



HAL
open science

Développement de nouveaux outils pour l'analyse métabolomique par spectrométrie de masse haute résolution : de l'acquisition de spectres MS/MS pour l'identification large-spectre de métabolites au marquage isotopique pour la quantification

Kathleen Rousseau

► To cite this version:

Kathleen Rousseau. Développement de nouveaux outils pour l'analyse métabolomique par spectrométrie de masse haute résolution : de l'acquisition de spectres MS/MS pour l'identification large-spectre de métabolites au marquage isotopique pour la quantification. Chimie analytique. Université Paris-Saclay, 2020. Français. NNT : 2020UPASF006 . tel-03793794

HAL Id: tel-03793794

<https://theses.hal.science/tel-03793794>

Submitted on 2 Oct 2022

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.

Développement de nouveaux outils pour l'analyse
métabolomique par spectrométrie de masse haute
résolution : de l'acquisition de spectres MS/MS pour
l'identification large-spectre de métabolites au
marquage isotopique pour la quantification

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 571, sciences chimiques : molécules, matériaux, instrumentation et
biosystèmes (2MIB)

Spécialité de doctorat: Chimie

Unité de recherche : Université Paris-Saclay, INRAE, CEA, Service de pharmacologie et Immunoanalyse,
91191, Gif-sur-Yvette, France.

Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Gif sur Yvette, le 23 Septembre 2020, par

Kathleen Rousseau

Composition du Jury

Guillaume VAN DER REST

Professeur, Université Paris-Saclay, CNRS, Institut de Chimie Physique,
Orsay, France

Président de Jury

Sandra ALVES

Maitresse de Conférence, HDR, Sorbonne Université, Faculté des Sciences
et de l'Ingénierie, Institut Parisien de Chimie Moléculaire (IPCM), Paris,
France

Rapporteuse &
Examinatrice

Gaud DERVILLY

Chercheuse HDR, Responsable Scientifique, Laboratoire d'Etude des Résidus
et Contaminants dans les Aliments (LABERCA), Ecole Nationale Vétérinaire,
Agroalimentaire et de l'Alimentation (Oniris), INRAE, Nantes, France

Rapporteuse &
Examinatrice

Patrick EMOND

Professeur, INSERM U1253, Université de Tours, Tours, France

Examineur

Christophe JUNOT

Directeur de Recherches, Département Médicaments et Technologies pour
la Santé (DMTS), CEA, INRAE, Université Paris-Saclay, MetaboHUB, Gif sur
Yvette, France

Directeur de thèse

Annelaure DAMONT

Cadre Scientifique des EPIC, Département Médicaments et Technologies
pour la Santé (DMTS), CEA, INRAE, Université Paris-Saclay, MetaboHUB, Gif
sur Yvette, France

Encadrante de thèse

François FENAILLE

Cadre Scientifique des EPIC, Département Médicaments et Technologies
pour la Santé (DMTS), CEA, INRAE, Université Paris-Saclay, MetaboHUB, Gif
sur Yvette, France

Encadrant de thèse

Remerciements

Tout d'abord, je remercie l'ensemble des membres de mon jury, Sandra Alves et Gaud Dervilly en tant que rapporteuses et Patrick Emond et Guillaume Van Der Rest en tant qu'examineur, pour avoir accepté d'évaluer mes travaux réalisés au Centre de l'Energie Atomique et des énergies alternatives (CEA) de Saclay dans le Laboratoire D'Etudes du Métabolisme des Médicament (LEMM) du Service de Pharmacologie et d'Immunoanalyse (SPI).

Ensuite, je remercie l'ensemble de mon équipe encadrante. Dans un premier temps, je remercie Christophe Junot, directeur de thèse et chef du SPI, pour m'avoir permis de réaliser cette thèse après mon stage de Master 2. Merci pour ta bonne humeur communicative. Ensuite, je remercie mes deux encadrants, Annelaure Damont et François Fenaille. Merci de m'avoir fait confiance pour cette thèse et d'avoir répondu à l'ensemble de mes questions et mes craintes. Tout d'abord Annelaure, merci de m'avoir donné ma chance pour ce projet malgré mes a priori et mes réticences sur la chimie organique au début de cette thèse. Je te remercie pour le temps que tu m'as consacré pour répondre à l'ensemble de mes questions et pour l'ensemble des discussions que nous avons pu avoir. François, je te remercie d'avoir été le premier à me faire confiance en acceptant ma candidature pour mon stage de Master 2. Merci pour tes conseils et ta rigueur qui m'ont fait grandir. Merci à tous les trois pour votre disponibilité tout au long de ma thèse.

Je remercie également l'ensemble des personnes avec qui j'ai pu travailler au cours de cette thèse. Merci à Pierre Barbier Saint Hilaire pour m'avoir formé au cours de mon stage de Master 2 sur les analyses métabolomiques par spectrométrie de masse. Merci pour le temps que tu m'as consacré sur la fin de ta thèse. Merci aux différents stagiaires qui ont travaillé sur les projets de cette thèse. Je remercie Yu Min Kiw et Jessica Michieletto qui ont développé la méthode d'échange H/D métallo-catalysée. Merci pour votre sérieux et les moments passés à côté du spectro. Je remercie également Leticia Lacalle-Bergeron de s'être expatriée quelques mois en France lors de sa thèse pour développer la méthode DDA itérative. Merci de m'avoir fait confiance, cela a été un véritable plaisir de travailler avec toi. Finalement, merci à Lou-Anne Altenburger pour le temps consacré à l'annotation des urines de souris lors de son stage de Master 2. Je te remercie de ton application dans ce projet malgré les conditions de travail complexes de cette année.

Je remercie également l'ensemble des partenaires des différents projets de cette thèse. Je remercie l'ensemble des partenaires du projet européen Human Biomonitoring for Europe (HBM4EU) et plus particulièrement Jean-Philippe Antignac, leader de notre groupe de travail, et Herbert Oberacher qui a pris en charge les études interlaboratoires. Merci à tous les deux d'avoir fait avancer ce projet. Je remercie également les équipes de Biognosys qui nous ont permis de tester le logiciel Spectronaut et plus particulièrement Maximilian Helf qui a permis d'adapter ce logiciel à la métabolomique. Merci également aux équipes de Thermo Fisher Scientific pour nous avoir mis à disposition les logiciels Acquire X et Compound Discoverer. Je remercie particulièrement l'équipe de San José d'avoir analysé nos échantillons avec Acquire X et Marie-Pierre Pavageau

pour nous avoir apporté son expertise sur Compound Discoverer. Je remercie également les équipes du SCBM pour nous avoir accueillis, Yu Min, Jessica et moi, au sein de leur laboratoire pour réaliser les réactions d'échanges H/D métallo-catalysée. Je remercie plus particulièrement Grégory Pieters et Sophie Feuillastre pour le temps consacré à ce projet. Finalement, je remercie l'ensemble des personnes qui ont travaillé sur le projet de marquage C-13 *in vivo* de souris et plus particulièrement les équipes des CEA de Marcoule et Cadarache qui nous ont permis d'avoir les échantillons. Merci pour votre inventivité.

Je remercie également l'ensemble du SPI pour m'avoir accueilli à bras grand ouvert dès le premier jour. Cela a été un grand plaisir de vous croiser tous les jours pendant un peu plus de trois ans. Merci à tous ceux du LEMM croisés lors de ces dernières années : Mikail Berdi, Arthur Viode, Noémie Butin, Samuel Terrier, Sandrine Livet, Aurélie Delvaux, Amir Louaar, Lydie Oliviera, Karl-Joe Zailah... Merci à l'ensemble de l'équipe MedDay : Sandine Aros, Fanny Leroux, Jennifer Hamoniaux, Stéphanie Oursel, Simon Broudin, Loic Mervant, Alexandre Seyer, Sylvain Dechaumet et Chloé Antoine. Merci à tous pour les discussions scientifiques et et pas que. Merci pour votre bonne humeur et l'aide que vous m'avez apporté tout au long de ma thèse. Merci à tous les habitués de la pause goûter et du petit déjeuner : Hélène Cazier, Anne-sophie Mierzala, Donatien Lefebvre, Marie Zemire, Jessica Michieletto, Daniel Leonco, Thais Hautbergue, Chloé Jacquemin, Yamina Adimy, Christelle Dubois, Céline Chollet, Nihel Bekhti, Marie Hautière... Merci pour les nombreuses discussions de travail mais aussi celles sur les avantages CEA auxquels nous avons droit et des soirées à prévoir. Les « afterworks » vont me manquer.

Finalement, je remercie mes proches et ma famille. Merci pour votre soutien. Je remercie tout particulièrement mes parents. Merci de m'avoir permis de faire les études que je voulais et de m'avoir soutenu dans les moments compliqués pour en arriver ici aujourd'hui même si vous n'avez jamais vraiment compris ce que je fais.

J'espère n'avoir oublié personnes, dans tous les cas, merci à toutes les personnes que j'ai pu rencontrer dans ce milieu professionnel ces dernières années. Discuter avec vous a été très enrichissant.

Préambule

La métabolomique est une science relativement jeune apparue dans les années 1990-2000^{1,2}. Elle fait partie des sciences « omiques » comme également la génomique, la transcriptomique et la protéomique. La métabolomique est le dernier maillon de la cascade de ces sciences (Figure 1).

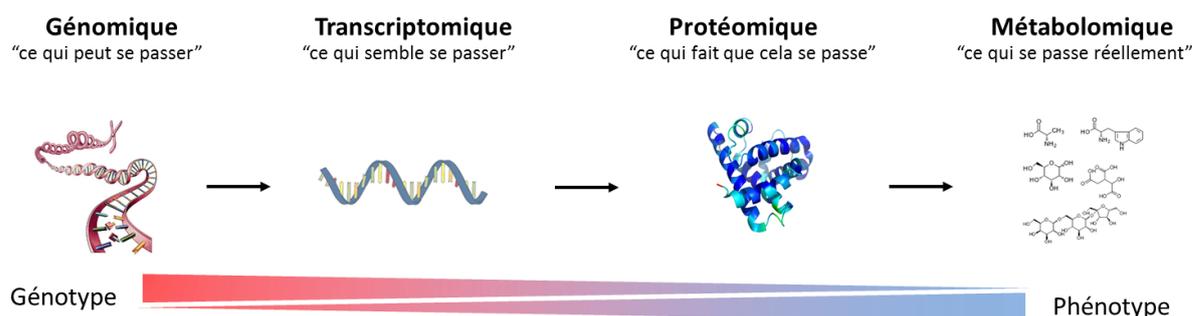


Figure 1 : La cascade des sciences "Omiques".

La métabolomique correspond à l'étude des molécules de masses moléculaires inférieures à 1.5 kDa présentes dans une matrice biologique donnée. Elle représente l'ultime réponse d'un organisme à un facteur perturbant son fonctionnement. Les métabolites étant de diversités chimiques très variées, il est très complexe de pouvoir détecter l'ensemble de ces composés en une seule analyse voire même avec une seule technique analytique. Deux méthodes analytiques sont principalement utilisées : la spectrométrie de masse (MS) et la résonance magnétique nucléaire (RMN). Cette dernière technique est robuste, plus directement quantitative et est non destructive mais en revanche souffre d'un manque de sensibilité. La spectrométrie de masse est moins robuste (notamment en raison de l'étape d'ionisation) que la RMN ce qui rend la quantification absolue des métabolites plus complexe. Cependant, en raison de sa plus grande sensibilité et de sa plus large couverture métabolique, la spectrométrie de masse est aujourd'hui la méthode analytique la plus utilisée pour les analyses métabolomiques. Pour avoir une plus large vision de la composition métabolique d'un fluide biologique, cette technique est souvent couplée à des méthodes séparatives telles que la chromatographie en phase liquide (LC) ou gazeuse (GC). Grâce à ces méthodes, des milliers de signaux sont détectés à partir d'un échantillon biologique complexe (par exemple un biofluide humain), mais il est aujourd'hui encore très difficile de tous les identifier avec certitude.

Dans ce contexte, ma thèse s'est focalisée sur l'amélioration de l'identification des métabolites et de l'aspect quantitatif des analyses métabolomiques par chromatographie liquide couplée à la spectrométrie de masse à haute résolution (LC-HRMS) respectivement au travers du développement de méthodes MS/MS sur un instrument de dernière génération (Orbitrap Fusion) et à la production de métabolites incorporant des isotopes

Préambule

stables (deutérium). Le premier objectif, initié lors de mon stage de Master 2, a été d'optimiser la méthodologie des analyses métabolomiques par spectrométrie de masse pour faciliter l'annotation voire l'identification de métabolites et gagner en temps d'analyse. En effet, lors des analyses métabolomiques actuelles, l'analyse de profil métabolique par LC-HRMS des échantillons est souvent découplée (en tout cas au début de ma thèse) de l'analyse par LC-HRMS/MS permettant de confirmer ou non l'identité des métabolites d'intérêt. Dans ces conditions, ne sont généralement fragmentés par MS/MS que les ions d'intérêt préalablement mis en évidence par l'analyse statistique des données de profilage. Jusqu'à récemment, les analyses MS/MS ne pouvaient pas être réalisées de façon non ciblée et suffisamment exhaustive au cours d'une même analyse en raison de la « faible » vitesse d'acquisition des analyseurs, notamment des analyseurs de type Orbitrap. A titre d'exemple, au sein du laboratoire, les analyses MS sont souvent réalisées avec un instrument de type Exactive (Thermo Fisher Scientific) qui ne permet pas l'acquisition de spectres MS/MS avec sélection d'ions précurseurs. Le délai entre les analyses MS et MS/MS correspond alors au temps nécessaire pour le traitement des données obtenues par l'analyse LC-HRMS. Afin de diminuer les biais analytiques induits par ce délai (dégradation de l'échantillon, changement des conditions analytiques pouvant provoquer des décalages de temps de rétention et une amélioration ou une diminution de la sensibilité), de plus en plus d'études s'intéressent au développement de méthodes permettant l'acquisition simultanée des profils métaboliques des échantillons (spectres MS) et de l'ensemble des spectres permettant la confirmation de l'identité des composés (spectres MS/MS). Ces développements sont aujourd'hui possible grâce aux améliorations technologiques ayant permis l'augmentation de la vitesse d'acquisition des analyseurs. Les méthodes de type « Data Dependent Acquisition » (DDA) et « Data Independent Acquisition » (DIA) apparaissent comme des solutions de choix pour augmenter le nombre de métabolites identifiés en une analyse en comparant les spectres MS/MS obtenus à des spectres MS/MS de référence présents dans des bases de données.

Le deuxième objectif de ma thèse a eu pour but de développer une nouvelle approche de production de métabolites marqués aux isotopes stables, afin de répondre à terme à des problématiques liées à la quantification de candidats biomarqueurs et à l'analyse de cohortes médicales. En effet, aujourd'hui, les cohortes comportent de plus en plus d'échantillons. Pour des raisons essentiellement analytiques (notamment encrassement progressif de la source du spectromètre de masse), elles sont alors divisées en plusieurs séries d'échantillons qui sont analysées sur des périodes allant de plusieurs jours à plusieurs mois. Dans ces conditions, les « batches » d'analyses sont souvent difficiles à normaliser et donc à comparer. De plus, pour certaines problématiques comme l'étude de maladies rares, le nombre d'échantillons disponibles à un instant « t » est faible. Différentes collectes d'échantillons sont alors effectuées dans le temps. Les analyses sont alors réalisées soit une fois tous les échantillons collectés soit après chaque collecte. Dans le dernier cas, afin de pouvoir comparer les différentes analyses espacées dans le temps et d'étudier les variations entre les échantillons, il est nécessaire de normaliser les signaux obtenus. En effet, il est primordial de prendre en compte uniquement les variations interindividuelles ou pathologiques et non les variations induites par la méthode analytique elle-même. Pour cela, la méthode de choix est la quantification absolue des métabolites. Le développement de méthodes ciblées de quantification absolue est souvent envisagé pour confirmer la pertinence biologique de candidats biomarqueurs révélés par les approches non ciblées, et déterminer des concentrations circulantes indispensables

Préambule

à leur utilisation ultérieure pour du diagnostic clinique. L'utilisation de standards internes marqués aux isotopes stables est, aujourd'hui, la méthode de choix pour quantifier les métabolites. Ces standards n'étant pas toujours disponibles commercialement ou à des prix parfois prohibitifs, des méthodes de marquage impliquant des isotopes stables sont développées au laboratoire.

Pour répondre à ces objectifs, ce manuscrit est divisé en deux parties principales. La première partie est une étude bibliographique en trois chapitres. Le premier chapitre présentera un état de l'art concernant la métabolomique, le déroulement d'une analyse métabolomique, les méthodes analytiques physico-chimiques utilisées et leur couplage avec les méthodes séparatives. Le deuxième chapitre se concentrera particulièrement sur les analyses par spectrométrie de masse pour les études métabolomiques. Tout d'abord, le processus d'annotation des métabolites sera détaillé puis les différents modes d'acquisitions MS/MS seront présentés. Le troisième chapitre s'intéressera aux méthodes quantitatives en métabolomique, notamment au travers de l'utilisation de standards internes incorporant des isotopes stables.

La deuxième partie de ce manuscrit présentera les résultats expérimentaux obtenus lors de cette thèse. Le premier chapitre détaillera les développements des méthodes d'acquisition MS/MS dits « Data Dependent Acquisition » et « Data Independent Acquisition » et l'application d'une méthode optimisée à l'analyse d'échantillons issus d'études inter-laboratoires dans le cadre d'un projet européen. Le second chapitre exposera les développements en terme de marquage de métabolites d'intérêt aux isotopes stables par échange hydrogène/deutérium (H/D), la caractérisation par LC-HRMS des composés deutérés ainsi obtenus et leur utilisation pour la quantification de métabolites dans un plasma de référence. Une étude préliminaire sur le marquage *in vivo* au carbone-13 (^{13}C) d'urine de souris sera finalement présentée. Pour cela, des urines de souris « marquées » au carbone-13 ont été analysées pour mieux caractériser le métabolome de la souris et étudier la cinétique de marquage.

Table des matières

Remerciements	- 3 -
Préambule.....	- 5 -
Table des matières	- 9 -
Liste des figures.....	- 13 -
Liste des tableaux.....	- 16 -
Liste des abréviations	- 18 -
Production scientifique.....	- 20 -
Partie A : Contexte scientifique et technologique	- 24 -
<i>Chapitre 1 : La métabolomique.....</i>	<i>- 26 -</i>
A – La métabolomique : contexte général	- 26 -
B – Déroulement d’une analyse métabolomique non ciblée	- 27 -
C – Les techniques d’analyse pour la métabolomique	- 32 -
1 – Vue d’ensemble des méthodes analytiques les plus utilisées	- 32 -
2 – La spectrométrie de masse	- 33 -
a – Les principales sources d’ionisation	- 34 -
b – Les principaux analyseurs.....	- 37 -
c – Les spectromètres de masse en tandem.....	- 40 -
d – La mobilité ionique	- 44 -
D – Le couplage avec les méthodes séparatives	- 45 -
1 – La chromatographie en phase gazeuse	- 45 -
2 – La chromatographie en phase liquide	- 46 -
3 – L’électrophorèse capillaire.....	- 49 -
4 – Comparaison de méthodes séparatives	- 49 -
<i>Chapitre 2 : L’acquisition de données en spectrométrie de masse pour la métabolomique.....</i>	<i>- 52 -</i>
A – L’annotation et l’identification des métabolites.....	- 52 -
B – Les différents modes d’acquisition MS/MS	- 59 -
1 – Le mode ciblé.....	- 60 -
2 – Les analyses en mode « Data Dependent Acquisition »	- 60 -
3 – Les analyses en mode « Data Independent Acquisition »	- 63 -
4 – Comparaison des différents modes d’acquisition MS/MS	- 66 -
<i>Chapitre 3 : La quantification en métabolomique.....</i>	<i>- 68 -</i>
A – La quantification : contexte général.....	- 68 -
1 – La quantification relative et la quantification absolue	- 68 -
2 – Les différentes approches de quantification par spectrométrie de masse.....	- 70 -
a – La quantification basée sur l’ion précurseur	- 70 -
b – La quantification basée sur les fragments.....	- 70 -
B – L’utilisation des isotopes stables pour la quantification des métabolites.....	- 72 -
1 – Standard interne marqué : définition, propriétés et limitations	- 72 -
2 – Les voies de production des standards internes marqués	- 73 -
a – Production in vivo de standards marqués	- 75 -
b – L’approche par marquage chimique (ou marquage in vitro)	- 76 -
3 – Utilisation des standards marqués pour la quantification	- 83 -

Table des matières

<i>Objectifs</i>	- 86 -
Partie B : Etudes expérimentales	- 88 -
<i>Chapitre 1 : Développement de nouvelles méthodes d'acquisition pour la métabolomique</i>	- 90 -
A – Développements des méthodes « Data Dependent Acquisition » et « Data Independent Acquisition » ..	90 -
1 – Contexte et objectifs	- 90 -
2 – Méthodologie suivie et introduction au travail expérimental	- 91 -
3 – Publication « Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics »	- 92 -
4 – Principaux résultats et conclusions	- 109 -
B – Application à des études d'intercomparaison de spectres de masse dans le cadre d'un projet européen (HBM4EU)	- 115 -
1 – Contexte du projet européen	- 115 -
2 – Création d'une base de données et harmonisation des protocoles d'acquisition	- 116 -
a – Contexte et objectifs	- 116 -
b – Méthodologie	- 117 -
c – Résultats	- 121 -
d – Conclusions et perspectives	- 126 -
C – Automatisation et amélioration de la méthode « Data Dependent Acquisition »	- 127 -
1 – Contexte et objectifs	- 127 -
2 – Méthodologie	- 129 -
3 – Résultats	- 130 -
4 – Conclusion et perspectives	- 132 -
D – Conclusions et perspectives	- 132 -
<i>Chapitre 2 : Production et utilisation de standards internes marqués (SIL-IS) pour la quantification en métabolomique</i>	- 134 -
A – Production de standards internes marqués (SIL-IS) par une méthode mono-étape métallo-catalysée	134 -
1 – Contexte et objectifs	- 134 -
2 – Principe de l'introduction de deutérium par réaction mono-étape métallo-catalysée	- 135 -
3 – Plan expérimental	- 135 -
4 – Matériel et méthodes	- 140 -
5 – Résultats	- 142 -
a - Utilisation du ruthénium comme catalyseur de l'incorporation de deutérium	- 142 -
b - Adaptation du procédé de marquage avec remplacement de Ru/C par Pd/C	- 147 -
6 – Bilan des marquages métallo-catalysés	- 149 -
B – Développement d'une nouvelle méthode de quantification : « multi-point internal calibration » ..	152 -
1 – Contexte et objectif	- 152 -
2 – Méthodologie	- 153 -
3 – Publication (en cours de rédaction)	- 155 -
4 – Principaux résultats et conclusion	- 168 -
C – Evaluation de l'efficacité de méthodes alternatives de deutération pour l'obtention de versions marquées de métabolites d'intérêt	- 169 -
1 - Méthodologie	- 169 -
2 – Résultats	- 171 -
3 – Conclusions et perspectives	- 180 -
D - Utilisation du marquage <i>In Vivo</i> pour la quantification et l'identification de métabolites	- 180 -
1 – Contexte et objectifs	- 180 -
2 – Matériels et méthodes	- 181 -
3 – Résultats et perspectives	- 183 -
E – Conclusion et perspectives	- 192 -
Conclusions et discussions générales	- 196 -
Références	- 202 -

Table des matières

Partie C : Annexes.....	- 219 -
<i>Annexe 1 : Revue sur les modes d'acquisitions en spectrométrie de masse</i>	<i>- 221 -</i>
<i>« Data Acquisitions workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? »</i>	<i>- 221 -</i>
<i>Annexe 2 : Données Supplémentaires de la Publication sur les Modes d'Acquisitions DDA et DIA</i>	<i>- 235 -</i>
<i>« Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics »</i>	<i>- 235 -</i>
<i>Annexe 3 : Publication HBM4EU sur la proposition de critères qualités pour la confection d'une base de données spectrale MS/MS.....</i>	<i>- 259 -</i>
<i>« A European proposal for quality control and quality assurance of tandem mass spectral libraries » .</i>	<i>- 259 -</i>
<i>Annexe 4 : Détails conditions d'échange H/D par la réaction mono-étape métallo-catalysée.....</i>	<i>- 289 -</i>
<i>Annexe 5 : Données supplémentaires sur la publication de la nouvelle méthode de quantification.....</i>	<i>- 293 -</i>
<i>« High-throughput metal-catalyzed deuterium labeling of small molecules for MS-based quantitative metabolomics and evaluation of an innovative multipoint calibration approach »</i>	<i>- 293 -</i>
<i>Annexe 6 : Liste des composés investigués avec la nouvelle méthode de production de standards internes marqués (SIL-IS)</i>	<i>- 299 -</i>

Liste des figures

FIGURE 1 : LA CASCADE DES SCIENCES "OMIQUES".	- 5 -
FIGURE 2 : DEROULEMENT D'UNE ANALYSE METABOLOMIQUE NON CIBLEE (ADAPTEE DE JUNOT ET AL.).	- 27 -
FIGURE 3 : EXEMPLE D'UNE SEQUENCE ANALYTIQUE.	- 29 -
FIGURE 4 : ILLUSTRATION DE L'IMPACT DE LA NORMALISATION DE QUATRE « BATCHES » DISTINCTS D'ANALYSE (COHORTE DE 826 ECHANTILLONS DE SERUM).	- 30 -
FIGURE 5 : EXEMPLES DE DECALAGE DE TEMPS DE RETENTION LORS DE L'ANALYSE D'UN MEME ECHANTILLON A 10 MOIS D'INTERVALLE..	- 31 -
FIGURE 6 : NOMBRE DE PUBLICATIONS EN METABOLOMIQUE AU COURS DE CES 15 DERNIERES ANNEES AVEC DES ANALYSES IMPLIQUANT LA RMN OU LA SPECTROMETRIE DE MASSE.	- 33 -
FIGURE 7 : SCHEMA D'UN SPECTROMETRE DE MASSE.	- 33 -
FIGURE 8 : SCHEMA D'UNE SOURCE ESI (EXTRAITE DE CECH ET AL.).	- 35 -
FIGURE 9 : RECAPITULATIF DES CONDITIONS OPTIMALES POUR L'UTILISATION DES SOURCES D'IONISATION A PRESSION ATMOSPHERIQUE (EXTRAITE DE SYAGE ET AL.).	- 36 -
FIGURE 10 : COMPARAISON ET COMPLEMENTARITE DES TROIS METHODES D'IONISATION A PRESSION ATMOSPHERIQUE ESI, APCI ET APPI POUR L'ANALYSE METABOLOMIQUE D'UN ECHANTILLON DE PLASMA HUMAIN (ADAPTEE DE TIAN ET AL.).	- 36 -
FIGURE 11 : SCHEMA D'UN ANALYSEUR QUADRIPOLAIRE (EXTRAITE DE WWW.SHIMADZU.COM).	- 38 -
FIGURE 12 : SCHEMA D'UN ANALYSEUR DE TYPE TEMPS DE VOL (EXTRAITE DE WWW.SHIMADZU.COM).	- 39 -
FIGURE 13 : SCHEMA D'UN ANALYSEUR DE TYPE ORBITRAP (EXTRAITE DE SCIGELOVA ET AL.).	- 40 -
FIGURE 14 : SCHEMA DE L'ORBITRAP FUSION (EXTRAITE DE PLANETORBITRAP.COM).	- 41 -
FIGURE 15 : SCHEMA D'UN INSTRUMENT DE MOBILITE IONIQUE A TEMPS DE DERIVE (ADAPTEE DE JÜNGER ET AL.).	- 44 -
FIGURE 16 : ANALYSE DE SERUM HUMAIN PAR TROIS TECHNIQUES CHROMATOGRAPHIQUES (EXTRAITE DE BOUDAH ET AL.).	- 48 -
FIGURE 17 : COUVERTURE DE 75 INTERMEDIAIRES DU METABOLISME CENTRAL PAR TROIS METHODES SEPARATIVES (EXTRAITE DE BÜSCHER ET AL.).	- 50 -
FIGURE 18 : LES ETAPES D'ANNOTATION ET D'IDENTIFICATION EN METABOLOMIQUE.	- 53 -
FIGURE 19 : REPARTITION DES COMPOSES SELON LEUR ORIGINE (ADAPTEE DE SINDELAR ET AL.).	- 53 -
FIGURE 20 : LES DIFFERENTS SIGNAUX MS PRODUITS PAR LE PROPIONYL-CoA (ADAPTEE DE DAMONT ET AL.).	- 54 -
FIGURE 21 : REPARTITION DES COMPOSES SELON LEUR ABONDANCE (ADAPTEE DE CHALECKIS ET AL.).	- 57 -
FIGURE 22 : LES DIFFERENTS MODES D'ACQUISITION DES SPECTRES MS/MS.	- 59 -
FIGURE 23 : PRINCIPE DE LA FRAGMENTATION EN MODE "STEPPED COLLISION ENERGY".	- 61 -
FIGURE 24 : COMPARAISON DES SPECTRES OBTENUS AVEC LES MODES DE FRAGMENTATION HCD ET CID POUR 1,2-DIDOCOSAHEXAENOYL-GLYCERO-3-PHOSPHOCHOLINE (ADAPTEE DE MULLARD ET AL.).	- 62 -
FIGURE 25 : EXEMPLE DE CYCLES AVEC UN SPECTRE MS (EN NOIR) ET PLUSIEURS SPECTRES MS/MS (EN ROUGE) LORS D'UN PIC CHROMATOGRAPHIQUE.	- 63 -
FIGURE 26 : LA QUANTIFICATION RELATIVE ET LA QUANTIFICATION ABSOLUE.	- 69 -
FIGURE 27 : COMPARAISON DES MODES D'ACQUISITION SRM (/MRM) ET PRM (ADAPTEE DE BURNUM-JOHNSON ET AL.).	- 71 -
FIGURE 28 : ISOTOPOLOGUES ET ISOTOPOMERES : ILLUSTRATION AVEC L'EXEMPLE DE LA SERINE MARQUEE AU DEUTERIUM.	- 73 -
FIGURE 29 : ILLUSTRATION DES DEUX TYPES D'APPROCHES (IN VIVO ET IN VITRO) POUR LA PRODUCTION DE COMPOSES MARQUES AUX ISOTOPES STABLES (EXTRAITE DE FREUND ET AL.).	- 74 -
FIGURE 30 : LES DIFFERENTS STANDARDS INTERNES POUR LES ANALYSES METABOLOMIQUES QUANTITATIVES (EXTRAITE DE KHAMIS ET AL.).	- 76 -
FIGURE 31 : ILLUSTRATION DES METHODES DE QUANTIFICATION RELATIVE ET ABSOLUE PAR DERIVATISATION CHIMIQUE (EXTRAITE DE HIGASHI ET AL.).	- 78 -
FIGURE 32 : STRATEGIE DE DERIVATISATION CHIMIQUE APPLIQUEE A L'ANALYSE DU METABOLOME.	- 79 -
FIGURE 33 : ILLUSTRATION D'UNE REACTION D'ECHANGE ¹² C/ ¹³ C PAR DECARBOXYLATION/CARBOXYLATION (EXTRAITE DE TORTAJADA ET AL.).	- 80 -
FIGURE 34 : REACTION D'ECHANGE H/D AVEC LE PLATINE OU LE PALLADIUM COMME CATALYSEUR (ADAPTEE DE SAJIKI ET AL.).	- 81 -
FIGURE 35 : EXEMPLES ILLUSTRANT LA REGIOSPECIFICITE DE PLUSIEURS CATALYSEURS METALLIQUES VIS-A-VIS DE LA DEUTERATION PAR REACTION D'ECHANGE H/D (EXTRAITE DE FUJOWARA ET AL.).	- 81 -
FIGURE 36 : EXEMPLE DE REACTION D'ECHANGE H/D AVEC LE RUTHENIUM COMME CATALYSEUR (EXTRAITE DE TAGLANG ET AL.).	- 82 -
FIGURE 37 : COMPARAISON DES METHODES DE QUANTIFICATION PAR GAMME EXTERNE AVEC SIL-IS ET DILUTION ISOTOPIQUE.	- 83 -
FIGURE 38 : EXEMPLE DE FRAGMENTATION EN MODE « STEPPED COLLISION ENERGY » POUR L'ANALYSE TYROSINE (EXTRAITE DE LA	

Liste des figures

PUBLICATION CI-DESSUS).....	- 110 -
FIGURE 39 : COMPARAISON DES SPECTRES MS/MS OBTENUS AVEC CEUX DE REFERENCE A L'AIDE DE MS-DIAL.....	- 111 -
FIGURE 40 : REPARTITION DES MASSES DES METABOLITES DETECTES DANS LE PLASMA NIST.....	- 112 -
FIGURE 41 : COMPARAISON DU NOMBRE DE COMPOSES QUANTIFIES PAR MS ET MS/MS PAR TRAITEMENT DE DONNEES MANUEL OU PAR SPECTRONAUT.....	- 114 -
FIGURE 42 : INFLUENCE DE LA GAMME D'ACQUISITION SUR LES SPECTRES DE FRAGMENTATION (ADAPTEE DE BARBIER SAINT HIALIRE ET AL.).....	- 125 -
FIGURE 43 : PRINCIPE DE LA METHODE DDA ITERATIVE.....	- 128 -
FIGURE 44 : NOMBRE DE COMPOSES IDENTIFIES PAR LA METHODE ACQUIRE X ET DDA CLASSIQUE.....	- 129 -
FIGURE 45 : METHODE DDA ACQUIRE X.....	- 130 -
FIGURE 46 : CONSTITUTION ITERATIVE DES LISTES D'INCLUSION ET D'EXCLUSION.....	- 131 -
FIGURE 47 : COMPARAISON DES SPECTRES MS OBTENUS PAR NOTRE LABORATOIRE ET PAR THERMOFISHER SCIENTIFIC.....	- 131 -
FIGURE 48 : REGIOSPECIFICITE ET STEREOSPECIFICITE DE LA REACTION D'ECHANGE H/D MONO-ETAPE METALLO-CATALYSEE (EXTRAITE DE TAGLANG ET AL.).....	- 135 -
FIGURE 49 : STRUCTURES DE 52 COMPOSES INVESTIGUES POUR LA REACTION D'ECHANGE H/D MONO-ETAPE METALLO-CATALYSEE (ADAPTEE DE LA PUBLICATION EN COURS DE REDACTION).....	- 137 -
FIGURE 50 : CAHIER DES CHARGES POUR LA PRODUCTION DE STANDARDS INTERNES MARQUES.....	- 138 -
FIGURE 51 : ARBRE DECISIONNEL POUR LE DEVELOPPEMENT DES ECHANGES H/D METALLO-CATALYSES (ADAPTEE DE LA PUBLICATION EN COURS DE REDACTION).....	- 139 -
FIGURE 52 : DISPOSITIFS UTILISES POUR LES REACTIONS D'ECHANGE H/D METALLO-CATALYSEES SOUS PRESSION DE D ₂	- 139 -
FIGURE 53 : PRINCIPE DES INVESTIGATIONS REALISEES SUR LE DISPOSITIF CAT24 AVEC LA REACTION D'ECHANGE H/D MONO-ETAPE METALLO-CATALYSEE.....	- 142 -
FIGURE 54 : EXEMPLES DE MASSIFS ISOTOPIQUES OBTENUS APRES MARQUAGE PAR ECHANGE H/D AVEC RU/C.....	- 143 -
FIGURE 55 : INFLUENCE DU TEMPS DE REACTION (ET DU NOMBRE DE RECHARGEMENT EN D ₂) POUR LA REACTION D'ECHANGE H/D MONO-ETAPE METALLO-CATALYSEE.....	- 144 -
FIGURE 56 : CINETIQUE D'INCORPORATION DES ATOMES DE DEUTERIUMS PAR CHIMIE EN FLUX CONTINU.....	- 145 -
FIGURE 57 : SPECTRES RMN ¹ H DE LA 4-HYDROXY-PROLINE (EN BAS) ET DE SON HOMOLOGUE MARQUE AU DEUTERIUM (EN HAUT).....	- 146 -
FIGURE 58 : EXEMPLES DE MARQUAGES OBTENUS PAR ECHANGE H/D METALLO-CATALYSE PAR Pd/C.....	- 147 -
FIGURE 59 : INFLUENCE DU PH POUR LA REACTION MONO-ETAPE METALLO-CATALYSEE.....	- 148 -
FIGURE 60 : CONVERSION DE L'ACIDE MALEIQUE EN ACIDE SUCCINIQUE DEUTERE DANS LES CONDITIONS REDUCTRICES DE REACTION D'ECHANGE H/D METALLO-CATALYSEE.....	- 149 -
FIGURE 61 : EXEMPLE D'ECHANGE EN RETOUR ("BACK EXCHANGE") OBSERVE AVEC L'ACIDE METHYLMALONIQUE.....	- 150 -
FIGURE 62 : REPARTITION DES COMPOSES MARQUES EN FONCTION DE L'HOMOGENEITE DE LEUR MARQUAGE.....	- 152 -
FIGURE 63 : PRINCIPE DE LA METHODE DE "MULTI-POINT INTERNAL CALIBRATION".....	- 153 -
FIGURE 64 : INFLUENCE DE LA RESOLUTION SUR LA SEPARATION DES COMPOSES MONO-C-13 DU STANDARD NON MARQUE ET DU MONO-DEUTERE DU LOT MARQUE. ILLUSTRATION AVEC LA CITRULLINE ET L'HISTIDINE.....	- 154 -
FIGURE 65 : COMPARAISON DES MESURES DE CONCENTRATIONS OBTENUES PAR TROIS METHODES DE QUANTIFICATION POUR SIX COMPOSES SUPPLEMENTES AU PLASMA NIST.....	- 168 -
FIGURE 66 : METHODE DDA POUR LES ANALYSES LC-HRMS/MS DE COMPOSES DEUTERES.....	- 170 -
FIGURE 67 : EXEMPLE DES MASSIFS ISOTOPIQUES OBTENUS POUR LE MARQUAGE DE L'ACIDE XANTHURENIQUE ANALYSE EN MODE D'IONISATION POSITIF.....	- 171 -
FIGURE 68 : EXEMPLES DE MARQUAGE EFFICACE OBTENUS POUR LA N-ACETYLGLYCINE ET L'ACIDE N-ACETYLASPARTIQUE.....	- 174 -
FIGURE 69 : EXEMPLES DE COMPOSES MARQUES COMPORTANT UN ISOTOPOLOGUE DEUTERE UNIQUE.....	- 175 -
FIGURE 70 : EXEMPLES DE MARQUAGE INCOMPLET DE L'ACIDE SUCCINIQUE ET DE LA TRANS-3-HYDROXYCOTININE.....	- 176 -
FIGURE 71 : REPARTITION DES 103 COMPOSES INVESTIGUES PAR LA METHODE DE DEUTERATION ALTERNATIVE SELON LA PROPORTION DE L'ISOTOPOLOGUE MAJORITAIRE PAR RAPPORT A L'ENSEMBLE DU MASSIF ISOTOPIQUE.....	- 177 -
FIGURE 72 : EXEMPLE D'EFFET ISOTOPIQUE LORS DE L'IONISATION OBSERVE POUR L'ACIDE HIPPIURIQUE MARQUE.....	- 178 -
FIGURE 73 : COMPARAISON DES SPECTRES MS/MS INDEPENDANTS DES IONS NON-DEUTERE, MONO-DEUTERE ET DI-DEUTERE... -	179 -
FIGURE 74 : DETAILS DU PRELEVEMENT DES ECHANTILLONS D'URINE DE SOURIS.....	- 182 -
FIGURE 75 : EXEMPLE DE MARQUAGE IN VIVO DE L'ASPARAGINE.....	- 185 -
FIGURE 76 : COMPARAISON DES CINETIQUES DE MARQUAGE DE L'ASPARAGINE POUR LES TROIS SOURIS SUR LES TROIS SYSTEMES CHROMATOGRAPHIQUES DANS LES DEUX MODES D'IONISATION.....	- 186 -
FIGURE 77 : EXEMPLES DE CINETIQUES DE MARQUAGE D'ACIDES AMINES.....	- 187 -
FIGURE 78 : EXEMPLE DE CINETIQUE DE MARQUAGE POUR L'ACIDE ADIPIQUE ET LA XANTHOSINE.....	- 189 -
FIGURE 79 : EXEMPLE DE DETERMINATION DE FORMULE BRUTE PAR COMPARAISON SIGNAL "FULL C-12" ET "FULL C-13".....	- 191 -
FIGURE 80 : COMPARAISON DES VERSIONS NON-MARQUEE ET MARQUEE D'UN MEME METABOLITE ET HYPOTHESE D'IDENTIFICATION DE	

Liste des figures

LA DECENOYLCARNITINE.	- 192 -
FIGURE 81 : PROTOCOLE ENVISAGEABLE POUR LA DIFFERENCIATION PAR LC-HRMS DES VRAIS METABOLITES PAR RAPPORT AUX CONTAMINANTS ET IONS DU BRUIT DE FOND DANS DES EXTRAITS DE PETITS MAMMIFERES.	- 194 -
FIGURE 82 : EXEMPLE D'UNE POSSIBLE FUTURE SEQUENCE ANALYTIQUE.	- 198 -

Liste des tableaux

TABLEAU 1 : COMPARAISON DES CARACTERISTIQUES DE QUELQUES SPECTROMETRES DE MASSE A HAUTE ET TRES HAUTE RESOLUTION. ...	43 -
TABLEAU 2 : LISTE DE BASES DE DONNEES MS (ADAPTE DE BLAZENOVIC ET AL.)	- 55 -
TABLEAU 3 : LES PRINCIPAUX ISOTOPES STABLES ET LEUR ABONDANCE NATURELLE.	- 56 -
TABLEAU 4 : LISTE DES BASES DE DONNEES SPECTRALES (ADAPTE DE BLAZENOVIC ET AL.)	- 56 -
TABLEAU 5 : LISTE NON-EXHAUSTIVE DE LOGICIELS UTILISES POUR LE TRAITEMENT DES DONNEES DDA ET DIA (ADAPTE DE FENAILLE ET AL.).....	- 65 -
TABLEAU 6 : BILAN DES MODES D'ACQUISITION MS/MS.	- 66 -
TABLEAU 7 : EXEMPLE DE REACTIONS DE DERIVATISATION SELON LA FAMILLE CHIMIQUE (ADAPTE DE ZHAO ET AL.)	- 79 -
TABLEAU 8 : TAILLES DES FENETRES D'ACQUISITION SWATH.	- 113 -
TABLEAU 9 : RECAPITULATIF DES MELANGES DE STANDARDS DE L'ETUDE "SCORE".	- 117 -
TABLEAU 10 : RECAPITULATIF DES MELANGES DE STANDARDS DE L'ETUDE "STUDY 1".	- 118 -
TABLEAU 11 : LISTE DES 19 STANDARDS DE L'ETUDE "STUDY 2".	- 119 -
TABLEAU 12 : RESULTATS DE L'ETUDE "SCORE" AVEC LE PARAMETRE « ¹² C SELECTION ».	- 121 -
TABLEAU 13 : RESULTATS DE L'ETUDE "STUDY 1" AVEC NOTRE MELANGE DE STANDARD.....	- 123 -
TABLEAU 14 : BILAN DES RESULTATS INTER-PLATEFORMES OBTENUS POUR LA "STUDY 2" (EXTRAIT DE OBERACHER ET AL.)	- 123 -
TABLEAU 15 : LISTE DES COMPOSES INVESTIGUES POUR LA REACTION D'ECHANGE H/D MONO-ETAPE METALLO-CATALYSEE (ADAPTE DE LA PUBLICATION EN COURS DE REDACTION)	- 136 -
TABLEAU 16 : BILAN DES MARQUAGES MONO-ETAPE METALLO-CATALYSES.	- 151 -
TABLEAU 17 : BILAN DES MASSIFS ISOTOPIQUES OBTENUS PAR LA METHODE ALTERNATIVE DE DEUTERATION.	- 172 -
TABLEAU 18 : LISTE DES 18 ACIDES AMINES D'INTERETS.	- 184 -
TABLEAU 19 : AUTRES COMPOSES D'INTERETS POUR L'ETUDE DU MARQUAGE DES URINES DE SOURIS.	- 188 -
TABLEAU 20 : BILAN DU NOMBRE DE VARIABLES OBTENUES APRES TRAITEMENT DES DONNEES SUR W4M	- 190 -
TABLEAU 21 : NOMBRE DE VARIABLES DETECTEES DANS LES ECHANTILLONS D'URINE DE SOURIS C-13.	- 190 -

Liste des abréviations

¹³C ou C-13 : carbone-13

¹⁵N ou N-15 : Azote-15

AGC : Automatic Gain Control, Contrôle Automatique du Gain

AIF : All Ion Fragmentation, Fragmentation de Tous les Ions

APCI : Atmospheric Pressure Chemical Ionization, Ionisation Chimique à Pression Atmosphérique

APPI : Atmospheric Pressure Photoionization, Photoionisation à Pression Atmosphérique

CASMI : Critical Assessment of Small Molecule Identification, Évaluation critique de l'identification des petites molécules

CCS : Collision Cross Section, Section Transversale de Collision

CE : Capillary Electrophoresis, Electrophorèse Capillaire

CE-MS : Capillary Electrophoresis coupled to Mass Spectrometry, Electrophorèse Capillaire couplée à la Spectrométrie de Masse

CI : Chemical Ionization, Ionisation Chimique

CID : Collision Induced Dissociation, Dissociation Induite par Collision (après activation résonante)

Da : Dalton

DDA : Data Dependent Acquisition, Acquisition des Données en mode Dépendent

DIA : Data Independent Acquisition, Acquisition des Données en mode Indépendent

DIMS : Direct Introduction Mass Spectrometry, Spectrométrie de Masse à Introduction Directe

DP score : Dot Product Score

DTIMS : Drift Time Ion Mobility Spectroscopy, Spectroscopie de Mobilité Ionique à Temps de Dérive

EI : Electron Ionization, Ionisation par Impact Electronique

ESI : ElectroSpray Ionization, Ionisation par ElectroSpray

FAIMS : Field Asymmetric Ion Mobility Spectroscopy, Spectroscopie de Mobilité Ionique à Champs Asymétrique

FDA : Food and Drug Administration, Administration des Aliments et des Médicaments

FIA : Flow Injection Analysis, Introduction Directe en flux

FT-ICR : Fourier Transform Ion Cyclotron Resonance, Transformée de Fourier à Résonance Cyclotronique Ionique

FWHM : Full Width at Half Maximum ; Largeur à Mi-Hauteur

GC : Gas Chromatography, Chromatographie en phase Gazeuse

GC-MS : Gas Chromatography - Mass Spectrometry, Chromatographie en phase Gazeuse couplée à la Spectrométrie de Masse

H/D : Hydrogène / Deutérium

HBM4EU : Human Biomonitoring For Europe, Biosurveillance Humaine Pour l'Europe

HCD : Dissociation Induite par Collision (après activation non résonante)

HILIC : Hydrophilic Interaction Liquid Chromatography, Chromatographie Liquide à Interaction Hydrophile

HMDB : Human Metabolome DataBase

IM-MS : Ion Mobility - Mass Spectrometry, Mobilité Ionique couplée à la Spectrométrie de Masse

LC : Liquid Chromatography, Chromatographie en phase Liquide

LC-HRMS : Liquid Chromatography coupled to High-Resolution Mass Spectrometry, Chromatographie Liquide couplée à la Spectrométrie de Masse à Haute Résolution

m/z : rapport masse sur charge

MIM : Multiple Ion Monitoring, Mesure d'Ions Multiples

MRM : Multiple Reaction Monitoring, Suivi de Réactions Multiples

MS : Mass Spectrometry, Spectrométrie de Masse

MS/MS : Spectrométrie de masse en tandem

NCE : Normalized Collision Energy, Energie de Collision Normalisée

NIST : National Institute of Standards and Technologies, Institut National des Normes et Technologies

NPLC : Normal Phase Liquid Chromatography, Chromatographie Liquide en Phase Normale

Pd/C : Palladium sur charbon actif

PFPP : PentaFluoroPhenylPropyl

Liste des abréviations

ppb : *Partie par billion*

ppm : *parties par million*

PRM : *Parallel Reaction Monitoring, Suivi de Réactions en Parallèles*

QC : *Quality Control, Contrôle Qualité*

QqQ : *Triple quadripôle*

QqTOF ou Q-TOF : *Quadrupole - Time Of Flight, Quadripôle - Temps De Vol*

RMN : *Nuclear Magnetic Resonance, Résonance Magnétique Nucléaire*

RPLC : *Reversed Phase Liquid Chromatographie, Chromatographie Liquide en Phase Inverse*

RT : *Retention Time, Temps de Rétention*

Ru/C : *Ruthénium sur charbon actif*

SIL-IS : *Stable Isotope Labeling - Internal Standard, Standard Interne Marqué aux Isotopes Stables*

SIM : *Single Ion Monitoring, Mesure d'un Seul Ion*

SRM : *Selected Reaction Monitoring, Suivi de Réaction Sélectionnée*

SWATH : *Sequential Window Acquisition of All THEoretical fragment-ion spectra*

TIMS : *Trapped Ion Mobility Spectroscopy, Spectroscopie de Mobilité Ionique à Piège*

TOF : *Time Of Flight, Temps de Vol*

TWIMS : *Traveling-Wave Ion Mobility Spectroscopy, Spectroscopie de Mobilité Ionique à Ondes Progressives*

UHPLC : *Ultra High Performance Liquid Chromatography, Chromatographie Liquide à Ultra-Haute Performance*

W4M : *Workflow4Metabolomics*

Production scientifique

Publications

- *Data Acquisition workflow 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
- *A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries*
Herbert Oberacher, Michael Sasse, Jean-Philippe Antignac, Yann Guitton, Laurent Debrauwer, Emilien L. Jamin, Tobias Schulze, Martin Krauss, Adrian Covaci, Noelia Caballero-Casero, Kathleen Rousseau, Annelaure Damont, François Fenaille, Marja Lamoree, Emma L. Schymanski
Environmental Sciences Europe, 2020, 32:43
- *Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics*
Pierre Barbier Saint Hilaire[#], Kathleen Rousseau[#], Alexandre Seyer, Sylvain Dechaumet, Annelaure Damont, Christophe Junot, François Fenaille
Metabolites, 2020, 10(4), 158
- *Quantification of plasma metabolites by multi-point internal calibration using homemade produced labeled internal standards*
Kathleen Rousseau, Jessica Michieletto, Yu Min Kiw, Sophie Feuillastre, Grégory Pieters, Christophe Junot, François Fenaille, Annelaure Damont
En cours de rédaction

Communications Orales

- *Quantitative mass spectrometry-based targeted metabolomics using fastly produced Deuterated internal standards*
Annelaure Damont, Yu Min Kiw, Kathleen Rousseau, Sophie Feuillastre, Grégory Pieters, Christophe Junot, François Fenaille
Congrès Metabolomics, Seattle, Etats-Unis, 2018
- *Convenient production of deuterium-labelled internal standards for large-scale quantitative metabolomics*
Kathleen Rousseau, Jessica Michieletto, Yu Min Kiw, Sophie Feuillastre, Gregory Pieters, Christophe Junot, François Fenaille, Annelaure Damont
Journée des doctorants de l'institut Joliot, CEA Saclay, France, 2019

Production scientifique

- *Convenient production of deuterium-labelled internal standards for large-scale quantitative metabolomics*
Kathleen Rousseau, Jessica Michieletto, Yu Min Kiw, Sophie Feuillastre, Gregory Pieters, Christophe Junot, François Fenaille, Annelaure Damont
Journée des doctorants de l'école doctorale 2MIB, Evry, France, 2019
- *Convenient production of deuterium-labelled internal standards for large-scale quantitative metabolomics*
Kathleen Rousseau, Jessica Michieletto, Yu Min Kiw, Sophie Feuillastre, Gregory Pieters, Christophe Junot, François Fenaille, Annelaure Damont
Congrès de Spectrométrie de Masse et Analyse Protéique, Strasbourg, France, 2019
- *Implementation of convenient one-step protocols for producing deuterium labeled metabolites: Assessment of its use for quantitative LC-HRMS-based targeted metabolomics*
Annelaure Damont, Yu Min Kiw, Kathleen Rousseau, Sophie Feuillastre, Grégory Pieters, Christophe Junot, François Fenaille
Symposium de Spectrométrie de Masse, Sherbrooke, Canada, 2019
- *Single injection LC-HRMS large-scale quantitative metabolomics using multipoint internal calibration and on-demand produced multideuterated metabolites*
Kathleen Rousseau, Jessica Michieletto, Yu Min Kiw, Sophie Feuillastre, Gregory Pieters, Christophe Junot, François Fenaille, Annelaure Damont
Congrès virtuel de l'American Society of Mass Spectrometry, 2020

Communications par Affiches

- *Development and comparison of Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA) LC-MS/MS workflow for untargeted metabolomics*
Kathleen Rousseau, Pierre Barbier Saint Hilaire, Annelaure Damont, François Fenaille, Christophe Junot
Congrès du Réseau Francophone de Métabolomique et Fluxomique, Liège, Belgique, 2018
- *Development and comparison of Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA) LC-MS/MS workflow for untargeted metabolomics*
Kathleen Rousseau, Pierre Barbier Saint Hilaire, Annelaure Damont, François Fenaille, Christophe Junot
Journée des doctorants de l'institut Joliot, CEA Saclay, France, 2018
- *¹³C-labeling of mice for quantitative metabolomics analyses*
Kathleen Rousseau, Annelaure Damont, Laurent Bellanger, Séverine Boity, Laurent Nussaume, Eric Ezan, Jean-Jacques Leguay, Christophe Junot, François Fenaille
Journées des doctorants de l'école doctorale 2MIB, Orsay, France, 2018
- *Implementation of convenient one-step protocols for producing deuterium labeled metabolites: Assessment of its use for quantitative LC-HRMS-based targeted metabolomics*
Annelaure Damont, Yu Min Kiw, Kathleen Rousseau, Sophie Feuillastre, Grégory Pieters, Christophe Junot, François Fenaille
Présentée par Annelaure Damont au congrès American Society of Mass Spectrometry, Atlanta, Etats-Unis, 2019
- *Application of novel automated data dependent acquisition (DDA) in untargeted metabolomics*

Production scientifique

Leticia Lacalle-Bergeron, Kathleen Rousseau, Tania Portolés-Nicolau, Juan V. Sancho, Christophe Junot, François Fenaille

Congrès 1st Iberian Meeting in Separation Sciences and Mass Spectrometry, Santiago de Compostela, Espagne, 2019

- *Exploring DIA data analysis for metabolomics with Spectronaut*

Maximilian J. Helf, Kathleen Rousseau, Oliver M. Bernhardt, Pierre Barbier Saint Hilaire, Tejas Gandhi, François Fenaille, Lukas Reiter

Congrès virtuel de l'American Society of Mass Spectrometry, 2020

*Partie A : Contexte
scientifique et technologique*

Chapitre 1 : La métabolomique

A – La métabolomique : contexte général

La métabolomique est une science relativement jeune, introduite dans les années 1990-2000 par Fiehn¹ et Nicholson². La métabolomique, qui correspondait aux analyses qualitatives des petites molécules, était alors différenciée de la métabonomique, qui correspondait aux analyses quantitatives. Aujourd'hui, seul le terme métabolomique est utilisé pour décrire l'analyse des molécules de poids moléculaires inférieurs à 1 kDa (ou 1.5 kDa selon les références).

