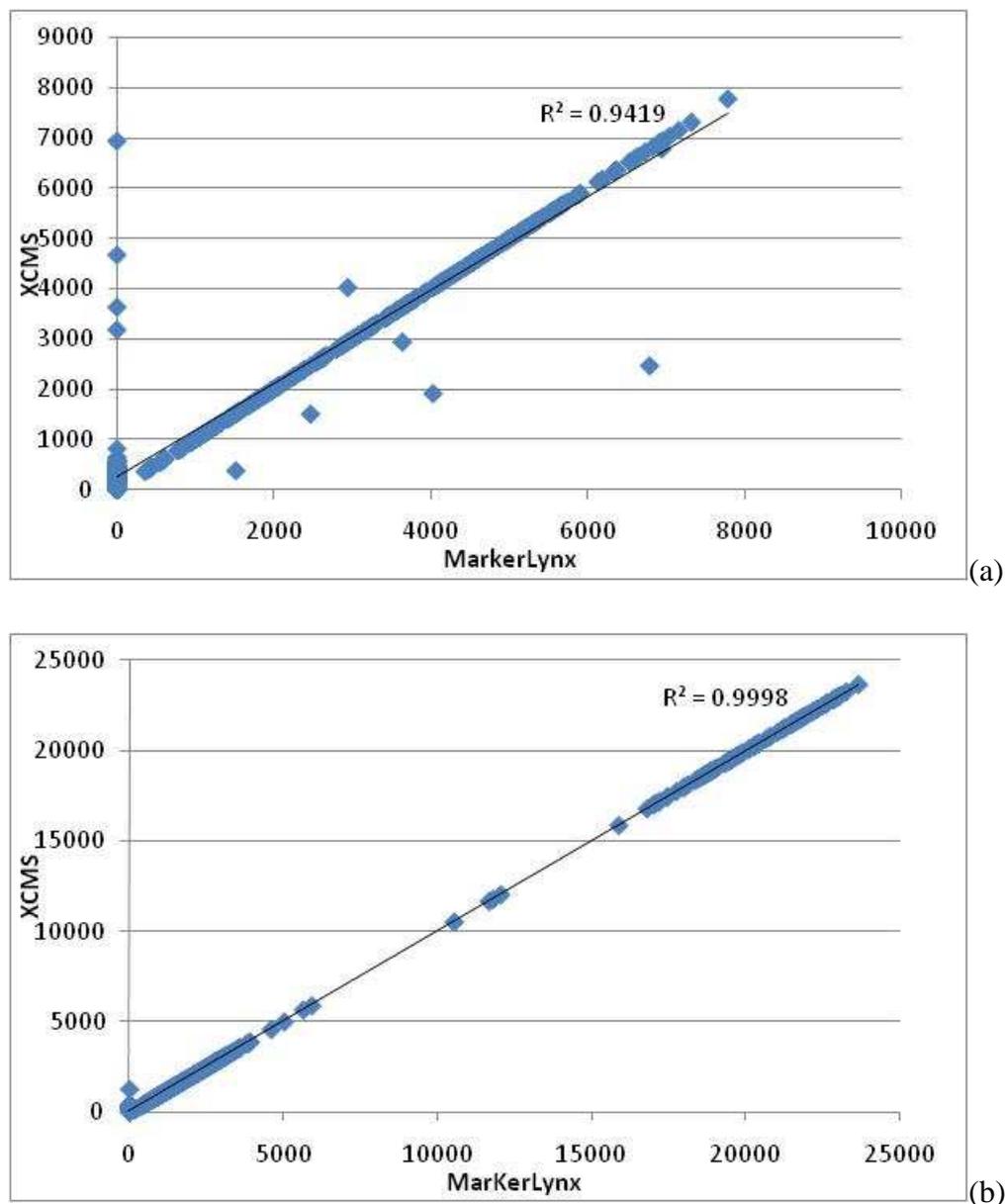


Figure. III.7 _ Corrélation entre les valeurs d'intensité (coups par seconde) extraites par Markerlynx (vertical) et XCMS (horizontal). **(a)** extraction des intensités de l'**acide hippurique**. **(b)** extraction des intensités d'un ion (m/z 304.29) d'intensité moyenne supérieure à 1000.

La Figure III. 7 montre la corrélation entre les valeurs d'intensité d'un ion mesurées par ces deux logiciels. Les valeurs d'intensités mesurées par les logiciels sont identiques pour des ions d'intensité moyenne ($R^2= 0.94$, pour l'acide hippurique) mais aussi pour des ions de plus hautes intensités ($R^2= 0.99$, pour l'ion m/z 304.29).

Néanmoins, il existe une limitation à cette analyse : en effet, plus la moyenne de l'intensité des pics est faible, plus MarkerLynx détecte des intensités nulles induisant une diminution de la valeur de R^2 . Cette corrélation n'est donc pas vérifiée pour les ions dont l'intensité moyenne est inférieure à 50.

2.2. Comparaison des valeurs des intensités obtenues par MarkerlynxTM et XCMS et dans les 26 séries d'analyse

L'objectif est de vérifier que la variation des intensités des différentes séries n'est pas influencée par les logiciels d'extraction et d'alignement. Dans la mesure où les souris ont été réparties de façon à équilibrer le phénotype obésité dans les 26 séries d'analyse, l'intensité totale calculée pour chaque série d'analyse doit être égale. Les Figure III. 8 représentent la somme totale des intensités mesurées par les logiciels dans une série d'analyse.

D'après la Figure III. 8 (a), la valeur de la somme des intensités de tous les ions extraits par les deux logiciels varie dans chaque série d'analyse. Nous observons qu'à partir de la série d'analyse numéro 6, la somme des intensités des ions extraits par XCMS est plus grande que celle de MarkerLynxTM. Pour les ions d'intensité moyenne supérieure à 5000, la somme des intensités mesurées par MarkerLynxTM est toujours supérieure à celle de XCMS. (Figure III. 8 (b)). Malgré ces différences, les profils de la somme des intensités extraites par MarkerlynxTM et XCMS sont similaires. Ces résultats permettent de vérifier que l'effet série n'est pas influencé par les logiciels.

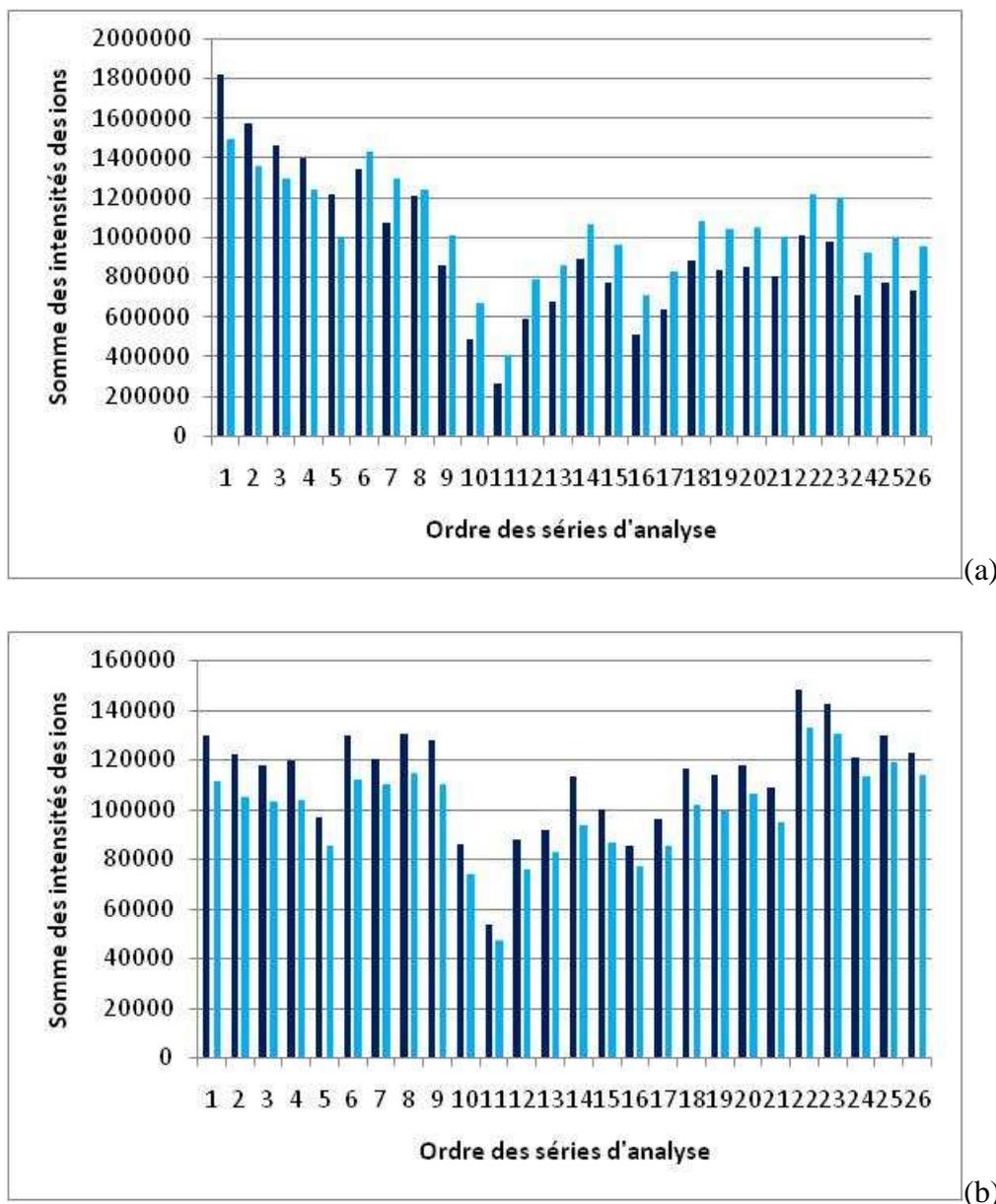


Figure III.8 _ Somme des intensités des ions extraits par MarkerLynx (bleu foncé) et XCMS (bleu clair) dans chaque série d'analyse. (a) Somme des intensités de tous les ions. (b) Somme des intensités des ions dont la moyenne de l'intensité est supérieure à 5000.

Toutefois, cette variation des intensités entre les séries d'analyse existe et influence l'analyse statistique et pourrait induire une interprétation biologique erronée. Par conséquent, il est important de minimiser ces variations dans le jeu de données. Pour se faire, les méthodes de prétraitements décrites au paragraphe matériel et méthode ont été appliquées suivant un plan factoriel. (Figure III. 1).

3. Choix des méthodes de prétraitements appliquées à l'OSC-PLS-DA.

Les différents prétraitements de données appliqués aux intensités ont été évalués par la qualité du modèle statistique OSC-PLS-DA. Sept méthodes de prétraitement ont été appliquées à l'ensemble des intensités issues de MarkerLynxTM ou XCMS suivant le plan factoriel. (Figure III. 1). Ce plan est appliqué lorsqu'on dispose de plusieurs répétitions pour chaque niveau de facteurs (méthodes de prétraitements). Le bénéfice est de pouvoir tester l'interaction entre les facteurs. L'analyse de la variance (ANOVA) permet de mesurer l'influence de chaque facteur indépendamment. (Tableau III. 3). Les variances de chacun des facteurs sont testées à l'aide d'un test de Fischer. Si $F < F_{\text{théorique}}$ et que p-value est inférieure à 0.05 alors le traitement appliqué a un effet significatif sur le Q^2_{cum} . (Tableau III. 3 colonne 2).

<i>Traitements appliqués</i>	<i>F(p-value)</i>	<i>Niveau 1</i>	<i>Q²_{Cum}</i>	<i>Niveau 2</i>	<i>Q²_{Cum}</i>	<i>Niveau 3</i>	<i>Q²_{Cum}</i>
Sélection des ions significatifs pour l'effet obésité	37.4 (0.0001)	Sélection des ions	0.70	Tous les ions	0.40		
Prétraitement statistique	8.8 (0.006)	Centré	0.62	Centré et réduit	0.48		
Correction des biais techniques	4.6 (0.02)	NormSerie	0.64	SSNorm	0.55	NormEchan	0.46
Logiciels	1.8 (ns)	XCMS	0.58	MarkerLynx	0.52		
Transformation logarithmique	0.6 (ns)	Données Log	0.57	Données brutes	0.53		

Tableau III. 3_ Résultats de l'analyse de variance appliquée indépendamment pour chaque traitement du plan factoriel. L'analyse de la variance nous donne deux valeurs : F (test de Fischer) et la p-value. Si la p-value < 0.05, alors le traitement est significatif sur le Q^2_{cum} . Chacune de ces matrices modifiées sont analysées par OSC-PLS-DA et validée par Q^2_{cum} . Plus cette valeur est proche de 1, plus le modèle est prédictif.

D'après le Tableau III. 3, trois facteurs ont un effet significatif sur le Q^2_{cum} : la sélection des ions significatifs, le prétraitement statistique et la correction des biais techniques. Chaque matrice transformée par la méthode de prétraitement a été analysée par

OSC-PLS-DA. La qualité du modèle de classification est vérifiée par la validation croisée dont l'indicateur est Q^2_{Cum} et l'aptitude qu'a le modèle à bien classer le phénotype dont l'indicateur est R^2Y . (Matériel et Méthodes 2.3).

Dans le Tableau III. 3, la valeur de Q^2_{Cum} validée et observée pour chaque matrice de données testée par le modèle OSC-PLS-DA a été relevée. La matrice de données contenant les ions sélectionnés par l'analyse de variance donne le meilleur modèle de prédiction. ($Q^2_{Cum} = 0.7$). Suite à ces résultats le modèle présenté basé sur l'étude des effets simples ne nous permet pas de conclure sur la combinaison des méthodes de prétraitements. Le plan factoriel permet de réaliser l'analyse de l'interaction des prétraitements nécessaire pour choisir la combinaison de méthode à appliquer aux données. (Tableau III. 4). Les résultats de l'analyse de variance appliquée à deux facteurs nous donnent la valeur de F et de p-value. Ensuite pour chaque interaction, le modèle de classification OSC-PLS-DA est appliqué. Enfin le test Posthoc nous permet de voir le facteur responsable de la discrimination des groupes.

Dans le Tableau III. 4, l'analyse de la variance nous permet de montrer que trois interactions sont significatives : l'effet du logiciel et de la transformation logarithme, l'effet du logiciel et du prétraitement des données et l'effet des méthodes de sélection des ions et de correction du biais technique. Bien que l'effet simple du logiciel ne soit pas significatif s'il est combiné avec une transformation logarithmique ou un prétraitement statistique, les deux effets deviennent significatifs. Afin d'obtenir le meilleur modèle de prédiction, il est préférable de sélectionner au préalable les ions significatifs pour l'effet obésité. Dans ce cas, la correction des biais techniques n'est pas utile. Dans le cas inverse, si aucune sélection n'est effectuée, c'est la normalisation par séries d'analyses qui donne les meilleurs résultats (0.57). Si on considère l'interaction logiciel et transformation logarithmique, alors les matrices de données issues de XCMS et transformées en logarithme (0.67) ou les données issues de MarkerLynxTM et centrée (0.65) permettent d'obtenir les meilleurs modèles de prédiction. Si on considère l'interaction logiciel et prétraitement statistique, c'est Markerlynx centré normé (0.39) qui se distingue et donne les moins bons résultats.

<i>Traitement 1</i>	<i>Traitement 2</i>	<i>F(p-value)</i>	<i>Interactions</i>	<i>Q² mean</i>	<i>Posthoc test</i>
Logiciel	Transformation logarithme	8 (0.008)	XCMS-log	0.67	a
			Mlynx-brutes	0.57	ab
			XCMS-brutes	0.50	b
			Mlynx-log	0.47	b
Logiciel	Prétraitement statistique	6.2 (0.02)	Mlynx cent	0.65	a
			XCMS cent	0.59	a
			XCMS UV	0.57	a
			Mlynx UV	0.39	b
Sélection des des ions significatifs pour l'effet obésité	Correction des biais technique	4.5 (0.02)	Ions signif – Norm Serie	0.71	a
			Ions signif – Norm échant	0.70	a
			Ions Signif – SsNorm	0.68	a
			Tous ions – Norm Série	0.57	ab
			Tous ions – SsNorm	0.43	b
			Tous ions – Norm échant	0.22	c

Tableau. III.4 _ L'analyse de la variance permet aussi de mettre en évidence les interactions significatives ($p\text{-value} < 0.05$) entre les différentes méthodes de traitements appliqués à la matrice de données. Pour chaque interaction, la valeur de Q^2_{Cum} confirmée par la validation du modèle devient un indicateur de l'OSC-PLS-DA. Le test Post-hoc permet d'isoler les combinaisons de méthodes de prétraitement qui optimise le Q^2_{Cum} . Les valeurs de Q^2_{Cum} présentant une même lettre ne sont pas significativement différentes (UV *Unit Variance*, centrage et réduction).

IV. Discussion

1. Choix du paramètre du signal

Deux paramètres peuvent caractériser le signal en spectrométrie de masse : la hauteur du pic chromatographique et l'aire sous la courbe de ce même pic. La concentration d'un composé est estimée grâce à la valeur de l'aire sous la courbe du pic chromatographique. Or, la réponse obtenue par le spectromètre de masse (QToF) montre que pour les signaux intenses ($I > 500$) les valeurs de hauteur de pic sont proportionnels aux valeurs d'aire. A l'inverse, pour les signaux de faible intensité ($I < 50$), cette proportionnalité n'existe plus en raison de la forme non gaussiennes des pics chromatographiques de faible intensité. Ainsi, la valeur de l'aire sous la courbe pour ces pics sera fausse contrairement à la hauteur du pic masse. De plus, dans cadres des études en nutrition, l'observation des métabolites de faible concentration est importante, c'est pourquoi, le paramètre choisi pour l'analyse les données de métabolomique ne sera pas l'aire mais la hauteur du pic. (Gika *et al.*, 2008)

2. Optimisation des paramètres des logiciels

Les paramètres testés et choisis dépendent de la qualité du chromatogramme ainsi que de l'empreinte spectrale des échantillons analysés. Pour fixer correctement les paramètres des logiciels d'extraction et d'alignement, il est important de vérifier les valeurs du bruit de fond du chromatogramme et du spectre de masse.

Dans notre étude les paramètres ont été optimisés pour des données obtenues par HPLC-MS. Néanmoins, pour chaque technique analytique ou matrice biologique, ces paramètres devront être optimisés. Pour XCMS, l'étude des différents paramètres montre que la valeur de la largeur du pic à mi-hauteur (fwmh) est une valeur clef pour ce logiciel aussi bien pour l'extraction que pour l'alignement. Pour MarkerLynxTM, cinq paramètres auront une influence, le pourcentage de bruit de fond à éliminer dans le spectre chromatographique, la valeur de seuil du bruit de fond à éliminer dans le spectre de masse, la valeur minimale de la largeur du pic de masse à 5% de la hauteur (PW5 en anglais *Peak Width at 5% Height*), l'intervalle de masse (MW en anglais *Mass Window*) et l'intervalle du temps de rétention (RTW en anglais *Retention Time Window*).

