



HAL
open science

Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse

Pierre Barbier Saint Hilaire

► **To cite this version:**

Pierre Barbier Saint Hilaire. Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse. Chimie analytique. Université Paris Saclay (COMUE), 2019. Français. NNT : 2019SACLS193 . tel-02435982

HAL Id: tel-02435982

<https://theses.hal.science/tel-02435982>

Submitted on 12 Jan 2020

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.

Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris Sud

École doctorale n°571 Sciences Chimiques : Molécules, Matériaux,
Instrumentation et Biosystèmes (2MIB)
Spécialité de doctorat : Chimie

Thèse présentée et soutenue à Gif-sur-Yvette, le 10 juillet 2019 par

Pierre BARBIER SAINT HILAIRE

Composition du Jury :

Pierre Chaminade Professeur, Université Paris Sud	Président
Patrick Emond Professeur, Université de Tours	Rapporteur
Gérard Hopfgartner Professeur, Université de Genève	Rapporteur
Estelle Pujos-Guillot Ingénieure de Recherche, INRA	Examineur
Christophe Junot Cadre scientifique des EPIC, CEA	Directeur de thèse
Nicolas Weiss Maitre de Conférences des Universités – Praticien Hospitalier, Hôpital Saint Antoine	Invité
François Fenaille Cadre scientifique des EPIC, CEA	Invité

RESUME

La métabolomique non ciblée est une approche relativement récente visant à caractériser le contenu métabolique d'un échantillon. Cette caractérisation à l'échelle du métabolite permet d'évaluer le plus finement possible les changements phénotypiques ayant lieu au sein de ces échantillons. De ce fait, la métabolomique tend à se développer dans différents domaines comme l'étude de cohortes cliniques. Au cours de cette thèse, une cohorte constituée de patients souffrant d'encéphalopathie hépatique (EH) a été étudiée. L'EH est une maladie dont les mécanismes sont encore mal compris mais ayant un impact important puisque les cas les plus graves causent près de 20 000 hospitalisations par an en France. L'utilisation de la métabolomique a permis de montrer que l'EH pouvait être liée à une altération du métabolisme énergétique au niveau cérébral. D'autres approches « -omiques » phénotypiques, comme la lipidomique ou la glycomique, ont également été utilisées pour appuyer cette hypothèse. Cela a également permis de montrer que ces différentes approches pouvaient se combiner à la métabolomique pour approfondir les dérèglements observés chez les patients. La réalisation de cette étude a également mis en évidence les facteurs limitants ou les étapes plus complexes d'une analyse métabolomique. La suite de ces travaux avait pour but d'évaluer ou de développer de nouvelles approches analytiques ou interprétatives afin d'augmenter la performance de la métabolomique non ciblée en termes de détection, d'annotation et d'identification des métabolites. Dans une première partie, nous avons montré que l'utilisation de nouveaux instruments comme l'Orbitrap Fusion, permettant d'atteindre de très hautes résolutions (500 000, $M/\Delta M$, largeur du pic à mi-hauteur), donne accès à la structure isotopique fine des métabolites (détection des isotopologues ^{13}C , mais aussi ^{18}O et double ^{13}C), ce qui est une réelle plus-value dans l'identification des composés inconnus. Dans une seconde partie ; toujours dans le cadre de l'identification de composés inconnus ; nous avons cherché à mieux comprendre les mécanismes de fragmentation des ions lors de la réalisation d'expériences de MS^2 . A partir de deux exemples, différents outils ont été utilisés, comme le marquage isotopique ou le calcul quantique, afin de faire ressortir les règles chimiques impliquées dans la fragmentation. La détermination de ces règles est indispensable pour pouvoir interpréter *de novo* les spectres de fragmentation. Enfin, une troisième partie a montré que les améliorations apportées aux instruments de type Orbitrap en termes de vitesse d'acquisition permettaient d'envisager la mise en place de nouveaux modes d'acquisition de données jusque-là essentiellement réservés aux analyseurs de type quadripôle-temps de vol. Ces modes d'acquisition (données dépendantes et données indépendantes) sont basés sur l'acquisition multi-événements de spectres MS et MS^2 . Ils permettent d'augmenter significativement le nombre d'expériences MS^2 donnant ainsi des informations structurales sur un grand nombre de métabolites beaucoup plus rapidement qu'avec les approches classiques. Toutes ces améliorations proposées dans le cadre de la métabolomique devraient rapidement être utilisables et applicables à l'étude de nouvelles cohortes médicales.

COMMUNICATIONS SCIENTIFIQUES

Publications scientifiques

- Evaluation of the High-Field Orbitrap Fusion for Compound Annotation in Metabolomics

Pierre Barbier Saint Hilaire, Ulli M. Hohenester, Benoit Colsch, Jean-Claude Tabet, Christophe Junot, François Fenaille

Analytical Chemistry, 2018, 90, 3030-3035

- Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand?

François Fenaille, Pierre Barbier Saint Hilaire, Kathleen Rousseau, Christophe Junot

J. Chrom A, 2017, 1526, 1-12

- Mechanistic study of competitive releases of H₂O, NH₃ and CO₂ from deprotonated aspartic and glutamic acids: Role of conformation

Pierre Barbier Saint Hilaire, Anna Warnet, Yves Gimbert, Ulli M. Hohenester, Gianluca Giorgi, Marie-Françoise Olivier, François Fenaille, Benoit Colsch, Christophe Junot, Jean-Claude Tabet

J. Chrom B, 2017, 1047, 64-74

- Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy

Nicolas Weiss, Pierre Barbier Saint Hilaire, Benoit Colsch, Foucauld Isnard, Suleiman Attala, Augustin Schaefer, Maria del Mar Amador, Marika Rudler, Foudil Lamari, Frédéric Sedel, Dominique Thabut, Christophe Junot

J Hepatol, 2016, 65(6):1120-1130

Communications orales

- Evaluation of the High-Field Orbitrap Fusion for determination of elemental composition in Metabolomics

13th conference of European Fourier Transform Mass Spectrometry, Freising, Avril 2018

- An integrated approach combining metabolomics, lipidomics and glycomics on plasma to highlight metabolism dysregulation in Hepatic Encephalopathy

13th conference of the Metabolomics Society, Brisbane, Juin 2017

- Metabolomics of cerebrospinal fluid and plasma for the characterization of hepatic encephalopathy

2nde Journée de l'école doctorale 2MIB, Saclay, Novembre 2016 – Prix de la meilleure présentation orale

- Apport de la technologie Orbitrap Fusion pour l'exploitation du massif isotopique en métabolomique

9^{èmes} Journées Scientifiques du Réseau Francophone de Métabolomique et de Fluxomique, Lille, Juin 2015

Communications par poster

- Role of the residual water into a linear ion trap cell on re-orientation of aspartate and glutamate anion dissociations

65th Conference on Mass Spectrometry and Allied Topics, Indianapolis, Juin 2017

- Relative Isotopic Abundance on Orbitrap Fusion: Input of Accurate ¹³C Measurement and ¹⁸O/¹³C₂ Separation and Measurement for Annotation in Metabolomics

65th Conference on Mass Spectrometry and Allied Topics, Indianapolis, Juin 2017

- Analyses métabolomiques du liquide céphalo-rachidien et du plasma pour la caractérisation de l'encéphalopathie hépatique

10^{èmes} Journées Scientifiques du Réseau Francophone de Métabolomique et de Fluxomique, Juin 2016 – Prix du public

- Origin of the regioselective $[\text{NH}_3+\text{CO}_2]$ concomitant losses from aspartate predicted by calculations and evidenced by ^{13}C and D labeling

64th Conference on Mass Spectrometry and Allied Topics, San Antonio, Juin 2016

- Origin of the specific ammonia loss from deprotonated aspartic acid under low energy collision-induced dissociation conditions

63th Conference on Mass Spectrometry and Allied Topics, Saint Louis, Juin 2015

TABLE DES MATIERES

RESUME	3
COMMUNICATIONS SCIENTIFIQUES	5
TABLE DES MATIERES	9
ABREVIATIONS	15
LISTE DES FIGURES	19
LISTE DES TABLEAUX	20
LISTE DES ANNEXES	21
Introduction	23
Les approches « -omiques »	25
1 Description générale	25
1.1 La génomique	25
1.2 L'étude du phénotype	26
1.2.1 Transcriptomique	26
1.2.2 Protéomique	27
1.2.3 Métabolomique	28
2 Les approches combinées	29
La métabolomique non ciblée	31
1 Développement de la métabolomique	31
2 Les méthodes d'analyse physico-chimiques utilisées en métabolomique	32
2.1 La résonance magnétique nucléaire (RMN)	32
2.1.1 Principe général	32
2.1.2 Intérêts et limites en métabolomique	33
2.2 La spectrométrie de masse (MS)	34
2.2.1 Principe général	34
2.2.2 Avantages et inconvénients par rapport à la RMN	34
2.3 Développement d'analyses combinant plusieurs techniques	35
3 Les différentes étapes d'une analyse métabolomique par spectrométrie de masse	36
3.1 Schéma général d'une analyse métabolomique par HRMS et points critiques	36
3.2 Préparation d'échantillons	38
3.2.1 Prélèvement et stockage des échantillons	38
3.2.1.1.1 Echantillons sanguins	38
3.2.1.1.2 Urine	40
3.2.1.1.3 Liquide céphalorachidien	41
3.2.2 Extraction des métabolites	42
3.2.2.1.1 Plasma/sérum	42
3.2.2.1.2 Urines	44

3.2.2.1.3	Liquide céphalorachidien	45
3.3	Acquisition des données	45
3.3.1	Les techniques séparatives en ligne	45
3.3.1.1.1	La chromatographie en phase gazeuse	45
3.3.1.1.2	La chromatographie liquide	47
3.3.1.1.3	La chromatographie en phase supercritique	48
3.3.1.1.4	L'électrophorèse capillaire	49
3.3.1.1.5	L'introduction directe	49
3.3.2	Les nouveaux instruments en spectrométrie de masse à haute résolution	50
3.3.3	Les différents modes d'acquisition de données	53
3.4	Travail sur le jeu de données	53
3.4.1	Validation du déroulement de l'analyse	54
3.4.1.1.1	Utilisation d'étalons	54
3.4.1.1.2	Utilisation de contrôles qualité (QC)	55
3.4.1.1.3	Utilisations d'échantillons « blancs »	56
3.4.2	Outils informatiques pour la détection automatique des signaux	56
3.4.3	Application de filtres analytiques	59
3.4.4	Normalisation des signaux	60
3.4.5	Annotation du jeu de données	61
3.4.6	Analyses statistiques	62
3.4.6.1.1	Outils statistiques univariés	63
3.4.6.1.2	Outils statistiques multivariés	64
3.4.7	Interprétation biologique	64
4	Les modes d'acquisition de données en spectrométrie de masse	66
4.1	L'acquisition par balayage MS	66
4.2	L'acquisition de données MS ²	66
4.2.1	Les différents modes de fragmentation	66
4.2.2	Les différents modes d'acquisition MS ²	69
4.2.2.1.1	Le mode ciblé	69
4.2.2.1.2	Le mode données dépendant	70
4.2.2.1.3	Le mode données indépendant	71
4.2.3	Choix du type d'analyseurs pour l'acquisition multi-événements de données	72
5	L'identification des métabolites	73
5.1	Identification des composés grâce aux bases de données internes	73
5.1.1	Utilisation de la masse précise et du temps de rétention	73
5.1.2	Confirmation par les spectres MS ²	75
5.2	Identification des composés absents des bases de données internes	75
5.2.1	Détermination de la formule brute	76
5.2.1.1.1	Apport de la masse précise	76
5.2.1.1.2	Apport des règles chimiques	77
5.2.1.1.3	Apport de l'abondance isotopique relative (RIA)	77
5.2.2	Prédiction des temps de rétention en LC	78
5.2.3	Prédiction et interprétation des spectres de fragmentation	80
5.2.3.1.1	Elucidation <i>de novo</i> des spectres MS ²	80
5.2.3.1.2	Utilisation des outils informatiques	81

Bilan	85
Présentation du projet de recherche	87
Résultats	91
Chapitre 1. Application à l'étude de l'encéphalopathie hépatique	93
1 L'encéphalopathie hépatique	93
1.1 Définition et classification	93
1.2 Physiopathologie	94
1.3 Diagnostic	95
1.4 Prévalence et impact	96
1.5 Traitements actuels	97
2 Objectifs des études	98
3 Aspects méthodologiques	98
4 Publication numéro 1	99
4.1 Principaux résultats	111
4.2 Conclusion de l'étude	112
5 Seconde étude	112
5.1 Principaux résultats	112
5.2 Conclusion de l'étude	117
6 Analyses des difficultés rencontrées lors de l'analyse de cohortes médicales	118
Chapitre 2. Développement d'outils d'aide à l'identification des métabolites	121
1 Les problématiques de l'annotation en métabolomique	121
2 Pertinence de l'utilisation des données d'abondance isotopique relative sur l'Orbitrap Fusion pour la détermination de la formule brute	122
2.1 Contexte	122
2.2 Objectifs de l'étude	123
2.3 Aspects méthodologiques	123
2.4 Publication numéro 2	124
2.5 Principaux résultats	131
2.6 Conclusion	131
3 Interprétation des mécanismes de fragmentation	132
3.1 Contexte	132
3.2 Objectifs de l'étude	133
3.3 Approche méthodologique	133
3.4 Publication numéro 3	134
3.5 Principaux résultats	146
3.6 Résultats supplémentaires	146
3.7 Conclusion	147
Chapitre 3. Développement de nouvelles méthodes d'acquisition	151
1 Contexte	151
2 Publication numéro 4	151
3 Implémentation des nouveaux modes d'acquisition sur l'Orbitrap Fusion	164
3.1 Objectifs de l'étude	164
3.2 Aspects méthodologiques	164

3.3	Publication numéro 5	165
3.4	Principaux résultats	192
3.5	Conclusion	194
	Chapitre 4. Discussion générale	197
	CONCLUSION GENERALE	207
	REFERENCES BIBLIOGRAPHIQUES	209
	ANNEXES	231

ABREVIATIONS

2D : bidimensionnelle

ADN : Acide DésoxyriboNucléique

AGC : Contrôle automatique du gain, *Automatic Gain Control*

AIF : Fragmentation des ions de toute la gamme de masse, *All Ion Fragmentation*

APCI : Ionisation chimique à pression atmosphérique, *Atmospheric Pressure Chemical Ionisation*

APPI : Ionisation par photoionisation à pression atmosphérique, *Atmospheric Pressure Photolonisation*

ARN : Acide RiboNucléique

CE : Electrophorèse capillaire, *Capillary Electrophoresis*

CID : Dissociation induite par collision après activation résonnante, *Collision Induced Dissociation*

DDA : Acquisition données dépendante, *Data Dependent Acquisition*

DIA : Acquisition données indépendante, *Data Independent Acquisition*

EH : Encéphalopathie Hépatique

ERMS : Spectrométrie de masse résolue en énergie, *Energy-Resolved Mass Spectrometry*

ESI : Ionisation par électrospray, *Électrospray Ionisation*

FIA : Introduction directe en flux, *Flow Injection Analysis*

FT-ICR : Résonance Cyclotronique des ions à transformée de Fourier, *Fourier Transform-Ion Cyclotronic Resonance*.

GC : Chromatographie en phase gazeuse, *Gas Chromatography*

FWMH : Largeur à mi-hauteur, *Full Width Mid Height*

HCD : Dissociation induite par collision après activation non-résonnante, *Higher-energy Collision Dissociation*

HILIC: Chromatographie par interaction hydrophile, *Hydrophilic Interaction Liquid Chromatography*

HPLC : Chromatographie liquide haute performance/pression, *High Performance/Pressure Liquid Chromatography*

HRMS : Spectrométrie de masse à haute résolution, *High Resolution Mass Spectrometry*

LC : Chromatographie en phase liquide, *Liquid Chromatography*

LCR : Liquide CéphaloRachidien

m/z : Rapport masse sur charge

MALDI : Désorption ionisation laser assistée par matrice, *Matrix Assisted Laser Desorption Ionisation*

MS : Spectrométrie de masse, *Mass Spectrometry*

MS² : Spectrométrie de masse en tandem, *Tandem Mass Spectrometry*

OPLS-DA : *Orthogonal Projection of Latent Structures Discriminant Analysis*

PLS-DA : *Partial Least Square Discriminant Analysis* ou *Projection of Latent Structures Discriminant Analysis*

PR : Pouvoir Résolutif

QC : Contrôle qualité, *Quality Control*

QSRR : relation quantitative entre la structure chimique et le temps de rétention, *Quantitative Structure Retention Relationship*

RIA : Abondance isotopique relative, *Relative Isotopic Abundance*

RMN : Résonance Magnétique Nucléaire

SFC : Chromatographie en phase supercritique, *Supercritical Fluid Chromatography*

SMILES : *Simplified Molecular Input Line Entry System*

SPE : Extraction sur phase solide, *Solid Phase Extraction*

SWATH : Acquisition séquentielle de toutes les fenêtres de masse, *Sequential Window acquisition of All Theoretical mass spectra*

TOF : Temps de vol, *Time Of Flight*

TUS : Somme des signaux d'intérêt, *Total Usefull Signal*

LISTE DES FIGURES

Figure 1 : La cascade des « -omiques ».....	26
Figure 2 : Nombre de publications concernant la métabolomique impliquant soit la spectroscopie par résonance magnétique nucléaire soit la spectrométrie de masse (Gowda and Raftery, 2015).....	32
Figure 3 : Schéma général d'une analyse métabolomique par LC-HRMS.....	37
Figure 4 : Déroulé d'un traitement de données par XCMS.....	58
Figure 5 : Arbre décisionnel dans le choix des tests univariés.....	63
Figure 6 : Extrait de la voie métabolique du tryptophane publiée dans KEGG (Ogata et al., 1999).....	65
Figure 7 : Exemple d'un spectre de fragmentation résonante (CID) de l'alanyltyrosine	67
Figure 8 : Exemple d'un spectre de fragmentation non résonante (HCD) de l'alanyltyrosine.....	68
Figure 9 : Les différents modes d'acquisition MS ²	69
Figure 10 : Arbre décisionnel du processus d'annotation	73
Figure 11 : Physiopathologie de l'encéphalopathie hépatique (Jalan and Lee, 2009).....	95
Figure 12 : Analyse multivariée de la seconde cohorte basée sur les 343 métabolites détectés lors de l'analyse métabolomique	113
Figure 13 : Distributions des certains métabolites entre les trois populations. Les patients de score West-Heaven 3 et 4 sont représentés en rouge.....	114
Figure 14 : Altération de la voie de biosynthèse des sphingolipides.	116
Figure 15 : Augmentation de la proportion de N-glycanes sous forme bissectine en lien avec une augmentation de la concentration en acetylhexosamine libre	117
Figure 16 : Structure de l'acide aspartique et de l'acide glutamique	133
Figure 17 : Chemin réactionnel des deux mécanismes de fragmentation de l'acide aspartique	147
Figure 18 : Formation des ions dans une source électrospray	233
Figure 19 : Diagramme résumant les gammes de masse et de polarité couvertes par les différents types de source (Forcisi et al., 2013).	234
Figure 20 : Principe de fonctionnement d'un analyseur quadripolaire	235
Figure 21 : Principe de fonctionnement d'un piège à ion linéaire segmenté à éjection radiale	236
Figure 22 : Principe de fonctionnement d'un analyseur à temps de vol avec réflectron	237
Figure 23 : Principe de fonctionnement d'un analyseur à résonance cyclonique d'ions ...	239
Figure 24 : Principe de fonctionnement d'un Orbitrap.....	240

LISTE DES TABLEAUX

Tableau 1 : Résumé des performances des différentes méthodes d'extraction des métabolites plasmatiques (Sitnikov et al., 2016).....	44
Tableau 2 : Résumé des performances de différents types d'analyseur	53
Tableau 3 : Nombre de formules brutes compatibles avec la masse de certains métabolites en fonction de la tolérance en masse acceptée.	76
Tableau 4 : Ratios classiques entre différents atomes à respecter pour l'annotation en métabolomique (Kind and Fiehn, 2007)	77
Tableau 5 : Abondance isotopique et différence de masse des formes isotopiques en métabolomique	78
Tableau 6 : Echelle de classification de West-Heaven (Blei and Córdoba, 2001)	94
Tableau 7 : Echelle de classification selon le Glasgow Coma Score (GCS) (Teasdale and Jennett, 1976)	96
Tableau 8 : Statistiques des différentes familles de lipides détectées en lipidomique.....	115

LISTE DES ANNEXES

Annexe 1 : Les différents types de sources en spectrométrie de masse	231
Annexe 2 : Les différents types d'analyseurs en spectrométrie de masse	235
Annexe 3 : Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy, Supplementary Information	242
Annexe 4 : Paramètres analytiques utilisés lors des analyses métabolomique, lipidomique et glycomique des échantillons de plasma de la seconde cohorte d'encéphalopathes.....	258
Annexe 5 : Evaluation of High-Filed Orbitrap Fusion for Compound Annotation in Metabolomics, Supplementary Information.....	262
Annexe 6 : Mechanistic study of competitive releases of H ₂ O, NH ₃ and CO ₂ from deprotonated aspartic and glutamic acids: Role of conformation, Supplementary Information.....	275
Annexe 7 : Evaluation of Data Dependent and Data Independent Acquisition workflows on an Orbitrap Fusion for metabolomics applications, Supplementary Information	280

Introduction

Les approches « -omiques »

1 Description générale

1.1 La génomique

La génomique est un domaine multidisciplinaire de la biologie moderne qui étudie l'ensemble des gènes d'un organisme, d'un organe ou d'un tissu. Le génome d'un organisme est constitué de l'ensemble de son ADN ainsi que des gènes qu'il renferme.

Les travaux de Watson et Crick sur la structure moléculaire de l'ADN en 1953 furent décisifs (Watson and Crick, 1953) et donnèrent naissance à la biologie moléculaire moderne puis à la génétique et enfin à la génomique. Leur découverte donnait à l'étude de caractères physiques héréditaires une dimension moléculaire, rendant tangibles les théories de Mandel, Morgan ou même Darwin (Potters, 2010).

Par opposition à la génétique, qui étudie des gènes spécifiques et leur hérédité, la génomique, elle, s'intéresse à la caractérisation et la quantification de l'ensemble des gènes qui sont responsables de la production des protéines d'un organisme.

La génomique se divise en deux branches : génomique structurale et génomique fonctionnelle. La génomique structurale, appelée également génétique moléculaire, s'intéresse à la cartographie physique des génomes et à la structure des gènes : localisation et annotation des gènes, nombre d'exons, d'introns, polymorphismes, etc... La mise au point du séquençage ADN à haut débit et l'essor de la bio-informatique ont permis une avancée considérable de la génomique structurale. La génomique s'intéresse également à la fonction de ces gènes, il s'agit de la génomique fonctionnelle. Un des moyens d'étudier ces fonctions est d'étudier les mécanismes de l'expression des gènes, en caractérisant le transcriptome (ensemble des ARN messagers) et le protéome (ensemble des protéines). Le développement des « puces à ADN » et l'automatisation des techniques de biologie moléculaire ont permis de révolutionner la génomique fonctionnelle.

Cette révolution de la génomique, amorcée par le séquençage du génome humain en 2003 (Chial, 2008), a permis de mieux comprendre la diversité naturelle des mécanismes biologiques à l'origine de la diversité des phénotypes. Ainsi la génomique nous éclaire sur le fonctionnement de certains organes et de certaines tumeurs et nous permet d'identifier et d'expliquer un certain nombre de prédispositions génétiques. Ces progrès s'annoncent très prometteurs pour le domaine

de la santé et devraient permettre d'améliorer le diagnostic, la prévention et le traitement d'un grand nombre de pathologies complexes (Pang, 2002).

1.2 L'étude du phénotype

Le génome regroupe l'ensemble de l'information génétique portée par un organisme. Cependant il ne s'agit que d'une information potentiellement exprimée. D'autres sciences se sont donc développées afin d'étudier non plus le génotype mais le phénotype des organismes. Ces nouvelles disciplines ont conduit à la mise en place de la cascade des « -omiques » (Figure 1).

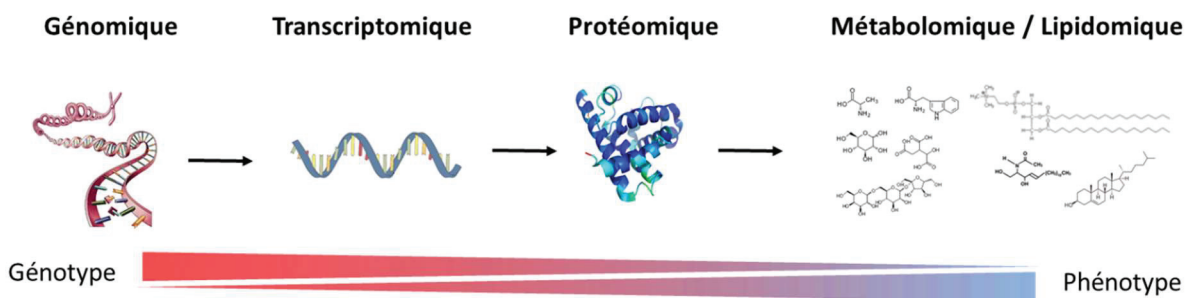


Figure 1 : La cascade des « -omiques »

1.2.1 Transcriptomique

La transcriptomique est l'étude de l'ensemble des éléments issus de la transcription de l'ADN en acide ribonucléique (ARN), appelé transcriptome. Le transcriptome est défini comme l'ensemble des transcrits présents dans une cellule à un moment donné et dans des conditions données. Ce terme fut proposé pour la première fois par Velculescu en 1997 (Velculescu et al., 1997). Il regroupe à la fois les ARN messagers, codant pour des protéines, et les ARN non-codants tels que les ARN de transfert, les ARN ribosomiaux, les micro-ARN et les longs ARN non-codants, intervenant entre autres dans la régulation de l'expression des gènes. Contrairement au génome qui est essentiellement statique, le transcriptome est dynamique et peut être modulé par des facteurs internes ou externes. Il s'agit donc de la première étape qui permet de faire le lien entre le génome d'un organisme et son phénotype exprimé à un instant donné.

Deux méthodes sont principalement utilisées pour étudier le transcriptome. Il s'agit tout d'abord des puces à ADN (Brown and Botstein, 1999). Cette méthode repose sur la propriété que possède l'ADN dénaturé de reformer spontanément sa double hélice lorsqu'il est placé face à un brin complémentaire (réaction d'hybridation). L'ensemble des transcrits issus d'un échantillon est transformé en ADN complémentaire par la technique de rétrotranscription, puis marqué. Chaque séquence d'ADN recherchée est fixée sur un support, appelé puce, et les ADN complémentaires

correspondants à chacun des gènes viennent s'hybrider, émettant alors un signal proportionnel au niveau d'expression. Chaque puce peut comporter plusieurs milliers de séquences, ce qui en fait une technique d'analyse à haut débit. Elle a ainsi permis d'établir les premiers catalogues des gènes exprimés pour plusieurs types de cellules et dans différentes conditions. Cependant, cette technique nécessite de connaître la séquence des gènes étudiés afin de pouvoir les synthétiser et les fixer sur la puce. De ce fait, elle permet de rechercher et de quantifier uniquement les transcrits dont la séquence est connue.

La seconde méthode détermine directement la séquence des ADN complémentaires synthétisés à partir du transcriptome. Elle est basée sur les techniques de séquençage haut-débit et est appelée RNA-seq. Cette technologie permet donc l'étude du transcriptome indépendamment de la connaissance du génome. De plus, sa gamme dynamique de niveau d'expression des transcrits est beaucoup plus importante que celles des techniques par hybridation (Wang et al., 2009).

L'étude du transcriptome permet de mieux comprendre la régulation de l'expression de l'ADN en révélant les séquences transcrites en ARN. Cependant d'autres niveaux de régulation existent, notamment lors de la traduction et de la maturation des protéines.

1.2.2 Protéomique

La protéomique est l'étude de l'ensemble des éléments issus de la traduction des ARN messagers, appelé protéome. Ce terme, introduit en 1995 par Wasinger (Wasinger et al., 1995), regroupe l'ensemble des protéines constituant un organisme vivant. A l'image du transcriptome, le protéome est un élément dynamique qui est modulé dans le temps. L'analyse protéomique s'est développée peu de temps après (Wilkins et al., 1996; James, 1997). Elle a pour but de décrire et de quantifier l'expression des protéines ainsi que ses changements sous l'influence de perturbations biologiques telles que la maladie ou la prise d'un traitement médicamenteux (Anderson and Anderson, 1998). Il s'agit donc de déterminer l'identité, la quantité et la fonctionnalité de ces protéines (Patterson and Aebersold, 2003). Elle permet d'étudier le produit de la traduction des ARN messagers mais également les modifications post-traductionnelles que peuvent subir les protéines comme des phosphorylations, des oxydations ou des glycosylations. Ces modifications peuvent faire varier l'action catalytique des protéines ou leur rôle dans la signalisation. De ce fait l'étude du protéome permet de comprendre un peu mieux les mécanismes en jeu au sein d'un organisme aussi bien à l'état physiologique que pathologique.

A l'heure actuelle, les analyses de protéomique sont principalement basées sur des technologies de spectrométrie de masse (Sokolowska et al., 2013). Face à la complexité du protéome, notamment en termes d'hétérogénéité chimique, différentes techniques séparatives sont utilisées en amont de la spectrométrie de masse. Il peut s'agir soit de techniques préparatoires,

comme la réalisation de gel d'électrophorèse bidimensionnelle (2D), soit de techniques séparatives couplées à la spectrométrie de masse comme la chromatographie liquide ou l'électrophorèse capillaire. La caractérisation des protéines peut se faire soit par une approche dite « top-down » où une protéine intacte est analysée, soit par une technique dite « bottom-up » où un mélange de protéines est digéré en peptides avant d'être analysé. La première approche représente un plus grand challenge analytique mais permet une meilleure couverture de la séquence d'acides aminés et une meilleure détection des modifications post-traductionnelles (Catherman et al., 2014).

L'analyse du protéome reflète au mieux l'expression du génome d'un organisme dans une condition donnée. Néanmoins, l'action des protéines peut être modulée en aval par l'intermédiaire de petites molécules, dites effectrices. L'analyse directe des molécules, qu'elles soient endogènes ou exogènes, permet donc de cerner plus précisément ce qu'il se passe au sein d'un organisme.

1.2.3 Métabolomique

Le métabolomique s'est développée à la suite des autres sciences « -omiques » à partir de la fin des années 1990 (Oliver et al., 1998; Nicholson et al., 1999). Son but est d'étudier le métabolome d'un organisme dans des conditions biologiques données, de l'identifier et le quantifier. Ce métabolome regroupe l'ensemble des composés de faible masse moléculaire (< 1000 Da ou < 1500 Da selon les définitions) qui peuvent être retrouvés au sein d'un échantillon biologique. Il s'agit des composés impliqués dans le métabolisme en tant que substrats, produits ou effecteurs. Ce métabolome est donc dynamique et inclut toutes les substances organiques naturellement produites par le métabolisme du système étudié. Il inclut également les xénométabolites. Il s'agit de métabolites provenant d'autres systèmes biologiques que ceux de l'hôte (Holmes et al., 2008). C'est notamment le cas des métabolites produits par la flore bactérienne constituant le microbiote intestinal. Une partie de ces métabolites bactériens peut passer la barrière intestinale et se retrouver au sein du métabolome de l'hôte. Les xénobiotiques (médicaments, pesticides, chimiques environnementaux, ...), ainsi que leurs produits de dégradation, font également partie du métabolome. Sont par contre exclues les structures polymérisées telles que les protéines ou les acides nucléiques. Un métabolite peut donc être considéré comme tout composé organique non directement issu de l'expression des gènes (Junot et al., 2014). On y retrouve des acides aminés, des sucres, des acides organiques, des acides gras, des nucléotides, des métabolites conjugués, des vitamines, des stéroïdes, etc... (Wishart et al., 2007). Cela inclut entre autres les lipides et les glycanes. Du fait de leurs propriétés physico-chimiques, ces deux catégories de métabolites font l'objet de techniques analytiques un peu particulières.

Quel que soit le type de métabolite étudié, deux approches peuvent être utilisées. La première est une approche ciblée dont le but est de détecter et surtout de quantifier de façon absolue un nombre limité de métabolites prédéterminés. Cette approche implique le plus souvent

une préparation d'échantillon et des conditions d'analyse spécifiques aux espèces étudiées. Elle est souvent plus sensible mais nécessite l'utilisation de standards de référence. La seconde approche est dite non ciblée et vise à obtenir un profil métabolique le plus exhaustif possible. On parle aussi d'approche globale. Cette approche se base sur une préparation d'échantillon la moins sélective possible et implique souvent l'utilisation de différents systèmes d'analyse afin de détecter le plus grand nombre de métabolites. Dans ce type d'approche, la quantification réalisée est en générale dite « relative », c'est-à-dire qu'elle permet une comparaison des échantillons entre eux sans qu'il existe de référence pour déterminer une concentration « absolue ». Ces méthodes d'analyse non ciblées en métabolomique seront détaillées dans la suite de ce manuscrit.

Le métabolome dans son ensemble est la dernière étape de la cascade des sciences « -omiques ». Il représente l'ultime réponse phénotypique d'un organisme à un facteur perturbant son fonctionnement (modification génétique, maladie, exposition à un toxique...). Il est donc, à ce jour, avec le protéome, le reflet le plus précis de la réponse de l'organisme à cette perturbation.

2 Les approches combinées

Au cours de ces trente dernières années, la biologie moléculaire et ses objectifs ont évolué. La volonté de comprendre le fonctionnement de chaque gène individuellement a glissé vers la volonté de comprendre le fonctionnement de l'ensemble des gènes et l'ensemble des produits qui résultent de leur expression (ARN, protéines, métabolites...). Cette approche scientifique repose sur une nouvelle façon d'envisager la biologie dans un contexte plus global, en prenant d'avantage de recul et en intégrant l'ensemble des éléments de la biologie ainsi que leurs interactions. Plutôt que de concentrer leurs efforts sur un gène et sur ses effets directs, les biologistes et biochimistes ont développé un arsenal de technologies et de méthodologies qui s'intègrent dans ce qu'on appelle désormais la « biologie des systèmes ». Ce concept moderne de biologie des systèmes émerge entre autres de l'idée que des « systèmes multi-dimensionnels, complexes, dynamiques et formés par des macromolécules, métabolites, cellules, organes et organismes en interaction les uns avec les autres, sont à l'origine de la plupart des processus biologiques fondamentaux » (Vidal, 2009). En effet, aucun élément n'agit seul en biologie et aucun mécanisme ne se produit sans interaction avec le reste de son environnement, tout est en lien, en opposition ou en synergie avec les autres éléments du système. Les gènes se régulent entre eux, les produits de l'expression des gènes se stimulent mutuellement ou s'inhibent les uns les autres, les cellules communiquent, les métabolites circulent d'un tissu vers un autre et tout est en perpétuelle interconnexion (Potters, 2010). La biologie des systèmes est donc l'étude du comportement d'une structure biologique complexe dans son ensemble. Elle repose sur la biologie moléculaire (avec un accent mis sur les transferts d'information), sur la physiologie et les capacités d'adaptation d'une cellule et d'un organisme, sur la biologie développementale pour définir une succession d'états physiologiques au cours de cette adaptation,

et enfin sur la biologie et l'écologie évolutives avec l'idée que chaque aspect d'un organisme est le résultat d'une sélection qu'il est rarement possible de comprendre à l'échelle moléculaire (Kirschner, 2005).

L'étude de la biologie des systèmes se concentre sur 4 axes : l'acquisition de mesures quantitatives, la mise au point de modèles mathématiques basés sur ces mesures, la reconstruction des réactions d'une cellule ou d'un organisme soumis à certaines conditions et l'élaboration de théories qui permettent d'expliquer la variabilité des réponses observées et leurs mécanismes. Cette approche a donné lieu à de nombreuses innovations dans plusieurs domaines scientifiques. Premièrement, le développement des techniques analytiques haut-débit, permettant l'acquisition d'un large éventail de données (Potters, 2010). Ensuite, la bio-informatique s'est développée, avec des outils permettant la gestion et l'intégration du nombre considérable de données générées (Rodrigo et al., 2008). En parallèle de cela, un certain nombre d'applications et d'outils ont dû être créés pour permettre une analyse minutieuse et précises de ces données et la création de modèles les plus précis possibles.

Les différentes sciences « -omiques » découlent donc directement de cette biologie des systèmes et de l'idée d'une approche intégrée de la physiologie. Chacun des domaines de recherche représente un niveau d'étude du système, un point de vue particulier, une pièce du puzzle de la biologie. Une fois combinés et intégrés au sein d'un modèle dynamique, ils offrent une vision plus complète, plus globale et plus proche du fonctionnement réel du vivant. Cette approche combinée des « -omiques » ouvre donc les portes de la biologie dite intégrative et prédictive. Il s'agit d'analyser l'activité des systèmes biologiques et leurs interactions dans un environnement donné et de pouvoir appréhender, dans leur ensemble, les paramètres influant sur un phénomène biologique. Un des objectifs majeurs ici est d'élucider les mécanismes pathologiques pour proposer de nouvelles cibles thérapeutiques potentielles et identifier de nouveaux traitements et de nouvelles méthodes de diagnostic (Cavazzana-Calvo and Debiais, 2011).

Parmi ces différentes sciences « -omiques », les travaux menés dans le cadre de cette thèse se focaliseront sur l'approche métabolomique, notamment du fait de sa position finale dans la cascade et des avantages que cela représente.

La métabolomique non ciblée

1 Développement de la métabolomique

Les premières approches qui conduisirent au développement de la métabolomique datent du début des années 1970 (Sumner et al., 2003). On parlait alors de « profilage métabolique » (Devaux et al., 1971; Horning and Horning, 1971), ce terme faisant référence à l'analyse qualitative et quantitative de matrices biologiques complexes. Ces premiers travaux portaient sur l'analyse de stéroïdes, d'acides organiques et de métabolites de médicaments dans les urines par chromatographie gazeuse couplée à la spectrométrie de masse. Peu après, l'idée d'utiliser le profilage métabolique dans le domaine médical commença à s'étendre (Laragh, 1971). A partir des années 1980, l'utilisation de la résonance magnétique nucléaire (RMN) prit le pas sur la spectrométrie de masse pour le profilage métabolique (Bales et al., 1984; Nicholson et al., 1984). Durant la fin des années 1980 et le début des années 1990, la recherche par profilage métabolique atteint un plateau, avec un nombre annuel de publications compris entre 10 et 15 (Sumner et al., 2003). Ce n'est qu'à la fin des années 1990 avec l'avènement du séquençage et de la génomique fonctionnelle que le concept de « métabolomique » prit son essor (Nicholson et al., 1999).

Au vu de la grande diversité chimique des métabolites, de nombreuses technologies analytiques ont été utilisées en métabolomique dans le but de détecter l'intégralité des métabolites. La spectroscopie infrarouge (Oliver et al., 1998) ou la chromatographie 2D sur couche mince (Tweeddale et al., 1998) font partie des premières méthodes utilisées, avec la RMN. Face à la complexité des échantillons étudiés, les techniques de détection des métabolites ont souvent été couplées à une technique séparative. Par exemple, des études menées par électrophorèse capillaire ont été couplées à une détection UV (Baggett et al., 2002), à une détection par fluorescence induite par laser (Arlt et al., 2001) ou plus couramment à une détection par spectrométrie de masse (Soga et al., 2002). D'autres techniques séparatives comme la chromatographie gazeuse ou liquide peuvent être couplées à des détections par spectrométrie de masse (Fiehn et al., 2000; Plumb et al., 2002).

A l'heure actuelle, les études scientifiques publiées en métabolomique impliquent principalement, deux techniques de détection des métabolites : la résonance magnétique nucléaire et la spectrométrie de masse (Figure 2). Cette dernière, souvent couplée à une technique chromatographique, est devenue majoritaire depuis quelques années (Gowda and Raftery, 2015).

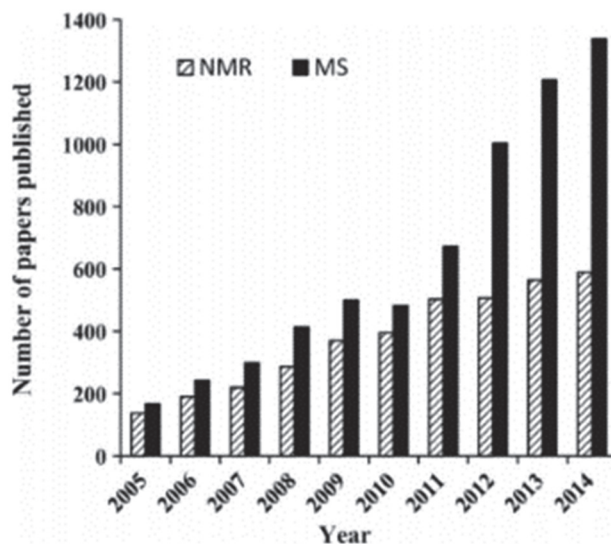


Figure 2 : Nombre de publications concernant la métabolomique impliquant soit la spectroscopie par résonance magnétique nucléaire soit la spectrométrie de masse (Gowda and Raftery, 2015).

2 Les méthodes d'analyse physico-chimiques utilisées en métabolomique

2.1 La résonance magnétique nucléaire (RMN)

2.1.1 Principe général

La RMN fut l'une des premières technologies utilisée pour l'analyse du métabolome (Nicholson and Wilson, 1989; Gartland et al., 1989). Cette technologie est basée sur les propriétés physiques de certains atomes qui possèdent un moment magnétique non nul. Pour des applications en métabolomique, il s'agit essentiellement du proton ^1H , du ^{13}C , du ^{15}N et du ^{31}P . Ces éléments possèdent un nombre de spin de $\frac{1}{2}$. Lorsqu'ils sont placés dans un champ magnétique intense, ces spins vont alors s'aligner parallèlement (ou antiparallèlement) au champ magnétique. Si on leur applique une impulsion de radiofréquence, les spins vont basculer perpendiculairement au champ magnétique. Après l'excitation, le système revient à l'équilibre en induisant un champ de radiofréquence qui est détecté et enregistré. L'application d'une transformée de Fourier permettra de convertir la fréquence enregistrée en spectre qui pourra alors être interprété.

Pour les applications en métabolomique, la RMN du ^1H est la plus utilisée. En effet, il s'agit de l'application de la RMN la plus sensible. Cela est dû à la forte présence d'atomes d'hydrogène dans les biomolécules et au fait que la forme ^1H représente plus de 99 % des atomes d'hydrogène. La RMN du ^{31}P est également fréquemment utilisée dans les études s'intéressant à l'état énergétique des cellules (Liu et al., 2011) car il s'agit de la seule forme de phosphore présente à l'état naturel. Les RMN ^{13}C et ^{15}N sont très peu sensibles du fait de la faible abondance de ces isotopes stables

(respectivement 1 % et 0,37 %). Cependant ces isotopes peuvent être utilisés dans des études d'analyse des flux de métabolites au cours desquels des métabolites marqués contenant ces formes isotopiques sont suivis et quantifiés (Lane and Fan, 2017).

2.1.2 Intérêts et limites en métabolomique

La RMN présente plusieurs intérêts pour les analyses de métabolomique. En effet, il s'agit d'une technique d'analyse non destructive, quantitative, robuste, nécessitant peu de préparation d'échantillon, comparable entre différentes plateformes et qui apporte de nombreuses informations structurales sur les métabolites étudiés.

Tout d'abord, la RMN est très informative et permet l'identification de nombreux métabolites. Cependant la grande complexité et l'abondance des métabolites présents dans un échantillon biologique peuvent poser problème. En effet cela est responsable de nombreux chevauchements entre les signaux sur le spectre obtenu. Pour pallier ce problème, des approches de RMN à deux dimensions ont été développées et sont utilisées en métabolomique. L'ajout d'une seconde dimension permet une meilleure résolution des signaux des différents métabolites et facilite leur identification en mélange biologique. Il peut s'agir de méthodes homo-nucléaires telles que la ^1H - ^1H COSY (Xi et al., 2006) ou la ^1H - ^1H TOCSY (Bingol et al., 2014) ou de méthodes hétéro-nucléaires telles que la ^1H - ^{13}C HSQC (Bingol et al., 2015b). Grâce à ces développements de méthodes, la RMN 2D est devenue la technique de choix pour l'élucidation structurale des métabolites.

Le second intérêt majeur de la RMN réside dans le fait qu'elle soit quantitative. Cela est particulièrement vrai pour la RMN unidimensionnelle où l'intensité des pics est corrélée linéairement à la concentration relative du métabolite. Pour obtenir une concentration absolue, il suffit d'ajouter un standard de concentration connue (Markley et al., 2017). Il n'en est pas de même pour la RMN 2D. Par exemple, pour la RMN ^1H - ^{13}C HSQC, l'intensité des pics n'est pas directement proportionnelle à la concentration. Différents outils ont été développés pour déterminer une quantification absolue directe à partir des spectres 2D. On peut notamment citer l'approche HSQC₀ (Hu et al., 2011).

Enfin, un autre avantage de la RMN réside dans sa très grande reproductibilité, notamment entre différents utilisateurs et même entre différents instruments (Bingol and Brüschweiler, 2015a). Cela facilite grandement la réalisation d'études inter-plateformes qui peuvent s'avérer intéressantes notamment lorsque le nombre d'échantillons à analyser devient très important.

En contrepartie, la principale limite de la RMN est son manque de sensibilité par rapport à la spectrométrie de masse (Dunn et al., 2013). De ce fait, la plupart des études en RMN sont focalisées sur la découverte de biomarqueurs plutôt que sur l'identification complète du métabolome. Les améliorations technologiques en cours devraient permettre d'améliorer la sensibilité des

instruments. Cela peut notamment passer par l'utilisation d'aimants à très hauts champs permettant d'atteindre des fréquences de résonance supérieures à 1 GHz, par l'utilisation de cryo-sondes et de bobines super conductrices à haute température (Ramaswamy et al., 2013) ou par le développement des techniques d'hyperpolarisation (Ardenkjær-Larsen et al., 2003).

2.2 La spectrométrie de masse (MS)

2.2.1 Principe général

La spectrométrie de masse est aujourd'hui la technique d'analyse physique la plus utilisée dans les études de métabolomique. Elle permet de mesurer le rapport masse/charge (m/z) de différentes espèces chimiques. Pour pouvoir être détectées, les composés présents dans les échantillons biologiques (principalement sous forme moléculaire) ont donc besoin d'être ionisés. Quel que soit son type, un spectromètre de masse sera toujours composé de trois éléments principaux : une source d'ionisation, un analyseur avec un système de détection des signaux et un système de traitement du signal qui permet de convertir les signaux enregistrés en rapport m/z . Le rôle de la source d'ionisation est d'ioniser et vaporiser l'échantillon qui arrive dans le spectromètre. Différents types de sources sont utilisés en fonction de la nature de l'échantillon et du type d'analyse à effectuer. L'analyseur sépare les différents ions formés au niveau de la source. Là encore, différents types d'analyseurs existent avec des caractéristiques qui leur sont propres. Le détecteur dépend du type d'analyseur utilisé et sert à enregistrer le signal des ions. Le traitement du signal se fait grâce à un logiciel d'acquisition généralement fourni par le constructeur de l'instrument. Les différents types d'instruments et leur fonctionnement seront détaillés dans la suite de ce manuscrit.

2.2.2 Avantages et inconvénients par rapport à la RMN

La spectrométrie de masse possède deux avantages principaux par rapport à la RMN. Le premier est sa plus grande sensibilité. En effet, là où la RMN est limitée à la détection en matrice des métabolites ayant une concentration autour du micromolaire (Simmler et al., 2014), la spectrométrie de masse permet de détecter des composés à des concentrations inférieures au nanomolaire (Scalbert et al., 2014). Cette sensibilité supérieure est un atout important car elle permet de détecter un nombre sensiblement plus important de composés (Simón-Manso et al., 2013). Cela en fait donc la technique de choix pour les analyses de métabolomique non ciblées. De plus, cela permet de travailler à partir d'une quantité réduite d'échantillon, de l'ordre de la dizaine de microlitres contre quelques millilitres en RMN.

Le second avantage est la vitesse d'acquisition d'un spectre de masse qui varie entre la dizaine de millisecondes et quelques secondes. Cela est à comparer avec la durée d'acquisition des spectres de RMN qui est de l'ordre de la minute en RMN 1D et de l'ordre du quart d'heure en RMN 2D

(Guenec et al., 2014). Cette vitesse d'acquisition permet de coupler la spectrométrie de masse à une technique séparative en ligne. Ce couplage rallonge considérablement le temps d'analyse par échantillon (10 à 30 minutes), mais il permet de séparer les différents métabolites de l'échantillon avant leur introduction dans le spectromètre de masse. Cela aboutit à l'acquisition de spectres plus propres et à la résolution temporelle de composés qui n'auraient pas pu être distingués lors d'une analyse directe. Les différentes techniques séparatives seront détaillées ci-après.

La principale limite que peut présenter la spectrométrie de masse est la nécessité d'ioniser les composés afin de pouvoir les détecter. Si différents types de sources d'ionisation existent, certains composés non ionisables ne pourront pas être détectés par cette technique. De plus, l'étape d'ionisation peut être difficile à contrôler ce qui rend la spectrométrie de masse moins reproductible que peut l'être la RMN et limite notamment les comparaisons de résultats entre différents laboratoires.

Du fait de sa meilleure couverture métabolique, la spectrométrie de masse sera utilisée pour l'ensemble des travaux réalisés dans le cadre de cette thèse. Les analyses métabolomiques utilisant cette technique seront présentées plus en détail ci-après.

2.3 Développement d'analyses combinant plusieurs techniques

D'autres approches visant à combiner les résultats obtenus par spectrométrie de masse et par RMN à partir des mêmes échantillons se sont développées (Crockford et al., 2006). Cette combinaison des deux techniques peut apporter une réelle plus-value sur deux aspects principaux : l'identification des composés et la détection de biomarqueurs via l'utilisation de modèles statistiques multivariés (Marshall and Powers, 2017). Pour l'identification rapide des métabolites, plusieurs approches ont récemment été développées. « NMR/MS translator » part des spectres RMN 1D ou 2D obtenus pour interroger les bases de données spectrales. La liste des candidats obtenue est utilisée pour simuler les spectres de masse de chaque candidat, en incluant les possibles adduits, fragments et isotopes. Les spectres MS générés *in silico* sont alors comparés aux spectres enregistrés. Les métabolites annotés en RMN sont donc confirmés par spectrométrie de masse. Cette approche a été validée sur des standards et a démontré son efficacité face à l'utilisation de la MS ou de la RMN seule (Bingol and Brüsweiler, 2015b). Une autre approche appelée « SUMMIT MS/NMR » part de l'acquisition MS à haute résolution. Pour chaque pic, les formules brutes pouvant correspondre à la masse sont calculées. Ces formules brutes sont ensuite utilisées pour interroger les bases de données chimiques. Les correspondances trouvées sont alors utilisées pour prédire les spectres RMN qui sont comparés aux spectres acquis (Bingol et al., 2015a). Cette approche a pu être validée sur des extraits bactériens (C. Wang et al., 2017). Dernièrement, une dernière approche appelée « ISEL NMR/MS² » reprend le principe de l'approche SUMMIT mais y ajoute une dimension supplémentaire en validant les identifications grâce à la RMN mais aussi à la MS² (Boiteau et al., 2018).

Le second aspect de la combinaison des techniques de MS et de RMN vise à améliorer l'interprétation des résultats, notamment au travers des analyses multivariées. La difficulté vient du fait que la plupart des méthodes sont uniquement adaptées à des données provenant d'une seule source. Il a toutefois pu être montré, à partir de quelques exemples, que les modèles intégrant MS et RMN permettent une meilleure séparation entre groupes d'échantillons et un plus haut niveau d'interprétabilité que l'utilisation isolée de la MS ou de la RMN (Chen et al., 2006; Gu et al., 2011). L'utilisation plus récente des méthodes d'analyses multi-blocs (Smilde et al., 2003; Westerhuis et al., 1998) pour des applications en métabolomique apparaît comme une solution idéale pour combiner les deux techniques analytiques. Les résultats obtenus par chacune des méthodes sont placés dans deux blocs séparés ce qui rend possible des corrélations à la fois intra et inter bloc. Il a été montré que l'utilisation de ces analyses multi-blocs permet de corrélérer les séparations observées dans chacune des analyses et aboutit à une meilleur séparation globale entre les groupes (Marshall et al., 2015).

L'utilisation des deux techniques complémentaires semble donc pouvoir apporter une réelle plus-value dans les analyses métabolomiques. Néanmoins cela nécessite l'accès à différents instruments qui ne sont pas forcément présents au sein d'un laboratoire et la durée d'acquisition des données s'en trouve fortement augmentée. De ce fait, même si elle se développe, cette approche reste aujourd'hui minoritaire et ne sera pas utilisée dans le cadre de ces travaux de thèse.

3 Les différentes étapes d'une analyse métabolomique par spectrométrie de masse

Comme expliqué ci-dessus, l'analyse par spectrométrie de masse présente de nombreux avantages qui en font la technique d'analyse la plus utilisée en métabolomique. C'est notamment le cas pour les analyses non ciblées où la spectrométrie de masse à haute résolution (HRMS, *High Resolution Mass Spectrometry*) est souvent couplée à une technique séparative. Les différentes étapes d'une analyse métabolomique non ciblée par spectrométrie de masse sont présentées et détaillées ci-après.

3.1 Schéma général d'une analyse métabolomique par HRMS et points critiques

Une analyse métabolomique se décompose en plusieurs étapes qui font appel à des disciplines diverses. Parmi ces étapes, certaines ont atteint une maturité de développement importante là où d'autres nécessitent encore la mise au point d'outils pour les améliorer. La Figure 3

présente le schéma général d'une analyse métabolomique par HRMS couplée à de la chromatographie liquide (LC-HRMS) telle qu'elle peut être réalisée au sein de notre laboratoire.

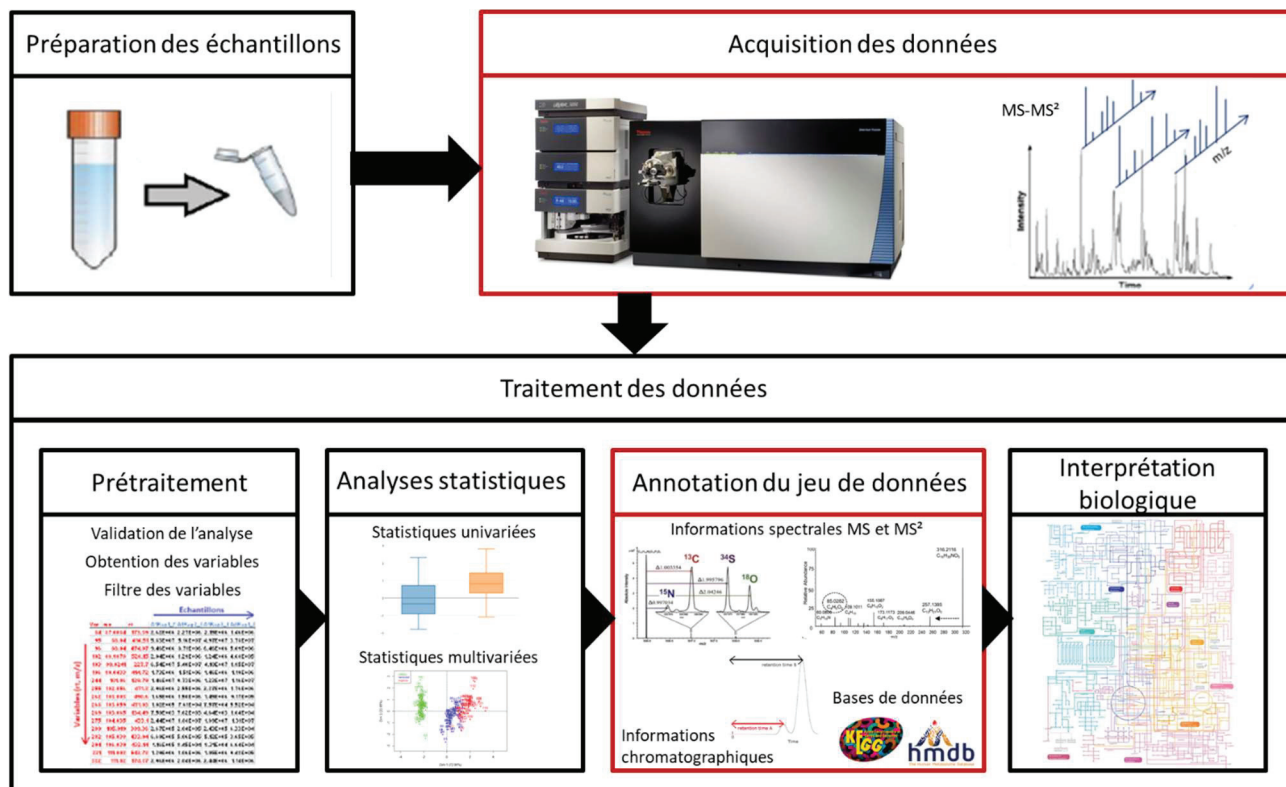


Figure 3 : Schéma général d'une analyse métabolomique par LC-HRMS

La première étape est la préparation des échantillons. Son objectif est d'extraire de façon la plus complète possible les métabolites présents dans un échantillon. Cette préparation est dépendante du type de matrice étudié. La seconde étape est l'acquisition des données. Elles sont obtenues par la séparation temporelle des analytes au niveau du système chromatographique, par leur introduction dans le spectromètre de masse grâce à leur ionisation au niveau d'une source et par l'acquisition des spectres de masse par le spectromètre. A ce niveau, différentes approches sont étudiées pour améliorer la qualité et la quantité d'information acquise, notamment grâce à l'utilisation de spectromètre de masse à haute résolution de dernière génération. Ces nouveaux spectromètres permettent d'envisager des modes d'acquisition plus complexes combinant analyses MS et MS². Après leur acquisition, les données sont traitées. Cette étape est cruciale pour extraire les informations pertinentes et utilisables à partir des données acquises. Elle peut être décomposée en différentes sous-étapes. Un des points les plus limitant actuellement en métabolomique est l'étape d'annotation du jeu de données et une partie des travaux de ce manuscrit y est consacré. L'étape d'analyse statistique permet d'identifier des variables discriminantes entre les groupes d'individus étudiés. Cette étape peut mettre en évidence de potentiels biomarqueurs et l'ensemble

de ces biomarqueurs constitue la signature métabolique liée à la perturbation étudiée. L'étape finale consiste en l'interprétation biologique de la signature métabolique obtenue. Cette étape est à l'heure actuelle la plus complexe.

3.2 Préparation d'échantillons

La préparation des échantillons comporte quatre étapes essentielles : la collecte d'échantillons, l'arrêt du métabolisme (afin d'éviter toute modification du métabolome après la collecte), la conservation de l'échantillon avant analyse et l'extraction des métabolites. Toutes ces étapes peuvent être source de biais et sont donc déterminantes dans la qualité globale d'une analyse métabolomique.

Une préparation d'échantillons pour une analyse de métabolomique doit répondre à différents critères : elle doit être non sélective, simple, rapide, reproductible, comprendre un nombre limité d'étapes et incorporer une étape d'arrêt du métabolisme (Vuckovic, 2012). La non-sélectivité de la méthode peut être caractérisée par sa capacité à préserver un nombre maximal de métabolites tout en éliminant les interférences telles que les protéines et les lipoprotéines. Sa capacité à solubiliser une large gamme de métabolites avec des taux élevés de recouvrement pour les différentes classes est aussi importante afin d'obtenir une sensibilité maximale. La simplicité et la rapidité de la méthode permet de minimiser la dégradation des métabolites ainsi que leur adsorption sur différentes matrices tout en évitant l'introduction de composés exogènes. Elle permet aussi une automatisation des méthodes et une préparation d'échantillons à haut débit. La reproductibilité est également un facteur essentiel pour permettre une comparaison des résultats d'analyse entre les différents échantillons. Enfin l'arrêt du métabolisme est important pour éviter des dégradations ou des surexpressions chimiques (oxydations) ou biologiques (action enzymatique) des métabolites après l'instant du prélèvement.

Tous ces points sont essentiels et la mise en œuvre de ces différentes étapes dépend du type d'échantillon étudié.

3.2.1 Prélèvement et stockage des échantillons

3.2.1.1.1 Echantillons sanguins

Le sang est la matrice la plus étudiée lors des analyses de métabolomique du fait de son rôle central dans l'organisme et de sa facilité de prélèvement. Sa composition métabolique est maintenue par homéostasie, cependant de nombreux facteurs peuvent entraîner des changements dynamiques, tels que par exemple le rythme circadien (Eckel-Mahan et al., 2012), le régime alimentaire ou la prise de médicaments. Pour limiter les effets de ces facteurs potentiellement confondants, il est

recommandé de prélever tous les échantillons au même moment de la journée et à distance équivalente des repas. Sur la base de ce qui est fait usuellement lors d'examens biochimiques, un prélèvement le matin après une période de jeûne de 12 h est recommandé pour les analyses de métabolomique (Yin et al., 2015).

A partir du sang, il est possible d'obtenir du plasma ou du sérum. Le plasma est obtenu par prélèvement sur un tube contenant un anticoagulant qui désactive les facteurs de coagulation ou limite leur action par chélation des ions Ca^{2+} . Une fois prélevé, le plasma est centrifugé afin d'éliminer les cellules sanguines. Le sérum est lui obtenu par prélèvement sur tube sec. La coagulation se fait alors naturellement pendant 30 à 60 minutes à température ambiante avant une étape de centrifugation.

Sérum et plasma ne présentent pas exactement la même composition. Cela est dû notamment à l'activation des plaquettes lors de la coagulation et à la nécessité de laisser les tubes à température ambiante pour que le processus de coagulation ait lieu. Le plasma a donc l'avantage d'être plus rapide à obtenir et l'absence de temps de coagulation limite le risque de conversion enzymatique ou de dégradation des métabolites les plus labiles (Denery et al., 2011). Cela confère une moindre variabilité aux échantillons de plasma (Teahan et al., 2006). En cas d'utilisation de sérum, il est essentiel de s'assurer que tous les échantillons soient laissés à coaguler pendant la même durée. En contrepartie, le sérum présente une plus faible teneur en protéines et conduit donc à moins de risques de suppression d'ionisation. De plus il a été montré que la concentration des métabolites était en générale plus importante dans le sérum (Yu et al., 2011). Malgré ces petites différences, il semble que les résultats obtenus en métabolomique non ciblée à partir de ces deux matrices soient très similaires en terme de couverture métabolique et de capacité discriminatoire (Wedge et al., 2011; Yu et al., 2011). Dans le cas d'analyse de plasma, la question du choix de l'anticoagulant doit se poser. Il peut s'agir d'héparine, d'EDTA ou de citrate et là encore, il n'existe pas de consensus. Certaines études ont montré qu'il n'y a pas de différences majeures entre les trois anticoagulants (Denery et al., 2011). On peut juste noter que le citrate est un des métabolites impliqués dans le cycle de Krebs et que pour l'étude de ce métabolite central, l'utilisation de tube citratés est à exclure. De manière globale, l'utilisation d'anticoagulant risque de causer une importante suppression d'ionisation des métabolites éluant au même temps de rétention (Barri and Dragsted, 2013; Yin et al., 2015).

Tous ces résultats suggèrent que les différents types de prélèvement peuvent être utilisés mais qu'il est essentiel de conserver un même et unique type d'échantillons et un même protocole de prélèvement pour tous les échantillons d'une étude (Hernandes et al., 2017).

Une fois le sang prélevé, une des étapes clés est la centrifugation de l'échantillon qui permet d'éliminer les cellules sanguines actives. Dans le cas du sérum, cette centrifugation ne peut intervenir

avant la fin de la coagulation. Ce type d'échantillon reflète donc notamment le métabolisme lié à l'activation plaquettaire et à l'activité des érythrocytes, des leucocytes et des enzymes circulantes durant cette étape de coagulation. D'où la nécessité de ne comparer que des échantillons de sérum ayant subi un temps de coagulation identique. Dans le cas du plasma, le temps et les conditions de stockage de l'échantillon entre son prélèvement et l'étape de centrifugation apparaît comme étant le premier facteur de variabilité dans la préparation des échantillons en métabolomique (Jobard et al., 2016). Cette étude a montré que si l'échantillon est laissé à température ambiante pendant plus d'une heure, le métabolome est altéré du fait de l'activité des érythrocytes. Plusieurs études ont montré que si l'échantillon est conservé sur glace ou à + 4°C, son métabolome peut être considéré comme stable pendant 6 h (Kamlage et al., 2014; Yin et al., 2013). Cette conservation à basse température des échantillons avant centrifugation est donc un point essentiel pour les analyses de métabolomique impliquant le plasma (Jobard et al., 2016).

Après l'élimination des cellules sanguines, le métabolome peut encore être affecté par la présence d'enzymes mais ces variations sont de bien moindre importance que celles observées directement après le prélèvement (Yin et al., 2015). Une étude a notamment montré qu'un échantillon de plasma pouvait rester stable à + 4 °C pendant 16h après centrifugation (Kamlage et al., 2014). Il est néanmoins conseillé de congeler le plus rapidement possible l'échantillon de plasma ou de sérum obtenu. Un stockage à -80° C doit être préféré à un simple stockage à - 20 °C (Vaught, 2006). En effet, il a été montré que certains métabolites tels que le glucose ou la proline présentent des variations significatives de concentration lors d'un stockage à -20°C (Pinto et al., 2014).

3.2.1.1.2 Urine

L'urine est un autre biofluide très utilisé en métabolomique. Cela est lié au fait qu'elle peut être obtenue en grande quantité de manière très simple et non invasive, ce qui facilite l'obtention de l'accord du patient. De plus les concentrations en métabolites sont souvent plus importantes que dans le plasma ce qui permet une meilleure sensibilité (Saude et al., 2007). Enfin, l'urine est très pauvre en protéines du fait de la filtration glomérulaire, ce qui simplifie le traitement des échantillons. Le prélèvement s'effectue en général dans un contenant en polypropylène (Fernández-Peralbo and Luque de Castro, 2012). L'addition de surfactant peut être envisagée pour minimiser les interactions entre la surface du contenant et certains métabolites apolaires (Silvester and Zang, 2012). Le principal facteur de variation lors du prélèvement est l'heure de la journée à laquelle il est réalisé. Cette heure peut être aléatoire, fixe ou il peut s'agir d'un prélèvement de l'ensemble des urines sur une journée. Ce dernier type de prélèvement est plus contraignant mais permet de s'affranchir des variations circadiennes de la concentration en métabolites (Slupsky et al., 2007) et permet donc une quantification plus adaptée à l'analyse globale du métabolome. En revanche cette approche ne

permet pas d'évaluer d'éventuelles modifications liées à ce rythme circadien d'expression des métabolites.

L'une des spécificités de l'urine est sa potentielle contamination bactérienne, notamment en cas d'obtention par miction naturelle. Pour éviter d'observer un développement bactérien qui viendrait modifier le métabolome urinaire, deux approches sont possibles. La première est basée sur l'ajout d'agents bactériostatiques tels que l'acide borique ou de l'azoture de sodium. Cependant, ces ajouts peuvent modifier l'échantillon, avec par exemple la formation de complexes, et engendrer une source de variabilité supplémentaire. A l'heure actuelle, la technique retenue est une filtration à 0.20 µm afin de retirer physiquement les bactéries de l'échantillon (Saude and Sykes, 2007; Bernini et al., 2011).

La question du stockage de l'urine a fait l'objet de nombreuses études (Roux et al., 2015). Certaines montrent que le stockage de l'urine à + 4°C pendant plusieurs jours altère le métabolome (Lauridsen et al., 2007; Saude and Sykes, 2007; van de Merbel, 2008). Il semblerait cependant qu'un stockage de courte durée (24 ou 48h) à + 4°C avant congélation n'ait pas d'impact significatif (Dunn et al., 2008; Gika et al., 2008). Il est cependant recommandé de congeler les échantillons d'urine dès que possible en les ayant préalablement aliquotés afin de limiter le nombre de cycle congélation/décongélation. Ce stockage peut se faire indifféremment à -20°C ou à - 80°C (Fernández-Peralbo and Luque de Castro, 2012).

3.2.1.1.3 Liquide céphalorachidien

Le dernier biofluide régulièrement étudié en métabolomique est le liquide céphalorachidien (LCR) (Wishart et al., 2008). Il présente l'avantage de baigner le cerveau et est donc le meilleur représentant possible des dérèglements métaboliques de ce dernier. La principale difficulté avec ce biofluide est l'obtention d'échantillons. En effet, il est prélevé par ponction lombaire, un acte invasif qui ne peut être réalisé qu'en milieu hospitalier et sous certaines conditions. Les patients souffrant notamment de troubles de la coagulation ne peuvent pas être inclus dans des études nécessitant des prélèvements de LCR à des fins de recherche (sauf si la ponction lombaire est prescrite pour raison médicale).

Lors du prélèvement et du stockage du LCR, différents paramètres peuvent affecter la qualité du métabolome. Il peut par exemple s'agir du temps et de la température de conservation entre le prélèvement et le stockage, du type de tube utilisé ou de la température de stockage. Il a été montré que parmi ces paramètres, la température de stockage est l'élément qui influence le plus la stabilité du métabolome (Wuolikainen et al., 2009). Un stockage à - 20 °C conduirait à des modifications du pH et de la pression partielle en CO₂, ce qui induirait des modifications notamment au niveau des acides organiques. Pour le LCR, un stockage à - 80°C apparait donc comme indispensable. Le type de

tube de prélèvement (polystyrène ou polypropylène) étant aussi un facteur de discrimination, il est recommandé de n'utiliser qu'un seul type de tube au sein d'une même étude. Il est également possible d'ajouter une étape de centrifugation du LCR avant congélation. Une fois centrifugé, le délai de conservation à température ambiante (jusqu'à 2h) avant la congélation ne semble pas affecter le métabolome (Rosenling et al., 2011).

3.2.2 Extraction des métabolites

Avant l'analyse proprement dite, l'échantillon doit être purifié afin d'éliminer au mieux les composés n'appartenant pas au métabolome étudié. Cela permet d'augmenter l'efficacité d'ionisation des molécules d'intérêt et donc d'augmenter la sensibilité de l'analyse. En revanche il est important de ne pas perdre de métabolites dans cette étape au risque de voir la qualité de la couverture métabolique diminuer. Il y a donc un compromis à trouver entre la pureté et l'exhaustivité du contenu métabolique de l'échantillon final. Si la spécificité de l'extraction peut être l'objectif pour des analyses ciblées sur certaines familles de molécules, la limitation de la perte d'information doit être le premier critère à prendre en compte pour les analyses non ciblées. Face à la diversité des propriétés physico chimiques des métabolites, l'exhaustivité de l'extraction passe le plus souvent par une simplification de l'extraction afin de limiter les risques de perte de métabolites.

Un autre critère à prendre en compte est la reproductibilité de l'extraction (Vuckovic, 2012). En effet, les analyses métabolomiques impliquent souvent une comparaison de données entre individus. Il est donc essentiel que la variabilité générée par l'extraction soit la plus faible possible afin de ne pas masquer les variations entre les différents échantillons.

3.2.2.1.1 Plasma/sérum

Le sang est un fluide très riche en protéines (Chang et al., 2007) ($60-80 \text{ g.L}^{-1}$) qu'il est nécessaire d'éliminer. En effet, une forte quantité de protéines peut conduire d'une part à un colmatage des colonnes, ce qui détériore la qualité de la séparation par chromatographie liquide, et d'autre part à une baisse de sensibilité dans la détection des métabolites par augmentation de l'effet matrice et de la suppression d'ionisation (Bruce et al., 2009).

Trois techniques de traitement de l'échantillon sont principalement utilisées : la précipitation par ajout de solvants organiques, l'extraction liquide-liquide et l'extraction sur phase solide (SPE, de l'anglais « solid phase extraction ») (Raterink et al., 2014).

La précipitation des protéines peut se faire par différentes techniques. Ces techniques sont toutes basées sur l'altération de la structure tridimensionnelle des protéines qui induit une baisse de solubilité et leur précipitation. Différents protocoles ont été testés tels que l'ajout d'acide, le chauffage ou l'ajout de solvants organiques. La nécessité de conserver intacte une large couverture

métabolique exclut l'utilisation d'acide ou de la chaleur. A l'inverse, l'utilisation de solvant organique, notamment le méthanol, semble être l'approche la plus robuste pour les études de métabolomique non ciblée (Michopoulos et al., 2009; Sitnikov et al., 2016; Want et al., 2006). Différentes études ont été menées sur la qualité de l'extraction en fonction du type de solvant utilisé. Le choix du solvant optimal ne fait pas consensus. Certains montrent qu'une extraction à partir d'un mélange méthanol : éthanol (1 : 1) ajouté au plasma dans des proportions (4 : 1) apporte de meilleurs résultats en terme de couverture métabolique (Bruce et al., 2009; Yang et al., 2013). L'apport du méthanol semble surtout se faire sur la récupération des composés hydrophobes comme les acides gras ou les phospholipides. D'autres conseillent l'utilisation d'acétonitrile dans des proportions (2 : 1) (Rico et al., 2014). Il semblerait qu'une tendance se dégage : l'acétonitrile et l'acétone permettent une meilleure précipitation des protéines alors que le méthanol et l'éthanol améliorent la couverture métabolique et la robustesse de la méthode. Une fois les protéines précipitées, elles sont éliminées par une étape de centrifugation. Lors de cette centrifugation les protéines forment un culot et le surnageant contenant les métabolites est récupéré pour analyse. La difficulté avec la centrifugation est qu'elle est compliquée à automatiser. Une autre possibilité pour éliminer les protéines précipitées est la filtration. Récemment, des solutions ont été développées sous forme de plaques à puits qui permettent un ajout de solvant puis une filtration par mise sous pression ou sous vide (Cao et al., 2014).

L'extraction liquide-liquide est basée sur la différence de solubilité des composés entre deux solvants non miscibles. Un solvant organique est ajouté à l'échantillon et un partage a lieu entre le milieu biologique (aqueux) et le solvant organique. Cela génère un mélange diphasique. Ce type d'extraction à base de chloroforme et de méthanol est très utilisé pour les études de lipidomique (Bligh and Dyer, 1959; Folch et al., 1951). L'extraction liquide-liquide utilisant du methyl-tertbutyl éther (MTBE) est devenue populaire au cours de ces dernières années y compris en métabolomique (Matyash et al., 2008). Cette étape d'extraction peut même directement être faite dans les flacons d'injection LC-MS. Deux phases se forment et la profondeur de l'aiguille de prélèvement définit la phase qui sera injectée, les deux phases pouvant donner lieu à deux analyses successives (Whiley et al., 2012). Elle permet une bonne couverture des métabolites et des lipides mais reste moins adaptée que la précipitation protéique pour l'extraction des composés les plus polaires et semble être moins reproductible (Sarafian et al., 2014).

L'extraction sur phase solide (SPE, *Solid Phase Extraction*) consiste à adsorber les composés à extraire sur une cartouche contenant une phase stationnaire, à laver cette phase afin d'éliminer les composés non désirés et enfin à éluer les composés d'intérêt hors de la colonne pour les analyser ensuite. La SPE permet donc de purifier l'échantillon et de le préconcentrer si le volume d'élution est inférieur au volume initial. Cette approche peut être très intéressante lors de l'analyse ciblée de métabolites présents à l'état de trace. En métabolomique non ciblée, la SPE est moins utilisée du fait

de sa sélectivité plus importante qui limite l'obtention d'une couverture métabolique globale. Cependant, elle apporte une plus grande reproductibilité à l'extraction, ce qui peut être un avantage pour l'observation de faibles variations de teneur en métabolites entre différentes conditions. Elle présente aussi l'avantage de pouvoir être couplée en ligne à un système chromatographique ce qui limite le nombre de manipulations de l'échantillon à réaliser et augmente la robustesse de la méthode. Une variante de la SPE, appelée « Turbulent Flow Chromatography », permet par exemple l'injection directe de sérum ou de plasma non traité. La colonne remplie de larges particules (25-50 μm) ne retient pas les macromolécules qui sont éliminées lors du passage la phase mobile avec un débit important générant des conditions de flux turbulents à l'intérieur de la colonne. Les petites molécules sont retenues à l'intérieur de la phase stationnaire poreuse (polystyrène - divinylbenzène) et peuvent être éluées en condition de flux laminaire dans la colonne chromatographique (Vuckovic, 2012). Cette technique permet une mesure des analytes comparable à ce qui peut être fait en SPE mais sa répétabilité semble être un peu moins bonne (Michopoulos et al., 2010).

Une étude récente menée par Sitnikov *et al.* a évalué ces différentes techniques d'extraction selon plusieurs critères : la couverture métabolique, le taux de recouvrement, la répétabilité et l'effet matrice lors de l'analyse. Le Tableau 1 résume les avantages de chaque technique d'extraction pour une analyse de métabolomique non ciblée par LC-MS. Ces résultats confirment la plus large couverture des méthodes de précipitation par solvant par rapport à la SPE ou l'extraction liquide-liquide, les meilleurs résultats étant obtenus avec l'utilisation de méthanol dans un ratio de quatre volumes de méthanol pour un volume de plasma (Sitnikov et al., 2016).

Méthode d'extraction	Couverture métabolique	Recouvrement	Répétabilité	Effet matrice
Méthanol/éthanol	+++	+++	+++	++
Méthanol	+++	+++	+++	++
Méthanol/MTBE	+++	+++	+++	++
MTBE	++	++	+	++++
SPE C ₁₈	++	+++	+	++++
SPE PEP ₂	++	+++	++	++
SPE IEX	+	+	++	+++

Tableau 1 : Résumé des performances des différentes méthodes d'extraction des métabolites plasmatiques (Sitnikov et al., 2016)

3.2.2.1.2 Urines

Du fait de sa faible teneur en protéines chez l'homme (Chang et al., 2007) (0.5-1 g.L⁻¹) la préparation des échantillons d'urine peut être simplifiée par rapport à ce qui doit être envisagé pour

un échantillon sanguin. L'approche utilisée est celle appelée « dilution et injection » (Khamis et al., 2017). Il s'agit simplement de diluer les échantillons afin de limiter l'effet matrice et la suppression d'ionisation avant de les injecter. La concentration en métabolites étant particulièrement importante dans les urines, le facteur de dilution doit être adapté à la sensibilité de l'instrument. S'il n'y a pas eu d'étape de filtration au moment de la collecte, une centrifugation doit être réalisée avant la dilution afin d'éliminer d'éventuels débris cellulaires. Pour des analyses plus spécifiques, d'autres protocoles peuvent être utilisés comme une hydrolyse enzymatique afin de déconjuguer les composés issus du métabolisme secondaire ou des étapes d'extraction sur phase solide ou liquide-liquide. Néanmoins, ces étapes ne sont pas conseillées pour les études non ciblées visant à obtenir la plus grande couverture possible du métabolome. Attention toutefois car selon les espèces, le volume de filtration glomérulaire varie ce qui peut conduire à des concentrations protéiques nécessitant une étape de précipitation.

3.2.2.1.3 Liquide céphalorachidien

Le liquide céphalorachidien est lui aussi relativement pauvre en protéines (0.15-0.6 g.L⁻¹). La nécessité d'une précipitation protéique est alors discutable. Du fait de cette pauvreté en protéines, le LCR est généralement analysé sans dilution préalable afin d'optimiser la sensibilité. Cette absence de dilution est une des justifications évoquées à la réalisation d'une étape de précipitation des protéines. La majorité des études en métabolomique sur le LCR réalise cette étape de précipitation des protéines au méthanol (Wang et al., 2016; Weiss et al., 2016; Wishart et al., 2008)..

3.3 Acquisition des données

Une fois les métabolites extraits, ils sont analysés. Cette analyse est constituée d'une étape de séparation des métabolites selon différents processus physico-chimiques suivie de leur analyse en spectrométrie de masse.

3.3.1 Les techniques séparatives en ligne

Ces techniques jouent un rôle essentiel dans les analyses métabolomiques, notamment en spectrométrie de masse. Elles permettent de séparer les composés en amont de leur analyse. Cela conduit à l'enregistrement de spectres moins chargés et donc à une augmentation de la sensibilité. De plus, cela apporte des informations complémentaires sur la nature des métabolites en fonction de leurs interactions avec le système séparatif.

3.3.1.1 La chromatographie en phase gazeuse

La chromatographie en phase gazeuse (GC, *Gas Chromatography*) est une technique séparative développée à la fin des années 1950. Le mélange à analyser est mis sous phase gazeuse

avant d'être entraîné dans une colonne chromatographique par un gaz (CO₂ en général). Il existe différents types de colonnes. Pour les colonnes dites « remplies », l'interaction entre l'échantillon et la phase stationnaire s'effectue à la surface des grains de silice. Pour les colonnes capillaires, l'interaction s'effectue à la surface du capillaire. La phase stationnaire peut être de deux types : soit un liquide (chromatographie gaz-liquide) sous forme de film adsorbé sur le support de la colonne, soit un solide directement adsorbant (chromatographie gaz-solide). Ce sont les interactions des analytes avec la phase stationnaire d'un côté et le gaz de l'autre qui conduisent à la séparation des composés.

Pour une approche métabolomique, la chromatographie en phase gazeuse présente différents avantages. Elle offre une haute capacité de séparation des pics et une forte sensibilité (Haggarty and Burgess, 2017). Le pouvoir séparatif de la GC est encore amélioré par l'utilisation de la GC-2D. Deux colonnes présentant des phases stationnaires différentes sont couplées en série ce qui augmente grandement la capacité de séparation des pics. Plusieurs études en métabolomique ont utilisé ces techniques 2D (Koek et al., 2008). De nouvelles colonnes chromatographiques ont également été développées afin d'optimiser la couverture métabolique par cette technique. Il s'agit de colonnes présentant une phase stationnaire liquide et ionique. Elles permettent de séparer efficacement les composés polaires mais aussi les composés apolaires (Anderson and Armstrong, 2005). La GC est souvent couplée à la spectrométrie de masse par une source d'ionisation par impact électronique. Avec ce type de source, la fragmentation des composés lors de leur ionisation est très reproductible. Cela permet de constituer des bases de données universelles de spectres de fragmentation, ce qui facilite grandement l'identification des métabolites détectés par cette technique. La contrepartie est qu'il est souvent difficile d'observer la masse de l'ion parent ce qui limite l'interprétation des spectres de masse des composés absents des bases de données.

La GC présente également des caractéristiques limitantes pour des analyses de métabolites. En effet, il faut que les composés analysés soient thermostables et volatiles. Ce qui représente un vrai handicap pour les composés polaires qui constituent une part importante du métabolome. Il est alors nécessaire de réaliser une étape de dérivation, c'est-à-dire de modifier chimiquement ces composés afin de les rendre volatiles et d'améliorer leur stabilité thermique ainsi que leur détectabilité. Cette étape est assez contraignante car elle peut induire des réactions de dégradation au sein du métabolome étudié et ainsi nuire à l'aspect quantitatif de l'analyse.

La GC est donc une technique séparative qui présente un fort potentiel pour l'analyse de composés apolaires de faible masse.

3.3.1.1.2 La chromatographie liquide

La chromatographie en phase liquide permet de séparer des composés chimiques dissouts dans un solvant. Il existe de nombreux types de chromatographie liquide en fonction du mécanisme de rétention mis en jeu. On peut citer la chromatographie d'adsorption, la chromatographie d'exclusion stérique, la chromatographie par échange d'ions ou encore la chromatographie de partage. Seule la chromatographie de partage, qui est de loin la plus utilisée pour des applications en métabolomique, sera détaillée ici.

Comme pour la chromatographie gazeuse, la séparation des composés est basée sur une différence d'affinité entre deux phases, l'une stationnaire, en générale fixée sur une colonne chromatographique, et l'autre mobile qui parcourt cette colonne.

Il existe aujourd'hui différents grands types de phases stationnaires. Historiquement, les colonnes chromatographiques sont remplies de petites particules sur lesquelles sont greffés des groupements chimiques dont la nature définit le type d'interactions avec les analytes. La taille et la nature de ces particules joue un rôle essentiel dans les caractéristiques de la chromatographie. En chromatographie liquide conventionnelle, il s'agit de particules pleines, greffées en surface et de taille variant de 3 à 5 μm . Ces particules de « grande » taille offrent une moindre résistance à l'écoulement de la phase stationnaire ce qui permet de travailler à basse pression. Cependant, les performances chromatographiques de ces phases stationnaires sont limitées. Afin de les améliorer, de nouvelles phases ont été développées avec des particules de plus petites tailles (inférieure à 2 μm) (Fekete et al., 2012). Cela entraîne une forte amélioration du nombre de plateaux théoriques mais également une plus forte résistance à l'écoulement dont résulte une hausse importante de la pression. On parle alors de chromatographie à ultra haute pression (autour de 1000 bars). Il a donc fallu développer des instruments capables de résister à de telles pressions. Ce type de chromatographie s'est développé à partir du milieu des années 2000 et est aujourd'hui le plus utilisé. La haute résolution chromatographique permise par ce type de particules conduit à une meilleure séparation des composés et une meilleure sensibilité qui est due à la diminution de la largeur des pics compensée par une augmentation de leur hauteur ce qui améliore le rapport signal sur bruit (Fekete et al., 2015). D'autres types de phases stationnaires ont récemment été développés afin d'améliorer les performances chromatographiques des colonnes. Il s'agit par exemple de particules superficiellement poreuses qui permettent de réduire le temps d'analyse tout en conservant la qualité de la séparation (Gritti et al., 2007). Existente aussi des colonnes dites monolithiques pour lesquelles la phase stationnaire est directement polymérisée à l'intérieur de la colonne. Avec l'apparition, à partir des années 2010, de colonnes monolithiques de seconde génération, des vitesses de séparation comparables à celles observées en chromatographie liquide haute performance (HPLC, *High Performance Liquid Chromatography*) peuvent être atteintes.

L'autre point essentiel en chromatographie liquide, qu'elle soit conventionnelle ou à ultra haute pression, est la nature du greffage sur les particules. Il existe deux types principaux de greffage. S'il s'agit de greffage de chaînes alkyles apolaires, on parle de chromatographie liquide en phase inverse. Ce type de greffage est particulièrement adapté à la séparation de composés apolaires. Il peut aussi s'agir de greffage de groupements polaires tel que des groupements amines ou nitriles. Cela permet d'obtenir une bonne séparation des composés polaires. On parle alors de chromatographie en phase normale. Les types de chromatographie les plus utilisés en métabolomique sont la chromatographie en phase inverse C₁₈ et la chromatographie d'interaction hydrophile (HILIC, *Hydrophilic Interaction Liquid Chromatography*). La complémentarité de ces deux techniques permet d'optimiser la couverture métabolique de l'analyse (Boudah et al., 2014).

De par la nature plutôt polaire des composés constituant le métabolome, la chromatographie en phase liquide est aujourd'hui la technique chromatographique la plus utilisée dans les analyses métabolomiques.

3.3.1.1.3 La chromatographie en phase supercritique

La chromatographie en phase supercritique (SFC, *Supercritical Fluid Chromatography*) est une technique séparative au cours de laquelle un fluide supercritique (fluide chauffé au-dessus de sa température critique et comprimé au-dessus de sa pression critique) est utilisé comme phase mobile. Ce fluide présente une densité, une viscosité et une diffusivité intermédiaire entre un liquide et un gaz (Laboureur et al., 2015).

La SFC est une technique mise en point dans les années 1980. Elle a été mise de côté pendant une vingtaine d'années avant de revenir sur le devant de la scène dans les années 2010 avec la fabrication de nouveaux systèmes par les constructeurs d'instruments. Elle reprend de nombreux aspects de la LC et de la GC. Elle présente une faible hauteur de plateau théorique même à relativement haut débit ce qui permet des analyses très résolutive à haut débit. De plus la polarité d'une phase mobile supercritique augmentant avec la densité, elle-même pouvant être grandement modifiée en jouant sur la pression et la température, de nombreux types de séparation peuvent être obtenus (Smith, 1999).

Le dioxyde de carbone est le principal fluide utilisé en SFC. Il présente l'avantage d'avoir une température et une pression critique relativement basses (31°C, 73.8 bars). De plus, il est chimiquement inerte, relativement peu toxique, facile à manipuler, peu cher et peu polluant. Dans son état supercritique, il a une faible polarité ce qui en fait un premier choix pour l'analyse de lipides (Lesellier, 2001; Señoráns and Ibañez, 2002). Il est également possible de modifier significativement la polarité de la phase mobile en ajoutant un solvant organique polaire tel que du méthanol (Bamba et al., 2012). Cela permet de couvrir une gamme de métabolites plus large.

3.3.1.1.4 L'électrophorèse capillaire

L'électrophorèse capillaire (CE, *Capillary Electrophoresis*) est une autre technique séparative utilisée en métabolomique, notamment pour les composés polaires et chargés. Contrairement à la LC et à la GC, la CE fonctionne sur un mode de séparation différent basé sur le rapport taille sur charge. Ce mode de séparation fait de la CE une technique très complémentaire des méthodes chromatographiques, avec la possibilité de séparer des composés qu'il est difficile de distinguer en LC (Ramautar et al., 2009). Elle est capable de séparer à la fois des petits ions organiques, des protéines et même des bactéries entières. Il existe différents types d'électrophorèses capillaires comme l'électrophorèse micellaire ou l'électrophorèse chirale mais cela implique l'utilisation d'additifs peu compatibles avec une détection en spectrométrie de masse. L'électrophorèse capillaire la plus utilisée en métabolomique reste donc l'électrophorèse de zone qui ne nécessite pas d'additif (Monton and Soga, 2007).

La CE est une technique rapide, avec un fort pouvoir séparatif et une faible consommation de solvant. De plus, elle ne nécessite qu'une très faible quantité d'échantillon (de l'ordre du nanolitre). Cette caractéristique peut aussi être un handicap car elle augmente la limite de détection de la méthode. Des techniques de préconcentration peuvent pallier ce problème.

3.3.1.1.5 L'introduction directe

L'introduction directe est une approche complètement différente puisqu'elle ne fait pas appel aux techniques séparatives. Les échantillons sont introduits directement dans le spectromètre de masse. Pour les approches métabolomiques, notamment pour l'analyse de cohortes d'échantillons biologiques, l'introduction directe se fait le plus souvent en mode « FIA » (*Flow Injection Analysis*). Dans ce mode, une quantité définie d'échantillon est injectée rapidement au sein d'un flux de solvant et portée par ce flux jusqu'au spectromètre de masse. Le signal observé a donc la forme globale d'un pic qui s'étend sur une dizaine de secondes (Draper et al., 2013).

Ce mode d'analyse a comme principaux avantages d'être rapide et facile à mettre en place. Le gain de temps par rapport à l'utilisation d'un système séparatif réduit le temps total d'analyse à une ou deux minutes ce qui permet d'analyser plusieurs centaines d'échantillons par jour. De plus, la préparation des échantillons pour ce type d'analyse est souvent plus sommaire ce qui permet de faire des analyses d'échantillons à haut débit. Cette approche a également l'avantage de nécessiter moins de travail d'optimisation que l'utilisation d'un système chromatographique (González-Domínguez et al., 2016) même si la nature ou le débit du solvant peuvent avoir une influence sur la qualité des résultats générés. Le prétraitement des analyses est aussi plus aisé, notamment car il n'y a pas de nécessité d'aligner des temps de rétention entre les échantillons. Attention cependant à la forme des pics qui peut être assez différente de celle d'un pic chromatographique classique ce qui les rend

souvent non détectables par la majorité des logiciels associés à une technique chromatographique. La détection automatique des pics doit donc être effectuée à l'aide d'algorithmes spécifiques à l'introduction directe (Delabrière et al., 2017).

En contrepartie, l'un des principaux inconvénients de l'introduction directe est lié au phénomène de suppression d'ionisation ou « effet matrice ». Il est dû à la présence simultanée de nombreux composés au niveau de la source d'ionisation. Une compétition va avoir lieu entre les différentes espèces présentes et seul un nombre réduit d'entre elles pourra s'ioniser pour ensuite être détecté. Le signal lié aux autres espèces est supprimé rendant leur détection impossible (Annesley, 2003). L'une des solutions pour limiter ce phénomène est d'optimiser la dilution de l'échantillon ce qui limite cette suppression d'ionisation et permet de détecter un nombre plus important de composés (Beckmann et al., 2008). Le second inconvénient principal de cette approche est le fait qu'il n'est pas possible de distinguer des isomères ou plus globalement des composés de masse trop proches. Cela a un impact limité en métabolomique ciblée si l'on sait qu'il n'y a pas d'ions isobares des ions suivis. Ainsi différentes méthodes ont pu être développées en introduction directe, notamment pour cibler des acylcarnitines, des phospholipides ou des sphingolipides (Römisch-Margl et al., 2012). Cependant, cela est beaucoup plus problématique pour les approches non ciblées. En conséquence, l'acquisition d'empreinte métabolique par introduction directe s'est surtout développée en métabolomique avec l'utilisation des spectromètres de masse à haute résolution (Chekmeneva et al., 2017; González-Domínguez, 2017; Madalinski et al., 2008). La haute résolution permet de distinguer des composés isobares de masse très proche et ainsi de pallier l'absence de technique séparative en amont et d'accroître la couverture métabolique lors d'acquisition en introduction directe.

L'utilisation adaptée de l'une ou l'autre de ces différentes techniques séparatives permet d'obtenir des données de qualité. L'autre élément clé dans l'acquisition des données lors d'une étude de métabolomique est le spectromètre de masse.

3.3.2 Les nouveaux instruments en spectrométrie de masse à haute résolution

Comme expliqué précédemment, un spectromètre de masse est composé d'une source d'ionisation et d'un analyseur couplé à un détecteur. Différents types de sources d'ionisation peuvent être utilisés en métabolomique comme l'ionisation chimique à pression atmosphérique (APCI, *Atmospheric-Pressure Chemical Ionisation*), la photoionisation à pression atmosphérique (APPI, *Atmospheric-Pressure Photoionisation*) ou l'ionisation par électrospray (ESI, *Électrospray Ionisation*). Cette dernière est la plus utilisée pour les analyses métabolomiques par LC-MS avec 90% des publications mentionnant l'utilisation de cette source contre 6 % pour l'APCI et 2% pour l'APPI (Forcisi

et al., 2013). Ces différentes sources ne sont pas détaillées dans le corps de ce manuscrit mais sont présentées en Annexe 1.

La partie analyseur et détecteur constitue le cœur du spectromètre de masse. Une fois formés au niveau de la source, les ions sont conduits jusqu'à l'analyseur via des optiques de transfert. Ils vont être analysés afin de déterminer leur rapport masse sur charge. En métabolomique non ciblée, l'intégralité du métabolome étant encore loin d'être connue, il est nécessaire de pouvoir identifier sans *a priori* les molécules détectées. Pour ce faire, l'utilisation d'analyseur à haute résolution est essentielle afin d'obtenir les informations les plus précises possibles. La haute résolution permet d'abord de séparer les nombreux composés isobares (composés ayant la même masse nominale mais des masses exactes différentes) présents dans les fluides biologiques complexes. Cela donne un signal spécifique pour chaque ion et non plus un signal regroupant plusieurs ions isobares. Grâce à cela, la précision de mesure de masse d'un ion isolé est grandement augmentée (de l'ordre du ppm), ce qui permet d'avoir accès à la formule brute des ions détectés.

Il existe différents types de spectromètres de masse à haute résolution. Les principaux sont les spectromètres de masse à Temps de Vol (TOF, *Time Of Flight*), les spectromètres à Résonance Cyclonique Ionique (FT-ICR, *Fourier Transform – Ion Cyclotron Resonance*) et les spectromètres de masse de technologie Orbitrap. Différents éléments permettent d'évaluer les performances de ces analyseurs : la résolution, la précision de mesure de masse et la vitesse d'acquisition.

La résolution correspond à la capacité de l'analyseur à séparer, sur le spectre de masse, deux ions présentant des masses proches. Cette résolution est donc en lien directe avec la largeur des pics de masse enregistrés. Elle est définie comme :

$$R = m/\Delta m$$

où m correspond au rapport masse sur charge mesuré et Δm à la largeur du pic

Afin d'uniformiser les définitions de la résolution, cette dernière est en générale calculée en prenant la largeur du pic à la moitié de sa hauteur. Cette résolution sera appelée FWHM (*Full Width Mid High*). Il est aussi important de faire la différence entre la résolution et le pouvoir résolutif (PR) d'un analyseur. La résolution peut être mesurée pour chaque pic de masse enregistré alors que le pouvoir résolutif correspond à un paramétrage de l'instrument. Il donne une idée de la résolution réellement atteinte mais sur la plupart des analyseurs la résolution réelle varie avec le rapport m/z du pic de masse étudié.

La précision de mesure de masse correspond à la capacité de l'analyseur à mesurer avec justesse la masse de l'ion. Elle s'exprime généralement sous la forme d'une erreur entre la masse

mesurée et la masse théorique et peut soit être donnée en dalton soit en partie par million (ppm). Cette dernière définition est généralement retenue pour les analyseurs à haute résolution.

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

où M_{exp} correspond à la masse mesurée et M_{th} à la masse théorique

La précision de mesure de masse sera un des éléments déterminant lors de l'analyse métabolomique, notamment dans l'identification des signaux d'intérêt.

Enfin, la vitesse d'acquisition correspond au temps nécessaire à l'enregistrement d'un spectre de masse. Elle s'exprime soit sous forme d'une fréquence (x spectres par seconde ou x Hertz) soit sous forme d'un temps d'acquisition en millisecondes. La vitesse d'acquisition est un élément important à prendre en compte, notamment lors d'analyse par couplage LC-MS. En effet, pour ce type d'analyse, le temps disponible pour enregistrer le signal d'un ion est déterminé par la largeur du pic chromatographique. La vitesse d'acquisition doit donc être adaptée au système chromatographique utilisé.

Le fonctionnement de chaque type d'instrument est détaillé en Annexe 2 mais chacun présente ses avantages et ses inconvénients. Le Tableau 2 présente les performances maximales atteintes par chacun des types d'analyseur. Les analyseurs à temps de vol sont les moins résolutifs mais présentent la plus haute vitesse d'acquisition ce qui a permis de développer de nouvelles approches d'acquisition (voire Paragraphe 4). A l'opposé, les FT-ICR nécessitent plus de temps pour enregistrer un spectre mais ils permettent d'atteindre la plus haute résolution et la meilleure précision de mesure de masse. Les Orbitrap présentent des caractéristiques (résolution, précision de mesure de masse, vitesse d'acquisition, coût) intermédiaires parmi les analyseurs à haute résolution. Bien que relativement récents, ils tendent donc à être de plus en plus utilisés pour les analyses métabolomiques. De par sa position centrale et son utilisation grandissante, c'est cette technologie qui sera utilisée tout au long de ce travail de thèse.

	Pouvoir Résolutif	Erreur de mesure de masse (ppm)	Vitesse d'acquisition	Coût
Temps de Vol	Jusqu'à 50 000 @ m/z 950	< 1	30 Hz à PR 50 000 100 Hz à PR 35 000	+
Orbitrap	Jusqu'à 1 000 000 @m/z 200	< 1	0.5 Hz à PR 1 000 000 40 Hz à PR 7 500	++
FT-ICR	Jusqu'à 20 000 000	< 0.25	1 Hz à PR 1 200 000 3 Hz à PR 150 000	+++

Tableau 2 : Résumé des performances de différents types d'analyseur ("Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer," 2018, "scimaX MRMS - Taking science to the max - Identification with confidence," 2018, "Spectrométrie de masse haute définition SYNAPT G2-Si : Waters," 2018)

Les spectromètres de masse modernes sont souvent composés de plusieurs types d'analyseurs. En générale, il s'agit d'un analyseur à basse résolution (souvent un quadripôle) couplé en série à un analyseur à haute résolution. Les analyseurs à basse résolution sont présentés en Annexe 2. Ce type de construction va permettre d'acquérir des données de profilage mais également des données de spectrométrie de masse en tandem. Cela consiste en la sélection et la fragmentation d'un nombre réduit d'ions dont les fragments seront analysés. Certains instruments dit « tribride » combinent même trois analyseurs différents en série ou en parallèle et donnent une flexibilité encore plus importante dans les types d'acquisition disponibles (Senko et al., 2013).

3.3.3 Les différents modes d'acquisition de données

Les nouveaux instruments disponibles permettent d'envisager de nouveaux modes de travail dans l'acquisition des données en spectrométrie de masse. Deux types d'information sont acquis au cours d'une analyse. Classiquement, les profils MS des échantillons sont enregistrés dans un premier temps. Ces informations sont traitées puis une seconde série d'acquisition est effectuée afin d'enregistrer des spectres de fragmentation qui vont servir pour l'identification des signaux d'intérêt. Cette acquisition séquentielle en deux temps présente un certain nombre d'inconvénients. Afin de pallier ces inconvénients, les nouveaux instruments permettent l'acquisition simultanée des spectres MS et MS². Ces modes d'acquisition sont en plein développement en métabolomique et seront détaillés en partie 4.

3.4 Travail sur le jeu de données

L'utilisation de spectromètres à haute résolution, notamment couplés à des systèmes chromatographiques, pour les analyses métabolomiques, conduit à la génération d'une quantité très importante de données. Une fois acquises, ces données doivent être traitées informatiquement afin d'en extraire l'information. En métabolomique non ciblée, l'utilisation d'outils bio-informatiques est

essentielle afin de pouvoir récupérer le plus d'information possible à partir des fichiers d'acquisition. Ce traitement de données se déroule en plusieurs étapes. La première étape est la validation du déroulement de l'analyse. Une fois l'analyse validée, les données brutes sont transformées en un tableau résumant les éléments identifiés lors de l'analyse. Une fois ces éléments obtenus, ils sont filtrés afin de ne garder que ceux ayant une pertinence analytique. Ces éléments sont ensuite annotés afin de leur attribuer un nom de métabolite. Des analyses statistiques sont ensuite réalisées pour identifier les éléments permettant de discriminer différents groupes au sein des échantillons. Enfin, ces résultats sont interprétés biologiquement. Toutes ces différentes étapes sont détaillées ci-après.

3.4.1 Validation du déroulement de l'analyse

Une fois les données acquises, différents éléments vont pouvoir être vérifiés afin de s'assurer du bon déroulement de l'analyse. Ces vérifications se basent sur deux éléments : l'utilisation d'étalons internes et l'analyse d'échantillons de contrôle qualité (QC).

3.4.1.1.1 Utilisation d'étalons

Les étalons sont des composés ajoutés aux échantillons afin de suivre la reproductibilité de l'expérience. Deux types d'étalons peuvent être utilisés : les étalons internes et les étalons externes (Katajamaa and Orešič, 2007). Les étalons internes sont ajoutés à l'échantillon avant toute autre manipulation. Ils vont donc être extraits au même titre que les autres métabolites puis injectés et analysés. Ce sont donc à la fois des marqueurs de la reproductibilité de l'étape de préparation de l'échantillon et de l'étape d'analyse en spectrométrie de masse. Les étalons externes sont ajoutés juste avant l'étape d'injection dans le système LC-MS. Ils vont donc permettre de vérifier le bon déroulement de cette dernière étape de l'analyse.