La métabolomique est une science pluridisciplinaire qui agrège des expertises en chimie analytique, informatique, biostatistiques et biochimie. En effet, la chimie analytique permet d'obtenir les empreintes métaboliques des échantillons et d'identifier les métabolites présents, le traitement du signal et les analyses statistiques permettent d'extraire l'information analytiquement pertinente et les sciences du vivant font le lien entre les résultats analytiques et le contexte biologique de l'étude. La métabolomique a de nombreux champs d'application comme la médecine, la nutrition, ou l'environnement³.

La métabolomique s'intéresse donc aux plus petits composés présents dans une matrice biologique donnée (plasma, urine, plantes, microorganismes...): les métabolites. Ils composent le métabolome défini comme l'ensemble des molécules de faibles poids moléculaires⁴. Le métabolome comprend notamment toutes les substances organiques provenant naturellement du métabolisme de l'organisme vivant étudié⁵. Les métabolites sont divisés en deux familles principales : les métabolites d'origine endogène directement produits par l'organisme (acides aminés, vitamines...) et ceux d'origine exogène appelés xénobiotiques provenant d'une source extérieure à l'organisme (médicaments, produits de l'alimentation, polluants environnementaux, toxines...). Les métabolites de sources endogènes peuvent être subdivisés en deux sous-groupes : les métabolites primaires⁶ et secondaires⁷. Les métabolites primaires sont ubiquitaires et nécessaires à la survie, au développement et à la reproduction d'un organisme ou d'une cellule tandis que les métabolites secondaires peuvent être propres à un règne ou à une espèce donnée et assurent donc des fonctions spécifiques. L'étude globale ou large spectre de ces composés permet de connaître la réponse d'un organisme à un facteur impactant son fonctionnement. L'analyse métabolomique a ainsi pour objectif d'aboutir à une meilleure compréhension des systèmes biologiques en mettant en évidence des interrelations métaboliques, *via* le suivi de plusieurs centaines de descripteurs moléculaires, qui n'auraient pu être détectées par des approches biochimiques conventionnelles. La complexité d'analyse de l'ensemble de ces métabolites est principalement due à la grande diversité chimique de ces composés. Cela a conduit au développement et à l'utilisation (potentiellement combinée) de différentes méthodes et stratégies analytiques. Les principales techniques d'analyse

métabolomique sont la RMN et la spectrométrie de masse. Le laboratoire où j'ai effectué ma thèse étant spécialisé en analyse métabolomique par couplage de la chromatographie liquide à la spectrométrie de masse, les paragraphes qui suivent traitent principalement de cette méthode d'analyse.

B – Déroulement d'une analyse métabolomique non ciblée

Les analyses métabolomiques comportent différentes étapes depuis la préparation des échantillons à l'identification des métabolites d'intérêt et l'interprétation biologique en passant par les analyses LC-HRMS⁸⁻¹⁰. La réalisation de ces différentes étapes nécessite du temps et chaque étape peut induire des biais particuliers qui peuvent perturber l'analyse métabolomique^{5,11}. Ces facteurs peuvent fausser l'interprétation biologique finale. La méthodologie de travail utilisée au laboratoire est présentée de façon schématique en Figure 2⁵.

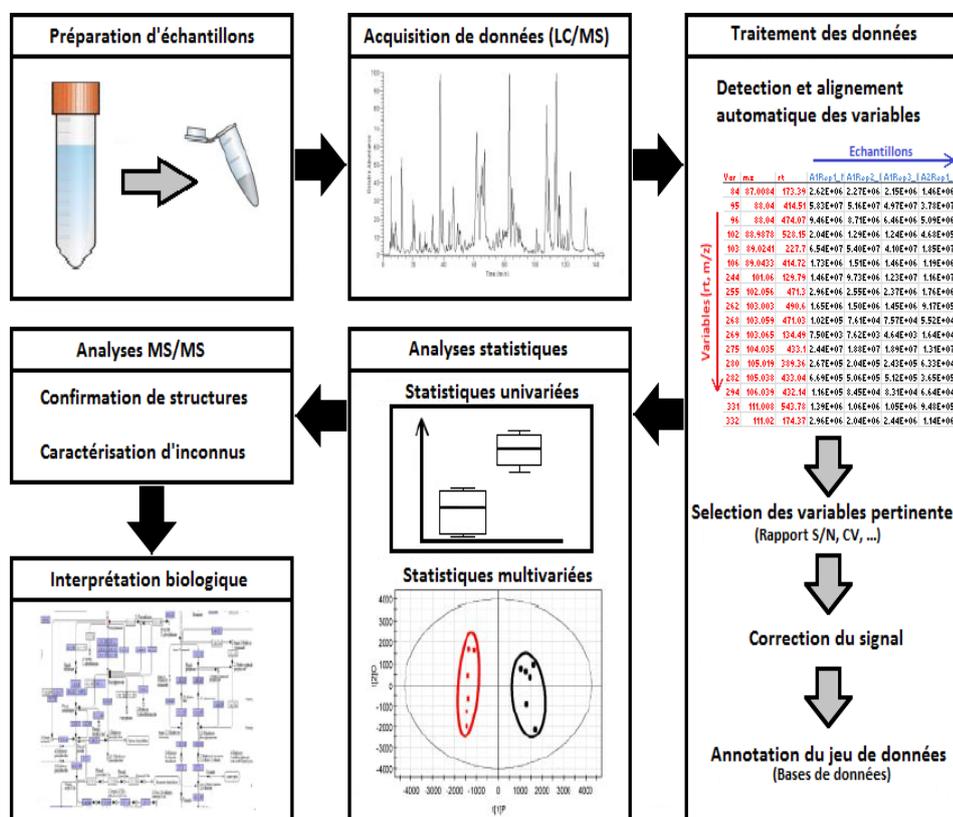


Figure 2 : Déroulement d'une analyse métabolomique non ciblée (adaptée de Junot et al.).

La première étape concerne la préparation d'échantillon. Le but est d'extraire un maximum de métabolites présents dans un échantillon d'intérêt et ce, de façon non biaisée. Cette étape est très dépendante du type d'échantillon analysé. Pour les échantillons de plasma ou sérum, cela consiste essentiellement à précipiter les protéines^{12,13}. De nombreuses études ont étudié l'impact du solvant organique utilisé pour la précipitation de ces protéines et se rejoignent souvent sur l'utilisation du méthanol pour obtenir le profil

métabolique le plus complet et le plus robuste^{14,15}. Tout comme le plasma, l'urine est un type d'échantillon très fréquent pour les analyses métabolomiques car leur prélèvement est non invasif^{15,16}. Cependant, les urines peuvent être plus ou moins diluées selon l'heure de la journée ou l'état d'hydratation du sujet ce qui, pour des contraintes de normalisation, rend plus complexe le profilage de ces échantillons. Cette normalisation peut être réalisée, par exemple, avec la créatinine urinaire ou la gravité spécifique¹⁷. En revanche, les urines contiennent peu de protéines (en raison de l'action naturelle de filtration des reins). La préparation de ces échantillons consiste donc souvent en une simple centrifugation pour éliminer les composants cellulaires puis une dilution avec de l'eau (ou avec les solvants utilisés pour l'analyse de ces échantillons)^{16,18,19}. D'autres types d'échantillons nécessitent des protocoles d'extraction des métabolites plus complexes. C'est le cas des tissus biologiques ou des organes prélevés par biopsie²⁰. Ces échantillons étant hétérogènes, lorsque l'ensemble du tissu doit être analysé et non une zone spécifique, une homogénéisation est souvent nécessaire avant l'extraction des métabolites (si l'objectif est d'analyser l'entièreté de l'échantillon et non une région spécifique). La quantité de tissu variant d'un échantillon à l'autre, le volume de solvant d'extraction est ajusté pour chaque échantillon. En général, pour la préparation des échantillons, il est conseillé d'extraire les métabolites en une étape pour limiter les risques de dégradation de l'échantillon ou de perte de métabolites. Dans le cas des tissus biologiques et des organes, les meilleures performances ont été obtenues avec un protocole biphasique en deux étapes²⁰⁻²². La première étape consiste à extraire les composés polaires avec un mélange eau/méthanol et la deuxième étape permet d'extraire les composés apolaires avec un mélange de solvants organiques tel que chloroforme/méthanol ou dichlorométhane/méthanol^{21,22}. La préparation de l'échantillon, le protocole d'extraction des métabolites et la normalisation des extraits à l'aide de la quantification des protéines totales de l'échantillon (dans le cas de tissus par exemple) sont donc des étapes primordiales pour obtenir l'empreinte métabolique de l'échantillon la plus complète et la plus robuste possible.

La deuxième étape d'une analyse métabolomique consiste à acquérir les données à proprement dites. Au sein de notre laboratoire, les échantillons sont analysés par chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS), généralement sur un instrument de type Exactive (Orbitrap) ne possédant pas la capacité de réaliser des expériences de MS/MS avec sélection d'ions parents. Cette analyse permet d'avoir une vision globale de l'échantillon et d'obtenir son empreinte métabolique. En raison de leur complémentarité, les colonnes C18 et HILIC sont les plus couramment utilisées pour couvrir au mieux la diversité physico-chimique du métabolome. L'ensemble des méthodes séparatives sont détaillées dans une prochaine section de ce chapitre. Les échantillons sont analysés aléatoirement selon une séquence prédéfinie comportant des échantillons de contrôle qualité (QC) (Figure 3)²³. Les QCs sont des mélanges de l'ensemble des échantillons à analyser et représentent donc des échantillons « moyens ». Ces QCs sont analysés tout au long de la séquence à intervalles réguliers et à différentes dilutions. Ils jouent un rôle crucial car permettent à la fois de mettre en évidence les signaux analytiquement pertinents et de corriger de possibles biais analytiques, comme les décalages de temps de rétention ou l'encrassement partiel de la source du spectromètre lors des analyses^{20,24,25}. Afin d'éviter les facteurs confondants liés à des paramètres physiologiques (sexe, âge, poids), les échantillons sont analysés selon un ordre aléatoire⁵ (Figure 3).

N° d'injection	Echantillon	N° d'injection	Echantillon
1	Blanc	17	Echantillon_4
2	QC_dil8	18	Echantillon_7
3	QC_dil8	19	Echantillon_19
4	QC_dil8	20	Echantillon_17
5	QC_dil4	21	Echantillon_36
6	QC_dil4	22	Echantillon_50
7	QC_dil4	23	Echantillon_45
8	QC_dil2	24	Echantillon_2
9	QC_dil2	25	Echantillon_22
10	QC_dil2	26	Blanc
11	QC_dil1	27	QC_dil1
12	QC_dil1	28	Echantillon_12
13	QC_dil1	29-37	...
14	Blanc	38	Blanc
15	QC_dil1	39	QC_dil1
16	Echantillon_31	40-x	...

dil1 ; dil2 ; dil4 ; dil8 = facteurs de dilution des QC

Figure 3 : Exemple d'une séquence analytique.

Les données ainsi obtenues sont ensuite traitées à l'aide d'outils informatiques en libre accès ou fournis par les constructeurs de spectromètres de masse²⁶⁻²⁸. Les plateformes telles que Workflow4Metabolomics (W4M)²⁹, MetaboAnalyst³⁰ ou encore XCMS Online³¹ permettent de détecter automatiquement les signaux LC-HRMS dans chaque échantillon, de regrouper les signaux similaires présents dans différents échantillons, d'aligner les temps de rétention et d'intégrer les pics. Au laboratoire, nous utilisons essentiellement la solution W4M développée au sein de l'infrastructure nationale de métabolomique et fluxomique, MetaboHUB. Après extraction et alignement des signaux, une matrice complexe de données est obtenue avec l'aire du signal mesurée dans chaque échantillon et pour chaque variable, c'est-à-dire pour chaque couple m/z et temps de rétention. À ce stade, la matrice contient des milliers de lignes qui correspondent à des milliers de variables. Il est donc nécessaire de trier cette matrice pour ne garder que les signaux d'intérêt. Pour cela, les signaux détectés dans les échantillons et les QCs sont comparés avec ceux détectés dans les blancs afin de supprimer les signaux non pertinents, non caractéristiques des échantillons analysés. L'analyse des QCs à différentes dilutions permet de savoir si les variables détectées proviennent des échantillons (dont le signal corrèle avec le facteur de dilution) ou s'il s'agit de signaux non pertinents^{32,33}. Un tri est également effectué avec la variation des aires des signaux dans les QCs. Une variable avec des aires fluctuantes dans les différents QCs, c'est-à-dire avec un coefficient de variation supérieur à 30 % dans notre cas, et avec une réponse analytique non proportionnelle en fonction des facteurs de dilution des QCs, sera supprimée de la liste. Les variables restantes sont ensuite annotées grâce à des bases de données internes puis publiques³⁴⁻³⁶. Le processus d'annotation sera détaillé dans le chapitre suivant.

Selon le nombre d'échantillons présents dans l'étude, une première étape de normalisation des signaux LC-HRMS peut être nécessaire. En effet, l'étude de grandes cohortes médicales implique un grand nombre d'échantillons généralement analysés sur plusieurs semaines. La normalisation (informatique) des données à partir des échantillons de QCs peut permettre de comparer les analyses effectuées lors de séquences d'acquisition distinctes ou effectuées à des temps différents (Figure 4)³⁷. La Figure 4 illustre la répartition de 826 échantillons de sérum analysés en quatre « batches » avant et après normalisation. Bien que généralement efficace, cette approche ne permet pas toujours de normaliser efficacement des données acquises avec des grands intervalles de temps³⁸. Dans ces conditions, la méthode de choix est l'utilisation d'étalons ou standards internes pour la normalisation des données voire la quantification large-spectre de métabolites³⁹⁻⁴¹.

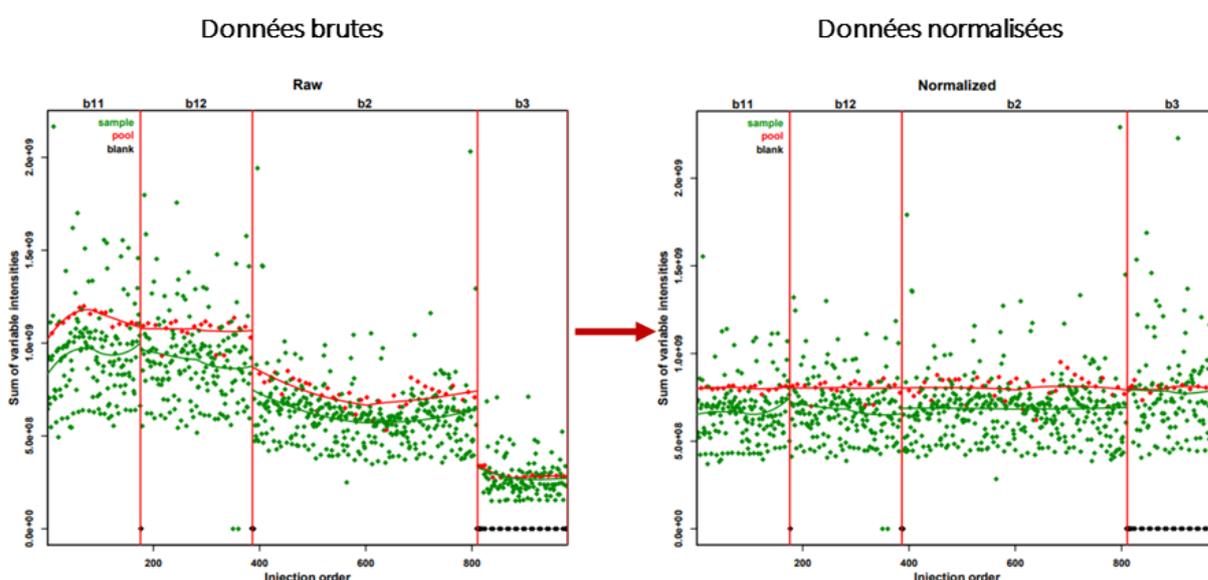


Figure 4 : Illustration de l'impact de la normalisation de quatre « batches » distincts d'analyse (cohorte de 826 échantillons de sérum).

Les points rouges correspondent aux QCs et les points verts correspondent aux échantillons.

Une fois les jeux de données normalisés et triés, les analyses statistiques permettent de mettre en évidence les métabolites discriminants par rapport au contexte biologique d'intérêt. Deux types d'analyses statistiques sont principalement utilisées : les analyses univariées et les analyses multivariées⁴². Les analyses statistiques univariées permettent de comparer deux à deux les variables. Les analyses statistiques multivariées permettent de visualiser les jeux de données et comparer des groupes de variables entre eux⁴³. Ces analyses peuvent être supervisées ou non. Les analyses non-supervisées permettent de visualiser les groupes sans a priori et donnent la tendance globale des variables. Les analyses supervisées orientent les statistiques selon certains facteurs précis⁴⁴.

La cinquième étape a pour objectif d'identifier les variables d'intérêt et de confirmer ou non l'annotation réalisée. Pour cela, des spectres MS/MS sont acquis afin de les comparer à des bases de données spectrales³⁴⁻³⁶. Cette étape sera détaillée dans le chapitre suivant.

Au laboratoire, les analyses de profilage sont réalisées à l'aide d'un Exactive (Thermo Fisher Scientifique) qui n'a pas la capacité d'acquérir, de façon ciblée, des spectres MS/MS. Les acquisitions MS/MS sont ensuite réalisées sur un instrument de type Q-Exactive ou Orbitrap Fusion. Dans ces conditions, les analyses MS et MS/MS sont réalisées avec deux spectromètres de masse différents à des temps également différents. Cela peut malheureusement s'accompagner de biais analytiques tels que des dérives de temps de rétention ou encore une dégradation des échantillons. Une illustration des dérives des temps de rétention est donnée en Figure 5 avec l'exemple du tryptophane et de l'acide quinolinique. Deux analyses réalisées dans des conditions « HILIC négatif » (forme [M-H]⁻ des deux composés) ont été effectuées à 10 mois d'intervalles et ont conduit à un décalage de temps de rétention d'une minute pour l'acide quinolinique⁴⁵.

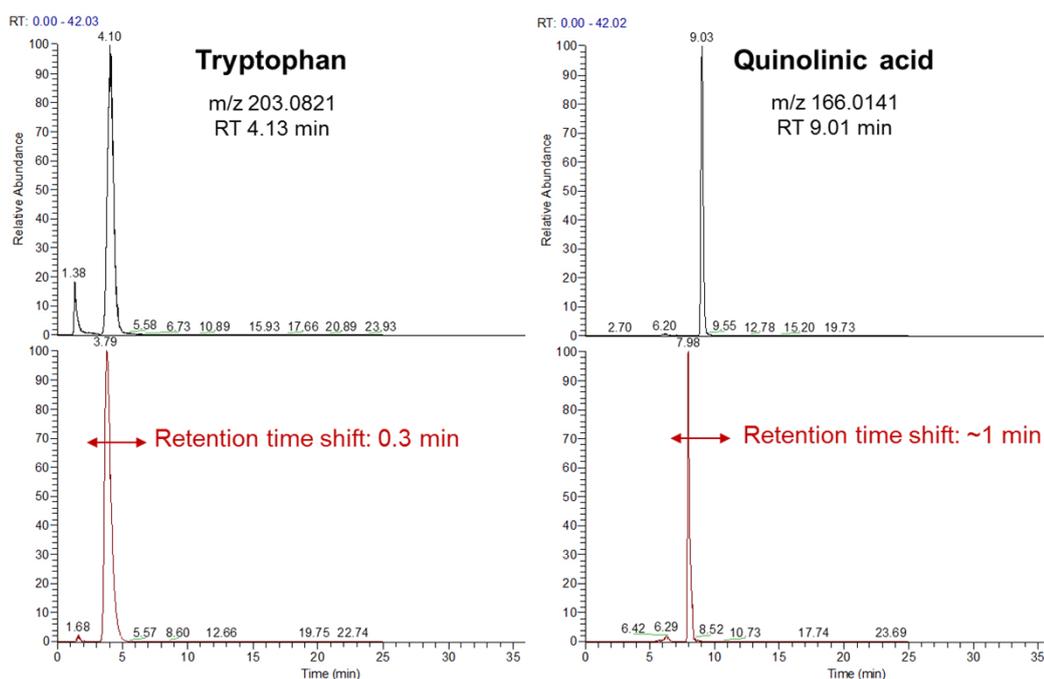


Figure 5 : Exemples de décalage de temps de rétention lors de l'analyse d'un même échantillon à 10 mois d'intervalle.

Dans le cadre de ma thèse, nous avons cherché à faire évoluer cette méthodologie vers une approche combinant les acquisitions MS et MS/MS grâce à l'Orbitrap Fusion. L'identification des métabolites par exploitation et interprétation des spectres MS/MS générés reste aujourd'hui un challenge lors des analyses métabolomiques⁴⁶, et sera aussi abordé dans la Partie B – Chapitre 1 de ce manuscrit.

La dernière étape d'une analyse métabolomique est l'interprétation biologique. Cette interprétation est basée sur les voies métaboliques. Grâce à des bases de données, il est possible de savoir dans quelle voie métabolique les métabolites d'intérêt sont impliqués et d'évaluer leur influence sur les autres métabolites de la même voie ou les autres voies liées^{47,48}. Le métabolome n'étant pas identifié dans sa totalité, les interactions entre l'ensemble des voies métaboliques ne sont également pas toutes connues. Aujourd'hui, il est estimé qu'environ 10 % des signaux détectés par LC-HRMS sont annotés alors que certaines bases de données comme

HMDB référence plus de 114 000 métabolites (prédits ou détectés)^{35,46,49}. Ces chiffres soulignent à eux seuls la difficulté d'identifier des métabolites mais également la marge de progression existante. L'interprétation biologique nécessite l'intervention d'experts dans le domaine d'application (généralement des médecins) pour comprendre les processus mis en jeu.

C – Les techniques d'analyse pour la métabolomique

1 – Vue d'ensemble des méthodes analytiques les plus utilisées

Au cours du développement des analyses métabolomiques, différentes techniques analytiques ont été utilisées pour détecter ces molécules de faibles poids moléculaires⁵⁰. Dans les années 1970, le terme « profilage métabolique » est apparu pour désigner les analyses qualitatives et quantitatives de mélanges complexes de substances physiologiques et chimiques. Ces premières analyses ont été réalisées par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS) pour détecter des stéroïdes, des acides (aminés et organiques) et/ou des métabolites de médicaments^{51,52}. Dans les années 1980, les premiers articles référençant des études de profilage métabolique par RMN sont publiés^{53,54}. De nombreuses méthodes ont été développées et testées depuis lors telles que la spectroscopie infrarouge⁵⁵ ou la chromatographie sur couche mince⁵⁶. Plus tard, les analyses ont été réalisées grâce à différents types de méthodes séparatives, comme l'électrophorèse capillaire (CE) couplée à un détecteur UltraViolet/visible⁵⁷ ou à un détecteur de fluorescence induite par laser (par exemple après dérivatisation chimique des acides aminés étudiés)⁵⁸ ou encore à un spectromètre de masse⁵⁹. Le couplage CE-MS sera plus particulièrement détaillé dans la partie suivante. Plus couramment, les chromatographies en phase liquide ou gazeuse sont couplées à un spectromètre de masse (LC-ESI-MS et GC-EI-MS)^{60,61}. Aujourd'hui, la RMN et la spectrométrie de masse sont les deux méthodes analytiques les plus utilisées en métabolomique (Figure 6).

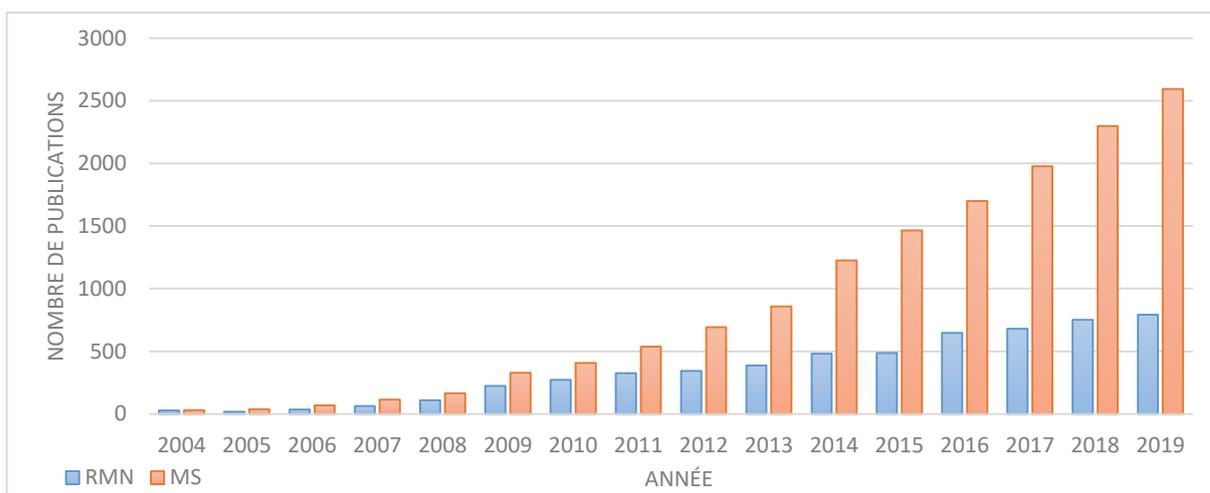


Figure 6 : Nombre de publications en métabolomique au cours de ces 15 dernières années avec des analyses impliquant la RMN ou la spectrométrie de masse.

(Références obtenues via PubMed et les recherches : metabolomics AND (nuclear magnetic resonance OR NMR) ; metabolomics AND (mass spectrometry OR MS))

2 – La spectrométrie de masse

La spectrométrie de masse est aujourd’hui la méthode analytique la plus utilisée pour les études métabolomiques. Les spectromètres de masse ont pour fonction de séparer les ions formés lors de l’étape d’ionisation en source, selon leur rapport de masse sur charge (m/z). Un spectromètre de masse comporte une source d’ionisation, un analyseur, un détecteur et un système de retraitement du signal (Figure 7).

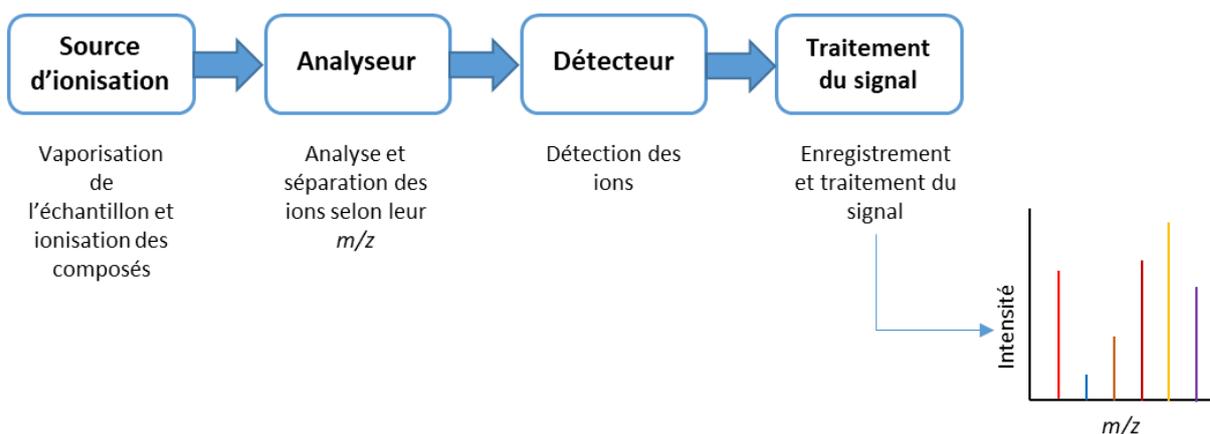


Figure 7 : Schéma d'un spectromètre de masse.

Les spectromètres de masse actuels sont très majoritairement des spectromètres de masse en tandem qui permettent de réaliser des expériences de fragmentation résonante ou non résonante d'ions d'intérêt. L'étude des fragments obtenus permet d'identifier le ou les composés dont les ions parents ont été préalablement sélectionnés.

Dans ce manuscrit, nous présenterons uniquement une description de quelques sources d'ionisation ainsi que d'analyseurs utilisés pour les analyses métabolomiques.

a – Les principales sources d'ionisation

La source d'ionisation permet d'ioniser des composés arrivant dans le spectromètre de masse et les rend donc détectables. Plusieurs types de sources existent et sont utilisées principalement en fonction de la méthode séparative avec laquelle elles sont couplées et des métabolites à détecter. Deux classes de sources d'ionisation se distinguent : les sources sous vide et les sources à pression atmosphérique. Dans les paragraphes qui suivent, nous présenterons très succinctement les sources d'ionisation par impact électronique et électrospray dont sont le plus souvent équipés les systèmes GC-MS ou LC-MS communément utilisés pour la métabolomique.

- Les sources d'ionisation sous vide

Les sources d'ionisation sous vide sont principalement utilisées avec un couplage avec la GC. L'ionisation par impact électronique (EI) est la méthode la plus ancienne qui a permis la création de base de données spectrales⁶². En effet, cette méthode d'ionisation possède l'avantage de produire des spectres très reproductibles d'un spectromètre de masse à l'autre. Cette reproductibilité est liée à la large distribution d'énergie interne déposée sur les ions et également à la cinétique de dissociation qui est en général bien plus rapide que le temps de séjour des ions dans la source. Les composés issus de la GC vont être bombardés par un faisceau d'électrons avec une énergie cinétique de 70 eV ce qui va former des ions radicaux. Il s'agit d'une méthode d'ionisation couramment qualifiée de « dure » car l'énergie du faisceau d'électrons induit une fragmentation importante. Il n'est donc pas toujours possible de détecter l'ion précurseur dans ces conditions⁶³. En revanche, les spectres obtenus sont hautement reproductibles contrairement aux spectres obtenus par LC-ESI-MS/MS. Les analytes les plus adaptés à l'EI sont les composés relativement volatiles naturellement (ou rendus volatiles par modification chimique) et thermiquement stables.

Une alternative plus douce à l'EI est l'ionisation chimique. Un gaz réactif est ajouté dans la source et est ionisé par un faisceau d'électrons. Les ions obtenus vont entrer en collision avec l'analyte et par transfert de proton vont l'ioniser⁶³. Ce sont des cations qui sont principalement formés par cette méthode.

- Les sources d'ionisation à pression atmosphérique

Les sources d'ionisation à pression atmosphérique sont principalement utilisées pour les couplages du spectromètre de masse avec la chromatographie liquide ou l'électrophorèse capillaire. L'ionisation chimique à pression atmosphérique (APCI) est l'une d'entre elles. Les analytes sont introduits dans un nébuliseur pneumatique pour former un aérosol. Ils sont emportés par un gaz présent dans la source vers une chambre de désolvatation où ils sont chauffés. L'analyte et la phase mobile se vaporisent sous l'effet de la chaleur et sont conduits jusqu'à la chambre de réaction. La vapeur obtenue passe à travers un flux d'électrons émis par des décharges Corona (décharges électriques de plusieurs kV). Les ions primaires obtenus à partir des molécules présentes dans l'atmosphère vont échanger leur charge avec les molécules du solvant (ions secondaires). L'ionisation de l'analyte va se faire soit par capture d'électron, soit par soustraction de proton avec les ions secondaires^{64,65}. Cette réaction est similaire à celle de l'ionisation chimique sous vide. Cette technique est bien adaptée pour les composés polaires et de masses inférieures à 1500 Da.

La photoionisation à pression atmosphérique (APPI) représente une autre méthode d'ionisation. La première application de cette source pour le couplage LC-MS date de 2000⁶⁶. La source APPI est très similaire à la source APCI où l'aiguille Corona est remplacée par une lampe à photons^{67,68}. L'APCI et l'APPI sont très sensibles aux conditions expérimentales tel que le choix du solvant ou des réactifs. Il est donc nécessaire d'avoir une bonne préparation d'échantillon pour ne pas avoir d'interférences⁶³. L'APCI et l'APPI sont considérées comme des méthodes douces d'ionisation.

Une troisième méthode d'ionisation atmosphérique, qui est la plus répandue aujourd'hui, est l'ionisation par électrospray (ESI⁶⁹). Un fort champ électrique est appliqué dans un capillaire où passe l'échantillon liquide. Cela induit une accumulation de charge à la surface du liquide et produit un cône de Taylor à l'extrémité du capillaire. Ce cône se décompose pour libérer des gouttelettes fortement chargées. Ces gouttelettes sont ensuite chauffées puis désolvatées (Figure 8⁷⁰).

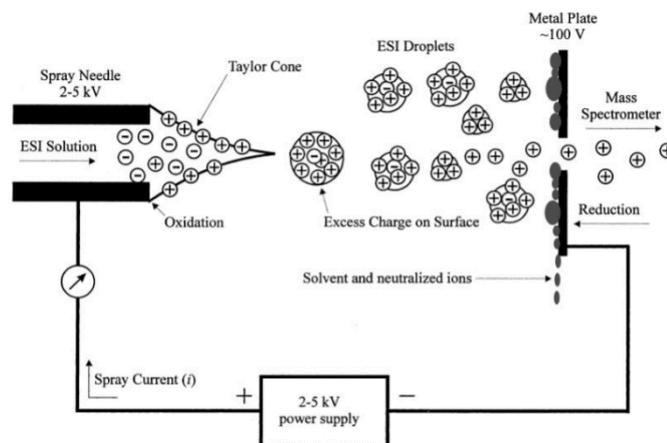


Figure 8 : Schéma d'une source ESI (extraite de Cech et al.).

Il s'agit également d'une méthode d'ionisation douce plus efficace pour les composés polaires et qui implique peu de fragmentation(s) en source⁶³.

La Figure 9⁷¹ résume les conditions optimales des trois sources sus-mentionnées à pression atmosphérique.

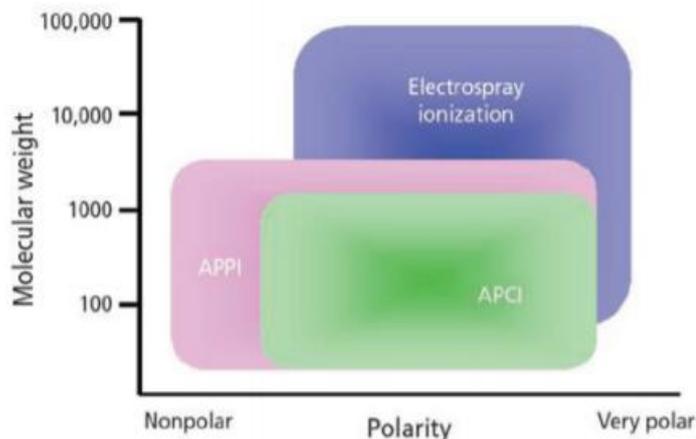


Figure 9 : Récapitulatif des conditions optimales pour l'utilisation des sources d'ionisation à pression atmosphérique (extraite de Syage et al.).

Appliquée à la métabolomique, l'ESI apparaît comme la méthode la plus polyvalente pour ioniser efficacement un grand nombre de métabolites. A titre d'exemple, la Figure 10 présente une comparaison des trois méthodes d'ionisation à pression atmosphérique présentées ci-dessus et montre que pour les centaines de signaux détectés dans le plasma humain avec l'ensemble de ces sources d'ionisation, 91 % d'entre eux sont détectés avec l'ESI alors que seulement 25 % et 29 % d'entre eux sont respectivement détectés avec l'APCI et l'APPI⁷² (Figure 10⁷²). A noter que 71% des métabolites ne sont détectés que par ESI.

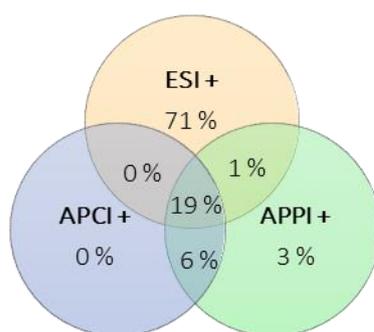


Figure 10 : Comparaison et complémentarité des trois méthodes d'ionisation à pression atmosphérique ESI, APCI et APPI pour l'analyse métabolomique d'un échantillon de plasma humain (adaptée de Tian et al.).

Pour ces raisons, l'ESI est la méthode d'ionisation la plus utilisée en métabolomique quel que soit le type d'échantillon analysé (plasma, urine, tissu...)^{18,72,73}.

b – Les principaux analyseurs

Les analyseurs peuvent être classés, comparés, en fonction de différentes caractéristiques telles que la résolution, la précision de mesure de masse, la sensibilité et la vitesse d'acquisition. La résolution en masse (ou le pouvoir résolutif) d'un analyseur correspond à sa capacité à séparer des ions de rapports m/z proches. La résolution est le plus souvent définie comme découlant de la largeur des pics à mi-hauteur (Equation 1). Elle peut également être caractérisée comme la capacité à séparer deux pics à 50 % de leur hauteur maximale (Equation 2).

$$R = \frac{m}{\Delta m_{FWHM}} \quad (\text{Equation 1})$$

$$R = \frac{m}{\Delta m} \quad (\text{Equation 2})$$

où m correspond au rapport m/z du composé d'intérêt, Δm_{FWHM} la largeur à mi-hauteur du pic mesuré (avec $FWHM$ pour « Full Width at Half Maximum ») et Δm la différence de masse la plus petite entre deux pics bien résolus à mi-hauteur.

La précision de mesure de masse, assimilée à une erreur de masse, correspond à la justesse avec laquelle l'analyseur va mesurer le rapport m/z de l'ion en fonction de la valeur théorique. Cette donnée est souvent exprimée en parties par million (ppm) et est obtenue selon l'équation 3.

$$\text{Erreur de masse} = \frac{|m(\text{théorique}) - m(\text{mesurée})|}{m(\text{théorique})} \times 10^6 \quad (\text{Equation 3})$$

où $m(\text{théorique})$ correspond au m/z théorique et $m(\text{mesurée})$ au m/z obtenu expérimentalement.

La vitesse d'acquisition d'un analyseur, aussi appelée fréquence d'acquisition, correspond au nombre de spectres que peut acquérir un analyseur pendant une durée déterminée. S'agissant d'une fréquence, cette valeur est donnée en Hertz ce qui correspond au nombre de spectres obtenus en une seconde. Pour les analyseurs à piègeage d'ions comme les analyseurs de type Orbitrap, la résolution est directement dépendante de la vitesse d'acquisition.

- *Les analyseurs basse résolution*

Les quadripôles et les pièges à ions linéaires ont été les premiers analyseurs utilisés⁷⁴. Un quadripôle est composé de quatre électrodes cylindriques ou hyperboliques reliées électriquement deux à deux. Ces électrodes sont soumises à deux potentiels, un continu et l'autre radiofréquence, déphasé de 180°. Cela crée un champ quadripolaire à l'intérieur de ces quatre électrodes. En entrant dans le quadripôle, les ions gardent leur vitesse longitudinale mais vont subir des oscillations à cause du champ quadripolaire (les trajectoires des ions peuvent être définies par les équations de Mathieu non développées dans ce manuscrit). Ces oscillations vont varier selon

les m/z des ions ce qui permettra de les séparer et les filtrer. Les ions avec des oscillations instables, ne pourront pas atteindre l'analyseur (Figure 11⁷⁵).

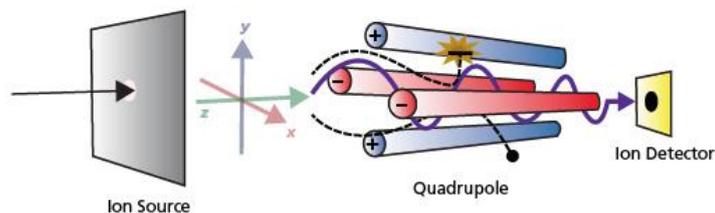


Figure 11 : Schéma d'un analyseur quadripolaire (extraite de www.shimadzu.com).

Les pièges à ions linéaires de type « Linear Trap Quadrupole (ou LTQ) », tels que ceux équipant les instruments de type LTQ-Orbitrap, sont composés de quatre électrodes à section hyperbolique. Les ions vont d'abord être piégés à l'intérieur des quatre électrodes puis, par l'application de radiofréquences plus ou moins élevées, les ions vont avoir des oscillations plus ou moins stables en fonction de leur rapport m/z . Cela implique que les ions avec les oscillations instables vont être éjectés du système pour être détectés^{74,76}.

- *Les analyseurs à haute résolution*

Les innovations en spectrométrie de masse sont actuellement majoritairement conduites par et pour les sciences « omiques » pour lesquelles des échantillons de plus en plus complexes doivent être analysés⁷⁷. Les analyseurs haute résolution permettent, comme leur nom l'indique, une meilleure résolution des composés et également une meilleure précision de mesure de masse. Il en existe deux familles principales : les analyseurs à Temps de vol (TOF) et ceux à Transformée de Fourier comme les analyseurs de type Orbitrap et ceux à Résonance Cyclonique des Ions (FT-ICR). Les analyseurs de type TOF font partie des premiers analyseurs utilisés en spectrométrie de masse. Cependant, dans les années 1960, par manque de technologies permettant d'enregistrer et traiter les spectres rapidement (de l'ordre de la milliseconde), leur résolution ainsi que leur sensibilité sont inférieures à celles des quadripôles qui leur sont préférés⁷⁸. En effet, la dispersion énergétique des ions à l'entrée du TOF entraîne une grande perte de résolution. Cette dispersion a pu être corrigée pour être réduite par refocalisation des ions sur le plan temporel et également sur le plan énergétique, via l'extraction retardée des ions⁷⁹ ou l'utilisation d'un réflectron⁸⁰. Lorsque les technologies pour améliorer la résolution, l'enregistrement et le traitement des spectres ont été développées, les analyseurs TOF apparaissent comme une très bonne solution pour le profilage et la quantification d'échantillons. Dans un spectromètre de type TOF, les ions sont séparés selon le temps passé dans le tube de vol (c'est-à-dire dans une région libre de champ électrique). Un paquet d'ions arrive au niveau d'une électrode d'accélération qui, grâce à une impulsion, va accélérer tous les ions avec la même énergie. La vitesse des ions dans le tube de dérive étant directement proportionnelle à l'énergie cinétique et à la masse des ions, les ions les plus lourds auront une vitesse de déplacement plus faible que les ions de petites masses, ils mettront donc plus de temps pour arriver à l'analyseur. La taille du tube de vol,

et donc la distance de vol des ions, a un impact direct sur la résolution des signaux. En effet, plus le tube de dérive sera grand, plus il sera possible de séparer des composés de m/z proches. Comme nous l'avons mentionné ci-dessus, pour améliorer la résolution et donc augmenter la distance de vol des ions, des réflectrons ont été développés. Les ions les plus énergétiques, les plus lourds, vont entrer plus en profondeur dans le réflectron et vont donc mettre plus de temps à être réfléchis. L'utilisation d'un réflectron permet d'augmenter la distance de vol des ions sans augmenter la taille du spectromètre de masse et de corriger les possibles écarts d'énergie de cinétique afin d'améliorer la résolution⁷⁸ (Figure 12⁷⁵). Les avantages actuels des TOF sont leur grande vitesse d'acquisition.

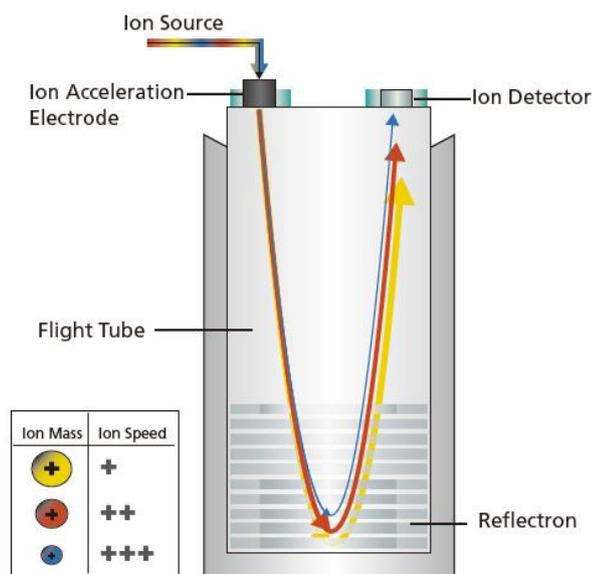


Figure 12 : Schéma d'un analyseur de type temps de vol (extraite de www.shimadzu.com).

Dans la partie suivante, une description des instruments à transformée de fourier est réalisée. Les pièges, tels que l'Orbitrap et le FT-ICR, sont des analyseurs dont le fonctionnement repose sur un principe comparable. Ces derniers sont soumis à l'accumulation d'un nombre important d'ions pour lesquels la détection n'est pas destructive. En effet, les ions sont détectés par l'intermédiaire d'un courant induit sans la nécessité d'être éjectés vers les détecteurs. L'analyseur de type Orbitrap est le plus récent, il a été décrit par Alexander Makarov 1999^{81,82}. Il s'agit d'un analyseur électrostatique à Transformée de Fourier. Cet analyseur, uniquement commercialisé par la société Thermo Fisher Scientific, est composé de deux électrodes : une électrode centrale en forme de fuseau et une électrode extérieure co-axiale (Figure 13⁸³). Les ions sont piégés à l'intérieur de l'analyseur par un champ électrostatique créé par ces deux électrodes. Les ions vont alors avoir une trajectoire stable qui combine la rotation autour de l'électrode centrale avec des oscillations harmoniques le long de celle-ci. Les oscillations axiales permettent d'obtenir le rapport m/z des ions grâce à une Transformée de Fourier rapide du courant détecté sur les deux moitiés de l'électrode externe^{77,83,84}. Les analyseurs de type Orbitrap sont assez récents et évoluent rapidement. Leurs caractéristiques se sont considérablement améliorées ces dernières années ce qui les rend de plus en plus comparables en terme de résolution spectrale aux analyseurs à très haute résolution tels que les analyseurs à Résonance Cyclonique des ions (FT-ICR). En effet, en 2005, les Orbitraps pouvaient atteindre

une résolution de 150 000 (à m/z 200) et une précision entre 2 et 5 ppm⁷⁷. Aujourd'hui, les Orbitraps peuvent atteindre une résolution de 1 000 000 (à m/z 200) avec une précision de mesure de masse inférieure au ppm^{85,86}.

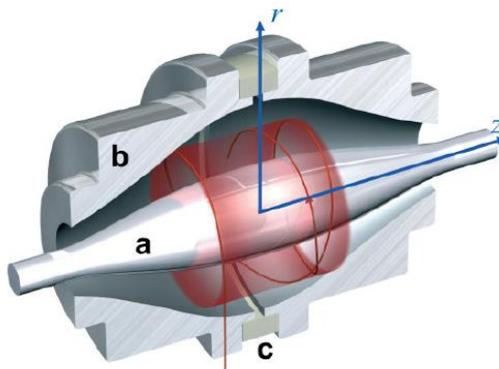


Figure 13 : Schéma d'un analyseur de type orbitrap (extraite de Scigelova et al).

Les analyseurs FT-ICR sont des analyseurs à très haute résolution qui ont vu le jour en 1974 sous l'impulsion de Melvin B. Comisarow et Alan G. Marshall⁸⁷. Cette technique consiste à soumettre les ions à un fort champ magnétique dans une cellule ICR (Ion Cyclotron Resonance). Ce champ entraîne les ions à adopter un mouvement circulaire dans la cellule dont la fréquence dépend du rapport m/z de l'ion. Une cellule cyclotronique se compose de trois paires de plaques qui sont assemblées en cube. Les plaques à chaque extrémité se comportent comme des plaques de piégeage pour maintenir les ions à l'intérieur de la cellule cyclotronique. Les deux plaques d'excitation sont reliées avec un émetteur de radiofréquences pour exciter les ions. Les plaques de détection enregistrent le courant des ions. Le fonctionnement de ces analyseurs est basé sur le principe du cyclotron dans un champ magnétique uniforme. Les ions sont piégés dans la cellule cyclotronique et se déplacent selon une trajectoire orbitale. Lorsqu'une impulsion radiofréquence est envoyée à la cellule, les ions se déplaceront selon une nouvelle orbite. Plus cette impulsion sera importante, plus l'orbite suivie par les ions sera grande et donc plus les ions se rapprocheront des plaques de détection. Ce signal est traité par Transformée de Fourier pour obtenir les spectres de masse^{88,89}. Cet analyseur n'est pas très utilisé en routine car très cher mais est surtout caractérisé par un temps d'acquisition relativement lent⁹⁰. De plus, sa faible vitesse d'acquisition le rend peu compatible avec le couplage avec des méthodes séparatives comme la chromatographie liquide. Pour cette raison, pour les analyses métabolomiques, il est souvent couplé avec un système d'introduction directe⁹¹.

c – Les spectromètres de masse en tandem

Afin d'améliorer leurs performances quantitatives et d'acquérir des spectres MS/MS nécessaires à l'identification de métabolites, les spectromètres de masse comportent plusieurs analyseurs. Un des analyseurs sert à filtrer les ions et l'autre permet de les analyser. L'un des spectromètres de masse en tandem le plus connu et le plus utilisé pour des approches de quantification ciblée est le triple quadripôle (QqQ). Comme son nom l'indique, il est composé de trois quadripôles en série. Le premier sert à filtrer les ions, le deuxième est une cellule

de collision qui permet de fragmenter les ions pour des analyses MS/MS et le troisième permet l'analyse des ions fragments⁹⁰. Ce type d'instrument permet d'augmenter la sensibilité et la spécificité de détection par rapport à un quadripôle simple. Ces instruments sont très utilisés pour les analyses ciblées quantitatives en raison de leur grande gamme dynamique^{90,92}.

Pour certains spectromètres de masse, les analyseurs présents sont différents. Ce sont des instruments dits « hybrides ». Les Q-TOFs sont des spectromètres de masse « hybrides »⁹³. Par comparaison, l'analyseur TOF du Q-TOF remplace le troisième quadripôle du triple quadripôle⁹⁴. Les Q-TOFs ont un pouvoir résolutif plus important que les triples quadripôles⁹⁰. Ce sont des spectromètres de masse très utilisés pour le profilage d'échantillon et avec l'amélioration des techniques, montrent de bonnes performances pour la quantification notamment grâce à leur grande vitesse d'acquisition⁹⁵⁻⁹⁷.

D'autres spectromètres de masse « hybrides » très populaires sont les instruments combinant un analyseur de type quadripôle, ou un piège à ions linéaire, et un analyseur de type Orbitrap⁹⁸⁻¹⁰⁰ (respectivement Q-Orbitrap ou Q-Exacte, et LTQ-Orbitrap). Ces instruments permettent donc d'accéder à la haute, voire très haute résolution avec une très bonne précision en masse et une grande gamme dynamique^{101,102}.

Afin d'améliorer encore les performances des spectromètres de masse « hybrides », des instruments avec trois analyseurs ont vu le jour¹⁰³. Un exemple de spectromètre de masse dit « tribride » est l'Orbitrap Fusion de Thermo Fisher Scientific (Figure 14¹⁰⁴). Les développements réalisés au cours de cette thèse ont été effectués avec ce spectromètre de masse.