L'optimisation de ces paramètres permet d'optimiser la matrice de données selon les critères établis.

3. Evaluation de la qualité des matrices de données obtenues.

Les deux logiciels étudiés donnaient des données similaires en termes d'intensité des ions détectés. Toutefois, Markerlynx™ détecte beaucoup plus d'ions que XCMS quelque soit les paramètres utilisés : 38% des ions sont communs aux deux logiciels, 19% sont spécifiques à XCMS et 47% sont spécifiques à MarkerLynx™. Cette différence peut être expliquée par la présence d'ions de faible intensité détectés exclusivement par MarkeLynx™. Le fait d'extraire plus d'ions peut être un atout dans les études nutritionnelles à condition de d'éliminer les signaux correspondant à des polluants, des artefacts ou du bruit de fond.

D'après les Figures III. 5 et 6, l'optimisation des paramètres des deux logiciels donne un nombre d'ions présents dans 95% des échantillons différent entre MarkerLynx™ (53) et XCMS (136). Cette différence est liée à une limitation de MarkerLynx™. Même si MarkerLynx™ présente des avantages en termes de détection des ions de faible intensité, lorsqu'un pic n'existe pas ou qu'il est de faible intensité, le logiciel lui attribue une valeur d'intensité nulle. Ainsi, la matrice de données MarkerLynx™ contient 60% de valeurs nulles alors que la matrice de données XCMS n'en contient aucune grâce à l'étape de recherche des données manquantes. Cette différence explique les valeurs obtenues dans les Figure III. 5 et 6. Dans la mesure où cette valeur nulle perturbe l'analyse statistique, il est nécessaire de remplacer systématiquement cette valeur par une valeur égale à 95% de la valeur minimale (différente de 0) d'intensité trouvé pour cet ion dans les échantillons biologiques. (Gika *et al.*, 2008)

Afin de définir la notion de matrice de données « propres » (Van Den Bergh *et al.*, 2006) nous avons élaboré certains critères d'évaluation. En outre, pour aider à la décision, il a fallu définir des seuils acceptables pour chaque critère. Ainsi, une matrice de donnée « propre » signifie qu'elle n'a aucun ion dédoublé et que plus de 50 ions sont présents dans plus de 95% des échantillons.

4. Prétraitement des données

En métabolomique, les signaux détectés par HPLC-MS contiennent aussi bien l'information biologique (les métabolites présents dans l'échantillon) que des polluants ou du bruit. Le meilleur modèle OSC-PLS-DA est obtenu en sélectionnant les ions dont l'effet du phénotype est significatif, ainsi les signaux correspondant au bruit (faible intensité) ou aux

polluants sont éliminés. En effet, contrairement à l'analyse de covariance sur mesures répétées en utilisant un modèle mixte, le modèle OSC-PLS-DA prend en compte toutes les données sans éliminer celles issues des variations multifactorielles liées au design expérimental (variation instrumentale, cinétique du régime). Cette élimination préalablement effectuées permet d'améliorer le modèle de prédiction (Smilde *et al.*, 2005 et Jansen *et al.*, 2005).

D'après le Tableau III. 4, le prétraitement statistique est lié au logiciel. Pour XCMS, il n'y a pas de différence entre centrage et centrage réduction (UV). Par contre pour MarkerLynxTM, le centrage réduction diminue significativement le Q^2_{cum} . Ceci s'explique par le fait que MarkerLynxTM extrait plus d'ions que XCMS (Environ 33% en plus). Ces ions sont de faibles intensités et appartiennent en partie au bruit de fond. Le centrage réduction homogénéise toutes les valeurs d'intensité ainsi, le bruit d'intensité plus faible, aura le même poids statistique que les ions intenses. Ce bruit spectral va perturber l'analyse statistique et donner un mauvais modèle de discrimination.

Transformer les intensités en logarithme (Van Velzen *et al.*, 2008) n'améliorent pas le modèle de prédiction appliqué à la matrice de données MarkerLynxTM contrairement à XCMS. La transformation logarithme comme la normalisation statistique qui consiste à centrer et réduire les données tend à homogénéiser les intensités de tous les ions et donc augmenter l'importance du bruit spectral plus abondant dans les pics extraits par MarkerLynxTM. En conséquence, le pouvoir discriminant du modèle (Q^2_{cum}) se trouve diminué.

Les deux logiciels montrent clairement qu'entre les séries d'analyse, l'intensité globale des ions extraits varie. La normalisation de ces intensités par la moyenne des intensités des ions dans chaque série d'analyse permet de palier en partie à ce problème si on ne sélectionne pas les ions par analyse de la variance. Néanmoins la correction du biais technique liée à l'échantillon n'améliore pas le modèle car celle-ci prend en compte en plus de la variation de la réponse instrumentale, la variation biologique intrinsèque à chaque échantillon.

V. Conclusion

Dans cette étude, nous avons développé une **stratégie pour le traitement des données métabolomiques**. Cette stratégie consiste à s'assurer que l'extraction et l'alignement des données est optimale et que les résultats obtenus puisse révéler l'information biologique. Des critères d'évaluation d'une matrice de données sont proposés pour permettre l'optimisation des paramètres des logiciels. D'après les critères testés, **les deux logiciels ont des performances équivalentes**. Pour XCMS, il faut optimiser la valeur de la largeur à mi-hauteur d'un pic de masse (fwhm). Pour MarkerLynxTM cinq paramètres doivent être optimisés pour chaque jeu de données : le pourcentage de bruit de fond à éliminer dans le spectre chromatographique (PPBN en anglais *Peak to Peak Baseline Noise*), la valeur de seuil du bruit de fond à éliminer (NEL en anglais *Noise Elimination Level*), la valeur minimale de la largeur du pic de masse à 5% de la hauteur (PW5 en anglais *Peak Width at 5% Height*), l'intervalle de masse (MW en anglais *Mass Window*) et l'intervalle du temps de rétention (RTW en anglais *Retention Time Window*). Le rapport signal sur bruit des pics chromatographiques et de masse ainsi que la largeur de ces pics seront des indicateurs forts pour déterminer les valeurs des paramètres d'extraction acceptables. Pour l'alignement, la variation du temps de rétention entre les échantillons analysés est la référence.

Les différents prétraitements appliqués aux matrices de données ont montré que cette étape pouvait améliorer la visualisation de la variation biologique dans le cas d'une intervention nutritionnelle. Les matrices issues des différents traitements ont pu être comparées par un modèle statistique de discrimination OSC PLS-DA. En raison de la qualité instrumentale (bruit important) de nos données, pour les deux logiciels, **la sélection des ions par ANOVA donne les meilleurs résultats en termes de prédiction du modèle**. Toutefois, cette stratégie devra être appliquée pour chaque technique d'analyse ou matrice biologique étudiée, afin de déterminer les paramètres des logiciels et les méthodes de prétraitement optimales.

Référence

- Dunn, W. B., Broadhurst, D., Brown, M., Baker, P. N., Redman, C. W. G., Kenny, L. C. and Kell, D. B. (2008). 'Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system.' *Journal of Chromatography B*. **871**, 288-298.
- Fardet, A., Llorach, R., Martin, J.-F., Besson, C., Lyan, B., Pujos-Guillot, E. and Scalbert, A. (2008). 'A Liquid Chromatography-Quadrupole Time-of-Flight (LC-QTOF)-based Metabolomic Approach Reveals New Metabolic Effects of Catechin in Rats Fed High-Fat Diets.' *Journal of Proteome Research*. **7**, 2388–2398.
- German, J. B., Hammock, B. D. and Watkins, S. M. (2005). 'Metabolomics: building on a century of biochemistry to guide human health.' *Metabolomics*. **1**, 3-9.
- Gibney, M. J., Walsh, M., Brennan, L., Roche, H. M., German, J. B. and van Ommen, B. (2005). 'Metabolomics in human nutrition : opportunities and challenges.' *American Journal of Clinical Nutrition*. **82**, 497-503.
- Gika, H., Macpherson, E., Theodoridis, G. and Wilson, I. D. (2008). 'Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples.' *Journal of Chromatography B*. **871**, 299-305.
- Idborg, H., Zamani, L., Edlund, P.-O., Schuppe-Koistinen, I. and Jacobsson, S. P.: 2005, 'Metabolic fingerprinting of rat urine by LC/MS: Part 2. Data pretreatment methods for handling of complex data. (2005).' *Journal of Chromatography B*. **828**, 14-20.
- Jansen, J. J., Hoefsloot, H. C. J., van der Greef, J., Timmerman, M. E., Westerhuis, J. A. and Smilde, A. K. (2005). 'ASCA: analysis of multivariate data obtained from an experimental design.' *JOURNAL OF CHEMOMETRICS*. **19**, 469–481.
- Kvalheim, O., Brakstad, F. and Liang, Y. (1994). 'PREPROCESSING OF ANALYTICAL PROFILES IN THE PRESENCE OF HOMOSCEDASTIC OR HETEROSCEDASTIC NOISE.' *Analytical chemistry*. **66**, 43 -51.
- Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., Ommen, B. v., Pujos-Guillot, E., Verheij, E., Wishart, D. and Wopereis, S. (2009). 'Mass-spectrometry-based metabolomics in nutrition – Current limitations and recommendations.' *Metabolomics*. **5**, 435–458.
- Smilde, A. K., van der Werf, M. J., Bijlsma, S., van der Werff-van der Vat, B. J. C. and Jellema, R. H. (2005). 'Fusion of Mass Spectrometry-Based Metabolomics Data.' *Analytical Chemistry*. **77**, 6729-6736.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006). 'XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification.' *Anal. Chem.* **78**, 779-787.

Snipen, L., Nyquist, O. L., Solheim, M., Aakra, Å. and Nes, I. F. (2009). 'Improved analysis of bacterial CGH data beyond the log-ratio paradigm.' *BMC Bioinformatics.* **10**, 91.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. and van der Werf, M. J. (2006). 'Centering, scaling, and transformations: improving the biological information content of metabolomics data.' *BMC Genomics.* **7**, 142.

van Velzen, E. J. J., Westerhuis, J. A., Duynhoven, J. P. M. v., Dorsten, F. A. v., Hoefsloot, H. C. J., Jacobs, D. M., Smit, S., Draijer, R., Kroner, C. I. and Smilde, A. K. (2008). 'Multilevel Data Analysis of a Crossover Designed Human Nutritional Intervention Study.' *Journal of Proteome Research.* **7**, 4483–4491.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J. v., Duijnhoven, J. P. M. v. and Dorsten, F. A. v. (2008). 'Assessment of PLSDA cross validation.' *Metabolomics.* **4**, 81–89.

Westerhuis, J. A., Velzen, E. J. J. v., Hoefsloot, H. C. J. and Smilde, A. K. (2008). 'Discriminant Q₂(DQ₂) for improved discrimination in PLSDA models.' *Metabolomics.* **4**, 293–296.

CHAPITRE IV : IDENTIFICATION ET CARACTERISATION

Chapitre IV

I. Introduction.

En métabolomique, la stratégie d'identification des biomarqueurs est basée sur l'utilisation de bases de données, ainsi que sur l'utilisation de spectrométrie de masse haute résolution en tandem. L'identification chimique consiste à rechercher des similitudes entre des données expérimentales (l'empreinte spectrale ou la masse monoisotopique détectée) et des métabolites référencés dans les bases de données. Pour confirmer l'identification, certains critères physico-chimiques entre l'analyte et le composé connu doivent être similaires (le temps de rétention, les fragments). (Milman *et al.*, 2005). Dans le cas où les masses détectées ne sont pas référencées, une étape de caractérisation de formules brutes est nécessaire avant l'identification.

L'objectif de ce chapitre est d'identifier les biomarqueurs plasmatiques de la prise de poids chez des volontaires soumis à un régime hyper-lipidique pendant 56 jours (ANR METAPROFILE). Les échantillons plasmatiques ont été analysés par UPLC-QToF (Pereira *et al.* 2009) afin d'obtenir des profils métaboliques à différents temps du régime. La comparaison de ces profils métaboliques plasmatiques a permis de sélectionner une liste d'ions témoins de la dérégulation du métabolisme induite par le régime.

La première étape de l'identification consiste à éliminer l'information redondante présente dans la matrice des marqueurs. Une méthode basée sur l'étude de corrélations a permis de classer les ions sélectionnés en quatre groupes qui sont les isotopes, les fragments, les adduits ou les ions moléculaires. Dans un deuxième temps, afin d'obtenir la masse exacte des ions sélectionnés, certains échantillons plasmatiques ont été re analysés par LTQ-Orbitrap (Scheltema *et al.*, 2008). Cette technique nous a également permis de réaliser des expériences de fragmentation (MS^2 ou MS^3) et d'accéder ainsi à des informations structurales. Aussi, l'utilisation combinée des masses exactes, des distributions isotopiques et des spectres de fragmentation des ions moléculaires sélectionnés et analysés par Orbitrap, ont permis d'identifier ou caractériser ces composés.

II. Matériels et Méthodes

1. Etude d'intervention chez l'Homme

Pour cette étude, deux groupes de sujets âgés de 18 à 35 ans ont été recrutés : un groupe de 40 sujets minces ($IMC < 25$) sans antécédents familiaux d'obésité, et un groupe de 40 sujets en surpoids ($25 < IMC < 30$) avec antécédents familiaux. Les volontaires ont été soumis à un régime hyperlipidique et hypercalorique pendant 56 jours. Ce régime a consisté en une surnutrition lipidique apportée par 20g de beurre, 40g d'amandes et 100g d'emmental correspondant à 760 kilocalories supplémentaires par jour. Les échantillons de plasma ont été prélevés sur héparine à différentes périodes du régime (0, 14^{ème} et 56^{ème} jour).

2. Traitement des échantillons

Les méthodes de préparation et d'analyse utilisées sont celles développées dans le deuxième chapitre.

2.1. Préparation des échantillons

Les plasmas ont été préparés par une méthode de précipitation protéique lente qui consistait à conserver pendant 30 minutes à $-20^{\circ}C$ un mélange de 200 μ L de plasma et de 400 μ L de méthanol. Une étape de centrifugation à 13000 tours par minute pendant 10 minutes a permis de séparer le précipité du surnageant. Ce surnageant est prélevé, évaporé sous azote, et repris dans un mélange eau/acétonitrile (50/50) contenant 0,1 % d'acide formique.

2.2. Analyse des échantillons

Analyse haut débit

131 échantillons plasmatiques ont été analysés par un couplage chromatographie liquide (UPLC) avec un spectromètre de masse ESI-QToF micro Waters (Saint Quentin en Yvelines, France) en mode positif et négatif avec une gamme de masse comprise entre m/z 70 et 1000. Pour la séparation, une colonne ACQUITY UPLC HSS T3 de 2,1 x 50 mm (Waters) contenant des particules de 1,8 μ m a été utilisée et maintenue dans un four à $30^{\circ}C$. 6 μ L d'échantillon plasmatique ont été injectés et élués à un débit de 0.4mL/minute utilisé en mode gradient linéaire (curve 6) réalisé comme suit : de 0 à 2 min: 100 A%, de 2 à 15 min: 0 à 100% B, de 15 à 22 min 100% B et de 22 à 26 minutes 100% A.

La tension du capillaire et la tension de cône ont été optimisées à 3 kV et à 30V, le temps de balayage et le temps de séjour ont été fixés à 1 et 0,1 seconde et le débit des gaz de nébulisation et de désolvatation (azote) ont été fixés à 50 et 500 L / h, respectivement.

Les échantillons ont été répartis dans quatre séries d'analyse de manière à équilibrer les 2 phénotypes (surpoids et mince) dans chaque série.

Analyse haute résolution

Certains échantillons ont été sélectionnés et ré-analysés par LTQ-Orbitrap (Thermo Fischer) afin d'obtenir une mesure de masse exacte (3ppm) et des spectres de fragmentations MS² et MS³ nécessaires à l'identification ou la caractérisation. La précision en masse est inférieure à 3 ppm compte tenu d'une calibration externe inférieure à 2 ppm. La calibrant utilisé est une solution de peptide fentomolaire.

3. Traitement des données

Pour déterminer les paramètres des logiciels d'extraction et d'alignement nous avons suivi la stratégie développée au Chapitre III.

3.1. Extraction et Alignement des données

D'après les résultats du Chapitre III, le choix des logiciels XCMS et MarkerLynxTM n'a pas d'effet significatif sur les résultats obtenus. Pour cette étude, le logiciel MarkerLynxTM est utilisé. Des tests préliminaires, utilisant les critères d'évaluation de la matrice de données élaborés au chapitre III, ont permis de fixer les cinq paramètres MarkerLynxTM comme suit :

- le pourcentage de bruit de fond à éliminer dans le spectre chromatographique (PPBN en anglais *Peak to Peak Baseline Noise*) à 30%,
- la valeur de seuil du bruit de fond à éliminer (NEL en anglais *Noise Elimination Level*) à 1,
- la valeur minimale de la largeur du pic de masse à 5% de la hauteur (PW5 en anglais *Peak Width at 5% Height*) est choisie automatiquement par le logiciel,
- l'intervalle de masse (MW en anglais *Mass Window*) a été fixé à 0,1 Th,
- et l'intervalle du temps de rétention (RTW en anglais *Retention Time Window*) 0.6 minute.

3.2. Sélection des candidats à l'identification

L'utilisation de méthodes univariées telles que l'analyse de covariance (ANCOVA) sur mesures répétées en utilisant le modèle mixte a permis de déterminer les métabolites discriminants le phénotype (surpoids ou mince), le régime et l'interaction des deux facteurs. Pour réaliser cette méthode univariée un programme a été élaboré sur le logiciel SAS version 9.1 (SAS Institute Cary, NC) basé sur la formule suivante :

$$x_{ijk} = x_0 + \alpha_i + \alpha_j + (\tilde{\alpha}\alpha)_{ij} + \alpha_k + (\tilde{\alpha}\alpha)_{ik} + \alpha_{s(i)} + \alpha_{ijsk}$$

- x_0 est la valeur du métabolite au départ;
- α_i est l'effet du phénotype;
- α_j est l'effet du régime ;
- $(\tilde{\alpha}\alpha)_{ij}$ est l'interaction entre le phénotype et le régime ;
- α_k effet de la série d'analyse ;
- $(\tilde{\alpha}\alpha)_{ik}$ est l'interaction entre le phénotype et la série d'analyse ;
- $\alpha_{s(i)}$ est l'effet aléatoire.
- α_{ijsk} est l'erreur aléatoire.

Ainsi, le programme permet de sélectionner un ion si l'effet du phénotype, du régime ou de l'interaction des deux est significatif et si les intensités mesurées de cet ion ne varient pas significativement entre les séries d'analyse.