Afin de ne pas fausser l'analyse des échantillons biologiques, les étalons sont des composés non naturellement présents dans les échantillons. Il est possible d'utiliser soit des composés exogènes comme des médicaments ou alors d'utiliser des composés marqués. Le marquage est généralement fait à l'aide d'isotopes stables, comme le ^{13}C ou l' ^{15}N . Contrairement aux isotopes radioactifs, ces composés ne se dégradent pas dans le temps et ne nécessitent pas de précautions d'usage particulières.

Ces étalons permettent de vérifier rapidement et en temps réel la stabilité du système chromatographique et le bon déroulement de l'étape d'injection des échantillons. Les temps de rétention chromatographiques de ces molécules peuvent être relevés sur l'intégralité des échantillons et peuvent servir à repérer une dérive chromatographique au cours de la séquence d'analyse. Ils peuvent également servir à normaliser les résultats et corriger les déviations liées à l'extraction des métabolites à partir de l'échantillons biologique. Cet aspect sera détaillé par la suite.

Idéalement, le jeu d'étalons utilisé doit être adapté aux conditions d'analyse. En effet les molécules choisies doivent s'ioniser correctement dans le mode d'ionisation utilisé et donner un signal facilement détectable sans pour autant masquer le reste de l'information (Zhou et al., 2012). La concentration de chaque étalon peut être adaptée en fonction de son efficacité d'ionisation. Les étalons doivent idéalement être répartis sur l'ensemble de la gamme de temps de rétention afin de s'assurer de la stabilité de l'ensemble des temps. Les molécules étant retenues différemment selon le type de colonne chromatographique utilisée, un jeu d'étalons adapté à la C₁₈ ne sera pas forcément retenu correctement pour une analyse en HILIC.

3.4.1.1.2 Utilisation de contrôles qualité (QC)

L'échantillon de contrôle qualité (QC) est injecté au début, à intervalles réguliers et à la fin de la séquence d'analyse. Il sert à suivre la stabilité temporelle de l'ensemble des performances de l'analyse LC-MS et à détecter d'éventuelles dérives analytiques (Sangster et al., 2006). L'utilisation de contrôles qualité est aujourd'hui largement répandue et est reconnue comme l'un des principaux outils permettant de valider la qualité d'une analyse métabolomique non ciblée (Dudzik et al., 2017). Le nombre d'injections de QC doit représenter environ 10 % du total nombre total d'échantillons lors de l'analyse de grosses cohortes (plus de 100 échantillons) (Sangster et al., 2006). Pour les petites études, un QC devra être analysé toutes les 5 à 10 injections.

Un QC doit être représentatif de la composition, à la fois qualitative et quantitative, des différents échantillons de l'étude. Selon le type d'échantillon, la quantité d'échantillon disponible et le nombre d'échantillons, différents types de QC peuvent être préparés. Pour des échantillons liquides (plasma, sérum, urines, ...) disponibles en quantité importante, il est recommandé de constituer un mélange de chaque échantillon puis d'extraire ce mélange (Sangster et al., 2006). Le mélange ainsi préparé est représentatif de la composition de l'ensemble des échantillons de l'étude (Liu et al., 2015). Dans le cas où la quantité d'échantillon disponible est limitée ou lorsqu'il s'agit de matrice difficile à homogénéiser (tissus, fèces, ...), il est possible d'utiliser un QC post extraction en mélangeant une quantité équivalente de chaque extrait d'échantillon (Dunn et al., 2012). Ces QC permettront de suivre les variations des performances associées à l'analyse LC-MS indépendamment des études précédentes.

En plus du suivi des performances du système analytique, l'utilisation de QC a plusieurs autres intérêts (Dunn et al., 2011). Les QC peuvent être utilisés en amont de la séquence d'analyse afin d'équilibrer la colonne avec un échantillon comparable à ceux qui seront analysés. Cela permet de s'assurer d'une meilleure stabilité des temps de rétention et d'obtenir des résultats plus reproductibles (Gika et al., 2007). Le nombre de ces injections d'équilibration varie entre 3 et 10 selon le type de colonne chromatographique utilisé. L'utilisation de QC va également permettre de corriger certaines dérives analytiques qui peuvent apparaître (baisse d'intensité du signal, ...) au

cours de la séquence ou d'intégrer des données issues de plusieurs lots analytiques différents. Enfin les QC vont également servir à filtrer les données. Ces différentes utilisations des QC seront détaillées ci-après.

3.4.1.1.3 Utilisations d'échantillons « blancs »

L'utilisation d'échantillons « blancs » est essentielle à la validation de l'analyse. Dans l'idéal, un blanc correspond à une matrice sans analytes d'intérêt. La préparation de cet « échantillon » doit être comparable à celle des échantillons réels en termes d'extraction, de stockage et d'analyse (Dudzik et al., 2017). En pratique il s'agit souvent simplement d'un mélange de phases mobiles dans des proportions identiques à celles dans lesquelles les échantillons sont repris avant d'être injectés (Bader et al., 2016). A minima, des échantillons « blancs » doivent être injectés en début et en fin d'analyse. Il n'est pas forcément recommandé d'analyser des échantillons « blancs » en cours de séquence car leur trop grande différence de composition avec les échantillons réels peut entraîner un déconditionnement du système conduisant à des décalages de temps de rétention ou des modifications de formes de pics (Gika et al., 2016). Ces échantillons « blancs » vont notamment servir à filtrer les données analytiques.

3.4.2 Outils informatiques pour la détection automatique des signaux

Une fois l'analyse validée, la première étape du traitement de données est la détection automatique des informations présentes dans les fichiers d'acquisition. Il s'agit à la fois d'informations spectrales avec la détection des différents signaux de masse présents à un instant t mais aussi d'informations chromatographiques avec la détection des pics chromatographiques correspondant aux différents ions détectés au cours du temps. L'objectif de cette étape est de transformer les signaux bruts en un format matriciel permettant de traiter facilement et de manière groupée les données de l'ensemble des échantillons.

Différents logiciels, soit libres (« *open source* ») soit commerciaux, permettent de réaliser cette étape. La majorité des logiciels commerciaux sont commercialisés par les constructeurs d'instruments. Ces logiciels présentent l'avantage d'être en général assez simples d'utilisation mais ils ne sont pas universels. Cela implique qu'il n'est pas possible d'appliquer le même processus à des fichiers venant d'instruments de marques différentes ce qui complexifie les études communes à plusieurs laboratoires. De plus, comme il s'agit de logiciels brevetés, les données de code ne sont pas accessibles et il n'est pas possible de savoir clairement quels sont les algorithmes utilisés. Enfin, ces logiciels sont souvent en retard par rapport à l'évolution des besoins.

Parmi les logiciels *open source*, XCMS (Smith et al., 2006) est l'un des plus utilisés. Il présente l'avantage d'être codé sous un format de langage R qui permet d'avoir accès au code source et de

comprendre, de suivre et éventuellement de modifier ou d'adapter les différentes étapes de l'analyse. Ce logiciel permet de traiter les données sous différents formats libres de fichiers tels que .NetCDF ou .mzXML. Comme il s'agit de formats libres, les données acquises en format constructeur doivent préalablement être converties sous l'un de ces formats. Il existe actuellement différents logiciels *open source* qui permettent de faire cette conversion (Kessner et al., 2008). De nouvelles versions de ces logiciels de conversion sont disponibles lorsque la structure des fichiers au format constructeur est modifiée et la dernière version du logiciel doit donc être utilisée pour prendre en compte ces potentielles modifications, notamment lors de changement d'instrument ou de version du logiciel d'acquisition.

Lors du traitement d'un jeu de donnée par XCMS, différents algorithmes sont utilisés (Figure 4). La première étape est la détection de pics sur chacun des échantillons. La seconde étape est l'alignement des pics identifiés sur les différents échantillons. En effet de petites variations des temps de rétentions peuvent être observées au cours de l'analyse. Une étape de correction de ces temps de rétention est donc réalisée pour permettre de grouper de façon optimale les signaux d'un même ion présents sur les différents échantillons. Suit enfin une étape de recherche des valeurs manquantes. En effet, les intensités des pics d'un composé pouvant varier d'un échantillon à l'autre, il est possible que certains pics ne soient pas détectés sur tous les échantillons lors de la première détection sans *a priori*. Ils sont alors recherchés de manière ciblée pour permettre de vérifier que la première détection n'a pas oublié un signal.

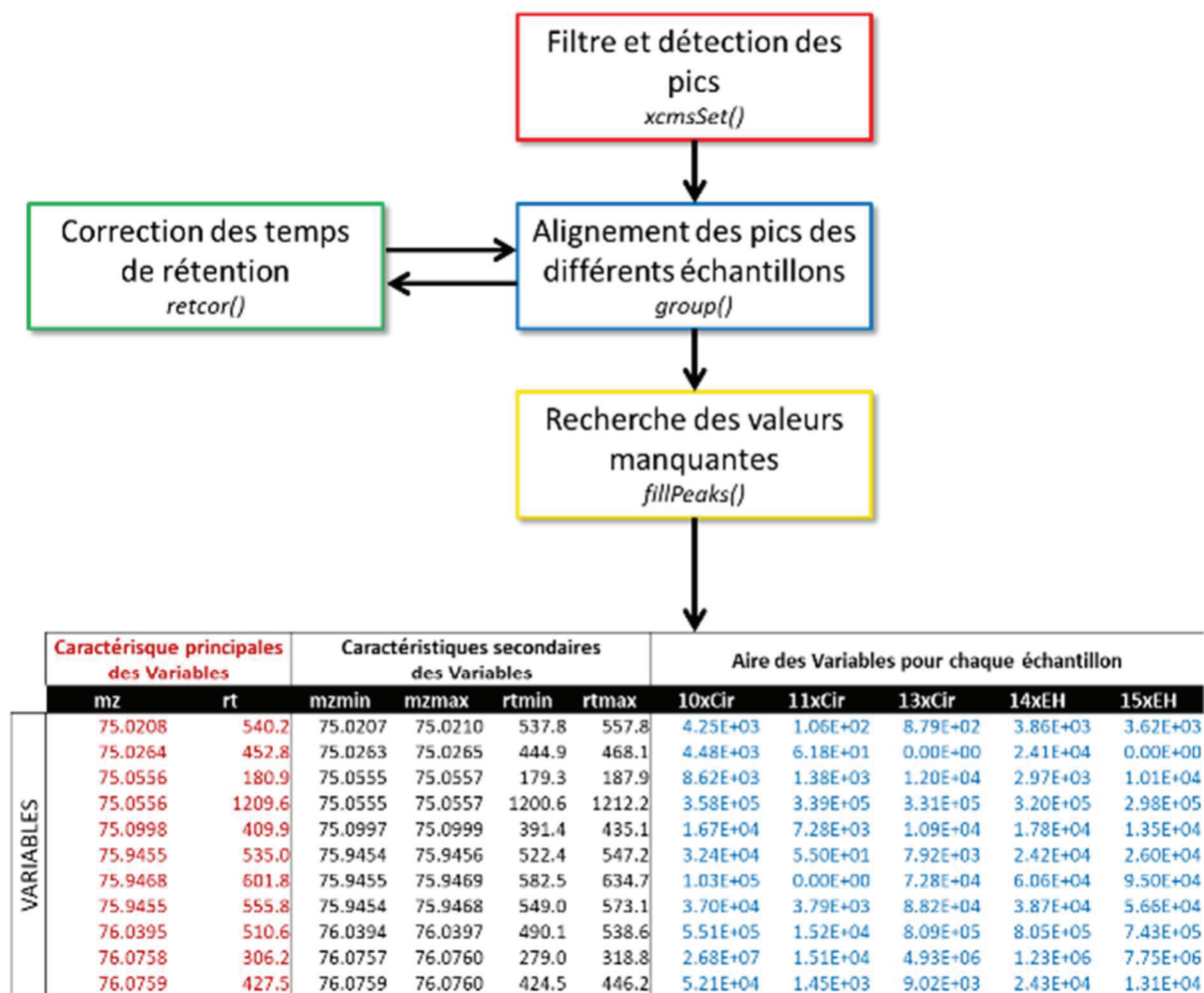


Figure 4 : Déroulé d'un traitement de données par XCMS

La première étape de détection de pics est l'étape la plus délicate du processus de traitement quel que soit le logiciel utilisé. Différentes études ont cherché à comparer les résultats obtenus par les différents logiciels (Rafiei and Sleno, 2015). Il apparaît de grandes différences dans la détection des pics avec à peine 10% des pics détectés dans les 4 logiciels testés. Une autre étude comparant XCMS et MZmine 2, un autre logiciel de prétraitement gratuit très utilisé, confirme ces divergences entre les deux logiciels sans pour autant pour conclure sur la supériorité de l'un par rapport à l'autre (Myers et al., 2017). Cela souligne l'importance de bien maîtriser l'algorithme de détection que l'on souhaite utiliser afin de pouvoir l'optimiser au mieux. Cette étape d'optimisation est indispensable pour pouvoir limiter le nombre de faux pics détectés et le nombre de vrais pics omis (Eliasson et al., 2012). Il existe maintenant des logiciels open source codés en R permettant d'optimiser automatiquement les différents paramètres des outils de détection automatique des signaux, notamment dans le cas d'XCMS (Libiseller et al., 2015).

Le résultat de cette étape d'obtention des variables est une matrice telle que présentée en Figure 4. Chaque ligne correspond à une variable définie par un couple m/z et temps de rétention. Pour chaque variable, différentes données sont présentées puis l'aire sous la courbe du signal correspondant à cette variable est indiquée pour chaque échantillon. C'est à partir de ces aires que sont réalisées toutes les étapes de validation de l'analyse et d'étude statistique.

Afin de réduire les temps de prétraitement des données, des solutions basées sur le *cloud* ont été développées. Le stockage et le traitement sur des serveurs à distance permet d'accroître les capacités de calculs par opposition à l'utilisation d'un ordinateur seul. Il existe notamment une version en ligne d'XCMS (XCMS Online) qui permet de traiter ses données sur des serveurs à distance (Gowda et al., 2014). Appuyé sur XCMS Online, XCMS Stream optimise encore le workflow de prétraitement en chargeant automatiquement les données au fur et à mesure de leur acquisition (Montenegro-Burke et al., 2017). Cela permet de commencer le prétraitement des données (conversion de format, détection de pics, ...) avant la fin de la séquence d'analyse et ainsi d'accéder plus rapidement aux résultats. Cela présente aussi l'avantage de pouvoir adapter la suite de l'acquisition des données en fonction des premiers résultats, ce qui peut s'avérer très intéressant, notamment lors d'analyse en mode données dépendant. Il est par exemple possible réaliser des acquisitions MS^2 dites *biology-dependant*. Lors de l'acquisition d'une cohorte comportant des témoins et des patients, le traitement en continu des données va permettre de déclencher l'enregistrement d'un spectre MS^2 correspondant à une variable dès qu'un seuil de significativité est atteint (Rinehart et al., 2014). Cela permet de focaliser l'acquisition de spectres MS^2 sur des éléments biologiquement pertinents.

3.4.3 Application de filtres analytiques

Une fois les données détectées et mises en forme, certaines variables sont éliminées car ne correspondant pas à de l'information pertinente. En effet, certaines proviennent de différentes sources de bruit au lieu de venir réellement de l'échantillon biologique ce qui risque de conduire à de fausses identifications. La majeure source de bruit est liée à des contaminations. Ces contaminants doivent être filtrés de la liste des variables à partir d'informations sur leur intensité en comparaison aux échantillons « blancs » ou à leur reproductibilité.

Pour ce faire, différents filtres sont appliqués. Un premier type de filtre est le calcul d'un ratio d'intensité entre les échantillons de QC et les échantillons « blancs ». Ce ratio est calculé pour chaque variable et celles pour lesquelles il est inférieur à une valeur seuil sont considérées comme provenant de contaminants (impuretés de solvant, relargage par les tubes, additifs, ...) (Dunn et al., 2017). Le seuil d'élimination des variables est ajusté à l'objectif de l'analyse (études non ciblées ou validation de biomarqueurs). Il varie donc selon les études entre 1.5 et 10. Un second type de filtre, basé sur les QC, vise à éliminer les variables pour lesquelles la variabilité analytique est trop importante et

risque de biaiser les comparaisons entre individus. Il est suggéré de ne conserver que les variables détectées sur au moins 50 % des injections de QC et qui présentent un coefficient de variation inférieur à un certain seuil en considérant l'ensemble des QC (Dunn et al., 2011). Ce seuil est généralement fixé à 20 % ou 30 % selon les techniques analytiques et les objectifs de l'étude. Un dernier filtre regroupant les deux aspects précédents est également couramment utilisé. Il s'agit d'injecter une gamme de dilution d'un QC. L'analyse de l'intensité de chaque variable sur cette gamme permet de ne conserver que les molécules dont les aires des pics chromatographiques corrélerent avec les facteurs de dilution (Eliasson et al., 2012). Un coefficient de corrélation est calculé et les variables présentant un coefficient inférieur à un seuil fixé sont éliminées. Ce seuil est en général fixé entre 0.5 et 0.8.

L'application de ces trois filtres est souvent redondante mais permet de limiter au maximum le risque de donner un sens biologique à une variable qui ne serait qu'artéfactuelle.

3.4.4 Normalisation des signaux

Les signaux conservés sont considérés comme pertinents et peuvent ensuite être corrigés afin de s'affranchir d'éventuelles dérives liées à la préparation d'échantillons. Les valeurs peuvent être normalisées pour pallier les différences dans la nature des échantillons ou dans la préparation de ces derniers.

Les outils de normalisation peuvent être de plusieurs types selon la nature de l'échantillon. Dans la majorité des cas, le plasma ne nécessite pas de normalisation. Il s'agit en effet d'un biofluide dont la composition est fortement régulée par l'organisme. De ce fait, l'utilisation d'un volume fixe de plasma entre les différents échantillons permet à elle seule une bonne comparaison des résultats. En revanche le cas de l'urine est plus complexe. En effet, de nombreux éléments vont pouvoir faire varier la concentration globale de l'urine, tels que le taux d'hydratation, le temps depuis la dernière miction ou l'état physiologique, notamment rénale, des individus. La normalisation de l'urine pour des analyses métabolomiques a été largement étudiée (Wu and Li, 2016). Il peut s'agir d'une normalisation pré-analyse ou d'une normalisation post-analyse. La normalisation pré-analytique consiste à diluer les échantillons spécifiquement en fonction de certains paramètres. L'élément le plus utilisé pour calculer le facteur de dilution spécifique est le taux de créatinine urinaire (Vogl et al., 2016). La créatinine présente l'avantage d'avoir un taux d'excrétion relativement stable chez les individus sains et d'être facilement mesurable. Cependant, chez les individus souffrant de différentes maladies ou présentant des conditions de vie particulières (exercice physique, ...), l'excrétion rénale de créatinine peut varier. D'autres approches peuvent être utilisées et consistent à normaliser par rapport à l'osmolarité de l'urine. Si l'osmolarité exacte n'est pas toujours disponible, une mesure de gravité spécifique ou de réfractométrie permet d'estimer cette osmolarité et conduire à une normalisation efficace (Edmands et al., 2014). Enfin, dans certaines études, le facteur de dilution pré-

analytique est calculé à partir du volume total d'urine produit en 24h (Zamora-Ros et al., 2011). Concernant la normalisation post-analytique, l'utilisation du TUS (*Total Usefull Signal*) est la technique la plus rependue. Elle consiste à normaliser l'aire de chaque signal en fonction de la somme des signaux pertinents enregistrés au sein de l'échantillon (Warrack et al., 2009). Dans le cas spécifique de l'analyse par introduction directe (FIA), une approche utilisant l'intensité de la suppression d'ionisation d'une molécule témoin (hexakis) comme marqueur de la concentration de l'urine a également été développée (Chen et al., 2015). Afin d'obtenir une normalisation optimale, ces différentes approches peuvent être combinées, ce qui semble conduire à de meilleurs résultats (Gagnebin et al., 2017).

Enfin, dans le cas d'échantillons solides tels que des fèces ou des tissus, il est parfois compliqué d'extraire exactement la même quantité de matériel initial. De ce fait, une normalisation est en général réalisée en rapportant les quantités de métabolites mesurées au poids de tissu ou de fèces extrait. Le cas des fèces chez les rongeurs caecotrophes (lapin, rat, souris, ...) est toutefois plus difficile à gérer car ces animaux produisent deux types des fèces dont les compositions sont relativement différentes, notamment en termes de teneur en eau ou en acides aminées (Sukemori et al., 2003). Il faudra donc faire attention de ne pas comparer des caecotrophes (crottes molles) chez certains animaux et des fèces vrai (crottes dures) chez d'autres.

3.4.5 Annotation du jeu de données

Une fois la liste de variables nettoyée pour ne conserver que les variables analytiquement pertinentes, ces variables doivent être annotées pour pouvoir conduire ensuite à une interprétation biologique des données. Cette étape consiste à attribuer un nom de métabolite aux variables détectées. Elle est cruciale au cours du processus métabolomique et reste aujourd'hui une des principales étapes limitantes lors d'une analyse par LC-HRMS (Gika et al., 2014). Ces limites sont dues à notre méconnaissance de l'intégralité du métabolome (Viant et al., 2017). Contrairement à ce qui est le cas en protéomique, il n'existe pas, en métabolomique, de base de données exhaustive des métabolites potentiellement retrouvés dans un échantillon. Différentes informations doivent donc être analysées pour chaque variable afin d'essayer de lui attribuer un nom de métabolite.

Pour homogénéiser la présentation des résultats entre les différentes études menées en métabolomique, un groupe de travail a défini en 2007 les différents niveaux de confiance dans l'annotation des métabolites et les critères correspondants à ces niveaux (Sumner et al., 2007). Des travaux sont actuellement en cours au sein de la *Metabolomics Society* pour faire évoluer ces niveaux (Creek et al., 2014). Certains se sont basés sur les niveaux existants pour les faire évoluer afin de correspondre au mieux aux cas pratiques rencontrés (Schymanski et al., 2014). D'autres travaux ont cherché à raffiner le système de score pour les composés putativement annotés se basant sur des éléments chromatographiques et des éléments de spectrométrie de masse tels que les spectres de

fragmentation ou le massif isotopique (Rochat, 2017). Cependant, ces approches sont compliquées et non consensuelles et les niveaux définis par Sumner *et al.* restent pour l'instant l'outil de référence. Ces niveaux sont au nombre de 4.

Le niveau 1 correspond aux composés identifiés. Pour qu'un composé puisse être considéré comme identifié, il faut qu'il soit validé par *a minima* deux critères indépendants et orthogonaux. Il peut par exemple s'agir du temps de rétention et de la masse précise ou de la masse précise et du spectre de fragmentation. Pour que ces critères soient validés, la comparaison doit être faite par rapport à un composé de référence analysé dans les mêmes conditions expérimentales. L'utilisation de valeurs rapportées dans la littérature à partir de standards analysés dans d'autres laboratoires est insuffisante pour considérer un composé comme identifié. Le *gold standard* est la comparaison de spectres de fragmentation.

Le niveau 2 correspond aux composés putativement annotés. Il s'agit en général de composés pour lesquels une annotation est très probable, mais pour laquelle le standard n'est pas disponible commercialement ou de composés qui sont annotés à partir de bases de données externes au laboratoire menant l'étude.

Le niveau 3 correspond aux composés dont l'annotation n'est pas complète, et/ou pour lesquels plusieurs annotations sont proposées, mais pour lesquels les informations disponibles permettent de les classer au sein d'une famille chimique.

Le niveau 4 correspond aux composés inconnus. Il s'agit des variables détectées pour lesquels les informations disponibles ne permettent pas de tirer de conclusions claires quant à la nature du composé.

Différentes approches sont mises en place pour annoter les variables. La première consiste à interroger les bases de données internes pour savoir si la variable correspond à un composé présent dans ces bases. Pour faire cette interrogation, deux critères principaux sont utilisés lors des analyses par LC-HRMS : le temps de rétention et la masse précise. Dans une seconde approche, les variables ne correspondant pas à un composé présent en base de données interne seront étudiés spécifiquement. Ces différentes approches ainsi que les différents éléments permettant d'identifier les métabolites détectés seront présentés en partie 5.

3.4.6 Analyses statistiques

L'étape suivante au cours d'une analyse métabolomique est l'utilisation des statistiques pour faire ressortir les métabolites pouvant être marqueurs d'un statut physiologique. Deux grandes catégories d'analyses statistiques peuvent être utilisées et sont présentées ci-après.

3.4.6.1.1 Outils statistiques univariés

Ces outils permettent de comparer les différents échantillons métabolites par métabolites. Il s'agit en général de test de comparaison des moyennes. Il peut s'agir de tests paramétriques (T-test, Weslch-Test, ANOVA, ...) ou non paramétriques (Mann-Withney-Wilcoxon, Kruskal-Wallis, ...). Le choix entre tests paramétriques et non paramétriques se fait en fonction de la taille de la population étudiée et de la tendance de la distribution du métabolite étudié au sein des différents groupes à se rapprocher d'une loi normale. La Figure 5 présente un arbre décisionnel pour le choix du test statistique univarié. Il est important de noter que la présence d'échantillons aberrants (*Outliers*) peut impacter de façon important le résultat des tests. En cas de présence d'un ou plusieurs échantillons aberrants, les tests non paramétriques sont à privilégier.

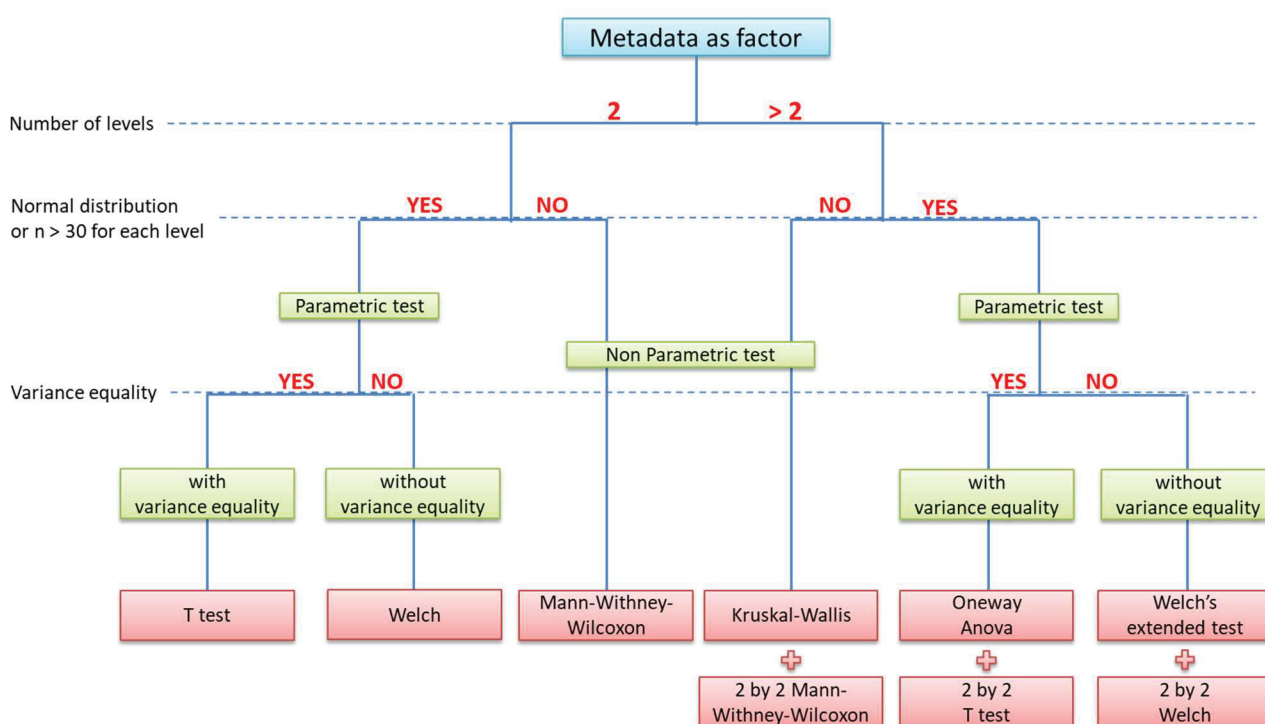


Figure 5 : Arbre décisionnel dans le choix des tests univariés.

Ces tests étant faits métabolite par métabolite, un grand nombre de tests doit être réalisé si l'on veut étudier un jeu de données standard en métabolomique non ciblées (classiquement entre quelques centaines et plusieurs milliers de variables détectées). La réalisation de ces tests multiples va augmenter fortement le risque d'obtention de faux positifs. Il est donc recommandé de corriger les p-Value obtenue à l'issue de ces tests multiples. Différentes procédures de corrections peuvent être utilisées. Actuellement, les corrections de Benjamini-Hochberg ou de Benjamini-Yekutiely semblent être les plus adaptées aux données métabolomiques (Vinaixa et al., 2012).

3.4.6.1.2 Outils statistiques multivariés

Une autre approche pour l'interprétation statistique est l'utilisation d'outils d'analyse multivariée. Ces outils vont traiter l'ensemble du jeu de données comme un seul élément. Cela permet une visualisation plus immédiate de la dispersion des échantillons. Ces approches sont basées sur l'utilisation de variables latentes qui sont calculées comme des combinaisons linéaires des variables initiales et permettent de mettre en avant des interactions existant entre ces différentes variables. Il existe deux types d'analyses statistiques multivariées. Les analyses statistiques multivariées non supervisées vont chercher à faire ressortir la variance qui existe dans le jeu de données sans *a priori* sur la nature de chaque échantillon. C'est par exemple le cas de l'analyse en composante principale. Les méthodes d'analyses statistiques multivariées supervisées vont chercher à mettre en évidence une hiérarchisation du jeu de données à partir de groupes d'appartenance prédéfinis pour chaque échantillon (Worley and Powers, 2013). Elles permettent de faire ressortir les métabolites qui vont avoir le plus de poids dans la séparation des échantillons en fonction de leurs groupes. Il existe différents types d'analyses statistiques multivariées supervisées mais les deux les plus utilisés sont la PLS-DA (*Partial Least Square Discriminant Analysis*) (Trygg et al., 2007) et l'OPLS-DA (*Orthogonal Projection to Latent Structure Discriminant Analysis*) (Bylesjo et al., 2006). L'objectif de l'OPLS-DA par rapport à la PLS-DA est d'isoler la variabilité du jeu de données qui est liée à l'élément de réponse de celle qui est liée à d'autres sources de variation. Elle permet en général une meilleure séparation entre les groupes d'échantillons.

Une fois identifiés, les métabolites permettant de discriminer les différents groupes vont pouvoir être interprétés biologiquement afin de chercher une cohérence entre ces biomarqueurs.

3.4.7 Interprétation biologique

L'interprétation biologique des résultats se base principalement sur la visualisation des différentes voies métaboliques connues et sur le positionnement des métabolites identifiés sur ces voies. Ces voies sont en général focalisées sur un nombre restreint de substrats initiaux ou produits finaux. Chaque voie va alors représenter les différentes étapes de biosynthèse ou de biodégradation de ce petit nombre de métabolites en les connectant par l'intermédiaire des enzymes capables de catalyser les réactions (Figure 6). Parmi les différentes bases de données de voies métaboliques, la plus utilisée est KEGG (Ogata et al., 1999) mais on peut également citer MetaCyc (Caspi et al., 2012) ou Reactome (Joshi-Tope et al., 2005). Le positionnement des métabolites par voies permet de les regrouper et d'analyser leurs évolutions conjointes au sein de ces voies.

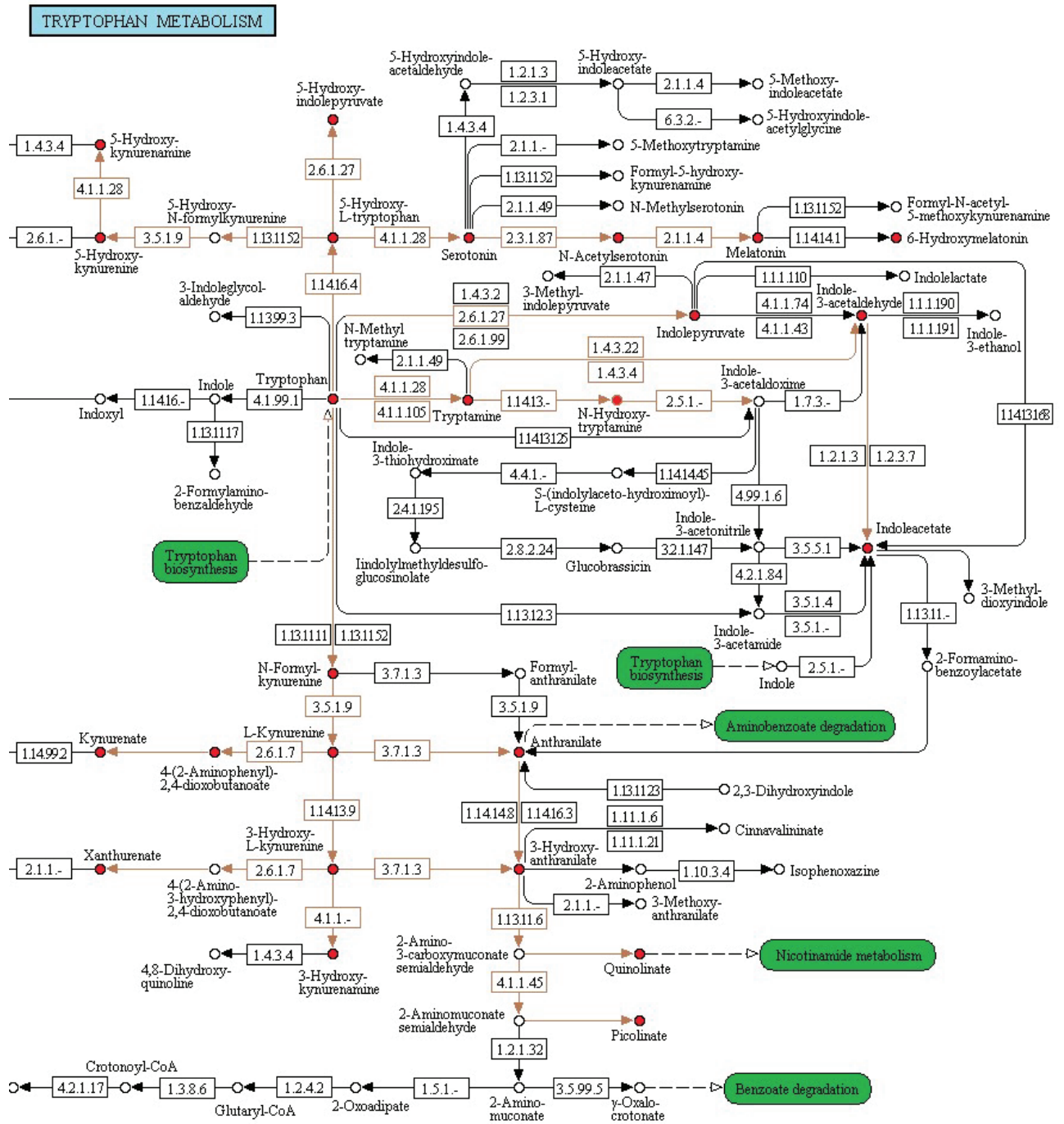


Figure 6 : Extrait de la voie métabolique du tryptophane publiée dans KEGG (Ogata et al., 1999)

Cependant, ces voies sont souvent traitées indépendamment les unes des autres. Cette approche ne tient donc pas compte des interactions qui peuvent exister entre elles ou de leur régulation par des éléments provenant d'autres voies. Les voies métaboliques étant globalement interconnectées, il en découle une perte d'information importante lors de l'interprétation. De nouvelles approches permettent aujourd'hui d'intégrer l'ensemble des données obtenues au sein du

réseau métabolique « global » de l'organisme étudié. Cela permet d'étudier les interactions qui existent entre les différentes voies et permet de faciliter la découverte de nouvelles voies. Cela est d'autant plus important qu'une partie importante du métabolisme est encore mal connue ou inconnue. Parmi les différents outils, la plateforme MetExplore (Cottret et al., 2010) ou le module Metscape de Cytoscape (Gao et al., 2010) permettent d'intégrer l'ensemble de ces informations pour obtenir une compréhension plus approfondie des phénomènes biologiques mis en jeu. Malgré le développement de ces outils, l'interprétation biologique repose encore beaucoup sur le savoir d'experts du métabolisme et fait encore aujourd'hui partie des étapes limitantes du processus d'analyse métabolomique.

Après avoir parcouru l'ensemble du processus mis en place lors d'une analyse métabolomique par LC-MS, nous allons détailler deux des étapes clés de ce processus : l'acquisition des données et l'identification de métabolites.

4 Les modes d'acquisition de données en spectrométrie de masse

4.1 L'acquisition par balayage MS

Jusqu'à maintenant, les protocoles d'analyse métabolomique non ciblée reposent essentiellement sur l'acquisition de profils par spectrométrie de masse à haute résolution couplée à la chromatographie liquide. L'obtention de ces données se fait par l'enregistrement simultané de toutes les masses présentes sur la gamme de masse choisie.

Ces données permettent de détecter et de quantifier (au moins de façon relative) les ions présents dans les différents échantillons. C'est grâce aux données acquises par cette méthode que les signaux permettant de discriminer différents groupes d'échantillons sont repérés.

4.2 L'acquisition de données MS²

L'acquisition de données de fragmentation en spectrométrie de masse en tandem va servir à identifier les différents signaux lors de l'étape d'annotation (voir partie 5). Ces données de fragmentation peuvent être de différents types et être acquises de différentes manières.

4.2.1 Les différents modes de fragmentation

En spectrométrie de masse utilisant des sources d'ionisation à pression atmosphérique, deux principaux modes de fragmentation peuvent être utilisés (Bennaceur et al., 2013). La fragmentation non résonante est actuellement la plus représentée sur les instruments à haute résolution

commercialisés. La fragmentation résonante n'est disponible que sur les instruments à piégeage d'ions, mais peut apporter des informations complémentaires.

Le processus de fragmentation se décompose en trois étapes. La première est la sélection du ou des ions à fragmenter. Cette sélection se fait par l'intermédiaire d'analyseurs basse résolution. Il peut s'agir soit d'un quadripôle soit d'un piège à ions linéaire. Les voltages appliqués à ces analyseurs sont ajustés pour ne conserver que les ions présents dans une fenêtre de masse définie. La seconde est l'activation des ions. Cette activation augmente l'énergie interne de l'ion. Il s'agit d'activation à basse énergie induite par collision, ce qui génère une excitation de nature principalement vibrationnelle (Horning et al., 1990). De ce fait, les mécanismes de fragmentation sont liés à la présence d'une charge sur les ions (Wysocki et al., 1988). La dernière étape est la décomposition de l'ion dit « parent » ou ion « précurseur » en divers ions « fils ».

Il existe deux modes d'activation des ions à basse énergie. La fragmentation résonante est une fragmentation lente par piégeage qui a donc lieu au sein d'un piège à ion (March et al., 1991). Elle sera appelée CID (*Collision Induced Dissociation*) dans la suite de ce manuscrit. Une radiofréquence correspondant à la fréquence de résonance de l'ion précurseur est appliquée au niveau des électrodes du piège. Cela conduit à exciter spécifiquement l'ion précurseur qui va alors acquérir de l'énergie cinétique. Cette énergie cinétique est convertie en énergie interne par collision avec des molécules de gaz neutre présentes dans le piège. Lorsqu'il accumule une énergie interne suffisante, l'ion précurseur va se dissocier pour former des ions fragments de masses différentes. La fréquence d'activation des ions étant liée à leur masse, les ions fragments ne seront pas activés à leur tour par la radiofréquence appliquée. En conséquence, seuls des ions de première génération seront observés. On parlera de fragmentation compétitive (Steiner et al., 1999). Les spectres MS² issus de ce type de fragmentation sont donc généralement constitués d'un nombre limité d'ions mais tous les ions de première génération sont observés.

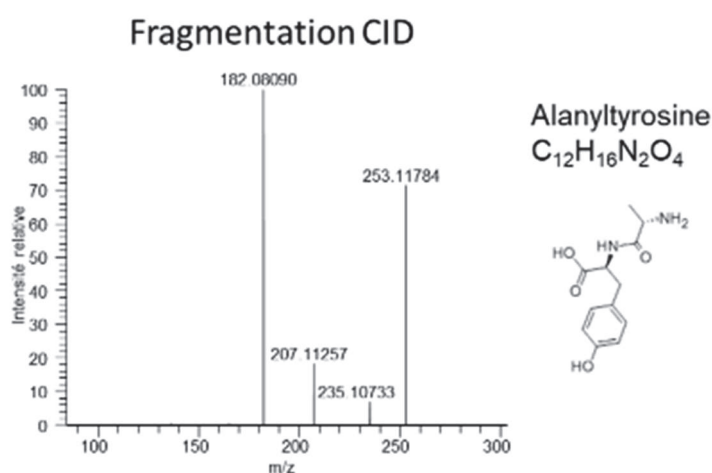


Figure 7 : Exemple d'un spectre de fragmentation résonante (CID) de l'alanyltyrosine

La fragmentation non résonante est une fragmentation en faisceau. Elle a lieu au niveau d'une cellule de collision (quadripôle, octopôle). Bien qu'il s'agisse toujours d'une dissociation induite par collision et qu'elle soit présente dans de nombreux types d'instruments, elle sera appelée HCD (*Higher-energy Collision Dissociation*) dans la suite de ce manuscrit, en référence à la nomenclature utilisée sur les instruments de type Orbitrap (Olsen et al., 2007). Le faisceau d'ions précurseurs accélérés à basse énergie est introduit dans la cellule de collision qui contient un gaz inerte à une pression suffisante pour que des collisions avec les ions la traversant se produisent. Les ions précurseurs entrant en collision avec le gaz accumulent de l'énergie interne jusqu'à se fragmenter. Les ions formés continuent leur parcours dans la cellule de collision, entrent en collision avec les molécules de gaz inerte et se fragmentent à leur tour. Des fragmentations consécutives vont donc avoir lieu. Les spectres MS² obtenus font donc apparaître les ions de plusieurs générations. Ils peuvent donc paraître plus informatifs mais l'absence de certains fragments de première génération, qui auront été refragmentés, peuvent être une gêne pour l'interprétation des mécanismes de fragmentation (Jedrychowski et al., 2011).

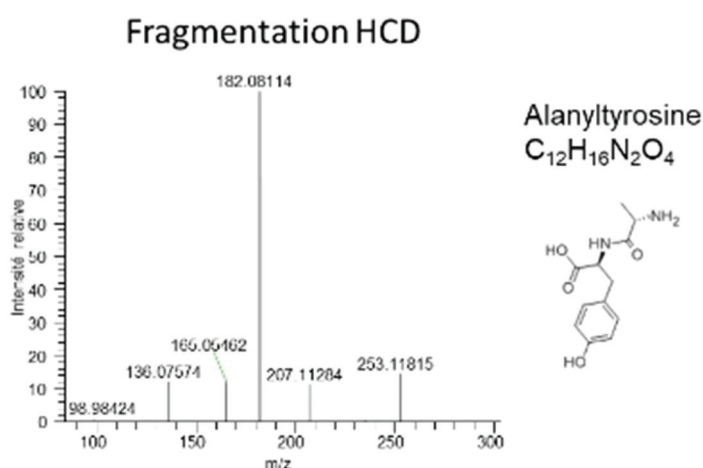


Figure 8 : Exemple d'un spectre de fragmentation non résonante (HCD) de l'alanyltirosine

Ce type de fragmentation non résonante est celui qui est le plus utilisé dans le cadre des analyses métabolomiques. Cela est d'abord dû au fait que la majorité des instruments utilisés pour mener des analyses en métabolomique sont équipés de cellules de collisions non résonantes. De plus, malgré le risque de disparition de certains ions fils, le plus grand nombre de fragments générés rend la signature plus spécifique de l'ion parent. Les gammes d'énergies permettant à la fois l'observation d'ions fragments et d'ion précurseurs sont nettement plus larges avec l'activation non résonante. Pour l'analyse de composés non identifiés et dont les énergies de fragmentation adaptées ne sont pas connues, cela s'avère être un avantage indéniable. Néanmoins, les deux modes d'activation vont apporter des informations complémentaires pour l'identification des molécules en métabolomique (Mullard et al., 2015).

4.2.2 Les différents modes d'acquisition MS²

L'acquisition de spectres de fragmentation est essentielle à l'identification des composés en métabolomique. Cette acquisition peut se faire selon différents modes (Figure 9).

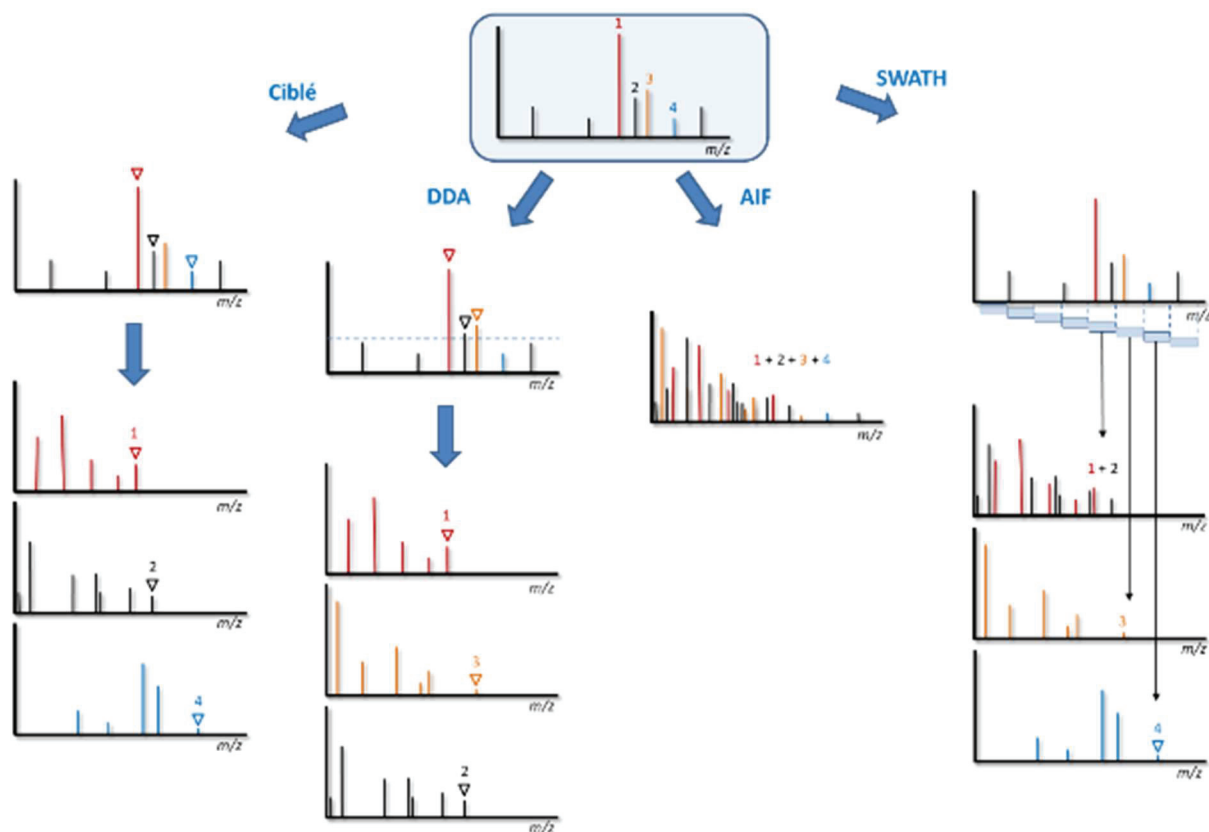


Figure 9 : Les différents modes d'acquisition MS²

4.2.2.1.1 Le mode ciblé

Le mode ciblé est actuellement encore le plus utilisé lors des analyses métabolomiques. Dans ce mode, les informations sur la masse et le temps de rétention sont renseignées et les ions ciblés sont spécifiquement isolés et fragmentés en fonction de ces informations. Cela signifie que les données acquises en mode profil MS ont été préalablement traitées et que les ions à fragmenter ont été identifiés à partir de ces données. Du fait de la grande quantité d'informations générées lors de l'analyse de profils MS en métabolomique, le traitement des données est une étape relativement longue. L'acquisition des spectres MS² doit donc se faire dans un second temps, parfois distant de plusieurs semaines de l'acquisition des profils MS. Cela implique de réinjecter les échantillons (souvent un QC est utilisé) une seconde fois, ce qui est à la fois consommateur d'échantillons et de temps. De plus, de nombreux biais peuvent apparaître ce qui risque de nuire à la pertinence et à la qualité des spectres MS² acquis. Tout d'abord, l'échantillon peut s'être modifié dans le laps de temps entre les deux injections. Même s'il est conservé à -80°C, il aura nécessairement subi un cycle de

recongélation-décongélation qui peut altérer sa teneur en différents métabolites. Si les métabolites visés lors des analyses MS² sont relativement instables, il se peut qu'ils ne soient plus observés lors de cette seconde injection. De plus, les conditions analytiques peuvent avoir évolué. La colonne chromatographique initialement utilisée peut avoir été impactée par l'analyse d'autres échantillons et les phases mobiles devront être réparées pour cette seconde injection. Il se peut donc que les performances du système chromatographique soient modifiées entre les deux analyses et que les temps de rétention de la seconde analyse ne correspondent plus exactement à ceux observés lors de la première. Enfin, l'efficacité de l'ionisation des composés en spectrométrie de masse est un élément manquant de robustesse ce qui risque de modifier les intensités des ions ciblés et de rendre l'acquisition d'un spectre MS² plus ardue.

Pour toutes ces raisons et grâce aux améliorations récemment apportées aux spectromètres, notamment au niveau du couple vitesse d'acquisition - résolution, des nouveaux modes d'acquisition commencent à être appliqués en métabolomique, à l'instar de ce qui a été mis en place en protéomique.

4.2.2.1.2 Le mode données dépendant

Le mode « données dépendant » (DDA, *Data-Dependent Acquisition*) va permettre l'acquisition de spectres MS et de spectres MS² sur une seule injection. L'instrument va enchaîner des cycles comprenant l'acquisition d'un spectre MS puis la sélection automatique d'un certain nombre d'ions qu'il va fragmenter pour acquérir les spectres MS² correspondants. Le nombre d'événements MS² réalisés dans chaque cycle est défini par l'utilisateur en lien avec le système chromatographique utilisé. En effet, si ce nombre est trop faible, peu de spectres MS² sont acquis et s'il est trop important, le temps entre chaque événement MS sera trop important pour permettre une définition correcte des pics chromatographiques et une quantification correcte. La sélection automatique des ions se fait selon un ordre établi en fonction de critères définis par l'utilisateur. Le premier est en général l'intensité : les ions seront sélectionnés par ordre décroissant d'intensité. Ensuite, d'autres éléments de décision peuvent être ajoutés comme par exemple des listes excluant certains ions contaminants ou une exclusion dynamique qui permet d'exclure les ions qui ont déjà été fragmentés. Cela permet de limiter le nombre d'acquisition de spectres non pertinents ou redondants (Neumann et al., 2013).

Ce type d'application commence à se développer en métabolomique et est surtout utilisé sur des instruments à haute vitesse d'acquisition tels que des instruments à temps de vol (Benton et al., 2015; Chen et al., 2011). Cependant certaines applications impliquent des Orbitrap utilisés à une résolution sous maximale (Mullard et al., 2015).

L'avantage de cette technique est qu'elle permet d'obtenir des spectres MS² relativement purs car les ions sont sélectionnés individuellement (sauf en présence de composés isobares). L'interprétation des spectres MS² obtenus est donc facilitée. Il existe cependant un risque que certains ions d'intérêt ne soient pas fragmentés (Hopfgartner et al., 2012). En effet, malgré l'application de différents critères permettant d'éliminer un certain nombre d'ions, les ions d'intérêt de faible intensité ont peu de chance d'être sélectionnés.

4.2.2.1.3 Le mode données indépendant

Pour pallier le risque de non-sélection de composés d'intérêt, un autre mode a été développé. Le mode « données indépendant » (DIA, *Data-Independent Acquisition*) consiste à fragmenter tous les ions sans critères de sélection. Deux approches se sont développées.

La première dite « All Ion Fragmentation » (AIF) consiste à fragmenter en une fois l'ensemble des ions présents sur le spectre. Pour les applications métabolomiques et face à la diversité structurale des molécules d'intérêt, deux énergies de fragmentation différentes sont alternées (Plumb et al., 2006; Wrona et al., 2005). L'objectif de la basse énergie est d'obtenir des informations sur les ions parents et celui de l'énergie haute est de recueillir des informations les ions fragments et les pertes de neutres associées. Le principal inconvénient de cette technique est la complexité des spectres MS² obtenus. De plus, il est difficile de pouvoir faire un lien direct entre les ions précurseurs et les différents fragments associés. Cela impose l'utilisation de logiciels de traitement de données très performants qui sont à l'heure actuelle peu adaptés aux problématiques de la métabolomique.

Une autre approche, dite « SWATH » (*Sequential Window acquisition of All Theoretical mass spectra*), consiste à découper la gamme de masse d'intérêt en plusieurs fenêtres de masse de petite taille (Gillet et al., 2012). Cela permet de réduire le nombre d'ions fragmentés simultanément et de faciliter l'interprétation des spectres. La taille des fenêtres de masse et donc le nombre de spectres MS² nécessaires pour couvrir toute la gamme de masse sont adaptés à la vitesse d'acquisition de l'instrument. Elle peut varier d'une vingtaine de dalton sur des instruments très rapides de type temps de vol (Arnhard et al., 2015) à une cinquantaine de Dalton pour des applications sur des Orbitrap (Zhou et al., 2017). Si le principe du SWATH, tel que développé pour la protéomique, consiste en l'utilisation de N fenêtres de masse de taille identique, cela peut être amélioré pour les applications métabolomiques. En effet, la distribution des principaux métabolites connus n'est pas homogène sur la gamme de masse. Ainsi, la taille des fenêtres peut être ajustée de façon à utiliser des fenêtres de plus petite largeur sur les gammes de masse où de nombreux métabolites sont attendus et des fenêtres de plus grande largeur là où peu de métabolites sont attendus (Zhang et al., 2015).

Cette technique est intermédiaire entre le mode All Ion Fragmentation et le mode données dépendant. Elle permet à la fois de fragmenter l'ensemble des ions présents et de diminuer le nombre d'ions fragmentés simultanément. Pour ces raisons, l'acquisition de données en mode SWATH est de plus en plus utilisée.

4.2.3 Choix du type d'analyseurs pour l'acquisition multi-événements de données

Au début des années 2000, les instruments à temps de vol ont été les premiers utilisés pour les analyses métabolomiques par couplage LC-MS. En effet, il s'agissait alors des seuls instruments à haute résolution ayant une vitesse d'acquisition suffisante pour permettre ce couplage. Avec l'apparition dans le milieu des années 2000 des premiers Orbitrap et grâce à la résolution nettement supérieure qu'ils peuvent atteindre, de nombreuses équipes se sont tournées vers cette technologie. Au cours des dernières années, de nombreuses améliorations ont été apportées aux instruments à temps de vol (augmentation de la résolution, amélioration de la gamme dynamique, introduction de la mobilité ionique). Avec le développement des approches multi-événements de type acquisition données dépendante ou indépendante, la vitesse d'acquisition des instruments à temps de vol couplée aux améliorations apportées, devient un argument important en leur faveur.

Aujourd'hui, deux grandes stratégies d'acquisition en métabolomique non ciblée par LC-MS se dégagent. Il s'agit soit d'utiliser la très haute résolution atteignable sur les instruments Orbitrap pour détecter et quantifier les métabolites grâce à la séparation des composés isobares et à la précision de mesure de masse et d'acquérir les informations MS² en mode données dépendant pour obtenir les informations structurales. La seconde approche consiste à exploiter la haute vitesse d'acquisition des instruments à temps de vol pour simultanément détecter, quantifier et obtenir les informations structurales grâce à des analyses en mode données indépendant. Comme expliqué ci-dessus, ces deux stratégies se développent de plus en plus mais restent relativement nouvelles dans le monde de la métabolomique. L'acquisition des données MS en mode profil puis l'obtention des informations structurales en mode ciblé représente encore aujourd'hui la majeure partie des analyses de métabolomique par LC-MS.

Au cours de ces travaux de thèse, nous avons développé de nouveaux protocoles d'acquisition DDA et DIA sur l'Orbitrap Fusion.

5 L'identification des métabolites

L'identification des métabolites est une étape clé pour pouvoir interpréter biologiquement les résultats obtenus. Lors d'études métabolomiques non ciblées, deux cas peuvent se présenter pour identifier un composé inconnu. Dans le premier cas, le composé inconnu correspond à un composé déjà présent en base de données interne et pour lequel les éléments d'identification tels que la masse, le temps de rétention et la fragmentation MS^2 sont déjà connus. Dans le second cas, le composé inconnu ne correspond à aucun élément présent en base de données interne. Les différents éléments d'identification (temps de rétention, spectre MS^2 , ...) devront être acquis et interprétés de façon à émettre une hypothèse d'annotation. Cette annotation de niveau 2 pourra être confirmée suite à l'achat ou la synthèse du composé cible et à son analyse dans des conditions identiques. L'ensemble de ce processus décisionnel est présenté en Figure 10 et détaillé par la suite.

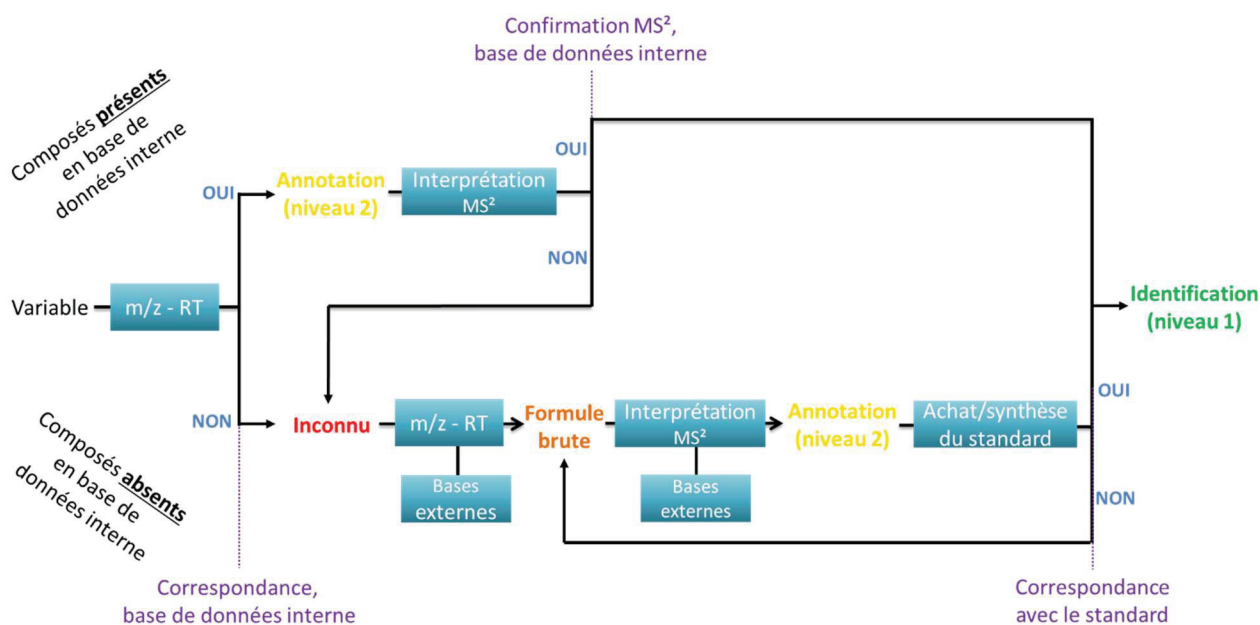


Figure 10 : Arbre décisionnel du processus d'annotation

5.1 Identification des composés grâce aux bases de données internes

5.1.1 Utilisation de la masse précise et du temps de rétention

L'interrogation des bases de données internes se fait à partir du couple masse précise / temps de rétention pour chaque variable détectée. Ce couple de valeurs est comparé à ceux enregistrés pour les standards de référence. L'efficacité de cette méthode d'annotation dépend de différents facteurs : le nombre de molécules dans la base de données, la construction de cette base et les tolérances acceptées en masse et en temps de rétention.

Plus la taille de la base de données est importante, plus le nombre de variables identifiées pourra être important. Cela augmente la possibilité de découvertes de potentiels biomarqueurs. En contrepartie, cela augmente également la possibilité de doubles annotations de certaines variables si la tolérance en masse ou en temps de rétention est trop importante. Cela est d'autant plus vrai pour les composés isomères qui peuvent avoir des temps de rétention très proches. Si une seule variable (correspondant à un des isomères d'un couple d'isomères) est détectée, que la base de données est de petite taille et qu'elle ne contient qu'un seul isomère, la variable sera identifiée comme étant l'isomère présent dans la base. Si le même cas se présente mais que la base de données est de plus grande taille et qu'elle contient les deux isomères, la variable sera doublement annotée et elle ne pourra donc pas être identifiée. Plus la base de données sera développée moins le risque de fausse identification sera élevé, mais le nombre total d'identifications confirmées ne sera pas proportionnel à la taille de la base du fait des doubles annotations générées.

Concernant la construction de la base, deux possibilités existent. La première est une approche ciblée qui vise uniquement à savoir si les molécules présentes dans la base de données sont détectées dans l'échantillon. Dans ce cas, la base de données n'est constituée que par la forme majoritaire attendue de la molécule (souvent l'ion pseudomoléculaire). La seconde approche correspond plus au principe de la métabolomique non ciblée et vise à identifier le maximum de variables détectés. Or, un métabolite donné peut être détecté sous différents types : l'ion pseudomoléculaire, des ions adduits ou des ions fragments si de la fragmentation en source est observée. De plus, pour chacun de ces types, différentes formes isotopiques peuvent être détectés : l'ion monoisotopique qui ne contient que la forme majoritaire de chaque élément, l'ion contenant un atome de carbone 13, l'ion contenant deux atomes de carbone 13, l'ion contenant un atome de soufre 34 (si la molécule contient du soufre) ... Il est alors nécessaire de construire une base de données exhaustive contenant les masses de toutes les formes de chaque métabolite présent dans la base.

Le dernier point est le choix de la tolérance dans la correspondance entre les valeurs de masse et de temps de rétention enregistrées dans la base et les valeurs mesurées dans l'échantillon. Plus cette tolérance sera importante plus le risque de fausses identifications sera élevé. A l'inverse, des critères trop stricts conduiront à la non-identification d'un nombre important de métabolites. La tolérance en masse est à fixer en fonction des performances des instruments. Sur les instruments à haute résolution, elle est en général fixée entre 1 et 5 ppm. Pour les analyseurs à basse résolution, une erreur de quelques centaines de Da est usuellement utilisée. La tolérance en temps de rétention est à fixer en fonction du système chromatographique utilisé. Les systèmes chromatographiques C₁₈ sont plus robustes que les systèmes HILIC. La tolérance sera donc plus faible en C₁₈ qu'en HILIC. De plus, soit la tolérance est fixe soit elle est modulée en fonction de temps de rétention. Plus le composé

sera attendu tard dans le gradient, plus la tolérance sera grande. En pratique, elle varie entre quelques secondes et quelques dizaines de secondes.

Tous ces paramètres permettent d'identifier de manière optimale les métabolites présents dans l'échantillon. Afin de s'assurer de l'absence de fausses identifications, des expériences MS² doivent être réalisées pour vérifier la concordance entre le spectre de fragmentation du métabolite étudié et celui du standard. Ces expériences MS² servent aussi à choisir l'identification correcte en cas d'annotations multiples.

5.1.2 Confirmation par les spectres MS²

L'étude de spectres MS² est un élément important pour confirmer les annotations en métabolomique par LC-HRMS. Les spectres MS² sont en général acquis en mode de fragmentation non résonnante. Comme détaillé précédemment, ce mode permet d'obtenir un nombre important de fragments différents, ce qui augmente la spécificité de la comparaison des spectres. De plus, ce mode est le plus répandu sur les différents instruments commerciaux, notamment de type Orbitrap. L'un des paramètres à déterminer est l'énergie de collision optimale pour chaque composé. Cette énergie est celle permettant une observation concomitante de l'ion parent (idéalement entre 10 et 30 % d'intensité relative) et des ions fragments. Une fois cette énergie déterminée pour chaque standard, les métabolites d'intérêt vont être fragmentés à une énergie proche de celle de référence puis les spectres vont être comparés pour valider ou invalider l'annotation.

A l'heure actuelle, la comparaison des spectres MS² est encore souvent faite manuellement par comparaison visuelle des spectres (masse des fragments, intensités relatives, ...). Des outils automatisés commencent à voir le jour pour évaluer la qualité des spectres MS² enregistré et pour les comparer (Kind et al., 2018; Mylonas et al., 2009). Ces outils sont appelés à s'étendre afin d'automatiser cette étape de validation d'identification par fragmentation.

Pour l'identification des composés à partir de bases de données internes, le niveau d'annotation seul n'est globalement pas une information suffisante pour déterminer la fiabilité d'une identification. Il est donc important de préciser la taille et la constitution de la base de données ainsi que les critères qui ont été utilisés pour réaliser ces annotations, notamment la confirmation ou non par fragmentation.

5.2 Identification des composés absents des bases de données internes

Si l'annotation par comparaison à un standard de référence permet d'identifier de façon fiable un certain nombre de métabolites, la majorité des composés détectés restera cependant non identifiée. Cela est d'une part dû à une connaissance encore incomplète du métabolome et d'autre part à la non disponibilité commerciale de tous les métabolites ce qui empêche de disposer de

standards à inclure dans les bases internes. Pour tous ces composés, une autre approche devra donc être utilisée pour essayer de les identifier.

5.2.1 Détermination de la formule brute

En métabolomique par LC-HRMS, la première étape dans l'annotation de composés inconnus est la détermination de la formule brute du dit composé. Cette détermination de la formule brute se base sur différentes informations.

5.2.1.1.1 Apport de la masse précise

La masse précise mesurée par les instruments à haute résolution apporte des informations essentielles pour déterminer une composition élémentaire. La masse exacte de chaque atome étant connue (M. Wang et al., 2017), la mesure de la masse exacte de l'ion d'intérêt doit suffire pour déterminer la formule brute. Les instruments modernes les plus performants peuvent atteindre des précisions de mesure de masse inférieure à 0.2 ppm (Hendrickson et al., 2015). En appliquant cette tolérance, cela permet d'aboutir en général à une unique formule brute pour les faibles masses (< 300 Da). Pour les masses plus importantes, même des précisions de mesure de masse inférieures à 0.2 ppm, peuvent correspondre à plusieurs formules brutes (Tableau 3).

Métabolite	Masse exacte	0.2 ppm	1 ppm	3 ppm	5 ppm	10 ppm	25 ppm
Hexanal	100.08881	1	1	1	1	1	2
Acide Dodécanoïque	200.1763	1	1	1	1	2	3
Acide Rétinoïque	300.20893	1	1	3	5	10	24
UDP	404.00219	4	20	56	93	188	> 400
ATP	506.99575	5	24	82	137	273	> 400

Tableau 3 : Nombre de formules brutes compatibles avec la masse de certains métabolites en fonction de la tolérance en masse acceptée. Les formules brutes potentielles sont de la forme $C_{0-50}H_{0-75}N_{0-10}O_{0-15}S_{0-3}P_{0-3}$

En pratique, lors de l'utilisation des instruments de type Orbitrap en métabolomique par LC-HRMS, les erreurs de masses moyennes sont de l'ordre du ppm. Des tolérances de 3 ppm (ou 5 ppm sur les anciennes générations d'instruments) sont donc utilisées. Pour ces tolérances, des règles complémentaires doivent donc être appliquées pour réduire au maximum le nombre de formules brutes potentielles.

5.2.1.1.2 Apport des règles chimiques

Ces règles sont basées à la fois sur des règles chimiques et sur des observations empiriques. Elles ont été présentées en 2007 et permettent d'éliminer des formules brutes proposées sur la base de la masse précise mais non réalistes chimiquement (Kind and Fiehn, 2007). Les points essentiels de ces règles sont les règles chimiques de Lewis et Senior et une limitation des ratios entre certains éléments chimiques. La règle de Lewis (ou règle de l'octet) énonce que les principaux atomes (carbone, oxygène, azote) doivent partager leurs électrons de façon à remplir leur couche électronique externe. Cependant, cette règle peut être trop restrictive pour certains composés et ne doit donc pas être utilisée si des composés radicalaires sont attendus. Le théorème de Senior permet d'élargir la règle de Lewis, notamment pour être applicable aux éléments hypervalents. En pratique, pour des applications métabolomiques, la règle de Lewis est bien adaptée aux molécules étudiées. Une autre règle pouvant être utilisée est la « règle de l'azote ». Selon cette règle, les ions ayant un m/z pair auront un nombre impair d'atome d'azote et *vice versa* (Karni and Mandelbaum, 1980). Il est toutefois important de noter que cette règle ne peut être appliquée que si les ions formés dans la source sont à nombre pair d'électrons. Elle est donc bien adaptée aux analyses utilisant une source électrospray mais peu adaptée en cas d'ionisation électronique par exemple. En plus de ces règles chimiques, différents ratios entre les éléments ont été étudiés. Kind et Fiehn ont analysé 45 000 formules brutes de composés et calculé les ratios entre plusieurs éléments tels que les ratios H/C, N/C, O/C, P/C, S/C... Pour chacun de ces ratios, ils ont déterminé une gamme permettant de regrouper 99,7 % des molécules connues (Tableau 4).

	H/C	N/C	O/C	P/C	S/C
Gamme couvrant 99,7% des formules brutes référencées	0.2-3.1	0-1.3	0-1.2	0-0.3	0-0.8

Tableau 4 : Ratios classiques entre différents atomes à respecter pour l'annotation en métabolomique (Kind and Fiehn, 2007)

Grace à ces ratios et aux règles chimiques, il est possible d'éliminer un certain nombre de formules brutes respectant la tolérance en masse fixée mais excentrique d'un point de vue chimique, comme par exemple $C_2H_{14}O_{10}N_8P_2S$ (écart de 0,36 ppm par rapport à la masse de l'UDP, Tableau 3).

5.2.1.1.3 Apport de l'abondance isotopique relative (RIA)

Si plusieurs formules brutes cohérentes remplissent les différents critères fixés ci-dessus, un autre niveau d'information peut être utilisé pour encore réduire le nombre de formules brutes potentielles. Il s'agit de l'abondance isotopique relative. Cette abondance isotopique est liée à l'existence de plusieurs formes isotopiques stables pour les différents éléments chimiques (Tableau 5). L'abondance de ces différentes formes étant connue, il est possible de générer un spectre de

masse théorique d'un composé incluant tous ses isotopologues. A l'inverse, à partir des intensités relatives des différents isotopologues mesurés sur un spectre enregistré, il est théoriquement possible de remonter au nombre d'atome de chaque élément constituant le composé étudié et de trancher entre différentes formules brutes.

	Hydrogène		Carbone		Azote		Oxygène		
Isotope	¹ H	² H	¹² C	¹³ C	¹⁴ N	¹⁵ N	¹⁶ O	¹⁷ O	¹⁸ O
Proba	0.9999	0.0001	0.9893	0.0107	0.9963	0.0037	0.9976	0.0004	0.002
Masse	1.00783	2.01410	12.00000	13.00335	14.00307	15.00011	15.99491	16.99913	17.99916
Delta de masse		1.00628		1.00335		0.99703		1.0042	2.00425
	Soufre		Chlore		Potassium				
Isotope	³² S	³³ S	³⁴ S	³⁵ Cl	³⁷ Cl	³⁹ K	⁴¹ K		
Proba	0.9504	0.0075	0.0421	0.7577	0.2423	0.9326	0.0673		
Masse	31.97207	32.97146	33.96787	34.96885	36.96593	38.96316	40.96128		
Delta de masse		0.9994	1.99580		1.99707		1.99812		

Tableau 5 : Abondance isotopique et différence de masse des formes isotopiques en métabolomique

Pour que l'abondance isotopique puisse être utilisée le plus efficacement possible dans l'annotation, différents éléments sont à prendre en compte. Le premier est la résolution avec laquelle le spectre est enregistré. En effet, certains isotopes ont des masses très proches comme par exemple l'oxygène 18 et le double carbone 13. Si la résolution n'est pas suffisante, les deux signaux sont mélangés et l'interprétation est beaucoup plus complexe et moins informative. Le second élément à considérer est la capacité de l'instrument à mesurer avec justesse cette abondance. En effet, si la justesse de mesure est insuffisante, la tolérance à appliquer pour éviter les faux négatifs sera telle que l'utilisation de l'abondance isotopique sera peu informative et ne permettra pas de trancher entre différentes formules brutes. Si tous les éléments sont réunis, l'abondance isotopique relative peut être au moins aussi informative que la précision de mesure de masse (Kind and Fiehn, 2006).

L'utilisation de l'abondance isotopique pour l'annotation a fait l'objet de nombreux travaux dans le cadre de cette thèse et les résultats associés sont présentés dans le chapitre 3.

5.2.2 Prédiction des temps de rétention en LC

Un autre élément qui peut aider à identifier un métabolite inconnu est l'utilisation d'outils de prédiction de temps de rétention. Ces outils sont utilisés en métabolomique depuis le début des années 2010 (Creek et al., 2011; Hagiwara et al., 2010). La prédiction des temps de rétention se base sur des modèles de relation entre structure chimique et temps de rétention chromatographique (QSRR : « *Quantitative Structure Retention Relationship* »). Pour établir un modèle QSRR, différentes étapes sont nécessaires (Cao et al., 2015) : avoir une représentation informatisée de la structure des molécules, calculer des descripteurs moléculaires à partir des structures, avoir une base de temps de rétention de standards acquis dans les conditions analytiques de l'étude, établir le modèle de rétention.

De manière générale, le format retenu pour la représentation informatisée de la structure des molécules est le SMILES (*Simplified Molecular Input Line Entry System*) pour sa simplicité d'utilisation par les outils informatiques.

Les descripteurs moléculaires servent à définir, par des valeurs chiffrées, les propriétés structurales et chimiques des molécules. Ils sont en général générés par différents outils de modélisations. ACD/Labs, Volsurf + ou Jchem for Excel font partie de ces outils. Une très grande variété de descripteurs peut être générée. Les plus utilisés sont le logP (coefficient de partition octanol/eau), le logD qui est une fonction pH dépendante du logP, le nombre de liaisons hydrogènes donneuses ou accepteuses, la masse ou le volume moléculaire ou encore le nombre de groupements phosphates. Le nombre de descripteurs moléculaires générés peut dépasser la centaine. Afin de faciliter la génération des modèles et pour éviter une redondance d'information trop importante, le nombre de descripteurs est souvent réduit à une dizaine en retirant les descripteurs fortement corrélés à d'autres (Bruderer et al., 2017). D'autres algorithmes plus complexes ont également été développés pour sélectionner au mieux les descripteurs moléculaires (Taraji et al., 2017).

Une fois les descripteurs sélectionnés, l'objectif est d'acquérir les temps de rétention d'un jeu d'apprentissage de standards de références qui servira à calculer le modèle. Il est essentiel que ces temps de référence soient acquis dans conditions identiques à celles de l'analyse. Si jamais le système chromatographique utilisé manque un peu de robustesse, les temps de rétention des standards utilisés devraient être acquis pour chaque séquence (Creek et al., 2011). Cela est relativement compliqué à mettre en place si le jeu de standards de référence est trop important. Afin de réduire la taille de ce jeu sans perdre en qualité d'information générée, l'utilisation de l'algorithme de Kennard et Stone a été proposé (Andries et al., 2009). Il permet la sélection d'un nombre réduit de standards au sein d'une liste principale en cherchant à couvrir l'ensemble de l'espace des descripteurs moléculaires défini par les éléments de la liste principale. Avec cette approche, le jeu de standards de référence peut être réduit à une dizaine. Il est toutefois important de bien s'assurer que la liste initiale et la liste réduite modélisent une rétention chromatographique comparable aux éléments à identifier (Amos et al., 2018).

Ces standards vont servir de référence pour calculer le modèle de prédiction des temps de rétention. Différents types de modèles peuvent être utilisés et donner de bons résultats selon les conditions d'analyse. Le plus utilisé est le modèle de régression linéaire multiple, notamment car il peut être utilisé à partir d'un nombre réduit de références, même si d'autres modèles plus complexes comme le *Random Forest* peuvent produire des résultats de qualité supérieure. Certains modèles sont basés sur plusieurs outils de modélisations tels que du *Random Forest* couplé à du *Support Vector Regression* (Wolfer et al., 2015). Les modèles créés sont ensuite évalués. Un des moyens les plus efficaces est l'utilisation d'un jeu de test de composés de standards dont les temps de rétention

sont connus. L'adéquation entre les temps expérimentaux et le temps prédit est évaluée. Différentes métriques peuvent être utilisées. L'utilisation d'un coefficient de corrélation ou le calcul d'une erreur moyenne absolue est très répandue mais la moyenne quadratique des pourcentages d'erreur semble donner les meilleurs résultats (Taraji et al., 2017).

Après validation du modèle, la prédiction des temps de rétention va permettre d'éliminer un nombre important de fausses annotations. Pour un composé inconnu, les temps de rétention de l'ensemble des candidats issus des précédentes étapes d'identification sont générés. Les candidats dont la fenêtre de temps de rétention prédit ne correspond pas au temps mesuré sont rejetés. La prédiction des temps de rétention est une valeur ajoutée par rapport à la détermination de la formule brute car elle va permettre d'éliminer des composés isomères du composé étudié. Elle permet d'éliminer entre 40 % et 80 % de faux positifs selon la finesse du modèle mis en place (Wolfer et al., 2015) ce qui est une aide précieuse dans l'identification des métabolites.

5.2.3 Prédiction et interprétation des spectres de fragmentation

L'utilisation des spectres de fragmentation est un élément essentiel dans l'identification des composés non présents dans les bases de données internes. Deux grands modes d'utilisation de ces spectres de fragmentation se distinguent : l'élucidation *de novo* des spectres MS² et l'utilisation d'outils informatiques permettent l'obtention de spectres fragmentés *in silico*.

5.2.3.1.1 Elucidation *de novo* des spectres MS²

L'élucidation *de novo* des spectres MS² est généralement très complexe et nécessite une très bonne connaissance des mécanismes de fragmentation. L'acquisition de cette compétence demande énormément de pratique et d'expérience. L'étude des mécanismes de fragmentation a fait l'objet d'un certain nombre d'articles pour des ionisations en mode positif et négatif (Holčapek et al., 2010; Levsen et al., 2007; Niessen, 2011; Weissberg and Dagan, 2011). Une revue récente liste les principales règles à retenir pour interpréter un spectre MS² (Vijlder et al., 2018).

1. En général, la localisation de la charge sur l'ion formé détermine la nature des ions fragments formés. On parle de mécanisme de fragmentation lié à la charge. Il est donc essentiel d'avoir une bonne connaissance des fonctions chimiques qui pourraient être protonées ou déprotonées. En électrospray, la capacité d'une fonction chimique à s'ioniser va dépendre de son acidité ou de sa basicité. Il n'est pour l'instant pas clair si c'est la basicité en solution ou en phase gazeuse qui joue le premier rôle. Il est également possible d'avoir des ions sous forme canonique ou zwitterion (2 charges d'un même signe et une charge du signe opposé sur la molécule). L'existence de ces formes zwitterion doit

être prise en compte car les charges étant placées sur des atomes différents, les fragments formés peuvent être différents de ceux issus de la forme canonique

2. La nature des ions formés et leur abondance relative dépend principalement du nombre et de la force des liaisons qui doivent être rompues ainsi que de la stabilité des produits et pertes générées. Une liaison C-C aliphatique sera plus facile à rompre qu'une liaison aromatique mais plus difficile qu'une liaison carbone-hétéroatome.
3. L'utilisation des différentes règles de détermination des formules brutes présentées ci-dessus s'applique également aux ion fragments, notamment la règle de l'azote.
4. Le principal mécanisme de fragmentation des ions formés par électrospray est la rupture d'une liaison carbone-hétéroatome avec migration de la charge sur le carbone en α . Un autre mécanisme principal est la formation de fragments de liaisons carbone-hétéroatome avec conservation de la charge sur l'hétéroatome par migration de protons. Dans certains cas, les deux fragments (carbone chargé ou hétéroatome chargé) peuvent être observés. On parlera alors d'ions complémentaires reconnaissables par le fait que la somme de leurs m/z est égale à la masse du parent + 1.
5. Il est aussi possible d'observer des ruptures de liaisons C-C adjacentes à un hétéroatome avec conservation de la charge sur l'hétéroatome et formation d'un ion iminium ou oxonium.
6. La rupture de liaisons conduit souvent à des réarrangements de protons. Ces mécanismes de réarrangement de protons ont été largement étudiés et ont fait l'objet de règles spécifiques (Tsugawa et al., 2016).

L'application de toutes ces « règles » et l'interprétation du spectre de fragmentation peut permettre de remonter à la structure de l'ion parent et de dessiner un chemin de fragmentation. Encore une fois, ce mode d'interprétation demande une grande connaissance du comportement des ions en phase gazeuse et demande une expertise très importante.

5.2.3.1.2 Utilisation des outils informatiques

L'une des premières utilisations des outils informatiques pour l'interprétation des spectres MS^2 est l'interrogation de base de données externes (Vinaixa et al., 2016). Les trois plus connues sont MassBank (Horai et al., 2010), METLIN (Smith et al., 2005) et mzCloud. L'intérêt de l'interrogation de bases de données spectrales est qu'aucun *a priori* n'est nécessaire pour les utiliser. Leur principal inconvénient est qu'elles sont encore largement incomplètes. De plus, les spectres présents dans ces bases sont enregistrés sur un instrument défini et dans des conditions analytiques fixées qui ne correspondent pas nécessairement à celles utilisées pour acquérir les spectres à interpréter. Or le spectre enregistré expérimentalement va dépendre de ces conditions. Il n'est donc pas toujours évident de reconnaître au sein d'une base le spectre d'un composé d'intérêt si les conditions

analytiques sont différentes. La fiabilité de l'utilisation de bases de données spectrales repose aussi beaucoup sur les algorithmes de comparaison de spectres. Ces algorithmes vont classer les spectres par score de ressemblance mais sans pouvoir affirmer qu'il s'agit bien du bon composé (Stein, 2012), notamment si ce dernier n'est pas présent dans la base spectrale.

Une autre approche pour pallier l'absence d'exhaustivité dans les bases est l'utilisation de bases de fragmentation générées *in silico*. Il existe différents modes de génération des spectres *in silico*. Le premier repose sur l'application de règles chimiques. Ces règles sont définies à partir de l'exploration de données expérimentales. En pratique, elles sont extraites de la littérature puis curées manuellement. Si actuellement les mécanismes de fragmentation des petites molécules à différentes énergies ne sont pas encore bien compris, l'étude approfondie des mécanismes de fragmentation et l'interprétation des données par des algorithmes basés sur ces règles semble se développer (Hufsky et al., 2014). Le développement du *machine learning* devrait aussi permettre d'accroître la connaissance des règles régissant la fragmentation (Kangas et al., 2012). Les principaux outils basés sur ce mode de génération de spectres sont Mass Frontier, ACD/MS Fragmenter et MOLGEN-MS.

Le second mode de génération est la fragmentation combinatoire. Il s'agit d'une rupture systématique de toutes les liaisons afin de générer tous les fragments possibles. L'avantage de ce mode de fragmentation est qu'il ne nécessite pas une connaissance approfondie des mécanismes de fragmentation. Son principal inconvénient est qu'il génère un nombre très important de fragments qui peuvent être dépourvus de sens chimique. Cela peut être amélioré par l'utilisation d'une fonction attribuant un « coût de clivage » pour chaque liaison de la molécule. Ce coût représente la difficulté de certaines liaisons à se rompre plus que d'autres mais sans pour autant s'assurer de la plausibilité de la rupture de cette liaison. L'autre gros inconvénient de cette approche est qu'elle ne prend pas du tout en compte les potentiels réarrangements intramoléculaires qui peuvent avoir lieu lors de la fragmentation. MetFrag (Wolf et al., 2010) est l'outil le plus utilisé basé sur ce mode de fragmentation.

Plus récemment, de nouveaux algorithmes, comme MetFusion (Gerlich and Neumann, 2013), ont été développés en combinant la génération de spectres *in silico* et l'interrogation de bases de données spectrales. Cela permet de combiner les avantages des deux approches pour augmenter la qualité de l'identification réalisée.

Tous ces différents outils basés sur l'utilisation de spectres MS² jouent un rôle très important dans l'annotation des composés en métabolomique par spectrométrie de masse. Il est toutefois important de rappeler, qu'à l'heure actuelle, pour passer d'une annotation niveau 2 à une identification niveau 1, le standard du composé proposé doit toujours être acheté ou synthétisé. Il doit alors être analysé pour vérifier la correspondance exacte avec le métabolite étudié.

Bilan

Au cours de cette première partie, nous avons tout d'abord présenté les différentes sciences « -omiques » avec leurs spécificités respectives. Cela a permis de mettre en avant l'intérêt que peut avoir la métabolomique pour observer ou retranscrire les changements ayant lieu lors de la perturbation de l'homéostasie d'un organisme. La métabolomique apparaît donc comme une voie d'intérêt pour mieux comprendre les dérèglements liés à diverses maladies et pour découvrir d'éventuels biomarqueurs permettant de stratifier des patients au sein de cohortes médicales.

Nous nous sommes ensuite attachés à détailler les différentes étapes d'une analyse de métabolomique non ciblée, en mettant en avant les avantages et les limites de ce type d'analyse. Lors de cette revue bibliographique, nous nous sommes particulièrement intéressés à deux étapes actuellement clés lors du déroulement d'une analyse : l'acquisition des données et leur annotation. Dans ces deux domaines, le perfectionnement des instruments de spectrométrie de masse permet aujourd'hui d'envisager des améliorations en termes de quantité et de qualité de données acquises. Ces améliorations devraient faciliter l'utilisation de cette technique pour la compréhension de phénomènes biologiques et médicaux. Cela devrait permettre à la métabolomique de continuer à faire ses preuves dans le cadre de l'étude de cohortes médicales et de poursuivre ainsi son essor.

Présentation du projet de recherche

L'objectif premier de ce projet de recherche est d'étudier une cohorte médicale en utilisant les outils de la métabolomique, mais aussi de la lipidomique et de la glycomique. Cette cohorte est composée de volontaires sains, de cirrhotiques et de patients souffrant d'encéphalopathie hépatique. Il s'agit d'une maladie causant des symptômes neurologiques pouvant aller jusqu'au coma et ayant pour origine une insuffisance hépatique (cirrhose dans la majorité des cas). L'encéphalopathie hépatique est sous diagnostiquée et pourrait toucher plus de 150 000 personnes en France avec un coût direct estimé à plus de 40 millions d'euros. Certaines des difficultés de prise en charge de patients souffrant de cette maladie sont liées à l'absence de connaissance précise des mécanismes biochimiques sous-jacents et l'absence de tests diagnostiques objectifs et rapides. La première partie de cette thèse consiste en l'étude d'échantillons (liquide céphalorachidien et plasma) de patients encéphalopathes par différentes techniques « -omiques » avec deux buts principaux. Le premier est d'identifier des biomarqueurs de la maladie et d'essayer de mieux comprendre sa physiopathologie. Le second est d'étudier la concordance des données générées par les différentes méthodes analytiques et de montrer comment ces données peuvent être combinées ensemble. Grâce à cette étude, nous avons également mis en évidence certaines étapes critiques lors de la réalisation concrète d'une analyse métabolomique. Cela nous a permis d'identifier des pistes d'amélioration dans ce type d'approche qui feront l'objet des deux parties suivantes.

Le second objectif de ce projet de recherche est donc de développer ou d'améliorer les outils permettant la réalisation efficace et robuste d'une analyse métabolomique. Cette partie se focalise sur une des étapes du déroulement de ce type d'étude : l'identification des métabolites d'intérêt. Deux points se référant à cette thématique ont été étudiés.

Le premier vise à évaluer les performances d'un Orbitrap à haut champ (Orbitrap Fusion) en termes de qualité d'acquisition de données. Cette évaluation est faite sur des éléments indispensables à la réalisation de l'étape d'identification : la précision de mesure du rapport m/z et la justesse de mesure de l'abondance isotopique relative. L'objectif est également de tester la capacité de l'instrument à détecter spécifiquement le signal lié à l'isotope ^{18}O et de montrer que la mesure de cet isotope peut permettre d'affiner les propositions d'annotation. Du fait de sa faible abondance et de sa proximité en masse avec d'autres isotopes (comme le $^{13}\text{C}_2$), ce travail ne peut être fait qu'à très haute résolution et n'a, à notre connaissance, encore jamais été réalisé sur un instrument de technologie Orbitrap. Tout cela nous a conduit à recommander un certain nombre de paramètres à utiliser sur cet instrument. Ce paramétrage vise à augmenter la qualité des données acquises, facilitant ainsi l'annotation des signaux d'intérêt.

Le second point étudié concerne un autre élément indispensable à l'identification des signaux : les spectres de fragmentation MS². En effet, il s'agit de l'élément le plus fiable pour identifier un composé en spectrométrie de masse. Pour utiliser cette information, deux approches existent. La première consiste à analyser, dans les mêmes conditions analytiques, le standard du métabolite dont on veut confirmer l'annotation et à comparer les spectres MS² du métabolite à confirmer et du standard. Si ces spectres sont identiques, cela confirme l'identification du métabolite. Cependant, ce standard n'est pas toujours disponible. Dans ce cas, on cherche à identifier *de novo* le métabolite à partir, notamment, de son spectre de fragmentation. Pour pouvoir remonter à la structure du métabolite à partir des fragments observés, il est indispensable de bien connaître les mécanismes mis en jeu lors de la fragmentation. Or, ces mécanismes chimiques, en phase gazeuse, sont encore mal connus. Nous allons donc chercher à mieux les comprendre. A titre d'exemple de ce qu'il est possible de faire, nous avons travaillé sur deux métabolites présentant une structure chimique très proche mais des profils de fragmentation différents : l'acide aspartique et l'acide glutamique. Le but est de rationaliser complètement les mécanismes de fragmentation afin de mieux comprendre les règles chimiques mises en jeu. En plus de l'interprétation *de novo* de spectres de fragmentation, ces éléments peuvent servir à améliorer les logiciels de fragmentation *in silico*. Tout cela permettra de faciliter l'identification de composés pour lesquels le standard n'est pas disponible.

Enfin, la troisième partie de mes travaux a pour objectif d'améliorer les méthodes d'acquisition de données en métabolomique. Comme cela a déjà été discuté, l'un des éléments clés pour l'identification des composés est l'obtention de spectres de fragmentation de qualité. Or, dans une analyse métabolomique, telle qu'elle se déroule traditionnellement, l'acquisition de ces spectres se fait dans un second temps. Un fois les signaux d'intérêt identifiés à partir du profil MS, les échantillons sont réanalysés dans le but d'acquérir les spectres de fragmentation de ces signaux. Cela a des conséquences en termes de durée globale d'analyse, de quantité d'échantillon requise et entraîne une complexification de l'interprétation liée à l'évolution de l'échantillon et du système analytique entre ces deux analyses. L'objectif de ce travail est donc de tirer profit des améliorations apportées aux dernières générations d'instruments pour développer de nouvelles méthodes analytiques. Ces méthodes doivent combiner l'acquisition du profil MS et des spectres de fragmentation MS² afin d'éviter la nécessité d'une analyse en deux étapes. Des approches telles que l'acquisition données-dépendante ou l'acquisition données-indépendante ont été développées et leur plus-value par rapport à l'approche traditionnelle non ciblée est évaluée pour des analyses métabolomiques non ciblées.

L'objectif global de ce projet de recherche est donc à la fois de montrer l'intérêt de la métabolomique et des autres techniques « -omiques » phénotypiques dans un contexte médical, mais également d'optimiser les outils permettant de faciliter ce type d'analyse, tout en assurant la rapidité et la qualité des résultats.

Résultats

Chapitre 1. Application à l'étude de l'encéphalopathie hépatique

1 L'encéphalopathie hépatique

1.1 Définition et classification

L'encéphalopathie hépatique (EH) correspond à un dysfonctionnement cérébral dû à une insuffisance hépatique ou un shunt porto-systémique (Ferenci et al., 2002). Elle se manifeste par un large spectre de dérèglements neurologiques ou psychiatriques non spécifiques allant de simples altérations sous-cliniques jusqu'au coma. Dans l'EH, il y a une relation de causalité entre l'insuffisance de la fonction hépatique et les troubles neuropsychiatriques.

Il existe différentes classifications complémentaires de l'EH. La première se base sur le facteur sous-jacent à la maladie. L'EH est dite de type A si elle est associée à une défaillance hépatique aigüe. Elle est dite de type B si elle est due à un shunt porto-systémique, qu'il soit naturel ou chirurgical (Transjugular Intrahepatic Portosystemic Shunt). Enfin une EH consécutive à une cirrhose est de type C. L'EH de type A est relativement différente de celles de type B ou C et devra avant tout être gérée médicalement comme une hépatite aigüe.

Le second type de classification est basé sur la gravité de la maladie (qui correspond à un continuum de symptômes). A son niveau le moins grave, l'EH altère uniquement certains tests psychomoteurs portés sur l'attention, la mémoire de travail ou les capacités visuospatiales. Au fur et à mesure que la maladie progresse, des troubles de la personnalité (apathie, irritabilité, désinhibition) peuvent être observés ainsi qu'une altération des fonctions motrices. Des dérèglements du cycle du sommeil sont également rapportés. Les patients vont progressivement développer une désorientation spatio-temporelle, agir de façon inappropriée et présenter un état de confusion (agitation, somnolence, stupeur). Les formes les plus graves conduiront à un coma. Afin de classier le degré d'avancée de la maladie, ce continuum a été divisé en différents niveaux au sein d'un score dit de West-Haven. Cette échelle se divise en 5 niveaux allant de 0 (ou EH minimale) à 4 (coma). L'ensemble de cette échelle est présenté en Tableau 6.

Score West Haven	Etat de conscience	Capacités intellectuelles	Comportement	Troubles neuromusculaires
Score 0 (EH minime)		Altération neuropsychologiques		Altération de la vitesse/coordination psychomotrice
Score 1	Troubles du sommeil	Troubles d'attention Difficultés à réaliser des opérations complexes	Phases d'euphorie ou de dépression	Tremblements
Score 2	Léthargie	Désorientation temporelle Amnésie à court terme Difficultés à réaliser des opérations simples	Comportement parfois inapproprié, Anxiété	Astaxis, Difficulté d'élocution,
Score 3	Somnolence	Confusion Désorientation marquée	Comportement étrange, colérique	Myoclonies, rigidité Nystagmus Signe de Babinsky
Score 4	Coma Absence d'ouverture des yeux Pas de réponse verbale Absence de réaction	Impossible à tester	Impossible à tester	Impossible à tester

Tableau 6 : Echelle de classification de West-Haven (Blei and Córdoba, 2001)

1.2 Physiopathologie

La physiopathologie de la maladie est encore méconnue. Néanmoins, l'ammoniaque semble jouer un rôle central dans l'établissement de la maladie (Ciećko-Michalska et al., 2012; Elwir and Rahimi, 2017). Le schéma ci-dessous (Figure 11, (Jalan and Lee, 2009)) présente les déséquilibres globaux liés à l'EH. Chez un individu sain, l'ammoniaque produite par le microbiote intestinal est absorbée passivement au niveau du tractus digestif et se retrouve dans la circulation sanguine. Or, cette ammoniaque est toxique pour l'organisme. Chez un individu sain, elle est transformée en urée en 5 étapes via le cycle de l'urée qui a principalement lieu au niveau du foie. Que ce soit chez un individu cirrhotique (dont la fonction hépatique est fortement diminuée) ou chez un individu présentant un shunt porto-systémique (et chez qui le sang passe directement de la veine porte à la veine cave sans passer par le foie), cette transformation de l'ammoniaque en urée est diminuée et l'ammoniaque s'accumule dans l'organisme. Un mode secondaire d'élimination de l'ammoniaque est sa conjugaison à l'acide glutamique pour former de la glutamine. Parmi les rares types cellulaires à être capables de réaliser cette transformation se trouvent les astrocytes. Au sein du cerveau, l'ammoniaque en excès conduit à une forte production de glutamine dans les astrocytes. Or, contrairement à l'acide glutamique dont elle est issue, la glutamine présente un fort pouvoir osmotique. L'accumulation de glutamine entraîne un stress osmotique et un gonflement des astrocytes, ce qui engendre un œdème cérébrale cytotoxique. Cet œdème cérébral serait la cause directe des symptômes observés dans l'EH.

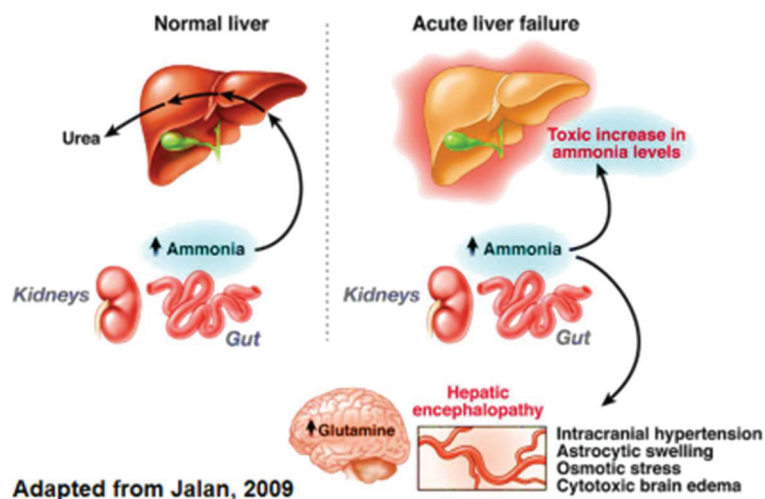


Figure 11 : Physiopathologie de l'encéphalopathie hépatique (Jalan and Lee, 2009)

1.3 Diagnostic

Le diagnostic de l'EH passe par l'observation des symptômes présentés ci-dessus, associés à une insuffisance hépatique sévère ou un shunt porto-systémique, et ce en l'absence d'autres causes possibles de dysfonctionnement cérébral. L'établissement de ce diagnostic est donc compliqué car il s'agit d'un diagnostic d'exclusion chez une population de patients présentant souvent des troubles du statut mental pouvant être liés à d'autres étiologies (prise de médicaments, abus d'alcool, usage de drogue, autres maladies psychiatriques...). Le diagnostic différentiel inclut également les neuro-infections (méningites), les désordres électrolytiques, l'épilepsie non convulsivante ou les accidents vasculaires cérébraux.

Pour effectuer le diagnostic clinique après exclusion des autres cause possibles, les cliniciens se basent sur des tests psycho-cognitifs pour attribuer au patient un score de West-Haven (Tableau 6). Cependant, il s'agit (comme la majorité des tests psycho-cognitifs) d'outils subjectifs à forte variabilité inter-observateur, notamment pour les EH de bas grade. Par contre, pour les grades intermédiaires, l'observation d'une désorientation ou d'un astérisis sont des marqueurs plus fiables d'une EH manifeste (Vilstrup et al., 2014). Pour les patients présentant un état de conscience fortement altéré, le Glasgow Coma Score peut également être utilisé (Tableau 7).

Glasgow Coma Scale		
Response	Scale	Score
Eye Opening Response	Eyes open spontaneously	4 Points
	Eyes open to verbal command, speech, or shout	3 Points
	Eyes open to pain (not applied to face)	2 Points
	No eye opening	1 Point
Verbal Response	Oriented	5 Points
	Confused conversation, but able to answer questions	4 Points
	Inappropriate responses, words discernible	3 Points
	Incomprehensible sounds or speech	2 Points
	No verbal response	1 Point
Motor Response	Obeys commands for movement	6 Points
	Purposeful movement to painful stimulus	5 Points
	Withdraws from pain	4 Points
	Abnormal (spastic) flexion, decorticate posture	3 Points
	Extensor (rigid) response, decerebrate posture	2 Points
	No motor response	1 Point
Minor Brain Injury = 13-15 points; Moderate Brain Injury = 9-12 points; Severe Brain Injury = 3-8 points		

Tableau 7 : Echelle de classification selon le Glasgow Coma Score (GCS) (Teasdale and Jennett, 1976)

Il est important de noter que malgré son rôle important dans la physiopathologie de l'EH, une augmentation du taux sanguin d'ammoniaque n'apporte aucun élément diagnostique, classifiant ou pronostique pour l'EH chez les patients souffrant d'insuffisance hépatique chronique. A l'heure actuelle, il n'existe donc pas de test objectif permettant de diagnostiquer un patient souffrant d'EH.

1.4 Prévalence et impact

La prévalence de l'EH chez les patients cirrhotiques est difficile à évaluer, notamment pour les EH de bas grades du fait de la subtilité des symptômes. Parmi les trois types d'EH, le plus fréquent est le type C, ce qui est essentiellement dû à la forte prévalence de la cirrhose. Il s'agit d'un vrai problème de santé publique dans les pays développés. En France, 700 000 cas de cirrhose sont recensés parmi une population à risque de 12 million de personnes (INSERM, 2012). La proportion de patients cirrhotiques développant une EH est actuellement d'au moins 20 % (Romero-Gómez et al., 2015). Cependant, l'EH n'est pas forcément diagnostiquée chez tous les patients hospitalisés et la prévalence de l'EH semble fortement sous-évaluée. Une étude rétrospective menée en Allemagne sur 19 hôpitaux a montré qu'à peine 20 % des 14 000 patients hospitalisés pour cirrhose étaient diagnostiqués comme encéphalopathes. Cependant, une étude prospective de 3 mois consistant à établir un diagnostic d'EH chez tous les cirrhotiques hospitalisés dans ces mêmes hôpitaux a montré qu'une EH clinique pouvait être diagnostiquée chez 36 % des patients (dont plus de 40 % avec un

score de West Haven égal ou supérieur à 2) (Labenz et al., 2017). Ce chiffre de 36 % semble mieux correspondre à la réalité.

En France, le coût des hospitalisations liées à l'encéphalopathie hépatique a été évalué aux alentours de 40 millions d'euros (Abergel et al., 2016). Cependant, il ne s'agit là que du coût d'hospitalisation et l'impact économique global lié aux coûts indirects (perte de revenus, baisse de la qualité de vie, ...) serait bien plus important.

1.5 Traitements actuels

Différents traitements sont actuellement utilisés dans le cadre de l'EH. Les traitements de première intention visent à limiter la production et l'absorption d'ammoniaque au niveau intestinal. Le lactulose est généralement le traitement initial de l'EH. En plus d'être un laxatif osmotique, il entraîne une acidification du contenu intestinal. Cette acidification conduit au piégeage, dans la lumière intestinale, de l'ammoniaque sous forme d'ion ammonium. Il aurait aussi un effet probiotique, favorisant la croissance de bactéries non ammoniogènes et inhiberait l'activité glutaminase de l'intestin. L'autre traitement utilisé en première intention est la Rifaximine. Il s'agit d'un antibiotique qui va limiter la croissance des bactéries intestinales aérobies, réduisant la production d'ammoniaque. Elle induit également une atrophie des villosités de l'intestin grêle conduisant à une baisse de l'absorption intestinale. Comme le lactulose, elle inhiberait aussi l'activité glutaminase de l'intestin.