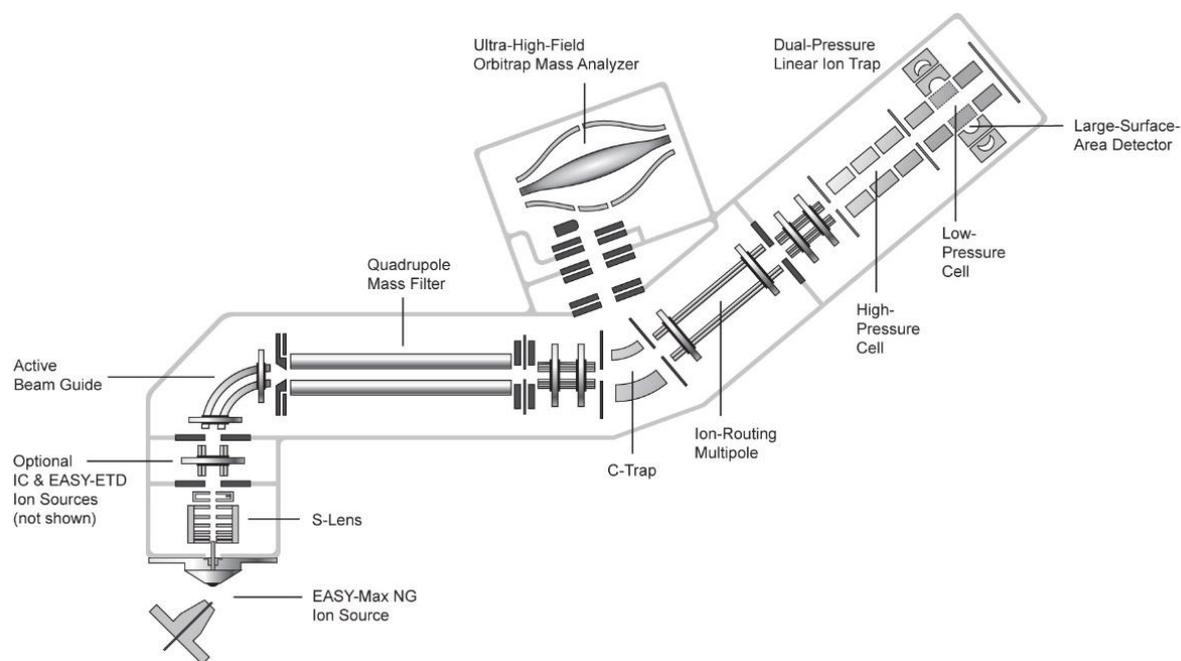


Figure 14 : Schéma de l'Orbitrap Fusion (extraite de planetorbitrap.com).

Ce spectromètre de masse possède donc trois analyseurs distincts : un quadripôle, un piège à ions linéaire ainsi qu'un analyseur de type Orbitrap. Le quadripôle sert de filtre d'ions, l'analyseur de technologie Orbitrap permet de détecter les ions à haute résolution, tandis que le piège à ions linéaire peut permettre à la fois d'isoler et fragmenter les ions d'intérêt mais aussi de les détecter à plus basse résolution. Un avantage important de cet instrument est sa capacité de fragmentation selon deux modes distincts mais complémentaires : la fragmentation de type « Higher-energy Collision Dissociation » (HCD) réalisée dans la cellule de collision et celle de type « Collision Induced Dissociation » (CID) réalisée dans le piège à ions. En effet, pour identifier les métabolites avec le plus haut degré de certitude, il est nécessaire d'acquérir les spectres MS/MS des métabolites d'intérêt présents dans l'échantillon d'étude et de les comparer aux spectres MS/MS des standards purs correspondants¹⁰⁵. La combinaison de différents modes de fragmentation permet d'obtenir des informations complémentaires pour un même composé (cet aspect est discuté et illustré dans le Chapitre 2 ci-après)¹⁰⁶⁻¹⁰⁸. La fragmentation HCD est une fragmentation dite « non résonante ». Les ions de l'analyte entrent en collision avec des molécules de gaz neutre (de l'azote sur cet instrument). Dans ces conditions, les ions vont subir plusieurs fragmentations successives ce qui permet de détecter plusieurs générations d'ions (ions parent, fils, petit-fils...). La fragmentation CID est une fragmentation dite « résonante ». Elle repose sur l'excitation spécifique d'un ion parent par application d'une radiofréquence qui lui est propre¹⁰⁶. Dans l'Orbitrap Fusion cette fragmentation a lieu dans le piège à ions linéaire et donne essentiellement des ions fils de « première génération ». La combinaison de ces différents éléments, analyseurs et modes de fragmentation a pour but d'améliorer au mieux la détection et l'identification des métabolites lors d'une analyse métabolomique.

A titre de comparaison, les caractéristiques de quelques spectromètres de masse « hybrides » à haute et très haute résolution sont présentées dans le Tableau 1. L'ensemble des instruments possède une haute précision de mesure de masse, inférieure au ppm. Cependant, ils diffèrent par leur pouvoir résolutif et leur vitesse d'acquisition. En effet, pour les analyseurs à piégeage d'ions, tels que les analyseurs de type Orbitrap, plus la résolution est élevée, plus la vitesse d'acquisition diminue. Lors d'analyses métabolomiques, il sera donc essentiel de faire des compromis afin d'avoir une bonne séparation en masse des composés et une vitesse d'acquisition suffisante. De plus, une vitesse d'acquisition élevée facilite le couplage avec les méthodes séparatives comme la chromatographie en phase liquide ou gazeuse.

Tableau 1 : Comparaison des caractéristiques de quelques spectromètres de masse à haute et très haute résolution.

(Les données correspondent aux spécifications des constructeurs : www.agilent.com, www.bruker.com et www.thermofisher.com)

Instruments	Précision en masse	Résolution	Vitesse d'acquisition
<i>Haute résolution</i>			
6545 Q-TOF (Agilent)	< 0.8 ppm	45 000 (m/z 2722)	50 Hz (à une résolution de 45 000 à m/z 2722)
tims TOF (Bruker)	< 1 ppm	50 000 (à m/z 1222)	100 Hz (à une résolution de 50 000 à m/z 1222)
Q-Exactive HF-X (Thermo Fisher Scientific)	< 1 ppm	240 000 (à m/z 200)	40 Hz (à une résolution de 7 500 à m/z 200)
<i>Très Haute Résolution</i>			
Orbitrap Fusion Lumos (Thermo Fisher Scientific)	< 1 ppm	500 000 (à m/z 200)	20 Hz (à une résolution de 15 000 à m/z 200)
SolariX MRMS (Bruker)	< 0.6 ppm	60 000 000 (à m/z 702)	1 Hz (à une résolution de 922 675 à m/z 200)

Pour conclure, il convient de noter que chaque instrument présente des avantages et des inconvénients qui lui sont propres. Les différents instruments disponibles peuvent donc apparaître comme complémentaires les uns des autres. En ce qui concerne les Q-TOFs, ils disposent, pour principal avantage, d'un équilibre entre leur précision en masse, leur sensibilité et leur vitesse d'acquisition. Par conséquent, ces instruments, de par leur polyvalence, sont fréquemment utilisés en métabolomique bien que leur résolution soit inférieure à 60 000. En outre, la vitesse d'acquisition de cet analyseur lui permet d'être combiné à des méthodes séparatives liquides ou gazeuses et lui confère ainsi des propriétés intéressantes pour l'étude de matrices ou de systèmes complexes dans les modes MS et MS/MS. De plus, les QTOFs récemment commercialisés disposent de pouvoirs résolutifs encore améliorés conséquents et pourront à terme rivaliser avec les Orbitraps. Les QTOFs récents sont souvent munis d'une cellule de mobilité ionique (voir ci-dessous), ce qui dans une certaine mesure leur permet de compenser leur « manque de résolution ».

L'Orbitrap se caractérise par des mesures de m/z particulièrement précises en routine (de l'ordre de 1 à 2 ppm) et par un pouvoir résolutif important (jusqu'à 500 000 voire 1 000 000), ce qui lui permet de rivaliser à ce niveau avec le FT-ICR. L'Orbitrap est facilement couplée avec des analyseurs de type LTQ ou quadripôle, ce qui lui confère une grande flexibilité d'utilisation et d'application. Le FT-ICR se distingue particulièrement par ses très hautes résolution et précision de mesure de masse (potentiellement quelques ppb), très utiles pour identifier sans ambiguïté des métabolites inconnus.

d – La mobilité ionique

La mobilité ionique permet de séparer les ions selon leur conformation, leur taille et leur charge. Par exemple, il est possible de séparer des isomères ce qui n'est pas toujours possible à réaliser avec les méthodes séparatives décrites par la suite^{109,110}. Les ions, sous forme gazeuse, arrivent dans un tube de dérive rempli de gaz et où ils sont soumis à un champ électrostatique. Lors du couplage avec la spectrométrie de masse (IM-MS), la cellule de mobilité ionique se situe entre la source d'ionisation et l'analyseur. L'IM-MS a été peu utilisée par le passé car elle était considérée comme insuffisamment résolutive et robuste contrairement aux autres méthodes séparatives^{109,111}. Aujourd'hui, pour l'IM-MS comme pour le couplage de la spectrométrie de masse avec d'autres méthodes séparatives, il est nécessaire d'avoir un analyseur avec une bonne vitesse d'acquisition. Pour cela, l'IM-MS est souvent couplé avec un analyseur de type TOF sur des systèmes de type Q-TOF¹⁰⁹. Le couplage avec des analyseurs de type Orbitrap est également possible jusqu'à des résolutions de 70 000 (à m/z 200)¹¹².

Il existe quatre principaux types de mobilité ionique : la mobilité ionique à temps de dérive (DTIMS), à ondes progressives (TWIMS), à champs asymétrique (FAIMS) et à piège (TIMS)^{109,113}. Les cellules DTIMS et TWIMS sont aujourd'hui les plus utilisées¹¹³. Les ions sont séparés selon le temps qu'ils mettent à traverser le tube de dérive. Cela dépend de la forme, de la taille et de la charge de l'ion considéré ainsi que de la nature du gaz présent dans le tube (Figure 15¹¹⁴).

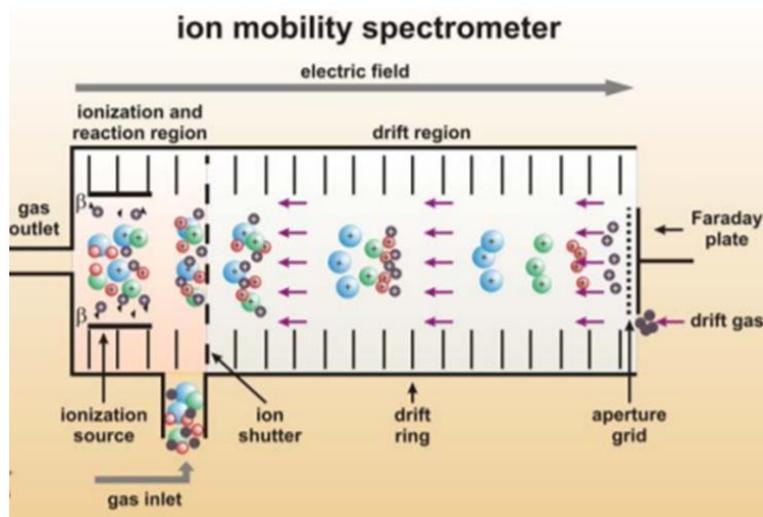


Figure 15 : Schéma d'un instrument de mobilité ionique à temps de dérive (adaptée de Jünger et al.).

Avec la cellule DTIMS, la section transversale de collision (« Collision Cross Section », CCS) peut être directement calculée à partir de l'équation de Mason-Schamp (non présentée ici). En revanche, avec la cellule TWIMS, il est nécessaire d'utiliser des étalons de structure proche du composé d'intérêt pour obtenir la CCS du métabolite d'intérêt¹¹⁵. La CCS représente la surface apparente moyenne du composé. Chaque composé possède sa propre CCS¹¹⁶.

Le couplage LC-IM-MS permet de couvrir une plus grande partie du métabolome que les méthodes LC-MS et IM-MS utilisées indépendamment¹¹⁷. La CCS ajoute une nouvelle dimension au couple m/z et RT ce qui contribue à améliorer l'identification des métabolites¹¹⁵. Cela fait de la mobilité ionique une technique analytique de plus en plus utilisée pour les analyses métabolomiques^{115,118}.

D – Le couplage avec les méthodes séparatives

Les analyses métabolomiques par spectrométrie de masse sans couplage avec des techniques séparatives sont réalisées par introduction directe en injectant l'échantillon soit à l'aide d'un pousse seringue (DIMS), soit *via* le système chromatographique (FIA)¹¹⁹ mais sans colonne chromatographique. Ces systèmes permettent l'analyse à haut débit des échantillons. Il est nécessaire de coupler ces systèmes d'injection à des spectromètres de masse à haute voire très haute résolution (type FT-ICR) afin de faciliter l'identification des composés et de distinguer les composés isobares^{91,120,121}. Cependant, ces techniques présentent certaines limitations notamment liées aux phénomènes de suppression d'ionisation. Un autre inconvénient majeur lié à l'introduction directe est l'impossibilité de distinguer les isomères. L'ajout d'une dimension séparative supplémentaire se révèle primordiale pour couvrir au mieux le métabolome d'un échantillon donné. Pour cela, différentes techniques sont utilisées telles que la chromatographie en phase gazeuse ou liquide, l'électrophorèse capillaire et la chromatographie en phase supercritique. La chromatographie en phase gazeuse, en phase liquide ainsi que l'électrophorèse capillaire sont spécialement détaillées dans ce manuscrit.

1 – La chromatographie en phase gazeuse

La chromatographie en phase gazeuse couplée à la spectrométrie de masse est l'une des techniques les plus utilisées pour les analyses métabolomiques¹²². L'échantillon est injecté puis chauffé afin d'être vaporisé puis dirigé vers la colonne. Les composés sous forme gazeuse vont interagir avec la phase stationnaire et seront plus ou moins retenus selon leur affinité avec cette dernière. Ils seront ensuite ionisés, dans la majorité des cas, *via* une source à impact électronique¹²².

Pour que les composés soient analysables dans ces conditions, il faut qu'ils soient volatils et thermiquement stables. Cependant, une grande partie des métabolites n'est pas volatile^{122,123}. Il est donc nécessaire d'ajouter une étape de dérivatisation chimique lors de la préparation des échantillons afin de les rendre volatiles et également plus stables thermiquement^{124,125}. Cette étape peut être longue et peut entraîner la dégradation d'une partie de l'échantillon. Pour réduire le temps d'analyse, des colonnes avec des diamètres plus petits (100 μm) ont été développées et ont montré de bonnes efficacités de séparation¹²⁶. Les pics chromatographiques étant très fins, le couplage est réalisé avec un analyseur ayant une vitesse d'acquisition importante, historiquement, avec les QqQ. Aujourd'hui le couplage GC-MS est envisageable avec des spectromètres de masse Q-TOF ou Q-Orbitrap à des résolutions médianes (entre 30 000 et 70 000 à m/z 200).

D'autre part, l'ajout d'une dimension supplémentaire de séparation avec la GCxGC-MS, qui correspond au couplage de deux colonnes GC, permet d'augmenter l'efficacité de séparation des composés. Souvent, la première colonne est apolaire et longue et la deuxième colonne est polaire et plus courte (selon les études, plus de 10 m pour la première et environ 2 m pour la deuxième)^{127,128}. Ces développements permettent d'augmenter le pouvoir résolutif de la séparation, la sensibilité ainsi que l'efficacité des analyses^{123,129}.

La robustesse des analyses par GC-MS a permis de constituer des bases de données spectrales robustes et interopérables, notamment grâce à la grande reproductibilité des spectres obtenus par impact électronique, permettant l'identification des métabolites¹³⁰. La GC-MS permet d'obtenir une haute capacité de séparation des pics ainsi qu'une grande sensibilité¹²³.

2 – La chromatographie en phase liquide

La chromatographie en phase liquide couplée à la spectrométrie de masse à haute résolution est aujourd'hui la technique analytique la plus utilisée pour les analyses métabolomiques. Par chromatographie en phase liquide, les molécules sont plus ou moins retenues selon leur affinité avec la phase stationnaire et les solvants de la phase mobile¹³¹. Il existe différents types de chromatographie liquide, avec notamment : la chromatographie liquide en phase inverse (RPLC), en phase normale (NPLC) et la chromatographie liquide d'interactions hydrophiles (HILIC). L'efficacité de séparation de la colonne dépend, de la longueur et du diamètre de la colonne, de la nature et de la taille des particules de la phase stationnaire ainsi que des phases mobiles.

Le principe de rétention de la chromatographie liquide en phase normale repose sur des interactions hydrophiles entre l'analyte et la phase stationnaire. Cette phase stationnaire est souvent composée de silice avec des groupement silanols (Si-OH). La chromatographie en phase normale permet donc de retenir les composés polaires. A cause des solvants apolaires utilisés pour les phases mobiles incompatibles avec le couplage avec la spectrométrie de masse, ce type de chromatographie est peu utilisée en métabolomique^{122,132}.

La chromatographie liquide en phase inverse est la plus connue et, historiquement, la plus utilisée car accessible et très robuste¹³³. Les colonnes de RPLC sont composées d'une phase stationnaire apolaire. Il s'agit de la même phase stationnaire en silice que celle utilisée pour la chromatographie liquide en phase normale sur laquelle des chaînes alkyles ont été greffées. Les composés sont alors retenus *via* des interactions hydrophobes. Selon la nature et la taille des chaînes alkyles greffées, la rétention des composés varie. Les mécanismes de RPLC sont bien connus et permettent de prédire l'ordre d'élution des composés^{133,134}. Il s'agit d'un facteur important pour l'identification de métabolites inconnus. L'utilisation de solvants aqueux ou organiques est un des points forts de cette technique chromatographique, en raison de leur compatibilité directe avec une détection par spectrométrie de masse¹³⁵. La RPLC avec une phase stationnaire greffée avec des chaînes alkyles à 18 carbones (RP-C18) est aujourd'hui très utilisée pour les analyses métabolomiques¹³⁵. Cette chromatographie permet de séparer les composés de faibles polarités à non-polaires. Cependant, elle ne permet pas de retenir suffisamment les composés très polaires comme certains acides aminés très présents dans les fluides biologiques¹²². Pour

séparer ces composés polaires, des colonnes pentafluorophenylpropyl (PFPP) peuvent être utilisées¹³⁶. Les interactions π - π entre la phase stationnaire pauvre en électrons et les analytes riches en électrons permettent de séparer les composés polaires tels que les acides aminés aromatiques¹³⁶.

La chromatographie liquide à interactions hydrophiles (HILIC) est également très utilisée en métabolomique pour les composés polaires (comme les acides aminés, sucres, acides organiques...) et souvent exploitée pour sa complémentarité avec la RPLC. Les colonnes HILIC sont composées d'une phase stationnaire de silice similaire à celle de la chromatographie liquide en phase normale avec par exemple un greffage diol, amino ou sulfobétaine afin d'avoir une couche d'eau adsorbée à la surface de cette phase. Contrairement à la chromatographie en phase normale, les phases mobiles sont directement compatibles avec le couplage avec la spectrométrie de masse. Le gradient débute avec un fort pourcentage de solvant organique (souvent de l'acétonitrile) qui diminue au cours du temps en faveur d'un solvant aqueux pouvant contenir des additifs^{137,138}. De nombreux mécanismes sont mis en jeu lors du processus de rétention des composés ce qui rend complexe la prédiction de l'ordre d'élution des métabolites contrairement à la RPLC. Cependant, il a été montré que les métabolites sont retenus selon leur polarité et leur degré de solvation¹³⁸. Un des inconvénients souvent évoqués pour cette technique est le temps assez long de rééquilibrage de la colonne entre deux échantillons afin d'obtenir les analyses les plus robustes possibles.

De nombreuses études ont comparé les capacités de détection et de rétention de ces différentes méthodes chromatographiques pour les analyses métabolomiques^{132,139}. Ces études ont montré que l'HILIC permettait de retenir le plus de composés lors du profilage métabolique de sérum humain (Figure 16¹³⁹). Cependant, il s'avère souvent nécessaire de combiner différentes techniques chromatographiques afin d'obtenir la couverture la plus large possible du métabolome. La nature chimique des métabolites influence l'ionisation de ces derniers. En effet, certains composés vont présenter une préférence pour l'ionisation positive ou négative, et certains composés s'ionisent très bien dans les deux modes tandis que d'autres non. Il a donc été montré que la combinaison PFPP-ESI(+)-MS et HILIC-ESI(-)-MS permettait de couvrir une grande partie du métabolome. Cependant, les colonnes C18 permettent une plus grande reproductibilité des analyses que les colonnes PFPP. De nombreuses études métabolomiques, notamment celles effectuées au sein de notre laboratoire, sont généralement effectuées en couplant les techniques C18-ESI(+)-MS et HILIC-ESI(-)-MS^{139,140}.

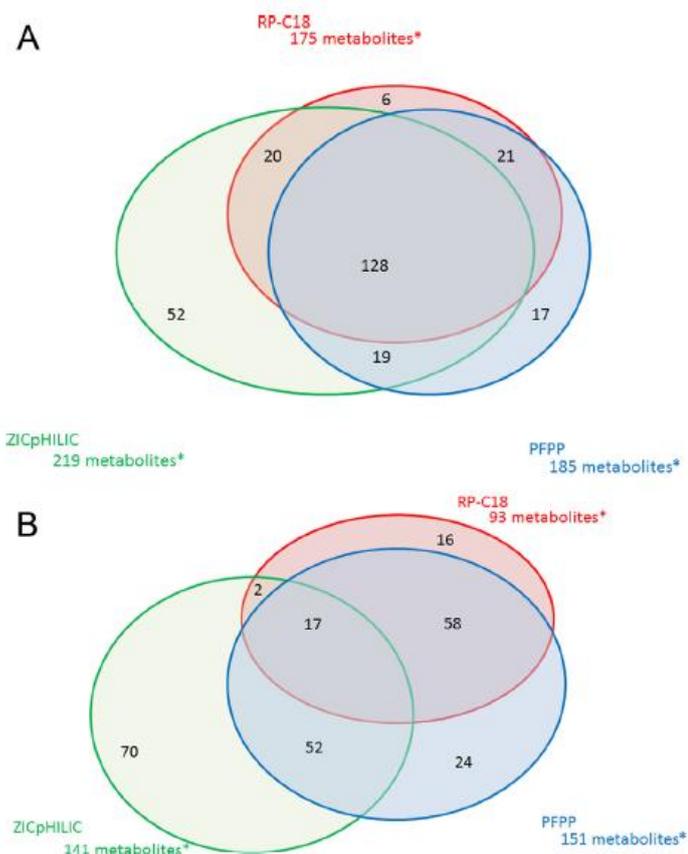


Figure 16 : Analyse de sérum humain par trois techniques chromatographiques (extraite de Boudah et al.).

(A : Nombre de métabolites détectés ; B : Nombre de métabolites retenus)

Pour obtenir une analyse chromatographique avec la meilleure séparation des composés, les gradients peuvent être relativement longs¹²². Utiliser différentes méthodes chromatographiques prend donc du temps pour analyser un seul échantillon, puisqu'il doit être injecté plusieurs fois. Pour diminuer ce temps d'analyse et gagner en efficacité de séparation, des colonnes avec des tailles de particules plus petites ont été développées. La chromatographie ultra-haute performance (UHPLC) est souvent utilisée¹⁴¹. Les particules des colonnes UHPLC ont un diamètre inférieur à 2 μm ce qui augmente l'efficacité de séparation des composés¹²². Ces colonnes permettent également de travailler à plus haute pression. L'UHPLC permet souvent une détection plus sensible et améliore l'efficacité de séparation des composés grâce à des pics très fins, ce qui nécessite le couplage avec un analyseur de masse ayant une grande vitesse d'acquisition¹⁴².

3 – L'électrophorèse capillaire

L'électrophorèse capillaire (CE) a été introduit par Arne Tiselius qui a obtenu le prix Nobel de Chimie en 1948 pour ces travaux sur le sujet¹⁴³. Le principe de séparation des composés est basé sur le rapport charge sur taille des ions. Un champ électrique est appliqué à chaque extrémité du capillaire contenant une solution d'électrolyte. Les composés fortement chargés avec un faible rayon dynamique seront les composés qui migreront le plus rapidement d'une extrémité à l'autre du capillaire¹⁴⁴.

Le couplage de l'électrophorèse capillaire avec la spectrométrie de masse a donné lieu à de nombreux développements et représente un des enjeux majeurs pour l'application de la CE-MS pour les analyses métabolomiques^{145,146}. Malgré de bonnes performances pour le profilage métabolique^{147,148}, la CE-MS est encore très peu utilisée en raison d'un certain manque de robustesse mais surtout à cause d'un couplage difficile, notamment au niveau du montage. En effet, il faut garder le contact électrique entre les deux électrodes tout en fixant la sortie du capillaire dans la source d'ionisation. De plus, les analyses de spectrométrie de masse se réalisant sous vide, la solution peut être « aspirée » par le spectromètre de masse à cause de la différence de pression ce qui peut perturber la séparation des composés. Un autre inconvénient est les faibles débits utilisés en électrophorèse capillaire qui sont peu compatibles avec la formation d'un spray stable en ESI^{148,149}.

La CE-MS permet de séparer les composés polaires et chargés. Les principaux avantages de cette technique analytique sont sa rapidité d'analyse, sa haute résolution, et donc sa haute capacité de séparation des composés, ainsi que le faible coût de la CE¹⁴⁴. De plus, seulement une faible quantité d'échantillon est nécessaire (de l'ordre du nanolitre)¹⁵⁰. Cependant, cet avantage est également un inconvénient car cela diminue fortement la sensibilité de la détection par spectrométrie de masse. Une étape de pré-concentration de l'échantillon peut ainsi être nécessaire¹⁵¹. De plus, le faible débit de l'électrophorèse capillaire rend l'ionisation instable. Un autre inconvénient de cette technique est son manque de reproductibilité et de robustesse¹⁴⁶. Malgré ces soucis instrumentaux, de nombreuses applications pour la métabolomique, avec l'analyse de matrices biologiques, ont été référencées par Ramautar et al. ces dernières années^{145,146,150,152}.

4 – Comparaison de méthodes séparatives

Les différentes méthodes séparatives décrites ci-dessus soulignent que de nombreuses techniques analytiques peuvent être utilisées pour les analyses métabolomiques. Cependant, chaque méthode ayant ses propres avantages et inconvénients, il peut apparaître nécessaire de coupler ces différentes méthodes afin de couvrir au mieux l'ensemble du métabolome.

Quelques études se sont intéressées à la comparaison et la complémentarité de l'électrophorèse capillaire, la chromatographie en phase gazeuse et la chromatographie en phase liquide (Figure 17¹⁵³), et soulignent généralement le caractère complémentaire des différentes plate-formes analytiques pour couvrir au mieux le métabolome d'un échantillon d'intérêt¹⁵³.

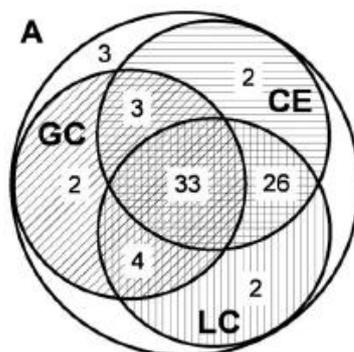


Figure 17 : Couverture de 75 intermédiaires du métabolisme central par trois méthodes séparatives (extraite de Büscher et al.).

GC : méthylation et silylation par *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide et *N*-tert-butyltrimethylsilyl-*N*-methyltrifluoroacetamide ; LC : HILIC et C18 ; CE : séparation cationique et anionique

Après avoir parcouru l'ensemble du processus mis en place lors d'une analyse métabolomique, notamment par LC-HRMS, nous allons détailler les deux points clés suivants : l'acquisition de données en spectrométrie de masse (notamment MS/MS) pour l'annotation et l'identification de métabolites, puis la problématique de la quantification en métabolomique.

Chapitre 2 : L'acquisition de données en spectrométrie de masse pour la métabolomique

Comme nous l'avons vu précédemment, 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 tous les m/z présents 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. Les signaux ainsi acquis permettant de discriminer différents groupes d'échantillons sont repérés grâce aux analyses statistiques. L'acquisition de données de fragmentation en spectrométrie de masse en tandem (MS/MS) va servir à identifier les métabolites correspondant aux différents signaux pertinents lors de l'étape d'annotation. Ces données de fragmentation peuvent être de différents types (par exemple CID ou HCD) et être acquises de différentes manières (par exemple avec plusieurs énergies de collision).

Ce chapitre présente tout d'abord la stratégie globale d'annotation et d'identification de métabolites à partir de données MS et MS/MS, puis traitera des modalités d'acquisition MS/MS.

A – L'annotation et l'identification des métabolites

Une fois la liste de variables nettoyée pour ne conserver que les variables analytiquement pertinentes, les variables doivent être annotées pour pouvoir ensuite permettre une interprétation biologique des données. Cette étape consiste à attribuer un nom de métabolite aux variables détectées. Elle est cruciale au cours du processus d'analyse métabolomique et reste aujourd'hui une des principales étapes limitantes lors d'une analyse métabolomique par LC-HRMS¹⁵⁴. Ces limites sont notamment dues à notre méconnaissance de l'intégralité du métabolome¹⁵⁵. En effet, on estime notamment que des centaines, voire des milliers de métabolites primaires sont présents de façon ubiquitaire chez l'ensemble des organismes vivants¹⁵⁶. A ceux-ci s'ajoute une multitude de métabolites secondaires et de xénobiotiques (aliments, médicaments, produits d'hygiène, polluants...), qui peuvent également être biotransformés par l'organisme hôte. Ainsi, le nombre et la diversité structurale des composés présents dans une matrice biologique donnée sont très importants¹⁵⁷. Un seul échantillon contiendrait donc plusieurs milliers de métabolites avec des structures et des concentrations potentiellement très variables¹⁵⁸. On estime que moins de 10 % de ces métabolites sont aujourd'hui annotés¹⁵⁹. Ainsi, les processus d'annotation et d'identification à partir de données MS et MS/MS représentent encore aujourd'hui des étapes extrêmement limitantes des analyses métabolomiques. La Figure 18 présente de façon schématique les modalités d'annotation d'un jeu de données par MS et MS/MS..

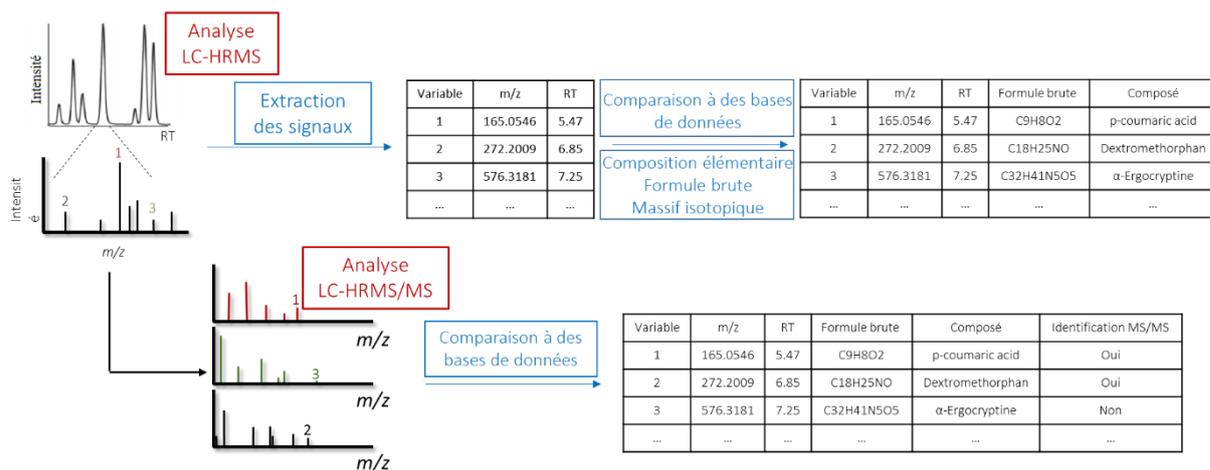


Figure 18 : Les étapes d'annotation et d'identification en métabolomique.

- Annotation à partir des données MS

Une analyse LC-HRMS génère souvent des milliers de signaux, qui peuvent être hiérarchisés selon leur origine, biologique ou non. Pour les composés n'étant pas « naturellement » présents dans l'échantillon ou non pertinents biologiquement parlant, il peut s'agir par exemple de contaminants, liés aux solvants ou aux matériels utilisés, ou d'artefacts (provenant par exemple d'erreurs informatiques ou du « bruit » électronique)¹⁶⁰ (Figure 19¹⁶⁰).

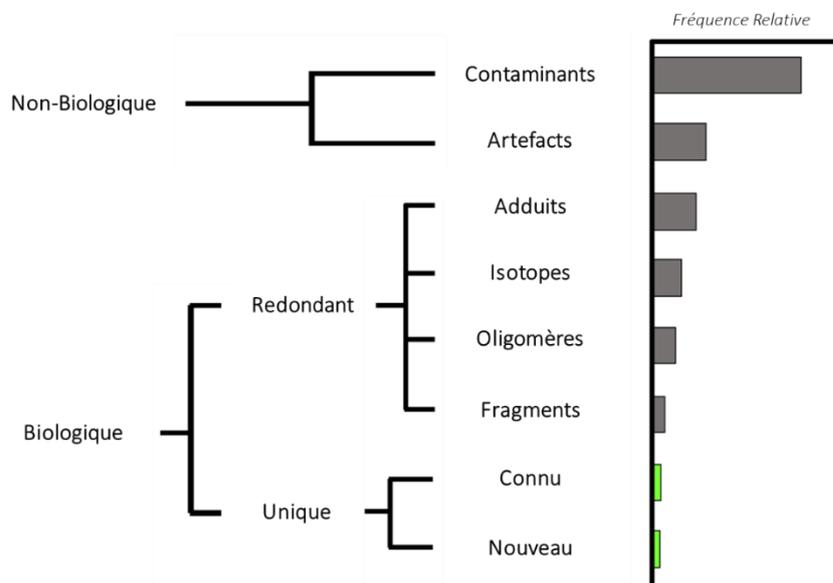


Figure 19 : Répartition des composés selon leur origine (adaptée de Sindelar et al.).

Ainsi, de nombreux signaux non identifiables sont présents au sein des empreintes métaboliques enregistrées, certains de ces signaux n'étant pas systématiquement éliminés par les processus de traitement des signaux présentés dans le chapitre précédent. L'ionisation ESI produit souvent plus d'un ion pour un métabolite

donné, ces ions peuvent correspondre à des fragments produits en source, des adduits et leurs isotopes, ce qui peut conduire à une surestimation du nombre de métabolites. A titre d'exemple, la Figure 20 présente le spectre ESI-MS obtenu pour le propionyl-coenzyme A avec la présence de nombreux ions. Domingo-Almenara et al. ont récemment rapporté que seuls 10 à 15 % des composés inclus dans la base de données METLIN (environ 15 000 spectres obtenus dans des conditions spécifiques LC-MS et MS/MS) ne se dissocient pas dans la source, alors que les espèces protonées et déprotonées ne représentent les espèces les plus intenses que dans environ 50 % des cas¹⁶¹. La présence de plusieurs ions pour un seul métabolite peut cependant être utile dans la mesure où elle peut confirmer la pertinence des compositions élémentaires car elle permet plusieurs mesures de masses précises liées à la même molécule. Toutefois, si elle n'est pas (correctement) prise en compte, cette redondance du signal complique le traitement des données. En effet, un ion fragment peut être par exemple considéré à tort comme un ion monoisotopique provenant d'un autre métabolite (Figure 20¹⁶²)^{160,162-164}. Tous ces différents éléments contribuent dans une certaine mesure à la difficulté d'annoter les variables détectées.

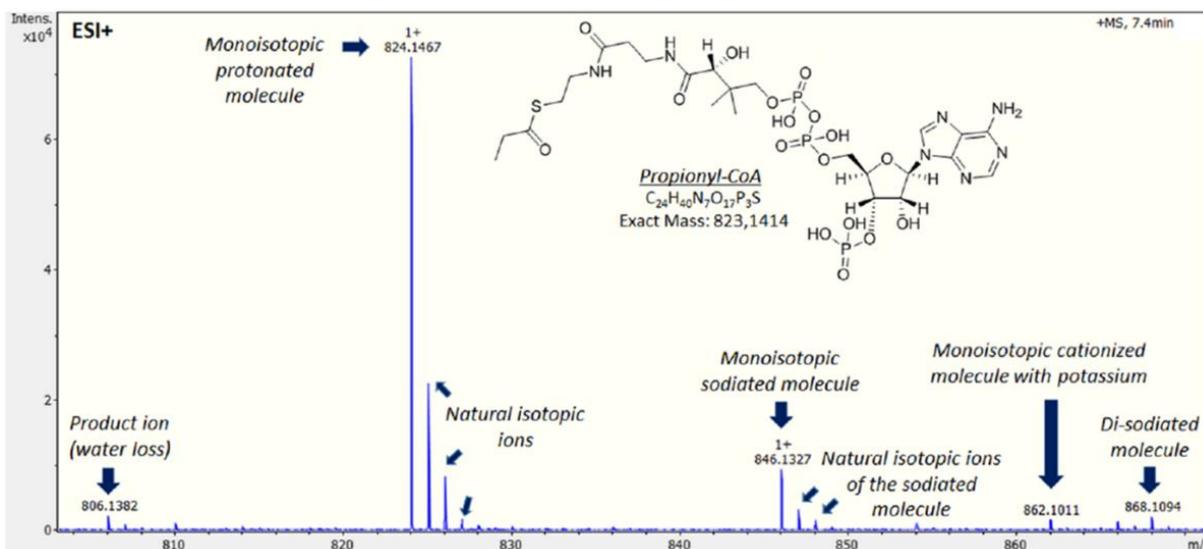


Figure 20 : Les différents signaux MS produits par le propionyl-CoA (adaptée de Damont et al.).

Différentes approches peuvent être 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. L'efficacité de cette méthode d'annotation dépend de plusieurs 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. Pour cela, une base de données d'environ 1000 métabolites a été composée au sein de notre laboratoire. Dans ces conditions, nous observons au laboratoire que moins de 10 % des variables détectées et présentes au sein des jeux de données peuvent être généralement annotées.

Dans une seconde approche, pour les variables ne correspondant pas à un composé présent dans la base de données interne, l'annotation pourra être réalisée par comparaison des *m/z* expérimentaux à ceux de bases de données publiques comprenant plusieurs millions de molécules chimiques. Cependant, peu d'entre

elles sont des métabolites endogènes ou des composés biologiquement pertinents¹⁶⁵. Les annotations retournées par interrogation de ces bases de données ne sont donc pas toujours pertinentes. Il sera donc préférable de privilégier l'utilisation de bases de données spécialisées (voir ci-dessous). Le Tableau 2 présente une liste non-exhaustive des bases de données disponibles¹⁶⁶.

Tableau 2 : Liste de bases de données MS (adapté de Blaženović et al.).

Base de données	Molécules Cibles
PubChem	Toutes les petites molécules
ChemSpider	Toutes les petites molécules
KEGG	Métabolites
MetaCyc	Métabolites
BRENDA	Enzymes
HMDB	Métabolites
CHEBI	Petites Molécules
UNPD	Métabolites
MINE	Métabolites
METLIN	Métabolites

KEGG : Kyoto Encyclopedia of Genes and Genomes

BRANDA : BRAunschweig ENzyme Database

HMDB : Human Metabolome DataBase

CHEBI : CHEmical Entities of Biological Interest

UNPD : Universal Natural Products Database

MINE : Metabolic In silico Network Expansion

METLIN : METabolite LINK

Dans le cas de métabolites non annotés, la haute précision de mesure de masse obtenue sur les spectromètres de masse à haute résolution pourra généralement permettre, à partir de m/z obtenus, d'avoir accès la composition élémentaire (formule brute) du composé. Les recherches classiques des formules brutes se basent sur la présence des éléments C, H, O, N, P et S et éventuellement Cl, Na et K, qui sont à l'origine d'adduits souvent détectés. Cependant, cela peut s'avérer insuffisant pour l'identification d'inconnus qui peuvent contenir des éléments moins courants¹⁵⁹. De plus, les spectromètres de masse à haute et très haute résolution donnent des masses précises avec une précision inférieure au ppm (particulièrement de FT-IR), ce qui limite le nombre de formules brutes pour un signal. Malgré cela, pour une masse donnée plusieurs compositions élémentaires sont tout de même possibles^{167,168}. L'étude du massif isotopique d'un composé permet alors de limiter le nombre de formules brutes possibles, en considérant notamment la présence d'hétéroatomes présentant des isotopes abondants (Tableau 3). Dans ces conditions, il est possible de limiter le nombre d'annotations pour un seul m/z ¹⁶⁹.

Tableau 3 : Les principaux isotopes stables et leur abondance naturelle.

Elément	Isotope	Abondance	Masse Isotopique	Isotope	Abondance	Masse isotopique	Ecart de masse entre isotope
H	¹ H	99.98%	1.0078	² H	0.015%	2.0141	1.0063
C	¹² C	98.93	12.0000	¹³ C	1.10%	13.0034	1.0034
N	¹⁴ N	99.64%	14.0031	¹⁴ N	0.36%	15.0001	0.9970
O	¹⁶ O	99.76%	15.9949	¹⁸ O	0.20%	17.9992	2.0042
S	³² S	94.93%	31.9721	³⁴ S	4.29%	33.9679	1.9958
Cl	³⁵ Cl	75.76%	34.9689	³⁷ Cl	24.24%	36.9659	1.9970
Br	⁷⁹ Br	50.69%	78.9183	⁸¹ Br	49.31%	80.1629	1.2446

- *Confirmation des annotations par acquisition de données MS/MS*

Dans un second temps, l'acquisition de spectres MS/MS permet de confirmer ou infirmer les annotations obtenues, voire élucider les structures de métabolites inconnus. Pour cela, les spectres MS/MS généralement obtenus en mode de fragmentation non résonant sont comparés à ceux présents dans des bases de données spécialisées¹⁷⁰. Dans le cas de comparaisons réalisées avec des bases de données externes ou publiques, un pré-requis est que des spectres acquis dans différents lieux et conditions, et avec des instruments différents soient comparables (par exemple Q-TOF vs Q-Orbitrap)¹⁷¹. Ces aspects seront plus particulièrement discutés dans la partie « Résultats » de ce manuscrit. Le Tableau 4 présente une liste non-exhaustive de bases de données de spectres MS/MS expérimentaux et/ou obtenus *in silico*^{166,170,172}.

Tableau 4 : Liste des bases de données spectrales (adapté de Blaženović et al.).

Base de données	Cibles	Nombre de Spectres MS/MS	Nombre de Composés
NIST	EI-MS, CID-MS/MS	193 120	9 344
WILEY	EI-MS, CID-MS/MS	20 000	1 200
METLIN	CID-MS/MS	72 268	14 034
MoNA	EI, MS/MS, MSn	194 000	68 700
MassBank	EI, MS/MS, MSn	22 000	2 800
mzCloud	MSn	182 000	2 800
GNPS	MS/MS	212 230	12 694
ReSpect	MS/MS, RT	9 000	4 000

NIST : National Institute of Standards and Technology
 MoNA : MassBank of North America
 GNPS : Global Natural Product Social Molecular Networking

La comparaison de spectres MS/MS expérimentaux avec ceux présents dans des bases de données peut être complexe. Or, pour que le spectre MS/MS soit le plus informatif possible, les ions précurseurs et fragments doivent être suffisamment intenses. Or, il a été rapporté que la grande majorité des ions pertinents devant être soumis à l'analyse par MS/MS sont souvent très peu abondants, ce qui complique encore le processus d'identification¹⁵⁹ (Figure 21¹⁵⁹).

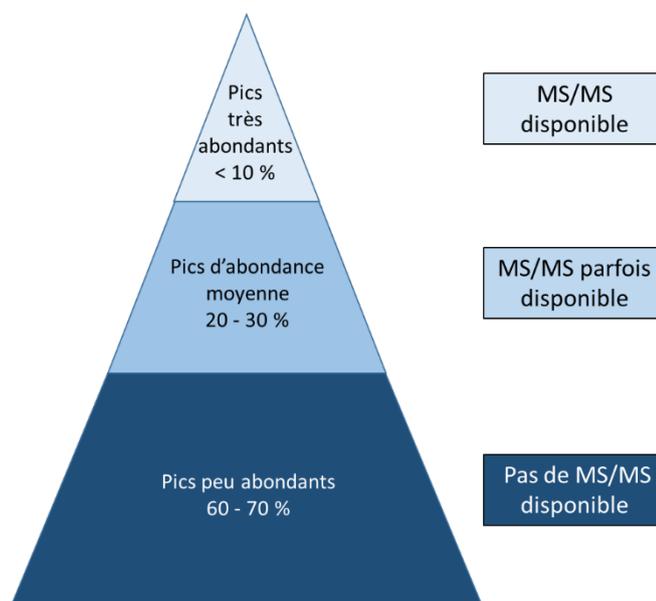


Figure 21 : Répartition des composés selon leur abondance (adaptée de Chaleckis et al.).

Comme discuté plus haut, il est préférable que l'ion utilisé pour la génération de la formule brute soit l'ion moléculaire. Or, cet ion peut être (très) minoritaire par rapport à un adduit ou un fragment en source¹⁶¹. Afin d'obtenir un spectre MS/MS de meilleure qualité et des informations complémentaires, il serait alors plus judicieux de fragmenter un ion adduit (ou un fragment en source). Malheureusement les bases de données ne contiennent souvent pas de spectres MS/MS obtenus expérimentalement sur des ions adduits.

Un autre point crucial est relatif à l'énergie de collision utilisée pour fragmenter les composés d'intérêt. En effet, pour comparer au mieux des spectres de fragmentation et identifier les métabolites d'intérêt, il faut que l'énergie de fragmentation utilisée pour la molécule de référence et l'endogène soient très proches. En effet, plus les énergies seront différentes moins les spectres seront comparables. De plus, les spectres MS/MS des composés de faible masse (< 150 Da) contiennent souvent peu de fragments qui ne leur sont pas forcément spécifiques. Cela peut complexifier fortement l'identification¹⁵⁹. A l'heure actuelle, la comparaison des spectres MS/MS 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 MS/MS enregistrés et pour les comparer plus efficacement et objectivement¹⁷⁰. Ces outils sont appelés à s'étendre afin d'automatiser cette étape de validation d'identification par fragmentation.

- *Niveaux d'annotation*

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^{105,173}.

Partie A : Contexte scientifique et technologique – Chapitre 2

- 1^{er} niveau : Les composés formellement identifiés (avec deux critères d'identification par rapport à un standard pur, par exemple m/z et RT, MS et MS/MS). Ce sont les composés pour lesquels un standard est disponible. Le standard et l'échantillon doivent être analysés dans les mêmes conditions analytiques et leur spectres MS/MS sont directement comparés.
- 2^{ème} niveau : Les composés putativement annotés. Ce sont les composés pour lesquels le standard n'est pas analysé dans les mêmes conditions que l'échantillon. Le spectre MS/MS de l'échantillon est alors comparé avec un spectre MS/MS présent dans une base de données mais pas nécessairement acquis avec les mêmes paramètres analytiques.
- 3^{ème} niveau : Les composés inconnus mais avec supposition de leur famille chimique. Ce sont les composés pour lesquels aucune correspondance ne peut être directement faites avec les bases de données. Des informations sur la classe du composé peuvent être supposées en fonction de similitudes au niveau des propriétés physico-chimiques avec des composés présents dans les bases de données.
- 4^{ème} niveau : Les composés inconnus.

Ces différents niveaux d'identification semblent aujourd'hui insuffisants pour décrire l'ensemble des métabolites car la stéréochimie des composés n'est pas prise en compte. Des réflexions sont actuellement en cours au sein de la « Metabolomics Society » pour faire évoluer ces niveaux. Sept niveaux d'identification pourraient être proposés¹⁷⁴. Le niveau d'identification le plus élevé correspondrait à une connaissance complète de la structure du métabolite avec la spécification de la stéréochimie et de la chiralité. Dans les niveaux intermédiaires, les informations comme la classe structurale, la formule chimique, les groupes fonctionnels et la stéréochimie seraient ajoutées séquentiellement.

Ces différents paramètres limitent fortement le nombre de composés identifiés lors d'une analyse métabolomique et de nombreux signaux ne sont toujours pas caractérisés⁴⁹. L'acquisition de spectres MS et MS/MS de qualité s'avère donc primordiale pour s'assurer de l'identification des métabolites. Cependant, il est encore compliqué d'acquérir ces spectres en une seule analyse sur l'ensemble des signaux pertinents, en raison des caractéristiques intrinsèques du spectromètre de masse telles que sa capacité à réaliser des acquisitions MS/MS ou encore la vitesse d'acquisition de l'analyseur. Les échantillons sont donc, généralement, tous analysés en MS et seuls les QCs sont utilisés pour les analyses MS/MS. Les spectres MS/MS obtenus sont donc représentatifs d'échantillons moyens et non d'un échantillon individuel¹⁵⁹. Le développement de méthodes permettant l'acquisition de spectres MS et MS/MS en une seule injection semblent ainsi primordial.

B – Les différents modes d'acquisition MS/MS

Comme décrit précédemment, l'identification des métabolites est souvent un point limitant dans les analyses métabolomiques non-ciblées. L'acquisition de spectres MS/MS informatifs est nécessaire pour confirmer l'identité des métabolites d'intérêt en les comparant avec des bases de données spectrales. Plusieurs modes d'acquisition de spectres MS/MS sont disponibles et sont résumés sur la Figure 22 : le mode ciblé, « données dépendantes » (DDA) et « données indépendantes » (DIA) (Figure 22).

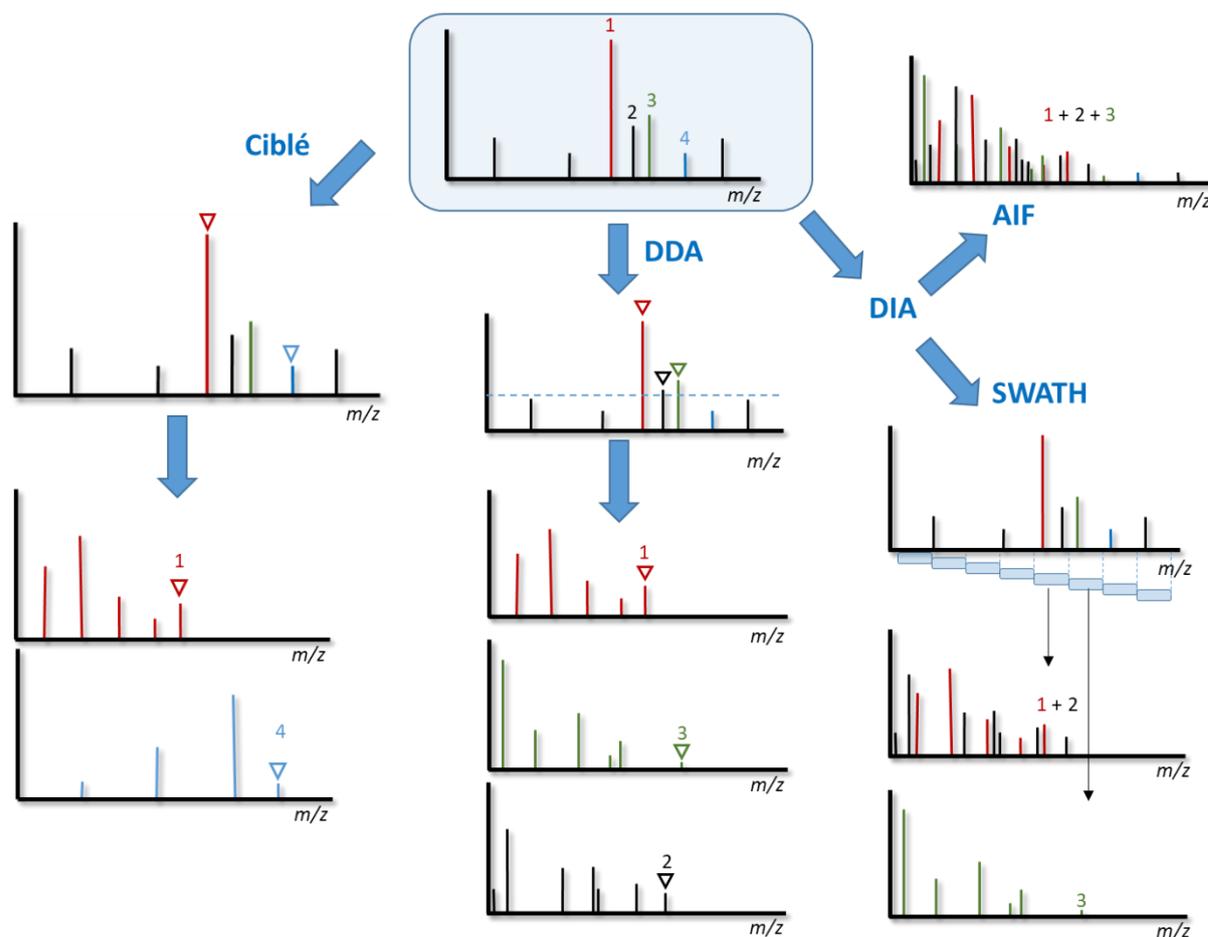


Figure 22 : Les différents modes d'acquisition des spectres MS/MS.