4. Stratégie d'identification

Deux stratégies d'identification sont appliquées aux ions sélectionnés par analyses statistiques.

La première stratégie consiste à soumettre aux bases de données, les données analytiques obtenues par Orbitrap (la masse exacte du pic monoisotopique, des adduits et/ou son spectre de masse (MS^2)). Un algorithme réalisé au laboratoire permet d'automatiser cette étape. Les bases de données utilisées sont des bases commerciales Kegg (www.genome.jp/kegg/), HMDB (Whishart, 2007) ainsi que celles développées au laboratoire. Pour accepter l'identification putative d'un pic proposée par la base de données, les composés comparés devaient avoir une variation des masses exactes inférieure à 3 ppm, des distributions isotopiques similaires et les fragments des spectre MS^2 réalisés. (Figure IV. 1). Un autre logiciel a été utilisé afin d'obtenir la distribution isotopique des métabolites présents dans la

base de données : HRMet (Draper et al., 2009). De plus, dans ce logiciel, une valeur de probabilité est affectée à chaque abondance relative d'un isotope.

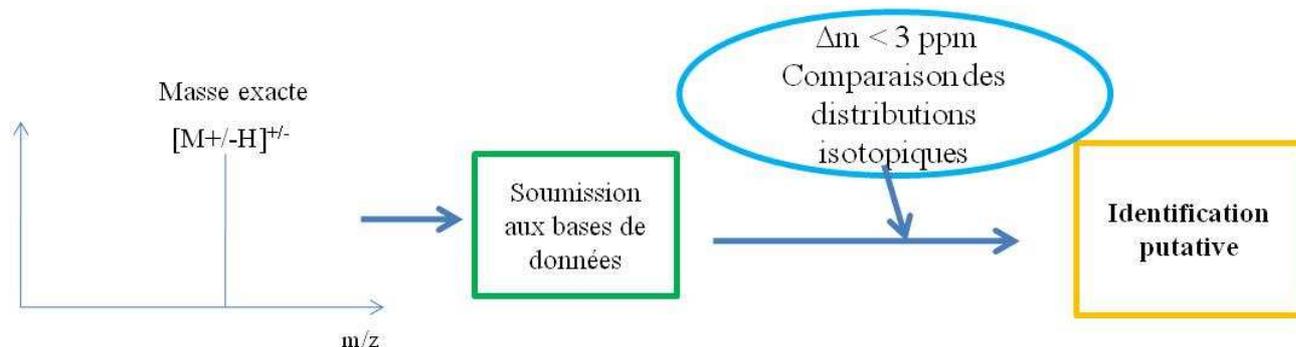


Figure. IV.1 _Schématisation de la stratégie d'identification putative (Sumner et al., 2007)).

Lorsqu'aucune proposition n'était proposée pour une masse mesurée, celle-ci était alors soumise à un logiciel fourni par la société Thermo Fischer « Qual Browser ». Ce logiciel propose des formules brutes à partir des masses exactes en s'appuyant sur les règles de caractérisation classiques telles que la règle de l'azote et les sept règles d'or élaborées par Kind et Fiehn (2007).

L'analyse de la distribution isotopique, ainsi que des fragments obtenus par CID nous a permis de confirmer ou non les formules proposées. (Figure IV. 2).

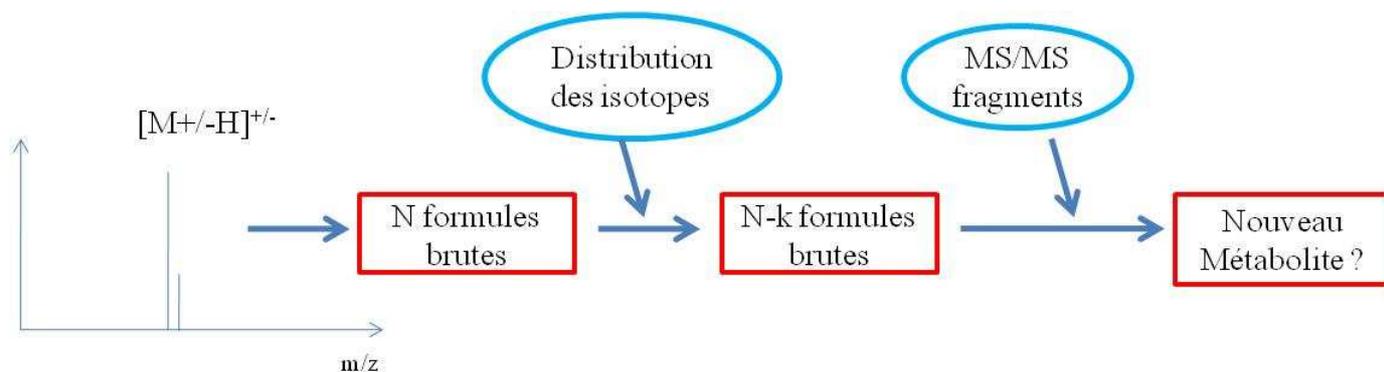


Figure. IV.2 _ Schématisation de la stratégie de caractérisation

La comparaison des distributions isotopiques a été réalisée à l'aide de la méthode des moindres carrés. (Loftus *et al.*, 2008). Le calcul ci-dessous permet d'obtenir une valeur de l'écart-type résiduel (a) qui doit tendre vers zéro.

$$a = (I_{[M+H+x]}^{\text{mesuré}} - I_{[M+H+x]}^{\text{théorique}})^2 + (I_{[M+H+x+1]}^{\text{mesuré}} - I_{[M+H+x+1]}^{\text{théorique}})^2 / 2 \dots + (I_{[M+H+x+n]}^{\text{mesuré}} - I_{[M+H+x+n]}^{\text{théorique}})^2 / n$$

III. Résultats

1. Ions sélectionnés en mode positif et négatif

1.1. Sélection des ions par ANCOVA.

L'analyse de covariance appliquée à chaque ion détecté par spectrométrie de masse en mode positif, a permis de sélectionner 108 ions, soit environ 2% des ions extraits. Parmi ces ions, 43 d'entre eux ont une réponse qui varie significativement en fonction du phénotype, 55 en fonction du régime et 18 en fonction de leur interaction. En mode négatif, 47 ions ont été sélectionnés et répartis comme suit : 15 ions significatifs pour l'effet du phénotype, 21 pour l'effet du régime et 4 pour l'effet de leur interaction. Dans la figure IV 3, les ions sélectionnés ont été répartis en fonction de leurs masses exactes et de leurs temps de rétention en mode positif (a) et en mode négatif (b).

En mode positif (a), une grande partie des ions sélectionnés ont été élués de la colonne entre 7 et 16 minutes, ce qui correspond à une phase mobile contenant entre 53% et 100% d'acetonitrile. Compte tenu du pourcentage élevé d'acetonitrile, les ions sélectionnés proviennent probablement de composés relativement peu polaires. Par exemple, d'après les résultats répertoriés dans le tableau 1 au chapitre II, les ions élués à 10 minutes et de valeurs m/z d'environ 500 sont potentiellement des lysophosphatidylcholines (lysoPC), molécules composées d'une longue chaîne de carbone avec une tête polaire.

En mode négatif (figure IV.3 (b)), les ions sélectionnés ont été élués tout au long de l'analyse et sont donc de polarité très variée. Toutefois, un nombre d'ions plus importants sont détectés vers 10 minutes, d'après le tableau II 1 présenté au chapitre II, les lysophosphatidyléthanolamine (PE) sont éluées à ce temps de rétention.

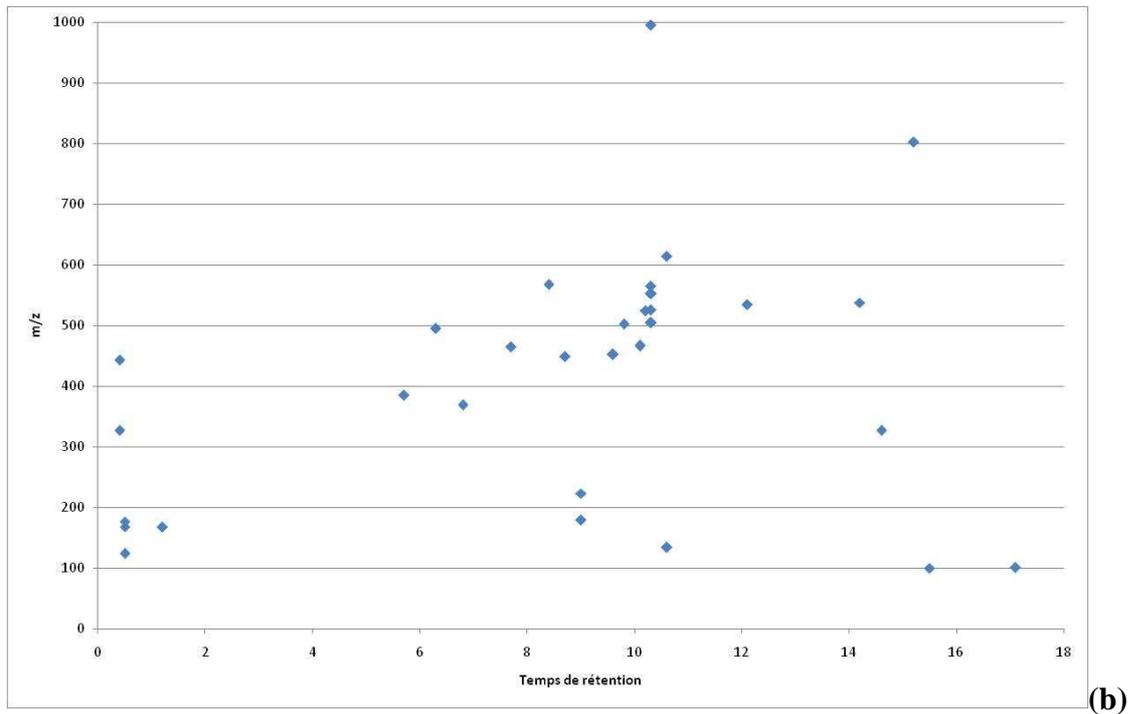
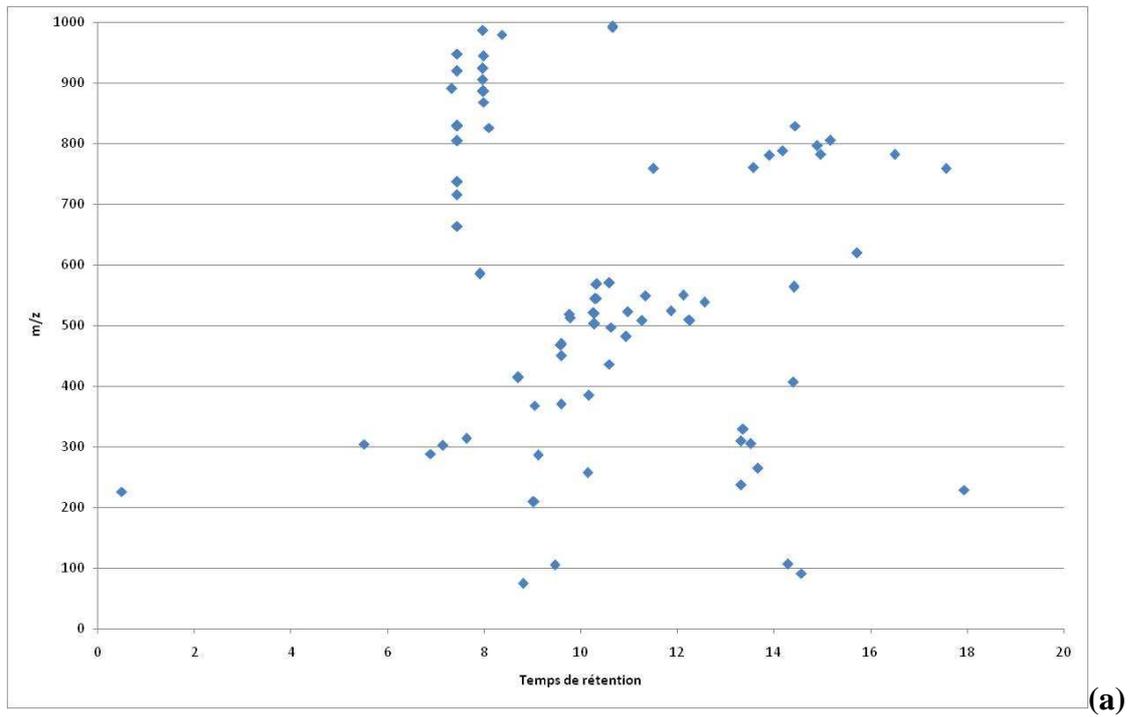


Figure. IV.3 _ Répartition des ions sélectionnés par analyse de la covariance selon leurs temps de rétention (axe x) et de leurs masses (axe y). En mode positif (a) et en mode négatif (b).

Dans une deuxième phase, afin d'éliminer l'information redondante, une étude de corrélation a été effectuée sur toutes les intensités des ions sélectionnés. Ainsi ces ions ont été répartis en

quatre groupes: les ions moléculaires potentiellement identifiables, leurs isotopes, leurs fragments et adduits.

1.2. Recherche des ions moléculaires, isotopes, adduits, et fragments en mode positif et négatif.

La liste des ions sélectionnés par ANCOVA a été analysée par l'algorithme développé au sein du laboratoire qui permet l'étude des corrélations. Pour cela, il calcule un coefficient de corrélation (R^2) entre deux variables indépendantes et quantitatives, soit les intensités de deux ions X et Y. La corrélation est acceptée si la valeur de R^2 est supérieur à 0.9, et que la probabilité de commettre l'erreur alpha (p) est inférieur à 0.05. Dans la figure IV 4, nous avons représenté le nombre d'isotopes, d'adduits (ajout de Na, K, CH_3COOH ...), de fragments et d'ions moléculaires ainsi trouvés dans la matrice des marqueurs.

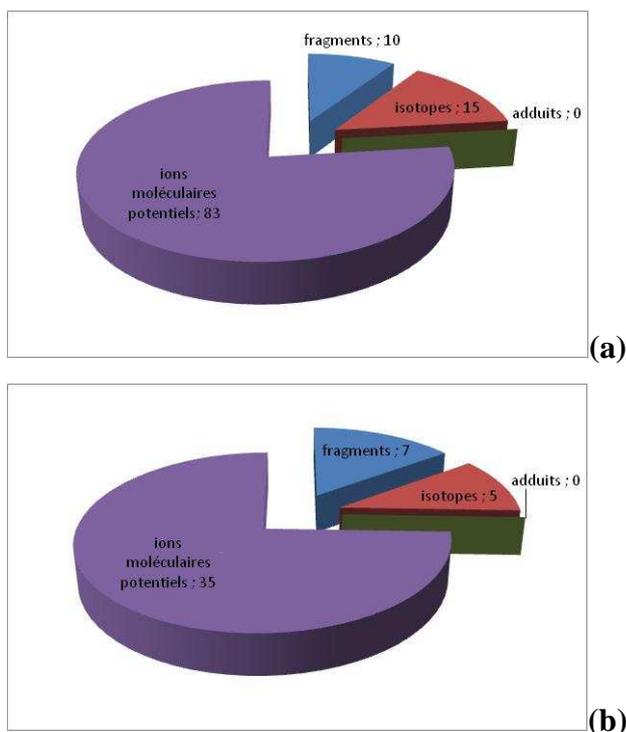


Figure. IV.4 __ L'étude des corrélations entre les intensités des ions sélectionnés par ANOVA ont permis de déterminer des relations entre ces ions permettant de les répartir en différentes classes : les ions moléculaires, les isotopes, les fragments et les adduits (a) Mode positif (b) Mode négatif.

D'après la figure IV 4, en mode positif, le nombre d'ions moléculaires $[\text{M}+\text{H}]^+$ potentiellement identifiables est de 83 ions. Pour aider à l'identification, 15 isotopes et 10 fragments des ions moléculaires ont été déterminés. En mode négatif, sur les 47 ions

sélectionnés, nous retrouvons 35 ions moléculaires potentiellement identifiables, ainsi que 7 fragments et 5 isotopes. Notons qu'aucun adduit n'a été identifié par le logiciel.

2. Identification des ions moléculaires

2.1 Requêtes dans les bases de données

L'ensemble des ions sélectionnés sont analysés par un deuxième algorithme développé au laboratoire qui recherche dans les bases de données (HMDB et KEGG) existantes les métabolites dont la masse monoisotopique est égale à la masse mesurée à ± 0.01 dalton près. Si tel est le cas, une identification est proposée. Pour affiner notre recherche, nous avons répertorié dans le tableau 1, les ions dont la variation entre la masse mesurée et la masse théorique était inférieure à 3 ppm.

Temps de rétention	Masse mesurée	Masse théorique	$\Delta m/z$ (ppm)	Identification par la base de données	Formule	Effet biologique
7.9	585.2710	585.2712	-0.34	Bilirubin	C ₃₃ H ₃₆ N ₄ O ₆	Diminue au cours du régime
9.58	468.3084	468.3090	-1.28	LysoPC(14:0)	C ₂₂ H ₄₆ NO ₇ P	Augmente au cours du régime
9.76	518.3242	518.3246	-0.77	LysoPC(18:3(9Z,12Z,15Z))	C ₂₆ H ₄₈ NO ₇ P	Augmente au cours du régime
10.27	520.3392	520.3403	-2.11	LysoPC(18:2(9Z,12Z))	C ₂₆ H ₅₀ NO ₇ P	Plus concentré chez les sujets minces
10.3	544.3389	544.3403	-2.57	LysoPC(20:4(5Z,8Z,11Z,14Z))	C ₂₈ H ₅₀ NO ₇ P	Diminue au cours du régime
10.31	568.3388	568.3403	-2.63	LysoPC(22:6(4Z,7Z,10Z,13Z,16Z,19Z))	C ₃₀ H ₅₀ NO ₇ P	Diminue au cours du régime et Plus concentré chez les sujets minces
10.59	570.3553	570.3559	-1.05	LysoPC(22:5(7Z,10Z,13Z,16Z,19Z))	C ₃₀ H ₅₂ NO ₇ P	Plus concentré chez les sujets minces
10.97	522.3550	522.3559	-1.72	LysoPC(18:1(9Z))	C ₂₆ H ₅₂ NO ₇ P	Plus concentré chez les sujets minces
11.34	548.3711	548.3710	0.18	LysoPC(20:2(11Z,14Z))	C ₂₈ H ₅₄ NO ₇ P	Plus concentré chez les sujets minces
11.5	758.5702	758.5699	0.39	PC(18:2(9Z,12Z)/16:0)	C ₄₂ H ₈₀ NO ₈ P	Plus concentré chez les sujets en surpoids
11.86	524.3709	524.3715	-1.14	LysoPC(18:0)	C ₂₆ H ₅₄ NO ₇ P	Diminue au cours du régime
12.12	550.3868	550.3872	-0.73	LysoPC(20:1(11Z)) PC(18:1(9Z)e/2:0)	C ₂₈ H ₅₆ NO ₇ P	Diminue au cours du régime
12.24	510.3915	510.3923	-1.56	LysoPC(O-18:0)	C ₂₆ H ₅₆ NO ₆ P	Diminue au cours du régime

12.56	538.3870	538.3867	0.55	LysoPE(22:0/0:0) LysoPE(0:0/22:0)	C ₂₇ H ₅₆ NO ₇ P	Diminue au cours du régime
13.35	329.2473	329.2474	-0.30	Docosahexaenoic acid 6alpha-Methylprogesterone 2alpha-Methylpregn-4-ene-3,20-dione	C ₂₂ H ₃₂ O ₂	Diminue au cours du régime
13.52	305.2469	305.2474	-1.63	Arachidonic acid Dihydroabietic acid 3beta-Methoxyandrost-5-en-16beta-ol 17beta-Hydroxy-4alpha-methyl-5alpha- androstan-3-one	C ₂₀ H ₃₂ O ₂	Diminue au cours du régime

Tableau. IV.1 _ Ce tableau contient tous les ions dont la variation entre la masse mesurée et la masse théorique est inférieure à 3 ppm. Les ions ont été détectés par Orbitrap et leurs masses ont été soumises aux bases de données en libre accès (Kegg et HMDB). PC (phosphatidylcholine), PE (phosphatidyléthanolamine)

D'après l'identification proposée par les bases de données, une grande partie des masses mesurées correspondent à des métabolites lipidiques (phospholipides ou acides gras) (Tableau IV. 1). De plus, plusieurs propositions de métabolites sont associées à la même masse, ce qui complique l'identification. Pour confirmer l'identification, des techniques complémentaires sont nécessaires comme la comparaison des distributions isotopiques mesurées et théoriques, ainsi que l'étude des spectres de fragmentation CID.