D'autres traitements supports peuvent être utilisés comme l'administration intraveineuse de L-Ornithine-L-Aspartate qui augmente l'activité de glutamine synthétase au niveau du foie et des muscles striés squelettiques ou encore l'administration de benzoate ou de phenylacétate de sodium qui se conjuguent à la glycine ou à la glutamine, déplaçant ainsi les équilibres vers la consommation d'ammoniaque. Il existe aussi des traitements qui corrigent les problèmes de neurotransmission comme l'administration orale d'acides aminés branchés, l'administration d'agents dopaminergiques (L-Dopa, bromocriptine) ou le flumazenil. Ce dernier est un antagoniste des récepteurs aux benzodiazépines qui améliore transitoirement le statut mental des patients.

En dernier recours, la transplantation hépatique est la seule option en cas de non réponse aux différents autres traitements. Cependant, en raison des risques liés à la réalisation d'un greffe, elle n'est utilisée que si l'EH impacte sévèrement la qualité de vie du patient et que cette dernière ne peut pas être améliorée par l'administration d'un traitement médical maximal.

2 Objectifs des études

L'EH est une maladie encore mal comprise et pour laquelle les mécanismes moléculaires impliqués sont encore en partie inconnus. Le premier objectif de ces études est donc d'appliquer les techniques de métabolomiques pour essayer de mieux comprendre les dérèglements impliqués dans l'encéphalopathie hépatique. Le second est d'essayer de trouver des biomarqueurs spécifiques de l'EH qui pourraient servir d'aide au diagnostic et permettraient d'avoir un outil plus robuste que les tests psychométriques qui font actuellement office de référence.

Afin de répondre à ces objectifs, deux études ont été réalisées. La première, publiée dans *Journal of Hepatology*, s'est principalement concentrée sur l'analyse métabolomique du LCR. La seconde, plus générale, s'est appuyée sur des outils de métabolomique, de glycomique et de lipidomique pour analyser de manière exhaustive du plasma.

3 Aspects méthodologiques

Au cours de la première étude, une cohorte composée de 27 sujets sains et de 14 encéphalopathes a été analysée en métabolomique non ciblée par LC-HRMS.

La matrice choisie était le liquide céphalo-rachidien en raison de sa plus grande proximité spatiale avec le cerveau. L'objectif était d'avoir une représentation la plus fidèle possible des dérèglements s'opérant à ce niveau.

Les profils ont été acquis dans trois conditions : colonne C₁₈ en ionisation positive par électrospray, colonne C₁₈ en ionisation négative par électrospray, colonne HILIC en ionisation négative par électrospray. Ces profils ont été enregistrés sur un Exactive (instrument de type Orbitrap). Les données ont ensuite été traitées et les signaux annotés par rapport à une base de données interne de près de 1000 composés. Les annotations basées sur le rapport m/z et sur le temps de rétention ont été confirmées par l'acquisition de spectres de fragmentation MS². Des analyses statistiques uni et multivariées ont été menées pour identifier les métabolites discriminants entre les deux populations. En plus de l'analyse du LCR de ces patients, une seconde cohorte a été montée afin de vérifier la spécificité de la signature observée vis-à-vis d'une signature de cirrhose sans EH. Pour des raisons éthiques, du plasma et non du LCR a été recueilli. Il n'est en effet pas envisageable de réaliser une ponction lombaire sur un patient cirrhotique sans autre indication médicale. Cette cohorte était composée de 9 volontaires sains, de 13 cirrhotiques non encéphalopathes et de 12 patients souffrant d'EH). Dans la première étude, seuls les métabolites initialement détectés dans le LCR ont été suivis.

La seconde étude se place dans la continuité directe de la précédente. Elle reprend la cohorte de plasma étudiée ci-dessus mais cette fois de manière plus complète en essayant de combiner plusieurs analyses « -omiques » : métabolomique, lipidomique et glycomique. Le but étant de continuer à approfondir la compréhension de l'EH, mais aussi de voir dans quelle mesure il est possible de combiner les résultats de ces différents types d'analyses.

L'analyse métabolomique sera menée par LC-HRMS sur l'Orbitrap Fusion. Une colonne C₁₈ et une colonne HILIC ont été utilisées et les analyses ont été conduites dans les deux modes d'ionisation.

Pour l'analyse lipidomique, une extraction liquide de type Bligh and Dyer a été réalisée avec récupération et mélange des deux phases. Ces échantillons ont ensuite été analysés par LC-HRMS sur une colonne C₈ en ionisation électrospray positive et négative sur un Q-Exactive Plus. Les différentes espèces lipidiques ont ensuite été identifiées sur la base de leur rapport m/z, de leur temps de rétention et de leur spectre de fragmentation.

L'analyse glycomique des N-glycanes a été réalisée sur un instrument de type MALDI-TOF. Les échantillons ont été traités à la PNGase F en condition dénaturante afin de déglycosyler les protéines. Après purification, les glycanes sont perméthylés et déposés sur une plaque MALDI. L'analyse se fait en ionisation positive. Les différents signaux seront annotés par rapport à leur rapport m/z, par rapport à des bases de données spécifiques.

Les différents paramètres analytiques utilisés lors de l'analyse du plasma sont détaillés en Annexe 4.

4 Publication numéro 1

Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy

Nicolas Weiss, Pierre Barbier Saint Hilaire, Benoit Colsch, Foucauld Isnard, Suleiman Attala, Augustin Schaefer, Maria del Mar Amador, Marika Rudler, Foudil Lamari, Frédéric Sedel, Dominique Thabut, Christophe Junot

Journal of Hepatology, 2016, 65, 1120-1130

Les données supplémentaires liées à cette publication sont présentées en Annexe 3.



Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy

Nicolas Weiss^{1,2}, Pierre Barbier Saint Hilaire³, Benoit Colsch³, Foucauld Isnard³, Suleiman Attala³, Augustin Schaefer¹, Maria del Mar Amador⁴, Marika Rudler^{1,5}, Foudil Lamari⁶, Frédéric Sedel⁷, Dominique Thabut^{1,5,†}, Christophe Junot^{3,*,†}

¹Brain Liver Pitié-Salpêtrière (BLIPS) Study Group, Groupement Hospitalier Pitié-Salpêtrière-Charles Foix, Paris, France; ²Unité de réanimation neurologique, Fédération de neurologie 1, pôle des maladies du système nerveux, Groupement Hospitalier Pitié-Salpêtrière-Charles Foix, Assistance Publique – Hôpitaux de Paris et Institut de neurosciences translationnelles IHU-A-ICM, Paris, France; ³CEA, iBiTec-S, Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, MetaboHUB-Paris, Université Paris Saclay, 91191 Gif-sur-Yvette cedex, France; ⁴Neurometabolic Unit and University Pierre and Marie Curie, Paris, Groupe Hospitalier Pitié Salpêtrière-Charles Foix, 83 Boulevard de l'Hôpital, 75013 Paris, France; ⁵Unité de Soins Intensifs d'Hépatogastroentérologie, Groupement Hospitalier Pitié-Salpêtrière-Charles Foix, Assistance Publique – Hôpitaux de Paris et Université Pierre et Marie Curie Paris 6, Paris, France; ⁶Department of Metabolic Biochemistry, Groupe Hospitalier Pitié Salpêtrière-Charles Foix, 83 Boulevard de l'Hôpital, 75013 Paris, France; ⁷Medday Pharmaceuticals, ICM-Brain and Spine Institute-iPEPS, Groupe Hospitalier Pitié Salpêtrière-Charles Foix, 83 Boulevard de l'Hôpital, 75013 Paris, France

See Editorial, pages 1077–1078

Background & Aims: Hepatic encephalopathy (HE) is a neurological complication observed in patients with liver disease and/or porto-systemic shunt. The proportion of cirrhotic patients developing overt HE is about 20%, and 60–80% of cirrhotic patients exhibit mild cognitive impairment potentially related to minimal HE. However, the pathophysiological mechanisms of HE remain poorly understood. In this context, metabolomics was used to highlight dysfunction of metabolic pathways in cerebrospinal fluid (CSF) samples of patients suffering from HE.

Methods: CSF samples were collected in 27 control patients without any proven neurological disease and 14 patients with symptoms of HE. Plasma samples were obtained from control patients, and from cirrhotic patients with and without HE. Metabolomic analysis was performed using liquid chromatography coupled to high-resolution mass spectrometry.

Results: Concentrations of 73 CSF metabolites, including amino acids, acylcarnitines, bile acids and nucleosides, were altered in HE patients. Accumulation of acetylated compounds, which could be due to a defect of the Krebs cycle in HE patients, is reported for the first time. Furthermore, analysis of plasma samples showed that concentrations of metabolites involved in ammonia, amino-acid and energy metabolism are specifically and significantly increased in CSF samples of HE patients. Lastly, several

drugs were detected in CSF samples and could partially explain worsening of neurological symptoms for some patients.

Conclusion: By enabling the simultaneous monitoring of a large set of metabolites in HE patients, CSF metabolomics highlighted alterations of metabolic pathways linked to energy metabolism that were not observed in plasma samples.

Lay summary: CSF metabolomics provides a global picture of altered metabolic pathways in CSF samples of HE patients and highlights alterations of metabolic pathways linked to energy metabolism that are not observed in plasma samples.

© 2016 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Hepatic encephalopathy (HE) encompasses a spectrum of neurological and neuropsychiatric abnormalities observed in patients with liver disease and/or porto-systemic shunt, after exclusion of other known brain diseases. According to the classification of the World Congress of Gastroenterology held in 1998 in Vienna, three HE categories should be defined: type A is associated with acute liver injury, type B is associated with a porto-systemic shunt without associated liver disease, and type C is associated with cirrhosis or portal hypertension [1]. The latter type is the most frequent since cirrhosis is a major public health problem in developed countries with, in France, 700,000 reported cirrhotic patients associated with 15,000 deaths per year and an at-risk population of 12 million people [2]. The prognosis of cirrhosis has improved dramatically in recent years by refinement of prophylactic strategies and better management of complications (gastrointestinal bleeding,

Keywords: Metabolomics; Cerebrospinal fluid; Hepatic encephalopathy; Cirrhosis; Liquid chromatography; High-resolution mass spectrometry; Blood-brain barrier.

Received 9 October 2015; received in revised form 11 July 2016; accepted 22 July 2016; available online 9 August 2016

* Corresponding author. Address: Laboratoire d'Etude du Métabolisme des Médicaments, DSV/iBiTec-S/SPI, bat 136, CEA-Saclay, Gif-sur-Yvette cedex 91191, France. Tel.: +33 1 69 08 43 66.

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

† These authors share senior authorship.



ELSEVIER

ascites). Furthermore, the recent and widespread indication of early trans-jugular intra-hepatic porto-systemic shunt has decreased mortality with, however, an increased occurrence of HE by direct hepatic shunting [3]. Thus, the proportion of cirrhotic patients developing overt HE has largely increased and is currently at least 20% [4].

The pathophysiological mechanisms of HE are still poorly understood. It is generally assumed that HE is caused by an association of increased ammonia levels and inflammation. Ammonia, the degradation product of nitrogen compounds, is metabolized by the liver into urea in healthy individuals. Ammonemia is increased in cirrhotic patients, and reflects altered liver function or porto-systemic shunting. Outside the liver, ammonia can be metabolized into glutamine by glutamine synthase in muscle cells and in astrocytes of the central nervous system. Glutamine is osmotically active and high glutamine levels are responsible for astrocyte swelling, and then afterwards brain edema, which are compensated for by the release from astrocytes of taurine and myoinositol, which act as osmoregulators [5].

However, although these mechanisms are well documented and accepted, there is no correlation between the severity of HE and ammonia or glutamine blood levels [6–8]. In animal models and in humans, some studies have shown that high ammonia levels induce HE only if systemic inflammatory response syndrome is present [9–11]. Thus, it is widely accepted that sepsis is able to trigger HE in cirrhotic patients as a result of altered nitrogen metabolism and also by releasing pro-inflammatory mediators [12].

Finally, impaired blood-brain barrier (BBB) permeability and transport have been pointed out as a mechanism involved in the occurrence of HE manifestations [13]. For example, in response to increased glutamine concentrations, an accelerated transport of methionine and aromatic amino acids (AAAs, including tryptophan, phenylalanine and tyrosine) across the BBB has been observed [14,15].

Cerebrospinal fluid (CSF) is a medium of choice to approximate intra-cerebral metabolism because of its proximity to neuronal and glial cells, and its biochemical analysis could thus be of value in understanding the pathophysiology of HE. However, there are few published studies regarding the metabolic composition of CSF in HE. In 1974, Vergara *et al.* reported an increase of ammonia concentration in CSF with large variations between patients [16]. They also detected elevated levels of glutamine, glutamate, and α -ketoglutarate (α KG). Increased concentrations of neurotransmitters such as serotonin and dopamine [17], amino acids like tryptophan, tyrosine, phenylalanine and methionine, and bile acids were also reported by other authors [14,18]. However, these data are sparse, and restricted to few metabolites.

By enabling the concomitant detection of a wide range of metabolites, metabolomics could address this issue. This is a data-driven and a multidisciplinary approach combining analytical chemistry for the acquisition of metabolic fingerprints, and biostatistics, informatics and biochemistry for mining and interpretation of data. Metabolomics has already been successfully applied to highlight disease biomarkers in many medical fields [19]. By using metabolomics, we aimed to provide a global picture of the metabolic phenotypes of HE patients.

In this study, we applied liquid chromatography coupled to high-resolution mass spectrometry-based metabolomics to CSF samples of HE patients. By these means, we observed multiple alterations of metabolic pathways. Some of them have already been described, whereas others are highlighted for the first time.

Patients and methods

This study was performed in the hepatology intensive care unit (ICU) of La Pitié-Salpêtrière Hospital between November 2012 and June 2013 for CSF samples, and in 2015 for plasma samples. Samples were collected after written informed consent, in accordance with the local ethics committee of La Pitié-Salpêtrière Hospital, Paris, France.

Patients

CSF study

We prospectively collected CSF samples from 14 patients admitted to the hepatology ICU for symptoms of HE. They underwent lumbar puncture to rule out meningitis or meningoencephalitis. Patients with a previous neurological history other than HE or with a do-not-resuscitate order were excluded. The severity of cirrhosis was evaluated by the Child-Pugh score [20] and the model for end-stage liver disease (MELD) score [21]. Leukocyte count, hemoglobin level, platelets, prothrombin time, international normalized ratio (INR), liver enzymes, creatinine level, sodium level, albumin, C-reactive protein (CRP), and ammonemia were recorded. HE was assessed using the West Haven (WH) score [1]. Overt HE, so-called HE patients in the present study, was defined by a WH score ranging from 2 to 4 [22].

Patient samples were compared with those of 27 CSF control samples collected from adult patients seen in the Neurometabolic Unit of La Pitié-Salpêtrière Hospital and free of any neurological disease after extensive etiological work-up.

Plasma study

Plasma samples were collected from 9 healthy control samples, 12 overt HE and 13 cirrhotic patients without overt HE. The severity of cirrhosis was evaluated by the Child-Pugh and the MELD scores, whereas the severity of HE was assessed by the WH score. Leukocyte count, hemoglobin level, platelets, prothrombin time, INR, liver enzymes, creatinine level, sodium level, albumin, CRP, and ammonemia were also recorded. Clinical data are given as absolute numbers and percentages for qualitative variables and median and interquartile range for quantitative variables.

Metabolomic analyses

Chemicals and reagents, the procedure for extraction of CSF and plasma metabolites, experimental settings for liquid chromatography coupled to high-resolution mass spectrometry, data processing tools and statistical analyses are detailed in the [Supplementary material](#).

Results

Between November 2012 and June 2013, 14 patients underwent lumbar puncture. Of the 14 CSF samples, 12 were from HE cirrhotic patients (type C) without trans-jugular intra-hepatic porto-systemic shunt, and 2 from non-cirrhotic HE patients: a liver transplant patient (PV2) who developed HE before transplantation and displayed delirium 4 months after transplantation, and a patient with a rare cause of portal hypertension (hepatoportal sclerosis) (MP11). Diagnosis of meningitis or meningoencephalitis was ruled out for all patients. CSF samples from HE patients were compared with those of 27 control patients. Patient characteristics are shown in [Table 1](#).

Three complementary liquid chromatography coupled to high-resolution mass spectrometry based methods were used to provide optimal coverage of the CSF metabolome. The annotation procedure and also MS/MS experiments led to the identification and relative quantification of 122 metabolites. Analytical information and identification status of annotated metabolites are displayed in [Supplementary Table 1](#) and in the [Supplementary material](#).

Research Article

Table 1. Clinical metadata.

	Control patients n = 27	HE patients n = 14
Age ^s	48 [33-62]	55 [53-58]
Sex (male/female)	14/13	10/4
Cirrhosis	n.a.	12 (86%)
Etiology of cirrhosis*		
Alcohol	n.a.	9
Viral	n.a.	2
NASH	n.a.	1
Ascites*	n.a.	12 (100%)
Flapping*	n.a.	8 (57%)
Child-Pugh score* ^s	n.a.	12 [10-14]
MELD * ^s	n.a.	16 [13-22]
West Haven score (WH)* ^s	n.a.	2 [2-4]
GCS ^s	n.a.	14 [11-15]
Blood		
Leukocytes (Giga/L) ^s	5500 [4580-8600]	8200 [6500-8700]
Platelets (Giga/L) ^s	171,000 [123,000-198,000]	87,000 [55,000-114,000]
Prothrombin time (%) ^s	89 [80-97]	52 [40-69]
INR ^s	1.0 [1.0-1.1]	1.8 [1.5-2.1]
Factor V (%) ^s	98 [90-98]	57 [45-89]
AST (U/L) ^s	36 [17-55]	64 [54-107]
ALT (U/L) ^s	34 [16-62]	44 [25-90]
Creatinine (μmol/L) ^s	80 [55-99]	62 [33-87]
Sodium (mmol/L) ^s	141 [137-145]	136 [133-141]
Albumin (g/L) ^s	38 [35-41]	27 [23-32]
CRP (mg/L)* ^s	n.a.	26 [6-39]
Bilirubin (mmol/L) ^s	15 [10-31]	26 [20-212]
Ammonemia (μmol/L)* ^s	n.a.	52 [33-70]
Lumbar puncture		
Elements (/mm ³) ^s	1 [1-5]	3 [1-7]
Red cells (mm ³) ^s	25 [5-35]	85 [21-3 622]
Protein (g/L) ^s	0.3 [0.2-0.4]	0.5 [0.3-0.6]
Glucose (mmol/L) ^s	4.0 [3.4-4.8]	3.6 [3.1-4.1]

^sValues are medians with interquartile ranges. *Only available in cirrhotic patients.

Metabolic information contained in the metabolic fingerprint enables the separation of HE from control patients

Data were initially processed using principle component analysis (PCA) in order to visualize the structure of the datasets in a two-dimensional space. The PCA score plots are shown in Fig. 1A–C for data obtained using the reverse phase column with detection in positive and negative modes, and using the hydrophilic interaction chromatography (HILIC) system, respectively. The three score plots show a clear discrimination between control and HE samples, with a wide variation of metabolic profiles related to HE patients.

The MP11 sample, which is part of the cluster corresponding to control samples for the 3 analytical conditions, corresponds to a patient with minimal HE at the time of lumbar puncture, but without cirrhosis and who had a non-cirrhotic cause of portal hypertension (i.e., hepatoportal sclerosis). Sample PV2 corresponds to a patient who received a liver transplant 2 months before. He experienced severe HE before transplantation. His neurological symptoms did not resolve after transplantation, and the diagnosis of neurological disorders was HE because of flapping without any other cause. Interestingly, this sample is closer to the control group than other HE patients (Fig. 1).

Several samples were outside the cluster of HE samples. One of them (ZG8) did not display hyperammonemia and had high concentrations of drugs in CSF, suggesting drug-induced encephalopathy. Another one (BM2) had cirrhosis, HE, status epilepticus, and high concentrations of an antiepileptic drug (levetiracetam), which could possibly modify its metabolite compared with other HE patients.

The metabolic signature of HE patients

Univariate statistical analyses (i.e., Mann-Whitney *U* test) were performed on the annotated and identified metabolites, from lists of integrated peaks obtained by manual peak integration.

The list of discriminating features, which is provided in Table 2A, indicates that 72 metabolites exhibit altered

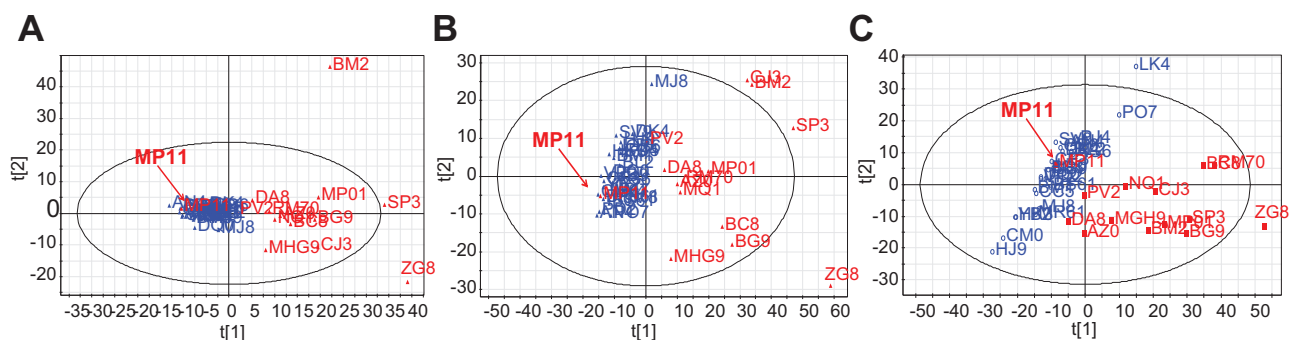


Fig. 1. PCA score plot obtained in CSF samples of control and HE patients. Samples of control and HE patients are displayed as open blue circles and red squares, respectively. Data were obtained using reverse phase chromatography with detection in the positive (A) and negative (B) ionization mode, and using HILIC with detection in the negative ionization mode (C).

concentrations in HE patients. Most of these metabolites belong to ammonia, energy metabolism, AAA and methylation pathways. Increased levels of bile acids were also observed.

Metabolites involved in ammonia metabolism

Glutamine and glutamic acid concentrations were increased by factors of 2.4 and 2.3 in HE patients, respectively. The synthesis

of glutamine is a major pathway for ammonia detoxification in the brain. Glutamate, which is produced by transamination of α KG, reacts with ammonia to form glutamine. This reaction occurs in astrocytes and is catalyzed by glutamine synthase. Besides glutamate and glutamine, the levels of pyroglutamic, phenylacetylglutamine and glutamyl-glutamine were also increased in HE samples.

Table 2. The metabolic signature of HE patients.

Biochemical pathway or chemical class	Metabolite	A		B					
		CSF		Plasma					
		Fold change EH vs. CTL	p value Mann Whitney	Fold change EH vs. CTL	Fold change Cir vs. CTL	Fold change EH vs. CIR	Mann Whitney-EH vs. CTL	Mann Whitney-CIR vs. CTL	Mann Whitney-Cir vs. EH
Acetylated compounds	Acetyl-glucosamine	2.26	4.3E-06	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	Acetyl-neuraminic acid	1.41	3.4E-02	2.04	1.22	1.68	1.9E-01	7.9E-01	1.2E-01
	Acetyl-serine	1.48	6.3E-03	2.42	1.49	1.62	6.6E-04	1.6E-01	1.5E-01
	Acetyl-tyrosine	6.84	4.3E-06	11.34	3.37	3.36	2.0E-04	1.2E-04	1.5E-01
	Acetyl-valine OR acetyl norvaline	1.99	2.8E-05	1.67	1.27	1.32	1.1E-01	3.6E-01	3.2E-01
	N4-Acetylcytidine	6.97	2.6E-04	6.08	3.55	1.71	6.8E-06	2.8E-05	3.0E-02
	N-Acetyl-allo-isoleucine/N-Acetyl-Leucine	2.20	6.0E-05	1.29	0.99	1.31	6.0E-01	6.0E-01	4.4E-01
	Acetyl-alanine	2.24	1.9E-05	2.21	1.53	1.44	2.3E-02	3.6E-02	2.9E-01
	N-Acetyl-L-histidine	3.35	4.2E-05	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	Acetyl-methionine	10.78	1.5E-05	3.07	1.58	1.95	2.8E-01	3.6E-01	4.1E-01
	N-Acetyl-L-phenylalanine	4.70	1.8E-05	9.19	2.66	3.46	1.8E-03	8.9E-03	9.8E-02
	O-Acetyl-L-homoserine	1.85	1.7E-03	2.70	1.22	2.21	1.8E-02	7.9E-01	3.0E-02
	4-Acetamidobutanoic acid	2.82	2.0E-05	7.83	3.22	2.43	3.3E-03	1.4E-02	6.8E-02
Glutamate/ Glutamine metabolism	Glutamic acid	2.30	2.0E-04	1.49	1.13	1.32	5.1E-01	6.0E-01	2.3E-01
	Glutamine	2.39	3.1E-05	0.89	1.00	0.89	9.7E-01	7.9E-01	9.4E-01
	Pyroglutamic acid	2.29	3.4E-06	2.78	2.89	0.96	3.3E-03	1.4E-02	8.1E-01
	Phenylacetyl-L-glutamine	1.98	5.0E-02	3.75	1.33	2.81	5.6E-03	3.6E-01	6.6E-03
	g-Glutamyl-glutamine	5.70	1.5E-04	1.94	2.61	0.74	2.2E-01	1.6E-01	8.9E-01
Methionine metabolism	Methionine	7.08	1.7E-05	3.24	2.10	1.54	9.5E-02	7.1E-02	5.0E-01
	Methionine sulfoxide	2.14	7.4E-06	9.03	24.98	0.36	3.0E-03	4.0E-03	9.4E-01
	Formylmethionine	7.24	1.3E-05	3.10	1.47	2.12	9.3E-03	2.6E-01	1.5E-01
	S-Adenosylmethioninamine	1.86	4.8E-03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Phenylalanine metabolism	Phenylalanine	3.04	4.4E-06	2.03	1.74	1.17	6.6E-04	8.0E-06	5.7E-01
	Tyrosine	5.12	3.6E-05	1.95	1.90	1.02	5.6E-03	5.1E-02	8.9E-01
	p-Hydroxyphenyllactic acid	8.42	1.3E-05	7.70	3.79	2.03	3.1E-04	5.6E-03	6.0E-02
	Glutamylphenylalanine	8.82	4.4E-06	4.76	2.40	1.98	3.1E-04	4.8E-05	4.0E-02
Tryptophan metabolism	Tryptophan	3.57	2.0E-05	0.84	0.79	1.05	5.8E-02	5.1E-02	4.1E-01
	Hydroxytryptophan	4.02	1.5E-06	3.68	2.17	1.69	1.5E-02	1.9E-03	2.7E-01
	Indolelactic acid	6.00	5.7E-06	2.37	1.03	2.29	4.6E-01	7.4E-01	2.5E-01
	5-Hydroxyindoleacetic acid	2.60	2.1E-03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	Kynurenine	5.91	2.6E-06	2.82	1.87	1.51	9.3E-03	5.1E-02	2.1E-01
	Aminomuconic acid	1.56	6.8E-05	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Fatty acid metabolism	Carnitine	1.83	4.6E-03	1.58	1.25	1.26	3.8E-01	2.6E-01	6.1E-01
	Acetyl-L-carnitine	2.27	1.6E-03	2.74	2.50	1.10	3.1E-04	5.5E-04	7.7E-01
	Propionylcarnitine	2.48	2.5E-03	1.84	1.11	1.65	8.2E-02	5.1E-01	8.7E-02
	3-Hydroxyisovalerylcarnitine	5.49	1.2E-02	1.96	1.21	1.62	1.3E-01	5.1E-01	2.9E-01
	Octanoylcarnitine	4.10	2.3E-04	1.71	0.90	1.89	3.5E-01	6.0E-01	6.9E-01
	2-Hydroxyvaleric acid	3.85	1.0E-03	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Amino-fatty acids	3-Amino-octanoic acid	3.45	1.6E-05	0.86	0.99	0.87	3.5E-01	3.9E-01	9.8E-01
Energy metabolism	Pyruvic acid	1.71	4.6E-03	0.84	1.60	0.53	8.6E-01	1.6E-01	4.0E-02
	Creatine	0.75	4.6E-03	0.78	1.36	0.57	1.7E-01	1.9E-01	5.0E-01

Research Article

Table 2 (continued)

Biochemical pathway or chemical class	Metabolite	A		B		Plasma			
		Fold change EH vs. CTL	p value Mann Whitney	Fold change EH vs. CTL	Fold change Cir vs. CTL	Fold change EH vs. CIR	Mann Whitney-EH vs. CTL	Mann Whitney-CIR vs. CTL	Mann Whitney-Cir vs. EH
Amino-acids and derivatives	Alanine	0.70	2.4E-03	0.65	0.74	0.87	1.2E-02	2.5E-02	8.5E-01
	Asparagine	1.68	1.0E-03	1.13	1.14	0.99	7.0E-01	3.9E-01	8.5E-01
	Aspartic acid	0.75	4.1E-03	1.54	1.56	0.99	3.8E-01	1.6E-01	1.0E+00
	Histidine	1.89	9.2E-04	0.88	0.91	0.96	5.1E-01	7.9E-01	5.7E-01
	Lysine	0.74	6.3E-03	0.92	0.89	1.03	7.5E-01	6.9E-01	9.8E-01
	Threonine	1.54	2.2E-02	1.55	1.27	1.22	3.5E-01	4.7E-01	6.1E-01
	Cystine	2.38	1.1E-04	2.01	1.73	1.16	7.3E-03	9.6E-02	4.1E-01
	Stachydrine	0.64	3.3E-03	1.78	1.08	1.64	4.6E-01	1.0E+00	4.7E-01
	Citrulline	1.88	3.3E-03	1.25	0.93	1.35	8.2E-02	6.5E-01	4.6E-02
	Taurine	0.43	1.1E-05	0.64	0.62	1.04	5.6E-03	1.9E-03	8.9E-01
Peptides	1-L-Leucyl-L-Proline	2.31	5.9E-04	4.58	2.94	1.56	1.3E-04	1.4E-03	8.7E-02
	Leu-Ala	5.45	3.4E-06	10.92	5.41	2.02	6.8E-06	8.0E-06	1.4E-01
Bile acids	Glycoursodeoxycholic acid OR isomers	93.88	4.8E-07	42.00	30.01	1.40	6.8E-06	7.1E-03	2.7E-01
	Glycocholic acid	241.78	5.7E-06	118.60	54.55	2.17	1.4E-05	5.5E-04	1.9E-01
	Taurocholic acid	3e+5	n.a.	332.60	109.77	3.03	6.8E-06	1.6E-05	3.2E-01
Nucleosides and derivatives	Dimethylguanosine	2.58	5.3E-05	4.38	3.23	1.36	6.8E-06	2.8E-05	9.8E-02
	Pseudouridine	1.72	1.0E-02	2.77	1.74	1.59	9.4E-04	5.1E-02	1.4E-01
	Succinyladenosine	1.73	6.7E-05	5.16	2.33	2.21	3.1E-04	1.4E-02	7.7E-02
Pyrimidine metabolism	Dihydrothymine	3.45	6.5E-06	2.85	2.44	1.17	8.2E-05	1.2E-04	6.9E-01
	Uracil	0.78	4.1E-02	0.51	0.52	0.98	3.1E-04	2.7E-04	8.9E-01
Monosaccharide	Anhydro-sorbitol	0.44	1.7E-04	0.37	0.26	1.39	1.3E-04	4.0E-06	3.5E-01
Alcohols and polyols	Quinic acid	0.43	1.7E-03	0.39	0.08	5.06	4.6E-01	2.5E-02	1.2E-01
	Dehydroquinic acid	2.34	4.3E-06	3.86	1.51	2.56	1.5E-02	1.6E-01	3.0E-02
	Gluconic acid	1.57	1.9E-03	2.63	2.36	1.11	1.1E-01	6.9E-01	4.1E-01
	Glyceric acid	1.43	3.1E-03	1.19	0.96	1.24	8.2E-02	7.4E-01	1.5E-01
	Xylonic acid OR isomers	2.25	6.8E-05	1.51	1.06	1.42	9.5E-02	1.0E+00	8.7E-02
Sugar alcohols	Adonitol/Arabitol	1.49	3.3E-02	4.21	1.40	3.00	3.4E-02	3.2E-01	5.2E-02
Steroid	Cortisol	3.55	1.6E-05	5.74	3.80	1.51	1.3E-04	1.8E-04	1.5E-01
Vitamin B6 metabolism	Pyridoxic acid	15.93	1.3E-04	108.75	71.94	1.51	9.4E-04	5.5E-04	3.5E-01
Xanthine metabolism	Theophylline/Paraxanthine	0.41	9.0E-03	0.56	0.47	1.20	1.3E-01	1.3E-01	7.3E-01
Miscellaneous	Levulinic acid OR isomers	0.82	3.8E-02	0.75	0.86	0.87	5.8E-02	3.9E-01	7.3E-01
	Phenol sulfate	13.69	3.5E-02	6.82	2.64	2.59	3.3E-03	3.9E-01	7.7E-02

Values are ratios of individual concentrations observed for HE patients to mean levels obtained for controls, except for taurocholic acid, which was not detected in the control group and for which values are areas of LC/HRMS peaks. Values with yellow, orange and red colors are above mean levels for controls plus one, two and three standard deviations, respectively. Values in blue are between mean levels for controls minus two and one standard deviations. *p* value have been obtained from Mann-Whitney test. NA: not available. ND: not determined.

Metabolites involved in energy metabolism (Table 2A and Fig. 2)

Concentrations of carnitine, acylcarnitines, hydroxy organic acids such as 2-hydroxyvaleric and 3-hydroxy-2-methylbutanoic acids (which are involved in the metabolism of fatty acids), and pyruvic acid were significantly increased, whereas the levels of detected Krebs cycle intermediates such as fumaric, malic and citric acids were unchanged. Creatine concentrations were significantly decreased in HE patients.

Acetylated metabolites

Concentrations of 13 acetylated compounds were increased, including acetyl-cytidine, acetyl-leucine, acetyl-valine, acetyl-alanine, acetyl-serine and acetamidobutanoic acid (a mono-

acetylation product of GABA), whereas the concentrations of the corresponding non-acetylated compounds were unchanged (serine, leucine and GABA), decreased (alanine) or not detected (cytidine).

Metabolism of AAAs

Concentrations of phenylalanine, tyrosine, and two other metabolites related to phenylalanine metabolism were increased in HE patients (Table 2A). Furthermore, the levels of tryptophan and 5 metabolites involved in its metabolic pathways were higher in HE patients than in controls. Two of these metabolites, indoleacetic acid and 5-hydroxyindoleacetic acid, are enzymatic degradation products of serotonin.

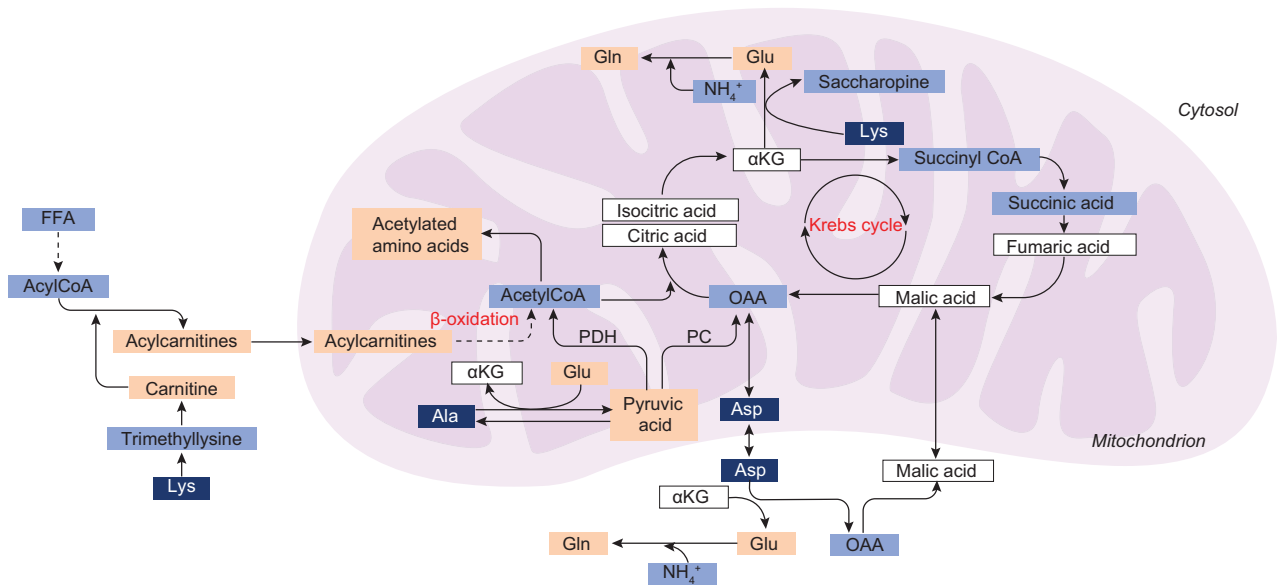


Fig. 2. Altered energy metabolism pathways in CSF samples of HE patients. Orange box: increased metabolite levels compared with the control group. Dark blue box: decreased metabolite levels compared with the control group. White box: metabolites that do not show any concentration increase compared with the control group. Light blue box: not detected.

Metabolites of the methyl cycle

Five metabolites involved in the methyl cycle were found at abnormal concentrations: methionine, methionine sulfoxide and formyl-methionine were increased by factors of 7.1, 2.1 and 7.2, respectively. Unfortunately, our experimental conditions made the detection of S-adenosyl methionine (SAM) impossible, but S-adenosyl-methioninamine (SAMN) and 5'-deoxy-5'-(methylthio)-adenosine, which are two other metabolites produced from SAM, were detected. The methionine to SAMN concentration ratios in HE patients and controls were 4 and 0.7%, respectively.

Presence of drugs and metabolites in CSF samples of HE patients

Interestingly and contrary to usual situations, differences between the chromatograms of HE and control patients could be observed directly by visual inspection of the LC-MS chromatograms. These can be attributed to large amounts of metabolites that are present in HE patients and absent in controls (Fig. 3A). Some of these high intensity features have been annotated by public metabolomic databases and could correspond to drugs, on the basis of interpretation of the mass spectra (presence of in-source fragment ions, and/or isotopes such as ³⁷Cl or ³⁴S for some of them, as shown in Supplementary Table 2). This is the case for metronidazole (BG9 sample), levetiracetam (BM2 sample), fluconazole (MGH9, NQ1, AZO and ZG8 samples) (Fig. 3B), diazepam and its metabolite N-desmethyldiazepam (BC8, MGH9, MP01, NQ1, ZG8 samples), piperacillin and tazobactam, its associated β-lactamase inhibitor (BG9, NQ1, ZG8 samples), norfloxacin (CJ3 and BC8 samples) and ciprofloxacin (AZO, BG9, ZG8, NQ1, SP3 samples). Overall, features putatively related to drugs and their metabolites were detected in 10 out of the 14 HE samples (Fig. 3C) and assessed by clinical metadata. Lastly, a significant correlation ($r_s = 0.7082, p = 0.0268$) between the number of drugs detected in CSF and the total concentration of plasma bilirubin was observed in patients, as displayed in Fig. 3D.

Correlations between metabolites of the HE metabolite and clinical scores or biological parameters

The results displayed in Supplementary Table 3 indicate that 26 out of 73 statistically significant metabolites correlated with clinical scores and biochemical parameters related to liver disease and HE. Correlations with MELD, the Child-Pugh score, bilirubinemia, prothrombin time and creatininemia were statistically significant for 13, 6, 12, 9 and 4 metabolites, respectively. Most of these metabolites are part of the methionine, phenylalanine, tryptophan, bile acid and fatty acid metabolism pathways.

Lysine and gluconic acid were the only metabolites whose concentrations correlated significantly with ammonemia. Surprisingly, although creatininemia is involved in the calculation of the MELD score, there was no overlap between metabolites correlating with creatininemia and those correlating with other parameters indicative of liver disease. Lastly, regarding the severity of HE, the levels of only three metabolites, namely carnitine, 5-hydroxyindoleacetic acid and uracil, were significantly positively and negatively correlated with the WH score and Glasgow coma scale (GCS) score, respectively, without being correlated with any liver function parameters, even MELD (Supplementary Table 3).

Evaluation of the impact of cirrhosis on the HE metabolite

As it was not possible to include CSF samples collected from cirrhotic patients due to ethical reasons, the impact of cirrhosis on the HE metabolite was assessed using plasma samples, collected from 9 healthy controls, 13 cirrhotic without HE, and 12 cirrhotic patients with HE patients. Patient characteristics are displayed in Supplementary Table 4.

The metabolic signature obtained from plasma samples is displayed in the Table 2B. Except for 6 metabolites, most of the significant metabolites of the HE metabolite in CSF were also

Research Article

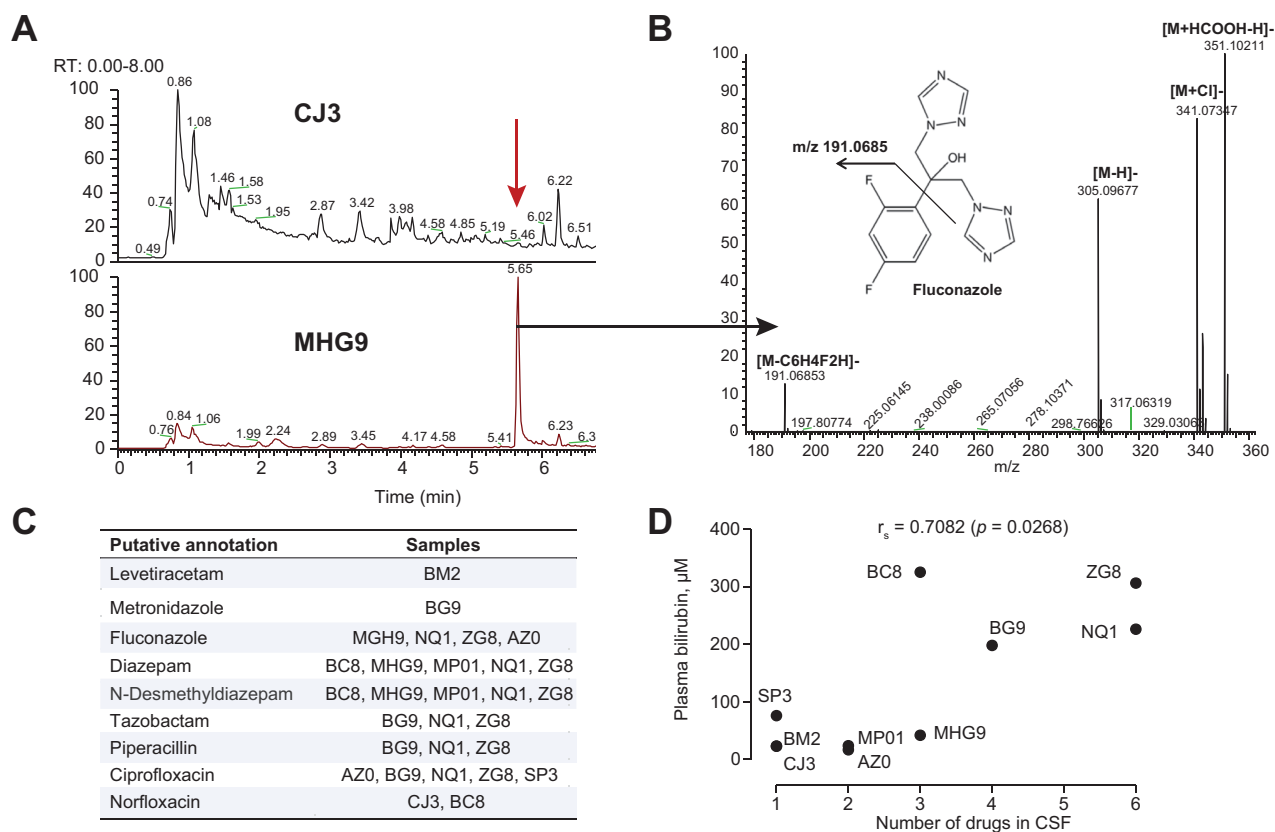


Fig. 3. Detection of drugs in CSF samples of HE patients. (A) Total ion current chromatogram recorded in the negative electrospray ionization (ESI) mode and displaying a high intensity peak at 5.7 min in the MHG9 sample, but not in the CJ3 sample. (B) ESI-mass spectrum related to the chromatographic peak. The ion at m/z 305.0968 is putatively annotated as fluconazole. This is reinforced by the co-elution of m/z 191.0685, which could correspond to a fragment ion (loss of the $C_6H_4F_2$ moiety), and m/z 341.0735 and 351.1015, which could correspond to chloride and formic acid adducts, respectively. All these ions are co-eluted and their intensities are correlated with each other. An ion annotated as fluconazole (m/z 307.1109) is also detected in the positive ion mode in reverse phase conditions at the same retention time. (C) Occurrence of the detected drug in HE patients. (D) Correlation of the number of detected drugs and the total bilirubin concentration measured in HE patients ($n = 10$). (This figure appears in colour on the web.)

detected in all plasma samples. Our results indicate that the same concentration trends are observed for all detected metabolites in plasma and CSF samples of HE patients. Compared to controls, 37 metabolites exhibited statistically significant concentration variations in plasma of cirrhotic patients with HE, and 26 metabolites were significantly altered in cirrhotic patients without HE. Thus, the levels of 11 plasma metabolites were specifically and significantly modified in HE patients compared to patients without HE (i.e., two acetylated derivatives, phenylacetylglutamine, formylmethionine, tyrosine, kynurenine, cysteine, pseudouridine, dehydroquinic acid, adonitol/arabitol and phenol sulfate).

Contrary to what is observed in CSF, glutamine, glutamic acid and methionine levels are not significantly altered in plasma samples of HE and cirrhotic patients without HE. Some metabolites such as those involved in phenylalanine metabolism, peptides, bile acids and nucleosides, displayed statistically significant concentration variations in plasma samples of all cirrhotic patients. At last, concentrations of metabolites involved in fatty acid and tryptophan metabolism, and also amino acids are far less impacted by HE and cirrhosis in plasma than in CSF samples of HE patients. Indeed, only 3 out of the detected 10 amino-acid derivatives exhibited statistically significant concentration variations in plasma samples. Furthermore, pyruvic acid

and creatine concentrations were not affected by HE and cirrhosis in plasma samples.

Discussion

Although CSF is only an approximation of extracellular fluid inside the brain, it is much easier to get than brain extracellular fluids, despite the major difficulties to obtain it in HE cirrhotic patients. Thus, our study emphasizes the relevance of using LC/HRMS-based metabolomics in order to provide a global picture of altered metabolic pathways in CSF samples of HE patients. This technique also enables direct exploration in HE patients of pathophysiological hypotheses previously only studied in animal models. Thus, metabolites of HE patients display altered levels of 73 metabolites of several major metabolite classes such as AAAs, which reflect alteration of the BBB and which are also involved in neurotransmission, or acylcarnitines, which are part of energy metabolism pathways. As shown in Table 3, the involvement of some of these metabolites in the development of HE is already well documented, as is the case with glutamic acid, glutamine, AAAs, methionine and bile acids. Thus, our study confirms these previous findings, but, more than that, our global metabolomic

Table 3. Comparison of the findings of this study with those of the literature.

Metabolic pathway or chemical class	Findings of this study	Previous findings	References
Glutamic acid and glutamine metabolism	Increased CSF concentrations of glutamine, glutamate and their derivatives: pyroglutamic acid, phenylacetylglutamine and glutamyl-glutamine dipeptide.	- Increased glutamine and glutamic acid levels in HE patients. - Phenylacetylglutamine accumulates in plasma samples of uremic patients, as a consequence of the alteration of nitrogen metabolism. - Gamma-glutamyl-glutamine has been detected in CSF samples of hyperammonemic patients.	[16] [43] [44]
AAAs and derivatives	Increased CSF concentrations of phenylalanine, tryptophan and derivatives: tyrosine, hydroxyphenyllactic acid, hydroxytryptophan, indoleacetic acid, 5-hydroxyindoleacetic acid, kynurenine and aminomuconic acid.	- Increased CSF concentrations of AAAs in HE patients. - Increased active transport of AAAs through the BBB in HE patients. - Urinary concentrations of hydroxyphenyllactic are increased in cirrhotic patients with HE.	[14] [15] [45]
Methionine and derivatives	Increased CSF concentrations of methionine, formyl-methionine, methionine sulfoxide and S-adenosylmethioninamine (SAMN).	- Increased CSF and plasma concentrations of methionine in HE patients.	[46]
Bile acids	Dramatic increases of CSF levels that could be explained by an increased BBB permeability in HE patients.	- Elevated concentrations of bile acids and particularly taurine conjugates in serum and CSF samples of hepatic failure patients.	[18]
Taurine	Decreased CSF concentrations in HE patients.	Conflicting results: - decreased taurine content in central nervous system (i.e., brain or CSF) in humans and rats. - increased in CSF samples in acute liver failure early in the progression of HE and prior to the onset of cerebral edema The osmoregulatory role of taurine in the pathophysiology of HE is not fully established.	[47-49] [6]

approach improves the mapping of metabolic pathways associated with these metabolites, as shown for AAA and ammonia metabolism, for example.

Our results also provide new findings regarding the alteration of brain energy metabolism in HE by investigating what happens in patients. The alteration of energy metabolism in HE has been mainly investigated using animal (rats or pigs) or cell culture (neurons) models [23]. By measuring the concentrations of intermediary metabolites, ammonia and amino acids in rat brain and plasma, and also the rate of whole-brain glucose utilization, Mans *et al.* showed that glucose was less used by the brain in rat that underwent portacaval shunting and hepatic artery ligation. However, they did not find any alterations in brain levels of metabolites of the glycolytic and Krebs pathways, except for pyruvate and lactate [24]. Furthermore, Shorey *et al.* observed that brain production from the Krebs cycle was reduced in animal models of HE [25], and the study of Muntz and Hurwitz showed that glycolysis was stimulated by hyperammonemia in rat brain [26]. Finally, Rama Rao *et al.* reported that α KG was used for the detoxification of ammonia, which could result in a blockage of the Krebs cycle at the level of its rate-limiting enzyme (i.e., α KG dehydrogenase) [23].

These previous studies are in line with our observations. Indeed, as shown in Fig. 2, we found unchanged concentrations of Krebs cycle intermediates in CSF, but increased concentrations of pyruvic acid, acylcarnitines and acetylated compounds. The dramatic increase of carnitine and acylcarnitine levels may reflect the reduced activity of the Krebs cycle, possibly associated with a compensatory increased activity of the fatty acid β -oxidation pathway. Although it is usually believed that β -oxidation is marginal in the central nervous system, as opposed to what occurs in the muscle and in the liver, recent studies show that β -oxidation

accounts for up to 20% of brain energy production in certain circumstances, such as fasting [27]. Most of the β -oxidation occurs in astrocytes, which are affected by hyperammonemia as they are responsible for the synthesis of glutamine from glutamate and α KG. Similarly, the increase in acetylated metabolites could be explained by an increased availability of acetylCoA produced by both pyruvate oxidation and β -oxidation, as a consequence of Krebs cycle impairment.

Otherwise, increased concentrations of carnitine and of some acylcarnitines have also been observed in plasma, although the concentration variations between control and HE patients were not statistically significant, except for acetylcarnitine (Table 2B). The data regarding carnitine and acylcarnitines in cirrhotic patients are controversial. Indeed, Tang *et al.* observed decreased levels of carnitines, such as propionylcarnitine and butyrylcarnitine with the progression of primary biliary cirrhosis. However, increased levels of acylcarnitines have been detected in plasma of cirrhotic patients, whereas carnitine levels were unchanged [28,29].

Interestingly, carnitine and acetylcarnitine supplementation has been reported to prevent ammonia toxicity and to improve its elimination in patients with hyperammonemic syndromes [30], and more particularly in patients with cirrhosis [31] and HE patients [32,33]. Increased carnitine concentrations in the CSF could be explained by competitive inhibition of transport proteins by drugs or bile acids [34] which are accumulated in CSF of HE patients, or by a compensatory increase linked to Krebs cycle impairment as discussed above.

The decrease of creatine in CSF samples of HE patients is another interesting feature which is reported for the first time. Interestingly, decreased levels of brain creatine have been observed in a rat model of chronic HE [35]. Furthermore, data

Research Article

obtained on a cell culture model indicate that axonal development is affected by ammonia, and that it can be normalized by creatine supplementation [36].

The contribution of energy metabolism to the physiopathology of HE has also been highlighted by the work of Jimenez *et al.*, which aimed at establishing a serum metabolic signature of minimal HE using nuclear magnetic resonance spectroscopy [37]. Although it is difficult to compare this work with our study because it did not focus on overt HE, some common concentration trends were observed, such as decreased concentrations of alanine (Table 2B), leucine and valine (data not shown). Otherwise, glucose, lactate, glycerol, choline, acetoacetate and glycine, whose levels were significantly altered in minimal HE patients, could not be detected using our method.

Our results also indicate that concentrations of 23 metabolites were influenced by the severity of liver disease, as assessed by the Child-Pugh and MELD scores, but also by total plasma bilirubin and creatinine concentrations, and prothrombin time. For some metabolites like bile acids, this is understandable, as patients with severe liver disease have increased bile acid concentrations in blood, provided there is an alteration of BBB permeability. However, these correlation data suffer from the small size of the cohort and have to be confirmed by others studies performed both on plasma and CSF samples. Furthermore, future studies could also be undertaken in order to assess whether or not it is possible to stratify the patients by degree of hepatic failure, cerebral dysfunction and also by the presence/absence of significant portal-systemic shunt according to the information contained in the metabolic fingerprints.

The specificity of the HE metabotype regarding cirrhosis was an important point to investigate. Our data show that HE patients exhibit similar but amplified metabolic disorders as those already observed on cirrhotic patients. However, the metabolic signature obtained in CSF is more complete than that obtained in plasma samples, regarding glutamine, methionine and tryptophan metabolism, for example. Furthermore, alterations of amino-acid, acyl-carnitines (which are involved in beta-oxidation of fatty acids), and polyol metabolisms were only observed in CSF, and not in plasma samples.

Otherwise, a NMR based metabolomic study aimed at distinguishing cirrhotic from control and HE patients has been published [38]. The authors established a 5 metabolite signature that could separate HE and cirrhotic patients, based on concentrations of tyrosine, phenylalanine, methionine, pyruvic acid and glycine, but which was not confirmed by our data.

Several drugs that are not all classically known to be neurotoxic or to trigger HE in cirrhotic patients were detected in the CSF. Chromatograms in 10 out of the 14 HE patients indicated drugs such as metronidazole, levetiracetam, fluconazole, diazepam and its metabolite N-desmethyldiazepam, piperacillin-tazobactam, norfloxacin and ciprofloxacin, which are more prone to accumulate in CSF samples of patients with severe liver disease (Fig. 2D). Although such drug accumulations could be explained by decreased liver metabolic activity (at least for metronidazole), they could also be at least partially explained by down-regulation of the expression of P-glycoprotein (P-gp) and Mrp1 (two of the main multidrug resistance proteins involved in drug efflux outside the brain) induced by exposure to bilirubin, as already observed in animals [39]. Moreover, several of the detected drugs, which are frequently indicated in cirrhotic patients, can decrease P-gp activity at the BBB by partial agonism, as is the case

with metronidazole, fluconazole or penicillins [40,41]. Thus, some of our HE patients could have some neurological symptoms explained by drug accumulation in the brain due to decreased expression of P-gp secondary to elevated concentrations of bilirubin or bile acids. Interestingly, sample ZG8, which had high concentrations of bile acids with very high bilirubinemia (306 μ M), also had elevated concentrations of P-gp agonists, suggesting drug-induced HE in this patient. Taken together, our data suggest that metabolomics could be an attractive approach to monitor accumulations of drugs in CSF samples of cirrhotic patients having neurological symptoms.

Our study presents some limitations. First of all, and as previously mentioned, correlations between the HE metabolic signature and clinical scores are limited by the small size of the group, due to ethical issue and stringent criteria to perform lumbar puncture outside an etiological work-up. Furthermore, the relevance of these correlations and also additional insights into the mechanisms involved in HE will have to be evaluated on animal and cellular models. At last, the specificity of the HE metabolic signature will also have to be investigated by analyzing biofluids of patients having liver diseases other than cirrhosis and without HE. This will be the aim of further studies.

In conclusion, metabotypes of HE patients display alterations in several major metabolite classes. Increased concentrations of pyruvic acid and acetylated metabolites, which have never been reported before in CSF to our knowledge, reinforce the hypothesis of an alteration of the activity of the Krebs cycle in HE patients. This is of major importance, as there are few treatments approved for HE, all of which act by decreasing blood ammonia levels, and none selectively address abnormalities inside the brain. The efficacy of those treatments is far from satisfactory, with 50% of patients considered as responders [42]. Hence, new therapeutic targets are mandatory. In that regard, CSF metabolomics could be a useful tool to assess the efficacy of new therapeutic strategy that target energy metabolism.

Financial support

This work was supported by the Commissariat à l'Energie Atomique, Assistance Publique des Hôpitaux de Paris, Medday Pharmaceuticals, the MetaboHUB infrastructure (ANR-11-INBS-0010 grant) and by a grant from the French Ministry of Health (P.H.R.C. "SPECMET2", grant number AOM 10300).

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Authors' contributions

FS, DT, NW, BC and CJ initiated the project. FS, DT, NW and MR procured the samples. NW, AS, FI, FS, MDMA, FL, MR and DT provided clinical and biochemical metadata. FI, PBS and BC performed metabolomic analyses. SA and PBS performed MS/MS experiments for metabolite identification. FI, PBS, SA, BC and CJ

performed data processing and statistical analyses. NW, BC, FL, MDMA, FL, FS, DT and CJ interpreted the data. NW, FS, DT and CJ wrote the manuscript. This manuscript has been seen and approved by all authors.

Acknowledgements

The authors would like to thank Professor Jean-Marie Saudubray for his help regarding the metabolic pathways displayed in Fig. 2.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jhep.2016.07.046>.

References

[1] Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K, Blei AT. Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* 2002;35:716–721.

[2] Inserm – Dossiers d’information. La cirrhose. 2012.

[3] Garcia-Pagan JC, Caca K, Bureau C, Laleman W, Appenrodt B, Luca A, et al. Early use of TIPS in patients with cirrhosis and variceal bleeding. *N Engl J Med* 2010;362:2370–2379.

[4] Romero-Gomez M, Montagnese S, Jalan R. Hepatic encephalopathy in patients with acute decompensation of cirrhosis and acute-on-chronic liver failure. *J Hepatol* 2015;62:437–447.

[5] Rovira A, Alonso J, Cordoba J. MR imaging findings in hepatic encephalopathy. *AJNR Am J Neuroradiol* 2008;29:1612–1621.

[6] Poh Z, Chang PE. A current review of the diagnostic and treatment strategies of hepatic encephalopathy. *Int J Hepatol* 2012;2012:480309.

[7] Blei AT, Cordoba J. Hepatic encephalopathy. *Am J Gastroenterol* 2001;96:1968–1976.

[8] Ong JP, Aggarwal A, Krieger D, Easley KA, Karafa MT, Van LF, et al. Correlation between ammonia levels and the severity of hepatic encephalopathy. *Am J Med* 2003;114:188–193.

[9] Shawcross DL, Shabbir SS, Taylor NJ, Hughes RD. Ammonia and the neutrophil in the pathogenesis of hepatic encephalopathy in cirrhosis. *Hepatology* 2010;51:1062–1069.

[10] Shawcross DL, Wright G, Olde Damink SW, Jalan R. Role of ammonia and inflammation in minimal hepatic encephalopathy. *Metab Brain Dis* 2007;22:125–138.

[11] Shawcross DL, Davies NA, Williams R, Jalan R. Systemic inflammatory response exacerbates the neuropsychological effects of induced hyperammonemia in cirrhosis. *J Hepatol* 2004;40:247–254.

[12] Coltart I, Tranah TH, Shawcross DL. Inflammation and hepatic encephalopathy. *Arch Biochem Biophys* 2013;536:189–196.

[13] Prakash R, Mullen KD. Mechanisms, diagnosis and management of hepatic encephalopathy. *Nat Rev Gastroenterol Hepatol* 2010;7:515–525.

[14] Cascino A, Cangiano C, Fiaccadori F, Ghinelli F, Merli M, Pelosi G, et al. Plasma and cerebrospinal fluid amino acid patterns in hepatic encephalopathy. *Dig Dis Sci* 1982;27:828–832.

[15] Watanabe A, Higashi T, Shiota T, Takei N, Sakata T, Fujiwara M, et al. Alterations in neutral amino acid transport across the blood-brain barrier in hepatic failure. *Res Exp Med (Berl)* 1983;183:11–18.

[16] Vergara F, Plum F, Duffy TE. Alpha-ketoglutarate: increased concentrations in the cerebrospinal fluid of patients in hepatic coma. *Science* 1974;183:81–83.

[17] Borg J, Warter JM, Schlienger JL, Imler M, Marescaux C, Mack G. Neurotransmitter modifications in human cerebrospinal fluid and serum during hepatic encephalopathy. *J Neurol Sci* 1982;57:343–356.

[18] Bron B, Waldram R, Silk DB, Williams R. Serum, cerebrospinal fluid, and brain levels of bile acids in patients with fulminant hepatic failure. *Cut* 1977;18:692–696.

[19] Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. *Nature* 2012;491:384–392.

[20] Child DL, Mathews JA, Thompson RP. Arthritis and primary biliary cirrhosis. *Br Med J* 1977;2:557.

[21] Kamath PS. Scoring severity of chronic liver disease. *J Gastroenterol Hepatol* 2002;17:S298–S299.

[22] Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen KD, et al. Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. *Hepatology* 2014;60:715–735.

[23] Rama Rao KV, Norenberg MD. Brain energy metabolism and mitochondrial dysfunction in acute and chronic hepatic encephalopathy. *Neurochem Int* 2012;60:697–706.

[24] Mans AM, DeJoseph MR, Hawkins RA. Metabolic abnormalities and grade of encephalopathy in acute hepatic failure. *J Neurochem* 1994;63:1829–1838.

[25] Shorey J, McCandless DW, Schenker S. Cerebral alpha-ketoglutarate in ammonia intoxication. *Gastroenterology* 1967;53:706–711.

[26] Muntz JA, Hurwitz J. Effect of potassium and ammonium ions upon glycolysis catalyzed by an extract of rat brain. *Arch Biochem Biophys* 1951;32:124–136.

[27] Jones LL, McDonald DA, Borum PR. Acylcarnitines: role in brain. *Prog Lipid Res* 2010;49:61–75.

[28] Amodio P, Angeli P, Merkel C, Menon F, Gatta A. Plasma carnitine levels in liver cirrhosis: relationship with nutritional status and liver damage. *J Clin Chem Clin Biochem* 1990;28:619–626.

[29] Krahenbuhl S, Reichen J. Carnitine metabolism in patients with chronic liver disease. *Hepatology* 1997;25:148–153.

[30] O’Connor JE, Costell M. New roles of carnitine metabolism in ammonia cytotoxicity. *Adv Exp Med Biol* 1990;272:183–195.

[31] Cecere A, Ciaramella F, Tancredi L, Romano C, Gattoni A. Efficacy of L-carnitine in reducing hyperammonaemia and improving neuropsychological test performance in patients with hepatic cirrhosis: results of a randomised trial. *Clin Drug Invest* 2002;22:7–14.

[32] Malaguarnera M, Pistone G, Elvira R, Leotta C, Scarpello L, Liborio R. Effects of L-carnitine in patients with hepatic encephalopathy. *World J Gastroenterol* 2005;11:7197–7202.

[33] Jiang Q, Jiang G, Shi KQ, Cai H, Wang YX, Zheng MH. Oral acetyl-L-carnitine treatment in hepatic encephalopathy: view of evidence-based medicine. *Ann Hepatol* 2013;12:803–809.

[34] Skowronska M, Albrecht J. Alterations of blood brain barrier function in hyperammonemia: an overview. *Neurotox Res* 2012;21:236–244.

[35] Rackayova V, Braissant O, McLin VA, Berset C, Lanz B, Cudalbu C. H and P magnetic resonance spectroscopy in a rat model of chronic hepatic encephalopathy: in vivo longitudinal measurements of brain energy metabolism. *Metab Brain Dis* 2015.

[36] Bachmann C. Mechanisms of hyperammonemia. *Clin Chem Lab Med* 2002;40:653–662.

[37] Jimenez B, Montoliu C, MacIntyre DA, Serra MA, Wassel A, Jover M, et al. Serum metabolic signature of minimal hepatic encephalopathy by (1)H-nuclear magnetic resonance. *J Proteome Res* 2010;9:5180–5187.

[38] Dabos KJ, Parkinson JA, Sadler IH, Plevris JN, Hayes PC. (1)H nuclear magnetic resonance spectroscopy-based metabonomic study in patients with cirrhosis and hepatic encephalopathy. *World J Hepatol* 2015;7:1701–1707.

[39] Gazzin S, Berengeno AL, Strazielle N, Fazzari F, Raseni A, Ostrow JD, et al. Modulation of Mrp1 (ABCC1) and Pgp (ABCB1) by bilirubin at the blood-CSF and blood-brain barriers in the Gunn rat. *PLoS One* 2011;6:e16165.

[40] Keppler D. Multidrug resistance proteins (MRPs, ABCs): importance for pathophysiology and drug therapy. *Handb Exp Pharmacol* 2011;201:299–323.

[41] Siniscalchi A, Mancuso F, Scornaienghi D, Ferreri G, De SG. Acute encephalopathy induced by oxcarbazepine and furosemide. *Ann Pharmacother* 2004;38:509–510.

[42] Bass NM, Mullen KD, Sanyal A, Poordad F, Neff G, Leevy CB, et al. Rifaximin treatment in hepatic encephalopathy. *N Engl J Med* 2010;362:1071–1081.

[43] Zimmerman L, Egestad B, Jornvall H, Bergstrom J. Identification and determination of phenylacetylglutamine, a major nitrogenous metabolite in plasma of uremic patients. *Clin Nephrol* 1989;32:124–128.

[44] Hammond JW, Potter M, Truscott R, Wilcken B. Gamma-Glutamylglutamine identified in plasma and cerebrospinal fluid from hyperammonaemic patients. *Clin Chim Acta* 1990;194:173–183.

[45] Muting D, Wuzel H, Bucsis L, Flasshoff HJ. Urinary p-hydroxyphenyllactic acid as indicator of hepatic encephalopathy in patients with hepatic cirrhosis. *Lancet* 1985;2:1365–1366.

Research Article

- [46] Higashi T. Impaired metabolism of methionine in severe liver diseases. II. Clinical and experimental studies on role of impaired methionine metabolism in pathogenesis of hepatic encephalopathy. *Gastroenterol Jpn* 1982;17:125–134.
- [47] Albrecht J, Zielinska M. The role of inhibitory amino acidergic neurotransmission in hepatic encephalopathy: a critical overview. *Metab Brain Dis* 2002;17:283–294.
- [48] Cordoba J, Gottstein J, Blei AT. Glutamine, myo-inositol, and organic brain osmolytes after portocaval anastomosis in the rat: implications for ammonia-induced brain edema. *Hepatology* 1996;24:919–923.
- [49] Butterworth RF. Taurine in hepatic encephalopathy. *Adv Exp Med Biol* 1996;403:601–606.

4.1 Principaux résultats

Cent vingt-deux métabolites ont pu être détectés dans le liquide céphalo-rachidien, parmi lesquels les concentrations de 72 étaient modifiées chez les patients souffrant d'EH. L'ensemble de ces métabolites permettent une distinction claire entre les encéphalopathes et les volontaires sains. Parmi les métabolites impactés par l'EH, se trouvent par exemple des acides aminés aromatiques impliqués dans la neurotransmission ou des acylcarnitines jouant un rôle dans le métabolisme énergétique. La voie du métabolisme de l'ammoniaque apparaît également impactée. Nous avons mis en évidence l'absence de modification de la concentration en intermédiaires du cycle de Krebs dans le LCR avec une augmentation des concentrations en acide pyruvique, en acylcarnitines et en composés acétylés. L'augmentation importante des concentrations de carnitine et d'acylcarnitines semble refléter une diminution de l'activité du cycle de Krebs, potentiellement compensée par une augmentation de la β -oxydation. Cette hypothèse est appuyée par le fait que la β -oxydation au niveau cérébral ait principalement lieu au sein des astrocytes qui sont les cellules les plus affectées par l'hyperamoniémie. L'accumulation de composés acétylés serait liée à l'augmentation de la disponibilité en acetylCoA produit à la fois par l'oxydation du pyruvate et la β -oxydation et moins consommé du fait de la baisse d'activité du cycle de Krebs. La contribution du métabolisme énergétique dans la physiopathologie de l'EH est mise en évidence pour la première fois chez l'homme dans le cadre de l'encéphalopathie hépatique.

Tous les patients souffrant d'encéphalopathie hépatique de type C sont par définition à un stade avancé de cirrhose. Il a donc fallu s'assurer que les variations observées entre les populations de volontaires sains et d'encéphalopathes étaient bien liées à l'EH et non uniquement à la cirrhose. Nous avons donc souhaité inclure un groupe de patients cirrhotiques non encéphalopathes dans l'étude. Or, en raison de la perte de fonction du foie (lieu de synthèse des facteurs de coagulation), les patients cirrhotiques présentent des troubles de la coagulation qui rendent plus risqué le recueil de LCR par ponction lombaire. Il n'a donc pas été possible, d'un point de vue éthique, de recueillir du LCR de patients cirrhotiques sans indication médicale neurologique. Pour cette seconde cohorte, des échantillons de plasma ont donc été recueillis et analysés.

Dans le cadre de cette première étude, nous nous sommes focalisés sur les métabolites identifiés comme discriminants dans le LCR. 66 des 72 métabolites modifiés dans le LCR étaient également modifiés dans le plasma et tous suivaient la même tendance. Cependant, du fait de la moins grande spécificité du plasma envers les dérèglements ayant lieux au niveau du cerveau, seuls 37 étaient également significativement modifiés dans le plasma. Parmi eux, 13 s'avèrent être des marqueurs spécifiques de l'EH au niveau plasmatique.

4.2 Conclusion de l'étude

Les profils métaboliques de patients souffrant d'encéphalopathie hépatique présentent des modifications spécifiques, notamment au niveau du LCR. L'augmentation des concentrations d'acides aminés aromatiques impliqués dans la neurotransmission et des acylcarnitines impliquées dans le métabolisme énergétique, font partie de ces modifications. L'augmentation des taux d'acide pyruvique et des composés acétylés pourrait signer une altération du fonctionnement du cycle de Krebs associé à une accumulation d'acétylCoA réactif. Cette altération du métabolisme énergétique est un élément nouveau pour la compréhension de la maladie. Sa pertinence doit maintenant être étudiée plus en profondeur, notamment sur des cohortes de plus grande taille. L'utilisation de modèles animaux pourrait également être un moyen de valider ces premières conclusions. Cela permettrait de réaliser des analyses métabolomiques sur des tissus, notamment au niveau du cerveau. Il serait ainsi possible de vérifier la concordance de la signature observée dans le LCR avec celle présente au sein même du cerveau. La confirmation des conclusions de cette première étude pourrait s'avérer être une piste intéressante pour développer de nouveaux traitements.

5 Seconde étude

A la suite de ces premiers résultats obtenus à partir de LCR, une seconde cohorte a été montée afin de vérifier la spécificité de la signature observée vis-à-vis d'une signature de cirrhotique. Etant donné les considérations éthiques chez les cirrhotiques, des prélèvements de plasma ont été utilisés. Lors de la première étude, seuls les 66 métabolites présents dans le LCR et dans le plasma ont été suivis. Cette seconde étude, qui se place dans la continuité directe de la précédente, reprend la cohorte étudiée ci-dessus mais cette fois de manière plus complète. L'ensemble des signaux identifiés en métabolomique est pris en considération et plusieurs analyses « -omiques » sont combinées : métabolomique, lipidomique et glycomique. Le but est donc de continuer à approfondir la compréhension de l'EH, et également de chercher à combiner les résultats de ces différents types d'analyses.

5.1 Principaux résultats

Le premier élément que nous avons voulu suivre est le recouvrement des couvertures métaboliques entre le plasma et le LCR. La première étude a permis d'identifier 122 métabolites dans le LCR. Au niveau du plasma, 342 métabolites ont pu être identifiés. Cette différence de couverture tient à deux éléments. Le premier est instrumental. En effet, les deux études n'ont pas été menées sur le même instrument (Exactive pour le LCR et Fusion pour le plasma). L'Orbitrap Fusion, plus récent, est probablement plus sensible que l'Exactive. Cependant, cet élément n'est pas la principale

raison pour expliquer cette différence dans le nombre de métabolites identifiés. En effet, l'utilisation de l'Exactive sur des échantillons plasmatiques lors d'autres études a permis de détecter plusieurs centaines de métabolites (Boudah et al., 2014). La source principale de la différence observée est donc la concentration intrinsèque des différentes matrices. Le plasma est en effet naturellement beaucoup plus concentré en métabolites que le LCR, ce qui permet d'y identifier un nombre plus important de métabolites. Il est aussi intéressant de noter que parmi les 122 métabolites initialement identifiés dans le LCR, 100 (82 %) sont également détectés dans le plasma.

L'analyse métabolomique du plasma a donc permis d'identifier 342 métabolites dont 157 qui sont statistiquement différents entre les encéphalopathes et les témoins. Parmi eux, 63 sont spécifiques de l'EH. Les analyses multivariées basées sur ces 342 métabolites identifiés permettent de mettre en évidence une bonne séparation entre les volontaires sains et les patients (Figure 12). La séparation entre patients uniquement cirrhotiques et encéphalopathes est plus limitée. La population d'encéphalopathes apparaît beaucoup plus hétérogène ce qui vient limiter la séparation entre ces deux groupes. Cependant il est intéressant d'observer que les encéphalopathes les plus graves (seulement deux patients de score de West Haven 3 ou 4) sont eux bien distincts de la population de cirrhotiques. Cela met en évidence une des limites de cette cohorte de validation. Du fait du faible nombre de patients présentant les stades les plus avancés, il semble difficile de pouvoir mettre en évidence une séparation entre les groupes. Cela peut être dû à la continuité clinique qui existe entre cirrhotiques et encéphalopathes de bas grade et au manque de critères objectivement quantifiables permettant de classifier les patients.

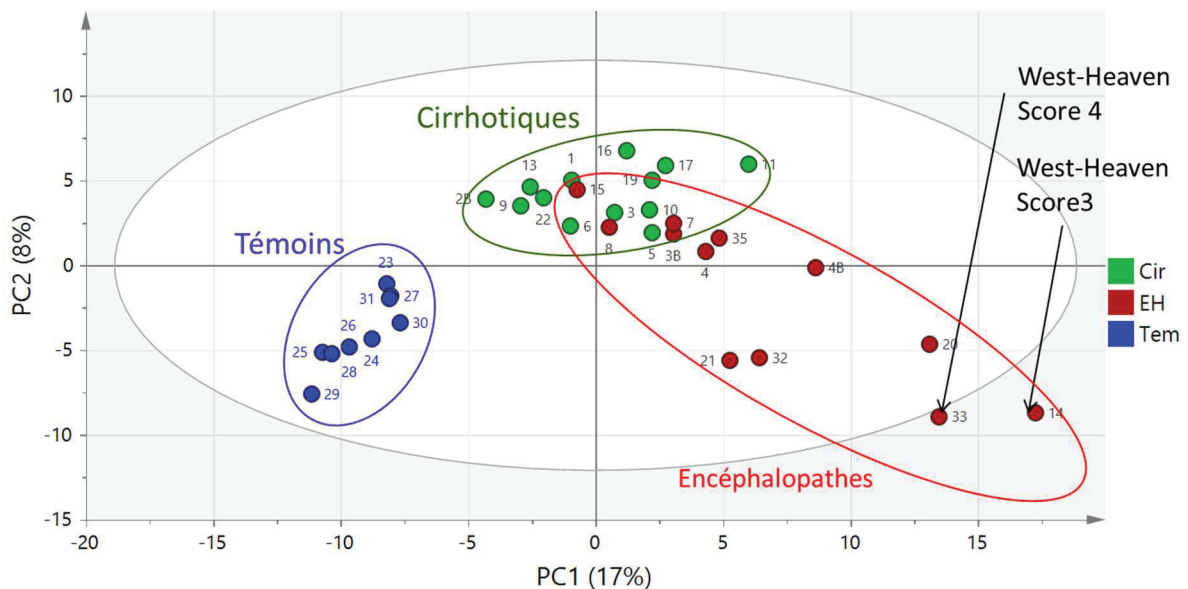


Figure 12 : Analyse multivariée (PLS-DA) de la seconde cohorte
Analyse sur les 342 métabolites détectés lors de l'analyse métabolomique combinant les quatre modes d'analyse LC-HRMS.

Parmi les métabolites les plus impactés, on retrouve les composés acétylés observés dans la première étude, qui sont également fortement augmentés dans le plasma (Figure 13). D'autres composés liés au métabolisme énergétique comme les adénosine-phosphate (AMP, ADP, ATP), sont également fortement impactés. Ils sont fortement diminués au niveau plasmatique, à la fois chez les cirrhotiques et chez les encéphalopathes (Figure 13). Les acylcarnitines à longues chaînes apparaissent spécifiquement diminuées chez les encéphalopathes (Figure 13). Tous ces résultats viennent corroborer l'hypothèse d'une dérégulation du métabolisme énergétique émise lors de la première étude.

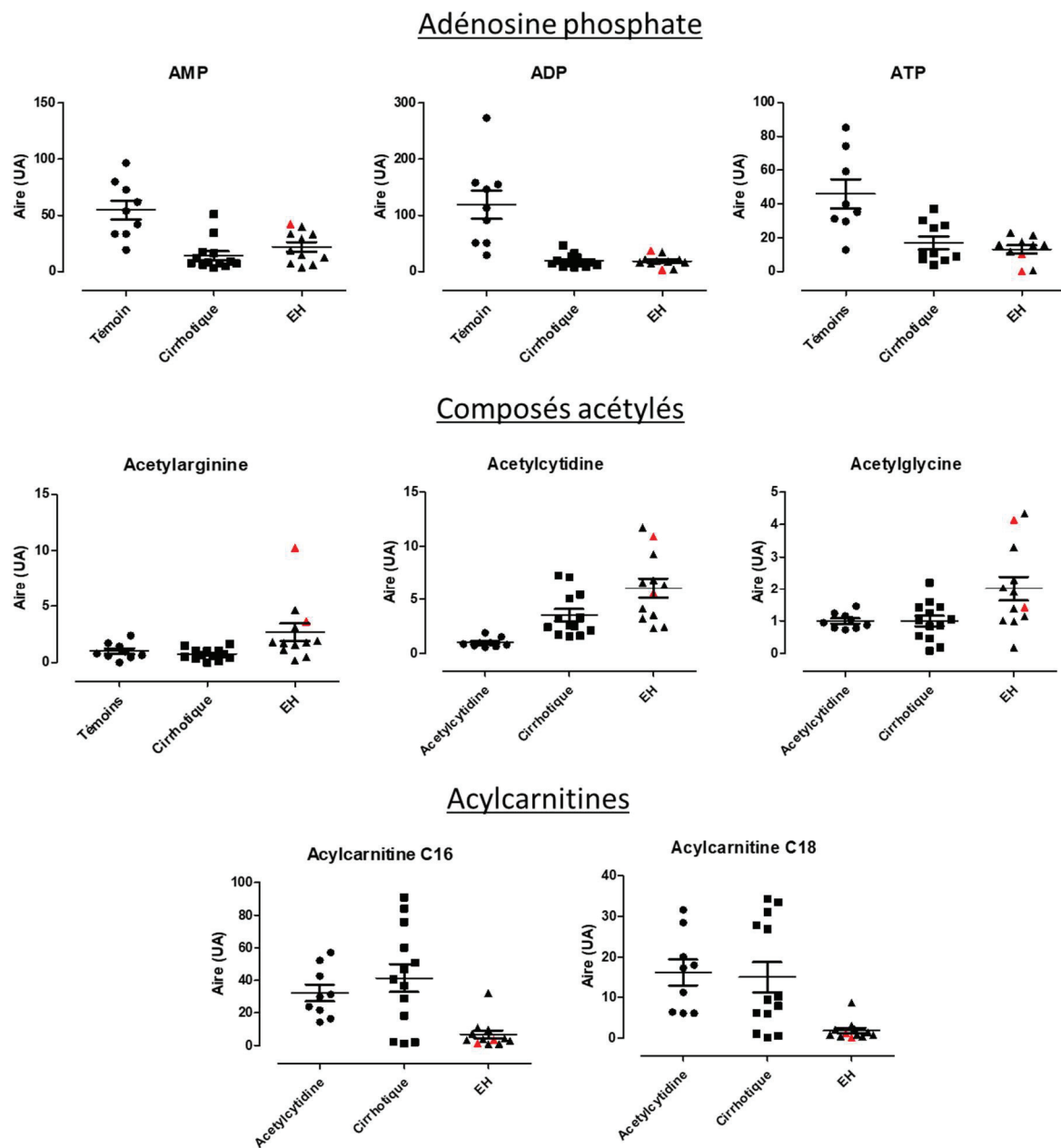


Figure 13 : Distributions des certains métabolites entre les trois populations. Les patients de score West-Haven 3 et 4 sont représentés en rouge.

L'analyse lipidomique montre une diminution drastique des populations de lipides chez les encéphalopathes (Tableau 8). Cette diminution est également observable chez les cirrhotiques. Cependant, pour certaines familles de lipides comme les phosphatidylcholines ou les sphingomyélines, cette diminution est significativement plus marquée chez les encéphalopathes.

	Detected lipids	Cir vs Ctl		HE vs Ctl		HE vs Cir	
		Mean ratio	p value	Mean ratio	p value	Mean ratio	p value
Glycerolipids							
Diacylglycerols (DG)	22	0.85	0.357	0.54	0.002	0.63	0.050
Triradylglycerols (TG)	216	0.90	0.512	0.67	0.034	0.75	0.318
Glycerophospholipids							
Lyso-Glycerophosphocholines (LPC)	32	0.49	0.002	0.25	< 0.001	0.51	0.042
Glycerophosphocholines (PC)	122	0.59	< 0.001	0.35	< 0.001	0.59	0.017
Glycerophosphoethanolamines (PE)	51	0.47	< 0.001	0.41	< 0.001	0.87	0.345
Glycerophosphoinositols (PI)	17	0.70	0.036	0.52	< 0.001	0.75	0.145
Glycerophosphoserines (PS)	20	0.35	< 0.001	0.24	< 0.001	0.68	0.091
Sphingolipids							
Ceramides (Cer)	25	0.48	< 0.001	0.33	< 0.001	0.7	0.07
Sphingomyelins (SM)	36	0.57	0.001	0.34	< 0.001	0.6	0.024
Sulfoglycosphingolipids (Su)	11	0.67	0.007	0.55	< 0.001	0.82	0.258
Steroids and Steroid Derivatives							
Cholesteryl esters (Chol.Ester)	42	0.54	< 0.001	0.24	< 0.001	0.45	0.027

Tableau 8 : Statistiques des différentes familles de lipides détectées en lipidomique

De façon intéressante, on observe en métabolomique des modifications de concentrations de précurseurs lipidiques comme l'éthanolamine phosphate, le glycerol-3-phosphate, la glycérophosphorylcholine, la sphingosine-1-phosphate ou la sphinganine, toutes ces espèces étant moins concentrées chez les encéphalopathes. Cette diminution des précurseurs lipidiques peut être reliée à une diminution globale des lipides observés en lipidomique. Parmi les familles de lipides significativement modifiées chez les encéphalopathes par rapport aux cirrhotiques, se trouvent notamment les (lyso)-glycérophosphocholines ou les sphingomyélines dont les précurseurs sont spécifiquement diminués. La voie de biosynthèse des sphingolipides est présentée en Figure 14. Cette diminution globale des concentrations d'espèces lipidiques peut être reliée à une transition du métabolisme énergétique d'une source d'énergie basée sur la consommation de sucres et lié au cycle de Krebs vers une source d'énergie lipidique via la β -oxydation.

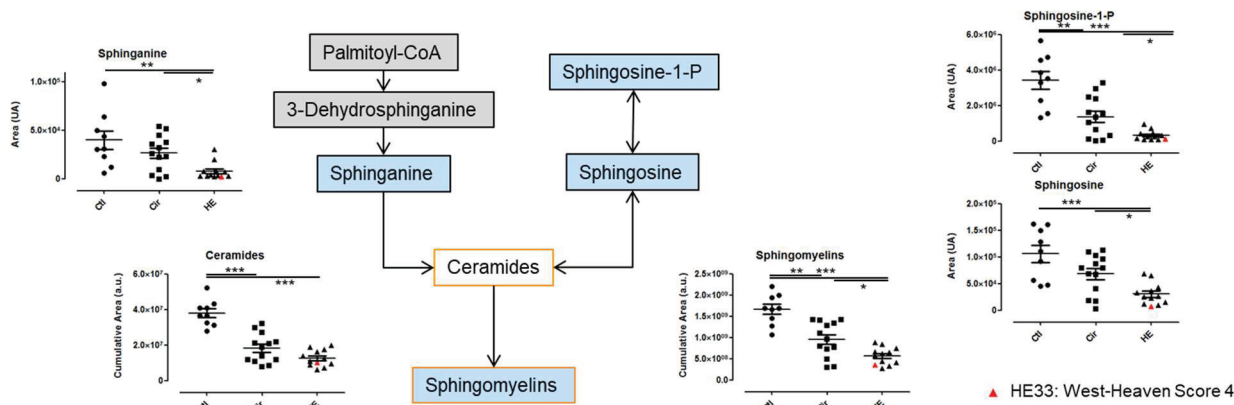


Figure 14 : Altération de la voie de biosynthèse des sphingolipides.

Lien entre la diminution des précurseurs lipidiques observés en métabolomiques (cadre noir) et les sphingolipides analysées en lipidomique (cadre orange). Les éléments sur fond gris ne sont pas détectés, sur fond blanc ne sont pas significativement modifiés entre cirrhotiques et encéphalopathes, sur fond bleu sont diminués chez les encéphalopathes par rapport aux cirrhotiques.

Concernant le métabolisme des sucres, l'analyse métabolomique montre par exemple une augmentation des concentrations de composés de la famille des acétylhexosamines. De son côté, l'analyse glycomique montre une augmentation de la proportion de glycanes sous forme bissectines, c'est-à-dire qui contiennent un acetylglucosamine supplémentaire (Figure 15). Ces deux éléments concordants montrent la complémentarité des toutes ces analyses « -omiques », qui permettent une description plus précise des phénotypes métaboliques des patients. L'interprétation biologique des variations de concentrations observées reste cependant pour l'instant limitée car il n'y a pas eu d'analyse protéomique effectuée et que l'on sait que les principales protéines glycosylées sont normalement produites au niveau du foie. Du fait de la cirrhose, il est probable que toute la machinerie de synthèse de ces protéines soit impactée. Même si nous avons travaillé sur des ratios entre forme bissectines et non bissectines, il est difficile de séparer la part liée aux modifications globales de la production protéique de celle plus spécifique de la glycosylation de ces protéines.

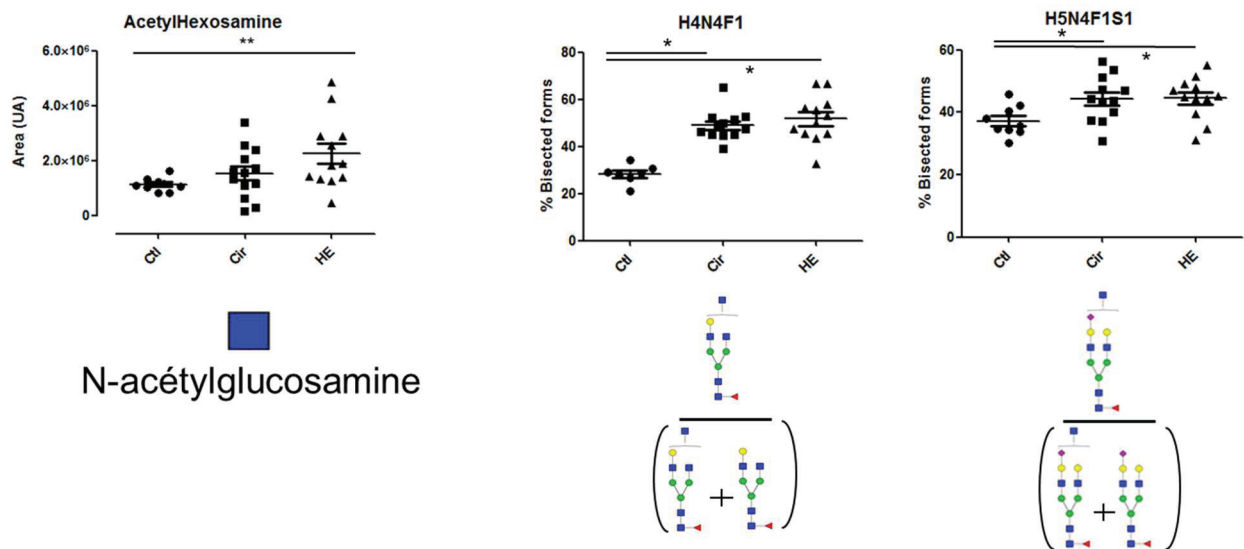


Figure 15 : Augmentation de la proportion de N-glycanes sous forme bissectine en lien avec une augmentation de la concentration en acetylhexosamine libre

5.2 Conclusion de l'étude

De manière générale, cette étude du phénotype plasmatique des patients souffrant d'encéphalopathie hépatique vient confirmer les dérèglements métaboliques observés dans la première étude. Le métabolisme énergétique apparaît fortement impacté par l'encéphalopathie hépatique et cet impact est visible au niveau plasmatique, en plus du LCR. L'utilisation conjointe des analyses métabolomique, lipidomique et glycomique permet d'obtenir une description plus précise des dérèglements métaboliques. Cependant, dès les premières analyses, nous avons constaté que la conception de la cohorte n'allait pas permettre de mettre en évidence de séparation claire entre patients cirrhotiques et encéphalopathes. En effet, cette cohorte présente une relativement faible puissance statistique, de par le nombre de patients, ce qui limite la mise en évidence d'une séparation complète entre les groupes de patients lors des analyses multivariées. Ce manque de distinction entre les groupes est également lié au faible grade des patients encéphalopathes (10 patients avec un score de West Haven 2, 1 score 3 et 1 score 4). Or, il apparaît sur la quasi-totalité des analyses que les patients de hauts scores (West-Haven 3-4) sont ceux qui apparaissent les plus impactés par les changements observés. L'absence de séparation claire ne remet donc pas à elle seule en cause l'intérêt des analyses sur le plasma et il serait prometteur de pouvoir compléter cette étude en incluant davantage de patients présentant une encéphalopathie hépatique avancée/grave afin de faire ressortir ces différences sous-jacentes. Cela permettra probablement d'identifier, avec une précision encore accrue, les dérèglements liés à l'EH.

Face aux difficultés liées à la conception de la cohorte et à l'interprétation clinique qu'il était possible d'en faire, nous avons étudié la complémentarité qui peut exister entre différentes approches que sont la métabolomique, la lipidomique et la glycomique. Nous avons pu montrer que les résultats apportés par ces différentes approches sont bien cohérents et qu'il est possible de les combiner pour augmenter la couverture métabolique et la compréhension des dérèglements liés à une maladie. Il s'agit d'une des premières études de cohorte médicale qui combine dans son interprétation des éléments de métabolomique, de lipidomique et de glycomique.

6 Analyses des difficultés rencontrées lors de l'analyse de cohortes médicales

La réalisation de ces deux études a permis de mettre en évidence un certain nombre de difficultés pouvant être rencontrées lors d'études métabolomiques.

La première vient de la définition de la cohorte et du recrutement des patients. En effet, il s'avère indispensable d'avoir une conception de la cohorte qui permette de répondre correctement à la question soulevée. Cela passe d'abord par une bonne définition des critères d'inclusion dans la cohorte. Si ces critères ne sont pas assez précis et que les différentes populations ne sont pas assez homogènes, la variabilité intragroupe risque de masquer la variabilité intergroupe que l'on cherche à observer. Il en va bien évidemment de même si les différentes populations sont trop similaires. Enfin, l'observation d'éventuelles différences ne pourra se faire que si le nombre d'échantillons est suffisamment important pour avoir une puissance statistique adéquate. Si dans une part importante des études de métabolomique, le recrutement de la cohorte est fait un amont par des tierces personnes (en générale des cliniciens dans le cadre des cohortes médicales), l'équipe en charge de l'analyse se doit d'être particulièrement attentive aux métadonnées des patients. Cela est nécessaire pour s'assurer que les résultats qui seront obtenus pourront être interprétés pour répondre à la question posée.

Les autres difficultés soulevées sont plus directement du ressort des métabolomistes. L'une des principales difficultés dans ce type d'analyses non ciblées est la caractérisation des variables permettant de discriminer les différentes populations. Cette étape d'annotation des données est une étape cruciale pour pouvoir donner un sens biologique aux études métabolomiques. Au cours du déroulé d'une analyse métabolomique, il s'agit d'une des étapes prenant le plus de temps. La qualité de cette annotation dépend d'abord de la qualité des données acquises, notamment en termes de précision de masse, de robustesse des temps de rétention et de mesure d'intensité relative intra spectre. Cela est vrai à la fois pour l'acquisition des profils initiaux mais également d'éventuels redosages d'échantillons ou de réinjections visant à obtenir des informations complémentaires

comme les spectres de fragmentation MS². Pour que cette étape soit la plus robuste possible, il est donc essentiel de maîtriser parfaitement les instruments analytiques et leur paramétrage. La qualité de l'annotation et le temps qu'elle nécessite dépend aussi également de la qualité des outils informatiques à disposition. Bien que de plus en plus de logiciels commerciaux soient disponibles pour réaliser cette étape, il est important de s'assurer que l'ensemble des éléments pouvant influencer la qualité de l'annotation soient accessibles et que les valeurs de paramétrages et les résultats obtenus soient parfaitement clairs. En effet, certains logiciels masquent parfois, dans un souci de facilité d'utilisation, un certain nombre de paramétrages devant pourtant être ajustés pour optimiser la qualité du résultat. De plus, ces outils s'appuient en général sur des bases de données universelles, ce qui peut parfois imposer l'utilisation d'un paramétrage moins restrictif au niveau du logiciel et donc conduire à une annotation de qualité moindre. Pour pallier ces difficultés, une partie de mon activité de thèse, non détaillée dans ce manuscrit, a porté sur l'analyse de près de 1000 composés standards. Cette analyse a été réalisée dans différentes conditions, à la fois chromatographiques et d'ionisation. Elle a conduit à l'acquisition de leur temps de rétention et à l'identification la forme ionique majoritaire présente, ce qui permet d'affiner la qualité de l'annotation. J'ai également développé des programmes informatiques en langage R permettant d'automatiser cette étape d'annotation.

La dernière étape particulièrement complexe au cours d'une analyse métabolomique est l'interprétation biologique et mécanistique des résultats obtenus. Il s'agit de l'étape la moins bien maîtrisée au niveau de l'ensemble de la communauté de métabolomique. Bien que quelques outils informatiques, notamment basés sur les réseaux métaboliques (**Cottret et al., 2010**), commencent à émerger, une grande partie de ce travail reste « manuel » et dépendant en grande partie des connaissances individuelles biologiques et biochimiques du responsable de l'étude (et de celles de son directeur de thèse...). Dans ce cadre, j'ai également pu développer quelques scripts permettant de classer automatiquement les molécules annotées selon les voies métaboliques dans lesquelles elles sont impliquées afin de faciliter ce travail d'interprétation. Cela a nécessité un gros travail bibliographique pour recouper les informations entre différentes bases de données tels que KEGG (**Kanehisa and Goto, 2000**) ou HMDB (**Wishart et al., 2007**).

L'ensemble des travaux présentés dans la suite de cette thèse consistera à évaluer les éléments permettant d'augmenter la quantité et la qualité des données générées ainsi que l'efficacité et la justesse de l'étape d'annotation des signaux d'intérêt.

Chapitre 2. Développement d'outils d'aide à l'identification des métabolites

Au cours de l'analyse métabolomique de cohortes, notamment médicales, l'interprétation biologique des résultats obtenus est souvent compliquée. Cette complexité est liée à différents éléments. Le premier est la large gamme dynamique des métabolites présents dans les échantillons biologiques. Malgré les progrès des instruments d'analyse en termes de sensibilité, une part importante des métabolites reste à ce jour invisible car présents à des concentrations inférieures aux limites de détection des instruments. Un autre point crucial est l'identification des signaux qui ont pu être détectés. Cette étape vise à attribuer un nom de métabolite à ces signaux. Il s'agit d'une étape clé pour permettre de transformer des données issues de la chimie analytique en données pouvant être interprétées biologiquement. Le troisième élément est l'interprétation biologique et mécanistique des données à partir des molécules apparaissant modifiées dans certains groupes d'échantillons. A l'heure actuelle, ces étapes sont les principaux freins à la transformation de résultats d'analyse de métabolomique en avancées médicales. L'objectif des travaux présentés ci-après est d'apporter des éléments d'information permettant d'améliorer l'exploitation des données issues des analyses métabolomiques par LC-HRMS. Ces travaux se focalisent sur l'étape d'identification des signaux détectés.

1 Les problématiques de l'annotation en métabolomique

L'annotation des données issues d'analyses métabolomiques par LC-HRMS se heurte à différentes limitations. La principale est l'absence de connaissance complète du métabolome. Contrairement à ce qui existe en génomique où l'ensemble de l'ADN humain a été séquencé, il n'existe pas de base de données exhaustive de métabolites. A l'heure actuelle, les bases de métabolites les plus complètes (HMDB (Wishart et al., 2018), CHEBI, LipidMaps) regroupent environ 50 000 métabolites. Pour ces métabolites, différentes informations sont disponibles. Certaines, telle que la masse exacte, peuvent être utilisées facilement. D'autres, telles que les temps de rétention ou les spectres de fragmentation sont liées à des conditions d'expérience spécifiques et ne sont donc pas généralisables à d'autres analyses. De plus, il reste encore une part importante des données obtenues lors d'une analyse métabolomique qui ne correspond à aucun élément déjà décrit et qui ne peut donc pas être annotée par l'interrogation de bases de données (da Silva et al., 2015). L'amélioration des outils d'annotation est donc un élément clé pour permettre l'augmentation de

l'apport de la métabolomique dans la compréhension des mécanismes régissant le métabolisme. Cela passe également par l'amélioration de la performance des spectromètres de masse et leur évaluation pour des applications en métabolomique, notamment pour l'étape d'identification des métabolites.

Ce chapitre se focalise principalement sur les outils d'identification des métabolites non directement annotés à partir des bases de données. Comme expliqué précédemment (partie 5.2 de l'introduction), différents éléments permettent d'apporter des informations sur ces composés. Ce chapitre se focalise sur deux de ces éléments d'identification : l'abondance isotopique relative et l'interprétation des spectres MS².

2 Pertinence de l'utilisation des données d'abondance isotopique relative sur l'Orbitrap Fusion pour la détermination de la formule brute

2.1 Contexte

Comme expliqué dans les précédentes parties de ce manuscrit, l'une des étapes les plus complexes pour l'interprétation des analyses métabolomiques en spectrométrie de masse à haute résolution est l'annotation des signaux d'intérêt (Alonso et al., 2015). Cette étape est d'autant plus importante pour les métabolites observés qui ne seraient pas encore présents dans les bases de données. Le premier objectif pour annoter un jeu de données est de déterminer la ou les formules brutes pouvant correspondre aux signaux observés. Cette détermination repose sur différents éléments. Le plus utilisé est le rapport m/z mesuré. Ce rapport m/z peut soit être utilisé pour interroger des bases de données, soit pour calculer *de novo* l'ensemble des formules brutes compatibles avec ce rapport (grâce à la connaissance de la masse exacte des différents éléments constituant le vivant). Grâce aux excellentes précisions de mesure de masse disponibles sur les spectromètres de masse à haute résolution, le rapport m/z peut parfois être suffisant pour permettre l'obtention d'une unique formule brute. Si cela est vrai pour les petits métabolites, cela devient plus complexe lorsque la masse mesurée augmente. En effet, le nombre de combinaisons possibles entre les différents atomes devient plus important, ce qui génère un nombre nettement plus grand de formules brutes compatibles avec la masse mesurée, même lorsque la précision de mesure de masse est de l'ordre du ppm (Kind and Fiehn, 2006). Pour aider à éliminer certaines des hypothèses générées par le seul rapport m/z, d'autres éléments peuvent être utilisés comme l'abondance isotopique relative (RIA, *Relative Isotopic Abundance*). L'utilisation du RIA est basée sur l'existence d'isotopes stables minoritaires des principaux atomes dont la proportion est connue et conservée (Tableau 5). Ainsi chaque métabolite va générer plusieurs signaux correspondants aux différentes formes isotopiques de la molécule (isotopologues). L'intensité relative entre ces différentes formes apporte des informations sur le nombre de chaque type d'atome (C, H, O, N, S, Cl, K, ...) composant

le métabolite à identifier. Cette approche utilisant le RIA pour la détermination de la formule brute a déjà beaucoup été étudiée et a fait ses preuves en métabolomique (Erve et al., 2009). Cependant la quasi-totalité des études menées se focalisent sur l'isotope minoritaire principal, à savoir le ^{13}C et ne cherchent pas à utiliser les signaux d'autres isotopes comme l' ^{18}O . Cela vient du fait que cet isotope ait une masse très proche de celle d'autres isotopes, ce qui implique l'utilisation d'une résolution très élevée pour individualiser le signal spécifique de l' ^{18}O , et de sa faible intensité relative. Jusqu'à l'arrivée des dernières générations d'Orbitrap, seuls les instruments de type FT-ICR permettaient d'atteindre une résolution suffisante (supérieure à 250 000) (Miura et al., 2010). Or, le temps d'acquisition des FT-ICR n'est pas compatible avec une analyse HPLC-HRMS. Les nouveaux Orbitrap (Orbitrap Fusion et Orbitrap Lumos) permettent d'atteindre de très hautes résolutions tout en conservant un temps de cycle compatible avec des analyses HPLC-HRMS. Nous avons donc souhaité étudier les performances de l'Orbitrap Fusion pour l'annotation et la détermination de formules brutes lors d'analyses métabolomiques.