(AIF : All Ion Fragmentation ; SWATH : « Sequential Window Acquisition of All Theoretical fragment-ion spectra »)

Une partie des travaux bibliographiques présentés ci-après a permis la publication d'une revue au début de cette thèse (2017) : « Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? » (Annexe 1)

1 – Le mode ciblé

Le mode ciblé est actuellement le plus utilisé lors des analyses métabolomiques. Ce mode d'acquisition nécessite une connaissance préalable de l'échantillon et des métabolites d'intérêt qu'il contient¹⁷⁵. En effet, les ions sont sélectionnés d'après une liste d'inclusion contenant le m/z et le RT de l'ion d'intérêt. Cette liste est obtenue après le traitement des données issues du profilage MS de l'échantillon. Seuls les ions d'intérêt par rapport à la problématique initiale sont généralement fragmentés pour être identifiés ou pour confirmer leur identité¹⁷⁶. 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 particulièrement longue. Ainsi, l'acquisition des spectres MS/MS se fait souvent dans un second temps, parfois distant de plusieurs semaines de l'acquisition des profils MS. Dans ces conditions, les analyses MS et MS/MS ne se déroulent pas dans les mêmes conditions analytiques. En effet, les échantillons ont généralement subi un cycle de congélation/décongélation supplémentaire ce qui peut modifier la concentration de certains métabolites, voire entraîner la disparition de certains d'entre eux particulièrement instables. Les conditions analytiques peuvent également être impactées, le système chromatographique ainsi que la colonne ont pu être utilisées dans l'intervalle de temps pour d'autres analyses avec d'autres échantillons, ce qui peut altérer leurs performances, avec des temps de rétention parfois drastiquement modifiés entre les deux analyses. Les performances du spectromètre de masse peuvent également être différentes avec par exemple une source plus ou moins encrassée ou une efficacité d'ionisation altérée. Ces différents points peuvent se matérialiser par une mauvaise détection de l'ion d'intérêt, une augmentation du bruit de fond ou encore une dérive des temps de rétention. De plus, la nécessité d'une nouvelle acquisition requiert d'avoir une quantité initiale d'échantillon non négligeable ce qui n'est malheureusement pas toujours possible. Ces différents facteurs sont des inconvénients majeurs dans le processus d'identification des métabolites. Pour toutes ces raisons et grâce aux améliorations récemment apportées aux spectromètres de masse, notamment au niveau du couple vitesse d'acquisition - résolution, des nouveaux modes d'acquisition commencent à être appliqués en métabolomique. Ces modes d'acquisition non ciblés permettent l'acquisition simultanée de données MS et MS/MS informatives et assez exhaustives.

2 – Les analyses en mode « Data Dependent Acquisition »

L'acquisition des données en mode données dépendantes (DDA) permet d'alterner automatiquement l'acquisition de spectres MS et MS/MS (Figure 22)¹⁷⁷. Dans ce cas, il n'est pas nécessaire d'avoir une connaissance préalable de l'échantillon et des métabolites d'intérêt car les ions sont sélectionnés selon certains critères. Le critère le plus utilisé pour ce mode d'acquisition est le seuil d'intensité mais il peut également s'agir du massif isotopique ou du défaut de masse^{178,179}. Le mode DDA peut ainsi permettre la détection, l'identification et la quantification (effectuée à partir de la trace MS) de métabolites en une seule analyse¹⁷⁷. Cela entraîne alors des avantages importants comme la réduction du temps d'analyse et une quantité d'échantillon nécessaire moindre par rapport au mode ciblé. Un des avantages majeurs des analyses DDA est la pureté des spectres obtenus¹⁸⁰. En

effet, un m/z spécifique est sélectionné avec une fenêtre d'isolement inférieure au Dalton (Da) afin d'être fragmenté. Cette fenêtre de masse restreinte limite la sélection d'interférents qui complexifieraient l'analyse du spectre MS/MS résultant. Cela permet généralement de s'assurer de la nature du métabolite sélectionné.

Ces caractéristiques font que la DDA est couramment utilisée en protéomique¹⁸¹. Cependant, elle est encore en cours d'implémentation pour des applications métabolomiques de routine. La principale différence entre la métabolomique et la protéomique est ici liée à l'énergie de fragmentation. En effet, en protéomique l'énergie optimale de fragmentation varie peu d'un peptide à l'autre (généralement une énergie de collision normalisée, NCE, entre 25 % et 30 %) contrairement à la métabolomique où elle peut varier énormément d'un métabolite à l'autre (par exemple de 10 % à 100 % en énergies HCD normalisées). Cela implique qu'il est difficile de définir, sans a priori, une énergie de collision « universelle » permettant l'acquisition de spectres MS/MS informatifs et plus facilement comparables pour l'ensemble des métabolites d'un échantillon donné⁶². Ceci s'applique bien sûr également aux modes d'acquisition ciblé et DIA. Afin de limiter ce problème, des acquisitions MS/MS sont souvent effectuées en appliquant différentes énergies de collision. Un spectre MS/MS moyen peut également être obtenu par sommation de différents spectres acquis à différentes énergies¹⁸². Ce mode de fragmentation est appelé « stepped collision energy » par le constructeur Thermo Fisher Scientific (Figure 23¹⁸³). La Figure 23 représente de façon schématique trois spectres de fragmentation indépendants à trois énergies de fragmentation : HCD 10 % ; HCD 30 % et HCD 50 %. Le spectre obtenu avec le mode de fragmentation « stepped collision energy » correspond alors à la somme des trois spectres de fragmentation précédents.

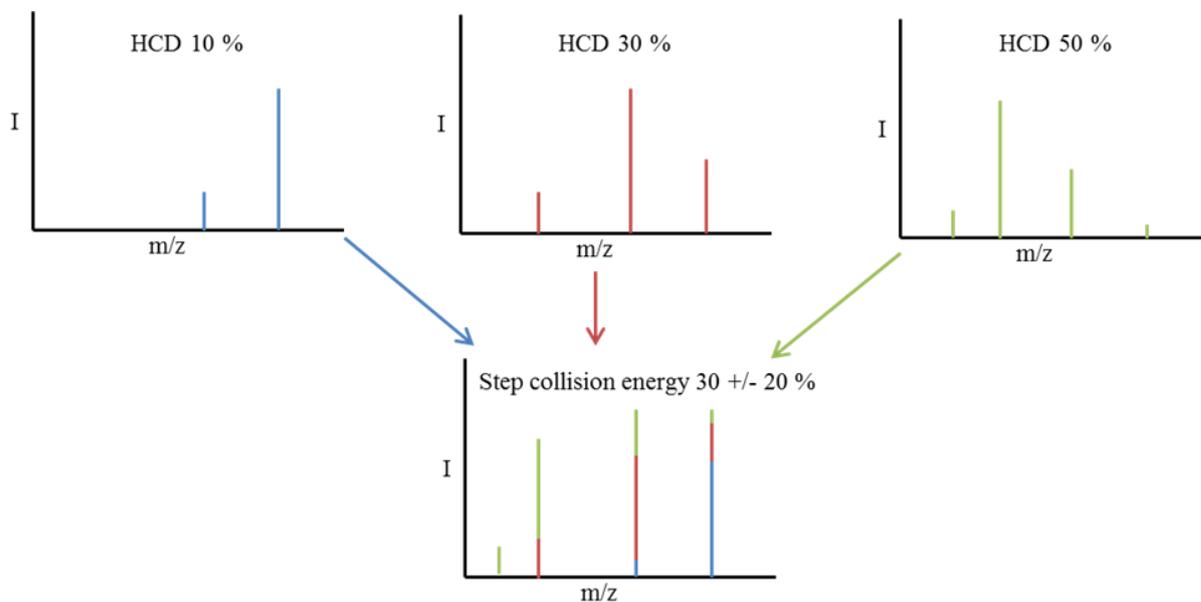


Figure 23 : Principe de la fragmentation en mode "Stepped Collision Energy".

De plus, la combinaison de spectres MS/MS obtenus avec différents modes de fragmentation complémentaires (HCD et CID) permet d'améliorer l'identification d'inconnus (Figure 23¹⁰⁶). Pour cela, les nouveaux spectromètres de masse récents, comme l'Orbitrap Fusion, comportant deux modes de fragmentation (CID et HCD), constituent des instruments intéressants pour la constitution de bases de données spectrales ou pour identifier des métabolites inconnus. La Figure 24 illustre la complémentarité de ces deux modes de fragmentation pour l'élucidation structurale d'un lipide complexe¹⁰⁷.

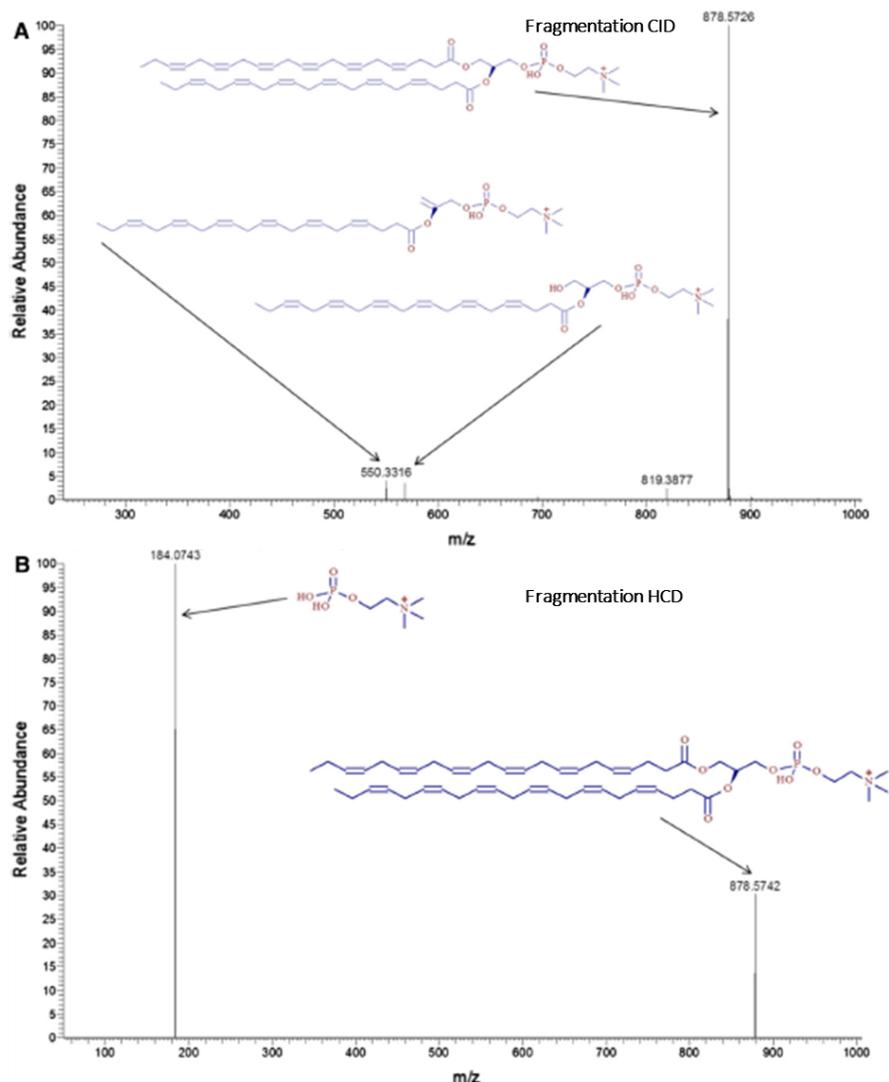


Figure 24 : Comparaison des spectres obtenus avec les modes de fragmentation HCD et CID pour 1,2-didocosahexaenoyl-glycero-3-phosphocholine (adaptée de Mullard et al.).

Un des facteurs importants à prendre en compte pour les analyses DDA (comme pour les analyses DIA décrites ci-après) est le temps cycle, correspondant au temps nécessaire à l'acquisition d'un spectre MS et de plusieurs spectres MS/MS, en particulier lors du couplage du spectromètre de masse avec la chromatographie liquide¹⁸⁴. En effet, le but de la DDA est d'obtenir des spectres de fragmentation pour un maximum de métabolites tout en ayant suffisamment de spectres MS pour permettre l'analyse quantitative, et donc suffisamment de points par pic chromatographique (autour de 10). Les analyses sont alors réalisées en « top-n »

où n représente le nombre de spectres MS/MS acquis entre chaque spectre MS, comme l'illustre la Figure 25. La vitesse d'acquisition des spectres MS et MS/MS est donc un paramètre crucial à considérer pour les analyses DDA¹⁷⁷.

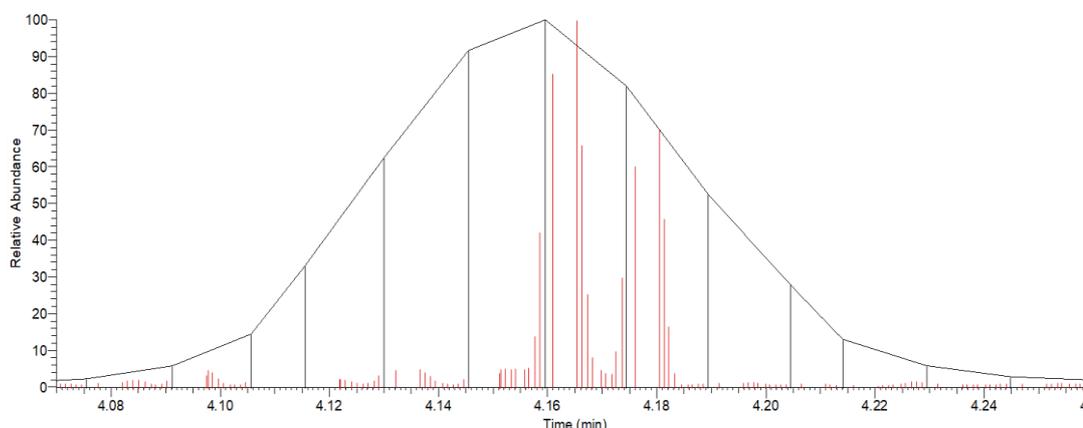


Figure 25 : Exemple de cycles avec un spectre MS (en noir) et plusieurs spectres MS/MS (en rouge) lors d'un pic chromatographique.

La sélection des ions est un facteur limitant pour les analyses DDA. En effet, les spectres MS/MS acquis correspondent aux spectres des ions répondant aux critères initiaux. Dans l'exemple du seuil d'intensité, seuls les « n » ions les plus intenses seront sélectionnés puis fragmentés. De plus, les ions d'intérêt peuvent co-éluer avec des ions plus intenses ce qui empêche leur sélection et donc leur identification. Ces ions les plus intenses peuvent provenir du bruit de fond, ce qui entraîne une perte de temps avec la fragmentation d'ions non pertinents. Une possibilité réside en la création de listes d'exclusion d'ions non pertinents, afin qu'ils ne soient pas sélectionnés.

D'une analyse à l'autre, les ions sélectionnés ne seront pas systématiquement les mêmes ce qui réduit l'aspect reproductible de ce mode d'acquisition¹⁸⁵. En effet, la reproductibilité de la fragmentation des ions est très dépendante de la reproductibilité de l'intensité des signaux mesurés. Si l'intensité des signaux varie d'une analyse à l'autre, l'ordre de sélection des ions varie également. Certains ions peuvent alors ne pas être fragmentés d'une analyse à l'autre.

3 – Les analyses en mode « Data Independent Acquisition »

L'acquisition des données en mode indépendant (DIA) consiste à fragmenter tous les ions détectés. Comme pour le mode DDA, il n'est pas nécessaire d'avoir une connaissance approfondie de l'échantillon, les analyses MS et MS/MS se déroulant simultanément¹⁸². Comme pour le mode DDA, le mode DIA est très utilisé en protéomique contrairement à la métabolomique^{186,187}.

Différents modes d'acquisition DIA existent (Figure 22). La première approche nommée « All-in-one » (MS^{ALL}) consiste à alterner l'acquisition de spectres MS/MS à basse puis haute énergies de fragmentation¹⁸⁸. Dans les deux cas, l'ensemble des ions sont fragmentés simultanément. La fragmentation à basse énergie permet de détecter les ions intacts et la fragmentation à haute énergie permet d'identifier les métabolites grâce aux fragments obtenus. Un peu plus tard, cette technique a été baptisée MS^E par Plumb et al¹⁸⁹. Ce mode d'acquisition est aussi appelé AIF pour « All Ion Fragmentation »¹⁹⁰.

Dans ces conditions, tous les ions, parents et fragments, sont acquis simultanément sur une grande gamme de masse. Les méthodes DIA sont donc couramment utilisées avec des instruments à haute résolution afin de distinguer au mieux l'ensemble des ions fragments^{184,191}. Cependant, les spectres de masse obtenus sont très complexes et il est difficile de faire le lien entre un ion fragment et son précurseur. Pour diminuer cet inconvénient, une méthode d'acquisition nommée SWATH pour « Sequential Window Acquisition of All Theoretical fragment-ion spectra » a été développée (Figure 22)^{192,193}. Cette méthode consiste à décomposer la gamme de masse totale en plusieurs gammes de masse plus petites. Généralement, les fenêtres font entre 20 Da et 50 Da, en fonction aussi de la matrice étudiée¹⁹⁴⁻¹⁹⁶. Comme pour les autres méthodes DIA, dans chaque sous fenêtre, tous les ions sont fragmentés puis détectés. Cela a pour but de diminuer le nombre d'ions dans chaque fenêtre et donc de faciliter l'interprétation des spectres de fragmentation. Comme pour le mode DDA, le temps de cycle a une grande importance pour les analyses SWATH, en raison de la multiplicité des fenêtres d'acquisition. En effet, la vitesse d'acquisition de l'analyseur et donc la résolution pour les analyseurs de type Orbitrap limite le nombre de sous fenêtres de masse¹⁸⁴. Pour cette raison, les Q-TOFs sont souvent les plus utilisés pour les analyses SWATH.

De même que pour le mode DDA, l'un des paramètres limitant l'identification des métabolites est l'énergie de fragmentation appliquée aux composés cibles. Grâce à la grande vitesse d'acquisition des analyseurs de type TOF, des rampes d'énergies de collision peuvent être réalisées afin d'obtenir un maximum de spectres de fragmentation pour chaque métabolite.

Le traitement des données obtenues par DIA est très complexe. Plusieurs logiciels existent pour les analyses protéomiques mais sont difficilement directement transposables à la métabolomique¹⁹⁷⁻¹⁹⁹. Certains logiciels utilisables pour des applications métabolomiques et lipidomiques sont listés dans le Tableau 5¹⁸¹.

Tableau 5 : Liste non-exhaustive de logiciels utilisés pour le traitement des données DDA et DIA (adapté de Fenaille et al.).

Nom	Application	Référence
msPurity	DDA, Evaluation de la pureté des ions précurseurs	T.N. Lawson et al., Anal. Chem. 89 (2017) 2432–2439
MetShot	DDA, Outils informatiques pour améliorer l'acquisition et le traitement des données	S. Neumann et al., Metabolomics 9(2013) 84–91
IE-Omics	DDA, Amélioration de l'acquisition des données grâce à la génération automatisée de listes d'exclusion	J.P. Koelmel et al., J. Am.Soc. Mass Spectrom. 28 (2017) 908–917
MS-DIAL	DIA, Déconvolution des spectres MS/MS et traitement des données. Applicable à la DDA.	H. Tsugawa et al., Nat. Methods 12(2015) 523–526
Met DIA	DIA, Traitement des données pour les analyses ciblées	H. Li at al., Anal. Chem. 88 (2016) 8757–8764
SWATH Tuner	DIA, Construction des fenêtres SWATH	Y. Zhang et al., J. Proteome Res. 14(2015) 4359–4371
NOFI	DIA, Le "non-outlier fragment ion" (NOFI) est un algorithme de classement qui attribue une faible priorité aux fragments d'ions affectés par des interférences lors de l'acquisition de DIA.	A. Bilbao et al., J. Proteome Res. 14 (2015) 4581–4593
MetaboDIA	DIA, Quantification par MS/MS avec les données DIA	G. Chen et al., Anal. Chem. 89(2017) 4897–4906
LipidMatch	DDA, DIA, Identification des lipides à l'aide de données non ciblées de spectrométrie de masse en tandem à haute résolution	J.P. Koelmel et al., BMC Bioinf. 18 (2017) 331

La fragmentation des protéines est connue et prédictible. Les bases de données sont alors plus faciles à générer *in silico* ce qui facilite la déconvolution des spectres DIA-MS/MS obtenus. En métabolomique, les principes de fragmentation dépendent des molécules et nécessitent l'acquisition préalable de spectres MS/MS de références à partir de standards¹⁹¹. Une comparaison est alors réalisée entre le spectre de référence et celui obtenu lors de l'étude d'intérêt. De solides bases de données sont donc nécessaires pour tenter d'interpréter les spectres. Afin de s'assurer au mieux de l'identité des métabolites, ces bases de données (internes et/ou standardisées) doivent contenir le RT ainsi que le m/z de l'ion parent et des fragments.

Comme pour le mode DDA, les analyses en mode DIA peuvent permettre de détecter, identifier et quantifier les métabolites en une seule analyse. Cependant, l'avantage du mode DIA est d'avoir des informations spectrales pour tous les ions détectés contrairement à la DDA¹⁹⁴.

4 – Comparaison des différents modes d'acquisition MS/MS

Le Tableau 6 dresse un premier bilan des trois modes d'acquisition de spectres MS/MS décrits précédemment. Une comparaison plus poussée sera présentée dans la partie « Résultats ».

Tableau 6 : Bilan des modes d'acquisition MS/MS.

Mode d'acquisition MS/MS	Ciblé	DDA	DIA
Fenêtre d'isolement du précurseur	< 1 Da	< 1 Da	Large - entre 10 et 100 Da
Pureté des Spectres MS/MS	Bonne	Bonne	Faible
Adaptation en routine (connaissance échantillon, liste inclusion / exclusion...)	Intermédiaire	Intermédiaire	Bonne
Quantification	MS/MS	MS	MS et MS/MS

Le point fort lié à l'acquisition des spectres MS/MS en mode ciblé est la sensibilité de la méthode ainsi que la pureté des spectres MS/MS obtenus mais pour identifier un grand nombre de métabolites, une très grande connaissance de l'échantillon est nécessaire. En mode DDA, le principal avantage est la pureté des spectres ainsi que l'alternance automatique de l'acquisition des spectres MS et MS/MS. Cependant, la couverture du métabolome est restreinte dû aux critères de sélection implémentés et parfois trop restrictifs. Concernant le mode d'acquisition DIA, le point fort est la couverture du métabolome mais les spectres MS/MS obtenus sont très complexes ce qui limite l'identification des métabolites, dont la sensibilité de détection est souvent moindre comparativement aux autres modes²⁰⁰. L'ensemble de ces méthodes présentent de bonnes caractéristiques pour leur utilisation pour la quantification de métabolites.

Un des objectifs de ma thèse a donc été de développer et de comparer les méthodes DDA et DIA pour l'identification large spectre de métabolites en se focalisant sur un échantillon biologique d'intérêt, le plasma humain. Ces approches permettent d'identifier les métabolites grâce à l'acquisition de spectre MS/MS et la validation de biomarqueurs, pour un futur diagnostic clinique, avec la quantification absolue réalisée sur les spectres MS.

Chapitre 3 : La quantification en métabolomique

A – La quantification : contexte général

Comme décrit précédemment, les analyses métabolomiques par LC-ESI-MS/MS sont souvent impactées par les effets matrice en raison de la coexistence et la co-elution de nombreux composés présents dans les matrices biologiques complexes. Pourtant, en métabolomique quantitative, il est primordial de pouvoir comparer les échantillons analysés de façon fiable et précise. A cet égard, la robustesse des analyses peut être garantie par la mise en place d'approches efficaces de normalisation des signaux pour les profilages non-ciblés (la quantification relative correspondant à la comparaison des niveaux d'abondance entre deux échantillons, métabolite par métabolite) ou par des stratégies de quantification absolue (détermination des concentrations de métabolites dans les extraits) de métabolites pour les analyses quantitatives ciblées^{38,201–203}.

La méthode de choix en terme de quantification, qu'elle soit relative ou absolue (définies ci-dessous), reste le recours aux isotopes stables (tels que carbone-13, deutérium, azote-15...), que cela soit sous forme de métabolites incorporant des isotopes stables constitutifs de leur structure ou après dérivatisation chimique avec des groupements isotopiquement alourdis. Pour que cette quantification soit la plus juste possible, le standard marqué utilisé doit avoir des propriétés physico-chimiques très similaires aux métabolites d'intérêt, l'idéal étant qu'il soit son homologue isotopiquement et constitutivement marqué. Cependant, l'ensemble des métabolites ne possède pas leurs versions marquées disponibles commercialement. Des développements, détaillés dans la prochaine section de ce chapitre, sont donc réalisés afin de normaliser les signaux grâce à une sélection de standards internes marqués aux isotopes stables pour les différentes analyses métabolomiques.

La quantification de métabolites fait face à de nombreuses contraintes. En effet, la gamme de concentration des métabolites dans une matrice étant très large, les méthodes de quantification doivent donc couvrir une grande gamme dynamique²⁰⁴. De plus, le nombre de métabolites détectés puis annotés dans une matrice biologique restant important, il est très complexe de quantifier un grand nombre de ces composés simultanément²⁰⁵. Couramment, une dizaine de métabolites sont quantifiés simultanément lors d'une analyse quantitative ciblée. De nombreuses méthodes analytiques ont été développées afin de répondre à ces besoins notamment avec l'utilisation de standards internes marqués.

1 – La quantification relative et la quantification absolue

Deux approches de quantification sont régulièrement mises en œuvre en métabolomique : la quantification relative et la quantification absolue (Figure 26). La quantification relative est, de loin, la plus

largement appliquée dans les études métabolomiques non ciblées (par exemple de découverte de biomarqueurs) et consiste à comparer les niveaux d'abondances d'un même métabolite d'un échantillon à un autre. C'est sur ces comparaisons de niveaux d'abondances de plusieurs centaines de signaux que repose la mise en évidence de biomarqueurs par les analyses statistiques suivant le pré-traitement des analyses LC-HRMS de cohortes d'échantillons. Par opposition, la quantification absolue consiste, quant à elle, à convertir l'abondance d'un ion en concentration. Cette concentration peut être exprimée en concentration massique (par exemple en $\mu\text{g/mL}$) ou en concentration molaire (par exemple en $\mu\text{mol/mL}$)²⁰⁵. La quantification absolue permet de comparer les échantillons injectés lors d'une même séquence, mais également lors d'analyses réalisées à des temps différents ou par des laboratoires différents. De plus, la concentration obtenue est indépendante de la technique analytique utilisée. La méthode de choix pour la quantification absolue est l'utilisation de standards internes marqués aux isotopes stables. Ces standards n'étant pas toujours disponibles et/ou coûtant très chers, ces analyses sont souvent réalisées sur un nombre restreint de composés, et concernent quelques dizaines de métabolites tout au plus²⁰⁶. Ces standards sont donc plus particulièrement utilisés pour la validation de biomarqueurs et ne sont pas souvent présents dans les approches de profilage métabolique.

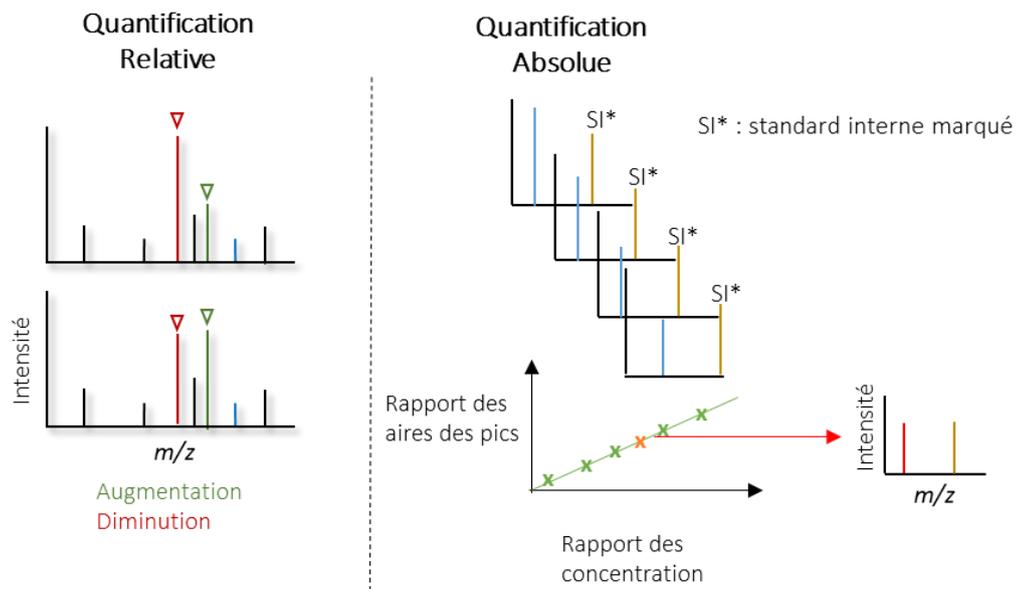


Figure 26 : La quantification relative et la quantification absolue.

2 – Les différentes approches de quantification par spectrométrie de masse

Différentes méthodes analytiques sont utilisées pour la quantification des métabolites. Elles peuvent dépendre de l'analyseur utilisé. La quantification des métabolites peut être réalisée grâce à l'ion précurseur ou via l'étude des fragments (c'est-à-dire respectivement à partir des spectres MS et MS/MS).

a – La quantification basée sur l'ion précurseur

La quantification de l'ion précurseur est possible à partir de spectres MS que ce soit sur les « Full Scan » ou dans des acquisitions appelées « Single Ion Monitoring » (SIM) ou « Multiple Ion Monitoring » (MIM)²⁰⁵. Ces deux derniers modes d'acquisition permettent l'obtention de spectres de masse plus sensibles que lors des expériences de profilage. Ces acquisitions impliquent préférentiellement un analyseur à haute, voire très haute, résolution afin de ne quantifier que le métabolite d'intérêt avec suffisamment de spécificité en limitant le recouvrement de signaux de composés isobares. Malgré des performances similaires au niveau de la sensibilité et de la gamme dynamique, ces techniques sont moins populaires que la quantification par MRM ou PRM (décrites ci-après)²⁰⁷. La quantification des métabolites basée sur la mesure des fragments apparaît plus spécifique, en particulier si on la compare aux performances de quantification obtenues en MS avec des équipements de résolution inférieure ou égale à 120 000²⁰⁸. La spécificité plus marquée de la spectrométrie de masse en tandem est, néanmoins de nos jours, plus discutable avec l'apparition d'Orbitrap à ultra haute résolution permettant la séparation d'isobares à 0.8 mDa près pour des ions de rapports m/z d'environ 200 (avec une résolution de 240 000). Notons également que si les approches par spectrométrie de masse en tandem (MS/MS) offrent une très bonne sensibilité (rapport signal/bruit) et spécificité aux analyses quantitatives, ces approches restent applicables à un nombre limité de molécules à quantifier par analyse (comme décrit dans la partie qui suit) et la quantification sur la base des spectres MS semble donc plus adaptée aux analyses quantitatives large spectre.

b – La quantification basée sur les fragments

La spectrométrie de masse en tandem (MS/MS) est une approche efficace pour augmenter la sensibilité (le rapport signal sur bruit) et la spécificité des analyses. Les analyses ciblées sont les plus utilisées pour la quantification absolue de métabolites. Plusieurs méthodes sont utilisées pour ces analyses ciblées : le mode SRM (« selected reaction monitoring ») et MRM (« multiple reaction monitoring ») à basse résolution, ou encore le mode PRM (« parallel reaction monitoring ») à haute résolution^{205,209} (Figure 27²¹⁰).

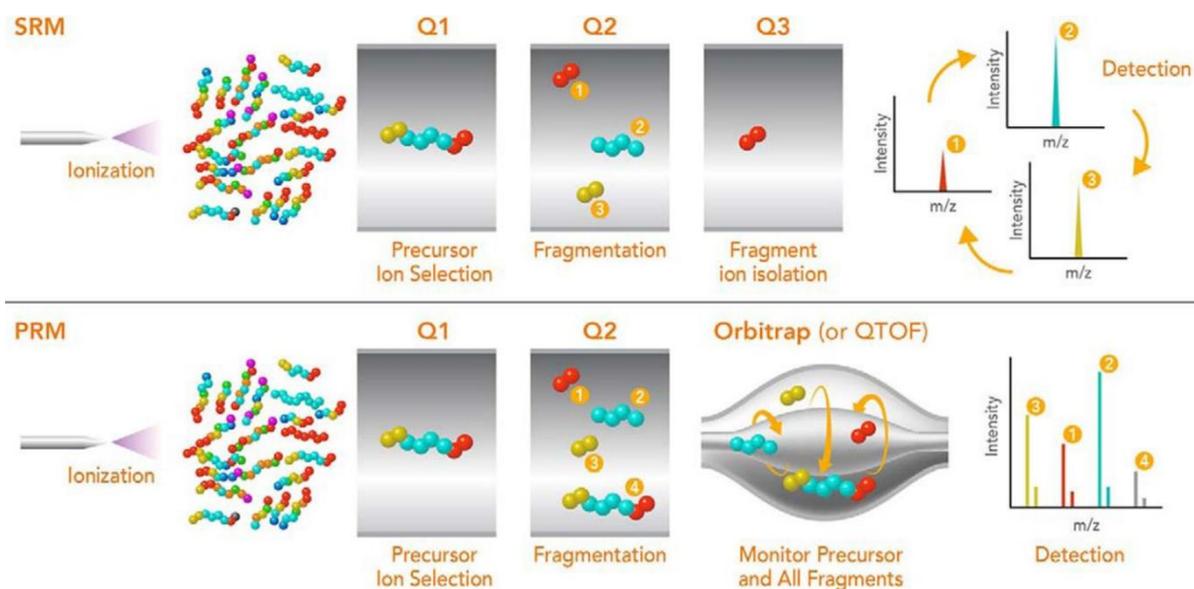


Figure 27 : Comparaison des Modes d'Acquisition SRM (/MRM) et PRM (adaptée de Burnum-Johnson et al.).

Le mode SRM repose sur l'isolement d'un ion précurseur. Cet ion est ensuite fragmenté et chaque fragment est sélectionné indépendamment pour être détecté. Le mode MRM est basé sur le même principe avec le suivi de plusieurs réactions de fragmentation. La quantification par MRM est la méthode de choix pour la quantification absolue des métabolites²⁰⁹. Elle requiert une bonne connaissance des composés à quantifier. En effet, il est nécessaire de connaître les fragments uniques du métabolite, c'est-à-dire, les fragments n'appartenant qu'au métabolite d'intérêt dans la fenêtre de temps de rétention correspondante. Ces modes d'acquisition sont principalement réalisés avec des triples quadripôles où le premier quadripôle permet d'isoler l'ion précurseur, le deuxième sert de cellule de collision et permet la fragmentation et le troisième sert d'analyseur pour les fragments générés^{175,211}. En PRM, de même qu'en SRM/MRM, un ion précurseur est sélectionné puis fragmenté. L'utilisation d'un analyseur à haute résolution en mode PRM permet cependant de suivre simultanément et à haute résolution tous les produits de fragmentation obtenus²¹². Pour ce dernier mode, les instruments de type Q-Orbitrap et Q-TOF sont utilisés²⁰⁵. La vitesse d'acquisition des spectres est donc moins rapide qu'en MRM ce qui limite le nombre de composés quantifiables en une injection et réduit sa sensibilité²⁰⁵.

Les analyses DIA permettent également de quantifier les métabolites grâce aux fragments si le nombre d'acquisitions MS/MS est suffisant au cours du pic chromatographique (au moins huit à dix spectres MS/MS par pic chromatographique). La fragmentation automatique de l'ensemble des composés est un grand avantage car elle permet la quantification d'un grand nombre d'entre eux en une seule injection. Aujourd'hui, les performances de quantification obtenue par DIA sont légèrement moins bonnes que celles obtenues par MRM, qui est considérée comme la méthode de choix pour la quantification mais permet de couvrir une gamme de métabolites beaucoup plus large²¹³. En revanche, il est d'autant plus nécessaire dans ce cas d'étudier des fragments spécifiques au métabolite d'intérêt.

B – L'utilisation des isotopes stables pour la quantification des métabolites

L'utilisation de composés marqués aux isotopes stables pour la quantification de métabolites est appliquée depuis de nombreuses années. En effet, ces composés présentent l'avantage de posséder les mêmes propriétés physico-chimiques que leurs homologues non marqués et permettent ainsi de s'affranchir de l'effet matrice²⁰⁴. Ces aspects de correction de l'effet matrice seront développés un peu plus loin dans cette partie du manuscrit.

1 – Standard interne marqué : définition, propriétés et limitations

Un standard interne marqué est un composé incorporant différents isotopes stables « lourds » et possédant les mêmes propriétés physico-chimiques que son homologue non marqué, en particulier vis-à-vis de la réponse analytique que ce soit au niveau de la chromatographie ou de l'ionisation en spectrométrie de masse. Comme présenté précédemment dans le Tableau 3, la faible abondance naturelle de certains isotopes stables tels que le carbone-13, l'azote-15 ou encore le deutérium, permet d'envisager leur utilisation pour le marquage de composés organiques d'intérêt en appliquant des méthodes d'enrichissement isotopique qui seront décrites par la suite²¹⁴. L'enrichissement isotopique peut ainsi conduire à différentes versions marquées d'un même métabolite selon l'isotope « lourd » utilisé et selon le degré d'enrichissement obtenu. Ainsi, chaque isotopologue (ou analogue isotopique de même formule brute que le métabolite non marqué aux isotopes près) sera généralement noté Mn où M représente l'ion moléculaire étudié et n le nombre d'isotopes stables lourds (carbone-13, azote-15 ou deutérium) incorporés dans sa structure. Une illustration des différents isotopologues deutérés de la sérine est présentée en Figure 28. Ces isotopologues sont différenciables par spectrométrie de masse et dans le cas de l'incorporation du deutérium, une différence de 1.0063 Da est observée pour chaque atome de deutérium supplémentaire dans la structure. Les isotopomères (ou isomères isotopiques) sont, quant à eux, des isomères de même formule brute (isotope(s) inclus) qui ne diffèrent que par le positionnement du ou des isotope(s) dans la structure comme illustrée en Figure 28, et sont donc confondus sous le même pic de masse.

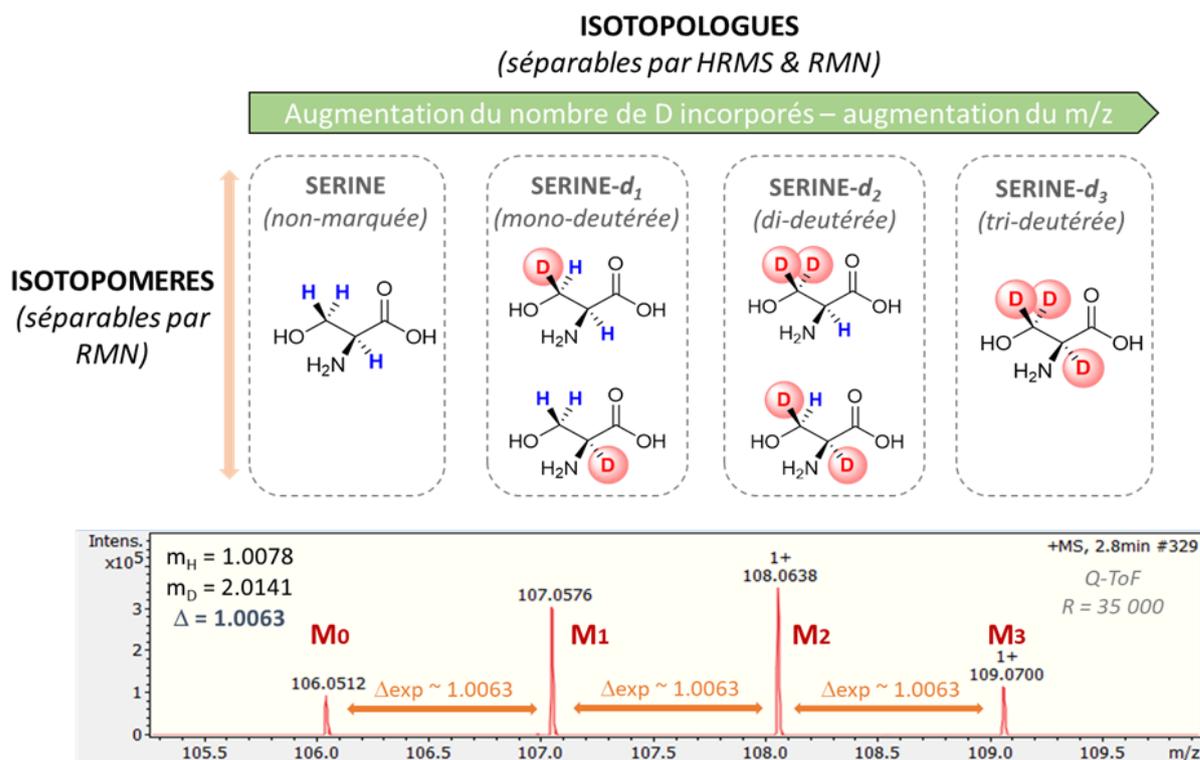


Figure 28 : Isotopologues et isotopomères : illustration avec l'exemple de la sérine marquée au deutérium.

2 – Les voies de production des standards internes marqués

Plusieurs approches permettant l'obtention de molécules marquées aux isotopes stables ont été développées jusqu'à présent et peuvent être regroupées en deux catégories : les approches par marquage métabolique (ou marquage *in vivo*) et les approches par marquage chimique (ou marquage *in vitro*). La Figure 29²¹⁵ illustre très bien ces deux types d'approches qui seront développées plus en détails dans la suite du manuscrit.

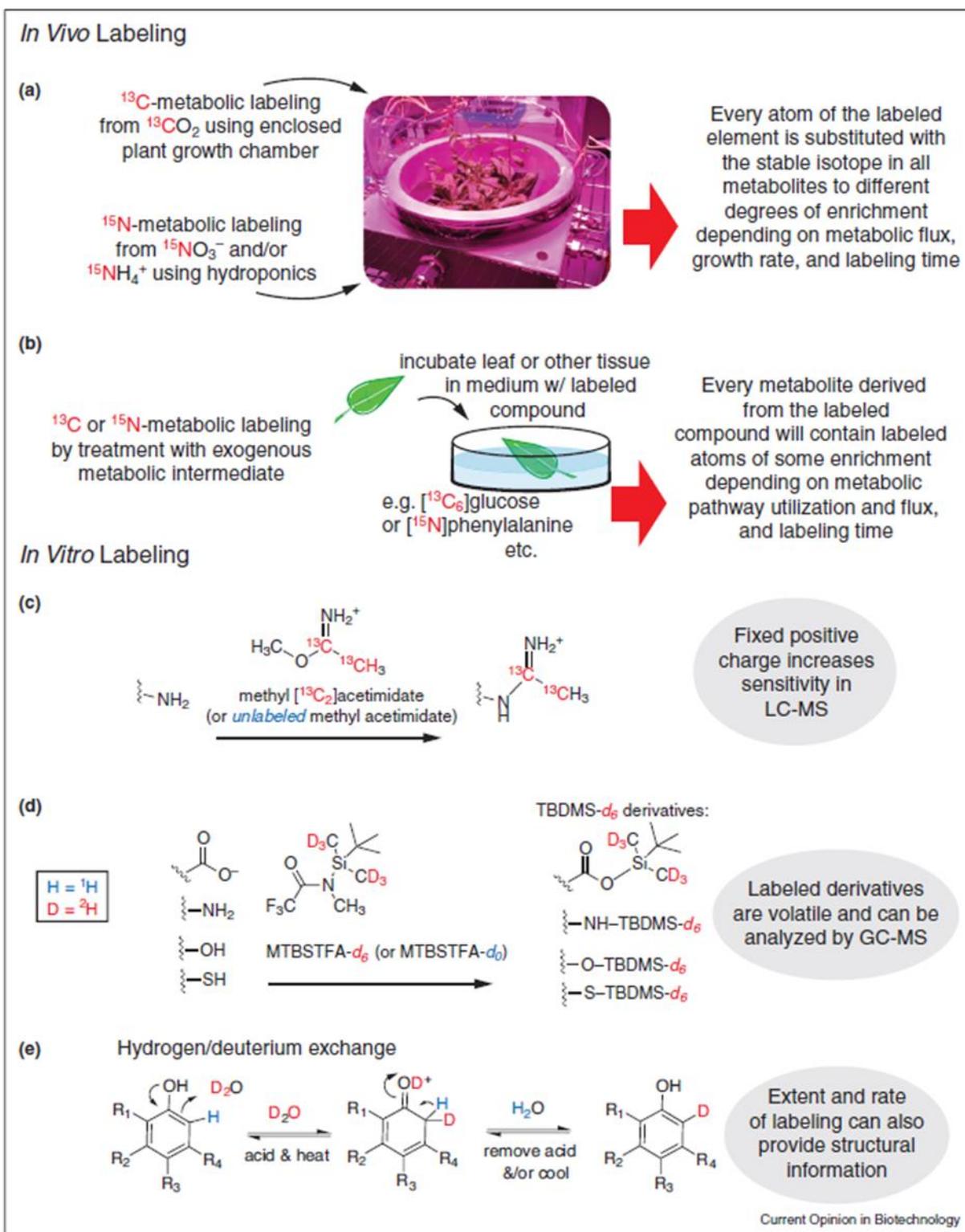


Figure 29 : Illustration des deux types d'approches (*in vivo* et *in vitro*) pour la production de composés marqués aux isotopes stables (extraite de Freund et al.).

L'approche *in vivo* consiste à produire des métabolites marqués, le plus souvent au carbone-13 ou à l'azote-15, par action métabolique (ou enzymatique) d'un organisme (plantes, champignons, petits mammifères, etc) alimentés en nutriments marqués. L'approche *in vitro* consiste, quant à elle, à produire chimiquement soit

par dérivatisation chimique soit par marquage direct (par exemple échange H/D stable) des métabolites appartenant le plus souvent à une famille chimique particulière.

a – Production in vivo de standards marqués

La synthèse chimique d'isotopologues contenant du ^{13}C ou ^{15}N est souvent longue et complexe. Une alternative à ce mode de production *in vitro* est la production *in vivo*, encore appelée production enzymatique ou métabolique. Cette méthode consiste à faire croître un organisme en présence de nutriments marqués (le plus souvent au ^{13}C ou/et au ^{15}N) afin que celui-ci produise l'ensemble des molécules constitutives de son métabolome dans une version marquée. Elle a été développée originellement pour l'analyse fluxomique, qui s'intéresse à la mesure de la dynamique du métabolisme et représente le résultat final de toutes les interactions métaboliques. Le marquage *in vivo* a principalement été développée avec des plantes²¹⁶ ou des microorganismes²¹⁷. A titre d'exemple, des métabolites de levure intégralement marqués au carbone-13 peuvent être obtenus par la mise en culture de levures dans un milieu nutritif constitué de glucose- $^{13}\text{C}_6$ comme seule source de carbone. De même, des versions marquées de métabolites de plantes peuvent être générées par la mise en croissance de ces végétaux sous atmosphère contrôlée de $^{13}\text{CO}_2$ ²¹⁸⁻²²⁰. Ce type d'approche *in vivo* présente l'avantage de donner accès à un nombre important de composés marqués simultanément. Il offre également la possibilité de mettre en évidence les voies métaboliques impliquées dans la biotransformation de chaque métabolite^{221,222}.

Après des études préalables de caractérisation et de quantification, les lots d'échantillons biologiques marqués peuvent être utilisés soit pour quantifier soit pour normaliser des signaux LC-HRMS obtenus lors d'études métabolomiques d'intérêt²¹⁹. Le choix de l'extrait biologique marqué, par rapport à l'échantillon d'étude, est donc crucial. Il peut s'avérer judicieux d'utiliser un échantillon marqué métaboliquement le plus proche possible de l'échantillon d'étude²²³. En effet, afin de quantifier un maximum de métabolites simultanément, il faut s'assurer que le recouvrement métabolique entre l'échantillon marqué et l'échantillon d'intérêt soit important et que les concentrations (ou abondances relatives) des métabolites de l'organisme marqué soient similaires à celles de l'échantillon d'étude. Des organismes plus gros que ceux précédemment cités, tels que des rongeurs, ont également été marqués pour des applications protéomiques. L'objectif est d'augmenter l'applicabilité d'utilisation de ces échantillons biologiques marqués aux différents types d'échantillons rencontrés^{218,224}. Pour cela, les rongeurs sont alimentés avec de la spiruline ^{13}C et/ou ^{15}N ²²⁵. Cependant, le coût important du marquage complet d'un organisme limite l'utilisation de ces échantillons comme standards internes marqués. Pour pallier cela, des alternatives ont été développées comme le marquage incomplet de l'organisme. La culture ou l'élevage de l'organisme dure moins longtemps ce qui limite les coûts sans réduire l'efficacité du standard marqué pour la quantification ou la normalisation des signaux d'intérêt²²⁴.

b – L'approche par marquage chimique (ou marquage in vitro)

Depuis de nombreuses années, différentes stratégies d'incorporation d'isotopes stables par voie chimique dans les structures de molécules à quantifier ont été développées. Compte tenu du grand nombre de composés constituant le métabolome et la grande variété de leurs structures chimiques, la production de composés marqués aux isotopes stables par synthèse totale est inenvisageable car trop longue et fastidieuse. Ainsi, l'ensemble des approches développées pour introduire un ou des isotopes stables dans les structures chimiques doivent répondre à deux critères principaux : la rapidité et l'efficacité d'incorporation des isotopes. Une méthode rapide de marquage se résume le plus souvent à l'application d'un protocole mono-étape, c'est pourquoi les méthodes de marquage chimique reposent principalement sur des techniques de dérivatisation chimique à l'aide de réactifs de couplage marqués ou sur des stratégies d'échange isotopique (le plus souvent hydrogène/deutérium). Dans certains cas où le marquage direct par dérivatisation chimique ou échange isotopique n'est pas possible, le recours à des analogues chimiques, dans lesquels seul un atome ou groupe d'atomes diffèrent des métabolites à quantifier, peut être envisagé. Ces composés présentent néanmoins l'inconvénient majeur de ne pas toujours offrir exactement les mêmes propriétés physico-chimiques que le composé à quantifier par LC-MS (différence en temps de rétention et/ou facteur de réponse non équivalent). La Figure 30²²⁶, résume l'ensemble des différents standards internes pouvant être utilisés pour la quantification de l'estradiol²²⁶.

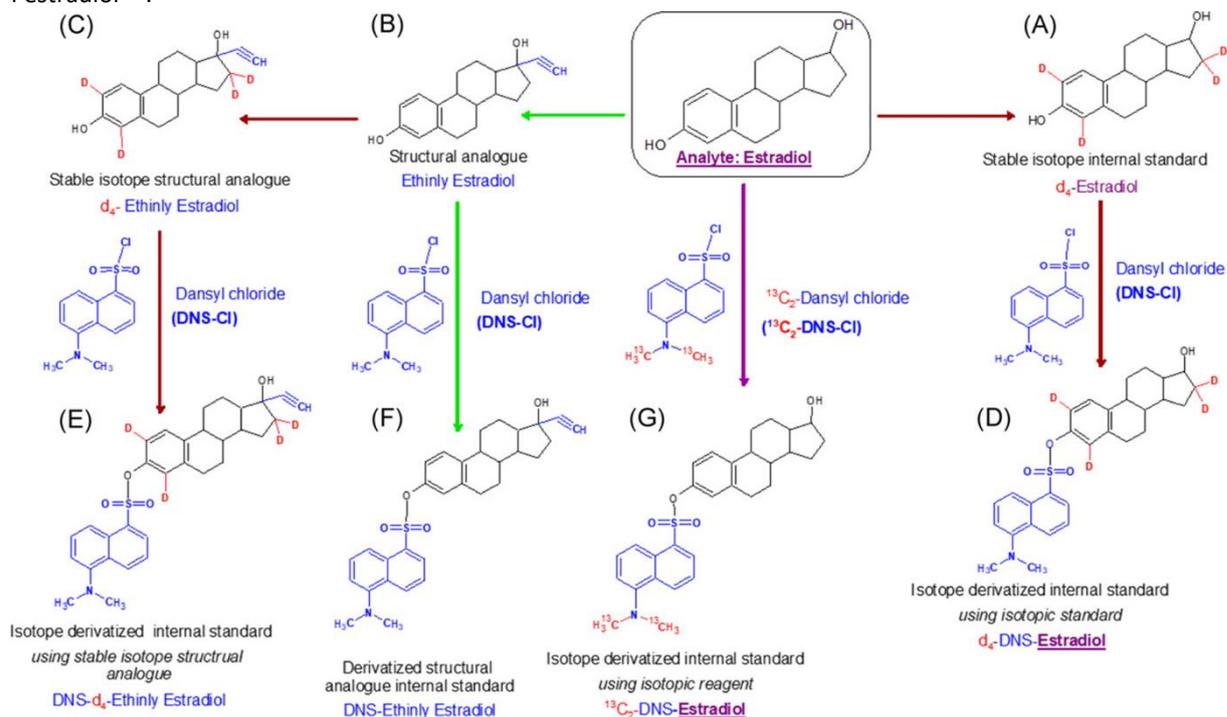


Figure 30 : Les différents standards internes pour les analyses métabolomiques quantitatives (extraite de Khamis et al.).

- *La dérivatisation chimique*

La dérivatisation chimique est utilisée depuis plusieurs décennies en métabolomique (et également en protéomique) quantitative^{227,228}. Cette technique repose sur la conversion d'une fonction chimique contenue par le métabolite d'intérêt en un groupement fonctionnel le plus souvent marqué avec un nombre suffisant d'isotopes stables (« lourds ») pour permettre un déplacement en masse suffisant pour le distinguer de son homologue non marqué. Ainsi le processus de dérivatisation chimique est souvent appliqué avec deux versions du même réactif, une version « légère » ou non-marquée et une version « lourde » c'est-à-dire comportant des isotopes stables « lourds ». La pratique de la dérivatisation chimique permet, selon qu'elle est appliquée exclusivement à des extraits biologiques à analyser ou également à des standards analytiques, de réaliser de la quantification relative ou absolue comme illustré sur la Figure 31.²²⁹ Par ailleurs, la Figure 32²³⁰, met également en évidence l'application large-spectre de la stratégie de dérivatisation chimique pour l'analyse du métabolome. Dans ce cas, l'utilisation simultanée de quatre réactifs de dérivatisation chimique pour cinq groupes distincts de fonctions chimiques (amines, phenols, hydroxyles, carbonyles, en particulier aldehyde et cétone, et acides carboxyliques) a été appliquée et a permis de démontrer que plus de 85% du métabolome plasmatique humain contenait au moins l'une de ces fonctions chimiques²³⁰.