En raison de la faible intensité des ions détectés en mode positif, seul le mode de fragmentation MS² a été utilisé sur l'Orbitrap. Dans les spectres MS² des ions précurseurs identifiés comme des phosphocholines ou lysophosphocholines (m/z 510.3915, 550.3868, 468.3084, 518.3242, 520.3392, 544.3389, 568.3388, 570.3553, 522.3550, 548.3711, 758.5702, 524.3709); un fragment caractéristique de la choline est détecté à m/z 184.0723 ainsi qu'une perte d'eau (18.0107 Th) (Figure IV. 5).

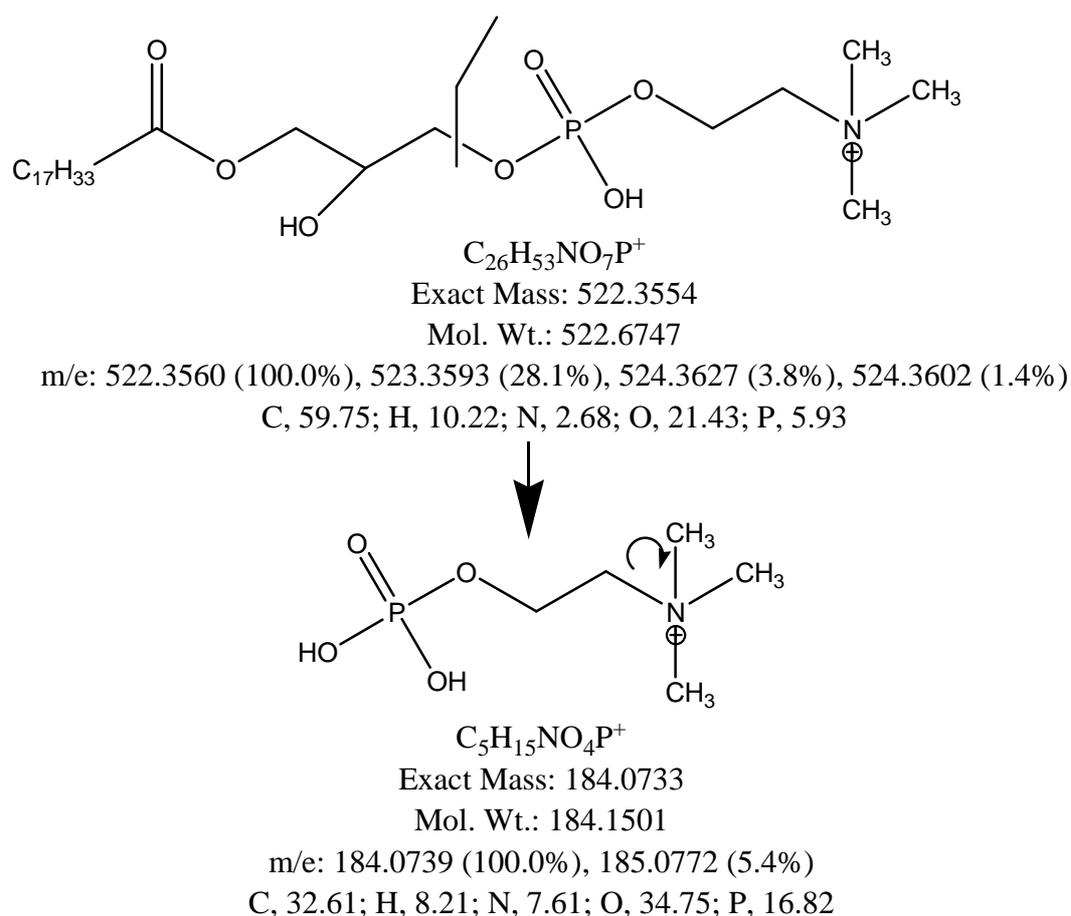


Figure. IV.5 _ Fragmentation caractéristique d'une lysophosphocholine. Perte de l'ion m/z 184,0733.

Par conséquent, pour ces douze ions, l'identification proposée par la base de données a permis de sélectionner tous les isomères de la famille des lysophosphocholines ou des phosphocholines. En effet à ce stade, il est difficile de différencier les isomères.

Pour les quatre ions restants, la fragmentation MS^2 n'a pas donné d'ions caractéristiques. Toutefois la formule brute proposée par la base de données peut être confirmée par la comparaison des distributions isotopiques mesurée et théorique. La distribution isotopique théorique du métabolite référencé est obtenue à l'aide d'un logiciel « Qual Browser ».

De plus, le calcul du rapport isotopique permet d'avoir une idée du nombre d'atomes de carbone présents dans le composé étudié.

Nombre d'atome de carbone = (intensité du pic $[M+1]$ / intensité du pic $[M]$) * (100/1.1) (1)

Pour trois des composés restant (la bilirubin, l'acide arachidonique, acide docosahexanoïque), la similitude des distributions isotopiques, nous permet de confirmer l'identification. Enfin, pour l'ion à m/z 538.3870 aucun fragment et isotope permet de confirmer son identification.

Pour toutes les masses non répertoriées dans les bases de données, une étape de caractérisation permettra de définir une formule brute.

2.2. Caractérisation des ions moléculaires $[M+H]^+$ en mode positif.

Dans le tableau IV. 2, nous avons inscrit le nombre de formules brutes proposées par le logiciel « Qual Browser » pour une masse donnée, ainsi que la formule brute retenue, grâce à la concordance des profils isotopiques théoriques et observés.

Ions moléculaires mesurés	Isotopes mesurés	Nombre de formules brutes proposées	Formule	$\Delta m/z$ (ppm)	Effet biologique
209.1224	210.1224 (12%) 211.1191 (0.7%)	3	X	X	Augmente au cours du régime
226.0489	227.0669 (5%)	3	C ₅ H ₁₂ N ₃ O ₅ S ⁺¹	1.41	Plus concentré chez les sujets en surpoids
286.2006	287.1684 (23%)	3	X	X	Augmente au cours du régime
314.2422	315.2299 (14%)	2	X	X	Diminue au cours du régime
406.3352	407.3302 (24.4%)	3	C ₂₂ H ₅₀ NOP ₂ ⁺¹	-2.5	Diminue au cours du régime
414.2969	415.3071 (26.6%)	7	C ₂₂ H ₄₄ N ₃ S ₂ ⁺¹	-0.52	Plus concentré chez les sujets minces
512.3402	513.3412 (18%)	27	C ₁₅ H ₄₃ O ₃ N ₁₅ P ⁺¹	-0.67	Diminue au cours du régime
524.3714	525.374 (24.3%) 526.377 (2.6%)	13	C ₂₆ H ₅₅ NO ₇ P ⁺¹ C ₁₉ H ₅₅ N ₇ O ₅ PS ⁺¹	-0.07 -0.67	Diminue au cours du régime
563.4244	564.4382 (37.7%)	36	C ₃₃ H ₅₉ N ₂ O ₃ S ⁺¹	0.55	Diminue au cours du régime
619.4894	620.4833 (48.7%) 621.4644 (12.6%)	37	X	X	Plus concentré chez les sujets en surpoids
780.5526	781.5699 (42%)	98	X	X	Plus concentré chez les sujets en surpoids
991.6549	992.6725 (43.8%) 993.6786 (5.9%) 994.6989 (0.6%)	133	X	X	Diminue au cours du régime

Tableau. IV.2 _ Caractérisation des masses mesurées par Orbitrap et non répertoriées dans les bases de données existantes. Tous les ions détectés et sélectionnés par ANCOVA sont de couleur noir tandis que ceux juste détectés par Orbitrap sont en rouge.

D'après le tableau IV. 2, six formules brutes ont pu être attribuées mais aucun spectre de fragmentation n'a permis de déterminer des pertes ou des ions caractéristiques.

Par conséquent, sur les 83 ions moléculaires potentiellement identifiables en mode positif, seuls 21 ions ont été identifiés ou caractérisés. Cette faible valeur s'explique en partie par le fait que de nombreux ions sélectionnés par analyse univariée avaient une réponse instrumentale trop faible pour pouvoir obtenir des informations spectrales permettant de confirmer l'identification (15 ions).

Une démarche similaire a été appliquée aux ions sélectionnés en mode négatif.

2.3. Requêtes dans les bases de données.

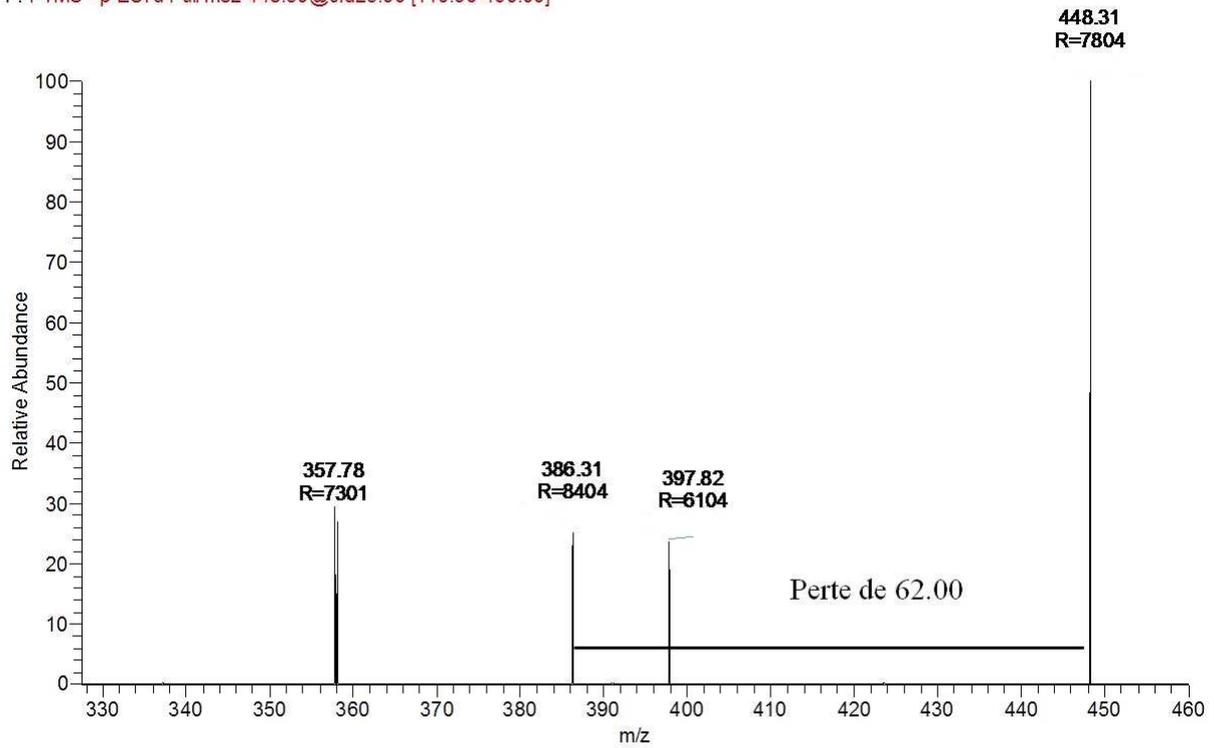
Dans un premier temps, la liste d'ions a été soumise aux différentes bases de données. En utilisant les mêmes critères que précédemment (variation de la masse inférieure à 3 ppm), montré au tableau IV. 3.

Temps de rétention	Masse mesurée (mode MS)	Masse théorique	$\Delta m/z$ (ppm)	Identification	Formule	Effet biologique
1.2	167.0209	167.0205	2.39	⁽¹⁾ Acide urique	C ₅ H ₄ N ₄ O ₃	Diminue au cours du régime
6.8	369.1733	369.1735	-0.50	⁽²⁾ 5a-Dihydrotestosterone sulfate Androsterone sulfate	C ₁₉ H ₃₀ O ₅ S	Augmente au cours du régime
7.7	464.3008	464.3012	-0.86	⁽³⁾ Glycocholic acid 3a,7b,12a-Trihydroxyoxocholanyl-Glycine Glycocholate	C ₂₆ H ₄₃ NO ₆	Plus concentré chez les sujets minces
8.4	567.3167	567.3169	-0.35	⁽³⁾ (3a,5b,7a)-23-Carboxy-7-hydroxy-24-norcholan-3-yl-b-D-Glucopyranosiduronic acid	C ₃₀ H ₄₈ O ₁₀	Plus concentré chez les sujets en surpoids
8.7	448.3057	448.3063	-1.33	⁽³⁾ Chenodeoxyglycocholic acid Glycoursodeoxycholic acid Glycochenodeoxycholate Glycodeoxycholate 3alpha,12alpha-Dihydroxy-5beta-cholan-24-oylglycine	C ₂₆ H ₄₃ NO ₅	Plus concentré chez les sujets minces
9.6	452.2787	452.2777	2.21	⁽⁴⁾ LysoPE(16:0/0:0) LysoPE(0:0/16:0)	C ₂₁ H ₄₄ NO ₇ P	Augmente au cours du régime
10.27	524.2788	524.2777	2.09	⁽⁴⁾ LysoPE(0:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)) LysoPE(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/0:0)	C ₂₇ H ₄₄ NO ₇ P	Diminue au cours du régime
14.2	537.4511	537.4519	-1.48	⁽⁵⁾ DG(16:1(9Z)/14:0/0:0) DG(16:0/14:1(9Z)/0:0) DG(14:1(9Z)/16:0/0:0) DG(14:0/16:1(9Z)/0:0)	C ₃₃ H ₆₂ O ₅	Plus concentré chez les sujets minces

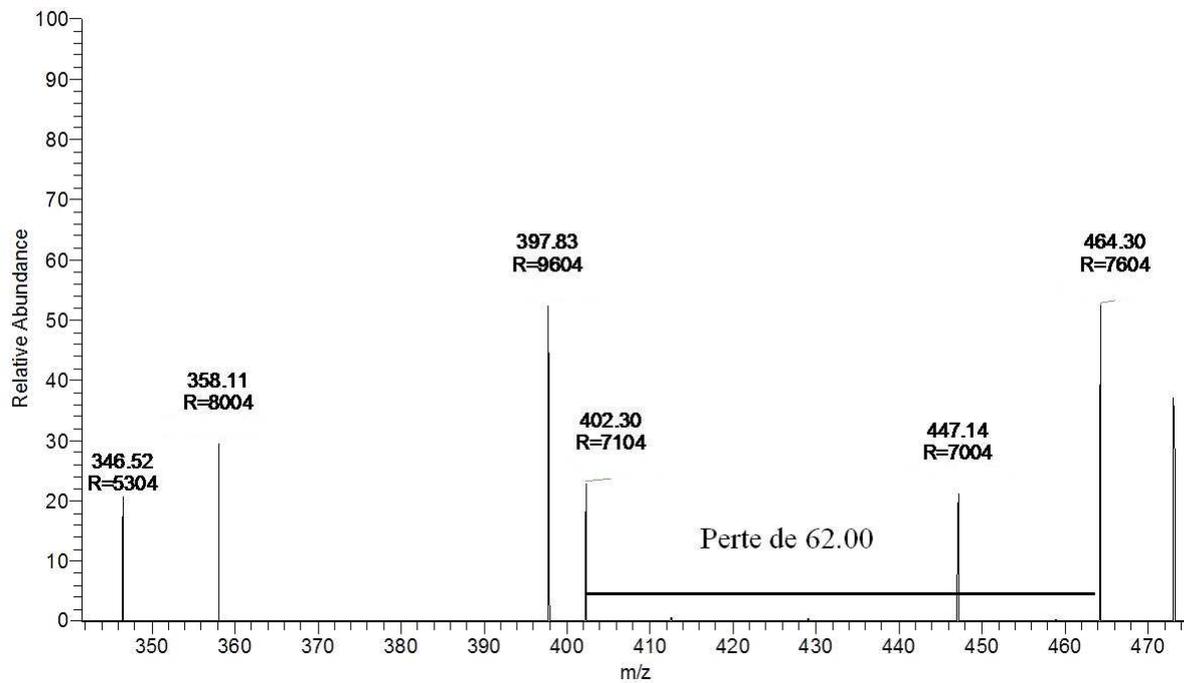
Tableau. IV.3 _ Ce tableau contient tous les ions dont la variation entre la masse mesurée et la masse théorique est inférieure à 3 ppm. Les ions ont été détectés par Orbitrap et leurs masses ont été soumises aux bases de données en libre accès (Kegg et HMDB). L'identification proposée par le moteur de recherche nous permet d'identifier 5 classes de métabolites : purines (1), des stéroïdes (2), des acides biliaires (3), des phospholipides (4) et des glycérolipides (5). (1DG (DiacylGlycérol), PE (PhosphatidylEthanolamine))