2.2 Objectifs de l'étude

Le but de ce projet est d'évaluer en détail les performances de l'Orbitrap Fusion dans un contexte d'annotation lors d'analyses métabolomiques par HPLC-HRMS. Ce projet se focalise sur la détermination de formules brutes. Les principaux éléments évalués sont donc la précision de mesure de masse, la précision de mesure de l'abondance isotopique relative du ^{13}C ainsi que celle de l' ^{18}O . Différents éléments paramétrables de l'Orbitrap Fusion sont étudiés (résolution, contrôle du nombre d'ions dans le piège, ...) afin de déterminer la configuration à recommander pour une utilisation optimale de l'instrument. Cette étude a fait l'objet d'un article publié dans la revue « Analytical Chemistry ».

2.3 Aspects méthodologiques

Afin d'évaluer les performances de l'Orbitrap Fusion, un jeu de 50 molécules a été composé. Ces molécules ont été sélectionnées selon différents critères. Le premier est leur bonne réponse en spectrométrie de masse à ionisation électrospray positive, condition dans laquelle est réalisée l'étude. Le second est le non chevauchement des pics chromatographiques en phase inverse afin de pouvoir identifier facilement les signaux liés à chaque molécule. Le dernier critère est la présence d'un ou de plusieurs atomes d'oxygènes dans la molécule afin de pouvoir étudier les signaux spécifiquement liés à cet atome et à ses différents isotopes. Ce jeu de molécules a été analysé en solvant mais aussi dopé dans un extrait de métabolites plasmatiques pour valider les résultats en conditions réelles. Différents paramètres de l'instrument ont été étudiés. Trois résolutions ont été utilisées (120 000, 240 000 et 500 000) pour évaluer l'impact de la résolution sur la justesse de mesure ainsi que trois niveaux d'AGC (*Automatic Gain Control* ; 5^e4 , 1^e5 et 5^e5) pour évaluer l'impact du nombre d'ions dans l'Orbitrap.

2.4 Publication numéro 2

Evaluation of High-Filed Orbitrap Fusion for Compound Annotation in Metabolomics,

Pierre Barbier Saint Hilaire, Ulli M. Hohenester, Benoit Colsch, Jean-Claude Tabet, Christophe Junot, François Fenaille

Analytical Chemistry, 2018, 90, 3030-3035

Les données supplémentaires liées à cette publication sont présentées en Annexe 5

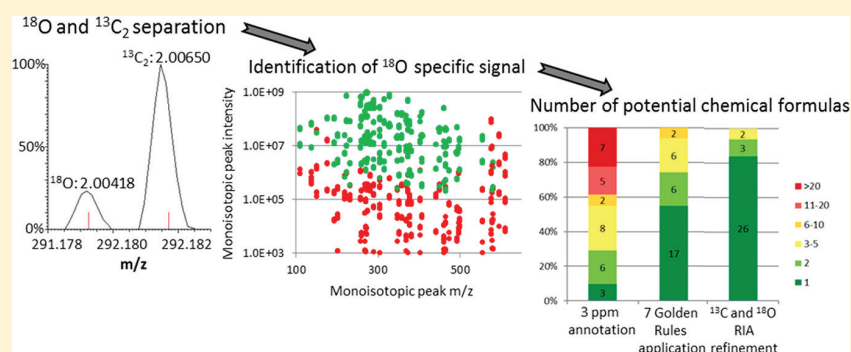
Evaluation of the High-Field Orbitrap Fusion for Compound Annotation in Metabolomics

Pierre Barbier Saint Hilaire,[†] Ulli M. Hohenester,[†] Benoit Colsch,[†] Jean-Claude Tabet,^{†,‡} Christophe Junot,[†] and François Fenaille^{*,†,§}

[†]Service de Pharmacologie et d'Immunoanalyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France

[‡]Sorbonne Universités, Campus Pierre et Marie Curie, IPCM, 4 place Jussieu, 75252 Paris Cedex 05, France

Supporting Information



ABSTRACT: Annotation of signals of interest represents a key point in mass spectrometry-based metabolomics studies. The first level of investigation is the elemental composition, which can be deduced from accurately measured masses and isotope patterns. However, accuracy of these two parameters remains to be evaluated on last generation mass spectrometers to determine the level of confidence that can be used during the annotation process. In this context, we evaluated the performance of the Orbitrap Fusion mass spectrometer for the first time and demonstrated huge potential for metabolite annotation via elemental composition determination. This work was performed using a set of 50 standard compounds analyzed under LC/MS conditions in solvent and biological media. Accurate control of the number of trapped ions proved mandatory to avoid space charge effects, ensure sub-ppm mass accuracy (using external calibration), and reliable measurement of isotopic patterns at 500,000 resolution. On the basis of the results, we propose standard optimized experimental conditions for performing robust and accurate untargeted metabolomics on the Orbitrap Fusion at high mass measurement and mass spectral accuracy.

Metabolomics aims at studying the whole metabolite set contained in biological media.^{1,2} The main objectives are to identify and quantify all the small molecules present in complex matrices in a comprehensive manner. Due to its ability to detect hundreds of metabolites in standard matrices,^{3–5} high resolution electrospray ionization mass spectrometry (HRMS) coupled to liquid chromatography is one of the most widely used analytical techniques in metabolomics.⁶

Despite the constant improvements of mass spectrometers in terms of sensitivity, mass accuracy, resolution, and acquisition speed, identification of unknown metabolites remains challenging.^{7–9} Accurate metabolite identification is crucial to translate relevant analytical data into meaningful biological information. The first level of investigations when using HRMS data relies on the robust and accurate determination of an elemental composition (EC) for a given MS peak. Altogether, mass accuracy, mass resolution, and isotopic pattern accuracy directly influence the ability of HRMS instruments to perform accurate EC determination. Recent Orbitrap instruments can routinely

achieve mass accuracies below 3 ppm (with external calibration), whereas Fourier transform ion cyclotron resonance (FTICR) mass spectrometers are characterized by an even better mass accuracy below 0.2 ppm.^{10,11} However, high mass accuracy alone is not sufficient to assign a unique and unambiguous EC to a measured mass.¹² For instance, Kind and Fiehn showed that a mass spectrometer capable of 3 ppm mass accuracy and 2% error for isotopic abundance patterns should outperform mass spectrometers providing a mass accuracy of 0.1 ppm but results in poor performances in terms of relative natural isotope abundance (RIA) measurement.¹² RIA has already been studied on different mass spectrometers, but these studies deal essentially with natural ¹³C₁ isotope. The RIA of this isotope appears to be rather commonly underestimated when measured on an Orbitrap^{13–15} as well as on FTICR

Received: December 22, 2017

Accepted: February 9, 2018

Published: February 9, 2018

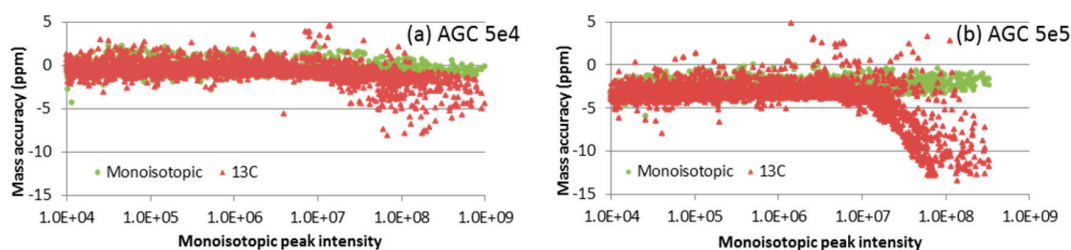


Figure 1. Evolution of mass accuracy of monoisotopic ion (green) and $^{13}\text{C}_1$ (red) peaks according to signal intensity. Data were acquired at a resolution of 500,000 with AGC target values of (a) 5e4 and (b) 5e5.

instruments.^{15,16} As an example, a mean RIA error standing between +5% and -50% is commonly observed, depending on ion intensity.^{14,16} Increased resolutions and low ion abundances also negatively impact the quality of the RIA measurement.¹⁴

Examining the fine isotopic distribution, in particular that of the $M + 2$ isotopes, can also bring valuable information for refining EC list. However, although not a real issue for low mass metabolites, it is estimated that a full width at half-maximum (fwhm) resolution over 300,000 is required to discriminate between ^{15}N and $^{13}\text{C}_1$ but also between ^{34}S , ^{18}O , and $^{13}\text{C}_2$ species for metabolites with a mass of about 500 Da.¹⁷ Until recently, only FTICR instruments could reach such required resolution, but the newly introduced high-field Orbitrap instruments, such as the Orbitrap Fusion, are now capable of achieving mass resolutions higher than 500,000 (at m/z 200, fwhm), for instance rendering ^{18}O and $^{13}\text{C}_2$ isotopic peaks distinguishable. Due to their natural abundance, ^{34}S isotopes are quite convenient to study as recently done by Blake et al. on an FTICR instrument.¹⁸ However, and to the best of our knowledge, no work has been published regarding specific ^{18}O RIA measurement for EC determination of unknown metabolites on FTICR or on an Orbitrap instrument.

The aim of this article is therefore to evaluate the effectiveness of the newly introduced high resolution Orbitrap Fusion for accurately measuring metabolite masses and relative isotopic abundances to expedite EC determination. A set of 50 standard molecules dissolved in a water–acetonitrile solution was first used in order to optimize analytical conditions and then after being spiked into a complex plasma metabolite extract to evaluate the potential impact of matrix interferences. In addition to being the first description of the Orbitrap Fusion for metabolomics, this report also provides optimized operating conditions for getting the best of this instrument in this context.

EXPERIMENTAL SECTION

Preparation of Standard and Serum Samples. A pool of 50 selected chemical compounds (Table S1, [Supporting Information](#)) was serially diluted with a solution of water/ACN (95:5) containing 0.1% formic acid to give concentrations ranging from 1 ng/mL to 10 $\mu\text{g}/\text{mL}$.

Serum metabolites were prepared using a dedicated protocol.¹⁹ Serum metabolites were resuspended with 1 mL of water/ACN (95:5) with 0.1% formic acid containing each one of the above-mentioned concentrations of standard compounds.

LC/HRMS Analysis. Chromatographic separation was performed on a Dionex Ultimate 3000 system equipped with a Hypersil Gold C_{18} column (ThermoFisher Scientific, San Jose, CA, U.S.A.). Analyses were performed on an Orbitrap Fusion (ThermoFisher Scientific) equipped with a heated electrospray ionization source operated in the positive mode.

The AGC target (5e4, 1e5, 5e5) and resolution (120,000, 240,000, and 500,000) were adapted to each experiment.

More detailed information regarding experimental conditions is given in the [Supporting Information](#).

RESULTS AND DISCUSSION

Our initial objective was to evaluate the ability of Orbitrap Fusion to perform accurate mass and RIA measurements in order to improve the relevance of metabolite identification. In that objective, a set of 50 standard molecules, with m/z values ranging between m/z 110 and m/z 775, was first analyzed by LC/HRMS. Preliminary investigations demonstrated that metabolite concentration and the number of singly charged ions injected into the Orbitrap mass analyzer had a huge negative impact on measurement accuracy. Therefore, we decided to investigate and optimize key acquisition parameters. Optimized conditions were then used to evaluate the performances of the Orbitrap Fusion for accurately and reliably measuring isotope profiles for metabolite identification.

Impact of Resolution and Automatic Gain Control (AGC) on Mass Accuracy. ECs of unknown metabolites are primarily deduced from accurate masses measured using HRMS. We hereafter report on the mass accuracy obtained on the Orbitrap Fusion under routine conditions and after external calibration performed on a weekly basis (as recommended by the manufacturer). We evaluated the impact of different resolutions and AGC target settings. AGC, for which the default value is 2e5 on the Orbitrap Fusion, monitors and defines the number of ions that enter the Orbitrap analyzer.

We used a monoisotopic ion and its $^{13}\text{C}_1$ m/z values measured on each individual mass spectrum, i.e., each acquired spectrum across the chromatographic peak, for the 50 molecules at the five concentrations in triplicate (3–10 mass spectra per compound depending on peak width and used resolution). Under these conditions, 16747, 8238, and 3989 charged molecular species were studied at a resolution of 120,000, 240,000, and 500,000, respectively. In all except one situation, with the highest AGC target setting, the mass accuracy of the investigated monoisotopic ions proved excellent and below 0.8 ppm on average (Table S2, [Supporting Information](#)). Mass accuracy also appeared unaffected by the resolution, but increasing the AGC target to 5e5 resulted in a dramatic degradation of the absolute mass accuracy to a 3 ppm error, whatever the peak intensity (Figure 1). Such systematic and rather unexpected mass inaccuracy is probably linked to the existence of space charge effects resulting from Orbitrap overfilling at such a high AGC target value. High AGC target values are known to negatively affect mass accuracy, due to frequency shifts of adjacent ions originating from space charge effects.²⁰ Space charge effects have been largely studied for Fourier transform ion cyclotron resonance mass spectrom-

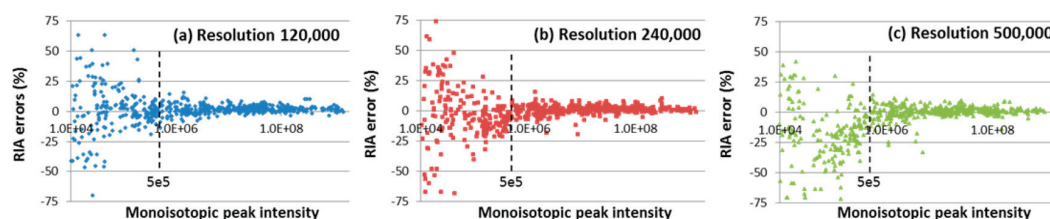


Figure 2. RIA errors of the $^{13}\text{C}_1$ isotope at three different resolutions: (a) 120,000, (b) 240,000 and (c) 500,000. In all cases, AGC target value was set to $5e4$.

eters²¹ but still remained rather poorly documented for high-field Orbitrap instruments²² especially when dealing with small molecules and metabolomics applications.

The impact of space charge effects was even more obvious when considering $^{13}\text{C}_1$ ion mass accuracy for especially high intensity peaks²³ (i.e., over $1e7$). At high ion intensities, Coulombic interaction between large ^{12}C and smaller $^{13}\text{C}_1$ ion clouds results in a shift of their oscillation frequencies toward the others, thus leading to a dramatic shift of mass difference of 1.0021 Da instead of the expected 1.0033 Da value between the two isotopes (Figure S1A, Supporting Information). Such a negative drift of $^{13}\text{C}_1$ mass accuracy (Figure 1) is reported for the first time and highlights the necessity to check all-isotope mass accuracies while optimizing new methods. This is particularly relevant in the context of metabolomic workflows where bioinformatics tools are set up to automatically pick up $^{13}\text{C}_1$ isotopes for further validation of metabolite annotation. Thus, inaccurate $^{13}\text{C}_1$ isotope measurement as a result of suboptimal acquisition conditions will undoubtedly lead to improper signal annotation and ultimately the identification of the corresponding metabolite. Of particular interest is that such particular behavior was not observed with other Exactive or Q-Exactive instruments as operated in our laboratory under routine conditions (Figure S2, Supporting Information), which tends to underline the potential particular sensitivity of compact high-field Orbitrap analyzers with regards to the number of ions injected.

Contrary to what has been observed by Tarasova et al. on an LTQ-Orbitrap Elite for shotgun proteomic experiments,²² sample complexity with several potential coeluting species does not seem to impact the ion coalescence phenomenon. For instance, another set of experiments performed on a serum metabolic extract spiked with the same 50 molecules and analyzed at a resolution of 500,000 and an AGC target value therefore set to $5e4$ demonstrated that the presence of a complex matrix does not impact mass accuracy and thus the relevance of our optimized parameters (Table S2, Supporting Information). In addition, the relationship between concentration and MS response of the 50 standard molecules proved linear over 3 to 4 orders of magnitude (from 1 to 10,000 ng/mL) with coefficients of determination $R^2 > 0.98$. Slopes of standard curves also proved highly similar for standards spiked in water and serum, illustrating the absence of dramatic matrix effects. In the opposite situation, a nonoptimal AGC target value yielded a poor linear response. Figure S3 gives the data obtained for three representative compounds.

Evaluation of the $^{13}\text{C}_1$ RIA Accuracy. The RIA can be extremely useful to confirm the EC of a detected metabolite. However, the accuracy of the RIA measurement has to be checked, especially when working with small molecules containing only a few carbon or oxygen atoms. Although studied using former generations of Orbitraps,^{13–15} the RIA

measurement accuracy of Orbitrap instruments equipped with a compact high-field trap has not been reported yet. In this work, as in previous ones,¹⁴ the RIA error is defined as

$$\text{RIA}_{\text{error}} = \frac{(\text{RIA}_{\text{measured}} - \text{RIA}_{\text{theoretical}})}{\text{RIA}_{\text{theoretical}}} \times 100$$

The signal of the $^{13}\text{C}_1$ isotope was observed 609, 598, and 569 times among the 750 possible cases (50 compounds, 5 concentrations, 3 replicates), for a resolution of 120,000, 240,000 and 500,000, respectively. $^{13}\text{C}_1$ RIA measurement was performed on all those observed signals, and the resulting data are shown in Figure 2.

These data can be separated into two distinct blocks according to the intensity of the monoisotopic peak. For low intensities, i.e., below $5e5$, the RIA measurement appeared impacted by the resolution and was not accurate with mean absolute RIA error values of 13.2, 16.2, and 24.7% for resolutions of 120,000, 240,000, and 500,000, respectively. Also, for resolutions of 120,000 and 240,000, the errors can be both positive and negative (mean RIA errors of +1.2 and -3.2% , respectively), whereas this error is essentially negative for a resolution of 500,000 (mean RIA error of -18.2%). This shift toward negative values has already been reported in previous work^{14,16} and can be linked to particular instability of small ion clouds in the Orbitrap.^{13,24}

For intensities above $5e5$, the RIA measurement becomes clearly more accurate with a mean absolute error below 5% for the three resolutions (2.8, 2.8, and 3.3% for resolutions of 120,000, 240,000, and 500,000, respectively; Table S3, Supporting Information). Under these conditions, ^{13}C RIA measurement error is comparable or even better than the ones previously reported on other ICR and Orbitrap instruments.^{14,15} The RIA of $^{13}\text{C}_1$ could thus be used quite confidently to improve molecular formula attribution.

The impact of the matrix was also evaluated by calculating the RIA error in a complex matrix. RIA error appears to be similar between pure molecules and spiked serum with a mean absolute RIA error of 3.9% for a resolution of 500,000 (Table S3, Supporting Information).

Altogether, these data demonstrate that an RIA error below 5% for signals $>5e5$ (noise level is at $\sim 1e3$ intensity under our conditions) allows the determination of the exact number of carbon atoms for molecules containing up to 20 carbons. If an ambiguity between ECs still exist, the ^{18}O isotope can also be used (when present) to refine the EC.

Detection of ^{18}O Isotopes and Evaluation of the ^{18}O RIA Accuracy. The detection and accurate measurement of resolved $M + 2$ isotopes can be highly informative for EC determination. When designing the test mixture, a special emphasis was placed on multioxygenated species to favor the observation of the corresponding $M + 2$ peak (the ^{18}O natural

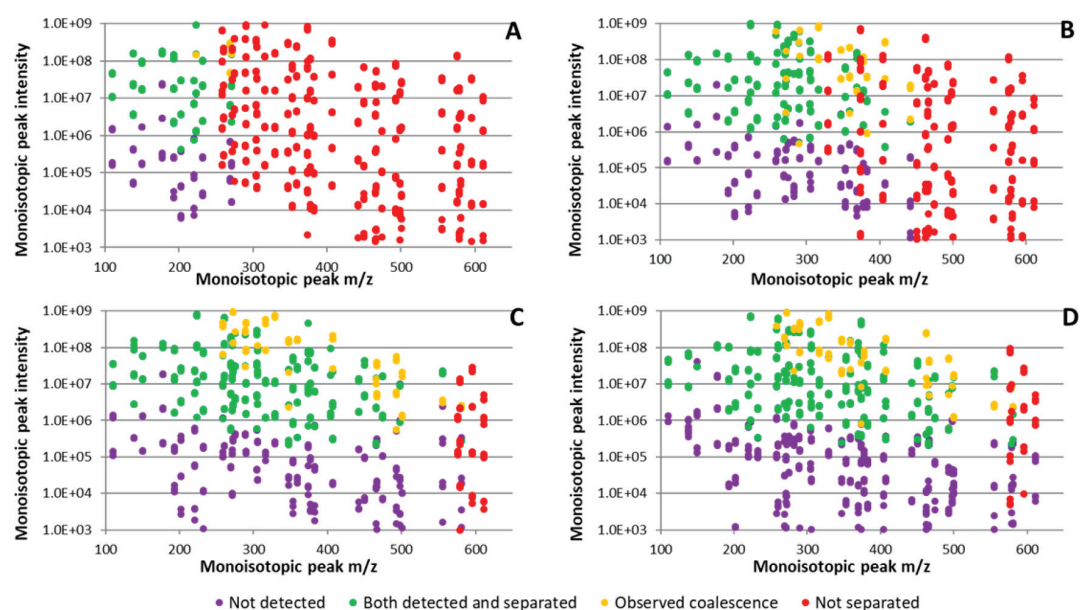


Figure 3. Separation of $^{18}\text{O}_1$ and $^{13}\text{C}_2$ isotopic m/z signals for the 50 standard compounds analyzed in buffer solutions at resolutions of (A) 120,000, (B) 240,000, and (C) 500,000; and D) after spiking in serum and acquisition at a 500,000 resolution. In all cases, AGC target value was set to 5e4. Purple dots, either one or both isotopes are not detected; green dots, both isotopes are detected and efficiently separated; orange dots, coalescence observed for high intensity signals; red dots, isotopes cannot be separated due to insufficient resolution.

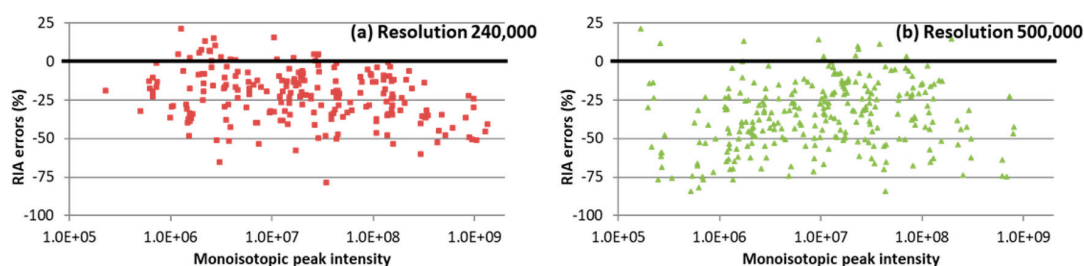


Figure 4. RIA errors for ^{18}O isotope at two different resolutions: 240,000 (a) and 500,000 (b). AGC target value was set to 5e4.

abundance being only 0.2%). The $^{18}\text{O}_1$ and $^{13}\text{C}_2$ isotopes are isobaric with a mass difference of only 0.00246 Da. Three different resolutions were compared: 120,000, 240,000, and 500,000 to study their impact on $M + 2$ fine isotope pattern and resolution.

The mass spectrum obtained at the apex of the chromatographic peak for each of the 50 standards was manually inspected. The presence and separation of ^{18}O and $^{13}\text{C}_2$ natural isotopes were checked.

Below a certain threshold of intensity, both isotopes cannot be detected due to their too weak abundances (Figure 3, purple dots). This threshold varies from $1\text{e}5$ to $1\text{e}6$ depending on the m/z of the monoisotopic ions and the numbers of carbon and oxygen atoms. For each tested resolution, an m/z threshold above which isotopes are not separated (at a given resolution) can be experimentally determined (Figure 3, red dots). This threshold is important to know for further defining the best resolution, regarding the mass range of interest or if a shorter duty cycle is required. For the three resolutions, i.e., 120,000, 240,000, 500,000, this threshold is at about m/z 250, m/z 350, and m/z 550, respectively. At high intensity, the m/z ratios of the two isotopes gradually shift and tend to merge into a single signal (Figure 3, orange dots). This phenomenon is mainly visible for intensities above $1\text{e}8$ for m/z between 250 and 400

or even lower intensities for higher m/z and is more pronounced at a resolution of 500,000 compared to 240,000. This coalescence of close m/z signals (Figure S1B₂₋₅, Supporting Information), has been studied or discussed in a number of publications using FTICR instruments,²⁵⁻²⁹ but only a few reports involved Orbitrap instruments. Gorshkov et al. showed the existence of the coalescence phenomenon on an Orbitrap Elite instrument by comparing two standard isobaric peptides spaced by 22 mDa (1059.5614 and 1059.5389 Da).²¹ Another study by Werner et al. confirmed this phenomenon on a Q-Exactive instrument.³⁰ They determined a threshold of intensity between $6\text{e}8$ and $1\text{e}9$ above which this coalescence appears, using an AGC target value of $1\text{e}6$ at a resolution of 35,000. With a lower AGC target value (i.e., $2\text{e}5$), they observed only very limited coalescence. Of note, almost no coalescence was observed for neutron encoded tandem mass tag (TMT) reported ions observed on an Orbitrap Elite, equipped with a compact high-field Orbitrap analyzer like the Orbitrap Fusion.³⁰

Taken as a whole, these data led us to define optimal conditions for studying molecules of interest in a metabolomic context at 500,000 resolution (Table S4, Supporting Information). A suitable dilution factor should also be determined for the molecules of interest such as the intensity

of the respective monoisotopic signals would not exceed 1e8. Of note, the resolution on Orbitraps decreases according to increasing m/z , which induces m/z thresholds above which $^{13}\text{C}_2$ and $^{18}\text{O}_1$ cannot be separated. For the Orbitrap Fusion, this threshold is between m/z 550 and m/z 600 at 500,000 resolution (Figure 3). Alternative mass spectrometry technologies such as FTICR or the newly introduced Orbitrap Lumos (1,000,000 resolution)³¹ should be considered for m/z values above this threshold.

These results were confirmed in a spiked serum extract at 500,000 resolution. Figure 3D shows that the different drawn conclusions (i.e., threshold in m/z and threshold in intensity) are also valid with molecules spiked in a complex matrix. Matrix complexity does not impact at all the reported phenomena. These results should therefore be directly transposable to real metabolomics studies.

RIA measurements of ^{18}O isotopes (when efficiently separated from $^{13}\text{C}_2$) appeared systematically undervaluated for resolutions of 240,000 and 500,000 (−20.6 and −33.0%, respectively), regardless of the ion intensity (Figure 4, Table S5, Supporting Information). As discussed above for $^{13}\text{C}_1$ isotopes, this is presumably linked to a nonideal movement of smaller ion clouds. Even if it might be difficult to define the exact number of oxygen atoms for every studied molecule, our results show that the observation of an ^{18}O signal on the spectrum and the subsequent measurement of its RIA can help estimating the number of oxygen atoms in the molecule (vide infra).

Another approach to improve ^{18}O RIA measurement of the $M + 2$ isotope could be to split the acquisition mass range or to use SIM acquisition as proposed by Su et al.³² But this would be hardly compatible with untargeted metabolomics due to the narrow chromatographic peaks obtained by UHPLC (peaks are few second wide) and the relatively high transient length needed to acquire mass spectra at a resolution of 500,000 (~1 s/mass spectrum).

Application to Elemental Composition Determination. The ability to determine the correct EC using different filtering criteria was evaluated using our 50-compound test mixture spiked in a serum extract at 1 $\mu\text{g}/\text{mL}$. This particular spiking concentration was selected, because it gave the highest number of ^{18}O isotopes detected (for 31 molecules over 50). The number of potential ECs within a 3 ppm window range was first calculated. The seven golden rules (as introduced by Kind and Fiehn)³³ were also evaluated and compared to the above optimized filters (considering either or both ^{13}C and ^{18}O isotopes) to reduce the number of candidates. Altogether the results demonstrated that our RIA filters outperformed the “7 golden rules” filter by providing a single correct EC for up to 26 over the 31 molecules tested. This represented a substantial ~30% improvement in terms of accurately determined elemental compositions (Figure S4, Supporting Information).

CONCLUSIONS

Performance of the Fusion instrument, equipped with a compact high-field Orbitrap analyzer, was studied for MS acquisition in an untargeted metabolomics context. Our data highlighted the presence of space charge effects occurring at high signal intensities, leading us to recommend the use of the smallest preset AGC target value, i.e., 5e4. Under these conditions and at 500,000 resolution, mass accuracy proved excellent with a mean absolute error of 0.63 ppm, while the absolute error regarding $^{13}\text{C}_1$ RIA was below 5% on average (when intensity of m/z signals exceeds 5e5). This isotope can

therefore be used with high reliability to refine the elemental composition list. The Orbitrap Fusion is also able to resolve the ^{18}O and $^{13}\text{C}_2$ isotopic signals, even for masses around 500 Da when using a resolution of 500,000. Last but not least, ^{18}O RIA proved helpful to determine a minimal number of oxygen atoms in the elemental composition of the molecule. Altogether, our results demonstrate significant potential of the high-field Orbitrap Fusion for enhancing confidence in metabolite annotation. This paper also underlines the potential need of new algorithms to check for the presence of space charge effects and further improve data quality and reliability.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b05372.

Experimental conditions; information on the set of 50 molecules studied in the positive mode; mass accuracy for monoisotopic ions and their corresponding ^{13}C ions in various experimental conditions; mean absolute ^{13}C RIA error values and mean ^{13}C RIA error values for low and high intensity signals in diverse conditions; recommended parameters for optimal measurements of ^{13}C and ^{18}O RIA on the Orbitrap Fusion; mean absolute ^{18}O RIA error values and mean ^{18}O RIA errors for all and high intensity signals in diverse conditions; mass spectra of atropine for five different concentrations showing space charge effect; Comparison of monoisotopic ion to $^{13}\text{C}_1$ mass differences on three different Orbitrap instruments; examples of calibration curves obtained for three compounds (formononetin; atropine; 21-doexcortisol) in diverse conditions; number of potential ECs per compound (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: 33-1-69-08-79-54; Fax: 33-169-08-59-07; E-mail: francois.fenaille@cea.fr.

ORCID

François Fenaille: 0000-0001-6787-4149

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Commissariat à l'Énergie Atomique et aux Énergies Alternatives and the MetaboHUB infrastructure (ANR-11-INBS-0010 grant).

REFERENCES

- (1) Roux, A.; Xu, Y.; Heilier, J.-F.; Olivier, M.-F.; Ezan, E.; Tabet, J.-C.; Junot, C. *Anal. Chem.* **2012**, *84*, 6429–6437.
- (2) Vanden Bussche, J.; Marzorati, M.; Laukens, D.; Vanhaecke, L. *Anal. Chem.* **2015**, *87*, 10927–10934.
- (3) Aretz, L.; Meierhofer, D. *Int. J. Mol. Sci.* **2016**, *17*, 632.
- (4) Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. *Metabolomics* **2009**, *5* (4), 435–458.
- (5) Fuhrer, T.; Zamboni, N. *Curr. Opin. Biotechnol.* **2015**, *31*, 73–78.
- (6) Nagana Gowda, G. A.; Raftery, D. J. *Magn. Reson.* **2015**, *260*, 144–160.

- (7) Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D. *J. Pharm. Biomed. Anal.* **2014**, *87*, 12–25.
- (8) Viant, M. R.; Kurland, I. J.; Jones, M. R.; Dunn, W. B. *Curr. Opin. Chem. Biol.* **2017**, *36*, 64–69.
- (9) Alonso, A.; Marsal, S.; Julià, A. *Front. Bioeng. Biotechnol.* **2015**, *3*, 23.
- (10) Marshall, A. G.; Hendrickson, C. L. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599.
- (11) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1626–1632.
- (12) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
- (13) Erve, J. C. L.; Gu, M.; Wang, Y.; DeMaio, W.; Talaat, R. E. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2058–2069.
- (14) Xu, Y.; Heilier, J.-F.; Madalinski, G.; Genin, E.; Ezan, E.; Tabet, J.-C.; Junot, C. *Anal. Chem.* **2010**, *82*, 5490–5501.
- (15) Weber, R. J. M.; Southam, A. D.; Sommer, U.; Viant, M. R. *Anal. Chem.* **2011**, *83*, 3737–3743.
- (16) Stoll, N.; Schmidt, E.; Thurow, K. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692–1699.
- (17) Miura, D.; Tsuji, Y.; Takahashi, K.; Wariishi, H.; Saito, K. *Anal. Chem.* **2010**, *82*, 5887–5891.
- (18) Blake, S. L.; Walker, S. H.; Muddiman, D. C.; Beck, K. R.; Hinks, D. J. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2269–2275.
- (19) Boudah, S.; Olivier, M.-F.; Aros-Calt, S.; Oliveira, L.; Fenaille, F.; Tabet, J.-C.; Junot, C. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2014**, *966*, 34–47.
- (20) Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.* **2008**, *27*, 661–699.
- (21) Gorshkov, M. V.; Fornelli, L.; Tsybin, Y. O. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1711–1717.
- (22) Tarasova, I. A.; Surin, A. K.; Fornelli, L.; Pridatchenko, M. L.; Suvorina, M. Y.; Gorshkov, M. V. *Eur. J. Mass Spectrom.* **2015**, *21*, 459–470.
- (23) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113–2120.
- (24) Ilchenko, S.; Previs, S. F.; Rachdaoui, N.; Willard, B.; McCullough, A.; Kasumov, T. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 309–312.
- (25) Vladimirov, G.; Hendrickson, C. L.; Blakney, G. T.; Marshall, A. G.; Heeren, R. M. A.; Nikolaev, E. N. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 375–384.
- (26) Mitchell, D. W.; Smith, R. D. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 4366–4386.
- (27) Gordon, E. F.; Muddiman, D. C. *J. Mass Spectrom.* **2001**, *36*, 195–203.
- (28) Aizikov, K.; Mathur, R.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 247–256.
- (29) Boldin, I. A.; Nikolaev, E. N. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3213–3219.
- (30) Werner, T.; Sweetman, G.; Savitski, M. F.; Mathieson, T.; Bantscheff, M.; Savitski, M. M. *Anal. Chem.* **2014**, *86*, 3594–3601.
- (31) Canterbury, J. D.; Senko, M. W.; Sharma, S.; Stratton, T.; Zabrouskov, V.; Denisov, E.; Makarov, A. Performance metrics and example applications: Orbitrap mass spectrometry at resolving power 1M. Presented at 65th ASMS Conference, Indianapolis, IN, U.S.A., 2017.
- (32) Su, X.; Lu, W.; Rabinowitz, J. D. *Anal. Chem.* **2017**, *89*, 5940–5948.
- (33) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, *8*, 105.

2.5 Principaux résultats

Le premier élément suivi est la précision de mesure de masse. Cette précision apparaît fortement influencée par l'AGC et ce, quelle que soit la résolution. On observe tout d'abord une déviation globale du signal au plus fort des AGC étudiés avec une erreur observée de 3 ppm contre moins de 0,8 ppm pour les deux autres valeurs d'AGC. La seconde observation importante est une dérive spécifique dans la mesure de masse du ^{13}C pour des signaux de forte intensité alors que la mesure de masse de l'ion monoisotopique reste juste. Cette observation est associée à des interactions entre nuages d'ions dans l'Orbitrap lorsque trop d'ions sont présents simultanément dans le piège. Ces observations nous conduisent à conseiller l'utilisation d'un faible AGC lors de la réalisation d'études métabolomiques sur l'Orbitrap Fusion. Un AGC de 5^4 a donc été utilisé pour le reste de ce projet.

Concernant la précision de mesure du RIA du ^{13}C , l'Orbitrap Fusion donne des résultats au moins aussi bons que ceux observés sur d'autres générations d'instruments (Weber et al., 2011; Xu et al., 2010) avec une erreur absolue moyenne inférieure à 3 %. Il faut tout de même noter que cette mesure n'est juste que lorsque le signal est suffisamment important pour que le signal du ^{13}C soit bien distinct du bruit. Ce seuil est situé à 5^5 en termes d'intensité de l'ion monoisotopique.

Pour ^{18}O , la première étape a été de déterminer les conditions dans lesquelles un signal spécifique de cet isotope pouvait être observé. Il a été montré que ce signal spécifique pouvait être séparé des autres jusqu'à des valeurs de rapport m/z de 550, 350 ou 250 pour des résolutions de 500 000, 240 000 et 120 000, respectivement. Un phénomène de coalescence de pic entre le signal du ^{18}O et celui du $^{13}\text{C}_2$ (distants de 2,46 mDa) a été mis en évidence pour des signaux d'intensité trop importante ($> 1^8$ pour l'ion monoisotopique). Dans ces conditions, le RIA de ^{18}O présente une erreur moyenne de l'ordre de -30%.

Nous avons enfin montré que l'utilisation du RIA à la fois pour le ^{13}C et pour ^{18}O permettait d'augmenter de plus de 30% le nombre de signaux auxquels il était possible d'assigner une formule brute unique par rapport à l'utilisation du seul rapport m/z précis et des règles chimiques telles que définies par Fiehn et Kind (Kind and Fiehn, 2007).

2.6 Conclusion

Cette étude démontre l'importance d'un paramétrage soigneusement choisi lors du transfert de méthodes métabolomiques sur de nouveaux instruments afin d'obtenir des résultats les plus justes possibles. Sur des instruments de technologie Orbitrap, les principaux paramètres à ajuster sont l'AGC et la résolution. L'optimisation de ces paramètres, notamment sur les dernières générations d'Orbitraps dits à « haut champs » électrique, apparaît essentielle pour bien contrôler

les phénomènes de charge d'espace qui peuvent exister. Ce contrôle est indispensable pour assurer la qualité des résultats générés, aussi bien en termes de précision de mesure de masse que de RIA. Au vu de l'impact potentiel de ces phénomènes démontré dans cette étude, nous ne pouvons qu'insister sur la nécessité d'évaluer ces interactions entre nuages d'ions au sein des spectromètres de masse à transformée de Fourier.

Ce travail démontre également pour la première fois que l'utilisation du RIA de l'¹⁸O apporte des éléments d'information importants permettant d'affiner l'annotation des signaux observés même si des progrès technologiques peuvent encore être fait pour réduire l'erreur de mesure du RIA de l'¹⁸O.

Enfin, si l'analyse du RIA permet d'améliorer et d'affiner l'attribution de formules brutes à des signaux, elle ne permettra jamais de distinguer d'éventuels métabolites isomères qui présentent exactement la même formule brute. Pour cela, d'autres techniques comme l'acquisition et l'interprétation de spectres de fragmentation devront être utilisées.

3 Interprétation des mécanismes de fragmentation

3.1 Contexte

Pour identifier les métabolites détectés lors d'analyses métabolomiques par HPLC-HRMS, un des moyens les plus utilisés et le plus robuste est l'interprétation de spectres de fragmentation (Creek et al., 2014). Le principe est de piéger les ions et d'augmenter leur énergie interne (soit par collision soit par résonance), ce qui entraîne leur dissociation en fragments de plus petites tailles. Ces fragments sont spécifiques du métabolite étudié et apporte donc des informations majeures quant à sa structure. Deux approches différentes peuvent être utilisées pour exploiter ces résultats. La première est la comparaison du spectre de fragmentation obtenu avec ceux présents dans des bases de données. Cette approche est la plus simple mais ne peut être utilisée que si le composé étudié est bien présent dans les bases de spectres MS² et s'il a été fragmenté dans des conditions comparables. Cette approche se limite donc le plus souvent à l'utilisation de base de données internes acquises sur le même instrument. La seconde se base sur l'utilisation d'outils informatiques permettant de générer automatiquement le spectre de fragmentation d'une molécule à partir de sa structure. Il existe des bases de données qui regroupent des spectres de fragmentation obtenus *in silico* et qui peuvent permettre d'identifier des composés pour lesquels il n'existe pas encore de spectre expérimental (Wolf et al., 2010). Cette approche nécessite d'avoir une connaissance parfaite des mécanismes chimiques mis en jeu lors de la fragmentation des molécules. Cela passe notamment par l'interprétation systématique de nombreux spectres de fragmentation obtenus à partir de différentes

classes de composés. À terme, une connaissance exhaustive des mécanismes mis en jeu permettrait d'envisager l'implémentation d'une solution logicielle qui pourrait automatiser l'identification des métabolites inconnus à partir de leurs spectres de fragmentation.

3.2 Objectifs de l'étude

Dans ce contexte, ce projet vise à étudier en profondeur la fragmentation de deux métabolites modèles : deux acides aminés (une des classes de métabolites les plus étudiées) dont la structure chimique est très proche : l'acide aspartique et l'acide glutamique (Figure 16). En effet ces deux acides aminés ne diffèrent que par un groupement méthylène sur la chaîne latérale. Différentes études ont déjà été menées sur ces espèces (Sekimoto et al., 2014) mais leur dissociation en ionisation négative par électrospray n'a pas été approfondie. Le but est donc d'observer et de rationaliser les différences au sein des spectres de fragmentation de ces deux composés.

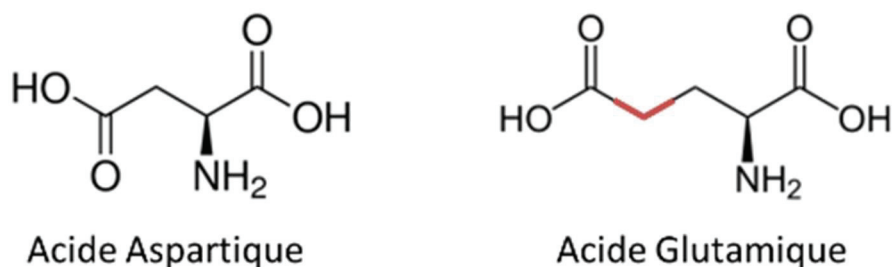


Figure 16 : Structure de l'acide aspartique et de l'acide glutamique

3.3 Approche méthodologique

Pour cette étude, les deux composés ont été analysés en introduction directe (infusion) afin de disposer du temps nécessaire pour la réalisation de nombreuses expériences de fragmentation. Des analyses de fragmentation MS² et MS³ ont été réalisées afin de comprendre au mieux la structure des ions formés. Les fragmentations ont été menées en mode résonant (CID) et non résonant (HCD) pour identifier les mécanismes consécutifs ou compétitifs de fragmentation. Pour chaque expérience, différentes énergies de collision ont été appliquées. Des courbes dites ERMS (*Energy Resolved Mass Spectrometry*) ont été tracées, ce qui permet de suivre l'évolution des différentes espèces formées lorsque l'énergie de collision augmente.

Enfin, une collaboration a été mise en place afin de réaliser des calculs quantiques pour appuyer et rationaliser les phénomènes observés expérimentalement.

3.4 Publication numéro 3

Mechanistic study of competitive releases of H₂O, NH₃ and CO₂ from deprotonated aspartic and glutamic acids: Role of conformation

Pierre Barbier Saint Hilaire, Anna Warnet, Yves Gimbert, Ulli M. Hohenester, Gianluca Giorgi, Marie-Françoise Olivier, François Fenaille, Benoît Colsch, Christophe Junot, Jean-Claude Tabet

Journal of Chromathography B, 2017, 1047, 64-74

Les données supplémentaires liées à cette publication sont présentées en Annexe 6.



Mechanistic study of competitive releases of H₂O, NH₃ and CO₂ from deprotonated aspartic and glutamic acids: Role of conformation



Pierre Barbier Saint Hilaire^a, Anna Warnet^{a,b,*}, Yves Gimbert^c, Ulli Martin Hohenester^a, Gianluca Giorgi^d, Marie-Françoise Olivier^a, François Fenaille^a, Benoît Colsch^a, Christophe Junot^a, Jean-Claude Tabet^{a,b}

^a CEA, iBiTec-S, SPI, LEMM, Metabohub Gif Sur Yvette, France

^b Université Paris VI (UPMC), CNRS UMR 7201, Paris Cedex 05, France

^c Université Grenoble Alpes et CNRS, DCM UMR 5250, Grenoble, France

^d University of Siena, Department of Biotechnology, Chemistry and Pharmacy, Via Aldo Moro, I-53100 Siena, Italy

ARTICLE INFO

Article history:

Received 25 February 2016

Received in revised form 22 July 2016

Accepted 22 August 2016

Available online 24 August 2016

Keywords:

Electrospray high-resolution mass spectrometry

Regioselective dissociation

Unexpected cleavage of aspartate anion

ABSTRACT

The aims of this study were to highlight the impact of minor structural differences (e.g. an amino acid side chain enlargement by one methylene group), on ion dissociation under collision-induced dissociation conditions, and to determine the underlying chemical mechanisms. Therefore, we compared fragmentations of deprotonated aspartic and glutamic acids generated in negative electrospray ionization. Energy-resolved mass spectrometry breakdown curves were recorded and MS³ experiments performed on an Orbitrap Fusion for high-resolution and high-mass accuracy measurements. Activated fragmentations were performed using both the resonant and non-resonant excitation modes (i.e., CID and HCD, respectively) in order to get complementary information on the competitive and consecutive dissociative pathways.

These experiments showed a specific loss of ammonia from the activated aspartate but not from the activated glutamate. We mainly focused on this specific observed loss from aspartate. Two different mechanisms based on intramolecular reactions (similar to those occurring in organic chemistry) were proposed, such as intramolecular elimination (i.e. E_i-like) and nucleophilic substitution (i.e. S_Ni-like) reactions, respectively, yielding anions as fumarate and α lactone from a particular conformation with the lowest steric hindrance (i.e. with antiperiplanar carboxyl groups). The detected deaminated aspartate anion can then release CO₂ as observed in the MS³ experimental spectra. However, quantum calculations did not indicate the formation of such a deaminated aspartate product ion without loss of carbon dioxide. Actually, calculations displayed the double neutral (NH₃+CO₂) loss as a concomitant pathway (from a particular conformation) with relative high activation energy instead of a consecutive process. This disagreement is apparent since the concomitant pathway may be changed into consecutive dissociations according to the collision energy i.e., at higher collision energy and at lower excitation conditions, respectively. The latter takes place by stabilization of the deaminated aspartate solvated with two residual molecules of water (present in the collision cell). This desolvated anion formed is an α lactone substituted by a methylene carboxylate group. The vibrational excitation acquired by [(D-H)-NH₃]⁻ during its isolation is enough to allow its prompt decarboxylation with a barrier lower than 8.4 kJ/mol. In addition, study of glutamic acid-like diastereomers constituted by a cyclopropane, hindering any side chain rotation, confirms the impact of the three-dimensional geometry on fragmentation pathways. A significant specific loss of water is only observed for one of these diastereomers. Other experiments, such as stable isotope labeling, need to be performed to elucidate all the observed losses from activated aspartate

* Corresponding author at: CEA, iBiTec-S, SPI, LEMM, Metabohub Gif Sur Yvette, France.

E-mail address: anna.warnet@upmc.fr (A. Warnet).

and glutamate anions. These first mechanistic interpretations enhance understanding of this dissociative pathway and underline the necessity of studying fragmentation of a large number of various compounds to implement properly new algorithms for *de novo* elucidation of unknown metabolites.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In metabolomics, the design of new reliable algorithms to elucidate *de novo* the structures of metabolites (from MS and MS² spectra) is based on (i) general chemical rules for the determination of elemental compositions from electrospray ionization (ESI) mass spectra [1,2], (ii) fragmentation rules collected from the literature, as proposed by commercial software tools such as ACD fragmenter [3] and Mass Frontier [4], (iii) *in silico* tools providing combinatorial fragmentation, such as MetFrag [5] and (iv) dissociation principles deduced from the systematic interpretation of many high-resolution fragmentation spectra for various classes of compounds [6,7]. Many chemical mechanisms are used to rationalize fragmentations but some need to be revised.

Multiple forms of poly-functional molecules generated at the end of the ESI process (i.e. after desorption of multiply-charged aggregates from droplets, followed by their desolvation) can co-exist and lead to co-fragmentations. This point is not often taken into account when interpreting MS² spectra, potentially leading to partial understanding of fragmentation mechanisms [8,9]. For molecules commonly studied in metabolomics and lipidomics, such as amino acids or phospholipids, both canonic (can) and zwitterion (zw) structures can potentially be produced through the ESI process [10–15]. Actually, if molecules in gas phase exist mainly in one form, this is not always the case for ions provided by ESI [16]. Molecules can be ionized in several charged tautomeric forms depending on the source and desolvation conditions. Furthermore, these structures can also be represented as folded or unfolded conformations [10,17], and stabilized by interactions involving mainly hydrogen bonds [18,19] or salt bridges (a common feature of compounds with permanent charges) [20–22]. Consequently, charged species with various structures can be obtained from a single compound under ESI conditions.

Moreover, such naked ions can be selected to be activated by collision (collision-induced dissociation, CID) as the generic name [23], the most widely used activation mode [24,25]. These collisional activations may occur from ion beams by fast heating [26,27] (as a non-resonant mode), either at low collision energy range (few tenths of eV) [27,28] or at high kinetic energy (several keV) [26,29]. At low energy, selected ions are vibrationally excited in their fundamental electronic state and subsequently dissociate by charge-promoted processes [30,31]. Under keV collision conditions, ions are vibrationally excited in both their fundamental and excited electronic levels [30] leading to both the charge-promote and charge-remote processes [31]. Alternatively, collisional activation can take place through slow heating [32,33] by ion storage experiments [34]. Such a resonant excitation mode is hereafter called CID [35] whereas the non-resonant excitation mode using ion beam collisions is called higher collision dissociation (HCD) [36]. CID mode favors competitive dissociations [33,37,38], in contrast to the HCD mode, which leads to consecutive dissociations [38]. In resonant excitation mode, a specific dipolar excitation with the motion frequency of the precursor ion is applied in the ion trap. This frequency cannot activate the product ion cloud characterized by higher motion frequencies. Consequently, only the competitive product ions of the first generation are detected. Exceptionally those of the second generation may sometimes appear if (i) enough

internal energy is carried by the first generation product ions and (ii) their dissociations required sufficiently low energy transition states and/or large frequency factors [39]. In the case of HCD mode, both survival precursor and product ions are activated by consecutive collisions and can dissociate into fragment ions of the second and *n*th generation. These two excitation modes yield complementary information for interpretation of mechanistic fragmentations [38,39]. Combined with sequential MSⁿ experiments, they allowed concomitant and consecutive processes to be distinguished in the case of linked double neutral losses.

Since the charge can be competitively located at different positions within poly-functional molecules, many different dissociative pathways can occur simultaneously. They all have to be considered as they can yield distinct product ions. Moreover, ion conformations in the gas phase may also play important roles in orientating the dissociation processes. For instance, if the charge is distant from the leaving group in a rigid conformation, the ion may isomerize prior to dissociation into an ion-induced dipole complex in which one (or several) reversible proton transfers may take place [40,41]. As significantly distinct dissociations may result from slight structural modifications, complete information on such processes is needed to understand ion cleavage mechanisms fully. Further investigations using quantum calculations can also provide relevant information to confirm or refute putative fragmentation mechanisms.

One of our objectives regarding metabolomics is to implement efficient software solutions to expedite the identification of unknown metabolites in biological samples from their MS² spectra. With that aim, we began the thorough manual inspection and the interpretation of the whole set of MS² spectra generated from our chemical database (including a thousand compounds) using high-resolution/high-mass accuracy mass spectrometers [42,43]. We are creating a mass spectral library with well-characterized fragment ions and neutral losses for more efficient and reliable database identification of unknown compounds.

Herein, we focus on the fragmentation of anions from acidic amino acids produced in ESI negative mode and submitted to low-energy collisions. Several fundamental studies have already been done on these anions in terms of (i) thermochemistry (e.g. acidity measurements) [19,20,44,45–47] and (ii) charged structures (canonic and zwitterion forms, charge location [10–15]). However, the dissociations of deprotonated amino acids (i.e., intact natural anions, non-modified, non-cationized/anionic species and non-engaged in complexes) have not yet been extensively investigated [48–50]. This work was accomplished essentially on deprotonated aspartic (D) and glutamic (E) acids, which only differ by a methylene group on the side chain. Such a minor structural modification suffices to generate a specific ammonia loss from [D–H][–] under low and high conditions [51]. A possible stepwise mechanism based on formation of the NH₂[–]/neutral complex has been proposed in the latter activation conditions [51]. However, this ion-neutral intermediate should be relatively unstable (due to the very weak gas phase acidity of ammonia), and is unlikely to occur. Thus, an alternative mechanism, which should be consistent with the observed consecutive processes, should be envisaged. In addition, specific water loss from the *cis* diastereoisomer of deprotonated cyclopropanic acidic amino acids was also observed and demonstrated the role of the rigid conformation on the rate constant.

Lastly, quantum chemical calculations were performed to support the mechanistic interpretation.

2. Material and methods

2.1. Chemicals and reagents

Aspartic acid (D), glutamic acid (E), isopropanol and deionized water (LC-MS grade) were from Sigma Aldrich (St. Quentin Fallavier France). The following diastereomers: (2R,1'S,2'S,3'R)-2-(2'-carboxy-3'-phenylcyclopropyl) glycine, (2R,1'S,2'S,3'S)-2-(2'-carboxy-3'-phenylcyclopropyl) glycine and (2R,1'R,2'S,3'R)-2-(2'-carboxy-3'-phenylcyclopropyl) glycine were generously provided by R. Pellicciari's group at the University of Perugia [52]. The stereochemistry of these compounds was expressed relatively to the phenyl substituent for $E_{\text{CycPh}}(c, t)$, $E_{\text{CycPh}}(t, c)$ and $E_{\text{CycPh}}(t, t)$.

2.2. Mass spectrometry

2.2.1. Experiments on aspartic and glutamic acids

Aspartic acid (D) and glutamic acid (E) were studied using a Dionex Ultimate 3000 UPLC system coupled to a high-resolution mass spectrometer Orbitrap Fusion [53] (Thermo Fisher Scientific, Courtaboeuf, France). Flow injection analyses were performed at a flow rate of 100 $\mu\text{L}/\text{min}$ with water/isopropanol, LC-MS grade, (50/50), as mobile phase. Energy-resolved mass spectrometry (ERMS) experiments were performed under both the CID (radial resonant excitation with a low mass cut-off [54], LMCO, 25% of the precursor ion m/z value) and HCD (non-resonant excitation without LMCO except the low mass limit of the Orbitrap analyzer at m/z 50) conditions (0% to 90% or 100% normalized collision energy [55], NCE, respectively). Sequential MS³ experiments were also performed under HCD conditions. In each case, multi-event spectra were recorded at a resolution of 60 k (full width at half maximum, FWHM, at m/z 200). Operating conditions for the ESI source were as follows: spray voltage -2.5 kV; sheath gas 20 au; aux gas 7 au; sweep gas 1 au, ion transfer tube temperature 300 °C and vaporizer temperature 175 °C. Nitrogen was used as source gas and collision gas in ion-routing multipole used for HCD and helium as both buffer and collision gas for CID in linear ion trap as LTQ.

2.2.2. Experiments on

2-(2'-carboxy-3'-phenylcyclo-propyl)glycine diastereomers

Electrospray measurements on 2-(2'-carboxy-3'-phenylcyclopropyl)glycine methanolic solutions (1×10^{-4} M) were performed via direct infusion at a flow rate of 5 $\mu\text{L}/\text{min}$ in an ESI source using an LCQ-DECA ion trap (Thermo Finnigan, Bremen, Germany). Operating conditions for the ESI source were as follows: spray voltage -4.5 kV; capillary temperature 200 °C; sheath gas (nitrogen) flow rate 0.75 L/min. MSⁿ product ion experiments were carried out inside the ion trap by isolating the precursor ion and then by applying a supplementary potential for CID; buffer and collision gas: He; collision energy: 20–40% arbitrary units.

2.3. Quantum calculations

DFT calculations [56,57] of molecular properties with Gaussian 09 software (Wallingford, USA) were performed using the B3LYP/6-31+G(d,p) methodology (corrected from ZPE with energy units in kJ/mol). Critical points (reactants, transition-state structures, intermediates, and products) were fully characterized as minima or first-order saddle points by diagonalizing the Hessian matrices of the structures at the corresponding level of optimization (see Supplementary material).

3. Results and discussion

3.1. Specific loss of ammonia from $[\text{D}-\text{H}]^-$ and consecutive processes

The HCD spectra (NCE of 15%, recorded at a resolution of 60 000, see Table S1, Fig. 1a and b) of both the $[\text{D}-\text{H}]^-$ (m/z 132.0300) and $[\text{E}-\text{H}]^-$ (m/z 146.0457) anions, display intense signals at m/z 88.0401 and m/z 102.0558 respectively, corresponding to a CO_2 loss (see Table S1). However, striking differences characterize these two HCD spectra. That one of the $[\text{D}-\text{H}]^-$ ion (see Fig. 1a) presents an intense signal associated with a rather unexpected NH_3 loss (ion at m/z 115.0034) and a second representing a formal loss (noted in inverted commas, i.e., ' $\text{C}_n\text{H}_m\text{N}_r\text{O}_s$ ') of 61.0163 u [CH_3NO_2], which could be attributed to the double (NH_3+CO_2) loss, ion at m/z 71.0136]. These losses occurred in competition with a less abundant loss of water. The opposite trend was observed in the HCD spectrum of the $[\text{E}-\text{H}]^-$ species since the signal at m/z 128.0351, corresponding to the H_2O loss, was very intense (see Fig. 1b).

The increase of the collision energy up to NCE=70% (see Fig. 1c and d) resulted in the complete disappearance of both the $[(\text{D}-\text{H})-\text{NH}_3]^-$ (m/z 115.0034) and $[(\text{D}-\text{H})-\text{H}_2\text{O}]^-$ (m/z 114.0194) as well as the $[(\text{E}-\text{H})-\text{H}_2\text{O}]^-$ (m/z 128.0351) ions. Instead, two series of new fragment ions were formed. The fragmentations of $[\text{D}-\text{H}]^-$ led to the formal 61.0163 u loss previously reported (ion at m/z 71.0137) and the ' $\text{C}_2\text{H}_4\text{O}_2$ ' loss yielding the product ion at m/z 72.0088. On the other hand, the activated $[\text{E}-\text{H}]^-$ precursor ion led to the m/z 74.0245 and m/z 71.0135 ions, respectively, due to the formal ' $\text{C}_3\text{H}_4\text{O}_2$ ' and ' $\text{C}_2\text{H}_5\text{NO}_2$ ' losses with the most abundant product ion at m/z 59.0135 corresponding to the formal neutral loss of ' $\text{C}_3\text{H}_5\text{NO}_2$ '. These neutral releases can result either from one or two neutral losses.

Interestingly, the CID spectra (NCE = 23% and 70%, see Fig. S1) of $[\text{D}-\text{H}]^-$ (m/z 132.0299) and $[\text{E}-\text{H}]^-$ (m/z 146.0456) compared with their respective HCD spectra (NCE = 15% and 70%, see Fig. 1) present particular features:

- at lower NCE values (CID at 23% and HCD at 15%), the collision spectra (Fig. 1 and Fig. S1) were similar in both activation modes except for (i) the relative peak intensities and (ii) the absence of a peak related to the loss of 61.0164 u from the aspartate dissociation. Indeed, its CID spectrum (Fig. S1a) displayed only peaks at m/z 115.0034, m/z 114.0194 and m/z 88.0401, respectively attributed to product ions formed from the NH_3 , H_2O and CO_2 neutral losses. For $[\text{E}-\text{H}]^-$, i.e. peaks at m/z 128.0350 $\{[(\text{E}-\text{H})-\text{H}_2\text{O}]^-\}$ and m/z 102.0558 $\{[(\text{E}-\text{H})-\text{CO}_2]^- \}$ were also only observed (Fig. S1b);
- at higher NCE values (CID at 70% and HCD at 70%), the CID spectra (Fig. S1c and d) still displayed the same peaks with very similar abundances as observed at NCE = 23% (except the disappearance of signals related to the precursor $[\text{D}-\text{H}]^-$ and $[\text{E}-\text{H}]^-$ ions). However, peaks displayed by HCD spectra at m/z 72.0088 and m/z 71.0137 for aspartate (Fig. 1c) and at m/z 74.0245, m/z 71.0135 and m/z 59.0135 for glutamate (Fig. 1d) were missed in their respective CID spectra (Fig. S1c and d). Thus, beyond NCE = 25%, the relative abundances of product ions became constant in CID and no other product ions were formed, unlike what was observed in the HCD mode.

The observed differences between resonant (CID) and non-resonant (HCD) excitation modes suggest that the neutral losses missed should be produced through consecutive processes.

Altogether, these data tend to suggest that the $[(\text{D}-\text{H})-\text{CH}_3\text{NO}_2]^-$ (m/z 71.0137) and $[(\text{D}-\text{H})-\text{C}_2\text{H}_4\text{O}_2]^-$ (m/z 72.0088) ions as well as the $[(\text{E}-\text{H})-\text{C}_2\text{H}_5\text{NO}_2]^-$ (m/z 71.0135) and $[(\text{E}-\text{H})-\text{C}_3\text{H}_4\text{O}_2]^-$ (m/z 74.0245) ions, were likely produced

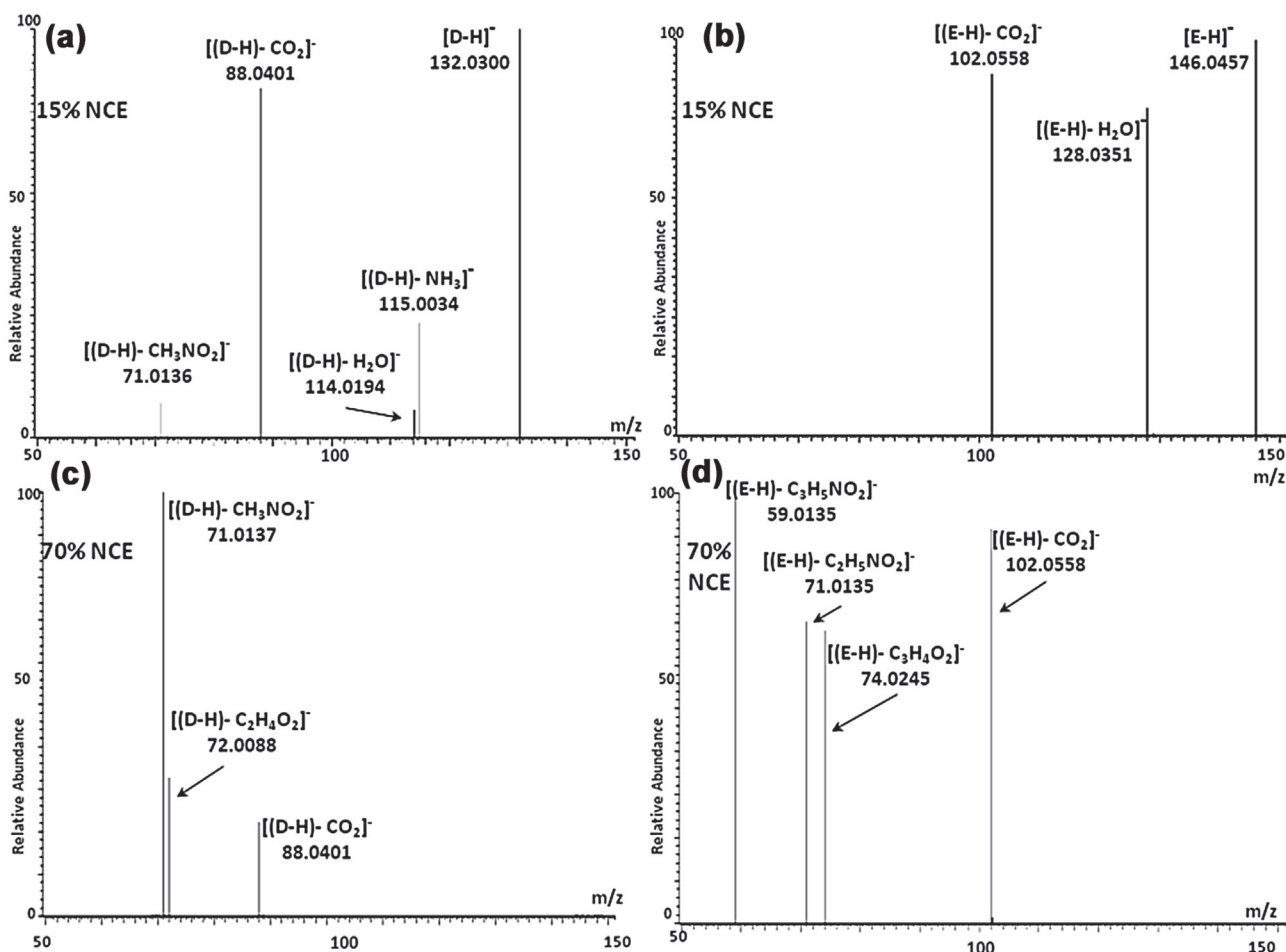


Fig. 1. Non-resonant collisional spectra (HCD mode) of: (a, c) [D-H]⁻ (*m/z* 132.0300) and (b, d) [E-H]⁻ (*m/z* 146.0457) recorded under (a, b) 15% of NCE and (c, d) 70% of NCE conditions for each selected anion.

by consecutive dissociations from [(D-H)-NH₃]⁻ (*m/z* 115.0034) and/or [(D-H)-H₂O]⁻ (*m/z* 114.0194) and [(E-H)-H₂O]⁻ (*m/z* 128.0351), respectively (see Fig. 1).

3.2. Sequential MS³ and ERMS experiments: HCD versus CID processes for confirmation of consecutive dissociations

To gain more information on consecutive dissociations, sequential MS³ experiments (in HCD mode) were performed on [(D-H)-NH₃]⁻ (*m/z* 115.0039) and [(D-H)-H₂O]⁻ (*m/z* 114.0194) species (Fig. S2a and b). They revealed that [(D-H)-NH₃]⁻ can promptly lose CO₂ (see Fig. S2a), in contrast to [(D-H)-H₂O]⁻, which releases the competitive C₂H₂O and CHNO neutral losses to yield the *m/z* 72.0088 and *m/z* 71.0136 product ions, respectively (see Fig. S2b). These fragmentations will not be studied further. Abundance of the *m/z* 70.0295 ion ([D-H)-(H₂O+CO₂)⁻) was significantly lower. The HCD spectrum (Fig. S2c) of [(E-H)-H₂O]⁻ (*m/z* 128.0351) revealed a distinct trend with the preferential releases of 'H₂CO₂' and 'C₂H₄O₃' (yielding the *m/z* 82.0295 and *m/z* 52.0189 ions, respectively), even if an ion resulting from the CO₂ release can be observed at *m/z* 84.0452 only at lower excitation energies.

The ERMS breakdown from the sequential (HCD) MS³ experiments of the [(D-H)-CO₂]⁻ (*m/z* 88.0401) and [(E-H)-CO₂]⁻ (*m/z* 102.0558) product ions (Fig. S3a and b), revealed that these ions were very stable and their respective dissociations did not exceed 5%. On the other hand, from these experiments (Fig. S3) on deprotonated aspartic acid, it appears that the 'CH₃NO₂' neutral

loss (61.01638 u), attributed to the consecutive (NH₃+CO₂) neutral release, occurs first from the NH₃ loss followed by the CO₂ release. Interestingly, the ERMS breakdown curves (in HCD mode) obtained for [D-H]⁻ roughly present similar trends on those published some years ago using a triple quadrupole instrument [58]; however, e.g. the consecutive losses of (H₂O+CO₂) yielding the *m/z* 70 ion (measured at low resolution) seems to be very more abundant than in the HCD mode where the *m/z* 70.0295 ion appears negligible. Furthermore, in these ERMS breakdown, the collision energy scales in center mass energy given in eV vs. normalized collision energy in% of ERMS, when triple quadrupole collision cell and in ion-routing multipole are respectively used, are difficult to be compared [39].

The extent of these competitive and consecutive processes depends strongly on the excitation energy and the activation mode (i.e., resonant as CID versus non-resonant as HCD processes). This dependency is reinforced by ERMS experiments of [D-H]⁻ and [E-H]⁻ species (see Fig. 2).

The progression of the ERMS breakdown curves from the HCD experiments (see Fig. 2) shows that:

- [(D-H)-NH₃]⁻ (*m/z* 115.0034), [(D-H)-H₂O]⁻ (*m/z* 114.0194) and [(D-H)-CO₂]⁻ (*m/z* 88.0401) relative abundances increase to an optimum then decrease (see Fig. a). A similar behavior was observed for [(E-H)-H₂O]⁻ (*m/z* 128.0351) and [(E-H)-CO₂]⁻ (*m/z* 102.0558) (see Fig. 2b);
- the formal 'CH₃NO₂' and 'C₂H₄O₃' releases present a different trend from [D-H]⁻, since relative abundances of [(D-H)-(NH₃+CO₂)⁻ (*m/z* 71.0136) and

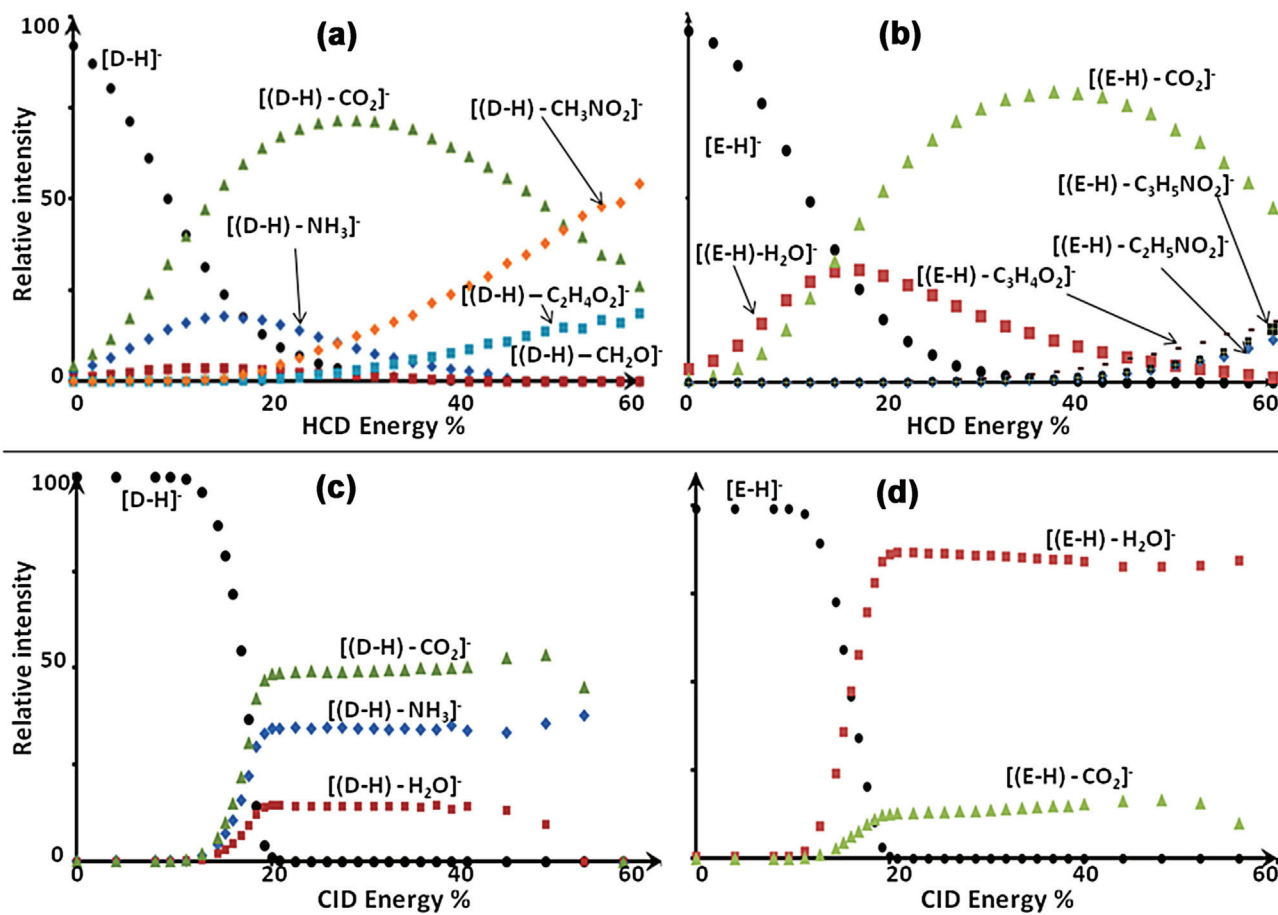


Fig. 2. HCD and CID ERMS breakdown curves of $[D-H]^-$ (m/z 132.0300) and $[E-H]^-$ (m/z 146.0457) in HCD mode ((a) and (b)) and CID mode ((c) and (d)). Only the product ions representing more than 2% of the relative abundance are reported in these graphs.

$[(D-H)-(H_2O+C_2H_2O)]^-$ (m/z 72.0088) increase throughout the NCE range. In the same way, the formal $[(E-H)-C_3H_4O_2]^-$ (m/z 74.0245), $[(E-H)-C_2H_5NO_2]^-$ (m/z 71.0135), and $[(E-H)-C_3H_5NO_2]^-$ (m/z 59.0135) product ions increase without reaching an optimum (see Fig. 2b).

As mentioned above, ERMS breakdown curves obtained for $[D-H]^-$ and $[E-H]^-$ under HCD conditions both exhibit preferential CO_2 neutral loss. This disappears at higher NCE values in favor of the formal ' CH_3NO_2 ' and ' $C_2H_4O_2$ ' losses [i.e., (NH_3+CO_2) and (H_2O+CH_2CO) , respectively] from $[D-H]^-$ contrasting with those of ' $C_3H_4O_2$ ', ' $C_2H_5NO_2$ ', and ' $C_3H_5NO_2$ ' (not discussed anymore) observed from $[E-H]^-$. Mainly, the occurrence of these specific processes requires stepwise pathways involving precursor anion isomerization into ion-dipole complexes prior to dissociation. Such stepwise processes have already been described, in particular, for dissociations of deprotonated amino acids [50,51].

The striking differences observed between ERMS breakdown curves performed in resonant (CID, Fig. 2) and in non-resonant (HCD, Fig. 2) excitation modes were rationalized by considering their respective excitation mode properties especially, those of ion trapping experiments (CID):

- even above 23% of NCE, none of the peaks related to consecutive dissociation of aspartate and glutamate were observed in their CID spectra. This can be explained by selective activation of the precursor ion by its specific motion frequency (slow heating due to the large numerous very low-energy collisions with the buffer gas). The product ion motions, being characterized

by higher frequencies (proportional to the m/z ratios), cannot be excited and are stored in the bottom of the pseudo-potential well. Consequently, the consecutive dissociations are unfavored (*vide supra*, see Introduction), except when the residual internal energy of direct product ions is high enough to allow consecutive processes, as shown by other studies [38,59,60]. However, consecutive processes may be observed when they are characterized by sufficiently low-energy transition states and/or large frequency factors [61]. Under HCD conditions, the ion cloud, consisting of survivor precursor ion and product ions, is activated by energetic consecutive collisions (fast heating) with the gas target. This may lead to favorable consecutive dissociation pathways (see Fig. 1):

- in resonant mode, the absence of variation of product ion relative abundances (despite the disappearance of precursor ion) beyond NCE = 25% signifies that the precursor ion is ejected from the pseudo potential well of the LTQ trapping field [32] prior to its decomposition. This happens when its kinetic energy increases;
- in CID the abundances of both the competitive NH_3 and H_2O losses, reinforced relative to that of CO_2 , can be due to the long time-window for ion dissociation in this mode and this effect can result in a decrease of kinetic shift effect [39,62] allowing fragmentations of relatively low rate constants (characterized by low activation energy and weak frequency factor). For the glutamate anion this should explain the inversion of the $\{[(E-H)-CO_2]^- \} / \{[(E-H)-H_2O]^- \}$ ion abundance ratio at lower collision energy. An additional interpretation can be proposed considering that residual molecules of water (or other solvent) in the ion trap may yield relaxation of the competitive fragment ions

by solvation/desolvation processes according to their structure [59].

3.3. Proposed interpretation of the specificity of the ammonia loss from aspartate anion

In order to fully interpret the specific release of ammonia from $[D-H]^-$, it is first necessary to investigate the structure of the deprotonated aspartic acid. To describe the neutral aspartic acid, both canonic (D_{can}) and zwitterion (D_{zw}) forms [20] may be proposed. However, the D_{zw} form (see Scheme 1) is only stable when solvated by two or more molecules of water [63], which should be possible in the desorbed charged aggregate or in the deprotonated dimeric anion [20]. As it cannot exist as a species free of solvent, even within a salt bridge form, it was no longer considered. Deprotonation of canonic aspartic acid may occur on either one of the carboxylic acid groups in a competitive manner, despite a possible proton transfer between these groups (transfer energy cannot exceed 21 kJ/mol [64,65]). Deprotonation may yield the two canonic $[D_{can(i)}-H]^-$ and $[D_{can(ii)}-H]^-$ forms (see Scheme 1) in the ESI source [20]. In the $[D_{can(i)}-H]^-$ form, a hydrogen bond, between the amino and carboxylic acid groups, should partially stabilize this anion (see Scheme 1). A more stable hydrogen bond between the two carboxylic groups could also be considered as solvating the proton of $[D_{can(ii)}-H]^-$ ion [66]. However, such a form does not survive collisional activation since the loss of CO_2 has already occurred.

The weaker hydrogen bond stabilization for the $[D_{can(i)}-OH]^-$ form maintains the conformation such as both the $-COO^-$ and $-COOH$ groups are antiperiplanar. This conformation presents the lowest steric hindrance and allows the approach of (i) the NH_2 group to the proton of carboxylic acid on the side chain and (ii) the $-COO^-$ site to the carbon atom geminated to the amino group as shown using a Newman representation (see Scheme 2).

Such a conformation provides evidence for NH_3 elimination without the involvement of a deprotonated ammonia molecule as previously proposed [67]. Indeed, this conformation, characterized by reduced steric hindrance, leads to a SN_i -like mechanism involving a nucleophilic attack by the alkoxide site to the carbon atom bearing the amino group. This group is close enough to the $-COOH$ site to undergo a proton transfer, which promotes the ammonia release. This reaction takes place either directly or by a stepwise process via the formation of an ion-dipole complex prior to NH_3 elimination. It results in the formation of a deprotonated α lactone carboxylic acid species at m/z 115 (see Scheme 2). In turn, this lactone would generate an ion at m/z 71 (after releasing CO_2) as deprotonated acrylate acid. Concomitant (NH_3+CO_2) loss can also take place to yield this ion at m/z 71. These interpretations corroborate the results of sequential MS^3 experiments: the stability of $[(D-H)-CO_2]^-$ and the reactivity of $[(D-H)-NH_3]^-$.

Alternatively, another pathway should be proposed. From the same conformation (ion m/z 132 in Scheme 2), in opposite of unlike a mobile proton of the enolizable methylene group (i.e. in α of $-CO_2H$) anti to the NH_2 group, should be close to the $-O^-$ of the carboxylate group (Scheme S1). This proton can migrate by a 1–4 transfer to neutralize the carboxylate group, allowing the release of ammonia and yielding the deprotonated fumaric acid at m/z 115. This ion can then potentially lose CO_2 giving rise also to formation of an acrylate anion. This mechanism could be likened to an E_i process.

3.4. Quantum calculation

Experimentally, in our study it was previously shown that both the $[(D-H)-NH_3]^-$ and $[(D-H)-'CH_3NO_2']^-$ product ions are detected in variable relative abundances, correlated with the applied excitation energy (see Fig. 3a). The mechanism of the formal $'CH_3NO_2'$ release must correspond to the (NH_3+CO_2) release

(as shown by the MS^3 experiment of the $[(D-H)-NH_3]^-$ ions). This double loss could either result from direct concomitant (NH_3+CO_2) losses or from a stepwise consecutive dissociation pathway via the formation of stable $[(D-H)-NH_3]^-$ ion.

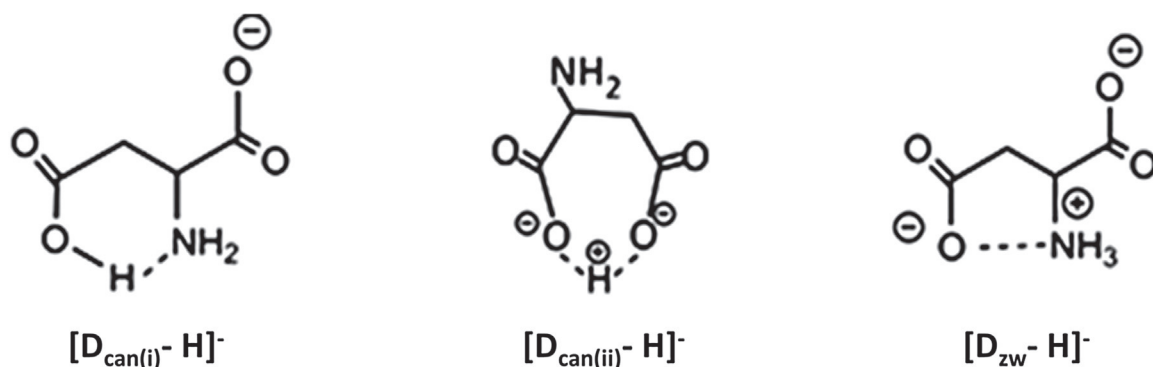
3.4.1. Concomitant double (NH_3+CO_2) loss

The double (NH_3+CO_2) loss from the $[D_{can(i)}-H]^-$ form (a) (see Fig. 3) is demonstrated by DFT calculations [B3LYP/6-31+G (d, p) level with values corrected from zero-point of energy] [55,56] to be exothermic by -3.7 kJ/mol. Furthermore, transition states were found for this studied dissociation. Concomitant pathway DFT calculations showed that the most stable canonic ion has its α amino carboxylic acid group deprotonated (a, Fig. 3). The process first starts by the $[D_{can(i)}-H]^-$ isomerization: (a) \rightarrow (b) step (Fig. 3). Instead of passing through a direct 1–6 transfer, the proton migration is assisted by the amino group acting as a relay group. Such a relay mechanism has already been described for the gas phase H/D exchanges [68,69]. This isomerization is endothermic by 17.5 kJ/mol (Fig. 3) and 24.6 kJ/mol were needed.

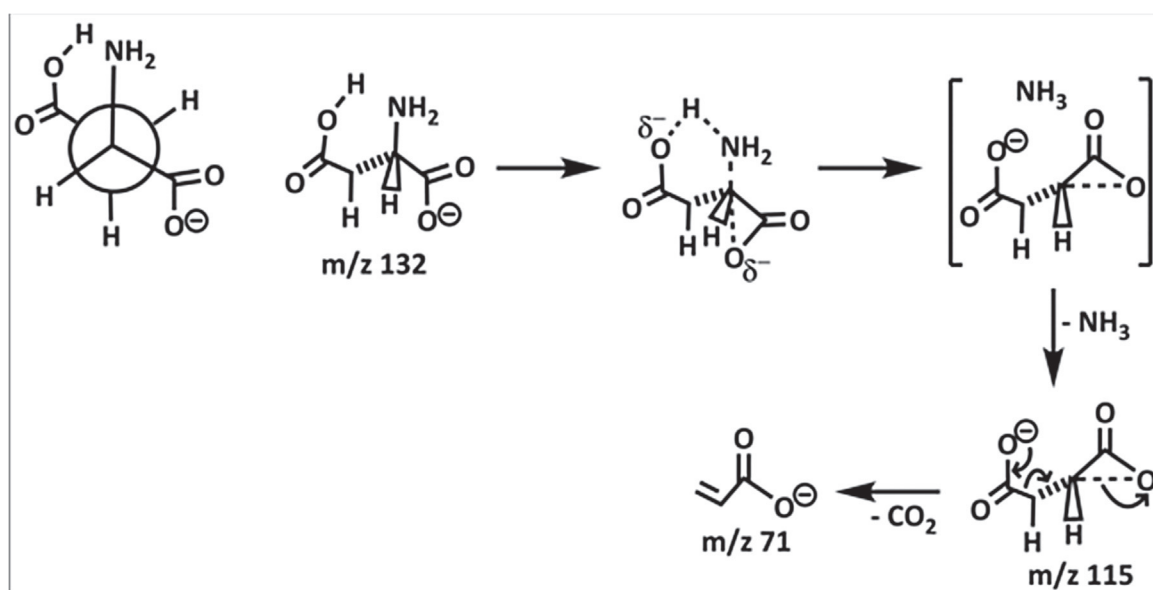
After the isomerization, the dissociation from the deprotonated form (b) might involve a deprotonated zwitterion-like form (TS_H). Gradually, while the NH_2 group is detached from the carbon atom, it carries with it the proton of the carboxylic acid in the α position; the NH_3 group begins to move away reaching an endothermic energy level of 144.2 kJ/mol (as transition state). This departure concomitantly induces a loss of CO_2 (rearrangement of carboxylate group in the β position) ultimately giving an ion at m/z 71, rather than just the loss of ammonia (ion at m/z 115). The overall exothermicity of this pathway was calculated as -3.7 kJ/mol, although the level of the higher activation energy is 161.7 kJ/mol from the initial state (a). These calculations apparently were in contradiction with the experimental results. In addition, they invalidated the proposed mechanisms, as both the E_i -like (Scheme S1) and SN_i (Scheme 2) processes explain the loss of ammonia. This gives rise to formation of enough stable ion at m/z 115 (such as either fumarate or α lactone carboxylate anions) to be detected, followed consecutively by that ion of carbon dioxide (ion at m/z 71). The questions that arise are (i) why is there loss of ammonia and (ii) why does the concomitant loss increase at higher energy?

3.4.2. Collision energy dictates fragmentation: consecutive versus concomitant processes

It is likely that some low-energy collisions allow the vibrational energy relaxation of the excited ions (slow heating) (33) to modify the fragmentation pathways [70]. Another explanation could be considered if the negative charge is stabilized by solvation with one or several molecules from residual solvent/water present in both the LTQ and ion-routing multipole collision cells (CID and HCD) [59,71–73]. To verify the validity of the latter proposition, the DFT calculations were performed considering solvation of the $[D_{zw}-H]^-$ form (impossible to minimize from the naked ion, as previously mentioned, see Scheme 1). Introduction of two water molecules yields converged $[(D_{zw}-H)+2H_2O]^-$ structure with an NH_3 group bonded to a C atom (see Fig. 4A). Starting from this bis-hydrated structure, the elongation of the C–N bond yields loss of NH_3 without concomitant loss of CO_2 . This bis-hydrated carboxylate played an important role in the stabilization of the resulting deamination product as illustrated in Fig. 4B. Indeed, the β CO_2^- group strongly interacts with a molecule of water, while the α CO_2^- efficiently stabilizes the cationic center created after the departure of NH_3 in a hydrated α lactone-like group organization corresponding to $[(D_{zw}-H)-(NH_3+2H_2O)]^-$. It is noteworthy that, taking into account the possible solvation, the suggested E_i -like process (Scheme S1) was not found to be possible by DFT calculations. Thus, this mechanism was rejected and will not be discussed further.



Scheme 1. Proposed structure of the deprotonated aspartic acid with hydrogen bonded $[D_{\text{can}}-H]^-$ canonic and salt bridged $[D_{\text{zw}}-H]^-$ zwitterion forms. The latter form does not converged and was not considered, except when it is bis-hydrated.



Scheme 2. Proposed S_Ni -like reaction assisting the ammonia elimination within an antiperiplanar conformation of the COOH/COO^- groups as shown in the Newman projection yielding a deprotonated carboxylic alpha lactone acid *via* an ion-dipole complex.

The two bis-hydrated anions from aspartate and deaminated aspartate (generated through the S_Ni pathway) were not detected in the direct HRMS analysis. In addition, for product ion selection during MS^3 experiments, after transferring the ion cloud (consisting of the survivor aspartate and product anions formed during the HCD process) from the ion routing multipole to ion trap (i.e. the LTQ cell), neither hydrated product ion (e.g., $[(D-H)-\text{NH}_3+2\text{H}_2\text{O}]^-$) was detected in HRMS. Such hydrated product anions are very likely fleeting and were promptly dehydrated during their injection either in the C-trap cell (direct HRMS analysis) or in the LTQ cell (MS^3 experiments).

The deaminated aspartate as acetate substituted by an α lactone appears as a weakly stabilized species. In the sequential MS^3 experiment on this $[(D-H)-\text{NH}_3]^-$ product ion (m/z 115), a prompt decarboxylation took place during its selection and without its excitation (NCE=0%), and the observed decarboxylation roughly 10% (Fig. S3c). Modeling to rationalize all the sequences involving hydrated form of $[D_{\text{zw}}-H]^-$ is in progress.

Finally, it appears from the calculations, that both the deamination and decarboxylation processes can take place through two pathways. The double $(\text{NH}_3+\text{CO}_2)$ loss may occur via a concomitant process (in one step) and/or through a stepwise process *via* a double hydration of the intermediate when it begins to release ammonia. The concomitant process is very likely favored in higher energy

conditions, whereas the consecutive pathway should be enhanced at lower energy collision, where the hydration reaction of the intermediate as ion-molecule reaction is favored. This investigation is in progress using D and ^{13}C labeling.

3.5. A feature in favor of the occurrence of a conformational effect

The proposed mechanisms (see Scheme 2 and Scheme S1) indicate that the relative three-dimensional position of the functional groups in these anions plays an important role. On the other hand, the conformation characterized by the lowest steric hindrance has the advantage of bringing interactive groups to enable formation of a relevant intermediary for the ammonia release from aspartate. To investigate this and the role of ion conformation further, we studied glutamic acid-like structures: 2-(2'-carboxy-3'-phenylcyclo-propyl) glycine diastereomers (see Fig. 5) characterized by a non-flexible structure due to the tight 3-membered ring and so presenting a stable cis/trans stereochemistry relative to the phenyl group on the ring.

Under CID conditions, the three diastereomeric $[\text{E}_{\text{CycPh}}(\text{t,c})-H]^-$, $[\text{E}_{\text{CycPh}}(\text{c,t})-H]^-$, and $[\text{E}_{\text{CycPh}}(\text{t,t})-H]^-$ anions (m/z 234) fragmented similarly to give, in particular, ions at m/z 190 (loss of CO_2), m/z 172 (loss of $\text{CO}_2+\text{H}_2\text{O}$), m/z 146 (loss of 2CO_2). While a very low or even zero loss of water (ion m/z 216) was observed for the *trans*

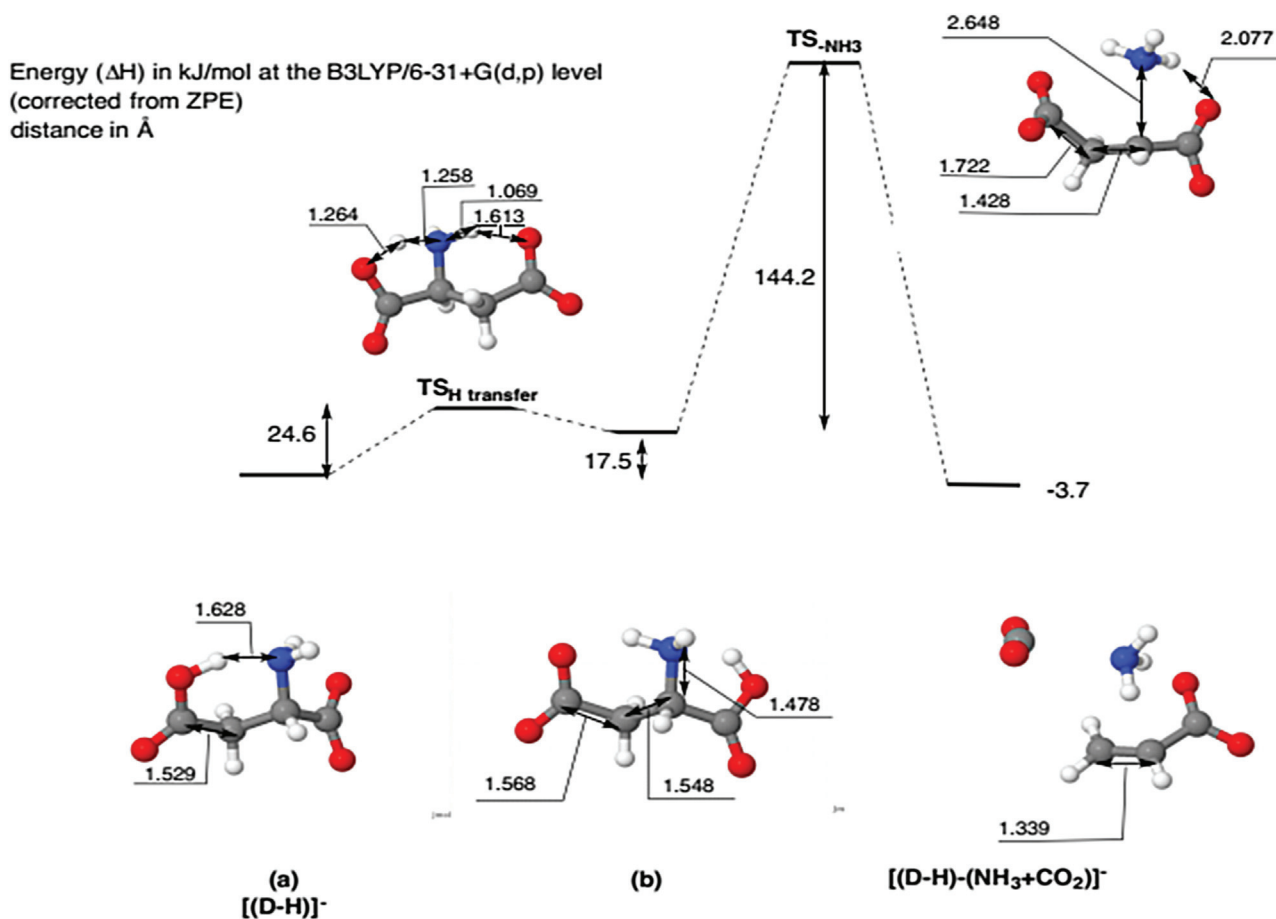


Fig. 3. Potential energy profiles for the most favorable $[D-H]^-$ cleavages (the concomitant (NH_3+CO_2) losses) relative to the initial state where carboxylate is located at the α position to the amine site (a) rather than in the β position at the side chain (b). This reaction proceeds by a weakly endothermic relay isomerization through a zwitterion-like low energetic TS followed by a second TS of 144.2 kJ/mol yielding the $[(D-H)-(NH_3+CO_2)]^-$ (See Table S2 for cartesian coordinates).

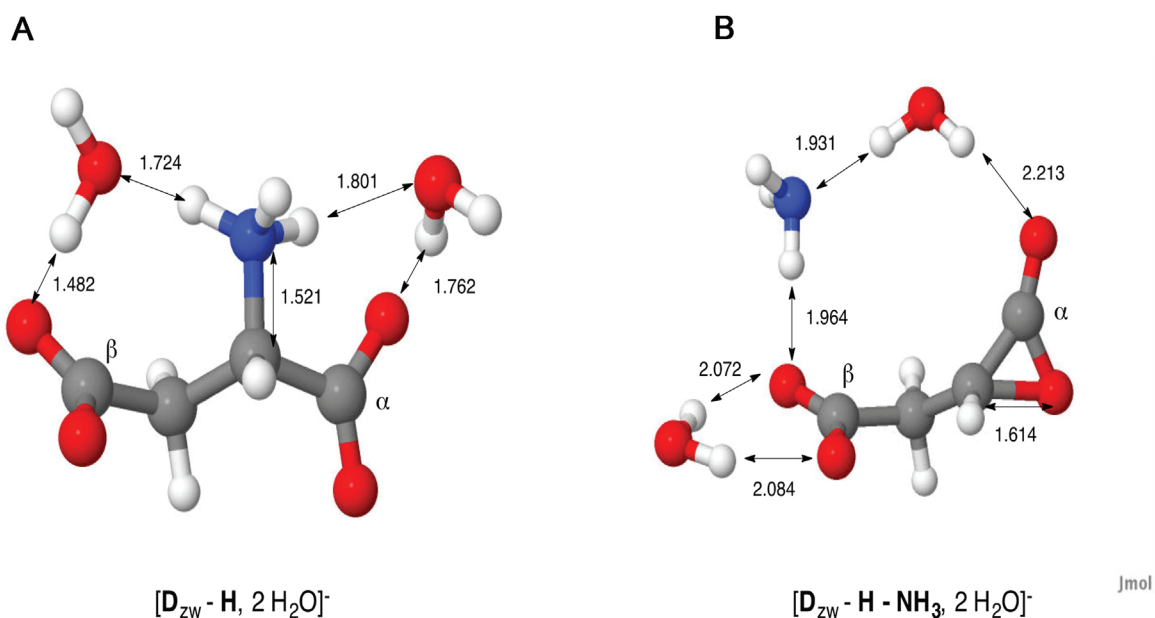


Fig. 4. Calculated stable doubly hydrated $[(D_{zw}-H)+2H_2O]^-$ product ion before (A) and after (B) deamination.

isomeric $[E_{CycPh}(t,c)-H]^-$ and $[E_{CycPh}(c,t)-H]^-$ species relative to the phenyl group (see Fig. 5a and b, respectively), the third diastereoisomeric structure (trans, trans) exhibited a prominent ion at

m/z 216 resulting from preferential loss of water (see Fig. 5c). Such particular behavior clearly demonstrates the drastic impact of the spatial vicinity of the two carboxylic acid groups on the fragmen-

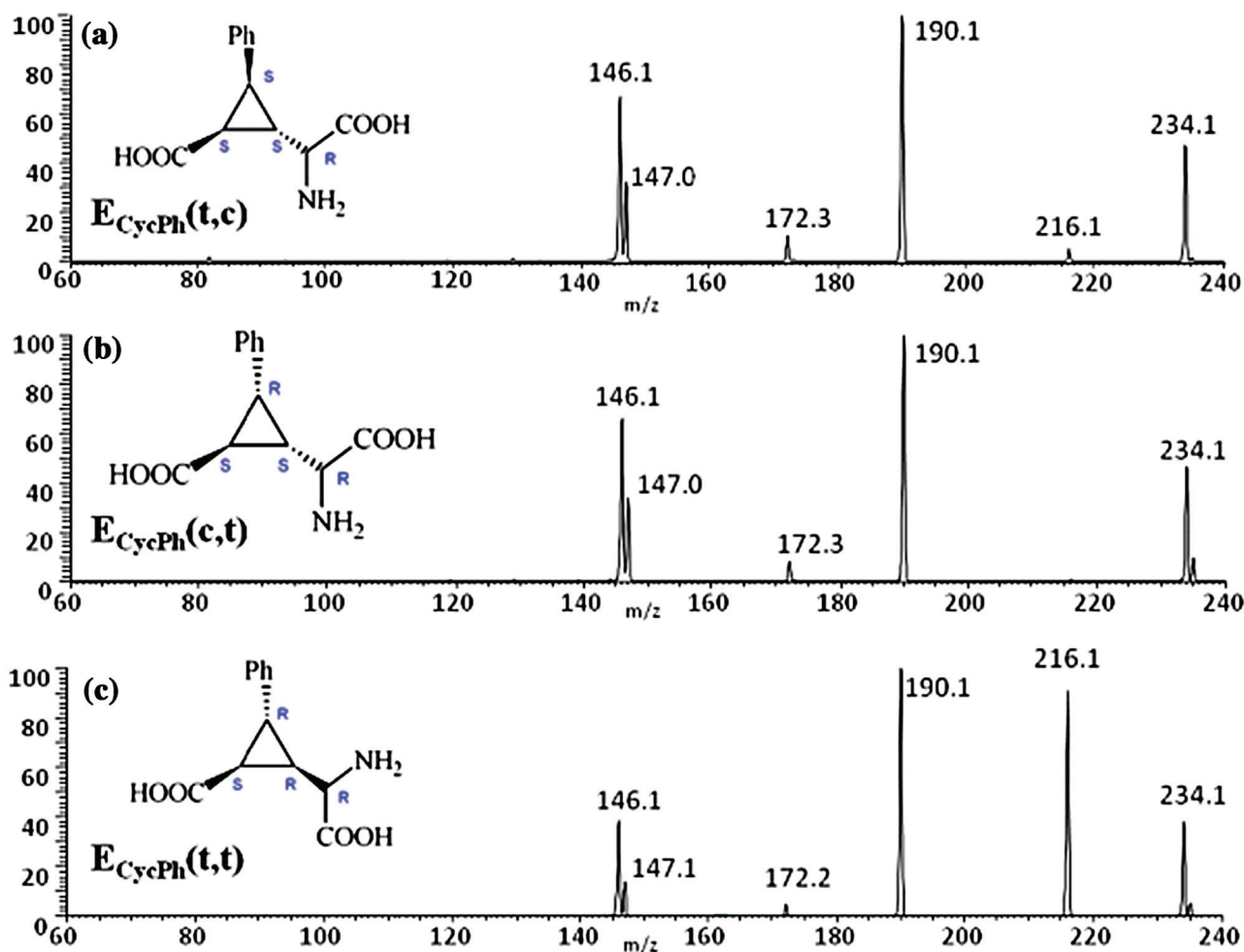


Fig. 5. CID spectra (NCE 31%) of the deprotonated (a) $E_{CycPh}(t,c)$, (b) $E_{CycPh}(c,t)$, and (c) $E_{CycPh}(t,t)$ modified amino acids. The $E_{CycPh}(t,c)$, $E_{CycPh}(c,t)$, and $E_{CycPh}(t,t)$ notations for stereoisomers are relative to the phenyl group [(t) and (c) are used for the trans and cis relative stereochemistry corresponding to the glycyl group and side chain carboxylic acid group, respectively].

tation pattern, and thus confirms the role of ion stereochemistry in this process.

4. Conclusion

The NH_3 , H_2O and CO_2 neutral losses are known to reflect competitive cleavages of deprotonated amino acids submitted to low-energy collision activation. Our study mainly focuses on a particular ammonia loss. This specific loss clearly distinguishes deprotonated aspartic acid from glutamic acid, which dissociates by water loss. Such strikingly different fragmentation behaviors for these two compounds, differing only by a methylene group, can be attributed to the occurrence of the specific $[(D-H)-NH_3]^-$ ion. Interestingly, this feature was consistently observed whatever the collisional activation mode (i.e., resonant, CID versus non-resonant, HCD, excitations). In contrast, the (NH_3, CO_2) loss specific to aspartate was only detected in the HCD mode. The latter indicates that in ion trapping experiments, aspartate anion apparently does not accumulate enough internal energy to yield such particular double loss without being ejected from the pseudo potential well. Quantum calculations demonstrated that the ammonia release concomitant with the loss of CO_2 on the side chain is preferentially from deprotonated aspartic acid.

More generally speaking, quantum calculations proved relevant and reliable in providing useful predictions of fragmentation path-

ways and overall were in strong agreement with experimental data obtained at high collision energies (e.g. loss of NH_3+CO_2). The observation of slightly divergent results between *in silico* predictions and experimental data often indicates inaccuracy of the model system used for calculation experiments. For instance, regarding the deaminated aspartate species that may be observed under low-energy conditions, calculations were necessary to highlight the involvement of residual molecules of water (present in the collision cell) in the dissociation process. Moreover, the role of ion conformation in the fragmentation was further confirmed by the analysis of rigid 2-(2'-carboxy-3'-phenylcyclopropyl) glycine diastereoisomeric species, with a preferential water loss from the relative (trans, trans) structure. Overall, our data tend to underline that quantum calculations should be used in conjunction with experimental data to provide working hypotheses for mechanistic studies.