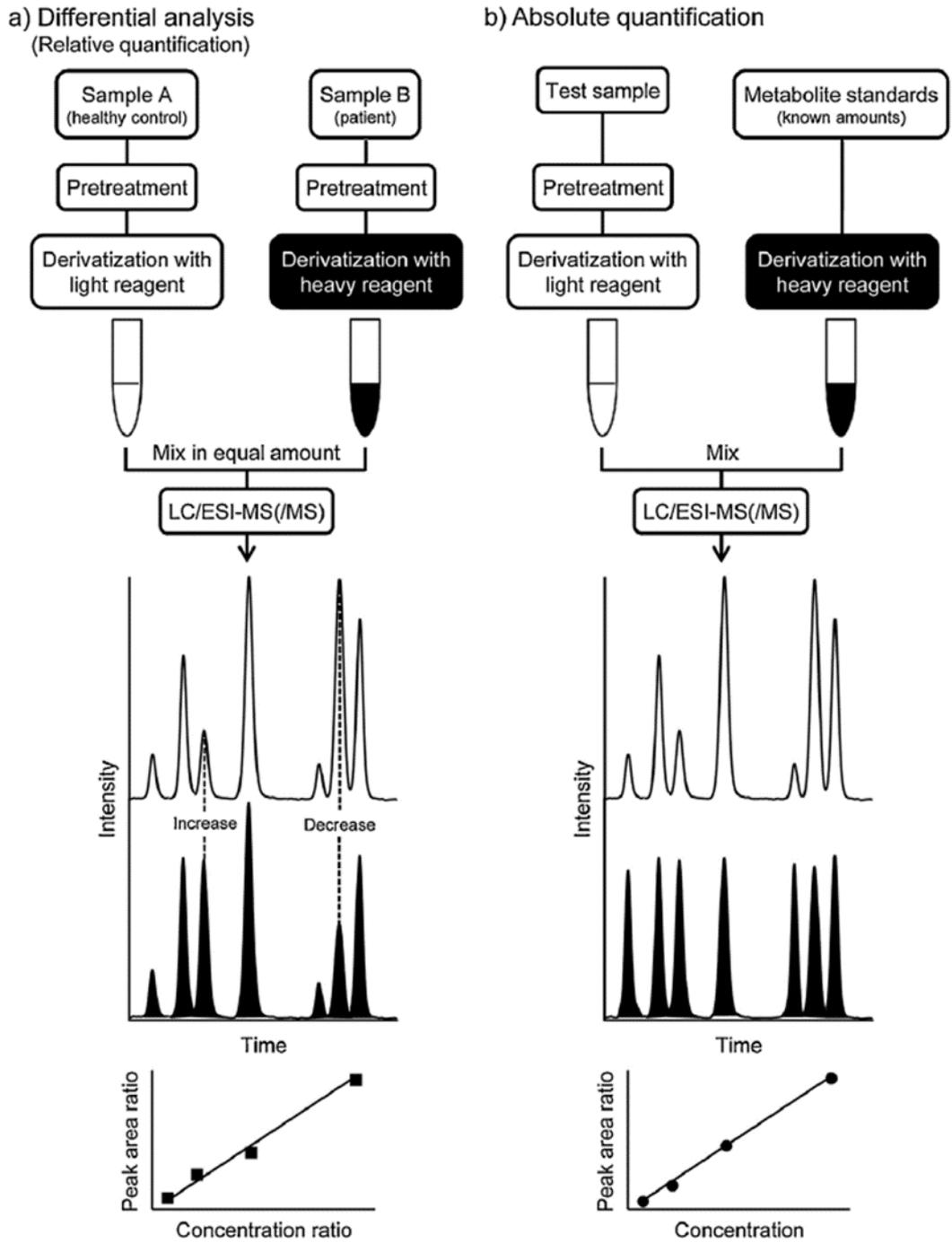


Figure 31 : Illustration des méthodes de quantification relative et absolue par dérivatisation chimique (extraite de Higashi et al.).

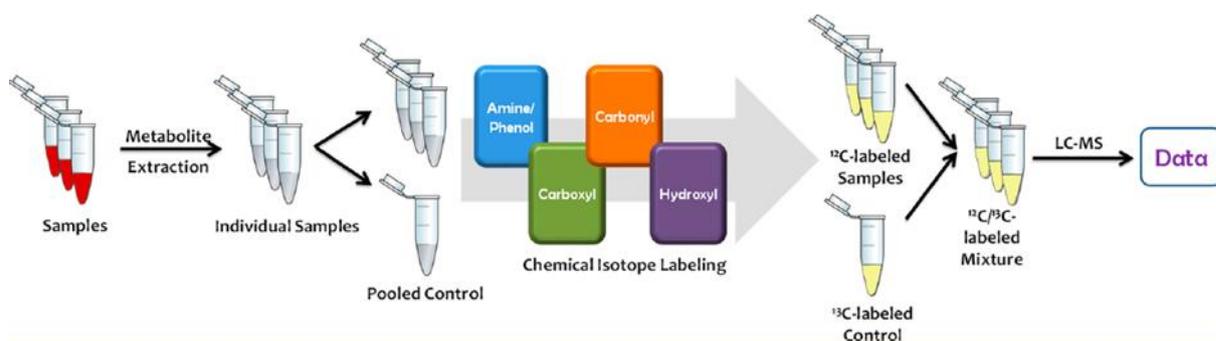


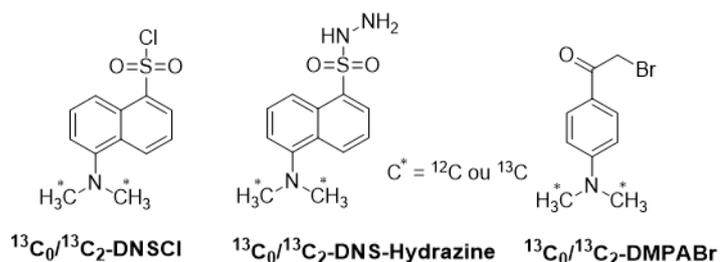
Figure 32 : Stratégie de dérivation chimique appliquée à l'analyse du métabolome.

Les fonctions chimiques le plus couramment modifiées chimiquement ainsi que des exemples non exhaustifs d'agents de couplage utilisés sont présentés dans le Tableau 7²³⁰.

Tableau 7 : Exemple de réactions de dérivation selon la famille chimique (adapté de Zhao et al.).

DNSCl : dansylchloride ; *DNS-Hydrazine* : dansyl-hydrazine ; *DMAP* : 4-diméthylaminopyridine ; *DmPABr* : diméthylaminophenacyl bromide

Fonction chimique	Réaction de dérivation	Exemple de couple de réactifs
Amine / Phénol	Dansylation	$^{13}\text{C}_0/^{13}\text{C}_2$ -DNSCl
Carbonyle	Formation d'hydrazone	$^{13}\text{C}_0/^{13}\text{C}_2$ -DNS-Hydrazine
Hydroxyle	Dansylation activée	$^{13}\text{C}_0/^{13}\text{C}_2$ -DNSCl + DMAP
Carboxyle	Estérification	$^{13}\text{C}_0/^{13}\text{C}_2$ -DmPABr



La dérivation chimique s'avère être une technique très intéressante pour des études où des métabolites appartenant à une même famille chimique sont ciblés. Cette technique peut également permettre d'augmenter les propriétés d'ionisation (ou de rétention) de certains composés qui, sans cela, sont en dessous des limites de détection avec les méthodes analytiques utilisées. La dérivation chimique est également une approche pouvant s'avérer efficace pour la quantification en mode SRM, en particulier pour les ions ayant de mauvaises propriétés de fragmentation lors de dissociation induite par collision (CID) à faible énergie. L'introduction d'un groupement chimique fragmentable permet d'augmenter la sensibilité et la spécificité de la méthode de quantification pour ces composés en permettant la formation d'un ion produit suffisamment intense à partir du métabolite modifié chimiquement^{231,232}.

- L'échange isotopique

Les réactions d'échange isotopique, c'est-à-dire de remplacement direct d'un atome par son isotope stable « lourd », ont également été largement investiguées pour la production de standards internes marqués pour des applications quantitatives par spectrométrie de masse. Ces réactions d'échange isotopique concernent presque essentiellement le remplacement d'atome(s) d'hydrogène par leur isotope stable « lourd » : le deutérium. On peut noter cependant que des développements pour l'échange isotopique d'autres atomes constitutifs des molécules organiques commencent à émerger avec notamment, et à titre d'exemple, la mise au point (impliquant notamment des collègues du CEA de Saclay) d'une stratégie mono-étape permettant d'échanger un atome de carbone-12 par son isotope carbone-13 spécifiquement sur des composés carboxylés par réaction de décarboxylation/carboxylation en présence de $^{13}\text{CO}_2$ comme illustré sur la Figure 33²³³.

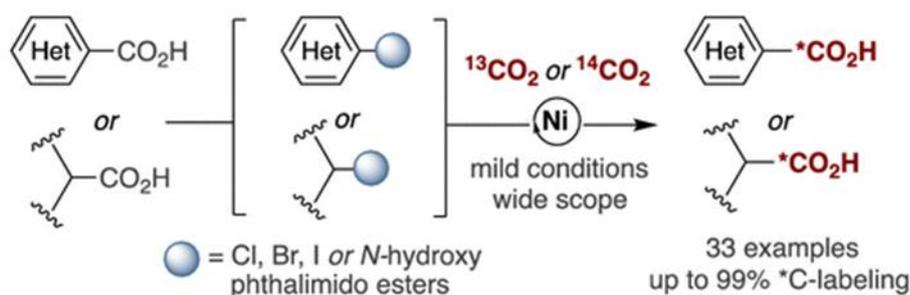


Figure 33 : Illustration d'une réaction d'échange $^{12}\text{C}/^{13}\text{C}$ par décarboxylation/carboxylation (extraite de Tortajada et al.).

Malgré ces récents développements, les réactions d'échanges H/D restent actuellement les seules réactions d'échange isotopique appliquées aux analyses métabolomiques quantitatives. Notons d'ailleurs que les hydrogènes sont plus ou moins labiles selon leur position dans les édifices moléculaires. En effet, des hydrogènes portés par des hétéroatomes tels que l'oxygène et l'azote sont facilement échangeables. Ainsi, l'utilisation de solvants deutérés lors de la préparation d'échantillons ou de phases mobiles utilisées lors d'analyse par chromatographie liquide permet d'échanger ces hydrogènes et d'obtenir des informations structurales pertinentes par spectrométrie de masse sur des composés présents en matrice complexe²³⁴. Ces positions de deutération sont considérées comme instables et échangeables en sens inverse (« back-exchange ») en présence de solvants protiques ou de traces d'eau non deutérée. Ainsi, nous parlerons dans cette section d'échange H/D sur des positions stables où les atomes d'hydrogène sont portés par des atomes de carbone permettant ainsi une utilisation des molécules marquées résultantes comme standards internes.

Les réactions d'échange H/D conduisant à des marquages stables sont généralement catalysées, soit par une base ou un acide, soit par un métal. Elles sont le plus souvent réalisées dans un solvant protique deutéré qui joue le rôle de source de deutérium en présence ou non de deutérium gaz (jouant également le rôle de source de deutérium). Dans l'exemple de la Figure 29, la catalyse acide permet l'incorporation d'un atome de deutérium sur un noyau aromatique. La majorité des catalyses, pour l'échange H/D stable, reste métalliques : elles peuvent

être homogènes²³⁵ (c'est-à-dire, réalisées à l'aide d'un catalyseur métallique miscible au milieu réactionnel) ou, à l'inverse, hétérogènes. Dans le dernier cas, le catalyseur métallique pourra être facilement éliminé par simple filtration du milieu réactionnel. Le palladium, le platine, le rhodium ou encore le ruthénium sont des catalyseurs très utilisés pour ces réactions d'échange H/D, qu'elles soient homogènes ou hétérogènes²³⁶⁻²³⁹. Il a été observé que les catalyseurs au platine permettent une plus grande sélectivité d'échanges des hydrogènes en positions aromatiques tandis que le palladium affiche une préférence pour les hydrogènes en positions aliphatiques (Figure 34)²⁴⁰.

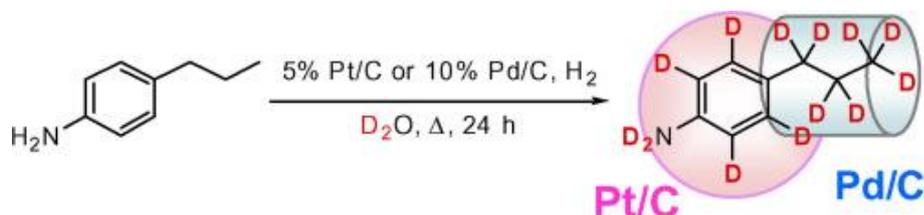


Figure 34 : Réaction d'échange H/D avec le Platine ou le Palladium comme catalyseur (adaptée de Sajiki et al.).

Les exemples présentés dans la Figure 35²⁴¹, mettent en évidence les régiospécificités de deutération propres à chaque catalyseur métallique. Notons que le rhodium, tout comme le palladium, semble bien adapté à la deutération de chaînes aliphatiques alors que le ruthénium présente la particularité de permettre une réaction d'échange H/D particulièrement orientée sur les hydrogènes portés par des atomes de carbone en alpha de groupements hydroxyles^{241,242}. Plus récemment, il a été montré que non seulement le ruthénium permettait un marquage régiospécifique en alpha des fonctions alcools mais également en alpha des fonctions amines dans des conditions douces (température et pression modérées)²³⁷.

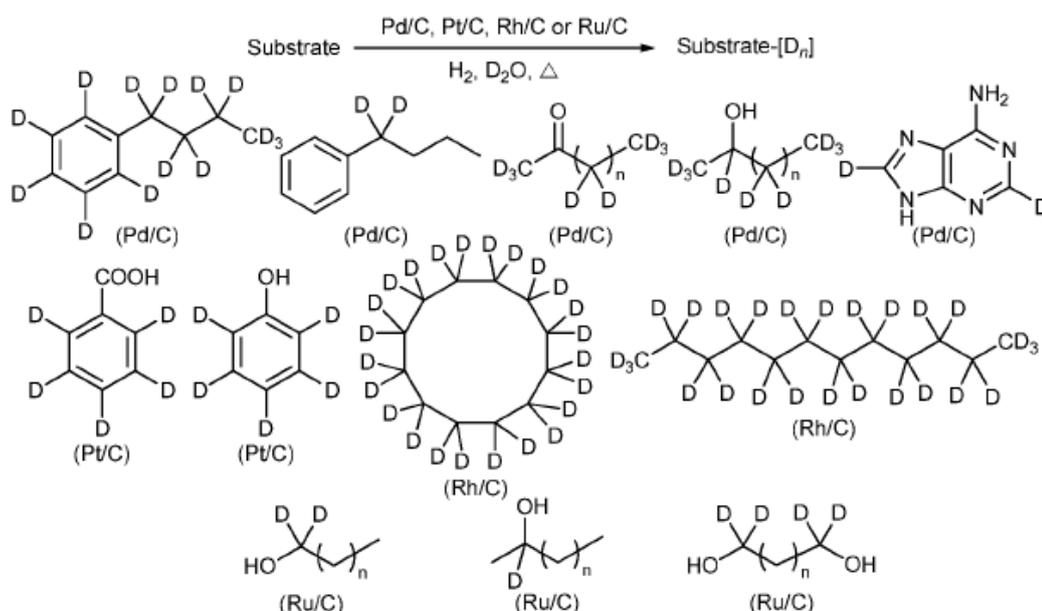


Figure 35 : Exemples illustrant la régiospécificité de plusieurs catalyseurs métalliques vis-à-vis de la deutération par réaction d'échange H/D (extraite de Fujowara et al.).

Des développements sur le sujet ont d'ailleurs été réalisés par des équipes du CEA avec des nanoparticules de ruthénium, équipées avec lesquelles nous avons collaboré au début de ce projet. Avec ce catalyseur, il a été montré que les hydrogènes portés par des carbones en alpha d'hétéroatome de fonctions amine et alcool, étaient échangés par des deutériums (Figure 36²⁴³). Notons que dans ces conditions, il a été montré que la chiralité de la molécule était conservée tout au long de la réaction²³⁷.

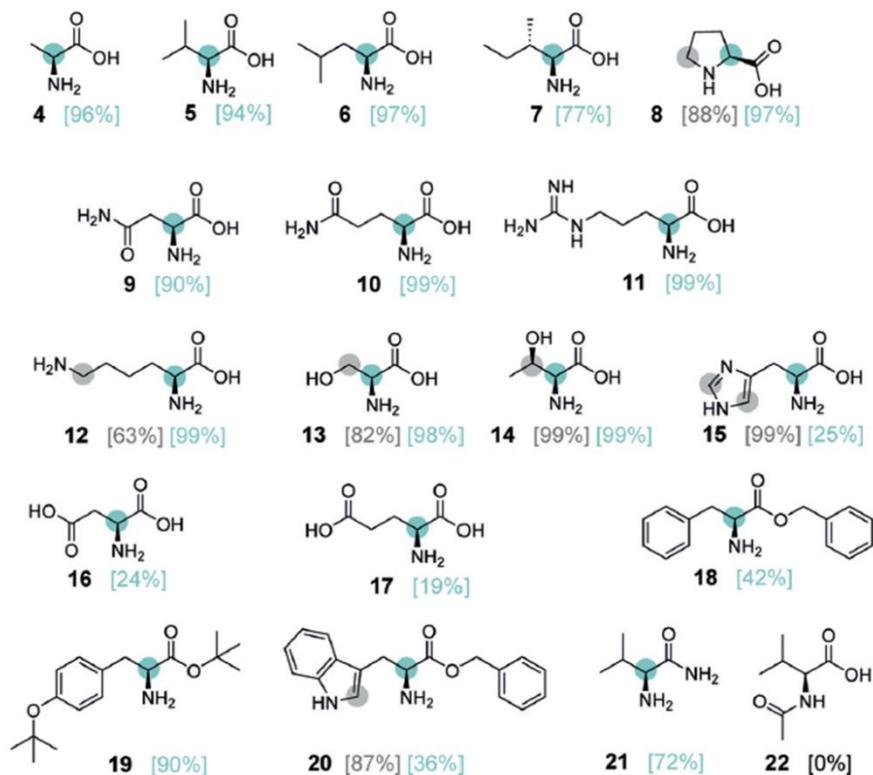


Figure 36 : Exemple de réaction d'échange H/D avec le Ruthénium comme catalyseur (extraite de Taglang et al.).

Ainsi, l'échange H/D stable constitue une approche économique et rapide lorsqu'elle peut s'appliquer directement sur la molécule cible, pour la production de standards marqués au deutérium utilisables en métabolomique quantitative. Cependant, l'utilisation de deutérium peut montrer des faiblesses pour les analyses métabolomiques quantitatives avec notamment des effets isotopiques tels que des décalages de temps de rétention, pour des composés à hauts degrés de marquage, qui ont pu être observés et qui font du deutérium un isotope de second choix pour certaines analyses²⁴⁴.

3 – Utilisation des standards marqués pour la quantification

La méthode de choix pour la quantification est l'utilisation d'isotopes marqués aux isotopes stables comme standards internes. Ces standards peuvent soit permettre de réaliser des gammes externes soit être supplémentés à l'échantillon d'intérêt pour de la dilution isotopique (Figure 37).

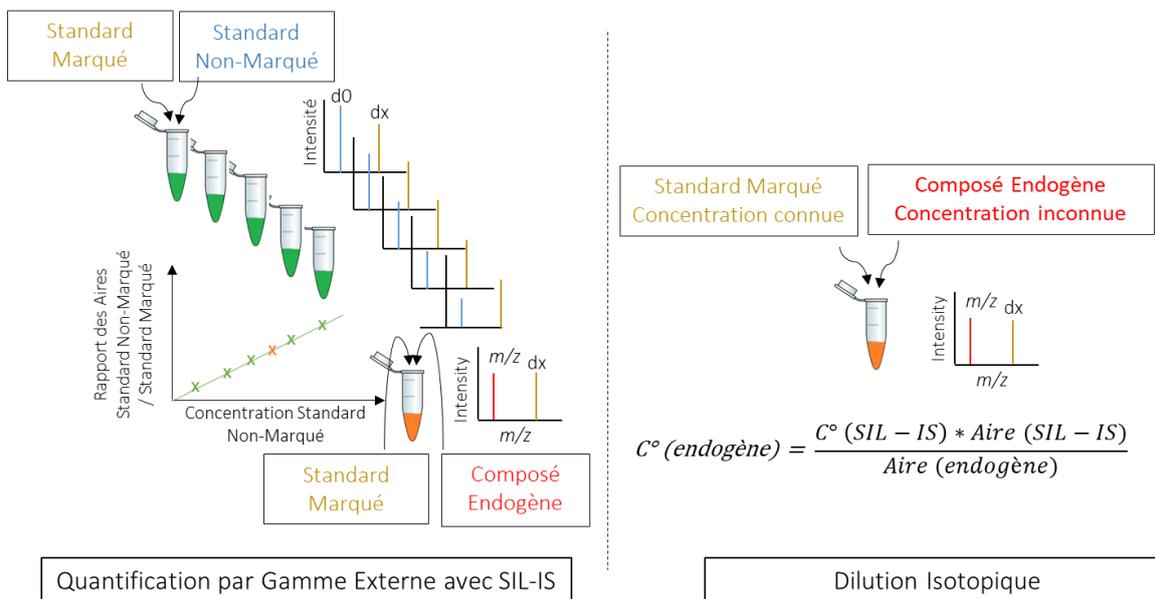


Figure 37 : Comparaison des méthodes de quantification par gamme externe avec SIL-IS et dilution isotopique.

La gamme externe consiste à analyser des concentrations croissantes connues de molécule(s) standard(s) correspondant(s) au(x) métabolite(s) d'intérêt de l'échantillon biologique. Ces gammes analysées avant et après l'échantillon permettent de détecter les potentielles dérives analytiques au cours du temps tels que les effets d'adsorption de métabolites sur la colonne chromatographique, la perte de sensibilité ou encore l'élargissement des pics chromatographiques. En l'absence de standards internes marqués, ces gammes externes doivent être réalisées dans une matrice similaire à celle de l'étude afin de limiter l'effet matrice qui peut induire des biais plus ou moins importants dans la quantification²⁴⁵. En effet, il a été prouvé que selon la matrice utilisée et le nombre d'ions arrivant dans la source d'ionisation, des effets de majoration ou plus souvent de suppression du signal sont détectés^{224,246}. Afin de s'affranchir de cet effet matrice, les gammes externes sont réalisées avec des concentrations croissantes connues de standards non marqués auxquels on supplémente leurs homologues marqués à une concentration constante et identique à celle qui sera ajoutée aux échantillons biologiques à quantifier²⁴⁷. L'établissement des gammes de calibration repose ensuite sur le calcul du rapport des réponses analytiques (aires) des standards marqués (SIL-IS) et non marqués du métabolite d'intérêt reporté en ordonnée contre la concentration du standard non-marqué dans la gamme reportée en abscisse. La gamme est dite linéaire tant que la courbe qui passe par l'ensemble des points de la gamme reste une droite. La concentration du métabolite endogène d'intérêt est alors obtenue par report sur la courbe de calibration du rapport des réponses

analytiques du SIL-IS et de l'endogène. Notons qu'il est préférable de compléter l'échantillon d'intérêt avec un standard marqué à une concentration proche de celle du métabolite endogène afin de limiter les biais²⁴⁷.

Un pré-requis pour la quantification est la linéarité de la réponse analytique. Cependant, à de hautes concentrations, des effets de saturation peuvent être détectés et les gammes ne sont alors plus linéaires²⁴⁸. Cet effet de saturation peut être induit dans la source d'ionisation. En effet, un nombre important d'ions arrive simultanément dans la source d'ionisation ce qui induit la production d'aggrégats ou de multimères²⁴⁹. Cet effet de saturation dépend également de l'analyseur et du détecteur qui équipe l'instrument utilisé pour l'analyse. A titre d'exemple, dans le cas de l'Orbitrap, selon la taille de celui-ci, le nombre d'ions pouvant y être analysés simultanément sans effet de charge varie^{169,250}. Pour ces raisons, l'administration des aliments et médicaments (FDA) stipule que : « Le modèle le plus simple qui décrit adéquatement la relation concentration-réponse devrait être utilisé. Le choix de la pondération et l'utilisation d'une équation de régression complexe doivent être justifiés. »²⁵¹. Il a d'ailleurs été prouvé que la quantification basée sur des gammes quadratiques était moins précise que celles réalisées à partir de gamme linéaire²⁵². Afin d'optimiser la précision de la quantification, il est préférable que la réponse analytique du métabolite dans l'échantillon d'intérêt soit comprise dans la gamme externe.

La quantification par dilution isotopique est aujourd'hui la méthode de référence pour la quantification absolue des métabolites^{253,254}. Elle correspond à un simple produit en croix entre les réponses analytiques de l'endogène et du SIL-IS ajouté. Afin d'obtenir une grande précision, le standard marqué préalablement quantifié doit être supplémenté dans l'échantillon d'intérêt à une concentration similaire à celle du métabolite endogène. Cette méthode de quantification est utilisée pour les analyses quantitatives en métabolomique mais est limitée par le nombre de standards marqués disponibles^{218,246}. De plus, la supplémentation de nombreux composés marqués dans l'échantillon peut créer des interférences entre les métabolites endogènes et les composés ajoutés.

Il semble intéressant de noter en conclusion de ce chapitre que les besoins en terme de quantification/normalisation par spectrométrie de masse allant croissants, les propositions commerciales mises en place pour faciliter ces analyses se multiplient également et offrent des solutions de plus en plus complètes. A ce titre, la gamme de produits marqués disponibles dans les catalogues de produits chimiques commerciaux (par exemple Sigma-Aldrich, Eurisotop, etc) est de plus en plus large. Les propositions de kits de quantification « clé en main » permettant de réaliser de la quantification rapide et large spectre continue d'émerger avec, à titre d'exemples, de nombreux kits proposés par Biocrates® dont le plus récent (MxP® Quant 500 kit) permet de quantifier jusqu'à 630 métabolites (en grande majorité des lipides) avec un QqQ ou encore le kit AbsoluteIDQ® p400 HR qui permet la quantification de plus de 400 métabolites par profilage métabolique sur un Q-Exactive. La mise à disposition d'extraits biologiques marqués se répand également pour répondre aux besoins de standardisation en métabolomique avec notamment un produit proposé par Eurisotop issu du marquage métabolique de levure au C-13²¹⁸. Néanmoins, ces solutions restent très coûteuses et assez peu flexibles en terme de domaine d'application. Ainsi, la capacité de produire à façon des librairies de composés marqués

spécifiquement adaptées aux besoins associés aux études métabolomiques du laboratoire, reste une idée enthousiasmante et constitue un aspect de cette thèse présenté dans la deuxième partie de ce manuscrit.

Objectifs

L'ensemble de ce travail de recherche bibliographique montre l'importance de l'identification des métabolites lors d'analyses métabolomiques. La qualité de cette identification est fortement liée à l'ensemble des étapes présentes lors d'une analyse métabolomique. Cela implique que l'identification des métabolites demeure encore actuellement un challenge analytique.

Le premier objectif de ma thèse a été d'améliorer le processus d'acquisition de spectres MS/MS. En effet, actuellement l'acquisition des spectres MS permettant d'obtenir le profil métabolique de l'échantillon d'intérêt et l'acquisition des spectres MS/MS nécessaire à l'identification des métabolites d'intérêts sont réalisés à l'aide de deux injections distinctes espacées de plusieurs semaines voire plusieurs mois. Ceci peut être dû à l'impossibilité d'acquérir des spectres MS/MS par le spectromètre de masse ou à des vitesses d'acquisition des analyseurs trop faibles. Entre les deux analyses, des biais analytiques peuvent se produire ce qui limite l'identification des métabolites. L'acquisition simultanée des spectres MS et MS/MS permet de s'affranchir de ces biais. Cependant, jusqu'à récemment, la vitesse d'acquisition des spectromètres de masse ne permettait pas ces acquisitions simultanées. Les développements réalisés sur les spectromètres de masse ont permis d'améliorer ces vitesses d'acquisition. Grâce à cela, mes premiers travaux de thèse avaient but d'exploiter au mieux les caractéristiques uniques de l'Orbitrap Fusion en termes de vitesse d'acquisition mais également d'acquisitions réalisées en parallèle à haute et basse résolutions. Ainsi, le développement de méthodes d'acquisition dites « Data Dependent Acquisition (DDA) » et « Data Independent Acquisition (DIA) » a été envisagé, dans un premier temps afin de multiplexer au maximum l'acquisition des spectres MS et MS/MS. Le traitement des données résultantes visant à identifier des métabolites par comparaison de spectres MS/MS expérimentaux à ceux stockés dans des bases de données n'est pas facilement réalisables de façon automatisée en raison d'un manque de logiciels efficaces (notamment pour les données DIA). De plus, selon les modalités de construction des bases de données, la comparaison des spectres acquis avec ceux de la base de données peut être plus ou moins complexe. Un autre des objectifs a donc été d'appliquer la méthode DDA développée pour des études interlaboratoires dans le cadre d'un projet européen afin d'évaluer la comparabilité de spectres MS/MS acquis sur différents spectromètres de masse. L'objectif final de ce consortium européen étant de réaliser une base de données spectrales MS/MS à l'échelle européenne. Finalement, la méthode DDA a pu être de nouveaux optimiser grâce à un nouveau logiciel, Acquire X, proposé par Thermo Fisher Scientific. Ce logiciel permet d'effectuer des analyses DDA itérative permettant de fragmenter un maximum d'ions présents dans l'échantillon. Cela permet d'avoir une connaissance approfondie de l'échantillon.

Dans un second temps, mes travaux de thèse avaient pour objectif de développer une méthode de quantification de métabolites à grande échelle à l'aide de standards internes marqués. En effet, le métabolome étant dynamique la comparaison de signaux provenant d'échantillons différents est complexe. De plus, les

Objectifs

analyses métabolomiques étant appliquées à des cohortes de plus en plus importantes, il est de plus en plus nécessaire de pouvoir comparer l'ensemble des analyses effectuées sur le grand nombre d'échantillons présents dans la cohorte d'intérêt. Pour cela, des étapes de normalisation sont nécessaires et réalisées par quantification. Les approches de quantification ciblées sont également pertinentes pour valider les candidats biomarqueurs mis en évidence par les approches non ciblées de métabolomique. La méthode de choix actuelle pour la quantification de métabolites est l'utilisation d'isotopes stables comme standards internes. Les SIL-IS n'étant pas tous disponibles commercialement, le premier objectif a été de développer des méthodes de production de standards internes marqués à grande échelle. Dans notre cas, l'objectif était de produire des standards marqués par échange H/D. Par ces méthodes, l'ensemble du métabolome humain ne peut être couvert. L'utilisation d'échantillons marqués *in vivo* permet de palier cet inconvénient. Un des objectifs de ma thèse a donc été d'analyser des échantillons d'urine de souris marqués *in vivo* au C-13. L'étude préliminaire de ces échantillons a permis d'évaluer la cinétique de marquage sur un nombre restreint de métabolites et d'investiguer l'apport de ce type d'échantillon marqué pour identifier des métabolites.

*Partie B : Etudes
expérimentales*

Chapitre 1 : Développement de nouvelles méthodes d'acquisition pour la métabolomique

A – Développements des méthodes « Data Dependent Acquisition » et « Data Independent Acquisition »

1 – Contexte et objectifs

Comme décrit précédemment, l'identification des métabolites représente un grand challenge dans les études métabolomiques. Dans ce contexte, l'acquisition d'un grand nombre de spectres de fragmentation de grande qualité est primordiale.

Classiquement, lors d'une analyse métabolomique effectuée au sein de notre laboratoire, l'échantillon d'étude est analysé deux fois. Une première fois pour réaliser son profilage métabolique (acquisition MS seulement) et une seconde fois pour confirmer l'identité par MS/MS (mode ciblé) des métabolites annotés discriminants de l'étude. Ces deux injections sont réalisées indépendamment, notamment en raison de l'impossibilité d'acquérir des spectres MS/MS ciblés avec le spectromètre de masse Exactive utilisé en routine au laboratoire pour les analyses métabolomiques. Néanmoins, quand un autre instrument type Q-Exactive est utilisé, le même mode opératoire est souvent suivi car ne sont généralement fragmentés par MS/MS (mode ciblé) que les ions d'intérêt préalablement mis en évidence par l'analyse statistique des données de profilage. Cela signifie que les données acquises en mode profil MS doivent être préalablement traitées pour que les ions à fragmenter soient identifiés. Le temps nécessaire au traitement des données entre ces deux analyses est relativement long (plusieurs semaines). Dans le cas plus général des analyseurs de type Orbitrap, les vitesses d'acquisition sont souvent trop faibles pour permettre l'acquisition des spectres MS et MS/MS (de façon suffisamment exhaustive pour être réalisée sans a priori) en une seule injection, en conservant suffisamment de points par pics lors de l'analyse MS pour réaliser une bonne quantification. Cette double injection induit des inconvénients majeurs. Tout d'abord, le temps nécessaire au traitement des données entre ces deux analyses peut être très long (plusieurs semaines). L'échantillon n'est plus injecté dans les mêmes conditions analytiques lors du profilage et de l'identification ce qui induit des biais analytiques comme le décalage de temps de rétention.

Selon la robustesse du système analytique utilisé, cet impact peut être plus ou moins important. Ensuite, la nécessité de faire deux injections requiert une quantité initiale d'échantillon non négligeable ce qui n'est pas toujours compatible avec les échantillons (typiquement pour le liquide céphalorachidien). Pour cela, les analyses MS/MS sont généralement réalisées sur les QCs placés en début ou fin de séquence d'analyse^{106,255}. Cette double injection est également très consommatrice de temps d'analyse.

Pour remédier à ces inconvénients, de nouveaux modes d'acquisition, « Data Dependent Acquisition » et « Data Independent Acquisition », ont été investigués dans le cadre de cette thèse. Très utilisés en protéomique, ils présentent encore quelques limitations pour la métabolomique telle que l'obtention d'une énergie de fragmentation optimale pour l'ensemble des métabolites¹⁹¹. Ce point sera discuté plus en détails ci-dessous. Ces modes d'acquisition permettent d'obtenir les spectres MS et MS/MS en une seule injection.

L'objectif de mes travaux était donc de développer et de comparer des méthodes d'acquisition DDA et DIA de type « SWATH » pour les analyses métabolomiques par LC-HRMS sur l'Orbitrap Fusion, en exploitant sa grande versatilité d'acquisition (fragmentation HCD et CID, détection basse et/ou haute résolution). Cette étude a pour but ultime de faciliter l'identification des métabolites tout en gagnant en temps d'analyse.

2 – Méthodologie suivie et introduction au travail expérimental

Un aperçu de la méthodologie est donné dans cette partie, un chapitre « matériel et méthodes » plus détaillé est donné dans la publication insérée juste après.

Pour répondre aux objectifs fixés, les développements ont été réalisés à partir d'un mélange de 47 standards et sur des échantillons de plasma de référence (SRM 1950) commercialisés par l'Institut National des Normes et Technologies (NIST). Dans un premier temps, les données ont été acquises avec l'Orbitrap Fusion couplé avec la chromatographie liquide en phase inverse (C18) et avec une ionisation en mode positif. Ces conditions sont les plus stringentes pour le développement des méthodes DDA et DIA dû à la faible largeur des pics chromatographiques et de la vitesse d'acquisition de l'Orbitrap.

La versatilité de l'Orbitrap Fusion avec ces trois analyseurs et ces deux modes de fragmentation offrent de nombreuses possibilités pour les développements méthodologiques. Les premiers développements se sont intéressés au mode d'acquisition DDA. Avec cette méthode, les ions sont automatiquement sélectionnés pour être fragmentés. Les développements avaient donc pour but d'optimiser la sélection de ces ions. Une première méthode nommée « HCD-only » a été développée afin d'optimiser l'ensemble des paramètres de sélection des ions.

Dans une analyse métabolomique, les signaux détectés peuvent être classés en deux catégories : les métabolites présents dans une base de données interne et les métabolites absents de cette base de données voire inconnus. Une deuxième méthode a donc été développée pour prendre en compte cette différence. Dans cette méthode, les ions correspondants peuvent être sélectionnés selon un arbre décisionnel comportant des listes d'inclusion et d'exclusion et des processus de fragmentation différents (HCD ou CID) selon la connaissance préalable ou non du métabolite (voir publication). Les méthodes DDA ne permettant pas de fragmenter et donc d'identifier l'ensemble des métabolites, les approches DIA de type « SWATH » ont ensuite été étudiées. Contrairement aux méthodes DDA, tous les ions détectés sont fragmentés. Les développements se sont alors principalement focalisés sur la taille des fenêtres de masse afin d'optimiser la sensibilité de la méthode.

Partie B : Etudes Expérimentales – Chapitre 1

Ensuite ces méthodes ont été comparées *via* le nombre de métabolites détectés et fragmentés ainsi que par la pureté et la qualité des spectres MS/MS obtenus. Finalement, ces méthodes multi-événements ont ensuite été comparées avec la méthode « Full Scan » classiquement utilisée pour les analyses métabolomiques, afin d'en évaluer l'aspect quantitatif. Pour cela, les échantillons de plasma NIST ont été supplémentés avec le mélange de standards à différentes concentrations.

3 – Publication « Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics »

Article

Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics

Pierre Barbier Saint Hilaire ^{1,†}, Kathleen Rousseau ^{1,†}, Alexandre Seyer ², Sylvain Dechaumet ² ,
Annelaure Damont ¹ , Christophe Junot ¹ and François Fenaille ^{1,*} 

¹ Département Médicaments et Technologies pour la Santé (DMTS), CEA, INRAE, Université Paris-Saclay, MetaboHUB, F-91191 Gif sur Yvette, France; pierre.barbiersainthilaire@gmail.com (P.B.S.H.); kathleen.rousseau@cea.fr (K.R.); annelaure.damont@cea.fr (A.D.); christophe.junot@cea.fr (C.J.)

² MedDay Pharmaceuticals SA, 24 Rue de la Pépinière, F-75008 Paris, France; alexandre.seyer@medday-pharma.com (A.S.); sylvain.dechaumet@medday-pharma.com (S.D.)

* Correspondence: francois.fenaille@cea.fr

† These authors contributed equally to this work.

Received: 11 March 2020; Accepted: 16 April 2020; Published: 18 April 2020



Abstract: Constant improvements to the Orbitrap mass analyzer, such as acquisition speed, resolution, dynamic range and sensitivity have strengthened its value for the large-scale identification and quantification of metabolites in complex biological matrices. Here, we report the development and optimization of Data Dependent Acquisition (DDA) and Sequential Window Acquisition of all Theoretical fragment ions (SWATH-type) Data Independent Acquisition (DIA) workflows on a high-field Orbitrap Fusion™ Tribrid™ instrument for the robust identification and quantification of metabolites in human plasma. By using a set of 47 exogenous and 72 endogenous molecules, we compared the efficiency and complementarity of both approaches. We exploited the versatility of this mass spectrometer to collect meaningful MS/MS spectra at both high- and low-mass resolution and various low-energy collision-induced dissociation conditions under optimized DDA conditions. We also observed that complex and composite DIA-MS/MS spectra can be efficiently exploited to identify metabolites in plasma thanks to a reference tandem spectral library made from authentic standards while also providing a valuable data resource for further identification of unknown metabolites. Finally, we found that adding multi-event MS/MS acquisition did not degrade the ability to use survey MS scans from DDA and DIA workflows for the reliable absolute quantification of metabolites down to 0.05 ng/mL in human plasma.

Keywords: metabolomics; high resolution mass spectrometry; Orbitrap Fusion; DDA; DIA

1. Introduction

Metabolomics aims at studying the whole set of metabolites found in a given biological system. Its objectives are therefore to identify and quantify all the small molecules present in complex matrices as comprehensively as possible. Nowadays, liquid chromatography coupled to high-resolution mass spectrometry (LC/HRMS) represents the most widely used approach for untargeted metabolomics [1]. LC/HRMS-based metabolomics workflows usually start with the acquisition of high-resolution full scan accurate mass spectra in order to obtain comprehensive metabolic profiles thanks to the detection of a few thousand signals in a semi-quantitative fashion [2]. The recorded signals then need to be annotated and corresponding metabolites identified to translate raw analytical data into meaningful biological information. However, a large portion of the detected signals remain structurally uncharacterized

and thus, metabolite annotation and identification still represent major bottlenecks in untargeted metabolomics [3,4].

The successful identification of metabolites requires a combination of information lines exploited in conjunction, including accurate LC retention time, high mass measurement accuracy (<1 ppm), isotope pattern and isotope fine structure measurement as well as high-quality fragmentation spectra [5,6]. High mass resolution (>100,000) allows for the measurement of isotope pattern and isotope fine structure (through the distinction of isobaric isotopes), which is particularly useful for confident compound annotation [6]. One of the most valuable elements to confirm metabolite annotation or reduce the list of possible annotations is the acquisition of fragmentation spectra and their comparison to reference MS/MS spectra included in spectral libraries [7].

Experimental approaches to acquire MS/MS data are generally referred to as Data Dependent or Data Independent Acquisition (DDA and DIA, respectively). In the most widely used DDA acquisition workflow, MS/MS spectra are automatically collected for precursor ions whose MS intensity exceeds a pre-defined threshold. In this mode, precursor ions are selected using a small isolation window (typically ≤ 1 Da wide), which leads to high-quality and higher purity MS/MS spectra. Nevertheless, the selection of precursor ions is a (semi)stochastic event suffering from low analytical reproducibility and favoring selection of the most abundant but sometimes biologically irrelevant ions. As a consequence, some biologically relevant precursor ions may not be selected for fragmentation if they do not match the defined selection criteria or if they are coeluted with more intense species. These issues can be partially overcome by the use and careful optimization of additional MS-to-MS/MS triggering criteria (e.g., isotope pattern or mass defect) [8–10], by optimizing the use of exclusion or inclusion lists prior to further MS/MS acquisitions [11–13], or by implementing efficient software tools for assessing precursor ion purity of acquired data [14].

Until very recently, the implementation of acquisition protocols combining MS and MS/MS acquisitions in one single run, especially those related to DIA, were rather limited to hybrid quadrupole-time-of-flight (Q-TOF) instruments due to their enhanced duty cycles [15,16]. Because of their lower acquisition speed, simultaneous MS and MS/MS data collection can rarely be performed at the highest mass resolution achievable on Orbitrap-based instruments, in order to keep enough data points per chromatographic peak, which is mandatory for further reliable peak integration and corresponding metabolite quantification from MS data [17]. It is therefore a rather common procedure in metabolomics, when using an Orbitrap mass analyzer, to acquire MS/MS data by DDA or targeted-DDA on pooled quality control (QC) samples at the beginning or at the end of an analytical batch [17,18]. Collected data can then be used to support metabolite annotation in biological samples, assuming that the ions fragmented in the pooled QC samples are identical to those present in each of the biological samples analyzed. While separating metabolite quantification from identity confirmation, this approach also requires additional acquisition time and sample volumes.

A DIA acquisition method would represent a viable alternative or complement to DDA to obtain valuable MS/MS data on all ions from all samples for further metabolite structure confirmation and quantification [19]. SWATH (accounting for Sequential Window Acquisition of all THEoretical fragment-ion spectra) [20], which is the most used DIA acquisition workflow, involves the selection of wide and consecutive m/z isolation windows (typically between 10 and 50 Da) to cover the whole mass range. All ions within these windows are then fragmented simultaneously. Due to the width of isolation windows, the direct link between a specific precursor ion and its corresponding fragments is often lost, which renders MS/MS data interpretation of unknown metabolites much more difficult than that of MS/MS data resulting from DDA protocols. All-Ion Fragmentation (AIF) is another form of DIA acquisition that has been successfully applied to metabolomics or lipidomic approaches [21–23]. The present paper dealing only with SWATH-like DIA as applied to low-mass metabolites (essentially below 700 Da), DIA accounts hereafter for SWATH-like approaches. Although some software solutions for handling MS/MS data obtained from DIA as well as DDA workflows have been published for

metabolomics such as MS-DIAL [24] or MetaboDIA [25], no comparative study has been carried out to objectively assess the advantages and limitations of these different tools.

While DIA strategies such as SWATH are often used in proteomics, very limited work has been reported on small molecules [26]. This particularly holds true when considering the use of the Orbitrap mass analyser, with very sporadic examples published [27,28]. In this context, the aim of this study was to evaluate the performance of DDA and SWATH-type DIA acquisition workflows when implemented on an Orbitrap FusionTM TribridTM mass spectrometer. We especially exploited the versatility of this mass spectrometer, comprising of quadrupole, linear ion trap and Orbitrap mass analyzers, in order to conciliate multiple modes of analysis and experimental flexibility. A set of 47 well defined standard compounds was first used for the development and preliminary evaluation of DDA and DIA strategies. Optimized methods were then used to study plasma metabolome using another set of 72 endogenous metabolites. Although reversed-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) provide complementary information and thus represent a relevant combination for expanding metabolome coverage [29,30], we decided to focus this study on RPLC, which is the most demanding in terms of MS-MS/MS acquisition speed as a result of the narrower chromatographic peaks, using routine workflows as used in our laboratory [29,31]. This report demonstrates the value and complementarity of DDA and SWATH-type DIA approaches for metabolomics, as implemented on an Orbitrap FusionTM.

2. Materials and Methods

2.1. Chemicals and Reagents

Chemical standards were from Sigma-Aldrich (Saint Quentin Fallavier, France). Liquid chromatography solvents, i.e., acetonitrile (ACN) and water, and extraction solvent (i.e., methanol) were LC-MS grade solvents (LC-MS Chromasolv, Sigma-Aldrich, Saint Quentin Fallavier, France). Mobile phase additive (i.e., formic acid) was from Sigma Aldrich. NIST SRM-1950 reference plasma was ordered from the National Institute of Standards and Technology (NIST, Gaithersburg, MD, USA).

2.2. Preparation of Standard Samples and Plasma Extracts

A pool of 47 chemical compounds (20 µg/mL for each compound) was prepared from solutions at 1 mg/mL (Supplementary Material, Table S1). The pool was then diluted in water:ACN (95:5) with 0.1% formic acid to reach 100, 10, 3, 1, 0.5, 0.25, 0.1, and 0.05 ng/mL.

NIST SRM-1950 reference human plasma was processed using a metabolomics-designed extraction adapted from a previously published one [29]. Briefly, 100 µL of plasma were diluted with 400 µL of methanol, sonicated for 15 min and centrifuged at 15,000× *g* for 15 min. After a 1h30 slow precipitation on ice, samples were centrifuged at 20,000× *g* for 15 min and supernatant was withdrawn and dried under nitrogen. Plasma metabolites were resuspended in water/ACN (95:5) with 0.1% formic acid. Some plasma samples were spiked post-extraction with the pool of 47 compounds at a final concentration ranging from 0.05 to 100 ng/mL for each metabolite.

2.3. LC/HRMS Analysis

Chromatographic separation was performed on a Dionex Ultimate 3000 (ThermoFisher Scientific, Courtaboeuf, France) with a Hypersil Gold C₁₈ column (150 × 2.1 mm, 1.9 µm; ThermoFisher Scientific) under gradient conditions, using mobile phases A (0.1% formic acid in water, *v/v*) and B (0.1% formic acid in ACN, *v/v*). Flow rate was set at 500 µL/min. Mobile phase A was held at 95% for 2 min followed by an increase to 100% of mobile phase 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 µL.

Mass spectrometry analyses were performed on an Orbitrap FusionTM instrument (ThermoFisher Scientific, Courtaboeuf, France) fitted with a HESI source. Source parameters were set as follows: spray voltage at + 3500 V, sheath gas at 50 (arbitrary units), auxiliary gas at 12 (arbitrary units),

sweep gas at 3 (arbitrary units), ion transfer tube temperature at 325 °C, and vaporizer temperature at 350 °C. Ion transfer parameters were set as follows: mass range at “normal”, S lens RF level at 60%. Scan parameters applied were as follows: acquisition time from 0 to 20 min, positive ionization mode, micro scan at “1”, data type at “profile”, AGC target at 5×10^4 . For acquisition in MS mode only, the data were acquired using the Orbitrap analyzer, at a resolution of 240,000 at m/z 200 (FWHM, a spectrum collected at this resolution requires ~0.5 s); maximum injection time was set at 400 ms, and the scan range from m/z 85 to m/z 1000. More detailed information regarding experimental conditions is given in the Supporting Information (Tables S2–S4).

2.4. Data Processing and Evaluation of the Quality of MS/MS Spectra

Raw instrument data (.RAW) were processed using Xcalibur v.2.2.44 (ThermoFisher Scientific) processing methods. The 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 $[M+H]^+$ m/z , fragment ion m/z and expected retention time (min) are shown in supplementary data (Supplementary Material, Table S1).

MS-DIAL software [24] (version 4.12) which is freely available (<http://prime.psc.riken.jp/compms/msdial/main.html>), was used to calculate MS/MS similarity between the evaluated and reference spectra. We merged all the reference MS/MS spectra from our in-house spectral library [29] into a single .msp file for further mass spectral matching, using the dot product (DP) scoring method included in the MS-DIAL software.

For both DDA and DIA data analyses, MS-DIAL parameters were set as follows: MS1, tolerance, 0.01 Da; MS2 tolerance, 0.025 Da (0.5 Da for ion trap detection); retention time begin, 0 min; retention time end, 100 min; minimum peak height, 1000; mass slice width, 0.05 Da; smoothing level, 3 scans; minimum peak width, 5 scans; sigma window value, 0.5.

3. Results and Discussion

Our first objective was to develop LC-MS/MS metabolomics workflows on an Orbitrap Fusion™ Tribrid™ instrument thanks to concomitant acquisitions of a MS profile for metabolite detection and quantification, and MS/MS spectra for metabolite identification. We especially aimed at exploiting the versatility of this mass spectrometer, comprising quadrupole, linear ion trap and Orbitrap mass analyzers, in order to conciliate multiple modes of analysis and experimental flexibility. In this context, a set of 47 standard molecules was first used to develop and optimize DDA and DIA (SWATH) acquisition methods, involving either or both non-resonant higher energy collision-induced dissociation (HCD), or ion trap resonant collision-induced dissociation (CID) conditions.

The 47 chemicals comprising this test set were those already used in a previous study [6] and well ionized in ESI+ mode while exhibiting large chemical, mass and retention time diversity when analyzed on a C₁₈ column (Supplementary Material, Table S1). As we observed in a previous study [29], such a C₁₈ column is not the most suited one to separate metabolite isomers, but is the most demanding in terms of MS-MS/MS acquisition speed (due to the narrow chromatographic peaks). Therefore, our test set was only made of non-isomeric chemicals, since isomer distinction was not at the center of the present study. Performances of the developed methods were then evaluated and compared using NIST SRM-1950 human plasma as a representative biofluid and considering another set of 72 endogenous metabolites annotated in this matrix (Supplementary Material, Table S5) [29].

These developments have been voluntarily carried out on a limited but well characterized set of exogenous and endogenous molecules, to allow further manual data analysis so as to avoid any potential bias induced by processing software tools. A significant part of the data treatment was manually performed while the evaluation of the quality of MS/MS data collected was achieved using the MS-DIAL software.

3.1. Implementation of a First “HCD-only” DDA Acquisition Protocol on Authentic Standards as a First Step toward the Collection of Meaningful MS/MS Data

The set of 47 authentic standard molecules was first used as a well known and calibrated mixture in order to optimize experimental conditions (e.g., precursor isolation filters and collision energy applied) and preliminarily evaluate the relevance of a DDA method. Prior to MS detection, metabolite separation was achieved on a C₁₈ UHPLC column. Under these conditions, chromatographic peaks are typically less than 10 s wide. In order not to downgrade the quality of MS1 quantification, sufficient data points (at least 8–10) should be obtained per chromatographic peak. Resolutions of the MS1 survey and MS/MS scans were thus kept at 240,000 and 30,000 (both at m/z 200, FWHM), respectively. In order to resolve co-selected or co-occurring isobaric precursor or fragment ions, which is a common phenomenon in metabolomics [14], and facilitate spectra identification, Orbitrap resolution was set at the highest value still compatible with the duty cycle. Acquisition of MS data was performed every 10 MS/MS spectra (i.e., a so-called top-10 method) to maintain a total cycle time below 1 s. For comparison, a previously reported and carefully optimized top-5 protocol implemented on an Orbitrap Velos instrument involved mass resolutions set at 30,000 and 7500 in the MS and MS/MS modes, respectively [17]. This highlights the substantial improvement in scan speed (or resolution) of the Orbitrap FusionTM compared to previous generations of instrument, essentially due to the presence of a high-field Orbitrap mass analyzer.