La comparaison avec les bases de données (HMDB et Kegg), nous a permis d'annoter huit pics correspondant à des métabolites de la famille des purines (1), des stéroïdes (2), des acides biliaires (7), des phospholipides (4) et des glycérolipides (4). Pour confirmer cette identification, les ions moléculaires ont été fragmentés. Les spectres MS² des ions moléculaires m/z 464.3008, 567.3167, 448.3057 contiennent une perte de 62.00 (R=7000) correspondant à la molécule CH₂O₃ (Δ m/z = 3.31 ppm). La comparaison des spectres MS² des acides biliaires présents dans la base de données HMDB, montre également une perte de 62. (Figure IV-6)

metapro-03-MS3 #3344 RT: 24.23 AV: 1 NL: 2.18E3
F: FTMS - p ESI d Full ms2 448.30@cid25.00 [110.00-460.00]



metapro-03-MS3 #3337 RT: 23.90 AV: 1 NL: 1.79E3
F: FTMS - p ESI d Full ms2 464.29@cid25.00 [115.00-475.00]



metapro-03_091118113311 #3734 RT: 25.75 AV: 1 NL: 4.63E3
F: FTMS - p ESI d Full ms2 567.31@cid25.00 [145.00-580.00]

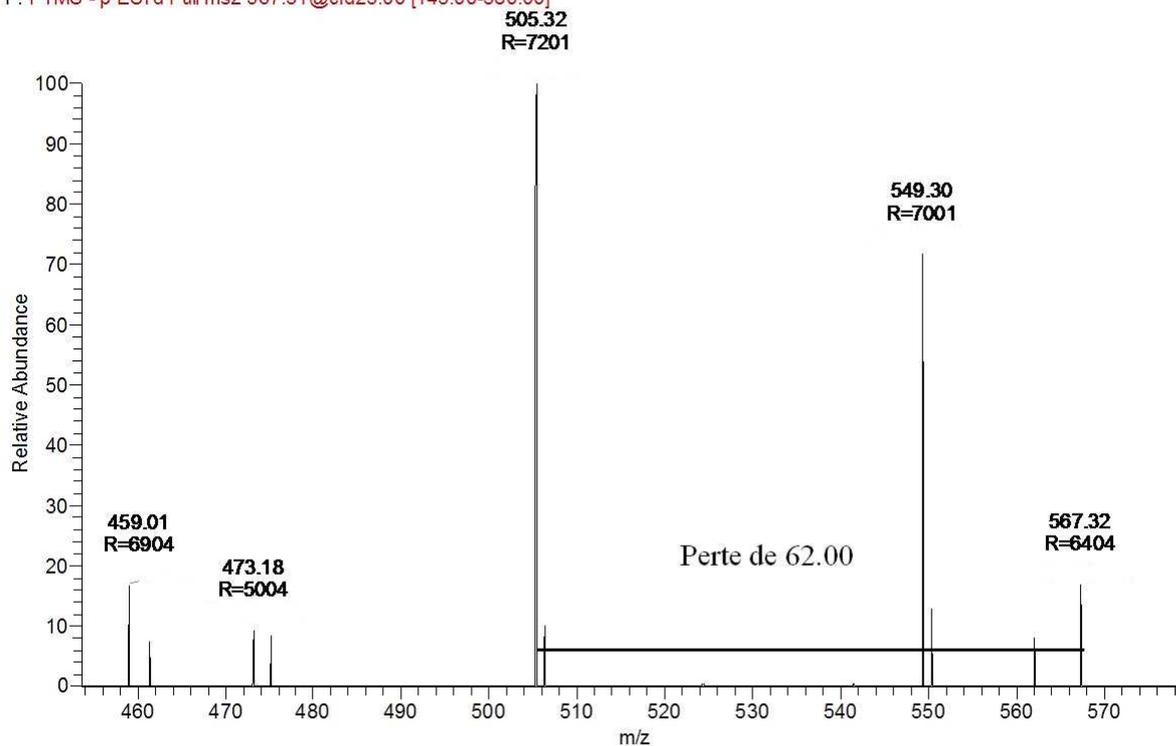


Figure. IV.6 _ Spectre MS² des ions pseudo-moléculaires m/z 448,31 ; 464,30 et 567,32. Notons une perte de 62,00.

Dans le spectre MS² de l'ion m/z 167.0209, deux fragments ont été détectés à m/z 124.0151 (C₄H₂N₃O₂⁻) et 96.0205 (C₃H₂N₃O⁻). (Figure IV-7).

metapro-03-MS3 #342 RT: 2.51 AV: 1 NL: 2.18E5
F: FTMS - p ESI d Full ms2 167.02@cid25.00 [50.00-180.00]

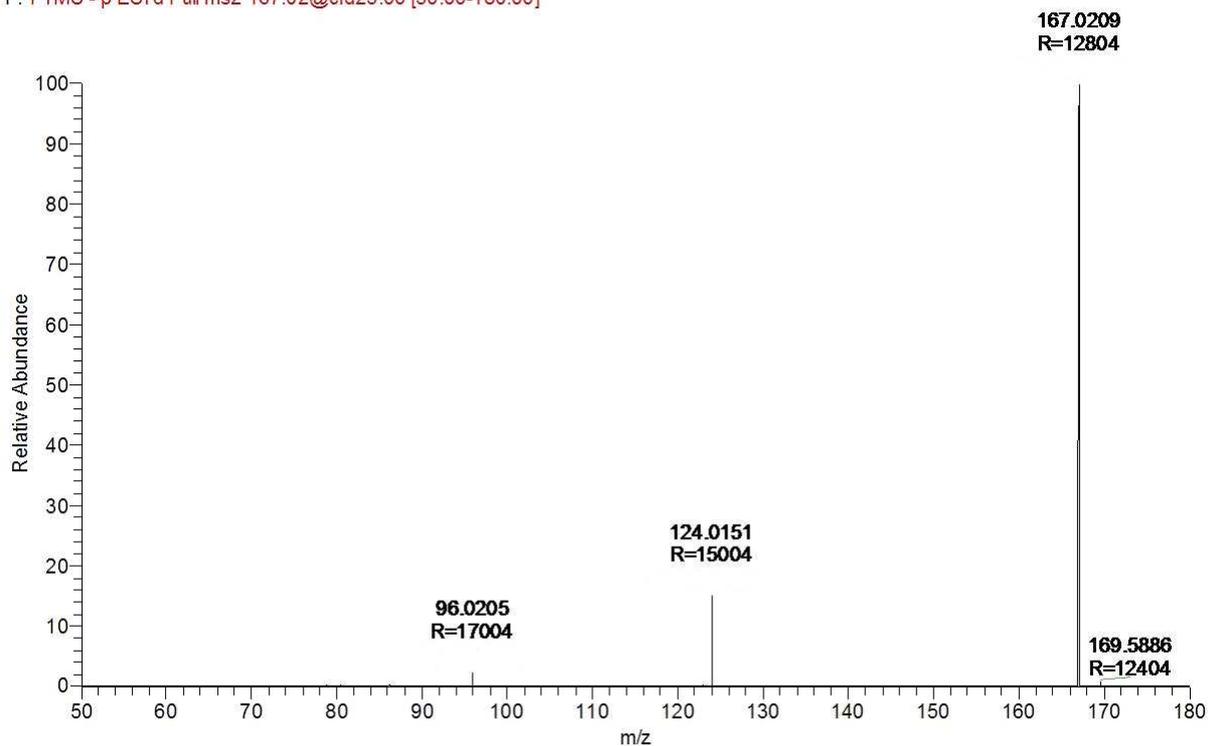


Figure. IV.7 _ Spectre MS² de l'ion pseudo moléculaire m/z 167,0209. Deux fragments sont détectés m/z 124,0151 et 96,0205.

Or ces fragments ne correspondent pas avec les fragments trouvés dans les bases de données. Toutefois, un chemin de fragmentation hypothétique est proposé. (Figure IV-8).

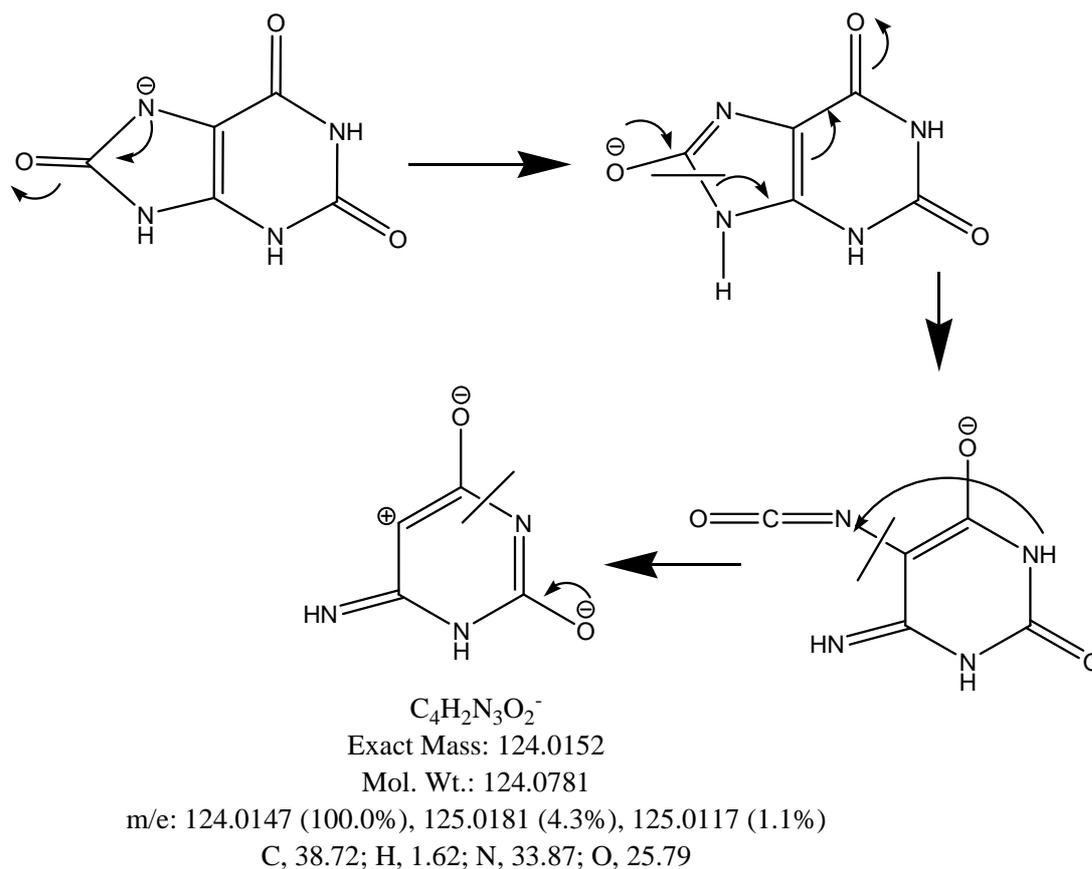
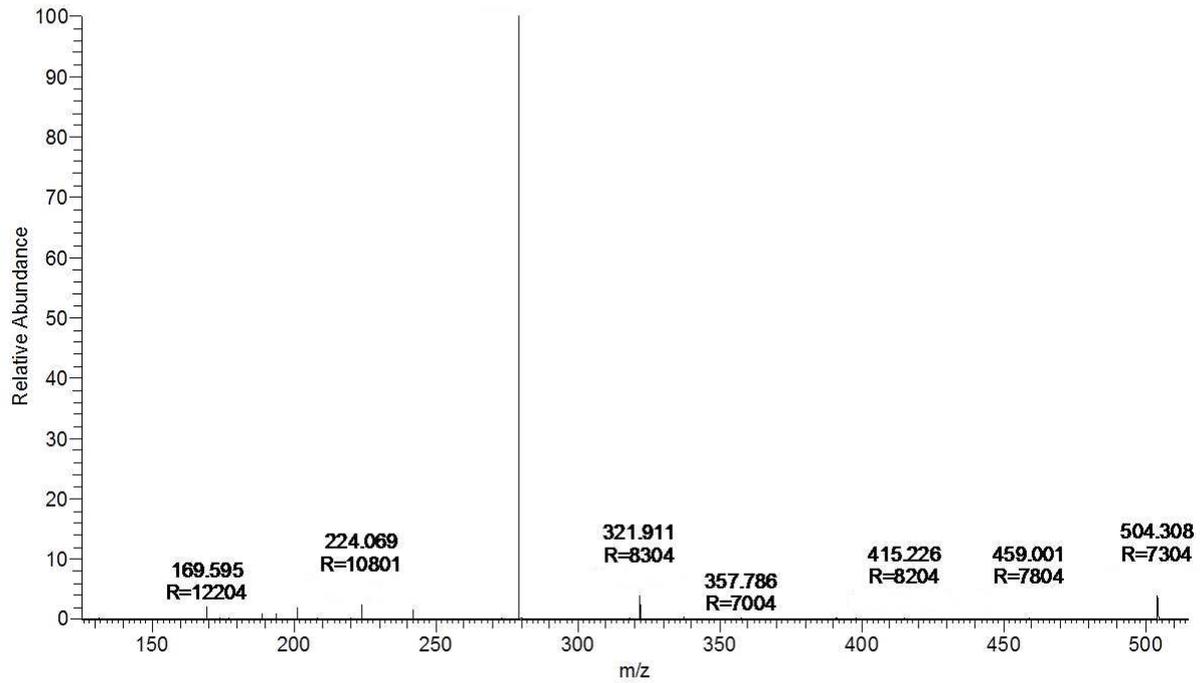


Figure. IV.8 _Chemin de fragmentation hypothétique de l'ion m/z 167.0209.

Le spectre MS² de l'ion m/z 369.1733 élué à 6.8 minutes contient de nombreux ions fragments difficilement interprétables (m/z 51.10 ou m/z 31.94). , la perte caractéristique des sulfates de stéroïdes (97.9596 Thomson correspondant à un groupement sulfaté (Hsu *et al.*, 2008)) n'est pas observée. Pour les trois derniers composés du tableau IV-3, les spectres MS² ne contiennent pas d'informations complémentaires. Toutefois les distributions isotopiques des ions mesurés et celles des métabolites référencés dans les bases de données sont similaires. Les écarts-type résiduels sont inférieurs à 10⁻⁴. Cette approche possède cependant une limitation dans le cas d'un faible nombre d'isotopes. L'analyse par LC-MS de ces standards permettra de valider cette identification. Pour toutes les masses non répertoriées dans les bases de données, seul l'ion m/z 564.3328 a pu être fragmenté. (Figure IV.9).

metapro-03-MS3 #3838 RT: 29.39 AV: 1 NL: 1.70E5
 F: FTMS - p ESI d Full ms3 564.33@cid25.00 504.31@cid25.00 [125.00-515.00]
 279.233
 R=10004



metapro-03-MS3 #3837 RT: 29.33 AV: 1 NL: 1.91E5
 F: FTMS - p ESI d Full ms2 564.33@cid25.00 [145.00-575.00]
 504.31
 R=7404

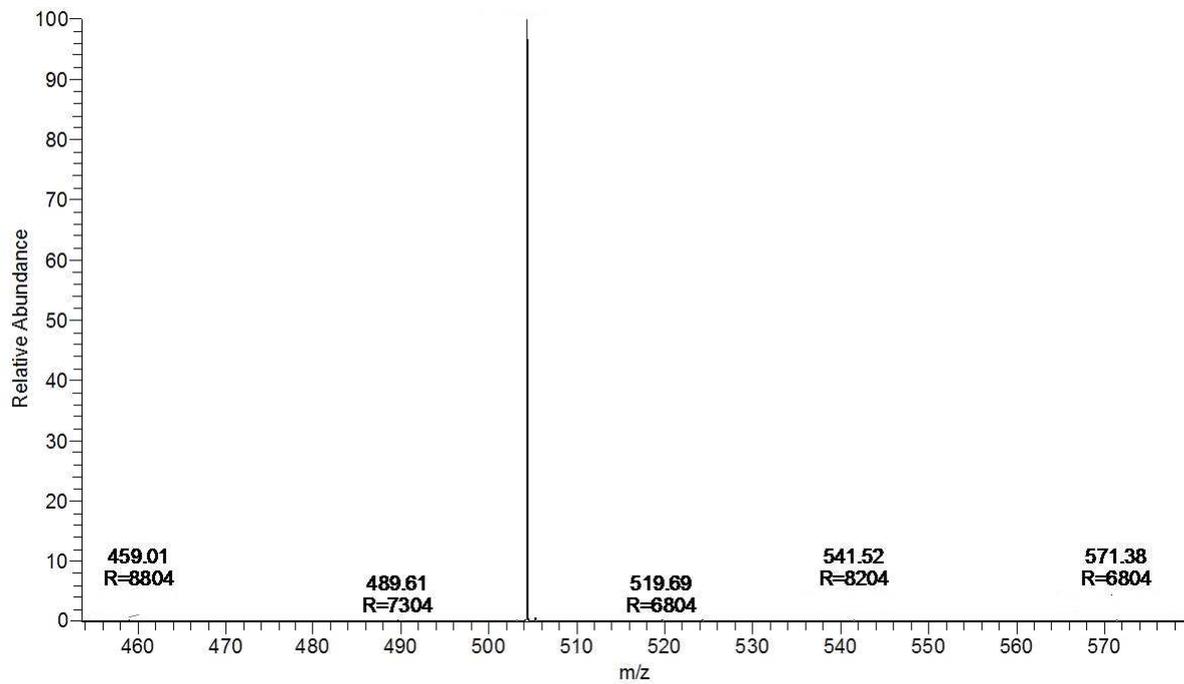


Figure IV.9 _ Spectre MS² de l'ion pseudo moléculaire m/z 564,3305. Le fragment m/z 504,31 est obtenu. Spectre MS³ de l'ion fragment m/z 504,31.

L'ion m/z 564.3305 se fragmente en MS^2 en un ion m/z 504,31. La perte de 60,02 est caractérisée par $C_2H_4O_2$. Le spectre MS^3 de l'ion m/z 504,31 donne un fragment à m/z 279.23 pouvant correspondre, d'après HMDB, à l'acide linoléique. Les autres ions n'ont pu être caractérisés en raison du manque d'information spectrale due à leurs faibles intensités (non détection des isotopes et fragments).

Par conséquent, sur les 35 ions moléculaires potentiellement identifiables, seuls 8 ions ont été identifiés. En mode négatif, une grande partie des ions détectés sont très peu intenses ce qui explique la difficulté de l'identification et de la caractérisation.

IV. Discussion

Dans cette étude, nous avons choisi d'utiliser une méthode statistique univariée plutôt que d'autres techniques statistiques telles que la PLS-DA (Loftus *et al.*, 2008 and Gika *et al.*, 2008), afin de pouvoir modéliser les différentes sources de variabilité et ainsi soustraire les effets liés à la série d'analyse.