In all mechanistic interpretations, knowledge of the CO_2 loss pathway will be very informative for the determination of the structure of the stabilized $[(D-H)-NH_3]^-$ product ion. For this purpose, the study of aspartic and glutamic acid residues incorporating one ^{13}C -labeled carboxylic acid group would be very informative and is currently in progress. This study shows the importance of knowledge of decomposition mechanisms when designing an efficient and reliable algorithm for structural elucidation of unknown metabolites.

Conflicts of interest statement

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgements

The authors thank Prof. R. Pellicciari and Prof. M. Marinuzzi, University of Perugia (Italy) for the gift of the 2-(2'-carboxy-3'-phenylcyclo-propyl) glycine samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jchromb.2016.08.036>.

References

- [1] T. Kind, O. Fiehn, Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, *BMC Bioinf.* 8 (2007) 105–125.
- [2] M. Holcapek, R. Jirasko, M. Lis, Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules, *J. Chromatogr. A* 1217 (2010) 3908–3921.
- [3] M. Heinonen, A. Rantanen, T. Mielikainen, J. Kokkonen, J. Kiuru, R.A. Ketola, J. Rousu, FiD: a software for ab initio structural identification of product ions from tandem mass spectrometric data, *Rapid Commun. Mass Spectrom.* 22 (2008) 3043–3052.
- [4] V. Kovacic, V. Patoprsty, P. Oksman, R. Mistrik, P. Kovac, Electron ionization mass spectrometric study of monomeric models of O-polysaccharides of *Vibrio cholerae* O:1, serotypes *Ogawa* and *Inaba*, *J. Mass Spectrom.* 38 (2003) 924–930.
- [5] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, *BMC Bioinf.* 11 (2010) 148–160.
- [6] C. Afonso, R.B. Cole, J.-C. Tabet, Even-Electron Dissociation in ESI and MALDI/MS: Fundamental, Instrumentation, Practicalities and Applications, in: D. Cole Richard (Ed.), Wiley Interscience, 2010, pp. 631–684.
- [7] E. Rathahao-Paris, S. Alves, C. Junot, J.-C. Tabet, High resolution mass spectrometry for structural identification of metabolites in metabolomics, *Metabolomics* 12 (2016) 10–25.
- [8] P. Kebarle, A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, *J. Mass Spectrom.* 35 (2000) 804–817.
- [9] Z. Tian, S.R. Kass, Does electrospray ionization produce gas-phase or liquid-phase structures? *J. Am. Chem. Soc.* 130 (2008) 10842–10843.
- [10] W.D. Price, R.A. Jockusch, E.R. Williams, Is arginine a zwitterion in the gas phase? *J. Am. Chem. Soc.* 119 (1997) 11988–11989.
- [11] T. Wyttenbach, M. Witt, M.T. Bowers, On the stability of amino acid zwitterions in the gas phase: the influence of derivatization, proton affinity, and alkali ion addition, *J. Am. Chem. Soc.* 122 (2000) 3458–3464.
- [12] S. Xu, J.M. Nilles, K.H. Bowen, Zwitterion formation in hydrated amino acid: dipole bound anions: how many water molecules are required? *J. Chem. Phys.* 119 (2003) 10696–10701.
- [13] N.C. Polfer, R.C. Dunbar, J. Oomens, Observation of zwitterion formation in the gas-phase H/D: exchange with CH₃OD: solution-phase structures in the gas phase, *J. Am. Soc. Mass Spectrom.* 18 (2007) 512–516.
- [14] T.K. Hwang, G.Y. Eom, M.S. Choi, S.W. Jang, J.Y. Kim, et al., Microsolvation of lysine by water computational study of stabilized zwitterion, *J. Phys. Chem. B* 115 (2011) 10147–10153.
- [15] J.Y. Kim, Y. Lee, S. Lee, Effects of microsolvation on the relative stability of zwitterionic vs canonical proline, *Chem. Phys. Lett.* 608 (2014) 177–185.
- [16] D. Schröder, M. Buděšínský, J. Roithová, Deprotonation of *p*-hydroxybenzoic acid: does electrospray ionization sample solution or gas-phase structures? *J. Am. Chem. Soc.* 134 (2012) 15897–15905.
- [17] L. Meng, Z. Lin, Comprehensive computational study of gas-phase conformations of neutral, protonated and deprotonated glutamic acids, *Comput. Theor. Chem.* 976 (2011) 42–50.
- [18] M.K. Drayß, P.B. Armentrout, J. Oomens, M. Schäfer, IR spectroscopy of cationized aliphatic amino acids: stability of charge-solvated structure increases with metal cation size, *Int. J. Mass Spectrom.* 297 (2010) 18–27.
- [19] Z. Li, M.H. Matus, H.A. Velazquez, C.J. Cassidy, D.A. Dixon, Gas-phase acidities of aspartic acid glutamic acid, and their amino acid amides, *Int. J. Mass Spectrom.* 265 (2007) 213–223.
- [20] F. Fournier, C. Afonso, A.E. Fagin, S. Gronert, J.C. Tabet, Can cluster structure affect kinetic method measurements? the curious case of glutamic acid's gas-phase acidity, *J. Am. Soc. Mass Spectrom.* 19 (2008) 1887–1896.
- [21] R.R. Julian, R. Hodyss, J.L. Beauchamp, Salt bridge stabilization of charged zwitterionic arginine aggregates in the gas phase, *J. Am. Chem. Soc.* 123 (2001) 3577–3583.
- [22] E. Pluharová, O. Marsalek, B. Schmidt, P. Jungwirth, Peptide salt bridge stability: from gas phase via microhydration to bulk water simulations, *J. Chem. Phys.* 137 (2012) 101–108.
- [23] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, Definitions of terms relating to mass spectrometry (IUPAC recommendations 2013), *Pure Appl. Chem.* 85 (7) (2013) 1515–1609.
- [24] A.G. Harrison, Collisional studies of gaseous negative ions. NATO ASI series. Series C Mathematical and Physical Sciences (Fundamental Gas Phase Ion Chemistry) 347 (1991) 289–313.
- [25] S. Broecker, S. Herre, B. Wüst, J. Zweigenbaum, F. Pragst, Development and practical application of a Library of CID accurate mass spectra of more than 2500 toxic compounds for systematic toxicological analysis with data-dependent acquisition, *Anal. Bioanal. Chem.* 400 (2001) 101–117.
- [26] F.W. McLafferty, P.F. Bente, R. Kornfeld, S.C. Tsai, I. Howe, Metastable ion characteristics. XXII. Collisional activation spectra of organic ions, *J. Am. Chem. Soc.* 95 (1973) 2120–2129.
- [27] R.A. Yost, C.G. Enke, Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation, *Anal. Chem.* 51 (1979) 1251–1264.
- [28] C. Basic, R.A. Yost, Collision induced dissociation breakdown surfaces for nalkylbenzene molecular ions in a quadrupole ion trap mass spectrometer, *Int. J. Mass Spectrom.* 194 (2000) 121–132.
- [29] H.I. Kunttamaa, R.G. Cooks, Internal energy distributions acquired through collisional activation at low and high energies, *Int. J. Mass Spectrom. Ion Process.* 64 (1985) 79–83.
- [30] S.R. Horning, M. Vincenti, R.G. Cooks, Angular dependence of internal energy distributions of activated. The contributions of vibrational and electronic excitation mechanisms in keV collisions, *J. Am. Chem. Soc.* 112 (1990) 119–126.
- [31] V.H. Wysocki, M.M. Ross, S.R. Horning, R.G. Cooks, Remote-site (charge-remote) fragmentation, *Rapid Commun. Mass Spectrom.* 2 (1988) 214–217.
- [32] P. Liere, S. Bouchonnet, R.E. March, J.C. Tabet, Cooling time and pressure effects on competitive thermalization/activation processes by resonance excitation on ITMS, *Rapid Commun. Mass Spectrom.* 9 (1995) 1594–1598.
- [33] S.A. Mcluckey, D.E. Goeringer, Slow heating methods in tandem mass spectrometry, *J. Mass Spectrom.* 32 (1997) 461–474.
- [34] R.E. March, F.A. Londry, R.L. Alfred, J.F.J. Todd, A.D. Penman, F. Vedel, M. Vedel, Resonance excitation of ions stored in a quadrupole ion trap. Part III. Introduction to the field interpolation simulation method, *Int. J. Mass Spectrom. Ion Process.* 110 (1991) 159–178.
- [35] M.P. Jedrychowski, E.L. Huttlin, W.E. Haas, E. Mathew, R. Rad, G. Ramin, S.P. Gygi, Evaluation of HCD and CID-type fragmentation within their respective detection platforms for Murine phosphoproteomics, *Mol. Cell. Proteom.* 10 (2011) 009910/1–009910/9.
- [36] R.A. Zubarev, A. Makarov, Orbitrap is the newest addition to the family of high-resolution mass spectrometry analyzers. With its revolutionarily new, miniature design, orbitrap combines high speed with excellent quantification properties, ranking favorably in many analytical applications, *Anal. Chem.* 85 (11) (2013) 5288–5296.
- [37] V. Steiner, C. Beaugrand, P. Liere, J.C. Tabet, Influence of trapping parameters on ion injection and dissociation efficiencies in a quadrupole mass filter/ion trap tandem instrument, *J. Mass Spectrom.* 34 (1999) 511–520.
- [38] C. Bennaceur, C. Afonso, S. Alves, A. Bossee, J.C. Tabet, Instrumental dependent dissociations of *n*-propyl/isopropyl phosphonate isomers: evaluation of resonant and non-resonant vibrational activations, *J. Am. Soc. Mass Spectrom.* 24 (2013) 1260–1270.
- [39] F. Ichou, A. Schwarzenberg, D. Lesage, S. Alves, C. Junot, et al., Comparison of the activation time effects and the internal energy distributions for the CID: PQD and HCD excitation modes, *J. Mass Spectrom.* 49 (2014) 498–508.
- [40] R.D. Bowen, D.H. Williams, Unimolecular reactions of isolated organic ions. The importance of ion-dipole interactions, *J. Am. Chem. Soc.* 102 (1980) 2752–2756.
- [41] F. Fournier, C. Salles, J.C. Tabet, L. Debrauwer, D. Rao, A. Paris, G. Borie, Tandem mass spectrometric investigation on 17 β -estradiol palmitate in negative ion chemical ionization, *Anal. Chim. Acta* 241 (1990) 265–272.
- [42] A. Roux, Y. Xu, J.F. Heilier, M.F. Olivier, E. Ezan, et al., Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-orbitrap mass spectrometer, *Anal. Chem.* 84 (2012) 6429–6437.
- [43] S. Boudah, M.F. Olivier, S. Aros-Calt, L. Oliveira, F. Fenaille, et al., Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 966 (2014) 34–47.
- [44] C.M. Jones, M. Bernier, E. Carson, K.E. Colyer, R. Metz, et al., Gas-phase acidities of the 20 protein amino acids, *Int. J. Mass Spectrom.* 267 (2007) 54–62.
- [45] Z. Tian, A. Pawlow, J.C. Poutsma, S.R. Kass, Are carboxyl groups the most acidic sites in amino acids? Gas-phase acidity H/D exchange experiments, and computations on cysteine and its conjugate base, *J. Am. Chem. Soc.* 129 (2007) 5403–5454.
- [46] M.L. Stover, V.E. Jackson, M.H. Matus, M.A. Adams, C.J. Cassidy, et al., Fundamental thermochemical properties of amino acids: gas-phase and aqueous acidities and gas-phase heats of formation, *J. Phys. Chem. B* 116 (2012) 2905–2916.

- [47] C.E. Plummer, M.L. Stover, S.S. Bokatzian, J.T.M. Davis, D.A. Dixon, et al., An experimental and computational study of the gas phase acidities of the common amino acid amides, *J. Phys. Chem. B* 119 (2015) 9661–9669.
- [48] S.S. Choi, O.B. Kim, Fragmentation of deprotonated amino acids in atmospheric pressure chemical ionization, *Int. J. Mass Spectrom.* 338 (2013) 17–22.
- [49] K. Sekimoto, N. Matsuda, M. Takayama, Formation of benzyl carbanion in collision-induced dissociation of deprotonated phenylalanine homologues, *Mass Spectrom.* 3 (2014), A0027/1-A0027/7.
- [50] K. Sekimoto, M. Sakakura, T. Kawamukai, H. Hike, T. Shiota, et al., Ionization characteristics of amino acids in direct analysis in real time mass spectrometry, *Analyst* 139 (2014) 2589–2599.
- [51] M. Eckersley, J.H. Bowie, R.N. Hayes, Collision induced dissociations of deprotonated amino acids. The occurrence of specific proton transfers preceding fragmentation, *Int. J. Mass Spectrom. Ion Process.* 93 (1989) 199–213.
- [52] R. Pellicciari, M. Marinozzi, B. Natalini, G. Costantino, R. Luneia, et al., Synthesis and pharmacological characterization of all sixteen stereoisomers of 2-(2'-carboxy-3'-phenylcyclo-propyl)glycine. focus on (2S,1'S,2'S,3'R)-2-(2'-carboxy-3'-phenylcyclopropyl)glycine, a novel and selective group II metabotropic glutamate receptors antagonist, *J. Med. Chem.* 39 (1996) 2259–2269.
- [53] O. Lange, E. Damoc, A. Wiegand, A. Makarov, Enhanced Fourier transform for orbitrap mass spectrometry, *Int. J. Mass Spectrom.* 369 (2014) 16–22.
- [54] R.E. March, Quadrupole ion trap mass spectrometry. A view at the turn of the century, *Int. J. Mass Spectrom.* 200 (2000) 285–312.
- [55] L.L. Lopez, P.R. Tiller, M.W. Senko, J.C. Schwartz, Automated strategies for obtaining standardized collisionally induced dissociation spectra on a benchtop ion trap mass spectrometer, *Rapid Commun. Mass Spectrom.* 13 (1999) 663–668.
- [56] L.W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B J. Chem. Phys.* 37 (1988) 785–789.
- [57] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* 98 (1993) 5648–5652.
- [58] L. Banu, V. Blagojevic, D.K. Bohme, Bond activation in complexes of Pb(II) with 15 deprotonated amino acids: correlation with gas-phase acidity, *Int. J. Mass Spectrom.* 316–318 (2012) 23–30.
- [59] A. Schwarzenberg, F. Ichou, R.B. Cole, X. Machuron-Mandard, C. Junot, et al., Identification tree based on fragmentation rules for structure elucidation of organophosphorus esters by electrospray mass spectrometry, *J. Mass Spectrom.* 48 (2013) 576–586.
- [60] M. Boutin, C. Bich, C. Afonso, F. Fournier, J.C. Tabet, Negative-charge driven fragmentations for evidencing zwitterionic forms from doubly-charged coppered peptides, *J. Mass Spectrom.* 42 (2007) 25–35.
- [61] H.M. Rosenstock, M.B. Wallenstein, A.L. Wahrhaftig, H. Eyring, Absolute rate theory isolated systems and the mass spectra of polyatomic molecules, *PNAS* 38 (1952) 667–678.
- [62] A. Pak, D. Lesage, Y. Gimbert, K. Vekey, J.-C. Tabet, Internal energy distribution of peptides in electrospray ionization: ESI and collision-induced dissociation spectra calculation, *J. Mass Spectrom.* 43 (2008) 447–455.
- [63] A.K. Pathak, Stabilizing the zwitterionic form of amino acids in the gas phase: an ab initio study on the minimum number of solvents and ions, *Chem. Phys. Lett.* 610–611 (2014) 345–350.
- [64] M. Meot-Ner Mautner, Update 1 of: strong ionic hydrogen bonds, *Chem. Rev.* 112 (2012) 22–103.
- [65] L.W. Sieck Meot-Ner, The ionic hydrogen bond and ion solvation. 5- OH...O-bonds. Gas-phase solvation and clustering of alkoxide and carboxylate anions, *J. Am. Chem. Soc.* 108 (1986) 7525–7529.
- [66] B. Jia, L.A. Angel, K.M. Ervin, Threshold collision-induced dissociation of hydrogen-bonded dimers of carboxylic acids, *J. Phys. Chem. A* 112 (2008) 1773–1782.
- [67] M.J. Maclean, S. Walker, T. Wang, P.C.H. Eichinger, P.J. Sherman, J.H. Bowie, Diagnostic fragmentations of adducts formed between carbanions and carbon disulfide in the gas phase. A joint experimental and theoretical study, *Org. Biomol. Chem.* 8 (2010) 371–377.
- [68] T. Wyttenbach, M.T. Bowers, Gas phase conformations of biological molecules: the H/D exchange mechanism, *J. Am. Soc. Mass Spectrom.* 10 (1999) 9–14.
- [69] Z. Tian, L. Lis, S.R. Kass, Hydrogen-deuterium exchange and selective labeling of deprotonated amino acids and peptides in the gas phase, *J. Am. Chem. Soc.* 130 (2008) 8–9.
- [70] P. Liere, V. Steiner, K.R. Jennings, R.E. March, J.C. Tabet, Influence of ion activation and thermalization effects on reaction rate constants in a quadrupole ion trap mass spectrometer, *Int. J. Mass Spectrom. Ion Process.* 167/168 (1997) 735–751.
- [71] C.G. Sinnige, L.J. de Koning, N.M.M. Nibbering, Water molecule assisted proton mobility in gaseous protonated GlyPro and ProGly, *Int. J. Mass Spectrom.* 195/196 (2000) 115–119.
- [72] J.W. Gauld, H.E. Audier, J. Fossey, L. Radom, Water-catalyzed interconversion of conventional and distonic radical cations: methanol and methyleneoxonium radical cations, *J. Am. Chem. Soc.* 118 (1996) 6299–6300.
- [73] H. Nedev, G. van der Rest, P. Mourgues, H.E. Audier, Reactions of CH_3CHO^+ and of CH_3COH^+ with water upon Fourier transform ion cyclotron resonance conditions, *Eur. Mass Spectrom.* 9 (2003) 319–326.

3.5 Principaux résultats

Cette étude a tout d'abord montré que, malgré leur grande similitude en termes de structure, les deux acides aminés présentent des spectres de fragmentation très différents suite à leur ionisation en mode négatif par électrospray. En effet, en plus d'une perte commune de CO_2 , la fragmentation de l'acide aspartique conduit à une perte importante de NH_3 qui n'est pas du tout observée lors de la fragmentation de l'acide glutamique. Accompagnant cette perte, une perte de CH_3NO_2 (ou $\text{NH}_3 + \text{CO}_2$) est également observée à plus haute énergie de fragmentation. L'acquisition de spectres MS^3 montre que cette double perte est consécutive à la perte de NH_3 alors que le fragment ayant initialement perdu CO_2 est, quant à lui, stable et ne se refragmente pas. L'utilisation de calculs quantiques met en évidence la nécessité d'avoir une solvatation de l'ion parent pour envisager la création de la forme désaminée de l'aspartate. C'est la conformation tridimensionnelle prise par l'aspartate lors de cette solvatation qui permet d'observer cette perte spécifique. Pour l'acide glutamique, la présence d'un groupement méthylène supplémentaire écarte le groupement $-\text{NH}_2$ du groupement carboxylique de la chaîne latérale ce qui bloque un échange de protons entre ces groupes. Cependant, à ce stade, les calculs quantiques expliquent bien la perte concomitante de $(\text{NH}_3 + \text{CO}_2)$ mais ne permettent pas de formaliser la perte de NH_3 seule et l'observation stable de la forme $[(\text{Asp-H})-\text{NH}_3]^-$.

3.6 Résultats supplémentaires

Suite à ces premiers travaux, d'autres expériences ont été réalisées pour tenter d'expliquer le désaccord au sujet du mécanisme mis en jeu pour la perte de NH_3 qui persiste entre la proposition issue des résultats expérimentaux et celle issue des calculs quantiques. En effet, les calculs prédisent une double perte $[\text{NH}_3 + \text{CO}_2]$ concomitante alors qu'expérimentalement, il semble s'agir d'un processus consécutif. L'utilisation de composé marqué au carbone 13 au niveau de l'atome du groupe carboxylique de la chaîne latérale a permis de montrer qu'à haute énergie de collision la perte de $^{13}\text{CO}_2$ est régiosélective à plus de 90%. Cependant, à plus basse énergie, cette régiosélectivité apparaît moins marquée. De plus, des expériences MS^3 réalisées sur l'acide aspartique ayant subi la perte d'ammoniac $[(\text{Asp-H})-\text{NH}_3]^-$ montre que la perte consécutive de CO_2 se fait à parts égales entre le groupement carboxylique latéral et celui de la chaîne principale. Cette information indique le passage par un intermédiaire symétrique lors de cette perte consécutive. Ces observations montrent que la double perte $[\text{NH}_3, \text{CO}_2]$ se fait par deux mécanismes distincts en fonction de l'énergie de collision : un mécanisme consécutif à basse énergie et un mécanisme compétitif à plus haute énergie. Suite à ces observations, de nouveaux calculs ont été effectués. Ils montrent qu'une double hydratation, possible uniquement à basse énergie, permettrait l'isomérisation de la forme zwitterion solvatée $\{[(\text{D-H})-\text{NH}_3]^- , \text{NH}_3, 2\text{H}_2\text{O}\}$. Dans ce cas, la forme désaminée est une α -lactone carboxylate, siège d'un échange de protons entre le groupement $-\text{NH}_3$ et les molécules d'eau, suivi de la perte de

NH₃. Cela conduit à un intermédiaire symétrique qui se décompose ensuite de façon non sélective avec la perte de CO₂. Le second mécanisme serait une double perte concomitante directe à partir de la forme canonique de l'aspartate par un second chemin réactionnel. Pour confirmer l'existence de ces échanges de protons, un marquage des hydrogènes labiles de l'acide aspartique par échange Hydrogène/Deutérium a été réalisé. Les échanges prévus sont observés expérimentalement à basse énergie de collision avec une perte prédominante de la forme NHD₂ (soit rééchange d'un deutérium avec un hydrogène au cours du mécanisme de fragmentation).

Suite à toutes ces observations, deux mécanismes dépendant de l'énergie de collision sont proposés et présentés en Figure 17. L'ensemble de ces résultats devrait faire l'objet d'une seconde publication.

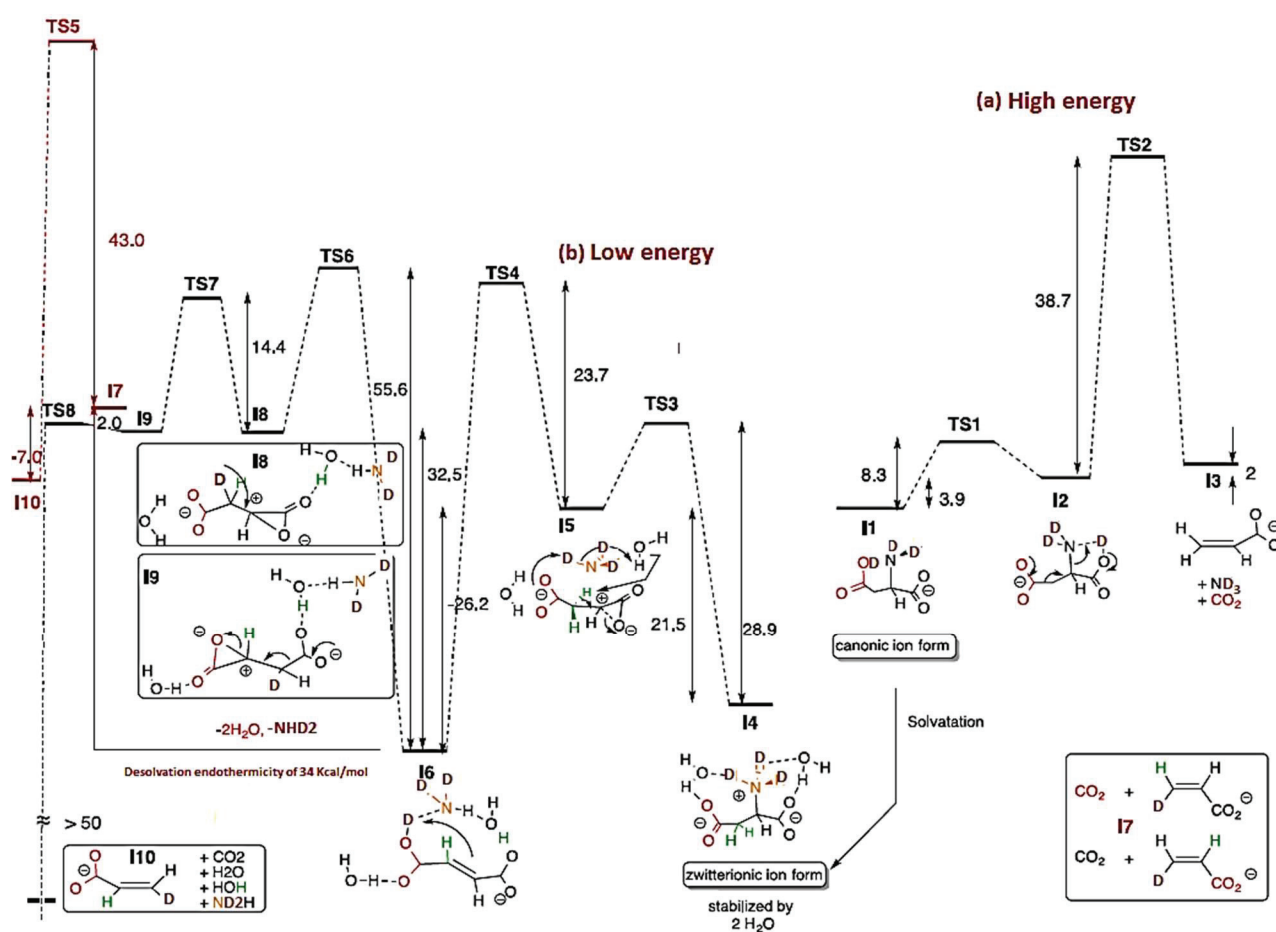


Figure 17 : Chemin réactionnel des deux mécanismes de fragmentation de l'acide aspartique

3.7 Conclusion

Ce travail a permis de mettre en évidence l'intérêt de différents outils tels que l'utilisation de MS³, de marquage isotopique stable, d'échanges Hydrogène/Deutérium ou encore du calcul quantique, pour aider à rationaliser les mécanismes mis en jeu lors de la fragmentation. Il souligne

également la complexité de ce genre d'approche. Cependant, si l'on veut un jour pouvoir concevoir des algorithmes permettant de retrouver automatiquement la structure d'un ion à partir de celle de ses fragments, ce travail devra être fait sur une grande diversité de molécules afin de cerner l'ensemble des mécanismes potentiellement impliqués. De plus, pour développer ce type de logiciel, il sera aussi important de développer de nouveaux modes d'acquisition de données qui feront de l'enregistrement des spectres fragmentation un élément central de ces méthodes. Cela permettra d'améliorer la qualité des spectres obtenus pour des applications en métabolomique.

Chapitre 3. Développement de nouvelles méthodes d'acquisition

1 Contexte

L'acquisition de spectres de fragmentation de haute qualité est un élément important dans le processus d'identification en métabolomique par LC-HRMS. Actuellement, l'acquisition de ces spectres MS² est en général découplée de l'acquisition du profil global MS. Une première injection de l'échantillon est réalisée pour acquérir le profil MS qui est analysé afin d'identifier les signaux d'intérêt. Une seconde injection est ensuite réalisée et permet d'enregistrer les spectres de fragmentation des composés d'intérêt. Cette approche présente différents handicaps. Tout d'abord, comme deux injections sont réalisées, elle est consommatrice d'échantillon et de temps. Elle nécessite de réanalyser un même échantillon à plusieurs reprises, ce qui lui fera nécessairement subir un cycle congélation/décongélation supplémentaire qui peut impacter sa stabilité et conduire à la disparition de certains signaux d'intérêt. Enfin, elle est tributaire de la robustesse du système chromatographique utilisé. Si le système n'est pas parfaitement robuste, un décalage dans les temps de rétention peut être observé entre les deux injections et conduire à des difficultés d'interprétation.

Pour pallier ces difficultés, de nouvelles méthodes d'acquisition sont en train de se développer en métabolomique. Ces protocoles, dérivés de la protéomique (Gillet et al., 2012), visent à acquérir en une seule injection, à la fois le profil MS et les différents spectres de fragmentation des métabolites détectés sur le profil. Ces méthodes ont d'abord été implémentées sur les instruments de type Temps de Vol (Benton et al., 2015) du fait de leur plus grande vitesse d'acquisition. Avec l'amélioration de la vitesse des instruments de type Orbitrap, il est maintenant envisageable d'y implémenter ce type d'approche sans répercussions sur la résolution utilisée.

2 Publication numéro 4

Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand?

François Fenaille, Pierre Barbier Saint Hilaire, Kathleen Rousseau, Christophe Junot

Journal of Chromatography A, 2017, 1526, 1-12



Review article

Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? ☆



François Fenaille^a, Pierre Barbier Saint-Hilaire^a, Kathleen Rousseau^a, Christophe Junot^{b,*}

^a Service de Pharmacologie et Immuno-Analyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France

^b Service de Pharmacologie et Immuno-Analyse (SPI), CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France

ARTICLE INFO

Article history:

Received 31 July 2017

Received in revised form 15 October 2017

Accepted 16 October 2017

Available online 18 October 2017

Keywords:

Metabolome

Metabolomics

Liquid chromatography

Mass spectrometry

Data dependent acquisitions

Data independent acquisitions

ABSTRACT

Typical mass spectrometry (MS) based untargeted metabolomics protocols are tedious as well as time- and sample-consuming. In particular, they often rely on “full-scan-only” analyses using liquid chromatography (LC) coupled to high resolution mass spectrometry (HRMS) from which metabolites of interest are first highlighted, and then tentatively identified by using targeted MS/MS experiments. However, this situation is evolving with the emergence of integrated HRMS based-data acquisition protocols able to perform multi-event acquisitions. Most of these protocols, referring to as data dependent and data independent acquisition (DDA and DIA, respectively), have been initially developed for proteomic applications and have recently demonstrated their applicability to biomedical studies. In this context, the aim of this article is to take stock of the progress made in the field of DDA- and DIA-based protocols, and evaluate their ability to change conventional metabolomic and lipidomic data acquisition workflows, through a review of HRMS instrumentation, DDA and DIA workflows, and also associated informatics tools.

© 2017 Elsevier B.V. All rights reserved.

Contents

1. Introduction	1
2. How to improve data acquisition workflows for metabolomics?	2
3. Data dependent acquisitions	3
4. Data independent acquisitions	5
4.1. All-Ion Fragmentation (AIF), MS ^{ALL} and MS ^E approaches	5
4.2. SWATH and SWATH-like workflows	6
4.3. Targeted MS/MS workflows: PRM vs SRM	7
5. Informatic tools for DDA and DIA workflows	8
6. Conclusion	9
Acknowledgments	9
References	9

1. Introduction

Metabolomics is a data driven and multidisciplinary approach combining analytical chemistry, biochemistry, informatics and biostatistics for analyzing the set of metabolites contained in biological media. A typical metabolomic workflow includes the design of analytical experiments, followed by sample preparation, data acquisition, data preprocessing, data annotation, statistical analysis, and data visualization for biological interpretations. Major

☆ Selected paper from the 12^{ème} Congrès Francophone sur les Sciences Séparatives et les Couplages Analytiques (SEP 2017), 29–30 March 2017, Paris, France.

* Corresponding author.

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

achievements have been realized in all these steps, which have made metabolomics a reliable tool now used in various biological applications.

Data acquisition still mostly relies on mass spectrometry and proton nuclear magnetic resonance spectroscopy (NMR). Regarding MS based methods, improvements have been achieved in the field of gas chromatography coupled to mass spectrometry (GC/MS), especially with the implementation of two-dimension GC for metabolomics [1]. High-resolution MS analyzers (*i.e.*, time of flight, Orbitrap and Fourier transform ion cyclotron resonance instruments) fitted with atmospheric pressure ionization sources are more and more often used with at least two complementary liquid chromatography (LC) stationary phases [2,3], and also in combination with other separative techniques such as capillary electrophoresis (CE) [4], and other tools, also including NMR [5] in order to improve metabolome coverage. Otherwise, the development of ion mobility coupled to high resolution mass spectrometry opens new avenues for the resolution of isobaric compounds and isomers in biofluids [6].

Beside the increased size of mass spectral libraries such as MassBank [7], Metlin [8], NIST (<http://www.nist.gov/srd/>), mzcloud (<https://www.mzcloud.org>), HMDB (Human Metabolome Database, <http://www.hmdb.ca>), MoNA (MassBank of North America, <http://mona.fiehnlab.ucdavis.edu>) or GNPS (Global Natural Product Social Molecular Network Platform, <http://gnps.ucsd.edu>), many computational tools for metabolite identification have been developed [9], such as tools for retention prediction [10–14], database searching algorithms that also perform *in silico* fragmentation of metabolites [15–22], and also other tools dedicated to annotation and interpretation of MS/MS spectra [23–26]. Other interesting approaches are the exploitation of shared molecular sub-structures from MS² spectra to group metabolites, thanks to an algorithm initially used for text mining [27], and also the use of molecular networking based on GNPS, which has already been used for the dereplication of natural products [28,29]. At last, data visualization tools relying on metabolic networks are popularizing [30,31].

Typical untargeted MS based-metabolomics protocols rely on “full-scan-only” analyses using LC-HRMS from which metabolites (or panels of metabolites) of interest are first deduced. Those hits are then tentatively identified by tandem mass spectrometry following multiple injections into the LC-MS/MS system using targeted MS/MS workflows. Such procedure is highly tedious as well as time- and sample-consuming, and relies essentially on analyst manual data analysis.

Until now, two main strategies for HRMS based-data acquisition were followed: (i) improving the mass resolving power of instruments in order to achieve detection and quantification using Fourier transform mass spectrometry (*i.e.*, FTICR and Orbitrap instruments) through accurate mass measurements and data dependent MS² for structural information, or (ii) exploiting the data acquisition speed of Q-TOF instruments in order to achieve simultaneous detection, quantification and structural analyses through data independent MS² experiments using Q-TOF instruments [32].

However, thanks to continuous improvements of HRMS analyzers with enhanced dynamic range, acquisition rate, and improved sensitivity, this situation is evolving with the emergence of integrated data acquisition protocols able to perform multi-event acquisitions enabling concomitant detection, quantification and structural characterization of biomolecules through MS² experiments within a single run. Most of these methods, referring to as data dependent and data independent acquisition (DDA and DIA, respectively), have been initially developed for proteomic applications. This is for example the case for SWATH (sequential window acquisition of all theoretical fragment-ion spectra)-MS which systematically fragments all precursor ions within user-

defined retention time and precursor ion mass to charge (m/z) windows [33]. These DIA based-proteomic approaches, which provide accurate peptide quantification over a dynamic range of more than 4.5 orders of magnitude, have recently demonstrated their applicability to analyses of few hundreds of samples from human cohorts [34].

Can these DIA protocols be routinely applied in the fields of metabolomics and lipidomics in a very near future? This question needs to be asked insofar as untargeted metabolomic, lipidomic and proteomic experiments are not exactly carried out in the same way. If data acquisition can be assumed as untargeted for all these approaches, proteomic data treatment and interpretation are mostly driven by genomic information and by the availability of large and reliable experimental and *in silico* spectral databases, whereas this is not the case for metabolomics. It is indeed still impossible to model the whole metabolomes of biological organisms, in part because of our limited knowledge regarding their metabolism, but also because of our inability to accurately model complex host-environment interactions. The situation tends to be intermediate for lipidomics, with the lipid map project [35] and the release of LipidBlast, an *in silico* tandem mass spectrometry database for lipid identification [36].

By writing this review, we would like to take stock of the progress made in the field of DDA- and DIA-based data acquisition protocols for LC-HRMS based metabolomics and lipidomics, and evaluate their ability to change conventional metabolomic data acquisition workflows.

2. How to improve data acquisition workflows for metabolomics?

High throughput metabolomics workflows are greatly benefiting from the constant improvements of mass spectrometers in terms of sensitivity, mass accuracy, resolution, and acquisition speed. Table 1 presents a non-exhaustive list of some high resolution mass spectrometers that can be used for metabolomics. In most common metabolomic workflows (at least those discussed in this article), such instruments are coupled to liquid chromatography using an electrospray ionization source. Commercially available QqTOF instruments traditionally have resolving powers between 35,000 and 60,000 (full width at half maximum, FWHM), while those of last generation Orbitrap-based can reach 1,000,000 FWHM (Orbitrap Fusion Lumos 1 M).

Orbitrap instruments often outperform Q-TOFs in terms of mass resolving power, but at the expense of longer accumulation times. Therefore such mass analyzers, when working at the highest resolving power, can be less suited for the accurate quantification of compounds eluting as narrow LC peaks [37]. In the opposite, Q-TOFs are often characterized by enhanced duty cycles with the ability to acquire up to the 100 MS/MS spectra per second, independently of the resolving power. Both types of instruments can routinely achieve mass accuracies below 1 ppm (considering internal calibration, Table 1).

The hyphenation of ion mobility with Q-TOF instruments provides another dimension of separation, particularly useful for resolving isobaric compounds and decomplexifying complex biological extracts. In addition, ion mobility allows for the measurement of collision cross-sections (CCS), which define ions in the gas phase according to their charge, shape, and size. Recent results indicate that metabolite identification can greatly benefit from the additional information provided by CCS [38].

Most (if not all) of the published metabolomic workflows involve metabolite fragmentation under low-energy collision-induced dissociation (CID) conditions (0–100 eV), although electron transfer dissociation (ETD) can be performed on most of the

Table 1
A selection of high resolution/high mass accuracy mass spectrometers used for metabolomics.

Instrument	Mass accuracy (ppm)	Mass resolving power	MS/MS acquisition speed (Hz)
Orbitrap-based instruments			
Q-Exactive HF-X (Thermo Scientific)	<1 ppm (internal calibration)	240,000 at m/z 200	40 Hz (at resolution of 7500 at m/z 200)
Orbitrap Fusion Lumos (Thermo Scientific)	<1 ppm (internal calibration)	500,000 at m/z 200 ^a	20 Hz (at resolution of 15,000 at m/z 200)
Q-TOF instruments			
TripleTOF 6600 (Sciex)	<1 ppm (internal calibration)	35,000 at m/z 956	100 Hz
timsTOF (Bruker)	<0.8 ppm (internal calibration)	50,000 at m/z 1222	100 Hz
6560 Q-TOF (Agilent)	<1 ppm (internal calibration)	40,000 at m/z 2722	30 Hz
Synapt G2-Si HDMS Q-TOF (Waters)	<1 ppm (internal calibration)	50,000 at m/z 956	30 Hz

Data were essentially collected from the websites of the different instrument providers, and might vary slightly. Acquisition rates can be different in MS and MS/MS modes.

^a The latest Orbitrap Lumos can provide 1,000,000 FWHM ultra-high resolution.

instruments listed in Table 1. The higher energy C-trap dissociation (HCD) mode, which is available on Orbitrap-based instruments, is a non-resonant form of CID which provides Q-TOF-like fragmentations. Of note, the Orbitrap Fusion Lumos is also equipped with a 213 nm UV laser to allow for ultraviolet photodissociation (UVPD) experiments in the ion trap, which proved useful for top-down proteomics but also for double bond locations within complex lipids [39,40].

3. Data dependent acquisitions

In DDA acquisition workflows, the MS instrument automatically switches from full-scan MS to MS/MS (Fig. 1) when precursor ions exceed a predefined threshold of intensity or any other predefined criteria such as isotope pattern [41,42], mass defect [43], or also the presence of diagnostic ion or characteristic neutral loss in the MS² spectra. The latter are especially used for detecting phase II metabolites such as sulfates [44], glucuronides [45], but also bile acid conjugates [46] and acylcarnitines [47] in biofluids. Although a few papers have been published, DDA acquisition workflows are not as popular in metabolomics as in proteomics, where they are the most widely used method. The main reason deals with metabolite identification issues, such as the incompleteness of spectral databases, and the occurrence of many unknown metabolite features and the difficulty to interpret related MS² spectra. However, DDA MS² approaches enabling concomitant detection, quantification and identification of metabolite features have been developed from the 2010s.

Chen et al. have developed a DDA MS² method on a Q-TOF instrument to investigate the human tear metabolome [48]. Within a total duty cycle time of 1.25 s, acquisitions included a survey scan and 20 MS/MS scans of 50 ms each. The DDA criteria were implemented and consisted of ion intensity, signal to noise ratio and isotope pattern matching threshold values. MS² data were then matched with information contained in Metlin [8], HMDB [49] and Massbank [7] databases, thus leading to the identification of 60 metabolites.

Benton et al. have developed an autonomous untargeted DDA-MS² method, named “Auto MS/MS acquisition”, on Q-TOF instruments [50]. Auto MS/MS acquisitions consisted of a first TOF MS scan performed at 4 Hz (4 spectra per second), followed by product ion scans acquired at a frequency of 20 Hz (20 spectra per second) using information dependent acquisitions using the following criteria: selection of the precursor at unit resolution, one fixed collision energy, single charged ions, an intensity threshold, isotope exclusion within 4 Da, 15 maximum candidate ions, an ion exclusion set at 9 s after two occurrences, and activation of a background subtraction tools available on the data acquisition software. MS/MS spectra were then compared to those of Metlin using a cosine similarity metric and a mirror plot. This workflow was subsequently evaluated on biological samples obtained from bacterial biofilm cultures grown in two conditions. The authors

found that 36% of features overlapped between conventional MS and Auto MS/MS detections.

Beside the incompleteness of available spectral databases, Benton et al. pointed out two limitations of DDA MS² procedures: (i) the limited MS/MS spectral coverage and (ii) the match percentages provided by scores. They addressed these two issues by using a mixture of 40 metabolites [50]. They obtained MS/MS spectra for 85% of metabolites. However, this result could be considered as rather optimistic especially when translated to “real life” since the impact of metabolite concentrations was not investigated. At last, authors concluded that the performance of DDA MS² based methods could be improved by implementing informatic workflows enabling the distinction between biologically relevant features and artifacts. They described a pilot study that enabled to select 836 metabolite features of interest out of 12420, which could be imported as a preferred ion list for further MS² experiments. However, this requires performing preliminary and extensive experiments on each biological matrix of interest.

Another limitation of DDA MS² based approaches is the presence of other ions than the targeted precursor in the isolation window for MS² experiments, which hampers data interpretation and also mass spectral library matching. To address this issue, Lawson et al. have developed a R-package (“ms-Purity”) that scores the contribution of the selected precursor ion by dividing the intensity of the precursor ion by the summed intensity of the ions detected in the isolation window [51]. Such software could be used for assessing precursor purity of previously acquired data, or for preparing future MS² experiments. By applying the software tool to 12 public DDA and 1 DIA metabolomics data sets stored in repositories, the authors found that lower scores (*i.e.*, below 0.5) are due to a too large number of consecutive MS² experiments (up to 26) that altered data reliability and also to large isolation windows, such as those used for DIA experiments.

As previously emphasized, DDA is an acquisition process generally producing good quality tandem mass spectra with minimal interferences since only ions within a selected m/z window (typically ~1 Da wide) are fragmented. However, the selection of precursor ions is a stochastic event, favoring the preferential selection of the most abundant/intense species. Such acquisition workflow is thus characterized by a low analytical reproducibility and under sampling especially for complex samples. In these conditions, some relevant precursor ions cannot be selected for fragmentation if they do not match the defined selection criteria or if they are coeluting with more intense species or isomers [52]. Another inconvenience is that huge amounts of MS/MS data from irrelevant ions might be generated. Also, acquiring many MS/MS spectra can result in insufficient data points in the MS mode for a given chromatographic peak, which can be detrimental when not using top speed HRMS instruments for quantitative LC-MS experiments.

Several kinds of approaches have been published to address these issues and for improving the spectral coverage and the per-

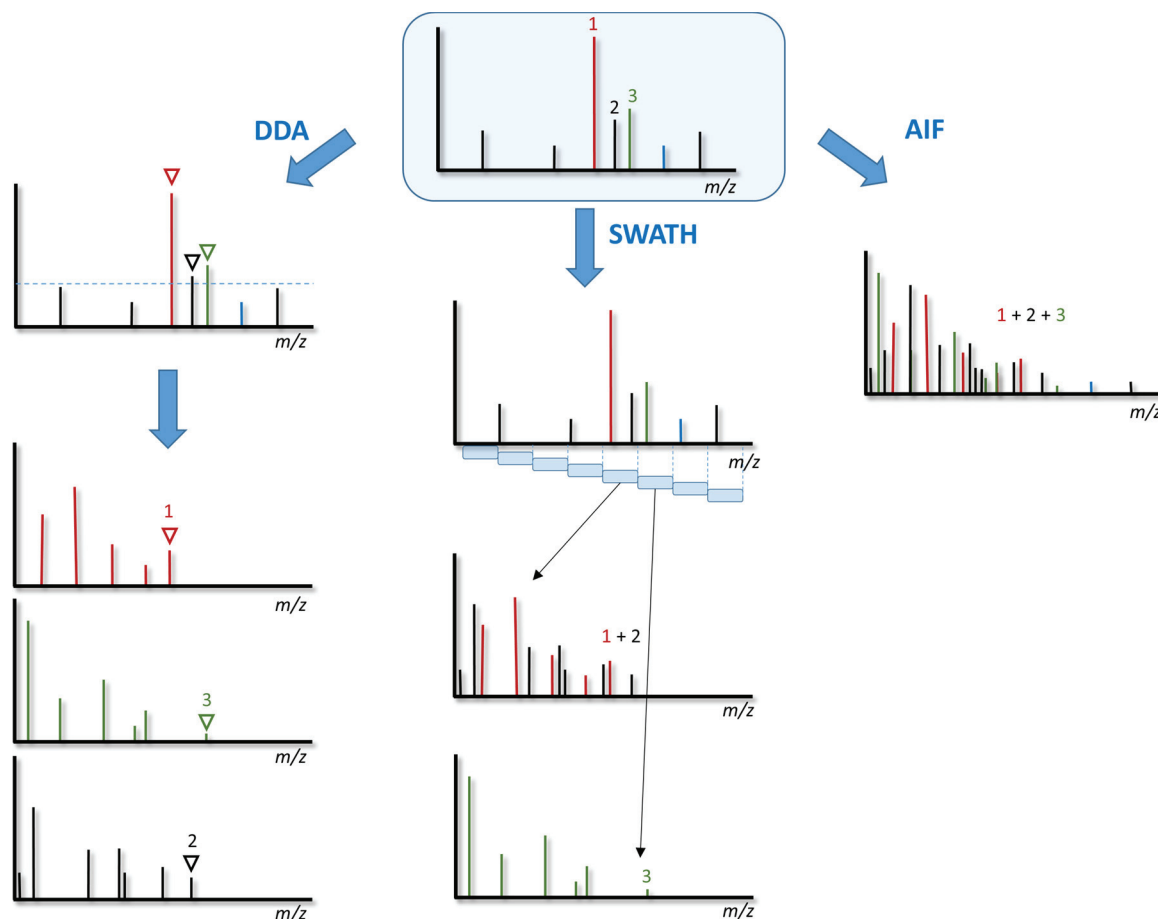


Fig. 1. Scheme illustrating the DDA, SWATH and AIF acquisition workflows.

performances of DDA MS²-based methods. They rely on (i) software tools that help to focus on relevant metabolite features [53,54], and (ii) advanced acquisition protocols for optimal MS/MS acquisitions [55–57].

Regarding software tools, Neumann et al. have released an R-package enabling the design of LC-MS/MS methods for data acquisition on relevant metabolite features selected by a preceding non targeted profiling experiment, by using various filtering criteria such as ion intensity threshold and ion type, in order to minimize the number of irrelevant data acquisitions [53]. In the field of lipidomics, Koemel et al. have developed “IE-Omics” (IE, for Iterative Exclusion) software tool also written in R language, and aimed at automatically exclude prior selected precursor and background ions of MS² experiments. In each sequential injection, unique precursors are fragmented until HRMS/MS spectra of all ions above a given ion intensity threshold are acquired. By these means, the number of metabolite identification was improved by 69% in a plasma sample using six sequential injections, compared to a conventional approach [54].

Regarding advanced DDA MS² acquisition protocols, Calderon-Santiago et al. have improved DDA by using gas-phase fractionation (GPF), which has been previously developed for proteomics applications [58]. It is based on automated iterative MS/MS spectra acquisitions on a sample. Here, multiple and small m/z ranges were designed in order to improve the number of MS/MS fragmentations on a human serum sample, using a Q-TOF instrument. The authors have evaluated 5 experimental protocols that divided the inclusion range for precursor selection into 1, 2, 3, 4 or 6 parts over a 100–1000 m/z range. They found that the method splitting the range of precursor ion masses into 4 intervals of 250 amu was the

optimal one, which provided MS/MS spectra for the highest number of detected features [55].

Such a GPF approach was improved by using multiple staggered narrower m/z segments, on a low resolution instrument (*i.e.*, a Q-TRAP mass spectrometer). The authors evaluated 5 methods with 2, 4, 6, 8 and 10 fractions over a 100–800 m/z range. When applied to the analysis of human urines and compared to targeted DDA MS² analyses, the method employing 8 segments of a 5 or 6 Da range performed best with 92% of MS² coverage over a set of 60 identified metabolites [56].

GPF based methods are attractive thanks to their capability to generate high number of MS/MS spectra. However, they are time consuming and require instruments offering very high data acquisition speed, such as Q-TOF or triple quadrupole instruments. Furthermore, additional studies are required in order to carefully evaluate the quality of the produced MS/MS spectra at a more global level.

Although most of the first published DDA approaches rely on Q-TOF instruments because of their unprecedented acquisition speed, they can also be performed efficiently on recent Orbitrap mass spectrometers. Mullard et al. have thus reported on the optimization of a new strategy for DDA MS² on a LTQ-Orbitrap Velos hybrid mass spectrometer [57]. They compared the information brought by CID and HCD mass spectra. They also evaluated the impact of mass resolution on the number and quality of MS² spectra, of the number of MS/MS experiments between full MS spectra, of repeat count, and of the precursor m/z range. By comparing MS² spectra recorded on 76 compounds from 39 metabolite classes, they found that CID and HCD provided complementary results and thus recommend the concomitant use of these two activation methods.

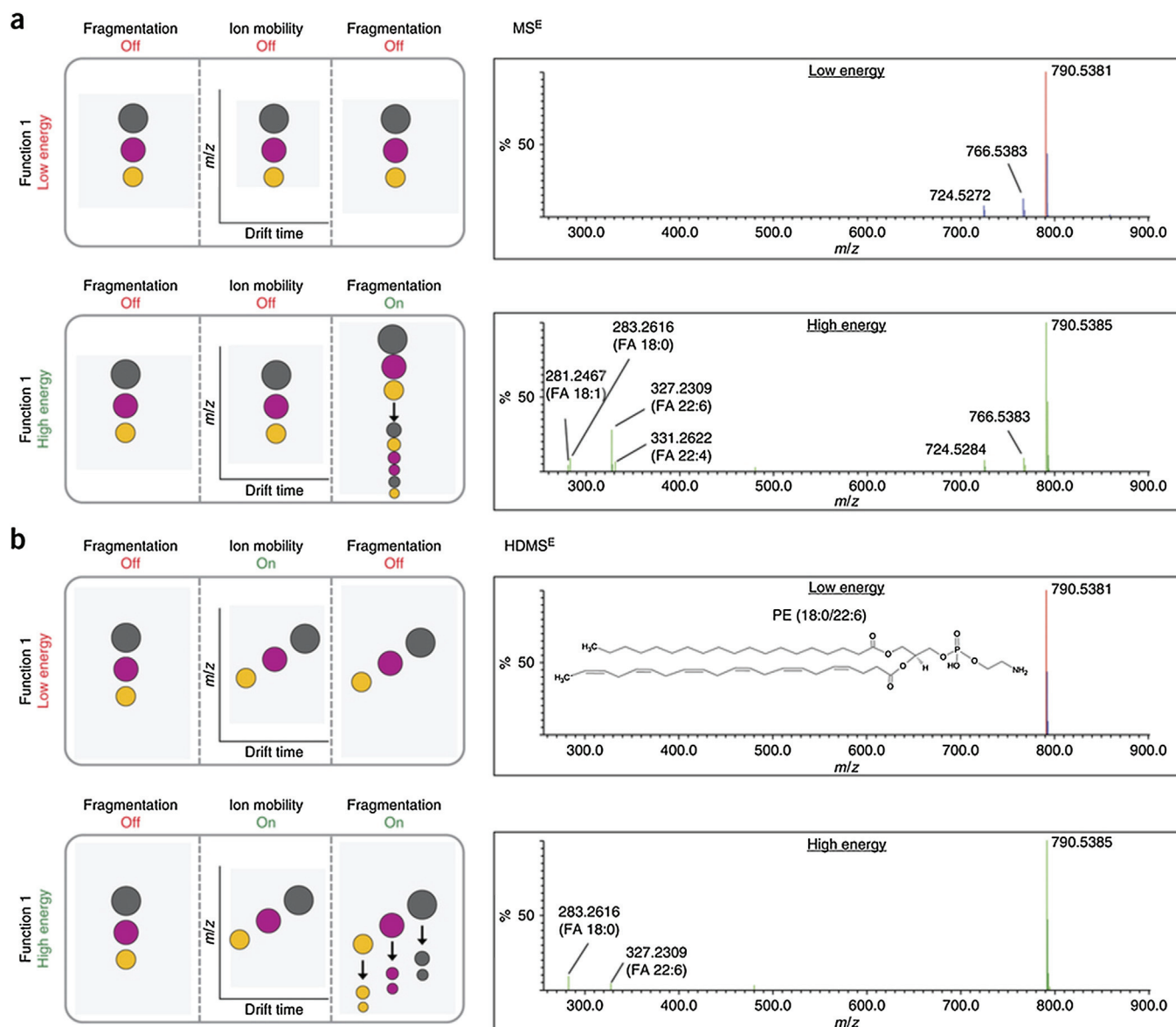


Fig. 2. Characterization of lipid species by combining ion mobility and MS^E. (a) MS^E-only data at low and high energies, providing complex mass spectra resulting from coelution of different phosphatidylethanolamine (PE) species. (b) Ion mobility efficiently separates the coeluting PE species, thus allowing identification of PE 18:0/22:6. The high energy is applied post ion mobility separation.

Adapted from Paglia and Astarita [6] with permission.

They also recommend narrowing the precursor m/z ranges to 100 or 300 u for acquisition of MS² spectra in order to increase the number of MS² experiments; and using mass resolutions of 30,000 and 7500 FWHM for full MS scan and MS² experiments, respectively. By applying this optimized DDA MS² method to the analysis of a human serum sample, they obtained MS/MS data for greater than 50% of metabolite features with a peak area above 200000 AU.

Such Orbitrap-MS based approaches should be improved by using new generations of instruments such as the Orbitrap-Fusion, as it has been published by Tu et al. in the field of proteomics. These authors investigated the performance of protein identification by CID/HCD fragmentations in combination with high/low resolution detection using a tribrid Orbitrap instrument on human cell extracts [59]. Although it was shown that the comparison of results obtained using CID or HCD fragmentations was greatly impacted by the type of data interpretation software tool, more peptide spectra matches, distinct peptides and protein groups were identified from MS² data generated in HCD-OT than HCD-IT when using Sequest or Mascot based software tools. Furthermore, the

limited peptide overlap between the CID-OT, CID-IT, HCD-OT and HCD-IT modes suggest that it could be useful to combine them, as emphasized by Mullard et al. in the field of metabolomics [57].

Finally, all these studies pave the way to the development and extensive use of integrated HRMS and DDA MS² workflows for the detection, quantification and identification of metabolite and lipid features in biological extracts. However, to encompass the limitations of DDA linked to the selection of precursor ions and under-sampling, data independent acquisition (DIA) workflows have also been developed and attracted a lot of attention since fragment ions can be collected for every precursor detected.

4. Data independent acquisitions

4.1. All-Ion Fragmentation (AIF), MS^{ALL} and MS^F approaches

In 2005, Wrona et al. [60] were the first to report an "All-in-One" approach to metabolite characterization exploiting the fast scanning and high mass accuracy properties of a Q-TOF instru-

ment (Fig. 1). The described acquisition method alternated between high and low collision energies throughout the LC gradient, and was successfully applied to the detection of *in vitro* metabolites of verapamil [60]. One year later, Plumb et al. introduced a similar LC–MS acquisition method, that was called MS^E and involved alternating scans acquired at either low or high collision energy in the collision cell to obtain precursor ion information and to collect full-scan accurate mass fragment, precursor ion and neutral loss information, respectively [61]. Robust and high resolution UPLC separation helped to relate fragment and parent ions. Interestingly, metabolomic data collected from rat urine using UPLC–MS^E were very similar to those obtained using a more conventional LC–MS/MS approach, thus confirming the value of the workflow [61]. Such UPLC–MS^E protocol also proved relevant to study the biliary metabolome of rats and dogs [62].

In another study, Vorkas et al. used a combination of RP- and HILIC-UHPLC–MS methodologies to achieve broader metabolome coverage in cardiovascular tissues [63]. The authors also combined DDA and MS^E acquisitions to specifically attribute fragment ions to precursor ions and highlight fragment ions signing particular metabolite classes, respectively.

Similar All-Ion Fragmentation (AIF, equivalent to MS^E) workflows have been implemented on Q-Orbitrap as well as benchtop standalone Orbitrap instruments for metabolite [64,65] or lipid [66–68] profiling. For instance, LC–MS combined with higher energy C-trap dissociation (HCD) all ion fragmentation yields both class-specific and lipid-specific diagnostic fragments, thus facilitating the identification of 381 unique lipids in rat liver mitochondria [69].

Although the mostly used MS^E or AIF workflows involve high performance UPLC–MS platforms and are characterized by high acquisition speed and detection sensitivity, their obvious intrinsic disadvantage is that the direct link between a specific precursor ion and its corresponding fragment ions does not exist. Also, these workflows targeting simultaneously all the precursor ions within a specified (wide) *m/z* window, result in highly complex MS/MS spectra that require highly efficient proper post-acquisition data processing software tools (see below). The introduction of ion mobility spectrometry (IMS) into such DIA workflows may solve some of these issues.

By coupling ion mobility with MS^E, Paglia and Astarita demonstrated that coeluting isobaric phosphatidylethanolamine species can be efficiently separated by ion mobility and characteristic MS/MS spectra can be obtained thus providing detailed information about the carried fatty acid groups [6]. Fig. 2 shows a relevant example of PE 18:0/22:6 structure confirmation by combining IM–MS^E. While in conventional MS^E workflow three coeluting lipid species were simultaneously selected and fragmented, yielding four fragment ions corresponding to distinct fatty acyl moieties, the introduction of ion mobility efficiently separates the ion at *m/z* 790.5381 from other coeluting ions, generating only two specific fragment ions corresponding to the deprotonated fatty acyl species C22:6 and C18:0 (Fig. 2). Thus, ion mobility separation prior to fragmentation leads to cleaner MS/MS spectra, somehow re-establishing the link between precursor and fragment ions and by the way facilitating data interpretation and improving considerably the detection specificity of such a DIA workflow.

4.2. SWATH and SWATH-like workflows

Another way of reducing the number and presence of interfering ions in DIA–MS/MS spectra makes use of a recently reported alternative DIA technique called SWATH [33] (Fig. 1). This approach can be viewed as an intermediate between narrow and wide precursor ion selections (DDA and MS^E approaches, respectively) and typically uses 20–50 Da consecutive precursor ion isolation windows

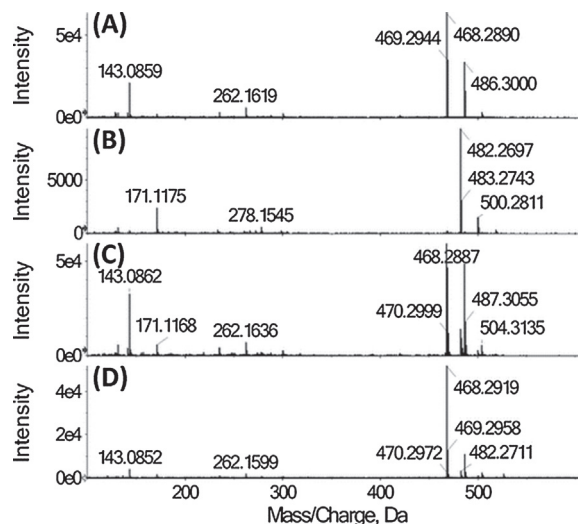


Fig. 3. Tandem mass spectra obtained from various manners (A) DDA of a terfenadine metabolite at *m/z* 504.31, (B) DDA of another terfenadine metabolite at *m/z* 518.29, (C) SWATH with a Q1 isolation window at *m/z* 499–525, and (D) AIF, high-energy TOF MS scan at *m/z* 100–1000.

Adapted from Zhu X et al. [78] with permission.

covering the whole mass range. Initially implemented in the field of proteomics on a fast scanning Q-TOF [33] and later on a Q-Exactive instrument [70], SWATH analyses have also recently found applications in the small molecule world and in metabolomics using both types of instruments [52,71–76]. Of note, although most of the published SWATH-based methods primarily and typically involve constant SWATH windows (e.g. 25 Da), the benefit of using variable SWATH windows depending on the ion population within a given mass range has been demonstrated by Zhang et al. [77].

Zhu et al. evaluated DDA, SWATH and AIF approaches on a Q-TOF instrument for drug metabolism studies, with a total of 227 metabolites from 8 different drugs monitored both in rat liver microsomes and urines [78]. Rather expectedly, they found that MS/MS spectra acquired under DDA conditions were of higher quality, which was particularly marked in the case of coeluting species with similar molecular masses (i.e., occurring within the same SWATH mass selection window, Fig. 3). However, DDA-based method was characterized by a lower acquisition hit rate than the other two methods, with MS/MS spectra that were missing for up to 30% of drug metabolites when analyzing urine samples. Both DIA approaches provided MS/MS spectra for all the metabolites of interest, while SWATH gave higher quality spectra [78].

More recently, Zhou et al. demonstrated that SWATH workflows for metabolomics analysis can be successfully implemented on a Q-Exactive instrument, although characterized by a lower acquisition rate than Q-TOFs [73]. Good quality MS/MS spectra were obtained using the SWATH approach under these conditions when compared to those resulting from DDA or AIF workflows, also well matching those included in the reference spectral database [73]. As a representative example, Fig. 4A–D depicts the MS/MS spectra generated for ceramide 22:0 by the DDA, DIA (with either 25 or 50 Da selection windows) and the AIF strategies. Using 50 Da-isolation windows, sufficient data points (~10) were obtained per chromatographic peak for using this strategy within quantitative contexts (Fig. 4E). In the same spirit, Colgrave et al. have shown that a SWATH strategy gave similar quantitative results that the gold standard multiple reaction monitoring (MRM; also called selected reaction monitoring, SRM) mass spectrometry approach for a set of about 30 peptides assessed across the four families of hordeins (gluten) in eight barley lines [79]. In different quantitative contexts, SWATH approaches often proved to be slightly less sensitive than

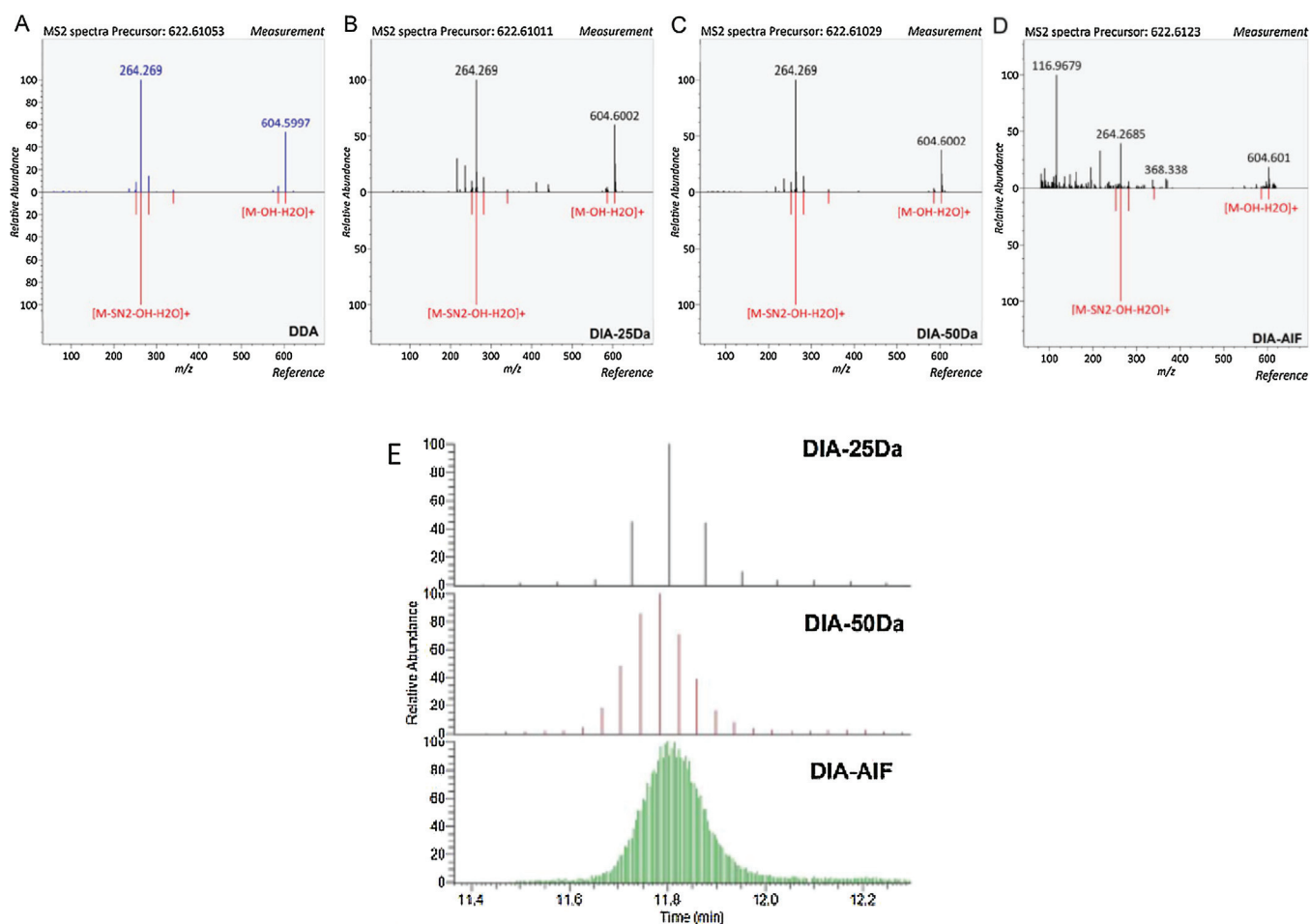


Fig. 4. Tandem mass spectra of ceramide 22:0 obtained (A) by the DDA strategy, and by the DIA strategy with (B) a 25 Da or (C) a 50 Da isolation window, and (D) using an AIF strategy. (E) Chromatographic traces obtained using the DIA–25 Da, DIA–50 Da, and AIF strategies. Adapted from Zhou J et al. [73] with permission.

SRM but inherently demonstrated increased flexibility and versatility [80], by the possible post-analysis selection of alternative precursors or fragments and the ability of performing large-scale peptide/metabolite quantification.

Metabolite quantification represents a prerequisite to the comparison of data generated in distinct laboratories with different analytical platforms and protocols. Also, rendering (large-scale) metabolomics workflows quantitative would significantly improve data accuracy and reliability, and enable more meaningful biological interpretations. Exhaustively quantifying the whole set of metabolites present in a given biological fluid constitutes a particularly tremendous if not impossible task, especially due to the vast number of unknown endogenous metabolites and the lack of authentic standards. To address these issues, Chen et al. have reported on the development of a novel method named “data-independent targeted quantitative metabolomics” (DITQM) [81]. Their strategy first implied the acquisition of high resolution MS/MS spectra on a Q-Orbitrap using 1 Da isolation window. About 30 precursor ions separated by ~25 Da from each other were fragmented per LC–MS/MS run, thus 26 distinct injections were required to cover the whole 80–850 m/z range. Under these conditions, they basically obtained high quality MS/MS spectra for almost all known and unknown metabolites they detected in human plasma (in the m/z 80–850 window). Resulting precursor and fragment ions were then used to design a multiplexed targeted quantitative MRM assay on a Q-Trap instrument for 1324 metabolites [81]. The DITQM and MRM workflows gave similar quantitative

measurements for 12 acylcarnitines in human plasma (as a proof-of-concept) thanks to authentic standards, which underlines the value of data-independent acquisition process.

As exemplified by the latter paper, shifting from untargeted-only metabolomics applications to combined large-scale untargeted/targeted workflows would be highly relevant for more sensitive, specific, and reproducible quantitative workflows. Such a pseudo targeted metabolomic approach has already been described a few years ago, and involved UHPLC–SRM, with the SRM ion pairs directly selected from an untargeted UHPLC/Q–TOF MS profiling method [82].

4.3. Targeted MS/MS workflows: PRM vs SRM

One of main driving forces in metabolomics is the large-scale accurate quantitation of metabolites on relevant biological samples, to evaluate changes in response to disease, treatment, environmental and genetic perturbations. Most of the large-scale targeted metabolomics workflows imply “only” relative – but already highly informative – quantification (or semi-quantification), because of its simplicity [83]. Absolute metabolite concentrations often require spike-in stable isotope-labeled analogues to correct for any changes in ionization efficiency across samples. Unfortunately, such molecules are not always available commercially or sometimes at a very prohibitive cost. To encompass these limitations, research groups have used metabolite extracts from microorganisms grown on various types labeled of

carbon or nitrogen sources [84–86]. Alternatively, labeled internal standards can also be enzymatically synthesized from commercially available precursors [87].

SRM-based workflows implemented on triple quadrupole or quadrupole-ion trap instruments still represent the leading choice for targeted metabolite quantitation [88]. Such strategies have been successfully applied for more than 10 years to large-scale metabolite quantification (see for representative examples [89–93]). For instance, in 2006, Bajad et al. reported an HILIC-SRM approach for reliable measurement in bacterial extracts of 141 metabolites, including components of central carbon, amino acid, and nucleotide metabolism [89]. Similarly, Wei et al. implemented a high throughput LC-SRM for target based metabolite profiling across different classes of metabolites (amino acids, sugars and nucleic acids, and organic acids) in rat plasma [90].

Method reproducibility and inter-laboratory precision represent critical steps toward the standardization of quantitative metabolomics workflows. Siskos et al. recently published the results of the first inter-laboratory assessment of a targeted metabolomics platform for the quantitative analysis of human serum and plasma [94]. This study involved 6 distinct laboratories using different instrumentation but a common protocol (a commercial kit for targeted metabolomics) for the measurement of 189 metabolites by LC-SRM or FIA-SRM. For 20 serum and plasma samples from healthy individuals, the median inter-laboratory coefficient of variation (CV) was 7.6%, with 85% of metabolites showing a median inter-laboratory CV of <20% [94], thus demonstrating the good reproducibility level of such approach.

Although SRM assays are still considered as the “gold standard” to quantitate metabolites (vide supra), parallel reaction monitoring (PRM, also termed high resolution multiple reaction monitoring or MRM^{HR}) strategies can represent more than a viable alternative. PRM is an MS/MS strategy commonly implemented on Q-Orbitrap instruments in the field of proteomics for the large-scale quantification of representative peptides [95–97]. However, until recently [83], there was almost no report of application of PRM for large-scale quantitative metabolomics.

In PRM-based workflows, the parallel acquisition of all target fragment ions in a single high resolution/high mass accuracy mass spectrum provides improved specificity so as to reliably discriminate the ions of interest from background ions. The PRM method has also fewer parameters that need to be optimized (maximum injection time/AGC target, “global” collision energy) [95], while the main advantage of SRM is the faster acquisition speed.

In 2016, Qiu et al. reported on an UHPLC-PRM approach implemented on a Q-Orbitrap instrument for the quantification of 25 metabolites from the central metabolic pathways in breast cancer cell lines [98]. This method exhibited good linearity (over

3–4 orders of magnitude), sensitivity (in the low ng/mL range), and reproducibility (relative standard deviation <5% for almost all the metabolites) [98]. The same year, Zhou et al. developed and evaluated the usefulness of PRM for quantifying up to 237 metabolites (from different metabolic pathways) in human cell lines [99]. The PRM approach yielded similar results than the same method implemented on a Q-ion trap (Q-Trap) instrument in terms of sensitivity, linearity and reproducibility but with improved specificity and flexibility regarding post-acquisition data treatment due to the presence of many fragment ions measured with a high mass accuracy. The authors also showed, on their particular set of 237 metabolites, that the PRM assay exhibited greater reproducibility and quantitative accuracy than HRMS-based (i.e. “full scan only”) quantification due to the higher selectivity of the PRM acquisition mode [99].

Of course, similar PRM-like workflows can be implemented on Q-TOF instruments. For instance, Li et al. recently used this strategy for the quantification of 65 metabolites in *E. coli* cell extracts [100], while Zhou et al. described a similar approach to quantify 222 lipids in human serum [101].

5. Informatic tools for DDA and DIA workflows

At the time of writing this review, most LC-MS based metabolomic experiments are acquired using high resolution mass spectrometry in the electrospray MS scanning mode. The two main free and open source software tools used by the metabolomic community for data processing, namely XCMS [102] and MZMine [103] do not handle DDA and DIA workflows. As a consequence, published DDA- and DIA-MS based metabolomic studies mainly rely on software tools provided by MS-manufacturers. Anyway, some free and open source algorithms and software tools have been developed for DDA- and DIA-MS based metabolomics and are listed in Table 2.

Regarding DDA-MS, data acquisition protocols are optimized so that the maximal numbers of MS/MS data are obtained [50,57]. In this context, metabolite features obtained from LC-HRMS are processed by using data processing tools provided by instrument vendors’ software, annotated using in house and/or public spectral databases, and the analysis of MS/MS spectra is driven by MS-annotation results in order to improve metabolite identification [48,50,55–57]. Anyway, some algorithms have been developed to improve the acquisition of MS/MS spectra by evaluating precursor purity [51], or by minimizing the acquisition of irrelevant spectra, as already mentioned [53,54]. At last, algorithms highlighting metabolite features exhibiting characteristic neutral losses in their CID spectra are also available, mainly for the study of phase II metabolites such as sulfate and glucuronide compounds [104].

Table 2
Non-exhaustive list of Informatic available tools for the acquisition and processing of DDA and DIA-MS data.

Name	Intended use	Source code/language	reference
msPurity	DDA. Evaluation of precursor ion purity	R package	[51]
MetShot	DDA. Informatic tools to improve acquisition and data processing	R package	[53]
IE-Omics	DDA. Improvement of data acquisition through automated exclusion list generation	R code	[54]
MS-DIAL	DIA. Deconvolution of MS ² spectra and data processing. Also supports DDA	Multiple. Some codes written in C language	[72]
Met DIA	DIA. Data processing for targeted analyses	R code	[110]
SWATH Tuner	DIA. Building of SWATH windows	R code	[77]
NOFI	DIA. ‘non-outlier fragment ion’ (NOFI) is a ranking algorithm that assigns low priority to fragment ions affected by interference during DIA acquisition.	R code	[113]
MetaboDIA	DIA. MS/MS-based quantification with DIA data.	R code	[114]
LipidMatch	DDA, DIA. Lipid identification using untargeted high-resolution tandem mass spectrometry data	R code	[115]

Regarding DIA, MS^E is the first approach that has been used in the field of metabolomics [61]. Other DIA-MS methods, such as SWATH-MS [33], or PRM [96] have been initially developed for proteomic applications, and dedicated open source software, such as OpenSWATH [105] and DIA-Umpire [106] have been released.

OpenSWATH is an open source software dedicated to targeted analysis of proteomic DIA data. It is based on an assay library containing precursor and fragment ion *m/z* and retention time values. The workflow includes data conversion of raw data into open file formats, retention time and chromatogram extraction based on assay library, peak-group scoring based on elution profiles of fragment ions and statistical analyses enabling the discrimination between true and false signals by using a false discovery rate (FDR) approach [105].

DIA-Umpire software is also an open source software dedicated to DIA-based proteomic data, which detects precursor and fragment ions and groups them into pseudo MS/MS spectra, in an untargeted way, without the need of an assay library [106]. These two computational frameworks have been designed for SWATH data, but they could handle other types of DIA data [105,106].

The proteomic data processing has benefited from achievements in genomics, which enables to model the proteome of living organisms, and of the possibility to generate *in silico* CID spectra of peptides using electrospray ionization [107,108]. Unfortunately, this still cannot be achieved for metabolites, due to their huge structural diversity and due to our limited knowledges regarding metabolic networks. As a consequence, the only way to process DIA-MS/MS based metabolomic data is to perform mathematical deconvolution of fragment ions and their association with precursor ions, thanks to their chromatographic elution profiles, as emphasized by Nikolskiy et al. [109].

This has been achieved with the development of the MS-DIAL open-source software pipeline, which performs precursor ion peaks spotting and deconvolution of MS/MS spectra on SWATH-MS data, and which could also handle data obtained with other DIA methods such as All-ion MS/MS or All Ion Fragmentation [72].

Li H. et al. have also published a new DIA data analysis approach (called MetDIA) for targeted extraction of metabolites that are present in spectral libraries. It is based on metabolite precursor ion peak detection and alignment by using XCMS software, followed by targeted chromatogram extraction, generation of peak-groups and pseudo MS² data from SWATH windows. MS² spectra are then scored for metabolite identification, thanks to information contained in spectral libraries. Two orthogonal scores have been implemented, based on peak-peak correlation, in order to evaluate to similarity of ion chromatograms between precursor and fragment ions, and based on spectrum-spectrum match, in order to evaluate the similarity between experimental and reference MS² spectra [110].

Zhu et al., in their comparative study of IDA, SWATH and MS^{All} have proposed two other kinds of scoring procedures. The first one relies on the MS² spectra quantity related to the number of precursor ions, whereas the second one is a quality score that indicates the number of common fragment ions between experimental and reference MS² spectra [78].

Last, Li Z. et al. have used a PRM workflow on a Q-TOF instrument for the targeted quantification of 65 metabolites. The Skyline software, initially developed for targeted proteomic experiments (design of MRM transitions, but also handling of DIA acquisitions) [111], was used to handle the data [100].

6. Conclusion

Thanks to recent technological advances in mass spectrometry, new relevant and reliable analytical workflows have recently

emerged and involve essentially acquisition of combined MS and MS/MS spectra at increased mass resolving power. Regarding DDA protocols, some authors propose to dedicate selected samples to MS/MS data acquisition in order to improve the quality and the quantity of information [55–57], whereas DIA protocols such as SWATH based approaches are performed on all biological samples of interest [74].

Current methodological developments indicate a global tendency toward convergence of untargeted and targeted metabolomics approaches, but most of those are more “proof-of-concept studies” or “works in progress”. Indeed, qualitative and quantitative comparisons of the various workflows implemented are still required to demonstrate their respective value for metabolomics, especially regarding their robustness, their sensitivity (number of detected peaks) and the reliability of data processing tools, compared to classical HRMS based protocols only relying on accurate mass measurements. Deeper investigations regarding DIA approaches are also needed to evaluate whether SWATH-like (HRMS/MS) approaches could be used for the large-scale robust and reliable measurement of any metabolite in any biofluid or tissue (and at which level of sensitivity).

In this context, the metabolomics/lipidomics community can greatly benefit from the availability of open access databases of PRM transitions/MS/MS spectra for metabolites/lipids identification and quantification, in a similar fashion to the SRMAtlas database for proteomics [112]. Ion mobility represents a technological improvement with a great potential in metabolite/lipid profiling and quantification. However, a critical and thorough evaluation of the added value of ion mobility and CCS measurement for metabolomics and lipidomics is still lacking and would be of high interest to the field. At last, the availability of open-source software tools and databases for exploiting in a convenient way data from ion mobility, as well as from DDA and DIA metabolomics approaches would allow people from the field to better evaluate those analytical strategies in comparison to conventional MS-only profiling methods.

Acknowledgments

This work was supported by CEA and the French Ministry of Research and National Research Agency as part of the French metabolomics and fluxomics infrastructure (MetaboHUB, ANR-11-INBS-0010 grant).

References

- [1] M. Hartonen, I. Mattila, A.-L. Ruskeepää, M. Orešič, T. Hyötyläinen, Characterization of cerebrospinal fluid by comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry, *J. Chromatogr. A* 1293 (2013) 142–149, <http://dx.doi.org/10.1016/j.chroma.2013.04.005>.
- [2] T. Zhang, D.J. Creek, M.P. Barrett, G. Blackburn, D.G. Watson, Evaluation of coupling reversed phase, aqueous normal phase, and hydrophilic interaction liquid chromatography with Orbitrap mass spectrometry for metabolomic studies of human urine, *Anal. Chem.* 84 (2012) 1994–2001, <http://dx.doi.org/10.1021/ac2030738>.
- [3] S. Boudah, M.-F. Olivier, S. Aros-Calt, L. Oliveira, F. Fenaille, J.-C. Tabet, C. Junot, Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* 966 (2014) 34–47, <http://dx.doi.org/10.1016/j.jchromb.2014.04.025>.
- [4] J. Saric, E.J. Want, U. Duthaler, M. Lewis, J. Keiser, J.P. Shockcor, G.A. Ross, J.K. Nicholson, E. Holmes, M.F.M. Tavares, Systematic evaluation of extraction methods for multiplatform-based metabolotyping: application to the *Fasciola hepatica* metabolome, *Anal. Chem.* 84 (2012) 6963–6972, <http://dx.doi.org/10.1021/ac300586m>.
- [5] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, *PLoS One* 6 (2011) e16957, <http://dx.doi.org/10.1371/journal.pone.0016957>.

- [6] G. Paglia, G. Astarita, Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry, *Nat. Protoc.* 12 (2017) 797–813, <http://dx.doi.org/10.1038/nprot.2017.013>.
- [7] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, MassBank: a public repository for sharing mass spectral data for life sciences, *J. Mass Spectrom.* 45 (2010) 703–714, <http://dx.doi.org/10.1002/jms.1777>.
- [8] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, *Ther. Drug Monit.* 27 (2005) 747–751.
- [9] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S.S. Mehta, G. Wohlgemuth, D.K. Barupal, M.R. Showalter, M. Arita, O. Fiehn, Identification of small molecules using accurate mass MS/MS search, *Mass Spectrom. Rev.* (2017), <http://dx.doi.org/10.1002/mas.21535>.
- [10] D.J. Creek, A. Jankevics, R. Breitling, D.G. Watson, M.P. Barrett, K.E.V. Burgess, Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction, *Anal. Chem.* 83 (2011) 8703–8710, <http://dx.doi.org/10.1021/ac2021823>.
- [11] J. Stanstrup, S. Neumann, U. Vrhovšek, PredRet: prediction of retention time by direct mapping between multiple chromatographic systems, *Anal. Chem.* 87 (2015) 9421–9428, <http://dx.doi.org/10.1021/acs.analchem.5b02287>.
- [12] M. Cao, K. Fraser, J. Huege, T. Featonby, S. Rasmussen, C. Jones, Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics, *Metabolomics Off. J. Metabolomic Soc.* 11 (2015) 696–706, <http://dx.doi.org/10.1007/s11306-014-0727-x>.
- [13] D. Abate-Pella, D.M. Freund, Y. Ma, Y. Simón-Manso, J. Hollender, C.D. Broeckling, D.V. Huhman, O.V. Krokhin, D.R. Stoll, A.D. Hegeman, T. Kind, O. Fiehn, E.L. Schymanski, J.E. Prenni, L.W. Sumner, P.G. Boswell, Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods, *J. Chromatogr. A* 2015 (1412) 43–51, <http://dx.doi.org/10.1016/j.chroma.2015.07.108>.
- [14] G.M. Randazzo, D. Tonoli, P. Strajhar, I. Xenarios, A. Odermatt, J. Boccard, S. Rudaz, Enhanced metabolite annotation via dynamic retention time prediction: steroidogenesis alterations as a case study, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* (2017), <http://dx.doi.org/10.1016/j.jchromb.2017.04.032>.
- [15] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, *BMC Bioinform.* 11 (2010) 148, <http://dx.doi.org/10.1186/1471-2105-11-148>.
- [16] M. Gerlich, S. Neumann, MetFusion: integration of compound identification strategies, *J. Mass Spectrom.* 48 (2013) 291–298, <http://dx.doi.org/10.1002/jms.3123>.
- [17] L. Ridder, J.J.J. van der Hooft, S. Verhoeven, R.C.H. de Vos, R.J. Bino, J. Vervoort, Automatic chemical structure annotation of an LC–MS(n) based metabolic profile from green tea, *Anal. Chem.* 85 (2013) 6033–6040, <http://dx.doi.org/10.1021/ac400861a>.
- [18] Y. Wang, G. Kora, B.P. Bowen, C. Pan, MIDAS: a database-searching algorithm for metabolite identification in metabolomics, *Anal. Chem.* 86 (2014) 9496–9503, <http://dx.doi.org/10.1021/ac5014783>.
- [19] F. Allen, A. Pon, M. Wilson, R. Greiner, D. Wishart, CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra, *Nucleic Acids Res.* 42 (2014) W94–99, <http://dx.doi.org/10.1093/nar/gku436>.
- [20] Mass Frontier Software, https://assets.thermofisher.com/TFS-Assets/CMD/brochures/BR63357_MassFrontier7_4-11S.L.pdf.
- [21] K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, Searching molecular structure databases with tandem mass spectra using CSI:FingerID, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 12580–12585, <http://dx.doi.org/10.1073/pnas.1509788112>.
- [22] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihiro, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software, *Anal. Chem.* 88 (2016) 7946–7958, <http://dx.doi.org/10.1021/acs.analchem.6b00770>.
- [23] Y. Ma, T. Kind, D. Yang, C. Leon, O. Fiehn, MS2Analyzer: a software for small molecule substructure annotations from accurate tandem mass spectra, *Anal. Chem.* 86 (2014) 10724–10731, <http://dx.doi.org/10.1021/ac502818e>.
- [24] T. Huan, C. Tang, R. Li, Y. Shi, G. Lin, L. Li, MyCompoundID MS/MS search: metabolite identification using a library of predicted fragment-ion-spectra of 383, 830 possible human metabolites, *Anal. Chem.* 87 (2015) 10619–10626, <http://dx.doi.org/10.1021/acs.analchem.5b03126>.
- [25] A. Aguilar-Mogas, M. Sales-Pardo, M. Navarro, R. Guimerà, O. Yanes, IMet: a network-based computational tool to assist in the annotation of metabolites from tandem mass spectra, *Anal. Chem.* 89 (2017) 3474–3482, <http://dx.doi.org/10.1021/acs.analchem.6b04512>.
- [26] W.M.B. Edmands, L. Petrick, D.K. Barupal, A. Scalbert, M.J. Wilson, J.K. Wickliffe, S.M. Rappaport, compMS2Miner: an automatable metabolite identification, visualization, and data-sharing r package for high-resolution LC–MS data sets, *Anal. Chem.* 89 (2017) 3919–3928, <http://dx.doi.org/10.1021/acs.analchem.6b02394>.
- [27] J.J.J. van der Hooft, J. Wandy, M.P. Barrett, K.E.V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 13738–13743, <http://dx.doi.org/10.1073/pnas.1608041113>.
- [28] P.-M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V.C. Pham, F. Roussi, M. Litaudon, J.-L. Wolfender, Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication, *Anal. Chem.* 88 (2016) 3317–3323, <http://dx.doi.org/10.1021/acs.analchem.5b04804>.
- [29] F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MZmine 2 data-preprocessing to enhance molecular networking reliability, *Anal. Chem.* (2017), <http://dx.doi.org/10.1021/acs.analchem.7b01563>.
- [30] C. Frainay, F. Jourdan, Computational methods to identify metabolic sub-networks based on metabolomic profiles, *Brief. Bioinform.* 18 (2017) 43–56, <http://dx.doi.org/10.1093/bib/bbv115>.
- [31] J.D. Quell, W. Römisch-Margl, M. Colombo, J. Krumsiek, A.M. Evans, R. Mohney, V. Salomaa, U. de Faire, L.C. Groop, F. Agakov, H.C. Looker, P. McKeigue, H.M. Colhoun, G. Kastenmüller, Automated pathway and reaction prediction facilitates in silico identification of unknown metabolites in human cohort studies, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* (2017), <http://dx.doi.org/10.1016/j.jchromb.2017.04.002>.
- [32] C. Junot, F. Fenaille, B. Colsch, F. Bécher, High resolution mass spectrometry based techniques at the crossroads of metabolic pathways, *Mass Spectrom. Rev.* 33 (2014) 471–500, <http://dx.doi.org/10.1002/mas.21401>.
- [33] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol. Cell. Proteomics* 11 (2012), <http://dx.doi.org/10.1074/mcp.O111.016717>, O111.016717.
- [34] G. Rosenberger, Y. Liu, H.L. Röst, C. Ludwig, A. Buil, A. Bensimon, M. Soste, T.D. Spector, E.T. Dermitzakis, B.C. Collins, L. Malmström, R. Aebersold, Inference and quantification of peptidofragments in large sample cohorts by SWATH-MS, *Nat. Biotechnol.* (2017), <http://dx.doi.org/10.1038/nbt.3908>.
- [35] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, LIPID MAPS online tools for lipid research, *Nucleic Acids Res.* 35 (2007) W606–612, <http://dx.doi.org/10.1093/nar/gkm324>.
- [36] T. Kind, K.-H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, *Nat. Methods* 10 (2013) 755–758, <http://dx.doi.org/10.1038/nmeth.2551>.
- [37] G. Hopfgartner, Can MS fully exploit the benefits of fast chromatography? *Bioanalysis* 3 (2011) 121–123, <http://dx.doi.org/10.4155/bio.10.191>.
- [38] G. Paglia, J.P. Williams, L. Menikarachi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B.O. Palsson, G. Astarita, Ion mobility derived collision cross sections to support metabolomics applications, *Anal. Chem.* 86 (2014) 3985–3993, <http://dx.doi.org/10.1021/acs500405x>.
- [39] T.P. Cleland, C.J. DeHart, R.T. Fellers, A.J. VanNispen, J.B. Greer, R.D. LeDuc, W.R. Parker, P.M. Thomas, N.L. Kelleher, J.S. Brodbelt, High-throughput analysis of intact human proteins using UVPD and HCD on an orbitrap mass spectrometer, *J. Proteome Res.* 16 (2017) 2072–2079, <http://dx.doi.org/10.1021/acs.jproteome.7b00043>.
- [40] E. Ryan, C.Q.N. Nguyen, C. Shiea, G.E. Reid, Detailed structural characterization of sphingolipids via 193 nm ultraviolet photodissociation and ultra high resolution tandem mass spectrometry, *J. Am. Soc. Mass Spectrom.* 28 (2017) 1406–1419, <http://dx.doi.org/10.1007/s13361-017-1668-1>.
- [41] H.-K. Lim, J. Chen, K. Cook, C. Sensenhauser, J. Silva, D.C. Evans, A generic method to detect electrophilic intermediates using isotopic pattern triggered data-dependent high-resolution accurate mass spectrometry, *Rapid Commun. Mass Spectrom.* 22 (2008) 1295–1311, <http://dx.doi.org/10.1002/rcm.3504>.
- [42] E. Rathahao-Paris, A. Paris, J. Bursztyka, J.-P. Jaeg, J.-P. Cravedi, L. Debrauwer, Identification of xenobiotic metabolites from biological fluids using flow injection analysis high-resolution mass spectrometry and post-acquisition data filtering, *Rapid Commun. Mass Spectrom.* 28 (2014) 2713–2722, <http://dx.doi.org/10.1002/rcm.7066>.
- [43] H. Zhang, D. Zhang, K. Ray, M. Zhu, Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry, *J. Mass Spectrom.* 44 (2009) 999–1016, <http://dx.doi.org/10.1002/jms.1610>.
- [44] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, E. Ezan, J.-C. Tabet, Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats, *J. Mass Spectrom.* 39 (2004) 655–664, <http://dx.doi.org/10.1002/jms.635>.
- [45] U. Lutz, R.W. Lutz, W.K. Lutz, Metabolic profiling of glucuronides in human urine by LC–MS/MS and partial least-squares discriminant analysis for classification and prediction of gender, *Anal. Chem.* 78 (2006) 4564–4571, <http://dx.doi.org/10.1021/ac0522299>.
- [46] R. Libert, D. Hermans, J.P. Draye, F. Van Hoof, E. Sokal, E. de Hoffmann, Bile acids and conjugates identified in metabolic disorders by fast atom bombardment and tandem mass spectrometry, *Clin. Chem.* 37 (1991) 2102–2110.

- [47] D.H. Chace, Mass spectrometry in the clinical laboratory, *Chem. Rev.* 101 (2001) 445–477.
- [48] L. Chen, L. Zhou, E.C.Y. Chan, J. Neo, R.W. Beuerman, Characterization of the human tear metabolome by LC–MS/MS, *J. Proteome Res.* 10 (2011) 4876–4882, <http://dx.doi.org/10.1021/pr2004874>.
- [49] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djombou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0—the human metabolome database in 2013, *Nucleic Acids Res.* 41 (2013) D801–807, <http://dx.doi.org/10.1093/nar/gks1065>.
- [50] H.P. Benton, J. Ivanisevic, N.G. Mahieu, M.E. Kurczy, C.H. Johnson, L. Franco, D. Rinehart, E. Valentine, H. Gowda, B.K. Ubhi, R. Tautenhahn, A. Gieschen, M.W. Fields, G.J. Patti, G. Siuzdak, Autonomous metabolomics for rapid metabolite identification in global profiling, *Anal. Chem.* 87 (2015) 884–891, <http://dx.doi.org/10.1021/ac5025649>.
- [51] T.N. Lawson, R.J.M. Weber, M.R. Jones, A.J. Chetwynd, G. Rodrı Guez-Blanco, R. Di Guida, M.R. Viant, W.B. Dunn, MS-Purity: automated evaluation of precursor ion purity for mass spectrometry-based fragmentation in metabolomics, *Anal. Chem.* 89 (2017) 2432–2439, <http://dx.doi.org/10.1021/acs.analchem.6b04358>.
- [52] G. Hopfgartner, D. Tonoli, E. Varesio, High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices, *Anal. Bioanal. Chem.* 402 (2012) 2587–2596, <http://dx.doi.org/10.1007/s00216-011-5641-8>.
- [53] S. Neumann, A. Thum, C. Böttcher, Nearline acquisition and processing of liquid chromatography–tandem mass spectrometry data, *Metabolomics* 9 (2013) 84–91, <http://dx.doi.org/10.1007/s11306-012-0401-0>.
- [54] J.P. Koelmel, N.M. Kroeger, E.L. Gill, C.Z. Ulmer, J.A. Bowden, R.E. Patterson, R.A. Yost, T.J. Garrett, Expanding lipidome coverage using LC–MS/MS data-dependent acquisition with automated exclusion list generation, *J. Am. Soc. Mass Spectrom.* 28 (2017) 908–917, <http://dx.doi.org/10.1007/s13361-017-1608-0>.
- [55] M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro, Enhanced detection and identification in metabolomics by use of LC–MS/MS untargeted analysis in combination with gas-phase fractionation, *Anal. Chem.* 86 (2014) 7558–7565, <http://dx.doi.org/10.1021/ac501353n>.
- [56] Z. Yan, R. Yan, Improved data-dependent acquisition for untargeted metabolomics using gas-phase fractionation with staggered mass range, *Anal. Chem.* 87 (2015) 2861–2868, <http://dx.doi.org/10.1021/ac504325x>.
- [57] G. Mullard, J.W. Allwood, R. Weber, M. Brown, P. Begley, K.A. Hollywood, M. Jones, R.D. Unwin, P.N. Bishop, G.J.S. Cooper, W.B. Dunn, A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications, *Metabolomics* 11 (2015) 1068–1080, <http://dx.doi.org/10.1007/s11306-014-0763-6>.
- [58] J. Blonder, M.C. Rodriguez-Galan, D.A. Lucas, H.A. Young, H.J. Issaq, T.D. Veenstra, T.P. Conrads, Proteomic investigation of natural killer cell microsomes using gas-phase fractionation by mass spectrometry, *Biochim. Biophys. Acta* 2004 (1698) 87–95, <http://dx.doi.org/10.1016/j.bbapap.2003.10.009>.
- [59] C. Tu, J. Li, S. Shen, Q. Sheng, Y. Shyr, J. Qu, Performance investigation of proteomic identification by HCD/CID fragmentations in combination with high/low-resolution detectors on a tribrid, high-field orbitrap instrument, *PLoS One* 11 (2016) e0160160, <http://dx.doi.org/10.1371/journal.pone.0160160>.
- [60] M. Wrona, T. Mauriala, K.P. Bateman, R.J. Mortishire-Smith, D. O'Connor, All-in-one analysis for metabolite identification using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry with collision energy switching, *Rapid Commun. Mass Spectrom.* 19 (2005) 2597–2602, <http://dx.doi.org/10.1002/rcm.2101>.
- [61] R.S. Plumb, K.A. Johnson, P. Rainville, B.W. Smith, I.D. Wilson, J.M. Castro-Perez, J.K. Nicholson, UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation, *Rapid Commun. Mass Spectrom.* 20 (2006) 1989–1994, <http://dx.doi.org/10.1002/rcm.2550>.
- [62] R.S. Plumb, P.D. Rainville, W.B. Potts, K.A. Johnson, E. Gika, I.D. Wilson, Application of ultra performance liquid chromatography–mass spectrometry to profiling rat and dog bile, *J. Proteome Res.* 8 (2009) 2495–2500, <http://dx.doi.org/10.1021/pr801078a>.
- [63] P.A. Vorkas, G. Isaac, M.A. Anwar, A.H. Davies, E.J. Want, J.K. Nicholson, E. Holmes, Untargeted UPLC–MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease, *Anal. Chem.* 87 (2015) 4184–4193, <http://dx.doi.org/10.1021/ac503775m>.
- [64] S.J. Perry, S. Nász, M. Saeed, A high-resolution accurate mass (HR/AM) approach to identification, profiling and characterization of in vitro nefazodone metabolites using a hybrid quadrupole Orbitrap (Q-Exactive): HR/AM approach to profiling nefazodone metabolites using a Q-Exactive, *Rapid Commun. Mass Spectrom.* 29 (2015) 1545–1555, <http://dx.doi.org/10.1002/rcm.7250>.
- [65] M. Navarro-Reig, J. Jaumot, B. Piña, E. Moyano, M.T. Galceran, R. Tauler, Metabolomic analysis of the effects of cadmium and copper treatment in *Oryza sativa* L. using untargeted liquid chromatography coupled to high resolution mass spectrometry and all-ion fragmentation, *Metallomics* 9 (2017) 660–675, <http://dx.doi.org/10.1039/C6MT00279j>.
- [66] S.S. Bird, V.R. Marur, M.J. Sniatynski, H.K. Greenberg, B.S. Kristal, Lipidomics profiling by high-resolution LC–MS and high-energy collisional dissociation fragmentation: focus on characterization of mitochondrial cardiolipins and monolysocardiolipins, *Anal. Chem.* 83 (2011) 940–949, <http://dx.doi.org/10.1021/ac102598u>.
- [67] H. Gallart-Ayala, F. Courant, S. Severe, J.-P. Antignac, F. Morio, J. Abadie, B. Le Bizec, Versatile lipid profiling by liquid chromatography–high resolution mass spectrometry using all ion fragmentation and polarity switching. Preliminary application for serum samples phenotyping related to canine mammary cancer, *Anal. Chim. Acta* 796 (2013) 75–83, <http://dx.doi.org/10.1016/j.aca.2013.08.006>.
- [68] S.S. Bird, I.G. Stavrovskaya, R.M. Gathungu, F. Touse, B.S. Kristal, Qualitative characterization of the rat liver mitochondrial lipidome using all ion fragmentation on an Exactive benchtop Orbitrap MS, *Methods Mol. Biol. Clifton NJ* 1264 (2015) 441–452, http://dx.doi.org/10.1007/978-1-4939-2257-4_36.
- [69] S.S. Bird, V.R. Marur, I.G. Stavrovskaya, B.S. Kristal, Qualitative characterization of the rat liver mitochondrial lipidome using LC–MS profiling and high energy collisional dissociation (HCD) all ion fragmentation, *Metabolomics* 9 (2013) 67–83, <http://dx.doi.org/10.1007/s11306-012-0400-1>.
- [70] R. Bruderer, O.M. Bernhardt, T. Gandhi, S.M. Miladinović, L.-Y. Cheng, S. Messner, T. Ehrenberger, V. Zanotelli, Y. Butscheid, C. Escher, O. Vitek, O. Rinner, L. Reiter, Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues, *Mol. Cell. Proteomics* 14 (2015) 1400–1410, <http://dx.doi.org/10.1074/mcp.M114.044305>.
- [71] K. Arnhard, A. Gottschall, F. Pitterl, H. Oberacher, Applying 'sequential windowed acquisition of all theoretical fragment ion mass spectra' (SWATH) for systematic toxicological analysis with liquid chromatography–high-resolution tandem mass spectrometry, *Anal. Bioanal. Chem.* 407 (2015) 405–414, <http://dx.doi.org/10.1007/s00216-014-8262-1>.
- [72] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat. Methods* 12 (2015) 523–526, <http://dx.doi.org/10.1038/nmeth.3393>.
- [73] J. Zhou, Y. Li, X. Chen, L. Zhong, Y. Yin, Development of data-independent acquisition workflows for metabolomic analysis on a quadrupole-orbitrap platform, *Talanta* 164 (2017) 128–136, <http://dx.doi.org/10.1016/j.talanta.2016.11.048>.
- [74] R. Bonner, G. Hopfgartner, SWATH acquisition mode for drug metabolism and metabolomics investigations, *Bioanalysis* 8 (2016) 1735–1750, <http://dx.doi.org/10.4155/bio-2016-0141>.
- [75] T. Cajka, R. Davis, K.J. Austin, J.W. Newman, J.B. German, O. Fiehn, J.T. Smilowitz, Using a lipidomics approach for nutritional phenotyping in response to a test meal containing gamma-linolenic acid, *Metabolomics* 12 (2016), <http://dx.doi.org/10.1007/s11306-016-1075-9>.
- [76] T. Bruderer, E. Varesio, G. Hopfgartner, The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition, *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* (2017), <http://dx.doi.org/10.1016/j.jchromb.2017.07.016>.
- [77] Y. Zhang, A. Bilbao, T. Bruderer, J. Luban, C. Strambio-De-Castillia, F. Lisacek, G. Hopfgartner, E. Varesio, The use of variable Q1 isolation windows improves selectivity in LC–SWATH–MS acquisition, *J. Proteome Res.* 14 (2015) 4359–4371, <http://dx.doi.org/10.1021/acs.jproteome.5b00543>.
- [78] X. Zhu, Y. Chen, R. Subramanian, Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography–quadrupole time-of-flight mass spectrometry, *Anal. Chem.* 86 (2014) 1202–1209, <http://dx.doi.org/10.1021/ac403385y>.
- [79] M.L. Colgrave, K. Byrne, M. Blundell, S. Heidelberger, C.S. Lane, G.J. Tanner, C.A. Howarth, Comparing multiple reaction monitoring and sequential window acquisition of all theoretical mass spectra for the relative quantification of barley gluten in selectively bred barley lines, *Anal. Chem.* 88 (2016) 9127–9135, <http://dx.doi.org/10.1021/acs.analchem.6b02108>.
- [80] T. Schmidlin, L. Garrigues, C.S. Lane, T.C. Mulder, S. van Doorn, H. Post, E.L. de Graaf, S. Lemeer, A.J.R. Heck, A.F.M. Altelaa, Assessment of SRM, MRM(3), and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer, *Proteomics* 16 (2016) 2193–2205, <http://dx.doi.org/10.1002/pmic.201500453>.
- [81] Y. Chen, Z. Zhou, W. Yang, N. Bi, J. Xu, J. He, R. Zhang, L. Wang, Z. Abliz, Development of a data-independent targeted metabolomics method for relative quantification using liquid chromatography coupled with tandem mass spectrometry, *Anal. Chem.* 89 (2017) 6954–6962, <http://dx.doi.org/10.1021/acs.analchem.6b04727>.
- [82] S. Chen, H. Kong, X. Lu, Y. Li, P. Yin, Z. Zeng, G. Xu, Pseudotargeted metabolomics method and its application in serum biomarker discovery for hepatocellular carcinoma based on ultra high-performance liquid chromatography/triple quadrupole mass spectrometry, *Anal. Chem.* 85 (2013) 8326–8333, <http://dx.doi.org/10.1021/ac4016787>.
- [83] J. Zhou, Y. Yin, Strategies for large-scale targeted metabolomics quantification by liquid chromatography–mass spectrometry, *Analyst* 141 (2016) 6362–6373, <http://dx.doi.org/10.1039/c6an01753c>.
- [84] A. Lafaye, J. Labarre, J.-C. Tabet, E. Ezan, C. Junot, Liquid chromatography–mass spectrometry and 15N metabolic labeling for quantitative metabolic profiling, *Anal. Chem.* 77 (2005) 2026–2033, <http://dx.doi.org/10.1021/ac048657g>.