Efficient DDA methods should enable the acquisition of high-quality and higher purity MS/MS spectra from metabolites of interest present in biological media (see the Supporting Information for information about method development). In this DDA method, non-resonant collision-induced dissociation mode (HCD) was chosen as unique mode of fragmentation. In sharp contrast with proteomics where a single normalized collision energy (NCE) of ~30% is suitable for almost all peptides [32,33], metabolite fragmentation is strongly NCE-dependent and thus no single collision energy can be applied to fit all metabolites [17]. Optimizing applied NCE therefore represents a prerequisite to efficient DDA acquisition protocol in metabolomics. Here, optimal NCE was empirically defined as yielding an MS/MS spectrum with precursor ion at a relative intensity of 15–45% of the base peak. These acquisition conditions often proved to yield the most informative MS/MS spectra, while maintaining the precursor ion detectable. This was also a mean to comparatively evaluate the NCE required to reach similar fragmentation level of given precursor ions. The 47 studied compounds have optimal NCEs varying from 10% to 100% (27% on average, Supplementary Material, Table S1), while ~500 additional compounds from our chemical database and detectable in the positive ion mode overall show optimal NCEs in the 10–50% range (~40% on average, Supplementary Material, Figure S4) [29]. Employing the stepped collision energy option represented a viable way to obtain information-rich MS/MS spectra by combining three different NCEs in a single fragmentation spectrum. A good compromise seems to set stepped NCE at $30 \pm 20\%$ (i.e., 10%, 30%, and 50%) to cover most of the compounds presented in Table S1 (Supplementary Material). During such experiments with stepped NCE, the precursor ion is fragmented at 10%, 30%, and 50% NCE, and the fragments from these three energies are combined and detected simultaneously. Under these conditions a single spectrum is obtained. As a representative example, fragmentation of alanyl-tyrosine is shown in Figure 1A with NCEs of 10%, 30%, and 50%, and with a stepped NCE combining those 3 values. Figure 1B shows the whole arrangement of filters, fragmentation and acquisition conditions.

Before working on the Orbitrap FusionTM, HCD spectra were routinely acquired in our laboratory on a Q-Exactive instrument starting from m/z 50 whatever the mass of the precursor ion, so as to highlight potential low-mass diagnostic fragment ions. Surprisingly, when translated to the Orbitrap FusionTM, such experimental conditions led to poor-quality HCD spectra for precursor ions with m/z values above ~250 (Figures S1–S3). This observation prompted us to investigate this aspect further by studying the effect of the starting mass on the HCD spectra. Figures S1–S3 display data obtained for 3 representative metabolites, and illustrate the poor detectability of fragment ions in the higher mass range (i.e., exceeding more than ~5 times the lower mass of the selected fragment ion window),

due to probable inefficient transmission of the whole ion range. To avoid any detrimental effects when acquiring MS/MS spectra with DDA or DIA workflows, the lower scan limit has been fixed at m/z 85.

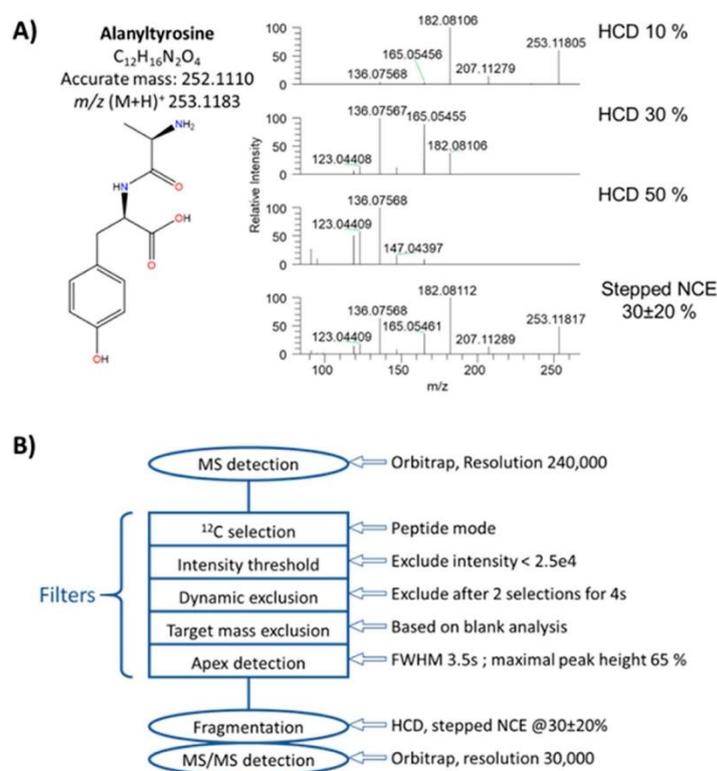


Figure 1. Implementation of an “HCD-only” DDA acquisition workflow. (A) MS/MS spectra of alanyltyrosine obtained upon HCD activation at 10%, 30% and 50%, and stepped HCD at 30% ± 20%. (B) Schematic overview of the “HCD-only” DDA method (More detailed information regarding experimental conditions and filters is given in Supplementary Material, Table S2).

The pool of 47 molecules was preliminarily analyzed at 100 ng/mL in triplicate by the implemented DDA method with a stepped NCE set at 30% ± 20%. At this rather high concentration, 43/47 compounds were detected in the MS1 survey scan. Limits of detection of 3'-dephosphocoenzyme A, 6-hydroxymelatonin, p-coumaric acid, and dodecanedioic acid were much higher than 100 ng/mL under the conditions used here, thus precluding their efficient detection (Supplementary Material, Table S1). All 43 compounds were efficiently fragmented in each of the three replicates, except ochratoxin A for which no MS/MS spectra were obtained in one replicate. More than 88% of these acquisitions returned a minimum of two MS/MS spectra per replicate for each compound, which highlights data consistency across replicate experiments. Since precursor ion selection occurs in a stochastic fashion within DDA workflows, such a good reproducibility is particularly remarkable. DP scoring was used to evaluate spectral similarity under these DDA conditions. DP scores ranged from 296 to 999 (mean value: 692; median value: 699), with 91% of metabolites having DP scores above 500 (Supplementary Material, Table S7). To better assess and visualize how spectral quality translates into DP score, Figure 2 presents eight MS/MS spectra with variable DP scores. The lowest DP score of 233 was obtained for a rather noisy spectrum of capryloylglycine, but this was the only score below 480. Then comes 4-pyridylacetic acid with a DP of 482 but with a good match between evaluated and reference spectra. Thus, despite fluctuations in the relative intensity of some fragment ions can be observed, a DP score > ~500 seems sufficient to retain high confidence in metabolite identification (Figure 2). Our results are in good agreement with those of Blazenovic et al. who found that a dot product score threshold of 400 gave the best results on a training dataset of more than 300 MS/MS spectra, within the frame of the Critical Assessment of Small Molecule Identification

(CASMI) contest [34]. Each 10-fold dilution of test mix concentration (from 100 to 10 and then 1 ng/mL) resulted as expected in a 15–20% drop of the MS/MS triggering efficiency. The global analysis of the recorded data demonstrates the overall efficiency of this primary DDA workflow that sets a solid basis for further methodological developments.

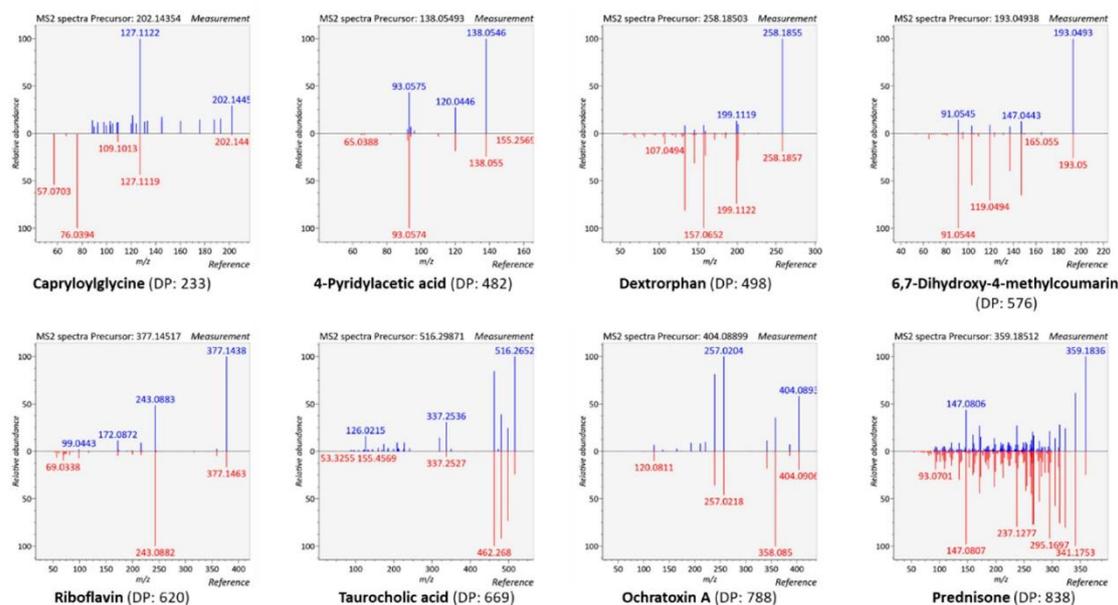


Figure 2. Head-to-tail comparison of evaluated versus reference MS/MS spectra. Evaluated MS/MS spectra were obtained using the HCD-only DDA acquisition workflow.

3.2. Collecting Meaningful MS/MS Data for Knowns and Unknowns using a DDA Workflow with Parallelized HCD and CID Fragmentations

In untargeted metabolomics, two main categories of compounds can be considered. The first group consists in known metabolites for which a standard compound is available. These metabolites are primarily identified based on their accurate mass and retention time, and then by comparison of their MS/MS spectrum with that of an authentic standard included in an in-house generated reference spectral library. The second group includes “unknown” metabolites that do not find any equivalent in the reference spectral libraries. Those two groups of metabolites can be handled differently when acquiring MS/MS data. In the first case, MS/MS acquisition is “only” to confirm the identity of an annotated metabolite by matching experimental and reference MS/MS spectra acquired under similar fragmentation conditions. Acquisition of MS/MS data at high resolution/high mass accuracy is not compulsory in this particular case. In contrast, for the second category of metabolites, the most complete MS/MS data and accurately measured masses are required to validate a putative annotation or to support a structure proposal based on observed fragment ions.

Thus, we designed another DDA method whose decision tree is displayed in Figure 3. This second DDA method exploits the versatile capabilities of the Orbitrap Fusion™: (i) metabolite confirmation by parallelizing acquisition of MS1 scans at high resolution in the Orbitrap and low resolution HCD-MS/MS data at a faster rate in the ion trap (red part of Figure 3), and (ii) for characterization of unknowns by acquiring complementary HCD and CID spectra at high resolution (green part of Figure 3).

In this method, we made use of the filters already optimized in the first DDA workflow described above (blue boxes in Figure 3). For metabolites present in our internal database, i.e., for which the corresponding standards are available (red part of the decision tree, Figure 3), the main goal was to obtain high quality MS/MS spectra as close as possible to those present in the spectral database, i.e., with fragments and precursor ion intensities around 20%. To better cover the optimal NCEs of the ~500 chemical compounds included in our chemical database and detectable in the positive ion mode (Supplementary Material, Figure S4) [29], two distinct stepped NCEs of 15% ± 10% and 45% ± 20%

were applied for improved reliability of the annotated metabolites identification. For metabolites absent of our chemical library (green part of the decision tree, Figure 3), each corresponding precursor ion was fragmented both under HCD and CID conditions to obtain complementary structural information [17,35], and using compromised stepped NCEs of $30\% \pm 20\%$ for HCD and 22% for CID. Of note, the multiplexed nature of this multi-event workflow prompted us to lower resolution of MS1 and MS2 events at 120,000 and 15,000, respectively.

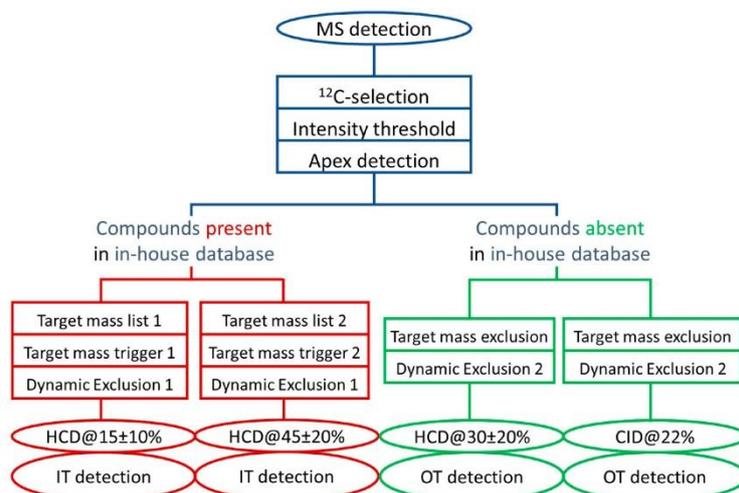


Figure 3. Schematic overview of the DDA method to collect meaningful MS/MS data for knowns and unknowns. See Table S3 for detailed information about the acquisition conditions. IT: Ion Trap, OT: Orbitrap.

We next evaluated the efficiency of this innovative dual DDA method by performing another series of experiments using a methanolic extract obtained from the NIST SRM-1950 plasma spiked with 43 of the above described test mixture of 47 compounds at 10 ng/mL and 100 ng/mL. The NIST plasma is commonly used in our laboratory as a reference sample for method development and data normalization. Up to 72 distinct endogenous metabolites can be routinely detected using our experimental conditions on a C_{18} column and MS detection in the positive ESI mode (Supplementary Material, Table S5). These compounds would be annotated at levels 1 or 3 (in case of isomers, which cannot be discriminated by their MS/MS spectra) according to the guidelines proposed by the Metabolomics Standards Initiative working group [36].

For further method evaluation, a working set of 115 molecules was first designed from those 72 endogenous metabolites plus the 43 authentic standards that have been previously used for method development and that demonstrated limits of detection below 100 ng/mL. Then, this working set was randomly split into two subsets of 69 and 46 molecules arbitrarily considered as “present” (i.e., further subjected to HCD MS/MS analyses at low resolution in the ion trap, as displayed in the red part of the decision tree of Figure 3, and referred to as “S1”) or “absent” (i.e., subjected to HCD and CID MS/MS analyses at high resolution in the Orbitrap, as displayed in the green part of the decision tree of Figure 3, and referred to as “S2”) of our chemical library, respectively. The 69 metabolites belonging to the S1 subset have been distributed into two target mass lists depending on their optimal NCE (Figure 3), and were also used to generate an exclusion list for the unknown screening approach.

Despite 95% of the 115 molecules could be detected at either spiking level, selection of precursor ions was on average ~65% effective, while culminating at more than 80% at the highest spiked concentration for both subsets of compounds (Figure 4A). MS/MS triggering rate also proved similar for compounds subjected to MS/MS experiments at low or high mass resolution. Between 80% and 90% of the molecules spiked at 100 ng/mL were subjected to MS/MS whatever the acquisition workflow. More surprisingly was the fact that only ~65% of the 72 endogenous metabolites triggered MS/MS, whether

acquisition was performed under high or low-resolution conditions. We observed that the median retention times of selected and unselected precursor ions were rather similar (1.12 min and 0.92 min, respectively), while their median intensities were significantly different (1.09×10^6 vs. 1.82×10^5 , respectively). Therefore, it seems that the less abundant of these endogenous compounds eluting close to the column dead volume were disadvantaged and thus untriggered due to limited acquisition rate of the mass spectrometer especially marked in this retention time region where many compounds coelute, thus leading to dramatic ion suppression effects and competition toward selection for MS/MS.

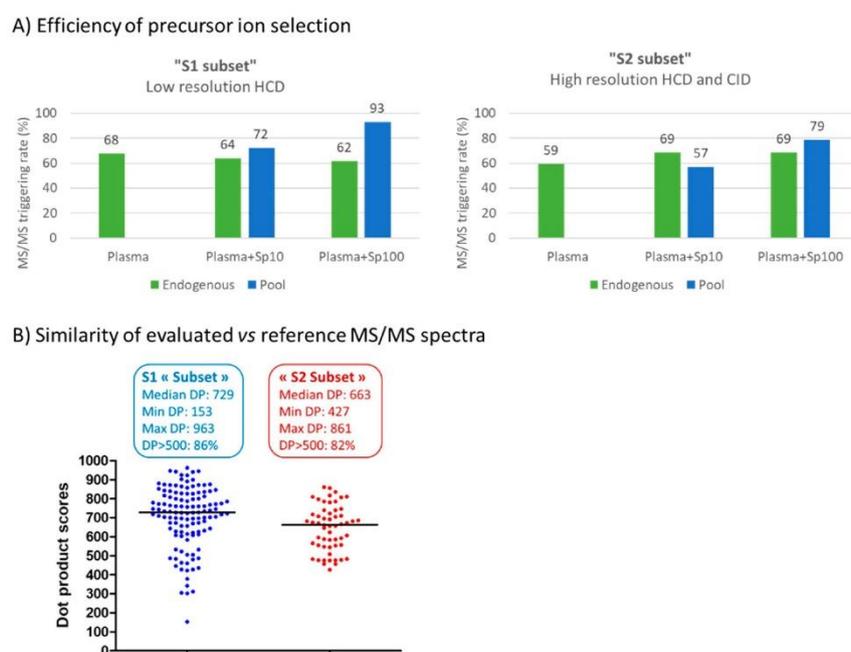


Figure 4. Performance characteristics of the low- and high-mass resolution DDA acquisition workflows on S1 and S2 subsets, respectively. (A) Efficiency of precursor ion selection and (B) distribution of dot product scores for S1 and S2 subsets. Plasma + Sp10 or Plasma + Sp100 correspond to plasma extracts spiked at 10 and 100 ng/mL, respectively.

Overall, MS/MS spectra obtained at low and high resolution were of high quality and proved comparable to the reference ones, with more than 80% of spectra characterized by a DP score > 500 (Figure 4B). Even though a single non-optimal NCE was used to fragment unknowns (i.e., $30\% \pm 20\%$) while maintaining a suitable cycle time, relevant high-resolution mass spectra were still recorded. More specifically regarding the endogenous metabolites monitored in the NIST plasma and belonging to the S2 group, valuable MS/MS spectra were obtained for 15/19 plasma metabolites, i.e., 78% of precursor ions triggering MS/MS acquisition. Manual inspection of the data revealed that this method provided both relevant HCD and CID MS/MS spectra for 10 metabolites, while HCD or CID spectra were retrieved for three and two compounds, respectively.

Altogether, these results emphasize the ability of the Orbitrap FusionTM to perform high quality DDA for metabolomics. Further software improvements, as those implemented in the AcquireX data dependent software suite (ThermoFisher Scientific), would be useful to improve the presented workflow, for instance regarding automatic exclusion of background signals or for implementing iterative DDA acquisitions in which fragmented metabolites are automatically included in an exclusion list to allow further acquisition of MS/MS data of minor metabolites.

As shown in this paper, applying different MS/MS acquisition parameters based on predefined exclusion or inclusion lists for proper selection of precursor ions can improve the number of metabolites with high-quality MS/MS spectra. Even exquisite optimization of such gas-phase fractionation procedure on fast-acquiring Q-TOFs, for instance with the joint use of several methods with different

mass ranges for selection of precursor ions [37], would not return useful MS/MS spectra for all the metabolites especially those presenting low intensity signals. Thus, we next investigated whether the implementation of DIA workflows might encompass those intrinsic limitations of DDA.

3.3. Development of a DIA Acquisition Workflow

For comparison purposes with the DDA method, we also implemented a DIA workflow for untargeted metabolomics applications. Such an approach has the benefit that any detected ion will trigger fragmentation (even those below the selection threshold for DDA) and thus its MS/MS mass spectrum will be collected. The method was built on a full scan MS event followed by a SWATH-like fragmentation, typically using consecutive precursor ion isolation windows with variable sizes (often 20–50 Da) to cover the mass range of interest. To guarantee sufficient data collection rate and chromatographic definition, a fixed number of 10 isolation windows was used and resulting MS/MS spectra were collected in the Orbitrap at a 15,000 resolution, while MS acquisition was performed at 120,000 (both at m/z 200, FWHM). In order to reduce the complexity of MS/MS spectra and the width of isolation windows, the studied mass range was limited to 100–600 Da which still well covers the panel of low-mass range metabolites detected in plasma samples (Supplementary Material, Figure S6, lipid species were not considered). The width of the isolation window drastically influences the number of precursor ion peaks isolated and simultaneously fragmented, and thus the complexity of the generated MS/MS spectrum. Therefore, we incrementally devised three distinct DIA workflows involving various isolation windows to obtain the more valuable data. During DIA method development, manual assessment of mass spectral data quality was first performed in a targeted way by specifically looking at the presence of three fragment ions shared with the corresponding spectra from our in-house database. This was also a first step toward the development of a targeted monitoring/quantification DIA-based approach (Section 3.4.2). Then, MS-DIAL software was used to evaluate the consistency of collected MS/MS spectra.

The first one made use of the DIA module included in the Orbitrap Fusion™ software, which automatically defines the number of windows (of equal sizes) according to the mass range studied (10 non-overlapping windows of 50 Da in this case) using a single fixed NCE (set at 30%). When evaluated on the pool of 43 compounds, this approach returned relevant MS/MS spectra for 56.8% of the compounds. When using the “Target MS²” acquisition mode of the Orbitrap Fusion software, it became possible to set a stepped NCE of 30% ± 20% and define overlapping 52-Da wide isolation windows. Under these conditions, compound selection and fragmentation were improved, especially at the borders of the isolation windows, thus raising the number of correctly detected compounds to 77.3%. Last, we devised a third acquisition scheme using variable but overlapping precursor isolation window widths (Supplementary Material, Table S4) [38], to partition the precursor ion density observed in plasma (Supplementary Material, Figure S6) almost equally across the whole mass range of interest. Under these conditions, up to 90% of the targeted metabolites were efficiently detected in the NIST plasma extract by their fragment ions (Supplementary Material, Table S1), which clearly underlines the advantages of using variable isolation windows. We next performed mass spectral deconvolution as well as similarity matching within the MS-DIAL software (see below), to more objectively compare the data with those from the above described DDA workflow.

3.4. Performance Evaluation of DDA and DIA Strategies

If DDA and DIA acquisition modes can bring valuable information, they both have their own specific drawbacks that are important to know before choosing the best methods toward the objectives of the study. The two strategies were compared on their ability to produce relevant MS/MS information and for quantitative performance characteristics.

3.4.1. Generation of Meaningful MS/MS Spectra to Confirm the Annotation of Plasma Metabolites

We next compared DDA and DIA approaches through their respective ability to detect the above-described 72 endogenous metabolites by analyzing in triplicate two independent metabolic extracts of NIST plasma. The methods were qualitatively compared first by manual evaluation and then more thoroughly through DP scores.

The DIA workflow provided comprehensive fragmentation of the 72 molecules, whereas only 61% of them were efficiently selected and fragmented by DDA. On average, 2.5 and 11.1 MS/MS spectra per compound were obtained with DDA and DIA approaches, respectively. In both cases, ~90% of the generated spectra proved of sufficient quality to enable manual metabolite confirmation while more than 10 data points/peak also yielded well defined chromatographic peaks for further quantification purposes. Figure 5 shows the data obtained for glycocholic acid as a representative example. These data demonstrate good comparability of MS/MS spectra obtained under DDA and DIA conditions both in terms of presence and relative intensity of the most intense fragment ions (despite the use of distinct fragmentation conditions), and illustrate the higher purity of MS/MS spectra acquired under DDA conditions (Figure 5).

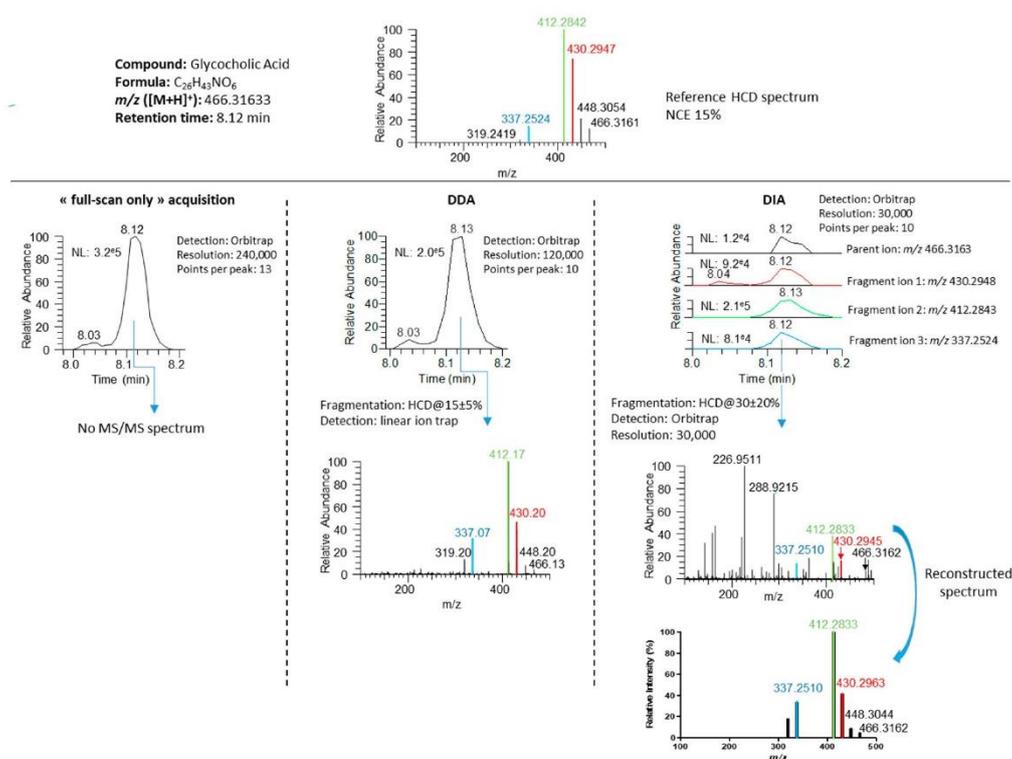


Figure 5. MS and MS/MS data of glycocholic acid obtained with full-scan only, DDA and DIA workflows. The reference MS/MS spectrum was acquired at an NCE of 15% while DDA and DIA acquisitions used NCEs of 15% ± 10% and 30% ± 20%, respectively. DIA MS/MS spectrum of glycocholic acid was manually reconstructed from relative fragment ion intensities.

Table S8 (Supplementary Material) compares the DP scores obtained for 34 metabolites consistently observed for the 6 replicates (two distinct plasma extractions, three technical replicates each) using both DDA and DIA workflows. DP scores returned by both approaches proved consistent between the two extractions. As expected, overall DIA DP scores proved significantly lower with median DP scores of 745 and 628 for DDA and DIA, respectively (Supplementary Material, Table S8). DIA is also characterized by 3 times more metabolites with DP scores below 500 (22 vs. 7). Unsurprisingly, DP scores from DIA measurements seem highly influenced by the quality of mass spectral deconvolution which in turn depends on initial signal intensity and spectral complexity (Figure 6). Even though the DIA can notably

benefit from improvements in the deconvolution process, the present results tend to confirm the great potential of such an approach for untargeted metabolomics, with numerous mass spectra exhibiting a rather astonishing spectral quality when compared to reference mass spectra (Supplementary Material, Figure S7).

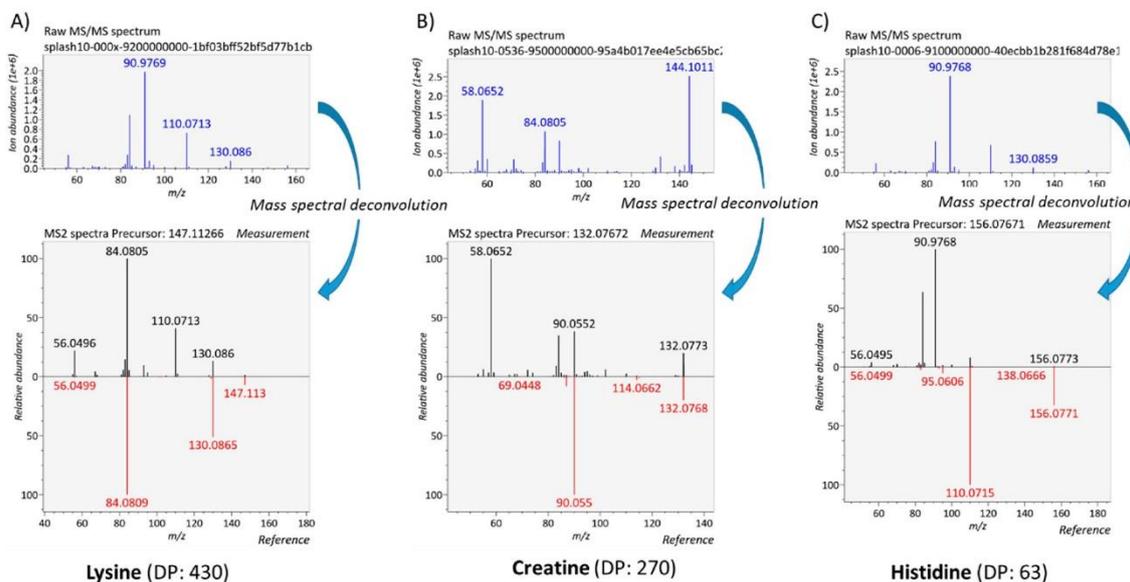


Figure 6. Deconvolution examples of DIA data. (A) Lysine with a good DP score, (B) creatine whose measured spectrum is contaminated by significant fragment ions arising from co-eluting and co-selected precursor ions but removed by an efficient deconvolution, and (C) histidine displaying a very low DP score due to the high abundance of fragment ions arising from other co-fragmented precursor ions.

DDA and DIA acquisition modes are therefore highly complementary and each of them can bring valuable information, depending on the required information. DDA can be used to generate high-quality mass spectra to confirm metabolite annotation or to feed a reference library gathering spectra from known compounds as well as endogenous compounds but still structurally unnamed. DIA approach might be the best choice for targeted or semi-targeted metabolomics projects, within which reference MS/MS spectra are generally available for the metabolites of interest. In this context, particular fragment ions can be targeted to confirm annotations, while fragment ion chromatograms can also be extracted for (semi)quantification purposes despite the lack of direct precursor-fragment link. Since all metabolites are fragmented in such a DIA workflow, all detected metabolites can be basically monitored.

3.4.2. Quantification of Human Plasma Metabolites from DDA and DIA Modes

One of the main driving forces in metabolomics, following metabolite identification, is the sensitive and robust metabolite quantification. We next intended to compare the quality of relative quantification from DDA and DIA methods by comparison to the reference MS1-only LC-HRMS approach. Then, to further mimic the complexity of real-life complex samples, the set of 43 compounds was spiked into two plasma extracts at concentrations ranging from 0.05 ng/mL to 10 ng/mL and each resulting sample was then analyzed in triplicate. Precursor ion-based quantification was performed from MS1 spectra obtained for both three sets of data, while fragment ion-based quantification was only achieved using DIA data. In the latter case, both precursor ion (when still present) and the three most intense fragments (as deduced from reference MS/MS spectra) were considered. To preliminarily compare measurement reproducibility of both workflows, we first evaluated the number of metabolites detected at a given concentration with a coefficient of variation (CV) on triplicate measurements from the two distinct plasma extracts below 30%.

As shown in Figure S8 (Supplementary Material), the acquisition of MS/MS events, either in DDA or in DIA has no drastic impact on the number of molecules being detected with CV < 30% at each concentration level. Importantly, this demonstrates that adding MS/MS events does not downgrade the quality and robustness of the measurement at the MS1 level, which is consistent with the results recently reported using a fast-acquiring Q-TOF instrument and a similar analytical strategy [25]. Overall, precursor ion-based monitoring from MS1 signals proved of equal sensitivity whatever the acquisition method, while the DIA-MS/MS strategy proved less sensitive with on average ~5 times less compounds detected. This might be related to incomplete efficiency of ion transmission and ion fragmentation using non-optimal NCE conditions. This is in sharp contrast with targeted parallel reaction monitoring (PRM)-based metabolomics approaches where NCE is carefully optimized in a compound-specific manner to reach exquisite sensitivity [39]. Nevertheless, at the highest concentrations, the average measurement reproducibility proved comparable using MS1 and MS/MS ion intensities.

Last, we studied in more detail the performances of the four workflows for the relative quantification of nine spike-in compounds with estimated limits of detection in the 0.05–0.25 ng/mL range (Supplementary Material, Table S9). Overall, MS-only and DDA-MS workflows shared similar detection sensitivities, while quantification based on DIA-MS (MS1 trace of DIA data) and DIA-MS/MS data proved ~2- and ~5-times less sensitive. Concentration-response relationship was demonstrated as linear over at least two orders of magnitude whatever the acquisition workflow with average coefficients of determination $R^2 > 0.97$ (Supplementary Material, Table S9). Similarly, all four workflows demonstrated similar good quantification accuracies as measured at 0.25 and 3 ng/mL concentration levels (with accuracies ranging from 81–108%, and 94–104%, respectively). Although significantly less sensitive, the DIA-MS/MS approach enables simultaneous metabolite structural confirmation and quantification. For example, DIA-MS/MS data proved useful to highlight potential misannotation of dextromethorphan (cough suppressant) in blank human plasma after analysis by the full-scan only method, while also enabling its quantification in the 0.05–10 ng/mL range (Figure 7).

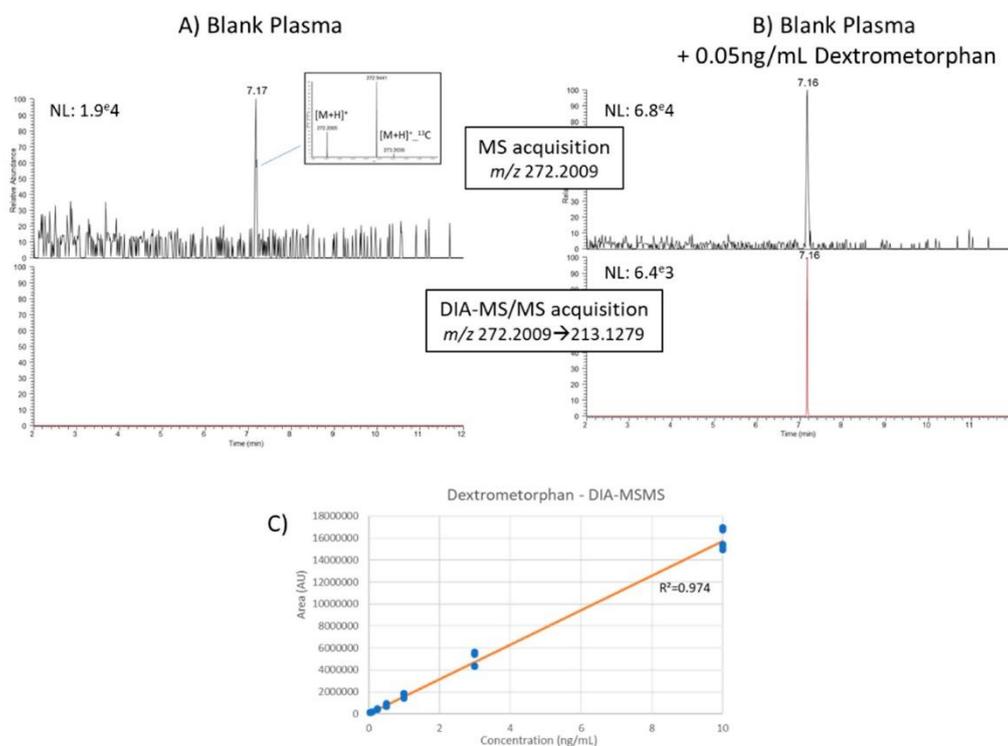


Figure 7. Example of Dextromethorphan. (A) A metabolite feature probably misidentified in blank plasma as dextromethorphan by the full-scan MS method and not confirmed by the DIA-MS/MS approach, (B) similar data as in (A) but after spiking 0.05 ng/mL of pure dextromethorphan in plasma. (C) calibration curve of dextromethorphan in human plasma as analyzed by DIA-MS/MS.

4. Conclusions

Performance of the Orbitrap FusionTM instrument, equipped with a compact high-field Orbitrap analyzer, was studied for the efficient MS/MS data collection in an untargeted metabolomics context. Our data indicated that both DDA and SWATH-type DIA approaches are highly valuable and proved complementary, DDA not surprisingly being more suited for acquiring high-quality and higher-purity spectra. The unprecedented versatility of the Orbitrap FusionTM enables parallelized DDA acquisition of both high resolution and low resolution spectra as well as concomitant collection of HCD and CID data, which is particularly relevant for identification of unknowns. Interestingly, comparative quantitative evaluation of those multi-event approaches revealed that the acquisition of MS/MS spectra, either in DDA or in DIA, does not downgrade the quality of the quantification at the MS1 level. While significantly less sensitive (~five times), fragment ion-based quantification from DIA-MS/MS data enables simultaneous metabolite structural confirmation and quantification of known metabolites. Most of the methodological developments reported here on the Orbitrap Fusion and drawn conclusions are also relevant for other Orbitrap-based platforms, such as Q-Exactive or LTQ-Orbitrap instruments. Both DDA and DIA approaches for metabolomics would greatly benefit from improvements in data acquisition speed on Orbitrap mass analyzers as well as in NCE management to better cover the wide heterogeneity of metabolites in terms of optimal fragmentation conditions. Given the results of this study, we conclude that combining metabolite quantification in the MS1 space and DIA-MS/MS for obtaining structural information of all detected compounds and querying high-quality spectral databases (generated by targeted DDA at multiple NCEs), can have great potential for future biological metabolomics applications.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2218-1989/10/4/158/s1>, Figure S1: Comparison of HCD spectra acquired for 4-pyridylacetic acid on the Orbitrap FusionTM or on a Q-ExactiveTM using variable *m/z* windows for fragment ions; Figure S2: Comparison of HCD spectra acquired for propranolol on the Orbitrap FusionTM or on a Q-ExactiveTM using variable *m/z* windows for fragment ions; Figure S3: Comparison of HCD spectra acquired for ochratoxin A on the Orbitrap FusionTM or on a Q-ExactiveTM using variable *m/z* windows for fragment ions; Figure S4: Distribution of optimal collision energy for 482 standard metabolites; Figure S5: Distribution of optimal collision energy for 72 standard metabolites identified in NIST plasma using C₁₈ UHPLC system coupled to (ESI+)-MS/MS; Figure S6: Mass distribution of metabolites identified in plasma; Figure S7: Head-to-tail comparison of evaluated versus reference MS/MS spectra. Evaluated MS/MS spectra were obtained using the DIA acquisition workflow; Figure S8: Number of metabolites reproducibly detected (with CV < 30%) using full-scan only (blue), DDA (red), DIA-MS (green) and DIA MS/MS workflows (purple). Each bar is the result of six independent measurements, Text S1. Optimisation of a first “HCD-only” DDA method; Table S1: List of the 47 compounds included in the test sample as analyzed under ESI+ conditions; Table S2: Acquisition parameters of the “reference” DDA workflow; Table S3: Acquisition parameters of the DDA workflow combining low- and high-mass resolution acquisitions; Table S4: Acquisition parameters of the DIA workflow; Table S5: List of the 72 metabolites annotated in plasma NIST in C18 LC condition and ESI+ ionization; Table S6: Example of an exclusion list generated from contaminant background ions; Table S7: Dot product scores obtained for the standard mixture by using the HCD-only DDA workflow (mean values of three replicate measurements); Table S8: Dot product scores obtained by DDA and DIA for a set of 34 metabolites (mean of three replicates per extraction); Table S9: Comparison of performance characteristics of MS-Only, DDA-MS, DIA-MS, and DIA-MS/MS workflows.

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

Funding: This work was supported by the Commissariat à l’Energie Atomique et aux Energies Alternatives and the MetaboHUB infrastructure (ANR-11-INBS-0010 grant).

Acknowledgments: Simon Broudin and Etienne Thévenot are warmly thanked for their valuable advice and help.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Gika, H.; Virgiliou, C.; Theodoridis, G.; Plumb, R.S.; Wilson, I.D. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2019**, *1117*, 136–147. [\[CrossRef\]](#)
2. Alonso, A.; Marsal, S.; Julia, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* **2015**, *3*, 23. [\[CrossRef\]](#)
3. da Silva, R.R.; Dorrestein, P.C.; Quinn, R.A. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12549–12550. [\[CrossRef\]](#)
4. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8*, 31. [\[CrossRef\]](#)
5. Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinform.* **2006**, *7*, 234. [\[CrossRef\]](#)
6. Barbier Saint Hilaire, P.; Hoheneiter, 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*, 3030–3035. [\[CrossRef\]](#)
7. 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.* **2019**, *120*, 115324. [\[CrossRef\]](#)
8. Rathahao-Paris, E.; Paris, A.; Bursztyka, J.; Jaeg, J.-P.; Cravedi, J.-P.; Debrauwer, L. Identification of xenobiotic metabolites from biological fluids using flow injection analysis high-resolution mass spectrometry and post-acquisition data filtering. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 2713–2722. [\[CrossRef\]](#)
9. Zhang, H.; Zhang, D.; Ray, K.; Zhu, M. Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. *J. Mass Spectrom.* **2009**, *44*, 999–1016. [\[CrossRef\]](#)
10. 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*, 4876–4882. [\[CrossRef\]](#)
11. Benton, H.P.; Ivanisevic, J.; Mahieu, N.G.; Kurczy, M.E.; Johnson, C.H.; Franco, L.; Rinehart, D.; Valentine, E.; Gowda, H.; Ubhi, B.K.; et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal. Chem.* **2015**, *87*, 884–891. [\[CrossRef\]](#)
12. Neumann, S.; Thum, A.; Böttcher, C. Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data. *Metabolomics* **2013**, *9*, 84–91. [\[CrossRef\]](#)
13. Koelmel, J.P.; Kroeger, N.M.; Gill, E.L.; Ulmer, C.Z.; Bowden, J.A.; Patterson, R.E.; Yost, R.A.; Garrett, T.J. Expanding Lipidome Coverage Using LC-MS/MS Data-Dependent Acquisition with Automated Exclusion List Generation. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 908–917. [\[CrossRef\]](#)
14. Lawson, T.N.; Weber, R.J.M.; Jones, M.R.; Chetwynd, A.J.; Rodriguez-Blanco, G.; Di Guida, R.; Viant, M.R.; Dunn, W.B. msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics. *Anal. Chem.* **2017**, *89*, 2432–2439. [\[CrossRef\]](#)
15. 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*, 471–500. [\[CrossRef\]](#)
16. 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. [\[CrossRef\]](#)
17. 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*, 1068–1080. [\[CrossRef\]](#)
18. Broadhurst, D.; Goodacre, R.; Reinke, S.N.; Kuligowski, J.; Wilson, I.D.; Lewis, M.R.; Dunn, W.B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018**, *14*, 72. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)

20. 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. Proteom.* **2012**, *11*, O111.016717. [[CrossRef](#)]
21. Naz, S.; Gallart-Ayala, H.; Reinke, S.N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C.E. Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal. Chem.* **2017**, *89*, 7933–7942. [[CrossRef](#)]
22. Tada, I.; Tsugawa, H.; Meister, I.; Zhang, P.; Shu, R.; Katsumi, R.; Wheelock, C.E.; Arita, M.; Chaleckis, R. Creating a Reliable Mass Spectral-Retention Time Library for All Ion Fragmentation-Based Metabolomics. *Metabolites* **2019**, *9*, 251. [[CrossRef](#)]
23. Gallart-Ayala, H.; Courant, F.; Severe, S.; Antignac, J.-P.; Morio, F.; Abadie, J.; Le Bizec, B. 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* **2013**, *796*, 75–83. [[CrossRef](#)]
24. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526. [[CrossRef](#)]
25. Chen, G.; Walmsley, S.; Cheung, G.C.M.; Chen, L.; Cheng, C.-Y.; Beuerman, R.W.; Wong, T.Y.; Zhou, L.; Choi, H. Customized Consensus Spectral Library Building for Untargeted Quantitative Metabolomics Analysis with Data Independent Acquisition Mass Spectrometry and MetaboDIA Workflow. *Anal. Chem.* **2017**, *89*, 4897–4906. [[CrossRef](#)]
26. Bonner, R.; Hopfgartner, G. SWATH data independent acquisition mass spectrometry for metabolomics. *TrAC Trends Anal. Chem.* **2019**, *120*, 115278. [[CrossRef](#)]
27. 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*, 128–136. [[CrossRef](#)]
28. Jia, W.; Shi, L.; Zhang, F.; Fan, C.; Chang, J.; Chu, X. Multiplexing data independent untargeted workflows for mycotoxins screening on a quadrupole-Orbitrap high resolution mass spectrometry platform. *Food Chem.* **2019**, *278*, 67–76. [[CrossRef](#)]
29. Boudah, S.; Olivier, M.-F.; Aros-Calt, S.; Oliveira, L.; Fenaille, F.; Tabet, J.-C.; Junot, C. 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.* **2014**, *966*, 34–47. [[CrossRef](#)]
30. Contrepois, K.; Jiang, L.; Snyder, M. Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol. Cell. Proteom.* **2015**, *14*, 1684–1695. [[CrossRef](#)]
31. Moreau, R.; Clària, J.; Aguilar, F.; Fenaille, F.; Lozano, J.J.; Junot, C.; Colsch, B.; Caraceni, P.; Trebicka, J.; Pavesi, M.; et al. Blood metabolomics uncovers inflammation-associated mitochondrial dysfunction as a potential mechanism underlying ACLF. *J. Hepatol.* **2020**, *72*, 688–701. [[CrossRef](#)] [[PubMed](#)]
32. Geyer, P.E.; Kulak, N.A.; Pichler, G.; Holdt, L.M.; Teupser, D.; Mann, M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2016**, *2*, 185–195. [[CrossRef](#)] [[PubMed](#)]
33. Bruderer, R.; Muntel, J.; Müller, S.; Bernhardt, O.M.; Gandhi, T.; Cominetti, O.; Macron, C.; Carayol, J.; Rinner, O.; Astrup, A.; et al. Analysis of 1508 Plasma Samples by Capillary-Flow Data-Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance. *Mol. Cell. Proteom.* **2019**, *18*, 1242–1254. [[CrossRef](#)]
34. Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S.S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy. *J. Cheminform.* **2017**, *9*, 32. [[CrossRef](#)]
35. 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*, e0160160. [[CrossRef](#)]

36. 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.; et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221. [[CrossRef](#)]
37. Calderón-Santiago, M.; Priego-Capote, F.; Luque de Castro, M.D. Enhanced detection and identification in metabolomics by use of LC-MS/MS untargeted analysis in combination with gas-phase fractionation. *Anal. Chem.* **2014**, *86*, 7558–7565. [[CrossRef](#)]
38. 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*, 4359–4371. [[CrossRef](#)]
39. 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*, 4478–4486. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Les données supplémentaires sont présentées en Annexe 2.

4 – Principaux résultats et conclusions

Ces travaux ont permis de développer de nouvelles méthodes d’acquisition permettant l’obtention simultanée de spectres MS et MS/MS. Les méthodes DDA et DIA ainsi développées ont pu être comparées grâce à un mélange de 47 standards et à un extrait de plasma NIST pour lequel 72 métabolites endogènes sont identifiés en routine (LC-HRMS, « C18 positif ») au sein du laboratoire.

Tout d’abord, après des développements concernant l’ensemble des filtres de sélection des ions, une méthode DDA a été obtenue en tenant compte qu’au cours d’une analyse métabolomique, certains métabolites sont connus et présents dans une base de données internes et d’autres sont inconnus. Le premier élément investigué fut l’énergie de collision. En effet, il est nécessaire d’utiliser un mode de fragmentation permettant l’obtention de spectres de fragmentation les plus informatifs possibles pour un maximum de métabolites. Le mode « stepped collision energy » a été investigué et a démontré son intérêt. Il permet d’obtenir un spectre MS/MS correspondant à la somme de trois spectres MS/MS acquis à trois énergies de collision. La Figure 38 montre l’exemple de l’alanyltyrosine. Le spectre MS/MS obtenu avec le mode « stepped collision energy » HCD 30 +/- 20 % correspond à la somme des spectres MS/MS obtenus avec des énergies de fragmentation HCD fixée à 10 %, 30 % et 50 %.

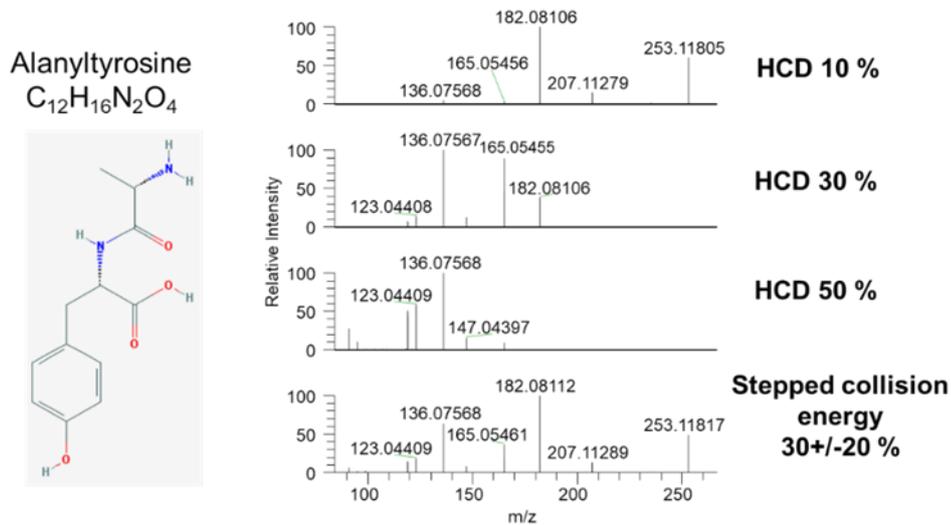


Figure 38 : Exemple de fragmentation en mode « stepped collision energy » pour l'analyltyrosine (extraite de la publication ci-dessus).

L'analyseur ainsi que le mode de fragmentation ont donc été choisis en fonction de la présence ou non des métabolites présents dans notre base de données. En effet, dans le cas des métabolites déjà présents dans notre base de données interne, l'identification se fait par comparaison des spectres MS/MS obtenus avec ceux présents dans la base de données. Pour cela, une haute précision en masse et donc un analyseur à haute résolution n'est pas forcément nécessaire. La détection des ions a donc pu être réalisée *via* le piège à ions. Ainsi, nous avons pu paralléliser les acquisitions. De plus, pour ces composés, l'énergie de fragmentation optimale est connue (dans notre cas, elle correspond à l'obtention du spectre MS/MS avec l'intensité relative de l'ion parent entre 10 % et 30 %). Les ions sont alors séparés en deux listes selon leur énergie de fragmentation optimale pour être fragmentés en mode HCD « stepped collision energy » 15 +/- 10 % ou 45 +/- 20 %. Ils sont ensuite détectés grâce au piège à ions linéaire. Les spectres MS/MS obtenus par « stepped collision energy » sont comparés à notre base de données internes comportant les spectres MS/MS des composés à 4 énergies de collision (HCD 10 %, 20 %, 40 % et 80 %) sur la base des *m/z* de l'ion précurseur et de trois fragments. Dans le cas des composés inconnus, l'identification peut être réalisée par interprétation manuelle des spectres MS/MS obtenus. Dans ce cas, il est donc nécessaire d'obtenir des spectres de fragmentation les plus informatifs possibles. Pour cela, les ions sont fragmentés à la fois en mode HCD (non résonnant) « stepped collision energy » et en mode CID (résonnant) afin d'obtenir des informations complémentaires¹⁰⁷. Dans ce cas, une haute précision en masse et donc un analyseur à haute résolution sont requis. Les métabolites inconnus sont alors détectés avec l'Orbitrap. Afin d'obtenir des temps de cycle (correspondant à l'acquisition d'un spectre MS et de plusieurs spectres MS/MS) concordants avec la durée d'un pic chromatographique, le nombre d'événements MS/MS entre deux « Full Scan » a été fixé à dix (top-10) et les résolutions à 120 000 et 15 000 (à *m/z* 200) respectivement pour les acquisitions MS et MS/MS. Les anciennes générations de spectromètres de masse avec des analyseurs de type Orbitrap (par exemple le LTQ-Orbitrap Velos, Thermo Fisher Scientific) ne permettaient que des acquisitions DDA top-5 probablement en raison de leur plus faible vitesse d'acquisition¹⁰⁶. Les résolutions d'acquisition des

spectres MS et MS/MS ont également été optimisées afin d'obtenir un temps de cycle d'une seconde. Avec cette méthode, 95 % des 115 métabolites de l'étude sont détectés avec le « Full Scan » et 65 % de ces composés sont sélectionnés pour être fragmentés. Cette sélection monte à 80 % à la plus haute concentration de dopage du plasma. L'efficacité de sélection est la même quel que soit le mode d'acquisition (basse ou haute résolution). La qualité des spectres MS/MS a ensuite été évaluée. Dans un premier par comparaison manuelle avec les spectres MS/MS de notre base de données. Cependant, cette méthode est très chronophage et ne permet pas d'obtenir une évaluation objective de la qualité des spectres obtenus (nombre et intensités relatives des ions fils, « pureté » des spectres). Un logiciel libre de droits, MS-DIAL, a donc été testé pour évaluer la qualité des spectres MS/MS¹⁹⁷. Notre base de données interne a été implémentée dans ce logiciel pour confronter de façon objective les spectres MS/MS obtenus à ceux de notre base de données. Qualitativement, nous avons considéré une bonne identification des composés lorsque le « dot product score » était supérieur à 500. Cette valeur est concordante à celle fixée lors projet CASMI dont le but est de proposer des critères pour aider à l'identification des petites molécules. En effet, dans le cadre de ce projet, le « DP score » optimal fut déterminé à 400¹⁷². Les deux méthodes de comparaison des spectres MS/MS (manuelle et avec l'aide d'un logiciel) ont montré que les spectres MS/MS obtenus sont de haute qualité et avec le logiciel MS-DIAL, 80 % des spectres obtenus avec la méthode DDA possèdent un « dot-product score » supérieur à 500 (Figure 39). Les spectres de fragmentation obtenus en mode « stepped collision energy » sont donc directement comparables avec ceux obtenus à une seule énergie de fragmentation.