Sur l'ensemble des métabolites sélectionnés (155 ions) seul 29 ont été identifiés ou caractérisés potentiellement. En raison de l'énorme diversité des structures chimiques des métabolites, l'étape d'identification est en effet limitante en métabolomique. Malgré la précision en masse des techniques utilisées, le nombre de métabolites présents dans les bases de données (7300 métabolites dans la base HMDB) ne permet pas dans tous les cas une identification. En réalité, bien qu'une variation en masse inférieure à 3 ppm soit une condition *sine qua non*, celle-ci n'est pas suffisante. Ainsi, il est indispensable d'appliquer d'autres critères pour la confirmation de l'identification, tel que la distribution isotopique et la fragmentation CID. Toutefois, cette stratégie ne permet pas toujours de différencier les isomères, comme nous pouvons les remarquer dans les tableaux IV 1 et 3. En effet, les isomères de la même famille de molécules ont souvent des fragments communs. Pour obtenir des informations complémentaires, il est envisagé d'utiliser des techniques complémentaires de fragmentations comme par exemple des fragmentations MS³ et des techniques d'ajout d'hydroxydes métalliques pour caractériser des isoméries de position. (Werner *et al.*, 2008). Néanmoins, pour améliorer l'efficacité de l'identification, il est essentiel de développer des bases de données contenant des spectres de masse (ou RMN) de tous les métabolites référencés (Whishart *et al.*, 2008). Toutefois, pour que cette amélioration permette une réelle avancée, les techniques utilisées pour les fragmentations doivent être standardisées. En effet, d'un spectromètre de masse à l'autre, les fragmentations CID d'une même molécule peuvent varier. Actuellement la seule manière efficace pour confirmer l'identification d'un ion détecté est d'analyser par le même instrument le composé standard quand il est disponible.

L'élucidation de nouvelles structures est très complexe et peut engendrer des conclusions erronées (Nicolaou *et al.*, 2005). Pour obtenir, une formule brute non ambiguë, il est nécessaire d'utiliser des spectromètres de masse à haute résolution afin d'avoir une précision en masse inférieure à 3 ppm. (0.001 Da). Plus la précision en masse est grande (inférieure à 3 ppm), plus le nombre de formules empiriques proposées est faible. (Scheltema *et al.*, 2008).

Ainsi, l'augmentation de la résolution et de la sensibilité des techniques analytiques utilisées permettront de caractériser de plus en plus d'ions inconnus. (Tableau IV. 4).

Exactitude de masse d'un spectromètre de masse	Nombre de formules proposées
3 ppm (Orbitrap)	23
1 ppm (FTICR)	7
0.1 ppm (pas encore)	1

Tableau. IV 4. Nombre de formules brutes proposées en fonction de la précision en masse du spectromètre de masse utilisé. (Kind *et al.*, 2006)

Les résultats obtenus (Tableau IV. 2) confirment l'étude de Kind *et al.* (2006) qui montre que, plus le poids moléculaire des composés est grand, plus le nombre de formules proposées l'est également. Toutefois l'utilisation de la distribution isotopique permet de diminuer drastiquement le nombre de candidats. (Tableau IV. 5).

Masse de la molécule étudiée (Da)	3 ppm	
	Sans isotope	Avec isotope
150	1	1
300	7	1
500	64	3

Tableau. IV.5. Nombre de formules brutes proposées pour différentes masses monoisotopiques avec ou sans ces isotopes détectés par un spectromètre de masse d'exactitude en masse de 3 ppm.

Enfin pour tous les ions non identifiables en raison de leur faible intensité, il serait envisageable de pré fractionner et/ou concentrer les échantillons afin d'obtenir de meilleurs empreintes spectrales et d'améliorer la réponse instrumentale en spectrométrie de masse.

V. Conclusion

Dans cette étude, 29 ions ont été identifiés ou caractérisés potentiellement, comme marqueurs de modifications métaboliques induites par un régime hyper-lipidique. Les ions identifiés ont été répartis en différentes catégories de métabolites : onze lysophosphocholines, une phosphocholine, quatre acides biliaires, deux acides gras, une purine, un glycérolipide, un stéroïde et deux lysophosphatidyléthanolamines. Il est toutefois nécessaire de confirmer l'identification et d'interpréter les modifications observées par l'étude des réseaux métaboliques impliqués.

De plus, l'augmentation de la précision en masse permettra dans l'avenir d'identifier de plus en plus de métabolites. Toutefois l'étude de matrices complexes induit des effets de charge qui peuvent conduire à la réduction du nombre d'ions transférés dans l'analyseur diminuant la capacité de reproductibilités des masses exactes mesurées. (Parsons *et al.*, 2007). Cette variation peut conduire à des erreurs dans l'identification. D'après Southam *et al.* (2007), l'utilisation du mode SIM en FT-ICR (en anglais *Selected Ion Monitoring*) permet d'augmenter d'un facteur 3000 la reproductibilité de la valeur de la masse mesurée. Ainsi, cette approche pourrait conduire à des identifications plus fiables.

Une alternative existe : elle consiste à développer des logiciels permettant de relier entre eux les métabolites par les réactions biochimiques identifiées dans l'organisme. Ces approches bioinformatiques permettraient peut être de simplifier cette étape d'identification. (Jourdan *et al.*, 2008).

Références

- Draper, J., Enot, D., Parker, D., Beckmann, M., Snowdon, S., Lin, W. and Zubair, H.: 2009, 'Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'.' *BMC Bioinformatics* **10**, 227.
- Gika, H. G., Theodoridis, G. A. and Wilson, A.: 2008, 'Hydrophilic interaction and reversed-phase ultraperformance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine.' *J. Sep. Sci.* **31**, 1598 – 1608.
- Jourdan, F., Breitling, R., Barrett, M. P. and Gilbert, D.: 2008, 'MetaNetter: inference and visualization of high-resolution metabolomic networks.' *Bioinformatics* **24**, 143-145.
- Kind, T. and Fiehn, O.: 2006, 'Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm.' *BMC Bioinformatics* **7**, 234.
- Loftus, N., Miseki, K., Iida, J., Gika, H. G., Theodoridis, G. and Wilson, I. D.: 2008, 'Profiling and biomarker identification in plasma from different Zucker rat strains via high mass accuracy multistage mass spectrometric analysis using liquid chromatography/mass spectrometry with a quadrupole ion trap-time of flight mass spectrometer.' *RCM* **22**, 2547–2554.*
- Milman, B. L.: 2005, 'Identification of chemical compounds.' *TrAC Trends in Analytical Chemistry* **24**, 493-508.
- Nicolaou, K., Snyder and Scott, A.: 2005, 'Chasing molecules that were never there: Misassigned natural products and the role of chemical synthesis in modern structure elucidation.' *Angew Chem Int Ed* **44**, 1012 - 1044.
- Parsons, H., Ludwig, C., Gunther, U. and Viant, M.: 2007, 'Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation.' *BMC Bioinformatics* **8**, 234.

Pereira, H., Martin, J.-F., Joly, C., Sébedio, J. L. and Pujos-Guillot, E.: 2009, 'Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma.' *Metabolomics*.

Scheltema, R., Kamleh, A., Wildridge, D., Ebikeme, C., Watson, D. G., Barrett, M. P., Jansen, R. C. and Breitling, R.: 2008, 'Increasing the mass accuracy of high-resolution LC-MS data using background ions – a case study on the LTQ-Orbitrap.' *Proteomics* **2008**, 4647–4656.

Southam, A. D., Payne, T. G., Cooper, H. J., Arvanitis, T. N. and Viant, M. R.: 2007, 'Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nano-electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method.' *Analytical Chemistry* **79**, 4595-4602.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reilly, M. D., Thaden, J. J. and Viant, M. R.: 2007, 'Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).' *Metabolomics* **3**, 211-221.

Werner, E., Croixmarie, V., Umbdenstock, T., Ezan, E., Chaminade, P., Tabet, J.-C. and Junot, C.: 2008, 'Mass Spectrometry-Based Metabolomics: Accelerating the Characterization of Discriminating Signals by Combining Statistical Correlations and Ultrahigh Resolution.' *Analytical Chemistry* **80**, 4918-4932.

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., MacInnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J. and Querengesser, L.: 2007, 'HMDB: the Human Metabolome Database.' *Nucl. Acids Res.* **35**, D521-526.

CONCLUSIONS ET PERSPECTIVES

Conclusions et perspectives

Dans le cadre d'une approche métabolomique, cette étude avait pour principal objectif de développer des méthodes analytiques permettant d'acquérir des profils métaboliques globaux de fluides biologiques et d'identifier des biomarqueurs potentiels de dysfonctionnements métaboliques occasionnés par une intervention nutritionnelle.

Dans un premier temps, une méthode robuste de détermination de profils plasmatiques a été développée et validée par UPLC-MS. D'après nos résultats, l'héparine est l'anticoagulant le plus approprié pour l'étude non ciblée du métabolome par LC/MS. La comparaison de trois méthodes de précipitation protéique lentes a montré que l'utilisation du méthanol avec un rapport plasma/méthanol 1/2 est la plus reproductible et extrait un maximum de métabolites (Chapitre II).

Les sources de variabilité analytiques ont été étudiées. Celles-ci peuvent être attribuées à la réponse instrumentale, mais également à l'étape d'extraction des pics durant le traitement des données. Pour comprendre l'origine de cette variabilité, les paramètres de deux logiciels d'extraction et d'alignement (MarkerLynx et XCMS) dans le cas d'une longue série d'analyses ont été optimisés (Chapitre III).

Ces logiciels génèrent un grand nombre de données contenant l'information biologique mais également du bruit. Le chapitre III de ce manuscrit propose une stratégie de traitement des données métabolomiques obtenues par LC-MS qui consiste d'une part, à s'assurer que l'extraction et l'alignement des données est optimale et d'autre part, que les méthodes de prétraitement appliquées permettent d'en extraire l'information biologique. Des critères d'évaluation d'une matrice de données sont proposés pour permettre l'optimisation des paramètres des logiciels. D'après les critères testés, **les deux logiciels (XCMS et MArkerLynxTM) ont des performances équivalentes**. Pour déterminer les valeurs critiques des paramètres de ces logiciels, le rapport signal sur bruit des pics chromatographiques et de masse, la largeur de ces pics ainsi que la variation du temps de rétention ont été des indicateurs forts.

L'étude des différents prétraitements appliqués aux matrices de données ont montré que la sélection des ions par ANCOVA donne les meilleurs résultats en termes de prédiction du modèle.

Enfin, ces deux avancées méthodologiques (méthode analytique et traitement de données) ont permis de traiter les échantillons plasmatiques issus de l'intervention nutritionnelle du projet METAPROFILE. Différentes familles de métabolites identifiés ont été proposées grâce aux bases de données : onze lysophosphocholines, une phosphocholine, quatre acides biliaires, deux acides gras, une purine, un glycérolipide, un stéroïde et deux lysophosphatidyléthanolamine. Toutefois, sur l'ensemble des ions candidats à l'identification seule 5 % ont été potentiellement identifiés ou caractérisés. Ce pourcentage montre la complexité de cette étape. La première difficulté est de réussir à distinguer dans la liste des marqueurs potentiels les ions moléculaires des adduits, des fragments ou des isotopes. Enfin, même si cette étape est franchie, la diversité chimique des métabolites ainsi que le manque d'exhaustivité dans les bases de données rend parfois impossible l'identification. Ainsi, aujourd'hui la seule façon de confirmer une identification est d'analyser directement le composé standard.

A partir de ces résultats, différentes perspectives peuvent être envisagées pour les étapes clés d'une étude métabolomique par spectrométrie de masse appliquée à la nutrition :

Couplages ESI-MS et métabolomique

Le couplage UPLC-QToF est une technique « haut débit » adaptée à l'obtention de profils métaboliques complets. Toutefois, les interventions nutritionnelles nécessitent la comparaison d'un grand nombre d'échantillons afin d'avoir une puissance statistique suffisante, compte tenu de la variabilité intra- et inter- individuelle. La prise en charge de telles séries d'échantillons impliquent des temps d'analyse importants, qui peuvent induire des variations instrumentales en termes de temps de rétention, d'intensité du signal ou de masse. Dans le cas d'une approche non ciblée, des méthodes analytiques de plus en plus courtes peuvent être envisagées comme première approche pour limiter ce phénomène. Néanmoins, les techniques utilisées ont certaines limitations. En effet, bien que **les meilleures performances** de l'UPLC soient obtenues à un haut débit, la faible vitesse de balayage du QToF rend ce couplage optimal impossible. **Ainsi, il serait préférable d'utiliser un couplage**

UPLC-TOF permettant d'obtenir de hautes performances en termes de sensibilité et de reproductibilité des temps de rétention et des masses mesurées. Des méthodes alternatives « shot gun » sont également envisagées par injection directe sur des spectromètres de masse très haute résolution (Orbitrap, FT-ICR) (Yang *et al.*, 2009). Ces méthodes nécessitent cependant de travailler sur matrices diluées pour limiter les phénomènes de suppression d'ionisation, ce qui limite la richesse de l'information accessible.

Traitement de données et métabolomique

L'amélioration des techniques analytiques engendrent des données de plus en plus complexes contenant certes l'information recherchée mais aussi du bruit. Dans les études métaboliques, les techniques les plus couramment utilisées sont la PLS-DA (Wiklund *et al.*, 2008), l'ACP (Gika *et al.*, 2008) ou l'ANOVA (Boccard *et al.*, 2007). Ces techniques statistiques permettent de révéler les ions discriminants pour un effet biologique étudié. Aujourd'hui, certains statisticiens mettent en doute la robustesse de ces modèles multivariées. (Werthesius *et al.*, 2008 (a)). Pour donner du poids à ces techniques, il est nécessaire de trouver des critères de validation capables de montrer que la discrimination biologique n'est pas due au hasard. (Werthesius *et al.*, 2008 (b)).

Il est essentiel de développer des algorithmes « intelligents » capables de trouver les faux positifs ou les faux négatifs contenus dans la liste des marqueurs potentiels. (Lu *et al.*, 2008).

Identification et métabolomique

Il est clair que l'amélioration de la précision en masse contribue à la détection de plus en plus de métabolites. Les avancées des techniques à haute résolution permettent des infusions directes sur couplage piège ionique et FTICR (en anglais *Fournier Transform Ion Cyclotron Resonance*) (Payne *et al.*, 2009) ou des couplages LC-LTQ-FT-Orbitrap. (Ding *et al.*, 2007 ; Dunn *et al.*, 2008 ; Scheltema *et al.*, 2008). Ces instruments sont adaptés à l'identification car d'une part, des spectres de masse sont obtenus avec une mesure de la masse exacte, et d'autre part le couplage avec le piège ionique permet d'effectuer des fragmentations MSⁿ.

Toutefois, il est important de s'assurer que la mesure de la masse exacte soit reproductible. Or, l'étude de matrices complexes induit des effets de charge dans le piège quadripolaire (LTQ) qui peuvent conduire à la réduction du nombre d'ions transférés dans

l'analyseur diminuant la reproductibilité des masses exactes mesurées. D'après Southam *et al.* (2007), l'utilisation du mode SIM (en anglais *Selected Ion Monitoring*) permet d'augmenter d'un facteur 3000, la reproductibilité de la valeur de la masse mesurée dans un FT-ICR. Une question subsiste sur l'obtention de la valeur de la masse exacte et des profils isotopiques qui peut, soit être déterminés à un temps donné, soit être une moyenne des mesures à des balayages successifs.

Pour simplifier l'étape de confirmation d'identification, il est nécessaire de développer des bases de données exhaustives contenant les spectres de fragmentations à différentes énergies de collision sur différents couplages LC-MS et la distribution isotopique à différentes résolutions. La standardisation des méthodes est une étape clé pour le partage des données analytiques. La diminution des coûts des nouvelles techniques à haute résolution (Orbitrap) permettra à de nombreuses équipes d'accéder à la spectrométrie de masse très haute résolution, facilitant le partage de données de qualité.

Une autre alternative à l'identification est de développer des outils bioinformatiques, qui permettent notamment d'étudier les données spectrales par la théorie des réseaux (Jourdan *et al.*, 2007).

En conclusion, l'enrichissement des bases de données, l'avancée des outils bioinformatiques et des techniques analytiques annoncent un avenir meilleur pour cette discipline qui est encore émergente.

Références

- Boccard, J., Grata, E., Thiocone, A., Gauvrit, J.-Y., Lantéri, P., Carrupt, P.-A., Wolfender, J.-L. and Rudaz, S. (2007), 'Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding.' *Chemometrics and Intelligent Laboratory Systems* **86**, 189–197.
- Ding, J., Sorensen, C. M., Zhang, Q., Jiang, H., Jaitly, N., Livesay, E. A., Shen, Y., Smith, R. D. and Metz, T. O. (2007), 'Capillary LC Coupled with High-Mass Measurement Accuracy Mass Spectrometry for Metabolic Profiling.' *Analytical Chemistry* **79**, 6081-6093.
- Dunn, W. B., Broadhurst, D., Brown, M., Baker, P. N., Redman, C. W. G., Kenny, L. C. and Kell, D. B. (2008), 'Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system.' *Journal of Chromatography B* **871**, 288-298.
- Gika, H., Macpherson, E., Theodoridis, G. and Wilson, I. D. (2008), 'Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples.' *Journal of Chromatography B* **871**, 299-305.
- Jourdan, F., Breitling, R., Barrett, M. P. and Gilbert, D. (2008), 'MetaNetter: inference and visualization of high-resolution metabolomic networks.' *Bioinformatics* **24**, 143-145.
- Lu, H., Dunn, W. B., Shen, H., Kell, D. B. and Liang, Y. (2008), 'Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS.' *Trends in Analytical Chemistry* **27**, 215-227.
- Payne, T. G., Southam, A. D., Arvanitis, T. N. and Viant, M. R. (2009), 'A Signal Filtering Method for Improved Quantification and Noise Discrimination in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Data.' *Journal of the American Society for Mass Spectrometry* **20**, 1087-1095.
- Scheltema, R., Kamleh, A., Wildridge, D., Ebikeme, C., Watson, D. G., Barrett, M. P., Jansen, R. C. and Breitling, R. (2008), 'Increasing the mass accuracy of high-resolution LC-MS data using background ions – a case study on the LTQ-Orbitrap.' *Proteomics* **2008**, 4647–4656.

Southam, A. D., Payne, T. G., Cooper, H. J., Arvanitis, T. N. and Viant, M. R. (2007), '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.' *Analytical Chemistry* **79**, 4595-4602.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J. v., Duijnhoven, J. P. M. v. and Dorsten, F. A. v. (2008), 'Assessment of PLSDA cross validation.' *Metabolomics* **4**, 81–89.

Westerhuis, J. A., Velzen, E. J. J. v., Hoefsloot, H. C. J. and Smilde, A. K. (2008), 'Discriminant Q2(DQ2) for improved discrimination in PLSDA models.' *Metabolomics* **4**, 293–296.

Wiklund, S., Johansson, E., Sjostrom, L., Mellerowicz, E., Edlund, U., Shockcor, J., Gottfries, J., Moritz, T. and Trygg, J. (2008), 'Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models.' *Anal Chem* **80**, 115 - 122.