- [85] L. Wu, M.R. Mashego, J.C. van Dam, A.M. Proell, J.L. Vinke, C. Ras, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ¹³C-labeled cell extracts as internal standards, *Anal. Biochem.* 336 (2005) 164–171, <http://dx.doi.org/10.1016/j.ab.2004.09.001>.
- [86] B.D. Bennett, E.H. Kimball, M. Gao, R. Osterhout, S.J. Van Dien, J.D. Rabinowitz, Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*, *Nat. Chem. Biol.* 5 (2009) 593–599, <http://dx.doi.org/10.1038/nchembio.186>.
- [87] S. Arrivault, M. Guenther, S.C. Fry, M.M.F.F. Fuenfgeld, D. Veyel, T. Mettler-Altmann, M. Stitt, J.E. Lunn, Synthesis and use of stable-Isotope-Labeled internal standards for quantification of phosphorylated metabolites by LC-MS/MS, *Anal. Chem.* 87 (2015) 6896–6904, <http://dx.doi.org/10.1021/acs.analchem.5b01387>.
- [88] T. Cajka, O. Fiehn, Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics, *Anal. Chem.* 88 (2016) 524–545, <http://dx.doi.org/10.1021/acs.analchem.5b04491>.
- [89] S.U. Bajad, W. Lu, E.H. Kimball, J. Yuan, C. Peterson, J.D. Rabinowitz, Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry, *J. Chromatogr. A* 1125 (2006) 76–88, <http://dx.doi.org/10.1016/j.chroma.2006.05.019>.
- [90] R. Wei, G. Li, A.B. Seymour, High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics, *Anal. Chem.* 82 (2010) 5527–5533, <http://dx.doi.org/10.1021/ac100331b>.
- [91] M. Yuan, S.B. Breitkopf, X. Yang, J.M. Asara, A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue, *Nat. Protoc.* 7 (2012) 872–881, <http://dx.doi.org/10.1038/nprot.2012.024>.
- [92] Z. Yan, R. Yan, Increase the accessibility and scale of targeted metabolomics: construction of a human urinary metabolome-wide multiple reaction monitoring library using directly-coupled reversed-phase and hydrophilic interaction chromatography, *Anal. Chim. Acta* 894 (2015) 65–75, <http://dx.doi.org/10.1016/j.jaca.2015.08.056>.
- [93] M. Yang, Z. Yu, S. Deng, X. Chen, L. Chen, Z. Guo, H. Zheng, L. Chen, D. Cai, B. Wen, Q. Wu, F. Liang, A targeted metabolomics MRM-MS study on identifying potential hypertension biomarkers in human plasma and evaluating acupuncture effects, *Sci. Rep.* 6 (2016) 25871, <http://dx.doi.org/10.1038/srep25871>.
- [94] A.P. Siskos, P. Jain, W. Römisch-Margl, M. Bennett, D. Achaintre, Y. Asad, L. Marney, L. Richardson, A. Koulman, J.L. Griffin, F. Raynaud, A. Scalbert, J. Adamski, C. Prehn, H.C. Keun, Interlaboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma, *Anal. Chem.* 89 (2017) 656–665, <http://dx.doi.org/10.1021/acs.analchem.6b02930>.
- [95] A.C. Peterson, J.D. Russell, D.J. Bailey, M.S. Westphall, J.J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics, *Mol. Cell. Proteomics* 11 (2012) 1475–1488, <http://dx.doi.org/10.1074/mcp.0112.020131>.
- [96] S. Gallien, E. Duriez, C. Crone, M. Kellmann, T. Moehring, B. Domon, Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer, *Mol. Cell. Proteomics* 11 (2012) 1709–1723, <http://dx.doi.org/10.1074/mcp.0112.019802>.
- [97] A. Bourmaud, S. Gallien, B. Domon, Parallel reaction monitoring using quadrupole-orbitrap mass spectrometer: principle and applications, *Proteomics* 16 (2016) 2146–2159, <http://dx.doi.org/10.1002/pmic.201500543>.
- [98] N. Qiu, D. Wu, X. Cui, G. Li, S. Fan, D. Chen, Y. Zhao, Y. Wu, Simultaneous determination of multiple intracellular primary metabolites by ultrahigh performance liquid chromatography coupled with a q exactive HF mass spectrometer, *Anal. Chem.* 88 (2016) 9647–9653, <http://dx.doi.org/10.1021/acs.analchem.6b02417>.
- [99] J. Zhou, H. Liu, Y. Liu, J. Liu, X. Zhao, Y. Yin, Development and evaluation of a parallel reaction monitoring strategy for large-scale targeted metabolomics quantification, *Anal. Chem.* 88 (2016) 4478–4486, <http://dx.doi.org/10.1021/acs.analchem.6b00355>.
- [100] Z. Li, Y. Li, W. Chen, Q. Cao, Y. Guo, N. Wan, X. Jiang, Y.J. Tang, Q. Wang, W. Shui, Integrating MS1 and MS2 scans in high-resolution parallel reaction monitoring assays for targeted metabolite quantification and dynamic (¹³C)-labeling metabolism analysis, *Anal. Chem.* 89 (2017) 877–885, <http://dx.doi.org/10.1021/acs.analchem.6b03947>.
- [101] J. Zhou, C. Liu, D. Si, B. Jia, L. Zhong, Y. Yin, Workflow development for targeted lipidomic quantification using parallel reaction monitoring on a quadrupole-time of flight mass spectrometry, *Anal. Chim. Acta* 972 (2017) 62–72, <http://dx.doi.org/10.1016/j.jaca.2017.04.008>.
- [102] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem.* 78 (2006) 779–787, <http://dx.doi.org/10.1021/ac051437y>.
- [103] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinf.* 11 (2010) 395, <http://dx.doi.org/10.1186/1471-2105-11-395>.
- [104] W. Dai, P. Yin, Z. Zeng, H. Kong, H. Tong, Z. Xu, X. Lu, R. Lehmann, G. Xu, Nontargeted modification-specific metabolomics study based on liquid chromatography-high-resolution mass spectrometry, *Anal. Chem.* 86 (2014) 9146–9153, <http://dx.doi.org/10.1021/ac502045j>.
- [105] H.L. Röst, G. Rosenberger, P. Navarro, L. Gillet, S.M. Miladinović, O.T. Schubert, W. Wolski, B.C. Collins, J. Malmström, L. Malmström, R. Aebersold, OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data, *Nat. Biotechnol.* 32 (2014) 219–223, <http://dx.doi.org/10.1038/nbt.2841>.
- [106] C.-C. Tsou, D. Avtonomov, B. Larsen, M. Tucholska, H. Choi, A.-C. Gingras, A.I. Nesvizhskii, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics, *Nat. Methods* 12 (2015) 258–264, <http://dx.doi.org/10.1038/nmeth.3255> (7 p. following 264).
- [107] P. Roepstorff, J. Fohlman, Proposal for a common nomenclature for sequence ions in mass spectra of peptides, *Biomed. Mass Spectrom.* 11 (1984) 601, <http://dx.doi.org/10.1002/bms.1200111109>.
- [108] K. Biemann, Contributions of mass spectrometry to peptide and protein structure, *Biomed. Environ. Mass Spectrom.* 16 (1988) 99–111.
- [109] I. Nikolskiy, N.G. Mahieu, Y.-J. Chen, R. Tautenhahn, G.-J. Patti, An untargeted metabolomic workflow to improve structural characterization of metabolites, *Anal. Chem.* 85 (2013) 7713–7719, <http://dx.doi.org/10.1021/ac400751j>.
- [110] H. Li, Y. Cai, Y. Guo, F. Chen, Z.-J. Zhu, MetDIA: targeted metabolite extraction of multiplexed MS/MS spectra generated by data-independent acquisition, *Anal. Chem.* 88 (2016) 8757–8764, <http://dx.doi.org/10.1021/acs.analchem.6b02122>.
- [111] J.D. Egerton, B. MacLean, R. Johnson, Y. Xuan, M.J. MacCoss, Multiplexed peptide analysis using data-independent acquisition and Skyline, *Nat. Protoc.* 10 (2015) 887–903, <http://dx.doi.org/10.1038/nprot.2015.055>.
- [112] U. Kusebauch, D.S. Campbell, E.W. Deutsch, C.S. Chu, D.A. Spicer, M.-Y. Brusniak, J. Slagel, Z. Sun, J. Stevens, B. Grimes, D. Shteynberg, M.R. Hoopmann, P. Blattmann, A.V. Ratushny, O. Rinner, P. Picotti, C. Carapito, C.-Y. Huang, M. Kapousouz, H. Lam, T. Tran, E. Demir, J.D. Aitchison, C. Sander, L. Hood, R. Aebersold, R.L. Moritz, Human SRMAtlas a resource of targeted assays to quantify the complete human proteome, *Cell* 166 (2016) 766–778, <http://dx.doi.org/10.1016/j.cell.2016.06.041>.
- [113] A. Bilbao, Y. Zhang, E. Varesio, J. Luban, C. Strambio-De-Castillia, F. Lisacek, G. Hopfgartner, Ranking fragment ions based on outlier detection for improved label-free quantification in data-independent acquisition LC-MS/MS, *J. Proteome Res.* 14 (2015) 4581–4593, <http://dx.doi.org/10.1021/acs.jproteome.5b00394>.
- [114] G. Chen, S. Walmsley, G.C.M. Cheung, L. Chen, C.-Y. Cheng, R.W. Beuerman, T.Y. Wong, L. Zhou, H. Choi, Customized consensus spectral library building for untargeted quantitative metabolomics analysis with data independent acquisition mass spectrometry and MetaboDIA workflow, *Anal. Chem.* 89 (2017) 4897–4906, <http://dx.doi.org/10.1021/acs.analchem.6b05006>.
- [115] J.P. Koelmel, N.M. Kroeger, C.Z. Ulmer, J.A. Bowden, R.E. Patterson, J.A. Cochran, C.W.W. Beecher, T.J. Garrett, R.A. Yost, LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data, *BMC Bioinf.* 18 (2017) 331, <http://dx.doi.org/10.1186/s12859-017-1744-3>.

3 Implémentation des nouveaux modes d'acquisition sur l'Orbitrap Fusion

3.1 Objectifs de l'étude

Le but de cette étude est de mettre au point différentes méthodes d'acquisition de données combinant l'acquisition de spectres MS et MS² sur l'Orbitrap Fusion pour l'analyse métabolomique par LC-HRMS. Deux méthodes basées respectivement sur l'acquisition données dépendante (DDA) et sur l'acquisition données indépendante (DIA) seront développées et optimisées. Ces deux approches seront ensuite comparées, avec la méthode MS seule et entre elles, afin d'évaluer leurs atouts et leurs faiblesses respectifs ainsi que la pertinence de leur utilisation en routine pour des analyses métabolomiques.

3.2 Aspects méthodologiques

Un jeu test de 50 composés ainsi qu'un extrait de plasma de référence (SRM 1950) commercialisé par le NIST (*National Institute for Standards and Technologies*) ont été utilisés pour développer et évaluer les méthodes. L'accent est d'abord mis sur l'approche DDA. Dans cette approche, un certain nombre de signaux sont automatiquement sélectionnés pour subir une fragmentation. Une des étapes clés dans la mise au point de cette méthode est la construction et l'optimisation de l'arbre décisionnel permettant la sélection des signaux à fragmenter. Dans cette approche, les métabolites sont traités différemment selon qu'il s'agisse de métabolites présents dans la base de données interne au laboratoire ou de molécules non connues, et ce afin d'acquérir les informations les plus pertinentes en vue de leur annotation. Dans un second temps, la méthode DIA a été optimisée, essentiellement en jouant sur la taille des différentes fenêtres d'isolement. Les différentes méthodes ont ensuite été comparées en termes de nombre de métabolites fragmentés, de pureté de spectres et de qualité des fragmentations observées. Enfin, la pertinence de l'utilisation des approches multi-événements et de l'acquisition DIA pour la quantification a été évaluée et comparée à la quantification réalisée lors d'une acquisition MS seule, grâce à des gammes de concentration et l'établissement d'une limite de quantification pour chaque métabolite du jeu de test une fois dopés dans l'extrait de plasma.

3.3 Publication numéro 5

Evaluation of Data Dependent and Data Independent Acquisition workflows on an Orbitrap Fusion for metabolomics applications

Pierre Barbier Saint Hilaire, Kathleen Rousseau, Annelaure Damont, Christophe Junot, François Fenaille

Draft version

Les données supplémentaires liées à cette publication sont présentées en Annexe 7

Evaluation of Data Dependent and Data Independent Acquisition Workflows on an Orbitrap Fusion for Metabolomics applications

Pierre BARBIER SAINT HILAIRE¹, Kathleen ROUSSEAU¹, Annelaure DAMONT¹,
Christophe JUNOT¹, François FENAILLE*,¹

¹ Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France.

ABSTRACT

The newly developed high field Orbitrap results in the improvement of acquisition speed without downgrading the high resolution. This gives new opportunities to implement both DDA and DIA protocols on Orbitrap instruments for metabolomics application. Furthermore, new tribrid Orbitrap instruments were developed. These instruments combine one high resolution Orbitrap detector with a low-resolution linear ion trap, enabling acquisition of numerous MS/MS spectra at low resolution in parallel with high resolution MS spectrum. In this context, we developed two acquisition methods on Orbitrap Fusion implementing all-in-one MS and MS/MS data acquisition. The suitability of these methods was evaluated for integrated metabolomics experiments. In the first one, MS/MS data are acquired in Data Dependent Acquisition mode, either on the orbitrap instrument or on the linear ion trap, depending on the presence of this compound in internal data base. For compound present in the database, MS/MS spectra are acquired at low resolution, which save time and enable spectra comparison. For compounds absent from the database, high resolution MS/MS spectra are acquired to perform fragment and thus parent ion identification. With this method we acquired high purity MS/MS spectrum (mean purity 87 %), enabling easy MS/MS spectrum comparison to standards or *de novo* interpretation even if this method suffers from a relatively low number of fragmented

metabolites (60 % of MS detected species in human plasma) and favorize high concentration compounds. The second method is a Data Independent Acquisition method. In these methods, there is no specific precursor selection. Therefore, MS/MS spectrum are of lower purity (33%) than with DDA, but all metabolites were fragmented, even those at low concentration. The comparison of these two approaches highlights their complementarity and leads us to recommend the use of the Data Dependent Acquisition mode for the annotation of unknown metabolites, where the high purity is essential to link fragment ions to the precursor ion, and the use of Data Independent Acquisition mode for the confirmation of already annotated compounds, where fragment ions are already known. Moreover, quantification based on DIA-MS/MS approach was evaluated. Due to high structural diversity in the metabolite's world and to the heterogeneity of optimal collision energy, the full scan HRMS acquisition still appears as more sensitive than a DIA approach. We therefore recommend to keep a high-resolution full scan MS acquisition in complement with potential use of DDA or DIA approaches.

▪ INTRODUCTION

Metabolomics aims at studying the whole metabolite content of biological media. Its objectives are to identify and quantify all the small molecules present in complex matrices in a comprehensive manner. LC/HRMS based metabolomics workflow usually starts with full scan MS profile acquisition in order to detect as many signals as possible and to perform relative quantification¹. These signals then need to be annotated in order to render possible biological interpretation and this step still represents a key challenge LC/HRMS-based metabolomics studies. To perform annotation, some information is acquired during the MS profile acquisition, such as the retention time, the exact mass² or the relative isotopic abundance³. But one of the most valuable elements to confirm metabolites annotation is the acquisition of fragmentation spectra and their comparison to reference MS/MS spectra⁴. To acquire these spectra, samples are generally reanalyzed on a target MS/MS mode to acquire MS/MS spectra in order to confirm these annotations. This doubles analysis time and is sample consuming. New types of acquisition protocols combining MS and MS/MS acquisition in one single run have been developed recently⁵. These protocols require high scan rates in order to acquire numerous MS/MS spectra without downgrading the full scan peak definition linked to the use of an LC separation. Until now, hybrid quadrupole time of flight instruments were the more suited for this kind of approach^{6,7}. But newly developed Orbitrap instruments with improved acquisition speed with high mass resolution are now compatible with these protocols⁸⁻¹⁰. MS and MS/MS combining protocols can roughly be divided in two categories. The first one is Data Dependent Acquisition (DDA). In this mode, the precursor ions for MS/MS will be automatically selected from the full scan MS profiles according to their intensities and other configurable filters. The isolation is performed within a small isolation window (typically 1 Da) which leads to high quality MS/MS spectra with low interference. However, as the precursor selection is automatic

and dependent of ion observed on the MS profile, the acquisition workflow suffers from low reproducibility and some interesting features might not be fragmented. To reduce these disadvantages, many selection tools or filters have been added to acquisition software. They need to be optimized carefully in order to select ions of interest rather than irrelevant ions. The second category is Data Independent Acquisition (DIA). In this mode, MS/MS is performed without regards to the MS profile. Two distinct approaches have been developed in DIA. The first one, called All Ion Fragmentation (AIF) or MS^E ^{11,12}, consists in the fragmentation of the whole m/z range in one spectrum. With this acquisition mode, all ions are fragmented which allow the fragmentation of lower intensity ions. But, as many are fragmented simultaneously, the direct link between a specific precursor ion and its corresponding fragments is lost. This main drawback makes MS/MS interpretation much more difficult to interpret, especially for untargeted metabolomics. With the increased acquisition speed of mass spectrometers, allowing the acquisition of more MS/MS events within the same cycle time, a new approach, called SWATH was developed aiming at reducing MS/MS spectra complexity ¹³. In SWATH acquisition, the whole m/z range is divided in small consecutive isolation windows (typically between 10 and 50 Da) and all ions within these windows are fragmented simultaneously. This tends to combine the benefits of AIF (reproducibility, fragmentation of all ions) with the ones of DDA (high MS/MS spectra quality). These protocols are already used for many applications in proteomics or in lipidomics.

All these acquisition protocols now tend to develop in the metabolomics field. Some of them have already been used in this field, mainly using Q-ToF instruments¹⁴ or Orbitrap at a reduced resolution ^{8,9}. Both DDA and DIA approaches tends to be more broadly used ^{15,16}. The aim of this article is to evaluate the performance of these MS-MS/MS protocols on new generation of Orbitrap Fusion ¹⁷, taking advantage of the enhanced acquisition rate to maintain high resolution full scan acquisition (resolution above 100.000) while acquiring MS/MS in the same analytical

run. This evaluation was performed on standard compounds and on plasma samples, a typical matrix for metabolomics applications. New methods, using either DDA or DIA mode, were developed, evaluated and compared, using criteria such as the number of MS² events, the diversity of fragmented metabolites or the purity of the acquired spectra. By the end, the possibility to use a DIA approach for quantification was compared to the standard MS procedure.

▪ **EXPERIMENTAL SECTION**

Chemicals and Reagents

Chemical standards were from Sigma Aldrich. Liquid chromatography solvents (i.e. ACN and water) and extraction solvent (i.e. methanol) were LC-MS grade solvents (LC-MS Chromasolv, Sigma Aldrich). Mobile phase additive (i.e. formic acid) was from Sigma Aldrich. Serum for metabolite extraction was from Biopredic.

Preparation of standards and serum

A pool of 50 chemical compounds (20 µg/mL for each compound) was prepared from solution at 1 mg/mL. The 50 compounds were chosen among the available chemical library in the laboratory according to their mass and their response in ElectroSpray Ionisation (ESI) in positive mode (Table S1, Supporting Information). The pool was then diluted in water:ACN (95:5) with 0.1 % formic acid to reach 100 ng/mL, 10 ng/mL, 3 ng/mL, 1 ng/mL, 0.5 ng/mL, 0.25 ng/mL, 0.1 ng/mL and 0.05 ng/mL.

NIST reference human plasma was prepared by metabolomics designed extraction. 100 µL of plasma were extracted by 400 µL of methanol, sonicated for 15 minutes and centrifuged at 15 000 G for 15 minutes. After a 1h30 slow precipitation on ice, samples were centrifuged at 20 000 G for 15 minutes and supernatant was dried under nitrogen. Plasma metabolites were

resuspended in water:ACN (95:5) with 0.1 % formic acid. Some plasma samples were spiked with the 50 compounds of the pool at a final concentration of 100 ng/mL and 10 ng/ml

LC-HRMS instruments and parameters

Chromatographic separation was performed on a Dionex Ultimate 3000 (ThermoFisher Scientific) with a Hypersyl Gold C18 column (150 x 2.1 mm, 1.9 μ m; ThermoFisher Scientific). Two solvents were applied. Solvent A: 0.1 % formic acid in water (v/v); solvent B: 0.1 % formic acid in ACN (v/v). Flow rate was set at 500 μ L/min. Solvent A is held at 95% for 2 min followed by an increase to 100 % solvent B over 11 min. These proportions were kept constant for 12.5 min before returning to 5% B for 4.5 min. Injection volume was set at 5 μ L.

Mass analyses were performed on an Orbitrap Fusion (ThermoFisher Scientific) with a HESI source. Source parameters applied were as follows; spray voltage = + 3500 V, sheath gas = 50 (arbitrary units), aux gas = 12 (arbitrary units), sweep gas = 3 (arbitrary units), ion transfer tube temperature = 325 °C, vaporizer temperature = 350 °C. Ion transfer parameters applied were as follows; mass range = normal, S lens RF level = 60 %. Scan parameters applied were as follows; acquisition time = 0-13 min, polarity = positive, micro scan = 1, data type = profile, AGC target = 5e4. For acquisition in MS mod only, the data are acquired in the orbitrap, resolution = 240 000, maximum injection time = 400 ms, scan range 75-1000 m/z. For MS and MS/MS acquisition different all parameters are described in the corresponding parts of the results section.

Data processing

Raw instrument data (.RAW) were processed using Xcalibur v.2.2.44 (ThermoFisher Scientific) processing methods. Genesis algorithm was used for detection (nearest RT, minimum peak height (S/N) = 3) and integration (smoothing point = 1, S/N threshold = 0.5,

enable valley detection = no, constrain peak width = no). Exact $[M+H]^+$ m/z, fragment ion m/z and expected retention time (min) are shown in supplementary data.

▪ RESULTS AND DISCUSSION

Our objective was to build and evaluate new LC/MS-based acquisition methods combining the acquisition of a MS profile and MS/MS spectra for LC-MS-based metabolomics applications. In that objective, a set of 50 standard molecules was used to develop and optimize DDA and DIA methods. These molecules were chosen as good responders in ESI + mode and with high chemical, mass and retention time diversity (Table S1, Supporting Information). In order not to downgrade the quality of acquired MS profile, the MS resolution was kept constant at 120 000 and number of MS/MS events between each MS scan was set to 10. With these parameters, the cycle time was below 1 second. This allowed the acquisition of enough data points for metabolite quantification in UHPLC (5 to 7 MS scans per chromatographic peak). As comparison, previous work on DDA acquisition performed on an Orbitrap instrument (Orbitrap Elite) was realized at a MS resolution of 30 000 with 5 MS/MS between each MS scan⁸. This highlights the improvement in scan speed of the Orbitrap Fusion compared to older instruments.

Performances of the newly developed methods were then compared using an already characterized NIST reference human plasma samples in which 72 metabolites have been previously identified by MS/MS under the same conditions (Table S2, Supporting Information).

Development of the DDA methods. For an efficient DDA method, combining MS and MS/MS events, two main elements had to be optimized: the general acquisition parameters and the selection of compounds to be fragmented. The first step of this work was to optimize the different filters available in the Orbitrap Fusion control software and to create a simple DDA method. We focused on the 5 available filters: Intensity, MonoIsotopic Precursor Selection (MIPS), Dynamic Exclusion, Target Mass Exclusion and Apex detection. The goal of these filters is to optimize the selection of the different precursor ions. Our goal was to select ions at

their highest intensity and to reduce the selection of redundant or non-relevant ions (e.g. lower intensity isotopes, mobile phase contaminants, ...). The first used filter was the “Intensity” filter, which was used to eliminate from selection low intensity ions leading to poor quality MS/MS spectra. We performed MS/MS spectra acquisition on peaks of various intensity in a plasma extract. It appears that a minimal intensity of 2.5×10^4 was necessary on Orbitrap Fusion to obtain informative MS/MS spectra with detectable fragments. The intensity threshold value was therefore set to this value. The second filter was the “Target mass exclusion”. This filter was used not to selected precursor ions coming from solvents used in mobile phase. Blank solvent samples were analyzed in order to define a list of mass present in the solvents. As UHPLC separation was performed in gradient mode, the mobile phase composition changes during elution time. The chromatographic run was therefore split in different time periods. A mean spectrum was established for each one of these time periods in order to fit the actual mobile phase compositions. In the software, exclusion lists are just based on m/z without considering any notion of intensity. This could lead to the exclusion of metabolites of high intensity if the exact same m/z is present at low intensity in the mobile phase mean spectra. To reduce the risk of excluding potential ions of interest, only the most intense m/z present in the mobile phase mean spectra of each period were placed in the exclusion list. All m/z ratios above 1×10^5 were excluded, which led to about 50 excluded mass per time period. These m/z values might correspond either to chemical contaminants or to clusters such as formic acid clusters. As contaminants strongly depend on solvents origin and ionization conditions, this list needs to be updated manually before running any analytical batch. The add on of an automatic selection of these contaminants based on a blank could be a way for improving automatization of acquisition workflows. An example of this list is shown in Table S2, Supporting Information. For target mass exclusion, a m/z tolerance of 5 ppm was used. The third filter used was the “Dynamic exclusion”. The goal of this filter was to limit the number of times the same ion is

selected in order to favor the diversity of selected m/z . A selected m/z was excluded after two selections (allowing a higher confidence on obtained mean MS/MS spectrum) for a period of time corresponding to the chromatographic peak width at mid high, i.e. 4s. The fourth filter was the “Apex detection” filter. This filter aims at selecting an m/z for fragmentation only when its intensity is maximal. 3.5 s for the mid high peak width and 65 % for the maximal peak height appeared to be optimal parameters to perform MS/MS acquisition at the apex of the peaks. The last filter was the “MIPS” filter (for MonoIsotopic Precursor Selection). For metabolomics applications the “small molecules” option was first used but it appeared that most of the ^{13}C ions were still selected, which is not desirable. The “peptide” option was then tested and gave optimal results for monoisotopic selection. It’s noteworthy that, if working well for common metabolites, this filter failed to perform correctly for some specific chemical compounds such as those bearing Chlorine or Bromine atoms. All these filters were combined, as shown in Figure 2.

For this first DDA method selected metabolites were fragmented only with one type of activation. Non-resonant collision induced dissociation mode in an hexapole (also named HCD on the Orbitrap Fusion) was chosen as the sole mode of fragmentation. Contrary to proteomics, metabolomics faces an important challenge in defining a universal optimal collision energy. This is mainly due to the structural heterogeneity of metabolite while in proteomics the fragmentation mainly focuses on the peptidic bond. In our lab, the optimal collision energy was determined for more than 500 metabolites (Supporting Information). Optimal energy was defined as the one allowing the observation of the parent ion at a relative intensity between 20% and 30%. Most of the studied metabolites had optimal energy ranging from 5% to 50%. To cover this wide range of energy, the stepped collision energy option was used. This option permits to combine three different collision energies in a single fragmentation spectrum. The batch of parent ions is split in 3 subgroups that are injected consecutively in the HCD cell. After

activation, all fragment ions will be ejected simultaneously and acquired in one unique spectrum. Due to this sequential injection, the three subgroups of the same parent ions will not receive the same activation energy and the obtained fragments will represent three different collision energies (Figure 1). This option is highly valuable for metabolomics applications due to the high heterogeneity of optimal collision energy. Based on our internal data base an HCD stepped collision energy of $30 \pm 20\%$ was selected.

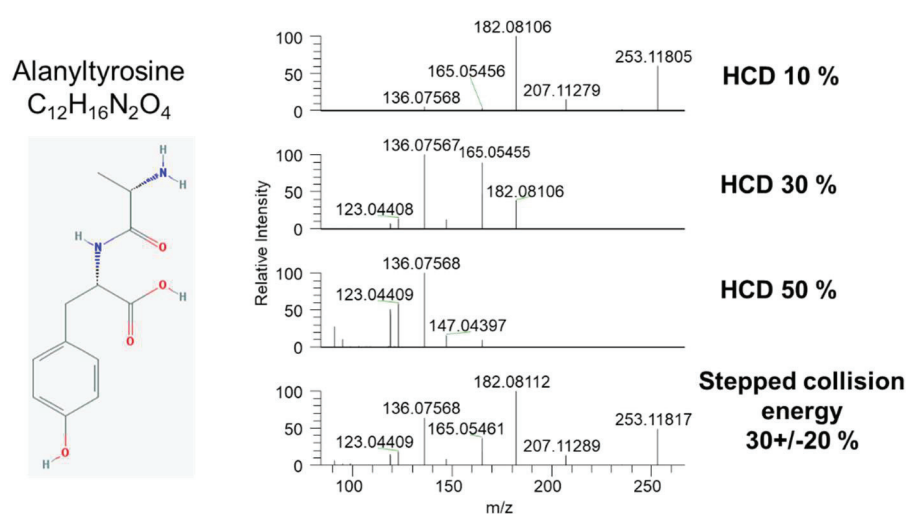


Figure 1: Example of the benefice of stepped collision energy mode for the obtention of highly informative MS/MS spectra

In order to provide higher specificity and facilitate fragment identification, MS/MS spectra were acquired in the Orbitrap at a resolution of 30 000. Figure 2 shows the whole arrangement of filters, fragmentation and acquisition. All chosen parameters are reported in the caption of the figure.

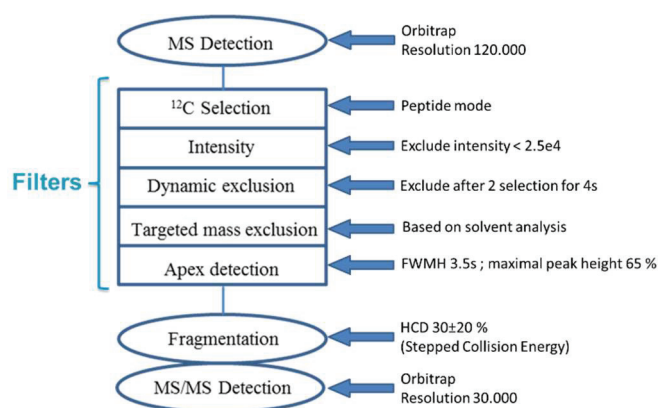


Figure 2. Diagram of the first DDA method. MS Detection: Orbitrap, resolution 120 000; MIPS: peptide; Intensity: above 2.5e4, Dynamic Exclusion: exclude after 2 repetitions with 1.5 s for 4 s; Target mass exclusion: see Supporting Information; Apex Detection: 3.5 s for the mid high peak width and 65 % for the maximal peak height; isolation: quadrupole, window 0.8s, fragmentation: HCD, 30±20 %, detection: Orbitrap, resolution 30 000.

To confirm all these parameters, this method was applied on the pool of 50 standard at 100 ng/mL in triplicates. In these conditions, 43/50 standards were detected in the MS profiles. Among these 43 compounds, only one compounds was not selected and fragmented at least once in one replicate and at least 2 MS/MS spectra were acquired for more than 88 % of compound-replicate pairs. Globally, 282 MS/MS spectra were acquired and 267 of them were comparable to the corresponding reference spectrum. For the 15 spectra left, the collision energy was not high enough for an efficient compound fragmentation.

These results demonstrate the necessity to optimize the filters' parameters using, as a first step, a simple DDA method. Once these filters and the acquisition parameters optimized, a more complex design was built for DDA application in metabolomics. In untargeted metabolomics, two categories of compounds can be detected. The first group consists in metabolites for which a standard compound is available. For these metabolites, information is available thanks to the analysis of this corresponding standard in the same specific analytical conditions. This information is mainly retention time and MS/MS spectra. The second group is constituted by metabolites that did not match with any standard present in the in-house m/z and retention time database. For these metabolites, no other information than the measured m/z is available. These

two groups of metabolites should not be treated identically in a metabolomics workflow. In the first case the goal of MS/MS acquisition is to confirm the putatively annotated metabolite by comparing the obtained MS/MS spectrum to the reference one. To perform this comparison, the spectra should be acquired in similar fragmentation conditions but high resolution is not compulsory. For the second category, as no reference compound is available, the most of complementary and accurate information is required to support a putative annotation or to propose a new annotation based on the reconstruction of precursor structure from observed fragments.

The goal of this second DDA method is to take into the specificities of these two categories of compounds and to design optimal data acquisition conditions taking into account these specificities. A new decision tree was therefore designed and presented in Figure 3.

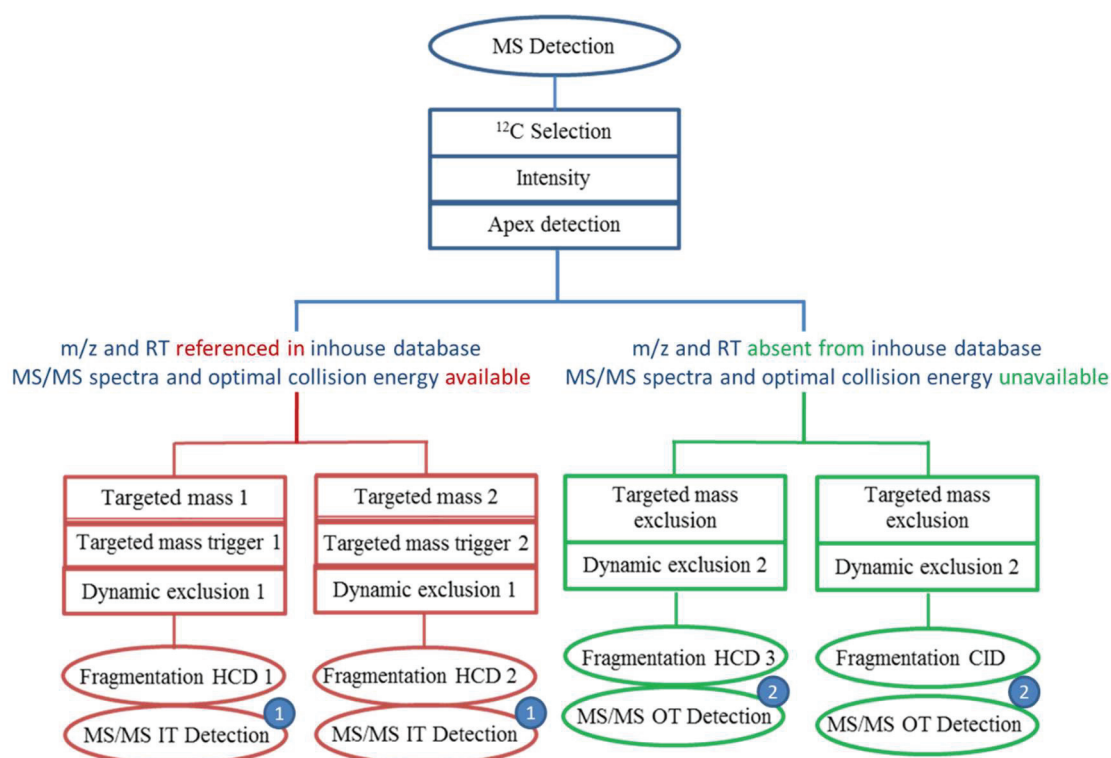


Figure 3. Diagram of the complete DDA method. MS Detection: Orbitrap, resolution 120 000; MIPS: peptide; Intensity: above $2.5e4$; Apex Detection: 3.5 s for the mid high peak width and 65 % for the maximal peak height; Target mass 1 and Target mass trigger 1: list of expected metabolites with low optimal fragmentation energy (including retention time); Target mass 2 and Target mass trigger 2: list of expected metabolites with high optimal fragmentation energy (including retention time); Dynamic exclusion 1: exclude after 3 repetitions within 3 s for 8 s; Fragmentation 1: HCD, stepped collision energy $15\pm 10\%$; Fragmentation 2: HCD, stepped collision energy $45\pm 20\%$; MS/MS detection (red): Ion trap, Normal speed. Target mass exclusion: list of expected metabolites presents in Target mass lists 1a and 1b and in blank exclusion list. Dynamic exclusion 2: exclude after 2 repetitions within 1.5 s for 4 s; Fragmentation 3: HCD, stepped collision energy $30\pm 20\%$; Fragmentation CID: CID, 22%; MS/MS detection (green): Orbitrap, resolution 15 000.

In this method some filters were common to all branches (i.e. MIPS, Intensity and Apex detection; in blue in Figure 3) and the used parameters were the same as in the first DDA method. For metabolites for which the standard compound is available and which are therefore present in our internal database (red part of decision tree, Figure 3), the main goal was to obtain high quality MS/MS spectra, i.e. with fragments and precursor ion intensity around 20%. As explained previously, optimal fragmentation energy is highly heterogeneous in metabolomics. To adapt at best the collisional energy to each compound, the set of metabolites present in the

database was split in two list, one with compounds with optimal fragmentation energy below 25% and the other with metabolite with optimal energy above 25%. For these two groups, stepped collision energy was used at $15\pm 10\%$ and $45\pm 20\%$ to cover within group versatility. As those compounds were of first interest, the dynamic exclusion parameters were modified in order to exclude precursors after 3 MS/MS acquisition. For these compounds, MS/MS spectra were compared to reference ones. Therefore, the masses of expected fragments were already known and the spectra were only used to confirm or infirm the presence of these fragments. Therefore, high resolution was not necessary and those data were acquired in the ion trap to reduce acquisition time. On the other hand, for “unknown metabolites” (i.e. metabolites not present in the reference database), no information was available according to their optimal collision energy or expected fragments. MS/MS spectra will be useful to predict precursor structure and identify corresponding metabolite. To enhance the quality of MS/MS spectra, the orbitrap was used in order to obtain accurate mass of fragments to facilitate identification. Moreover, each selected precursor was fragmented both in HCD and in CID mode for the complementarity of obtained information^{8,18}.

To evaluate the efficiency of this new DDA method, a plasma extract, spiked with our set of 50 molecules at 10 ng/mL and 100 ng/mL, was used. The sample was a NIST plasma already studied in our lab. In this sample, 72 endogenous metabolites were already characterized by ESI+ C₁₈ UHPLC/HRMS-MS/MS in the exact same conditions (Table S2, Supporting Information). With the addition of 43 detected spiked compounds (see above), our work set is constituted of 115 molecules. This set was randomly split in two subsets. The first one (69 molecules) was randomly considered as “expected” (i.e. with available information in the database) and added either to Target mass list 1 or to target mass list two according to their optimal fragmentation energy. This first set was therefore oriented to the red part of the decision

tree (Figure 3). The second set (46 molecules) was considered as “unknown molecules” and therefore oriented to the green part of the decision tree.

For each molecule, we looked at its detection in the MS profile, its selection through the DDA decision tree and the quality of the acquired MS/MS spectra (i.e. presence of fragments and positive comparison to reference spectrum). All these results are summarized in Table 1.

Table 1. Detection, selection and fragmentation results with the second designed DDA method.

	Detection MS			Selection MS/MS		
	Blank plasma	Spiked Plasma 10 ng/mL	Spiked Plasma 100 ng/mL	Blank plasma	Spiked Plasma 10 ng/mL	Spiked Plasma 100 ng/mL
Total	100 % (72/72)	96.5 % (111/115)	97.4 % (112/115)	63.9 % (46/72)	66.1 % (76/115)	72.2 % (83/115)
“Expected”	100 % (40/40)	97.1 % (67/69)	98.6 % (68/69)	67.5 % (27/40)	66.7 % (46/69)	72.5 % (50/69)
“Unknown”	100 % (32/32)	95.7 % (44/46)	95.7 % (44/46)	59.4 % (19/32)	65.2 % (30/46)	71.7 % (33/46)
	HCD Fragmentation			CID Fragmentation		
	Blank plasma	Spiked Plasma 10 ng/mL	Spiked Plasma 100 ng/mL	Blank plasma	Spiked Plasma 10 ng/mL	Spiked Plasma 100 ng/mL
Total	82.6 % (38/46)	82.9 % (63/76)	85.5 % (71/83)	NA	NA	NA
“Expected”	92.6 % (25/27)	93.5 % (43/46)	94.0 % (47/50)	NA	NA	NA
“Unknown”	68.4 % (13/19)	66.7 % (20/30)	72.7 % (24/33)	63.2 % (12/19)	53.3 % (16/30)	60.6 % (20/33)

The selection of precursor ions was not as efficient as in plasma as in solvent (previous method) but seemed to increase with the concentration of spiked compounds (66.1 % at 10 ng/mL vs. 73.0 % at 100 ng/mL). This apparent lack of selection might be linked to the complexity of the matrix. In plasma, many other ions are present, some ions of interest might detect but not selected because appearing at a too low intensity relative to other endogenous unknown

compounds. As precursor selection is performed preferentially from higher intensity ions to lower intensity ions, low intensity ions were disadvantaged. Fragmentation quality appears to be different between “expected” compounds and “unknown” compounds. For “expected” compounds, more than 92% of MS/MS spectra are interpretable and comparable to reference ones whereas this number decreased to around 70 % (HCD fragmentation) for unknown compounds. This difference was linked to the applied collision energy. For “expected” compounds, the collision energy was adapted to the optimal one ($15\pm 10\%$ or $45\pm 20\%$ for optimal energies below 25% or above 25%, respectively). For compounds considered as “unknown”, the optimal collision energy was considered as unavailable and a standard collisional energy of $30\pm 20\%$ was applied for all of them. For these “unknown”, two or more arbitrary collision energy could have been applied on each m/z. But this would have either increased the cycle time and reduce peak definition for MS quantification or reduce the number of different ions that could have been selected.

Globally this optimal DDA method proved to be efficient for combining MS and MS/MS acquisition within one analytical run for acquisition of metabolomics data. On NIST plasma analysis, this method gives valuable MS/MS spectra for more than 60 % of “expected” metabolites and brings structural information for around 40 % of the followed “unknown” metabolites through both HCD and CID fragmentation modes. These data are comparable with a previous study based on Orbitrap instruments where 30 to 40% of the most intense metabolite features were selected for MS/MS experiments⁸. Of note, in our work, the criteria were not only the selection of the feature but also the matching of the spectra with the reference one. As all endogenous molecules detected in plasma with our previous MS-only method were detected in the DDA method, this MS/MS information was acquired without downgrading the sensibility of our MS profiles. Altogether these results demonstrate the ability of Orbitrap Fusion to perform high quality Data Dependent Acquisition for metabolomics study.

Development of a DIA method. The development of the DIA method was performed for untargeted metabolomics application. We therefore choose to keep a profile MS acquisition within the methods. The method was built with a full scan MS event followed by a SWATH like fragmentation of our mass range. As in these approaches, the isolation step selected a relatively large mass range (whereas the isolation window is less than one Dalton in a DDA approach), the acquired MS/MS spectra were much more complex. To deal with such spectral complexity and to help for interpretation, we use of the Orbitrap analyzer at a resolution of 15 000, rather than the ion trap mass analyzer. The optimization of this method mainly consisted in the choice of the isolation windows width and the fragmentation energy. To keep a correct peak definition on the MS profiles (i.e. more than 6 points per peak), the number of MS/MS scans between each MS scan was set to 10 (as in the DDA method). In order to reduce the complexity of MS/MS spectra and the isolation windows width, the mass range was limited to 100-600 Da. Three different DIA methods were tested and compared. The first one was acquired via the DIA module of the Orbitrap Fusion software. With this module, the window limits are automatically calculated, taking into account the number of desired windows and the studied mass range. 10 windows of 50 Da were then automatically defined (50-100, 100-150, ..., 550-600). This method was tested on our set of standards and corresponding MS/MS fragments were observed for 56.8 % of the detected compounds. This relatively low number can be explained by the fact that the “stepped collision energy” mode was not available in the DIA module forcing the use a of fixed collision energy value, set at 30%. We therefore developed a second DIA method using the “Target MS²” module in which the “stepped collision energy” was available we defined the same MS/MS windows design, with 10 windows of 52 Da to ensure a correct detection of compounds at the border of two windows. With the method, and the used of the “stepped collision energy” at 30±20 %, the number of correctly detected

spectra rose to 77.3%. For the others, either the collision energy was unsuitable to produce correct fragmentation or signal intensity of the fragments was too low to be detected. Finally, we wanted to further optimize this method to make it compatible with a real sample metabolomics analysis. The NIST plasma sample with the 72 endogenous compounds identified was used. One major difficulty observed in real samples, compare to the our 50-compounds standard set in solvent, is the complexity of acquired MS/MS spectra. Another important point is the fact that the distribution of endogenous metabolites masses is not homogenous along the whole mass range (Figure S2, Supporting Information). To analyze the NIST plasma, inhomogeneous isolation windows were defined with smaller windows in the low mass range and larger windows in the high mass range where the number of metabolites is lower in metabolomics applications¹⁹. The new mass range splitting was: 98-132, 129.5-145.5, 144.5-160.5, 159.5-175.5, 174.5-190.5, 189-211, 209-262, 259-301, 300-400, 398-602. With these optimized windows and the “stepped collision energy”, we detected 97.5 % of the known endogenous metabolites and found the corresponding fragments for 89.7 % of them.

	No stepped collision energy Fix windows	Stepped collision energy Fix windows	Stepped collision energy Adjusted windows
Percentage of correct fragmentation in DIA	56.8 %	77.3 %	89.7 %

Table 2. Comparison of the impact of different DIA parameters on the number of correctly fragmented metabolites

Which MS-MS/MS protocol for which application in metabolomics? In the first parts of our work, we demonstrate the efficiency of both DDA and DIA approach for real sample metabolomics application. If these two acquisition modes can bring valuable information, they both have some drawbacks that are important to know to choose the best methods toward the objectives of the study. In order to highlight the comparative benefits and drawbacks of both

methods, two extracts of NIST plasma were analyzed in triplicate and the 72 known endogenous metabolites were followed. The methods were compared using three criteria: the number of acquired MS/MS events during 10 minutes (from 2 min to 12 min, i.e. the chromatographic part where metabolites are identified), the percentage of molecules for which the correct fragments are found on the corresponding MS/MS event and the purity on the acquired spectrum (i.e. the intensity of corresponding fragments on the spectra divided by the total ion current). Results are shown in Table 3.

Table 3. Comparison of DDA and DIA results based on the 72 metabolites characterized in NIST reference human plasma under identical conditions.

	DDA	DIA
Number of event in 10 min	694	723
Number of fragmented molecules	44/72 (61.1%)	72/72
Number of interpretable and correct fragmentation	38/44 (86.3%)	66/72 (91.7%)
Mean number of MS/MS spectra per fragmented molecule	2.54 (sd=0.76)	11.1 (sd=8.1)
Mean purity of MS/MS spectrum	89.0 % (sd=11.4)	33.6 % (sd=29.4)

These results showed that both the number of events and the percentage of correct fragmentation were quite comparable between the two methods. The main differences were found in the number of fragmented molecules and in the purity of MS/MS spectra. As in DIA method, no specific ion selection is performed, all molecules can be fragmented. To the contrary, in the DDA approach, only 60 % of the studied molecules are selected and fragmented, even if globally, more than 3750 different signals were selected and fragmented. If this number was relatively low compared to the DIA method, it was compensated by a much higher mean spectrum purity (89.0 % compared to 33.6 % for the DIA methods). Due to these differences, each of the two approaches might give better results in specific cases. DIA approach is recommended for targeted or semi-targeted metabolomics studies. Indeed, in such case number of followed metabolites is reduced and reference MS/MS spectra are generally

available. Thanks to this reference information, even if the MS/MS spectra is quite complex due to simultaneous selection of several precursors, it is possible to check specifically some particular fragments in order to confirm annotations. The great advantage of this approach is that all the annotated metabolites are fragmented and the detection of specific fragments is possible for each of them. On the contrary, for untargeted metabolomic approach, the lack of MS/MS spectrum is highly deleterious. If no reference MS/MS spectrum is available, complex acquired MS/MS spectrum will mostly be impossible to interpret and the useful information will remain hidden. Therefore, fragmenting a lower number of species but with high quality MS/MS spectra should be more informative and DDA mode should be prioritized.

DDA and DIA acquisition modes are therefore highly complementary and each of them will bring valuable information, depending on the type of metabolomic approach.

Relative quantification on full scans HRMS spectrum acquired with HRMS-MS/MS protocols. We finally wished to evaluate the opportunity to use the newly developed HRMS-MS/MS methods for relative quantification of real samples instead of the classically used HRMS only acquisition mode. The first point was to assess the impact of the presence on MS/MS acquisitions on relative quantification based on HRMS profiles. The set of 50 compounds was spiked into two NIST plasma extracts at concentrations ranging from 0.05 ng/mL to 10 ng/mL and each sample was analyzed in triplicates. The first step was to define the criteria for determination of the limit of relative quantification for HRMS. In fact, due to noise suppression during FID acquisition and Fourier transform on Orbitrap instruments, the usual signal to noise ratio cannot be used efficiently on this type of instrument. For MS, limit of relative quantification was therefore defined as the lower concentration for which the coefficient of variation was below 30% (among 6 replicates) at this concentration and for all higher concentrations. As some molecules of the standard set are also present endogenously in

plasma, a second criterion was set using the ratio of area observed in spiked plasma divided by the one observed in blank plasma. Spike concentration were considered as quantified for area ratio above 5. With both these criteria, a limit of spike relative quantification was determined for 30/50 compounds. As shown in Figure 4, the acquisition of MS/MS events, either in DDA or in DIA has no impact on the limits of relative quantification with the same number (± 1) of molecules quantified for each concentration. This confirms that adding MS/MS events does not downgrade the quality of the full scan HRMS acquisition.

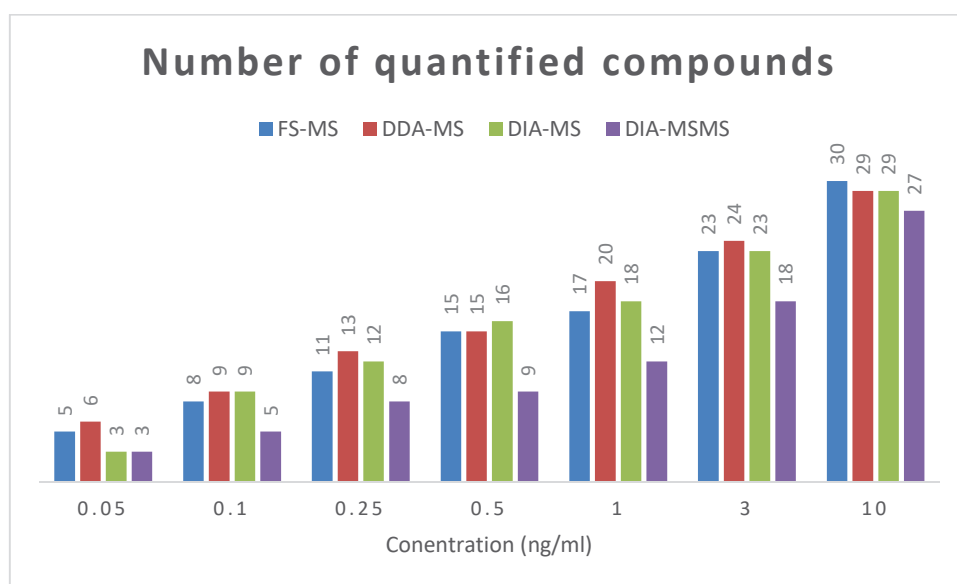


Figure 4. Number of relatively quantified molecules based on HRMS profiles (full scan method only in blue, DDA method in red and DIA method in green) and DIA MS/MS profiles (purple).

Linearity of the signal was also assessed for every compound for which at least 4 levels of concentration were available; i.e. a relative limit of quantification was equal or below 1 ng/mL, with 6 points per concentration level. For all acquisition methods, the linearity proved rather correct with a $R^2 \geq 0.98$ for all compounds except for 7 α -Hydroxytestosterone on the Full scan only acquisition ($R^2=0.97$). (Figure S3, Supporting Information).

Relative quantification on MS/MS DIA spectrum vs HRMS spectrum. We also wished to evaluate the opportunity to use DIA MS/MS spectrum for relative quantification. For MS/MS, attention was paid to the precursor ion (if still present) and the three most intense fragments (defined based on reference MS/MS spectra). Relative quantification criteria for MS/MS was based on the most intense of these 4 ions. It was defined as the lower concentration for which we observe a peak of at least 5 points, a coefficient of variation for measured area below 30% and a spiked to blank plasma area ratio above 5. With these three criteria, a limit of relative quantification was determined for 27/50 compounds. The number of detected molecules at each concentration for both approaches are shown in Figure 4. At all concentrations, the number of relatively quantified spiked molecules was higher for the HRMS methods than for the MS/MS-DIA method. The sensitivity of the full scan HRMS methods appears slightly higher than the one of the MS/MS method, when applied to real-life plasma samples. This might be due to different factors. The first one is the transmission efficiency through the ion optics and the storage of ions in the collision cell for fragmentation, that leads to a loss of some ions. The second point is that to perform DIA in a LC-MS/MS method, the resolution of the MS/MS spectra must be downgraded (compared to HRMS) to keep a sufficient number of points per chromatographic peak. The last point is the heterogeneity of optimal fragmentation energy in metabolomics (Figure S1, Supporting Information). In proteomics, the collision energy for peptides' fragmentation is quite homogeneous and an increased energy is generally required to fragment higher m/z peptide. For these applications, it is therefore quite easy to observe good fragmentation patterns for a defined energy. In metabolomics, the optimal activation energies are not highly correlated to m/z ⁸. Even using the “stepped collision energy”, a selected energy often leads to total fragmentation of some metabolites while others are not fragmented. All these drawbacks of DIA MS/MS quantification are not balanced by the fact that a reduced number of ions is present in the Orbitrap during spectra acquisition, which should have led to

an increase sensitivity. The fragment-based relative quantification using DIA approach therefore appears less suited to metabolomics approach than proteomics.

Nevertheless, the signal linearity appears as good for MS/MS acquisition than for HRMS profiles with $R^2 \geq 0.98$ for all compounds expected for 7α -Hydroxytestosterone ($R^2=0.97$) (Figure 3, Supporting information).

In conclusion, both Data Dependent Acquisition and Data Independent Acquisition (SWATH-like) can be efficiently used in LC-MS/MS based metabolomics study on Orbitrap Fusion in order to acquire MS/MS spectra. Both methods have their own advantages and drawbacks. For targeted or semi-targeted metabolomics, either DDA at low resolution using linear ion trap or DIA at higher resolution using the Orbitrap showed similar levels of efficiency in terms of annotation confirmation while for untargeted metabolomics, DDA allows acquisition of high purity MS/MS spectra which is compulsory to help identification of unknowns. In terms of quantification, the addition of MS/MS events, either DDA or DIA, in the acquisition method does not impact the HRMS profile. Nevertheless, due the heterogeneity of metabolite structure and optimal collision energy, the HRMS acquisition still appears more sensitive than a MS/MS SWATH approach. To be competitive for quantification, the SWATH acquisition in metabolomics will need even faster Orbitrap to enable the use of multiple collision energies over the whole mass range.

- (1) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* **2015**, *3*. <https://doi.org/10.3389/fbioe.2015.00023>.
- (2) Kind, T.; Fiehn, O. Metabolomic Database Annotations via Query of Elemental Compositions: Mass Accuracy Is Insufficient Even at Less than 1 Ppm. *BMC Bioinformatics* **2006**, *7*, 234. <https://doi.org/10.1186/1471-2105-7-234>.
- (3) Barbier Saint Hilaire, P.; Hohenester, U. M.; Colsch, B.; Tabet, J.-C.; Junot, C.; Fenaille, F. Evaluation of the High-Field Orbitrap Fusion for Compound Annotation in Metabolomics. *Anal. Chem.* **2018**, *90* (5), 3030–3035. <https://doi.org/10.1021/acs.analchem.7b05372>.
- (4) Nash, W. J.; Dunn, W. B. From Mass to Metabolite in Human Untargeted Metabolomics: Recent Advances in Annotation of Metabolites Applying Liquid Chromatography-Mass Spectrometry Data. *TrAC Trends Anal. Chem.* **2018**. <https://doi.org/10.1016/j.trac.2018.11.022>.
- (5) Fenaille, F.; Barbier Saint-Hilaire, P.; Rousseau, K.; Junot, C. Data Acquisition Workflows in Liquid Chromatography Coupled to High Resolution Mass Spectrometry-Based Metabolomics: Where Do We Stand? *J. Chromatogr. A* **2017**, *1526*, 1–12. <https://doi.org/10.1016/j.chroma.2017.10.043>.
- (6) Junot, C.; Fenaille, F.; Colsch, B.; Bécher, F. High Resolution Mass Spectrometry Based Techniques at the Crossroads of Metabolic Pathways. *Mass Spectrom. Rev.* **2014**, *33* (6), 471–500. <https://doi.org/10.1002/mas.21401>.
- (7) Wang, Y.; Feng, R.; Wang, R.; Yang, F.; Li, P.; Wan, J.-B. Enhanced MS/MS Coverage for Metabolite Identification in LC-MS-Based Untargeted Metabolomics by Target-Directed Data Dependent Acquisition with Time-Staggered Precursor Ion List. *Anal. Chim. Acta* **2017**, *992*, 67–75. <https://doi.org/10.1016/j.aca.2017.08.044>.
- (8) Mullard, G.; Allwood, J. W.; Weber, R.; Brown, M.; Begley, P.; Hollywood, K. A.; Jones, M.; Unwin, R. D.; Bishop, P. N.; Cooper, G. J. S.; et al. A New Strategy for MS/MS Data Acquisition Applying Multiple Data Dependent Experiments on Orbitrap Mass Spectrometers in Non-Targeted Metabolomic Applications. *Metabolomics* **2015**, *11* (5), 1068–1080. <https://doi.org/10.1007/s11306-014-0763-6>.
- (9) Zhou, J.; Li, Y.; Chen, X.; Zhong, L.; Yin, Y. Development of Data-Independent Acquisition Workflows for Metabolomic Analysis on a Quadrupole-Orbitrap Platform. *Talanta* **2017**, *164* (Supplement C), 128–136. <https://doi.org/10.1016/j.talanta.2016.11.048>.
- (10) Zhou, J.; Liu, H.; Liu, Y.; Liu, J.; Zhao, X.; Yin, Y. Development and Evaluation of a Parallel Reaction Monitoring Strategy for Large-Scale Targeted Metabolomics Quantification. *Anal. Chem.* **2016**, *88* (8), 4478–4486. <https://doi.org/10.1021/acs.analchem.6b00355>.
- (11) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MSE; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20* (13), 1989–1994. <https://doi.org/10.1002/rcm.2550>.
- (12) Wrona, M.; Mauriala, T.; Bateman, K. P.; Mortishire-Smith, R. J.; O'Connor, D. 'All-in-One' Analysis for Metabolite Identification Using Liquid Chromatography/Hybrid Quadrupole Time-of-Flight Mass Spectrometry with Collision Energy Switching. *Rapid Commun. Mass Spectrom.* **2005**, *19* (18), 2597–2602. <https://doi.org/10.1002/rcm.2101>.
- (13) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717. <https://doi.org/10.1074/mcp.O111.016717>.
- (14) Chen, L.; Zhou, L.; Chan, E. C. Y.; Neo, J.; Beuerman, R. W. Characterization of The Human Tear Metabolome by LC-MS/MS. *J. Proteome Res.* **2011**, *10* (10), 4876–4882. <https://doi.org/10.1021/pr2004874>.
- (15) Bonner, R.; Hopfgartner, G. SWATH Acquisition Mode for Drug Metabolism and Metabolomics Investigations. *Bioanalysis* **2016**, *8* (16), 1735–1750. <https://doi.org/10.4155/bio-2016-0141>.
- (16) Zhu, X.; Chen, Y.; Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MSAll Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2014**, *86* (2), 1202–1209. <https://doi.org/10.1021/ac403385y>.
- (17) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; et al. Novel Parallelized Quadrupole/Linear Ion

Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* **2013**, *85* (24), 11710–11714. <https://doi.org/10.1021/ac403115c>.

- (18) Tu, C.; Li, J.; Shen, S.; Sheng, Q.; Shyr, Y.; Qu, J. Performance Investigation of Proteomic Identification by HCD/CID Fragmentations in Combination with High/Low-Resolution Detectors on a Tribrid, High-Field Orbitrap Instrument. *PLOS ONE* **2016**, *11* (7), e0160160. <https://doi.org/10.1371/journal.pone.0160160>.
- (19) Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. The Use of Variable Q1 Isolation Windows Improves Selectivity in LC–SWATH–MS Acquisition. *J. Proteome Res.* **2015**, *14* (10), 4359–4371. <https://doi.org/10.1021/acs.jproteome.5b00543>.

3.4 Principaux résultats

Au cours de cette étude, de nouveaux modes d'acquisition de données, basés sur l'enregistrement de spectres MS et MS² lors d'une analyse unique, ont été développés. Ce travail a principalement été mené sur un mélange de composés standards (dont 43 métabolites sont détectés dans nos conditions d'analyse) et sur un extrait de métabolites plasmatiques fournis par le NIST au sein duquel 72 composés ont pu être formellement identifiés dans ces mêmes conditions (séparation sur colonne C₁₈ et mode d'ionisation positif). Deux approches ont été étudiées et comparées.

La première approche optimisée est l'acquisition données dépendante (DDA). Après une première phase de développement, nous avons mis au point une méthode correspondant au mieux à la réalité d'une analyse de métabolomique non ciblée. Cette méthode est basée sur l'idée qu'au sein d'un échantillon, un certain nombre de composés connus sont potentiellement attendus et que de nombreux autres composés non connus sont également détectés. Or ces deux groupes de composés ne sont pas traités à l'identique. Pour le premier cas, le but est de confirmer des annotations basées sur le rapport m/z et sur le temps de rétention en comparant un spectre de fragmentation avec un spectre de référence. Dans ce cas, une haute précision de mesure de masse est moins indispensable et la priorité peut être mise sur la vitesse d'acquisition de ces spectres. De plus, l'énergie de collision optimale des composés (qui est très hétérogène en métabolomique) est connue et peut donc être ajustée. Dans le cas des composés inconnus, leur identification doit être faite uniquement à partir des données acquises et la justesse et la complémentarité de ces informations sont essentielles. L'utilisation de la haute résolution est donc primordiale et l'utilisation des différents modes de fragmentation disponibles (résonnant et non-résonnant) permet d'apporter des informations complémentaires pour l'identification des composés. C'est à partir de ce postulat que la méthode DDA a été développée. Différents filtres permettant une sélection optimale des ions à fragmenter ont été optimisés. Le *Monoisotopic Precursor Selection* (réglé en mode peptide), le filtre d'intensité (fixé à 2^e) ou l'*Apex Detection* sont communs aux composés potentiellement attendus et aux inconnus. L'exclusion dynamique est paramétrée différemment selon les cas. Pour les composés connus qui seront enregistrés avec une plus haute vitesse d'acquisition, l'exclusion du composé se fait après 3 acquisitions afin d'améliorer la qualité du spectre MS² moyen obtenu en augmentant le rapport signal sur bruit alors que ce nombre d'acquisition est limité à 2 pour les composés inconnus qui seront enregistrés à une plus haute résolution et ce afin de maximiser le nombre de composés différents qui seront fragmentés. Les composés connus sont subdivisés en deux groupes en fonction de leur énergie de fragmentation optimale et l'énergie de fragmentation HCD sera donc adaptée à chaque sous-groupe. Pour les composés inconnus, il n'est pas possible d'optimiser ce paramètre et une énergie moyenne sera utilisée. De plus, chaque composé sélectionné sera fragmenté à la fois en mode HCD (non résonnant) et CID (résonnant) pour acquérir le plus d'informations permettant son identification. L'enregistrement des spectres se fera

à haute vitesse au sein du piège linéaire pour les composés connus et à haute résolution dans l'Orbitrap pour les composés inconnus. Il est important de noter l'apport du mode *stepped collision energy* est utilisé ici pour l'ensemble des acquisitions MS² en HCD. Ce mode permet d'acquérir des spectres composites issus de la fragmentation de l'ion parent à 3 énergies de collisions. Pour les composés inconnus, une énergie de 30%±20% a été sélectionnée alors que pour les composés connus la fragmentation des deux sous-groupes se fait à 15%±10% et 45%±20%.

Avec cette méthode, plus de 95% des composés détectés lors de l'analyse MS seule sont détectés montrant que l'ajout d'évènements MS² n'impacte que très faiblement la trace MS. Concernant les composés attendus, leur taux de sélection est d'environ 70% avec des spectres correspondant à ceux de référence dans plus de 90% des cas. Pour les composés inconnus, le taux de sélection est comparable à celui des composés attendus (environ 65%) mais le taux de spectres correspondant à la référence est plus bas (60 à 70%). Cela s'explique par l'impossibilité d'adapter l'énergie de collision à l'énergie optimale puisque cette dernière n'est pas connue. Les 30 à 35 % de composés non sélectionnés sont ceux ayant la plus faible intensité et qui ne sont donc pas en tête de liste lors de la sélection automatique des masses à fragmenter. Ce type d'approche apparaît donc utilisable pour des analyses métabolomiques en routine, sans perte d'information par rapport à l'utilisation de la MS seule et avec l'enregistrement de spectres MS² de bonne qualité pour un nombre significatif de composés.

Nous avons ensuite cherché à développer une autre méthode d'acquisition basée sur le mode données indépendant (DIA). Dans ce mode, les principaux paramètres à ajuster sont la taille et la répartition des fenêtres de masse. Pour conserver une trace MS correcte sans perte de sensibilité et avec une définition des pics correcte, nous avons décidé de nous limiter à 10 fenêtres (soit 10 évènements MS² entre chaque spectre MS, comme pour la méthode DDA). Afin de ne pas avoir de fenêtres de trop grande taille, ce qui compliquerait d'autant plus l'interprétation des spectres, nous avons initialement fixé leur taille à 50 m/z pour une couverture MS² de 100 à 600 daltons. Dans ces conditions et avec l'utilisation du mode *stepped collision energy*, nous avons détecté en MS 97,5% des composés endogènes connus dans notre plasma de référence et détecté les fragments correspondants pour 77% des molécules. Suite à l'observation de spectres MS² plus chargés pour les fenêtres de basse masse, la distribution des composés endogènes a été étudiée. Comme attendu, cette distribution n'est pas homogène sur l'ensemble de la gamme de masse. De nouvelles fenêtres ajustées sur cette distribution (plus étroites pour les basses masses et plus larges pour les hautes masses) ont alors été mises en place. Avec cette méthode optimisée, les fragments caractéristiques de 90% des composés endogènes préalablement identifiés dans le plasma NIST ont pu être détectés.

Globalement, les approches DDA et DIA utilisées sur l'Orbitrap Fusion apportent toutes deux des informations complémentaires par rapport à une méthode MS seule. Cela se fait sans

détérioration du profil MS qui peut toujours être utilisé pour la détection et la quantification des composés. Chaque méthode combinant MS et MS² présente des avantages par rapport à l'autre. La DDA permet d'obtenir des spectres MS² d'une plus grande pureté (87% vs 33%) puisque la fenêtre de sélection est beaucoup plus étroite (0.8 Da vs 30 à 100 Da). Cette approche convient donc bien à l'élucidation structurale de composés inconnus. La DIA permet une meilleure couverture MS² avec une fragmentation de l'ensemble des composés sur une gamme de 100 à 600 alors que la méthode DDA ne fournit des spectres MS² que pour 60 à 70 % des composés. Si les spectres sont plus complexes et difficiles à interpréter, plus de composés sont fragmentés et elle permet tout de même de valider facilement une annotation par homologie de fragmentation avec des spectres de référence. Dans ce contexte, nous recommanderions une utilisation de type DIA lors de la réalisation d'analyses métabolomiques semi-ciblées (dont l'objectif principal est d'identifier des composés présents dans une base interne incluant des spectres MS² de référence), ce qui est en général le cas au sein de notre laboratoire. Par contre, pour des analyses métabolomiques complètement non ciblées, nous recommanderions plutôt l'utilisation d'une approche DDA qui permet l'obtention de spectres MS² plus purs et qui combine fragmentation HCD et CID afin d'avoir le plus d'information hautement qualitative possible pour une identification *de novo* des composés d'intérêt.

La possibilité de quantifier les composés directement sur les traces MS² a aussi été étudiée en mode DIA (cela n'est pas envisageable en mode DDA puisque seul est nombre restreint de spectres est acquis pour un composé donné ce qui rend impossible l'obtention d'un pic chromatographique intégrable). Il est apparu que la quantification reste plus fiable à partir du profil MS (acquis seul ou conjointement avec des événement MS²) que directement sur les traces MS², contrairement à ce qui est couramment fait en protéomique. Cela est en majeure partie dû au fait que la nature chimique des liaisons rompues lors de la fragmentation de métabolites est beaucoup plus variée que lors de la fragmentation de peptides qui a essentiellement lieu au niveau de la liaison peptidique. Cette grande variabilité rend impossible le choix d'une énergie de collision unique (même avec le mode *stepped collision energy*) qui permettrait d'obtenir des fragments d'intensité suffisante et uniforme pour tous les métabolites. Pour envisager ce type de quantification, il faudrait donc démultiplier le nombre d'énergies de fragmentation appliquées ce qui multiplierait d'autant le nombre d'événements MS² à réaliser et qui n'est pour l'instant pas compatible avec une analyse LC-HRMS sur les instruments de type Orbitrap. De ce fait, le maintien d'une trace MS haute résolution est fortement recommandé.

3.5 Conclusion

Grâce aux améliorations apportées par le constructeur aux dernières générations d'Orbitrap, ce type d'instrument a acquis une vitesse suffisante pour permettre l'implémentation de méthodes combinant MS et MS². Au vu des résultats obtenus, ce type d'approche doit pouvoir devenir une référence pour les analyses métabolomiques et remplacer les protocoles d'acquisition actuels. Cela

permettra d'améliorer la capacité d'annotation et d'interprétation des données générées, par exemple lors d'études de cohortes médicales.

Chapitre 4. Discussion générale

Le travail de thèse présenté dans ce manuscrit se décompose en trois étapes. Tout d'abord, le travail bibliographique effectué présente un état des lieux de la métabolomique. La seconde partie met en application ces techniques métabolomiques sur une cohorte médicale visant à mieux caractériser et à trouver des biomarqueurs de l'encéphalopathie hépatique à partir de l'analyse d'échantillons de plasma et de liquide céphalo-rachidien. Enfin, la dernière partie présente différents développements visant à améliorer la qualité des informations obtenues et à faciliter l'interprétation des analyses de métabolomique.

Au cours de l'étude bibliographique, l'ensemble des étapes importantes d'une analyse métabolomique a été parcouru, de la collecte et la préparation des échantillons à l'interprétation biologique des résultats obtenus. Ce travail constitue à la fois un état des lieux de la métabolomique telle qu'elle est aujourd'hui appliquée en routine et une mise en évidence des tendances novatrices en métabolomique par chromatographie liquide couplée à la spectrométrie de masse, d'un point de vue instrumental, analytique ou bio-informatique. Parmi ces éléments, l'apparition de nouveaux instruments, notamment de type Orbitrap, apparaît comme un facteur d'évolution important. La résolution de ces nouveaux instruments a été accrue et ils peuvent maintenant travailler à de très hautes résolutions (1 000 000, $M/\Delta M$, FWHM, pour l'Orbitrap Lumos). Cela permet d'envisager une meilleure séparation des composés de masses moléculaires très proches et ainsi de mesurer plus finement le massif isotopique des métabolites d'intérêt. Ces éléments apportent des informations complémentaires qui peuvent être utilisées pour améliorer l'identification des métabolites d'intérêt. De plus, ces hautes résolutions sont combinées à des vitesses d'acquisition améliorées qui permettent d'envisager la mise au point de protocoles d'acquisition multi-événements sur ce type d'instruments. Cette première partie a également démontré le potentiel de la métabolomique dans l'amélioration de la compréhension de différentes maladies. En effet, de par son approche phénotypique, elle permet d'être le reflet le plus direct des dérèglements impliqués.

Nous avons tout d'abord voulu exploiter ce potentiel dans le cadre de l'encéphalopathie hépatique. Il s'agit d'une maladie sous diagnostiquée dont l'incidence médicale et économique est non négligeable. Son diagnostic est complexe et il n'existe pas de biomarqueurs objectifs permettant de détecter cette maladie et de stratifier les patients. En l'absence d'éléments objectifs de diagnostic et du fait d'une compréhension encore limitée des mécanismes mis en jeu dans l'EH, la métabolomique devrait pouvoir apporter une vraie plus-value. Nous avons donc choisi d'étudier cette maladie en utilisant les techniques « -omiques » phénotypiques disponibles au sein du laboratoire. Les objectifs de cette étude étaient multiples. Il s'agissait tout d'abord d'utiliser la métabolomique sur des échantillons de LCR pour caractériser au mieux les dérèglements

métaboliques ayant lieu au niveau cérébral chez les encéphalopathes. A cette fin, une première cohorte a été constituée. Elle est composée de volontaires sains et d'encéphalopathes. L'analyse de cette cohorte a permis de mettre en évidence un certain nombre de voies métaboliques altérées chez les encéphalopathes.

Les différents éléments observés nous ont conduit à proposer un mécanisme de l'EH en lien avec une altération du métabolisme énergétique. En effet, l'accumulation constatée de composés acétylés pourrait être liée à une diminution de la dégradation de l'acétylCoA par le cycle de Krebs. Si cette première étude apporte un nouveau regard sur l'EH et sur de potentiels nouveaux modes d'action thérapeutique, elle souffre de l'absence d'un groupe de patient cirrhotiques non encéphalopathes. Comme les encéphalopathes présentent tous une altération de la fonction hépatique, nous avons décidé de travailler sur du LCR, biofluide que nous pensions le plus à même de refléter les modifications métaboliques spécifiques du cerveau en s'affranchissant ainsi de l'impact des modifications hépatiques. Après l'analyse des premiers résultats, nous avons tout de même souhaité valider cette hypothèse en analysant un groupe de patients cirrhotiques sans signaux neurologiques pour s'assurer de pouvoir isoler la part de la signature proposée liée à l'EH de celle liée à la cirrhose. Cependant, nous nous sommes trouvés dans l'incapacité d'obtenir des échantillons de LCR pour ce groupe. Du fait des troubles de la coagulation que présentent les patients cirrhotiques, il n'est éthiquement pas envisageable de réaliser une ponction lombaire chez ces individus sans que cela n'ait de rationnel médical. Face à cette problématique, nous avons dû modifier notre choix initial de matrice biologique. Nous avons donc décidé d'utiliser des échantillons plasmatiques et avons constitué une seconde cohorte.

Cette cohorte a tout d'abord été analysée en métabolomique. Cela a permis de confirmer la signature métabolique de l'EH observée au niveau du LCR. L'évolution des différents métabolites observés est concordante entre ces deux matrices. Il apparaît néanmoins que les différences entre les groupes sont plus marquées au niveau du LCR que du plasma. Cela vient confirmer l'intérêt de l'analyse du LCR en métabolomique pour l'étude des maladies cérébrales. Cependant, le nombre de métabolites détectés y est nettement plus faible malgré une dilution analytique moins importante que pour le plasma. Cette plus faible couverture métabolique serait la contrepartie de sa plus grande spécificité et serait liée à des concentrations en métabolites circulants dans le LCR beaucoup plus faibles. Le choix de la matrice à partir de laquelle réaliser les analyses métabolomiques dépendra donc du type de maladie étudiée. Dans la majorité des cas, le plasma sera le choix numéro un car il présente une couverture métabolique plus importante, ce qui est un avantage indéniable pour la compréhension des mécanismes moléculaires mis en jeu. De plus, il s'agit d'une matrice plus facilement disponible. Le LCR devra, lui, sera privilégié pour les études ciblant les maladies causant des dérèglements cérébraux afin de pouvoir les observer le plus finement possible.

En plus des analyses classiques de métabolomique, nous avons souhaité utiliser cette cohorte pour essayer de combiner différentes analyses « -omiques » phénotypiques. Pour ce faire nous avons également analysé la cohorte plasmatique en lipidomique et en glycomique. Le but premier de ces analyses complémentaires était bien évidemment de continuer à approfondir la caractérisation de l'EH. Mais nous avons également comme objectif de voir comment les résultats obtenus par ces différentes techniques pouvaient être interprétés, non pas individuellement, mais en lien les uns avec les autres.

Concernant la métabolomique et la lipidomique, le lien s'est fait assez naturellement de par les similarités analytiques qui existent entre ces techniques et leur maturité au sein du laboratoire. La métabolomique permet de s'intéresser aux précurseurs des lipides, souvent de plus faible masse moléculaire et plus hydrophiles. Ont notamment pu être identifiées des espèces tels que la glycérophosphocholine, la phosphoéthanolamine ou la sphingosine phosphate. Tous les précurseurs identifiés étaient présents en quantité diminuée chez les encéphalopathes. La lipidomique permet l'extraction et l'identification de composés plus hydrophobes. Ces analyses ont montré que les concentrations de l'ensemble des familles lipidiques étaient également diminuées. Plus particulièrement, les glycosylglycérophosphocholines, les glycérophosphocholines et les sphingomyélines sont en concentration significativement inférieure chez les encéphalopathes par rapport aux cirrhotiques. Ces résultats concordants viennent confirmer l'apport complémentaire que peuvent avoir ces deux techniques pour l'analyse de cohortes médicales.

Concernant le glycomique, l'interprétation des résultats s'est avérée plus complexe. Cela est notamment lié à une expertise moins importante dans ce domaine, à des difficultés propres à la technique MALDI et à des problématiques biologiques. En effet, il s'agissait de l'une des premières fois que cette approche était envisagée en lien direct avec la métabolomique. De plus, l'ionisation par MALDI rend plus difficile la normalisation inter spectres. La quantité d'ions formés à chaque tir laser est relativement variable ce qui limite la comparaison directe des intensités observées entre spectres, et donc entre les échantillons. Pour pallier cette difficulté, nous avons donc uniquement raisonné en termes de ratios entre différents ions d'intérêt. Enfin, la méthode que nous avons utilisée cherche à identifier les N-glycanes greffés sur les protéines. Or la plupart des protéines plasmatiques glycosylées sont produites au niveau hépatique et leur proportion peut donc être fortement impactée chez les patients cirrhotiques ou, *a fortiori*, encéphalopathes. Il est donc compliqué de faire la part des choses entre les modifications liées à une altération globale de la synthèse protéique et celles liées plus spécifiquement à une altération de la glycosylation.

Pour évaluer la complémentarité entre glycomique et métabolomique, nous avons donc fait le choix de nous focaliser sur certains glycanes exprimés nativement principalement sur des protéines d'origine non hépatiques comme les immunoglobulines. Nous avons calculé des ratios de glycanes

entres formes proches ne différant que par la présence d'un sucre supplémentaire. Nous avons ainsi pu montrer que les formes bissectines des glycanes H4N4F1 ou H4N4F1S1 étaient surreprésentées par rapport aux formes non bissectines chez les cirrhotiques et les encéphalopathes. Ces formes bissectines correspondent à la présence d'un groupement acétylglucosamine additionnel. Ces résultats peuvent par exemple être liés à une augmentation des acétylhexosamines observée chez les encéphalopathes en métabolomique. Même s'il ne s'agit là que d'une première preuve de concept qui demande à être approfondie et renouvelée, nous avons pu montrer que les résultats de glycomique pouvaient être rapprochés de façon assez intéressante des résultats obtenus en métabolomiques.

Globalement, toute cette étude a permis de proposer un nouveau schéma des dérèglements moléculaires associés à l'encéphalopathie hépatique. Ce schéma est basé sur un dérèglement du métabolisme énergétique glucidique et lipidique. Ces nouvelles découvertes ouvrent la voie à de nouvelles cibles thérapeutiques contre cette maladie ciblant le métabolisme énergétique. Dans ce contexte, il serait intéressant de réaliser des analyses métabolomiques sur des modèles animaux d'encéphalopathie hépatique afin de voir si les signatures métaboliques obtenues chez les patients sont retrouvées au niveau cérébral. Cela permettrait par exemple de valider l'utilisation des composés acétylés en tant que marqueurs du métabolisme énergétique cérébral. Cette étude a aussi permis de montrer la complémentarité qui peut exister entre les différentes approches « -omiques » dans ce contexte de cohortes médicales. Enfin, elle a également permis de se rendre compte concrètement des étapes complexes lors d'une analyse de métabolomique et a permis de pointer du doigt certains éléments sur lesquels des améliorations analytiques pouvaient être apportées. Ces améliorations portent sur les outils permettant l'identification des composés inconnus, sur l'amélioration des modes d'acquisition de données ou encore sur l'interprétation des résultats en termes de réseaux métaboliques.

A partir de ces observations, nous avons souhaité effectuer un travail instrumental afin d'essayer d'améliorer l'efficacité des analyses métabolomiques par spectrométrie de masse. Nous avons décidé de nous focaliser tout d'abord sur l'étape d'annotation sous deux aspects différents. Le premier est d'évaluer les performances de la dernière génération d'Orbitrap disponible au laboratoire : l'Orbitrap Fusion. Nous avons cherché à optimiser les paramètres d'utilisation de cet instrument dans un contexte de métabolomique, en essayant d'obtenir des résultats facilitant le plus possible l'étape d'annotation. Pour ce faire, nous avons fixé comme critères (i) la précision de mesure de masse, pour l'ion monoisotopique mais également pour l'ion contenant un atome de ^{13}C , (ii) la capacité de l'instrument à résoudre la structure isotopique fine et enfin (iii) la précision de mesure de l'abondance isotopique relative, à la fois pour l'ion contenant un atome de ^{13}C et pour celui contenant un atome de ^{18}O .

Ce travail d'optimisation nous a conduit à proposer des valeurs pour un certain nombre de paramètres lors de l'utilisation de l'instrument en métabolomique non ciblée. Différents éléments sont pris en compte dans ces recommandations. Ayant choisi de nous intéresser principalement à la précision de mesure, nous avons en conséquence recommandé l'utilisation d'une valeur basse d'AGC. Cependant, pour d'autres types d'approches métabolomique, d'autres critères d'optimisation peuvent être utilisés. Par exemple, sur des approches de métabolomique ciblée quantitative, le critère d'optimisation peut être la sensibilité. Dans ce cas, l'utilisation d'un AGC plus élevé peut pleinement se justifier malgré une perte de précision de mesure. Ce paramétrage doit donc être réfléchi en fonction de la finalité de l'étude à effectuer. Il en va de même pour le choix de résolution qui peut être adapté en fonction du système séparatif en amont ou de la masse des molécules ciblées. Néanmoins, pour des études non ciblées, il nous est apparu plus pertinent de favoriser la précision de mesure. En utilisant ces paramètres (AGC 5^e4 et résolution de 500 000), nous avons pu montrer que les instruments de type Orbitrap permettaient d'isoler spécifiquement le signal de l'¹⁸O et que la mesure du RIA de cet ion permettait d'affiner le choix de la formule brute d'une espèce inconnue. Sur l'Orbitrap Fusion, cela est possible uniquement pour des masses allant jusqu'à 600 Da. L'Orbitrap Lumos peut atteindre une résolution deux fois plus importante que celle de l'Orbitrap Fusion. Il serait donc intéressant de renouveler ce type d'expérience sur cet instrument afin de voir jusqu'à quelle masse il est possible d'isoler le signal de l'¹⁸O. Cependant, cette résolution accessible d'un million (à m/z 200) se fait au détriment d'un doublement du temps d'acquisition (environ 2.1s par scan) ce qui rend compliqué son couplage à un système chromatographique de type HPLC. Ce temps d'acquisition est réellement un élément central dans l'évolution des possibilités offertes par les instruments de type Orbitrap.

La capacité des instruments de dernière génération à travailler à haute résolution avec des vitesses d'acquisition améliorées permet d'envisager la mise en place de protocoles d'acquisition mixtes combinant des analyses MS et MS². Ce type de protocole est inspiré de ce qui commence à être un standard d'analyse en protéomique, principalement sur des instruments de type temps de vol. Nous avons montré que l'implémentation de ce type de protocole, qu'il s'agisse de DDA ou de DIA, peut être effectuée sur des instruments Orbitrap dans un contexte de métabolomique non ciblée. Cependant, nous avons pu constater que ces protocoles étaient plus compliqués à mettre en œuvre en métabolomique qu'en protéomique. La principale raison est la grande diversité des structures identifiables en métabolomique. De ce fait, les liaisons rompues sont de natures très diverses, là où la protéomique se focalise principalement sur la liaison peptidique. Les énergies de collision adaptées peuvent être très variables pour différents métabolites et il est impossible de trouver une énergie unique permettant l'obtention de spectres de fragmentation interprétables pour toutes les molécules.

Pour pallier cette difficulté, les concepteurs de l'Orbitrap ont mis en place une fonction « *stepped collision energy* » qui permet d'obtenir un spectre composite entre plusieurs énergies. Cela constitue un véritable apport dans la mesure où il est possible d'augmenter le nombre de molécules présentant des spectres interprétables pour un triplet fixe d'énergies de collision. Mais cela a également tendance à diluer l'information présente sur les spectres. En effet, si un composé présente une énergie de collision idéale forte, l'intensité des fragments observés sera réduite sur le spectre composite car la proportion d'ion parent soumis à une énergie moyenne ou faible ne sera pas fragmentée. Pour continuer à améliorer l'efficacité de ces protocoles, l'idéal serait de pouvoir appliquer différentes énergies de collision individuellement pour chaque ion fragmenté. Cela nécessiterait donc l'acquisition de plusieurs spectres pour un même ion et multiplierait d'autant le temps entre deux acquisitions de profil MS. A l'heure actuelle, ce n'est pas compatible en routine avec la vitesse d'acquisition des instruments de type Orbitrap (dont les vitesses d'acquisition de spectres ne dépassent pas 20 Hz) pour une utilisation en couplage HPLC-MS. Cependant si les instruments continuent d'être améliorés, il serait possible d'envisager des cycles incluant plus de 50 spectres MS² entre deux full scan avec un temps de cycle compatible avec un couplage HPLC tout en préservant la haute résolution indispensable aux approches métabolomiques non ciblées. Tous ces travaux instrumentaux montrent que les progrès réalisés sur les instruments permettent maintenant d'envisager l'acquisition de données plus informatives, plus complètes et plus précises pour les analyses métabolomiques. Parmi les deux approches testées, la DDA et la DIA, chacune présente des avantages et des inconvénients ce qui nous a conduit à recommander leur utilisation différenciée en fonction de l'approche métabolomique. Ainsi pour des approches semi-ciblées dont le but est de détecter et confirmer un certain nombre d'espèces dont les standards sont disponibles, nous privilégierions une acquisition de type DIA qui permet de fragmenter l'ensemble des composés, même ceux ayant une faible intensité. Pour des approches réellement non ciblées, nous privilégierions plutôt une acquisition de type DDA qui favorise l'acquisition de spectres de plus grande pureté ce qui facilite grandement leur interprétation.

Outre cet aspect instrumental, nous avons également cherché à mieux comprendre les mécanismes mis en jeu lors de la fragmentation en spectrométrie de masse. Cette meilleure compréhension est indispensable pour alimenter de façon fiable les algorithmes utilisés pour réaliser de la fragmentation *in silico* ou pour déterminer des structures *de novo* à partir des spectres de fragmentation. Bien que nous n'ayons travaillé que sur deux molécules, nos travaux permettent de mettre en évidence les différents outils pouvant être utilisés pour rationaliser les mécanismes de fragmentation.

Tout d'abord, il existe une vraie complémentarité entre les deux types d'activation en dissociation induite par collision. L'activation résonante va spécifiquement conduire à la fragmentation de l'ion parent et de lui seul. Ce mode d'activation permet de mettre en évidence les

fragments issus de mécanismes compétitifs. Il permet d'identifier facilement les ions de première génération. A l'inverse, l'activation non résonante conduit à une fragmentation simultanée de tous les ions présents dans le piège. Les ions parents sont évidemment fragmentés mais les ions fils issus de ces ions parents sont également fragmentés. Cela conduit à des spectres de fragmentation plus complexes et permet de mettre en évidence les fragmentations consécutives. La prise en compte de ces deux types de mécanismes est un atout important pour la compréhension des règles de fragmentation, d'autant plus que la réalisation de courbes ERMS (*Energy Resolved Mass Spectrometry*) montre que les deux types de mécanismes peuvent être impliqués dans des proportions différentes selon l'énergie utilisée.

L'autre point important est l'utilisation de marquages isotopiques, qu'il s'agisse de marquages stables ou labiles. Les marquages stables (^{13}C , ^{15}N , ^2H non lié à un hétéroatome, ...) sont réalisés par synthèse chimique. Ils permettent de savoir quelle partie de la molécule est perdue lors de la fragmentation et où a lieu la coupure. Dans notre exemple, ce type d'approche a permis de déterminer de quel groupement $-\text{COOH}$ était originaire la perte de CO_2 . Les marquages labiles ou échanges H/D se font à partir d'une molécule non marquée. Cette molécule est diluée dans de l'eau lourde (D_2O). Des échanges vont avoir lieu entre les atomes d'hydrogène liés aux hétéroatomes de la molécule et les atomes de deutérium de l'eau lourde. Un état d'équilibre sera atteint et, du fait de la grande prédominance des molécules d'eau lourde, la molécule d'intérêt sera très majoritairement présente sous la forme ayant échangé tous ses H labiles contre des deutériums. L'utilisation de ces échanges permet de dénombrer les atomes d'hydrogène labiles. Elle peut également mettre en évidence soit des rétro-échanges pouvant avoir lieu au sein de la cellule de collision du spectromètre de masse qui signent une solvatation de la molécule par de l'eau résiduelle, soit des transferts de protons intramoléculaires. Toutes ces informations sont des éléments précieux pour la compréhension des mécanismes de fragmentation. Selon le type de molécule étudié, la réalisation de ces expériences peut s'avérer précieuse. La combinaison de toutes ces expériences avec du calcul quantique permet d'approfondir encore plus ces mécanismes. Si les calculs arrivent à être paramétrés de manière à reproduire les résultats expérimentaux observés, ils donnent accès aux chemins réactionnels mis en jeu. L'obtention de ces chemins et la proposition d'états de transition donnent une vue encore plus fine sur les règles régissant la fragmentation.

C'est grâce à la combinaison de tous ces outils et à leur utilisation à une bien plus grande échelle que l'on pourra efficacement implémenter des règles fiables permettant aux logiciels de fragmentation de fournir des résultats correspondant aux réalités expérimentales. En effet, si les logiciels de fragmentation *in silico* commencent à voir le jour, ils ne sont pour l'instant construits qu'à partir de règles simples et incomplètes. Bien que souvent informatifs, les résultats actuels ne correspondent pas réellement aux spectres expérimentalement observés. A l'inverse, l'approche plus « scientifique » utilisée ici permet de présenter des résultats bien plus proches de la réalité.

Cependant il s'agit d'un travail de très longue haleine qui nécessitera de très gros moyens humains et ne pourra se faire qu'au travers de consortia regroupant un nombre important d'équipes de recherche. Le développement des « intelligences artificielles » et du *machine learning* devrait permettre de faire converger ces deux approches. L'interprétation par l'homme d'un nombre réduit de spectres et la définition de premières règles devrait permettre ensuite à la machine de les valider rapidement à une échelle bien plus importante pour ensuite les injecter dans les logiciels de fragmentation *in silico*. Une fois en place, les règles initiales continueront à être raffinées automatiquement au fur et à mesure que de nouveaux cas seront présentés. Cette combinaison du savoir humain et de la puissance de la machine devrait permettre d'avoir, à terme, des outils de prédiction de fragmentations bien plus performants.

CONCLUSION GENERALE

Tous ces différents travaux montrent l'étendue des possibilités offertes par la métabolomique non ciblée, à la fois à l'heure actuelle mais aussi dans un futur proche. Dans le cadre d'études de cohortes médicales, la métabolomique a un réel potentiel pour apporter des informations utiles dans la compréhension des mécanismes moléculaires mis en jeu dans des maladies encore mal comprises. De plus, elle peut être combinée à d'autres approches « -omiques » phénotypiques pour permettre d'étudier en profondeur ces maladies et proposer des biomarqueurs qui pourront servir à envisager de nouvelles cibles thérapeutiques et à les valider dans le contexte de modèles cellulaires ou animaux mais également lors d'études cliniques. Cela est vrai avec les approches actuelles mais les progrès technologiques permettent d'envisager des protocoles analytiques et de traitement de données encore plus performants. Ces nouvelles approches, comme l'acquisition multi évènements ou le travail en couplage HPLC-spectrométrie de masse à très haute résolution doivent permettre d'augmenter encore le nombre de métabolites identifiés ainsi que la fiabilité de ces annotations. Au cours de ces travaux de thèse, nous avons développé de nouvelles approches permises par l'amélioration de la performance des instruments. Nous avons démontré leurs intérêts réels et la plus-value qu'elles peuvent avoir dans un contexte de métabolomique. Leur utilisation n'est pas encore faite en routine mais leur mise en place devrait prochainement s'imposer dans le monde de la métabolomique. A l'avenir, elles seront probablement associées à des approches de *data-streaming* qui permettent de traiter en temps réel les données et d'influencer l'acquisition des données en fonction des premiers résultats. Elles pourront alors être plus efficacement utilisées dans le cadre de l'analyse de cohortes médicales. Cela facilitera l'utilisation d'analyses métabolomiques et permettra d'accroître l'impact de ces techniques « -omiques » phénotypiques, notamment dans le domaine de la santé. Afin d'augmenter encore l'apport de ces données, il sera nécessaire de toujours plus fiabiliser les données acquises et de standardiser leur mise en forme. Cette uniformisation facilitera grandement leur intégration avec d'autres types de données dans le cadre de projets de médecine translationnels. A l'avenir, il s'agira d'un passage obligatoire pour permettre de nouveaux grands progrès médicaux basés sur la découverte de biomarqueurs diverses et complémentaires. Pour parvenir à un partage efficace de données et à l'utilisation de données métabolomiques issues de plusieurs études, conjointement avec des données d'autres sciences « -omiques », la quantification absolue des métabolites à grande échelle apparaît comme un des futurs grands challenges à relever au sein de la communauté.

REFERENCES BIBLIOGRAPHIQUES

- Abergel, A., Leurs, I., Hagege, H., Benamouzig, R., Bureau, C., Blein, C., Amaz, C., Ribot-Mariotte, E., 2016. Le fardeau des hospitalisations pour l'encéphalopathie hépatique en France : étude des données françaises du programme de médicalisation des systèmes d'information (PMSI). *Rev. d'Épidémiologie Santé Publique* 64, S300. <https://doi.org/10.1016/j.respe.2016.10.029>
- Alonso, A., Marsal, S., Julià, A., 2015. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* 3, 1–20. <https://doi.org/10.3389/fbioe.2015.00023>
- Amos, R.I.J., Haddad, P.R., Szucs, R., Dolan, J.W., Pohl, C.A., 2018. Molecular modeling and prediction accuracy in Quantitative Structure-Retention Relationship calculations for chromatography. *Trends Anal. Chem.* 105, 352–359. <https://doi.org/10.1016/j.trac.2018.05.019>
- Anderson, J.L., Armstrong, D.W., 2005. Immobilized Ionic Liquids as High-Selectivity/High-Temperature/High-Stability Gas Chromatography Stationary Phases. *Anal. Chem.* 77, 6453–6462. <https://doi.org/10.1021/ac051006f>
- Anderson, N.L., Anderson, N.G., 1998. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19, 1853–1861. <https://doi.org/10.1002/elps.1150191103>
- Andries, J.P.M., Claessens, H.A., Heyden, Y.V., Buydens, L.M.C., 2009. Strategy for reduced calibration sets to develop quantitative structure–retention relationships in high-performance liquid chromatography. *Anal. Chim. Acta* 652, 180–188. <https://doi.org/10.1016/j.aca.2009.06.019>
- Annesley, T.M., 2003. Ion Suppression in Mass Spectrometry. *Clin. Chem.* 49, 1041–1044. <https://doi.org/10.1373/49.7.1041>
- Ardenkjær-Larsen, J.H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M.H., Servin, R., Thaning, M., Golman, K., 2003. Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10158–10163. <https://doi.org/10.1073/pnas.1733835100>
- Arlt, K., Brandt, S., Kehr, J., 2001. Amino acid analysis in five pooled single plant cell samples using capillary electrophoresis coupled to laser-induced fluorescence detection. *J. Chromatogr. A* 926, 319–325. [https://doi.org/10.1016/S0021-9673\(01\)01052-4](https://doi.org/10.1016/S0021-9673(01)01052-4)
- Arnhard, K., Gottschall, A., Pitterl, F., Oberacher, H., 2015. Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.* 407, 405–414. <https://doi.org/10.1007/s00216-014-8262-1>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2016. General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Anal. Chim. Acta* 935, 173–186. <https://doi.org/10.1016/j.aca.2016.06.030>
- Baggett, B.R., Cooper, J.D., Hogan, E.T., Carper, J., Paiva, N.L., Smith, J.T., 2002. Profiling isoflavonoids found in legume root extracts using capillary electrophoresis. *Electrophoresis* 23, 1642–1651. [https://doi.org/10.1002/1522-2683\(200206\)23:11<1642::AID-ELPS1642>3.0.CO;2-Y](https://doi.org/10.1002/1522-2683(200206)23:11<1642::AID-ELPS1642>3.0.CO;2-Y)
- Bales, J.R., Higham, D.P., Howe, I., Nicholson, J.K., Sadler, P.J., 1984. Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. *Clin. Chem.* 30, 426–432.