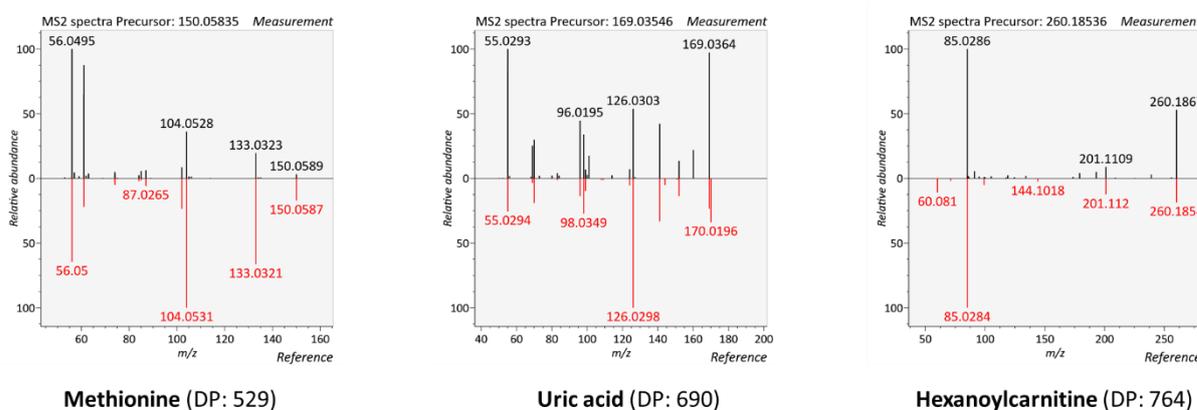


Figure 39 : Comparaison des spectres MS/MS obtenus avec ceux de référence à l'aide de MS-DIAL.

Les travaux récents de Blaženović et al. montrent qu'il est possible d'annoter l'ensemble des métabolites fragmentés en DDA moyennant une suite d'outils bio-informatiques et des bases de données très complètes (notamment MoNA, NIST, LipidBlast)²⁵⁶. Dans ces conditions, environ 750 métabolites ont été annotés aux niveaux 1 et 2 (selon Sumner et al.¹⁷³) et plusieurs centaines au niveau 3. Comme l'ont souligné nos travaux, les méthodes DDA (optimisées) permettent d'identifier un grand nombre de métabolites lorsque ceux-ci sont sélectionnés et fragmentés mais tous les métabolites ne sont pas fragmentés dans ces conditions expérimentales. Fragmenter et obtenir des spectres pertinents pour un maximum d'ions apparaît comme indispensable.

Pour maximiser le nombre de métabolites fragmentés, une méthode DIA de type « SWATH » a été développée. Ces développements se sont focalisés sur la taille des fenêtres de masse afin d’avoir une bonne sensibilité. Tout comme pour la méthode DDA, le nombre de fenêtres et donc d’événements MS/MS a été limité pour être compatible avec l’analyse chromatographique. Ce nombre de fenêtres a donc été fixé à dix sur une gamme de masses de 100 à 600 Da, correspondant à la gamme de masses où la majorité des métabolites sont détectés avec la méthode LC-HRMS « C18 positif » utilisée au laboratoire (Figure 40).

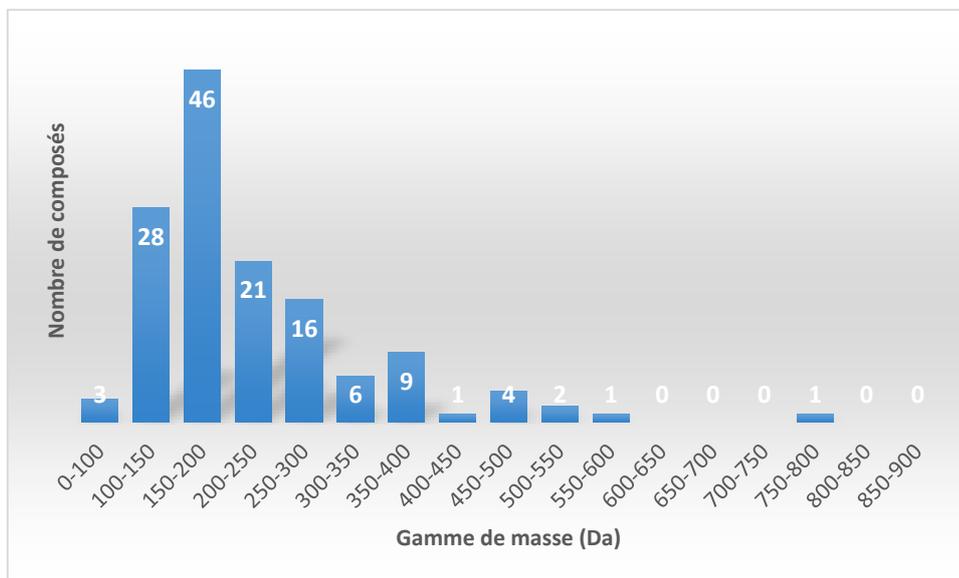


Figure 40 : Répartition des masses des métabolites détectés dans le plasma NIST.

Dans un premier temps, des tests ont été réalisés en gardant des fenêtres de même taille (50 Da). Cependant, dans un échantillon biologique, la distribution des masses des métabolites annotés n’est pas homogène. En effet, il y a plus de métabolites avec de petites masses (inférieures à 300 Da) que de métabolites avec de hautes masses (supérieures à 300 Da) (Figure 40). Les tailles des fenêtres de masses ont donc été ajustées selon la répartition des métabolites du plasma NIST. Les largeurs des fenêtres varient alors entre 16 et 102 Da, et les fenêtres sont chevauchantes afin de détecter au mieux des métabolites dont les masses correspondraient aux limites de ces fenêtres (Tableau 8). Avec cette méthode, 90 % des métabolites ont pu être identifiés grâce à la détection de leurs fragments. Ces résultats sont consistants avec ceux de Zhang et al., qui ont également montré que l’utilisation de fenêtres de tailles variables permettait d’améliorer la sélectivité, l’identification et la quantification de métabolites et de peptides¹⁹⁶.

Tableau 8 : Tailles des fenêtres d'acquisition SWATH.

N° de fenêtres	Fenêtres « SWATH » classique (en m/z)	Fenêtres « SWATH » avec des fenêtres ajustées (en m/z)
1	100.3 - 150.3	98 - 132
2	150.3 - 200.3	129.5 - 145.5
3	200.4 - 250.4	144.5 - 160.5
4	250.4 - 300.4	159.5 - 175.5
5	300.4 - 350.4	174.5 - 190.5
6	350.4 - 400.4	189 - 211
7	400.5 - 450.5	209 - 262
8	450.5 - 500.5	259 - 301
9	500.5 - 550.5	300 - 400
10	550.5 - 600.5	398 - 602

L'aspect quantitatif des méthodes DDA et DIA a ensuite été évalué par comparaison avec la méthode « Full Scan » traditionnellement utilisée pour les études métabolomiques de profilage semi-quantitatif. Le but étant de savoir si l'acquisition simultanée de spectres MS et MS/MS avait un impact sur la quantification des métabolites. Cette quantification a été réalisée à partir des spectres MS pour les trois méthodes d'intérêt et également à partir des spectres MS/MS dans le cas de la méthode DIA (le nombre d'acquisition de spectres MS/MS d'un même métabolite avec la méthode DDA ne permet pas de faire ce type de quantification). La quantification basée sur les spectres MS est similaire quel que soit le mode d'acquisition utilisé. Cela montre que l'ajout de spectres MS/MS n'altère pas la qualité et la robustesse des acquisitions MS ce qui est en accord avec des données obtenues précédemment avec des analyses réalisées à l'aide d'un Q-TOF²⁵⁷. Dans le cas des analyses DIA, cela signifie que la quantification et l'identification des métabolites peuvent être réalisées en une seule et même injection. Néanmoins, la sensibilité de la quantification en DIA-MS/MS apparaît comme environ cinq fois moindre comparativement aux autres approches (dans nos conditions en tout cas).

Les données acquises par les méthodes DDA et DIA ont été traitées manuellement dans un premier temps. Ce traitement manuel est long et fastidieux ce qui nous a encouragé à utiliser des méthodes de traitement de données plus automatisées et objectives. Nous avons donc utilisé le logiciel MS-DIAL, libre d'accès, pour évaluer objectivement la qualité de nos spectres MS/MS. En parallèle, nous avons collaboré avec les équipes de Biognosys (Zurich, Suisse) pour tester le logiciel Spectronaut qu'ils ont développé initialement pour traiter les données DIA de protéomiques²⁵⁸. Après des premiers tests réalisés au sein du laboratoire, les équipes de Biognosys ont adapté le logiciel Spectronaut aux analyses métabolomiques. Ces tests ont été effectués sur les données acquises avec la méthode DIA des plasmas supplémentés avec le mélange de standards. Les résultats obtenus avec le logiciel Spectronaut sont équivalents à ceux que nous avons obtenus par traitement des données manuel (Figure 41). Ce logiciel semble donc très prometteur pour les futurs traitements de données liés aux acquisitions DIA en métabolomique. Des essais pourront également être effectués sur les données acquises en mode DDA.

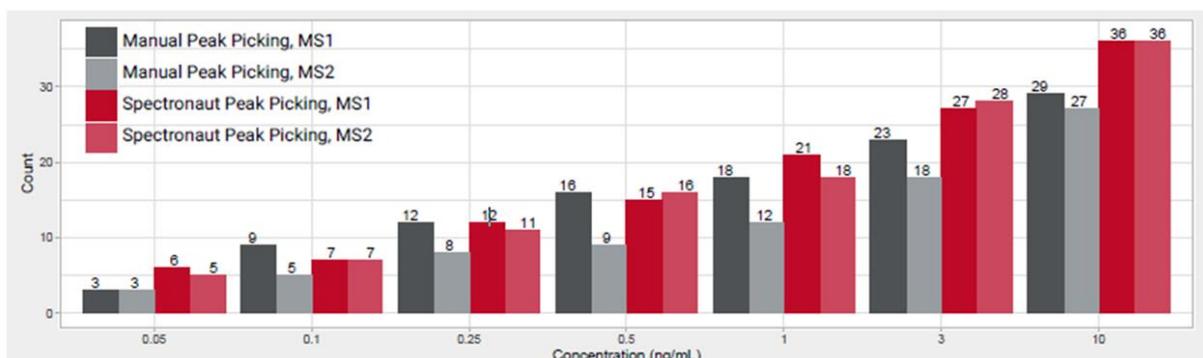


Figure 41 : Comparaison du nombre de composés quantifiés par MS et MS/MS par traitement de données manuel ou par Spectronaut.

Ces travaux ont été présentés sous forme de poster au congrès virtuel de l'American Society of Mass Spectrometry en Juin 2020.

Pour conclure, les méthodes DDA et DIA de type « SWATH » développées montrent de bonnes caractéristiques et une certaine complémentarité. La méthode DDA répond bien au challenge lié à l'identification des métabolites. En effet, les spectres MS/MS obtenus sont très purs ce qui permet de les comparer aisément à des bases de données. De plus, les spectres MS/MS obtenus permettent d'améliorer l'élucidation structurale des métabolites inconnus. La méthode DIA répond à deux challenges rencontrés lors d'analyses métabolomiques qui sont l'identification et la quantification des métabolites. De même que pour les acquisitions en mode DDA, les spectres MS/MS obtenus sont comparés à des bases de données. Cependant, la faible pureté des spectres MS/MS rend complexe cette comparaison. Il serait judicieux d'analyser tous les échantillons en mode DIA afin de se donner l'opportunité de quantifier les métabolites d'intérêt tout en s'assurant de leur identification. Le point limitant des acquisitions DIA est essentiellement lié au processus de déconvolution des données afin de relier un ion précurseur à des ions fragments. Cet aspect représente un champ d'investigations important et des données récentes rapportant la valeur ajoutée d'un outil de corrélation-déconvolution (CorrDec) appliquée à une large série d'échantillons sont très encourageantes²⁵⁹. La méthode DDA permettrait quant à elle d'identifier de nouveaux métabolites. Pour cela, et afin de ne pas allonger les temps d'analyses d'une séquence, les acquisitions DDA pourront être réalisées sur les QCs. Dans le cas des méthodes DDA et DIA, la qualité de la comparaison des spectres MS/MS obtenus en mode « stepped collision energy » avec ceux de la base de données acquis avec une seule énergie de fragmentation est à évaluer.

Pour être appliquée aux analyses métabolomiques de routine au laboratoire, il serait également intéressant d'évaluer et d'optimiser les méthodes DDA et DIA pour les analyses HILIC en mode d'ionisation négatif.

B – Application à des études d'intercomparaison de spectres de masse dans le cadre d'un projet européen (HBM4EU)

1 - Contexte du projet européen

Le projet HBM4EU (Human Biomonitoring for Europe) est un projet européen qui a démarré en 2017 et qui représente l'effort conjoint de 30 pays ainsi que de l'Agence Européenne pour l'Environnement et de la Commission Européenne²⁶⁰. La biosurveillance humaine consiste à mesurer les substances chimiques environnementales et leurs métabolites dans le corps humain à l'aide d'analyse d'échantillons biologiques. Ces analyses sont importantes pour mesurer les effets liés à l'exposition à ces substances chimiques²⁶¹. Ce projet a pour but de coordonner et de faire progresser la biosurveillance humaine en Europe en mettant en évidence des preuves de l'exposition des citoyens aux substances chimiques et les effets possibles sur leur santé. Les résultats obtenus seront mis à disposition des responsables politiques afin qu'ils puissent prendre les décisions nécessaires à la préservation de la santé des citoyens. Les principaux objectifs de ce projet sont les suivants :

- Harmoniser les procédures de biosurveillance humaine dans l'ensemble des pays participants pour donner aux responsables politiques des données comparables sur l'exposition interne de l'homme aux produits chimiques et aux mélanges de produits chimiques au niveau de l'Union Européenne.
- Relier les données sur l'exposition interne aux substances chimiques à l'exposition externe globale et identifier les voies d'exposition et les sources en amont.
- Fournir les outils les plus pertinents pour détecter les nouvelles substances chimiques et identifier les mélanges de substances chimiques les plus préoccupantes.
- Adapter les méthodes d'évaluation des risques chimiques pour utiliser les données de biosurveillance humaine et tenir compte de la contribution des multiples voies d'exposition externes à la charge corporelle totale de substances chimiques.
- Intégrer des informations sur les voies d'exposition dans la conception de mesures politiques ciblées visant à réduire l'exposition.

Aujourd'hui, afin d'améliorer l'identification des substances chimiques (contaminantes) et de leurs métabolites potentiels au sein de différents laboratoires, il apparaît nécessaire de construire une base de données spectrales (MS/MS) robustes issues d'un très grand nombre de métabolites. La quantité de composés ciblés étant très importante (plusieurs centaines de molécules), cette base de données doit être alimentée par différents laboratoires Européens unissant leurs efforts. Néanmoins, un des prérequis à la constitution d'une telle base de données est que les spectres MS/MS soient comparables d'un laboratoire à l'autre. Les abondances relatives des ions fragments et les voies de dissociation observées dans le cadre des expériences en basse énergie sont régies principalement par des aspects cinétiques²⁶² (énergie interne et temps de résidence des ions) et des

paramètres instrumentaux (pression, nature du gaz de collision, etc.). Suivant ces considérations, la création d'une base de données de spectres MS/MS, indépendante de la plate-forme instrumentale, n'est pas facile. Ainsi, dans un premier temps, nous avons été impliqués dans l'harmonisation des méthodologies analytiques permettant l'acquisition de spectres MS/MS afin d'évaluer la comparabilité des spectres MS/MS obtenus.

2 – Création d'une base de données et harmonisation des protocoles d'acquisition

a – Contexte et objectifs

L'étude des effets d'exposition est nécessaire pour évaluer leur impact sur la santé de l'ensemble des citoyens. Ces marqueurs d'expositions sont très nombreux (plusieurs milliers selon l'exposition possible). Il est donc très complexe pour un laboratoire seul d'identifier l'ensemble de ces marqueurs à partir d'une base de données spectrale générée dans une seule unité. La création de consortium entre différents laboratoires s'avère donc primordiale pour augmenter la couverture de ces métabolites d'intérêt. La création d'une base de données à l'échelle Européenne permettrait d'augmenter considérablement le nombre de spectres de masse disponibles et donc de faciliter l'identification des métabolites pour l'ensemble des partenaires.

Dans les analyses métabolomiques courantes, les métabolites sont classés en trois catégories : les composés connus, suspects ou inconnus. Pour les composés connus, les méthodes d'acquisition sont généralement ciblées et quantitatives contrairement à l'étude des composés inconnus. Cela a conduit les laboratoires à développer des méthodes de profilage non ciblées comme les méthodes DDA. En effet, comme montré dans la partie précédente, le mode d'acquisition DDA permet d'obtenir des spectres MS/MS de haute qualité ce qui les rend plus facilement comparables à ceux présents dans des bases de données qui se doivent d'être interopérables. Les laboratoires ne possédant pas les mêmes spectromètres de masse, l'harmonisation des méthodologies analytiques entre ces laboratoires est cruciale que ce soit au niveau de la préparation d'échantillon ou de l'acquisition comme du traitement des données²⁶³. La mise en place de critères d'assurance qualité et de contrôle qualité est primordiale. Des premières études ont montré que des spectres MS/MS acquis avec différents spectromètres de masse (Q-TOF, LIT, 3-D quadrupole IT et LIT-Orbitrap) sont comparables à ceux d'une base de données réalisée à partir de spectres MS/MS acquis avec un Q-TOF (Wiley Registry of Tandem Mass Spectral Data, MSforID)²⁶⁴. Ce projet s'inscrit dans le même objectif que le projet CASMI qui consiste à aider les différents laboratoires à identifier les métabolites en travaillant sur les spectres MS/MS obtenus dans chaque laboratoire¹⁷².

L'objectif premier des différentes études de ce projet est d'évaluer la comparabilité des spectres MS/MS produits par les différents laboratoires sur des spectromètres de masses différents (essentiellement Q-TOF et Q-Orbitrap). Ces spectres MS/MS sont comparés à une base de données déjà mise en place avec un spectromètre de masse Q-TOF. Ces analyses doivent permettre l'élaboration de critères d'assurance qualité et de contrôle qualité pour permettre la construction d'une nouvelle base de données spectrale à plus grande échelle.

b – Méthodologie

Trois études interlaboratoires ont été réalisées.

La première, l'étude « SCORE » consistait à acquérir des spectres MS/MS en mode DDA sur un échantillon blanc et un mélange de standards connus envoyé par l'un des partenaires (Institut de médecine légale et de métabolomique, Université médicale d'Innsbruck, Autriche) et sur un autre échantillon blanc et un mélange de standards issus de notre laboratoire (Tableau 9). Les composés étudiés sont principalement des herbicides, insecticides et composés pharmaceutiques.

Tableau 9 : Récapitulatif des mélanges de standards de l'étude "SCORE".

Mélange de Standards du Laboratoire Partenaire			Mélange de Standards de Notre Laboratoire		
Composé	Formule Brute	[M+H] ⁺	Composé	Formule Brute	[(M+H)] ⁺
6-Mercaptopurine	C5H4N4S	153.02294	Amitriptyline	C20H23N	278.19033
Alachlor	C14H20ClNO2	270.12552	Amoxicilline	C16H19N3O5S	366.11182
Bayrepel	C12H23NO3	230.17505	Atenolol	C14H22N2O3	267.17031
Boscalid	C18H12Cl2N2O	343.03993	Atropine	C17H23NO3	290.17507
Caffeine	C8H10N4O2	195.08764	Caffeine	C8H10N4O2	195.08766
Carbamazepine	C15H12N2O	237.10223	Clindamycine	C18H33ClN2O5S	425.18715
Carbetamide	C12H16N2O3	237.12336	Cotinine	C10H12N2O	177.10224
Chlorbromuron	C9H10BrClN2O2	292.96874	Diltiazem	C22H26N2O4S	415.16860
Chloridazon	C10H8ClN3O	222.04280	Diphenhydramine	C17H21NO	256.16959
Chlorpropham	C10H12ClNO2	214.06292	Doxepine	C19H21NO	280.16959
Dapson	C12H12N2O2S	249.06922	Epinephrine	C9H13NO3	184.09682
Diazinon	C12H21N2O3PS	305.10831	Indomethacin	C19H16ClNO4	358.08407
Diclofenac	C14H11Cl2NO2	296.02395	Nadolol	C17H27NO4	310.20128
Ensulizole	C13H10N2O3S	275.04848	Oxytetracycline	C22H24N2O9	461.15546
Fenofibrate	C20H21ClO4	361.12010	Paracetamol	C8H9NO2	152.07061
Linuron	C9H10Cl2N2O2	249.01920	Perindopril	C19H32N2O5	369.23840
Metconazole	C17H22ClN3O	320.15240	Ranitidine	C13H22N4O3S	315.14854
Metformin	C4H11N5	130.10871	Resveratrol	C14H12O3	229.08592
Metobromuron	C9H11BrN2O2	259.00770	Tetracycline	C22H24N2O8	445.16054
Monuron	C9H11ClN2O	199.06326	Tramadol	C16H25NO2	264.19580
Naproxen	C14H14O3	231.10156			
Phenytoin	C15H12N2O2	253.09714			
Picoxystrobin	C18H16F3NO4	368.11040			
Profenofos	C11H15BrClO3PS	372.94246			
Quinoxifen	C15H8Cl2FNO	308.00396			

Partie B : Etudes Expérimentales – Chapitre 1

Une autre étude nommée « Study 1 » consistait à acquérir des spectres MS/MS en mode DDA sur des échantillons de plasma envoyés par l'un des partenaires. Le but étant de s'affranchir des variations liées à la préparation de l'échantillon. Deux de ces échantillons étaient supplémentés avec deux mélanges de standards distincts dont nous ne connaissions pas la composition. Les analyses ont donc été réalisées à l'aveugle. Un troisième échantillon était un plasma vierge. Une partie de cet échantillon a été analysée telle quelle et une autre partie a été supplémentée avec un mélange de standards de notre laboratoire (Tableau 10).

Pour ces deux études, le laboratoire partenaire avait pour objectif d'identifier les standards que nous avons ajoutés à partir de nos données brutes. Afin de ne pas avoir de biais au niveau du traitement des données et de l'identification des composés, l'ensemble des spectres obtenus a été envoyé au partenaire Autrichien (Herbert Oberacher) qui s'est chargé de l'entièreté du traitement des données.

Tableau 10 : Récapitulatif des mélanges de standards de l'étude "Study 1".

Mélange de Standards 1 du Laboratoire Partenaire	Mélange de Standards 2 du Laboratoire Partenaire	Mélange de Standards de Notre Laboratoire	La-
2-Methylidene-1,5-dimethyl-3,3-diphenylpyrrolidine	3,4-Methylenedioxyamphetamine	Adenosine	
3,4-Methylenedioxyamphetamine	Acetaminophen	Allopurinol	
6-o-Monoacetylmorphine	Amphetamine	Amantadine	
Acetaminophen	Benzoylcegonine	Cimetidine	
Amphetamine	Citalopram	Clenbuterol	
Benzoylcegonine	Cocaine	Dextromethorphan	
Carbamazepine	Fragment of 3,4-Methylenedioxy-methamphetamine	Diclofenac	
Cocaine	Glutamine	Dopamine	
Codeine	Histidine	Gabapentine	
Fragment of 3,4-Methylenedioxy-methamphetamine	Hydroxyzine	Hydrochlorothiazide	
Lidocaine	Ketamine	Lincomycin	
Methadone	Levamisole	Metformin	
Methamphetamine	Lidocaine	Mianserine	
Metoprolol	Lysergic acid diethylamide	Ofloxacin	
Morphine	Lysine	Pantoprazole	
Oxazepam	Methamphetamine	Pheniramine	
Tramadol	Mirtazapine	Sotalol	
Trimethoprim	N,N-Dimethyltryptamine	Theophylline	
Venlafaxine	Olanzapine	Timolol	
	Phenacetin	Warfarin	
	Prothipendyl		
	Quetiapine		
	Sertraline		
	Trazodone		
	Venlafaxine		
	Zolpidem		

Partie B : Etudes Expérimentales – Chapitre 1

Pour ces deux études, les analyses ont été réalisées par LC-HRMS/MS sur l'Orbitrap Fusion. Les paramètres chromatographiques sont les suivants :

Système LC : UHPLC Dionex Ultimate 3000 (Thermo Fisher Scientific, Courtaboeuf, France).

Colonne : Hypersil Gold C18, 2.1 x 150 mm, 2.1 μ m (Thermo Fisher Scientific, Courtaboeuf, France).

Phases Mobiles : Phase A : Eau + 0.1 % d'acide formique ; Phase B : Acétonitrile + 0.1 % d'acide formique.

Gradient : 0 – 2 min : 5 % phase B ; 2 – 13 min : gradient linéaire de 5 % à 100 % de phase B ; 13 – 25.5 min : 100 % de phase B ; 25.5 – 30 min : 5 % de phase B.

Volume d'injection : 10 μ L.

Débit : 500 μ L/min.

Température du four : 30 °C.

Température de passeur d'échantillon : 4°C.

Les acquisitions des spectres MS et MS/MS ont été réalisées avec la méthode DDA « HCD-only » décrite précédemment dans la publication « Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics ». Cette méthode a dû être adaptée à l'acquisition de composés halogénés (aspect détaillé ci-dessous).

Une troisième étude nommée « Study 2 » est une étude préliminaire à la constitution d'une large base de données MS/MS au niveau européen. Pour cela, un set de 19 standards nous a été envoyé (Tableau 11).

Tableau 11 : Liste des 19 standards de l'étude "Study 2".

Composé	Formule Brute	[M+H] ⁺	Pool	Temps de Rétention
Amiloride	C6H8ClN7O	230.0551	3	2.86
Antipyrine	C11H12N2O	189.1022	3	5.65
Cyclizine	C18H22N2	267.1855	1	7.25
Desipramine	C18H22N2	267.1855	3	7.75
Dibucaine	C20H29N3O2	344.2332	3	7.94
Dixyrazine	C24H33N3O2S	428.2366	1	7.82
Dosulepin	C19H21NS	296.1467	2	7.74
Ethambutol	C10H24N2O2	205.191	1	0.71
Etilefrine	C10H15NO2	182.1176	2	1.55
Etofylline	C9H12N4O3	225.0982	1	4.41
Mefruside	C13H19ClN2O5S2	383.0496	1	8.19
Metoclopramide	C14H22ClN3O2	300.1474	2	5.60
Nylidrin	C19H25NO2	300.1958	3	6.73
Phentermine	C10H15N	150.1277	2	5.16
Phénytoin	C15H12N2O2	253.0971	2	7.79
Sulfamethoxazole	C10H11N3O3S	254.0593	2	6.42

Partie B : Etudes Expérimentales – Chapitre 1

Sulfamoxole	C ₁₁ H ₁₃ N ₃ O ₃ S	268.075	3	5.34
Sulthiame	C ₁₀ H ₁₄ N ₂ O ₄ S ₂	291.0467	1	5.90
Tetracycline	C ₂₂ H ₂₄ N ₂ O ₈	445.1605	1	5.55

Dans un premier temps, les analyses LC-HRMS/MS ont été réalisées avec les méthodes traditionnellement utilisées au sein du laboratoire. La méthode chromatographique est la même que celle décrite pour les études « SCORE » et « Study 1 ». Les analyses ont été réalisées avec l'Orbitrap Fusion doté d'une source ESI. Les analyses ont été réalisées en mode positif. Au niveau de la source, une tension de +3500 V est appliquée au capillaire métallique de nébulisation. Le gaz utilisé est de l'azote. Il est chauffé à 350°C dans la source d'ionisation. Le débit pour les « sheath gaz » est de 60 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 3 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C. Pour l'analyse des « Full Scan », la fenêtre d'analyse s'étend de 85 à 1000 m/z à une résolution de 240 000 (à m/z 200). L'« AGC target », qui permet de compter le nombre d'ions accumulés, est fixé à 5.10^4 ions et le temps maximal d'injection à 400 ms. Pour l'analyse des événements MS/MS, l'isolement des ions se fait au niveau du quadripôle avec une fenêtre d'isolement de 0.8 Da et la détection se fait avec l'Orbitrap à une résolution de 30 000 (à m/z 200). L'« AGC target » est fixé à 5.10^4 ions et le temps maximal d'injection à 54 ms. Une liste d'inclusion est créée puis importée avec le m/z du composé, une gamme de temps de rétention, la gamme d'acquisition ainsi que l'énergie de fragmentation. Dans les analyses courantes de notre laboratoire, les composés sont fragmentés à quatre énergies de collision en mode HCD : 10 %, 20 %, 40 % et 80 %. Compte tenu du nombre de standards et leur répartition au cours du temps d'analyse chromatographique ainsi que du temps nécessaire à l'acquisition des spectres MS et MS/MS, les standards ont été répartis en trois mélanges distincts afin d'obtenir suffisamment d'acquisitions pour chacun des événements (Tableau 11).

Afin de répondre au mieux aux critères de cette étude, de nouvelles acquisitions ont été également réalisées en mode FIA. En effet, afin d'éviter toute interférence et pour obtenir une base de données robuste, chaque standard est injecté individuellement. Cela permet de s'affranchir du couplage avec une méthode séparative. Dix microlitres d'échantillon concentré à 1 $\mu\text{g/mL}$ ont été injectés. Le débit de solvant a été fixé à 100 $\mu\text{L/min}$. Le temps total d'élution est alors de 3 min avec une proportion de phase mobile de 95/5 phase A/phase B tout au long de l'analyse. Comme précédemment, une liste d'inclusion avec le m/z du composé et ses énergies de fragmentation a été importée. Afin d'obtenir le spectre MS/MS le plus informatif, chaque standard a été fragmenté avec des énergies allant de 0 % à 100 % (avec un pas de 5 %) en mode HCD. Les gammes d'acquisition débutent toutes à 50 Da. Les autres paramètres sont identiques aux premières acquisitions de spectres MS/MS. Les spectres MS/MS obtenus sont comparés à une base de données acquise avec un Q-TOF. Cette comparaison est réalisée par le partenaire Autrichien.

c – Résultats

- Etude « SCORE »

Tout d'abord, le mélange de standards envoyé par le laboratoire a été analysé avec la méthode « HCD-only » détaillée précédemment. Les résultats obtenus au niveau de la détection des standards étaient globalement très bons avec la détection de 24 composés sur 25. Cependant, 12 des composés détectés n'étaient pas sélectionnés pour être fragmentés (Tableau 12).

Tableau 12 : Résultats de l'étude "SCORE" avec le paramètre « ¹²C Selection ».

Composé	Formule Brute	[M+H] ⁺	RT (min)	Sélection
Metformin	C4H11N5	130.10871	0.79	Oui
Caffeine	C8H10N4O2	195.08764	4.80	Oui
6-Mercaptopurine	C5H4N4S	153.02294	0.93	Oui
Ensulizole	C13H10N2O3S	275.04848	4.49	Oui
Dapson	C12H12N2O2S	249.06922	6.03	Oui
Bayrepep	C12H23NO3	230.17505	8.41	Oui
Carbetamide	C12H16N2O3	237.12336	7.15	Oui
Phénytoin	C15H12N2O2	253.09714	7.77	Oui
Carbamazépine	C15H12N2O2	237.10223	7.68	Oui
Metconazole	C17H22ClN3O	320.1524	10.14	Oui
Diazinon	C12H21N2O3PS	305.10831	10.76	Oui
Picoxystrobin	C18H16F3NO4	368.1104	10.64	Oui
Naproxen	C14H14O3	231.10156	8.79	Non
Chloridazon	C10H8ClN3O	222.0428	6.08	Non
Diclofenac	C14H11Cl2NO2	296.02395	9.89	Non
Monuron	C9H11ClN2O	199.06326	7.38	Non
Metobromuron	C9H11BrN2O2	259.0077	8.65	Non
Linuron	C9H10Cl2N2O2	249.0192	9.37	Non
Chlorbromuron	C9H10BrClN2O2	292.96874	9.53	Non
Alachlor	C14H20ClNO2	270.12552	10.11	Non
Profenofos	C11H15BrClO3PS	372.94246	10.60	Non
Boscalid	C18H12Cl2N2O	343.03993	9.71	Non
Quinoxifén	C15H8Cl2FNO	308.00396	11.24	Non
Fenofibrate	C20H21ClO4	361.1201	11.91	Non
Chlorpropham	C10H12ClNO2	214.06292	ND	NA

Un point commun à l'ensemble des composés détectés mais non sélectionnés est la présence d'au moins un atome d'halogène (chlore et/ou brome). Dans la méthode DDA « HCD-only », le paramètre « ¹²C Selection » permet de limiter la sélection d'isotopologues pour un même composé. Cependant, ce paramètre n'est efficace que pour les massifs isotopiques les plus courants ne comportant que des atomes non halogénés (¹²C/¹³C ; ¹⁶O/¹⁸O). Les composés halogénés ayant des massifs isotopiques plus spécifiques ne seraient alors pas

sélectionnés dans ces conditions. La suppression de ce paramètre impliquerait également la possible sélection de différents isotopologues d'un même métabolite (pour les composés les plus intenses). Cela peut donc limiter la fragmentation et donc l'identification de composés d'intérêt.

Les 25 standards ont donc été de nouveau analysés avec la méthode DDA « HCD-only » sans le paramètre « ^{12}C Selection ». Dans ce cas, 24 des 25 composés sont toujours détectés et 23 d'entre eux sont sélectionnés pour être fragmentés. Seul le naproxem n'est pas fragmenté car il n'est pas suffisamment intense. Ces données ont été envoyées au laboratoire partenaire pour comparaison des spectres MS/MS obtenus en mode « stepped collision energy » HCD 30 +/- 20 % avec leur base de données spectrale. Vingt des 23 composés fragmentés ont pu être correctement identifiés. Ces résultats sont très encourageants concernant l'intérêt de la fragmentation par « stepped collision energy » qui permet de gagner du temps au niveau développements méthodologiques (une seule injection réalisée sans connaissance préalable des énergies de fragmentation optimales) tout en gardant une bonne qualité spectrale. De plus, ce bon niveau d'identification montre que les spectres MS/MS acquis avec des spectromètres de masse de technologies différentes (Q-Orbitrap vs Q-TOF) sont directement comparables.

L'analyse de notre mélange de standards avec le paramètre « ^{12}C Selection » a permis la détection de 19 composés sur 20 (le resveratrol n'est pas détecté) et la sélection de 17 de ces composés. La suppression de ce paramètre a permis la sélection supplémentaire de la clindamycine. Seule l'indomycine n'est pas fragmentée en raison d'une trop faible intensité. Ces données ont également été envoyées au laboratoire partenaire qui a pu identifier 15 des 18 composés sélectionnés.

- « Study 1 »

Les échantillons de plasma de l'étude « Study 1 » ont été analysés avec la méthode « HCD-only » sans le paramètre « ^{12}C Selection ». Ces analyses ont été réalisées à l'aveugle. L'ensemble des données ont été traitées par le laboratoire partenaire. Dans ce cas, 17 des 19 composés du mélange 1 ont été fragmentés et 16 d'entre eux ont pu être identifiés. Le tramadolol n'a pas pu être identifié car l'ion fragment diagnostique à m/z 58 est en dehors de notre gamme d'acquisition débutant à 85 Da. Pour le deuxième set de standards, 25 des 26 composés ont été fragmentés et identifiés. Concernant l'échantillon de plasma supplémenté avec notre mélange de standards, les analyses DDA « HCD-only » ont été réalisées avec et sans le paramètre « ^{12}C Selection ». La suppression de ce paramètre a permis la sélection du clenbutérol. Au total, 16 des 20 composés sont détectés et 13 d'entre eux sont fragmentés (Tableau 13). L'allopurinol, la gabapentine ainsi que la metformin ne sont pas sélectionnés car ils ne sont pas assez intenses comparativement aux autres ions co-éluants. Sur les 13 composés fragmentés, 12 ont pu être correctement identifiés par le laboratoire partenaire. Etonnament, le spectre MS/MS obtenu pour la theophylline n'avait pas suffisamment de fragments communs avec le spectre de référence pour être correctement identifié, ceci potentiellement à cause de la co-sélection et la co-fragmentation d'un composé isobare dans nos conditions.

Tableau 13 : Résultats de l'étude "Study 1" avec notre mélange de standard.

Composé	Formule Brute	[M+H] ⁺	Avec Paramètre ¹² C Selection"		Sans Paramètre ¹² C Selection"	
			Détection	Sélection	Détection	Sélection
Adenosine	C10H13N5O4	268.10403	ND	NA	ND	NA
Allopurinol	C5H4N4O	137.04579	Oui	Oui	Oui	Non
Amantadine	C10H17N	152.14338	Oui	Oui	Oui	Oui
Cimétidine	C10H16N6S	253.12299	Oui	Oui	Oui	Oui
Clenbutérol	C12H18Cl2N2O	277.08690	Oui	Non	Oui	Oui
Dextrométhorphan	C18H25NO	272.20089	Oui	Oui	Oui	Oui
Diclofenac	C14H11Cl2NO2	296.02396	ND	NA	ND	NA
Dopamine	C8H11NO2	154.08626	ND	NA	ND	NA
Gabapentine	C9H17NO2	172.13321	Oui	Oui	Oui	Non
Hydrochlorothiazide	C7H8ClN3O4S2	297.97175	ND	NA	ND	NA
Lincomycine	C18H34N2O6S	407.22104	Oui	Oui	Oui	Oui
Metformine	C4H11N5	130.10873	Oui	Non	Oui	Non
Miansérine	C18H20N2	265.16993	Oui	Oui	Oui	Oui
Ofloxacine	C18H20FN3O4	362.15106	Oui	Oui	Oui	Oui
Pantoprazole	C16H15F2N3O4S	336.11542 (-SO)	Oui	Oui	Oui	Oui
Phéniramine	C16H20N2	241.16993	Oui	Oui	Oui	Oui
Sotalol	C12H20N2O3S	273.12674	Oui	Oui	Oui	Oui
Theophylline	C7H8N4O2	181.07201	Oui	Oui	Oui	Oui
Timolol	C13H24N4O3S	317.16419	Oui	Oui	Oui	Oui
Warfarin	C19H16O4	309.11214	Oui	Oui	Oui	Oui

- « Study 2 »

L'ensemble des 19 composés a pu être identifiée à partir de la base de données du laboratoire partenaire avec une acquisition selon la méthode couramment utilisée par notre laboratoire (avec 4 énergies de collision : 10 %, 20 %, 40 % et 80 % HCD). Cela montre que la stratégie de notre laboratoire consistant à enregistrer des données MS/MS pour les métabolites à partir de spectres de fragmentation obtenus avec uniquement quatre énergies de collision est pertinente. Le Tableau 14²⁶⁵ présente les résultats de la « Study 2 » obtenus par l'ensemble des laboratoires. Les résultats obtenus avec l'Orbitrap Fusion (Q-Orbitrap-LIT dans le tableau) montre de très bonnes performances comparativement aux autres instruments tout en ayant peu de spectres MS/MS acquis. Globalement, l'ensemble des instruments montre une bonne efficacité de détection par rapport au jeu de molécules testées.

Tableau 14 : Bilan des résultats inter-plateformes obtenus pour la "Study 2" (extrait de Oberacher et al.)

Laboratory	Instrument type	CE range	Number of spectra	Sensitivity [%] of the test set	Sensitivity [%] of the library
1	Q-Orbitrap-LIT	10–80%	60	95.0	92.8
2	Q-Orbitrap	10–80%	225	98.7	94.2
3	Q-Orbitrap	15–90%	82	73.2	78.5
4	LIT-Orbitrap with HCD	5–80%	236	81.8	94.7
5	LIT-Orbitrap with CID	10–130%	518	82.4	83.2
	LIT-Orbitrap with HCD	10–185%	343	93.4	93.8
6	QqTOF	5–50eV	150	94.0	79.9
7	QqTOF	5–50 eV	150	100	96.2

Afin d'avoir une base de données la plus robuste possible, il peut s'avérer nécessaire d'augmenter le nombre de spectres MS/MS acquis avec un maximum d'énergies de collision. De plus, au sein du laboratoire, lors d'analyses MS/MS ciblées, les gammes d'acquisitions sont adaptées à chaque composé afin de transmettre l'ensemble des ions et de détecter le plus de fragments possibles. Cependant, si la gamme d'acquisition débute à un m/z trop élevé (> 70 Da), certains ions diagnostiques peuvent ne pas être détectés (cas du tramadolol dans la « Study 1 »). De nouvelles données ont donc été acquises par une analyse chromatographique FIA, avec des énergies de fragmentation allant de 0 % à 100 % HCD avec un pas de 5 % et avec une plage de balayage commençant à 50 Da. Contre toute attente, l'identification des composés a diminué. En effet, avec cette étude il a été montré que la taille de la gamme de m/z suivie a un impact sur la transmission des ions. En effet, si la masse inférieure de cette gamme est plus de cinq fois inférieure à la masse du composé d'intérêt, les ions de plus hauts m/z seront peu ou pas transmis. Par exemple, pour une masse inférieure de 50 Da, seuls les composés de m/z inférieur à 250 Da seront correctement transmis. Ce problème de transmission d'ions n'est pas (ou est moins) présent sur les instruments de type Q-Exactive (Figure 42¹⁸³). Cet inconvénient de l'Orbitrap Fusion avait déjà été mis en évidence lors de tests préliminaires lors des développements associés aux méthodes DDA et DIA (voir article ci-dessus).

Partie B : Etudes Expérimentales – Chapitre 1

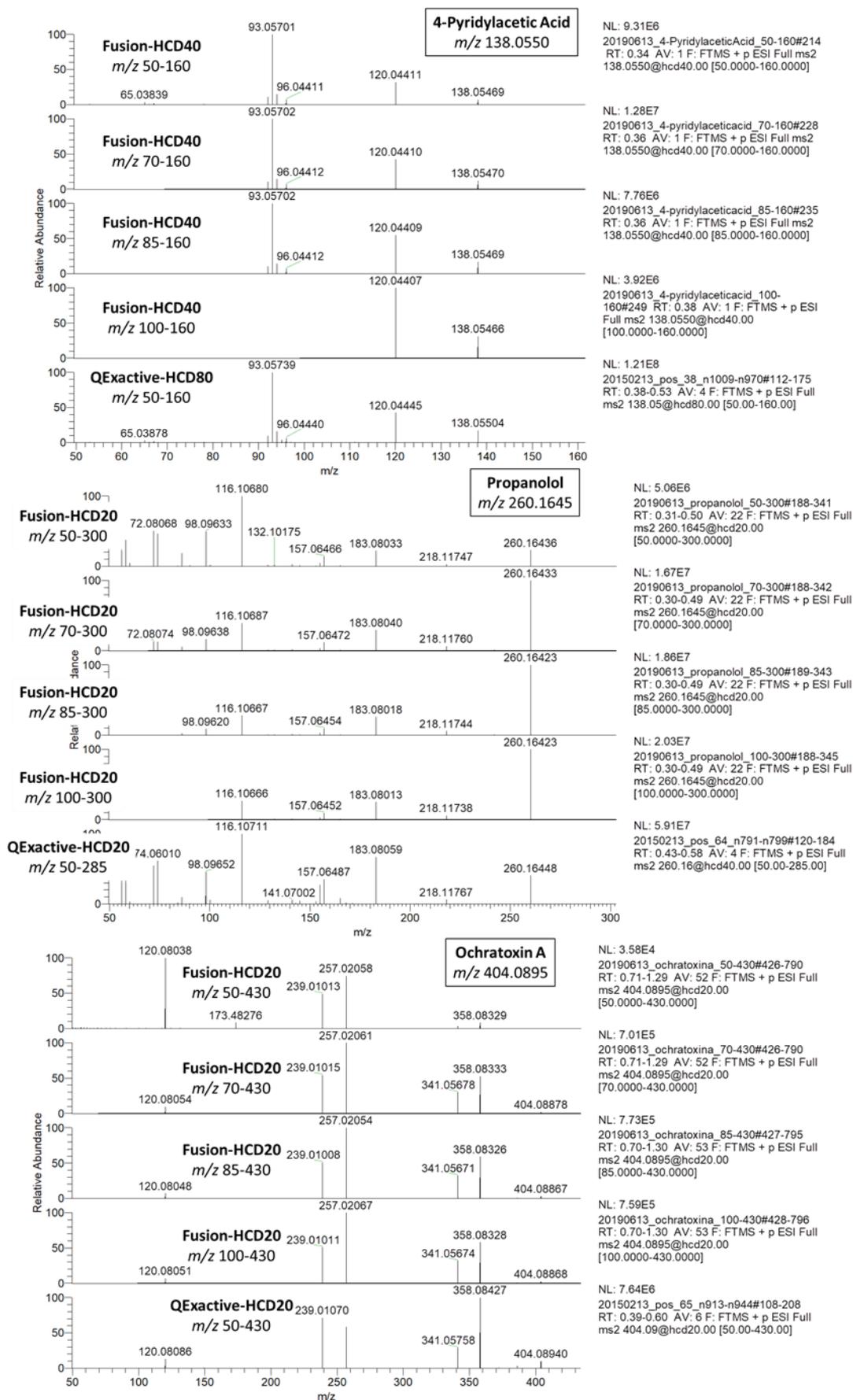


Figure 42 : Influence de la gamme d'acquisition sur les spectres de fragmentation (adaptée de Barbier Saint Hialire et al.).

Partie B : Etudes Expérimentales – Chapitre 1

La Figure 42 présente trois exemples d'influence de la gamme d'acquisition sur les intensités des ions fragments. Pour les composés de plus faible m/z (< 250 Da, exemple de l'acide 4 pyridylacétique), la variation de la gamme de masse n'influe pas sur l'intensité des ions (parents et fragments). Plus la valeur de m/z est élevée, plus cette variation a un impact sur l'intensité des fragments (cas du propanolol et de l'ochratoxin A). En effet, nous pouvons voir que lorsque la gamme d'acquisition débute à des m/z de plus en plus élevés, les signaux des hauts rapports m/z du spectre de fragmentation sont de plus en plus intenses jusqu'à se stabiliser lorsque le m/z du début de la gamme d'acquisition correspond à $1/5^{\text{ème}}$ du m/z de l'ion parent.

Les analyses ont donc également été effectuées avec une gamme d'acquisition commençant à 70 Da. Avec cette méthode, les 19 composés ont été identifiés.

d – Conclusions et perspectives

L'objectif de ce projet est à terme de créer une base de données spectrale Européenne afin de faciliter l'identification des métabolites pour l'ensemble des laboratoires. Pour cela, dans un premier temps, les études « SCORE », « Study 1 » et « Study 2 » avaient pour but d'évaluer les méthodes d'acquisition de spectres MS/MS de différents laboratoires. En effet, plusieurs centaines de standards vont devoir être analysés. Ce travail demande énormément de temps ce qui implique qu'un seul laboratoire ne peut pas à lui seul acquérir l'ensemble des données. Cette mission doit donc être partagée entre différents laboratoires. Cela nécessite que les spectres MS/MS soient comparables d'un laboratoire à l'autre et d'un spectromètre de masse à l'autre.

Les études « SCORE » et « Study 1 » ont prouvé que la méthode DDA « HCD-only » développée est efficace pour la détection et l'identification de composés d'origines variées. Moyennant quelques optimisations, cette méthode s'est avérée très efficace pour l'étude des xénobiotiques.

L'étude « Study 2 » a permis de s'assurer que la méthode d'acquisition des spectres MS/MS permettait d'obtenir des spectres de fragmentation robustes et comparables avec une base de données externe au laboratoire. La comparaison de spectres MS/MS obtenus par les différents laboratoires, avec des instruments de technologies différentes (Q-Orbitrap, Q-TOF, ...) a montré que les spectres MS/MS (tous obtenus en mode non résonnant) étaient comparables quel que soit l'instrument utilisé. Cette étude vient donc conforter les résultats obtenus sur des projets antérieurs concernant la comparabilité des spectres MS/MS acquis sur différents spectromètres de masse²⁶⁴. Cela a également permis l'établissement de critères d'assurance qualité et de contrôle qualité. Cette étude a permis la publication d'une proposition de critères d'assurance qualité et de contrôle qualité : « A European proposal for quality control and quality assurance of tandem mass spectral libraries » (Annexe 3²⁶⁵). En effet, l'ensemble des acquisitions devaient répondre à un cahier des charges initial que ce soit au niveau chromatographique avec l'obligation d'analyser les échantillons avec une colonne C18 ou au niveau de l'acquisition des données avec l'utilisation d'un analyseur à haute résolution, des gammes de balayage et des énergies de fragmentation bien définies. Avec ces critères, l'ensemble des laboratoires

partenaires devrait pouvoir acquérir de nombreux spectres MS/MS et ainsi créer une grande base de données spectrale au niveau européen utilisable par, a minima, tous les membres du consortium.

C – Automatisation et amélioration de la méthode « Data Dependent Acquisition »

1 – Contexte et objectifs

Les analyses effectuées avec les méthodes DDA développées ont montré une certaine reproductibilité dans la sélection des ions. Un même échantillon analysé plusieurs fois avec la même méthode DDA donnera un bon recouvrement des ions sélectionnés et une fragmentation comparable pour toutes les analyses. Cependant, seuls les ions les plus intenses sont généralement identifiés dans ces conditions. Pour augmenter le nombre de composés identifiés au cours de multiples analyses, des listes d'exclusion peuvent être réalisées à partir des ions identifiés lors des analyses précédentes. Dans nos conditions, ces listes sont faites manuellement comme la liste d'exclusion correspondant aux ions de l'échantillon blanc et la liste d'inclusion correspondant aux composés couramment identifiés dans une matrice donnée. Ces deux dernières listes sont réalisées grâce à des analyses réalisées préalablement aux analyses d'intérêt. La création de ces listes est très consommatrice de temps pour l'expérimentateur, et doivent être re-validées avant chaque nouvelle série d'analyses.

Le constructeur Thermo Fisher Scientific a récemment développé un logiciel (nommé Acquire X) permettant d'automatiser la méthode DDA de façon itérative. Pour cela, une liste d'exclusion est automatiquement implémentée à la méthode DDA après une première analyse de l'échantillon blanc en mode « Full Scan ». Ensuite, l'échantillon d'intérêt est également analysé en mode « Full Scan » et le logiciel crée automatiquement une liste d'inclusion à partir de tous les ions détectés dans cet échantillon. Lors de la première analyse DDA, tous les ions les plus intenses sont fragmentés selon les critères DDA sus-mentionnés. Pour la deuxième analyse, toutes les variables (m/z et RT) des ions fragmentés lors la première analyse DDA sont implémentés automatiquement dans une liste d'exclusion ce qui diminue d'autant la taille de la liste d'inclusion. Cela permet alors de fragmenter des ions d'intensités plus faibles. Cette itération est réalisable plusieurs fois jusqu'à la fragmentation de tous les ions présents dans l'échantillon (Figure 43).

Partie B : Etudes Expérimentales – Chapitre 1

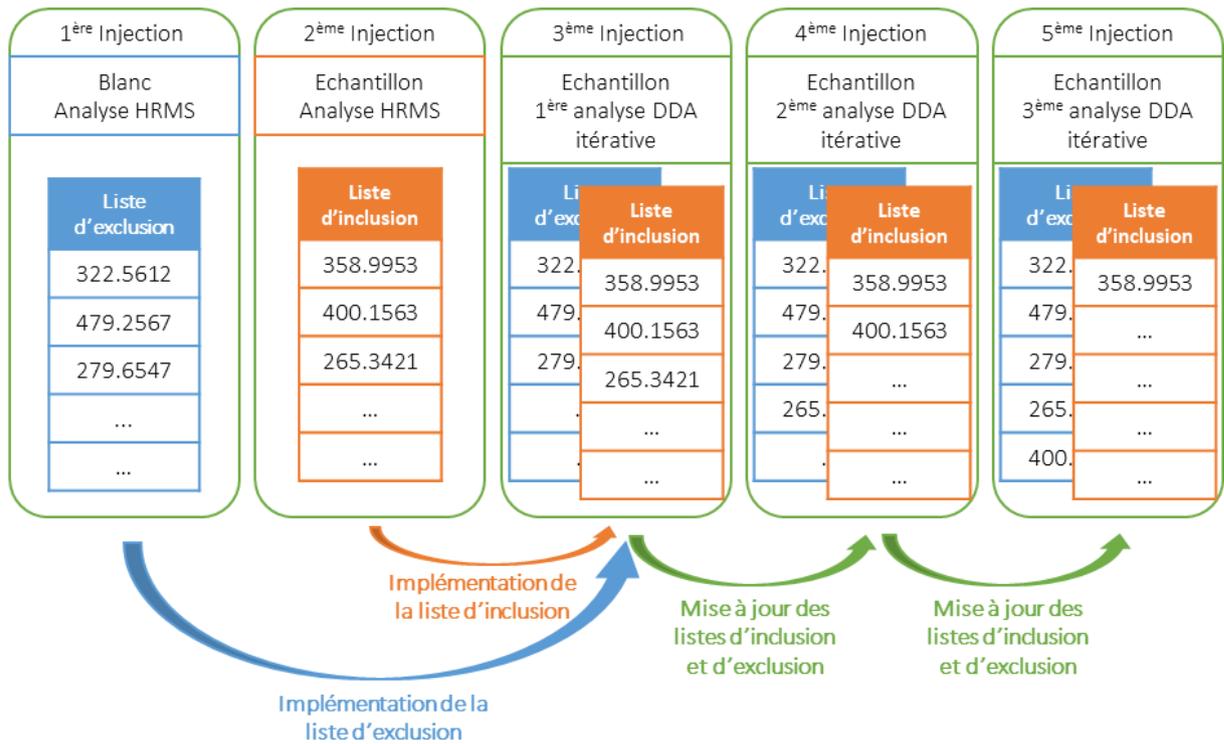


Figure 43 : Principe de la méthode DDA itérative.