Yang, K., Cheng, H., Gross, R. W. and Han, X. (2009), 'Automated Lipid Identification and Quantification by Multidimensional Mass Spectrometry-Based Shotgun Lipidomics.' *Analytical Chemistry* **81**, 4356-4368.

NOUVEAU CHAPITRE DE THESE

Mentor Nicolas Baleyrier

Chapitre V

I. Enjeux scientifiques de la thèse.

1. Contexte

La prise de poids est l'état d'une personne souffrant d'une augmentation importante de cellules adipeuses, qui se traduit par un excès de poids, réparti de façon généralisée dans les diverses zones grasses de l'organisme pouvant conduire à l'obésité. L'obésité a été reconnue comme une maladie en 1997 par l'OMS (Organisation Mondiale de la Santé). Sa prévention est un problème de santé publique dans les pays développés et en voie de développement. En effet, la prévalence de l'obésité dans le monde a doublé pour atteindre 31% en moins de 20 ans. **Des études épidémiologiques ont établi de fortes associations entre l'obésité et l'apparition de certaines pathologies comme les maladies cardiovasculaires ou le diabète de type 2. L'obésité est également associée à une augmentation du risque de développer un syndrome métabolique. Le syndrome métabolique consiste à un dérèglement du métabolisme.** Le métabolisme est l'ensemble des transformations moléculaires et énergétiques qui se déroulent de manière ininterrompue dans les cellules d'un organisme vivant (animaux, plantes,..). C'est un processus ordonné, qui fait intervenir des processus de de synthèse organique (anabolisme) et de dégradation (catabolisme). Toutes ces réactions provoquent la synthèse ou la perte de métabolites qui seront impliqués dans des phénomènes biologiques, puis expulsé de la cellule. Hors de la cellule, les métabolites sont localisés dans les fluides biologiques (urine, plasma, salive). Le diabète de type 2 (DT 2), les troubles cardiovasculaires et l'accident vasculaire cérébral(AVC) font parti des syndromes métaboliques décrits. Considérant l'importance de ce problème de santé publique, de nombreux projets se sont développés autour de la thématique de l'obésité et de ses complications.

2. Projet METAPROFILE financé par l'Agence Nationale pour la Recherche

Les objectifs de ces projets, financés par des programmes de recherche nationaux ou internationaux, sont de **mieux comprendre l'étiologie de l'obésité afin de définir des stratégies nutritionnelles permettant d'améliorer la santé.** Certains gènes et certaines

voies métaboliques spécifiquement impliqués dans le développement de ces maladies ont été découverts en utilisant des approches classiques et ciblées. Contrairement à ces derniers, le projet financé par l'ANR (Allocation National de Recherche) nommé **METAPROFILE**, dans lequel s'inscrit ma thèse, a pour principal objectif d'identifier des marqueurs biologiques en utilisant une approche globale afin de détecter le plus tôt possible les dérégulations métaboliques associées à la surnutrition et prédisposant au développement de désordre métabolique. Cette approche globale est d'intérêt dans le domaine de la nutrition puisque l'analyse du métabolome (ensemble des métabolites) permettra d'évaluer l'effet d'un régime, non plus sur une voie métabolique ciblée, mais au niveau du métabolisme global. **La métabolomique est une approche** qui permet de détecter des métabolites présents dans les fluides biologiques comme l'urine, le plasma ou la salive. Un grand nombre de méthodes basées sur des techniques comme **la résonance magnétique nucléaire (RMN) et la chromatographie liquide (LC) ou gazeuse (GC) couplée à la spectrométrie de masse (MS)** ont été développées pour analyser les centaines, voir les milliers de métabolites présents dans les fluides biologiques. **Dans notre projet, les protocoles biologiques ont été menés chez l'homme et la souris soumis à un régime hyper-lipidique.** Au cours de ces études cliniques, les matrices biologiques plasmatiques et urinaires ont été prélevées à différents temps cinétiques puis analysées par RMN et LC/MS. Ces analyses nous ont permis d'obtenir des empreintes métaboliques caractéristiques du phénotype de l'individu à un temps. Ces empreintes ont été ensuite analysées par des techniques statistiques univariées ou multivariées capables de déterminer les biomarqueurs modifiés au cours du protocole. La comparaison des empreintes métaboliques plasmatiques et urinaires des hommes et des souris nous ont permis d'identifier une liste de métabolites pertinents. L'identification de ces métabolites a été utilisée pour définir les voies métaboliques impliquées dans l'apparition de l'obésité.

Ce projet ANR (7) a impliqué **plusieurs partenaires** ayant chacun des objectifs et des livrables bien précis.

Objectifs	Partenaires
Intervention nutritionnelle sur la souris	INSERM Toulouse (1)
Intervention nutritionnelle sur l'homme	CRNH Lyon (2) LNH Clermont-Ferrand (3)
Analyse des fluides biologiques par RMN	INRA (4) Toulouse
Analyse des fluides biologiques par MS	INRA (4) Clermont-Ferrand/Theix
Analyse statistique	INRA (4) Clermont-Ferrand/Theix INRA (4) Toulouse
Partage des données générées	Plateforme intégrative de Bordeaux.

II. Le sujet de thèse.

1. Choix de la thèse

Suite à des études fondamentales en spectrométrie de masse, je souhaitais appliquer mes compétences analytiques à une problématique d'intérêt général. La thèse recherchée devait donc, utiliser la spectrométrie de masse dans des domaines du vivant. Après quelques semaines de recherche, le sujet de thèse proposé par l'INRA qui avait pour objectif de développer une méthode permettant la détection d'empreintes de métabolites présents dans des fluides biologiques (plasma et urine) par chromatographie liquide couplée à la spectrométrie de masse (LC/MS) semblait correspondre à mes attentes. En effet, je continuais à approfondir mes connaissances en analytique tout en me formant sur de nouvelles disciplines la nutrition et les statistiques.

2. Moyens humains mis en place en place pour réaliser ce projet

Cette thèse a été réalisée sur la Plateforme d'Exploration du Métabolisme (PFEM) dans l'équipe « Plateforme du métabolisme et spectrométrie de masse » (PMSM) appartenant à l'Unité de Nutrition Humaine (UNH) à l'INRA de Clermont-Ferrand/ Theix. Cette équipe est composée de 8 titulaires dirigés par Estelle Pujos-Guillot, ingénieur de recherche : 4 analystes spécialisés en spectrométrie de masse, une responsable qualité, un statisticien et un

bioinformaticien. Un directeur de recherche, Jean Louis Sébedio, oriente les travaux scientifiques de la plateforme.

La réalisation de ce projet a mis à contribution 4 personnes :

Personnes impliquées de février 2007 à Janvier 2010:

- 1 thésarde à temps complet : Hélène Pereira.
- 1 Ingénieur Statisticien à 31% de son temps : Jean François Martin, Ingénieur d'Etude.
- 1 encadrante à 22% de son temps : Estelle Pujos-Guillot, Ingénieur de Recherche et responsable de l'équipe PMSM.
- 1 directeur de thèse à 21% de son temps : Jean Louis Sébedio, Directeur de Recherche.

3. Le réseau développé

Les trois années de ma thèse m'ont permis de développer mon réseau professionnel essentiellement dans le domaine de la recherche en métabonomique et en spectrométrie de masse.

En métabonomique, discipline en pleine essor, j'ai rencontré des scientifiques à travers différents réseaux centrés sur cette nouvelle technique : des réseaux Nationaux (Réseau Français de Métabolomique et de Fluxomique ; RFMF), Européens (Nugo (8) ou Metabomeeting), et internationaux (Metabolomics Society).

Par ailleurs, ma spécialité initiale, la spectrométrie de masse, m'amena naturellement à faire partie de la Société Française de Spectrométrie de Masse (SFSM) qui organise chaque année des congrès à travers la France.

Au cours de ma thèse, j'ai participé à ces différentes rencontres dans le cadre de congrès (SFSM, RFMF, Metabolomics Society), de groupes de travail (Nugo, Metabomeeting) qui m'ont permis de mieux comprendre le contexte biologique de mes travaux et de confirmer ou infirmer certaines orientations que j'ai pu prendre.

J'ai également participé activement à l'organisation de la journée INR'Avenir 2008, qui consistait à réunir les acteurs du monde professionnel (privé et public) et les jeunes chercheurs. Pour réaliser cet événement traditionnel à l'INRA de Clermont-Ferrand/ Theix, nous avons constitué un groupe de 5 doctorants et d'un bioinformaticien. Afin de mener au

mieux cette journée nous avons suivi 3 formations, très constructives pour l'élaboration de ce projet, « élaboration d'un projet événementiel », « communication » et « prise de parole en public ». Cette journée a été un franc succès puisque le taux de participation a été conséquent, près d'une centaine d'étudiants et professionnels ont répondu positivement à notre invitation. La préparation de cette journée m'a permis d'élargir mon réseau aux entreprises agro-alimentaires et aux anciens doctorants de l'INRA.

III. Déroulement du projet.

1. Préparation et orientation du projet

Ce projet étant imbriqué dans un projet de recherche global METAPROFILE, tous les éléments de préparation avaient été formalisés avant mon arrivée : ainsi des objectifs précis avaient été élaborés. La finalité du projet global étant de répondre à une question biologique, l'apport analytique devait être préparé. En effet, la nouveauté de cette approche nécessitait l'élaboration d'une stratégie analytique rigoureuse afin d'exploiter au mieux les résultats biologiques obtenus.

Les principaux livrables qui incombent à mon projet étaient :

- valider les méthodes analytiques globales pour l'analyse de matrice complexe (le plasma) par spectrométrie de masse,
- déterminer des signatures biologiques à partir de métabolites observés par LC/MS chez les souris et les hommes,
- identifier des métabolites responsables de la prise de poids.

Chaque livrable avait une date de livraison prévisionnelle.

2. Conduite du projet

J'avais la responsabilité opérationnelle de la conduite du projet sous le contrôle du Directeur de Thèse et de l'Encadrante de thèse. Toutefois, deux structures formelles ont eu un rôle fondamental dans la conduite et la réorientation du projet au cours des trois années:

- Le comité de direction du projet ANR Méta-profile
- Le comité de thèse

Le comité de direction du projet ANR Méta-profile

Dans le cadre de l'ANR, des réunions d'avancement de projet ont été réalisées tous les 6 mois. A chaque réunion, les partenaires devaient présenter les avancées scientifiques et les problèmes rencontrés le cas échéant. Un représentant de l'ANR suivait ces avancées. Ces réunions m'ont permis de faire un bilan régulier et formalisé de mes recherches et ainsi d'envisager la faisabilité temporelle de certains livrables. A titre illustratif, le premier

délivrable devait être prêt en septembre 2007, or en raison de problèmes techniques, interruption de l'appareil d'analyse, et difficultés de l'exploitation des résultats, cette étape a été retardée. Néanmoins, grâce aux différentes expertises des membres de cette équipe, aux échanges d'idées et de connaissance, le premier livrable a été réalisé. Les autres livrables doivent être finalisés pour la fin de l'année 2009. Toutes les données sont actuellement en cours de traitement.

Le comité de thèse

Un comité de thèse a été réuni, composé de chercheurs spécialisés dans l'analyse chimique par spectrométrie de masse ou chromatographie liquide :

- Jean Philippe Antignac du LABERCA (Nantes),
- Christophe Junot du CEA (Saclay),
- Bruno Combourieu de l'Université Blaise Pascal (Clermont Ferrand)
- Olivier Berdeaux (INRA Dijon).

Ce comité s'est réuni à deux reprises, en fin de première année et en la fin de deuxième année de thèse. Lors de ces discussions, les échanges ont été très enrichissants et ont permis une orientation objective du projet de recherche. La réalisation de comptes-rendus annuels fut essentielle pour permettre des prises de décisions collégiales et des choix sur les orientations du projet scientifique.

La participation aux différents groupe de travail organisé par les réseaux européens (Nugo ou Metabomeeting), m'ont permis de me positionner par rapport aux avancements de ma thèse. En effet, les discussions abordées lors de ces réunions m'ont permis de réaliser que les problèmes que je rencontrais, étaient semblables à ceux d'autres équipes. Ces groupes de travail ont permis de s'échanger sur certains points techniques concrets qui restaient ambigus dans la littérature, et ainsi de mieux guider mon projet.

3. Coût du projet (Annexe 1)

Ce projet est financé entièrement par l'ANR Metaprofile, la dotation totale versée directement à l'équipe « plateforme de spectrométrie de masse et de métabolisme » (PMSM) chaque année. Le montant de cette dotation est réparti analytiquement par postes de dépenses : les déplacements, la sous-traitance des appareils, le financement de la bourse de thèse, l'équipement et la maintenance.

La consolidation du coût de la thèse fait en annexe, quantifie le projet à 278 000 euros pour les 3 années de recherche, ce qui est concordant avec le budget prévu initialement et les reporting analytiques effectués auprès de l'ANR pour la globalité du projet Metaprofile.

IV. Principales compétences développées au cours de la thèse

1. Compétences Scientifiques

Chromatographie Liquide	Maintenance et développement de méthodes sur une UPLC Acquity et HPLC Alliance 2600 de chez Waters
Spectrométrie de Masse	Maintenance, diagnostic et réparation de la source (<i>electrospray</i>) et analyseur (Quadripôle couplé à un temps de vol) de chez Waters. Elaboration de méthode analytique.
Préparation d'échantillons	Développement méthodologique de la préparation d'échantillons sur des matrices biologiques complexes (plasma & urine)
Traitement des données/ interprétation des résultats.	Utilisation de logiciels de traitement du signal (Masslynx). Utilisation de logiciels spécifiques (XCMS, Markerlynx) qui permettent de comparer les données de spectrométrie de masse. Utilisation de logiciels d'analyse statistiques (ExcelStat, SAS Simca-P). Utilisation de base de données afin d'interpréter biologiquement les résultats. (HMDB, Kegg, Metlin)

2. Connaissances méthodologiques

Veille scientifique et bibliographique	Rechercher, sélectionner, traiter et organiser l'information scientifique
Conduite de projet	Gestion du temps, planification des tâches, classification des priorités. Gestion des moyens matériels. Organisation et animation des réunions du comité de thèse : Rapport d'avancement semestriel, livraison des livrables.
Méthodes d'analyse.	Mettre au point des méthodologies pour recueillir des résultats : Détermination de critères de validation nécessaire à la prise de décision, Utilisation de plan factoriel afin d'évaluer les logiciels utilisés. Mise en place d'une méthodologie rigoureuse pour tester tous les paramètres analytiques afin de valider la méthode mise en place.
Démarche qualité	La démarche qualité ISO 9000 mise en place au sein de la plateforme de spectrométrie de masse m'a conduit à mettre à jour des fiches existantes, et créer des fiches d'instruction propre aux appareils utilisés.
Valorisation scientifique	Hierarchisation des idées. Illustration scientifique claire des résultats. Rédaction de manuscrit (français et anglais). (Annexe 2).

3. Qualités personnelles exprimées

Autonomie	Elaboration de stratégie, choix des sujets à aborder. Prise de décision.
Organisation & Rigueur	Rédaction quotidienne des actions à faire. Organisation et animation de réunion. Compte rendu détaillé des points importants et à développer le cas échéant.
Communication	Elaboration de poster et de communication orale de mes travaux au sein de l'équipe, de l'unité et de congrès internationaux et de collègues doctorants provenant d'autres domaines.
Créativité & Persévérance	La métabolomique est une discipline nouvelle et la démarche à développer nécessitait de formuler des hypothèses en évitant les idées préconçues. Analyser le besoin en métabolomique et mettre en place des stratégies pour obtenir une méthode plus rigoureuse était compliquée et contenait beaucoup de point d'incertitude. Il a été nécessaire de reformuler les problèmes lié à la métabolomique par la décomposition des éléments du problème
Motivation	Malgré mes difficultés à rédiger, à prendre la parole en public et à parler anglais, j'ai continué à m'inscrire pour des oraux ou poster dans différents congrès (nationaux et internationaux). J'ai mené de front : mon activité dans l'association des non titulaires de l'INRA et ma thèse.

4. Réseau professionnel développé

ANR METAPROFILE	Développement de liens avec des biologistes : Remy Bircellin, Alain Paris, une bioinformaticienne: Hélène Dumazet et des analystes.
Société Française de Spectrométrie de Masse.	Conservation des liens avec le laboratoire de Jean Claude Tabet et Carlos Afonso (Paris) et de Jean François Muller (Metz).
Réseau Français de Métabolomique & Fluxomique	Développement de liens avec les acteurs de la métabolomique en France au CEA Saclay, LABERCA de Nantes, Plateforme intégrative de Bordeaux, INRA de Toulouse.
Nugo	Développement de liens avec de grands instituts de nutrition en Europe : TNO (Pays Bas).
Ancien doctorants de l'INRA	A l'aide du directeur du site de Theix, Mr Becker, nous mettons en place un annuaire des anciens doctorants afin d'aider les jeunes doctorants à rentrer dans le monde professionnel.

V. Conclusion

Cette thèse m'a permis de découvrir des domaines scientifiques nouveaux : la biologie, l'épidémiologie, la réalisation d'études cliniques et les statistiques multivariées.

Cette expérience a contribué au développement de ma rigueur, en adoptant des méthodes de travail efficaces et ma créativité en faisant évoluer ce projet avec de nouvelles idées ou approches. Aidée par la recherche bibliographique, par les discussions et les échanges avec différents partenaires et chercheurs, mon esprit d'analyse et de synthèse m'ont permis d'avoir une gestion claire du projet, qualité nécessaire à des postes à responsabilité.

Une partie de ma mission consistait à présenter mes avancées scientifiques. Les méthodes de communication, de la construction du raisonnement au développement de l'argumentaire, m'ont permis d'apprendre à exposer clairement des idées. Ainsi cette expérience m'a permis de prendre confiance en moi et d'améliorer mon expression orale et écrite. L'échange scientifique et le travail en équipe ont été essentiels dans mon épanouissement au sein de cette équipe.

Je souhaiterais continuer à appliquer mes connaissances analytiques dans des domaines tels que les sciences du vivant, la pharmacologie ou l'environnement. Par ailleurs, les problèmes techniques rencontrés au cours de ma thèse m'ont donné envie de participer au développement instrumental de ces outils et d'élargir mes connaissances analytiques, sur de nouveaux appareils tels que l'Orbitrap ou le FTICR. Le choix entre laboratoire public ou privé pour mon futur emploi se fera essentiellement en fonction des objectifs du poste proposé et des opportunités qui vont réellement se concrétiser. Idéalement, je recherche un poste pour lesquels des challenges dans le domaine de la spectrométrie de masse sont importants pour capitaliser et développer mes connaissances et expertises déjà construites lors de ces trois dernières années.