- Bamba, T., Lee, J.W., Matsubara, A., Fukusaki, E., 2012. Metabolic profiling of lipids by supercritical fluid chromatography/mass spectrometry. *J. Chromatogr. A* 1250, 212–219. <https://doi.org/10.1016/j.chroma.2012.05.068>
- Barbier Saint Hilaire, P., 2012. Etude expérimentale du métabolisme du coumatétralyl chez le rat. Implications dans les mécanismes de résistance aux anticoagulants (Thèse). Université Claude Bernard Lyon 1.
- Barri, T., Dragsted, L.O., 2013. UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: Effect of experimental artefacts and anticoagulant. *Anal. Chim. Acta* 768, 118–128. <https://doi.org/10.1016/j.aca.2013.01.015>
- Beckmann, M., Parker, D., Enot, D.P., Duval, E., Draper, J., 2008. High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat. Protoc.* 3, 486–504. <https://doi.org/10.1038/nprot.2007.500>
- Bennaceur, C., Afonso, C., Alves, S., Bossée, A., Tabet, J.-C., 2013. Instrumental Dependent Dissociations of n-Propyl/Isopropyl Phosphonate Isomers: Evaluation of Resonant and Non-Resonant Vibrational Activations. *J. Am. Soc. Mass Spectrom.* 24, 1260–1270. <https://doi.org/10.1007/s13361-013-0656-3>
- Benton, H.P., Ivanisevic, J., Mahieu, N.G., Kurczyk, M.E., Johnson, C.H., Franco, L., Rinehart, D., Valentine, E., Gowda, H., Ubhi, B.K., Tautenhahn, R., Gieschen, A., Fields, M.W., Patti, G.J., Siuzdak, G., 2015. Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling. *Anal. Chem.* 87, 884–891. <https://doi.org/10.1021/ac5025649>
- Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S., Turano, P., 2011. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J. Biomol. NMR* 49, 231–243. <https://doi.org/10.1007/s10858-011-9489-1>
- Bingol, K., Brüsweiler, R., 2015a. Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics. *Curr. Opin. Clin. Nutr. Metab. Care* 18, 471–477. <https://doi.org/10.1097/MCO.0000000000000206>
- Bingol, K., Brüsweiler, R., 2015b. NMR/MS Translator for the Enhanced Simultaneous Analysis of Metabolomics Mixtures by NMR Spectroscopy and Mass Spectrometry: Application to Human Urine. *J. Proteome Res.* 14, 2642–2648. <https://doi.org/10.1021/acs.jproteome.5b00184>
- Bingol, K., Brusweiler-Li, L., Li, D.-W., Brüsweiler, R., 2014. Customized Metabolomics Database for the Analysis of NMR 1H–1H TOCSY and 13C–1H HSQC-TOCSY Spectra of Complex Mixtures. *Anal. Chem.* 86, 5494–5501. <https://doi.org/10.1021/ac500979g>
- Bingol, K., Brusweiler-Li, L., Yu, C., Somogyi, A., Zhang, F., Brüsweiler, R., 2015a. Metabolomics Beyond Spectroscopic Databases: A Combined MS/NMR Strategy for the Rapid Identification of New Metabolites in Complex Mixtures. *Anal. Chem.* 87, 3864–3870. <https://doi.org/10.1021/ac504633z>
- Bingol, K., Li, D.-W., Brusweiler-Li, L., Cabrera, O.A., Megraw, T., Zhang, F., Brüsweiler, R., 2015b. Unified and Isomer-Specific NMR Metabolomics Database for the Accurate Analysis of 13C–1H HSQC Spectra. *ACS Chem. Biol.* 10, 452–459. <https://doi.org/10.1021/cb5006382>
- Blej, A.T., Córdoba, J., 2001. Hepatic encephalopathy. *Am. J. Gastroenterol.* 96, 1968–1976. <https://doi.org/10.1111/j.1572-0241.2001.03964.x>
- Bligh, E.G., Dyer, W.J., 1959. A Rapid Method of Total Lipid Extraction and Purification. *Can. J. Biochem. Physiol.* 37, 911–917. <https://doi.org/10.1139/o59-099>
- Boiteau, R.M., Hoyt, D.W., Nicora, C.D., Kinmonth-Schultz, H.A., Ward, J.K., Bingol, K., 2018. Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR and MS/MS Prediction. *Metabolites* 8, 8. <https://doi.org/10.3390/metabo8010008>

- Boudah, S., Olivier, M.-F., Aros-Calt, S., Oliveira, L., Fenaille, F., Tabet, J.-C., Junot, C., 2014. Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry. *J. Chromatogr. B* 966, 34–47. <https://doi.org/10.1016/j.jchromb.2014.04.025>
- Brown, P.O., Botstein, D., 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21, 33–37. <https://doi.org/10.1038/4462>
- 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. <https://doi.org/10.1021/ac8024569>
- Bruderer, T., Varesio, E., Hopfgartner, G., 2017. The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. *J. Chromatogr. B* 1071, 3–10. <https://doi.org/10.1016/j.jchromb.2017.07.016>
- Bylesjo, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., Trygg, J., 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* <https://doi.org/10.1002/cem.1006>
- Cao, D., Li, W., Zhao, X., Ye, X., Sun, F., Li, J., Song, F., Fan, G., 2014. Development and validation of a rapid and high-sensitivity liquid chromatography–tandem mass spectrometry assay for the determination of neostigmine in small-volume beagle dog plasma and its application to a pharmacokinetic study. *Biomed. Chromatogr.* 28, 354–361. <https://doi.org/10.1002/bmc.3028>
- Cao, M., Fraser, K., Huege, J., Featonby, T., Rasmussen, S., Jones, C., 2015. Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. *Metabolomics* 11, 696–706. <https://doi.org/10.1007/s11306-014-0727-x>
- Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P.D., 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742–753. <https://doi.org/10.1093/nar/gkr1014>
- Catherman, A.D., Skinner, O.S., Kelleher, N.L., 2014. Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.* 445, 683–693. <https://doi.org/10.1016/j.bbrc.2014.02.041>
- Cavazzana-Calvo, M., Debais, D., 2011. *Les biomédicaments*, PUF. ed, Que sais-je ? PUF.
- Chang, M.S., Ji, Q., Zhang, J., El-Shourbagy, T.A., 2007. Historical review of sample preparation for chromatographic bioanalysis: pros and cons. *Drug Dev. Res.* 68, 107–133. <https://doi.org/10.1002/ddr.20173>
- Chekmeneva, E., dos Santos Correia, G., Chan, Q., Wijeyesekera, A., Tin, A., Young, J.H., Elliott, P., Nicholson, J.K., Holmes, E., 2017. Optimization and Application of Direct Infusion Nano-electrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology. *J. Proteome Res.* 16, 1646–1658. <https://doi.org/10.1021/acs.jproteome.6b01003>
- Chen, G., Liao, H., Tseng, Y.J., Tsai, I.-li., Kuo, C., 2015. A matrix-induced ion suppression method to normalize concentration in urinary metabolomics studies using flow injection analysis electrospray ionization mass spectrometry. *Anal. Chim. Acta* 864, 21–29. <https://doi.org/10.1016/j.aca.2015.01.022>

- Chen, H., Pan, Z., Talaty, N., Raftery, D., Cooks, R.G., 2006. Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. *Rapid Commun. Mass Spectrom.* 20, 1577–1584. <https://doi.org/10.1002/rcm.2474>
- Chen, L., Zhou, L., Chan, E.C.Y., Neo, J., Beuerman, R.W., 2011. Characterization of The Human Tear Metabolome by LC–MS/MS. *J. Proteome Res.* 10, 4876–4882. <https://doi.org/10.1021/pr2004874>
- Chial, H., 2008. DNA Sequencing Technologies Key to the Human Genome Project. *Nat. Educ.* 1, 219.
- Ciećko-Michalska, I., Szczepanek, M., Słowik, A., Mach, T., 2012. Pathogenesis of Hepatic Encephalopathy. *Gastroenterol. Res. Pract.* 2012, 642108. <https://doi.org/10.1155/2012/642108>
- Comisarow, M.B., Marshall, A.G., 1974. Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* 25, 282–283. [https://doi.org/10.1016/0009-2614\(74\)89137-2](https://doi.org/10.1016/0009-2614(74)89137-2)
- Commisso, M., Anesi, A., Dal Santo, S., Guzzo, F., 2017. Performance comparison of electrospray ionization and atmospheric pressure chemical ionization in untargeted and targeted liquid chromatography/mass spectrometry based metabolomics analysis of grapeberry metabolites. *Rapid Commun. Mass Spectrom.* 31, 292–300. <https://doi.org/10.1002/rcm.7789>
- Cottret, L., Wildridge, D., Vinson, F., Barrett, M.P., Charles, H., Sagot, M.-F., Jourdan, F., 2010. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* 38, W132–W137. <https://doi.org/10.1093/nar/gkq312>
- Creek, D.J., Dunn, W.B., Fiehn, O., Griffin, J.L., Hall, R.D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E.L., Sumner, L.W., Trengove, R., Wolfender, J.-L., 2014. Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* 10, 350–353. <https://doi.org/10.1007/s11306-014-0656-8>
- Creek, D.J., Jankevics, A., Breitling, R., Watson, D.G., Barrett, M.P., Burgess, K.E.V., 2011. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal. Chem.* 83, 8703–8710. <https://doi.org/10.1021/ac2021823>
- Crockford, D.J., Holmes, E., Lindon, J.C., Plumb, R.S., Zirah, S., Bruce, S.J., Rainville, P., Stumpf, C.L., Nicholson, J.K., 2006. Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies. *Anal. Chem.* 78, 363–371. <https://doi.org/10.1021/ac051444m>
- da Silva, R.R., Dorrestein, P.C., Quinn, R.A., 2015. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12549–12550. <https://doi.org/10.1073/pnas.1516878112>
- Delabrière, A., Hohenester, U.M., Colsch, B., Junot, C., Fenaille, F., Thévenot, E.A., Stegle, O., 2017. proFIA: a data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry. *Bioinformatics* 33, 3767–3775. <https://doi.org/10.1093/bioinformatics/btx458>
- Denery, J.R., Nunes, A.A.K., Dickerson, T.J., 2011. Characterization of Differences between Blood Sample Matrices in Untargeted Metabolomics. *Anal. Chem.* 83, 1040–1047. <https://doi.org/10.1021/ac102806p>
- Devaux, P.G., Horning, M.G., Horning, E.C., 1971. Benzylxime Derivatives of Steroids. A New Metabolic Profile Procedure for Human Urinary Steroids Human Urinary Steroids. *Anal. Lett.* 4, 151–160. <https://doi.org/10.1080/00032717108059686>

- Draper, J., Lloyd, A.J., Goodacre, R., Beckmann, M., 2013. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics* 9, 4–29. <https://doi.org/10.1007/s11306-012-0449-x>
- Dudzic, D., Barbas-Bernardos, C., García, A., Barbas, C., 2017. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J. Pharm. Biomed. Anal.* 147, 149–173. <https://doi.org/10.1016/j.jpba.2017.07.044>
- Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., Nicholls, A.W., Wilson, I.D., Kell, D.B., Goodacre, R., Consortium, T.H.S.M. (HUSERMET), 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 6, 1060–1083. <https://doi.org/10.1038/nprot.2011.335>
- Dunn, W.B., Broadhurst, D., Ellis, D.I., Brown, M., Halsall, A., O'Hagan, S., Spasic, I., Tseng, A., Kell, D.B., 2008. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. *Int. J. Epidemiol.* 37, i23–i30. <https://doi.org/10.1093/ije/dym281>
- Dunn, W.B., Broadhurst, D.I., Edison, A., Guillou, C., Viant, M.R., Bearden, D.W., Beger, R.D., 2017. Quality assurance and quality control processes: summary of a metabolomics community questionnaire. *Metabolomics* 13, 50. <https://doi.org/10.1007/s11306-017-1188-9>
- Dunn, W.B., Erban, A., Weber, R.J.M., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., Viant, M.R., 2013. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 9, 44–66. <https://doi.org/10.1007/s11306-012-0434-4>
- Dunn, W.B., Wilson, I.D., Nicholls, A.W., Broadhurst, D., 2012. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* 4, 2249–2264. <https://doi.org/10.4155/bio.12.204>
- Dwivedi, P., Wu, P., Klopsch, S.J., Puzon, G.J., Xun, L., Hill, H.H., 2008. Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics* 4, 63–80. <https://doi.org/10.1007/s11306-007-0093-z>
- Eckel-Mahan, K.L., Patel, V.R., Mohney, R.P., Vignola, K.S., Baldi, P., Sassone-Corsi, P., 2012. Coordination of the transcriptome and metabolome by the circadian clock. *Proc. Natl. Acad. Sci.* 109, 5541–5546. <https://doi.org/10.1073/pnas.1118726109>
- Edmands, W.M.B., Ferrari, P., Scalbert, A., 2014. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Anal. Chem.* 86, 10925–10931. <https://doi.org/10.1021/ac503190m>
- Eliasson, M., Rännar, S., Madsen, R., Donten, M.A., Marsden-Edwards, E., Moritz, T., Shockcor, J.P., Johansson, E., Trygg, J., 2012. Strategy for Optimizing LC-MS Data Processing in Metabolomics: A Design of Experiments Approach. *Anal. Chem.* 84, 6869–6876. <https://doi.org/10.1021/ac301482k>
- Elwir, S., Rahimi, R.S., 2017. Hepatic Encephalopathy: An Update on the Pathophysiology and Therapeutic Options. *J. Clin. Transl. Hepatol.* 5, 142–151. <https://doi.org/10.14218/JCTH.2016.00069>
- Erve, J.C.L., Gu, M., Wang, Y., DeMaio, W., Talaat, R.E., 2009. Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination. *J. Am. Soc. Mass Spectrom.* 20, 2058–2069. <https://doi.org/10.1016/j.jasms.2009.07.014>

- Fekete, S., Oláh, E., Fekete, J., 2012. Fast liquid chromatography: The domination of core-shell and very fine particles. *J. Chromatogr. A* 1228, 57–71. <https://doi.org/10.1016/j.chroma.2011.09.050>
- Fekete, S., Veuthey, J.-L., Guillarme, D., 2015. Comparison of the most recent chromatographic approaches applied for fast and high resolution separations: Theory and practice. *J. Chromatogr. A* 1408, 1–14. <https://doi.org/10.1016/j.chroma.2015.07.014>
- Ferenci, P., Lockwood, A., Mullen, K., Tarter, R., Weissenborn, K., Blei, A.T., 2002. Hepatic encephalopathy—Definition, nomenclature, diagnosis, and quantification: Final report of the Working Party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* 35, 716–721. <https://doi.org/10.1053/jhep.2002.31250>
- Fernández-Peralbo, M.A., Luque de Castro, M.D., 2012. Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. *Trends Anal. Chem.* 41, 75–85. <https://doi.org/10.1016/j.trac.2012.08.011>
- Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161. <https://doi.org/10.1038/81137>
- Folch, J., Ascoli, I., Lees, M., Meath, J.A., LeBaron, F.N., 1951. Preparation of Lipide Extracts from Brain Tissue. *J. Biol. Chem.* 191, 833–841.
- Forcisi, S., Moritz, F., Kanawati, B., Tziotis, D., Lehmann, R., Schmitt-Kopplin, P., 2013. Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *J. Chromatogr. A, State-of-the art of (UHP)LC–MS(-MS) techniques and their practical application* 1292, 51–65. <https://doi.org/10.1016/j.chroma.2013.04.017>
- Fourier Transform Ion Cyclotron Resonance Mass Spectrometry | Johnson Lab [WWW Document], 2019. . Yale Univ. URL <https://jlab.chem.yale.edu/research/techniques/fourier-transform-ion-cyclotron-resonance-mass-spectrometry> (accessed 3.24.19).
- Gagnebin, Y., Tonoli, D., Lescuyer, P., Ponte, B., de Seigneux, S., Martin, P.-Y., Schappler, J., Boccard, J., Rudaz, S., 2017. Metabolomic analysis of urine samples by UHPLC-QTOF-MS: Impact of normalization strategies. *Anal. Chim. Acta* 955, 27–35. <https://doi.org/10.1016/j.aca.2016.12.029>
- Gao, J., Tarcea, V.G., Karnovsky, A., Mirel, B.R., Weymouth, T.E., Beecher, C.W., Cavalcoli, J.D., Athey, B.D., Omenn, G.S., Burant, C.F., Jagadish, H.V., 2010. Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. *Bioinforma. Oxf. Engl.* 26, 971–973. <https://doi.org/10.1093/bioinformatics/btq048>
- Gartland, K.P., Bonner, F.W., Nicholson, J.K., 1989. Investigations into the biochemical effects of region-specific nephrotoxins. *Mol. Pharmacol.* 35, 242–250.
- Gerlich, M., Neumann, S., 2013. MetFusion: integration of compound identification strategies. *J. Mass Spectrom.* 48, 291–298. <https://doi.org/10.1002/jms.3123>
- Gika, H.G., Theodoridis, G.A., Plumb, R.S., Wilson, I.D., 2014. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* 87, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>
- Gika, H.G., Theodoridis, G.A., Wilson, I.D., 2008. Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry fingerprinting of human urine. *J. Chromatogr. A* 1189, 314–322. <https://doi.org/10.1016/j.chroma.2007.10.066>
- Gika, H.G., Theodoridis, G.A., Wingate, J.E., Wilson, I.D., 2007. Within-Day Reproducibility of an HPLC–MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J. Proteome Res.* 6, 3291–3303. <https://doi.org/10.1021/pr070183p>

- Gika, H.G., Zisi, C., Theodoridis, G., Wilson, I.D., 2016. Protocol for quality control in metabolic profiling of biological fluids by U(H)PLC-MS. *J. Chromatogr. B* 1008, 15–25. <https://doi.org/10.1016/j.jchromb.2015.10.045>
- Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 11, O111.016717. <https://doi.org/10.1074/mcp.O111.016717>
- González-Domínguez, R., 2017. Metabolomic Fingerprinting of Blood Samples by Direct Infusion Mass Spectrometry: Application in Alzheimer's Disease Research. *J. Anal. Test.* 1, 1–9. <https://doi.org/10.1007/s41664-017-0018-4>
- González-Domínguez, R., Sayago, A., Fernández-Recamales, Á., 2016. Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis* 9, 131–148. <https://doi.org/10.4155/bio-2016-0202>
- Gordon, E.F., Muddiman, D.C., 2001. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass spectrometry: experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *J. Mass Spectrom.* 36, 195–203. <https://doi.org/10.1002/jms.121>
- Gowda, G.A.N., Raftery, D., 2015. Can NMR solve some significant challenges in metabolomics? *J. Magn. Reson.* 260, 144–160. <https://doi.org/10.1016/j.jmr.2015.07.014>
- Gowda, H., Ivanisevic, J., Johnson, C.H., Kurczy, M.E., Benton, H.P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevalo, B., Westenskow, P.D., Wang, J., Arkin, A.P., Deutschbauer, A.M., Patti, G.J., Siuzdak, G., 2014. Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. *Anal. Chem.* 86, 6931–6939. <https://doi.org/10.1021/ac500734c>
- Gritti, F., Cavazzini, A., Marchetti, N., Guiochon, G., 2007. Comparison between the efficiencies of columns packed with fully and partially porous C18-bonded silica materials. *J. Chromatogr. A* 1157, 289–303. <https://doi.org/10.1016/j.chroma.2007.05.030>
- Gu, H., Pan, Z., Xi, B., Asiago, V., Musselman, B., Raftery, D., 2011. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer. *Anal. Chim. Acta* 686, 57–63. <https://doi.org/10.1016/j.aca.2010.11.040>
- Guenec, A.L., Giraudeau, P., Caldarelli, S., 2014. Evaluation of Fast 2D NMR for Metabolomics. *Anal. Chem.* 86, 5946–5954. <https://doi.org/10.1021/ac500966e>
- Haggarty, J., Burgess, K.E., 2017. Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome. *Curr. Opin. Biotechnol., Analytical biotechnology* 43, 77–85. <https://doi.org/10.1016/j.copbio.2016.09.006>
- Hagiwara, T., Saito, S., Ujiie, Y., Imai, K., Kakuta, M., Kadota, K., Terada, T., Sumikoshi, K., Shimizu, K., Nishi, T., 2010. HPLC Retention time prediction for metabolome analysis. *Bioinformatics* 5, 255–258.
- Hendrickson, C.L., Quinn, J.P., Kaiser, N.K., Smith, D.F., Blakney, G.T., Chen, T., Marshall, A.G., Weisbrod, C.R., Beu, S.C., 2015. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* 26, 1626–1632. <https://doi.org/10.1007/s13361-015-1182-2>

- Hernandes, V.V., Barbas, C., Dudzik, D., 2017. A review of blood sample handling and pre-processing for metabolomics studies. *Electrophoresis* 0, 1-10. <https://doi.org/10.1002/elps.201700086>
- Holčapek, M., Jirásko, R., Lísa, M., 2010. Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules. *J. Chromatogr. A* 1217, 3908–3921. <https://doi.org/10.1016/j.chroma.2010.02.049>
- Holmes, E., Wilson, I.D., Nicholson, J.K., 2008. Metabolic Phenotyping in Health and Disease. *Cell* 134, 714–717. <https://doi.org/10.1016/j.cell.2008.08.026>
- Hopfgartner, G., Tonoli, D., Varesio, E., 2012. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal. Bioanal. Chem.* 402, 2587–2596. <https://doi.org/10.1007/s00216-011-5641-8>
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, Kenichi, Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, Ken, Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. <https://doi.org/10.1002/jms.1777>
- Horning, E.C., Horning, M.G., 1971. Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites. *Clin. Chem.* 17, 802–809.
- Horning, S.R., Vincenti, M., Cooks, R.G., 1990. Angular dependence of internal energy distributions of activated Fe(CO)₅ and W(CO)₆ ions. The contributions of vibrational and electronic excitation mechanisms in kiloelectron volt collisions. *J. Am. Chem. Soc.* 112, 119–126. <https://doi.org/10.1021/ja00157a019>
- Hu, K., Ellinger, J.J., Chylla, R.A., Markley, J.L., 2011. Measurement of Absolute Concentrations of Individual Compounds in Metabolite Mixtures by Gradient-Selective Time-Zero ¹H–¹³C HSQC with Two Concentration References and Fast Maximum Likelihood Reconstruction Analysis. *Anal. Chem.* 83, 9352–9360. <https://doi.org/10.1021/ac201948f>
- Hufsky, F., Scheubert, K., Böcker, S., 2014. Computational mass spectrometry for small-molecule fragmentation. *Trends Anal. Chem.* 53, 41–48. <https://doi.org/10.1016/j.trac.2013.09.008>
- INSERM, 2012. Dossier d'information - La cirrhose [WWW Document]. Inserm. URL <https://www.inserm.fr/information-en-sante/dossiers-information/cirrhose> (accessed 6.10.18).
- Jalan, R., Lee, W.M., 2009. Treatment of Hyperammonemia in Liver Failure: A Tale of Two Enzymes. *Gastroenterology* 136, 2048–2051. <https://doi.org/10.1053/j.gastro.2009.04.016>
- James, P., 1997. Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* 30, 279–331.
- Jedrychowski, M.P., Huttlin, E.L., Haas, W., Sowa, M.E., Rad, R., Gygi, S.P., 2011. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Mol. Cell. Proteomics* 10. <https://doi.org/10.1074/mcp.M111.009910>
- Jobard, E., Trédan, O., Postoly, D., André, F., Martin, A.-L., Elena-Herrmann, B., Boyault, S., 2016. A Systematic Evaluation of Blood Serum and Plasma Pre-Analytics for Metabolomics Cohort Studies. *Int. J. Mol. Sci.* 17, 2035. <https://doi.org/10.3390/ijms17122035>
- Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E., Stein, L., 2005. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 33, D428-432. <https://doi.org/10.1093/nar/gki072>
- Junot, C., Fenaille, F., Colsch, B., Bécher, F., 2014. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev.* 33, 471–500. <https://doi.org/10.1002/mas.21401>

- Kamlage, B., Maldonado, S.G., Bethan, B., Peter, E., Schmitz, O., Liebenberg, V., Schatz, P., 2014. Quality Markers Addressing Preanalytical Variations of Blood and Plasma Processing Identified by Broad and Targeted Metabolite Profiling. *Clin. Chem.* 60, 399–412. <https://doi.org/10.1373/clinchem.2013.211979>
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Kangas, L.J., Metz, T.O., Isaac, G., Schrom, B.T., Ginovska-Pangovska, B., Wang, L., Tan, L., Lewis, R.R., Miller, J.H., 2012. In silico identification software (ISIS): a machine learning approach to tandem mass spectral identification of lipids. *Bioinformatics* 28, 1705–1713. <https://doi.org/10.1093/bioinformatics/bts194>
- Karni, M., Mandelbaum, A., 1980. The ‘even-electron rule.’ *Org. Mass Spectrom.* 15, 53–64. <https://doi.org/10.1002/oms.1210150202>
- Katajamaa, M., Orešič, M., 2007. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* 1158, 318–328. <https://doi.org/10.1016/j.chroma.2007.04.021>
- Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. *Bioinforma. Oxf. Engl.* 24, 2534–2536. <https://doi.org/10.1093/bioinformatics/btn323>
- Khamis, M.M., Adamko, D.J., El-Aneed, A., 2017. Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom. Rev.* 36, 115–134. <https://doi.org/10.1002/mas.21455>
- Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 8, 105. <https://doi.org/10.1186/1471-2105-8-105>
- Kind, T., 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. <https://doi.org/10.1186/1471-2105-7-234>
- Kind, T., Tsugawa, H., Cajka, T., Ma, Y., Lai, Z., Mehta, S.S., Wohlgemuth, G., Barupal, D.K., Showalter, M.R., Arita, M., Fiehn, O., 2018. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* 37, 513–532. <https://doi.org/10.1002/mas.21535>
- Kingdon, K.H., 1923. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Phys. Rev.* 21, 408–418. <https://doi.org/10.1103/PhysRev.21.408>
- Kirschner, M.W., 2005. The Meaning of Systems Biology. *Cell* 121, 503–504. <https://doi.org/10.1016/j.cell.2005.05.005>
- Koek, M.M., Muilwijk, B., van Stee, L.L.P., Hankemeier, T., 2008. Higher mass loadability in comprehensive two-dimensional gas chromatography–mass spectrometry for improved analytical performance in metabolomics analysis. *J. Chromatogr. A* 1186, 420–429. <https://doi.org/10.1016/j.chroma.2007.11.107>
- Labenz, C., Wörns, M.-A., Schattenberg, J.M., Huber, Y., Galle, P.R., Labenz, J., ALGK study Group, 2017. Epidemiology of hepatic encephalopathy in German hospitals – the EpHE study. *Z. Für Gastroenterol.* 55, 741–747. <https://doi.org/10.1055/s-0043-114671>
- Laboureur, L., Ollero, M., Touboul, D., 2015. Lipidomics by Supercritical Fluid Chromatography. *Int. J. Mol. Sci.* 16, 13868–13884. <https://doi.org/10.3390/ijms160613868>
- Lane, A.N., Fan, T.W.-M., 2017. NMR-based Stable Isotope Resolved Metabolomics in systems biochemistry. *Arch. Biochem. Biophys.* 628, 123–131. <https://doi.org/10.1016/j.abb.2017.02.009>

- Laragh, J.H., 1971. Biochemical Profiling and the Natural History of Hypertensive Diseases: Low-Renin Essential Hypertension, a Benign Condition. *Circulation* 44, 971–974. <https://doi.org/10.1161/01.CIR.44.6.971>
- Lau, S.K.P., Lee, K.-C., Lo, G.C.S., Ding, V.S.Y., Chow, W.-N., Ke, T.Y.H., Curreem, S.O.T., To, K.K.W., Ho, D.T.Y., Sridhar, S., Wong, S.C.Y., Chan, J.F.W., Hung, I.F.N., Sze, K.-H., Lam, C.-W., Yuen, K.-Y., Woo, P.C.Y., 2016. Metabolomic Profiling of Plasma from Melioidosis Patients Using UHPLC-QTOF MS Reveals Novel Biomarkers for Diagnosis. *Int. J. Mol. Sci.* 17, 307. <https://doi.org/10.3390/ijms17030307>
- Lauridsen, M., Hansen, S.H., Jaroszewski, J.W., Cornett, C., 2007. Human Urine as Test Material in 1H NMR-Based Metabonomics: Recommendations for Sample Preparation and Storage. *Anal. Chem.* 79, 1181–1186. <https://doi.org/10.1021/ac061354x>
- Lei, Z., Huhman, D.V., Sumner, L.W., 2011. Mass Spectrometry Strategies in Metabolomics. *J. Biol. Chem.* 286, 25435–25442. <https://doi.org/10.1074/jbc.R111.238691>
- Lesellier, E., 2001. Analysis of non-saponifiable lipids by super-/subcritical-fluid chromatography. *J. Chromatogr. A* 936, 201–214. [https://doi.org/10.1016/S0021-9673\(01\)01204-3](https://doi.org/10.1016/S0021-9673(01)01204-3)
- Levsen, K., Schiebel, H.-M., Terlouw, J.K., Jobst, K.J., Elend, M., Preiss, A., Thiele, H., Ingendoh, A., 2007. Even-electron ions: a systematic study of the neutral species lost in the dissociation of quasi-molecular ions. *J. Mass Spectrom.* 42, 1024–1044. <https://doi.org/10.1002/jms.1234>
- Libiseller, G., Dvorzak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., Neumann, S., Trausinger, G., Sinner, F., Pieber, T., Magnes, C., 2015. IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics* 16, 118. <https://doi.org/10.1186/s12859-015-0562-8>
- Liu, J., Litt, L., Segal, M.R., Kelly, M.J., Yoshihara, H.A., James, T.L., 2011. Outcome-Related Metabolomic Patterns from 1H/31P NMR after Mild Hypothermia Treatments of Oxygen—Glucose Deprivation in a Neonatal Brain Slice Model of Asphyxia. *J. Cereb. Blood Flow Metab.* 31, 547–559. <https://doi.org/10.1038/jcbfm.2010.125>
- Liu, X., Zheng, P., Zhao, X., Zhang, Y., Hu, C., Li, J., Zhao, J., Zhou, J., Xie, P., Xu, G., 2015. Discovery and Validation of Plasma Biomarkers for Major Depressive Disorder Classification Based on Liquid Chromatography–Mass Spectrometry. *J. Proteome Res.* 14, 2322–2330. <https://doi.org/10.1021/acs.jproteome.5b00144>
- Madalinski, G., Godat, E., Alves, S., Lesage, D., Genin, E., Levi, P., Labarre, J., Tabet, J.-C., Ezan, E., Junot, C., 2008. Direct Introduction of Biological Samples into a LTQ-Orbitrap Hybrid Mass Spectrometer as a Tool for Fast Metabolome Analysis. *Anal. Chem.* 80, 3291–3303. <https://doi.org/10.1021/ac7024915>
- Makarov, A., 2000. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* 72, 1156–1162. <https://doi.org/10.1021/ac991131p>
- Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S., 2006. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.* 78, 2113–2120. <https://doi.org/10.1021/ac0518811>
- Makarov, A., Denisov, E., Lange, O., 2009. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* 20, 1391–1396. <https://doi.org/10.1016/j.jasms.2009.01.005>
- Makarov, A., Muenster, H., 2006. Breakthrough Technology for Bottom-Up Proteomics and Small Molecule Research [WWW Document]. *Am. Lab.* URL <http://americanlaboratory.com/914-Application-Notes/18815-Breakthrough-Technology-for-Bottom-Up-Proteomics-and-Small-Molecule-Research/> (accessed 3.24.19).

- March, R.E., Londry, F.A., Alfred, R.L., Todd, J.F.J., Penman, A.D., Vedel, F., Vedel, M., 1991. Resonance excitation of ions stored in a quadrupole ion trap Part III. Introduction to the field interpolation simulation method. *Int. J. Mass Spectrom. Ion Process.* 110, 159–178. [https://doi.org/10.1016/0168-1176\(91\)80026-J](https://doi.org/10.1016/0168-1176(91)80026-J)
- Markley, J.L., Brüschweiler, R., Edison, A.S., Eghbalnia, H.R., Powers, R., Raftery, D., Wishart, D.S., 2017. The future of NMR-based metabolomics. *Curr. Opin. Biotechnol., Analytical biotechnology* 43, 34–40. <https://doi.org/10.1016/j.copbio.2016.08.001>
- Marshall, D.D., Lei, S., Worley, B., Huang, Y., Garcia-Garcia, A., Franco, R., Dodds, E.D., Powers, R., 2015. Combining DI-ESI-MS and NMR datasets for metabolic profiling. *Metabolomics* 11, 391–402. <https://doi.org/10.1007/s11306-014-0704-4>
- Marshall, D.D., Powers, R., 2017. Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. *Prog. Nucl. Magn. Reson. Spectrosc.* 100, 1–16. <https://doi.org/10.1016/j.pnmrs.2017.01.001>
- Mass Spectrometry Facility | Quadrupole [WWW Document], 2019. . Univ. Bristol. URL <http://www.chm.bris.ac.uk/ms/quadrupole.xhtml> (accessed 3.24.19).
- Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A., Schwudke, D., 2008. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* 49, 1137–1146. <https://doi.org/10.1194/jlr.D700041-JLR200>
- Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M., Makarov, A., 2012. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol. Cell. Proteomics* 11, O111.013698. <https://doi.org/10.1074/mcp.O111.013698>
- Michopoulos, F., Edge, A.M., Theodoridis, G., Wilson, I.D., 2010. Application of turbulent flow chromatography to the metabonomic analysis of human plasma: Comparison with protein precipitation. *J. Sep. Sci.* 33, 1472–1479. <https://doi.org/10.1002/jssc.200900789>
- 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. <https://doi.org/10.1021/pr801045q>
- Miura, D., Tsuji, Y., Takahashi, K., Wariishi, H., Saito, K., 2010. A Strategy for the Determination of the Elemental Composition by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Based on Isotopic Peak Ratios. *Anal. Chem.* 82, 5887–5891. <https://doi.org/10.1021/ac902931x>
- Montenegro-Burke, J.R., Aisporna, A.E., Benton, H.P., Rinehart, D., Fang, M., Huan, T., Warth, B., Forsberg, E., Abe, B.T., Ivanisevic, J., Wolan, D.W., Teyton, L., Lairson, L., Siuzdak, G., 2017. Data Streaming for Metabolomics: Accelerating Data Processing and Analysis from Days to Minutes. *Anal. Chem.* 89, 1254–1259. <https://doi.org/10.1021/acs.analchem.6b03890>
- Monton, M.R.N., Soga, T., 2007. Metabolome analysis by capillary electrophoresis–mass spectrometry. *J. Chromatogr. A* 1168, 237–246. <https://doi.org/10.1016/j.chroma.2007.02.065>
- Mullard, G., Allwood, J.W., Weber, R., Brown, M., Begley, P., Hollywood, K.A., Jones, M., Unwin, R.D., Bishop, P.N., Cooper, G.J.S., Dunn, W.B., 2015. A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications. *Metabolomics* 11, 1068–1080. <https://doi.org/10.1007/s11306-014-0763-6>

- Myers, O.D., Sumner, S.J., Li, S., Barnes, S., Du, X., 2017. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* 89, 8689–8695. <https://doi.org/10.1021/acs.analchem.7b01069>
- Mylonas, R., Mauron, Y., Masselot, A., Binz, P.-A., Budin, N., Fathi, M., Viette, V., Hochstrasser, D.F., Lisacek, F., 2009. X-Rank: A Robust Algorithm for Small Molecule Identification Using Tandem Mass Spectrometry. *Anal. Chem.* 81, 7604–7610. <https://doi.org/10.1021/ac900954d>
- Neumann, S., Thum, A., Böttcher, C., 2013. Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. *Metabolomics* 9, 84–91. <https://doi.org/10.1007/s11306-012-0401-0>
- 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. <https://doi.org/10.1080/004982599238047>
- Nicholson, J.K., O’Flynn, M.P., Sadler, P.J., Macleod, A.F., Juul, S.M., Sönksen, P.H., 1984. Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochem. J.* 217, 365–375.
- Nicholson, J.K., Wilson, I.D., 1989. High resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* 21, 449–501. [https://doi.org/10.1016/0079-6565\(89\)80008-1](https://doi.org/10.1016/0079-6565(89)80008-1)
- Niessen, W.M.A., 2011. Fragmentation of toxicologically relevant drugs in positive-ion liquid chromatography-tandem mass spectrometry. *Mass Spectrom. Rev.* 30, 626–663. <https://doi.org/10.1002/mas.20332>
- Nordström, A., Want, E., Northen, T., Lehtiö, J., Siuzdak, G., 2008. Multiple Ionization Mass Spectrometry Strategy Used To Reveal the Complexity of Metabolomics. *Anal. Chem.* 80, 421–429. <https://doi.org/10.1021/ac701982e>
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27, 29–34. <https://doi.org/10.1093/nar/27.1.29>
- Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F., 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 373–378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1)
- Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M., 2007. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 4, 709–712. <https://doi.org/10.1038/nmeth1060>
- Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer [WWW Document], 2018. URL <https://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFADBMBHQ> (accessed 10.23.18).
- Pang, T., 2002. The Impact of Genomics on Global Health. *Am. J. Public Health* 92, 1077–1079. <https://doi.org/10.2105/AJPH.92.7.1077>
- Patterson, S.D., Aebersold, R.H., 2003. Proteomics: the first decade and beyond. *Nat. Genet.* 33, 311–323. <https://doi.org/10.1038/ng1106>
- Pinto, J., Domingues, M.R.M., Galhano, E., Pita, C., Almeida, M. do C., Carreira, I.M., Gil, A.M., 2014. Human plasma stability during handling and storage: impact on NMR metabolomics. *Analyst* 139, 1168–1177. <https://doi.org/10.1039/C3AN02188B>
- Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., Nicholson, J.K., 2006. UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* 20, 1989–1994. <https://doi.org/10.1002/rcm.2550>

- Plumb, R.S., Stumpf, C.L., Gorenstein, M.V., Castro-Perez, J.M., Dear, G.J., Anthony, M., Sweatman, B.C., Connor, S.C., Haselden, J.N., 2002. Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. *Rapid Commun. Mass Spectrom.* 16, 1991–1996. <https://doi.org/10.1002/rcm.813>
- Potters, G., 2010. Systems Biology of the Cell. *Nat. Educ.* 3, 33.
- Rafiei, A., Sleno, L., 2015. Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. *Rapid Commun. Mass Spectrom.* 29, 119–127. <https://doi.org/10.1002/rcm.7094>
- Ramaswamy, V., Hooker, J.W., Withers, R.S., Nast, R.E., Brey, W.W., Edison, A.S., 2013. Development of a ¹³C-Optimized 1.5-mm High Temperature Superconducting NMR Probe. *J. Magn. Reson.* 235, 58–65. <https://doi.org/10.1016/j.jmr.2013.07.012>
- Ramautar, R., Somsen, G.W., de Jong, G.J., 2009. CE-MS in metabolomics. *Electrophoresis* 30, 276–291. <https://doi.org/10.1002/elps.200800512>
- Raterink, R.-J., Lindenburg, P.W., Vreeken, R.J., Ramautar, R., Hankemeier, T., 2014. Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics. *Trends Anal. Chem.* 61, 157–167. <https://doi.org/10.1016/j.trac.2014.06.003>
- Rico, E., González, O., Blanco, M.E., Alonso, R.M., 2014. Evaluation of human plasma sample preparation protocols for untargeted metabolic profiles analyzed by UHPLC-ESI-TOF-MS. *Anal. Bioanal. Chem.* 406, 7641–7652. <https://doi.org/10.1007/s00216-014-8212-y>
- Rinehart, D., Johnson, C.H., Nguyen, T., Ivanisevic, J., Benton, H.P., Lloyd, J., Arkin, A.P., Deutschbauer, A.M., Patti, G.J., Siuzdak, G., 2014. Metabolomic data streaming for biology-dependent data acquisition. *Nat. Biotechnol.* 32, 524–527. <https://doi.org/10.1038/nbt.2927>
- Rochat, B., 2017. Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data. *J. Am. Soc. Mass Spectrom.* 28, 709–723. <https://doi.org/10.1007/s13361-016-1556-0>
- Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H., Tsai, P., 2008. The perils of plenty: what are we going to do with all these genes? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 3893–3902. <https://doi.org/10.1098/rstb.2008.0173>
- Romero-Gómez, M., Montagnese, S., Jalan, R., 2015. Hepatic encephalopathy in patients with acute decompensation of cirrhosis and acute-on-chronic liver failure. *J. Hepatol.* 62, 437–447. <https://doi.org/10.1016/j.jhep.2014.09.005>
- Römisch-Margl, W., Prehn, C., Bogumil, R., Röhring, C., Suhre, K., Adamski, J., 2012. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8, 133–142. <https://doi.org/10.1007/s11306-011-0293-4>
- Rosenling, T., Stoop, M.P., Smolinska, A., Muilwijk, B., Coulier, L., Shi, S., Dane, A., Christin, C., Suits, F., Horvatovich, P.L., Wijmenga, S.S., Buydens, L.M.C., Vreeken, R., Hankemeier, T., Gool, A.J. van, Luider, T.M., Bischoff, R., 2011. The Impact of Delayed Storage on the Measured Proteome and Metabolome of Human Cerebrospinal Fluid. *Clin. Chem.* 57, 1703–1711. <https://doi.org/10.1373/clinchem.2011.167601>
- Roux, A., Thévenot, E.A., Seguin, F., Olivier, M.-F., Junot, C., 2015. Impact of collection conditions on the metabolite content of human urine samples as analyzed by liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy. *Metabolomics* 11, 1095–1105. <https://doi.org/10.1007/s11306-014-0764-5>
- Sana, T.R., Waddell, K., Fischer, S.M., 2008. A sample extraction and chromatographic strategy for increasing LC/MS detection coverage of the erythrocyte metabolome. *J. Chromatogr. B* 871, 314–321. <https://doi.org/10.1016/j.jchromb.2008.04.030>

- Sangster, T., Major, H., Plumb, R., Wilson, A.J., Wilson, I.D., 2006. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst* 131, 1075–1078. <https://doi.org/10.1039/B604498K>
- Sarafian, M.H., Gaudin, M., Lewis, M.R., Martin, F.-P., Holmes, E., Nicholson, J.K., Dumas, M.-E., 2014. Objective Set of Criteria for Optimization of Sample Preparation Procedures for Ultra-High Throughput Untargeted Blood Plasma Lipid Profiling by Ultra Performance Liquid Chromatography–Mass Spectrometry. *Anal. Chem.* 86, 5766–5774. <https://doi.org/10.1021/ac500317c>
- Saude, E.J., Adamko, D., Rowe, B.H., Marrie, T., Sykes, B.D., 2007. Variation of metabolites in normal human urine. *Metabolomics* 3, 439–451. <https://doi.org/10.1007/s11306-007-0091-1>
- Saude, E.J., Sykes, B.D., 2007. Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* 3, 19–27. <https://doi.org/10.1007/s11306-006-0042-2>
- Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L.O., Draper, J., Rappaport, S.M., Hooft, J.J. van der, Wishart, D.S., 2014. The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* 99, 1286–1308. <https://doi.org/10.3945/ajcn.113.076133>
- Schwartz, J.C., Senko, M.W., Syka, J.E.P., 2002. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 13, 659–669. [https://doi.org/10.1016/S1044-0305\(02\)00384-7](https://doi.org/10.1016/S1044-0305(02)00384-7)
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* 48, 2097–2098. <https://doi.org/10.1021/es5002105>
- scimaX MRMS - Taking science to the max - Identification with confidence [WWW Document], 2018. URL <https://www.bruker.com/products/mass-spectrometry-and-separations/mrms/scimax/overview.html> (accessed 10.23.18).
- Sekimoto, K., Sakakura, M., Kawamukai, T., Hike, H., Shiota, T., Usui, F., Bando, Y., Takayama, M., 2014. Ionization characteristics of amino acids in direct analysis in real time mass spectrometry. *Analyst* 139, 2589–2599. <https://doi.org/10.1039/C3AN02193A>
- Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D., Bui, H., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S., Zabrouskov, V., 2013. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* 85, 11710–11714. <https://doi.org/10.1021/ac403115c>
- Señoráns, F.J., Ibañez, E., 2002. Analysis of fatty acids in foods by supercritical fluid chromatography. *Anal. Chim. Acta* 465, 131–144. [https://doi.org/10.1016/S0003-2670\(02\)00208-8](https://doi.org/10.1016/S0003-2670(02)00208-8)
- Silvester, S., Zang, F., 2012. Overcoming non-specific adsorption issues for AZD9164 in human urine samples: Consideration of bioanalytical and metabolite identification procedures. *J. Chromatogr. B* 893, 134–143. <https://doi.org/10.1016/j.jchromb.2012.03.004>
- Simmler, C., Napolitano, J.G., McAlpine, J.B., Chen, S.-N., Pauli, G.F., 2014. Universal quantitative NMR analysis of complex natural samples. *Curr. Opin. Biotechnol., Analytical biotechnology* 25, 51–59. <https://doi.org/10.1016/j.copbio.2013.08.004>
- Simón-Manso, Y., Lowenthal, M.S., Kilpatrick, L.E., Sampson, M.L., Telu, K.H., Rudnick, P.A., Mallard, W.G., Bearden, D.W., Schock, T.B., Tchekhovskoi, D.V., Blonder, N., Yan, X., Liang, Y., Zheng, Y., Wallace, W.E., Neta, P., Phinney, K.W., Remaley, A.T., Stein, S.E., 2013. Metabolite Profiling of a NIST Standard Reference Material for Human Plasma (SRM 1950): GC-MS, LC-MS, NMR, and Clinical Laboratory Analyses, Libraries, and Web-Based Resources. *Anal. Chem.* 85, 11725–11731. <https://doi.org/10.1021/ac402503m>

- Sitnikov, D.G., Monnin, C.S., Vuckovic, D., 2016. Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS. *Sci. Rep.* 6, srep38885. <https://doi.org/10.1038/srep38885>
- Slupsky, C.M., Rankin, K.N., Wagner, J., Fu, H., Chang, D., Weljie, A.M., Saude, E.J., Lix, B., Adamko, D.J., Shah, S., Greiner, R., Sykes, B.D., Marrie, T.J., 2007. Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles. *Anal. Chem.* 79, 6995–7004. <https://doi.org/10.1021/ac0708588>
- Smilde, A.K., Westerhuis, J.A., Jong, S. de, 2003. A framework for sequential multiblock component methods. *J. Chemom.* 17, 323–337. <https://doi.org/10.1002/cem.811>
- Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>
- 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. <https://doi.org/10.1021/ac051437y>
- Smith, R.M., 1999. Supercritical fluids in separation science – the dreams, the reality and the future. *J. Chromatogr. A* 856, 83–115. [https://doi.org/10.1016/S0021-9673\(99\)00617-2](https://doi.org/10.1016/S0021-9673(99)00617-2)
- Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., Nishioka, T., 2002. Simultaneous Determination of Anionic Intermediates for *Bacillus subtilis* Metabolic Pathways by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 74, 2233–2239. <https://doi.org/10.1021/ac020064n>
- Sokolowska, I., Wetie, A.G.N., Woods, A.G., Darie, C.C., 2013. Applications of Mass Spectrometry in Proteomics. *Aust. J. Chem.* 66, 721–733. <https://doi.org/10.1071/CH13137>
- Spectrométrie de masse haute définition SYNAPT G2-Si : Waters [WWW Document], 2018. URL http://www.waters.com/waters/fr_FR/SYNAPT-G2-Si-High-Definition-Mass-Spectrometry/nav.htm?cid=134740622&locale=fr_FR (accessed 10.23.18).
- Stein, S., 2012. Mass Spectral Reference Libraries: An Ever-Expanding Resource for Chemical Identification. *Anal. Chem.* 84, 7274–7282. <https://doi.org/10.1021/ac301205z>
- Steiner, V., Beaugrand, C., Liere, P., Tabet, J.-C., 1999. Influence of trapping parameters on ion injection and dissociation efficiencies in a quadrupole mass filter/ion trap tandem instrument. *J. Mass Spectrom.* 34, 511–520. [https://doi.org/10.1002/\(SICI\)1096-9888\(199905\)34:5<511::AID-JMS802>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1096-9888(199905)34:5<511::AID-JMS802>3.0.CO;2-C)
- Sukemori, S., Ikeda, S., Kurihara, Y., Ito, S., 2003. Amino acid, mineral and vitamin levels in hydrous faeces obtained from coprophagy-prevented rats. *J. Anim. Physiol. Anim. Nutr.* 87, 213–220. <https://doi.org/10.1046/j.1439-0396.2003.00415.x>
- 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., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 3, 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Sumner, L.W., Mendes, P., Dixon, R.A., 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry, Plant Metabolomics* 62, 817–836. [https://doi.org/10.1016/S0031-9422\(02\)00708-2](https://doi.org/10.1016/S0031-9422(02)00708-2)
- Taraji, M., Haddad, P.R., Amos, R.I.J., Talebi, M., Szucs, R., Dolan, J.W., Pohl, C.A., 2017. Error measures in quantitative structure-retention relationships studies. *J. Chromatogr. A* 1524, 298–302. <https://doi.org/10.1016/j.chroma.2017.09.050>

- Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J.K., Bevan, C., Keun, H.C., 2006. Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma. *Anal. Chem.* 78, 4307–4318. <https://doi.org/10.1021/ac051972y>
- Teasdale, G., Jennett, B., 1976. Assessment and prognosis of coma after head injury. *Acta Neurochir. (Wien)* 34, 45–55. <https://doi.org/10.1007/BF01405862>
- Tian, Q., Duncan, C.J.G., Schwartz, S.J., 2003. Atmospheric pressure chemical ionization mass spectrometry and in-source fragmentation of lutein esters. *J. Mass Spectrom.* 38, 990–995. <https://doi.org/10.1002/jms.514>
- Trygg, J., Holmes, E., Lundstedt, T., 2007. Chemometrics in Metabonomics. *J. Proteome Res.* 6, 469–479. <https://doi.org/10.1021/pr060594q>
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., Saito, K., Fiehn, O., Arita, M., 2016. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* 88, 7946–7958. <https://doi.org/10.1021/acs.analchem.6b00770>
- Tweeddale, H., Notley-McRobb, L., Ferenci, T., 1998. Effect of Slow Growth on Metabolism of *Escherichia coli*, as Revealed by Global Metabolite Pool (“Metabolome”) Analysis. *J. Bacteriol.* 180, 5109–5116.
- van de Merbel, N.C., 2008. Quantitative determination of endogenous compounds in biological samples using chromatographic techniques. *Trends Anal. Chem., Advanced MS Analysis of Metabolites and Degradation Products - I* 27, 924–933. <https://doi.org/10.1016/j.trac.2008.09.002>
- Vaught, J.B., 2006. Blood Collection, Shipment, Processing, and Storage. *Cancer Epidemiol. Prev. Biomark.* 15, 1582–1584. <https://doi.org/10.1158/1055-9965.EPI-06-0630>
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B., Kinzler, K.W., 1997. Characterization of the Yeast Transcriptome. *Cell* 88, 243–251. [https://doi.org/10.1016/S0092-8674\(00\)81845-0](https://doi.org/10.1016/S0092-8674(00)81845-0)
- Velenosi, T.J., Hennop, A., Feere, D.A., Tieu, A., Kucey, A.S., Kyriacou, P., McCuaig, L.E., Nevison, S.E., Kerr, M.A., Urquhart, B.L., 2016. Untargeted plasma and tissue metabolomics in rats with chronic kidney disease given AST-120. *Sci. Rep.* 6, srep22526. <https://doi.org/10.1038/srep22526>
- Viant, M.R., Kurland, I.J., Jones, M.R., Dunn, W.B., 2017. How close are we to complete annotation of metabolomes? *Curr. Opin. Chem. Biol.* 36, 64–69. <https://doi.org/10.1016/j.cbpa.2017.01.001>
- Vidal, M., 2009. A unifying view of 21st century systems biology. *FEBS Lett.* 583, 3891–3894. <https://doi.org/10.1016/j.febslet.2009.11.024>
- Vijlder, T.D., Valkenburg, D., Lemièrre, F., Romijn, E.P., Laukens, K., Cuyckens, F., 2018. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* 37, 607–629. <https://doi.org/10.1002/mas.21551>
- Vilstrup, H., Amodio, P., Bajaj, J., Cordoba, J., Ferenci, P., Mullen, K.D., Weissenborn, K., Wong, P., 2014. Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study Of Liver Diseases and the European Association for the Study of the Liver. *Hepatology* 60, 715–735. <https://doi.org/10.1002/hep.27210>
- Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., Yanes, O., Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., Yanes, O., 2012. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* 2, 775–795. <https://doi.org/10.3390/metabo2040775>

- Vinaixa, M., Schymanski, E.L., Neumann, S., Navarro, M., Salek, R.M., Yanes, O., 2016. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *Trends Anal. Chem.* 78, 23–35. <https://doi.org/10.1016/j.trac.2015.09.005>
- Vogl, F.C., Mehrl, S., Heizinger, L., Schlecht, I., Zacharias, H.U., Ellmann, L., Nürnberger, N., Gronwald, W., Leitzmann, M.F., Rossert, J., Eckardt, K.-U., Dettmer, K., Oefner, P.J., GCKD Study Investigators, 2016. Evaluation of dilution and normalization strategies to correct for urinary output in HPLC-HRTOFMS metabolomics. *Anal. Bioanal. Chem.* 408, 8483–8493. <https://doi.org/10.1007/s00216-016-9974-1>
- Vuckovic, D., 2012. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal. Bioanal. Chem.* 403, 1523–1548. <https://doi.org/10.1007/s00216-012-6039-y>
- Wang, C., He, L., Li, D.-W., Bruschiweiler-Li, L., Marshall, A.G., Brüschiweiler, R., 2017. Accurate Identification of Unknown and Known Metabolic Mixture Components by Combining 3D NMR with Fourier Transform Ion Cyclotron Resonance Tandem Mass Spectrometry. *J. Proteome Res.* 16, 3774–3786. <https://doi.org/10.1021/acs.jproteome.7b00457>
- Wang, M., Audi, G., Kondev, F.G., Huang, W.J., Naimi, S., Xu, X., 2017. The AME2016 atomic mass evaluation (II). Tables, graphs and references. *Chin. Phys. C* 41, 030003. <https://doi.org/10.1088/1674-1137/41/3/030003>
- Wang, S., Fu, P., Liu, L., Wang, L., Peng, C., Zhang, W., Liu, R., 2016. Rapid discovery and identification of multiple absorbed chemical constituents and metabolites in rat cerebrospinal fluid after oral administration of Jitai tablets by a LC-MS based metabolomics approach. *RSC Adv.* 6, 32811–32822. <https://doi.org/10.1039/C6RA01382A>
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. <https://doi.org/10.1038/nrg2484>
- Want, E.J., O'Maille, G., Smith, C.A., Brandon, T.R., Uritboonthai, W., Qin, C., Trauger, S.A., Siuzdak, G., 2006. Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Anal. Chem.* 78, 743–752. <https://doi.org/10.1021/ac051312t>
- Warrack, B.M., Hnatyshyn, S., Ott, K.-H., Reily, M.D., Sanders, M., Zhang, H., Drexler, D.M., 2009. Normalization strategies for metabolomic analysis of urine samples. *J. Chromatogr. B* 877, 547–552. <https://doi.org/10.1016/j.jchromb.2009.01.007>
- Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., Humphery-Smith, I., 1995. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16, 1090–1094. <https://doi.org/10.1002/elps.11501601185>
- Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738. <https://doi.org/10.1038/171737a0>
- Weber, R.J.M., Southam, A.D., Sommer, U., Viant, M.R., 2011. Characterization of Isotopic Abundance Measurements in High Resolution FT-ICR and Orbitrap Mass Spectra for Improved Confidence of Metabolite Identification. *Anal. Chem.* 83, 3737–3743. <https://doi.org/10.1021/ac2001803>
- Wedge, D.C., Allwood, J.W., Dunn, W., Vaughan, A.A., Simpson, K., Brown, M., Priest, L., Blackhall, F.H., Whetton, A.D., Dive, C., Goodacre, R., 2011. Is Serum or Plasma More Appropriate for Intersubject Comparisons in Metabolomic Studies? An Assessment in Patients with Small-Cell Lung Cancer. *Anal. Chem.* 83, 6689–6697. <https://doi.org/10.1021/ac2012224>
- Weickhardt, C., Moritz, F., Grottemeyer, J., 1996. Time-of-flight mass spectrometry: State-of-the-art in chemical analysis and molecular science. *Mass Spectrom. Rev.* 15, 139–162. [https://doi.org/10.1002/\(SICI\)1098-2787\(1996\)15:3<139::AID-MAS1>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-2787(1996)15:3<139::AID-MAS1>3.0.CO;2-J)

- Weiss, N., Barbier Saint Hilaire, P., Colsch, B., Isnard, F., Attala, S., Schaefer, A., Amador, M. del M., Rudler, M., Lamari, F., Sedel, F., Thabut, D., Junot, C., 2016. Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy. *J. Hepatol.* 65, 1120–1130. <https://doi.org/10.1016/j.jhep.2016.07.046>
- Weissberg, A., Dagan, S., 2011. Interpretation of ESI(+)-MS-MS spectra—Towards the identification of “unknowns.” *Int. J. Mass Spectrom.* 299, 158–168. <https://doi.org/10.1016/j.ijms.2010.10.024>
- Werner, P., Meiss, E., Scheja, L., Heeren, J., Fischer, M., 2017. Metabolite profiling: development and application of an UHR-QTOF-MS(/MS) method approach for the assessment of metabolic changes in high fat diet fed mice. *Metabolomics* 13, 44. <https://doi.org/10.1007/s11306-017-1181-3>
- Westerhuis, J.A., Kourti, T., MacGregor, J.F., 1998. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* 12, 301–321. [https://doi.org/10.1002/\(SICI\)1099-128X\(199809/10\)12:5<301::AID-CEM515>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S)
- Whiley, L., Godzien, J., Ruperez, F.J., Legido-Quigley, C., Barbas, C., 2012. In-Vial Dual Extraction for Direct LC-MS Analysis of Plasma for Comprehensive and Highly Reproducible Metabolic Fingerprinting. *Anal. Chem.* 84, 5992–5999. <https://doi.org/10.1021/ac300716u>
- Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L., 1996. Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. *Biotechnol. Genet. Eng. Rev.* 13, 19–50. <https://doi.org/10.1080/02648725.1996.10647923>
- Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., Scalbert, A., 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 46, D608–D617. <https://doi.org/10.1093/nar/gkx1089>
- Wishart, D.S., Lewis, M.J., Morrissey, J.A., Flegel, M.D., Jeroncic, K., Xiong, Y., Cheng, D., Eisner, R., Gautam, B., Tzur, D., Sawhney, S., Bamforth, F., Greiner, R., Li, L., 2008. The human cerebrospinal fluid metabolome. *J. Chromatogr. B* 871, 164–173. <https://doi.org/10.1016/j.jchromb.2008.05.001>
- 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., Querengesser, L., 2007. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35, D521–D526. <https://doi.org/10.1093/nar/gkl923>
- Wolf, S., Schmidt, S., Müller-Hannemann, M., Neumann, S., 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* 11, 148. <https://doi.org/10.1186/1471-2105-11-148>
- Wolfer, A.M., Lozano, S., Umbdenstock, T., Croixmarie, V., Arrault, A., Vayer, P., 2015. UPLC–MS retention time prediction: a machine learning approach to metabolite identification in untargeted profiling. *Metabolomics* 12, 8. <https://doi.org/10.1007/s11306-015-0888-2>
- Wolff, M.M., Stephens, W.E., 1953. A Pulsed Mass Spectrometer with Time Dispersion. *Rev. Sci. Instrum.* 24, 616–617. <https://doi.org/10.1063/1.1770801>
- Worley, B., Powers, R., 2013. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* 1, 92–107. <https://doi.org/10.2174/2213235X11301010092>

- Wrona, M., Mauriala, T., Bateman, K.P., Mortishire-Smith, R.J., O'Connor, D., 2005. "All-in-one" analysis for metabolite identification using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry with collision energy switching. *Rapid Commun. Mass Spectrom.* 19, 2597–2602. <https://doi.org/10.1002/rcm.2101>
- Wu, Y., Li, L., 2016. Sample normalization methods in quantitative metabolomics. *J. Chromatogr. A* 1430, 80–95. <https://doi.org/10.1016/j.chroma.2015.12.007>
- Wuolikainen, A., Hedenström, M., Moritz, T., Marklund, S.L., Antti, H., Andersen, P.M., 2009. Optimization of procedures for collecting and storing of CSF for studying the metabolome in ALS. *Amyotroph. Lateral Scler.* 10, 229–236. <https://doi.org/10.1080/17482960902871009>
- Wysocki, V.H., Ross, M.M., Horning, S.S., Cooks, R.G., 1988. Remote-site (charge-remote) fragmentation. *Rapid Commun. Mass Spectrom.* 2, 214–217. <https://doi.org/10.1002/rcm.1290021009>
- Xi, Y., Ropp, J.S. de, Viant, M.R., Woodruff, D.L., Yu, P., 2006. Automated screening for metabolites in complex mixtures using 2D COSY NMR spectroscopy. *Metabolomics* 2, 221–233. <https://doi.org/10.1007/s11306-006-0036-0>
- Xu, Y., Heilier, J.-F., Madalinski, G., Genin, E., Ezan, E., Tabet, J.-C., Junot, C., 2010. Evaluation of Accurate Mass and Relative Isotopic Abundance Measurements in the LTQ-Orbitrap Mass Spectrometer for Further Metabolomics Database Building. *Anal. Chem.* 82, 5490–5501. <https://doi.org/10.1021/ac100271j>
- Yang, Y., Cruickshank, C., Armstrong, M., Mahaffey, S., Reisdorph, R., Reisdorph, N., 2013. New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome. *J. Chromatogr. A* 1300, 217–226. <https://doi.org/10.1016/j.chroma.2013.04.030>
- Yin, P., Lehmann, R., Xu, G., 2015. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal. Bioanal. Chem.* 407, 4879–4892. <https://doi.org/10.1007/s00216-015-8565-x>
- Yin, P., Peter, A., Franken, H., Zhao, X., Neukamm, S.S., Rosenbaum, L., Lucio, M., Zell, A., Häring, H.-U., Xu, G., Lehmann, R., 2013. Preanalytical Aspects and Sample Quality Assessment in Metabolomics Studies of Human Blood. *Clin. Chem.* 59, 833–845. <https://doi.org/10.1373/clinchem.2012.199257>
- Yu, Z., Kastenmüller, G., He, Y., Belcredi, P., Möller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., Polonikov, A., Dahmen, N., Prokisch, H., Xie, L., Li, Y., Wichmann, H.-E., Peters, A., Kronenberg, F., Suhre, K., Adamski, J., Illig, T., Wang-Sattler, R., 2011. Differences between Human Plasma and Serum Metabolite Profiles. *PLOS ONE* 6, e21230. <https://doi.org/10.1371/journal.pone.0021230>
- Zamora-Ros, R., Rabassa, M., Cherubini, A., Urpi-Sarda, M., Llorach, R., Bandinelli, S., Ferrucci, L., Andres-Lacueva, C., 2011. Comparison of 24-h volume and creatinine-corrected total urinary polyphenol as a biomarker of total dietary polyphenols in the Invecchiare InCHIANTI study. *Anal. Chim. Acta* 704, 110–115. <https://doi.org/10.1016/j.aca.2011.07.035>
- Zhang, Y., Bilbao, A., Bruderer, T., Luban, J., Strambio-De-Castillia, C., Lisacek, F., Hopfgartner, G., Varesio, E., 2015. The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* 14, 4359–4371. <https://doi.org/10.1021/acs.jproteome.5b00543>
- Zhou, B., Feng Xiao, J., Tuli, L., W. Resson, H., 2012. LC-MS-based metabolomics. *Mol. Biosyst.* 8, 470–481. <https://doi.org/10.1039/C1MB05350G>

- Zhou, J., Li, Y., Chen, X., Zhong, L., Yin, Y., 2017. Development of data-independent acquisition workflows for metabolomic analysis on a quadrupole-orbitrap platform. *Talanta* 164, 128–136. <https://doi.org/10.1016/j.talanta.2016.11.048>
- Zubarev, R.A., Makarov, A., 2013. Orbitrap Mass Spectrometry. *Anal. Chem.* 85, 5288–5296. <https://doi.org/10.1021/ac4001223>

ANNEXES

Annexe 1 : Les différents types de sources en spectrométrie de masse

Lors d'un couplage LC-MS, les principales sources d'ionisation sont des sources dites à pression atmosphérique. Ce type de source simplifie grandement le couplage direct de la LC au spectromètre de masse. Cependant, cela implique l'existence d'un système de pompage différentiel par étape entre la source et l'analyseur dans lequel la pression est très nettement inférieure. Différentes pompes vont donc devoir être présentes afin d'assurer différents niveaux de vide entre les différentes parties du spectromètre de masse. Afin de maintenir les pressions adéquates dans chacun des compartiments, ces derniers ne sont reliés entre eux que par d'étroites ouvertures. Cela entraîne une déperdition d'ions lors de leur transfert au travers de ce gradient de pression. Cependant, ce type de source a l'avantage de présenter un rendement d'ionisation plus important que ce qui est le cas pour les sources d'ionisation sous vide. Cela est dû à la stabilisation thermique des ions formés et vient compenser le faible taux de transmission des ions. Les principales sources utilisées sont présentées ci-après.

1 L'ionisation chimique à pression atmosphérique

Cette technique d'ionisation est basée sur des réactions ion-molécule en phase gazeuse. Le flux sortant de la LC est introduit dans un nébuliseur pneumatique où, avec l'assistance d'un gaz à haute vitesse, un aérosol se forme. Les gouttelettes ainsi produites sont emportées par le gaz au travers d'une chambre de désolvatation. Dans ce tube chauffé, la phase mobile et l'échantillon se vaporisent sous l'effet de la chaleur et continuent leur chemin dans le flux gazeux jusqu'à la chambre de réaction. A ce niveau, des décharges électroniques (décharge de Corona ou émetteur de particules β^-) vont conduire à la formation d'ions réactifs à partir du gaz de solvant. Ces ions vont réagir à leur tour avec d'autres molécules de solvant pour former des ions secondaires. Les ions de la substance à analyser se formeront par réaction chimiques avec ces ions secondaires. En ionisation en mode positif, des ions protonés de l'espèce moléculaire vont se former par transfert de protons à partir des ions du gaz de solvant. En mode négatif, les ions de l'espèce moléculaire sont obtenus par transfert d'électrons ou par abstraction de protons par le gaz de solvant. Dans ces deux modes, la formation d'adduits entre molécules et ions pseudo-moléculaires est possible.

L'APCI fonctionne particulièrement bien pour les composés polaires et ioniques de faible poids moléculaire (< 1500 Da) même si elle est également capable d'ioniser des composés plus

apolaires. Elle aboutit à une ionisation où prédominent les ions monochargés de l'espèce moléculaire avec pas ou peu de fragmentation en source.

2 La photoionisation à pression atmosphérique

Dans une source APPI, le flux de solvant est vaporisé à l'aide d'un nébuliseur chauffant comparable à celui utilisé pour une source APCI. Après vaporisation, les molécules en phase gazeuse vont interagir avec des photons émis par une lampe à décharge. Ces photons vont induire une série de réactions en phase gazeuse et conduire à l'ionisation des molécules. Le choix du type de lampe à décharge est important. En effet, si l'énergie des photons émis est plus haute que l'énergie d'ionisation des molécules à analyser mais inférieure à celle de l'air ou des solvants, l'ionisation sera sélective des ions d'intérêt ce qui réduit le bruit de fond. Si des phases mobiles eau/acétonitrile sont utilisées, une lampe à argon peut être choisie. Si les phases mobiles contiennent du méthanol, qui possède une énergie d'ionisation plus faible, une lampe à krypton qui fournit des photons moins énergétiques est à privilégier.

L'ionisation en APPI présente une faible efficacité du fait de l'absorption des photons pour les molécules de solvant, qui vont s'exciter sans toutefois s'ioniser. Pour améliorer ce rendement d'ionisation, un dopant, en quantité relativement importante par rapport à l'analyte, peut être utilisé. Ce dopant doit être facilement photoionisable. Les plus utilisés sont l'acétone et le toluène. Ce dopant est particulièrement important lors d'une ionisation en mode négatif. Quel que soit le mode d'ionisation choisi, les ions moléculaires et pseudomoléculaires seront observés.

Cette technique d'ionisation représente une bonne alternative aux ionisations APCI ou ESI (voir ci-après) pour certains composés qui s'ionisent difficilement par ces méthodes. Il s'agit principalement des composés non polaires tels que les flavonoïdes, les stéroïdes ou certains médicaments.

3 L'ionisation par électrospray

Ce mode d'ionisation est aujourd'hui le plus utilisé, notamment pour le couplage LC-MS. Il a été développé dans le milieu des années 1980. Le spray est formé à pression atmosphérique par l'application d'un très fort champ électrique sur le flux de solvant traversant un capillaire métallique. Le champ électrique est obtenu par application d'une différence de potentiel de 2 à 6 kV entre le capillaire et une contre électrode. A l'extrémité du capillaire et sous l'effet de la différence de potentiel, les ions s'accumulent à la surface du liquide et la déforment. Cela conduit à la formation d'un cône de Taylor. Lorsque la pression est suffisante et que la force exercée est supérieure à la tension de surface, l'extrémité de ce cône se rompt pour former des gouttelettes fortement chargées. Ces dernières sont constituées de molécules de solvant et d'analyte. L'évaporation du

solvant conduit à la diminution de la taille de ces gouttelettes ce qui augmente la densité de charge. Passé un certain seuil, les forces de Coulomb, à caractère répulsif, dépassent les forces de cohésion des gouttelettes et provoquent l'explosion de ces dernières. De nouvelles gouttelettes se forment. Le processus d'évaporation et d'explosion en gouttelettes de plus petite taille se déroule en cascade jusqu'au moment où le champ électrique à leur surface devient suffisant pour engendrer la désorption d'agrégats chargés. Ces agrégats sont ensuite accélérés et finissent de se désolvater dans différents canaux de transmission d'ions au sein du spectromètre de masse. Cette technique d'ionisation conduit essentiellement à la formation d'ions pseudomoléculaires avec un faible taux de fragmentation en source et donne donc directement accès à la masse des analytes.

Ce type de source est une source à courant constant. Si plusieurs analytes sont présents simultanément dans la source, le courant sera donc partagé entre eux. Il va donc se mettre en place un système de compétition entre les différents analytes. S'ils sont peu nombreux, tous vont s'ioniser avec un ratio correspondant à celui de leur constante de désorption. S'ils sont plus nombreux, et que le courant maximum constant est atteint, les ions ayant les constantes de désorption les plus élevés vont fortement s'ioniser et ceux ayant de faibles constantes vont très peu pouvoir se former au point de parfois ne plus être visibles sur le spectre de masse. Il s'agit d'un phénomène de suppression d'ionisation qui justifie les étapes de préparation d'échantillon visant à les purifier avant leur analyse.

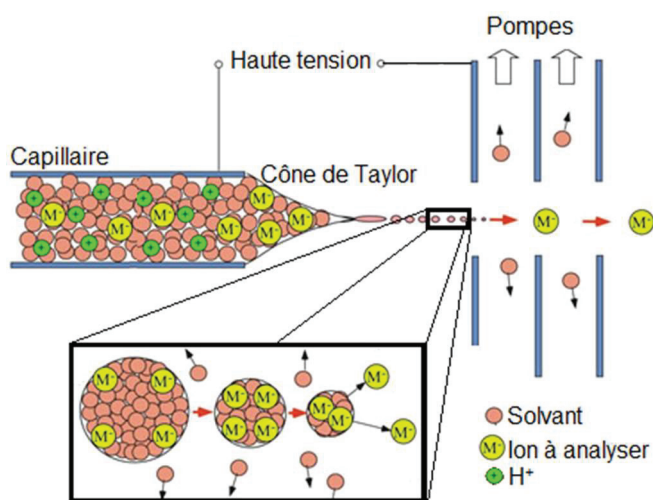


Figure 18 : Formation des ions dans une source électrospray (d'après Barbier Saint Hilaire, 2012)

4 Caractéristiques comparatives des différentes sources

L'ionisation par électrospray est la plus utilisée pour les analyses de métabolomique par LC-MS avec 90% des publications mentionnant l'utilisation de cette source contre 6 % pour l'APCI et 2% pour l'APPI (Forcisi et al., 2013). Chacun de ces types de source permet d'ioniser des molécules

différentes. De manière générale, il est admis que les sources ESI sont idéales pour les composés semi-polaires et polaires alors que les sources APCI fonctionnent mieux pour les composés moins polaires et neutres (Lei et al., 2011). Etant donné la nature majoritairement polaire du métabolome, les sources ESI permettent une meilleure couverture du métabolome (Nordström et al., 2008) même si l'APCI permet d'atteindre une meilleure sensibilité pour certains métabolites plasmatiques (Tian et al., 2003). L'APPI peut être utilisée pour ioniser simultanément les petites molécules polaires et apolaires avec une forte sensibilité et permettre la détection de métabolites plus variés. Cependant, une récente étude a montré que les sources APCI pouvaient aussi permettre d'ioniser des composés polaires (Commisso et al., 2017). Les caractéristiques des différentes sources, eu égard à la polarité des composés, sont résumées dans la Figure 19. L'utilisation simultanée des deux types de source pour les analyses métabolomiques non ciblées permet d'accroître le nombre de signaux détectés dans le sérum de 20% (Nordström et al., 2008) et de 34% dans le métabolome des érythrocytes (Sana et al., 2008). Cette approche peut donc être recommandée pour optimiser la couverture de l'ensemble du métabolome.

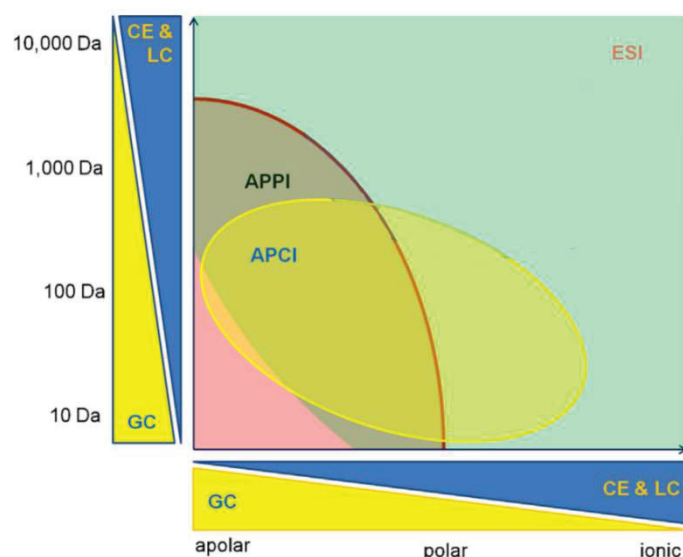


Figure 19 : Diagramme résumant les gammes de masse et de polarité couvertes par les différents types de source (Forcisi et al., 2013).

De plus, les sources ESI sont particulièrement sensibles au phénomène de suppression d'ionisation notamment lors de l'analyse de matrices biologiques. Ce phénomène va masquer les espèces s'ionisant les moins bien lorsque de nombreux analytes entrent simultanément dans la source. Ceci est moins le cas pour l'APCI et encore moins pour l'APPI. Cela confère une plus faible gamme dynamique aux sources ESI. Enfin l'APCI tend à générer plus de fragmentation en source alors que l'ESI génère plus d'ions sous forme d'adduits (Commisso et al., 2017).

Annexe 2 : Les différents types d'analyseurs en spectrométrie de masse

Si les analyses métabolomiques non ciblées sont essentiellement réalisées avec des instruments à haute résolution, certains instruments incluent également des analyseurs à basse résolution, soit pour isoler rapidement certains ions, soit pour favoriser la vitesse d'acquisition ou de fragmentation. Ces différents analyseurs sont présentés ci-après.

1 Analyseurs à basse résolution

1.1 Quadripôle

Un quadripôle est un analyseur de masse qui fonctionne comme un filtre de masse. Il est constitué de quatre électrodes à sections cylindriques ou hyperboliques (Figure 20). Deux potentiels déphasés de 180° , l'un continu, l'autre alternatif sont appliqués aux électrodes, créant un champ quadripolaire. Au sein de ce champ, seuls les ions présentant une trajectoire stable sont conduits jusqu'à l'extrémité de l'analyseur, les autres étant arrêtés lorsque leur trajectoire instable les conduit à entrer en collision avec les électrodes. Le domaine de stabilité des ions au sein du quadripôle dépend de leur rapport m/z et des tensions appliquées aux électrodes permettant, ainsi de sélectionner uniquement les ions appartenant à une gamme de masse définie.

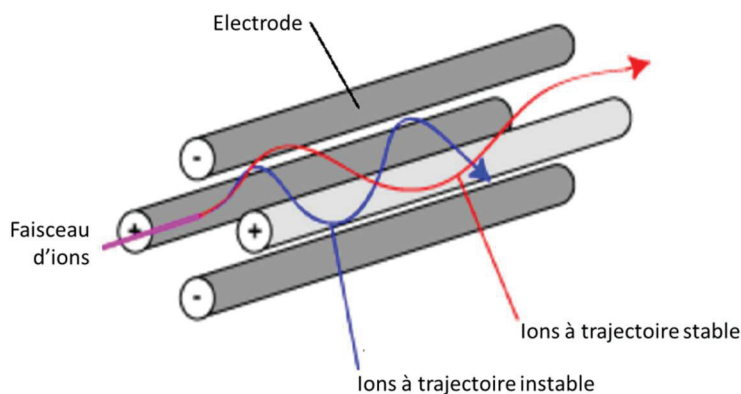


Figure 20 : Principe de fonctionnement d'un analyseur quadripolaire (adapté de "Mass Spectrometry Facility | Quadrupole," 2019)

Dans ce type d'analyseur, l'augmentation de la résolution entraîne une diminution du nombre d'ions atteignant le détecteur. Comme ils ne permettent pas d'atteindre une résolution suffisante pour déterminer une composition élémentaire, ils sont généralement utilisés à un pouvoir résolutif minimal permettant la séparation d'ions différant d'une unité de masse. Avec ce type d'analyseur, la largeur de pic est fixe, la résolution croit donc avec le rapport m/z des ions analysés.

1.2 Piège à ions linéaire

Les pièges à ions linéaires sont des analyseurs constitués de quatre électrodes à section hyperbolique entre lesquelles les ions sont piégés. Pour éviter qu'ils ne sortent du quadripôle, soit deux électrodes d'entrée et de sortie « ferment » les extrémités, soit les électrodes sont segmentées en trois parties permettant l'application de potentiels spécifiques et le piégeage des ions (Figure 21).

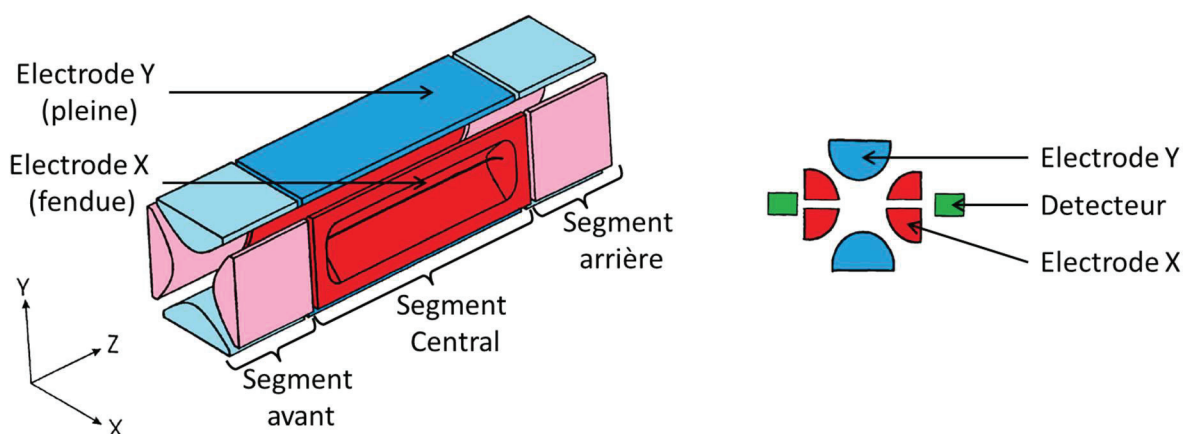


Figure 21 : Principe de fonctionnement d'un piège à ion linéaire segmenté à éjection radiale (adapté de Schwartz et al., 2002)

Les ions sont décélérés par collision avec un gaz inerte présent à faible pression. Ils sont piégés dans un mouvement d'oscillation entre les électrodes centrales. Ils sont ensuite éjectés du piège, par modulation des radiofréquences appliquées, puis détectés. Une gamme de radiofréquence est appliquée au piège afin de balayer toute la gamme de masse étudiée. Deux modes d'éjection sont utilisés en fonction de la conception du piège : l'éjection axiale ou l'éjection radiale. Dans ce dernier cas, l'extraction sélective des ions de différentes masses se fait par application d'une tension alternative entre les deux électrodes fendues. Une fois éjectés par l'une des fentes, les ions sont détectés par l'un des deux détecteurs situés au niveau des électrodes fendues.

Les pièges à ions linéaires ont l'avantage de pouvoir piéger une quantité importante d'ions grâce au volume important disponible et à la focalisation axiale qui a lieu. Ce piège présente une très bonne efficacité de capture (autour de 60 %). Il présente une vitesse de balayage élevée (pouvant atteindre 66 kDa/s). Tous ces éléments confèrent au piège à ions linéaire une gamme dynamique importante qui permet la détection d'espèces minoritaires lors des analyses métabolomiques.

2 Analyseurs à haute résolution

Lors d'une analyse métabolomique non ciblée, il est nécessaire de pouvoir identifier sans *a priori* les molécules détectées. Pour ce faire, l'utilisation d'analyseurs à haute résolution est

essentielle afin d'obtenir les informations les plus précises possibles. La haute résolution permet d'abord de séparer les nombreux composés isobares (composés ayant la même masse nominale mais des masses exactes différentes) présents dans les fluides biologiques complexes. Cela donne un signal spécifique pour chaque ion et non plus un signal composé de plusieurs ions isobares. Grâce à cela, la précision de mesure de masse d'un ion isolé est grandement augmentée (de l'ordre du ppm), ce qui permet d'avoir accès à la formule brute des ions détectés.

2.1 Analyseur à Temps de Vol

La technologie des analyseurs à temps de vol est relativement ancienne (Wolff and Stephens, 1953). Dans les instruments à temps de vol récents, les ions sont introduits orthogonalement au tube de vol et sont accélérés vers le tube par l'application d'une différence de potentiel entre une électrode et une grille d'extraction. Tous les ions de même charge reçoivent alors la même énergie cinétique et se déplacent donc avec une vitesse inversement proportionnelle au carré de leur rapport m/z . Après un parcours en vol libre dans le tube de vol, les ions entrent dans un réflectron qui, par application d'un champ électrique graduel, va inverser la direction des ions (Figure 4). Cela permet de corriger une éventuelle disparité dans les énergies cinétiques des ions et d'augmenter la distance de vol (5 mètres pour le MAXis3G-TOF). Ces deux éléments permettent d'améliorer la résolution (Weickhardt et al., 1996). De plus, le réflectron va permettre de focaliser temporellement les ions au niveau du détecteur. L'impact des ions au niveau du détecteur est enregistré, amplifié et converti en signal.

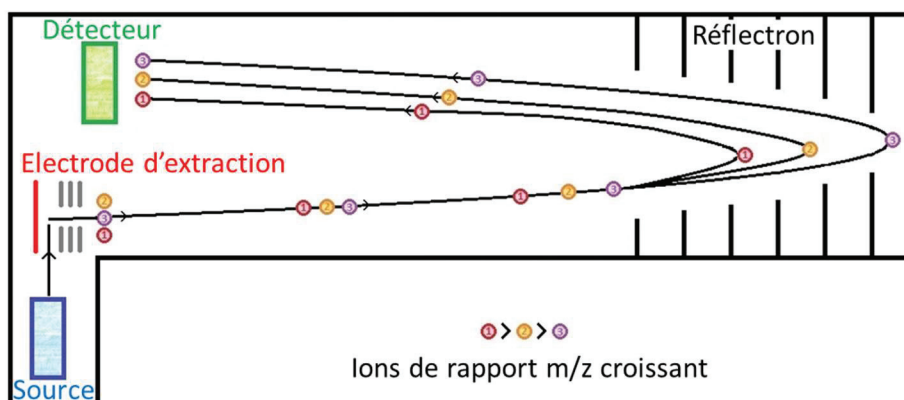


Figure 22 : Principe de fonctionnement d'un analyseur à temps de vol avec réflectron

Les analyseurs à temps de vol présentent une sensibilité supérieure à celle des autres analyseurs à haute résolution et comparable à celle d'un quadripôle. En termes de résolution, cette technologie n'est pas aussi performante que celles des autres instruments à haute résolution mais reste cependant intéressante avec des résolutions pouvant atteindre 100 000 (à m/z 400). Il est intéressant de noter que la résolution des ions croît avec la racine carrée de leur rapport m/z . Le principal avantage de ce type d'analyseur est sa très haute vitesse d'acquisition de données (jusqu'à

100 scans par secondes). Cela lui confère une place importante dans le cadre de protocoles d'acquisition complexes mélangeant différents types d'acquisition. Cette haute vitesse d'acquisition lui permet également aux analyseurs à temps de vol d'être couplés à de nouvelles techniques séparatives en ligne comme la mobilité ionique (Dwivedi et al., 2008). Leurs inconvénients majeurs sont une gamme dynamique plus limitée, liée à l'utilisation de digitaliseurs rapides, et une moindre stabilité temporelle de la calibration (Forcisi et al., 2013) qui implique l'utilisation d'une calibration interne. Ces instruments sont fréquemment utilisés pour les analyses métabolomiques (Lau et al., 2016; Velenosi et al., 2016; Werner et al., 2017).

2.2 Analyseur à Résonance Cyclonique Ionique (FT-ICR)

Les analyseurs à résonance cyclonique ionique se basent sur l'utilisation de la transformée de Fourier. L'application de cette transformée de Fourier à la spectrométrie de masse a été décrite pour la première fois en 1974 (Comisarow and Marshall, 1974). Ces analyseurs fonctionnent par un système de piégeage d'ion et d'enregistrement du courant induit par les ions en mouvement dans le piège.

Dans un analyseur à résonance cyclonique ionique, les ions sont piégés par un champ magnétique au sein d'une cellule constituée de deux paires de plaques (plaques émettrices et plaques réceptrices). Ces plaques sont situées dans un plan parallèle au champ magnétique (Figure 23). Soumis à ce champ magnétique, les ions vont adopter une trajectoire circulaire (principe du cyclotron). Ils sont introduits dans l'axe du champ et vont être maintenus dans le piège grâce à l'application d'un potentiel répulsif sur deux plaques supplémentaires (plaques de piégeage). Ces plaques sont nécessaires car le champ magnétique ne permet pas de bloquer le mouvement des ions selon son propre axe. Les ions prennent alors un mouvement circulaire dans un plan perpendiculaire à l'axe du champ magnétique. Les ions de même rapport m/z ont alors la même fréquence cyclotronique mais vont se mouvoir indépendamment dans le centre du piège. L'application d'une radiofréquence par les plaques émettrices conduit les ions résonants à entrer en phase avec le pulse d'excitation et à prendre un mouvement cohérent. En absorbant l'énergie fournie, les ions voient le rayon de leur orbite augmenter et ils se rapprochent des parois de la cellule de piégeage. Si plusieurs rapports m/z doivent être enregistrés, un balayage rapide d'une large gamme de fréquence d'excitation est réalisé. Les paquets formés par les différents ions, se placent tous sur la même orbite mais ont des fréquences de rotation dépendant de leur rapport m/z . Ces ions passent alors à proximité des plaques réceptrices et induisent un courant électrique qui est amplifié et enregistré (« transient »). Les différentes fréquences sont déterminées à partir du courant induit par transformée de Fourier. Ces fréquences sont alors converties en rapport m/z grâce à l'équation de calibration.

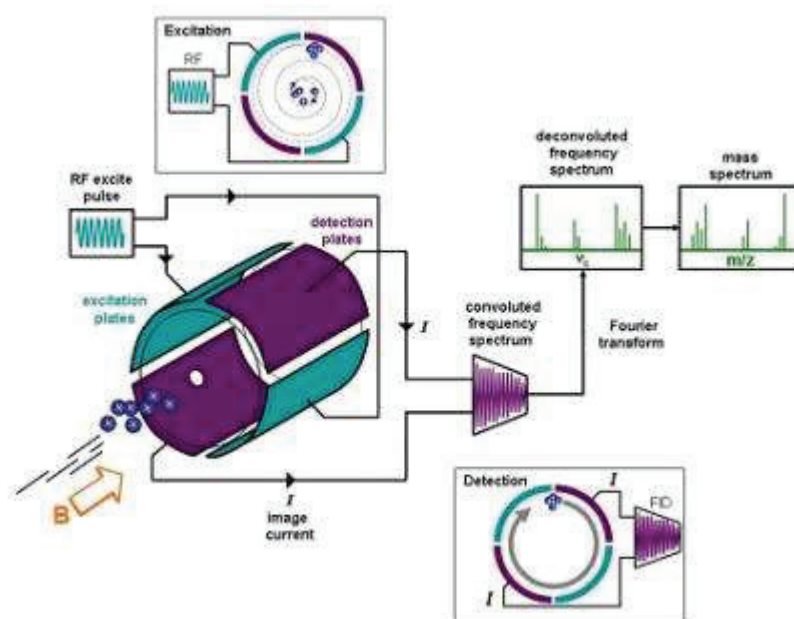


Figure 23 : Principe de fonctionnement d'un analyseur à résonance cyclotronique d'ions ("Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," 2019)

Dans les analyseurs à transformée de Fourier, la résolution obtenue dépend du temps de libre parcours moyen des ions et du temps d'observation du courant qu'ils induisent. Ces temps sont limités par la décroissance du signal qui est la conséquence de collisions entre les ions et des molécules de gaz résiduelles dans le piège. Ces collisions entraînent une perte de cohérence des paquets d'ions et leur ralentissement, ce qui a pour conséquence de diminuer le rayon de leur orbite et de les faire s'éloigner des plaques réceptrices. La fréquence des ions étant inversement proportionnelle à leur rapport m/z , le libre parcours des ions et la résolution diminuent lorsque ce rapport m/z augmente. Les analyseurs à transformée de Fourier apparaissent donc particulièrement adaptés à l'analyse de petites molécules et à des études de métabolomique. Cependant, les plus faibles masses sont plus difficilement excitées car cela nécessite l'utilisation d'un générateur à très haute fréquence. Les FT-ICR modernes sont généralement limités à une basse masse de 125 Da (Forcisi et al., 2013).

Les analyseurs FT-ICR sont ceux possédant le plus haut pouvoir résolutif (jusqu'à 2 000 000). De plus, l'exactitude de mesure de masse apporte une précision de l'ordre de 0,1 ppm. Ces hautes performances sont un réel avantage pour l'analyse de mélange complexe en métabolomique. Cependant, elles sont atteintes au prix d'un temps d'acquisition beaucoup plus important que pour les autres analyseurs (de l'ordre de la dizaine de secondes) ce qui rend les FT-ICR difficilement compatibles avec un couplage à un système chromatographique. De ce fait, les études de métabolomiques réalisées par FT-ICR le sont généralement par introduction directe. Cela permet d'infuser un échantillon en continu pendant quelques dizaines de secondes voire quelques minutes lorsque cela est nécessaire à l'acquisition complète d'un spectre de masse. De plus, les FT-ICR sont

soumis à des problèmes d'effet de charge (Gordon and Muddiman, 2001). Lorsque trop d'ions sont présents dans le piège, la répulsion entre ions va fortement perturber le mouvement des ions de plus faible abondance ce qui impacte la précision de mesure de masse. La gamme dynamique est donc très limitée. Enfin, les performances des FT-ICR étant liées à l'intensité du champ magnétique généré, les aimants utilisés sont généralement des aimants supraconducteurs (jusqu'à 21T) ce qui rend ces instruments très chers à l'achat et à l'entretien.

2.3 Analyseur Orbitrap

La technologie de l'Orbitrap est la plus récente parmi tous les analyseurs. Elle est basée sur le piégeage d'ions (comme le FT-ICR ou le piège ion) et sur l'enregistrement d'un courant induit converti en gamme de fréquences par transformée de Fourier.

Cet analyseur trouve sa source dans le principe de piégeage orbital des ions démontré par Kingdon en 1923. Il utilise un câble chargé placé à l'intérieur d'un cylindre métallique clos. Les ions formés par l'application d'une décharge à l'intérieur du cylindre sont attirés par le câble central. Kingdon a montré que si les ions formés possèdent une vitesse tangentielle suffisante, ils vont se mettre en orbite autour du câble (Kingdon, 1923). Ce n'est qu'à partir des années 2000 que Makarov réussit à améliorer suffisamment ce principe pour créer le premier prototype d'Orbitrap en remplaçant notamment le fin câble central par une électrode en forme de fuseau (Makarov, 2000). L'apport de développement complémentaire, comme l'ajout d'une « C-trap » qui permet une injection pulsée des ions dans l'Orbitrap à partir d'une source d'ions externe, a conduit à la première commercialisation d'un analyseur de masse Orbitrap en 2005 (Makarov et al., 2006).

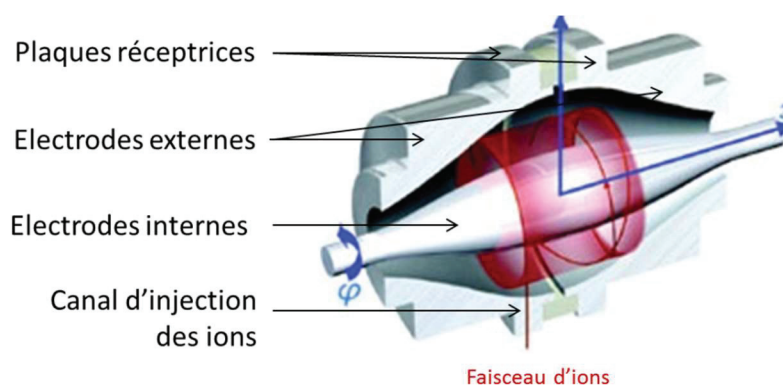


Figure 24 : Principe de fonctionnement d'un Orbitrap (adapté de Makarov and Muenster, 2006)

Un Orbitrap est composé de trois électrodes (Figure 24). Les deux électrodes externes possèdent une forme de coupe et sont placées face à face. Elles sont séparées par un fin anneau isolant. La troisième électrode est centrale en forme de fuseau va maintenir le piège assemblé et aligné. Quand un voltage est appliqué aux trois électrodes, il se forme un champ électrique linéaire selon l'axe de l'électrode centrale. Les ions sont injectés dans l'espace entre les électrodes

extérieures et l'électrode centrale selon une trajectoire tangentielle à l'électrode centrale à travers un orifice situé sur l'une des électrodes externes. La différence de potentiel entre les électrodes externes et l'électrode centrale engendre un champ électrique radial qui tend à attirer les ions vers l'électrode centrale alors que leur vitesse tangentielle crée une force centrifuge opposée. Les ions vont alors adopter une orbite en spirale circulaire dans le piège. Dans le même temps, le champ électrique axial linéaire pousse les ions vers la partie large du piège, entraînant une oscillation axiale des ions. Le vide poussé dans l'Orbitrap permet aux ions de parcourir librement des dizaines de kilomètres. Les électrodes externes sont alors utilisées comme plaque réceptrices pour la détection du courant induit par les oscillations axiales. L'image de ce courant induit est transformée en domaines de fréquence comme c'est le cas pour un analyseur FT-ICR.

En 2011, un nouveau modèle d'analyseur Orbitrap à haut champ est apparu (**Michalski et al., 2012**). Il est plus compact que le précédent. Piéger les ions dans un espace plus restreint permet d'augmenter la fréquence d'oscillation axiale des ions et d'augmenter ainsi la résolution. La position du canal d'injection des ions a également été modifiée. Cette position définit l'amplitude des oscillations axiales ce qui permet d'établir une relation entre la phase d'oscillation et le rapport m/z . Une version améliorée de la transformée de Fourier (dite « enhanced Fourier Transform ») peut alors être utilisée ce qui double la résolution.

La détection des ions est basée sur l'enregistrement d'un courant induit et l'application d'une transformée de Fourier comme c'est le cas pour un FT-ICR. Si dans les deux cas, la résolution décroît avec l'augmentation du rapport m/z , la relation qui lie ces deux paramètres n'est pas la même (**Zubarev and Makarov, 2013**). Dans un FT-ICR où les ions sont piégés par un champ magnétique, la résolution est inversement proportionnelle au rapport m/z . Dans un Orbitrap, le piégeage est lié à un champ électrique et la résolution décroît avec la racine carrée du rapport m/z . Cette plus faible décroissance permet à l'Orbitrap d'être plus résolutif qu'un FT-ICR pour de hauts rapports m/z (**Makarov et al., 2009**). De plus, le piégeage par un champ électrique permet de s'affranchir de l'utilisation d'un aimant supraconducteur. Cela conduit à la conception d'analyseur de plus petite taille et à moindre coût.

Les analyseurs Orbitrap présentent donc des caractéristiques (résolution, précision de mesure de masse, vitesse d'acquisition, coût) intermédiaires parmi les analyseurs à haute résolution. Bien que relativement récents, ils tendent donc à être de plus en plus utilisés pour les analyses métabolomiques.

Annexe 3 : Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy, Supplementary Information

Nicolas Weiss, Pierre Barbier Saint Hilaire, Benoit Colsch, Foucauld Isnard, Suleiman Attala, Augustin Schaefer, Maria del Mar Amador, Marika Rudler, Foudil Lamari, Frédéric Sedel, Dominique Thabut, Christophe Junot

Journal of Hepatology, 2016, 65, 1120-1130

Cerebrospinal fluid metabolomics highlights dysregulation of energy metabolism in overt hepatic encephalopathy

Nicolas Weiss, Pierre Barbier Saint Hilaire, Benoit Colsch, Foucault Isnard, Suleiman Attala, Augustin Schaefer, Maria del Mar Amador, Marika Rudler, Foudil Lamari, Frédéric Sedel, Dominique Thabut, Christophe Junot

Table of contents

Supplementary materials and methods.....	2
Supplementary Table 1.....	9
Supplementary Table 2.....	11
Supplementary Table 3.....	12
Supplementary Table 4.....	13
Supplementary references.....	14

Supplementary materials and methods

Chemicals and reagents.

All analytical grade reference compounds were from Sigma (Saint Quentin Fallavier, France). The standard mixtures used for the external calibration of the MS instrument (Calmix-positive, for the positive ion mode, consisting of caffeine, L-methionyl-arginyl-phenylalanyl-alanine acetate, and Ultramark 1621, and Calmix-negative, for the negative ion mode, consisting of same mixture plus sodium dodecyl sulfate and sodium taurocholate) were from Thermo Fisher Scientific (Courtaboeuf, France). Acetonitrile (ACN) was from SDS (Peypin, France), formic acid from Merck (Briare-le-Canal, France) and methanol from VWR Chemicals (Fontenay-sous-Bois, France). Deionized water was filtered through a Millipore Milli-Q water purification system.

Metabolite extraction.

Plasma samples. Plasma samples were deproteinized by using methanol as previously described (1).

CSF samples. After thawing at room temperature, the CSF samples were immediately processed. CSF samples (100 μ L) were deproteinized in microcentrifuge tubes by adding 300 μ L of methanol, thoroughly mixed on a vortex mixer, incubated in an ultrasonic bath for 10 minutes and then left on ice for 90 minutes. After having been vortexed again, the samples were centrifuged at 3600 RPM for 15 minutes at 4°C. The supernatant was evaporated to dryness under nitrogen at 30°C using a Turbovap (Caliper Life Science Inc., Roissy, France). Then, 100 μ L of either H₂O/ACN, 95/5, v/v, with 0.1% formic acid or H₂O/ACN, 40/60, v/v was added to the residue to reconstitute the initial volume of CSF for RP and HILIC analyses, respectively. The tubes were vortexed again and incubated in the ultrasonic bath for

10 minutes and centrifuged for another 5 minutes. The supernatant was transferred to 0.2 mL vials. Internal standard solution (5 μ L; mixture of 9 authentic chemical standards covering the mass range of interest: ^{13}C -glucose, ^{15}N -aspartate, ethylmalonic acid, amiloride, prednisone, metformin, atropine sulfate, colchicine, imipramine) was added to all samples in order to check for consistency of analytical results in terms of signal and retention time stability throughout the experiment. In addition, a quality control (QC) sample was obtained by pooling 20 μ L of each original CSF sample. It was injected every 10 to 12 samples in order to evaluate the analytical error for each metabolite.

Liquid chromatography coupled to high-resolution mass spectrometry.

The ultra-high performance liquid chromatographic (UHPLC) separation was performed on a Hypersil GOLD C8 1.9 μm , 2.1 mm x 150 mm column at 30°C (Thermo Fisher Scientific, Les Ulis, France), and HPLC chromatographic separations were performed on a Sequant ZICpHILIC 5 μm , 2.1 x 150 mm at 15°C (Merck, Darmstadt, Germany). All chromatographic systems were equipped with an on-line prefilter (Thermo Fisher Scientific, Courtaboeuf, France). Experimental settings for each LC/MS condition are described below.

Mobile phases for RP and PFPP columns were 100% water in A and 100% ACN in B, both containing 0.1% formic acid. Regarding HILIC, phase A consisted of an aqueous buffer of 10 mM ammonium carbonate in water with ammonium hydroxide to adjust basicity to pH 10.5, whereas acetonitrile was used as solvent B. Chromatographic elutions were achieved under gradient conditions as follows:

(i) RP-based system: the flow rate was set at 500 $\mu\text{L}/\text{min}$. The elution consisted of an isocratic step of 2 minutes at 5% phase B, followed by a linear gradient from 5 to

100% of phase B for the next 11 minutes. These proportions were kept constant for 12.5 min before returning to 5% B for 4.5 min.

(ii) HILIC-based system: the flow rate was 200 $\mu\text{L}/\text{min}$. Elution started with an isocratic step of 2 min at 80% B, followed by a linear gradient from 80 to 40% of phase B from 2 to 12 min. The chromatographic system was then rinsed for 5 min at 0% B, and the run ended with an equilibration step of 15 min.

LC-MS analyses were performed using a NEXERA-Shimadzu liquid chromatography system (Champs-sur-Marne, France) coupled to an Exactive mass spectrometer from Thermo Fisher Scientific (Courtaboeuf, France) fitted with an electrospray source operated in the positive and negative ion modes. The software interface was Xcalibur (version 2.1) (Thermo Fisher Scientific, Courtaboeuf, France). The mass spectrometer was calibrated before each analysis in both ESI polarities using the manufacturer's predefined methods and recommended calibration mixture provided by the manufacturer (external calibration).

The Exactive mass spectrometer was operated with capillary voltage at -3 kV in the negative ionization mode and 5 kV in the positive ionization mode and capillary temperature at 280°C. The sheath gas pressure and the auxiliary gas pressure were set, respectively, at 60 and 10 arbitrary units with nitrogen gas. The mass resolution power of the analyzer was 50,000 $m/\Delta m$, full width at half maximum (FWHM) at m/z 200, for singly charged ions. The detection was achieved from m/z 85 to 1000 in the positive ionization mode, from m/z 95 to 1000 for RP chromatography in the negative ionization mode and from m/z 85 to 1000 for HILIC conditions in the two ionization modes.

LC/ESI-MS/MS experiments were performed using a Dionex Ultimate chromatographic system (Thermo Fisher Scientific, Courtaboeuf, France) combined

with a Q-Exactive spectrometer (Thermo Fisher Scientific, Courtaboeuf, France). Non-resonant induced dissociation experiments using higher-energy C-trap dissociation (HCD) were performed at normalized collision energies of 10, 12 and 23%.

Data treatment and handling.

All raw data were manually inspected using the Qualbrowser module of Xcalibur version 2.1 (Thermo Fisher Scientific, Courtaboeuf, France). Raw files were first of all converted to netCDF format using the Xcalibur software. Automatic peak detection and integration were performed using the XCMS software package (2), which returned three data matrices (*i.e.*, corresponding to the three detection conditions) containing m/z and retention time values of features together with their concentrations expressed in arbitrary units (*i.e.*, areas of chromatographic peaks).

XCMS features were thereafter filtered according to the following criteria: (i) the correlation between dilution factors of QC samples and areas of chromatographic peaks (filtered variables should exhibit coefficients of correlation above 0.7 in order to account for metabolites occurring at low concentrations and which are not detected anymore in the most diluted samples), (ii) repeatability (the coefficient of variations obtained for chromatographic peak areas of QC samples should be below 30%) and (iii) ratio of chromatographic peak area of biological to blank samples above a value of 1.5. Finally, chromatographic peak areas of each variable present in the XCMS peak lists were normalized using the LOESS algorithm (3) in order to remove analytical drift induced by clogging of the ESI source observed in the course of analytical runs.

Metabolite annotation and identification.

Features were annotated by matching their accurate measured masses ± 10 ppm with theoretical masses contained in biochemical and metabolomic databases by using an informatics tool developed in R language. The databases used were the Kyoto Encyclopedia of Genes and Genomes (KEGG) (4), the Human Metabolome Database (HMDB) (5) and METLIN (6). Features were also annotated by our spectral database according to accurate measured masses and chromatographic retention times (1, 7).

To be identified, ions had to match at least 2 orthogonal criteria (accurate measured mass, isotopic pattern, MS/MS spectrum and retention time) to those of an authentic chemical standard analyzed under the same analytical conditions, as proposed by the Metabolomics Standards Initiative (8). In the absence of an available authentic chemical standard, other metabolites of interest were putatively annotated, based on accurate measured mass and interpretation of the MS/MS spectra when available.

Statistical analyses.

Chromatographic peak areas contained in the 3 filtered and LOESS-processed XCMS data matrices (i.e., RP-UHPLC-HRMS with MS detection in both positive and negative ESI modes, and HILIC-HPLC-HRMS with MS detection in the negative ESI mode) were first mean-centered and scaled to unit variance before performing principal component analysis using SIMCA-P12 (Umetrics, Umea, Sweden). Statistical significance was assessed by using a Mann-Whitney test at a 0.05 threshold using the SPSS software (SPSS Inc., Chicago, USA). Graphic representations of the data were obtained using the GraphPad Prism software (GraphPad Software, Inc., La Jolla, CA).

Correlations between metabolites of the HE metabotype and clinical scores or biochemical parameters related to liver disease (*i.e.*, MELD and Child-Pugh scores, plasma bilirubin, creatinine and ammonia concentrations, and prothrombin time, creatininemia), and also the severity of HE (*i.e.*, West Haven score and GCS score) were evaluated by calculating Pearson correlation coefficients.

Determination and visualization of the HE metabotype.

Features corresponding to identified or putatively annotated ions were selected. The areas of their chromatographic peaks were integrated by using the Qualbrowser module of Xcalibur version 2.1 (Thermo Fisher Scientific, Courtaboeuf, France) in order to confirm the XCMS results. Integrations were then manually reviewed in each biological sample. The resulting datasets consisted of 3 matrices containing *m/z* and retention time values of features together with their concentrations expressed in arbitrary units (*i.e.*, areas of chromatographic peaks). Data were expressed as ratios of individual levels observed for EH patients and controls to control mean levels.

The CSF metabolome as analyzed by liquid chromatography coupled to high-resolution mass spectrometry.

Three complementary methods (involving a reverse phase (RP) column with detection in both positive and negative electrospray modes, and a hydrophilic interaction liquid chromatography (HILIC) column with detection in the negative electrospray mode) were used to provide optimal coverage of the CSF metabolome. Among the several thousand signals per chromatogram obtained for each CSF sample, data processing led to the annotation of few hundred features per method.

The annotation procedure and also MS/MS experiments led to the identification and relative quantification of 122 metabolites. Analytical information and

identification status (see methods for additional information) of annotated metabolites are displayed in Table 1s. The CSF metabolome, as detected in our experimental conditions, was composed of 38% amino acids and derivatives. Other detected species were organic acids, nucleotides and conjugates, carbohydrates and conjugates, bile acids, ketones, and alcohols. About 40% of these metabolites are not reported as present in CSF according to the human metabolome database (HMDB, <http://www.hmdb.ca>) (Table 1s). Although our LC/MS approach does not provide absolute quantification, concentrations of detected metabolites could be estimated to range from a few micromoles to hundreds of micromoles per liter, according to HMDB (Table 1s).

Supplementary Table 1.

MetaboliteID	N°	Subclass	Class	Identification status	Formula	M	m/z	HMDB	KEGG	ESI	Reported in CSF	Reported concentration in CSF**
Dimethylguanoxine	2	-	Purine Nucleosides and Analogues	a,c	C12H17N5O5	311	1229.6	2.89	-	[M-H] ⁻	Pos	no
Stachyline	3	Monoterpenes	Prenyl Lipids	a,b,d	C7H13NO2	143	196.282	0.87	-	[M-H] ⁻	Pos	no
1-L-tryptophan	4	-	Peptides	a,c	C11H16N2O2	228	147.918	1.08	-	[M-H] ⁻	Pos	no
2-hydroxy-2-methylbutyric acid	7	-	-	a,b	C5H10O3	116	102.078	2.45	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
2-hydroxy-3-methylbutyric acid	9	-	-	a,b	C5H10O3	116	102.078	2.8	-	[M-H] ⁻	Pos	yes
2-Hydroxyvaleric acid	11	Beta Hydroxy Acids and Derivatives	Hydroxy Acids and Derivatives	a,b,c	C5H10O3	116	102.078	3.43	2.6	[M-H] ⁻	Neg	no
2-methylbutyrcarnitine	12	Acyl Carnitines	Fatty Acid Esters	a	C12H23NO4	245	162.704	4.74	-	[M-H] ⁻	Pos	no
2R-aminobutyric acid	16	-	Fatty Acids and Conjugates	a	C4H9NO2	271	121.282	9.17	-	[M-H] ⁻	Pos	no
1,4-Dihydroxybutyric acid	17	Beta Hydroxy Acids and Derivatives	Hydroxy Acids and Derivatives	a,c	C4H8O4	120	102.024	7.67	-	[M-H] ⁻	Neg	yes
2-aminocinnamic acid	19	Amino Fatty Acids	Fatty Acids and Conjugates	a,c	C8H13NO2	159	125.922	5.12	-	[M-H] ⁻	Pos	no
3-Hydroxy-2-methylbutyric acid	22	Beta Hydroxy Acids and Derivatives	Hydroxy Acids and Derivatives	a	C5H10O3	116	102.078	1.98	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
3-Hydroxyisovalerylcarnitine	23	Acyl Carnitines	Fatty Acid Esters	a,c	C12H23NO4	245	162.722	1.62	-	[M-H] ⁻	Pos	yes
3-Hydroxy-2-methylbutanoic acid	24	Beta Hydroxy Acids and Derivatives	Hydroxy Acids and Derivatives	a,b	C5H10O3	116	102.078	1.99	-	[M-H] ⁻	Neg	yes
4-Acetylglutamic acid	29	Amino Fatty Acids	Fatty Acids and Conjugates	a,b,d	C8H13NO2	143	107.928	1.42	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
5-Aminovaleric acid	32	Delta Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C5H11NO2	117	107.922	0.95	-	[M-H] ⁻	Pos	no
5-deoxy-5-(methylthio)adenosine	33	-	Purine Nucleosides and Analogues	a,d	C11H15NSO3S	297	189.595	3.94	-	[M-H] ⁻	Pos	no
5-hydroxyindoleacetic acid	35	Indolyl Carboxylic Acids and Derivatives	Indoles	a,b,c	C6H9NO3	191	153.028	4.55	2.4	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
5-Methylthio-D-ribose	38	Pentoses	Monosaccharides	a	C6H12O5S	180	164.029	1.15	-	[M-H] ⁻	Pos	no
Leucine isoleucine	39	Amino Fatty Acids	Fatty Acids and Conjugates	a,b	C8H13NO2	131	104.628	1.43	4.2	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Acetyl-galactosamine	42	Hexenes	Monosaccharides	a,b,c	C8H13NO6	256	199.137	0.85	-	[M-CT]	Neg	no
Acetyl-L-carnitine	45	Acyl Carnitines	Fatty Acid Esters	a,b,d	C9H17NO4	203	115.754	1	-	[M-H] ⁻	Pos	yes
Acetyl-sebacic acid	47	Sugar Amino Acids and Derivatives	Sugar Acids and Derivatives	a,b,c,d	C11H19NO9	209	199.084	0.85	7.62	[M-H] ⁻	Neg	yes
Acetyl-serine	51	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C5H9NO4	147	103.154	5.11	-	[M-H] ⁻	Neg	no
Acetyl-tyrosine	52	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C11H13NO4	223	184.454	4.15	-	[M-H] ⁻	Neg	no
Acetyl-omega-OR acetyl-auronine	53	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,d	C7H13NO3	159	105.928	4.07	-	[M-H] ⁻	Neg	no
Adipic acid	54	Dicarboxylic Acids and Derivatives	Carboxylic Acids and Derivatives	a,b	C8H16O4	146	157.084	3.68	-	[M-H] ⁻	Neg	yes
Adonitol / Arabitol	56	-	-	a,b,c	C5H12O6	187	173.741	0.86	6.5	[M-CT]	Neg	yes
Alanine	58	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C3H7NO2	89	87.072	7.69	-	[M-H] ⁻	Pos / Neg	yes
Allantoin	60	Imidazolidinones	Imidazolidinones	a,c	C4H8N2O3	158	143.988	-	7.08	[M-H]	Neg	yes
Alfose	61	Hexenes	Monosaccharides	a,b	C6H12O6	180	163.176	0.86	-	[M-H] ⁻	Pos	no
Antimonic acid	65	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a	C6H13NO4	157	107.974	1.64	-	[M-H] ⁻	Neg	no
Anhydro-sorbitol	67	Hexenes	Monosaccharides	a,c,d	C6H12O5	166	166.043	-	5.88	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Arginine	69	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C6H14N4O2	174	111.752	-	10	[M-H]	Neg	yes
Ascorbic acid	71	Furanones	Dihydrofurans	a,c	C6H8O6	176	176.028	-	8.22	[M-H]	Neg	yes
Asparagine	72	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C4H8N2O3	132	135.418	-	7.64	[M-H]	Neg	yes
Aspartic acid	74	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C4H7NO4	133	137.074	0.85	8.3	[M-H]	Neg	yes
Beta-cortisol	76	Gluco甾oidal corticoids, Progestins and Derivatives	Steroids and Steroid Derivatives	a	C21H30O5	368	256.273	6.16	-	[M-H]	Neg	no
Carnitine	79	Acyl Carnitines	Fatty Acid Esters	a,b,d	C7H13NO3	161	105.192	0.84	-	[M-H] ⁻	Pos	yes
cis-Ascorbic acid	81	Tricarboxylic Acids and Derivatives	Carboxylic Acids and Derivatives	a,b,d	C6H8O6	176	176.038	1.14	-	[M-H] ⁻	Pos	yes
Citric acid	84	Tricarboxylic Acids and Derivatives	Carboxylic Acids and Derivatives	a,b,c,d	C6H8O7	192	192.022	1.65	9.4	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Citraline	86	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C6H13NO3	175	105.968	-	7.91	[M-H]	Neg	yes
Cortisol	87	Gluco甾oidal corticoids, Progestins and Derivatives	Steroids and Steroid Derivatives	a,b	C21H30O5	362	250.23	6.96	-	[M-H] ⁻	Pos	yes
Cytidine	88	Pyrimidines and Pyrimidine Derivatives	Nucleosides and Nucleotides	a,b	C10H12N2O5	226	194.026	1.07	-	[M-H] ⁻	Pos	yes
Creatine	91	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c	C4H9N3O2	131	109.672	0.86	7.7	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Creatinine	94	-	Lactams	a,b,c,d	C4H7NO	115	108.916	0.85	3.85	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Cytosine	96	Pyrimidines and Pyrimidine Derivatives	Amino Acids and Derivatives	a,c,d	C5H7N3O2	240	123.048	-	8.34	[M-H]	Neg	yes
Dihydroxybutyric acid	97	Cyclic Methylols and Derivatives	Aliphatic Homocyclic Compounds	a,c	C5H10O4	196	107.776	1.2	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
Deoxythymine	100	Pentoses	Monosaccharides	a,b	C5H10O4	134	157.084	1.36	-	[M-H]	Neg	no
Dihydrothymine	101	Pyrimidines and Pyrimidine Derivatives	Dianions	a,b,c,d	C5H8NO2	128	105.372	1.41	2.4	[M-H] ⁻	Pos	yes
Formylmethionine	112	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C8H15NO3S	177	147.963	3.69	-	[M-H]	Neg	no
Fumarate	113	Unsaturated Fatty Acids	Fatty Acids and Conjugates	a,c	C4H4O4	116	109.954	-	8.46	[M-H]	Neg	yes
Furic acid	114	Furic Acid Derivatives	Furans	a	C5H4O3	132	116.038	1.07	-	[M-H]	Neg	no
Gluconic acid	116	-	-	a,c,d	C6H12O7	196	196.022	-	7.77	[M-H]	Neg	no
Gluamic acid	121	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c	C5H9NO4	147	103.154	0.91	8.25	[M-H]	Neg	yes
Glutamine	124	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C5H11NO2	146	109.148	0.84	7.59	[M-H]	Neg	yes
Glutarylphenylalanine	127	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C14H18N2O5	294	123.171	4.59	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
Glyceric acid	129	-	Sugar Acids and Derivatives	a,b,c,d	C3H6O4	106	106.084	0.86	5.9	[M-H]	Neg	yes
Glycolic acid	133	Bile Acids, Alcohols and Derivatives	Steroids and Steroid Derivatives	a,b	C2H4O3	46	309.066	7.84	-	[M-H]	Neg	no
Glyoxyloxydicarboxylic / Glyoxyloxydicarboxylic acid	134	Bile Acids, Alcohols and Derivatives	Steroids and Steroid Derivatives	a,b	C2H4O3	46	311.422	8.75	-	[M-H]	Neg	no
Hexose	135	Hexenes	Monosaccharides	a	C6H12O6	180	163.176	0.85	-	[M-CT]	Neg	yes
Hippuric acid	137	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C9H9NO3	179	105.248	4.73	-	[M-H]	Neg	yes
Histidine	139	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C6H9NO2	155	109.672	-	7.49	[M-H]	Neg	yes
Homouric acid	140	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C7H14N2O2	188	123.252	0.85	-	[M-H] ⁻	Pos	yes
Homoserine	141	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C4H9NO2S	135	105.989	-	4.41	[M-H]	Neg	yes
Hydroxyquinone	145	Benzenetriols and Derivatives	Phenols and Derivatives	a	C6H6O3	126	131.038	4.03	-	[M-H] ⁻	Pos	no
Hydroxytyrosin	147	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,d	C11H15NO3	220	147.918	1.59	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	not in adult
Indolelactic acid	151	-	-	a,b,d	C11H13NO3	205	173.028	5.93	-	[M-H]	Neg	no
Indoxyl sulfate	152	-	Indoles	a,c	C8H7NO4S	213	169.578	4.77	2.68	[M-H]	Neg	no
Inosine	157	Purine 2'-deoxyribonucleosides and analogues	Purine Nucleosides and Analogues	a,b,c	C10H12N4O5	268	180.769	1.24	5.11	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Isoquinoline N-oxide	159	-	Isoquinolines	a	C9H7NO	145	102.736	5.09	-	[M-H] ⁻	Pos	no
Lactic acid	162	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C3H6O3	90	103.038	-	4.1	[M-H]	Neg	yes
Leu-Ala	163	-	Peptides	a,c	C9H15NO2	202	131.748	0.94	-	[M-H] ⁻	Pos	no
Levulinic acid OR isomers	167	-	-	a	C5H8O3	116	104.748	1.94	-	[M-H]	Neg	-
Kynurenine	168	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,d	C10H12N2O3	208	147.918	2.22	-	[M-H] ⁻	Pos	yes
Lysine	170	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c,d	C6H12NO2	146	115.272	-	9.55	[M-H]	Neg	yes
Malic acid	172	Dicarboxylic Acids and Derivatives	Carboxylic Acids and Derivatives	a,b,c	C4H6O5	134	123.523	0.92	8.59	[M-H]	Neg	yes
Methionine	174	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C5H11NO2S	195	151.089	1.07	4.27	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
Methionine sulfonide	175	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C5H11NO3S	165	145.963	1.07	-	[M-H] ⁻	Pos	no
Methylglucosamine	176	Branched Fatty Acids	Fatty Acids and Conjugates	a,c	C8H18O4	144	142.284	-	7.66	[M-H]	Neg	no
Methylthymine	177	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C7H11NO2	160	121.172	-	9.18	[M-H]	Neg	yes
N-Acetyltyrosine	179	-	Pyrimidine Nucleosides and Analogues	a,d	C11H15NO3	285	196.086	2.09	-	[M-H] ⁻	Pos	no
N-acetylserine / N-acetylthreonine	180	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,d	C8H13NO3	173	105.192	5.2	-	[M-H]	Neg	no
N-Acetylparaglutamic acid	182	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C11H16NO6	234	190.668	1.18	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	yes
N-Acetylthreonine	186	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b	C4H7NO3	117	104.748	0.96	-	[M-H] ⁻	Pos	no
Acetyl-tyrosine	188	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C5H9NO4	131	105.248	1.29	3.73	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
N-Acetylserine	192	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C5H9NO3S	175	104.802	-	8.08	[M-H]	Neg	no
N-Acetyl-L-histidine	194	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,c	C8H11NO3	197	108.008	0.87	-	[M-H] ⁻ AND [M-H] ⁺	Pos / Neg	no
Acetyl-methionine	196	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C11H17NO3S	191	161.135	4.14	2.3	[M-H]	Neg	no
N-Acetyl-L-phenylalanine	200	Alpha Amino Acids and Derivatives	Amino Acids and Derivatives	a,b,c,d	C11H13NO3	207	109.672	5.72				

^aBased on accurate mass; ^bBased on C8 column retention time similarity with a standard; ^cBased on HILIC column retention time similarity with a standard; ^dBased on MS² spectrum similarity with a standard; ^eBased on interpretation of MS² spectra; ^{**}according to HMDB (<http://metabolomics.ca>).

Supplementary Table 2.

Putative annotation	HMDB	Metlin	KEGG	Elemental composition	Samples	LC/HRMS condition	ESI mode	Retention time (min)	m/z	MS interpretation
Levetiracetam	HMDB15333	66750	C07841	$C_8H_{14}N_2O_2$	BM2	RP	positive	3.9	171.1129	[M+H] ⁺
									193.0948	[M+Na] ⁺
									363.200	[2M+Na] ⁺
									154.0864	[M+H-NH3] ⁺
									126.0914	[M+H-HCONH2] ⁺
Metronidazole	HMDB15052	573	C07203	$C_6H_9N_3O_3$	BG9	RP	positive	2.8	172.0716	[M+H] ⁺
						HILIC	negative	2.1	128.0455	[M+H-C2H4O] ⁺
									170.0566	[M-H] ⁻
Fluconazole	HMDB14342	2708	$C_{13}H_{12}F_2N_6O$	MGH9, NQ1, ZG8, AZO	RP	positive	5.7	307.1109	[M+H] ⁺	
								305.0968	[M-H] ⁻	
								341.0735	[M+Cl] ⁻	
								343.0705	[M+Cl] ⁻ (37Cl)	
								351.1021	[M+HCOOH-H] ⁻	
								191.0685	[M-H-C6H4F2] ⁻	
								305.0968	[M-H] ⁻	
								341.0735	[M+Cl] ⁻	
								191.0685	[M-H-C6H4F2] ⁻	
								191.0685	[M-H-C6H4F2] ⁻	
Diazepam	HMDB14967	3521	C06948	$C_{16}H_{13}ClN_2O$	BC8, MHG9, MP01, NQ1, ZG8	RP	positive	8.6	285.0787	[M+H] ⁺
									287.0757	[M+H] ⁺ (37Cl)
N-Desmethyldiazepam	HMDB60538	896	C07486	$C_{15}H_{11}ClN_2O$	BC8, MHG9, MP01, NQ1, ZG8	RP	positive	7.8	271.0627	[M+H] ⁺
									273.0594	[M+H] ⁺ (37Cl)
									301.0599	[M+H] ⁺
Tazobactam	HMDB15544	581	D00660	$C_{10}H_{12}N_4O_5S$	BG9, NQ1, ZG8	RP	positive	2.7	303.0557	[M+H] ⁺ (34S)
									540.1531	[M+H] ⁺
Piperacillin	HMDB14464	1964	D08380	$C_{23}H_{27}N_5O_7S$	BG9, NQ1, ZG8	RP	positive	7	518.1702	[M+Na] ⁺
Ciprofloxacin	HMDB14677	1772	C05349	$C_{17}H_{18}FN_3O_3$	AZO, BG9, ZG8, NQ1, SP3	RP	positive	5.2	332.1405	[M+H] ⁺
Norfloxacin				$C_{16}H_{18}FN_3O_3$	CJ3, BC8	RP	positive	5.1	320.1406	[M+H] ⁺

Supplementary Table 3. Correlation between the metabolic signature of HE patients and clinical parameters.

		r_s^a							
		NH3	MELD ^b	ChildP	Bili	PT	Creat	W-H	GCS ^c
Acetylated compounds	Acetyl-methionine	0.027	0.598	0.682	0.575	-0.424	0.134	0.166	-0.110
Methionine metabolism	Methionine	0.115	0.689	0.364	0.666	-0.592	0.115	0.034	-0.201
	Methionine sulfoxide	-0.018	0.594	0.318	0.561	-0.715	-0.007	0.240	-0.258
	Formylmethionine	0.055	0.516	0.571	0.509	-0.581	0.116	0.530	-0.377
Phenylalanine metabolism	p-hydroxyphenyllactic acid	0.191	0.708	0.618	0.424	-0.662	0.077	0.291	-0.191
Tryptophan metabolism	Tryptophan	0.036	0.671	0.507	0.514	-0.529	0.240	0.495	-0.395
	Indolelactic acid	-0.227	0.671	0.516	0.377	-0.403	0.297	0.180	-0.063
	5-hydroxyindoleacetic acid	0.155	0.105	0.502	-0.083	0.221	-0.446	0.562	-0.552
Amino-acids and derivatives	Lysine	-0.648	0.500	0.216	0.035	0.273	0.203	0.152	-0.168
	Stachydrine	-0.455	-0.279	-0.037	0.058	0.011	0.697	-0.317	0.426
	Citrulline	-0.297	0.024	-0.420	-0.004	-0.018	0.753	-0.180	0.190
Peptides	Leu-Ala	-0.027	0.644	0.553	0.627	-0.375	0.226	0.111	-0.135
Bile acids	Glycocholic acid	0.309	0.612	0.521	0.688	-0.806	0.033	0.197	-0.173
	Glycoursodeoxycholic acid OR isomers	0.027	0.703	0.470	0.704	-0.806	0.130	0.053	-0.061
	Taurocholic acid	-0.045	0.498	0.606	0.567	-0.627	-0.022	0.203	-0.157
Fatty acid metabolism	carnitine	-0.236	0.434	0.327	0.402	-0.410	-0.068	0.763	-0.834
	Acetyl-L-carnitine	-0.436	0.525	0.272	0.572	-0.399	0.385	0.157	-0.094
	Octanoylcarnitine	-0.109	0.662	0.535	0.759	-0.399	0.301	0.009	-0.018
	2-Hydroxyvaleric acid	-0.136	0.612	0.802	0.561	-0.634	0.160	0.129	-0.110
Amino-fatty acids	3-amino-octanoic acid	0.145	0.543	0.438	0.440	-0.781	-0.178	0.361	-0.227
Nucleosides and derivatives	Pseudouridine	-0.155	0.151	0.014	0.121	-0.053	0.552	-0.301	0.424
Pyrimidine metabolism	Dihydrothymine	-0.055	0.626	0.691	0.657	-0.543	0.279	-0.018	-0.016
	Uracil	-0.055	0.140	-0.278	0.193	-0.223	0.610	-0.670	0.564
Alcohols and polyols	Quinic acid	-0.539	-0.720	-0.803	-0.613	0.642	0.341	-0.387	0.503
Sugar acids and derivatives	Gluconic acid	-0.709	0.256	-0.111	-0.004	0.205	0.319	-0.026	0.073
Sugar alcohols	Adonitol	-0.188	0.778	0.673	0.655	-0.446	0.033	0.134	-0.134
Vitamin B6 metabolism	Pyridoxic acid	-0.358	0.805	0.512	0.301	-0.369	-0.236	0.335	-0.330

^a Pearson correlation coefficient, ^b MELD (Model for End-stage Liver Disease) score, ^c Glasgow Coma Scale. Red/blue: $p < 0.005$, Orange/green: $0.005 < p < 0.05$. Red/orange: positive correlation. Blue/green: negative correlation. Abbreviations: NH3, ammonia; ChildP: Child-Pugh score; Bili, bilirubin; PT, prothrombin time; Creat, creatinine; W-H, West Haven scale.

Supplementary Table 4. Clinical metadata of plasma samples

	Healthy controls n = 9	Cirrhotic patients n = 13	HE patients n = 12
Age\$	29 [26-44]	56 [55-64]	59 [40-68]
Sex (male/female)	2/7	9/4	7/5
Cirrhosis	NA	13 (100%)	13 (100%)
Etiology of cirrhosis*			
Alcohol	NA	5 (38%)	7 (58%)
Viral	NA	4 (31%)	4 (33%)
NASH	NA	3 (23%)	0 (0%)
Ascites*	NA	8 (62%)	7 (58%)
Flapping*	NA	3 (23%)	8 (67%)
Child-Pugh score*\$	NA	10 [7-11]	12 [10-13]
MELD *\$	NA	17 [14-29]	20 [16-31]
West Haven score (WH) *\$	NA	0 [0-1]	2 [2-3]
GCS \$	NA	15 [15-15]	15 [14-15]
Blood			
Leukocytes (Giga / L)\$	6 110 [5 325-7 195]	4 790 [4 030-8 575]	7 590 [4 632-12 175]
Platelets (Giga / L)\$	264 000 [207 000-323 000]	76 000 [57 000-110 000]	85 000 [67 000-155 000]
Prothrombin time (%)\$	98 [82-100]	49 [37-67]	39 [23-57]
INR\$	1.2 [1.0-1.3]	1.8 [1.3-2.5]	2.1 [1.4-3.0]
Factor V (%)\$	100 [66-100]	48 [37-77]	45 [28-76]
AST (U / L)\$	25 [23-30]	44 [35-64]	66 [50-88]
ALT (U / L)\$	15 [12-27]	25 [19-39]	34 [23-50]
Creatinine (µmol / L)\$	73 [58-78]	65 [52-96]	83 [62-147]
Sodium (mmol / L)\$	140 [139-141]	137 [133-142]	136 [134-140]
Albumin (g / L)\$	44 [42-47]	32 [26-34]	26 [23-34]
CRP (mg / L)*\$	5 [3-14]	15 [5-57]	13 [7-52]
Bilirubin (mmol/L) \$	6 [5-11]	45 [26-145]	185 [33-476]
Ammonemia (µmol / L)*\$	34 [29-39]	66 [48-76]	88 [58-122]

\$: Values are medians with interquartile ranges. *: only available in cirrhotic patients.

Supplementary references

- (1) Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 2006 Feb 1;78(3):779-787.
- (2) Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2011;6(7):1060-1083.
- (3) Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000 Jan 1;28(1):27-30.
- (4) Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007 Jan;35(Database issue):D521-D526.
- (5) Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005 Dec;27(6):747-751.
- (6) Roux A, Xu Y, Heilier JF, Olivier MF, Ezan E, Tabet JC, et al. Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-Orbitrap mass spectrometer. *Anal Chem* 2012 Aug 7;84(15):6429-6437.
- (7) Boudah S, Olivier MF, ros-Calt S, Oliveira L, Fenaille F, Tabet JC, et al. Annotation of the human serum metabolome by coupling three liquid

chromatography methods to high-resolution mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 2014 Sep 1;966:34-47.

- (8) Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 2007 Sep;3(3):211-221.

Annexe 4 : Paramètres analytiques utilisés lors des analyses métabolomique, lipidomique et glycomique des échantillons de plasma de la seconde cohorte d'encéphalopathes

1. Analyse métabolomique :

Préparation des échantillons plasmatiques : prise d'essai de 25 μL de sérum, ajout de 75 μL de méthanol, sonication au bain à ultrasons (avec ajout de glace) pendant 15 min, centrifugation à 15 000 g pendant 15 min, précipitation lente des protéines sur glace durant 1h30, centrifugation 15 min à 20 000 g. Récupération du surnageant et évaporation à sec sous flux d'azote, reprise de l'extrait sec par 250 μL d'un mélange eau/ACN (95/5, v/v), agitation, sonication au bain à ultrasons (avec ajout de glace) pendant 15 min, centrifugation à 20 000 g pendant 15 min à 4°C. Transfert du surnageant en vial.

Conditions chromatographiques :

Utilisation d'un système Dionex Ultimate 3000 (ThermoFisher Scientific, Les Ulis, France).

C18 : colonne Hypersil Gold C18 (150 x 2.1 mm, 1.9 μm ; ThermoFisher Scientific) thermostaté à 30°C. Phase mobile A constituée d'eau Chromasolv LC-MS (SigmaAldrich) avec ajout de 0,1 % d'acide formique (v/v). Phase mobile B constituée d'acétonitrile Chromasolv LC-MS (SigmaAldrich) avec ajout de 0,1 % d'acide formique (v/v). Débit de 500 $\mu\text{L}/\text{min}$. Gradient : 5 % de phase mobile B pendant 2 min puis augmentation linéaire jusqu'à 100% de phase mobile B sur 11 min. Plateau à 100 % pendant 12,5 min puis retour instantané aux conditions initiales et équilibration à 5 % de phase mobile B pendant 4.5 min. Volume d'injection de 5 μL .

HILIC : colonne SeQuant ZICpHILIC de dimension 2.1 x 150mm, 5 μm , thermostatée à 15 °C (Merck, Darmstadt, Allemagne). Phase mobile A composée d'une solution de carbonate d'ammonium (10 mM) en eau Chromasolv LC-MS (SigmaAldrich) ajustée à un de pH 10.5 avec de l'hydroxyde d'ammonium. Phase mobile B constituée d'acétonitrile Chromasolv LC-MS (SigmaAldrich). Débit de 200 $\mu\text{L}/\text{min}$. Gradient : 80% de phase mobile B pendant 2 min puis diminution linéaire jusqu'à 40 % de phase mobile B sur 10 min suivie d'un plateau à 0 % de phase mobile B pendant 5 min puis retour instantané aux conditions initiales et équilibration à 80 % de phase mobile B pendant 15 min. Volume d'injection de 10 μL .

Conditions HRMS :

Utilisation de l'Oritrapp Fusion (ThermoFisher Scientific, Les Ulis, France) équipé d'une source Electrospray chauffée. Paramètres de la source en conditions C₁₈ : *Spray voltage*, + 3 500 V en mode positif, - 2 500 V en mode négatif ; *sheath gas*, 60 (unités arbitraires); *auxiliary gas*, 12 (unités

arbitraires); *sweep gas*, 3 (unités arbitraires); température de l'*ion transfer tube*, 325 °C; température du *vaporizer*, 350 °C. Paramètres de la source en conditions HILIC : *Spray voltage*, + 4 000 V en mode positif, - 2 800 V en mode négatif; *sheath gas*, 40 (unités arbitraires) ; *auxiliary gas*, 12 (unités arbitraires) ; *sweep gas*, 1 (unités arbitraires) ; température de l'*ion transfer tube*, 325 °C; température du *vaporizer*, 100 °C. Paramètres de transfert des ions : *normal mass range*; activation du *quadrupole isolation*; *scan range*, 85-1000 m/z; *S lens RF level*, 50 %. Paramètres d'acquisition des spectres sur l'Orbitrap : *acquisition time*, 0-13 min ; *maximum injection time*, 100 ms; 1 micro scan; acquisition en mode profile, AGC target, 5^e4 ; *resolving power*, 240 000 (FWHM à m/z 200).

2. Analyse lipidomique :

Préparation des échantillons plasmatiques :

Prise d'essai de 50 µL de sérum, ajout de 250 µL de mélange Chloroforme/Méthanol (1/1, v/v), agitation au vortex à puissance maximale pendant 1 minute, sonication à la sonde à sonication : 2×3×10 secondes (bien refroidir dans un bain de glace entre chaque sonication). Continuer l'extraction sous agitation rotative pendant 2 heures. Ajout de 37,5 µL d'eau pure. Le ratio : Chloroforme/Méthanol/(Plasma+H2O) doit être de 1/1/0.7 (v/v/v). Agitation au vortex à puissance maximale pendant 2×30 secondes. Centrifugation à 4 000 tours/min pendant 15 minutes. Prélèvement de la phase supérieure (aqueuse) et de la phase inférieure. Ces deux phases sont poolées et évaporées sous azote. Reprise avec 125 µL de Chloroforme/Méthanol (1/1, v/v), agitation au vortex 2×30 secondes, centrifugation à 12 000 g pendant 15 minutes, récupération du surnageant et évaporation sous azote. Reprise des échantillons avec 50 µL de Chloroforme/Méthanol (1/1, v/v), agitation au vortex 2×30 secondes. Dilution des échantillons au 1/100^{ème} dans une solution de MeOH/IPA/H2O 65:35:5. Transfert en vial.

Conditions chromatographiques :

Utilisation d'un système Dionex Ultimate 3000 (ThermoFisher Scientific, Les Ulis, France). Colonne Kinetex C₈ (150 x 2.1 mm, 2.6 µm ; Phenomenex, Sydney, NSW, Australie) thermostaté à 60°C. Phase mobile A composée d'un mélange de H2O/MeOH 60/40 (v/v) additionné de 0.1% d'acide formique. Phase mobile B composée d'un mélange de IPA/MeOH 90/10 (v/v) additionné de 0.1% d'acide formique. Ajout de formate d'ammonium (10 mM) en mode positif. Débit de 400 µL/min. Gradient : 0.0min, 68 % de A, 32 % de B ; 2.5min, 68 % de A, 32 % de B ; 3.5min, 55 % de A, 45 % de B ; 5.0min, 48 % de A, 52 % de B ; 7.0min, 42 % de A, 58 % de B ; 10.0min, 34 % de A, 66 % de B ; 13.0min, 30 % de A, 70 % de B ; 16.0min 25 % de A, 75 % de B ; 20.0min, 20 % de A, 80 % de B ; 23.0min, 15 % de A, 85 % de B ; 26.0min, 10 % de A, 90 % de B ; 28.0min 5 % de A, 95 % de B ; 30.0min, 3 % de A, 97 % de B ; 32.0min, 68 % de A, 32 % de B ; 40.0min, 68 % de A, 32 % de B. Volume d'injection de 10 µL.

Conditions HRMS :

Utilisation du Q Exactive (ThermoFisher Scientific, Les Ulis, France) équipé d'une source Electrospray. Paramètres de la source : *Spray voltage*, + 3700 V en mode positif, - 3100 V en mode négatif ; *sheath gas*, 15 (unités arbitraires); *auxiliary gas*, 10 (unités arbitraires); température de l'*ion transfer tube*, 360 °C. Paramètres de transfert des ions : activation du *quadrupole isolation*; *scan range*, 150-2000 m/z; *S lens RF level*, 60 %. Utilisation de l'orbitrap à un pouvoir résolutif de 140 000 (FWHM à m/z 200).

3. Analyse glycomique :

Préparation des échantillons plasmatiques :

5 µL de plasma sont dilués dans un tampon phosphate sodique à 20 mM (pH 7.4) et une solution de dithiothreitol à 10 mM. La solution est chauffée à 95°C pendant 5 minutes. L'échantillon est ensuite digéré avec 2 µL de N-glycosidase F (1 U/µL, Roche Diagnostics, Meylan, France) à 37°C pendant une nuit. Les glycosylamines résiduelles sont convertis en glycanes par acidification (5µL d'HCl à 1 M, 45 min à 37°C). Après l'acidification, les protéines sont précipitées dans 150 µL d'éthanol à -20°C pendant 1h. Les surnageants contenant les glycanes sont évaporés à sec puis repris dans 200 µL d'acide trifluoroacétique (2,5 %). Les N-glycanes sont alors purifiés en utilisant une cartouche SPE garnie de carbone graphite poreux (Thermo Scientific, les Ulis, France), conformément aux recommandations du fabricant. Les glycanes sont élués à l'aide d'une solution d'ACN à 50 % contenant 0.1 % d'acide formique puis évaporés à sec. Les glycanes sont repris dans une 40 µL de DMSO, 60 µL de suspension de NaOH/DMSO et 60 µL d'iodure de méthyl. Les échantillons sont laissés à incuber dans un bain à ultrasons à température ambiante pendant 90 min. La réaction est arrêtée par l'ajout de 200 µL d'acide acétique à 2,5 %. Les glycanes perméthylés sont extraits par deux fois 120 µL de CHCl₃ puis lavés cinq fois par 250 µL d'eau. Après évaporation, les glycanes sont repris dans 150 µL d'une solution de méthanol à 50 %. Les N-glycanes sont purifiés sur une colonne C₁₈ spin (Thermo Scientific) conformément aux recommandations du fabricant. L'élution est réalisée avec 3 fois 20 µL d'une solution d'ACN à 80 %. Les extraits sont finalement évaporés à sec avant leur analyse en spectrométrie de masse.

Conditions HRMS :

Les extraits perméthylés sont repris dans 10 µL d'une solution de méthanol à 50 %. 0,5 µL de chaque échantillon sont déposés sur une plaque MALDI et mélangés avec 0,5 µL d'une solution d'acide 2,5-dihydroxybenzoic (10 mg/mL dans une solution à 50 % de méthanol contenant 10 mM d'acétate de

sodium). Les glycanes sont analysés sur un UltrafleXtreme (Bruker Daltonics, Bremen, Germany) en mode d'ionisation positif avec utilisation du reflectron.

Annexe 5 : Evaluation of High-Filed Orbitrap Fusion for Compound Annotation in Metabolomics, Supplementary Information

Pierre Barbier Saint Hilaire, Ulli M. Hohenester, Benoit Colsch, Jean-Claude Tabet, Christophe Junot, François Fenaille

Analytical Chemistry, 2018, 90, 3030-3035

SUPPLEMENTAL INFORMATION

Evaluation of the high-field Orbitrap Fusion for compound annotation in metabolomics

Pierre BARBIER SAINT HILAIRE¹, Ulli M. HOHENESTER¹, Benoit COLSCH¹, Jean-Claude TABET^{1,2}, Christophe JUNOT¹, François FENAILLE*¹

¹ Service de Pharmacologie et d'Immunoanalyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France.

² Sorbonne Universités, Campus Pierre et Marie Curie, IPCM, 4 place Jussieu, 75252 Paris Cedex 05, France.

* E-mail: francois.fenaille@cea.fr

Text S1. Experimental conditions

Table S1. Information on the studied set of 50 molecules

Table S2. Mass accuracy for monoisotopic and ¹³C ions in various conditions

Table S3. Mean absolute ¹³C RIA errors and mean ¹³C RIA errors in diverse conditions

Table S4. Recommended parameters for optimal measurement of ¹³C and ¹⁸O RIA on Orbitrap Fusion

Table S5. Mean absolute ¹⁸O RIA errors and mean ¹⁸O RIA errors in diverse conditions

Figure S1. Spectra of atropine for 5 different concentrations showing space charge effect

Figure S2. Comparison of monoisotopic to ¹³C delta on three different Orbitrap instruments

Figure S3. Example of calibration curves

Figure S4. Number of potential EC by compound

Text S1. Experimental conditions

Chemicals and Reagents. Chemical standards were from Sigma Aldrich (Saint Quentin Fallavier, France). Liquid chromatography solvents (acetonitrile, ACN and water) and extraction solvent (methanol) were LC-MS grade solvents (LC-MS Chromasolv, Sigma Aldrich). Formic acid was also from Sigma Aldrich. Serum for metabolite extraction was purchased from Biopredic (Rennes, France).

Preparation of standards and serum. A pool of 50 compounds (20 µg/mL each) was prepared from mother solutions at 1 mg/mL. The 50 compounds were chosen among the chemical library available in the laboratory^{1,2} according to their molecular mass and their detectability in the positive ElectroSpray Ionisation (ESI) mode (Table S1, Supporting Information). The pool was then serially diluted with a solution Water/ACN (95:5) containing 0.1% formic acid to give working solutions with concentrations ranging from 1 ng/mL to 10 µg/mL.

Serum metabolites were prepared using a dedicated protocol². Briefly, 100 µL of serum was mixed with 400 µL of methanol, and the resulting mixture was sonicated during 15 min and then centrifuged at 15,000 g for 15 min. After a 1.5 h of slow precipitation on ice, samples were centrifuged at 20,000 g for 15 min and the supernatant was withdrawn and dried under a nitrogen stream. Serum metabolites were resuspended in 1 mL of Water/ACN (95:5) with 0.1% formic acid containing either 10 µg/mL, 1µg/mL, 100 ng/mL, 10 ng/mL and 1 ng/mL of standard compounds or none of them.

LC-HRMS analysis. Chromatographic separation was performed on a Dionex Ultimate 3000 system (ThermoFisher Scientific, Les Ulis, France) equipped with a Hypersil Gold C₁₈

column (150 x 2.1 mm, 1.9 μm ; ThermoFisher Scientific). Mobile phases were 0.1% formic acid in water (v/v; A) and 0.1% formic acid in ACN (v/v; B). Flow rate was set at 500 $\mu\text{L}/\text{min}$. Solvent B was maintained at 5% for the first 2 min and then linearly increased to 100% over 11 min. These proportions were kept constant for 12.5 min before returning to 5% B in 4.5 min. Injection volume was set at 5 μL .

Mass analyses were performed on an Orbitrap Fusion (ThermoFisher Scientific) equipped with a Heated ElectroSpray Ionisation source. Source parameters were fixed as follows: spray voltage, + 3500 V; sheath gas, 50 arbitrary units; auxiliary gas, 12 arbitrary units; sweep gas, 2 arbitrary units; ion transfer tube temperature, 325°C; vaporizer temperature, 300°C. Ion transfer parameters applied were as follows: normal m/z range; quadrupole isolation set to on; scan range, 75-1000 m/z ; S lens RF level, 60%. Scan parameters applied were as follows: acquisition time, 0-13 min; ion polarity, positive; maximum injection time, 35 msec; data type, profile. AGC target (5e4, 1e5, 5e5) and resolving power (120,000, 240,000, and 500,000) was adapted to each experiment.

Data processing. Raw data files (.RAW) were manually processed using Xcalibur v.2.2.44 (ThermoFisher Scientific) processing methods. Genesis algorithm was used for detection (nearest RT, minimum peak height (S/N) = 3) and integration (smoothing point = 1, S/N threshold = 0.5, enable valley detection = no, constrain peak width = no). Accurate $[\text{M}+\text{H}]^+$ m/z , expected retention time (min) and retention time window (sec) are shown in supplementary data (Table S1, Supporting Information).

- (1) Aros-Calt, S.; Muller, B. H.; Boudah, S.; Ducruix, C.; Gervasi, G.; Junot, C.; Fenaille, F. *J. Proteome Res.* **2015**, *14* (11), 4863–4875.
- (2) Boudah, S.; Olivier, M.-F.; Aros-Calt, S.; Oliveira, L.; Fenaille, F.; Tabet, J.-C.; Junot, C. *J. Chromatogr. B* **2014**, *966*, 34–47.

Table S1. Information on the set of 50 molecules studied in the positive mode.

Compound	Chemical formula	Accurate mass	[M+H] ⁺	Retention time
2-Aminophenol	C6H7NO	109.05276	110.06004	1.04
4-Pyridylacetic acid	C7H7NO2	137.04768	138.05496	0.81
Triethanolamine	C6H15NO3	149.10519	150.11247	0.76
4-Methylumbelliferone	C10H8O3	176.04734	177.05462	6.51
6,7-Dihydroxy-4-methylcoumarin	C10H8O4	192.04226	193.04954	5.62
Capryloylglycine	C10H19NO3	201.13649	202.14377	7.62
Pantothenic acid	C9H17NO5	219.11067	220.11795	2.27
Flavone	C15H10O2	222.06808	223.07536	9.17
(R)-Butyryl carnitine	C11H21NO4	231.14706	232.15434	2.97
6-Hydroxymelatonin	C13H16N2O3	248.11609	249.12337	5.43
Dextrophan	C17H23NO	257.17796	258.18524	5.72
(±)-Propranolol	C16H21NO2	259.15723	260.16451	6.83
Formononetin	C16H12O4	268.07356	269.08084	8.16
Genistein (isoflavone)	C15H10O5	270.05282	271.06010	7.57
Dextromethorphan	C18H25NO	271.19361	272.20089	6.92
19-Nortestosterone	C18H26O2	274.19328	275.20056	8.51
D-Sphingosine (-H2O)	C18H35NO	281.27186	282.27914	10.02
Testosterone	C19H28O2	288.20893	289.21621	8.85
Atropine	C17H23NO3	289.16779	290.17507	5.46
(-)-Scopolamine	C17H21NO4	303.14706	304.15434	4.83
7 α -Hydroxytestosterone	C19H28O3	304.20384	305.21112	6.79
Decanoylcarnitine	C17H33NO4	315.24096	316.24824	8.32
Stanozolol	C21H32N2O	328.25146	329.25874	9.07
21-Deoxycortisol	C21H30O4	346.21441	347.22169	7.84
4-Methylumbelliferyl β -D-glucuronide	C16H16O9	352.07943	353.08671	5.21
Prednisone	C21H26O5	358.17802	359.18530	7.13
Curcumin	C21H20O6	368.12599	369.13327	9.21
Cholic acid (-2H2O)	C24H36O3	372.26644	373.27372	8.73
Finasteride	C23H36N2O2	372.27768	373.28496	9.10
Riboflavin	C17H20N4O6	376.13828	377.14556	5.17
trans-Zeatin glucoside	C16H23N5O6	381.16483	382.17211	2.68
Ochratoxin A	C20H18ClNO6	403.08226	404.08954	9.20
Lincomycin	C18H34N2O6S	406.21376	407.22104	4.77
Folic acid	C19H19N7O6	441.13968	442.14696	4.63
glycodeoxycholate	C26H43NO5	449.31410	450.32138	9.03
Psychosine	C24H47NO7	461.33525	462.34253	9.34
Quercetin-3-beta-D-glucoside	C21H20O12	464.09548	465.10276	5.86
Glycocholic acid	C26H43NO6	465.30905	466.31633	7.97
Folinic acid	C20H23N7O7	473.16590	474.17318	4.51
Deoxycorticosterone 21-glucoside	C27H40O8	492.27232	493.27960	7.44
Taurocholic acid (-H2O)	C26H43NO6S	497.28111	498.28839	7.53
Taurodeoxycholic acid	C26H45NO6S	499.29675	500.30403	8.43
D-Pantethine	C22H42N4O8S2	554.24441	555.25169	5.52
α -Ergocryptine	C32H41N5O5	575.31077	576.31805	7.30
Apigenin 7-O-neohesperidoside	C27H30O14	578.16356	579.17084	6.13
Naringin	C27H32O14	580.17920	581.18648	6.18
Stercobilin	C33H46N4O6	594.34174	595.34902	7.00
Rutin	C27H30O16	610.15338	611.16066	5.70
3'-Dephosphocoenzyme A	C21H35N7O13P2S	687.14888	688.15616	2.04
Angiotensin IV	C40H54N8O8	774.40646	775.41374	5.89

Table S2. Mass accuracy for monoisotopic ions and their corresponding ^{13}C ions in various experimental conditions. The table gathers the values obtained for the 50 molecules studied in the positive mode. The percentages refer to the number of signals with a mass error below the defined limit among the total number of signals.

		Monoisotopic ion				$^{13}\text{C}_1$ ion			
Resolution & AGC		Number of signals (I > 5e3)	Mean mass error (ppm)	Mass error < 2ppm	Mass error < 1ppm	Number of signals (I > 1e7)	Mean Mono to $^{13}\text{C}_1$ mass error (ppm)	Mono to $^{13}\text{C}_1$ mass error < 3 ppm	Mono to $^{13}\text{C}_1$ mass error < 1 ppm
Resolution	120,000 & 5e4	16 747	0.79	99.2%	71.1%				
	240,000 & 5e4	8238	0.77	99.6%	73.4%				
	500,000 & 5e4	3989	0.63	99.6%	83.6%	895	0.90	91.3%	73.3%
AGC Target	500,000 & 1e5	3912	0.77	99.5%	72.3%	775	1.75	79.7%	52.9%
	500,000 & 5e5	3739	3.00	10.1%	1.3%	702	3.63	54.9%	28.8%
Serum	500,000 & 5e4	4199	0.83	99.5%	69.5%	809	0.84	91.8%	75.8%

Table S3. Mean absolute ^{13}C RIA error values and mean ^{13}C RIA error values for low and high intensity signals in diverse conditions. Mean absolute RIA error values are calculated as the mean of the absolute values of the RIA errors for the n signals whereas mean RIA errors are calculated as the mean of positive and negative values of the RIA errors for the n signals.

		Intensity < 5e5	Intensity > 5e5
Solvent Resolution 120,000 AGC 5e4	Number of signals	210	399
	Mean absolute RIA error (Mean RIA error)	13.2% (1.2%)	2.8% (1.9%)
Solvent Resolution 240,000 AGC 5e4	Number of signals	203	395
	Mean absolute RIA error (Mean RIA error)	16.2% (-3.2%)	2.8% (1.3%)
Solvent Resolution 500,000 AGC 5e4	Number of signals	180	389
	Mean absolute RIA error (Mean RIA error)	24.7% (-18.2%)	3.3% (1.0%)
Serum Resolution 500,000 AGC 5e4	Number of signals	154	375
	Mean absolute RIA error (Mean RIA error)	25.7% (-19.1%)	3.9% (0.9%)

Table S4. Recommended parameters for optimal measurements of ^{13}C and ^{18}O RIA on the Orbitrap Fusion

Mass range	85<m/z<250	250<m/z<350	350<m/z<500	500<m/z
Resolution	120,000	240,000	500,000	N/A
AGC Target	5e4			
Intensity	5e5 – 1e8			

Table S5. Mean absolute ^{18}O RIA error values and mean ^{18}O RIA errors for all and high intensity signals in diverse conditions. Mean absolute RIA error values are calculated as the mean of the absolute values of the RIA errors for the n signals whereas mean RIA errors are calculated as the mean of positive and negative values of the RIA errors for the n signals.

		^{18}O	
		All intensities	Intensities > 10^7
Solvent Resolution 240,000	Number of signals	216	135
	Mean absolute RIA error (Mean RIA error)	24.5 % (-20.6 %)	24.7 % (-24.2 %)
Solvent Resolution 500,000	Number of signals	275	166
	Mean absolute RIA error (Mean RIA error)	37.7 % (-33.0 %)	34.8 % (-29.3 %)
Serum Resolution 500,000	Number of signals	261	133
	Mean absolute RIA error (Mean RIA error)	38.5 % (-35.0 %)	32.9 % (-31.3 %)

Figure S1

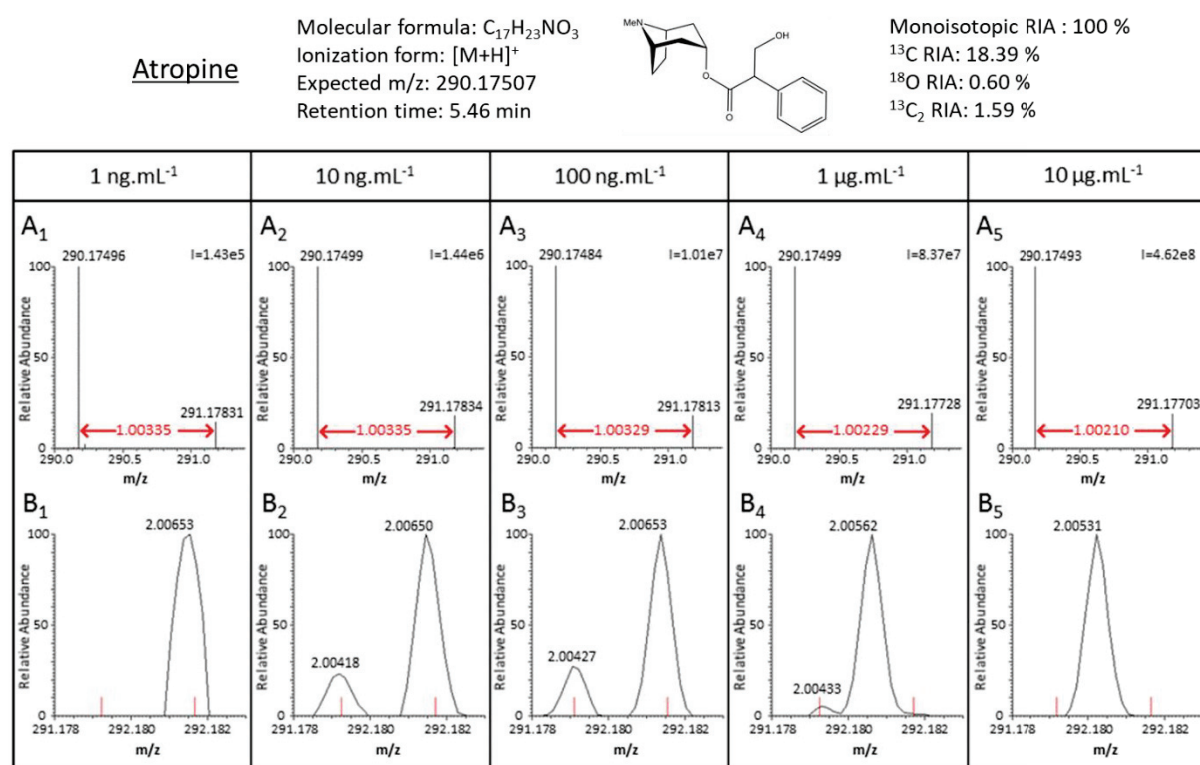


Figure S1. Mass spectra of atropine (expected theoretical m/z 290.17507) for 5 different concentrations showing space charge effect. A: Monoisotopic to ^{13}C delta decreasing while intensity increases (expected value: 1.00335). B: M+2 isotopic patterns. Masses are relative to the monoisotopic ones. Theoretical positions are shown in red (2.00425 for ^{18}O and 2.00671 for $^{13}C_2$). B₁: for 1 ng/mL, only $^{13}C_2$ signal is present. B₂-B₃: for 10 ng/mL and 100 ng/mL, signals of both ^{18}O and $^{13}C_2$ are present and well resolve. B₄-B₅: for 1 µg/mL and 10 µg/mL, coalescence of peaks is observed. All spectra are from the apex of each chromatographic peak. Resolution is set to 500,000 and AGC target to 5e4. I: Intensity.

Figure S2

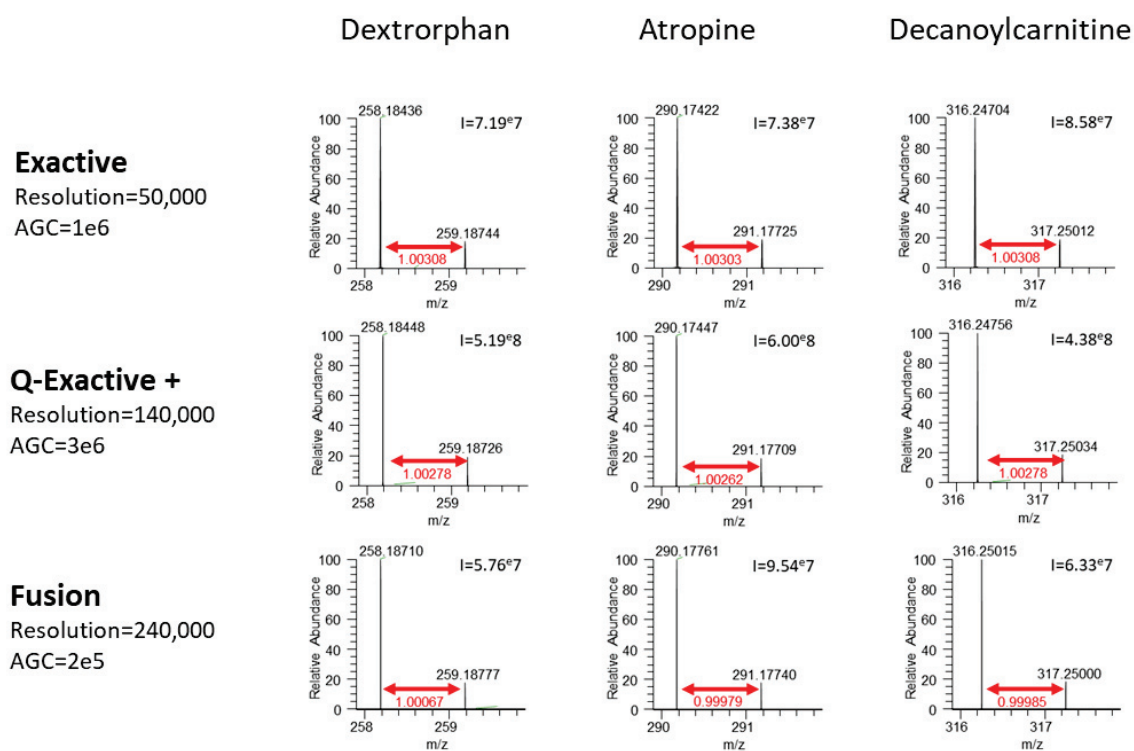


Figure S2. Comparison of monoisotopic ion to $^{13}\text{C}_1$ mass differences on three different Orbitrap instruments. The three standards are analyzed under the same HPLC conditions and injected at the same concentration. I: Intensity

Figure S3

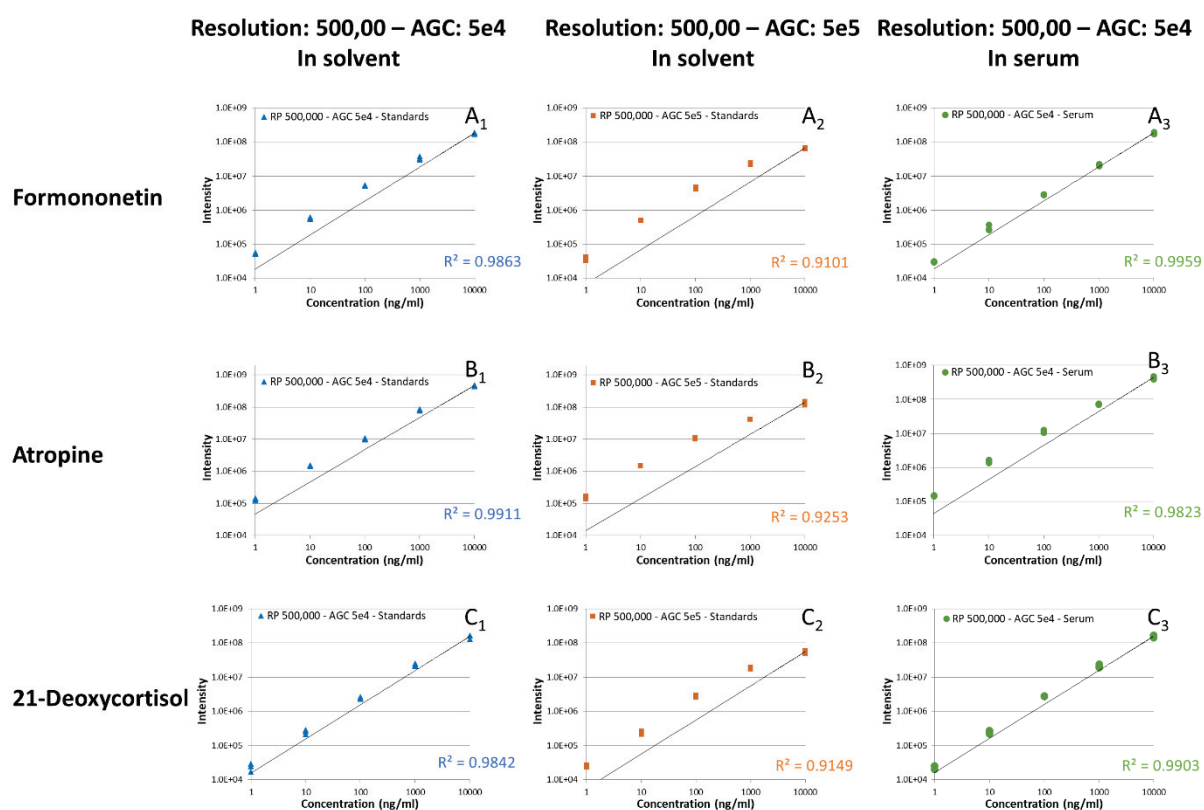


Figure S3. Examples of calibration curves obtained for three compounds (A, Formononetin; B, Atropine; C, 21-Doexycortisol) in diverse conditions (₁: resolution 500,000, AGC 5e4, in solvent; ₂: resolution 500,000, AGC 5e5, in solvent; ₃: resolution 500,000, AGC 5e4, in serum). Concentrations ranged from 1 to 10,000 ng/mL.

Figure S4

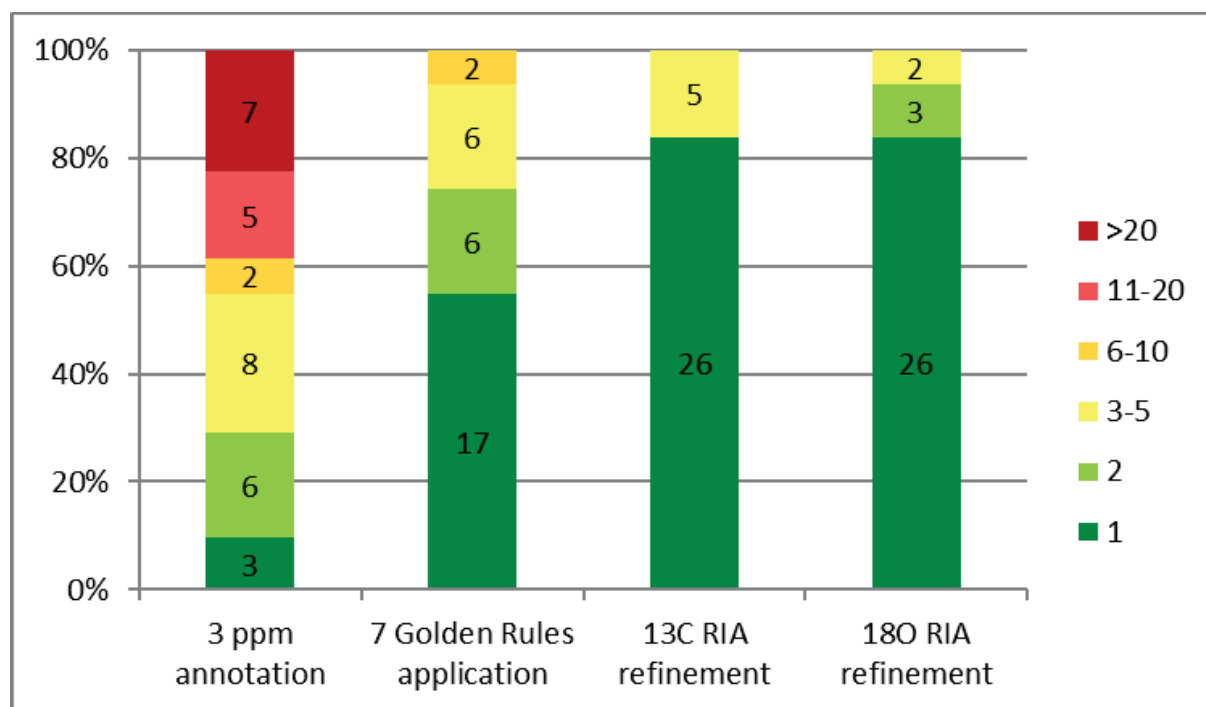


Figure S4. Number of potential ECs per compound. This is shown for the 31 molecules spiked in a metabolic plasma extract at 1 $\mu\text{g/mL}$ and with detected ^{18}O signals. First column represents the number of potential ECs using only the accurate mass measurement with a 3 ppm tolerance. Potential numbers of atoms were as follows $\text{C}_{0-50}\text{H}_{0-75}\text{N}_{0-10}\text{O}_{0-15}\text{S}_{0-3}\text{P}_{0-3}$. For the second column, the seven golden rules as defined by Fiehn and Kind³ were considered. For the third column, ^{13}C RIA measurements were taken into account; all ECs that did not correspond to the measured RIA (with a -10% +10% tolerance) were rejected. For the fourth column, ^{18}O RIA measurement was taken into account; all EC that did not correspond to the measured RIA (with a -75% +10% tolerance) were rejected.

(3) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2007**, *8*, 105.

**Annexe 6 : Mechanistic study of competitive releases of H₂O, NH₃ and CO₂
from deprotonated aspartic and glutamic acids: Role of conformation,
Supplementary Information**

Pierre Barbier Saint Hilaire, Anna Warnet, Yves Gimbert, Ulli M. Hohenester, Gianluca Giorgi,
Marie-Françoise Olivier, François Fenaille, Benoît Colsch, Christophe Junot, Jean-Claude Tabet

Journal of Chromathography B, 2017, 1047, 64-74

Figure S1.

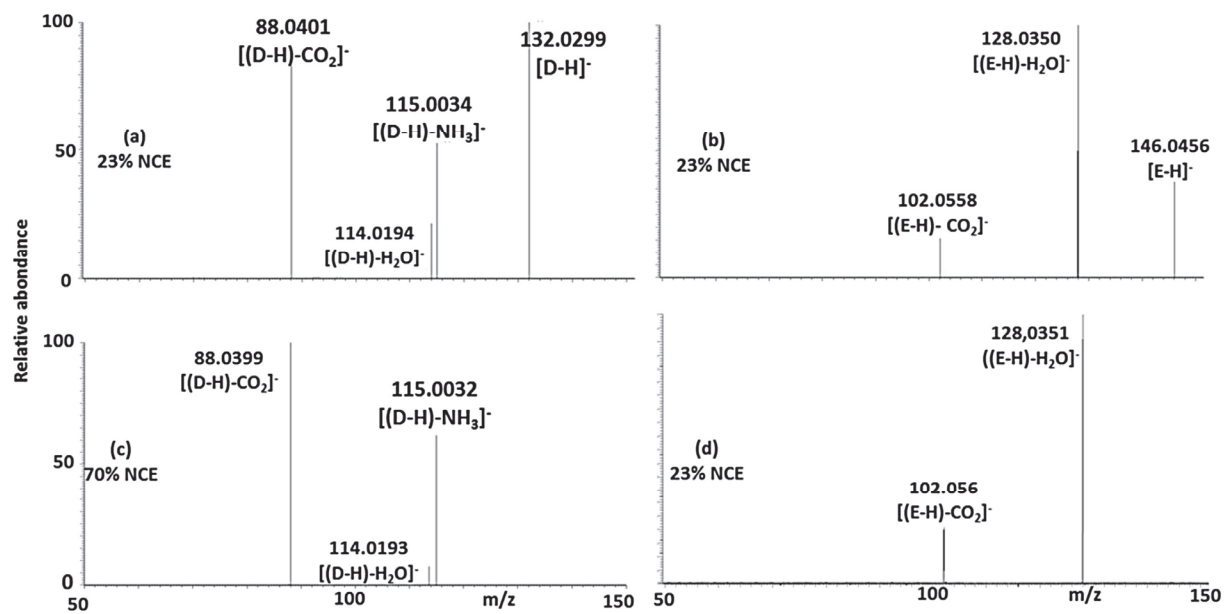


Figure S2.

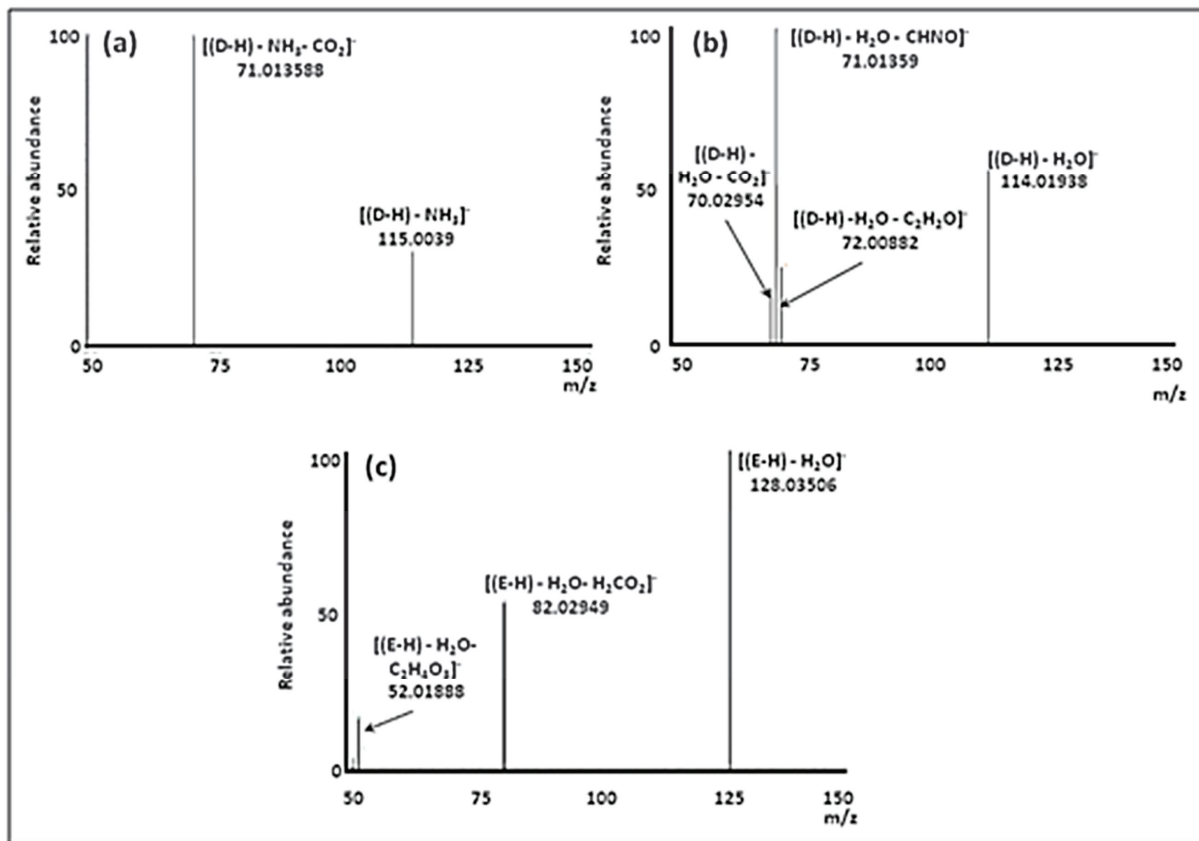


Figure S3.

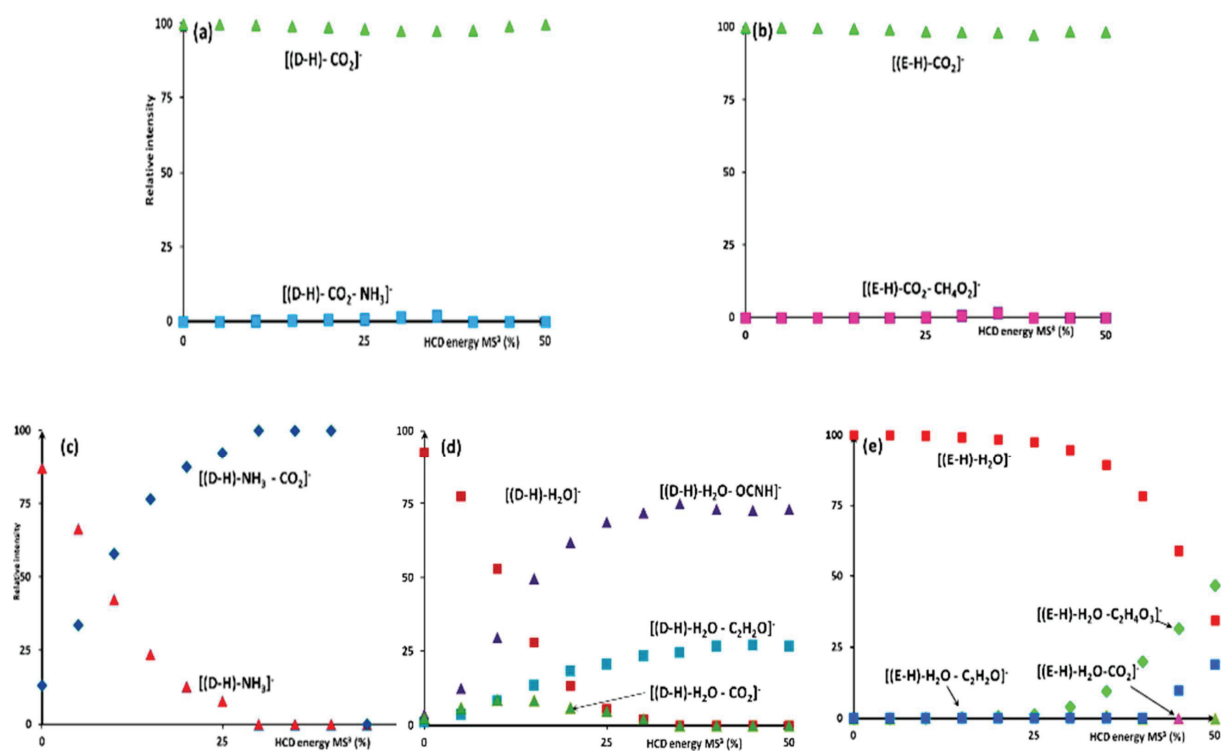


Figure S4.

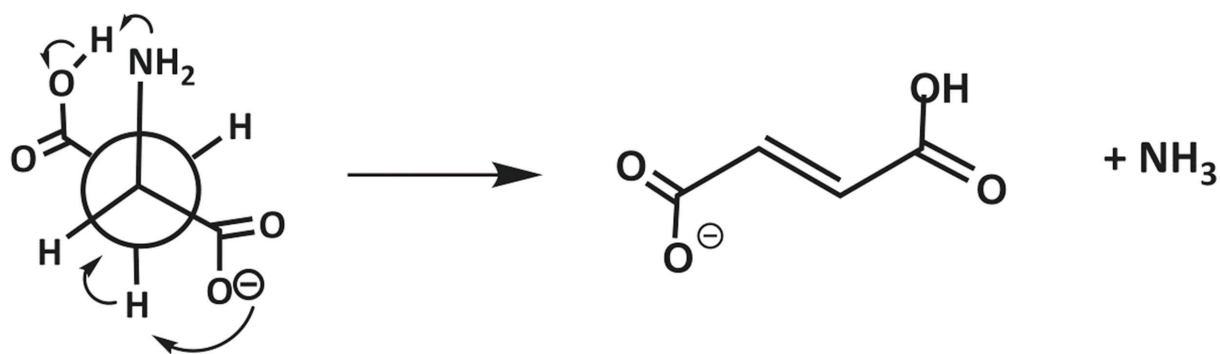


Table S1. High resolution measurements by FT/MS of main product ions* generated under HCD mode conditions from (a) the deprotonated $[D-H]^-$ (m/z 132.0300) and $[E-H]^-$ (m/z 146.0456) amino-acids in MS/MS and (b) the following product ions: $[(D-H)-NH_3]^-$ (m/z 115.0034), $[(D-H)-H_2O]^-$ (m/z 114.0194) for aspartate and $[(E-H)-H_2O]^-$ (m/z 128.0351) for glutamate in sequential MS³ experiments the consecutive losses from the CO₂ release are less 5% and thus, too weak to be reported).

Selected precursor ions from the ESI ion source (sequential MS ² experiment)	Product ions				
	Selected product ions (sequential MS ³ experiments)	m/z values (experimental)	Elemental composition	m/z values (calculated.)	Formal loss from the selected parent (a) or the product ion (b)
$[D-H]^{-a)}$		132.02998	C ₄ H ₆ NO ₄	132.02968	
		115.0034	C ₄ H ₃ O ₄	115.00313	NH ₃
		114.01943	C ₄ H ₄ NO ₃	114.01912	H ₂ O
		88.04015	C ₃ H ₆ NO ₂	88.03985	CO ₂
		72.00883	C ₂ H ₂ NO ₂	72.00855	C ₂ H ₄ O ₂
		71.01354	C ₃ H ₃ O ₂	71.01330	CH ₃ NO ₂
		59.01357	C ₂ H ₃ O ₂	59.01330	C ₂ H ₃ NO ₂
	$[D-H]^-$				
$[D-H]^-$	$[(D-H)-NH_3]^{-b)}$	115.0033	C ₄ H ₃ O ₄	115.00313	
		71.01358	C ₃ H ₃ O ₂	71.01330	CO ₂
	$[D-H]^-$				
$[D-H]^-$	$[(D-H)-H_2O]^{-b)}$	114.01937	C ₄ H ₄ NO ₃	114.01912	
		72.00882	C ₂ H ₂ NO ₂	72.00855	C ₂ H ₂ O
		71.01358	C ₃ H ₃ O ₂	71.01330	CHNO
		70.02956	C ₃ H ₄ NO	70.02929	CO ₂
	$[D-H]^-$				
Selected precursor ions from the ESI ion source (sequential MS ² experiment)	Product ions				
	Selected product ions (sequential MS ³ experiment)	m/z values (experimental)	Elemental composition	m/z values (calculated)	Formal loss from the selected parent (a) or the product ion (b)
$[E-H]^{-a)}$		146.04564	C ₅ H ₈ NO ₄	146.04533	
		128.03512	C ₅ H ₆ NO ₃	128.03477	H ₂ O
		102.05581	C ₄ H ₈ NO ₂	102.05550	CO ₂
		85.02927	C ₄ H ₅ O ₂	85.02895	CH ₃ NO ₂
		84.04523	C ₄ H ₆ NO	84.04494	CH ₂ O ₃
		82.02957	C ₄ H ₄ NO	82.02929	CH ₄ O ₃
		74.02450	C ₂ H ₄ NO ₂	74.02420	C ₃ H ₄ O ₂
		71.01346	C ₃ H ₃ O ₂	71.01330	C ₂ H ₅ NO ₂
		59.01359	C ₂ H ₃ O ₂	59.01330	C ₃ H ₅ NO ₂
	$[E-H]^-$				
$[E-H]^-$	$[(E-H)-H_2O]^{-b)}$	128.03501	C ₅ H ₆ NO ₃	128.03477	
		84.04518	C ₄ H ₆ NO	84.04494	CO ₂
		82.02951	C ₄ H ₄ NO	82.02929	H ₂ CO ₂
		52.01888	C ₃ H ₂ N	52.01872	C ₂ H ₄ O ₃

*Signals provided in the ERMS experiments (see later).

Table S2. Enthalpies and Gibbs free energies (kJ/mol) for intermediates and transition states at B3LYP/6-31+G(d,p) level described in the concomitant loss of NH₃ and CO₂ pathway, starting from [(D-H)]- taken as 0 energy.

B3LYP/6-31+G(d,p)	(a) [(D-H)]-	TS _{H transfer}	(b)	TS _{NH₃}	[(D-H)-(NH ₃ +CO ₂)] ⁻
ΔH(kJ/mol)		24.66	17.55	144.21	-15.71
ΔG (kJ/mol)		25.91	17.55	138.36	-18.39

(a) [(D-H)]⁻			TS_{H transfer}				
C	-0.519316	0.137798	-0.467734	C	0.486069	-0.018995	0.534267
H	-0.399048	-0.091528	-1.532231	H	0.408222	-0.307216	1.587457
C	0.526031	-0.666245	0.337019	C	-0.594308	-0.749780	-0.291511
H	0.403764	-1.731561	0.137719	H	-0.569715	-1.811771	-0.042473
H	0.328993	-0.511770	1.408587	H	-0.344619	-0.654831	-1.357564
N	-0.292633	1.595832	-0.256407	N	0.356757	1.461186	0.360958
H	-0.626986	2.103809	-1.072393	H	-0.637598	1.617637	0.000888
O	-0.955966	1.829843	0.494728	H	0.467373	1.949528	1.246148
C	-1.966196	-0.286839	-0.017618	C	1.925194	-0.315657	0.009840
O	-2.539963	0.537944	0.750036	O	2.457768	-1.414060	0.156286
O	-2.361366	-1.394925	-0.446623	O	2.412479	0.740216	-0.582763
C	1.989628	-0.292445	0.100375	C	-2.042702	-0.192379	-0.110219
O	2.893648	-1.112906	0.087006	O	-2.960388	-1.014609	0.065480
O	2.231865	1.017925	-0.049753	O	-2.146051	1.083683	-0.202518
H	1.323316	1.492463	-0.079139	H	1.423056	1.477369	-0.307304
(b)			TS_{NH₃}				
C	-0.518965	0.137667	-0.467058	C	-0.565175	-0.544918	-0.398107
H	-0.398012	-0.091805	-1.531433	H	-0.405515	-0.634285	-1.468156
C	0.525991	-0.665586	0.338735	C	0.424738	-1.088023	0.476589
H	0.403096	-1.731164	0.141227	H	0.703951	-2.108596	0.203502
H	0.329284	-0.508928	1.410040	H	0.187314	-0.988639	1.535566
N	-0.292299	1.596049	-0.256139	N	0.160912	1.993613	-0.606826
H	1.323502	1.492279	-0.079781	H	0.978032	2.017784	-0.001729
H	-0.628509	2.103523	-1.071694	H	0.235645	2.702486	-1.328621
C	-1.966189	-0.286856	-0.017874	C	-1.940361	-0.136443	0.041122
O	-2.361217	-1.394687	-0.447857	O	-2.592992	-1.186389	-0.239673
O	-2.540366	0.537451	0.749959	O	-2.279987	0.933238	0.567780
C	1.989531	-0.292619	0.100496	C	1.941142	-0.339848	0.147014
O	2.893396	-1.113177	0.086764	O	2.429359	-0.716550	-0.919506
O	2.231883	1.017717	-0.050974	O	2.281512	0.419171	1.062236
H	-0.955037	1.829683	0.495694	H	-0.691003	2.115589	-0.059190
[(D-H)-(NH₃+CO₂)]⁻							
C	3.195268	0.111701	0.065143				
H	3.817537	1.005208	0.074897				
C	3.733258	-1.109337	0.155874				
H	4.807641	-1.259099	0.242873				
H	3.090238	-1.985176	0.145402				
N	-1.428066	0.500069	-0.291085				
H	-1.234619	1.314341	0.291915				
H	-0.537102	-0.117087	-0.172104				
C	1.707648	0.373758	-0.056857				
O	0.936983	-0.646244	-0.073962				
O	1.351911	1.578084	-0.132499				
C	-2.832750	-0.179877	0.080882				
O	-3.101306	-0.041480	1.274334				
O	-3.353865	-0.693777	-0.911441				
H	-1.437559	0.791207	-1.267094				

**Annexe 7 : Evaluation of Data Dependent and Data Independent
Acquisition workflows on an Orbitrap Fusion for metabolomics applications,
Supplementary Information**

Pierre Barbier Saint Hilaire, Kathleen Rousseau, Annelaure Damont, Christophe Junot,
François Fenaille

Draft version

Evaluation of Data Dependent and Data Independent Acquisition

Workflows on an Orbitrap Fusion for Metabolomics applications

Pierre BARBIER SAINT HILAIRE¹, Kathleen ROUSSEAU¹, Annelaure DAMONT¹, Christophe JUNOT¹, François FENAILLE^{*,1}

¹ Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, F-91191 Gif-sur-Yvette, France.

* E-mail: francois.fenaille@cea.fr

Table S1. List of our standard set of 50 compounds

Table S2. List of the 72 metabolites annotated in plasma NIST in C18 LC condition and ESI+ ionization

Table S3. Example of exclusion list used for filtering mobile phase ions for DDA application

Figure S1. Distribution of optimal collision energy for 482 standard metabolites

Figure S2. Mass distribution of metabolites identified in plasma.

Figure S3. Example of calibration curves for Propranolol and Lincomycin in plasma.

Table S1. List of our standard set of 50 compounds

Compound	Composition	Exact Mass	[(M+H)]+	Retention Time (min)
Triethanolamine	C6H15NO3	149.10519	150.11247	0.75
4-Pyridylacetic acid	C7H7NO2	137.04768	138.05496	0.83
Nicotinic acid	C6H5NO2	123.03203	124.03931	0.93
2-Aminophenol	C6H7NO	109.05276	110.06004	1.12
Isoleucine	C6H13NO2	131.09463	132.10191	1.26
Ala-Tyr	C12H16N2O4	252.11101	253.11829	1.68
3'-Dephosphocoenzyme A	C21H35N7O13P2S	687.14888	688.15616	2.05
Pantothenic acid	C9H17NO5	219.11067	220.11795	2.42
trans-Zeatin glucoside	C16H23N5O6	381.16483	382.17211	2.93
Folic acid	C19H19N7O6	441.13968	442.14696	4.64
Lincomycin	C18H34N2O6S	406.21376	407.22104	4.88
(-)-Scopolamine	C17H21NO4	303.14706	304.15434	5.00
Riboflavin	C17H20N4O6	376.13828	377.14556	5.19
6-Hydroxymelatonin	C13H16N2O3	248.11609	249.12337	5.40
D-Pantethine	C22H42N4O8S2	554.24441	555.25169	5.55
Atropine	C17H23NO3	289.16779	290.17507	5.60
6,7-Dihydroxy-4-methylcoumarin	C10H8O4	192.04226	193.04954	5.66
Rutin	C27H30O16	610.15338	611.16066	5.73
p-Coumaric acid	C9H8O3	164.04734	165.05462	5.77
Dextrorphan	C17H23NO	257.17796	258.18524	5.89
Apigenin 7-O-neohesperidoside	C27H30O14	578.16356	579.17084	6.18
Naringin	C27H32O14	580.17920	581.18648	6.24
4-Methylumbelliferone	C10H8O3	176.04734	177.05462	6.60
7 α -Hydroxytestosterone	C19H28O3	304.20384	305.21112	6.86
(\pm)-Propranolol	C16H21NO2	259.15723	260.16451	7.10
Dextromethorphan	C18H25NO	271.19361	272.20089	7.20
Prednisone	C21H26O5	358.17802	359.18530	7.24
Taurocholic acid (-H2O)	C26H43NO6S	497.28111	516.29900	7.46
Deoxycorticosterone 21-glucoside	C27H40O8	492.27232	493.27960	7.57
α -Ergocryptine	C32H41N5O5	575.31077	576.31805	7.59
Capryloylglycine	C10H19NO3	201.13649	202.14377	7.78
21-Deoxycortisol	C21H30O4	346.21441	347.22169	7.99
Glycocholic acid	C26H43NO6	465.30905	466.31633	8.12
Taurodeoxycholic acid	C26H45NO6S	499.29675	500.30403	8.28
Formononetin	C16H12O4	268.07356	269.08084	8.33
Dodecanedioic acid	C12H22O4	230.15181	231.15909	8.34
19-Nortestosterone	C18H26O2	274.19328	275.20056	8.65
Cholic acid (-2H2O)	C24H36O3	372.26644	373.27372	8.89
Testosterone	C19H28O2	288.20893	289.21621	9.01
glycodeoxycholate	C26H43NO5	449.31410	450.32138	9.19
Stanozolol	C21H32N2O	328.25146	329.25874	9.22
Finasteride	C23H36N2O2	372.27768	373.28496	9.26
Flavone	C15H10O2	222.06808	223.07536	9.35
Ochratoxin A	C20H18ClNO6	403.08226	404.08954	9.39
Curcumin	C21H20O6	368.12599	369.13327	9.39
Psychosine	C24H47NO7	461.33525	462.34253	9.84
D-Sphingosine (-H2O)	C18H35NO	281.27186	300.28970	10.63
Pyrogallol	C6H9O3	129.05517	130.06245	
1,7-Dimethyluric acid	C7H8N4O3	196.05960	197.06690	
Stercobilin	C33H46N4O6	594.34174	595.34902	

Table S2. List of the 72 metabolites annotated in plasma NIST in C18 LC condition and ESI+ ionization

Name	Formule	Exact Mass	Retention Time (min)
Uracil	C4H4N2O2	112.02728	1.06
Creatinine	C4H7N3O	113.05891	0.83
Proline	C5H9NO2	115.06333	0.85
Betaine	C5H11NO2	117.07898	0.82
Valine	C5H11NO2	117.07898	0.93
Threonine / D-allo-Threonine	C4H9NO3	119.05824	0.84
Nicotinamide	C6H6N2O	122.04801	0.99
Pyroglutamic-acid	C5H7NO3	129.04259	1.08
Pipecolinic-acid	C6H11NO2	129.07898	0.94
Creatine	C4H9N3O2	131.06948	0.84
Isoleucine	C6H13NO2	131.09463	1.28
Leucine	C6H13NO2	131.09463	1.38
5-Hydroxyindole	C8H7NO	133.05276	4.92
Hypoxanthine	C5H4N4O	136.03851	0.99
Trigonelline	C7H7NO2	137.04768	0.84
4-Imidazoleacrylic acid	C6H6N2O2	138.04293	0.93
Methylimidazoleacetic-acid	C6H8N2O2	140.05858	0.86
Stachydrine	C7H13NO2	143.09463	0.85
4-Guanidinobutyric-acid	C5H11N3O2	145.08513	0.93
Glutamine	C5H10N2O3	146.06914	0.81
Lysine	C6H14N2O2	146.10553	0.72
L-Glutamic-acid	C5H9NO4	147.05316	0.82
Methionine	C5H11NO2S	149.05105	1.04
Acetaminophen-(4-Acetamidophenol)	C8H9NO2	151.06333	2.88
Histidine	C6H9N3O2	155.06948	0.76
Carnitine	C7H15NO3	161.10519	0.82
1-methyl-guanine / 7-methylguanine	C6H7N5O	165.06506	0.97
Phenylalanine	C9H11NO2	165.07898	2.10
7-Methylxanthine	C6H6N4O2	166.04908	1.45
3-Methylxanthine	C6H6N4O2	166.04908	1.60
1-Methylxanthine	C6H6N4O2	166.04908	1.74
Uric-acid	C5H4N4O3	168.02834	0.94
1-Methylhistidine	C7H11N3O2	169.08513	0.78
Arginine	C6H14N4O2	174.11168	0.77
L-Citrulline	C6H13N3O3	175.09569	0.81
Cotinine	C10H12N2O	176.09496	1.08
Hippuric-acid	C9H9NO3	179.05824	4.93
paraxanthine / Theophylline	C7H8N4O2	180.06473	3.95
Theobromine	C7H8N4O2	180.06473	2.52
Tyrosine	C9H11NO3	181.07389	1.16
3-Amino-3-(4-hydroxyphenyl)propanoic	C9H11NO3	181.07389	1.17
1-Methyluric-acid	C6H6N4O3	182.04399	1.46
4-Pyridoxic-acid	C8H9NO4	183.05316	1.37
N-acetyl-L-glutamine	C7H12N2O4	188.07971	0.94
N6-Acetyl-L-lysine	C8H16N2O3	188.11609	0.90
N6,N6,N6-Trimethyl-L-lysine	C9H20N2O2	188.15248	0.76
Trans-3-Hydroxy-cotinine	C10H12N2O2	192.08988	0.91
(S)-Cotinine-N-oxide	C10H12N2O2	192.08988	1.52
Caffeine	C8H10N4O2	194.08038	4.89
Acetyl-L-carnitin	C9H17NO4	203.11576	1.01
Tryptophan	C11H12N2O2	204.08988	4.24
Panthenol	C9H19NO4	205.13141	2.48
L-Kynurenine	C10H12N2O3	208.08479	2.11
N-alpha-acetyl-L-arginine	C8H16N4O3	216.12224	0.90

Propionylcarnitine	C10H19NO4	217.13141	1.54
Pantothenic-acid	C9H17NO5	219.11067	2.45
(R)-Butyryl-carnitine	C11H21NO4	231.14706	3.48
L-Cystine	C6H12N2O4S2	240.02385	0.84
L-a-Glycerophosphorylcholine	C8H20NO6P	257.10282	0.77
Hexanoylcarnitine	C13H25NO4	259.17836	6.19
Phenylacetyl-L-glutamine	C13H16N2O4	264.11101	5.07
N-Acetyl-L-carnosine	C11H16N4O4	268.11715	0.88
1-Methyladenosine	C11H15N5O4	281.11240	0.92
Octanoylcarnitine	C15H29NO4	287.20966	7.63
5-Deoxy-5-(methylthio)adenosine	C11H15N5O3S	297.08956	4.16
Decanoylcarnitine	C17H33NO4	315.24096	8.79
Acetaminophen-glucuronide	C14H17NO8	327.09542	1.56
Sphingosine-1-phosphate	C18H38NO5P	379.24876	10.24
glycochenodeoxycholic-acid	C26H43NO5	449.31412	9.12
glycodeoxycholate	C26H43NO5	449.31412	9.29
Glycocholic-acid	C26H43NO6	465.30904	8.21
Stercobilin	C33H46N4O6	594.34173	7.16

Table S3. Example of exclusion list used for filtering mobile phase ions for DDA application

m/z	t start (min)	t stop (min)
90.97677	5	20
99.51234	8	14
103.95569	8	14
113.96374	0	14
114.0914	0	14
116.97196	8	14
121.9662	8	20
123.96438	8	20
125.98627	5	8
130.15904	0	14
139.98788	8	14
144.98213	8	20
146.98031	8	20
149.02328	0	14
150.02661	0	5
158.15388	0	5
158.96401	5	20
189.05138	0	8
194.11747	0	14
195.08755	0	5
199.18038	0	14
220.93438	5	14
226.95133	0.8	20
245.11387	0	5
268.24527	0	0.8
274.27385	8	14
279.15883	0	0.8
282.22136	0	14
288.9217	0	14
301.14073	0	5
306.89443	8	14
362.92603	5	20
415.21113	8	14
424.89646	5	14
427.37781	0	5
430.91334	5	20
476.30604	5	8
492.88401	8	14
498.90092	8	20
566.88831	8	20
634.87566	8	20
702.86305	8	20
703.57444	8	14
758.56899	8	14
759.57238	8	14
760.58458	8	14
761.58797	8	14
770.85044	8	20
782.56788	8	20
783.57138	8	14
784.58444	8	14
786.60024	8	14
787.60364	8	14
804.55078	8	14

806.56801	8	14
808.5826	8	14
810.60005	8	14
265.96213	14	20
838.83751	14	20

Figure S1.

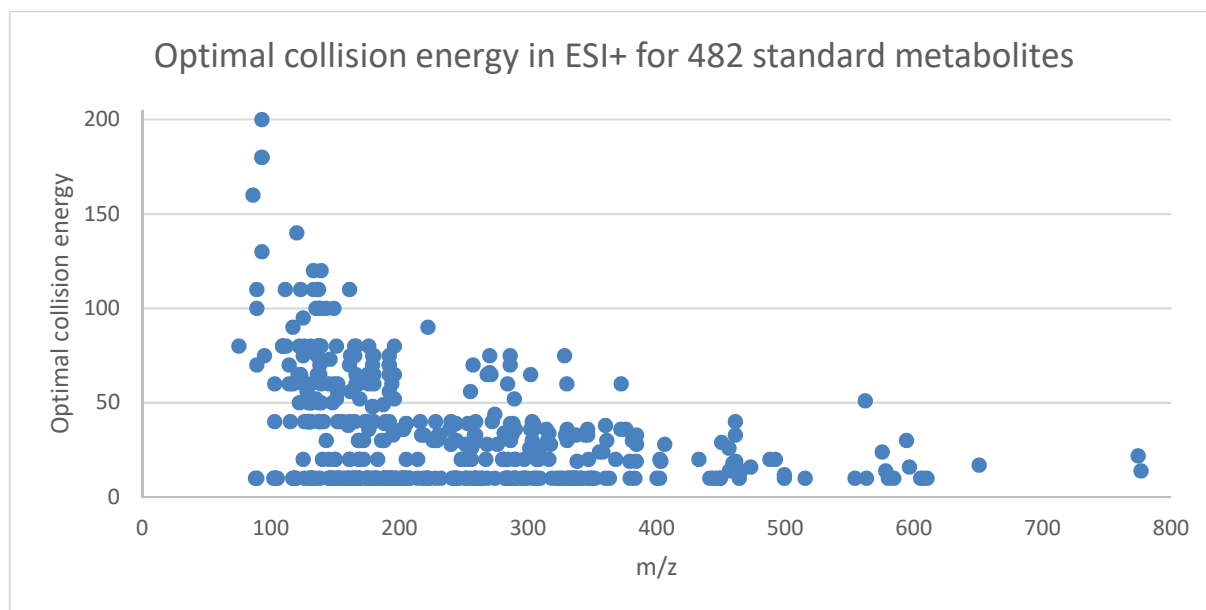


Figure S1. Distribution of optimal collision energy for 482 standard metabolites.

These energies were determined in ESI+ on a Q-Exactive instrument. Collision energy was considered as optimal if the parent ion presents a relative abundance between 20 and 30%.

These energies were determined in ESI+ on a Q-Exactive instrument. Collision energy was considered as optimal if the parent ion presents a relative abundance between 20 and 30%.

Figure S2.

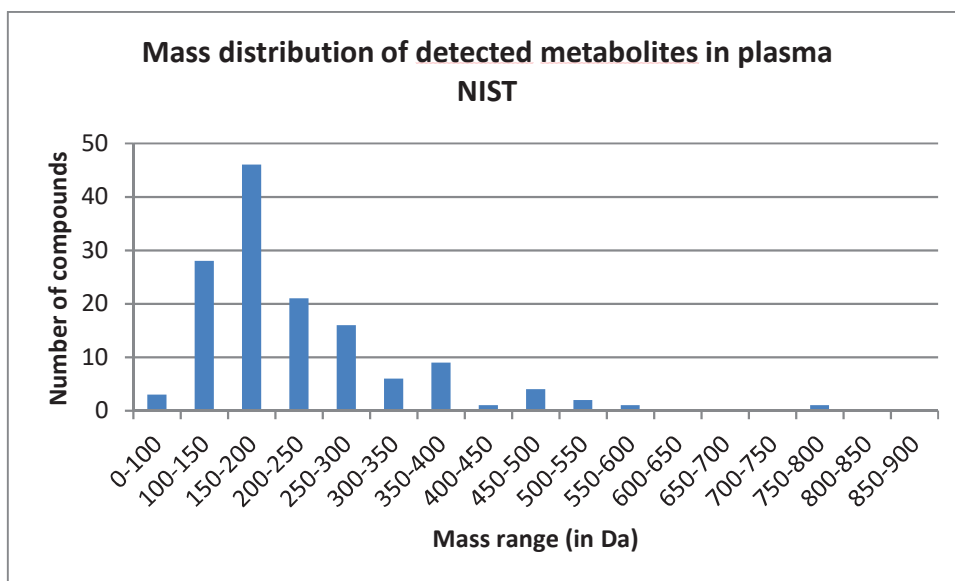


Figure S2. Mass distribution of metabolites identified in plasma.

Figure S3.

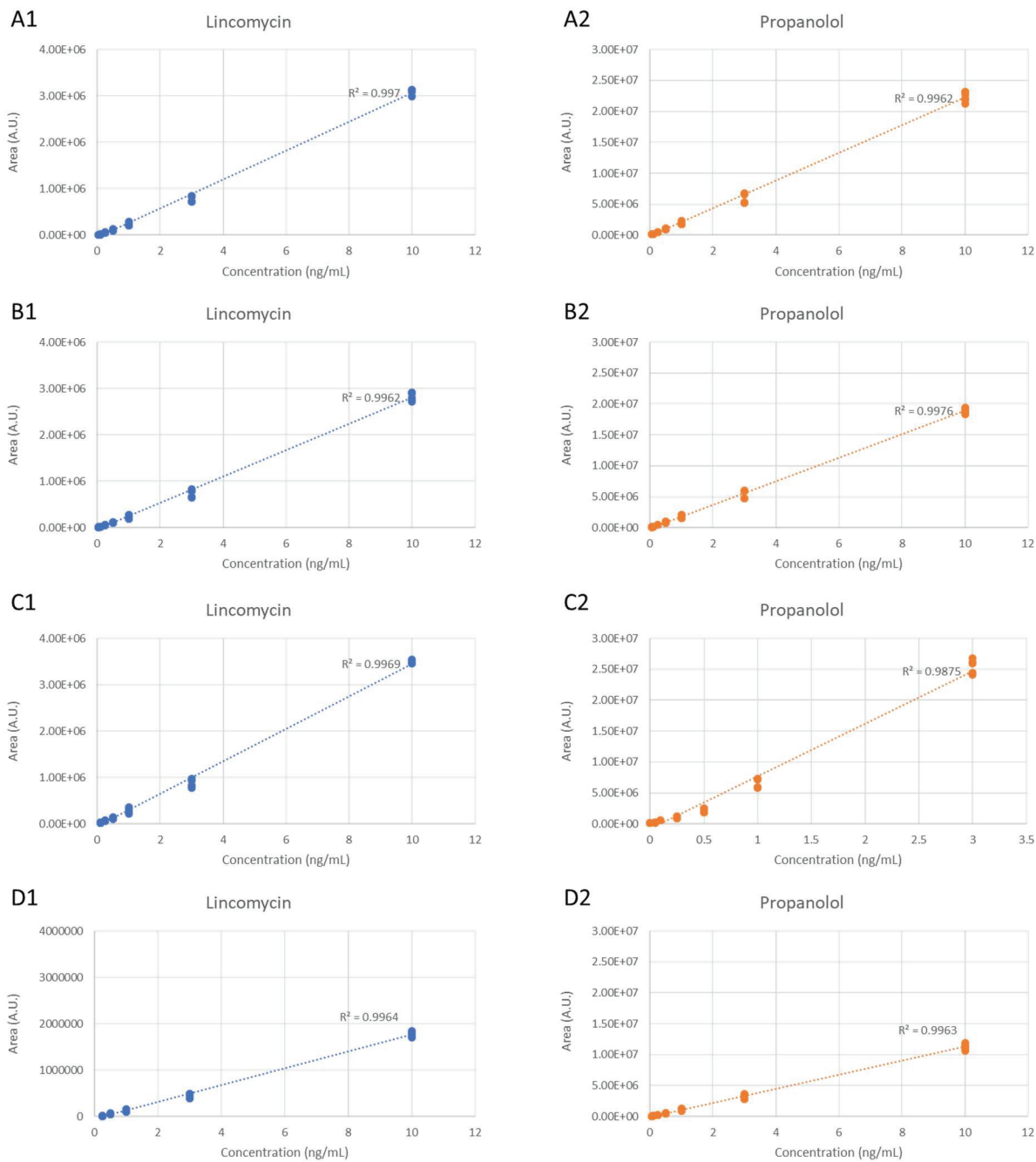


Figure S3. Example of calibration curves for Propranolol and Lincomycin in plasma.

Standards of these two compounds were spiked in plasma. No signal corresponding to these compounds was observed in blank plasma. A. Quantification with Full MS acquisition protocol. Limit of quantification: 0.05 for Propranolol and Lincomycin. B. Quantification with DIA acquisition protocol. Limit of quantification: 0.05 and 0.25 for Propranolol and Lincomycin, respectively.

Titre : Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse.

Mots clés : Métabolomique, Spectrométrie de masse, Identification, Cohortes

La métabolomique non ciblée est une approche visant à caractériser le contenu métabolique d'un échantillon. Cela permet d'évaluer les changements phénotypiques ayant lieu au sein d'échantillons biologiques. Au cours de cette thèse, une cohorte de patients souffrant d'encéphalopathie hépatique (EH) a été étudiée en métabolomique, en association avec de la lipidomique et de la glycomique. L'EH est une maladie dont les mécanismes sont encore mal compris mais avec un impact important médicale et sociétal. Ces travaux ont montré que l'EH pouvait être liée à une altération du métabolisme énergétique observable dans le sang et dans le liquide céphalorachidien. Face aux limites mises en exergue lors de cette étude, la suite de ces travaux a visé à développer de nouvelles approches analytiques en métabolomique.

Nous ainsi avons montré que l'utilisation de nouveaux instruments de type Orbitrap-Fusion, permettant d'atteindre de très hautes résolutions, donne accès à la structure isotopique fine des métabolites ce qui facilite leur identification. L'étude de la fragmentation MS² des ions a permis de mieux comprendre des mécanismes mis en jeu. Enfin, nous avons montré que les améliorations apportées aux instruments, notamment en termes de vitesse d'acquisition permettaient d'envisager la mise en place de nouveaux modes d'acquisition données dépendantes et données indépendantes basés sur l'acquisition multi-événements de spectres MS et MS². Toutes ces améliorations proposées dans le cadre de la métabolomique devraient rapidement être utilisables pour pouvoir être appliqués à de nouvelles cohortes médicales.

Title: Application of metabolomics for the study of hepatic encephalopathy and development of new data acquisition protocols in mass spectrometry

Keywords: Metabolomics, Mass spectrometry, Identification, Cohorts

Abstract: Untargeted metabolomics aims at studying the whole metabolite content of biological media. This allows to assess phenotypic changes in these samples. During this work, a cohort of patients suffering from hepatic encephalopathy (HE) was studied in metabolomics, in association with lipidomics and glycomics. HE is a disease still poorly understood but with high medical and social impact. This work, performed on cerebrospinal fluid and blood, demonstrates that HE might be related to an alteration of energy metabolism. In face of some highlighted limitations, we aim at developing new analytical technics for metabolomics.

We demonstrate that the use of new instrument such as Orbitrap Fusion, which can reach very high resolution, gives access to isotopic fine structure of metabolites, facilitating their identification. The study of ion MS² fragmentation enabled to better understand the involved mechanism. Finally, we demonstrate the progress of this instrument in terms of acquisition speed allows to consider the implementation of new acquisition mode based on multi event MS and MS² acquisition, such as data dependent or data independent acquisitions. All these improvement in metabolomics should rapidly be applicable to medical cohort analyses.