Lexique

1. Structures & programmes

INSERM Institut National de la Santé et de la Recherche Médicale (1)

CRNHL Centre de Recherche en Nutrition Humaine de Lyon(2)

LNH Laboratoire de Nutrition Humaine (3)

INRA Institut National de Recherche Agronomique (4)

CEA Commissariat à l'Energie Atomique (5)

LABERCA Laboratoire d'Etudes des Résidus et Contaminants dans les Aliments (6)

ANR allocation Nationale de Recherche(7)

NUGO The European Nutrigenomics Organisation (8)

PMSM Plateforme Métabolisme et Spectrométrie de Masse

PFEM Plateforme d'Exploration du métabolisme

UNH Unité de Nutrition Humaine

RFMF Réseau Français de Métabolomique et de Fluxomique ;

SFSM Société Française de Spectrométrie de Masse

2. Termes scientifiques et techniques

RMN Résonance Magnétique Nucléaire

LC Chromatographie Liquide

GC Chromatographie Gazeuse

MS Spectrométrie de Masse

Annexe 1 Evaluation du coût de la thèse

	Nature de la dépense	Détails *		Coûts totaux (euros TTC)			
				Nombre d'unités	Coût unitaire moyen en €	Quote-part utilisation	Total
1	Ressources Humaines	Salaire Brut /mois	Charges	Nombre de mois	Salaire chargé/ mois	Quote-part utilisation	Total
1.1	Hélène Pereira	1664	499	36	2163	1	77868
1.2	Encadrante 1 : Estelle Pujos Guillot	2500	630	36	3130	0.22	24789.6
1.3	Encadrant 2 : Jean Louis Sébédio	5800	1740	36	7540	0.21	57002.4
1.4	Ingénieur : Jean François Martin	3200	840	36	4040	0.31	45086.4
	Sous-total Ressources Humaines						204746.4

2	Consommables			Nombre d'unités	Coût unitaire	Quote-part utilisation	Total
2.1	Fournitures expérimentales						
	1-Produits chimiques			indéfini	indéfini	1	2531.2
	2-Vials, inserts, verrerie, filtres			indéfini	indéfini	1	6533.4
	3-Seringue			1	102	1	102
	4-Colonnes			4	594	1	2371.1
	5-Gaz			4	120	1	480.8
2.2	Fournitures de bureau			indéfini	indéfini	1	478.4
	Sous-total Consommables						12496.9

3 Infrastructures		Coût annuel de l'équipe PMSP	Coût mensuel de l'équipe PMSM	Nombre de mois	Coût unitaire mensuel	Quote-part utilisation de l'équipe	Total
3.1	Entretien, gardiennage, secrétariat	3000	250	36	18	0.07	53
3.2	Loyers des locaux	792	66	36	5	0.07	167
3.3	Electricité, eau, chauffage	18114	1509.5	36	106	0.07	3800
3.4	Autres						
	1-Essence	1884	157	36	11	0.07	396
	2-Traitements des déchets	252	21	36	2	0.07	53
	3-Mise en place des normes	5988	499	36	35	0.07	1363
	4-Hygiène et sécurité	960	80	36	6	0.07	202
	5-Matériels moyens	18459	1538.25	36	108	0.07	3875
	Sous-total Infrastructures						9909

Nouveau Chapitre de la thèse

4	Matériel (amortissements)	Taux d'amortissement en début de thèse	Durée de l'amortissement	Année de prise en compte	Coût unitaire	Quote-part utilisation	Total
4.1	Matériel d'expérimentation						
	1-Spectrometrie de masse couplée à une chromatographie liquide-acheté en 2005	40%	5 ans	3	250000	0.25	37500
	2-Centrifugeuse-acheté en 2008	0%	3 ans	2	6100	0.07	285
	3-Evaporateur sous azote-acheté en 2008	0%	3 ans	2	2500	0.07	118
	4-Pousse seringue-acheté en 2007	0%	3 ans	3	2000	0.25	500
4.2	Matériel informatique						
	1-Ordinateur de bureau-acheté en 2007	0%	3 ans	3	1890	1	1890
	2-Disque dure-acheté en 2007	0%	1 an	1	396	1	396
	3-Logiciels-acheté en 2007	0%	1 an	1	252	1	252
	Sous-total Matériel						40941

Nouveau Chapitre de la thèse

5 Déplacements	Transports/ Hébergements	Inscription	Nombre de jours	Coût unitaire	Quote-part utilisation	Total
5.1 Réunion de thèse						
1-Réunion ANR(*3)	250	0	3	83	1	250
2-Comité de thèse(*2)	1000	0	2	500	1	1000
5.2 Congrès en France						
1-Congrès de la Société Française de Spectrométrie de Masse (2007)	100	150	3	83	1	250
Congrès de la Société Française de Spectrométrie de Masse (2008)	100	150	3	83	1	250
2-Congrès de Métabolomique et Fluxomique(2008)	100	150	3	83	1	250
3-Metabomeeting (2008)	50	bourse	2	25	1	50
5.3 Congrès à l'étranger						
1-Nugo Workshop (Copenhague-2009)	400	0	4	100	1	400
2-Metabolomics Society (Edmonton-2009)	300	1200	5	300	1	1500
Sous-total Déplacements						3950

6 Formation	Transports/ Hébergements	Inscription	Nombre de jours	Coût unitaire moyen en €	Quote-part utilisation	Total
6.1 Formation						
1-Validation de Méthode	400	500	5	80	1	900
2-Anglais	0	120	5	24	1	120
3-Prendre la parole en public/ animation de débats	0	220	1	220	1	220
4-Gérer un projet et communication	0	280	2	140	1	280
5-Ecole d'été	80	700	5	156	1	780
6-Applied Biosystem (analyse des lipides)	50	0	1	50	1	50
6.2 Autres frais (Inscription à l'Université, Sécu.)	0	1350	indéfini	indéfini	1	1350
Sous-total Formation						3700

7	Documentation et communication	Coût annuel de l'équipe PMSP	Coût mensuel de l'équipe PMSM	Nombre de mois	Coût unitaire	Quote-part utilisation	Total
7.1	Affranchissements, Internet, téléphone	6456	538	36	38	0.07	1356
7.2	Impression Poster	62	5	36	5	1	180
7.3	Impressions	992	83	36	6	0.07	210
7.4	Documentation (périodiques, livres, bases de données, bibliothèque, etc.)	466	39	36	3	0.07	100
	Sous-total Documentation et communication						1846
8	TOTAL						277589.3

Annexe 2 Valorisation Scientifique

Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. Publication. *Hélène Pereira, Jean François Martin, Charlotte Joly, Jean Louis Sébédio, Estelle Pujos-Guillot.* *Metabolomics* (acceptée).

A strategy to analyse large batches of samples in metabolomic study. *Hélène Pereira, Jean François Martin, Jean Louis Sébédio, Estelle Pujos-Guillot.* (en cours).

Mise en place d'une méthode d'analyse de plasma dans le cadre d'une étude en métabonomique. Communication Orale et poster. *Hélène Pereira, Jean François Martin, Jean Louis Sébédio, Estelle Pujos-Guillot.* Congrès du réseau français de métabolomique et fluxomique.(Bordeaux). Février 2008.

Identification de biomarqueurs précoces des perturbations métaboliques associées à la prise de poids par spectrométrie de masse par une approche en métabolomique. Communication Orale. *Hélène Pereira, Jean François Martin, Jean Louis Sébédio, Estelle Pujos-Guillot.* Animation Scientifique INRA de Clermont-Fd-Theix. Juin 2008 .

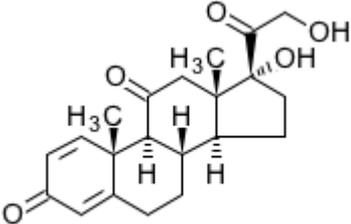
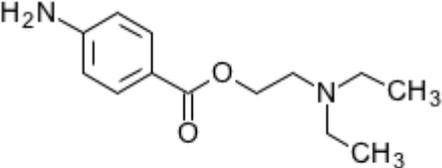
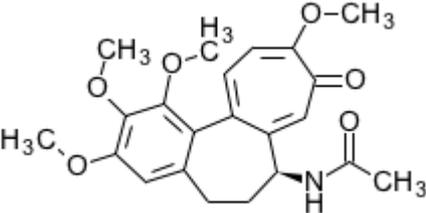
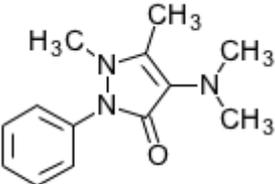
Untargeted analysis by LC/ESI/Qtof of urines of mice fed with a hight fat diet. Communication orale. *Hélène Pereira, Jean François Martin, Jean Louis Sébédio, et Estelle Pujos Guillot.* Société Française de Spectrométrie de Masse. (Grenoble). Septembre 2008.

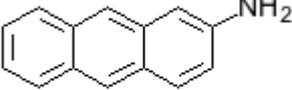
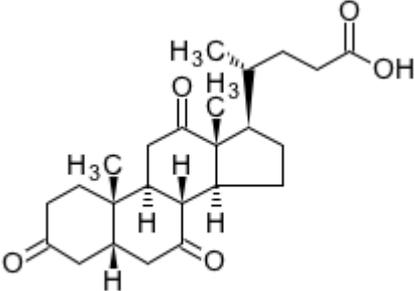
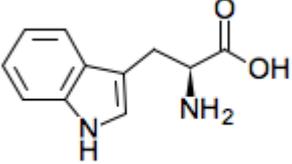
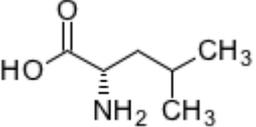
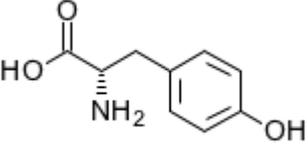
Metabolomics on large batches of samples: problems and solutions for extracting valuable information. Communication orale et poster. *Hélène Pereira, Jean François Martin, Remy Burcelin, Alain Paris, Jean Louis Sébédio, Estelle Pujos-Guillot.* Workshop Nugo (Copenhague). Février 2009.

Methods development for metabolic profiling of plasma samples for nutrition studies. Poster. *Hélène Pereira, Jean François Martin, Remy Burcelin, Alain Paris, Jean Louis Sébédio, Estelle Pujos-Guillot.* Congrès Metabolomics Society (Edmonton Canada). Septembre 2009

ANNEXE

ANNEXE I (CHAPITRE II)

Prednisone	 <p>The chemical structure of Prednisone is a corticosteroid. It features a four-ring steroid nucleus with a ketone group at C3, a double bond between C4 and C5, and methyl groups at C10 and C13. At C17, there is a side chain consisting of a two-carbon ethyl group and a two-carbon chain ending in a hydroxyl group and a hydroxymethyl group.</p> <p>C07370</p>
Procaine	 <p>The chemical structure of Procaine is an ester. It consists of a benzene ring with an amino group (-NH₂) at the para position and a procaine ester group (-COOCH₂CH₂N(CH₂CH₃)₂) at the other para position.</p> <p>C07375</p>
Colchicine	 <p>The chemical structure of Colchicine is a complex polycyclic alkaloid. It features a decalin core with a lactone ring fused to one of the decalin rings. The structure is substituted with several methoxy groups (-OCH₃) and an acetamido group (-NHCOCH₃).</p> <p>C07592</p>
4 Aminopyridine	 <p>The chemical structure of 4 Aminopyridine is a pyridine ring with an amino group (-NH₂) at the 4-position. The pyridine ring is substituted with a methyl group at the 2-position and a dimethylamino group (-N(CH₃)₂) at the 3-position.</p> <p>C07539</p>

Aminoanthracène	 C14417
Acide dehydrocholic	 C13154
Tryptophane	 D00020
Leucine	 D00030
Tyrosine	 D00022

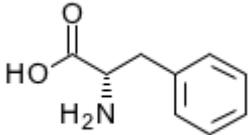
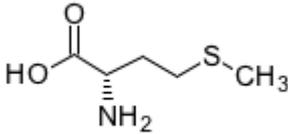
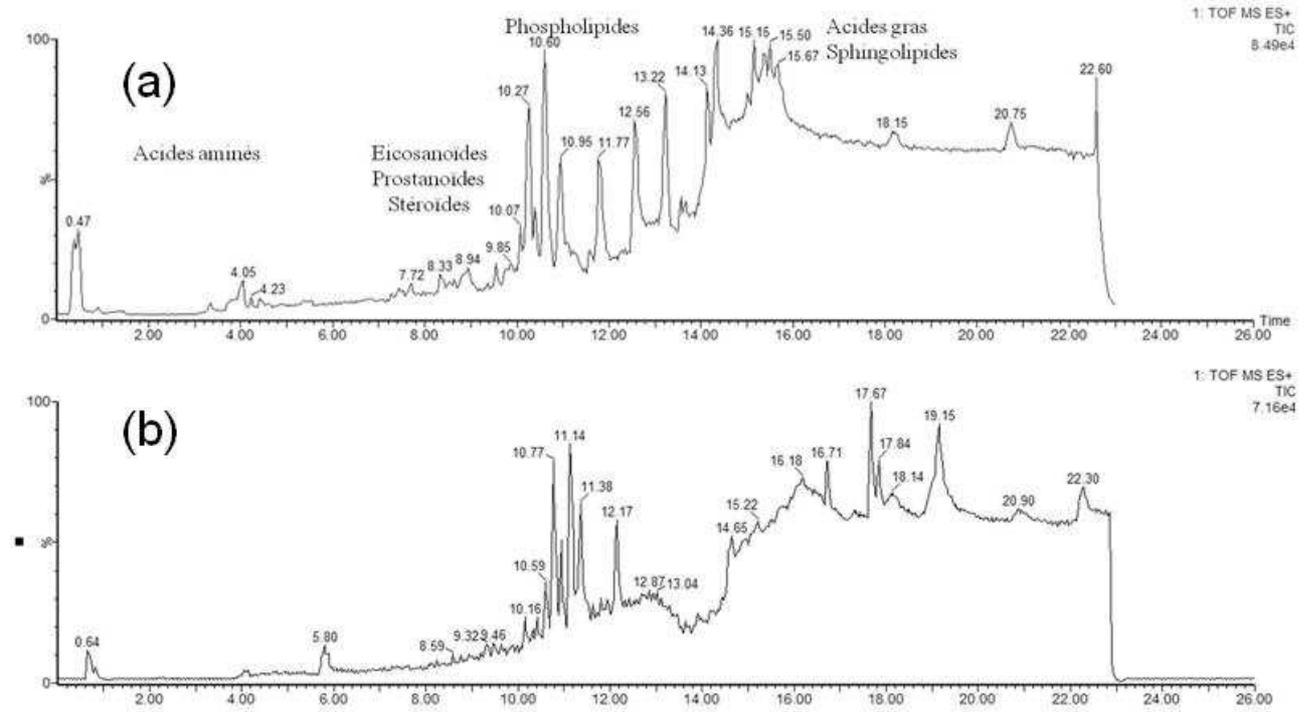
Phénylalanine	 <p>D00021</p>
Méthionine	 <p>D00019</p>

Tableau 1_ Structure des standards biologiques et de xénobiotiques étudiés.

ANNEXE 2 (CHAPITRE II)



ANNEXE 3 (CHAPITRE III)

Individus ayant suivi le régime hyper-lipidique pendant 56-semaines

	Temps de rétention	Masse	angfr_0OK	angfr_14OK	angfr_28OK	angfr_56OK	barri_14OK	barri_56OK
	Ions extraits et alignés	22.64	70.01	260.2463	271.3848	307.8416	0	0
0.47		70.07	366.5503	336.2773	320.1655	364.5176	196.4729	206.6116
15.47		71.09	208.6075	216.7085	209.62	0	150.8861	215.6961
8.81		71.59	94.6752	0	49.2125	54.4254	90.1322	0
0.49		72.08	419.2585	292.6716	256.1985	248.0752	258.2668	180.2504
8.8		72.12	0	0	0	0	0	0
22.73		72.94	0	233.4347	227.4534	276.4502	301.8225	275.4391
14.64		74.06	0	0	0	0	0	0
15.31		74.06	0	0	0	0	0	0
16.26		74.07	0	0	0	50.9919	0	0
17.02		74.07	0	0	0	0	0	0
17.95		74.08	73.6182	0	0	0	0	0
21.42		74.09	58.0819	0	0	29.1164	0	0
20.8		74.09	44.0167	0	0	0	0	0
22.2		74.09	0	0	0	0	0	0
15.73		457.37	0	0	0	0	0	0
9.26		458.35	0	0	0	0	0	0
4.95		459.29	113.0856	151.8068	106.6606	85.0652	122.2581	101.6114

Individu avec un IMC initial : 25

Individu avec un IMC initial : 25

Représentation d'une matrice de donnée

En enfin, je terminerai par la phrase de mon collègue Monsieur Bernard LYAN si souvent répétée durant ces trois années de thèse : « Ce n'est pas que je m'ennuis ici, mais le devoir m'appelant ailleurs je m'en vais de ce pas le rejoindre. »... Merci Bernard !

RESUME EN FRANCAIS:

L'étude du métabolome est un challenge analytique important, de par la grande diversité chimique des métabolites présents à des concentrations très variables dans les matrices biologiques. Ce travail présente une stratégie analytique mettant en œuvre le couplage chromatographie liquide / spectrométrie de masse afin d'acquérir des profils métaboliques de fluides biologiques et d'identifier des biomarqueurs potentiels de dysfonctionnements métaboliques. Une attention particulière a été donnée à l'étude des sources de variabilité analytique, depuis l'étape de préparation d'échantillons jusqu'au traitement des données. Un protocole optimisé a été proposé et appliqué à une intervention nutritionnelle (ANR METAPROFILE). L'identification a été basée sur l'utilisation des bases de données, ainsi que sur l'interprétation des informations spectrales (profils isotopiques et fragmentations MSn).

TITRE EN ANGLAIS :

Development of a metabolomic approach by mass spectrometry. Application to Nutrition.

RESUME EN ANGLAIS :

The study of small molecules, defined as the metabolome, is an important analytical challenge due to the wide variety of metabolites with different physico-chemical properties, which are present at different concentrations. We report, here, an analytical strategy using mass spectrometry coupled to liquid chromatography (LC/MS) to acquire metabolic profiles from biofluids and identify potential biomarkers of metabolic dysfunctions. Special attention was given to the analytical variability from sample preparation to data processing. An optimized workflow was proposed and applied to a nutritional intervention (ANR METAPROFILE). The biomarker identification process included database queries and study of spectral informations from isotopic patterns to MSn fragmentations.

DISCIPLINE :

Ecole Doctorale des Sciences Fondamentales – Spécialité Chimie

MOTS CLES :

Métabolomique, Spectrométrie de masse, Métabolisme, Nutrition

INTITULE ET ADRESSE DU LABORATOIRE :

Plateforme d'Exploration du Métabolisme - Unité de Nutrition Humaine UMR1019

Centre de Clermont-Ferrand / Theix – 63122 Saint Genès-Champanelle