Cette méthodologie de travail a été étudiée en collaboration avec Thermo Fisher Scientific (San Jose, Etats-Unis). Dans un premier, nous avons utilisé un échantillon de plasma NIST supplémenté avec le mélange de standards utilisé précédemment pour le développement des méthodes DDA et DIA. Grâce à la méthode DDA itérative, 419 composés ont été identifiés au total (188 dans la 1^{ère} injection, 91 dans la 2^{ème} injection, 67 lors de la 3^{ème} injection et 73 lors de quatrième injection) sur leur instrument (Orbitrap ID-X Tribrid). A noter que les métabolites ont été arbitrairement considérés comme identifiés quand un « DP score » > 500 était obtenu (logiciel MS-DIAL, base de données : MoNA ; voir ci-dessous). Par comparaison avec la méthode DDA conventionnelle, seuls 135 composés sont identifiés (Figure 44). Ces résultats montrent l'intérêt d'une telle stratégie d'analyse.

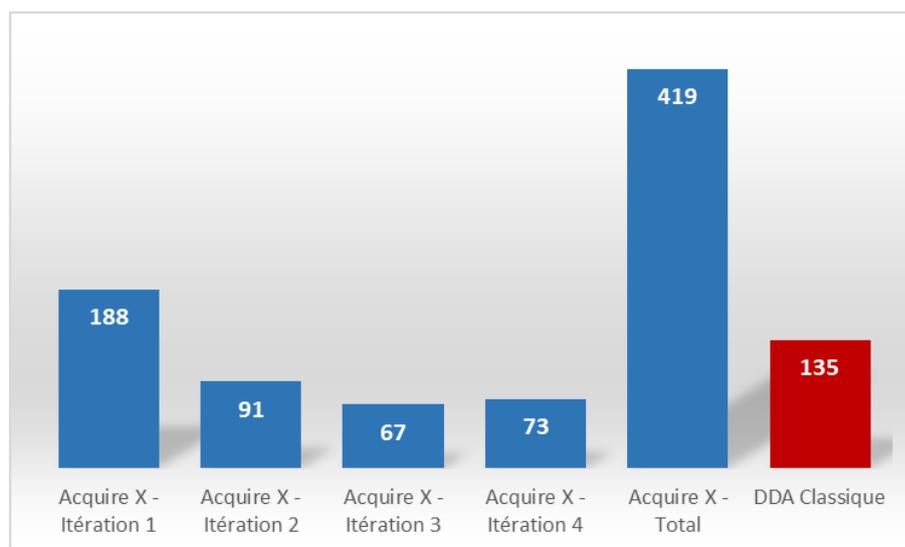


Figure 44 : Nombre de composés identifiés par la méthode Acquire X et DDA classique.

Nous avons donc souhaité implémenter une telle approche DDA itérative au sein de laboratoire afin d'optimiser la fragmentation et l'identification des métabolites présents dans l'échantillon d'intérêt sans connaissance préalable de celui-ci.

2 – Méthodologie

Les développements ont été réalisés sur le même mélange de standards que celui utilisé pour le développement des méthodes DDA et DIA. La méthode chromatographique C18 et les paramètres de source ESI positif sont les mêmes que ceux utilisés pour les développements DDA et DIA. Pour la création des listes d'exclusion et d'inclusion, le blanc et l'échantillon d'intérêt sont analysés en mode « Full Scan ». La détection est réalisée avec l'Orbitrap à une résolution de 500 000 (à m/z 200) avec une gamme de masse de 85 à 1000 Da. Pour les acquisitions DDA, les spectres MS sont acquis avec l'Orbitrap à une résolution de 120 000 (à m/z 200) sur une gamme de masse pré-filtrée par le quadripôle de 85 à 1000 Da. Les spectres MS/MS sont également acquis avec l'Orbitrap à une résolution de 15 000 (à m/z 200) avec une masse de début de gamme de 85 Da. Les ions sont sélectionnés avec le quadripôle avec une fenêtre d'isolement de 0.8 Da et sont fragmentés en mode HCD « stepped collision energy » à 30 +/- 20 %. Les ions sont sélectionnés avec un paramètre de seuil d'intensité fixé à $2.5 \cdot 10^4$. Les analyses ont été réalisées avec ou sans « dynamic exclusion ». Lorsque celle-ci est activée, les ions sont exclus pendant quatre secondes s'ils ont été fragmentés deux fois dans la même seconde. Les paramètres de liste d'inclusion et d'exclusion ont été ajoutés avec des listes réalisées avec le couple m/z et RT de l'ion détecté. L'inclusion ou l'exclusion est réalisée avec une erreur maximale de trois ppm par rapport au

m/z de référence détecté. Le nombre d'événements MS/MS entre les acquisitions MS est fixé à 10 (top-10) (Figure 45).

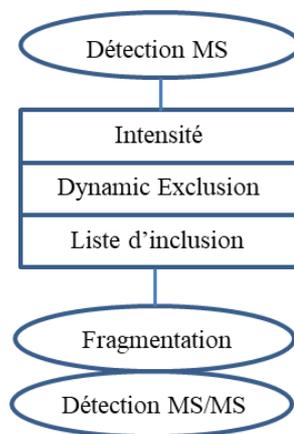


Figure 45 : Méthode DDA Acquire X.

Un facteur de dérogation de l'exclusion a été fixé à trois. C'est-à-dire que si un ion présent dans la liste d'exclusion est détecté dans l'échantillon d'intérêt à une intensité trois fois supérieure à celle du blanc, cet ion est sélectionné pour être fragmenté car il est considéré comme provenant majoritairement de cet échantillon. Les ions préférentiellement fragmentés sont les $[M+H]^+$.

Afin d'évaluer l'efficacité de cette méthode, des échantillons de plasma NIST supplémentés avec le même mélange de standards ont été analysés. Les concentrations finales de ce mélange dans le plasma sont de 0 – 0.1 – 1 – 10 ng/mL. Ces analyses ont été réalisées en triplicat.

3 – Résultats

Dans un premier temps, le mélange de standards a été analysé à 1 $\mu\text{g/mL}$ avec la méthode DDA « itérative » avec et sans « Dynamic Exclusion ». A cette concentration, les 47 composés sont détectés et fragmentés dès la première analyse DDA. Ces analyses n'ont pas montré de différences significatives avec et sans « Dynamic Exclusion » au niveau du nombre d'ions présents dans les listes d'inclusion et d'exclusion. Au bout de trois injections en mode DDA, seuls les ions d'intensité très faible (inférieur au seuil d'intensité de la méthode DDA) restent à fragmenter.

Ce mélange de standards a ensuite été supplémenté aux concentrations suivantes : 0 – 0.1 – 1 et 10 ng/mL au plasma NIST pour évaluer la méthode sur une matrice plus complexe. Initialement, la liste d'inclusion comporte en moyenne 1500 variables contre 440 lors de la deuxième analyse DDA et 340 à la troisième analyse DDA (Figure 46).

Partie B : Etudes Expérimentales – Chapitre 1

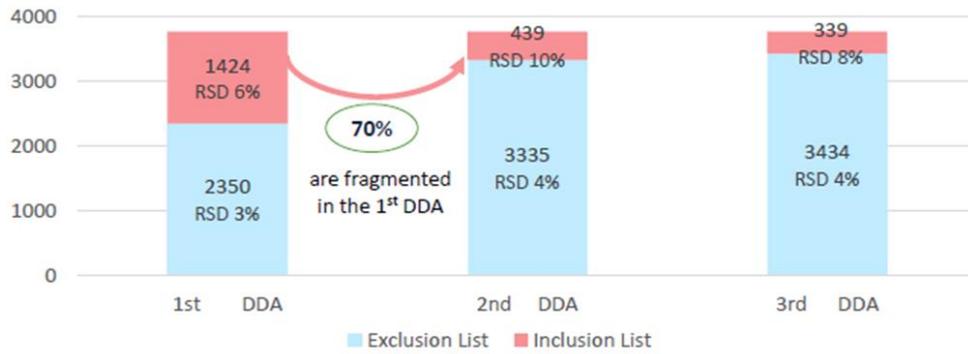


Figure 46 : Constitution itérative des listes d'inclusion et d'exclusion.

Cela signifie qu'environ 70 % des variables sont fragmentées dès la première analyse DDA que ce soit avec ou sans le paramètre « Dynamic Exclusion ». Les ions sélectionnés lors de la deuxième et la troisième analyse DDA sont principalement ceux présents dans le volume mort où de nombreux ions co-éluent. Cependant, concernant cette méthode DDA « itérative » utilisée, pour le plasma vierge, seuls 118 composés sont annotés lors de la première analyse MS/MS et aucune des variables fragmentées lors des deuxièmes et troisième injection ne sont annotées. Nous supposons que cet écart important en terme de nombre de métabolites annotés pour les injections 2 et 3 en comparaison des tests réalisés avec Thermo Fisher Scientific est la présence de contaminants dans notre cas (Figure 47). Ces derniers étant très intenses, une perte de sensibilité est possible au niveau des autres métabolites ce qui peut limiter leur sélection pour être fragmentés.

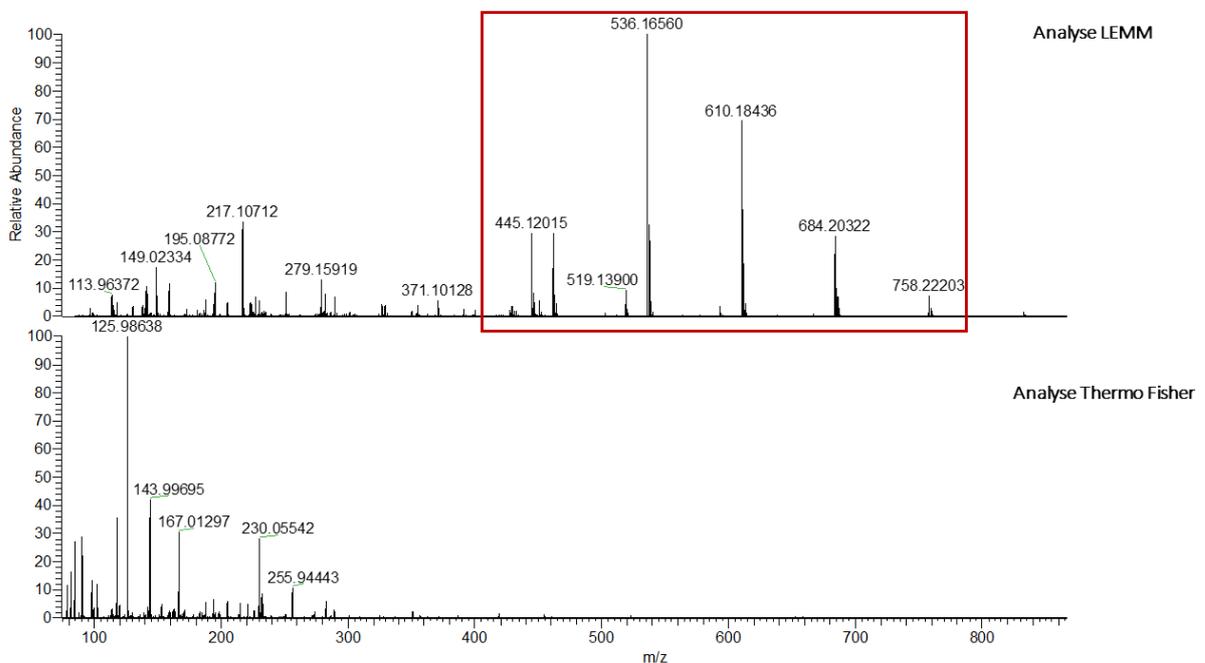


Figure 47 : Comparaison des spectres MS obtenus par notre laboratoire et par ThermoFisher Scientific.

Les annotations ont été réalisées avec MS-DIAL (« DP score » > 500 considéré) à l'aide de la base de données MoNA qui référence les spectres MS/MS de plus de 60 000 composés acquis sur un Q-TOF. Le traitement de données a également été initié avec le logiciel Compound Discoverer (Thermo Fisher Scientific)^{266,267}. Il s'agit dans ce cas d'un logiciel sous licence constructeur. La prise en main de ce logiciel est plus complexe que MS-DIAL et les paramètres de traitement doivent encore être optimisés.

Ces développements préliminaires ont été effectués lors du stage de trois mois de Leticia Lacalle-Bergeron, doctorante Espagnole que j'ai encadrée, et nécessitent encore d'être poursuivis. Ces travaux préliminaires ont permis la présentation d'un poster au congrès « 1st Iberian Meeting in separation science and mass spectrometry » en 2019.

4 – Conclusion et perspectives

Les résultats obtenus avec la méthode d'acquisition DDA itérative (telle qu'implémentée de façon préliminaire au laboratoire) sont très similaires à ceux obtenus avec la méthode DDA avec l'arbre décisionnel ramifié développé auparavant (voir article plus haut). La collaboration avec la société Thermo Fisher Scientific de San Jose nous a permis de voir que la méthode DDA itérative permettrait d'annoter beaucoup plus de composés qu'avec les méthodes DDA classiques. Cette approche a pour principal avantage de faire gagner un temps considérable grâce à la création automatique des listes d'inclusion et d'exclusion. Cependant, nos premiers résultats montrent que des phases d'optimisation sont encore nécessaires afin de disposer d'une méthode efficace et applicable en routine au laboratoire. Des premiers tests ont été réalisés en ajustant l'« AGC Target » ainsi que les « RF lens » qui permettent de focaliser les ions après la source. Avec ces nouveaux paramètres, l'abondance relative de certains ions a augmenté significativement diminuant ainsi l'impact des contaminants sur nos analyses. Ces tests préliminaires devront être poursuivis pour optimiser au mieux cette méthodologie d'acquisition.

D – Conclusions et perspectives

Les méthodes DDA et DIA développées ont montré de très bonnes performances pour l'acquisition simultanée des spectres MS et MS/MS sans perte de sensibilité au niveau du profilage MS. L'application au projet européen HBM4EU a permis d'évaluer les limites de l'approche DDA, tout en permettant son optimisation et démontrant son efficacité. Des résultats préliminaires mais très prometteurs soulignent également l'apport d'une approche DDA « itérative » pour l'annotation d'un plus grand nombre de métabolites humains plasmatiques.

L'application de ces méthodes à une cohorte médicale permettrait d'évaluer leur utilité pour les analyses métabolomiques de routine. Pour cela, il serait intéressant d'analyser l'ensemble des échantillons avec

la méthode DIA de type « SWATH » décrite précédemment et d'analyser des contrôles qualité avec la méthode DDA « itérative ». Grâce à cette dernière, l'analyse de l'échantillon se ferait en profondeur et permettrait potentiellement d'identifier des composés encore inconnus et la méthode DIA de type « SWATH » permettrait d'identifier et de quantifier les métabolites d'intérêt. Combiner les informations obtenues au niveau des ions précurseurs et fragments par DIA pourrait également permettre d'améliorer la comparaison de profils métaboliques²⁶⁸.

Chapitre 2 : Production et utilisation de standards internes marqués (SIL-IS) pour la quantification en métabolomique

A – Production de standards internes marqués (SIL-IS) par une méthode mono-étape métallo-catalysée

1 – Contexte et objectifs

Classiquement en métabolomique, la quantification absolue des métabolites est réalisée à l'aide de standards internes marqués avec des isotopes stables. Ces isotopes stables sont le plus souvent du carbone-13, de l'azote-15 ou encore du deutérium. Cependant, tous les métabolites connus ne possèdent pas un isotopologue marqué disponible commercialement. De plus, une partie des standards internes marqués disponibles reste encore très coûteux ce qui limite leur utilisation systématique.

Comme illustré dans la Partie 1 – Chapitre 3 de cette thèse, la synthèse chimique est une solution attrayante pour la production à façon de ses propres molécules marquées pour répondre, le plus souvent, à des besoins de quantification ou de normalisation lors d'analyse métabolomique par LC-HRMS. Cependant, la production de standards marqués par synthèse totale ou multi-étapes est généralement longue et fastidieuse. Une alternative consiste à explorer les réactions de marquage direct. Ces réactions sont le plus souvent effectuées pour introduire des atomes de deutérium par réaction d'échange hydrogène/deutérium (H/D) à des positions stables de la structure chimique de la molécule, par opposition aux réactions d'échanges H/D sur des hydrogènes labiles portés le plus souvent par des hétéroatomes comme l'oxygène ou l'azote.

L'objectif de cette étude a été de produire facilement et à grande échelle des standards internes marqués avec des isotopes stables par échange hydrogène/deutérium. Pour cela, une approche de marquage au deutérium mono-étape métallo-catalysée par activation de la liaison C-H a été explorée.

2 – Principe de l'introduction de deutérium par réaction mono-étape métallo-catalysée

Les réactions de deutération métallo-catalysée pour la production de molécules marquées ont été assez largement décrites et font intervenir le plus souvent des métaux tels que le palladium, le platine ou encore le ruthénium sous forme de catalyseurs homogènes ou hétérogènes^{269–272} (voir partie introductive, chapitre 3 pour plus de détails). La réaction à laquelle nous nous sommes particulièrement intéressée repose sur l'utilisation de ruthénium comme catalyseur et a été assez largement investiguée par un laboratoire partenaire au CEA pour le marquage stéréosélectif et stéréorétentif de divers composés biologiquement actifs : médicaments, acides aminés, peptides...^{237,243}. Cette réaction de deutération repose sur une activation de la liaison C-H en alpha de fonctions amine ou alcool par coordination du ruthénium aux substrats à deutérer *via* le doublet non liant de leurs atomes d'azote ou d'oxygène. L'échange des atomes d'hydrogène par des atomes de deutérium en alpha des groupements amino ou hydroxyl est ainsi permis en présence d'un solvant deutéré, tel que D₂O, et de D₂ gaz qui jouent le rôle de sources d'atome de deutérium. Ainsi, le principal avantage de ce procédé de deutération, qui a été mis en évidence par nos collègues du CEA sur des acides aminés, sont les caractères régio- et stéréospécifique de l'incorporation de deutérium (Figure 48)²⁴³.

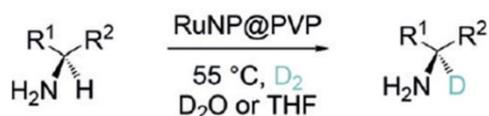


Figure 48 : Régiospécificité et stéréospécificité de la réaction d'échange H/D mono-étape métallo-catalysée (extraite de Taglang et al.).

Np : nanoparticule ; *PVP* : polyvinylpyrrolidone ; *THF* : tétrahydrofurane

Du point de vue de la production de standards internes marqués pour la métabolomique quantitative par LC-HRMS, les deux avantages majeurs de ce procédé sont le caractère mono-étape de la synthèse et le fait que le composé (ou substrat) de départ soit le métabolite d'intérêt. Ainsi, le marquage au deutérium de séries de métabolites semble envisageable de façon relativement simple et rapide.

3 – Plan expérimental

Sur la base des nombreuses études métabolomiques préalablement réalisées au laboratoire, une liste d'une cinquantaine de métabolites a pu être définie. En effet, les métabolites détectés de façon récurrente lors de ces études dans les conditions analytiques employées en routine au laboratoire semblent constituer une bonne base de départ pour l'investigation du procédé de deutération préalablement présenté. Parmi ces métabolites, se trouve principalement des acides aminés et dérivés, des acides organiques ainsi que quelques

polyols et molécules hétéroaromatiques. La liste des 52 composés choisis ainsi que leur structures sont présentées ci-dessous (Tableau 15 et Figure 49).

Tableau 15 : Liste des composés investigués pour la réaction d'échange H/D mono-étape métallo-catalysée (adapté de la publication en cours de rédaction).

Compound	N°	Formula	Accurate mass	Smiles	Chemical class
α -Amino adipic Acid	1	C6H11NO4	161.0688	<chem>O=C(O)C(N)CCCC(O)=O</chem>	Amino acid
2-Aminobutyric Acid	2	C4H9NO2	103.0633	<chem>CCC(C(O)=O)N</chem>	Amino acid
3-Amino-isobutyric Acid	3	C4H9NO2	103.0633	<chem>CC(C(O)=O)CN</chem>	Amino acid
Guanidinoacetic Acid	4	C3H7N3O2	117.0538	<chem>N=C(NCC(O)=O)N</chem>	Organic acid
Guanidinosuccinic Acid	5	C5H9N3O4	175.0593	<chem>N=C(N)NC(CC(O)=O)C(O)=O</chem>	Organic acid
3-Hydroxy-3-Methylglutaric acid	6	C6H10O5	162.0528	<chem>CC(CC(O)=O)(O)CC(O)=O</chem>	Organic acid
3-Methylhistamine	7	C6H11N3	125.0953	<chem>CN1C=NC=C1CCN</chem>	Organic amine
3-Methylhistidine	8	C7H11N3O2	169.0851	<chem>CN1C=NC=C1CC(C(O)=O)N</chem>	Amino acid
7-Methylguanine	9	C6H7N5O	165.0651	<chem>CN1C=NC(NC(N)=N2)=C1C2=O</chem>	Heteroaromatic
Adipic Acid	10	C6H10O4	146.0579	<chem>O=C(O)CCCC(O)=O</chem>	Organic acid
Alanine	11	C3H7NO2	89.0477	<chem>CC(C(O)=O)N</chem>	Amino acid
Arginine	12	C6H14N4O2	174.1117	<chem>NC(CCCNC(N)=N)C(O)=O</chem>	Amino acid
Asparagine	13	C4H8N2O3	132.0535	<chem>NC(CC(N)=O)C(O)=O</chem>	Amino acid
Aspartic Acid	14	C4H7NO4	133.0375	<chem>NC(CC(O)=O)C(O)=O</chem>	Amino acid
Citric Acid	15	C6H8O7	192.0270	<chem>O=C(O)CC(C(O)=O)(O)CC(O)=O</chem>	Organic acid
Citrulline	16	C6H13N3O3	175.0957	<chem>NC(CCCNC(N)=O)C(O)=O</chem>	Amino acid
Diglycolic Acid	17	C4H6O5	134.0215	<chem>O=C(O)COCC(O)=O</chem>	Organic acid
Glutaconic Acid	18	C5H6O4	130.0266	<chem>O=C(O)\C=C\CC(=O)O</chem>	Organic acid
Glutamic Acid	19	C5H9NO4	147.0532	<chem>NC(CCC(O)=O)C(O)=O</chem>	Amino acid
Glutamine	20	C5H10N2O3	146.0691	<chem>NC(C(O)=O)CCC(N)=O</chem>	Amino acid
Glycerol-3-Phosphate	21	C3H9O6P	172.0137	<chem>OCC(COP(O)(O)=O)O</chem>	Polyol & derivative
Glycine	22	C2H5NO2	75.0320	<chem>NCC(O)=O</chem>	Amino acid
Histidine	23	C6H9N3O2	155.0695	<chem>NC(CC1=CN=CN1)C(O)=O</chem>	Amino acid
Hydroxyglutaric Acid	24	C5H8O5	148.0372	<chem>C(CC(=O)O)C(C(=O)O)O</chem>	Organic acid
Imidazolelactic Acid	25	C6H8N2O3	156.0535	<chem>OC(C(O)=O)CC1=CN=CN1</chem>	Organic acid
Isocitric Acid	26	C6H8O7	192.0270	<chem>O=C(O)CC(C(O)=O)C(O)C(O)=O</chem>	Organic acid
Lactic Acid	27	C3H6O3	90.0317	<chem>CC(O)C(O)=O</chem>	Organic acid
Leucine	28	C6H13NO2	131.0946	<chem>CC(C)CC(C(O)=O)N</chem>	Amino acid
Lysine	29	C6H14N2O2	146.1055	<chem>NCCCC(N)C(O)=O</chem>	Amino acid
Maleic Acid	30	C4H4O4	116.0110	<chem>O=C(O)/C=C\C(O)=O</chem>	Organic acid
Malic Acid	31	C4H6O5	134.0215	<chem>OC(CC(O)=O)C(O)=O</chem>	Organic acid
Methylimidazoleacetic Acid	32	C6H8N2O2	140.0586	<chem>CN(C=N1)C=C1CC(O)=O</chem>	Organic acid
Methylmalonic Acid	33	C4H6O4	118.0266	<chem>CC(C(O)=O)C(O)=O</chem>	Organic acid
N-Methyl-D-Aspartic Acid	34	C5H9NO4	147.0532	<chem>CNC(CC(O)=O)C(O)=O</chem>	Amino acid
O-Acetylserine	35	C5H9NO4	147.0532	<chem>CC(OCC(C(O)=O)N)=O</chem>	Amino acid derivative
Ornithine	36	C5H12N2O2	132.0899	<chem>NCCCC(C(O)=O)N</chem>	Amino acid
Pantothenic Acid	37	C9H17NO5	219.1107	<chem>CC(CO)(C(O)C(NCCC(O)=O)=O)C</chem>	Organic acid
Phosphoethanolamine	38	C2H8NO4P	141.0191	<chem>NCCOP(O)(O)=O</chem>	Organic amine
Pipecolic Acid	39	C6H11NO2	129.0790	<chem>O=C(O)C1CCCCN1</chem>	Amino acid
Proline	40	C5H9NO2	115.0633	<chem>O=C(C1CCCN1)O</chem>	Amino acid
Pyroglutamic Acid	41	C5H7NO3	129.0426	<chem>O=C1CCC(N1)C(O)=O</chem>	Organic acid
Sarcosine	42	C3H7NO2	89.0477	<chem>CNCC(O)=O</chem>	Amino acid
Serine	43	C3H7NO3	105.0426	<chem>NC(CO)C(O)=O</chem>	Amino acid
Sorbitol	44	C6H14O6	182.0790	<chem>OCC(O)C(O)C(O)C(O)CO</chem>	Polyol & derivative
Succinic Acid	45	C4H6O4	118.0266	<chem>O=C(O)CCC(O)=O</chem>	Organic acid
Taurine	46	C2H7NO3S	125.0147	<chem>NCCS(O)(=O)=O</chem>	Organic amine
Taurocholic Acid	47	C26H45NO7S	515.2917	<chem>OC1C2C3CCC(C(CCC(NCCS(O)(=O)=O)=O)=O)C)C3(C)C(O)CC2C4(C)CCC(O)CC4C1</chem>	Bile acid derivative
Threonic Acid	48	C4H8O5	136.0372	<chem>OCC(C(C(O)=O)O)O</chem>	Organic acid
Threonine	49	C4H9NO3	119.0582	<chem>CC(C(C(O)=O)N)O</chem>	Amino acid
4-Hydroxyproline	50	C5H9NO3	131.0582	<chem>OC1CNC(C(O)=O)C1</chem>	Amino acid
Tyrosine	51	C9H11NO3	181.0739	<chem>NC(C(O)=O)CC1=CC=C(C=C1)O</chem>	Amino acid
Valine	52	C5H11NO2	117.0790	<chem>CC(C)C(C(O)=O)N</chem>	Amino acid

Partie B : Etudes Expérimentales – Chapitre 2

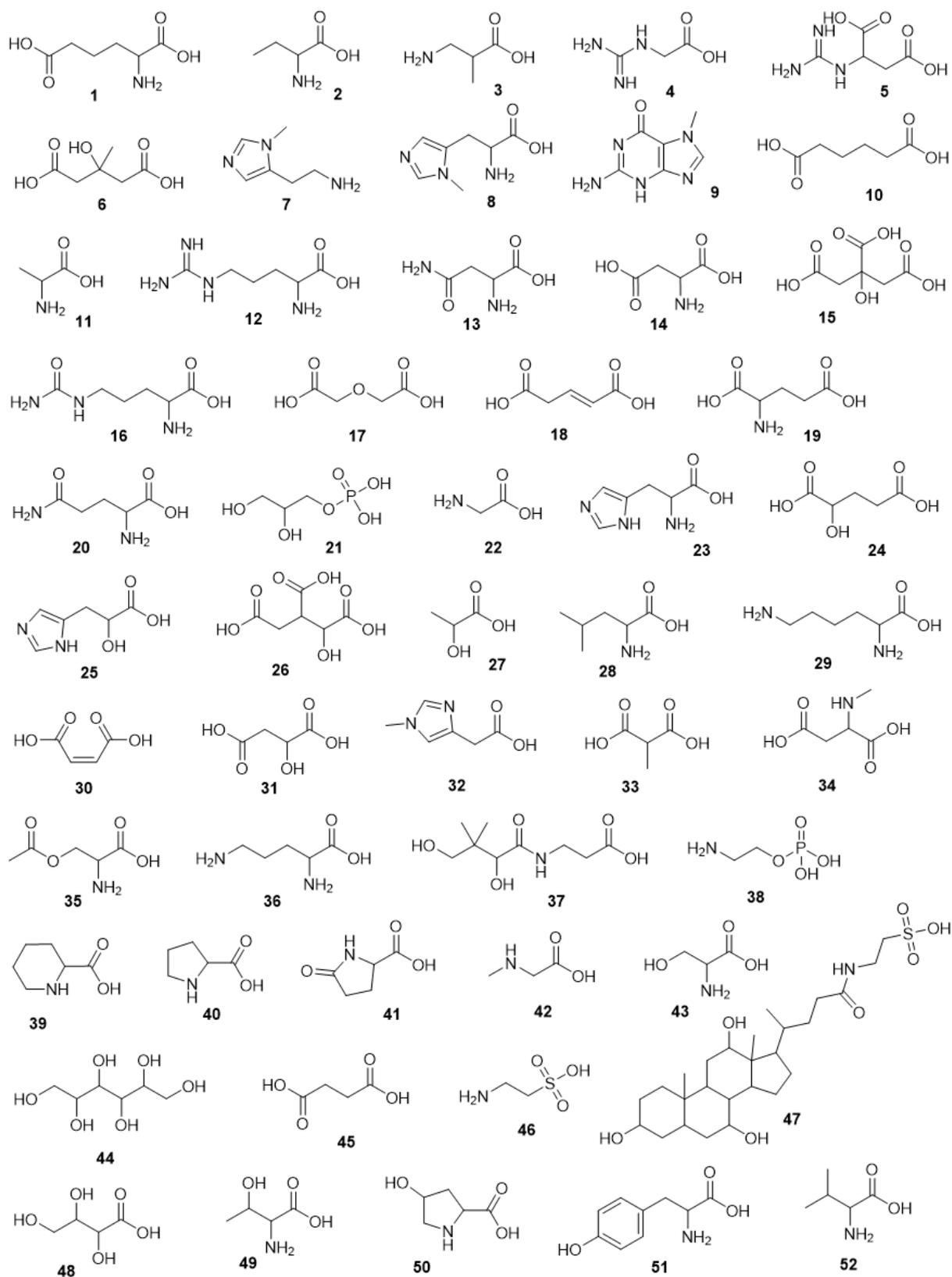


Figure 49 : Structures de 52 composés investigués pour la réaction d'échange H/D mono-étape métallo-catalysée (adaptée de la publication en cours de rédaction).

Les efficacités de marquage ont été évaluées systématiquement par LC-HRMS en se fixant des critères d'évaluation (cahier des charges) en cohérence avec l'utilisation future de ces composés marqués en tant que standards internes. Pour cela, nous avons considéré que deux atomes de deutérium devaient au minimum être incorporés dans la structure pour éviter toute superposition avec le massif isotopique du composé non-marqué, et en particulier, la contribution naturelle de l'isotopologue C-13 du composé non marqué. La distribution isotopique doit également être la plus étroite possible (c'est-à-dire la plus homogène) avec idéalement un isotopologue très majoritaire comptant pour plus de 90 % du massif isotopique total afin d'avoir une pureté isotopique se rapprochant des composés commerciaux, et d'éviter d'introduire des effets matrices délétaires consécutivement à l'ajout de quantités importantes de standards marqués afin de pallier la multiplicité d'isotopologues. Il ne faut pas ou peu de dégradation lors du marquage de façon à pouvoir récupérer un produit brut marqué directement utilisable en tant que SI sans avoir à le purifier. Le marquage doit être stable : c'est-à-dire qu'une fois incorporés, les atomes de deutérium ne doivent pas de nouveau s'échanger avec des atomes d'hydrogène (Figure 50).

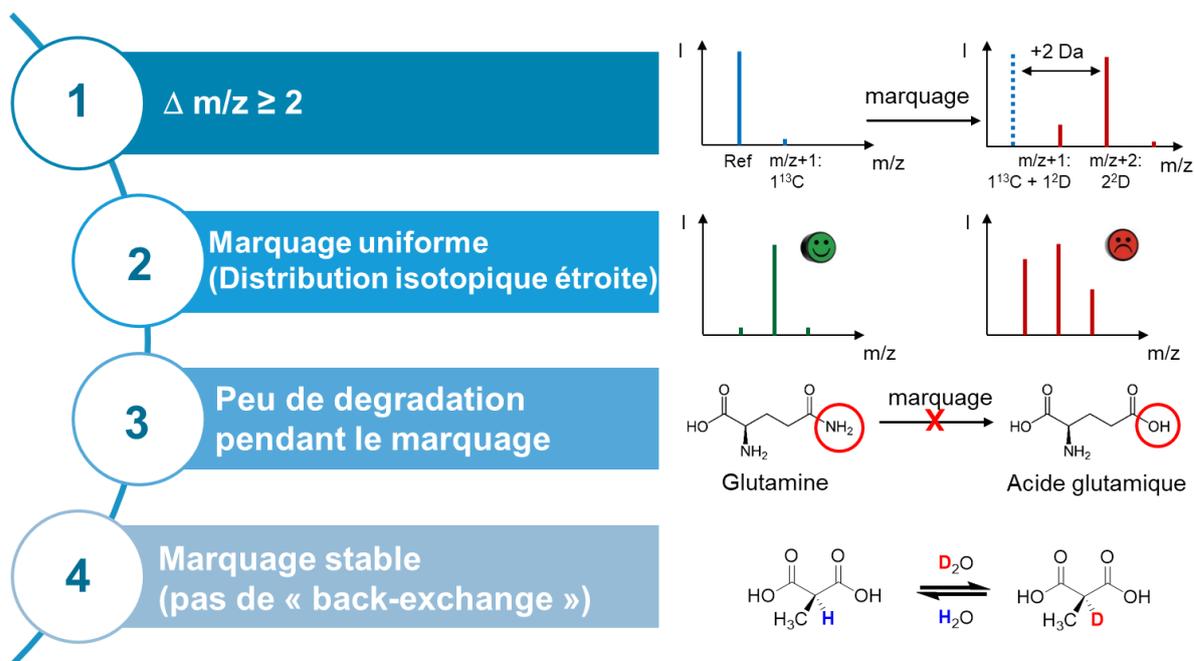


Figure 50 : Cahier des charges pour la production de standards internes marqués.

Les paramètres de la réaction d'échange H/D mono-étape métallo-catalysée ont donc été optimisés pour répondre à ce cahier des charges. Un certain nombre des composés de l'étude (41 molécules sur 52) présentent les fonctions amines et/ou alcools adaptées à l'incorporation de deutérium selon le procédé développé par le laboratoire partenaire au CEA. Pour les autres composés, tels que les acides organiques, ne possédant pas les groupements requis pour la chélation du ruthénium, un procédé de marquage analogue a été envisagé en employant comme catalyseur du palladium en remplacement du ruthénium, cependant beaucoup moins chimio- et régio-sélectif. De plus, des ajustements et/ou modifications dans les conditions réactionnelles originellement décrites par Taglang et al. et Pieters et al. ont été envisagées dans l'idée de favoriser

l'incorporation de deutérium au sein des diverses structures chimiques^{237,243}. Ainsi, un arbre décisionnel comme présenté ci-dessous a pu être appliqué (Figure 51).

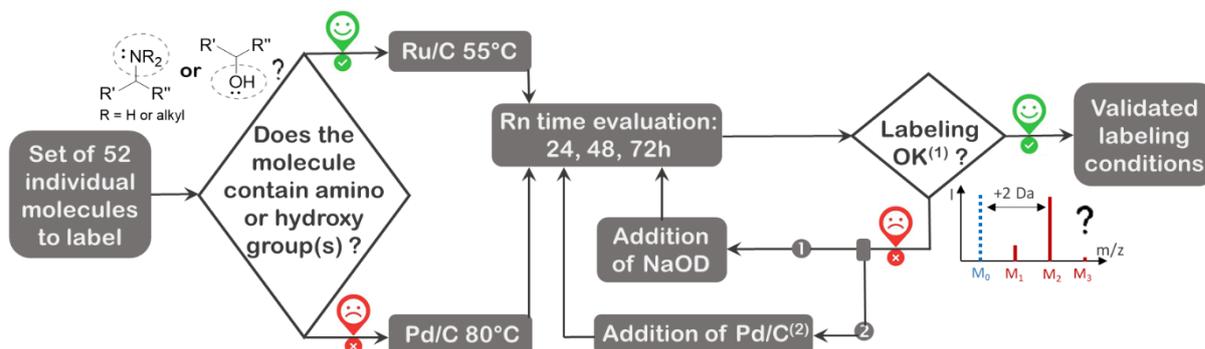


Figure 51 : Arbre décisionnel pour le développement des échanges H/D métallo-catalysés (adaptée de la publication en cours de rédaction).

Enfin, il est important de noter que ces réactions d'échange H/D sont traditionnellement réalisées à l'aide d'un système, nommé tube de Fisher-Porter et présenté en Figure 52(A), spécifiquement dédié aux réactions d'hydrogénation catalytique car permettant une mise sous atmosphère de D₂ (ou de H₂). Ce dispositif présente l'avantage par rapport aux autoclaves en acier de permettre un suivi visuel de la réaction. Cependant, ce dispositif ne permettant la mise en œuvre que d'une réaction à la fois et compte tenu du nombre important de composés que nous avons prévu d'étudier et des ajustements de conditions opératoires envisagés qui en découlent, la mise en place et l'utilisation d'un système alternatif, permettant la parallélisation de 24 réactions simultanées ont été évaluées. Ainsi, nous avons fait l'acquisition du système manuel de type autoclave CAT 24, illustré Figure 52 (B), commercialisé par HEL pour l'évaluation rapide et multiplexée de plusieurs catalyseurs et de multiples conditions expérimentales (température, pH...).

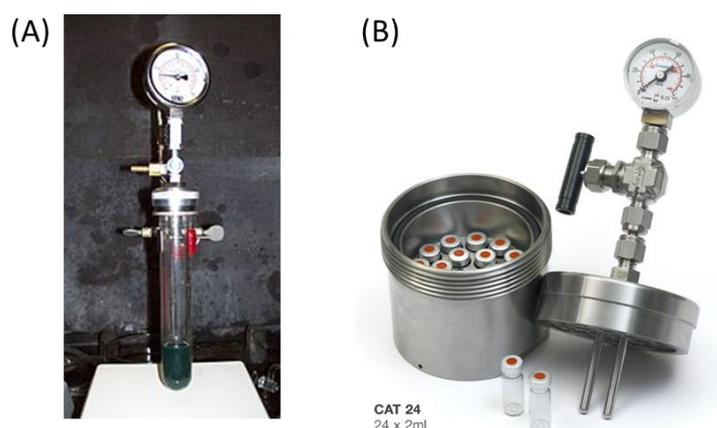


Figure 52 : Dispositifs utilisés pour les réactions d'échange H/D métallo-catalysées sous pression de D₂.

(A) Tube de Fisher-Porter (https://en.wikipedia.org/wiki/Fisher-Porter_tube)

(B) Dispositif CAT24 commercialisé par HEL (<https://helgroup.com/products/catalytic-processes/catalyst-screening-choice-of-multi-sample-manual-systems>)

L'ensemble des réactions de marquage ont été réalisées expérimentalement dans le laboratoire partenaire (SCBM/LMT : Laboratoire de marquage au Tritium dans le Service de Chimie Bioorganique et de Marquage) sur le site de Saclay dans la mesure où ils disposent des équipements, installations et réactifs (sorbonne de synthèse, bouteille de D₂ et système de mise sous atmosphère de D₂ adapté, solvants et réactifs deutérés, etc), nécessaires au bon déroulement des réactions d'échange H/D. Une grande partie de ces premiers développements est issue de travaux de deux étudiantes de Master 2 (Yu Min Kiw et Jessica Michieletto) accueillies successivement au LEMM. L'ensemble des analyses LC-HRMS réalisées pour évaluer les efficacités de marquage ainsi que l'ensemble des mises au point et développements analytiques pour l'application à la métabolomique quantitative ont, quant à eux, été réalisés dans mon laboratoire d'accueil.

4 – Matériel et méthodes

- Synthèse

Cinquante-deux composés (Sigma Aldrich, Saint Quentin Fallavier, France) correspondants à des métabolites d'intérêt ont été testés. Le protocole général de mise en œuvre de la réaction d'échange H/D sur le système CAT24 est décrit dans la publication, en cours de rédaction, incluse ci-après dans le manuscrit, et a été adapté de la publication Taglang et al et Pieters et al.^{237,243}. Au total, quarante-six conditions de synthèse ont pu être évaluées grâce au réacteur CAT 24 (HEL, Equilabo, France) en faisant varier la nature du catalyseur, la température, le pH, le temps de réaction ainsi que la charge catalytique dans le but d'arriver à un marquage optimal (Annexe 4).

- Préparation des échantillons pour évaluation des efficacités de marquage

Dix à vingt microlitres du milieu réactionnel ont été directement prélevés à l'aide d'une seringue et dilués dans 500 µL d'eau. La solution obtenue a ensuite été filtrée pour éliminer le catalyseur avec une seringue Acridisc® (membrane PVDF, diamètre 25 mm, taille de pores 0.45 µm). Ces solutions ont été diluées dans la phase mobile pour obtenir une concentration finale pour injection de 1 à 5 µg/mL. Les standards non marqués ont également été injectés pour comparer les massifs isotopiques ainsi que les réponses chromatographiques avec ceux des produits marqués.

- Analyses LC-HRMS

Tous les composés marqués produits ainsi que leurs homologues non-marqués ont été analysés par chromatographie liquide couplée à la spectrométrie de masse à haute résolution. Deux méthodes séparatives ont été utilisées avec le système chromatographique Dionex Ultimate 3000 (Thermo Fisher Scientific). Tout d'abord, les composés ont été analysés avec une colonne Discovery® HS-F5 PFPP, 2.1 x 250 mm, 5 µm (Supelco, Merck, Darmstadt, Allemagne) pour la partie chromatographique. La phase mobile A est constituée d'eau avec

Partie B : Etudes Expérimentales – Chapitre 2

0.1 % d'acide formique et la phase B est composée d'acétonitrile avec 0.1 % d'acide formique. Le débit est fixé à 250 $\mu\text{L}/\text{min}$ et le compartiment à colonne est maintenu à une température de 30°C. Le gradient est le suivant : de 0 à 2 min gradient isocratique à 2 % de phase B, de 2 à 10 min gradient linéaire de 2 % à 5 % de phase B, de 10 à 16 min gradient linéaire de 5 % à 35 % de phase B, de 16 à 20 min gradient linéaire de 35 % à 100 % de phase B, de 20 à 24 min gradient isocratique à 100 % de phase B et de 24 à 30 min gradient isocratique à 2 % de phase B. Les composés ont également été analysés avec une colonne SeQuant® ZIC-p-HILIC, 2.1 x 150 mm, 5 μm (Merck, Darmstadt, Allemagne). Le compartiment à colonne est maintenu à une température de 15°C et le débit est fixé à 200 $\mu\text{L}/\text{min}$. La phase A est une solution aqueuse à 10 mM de carbonate d'ammonium ajustée à pH 10.5 avec de l'hydroxyde d'ammonium et la phase B est uniquement composée d'acétonitrile. Le gradient est le suivant : de 0 à 2 min gradient isocratique à 80 % de phase B, de 2 à 12 min gradient linéaire de 80 % à 40 % de phase B, de 12 à 17 min gradient isocratique à 0 % de phase B. Une phase d'équilibrage de 25 min à 80 % de phase B est ensuite réalisée.

Les analyses par spectrométrie de masse à haute résolution ont été réalisées avec un Q-TOF Impact HD (Bruker, Wissembourg, France) et un Orbitrap Fusion Tribrid (Thermo Fisher Scientific, Courtaboeuf, France). Tout d'abord, concernant l'Impact HD, une source ESI est utilisée avec les paramètres suivants : pour les analyses en mode d'ionisation positif, une tension de +4000 V est appliquée au capillaire maintenu à une température de 200 °C. Le débit de gaz est de 10 L/min. La pression du nébuliseur est fixé à 2.1 bar. Les paramètres d'acquisition sont les suivants : la gamme de masse étudiée est de 50 à 400 m/z et la fréquence d'acquisition est de 2 Hz. Pour les analyses MS, une énergie de collision de 4 eV est appliquée au niveau du quadripôle et une énergie de 2 eV est appliqué au niveau de la cellule de collision pour le transfert des ions. Pour les analyses en mode d'ionisation négatif, les paramètres de source sont les suivants : la tension du capillaire est fixée à -4500V avec une température de 200°C. Le débit de gaz est fixé à 8 L/min et la pression de nébuliseur est de 1.7 bar. Une énergie de 8 eV est fixée au niveau du quadripôle et une énergie de 2 eV est appliquée dans la cellule de collision. Les spectres MS sont acquis de 50 à 400 m/z à une fréquence de 1 Hz. Concernant l'Orbitrap Tribrid Fusion, une source ESI est également installée avec les paramètres suivants : une tension de + 4000 V ou – 2800 V est appliquée au capillaire métallique de nébulisation qui est chauffé à 100°C, le débit pour les « sheath gaz » est de 40 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 1 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C. Les acquisitions sont réalisées en mode « Full Scan », avec une fenêtre de balayage de 50 à 250 m/z à une résolution de 240 000 (à m/z 200). L'« AGC target », est fixé à $5 \cdot 10^4$ ions et le temps maximal d'injection à 400 ms.

Les données ont été traitées avec les logiciels constructeurs : DataAnalysis v.4.2 (Bruker) pour l'Impact HD et Xcalibur v.4.1 (ThermoFisher Scientific) pour l'Orbitrap Fusion.

5 – Résultats

Les tests de marquage ont été réalisés par série de 24 réactions en parallèle grâce au dispositif CAT24 selon le schéma général présenté en Figure 53. Les positions de deutération spécifiées correspondent aux positions de marquage attendues pour les quatre exemples de la Figure 53.

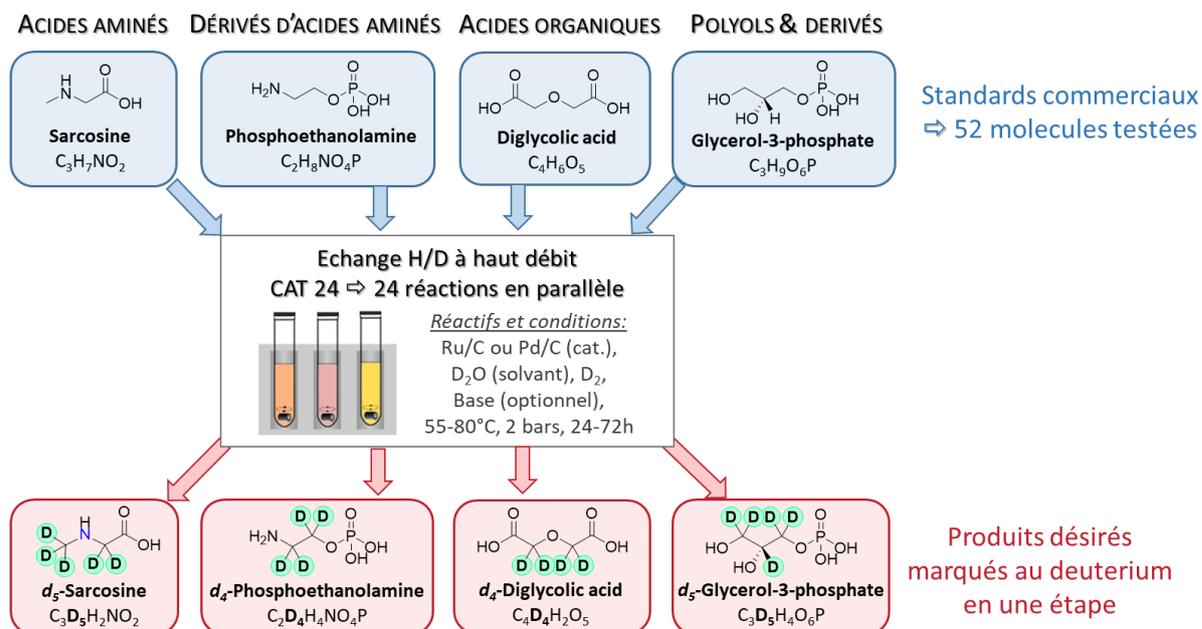


Figure 53 : Principe des investigations réalisées sur le dispositif CAT24 avec la réaction d'échange H/D mono-étape métallo-catalysée.

Les résultats de marquage obtenus sont présentés dans la suite du document de façon globale, les investigations réalisées pour chaque composé ne seront pas décrites au cas par cas. L'objectif étant de répondre au cahier des charges fixé initialement (voir partie A – 3 de ce chapitre) et en particulier d'obtenir des produits marqués ne présentant plus d'isotopologue non-deutéié (M0) et si possible présentant une incorporation minimum de deux atomes de deutérium. De plus, une distribution isotopique homogène est préférée. Les analyses LC- HRMS ont permis également de réaliser une première évaluation de la stabilité des produits marqués obtenus lors de la réaction de marquage. Enfin, la stabilité du marquage a pu également être appréciée en analysant le même échantillon à plusieurs heures, jours, voire semaines d'intervalle.

a - Utilisation du ruthénium comme catalyseur de l'incorporation de deutérium

Dans un premier temps, nous avons testés les conditions décrites par Pieters et al. et Taglang et al.^{237,243} en remplaçant les nanoparticules de ruthénium RuNp@PVP (Np : nanoparticule, PVP : polyvinylpyrrolidone) par le catalyseur commercial Ru/C, plus facile à se procurer et à manipuler. Des prélèvements du milieu réactionnel ont été réalisés à 24h, 48h et 72h. Dans ces conditions, seuls les composés

avec des fonctions amine et alcool ont été testés. L'analyse LC-HRMS des composés de référence et des produits de la réaction de marquage, et la comparaison de leur massif isotopique respectif, permet de mettre en évidence que pour 23 des composés testés, ces conditions sont celles qui ont été retenues car, à défaut de répondre totalement au cahier des charges initial, elles ont permis d'obtenir la meilleure incorporation de deutérium. Pour 10 d'entre eux, l'isotopologue monoisotopique M0 a totalement disparu (ou a une intensité relative inférieure à 1 %). Les résultats obtenus pour l'acide 2-aminobutyrique et pour l'acide thréonique sont donnés en exemple Figure 54. Pour les 13 autres composés, l'abondance de l'isotopologue monoisotopique M0 est non négligeable.

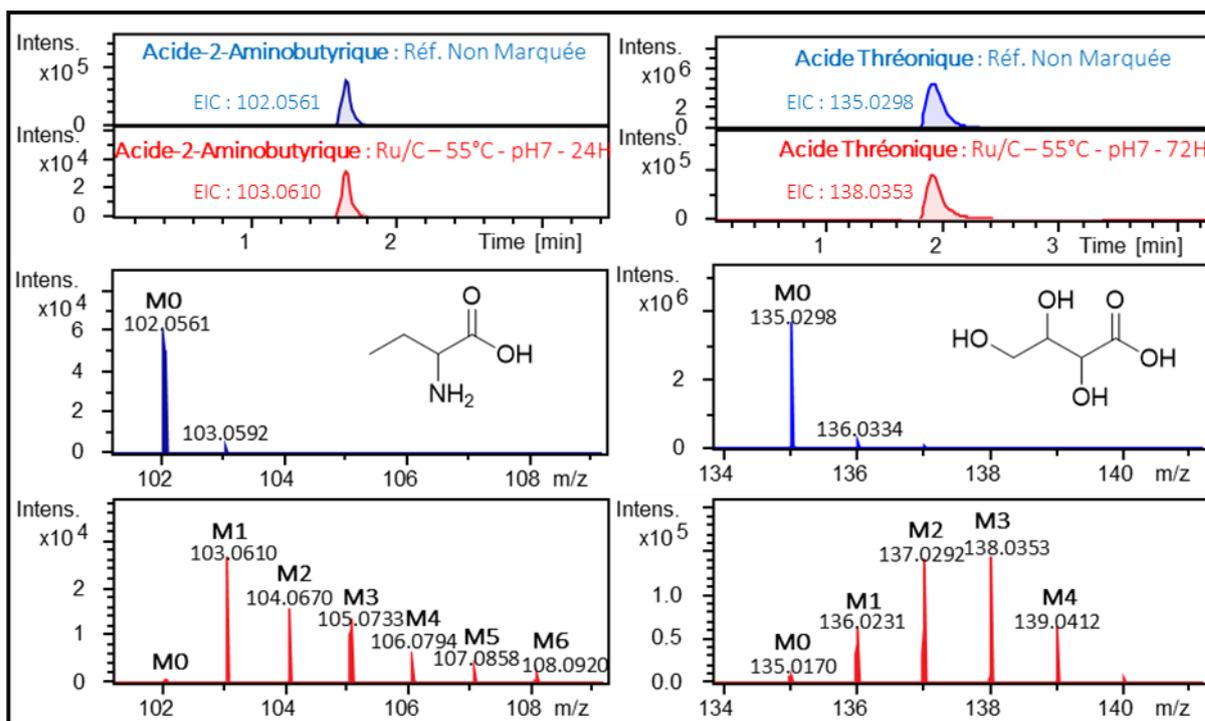


Figure 54 : Exemples de massifs isotopiques obtenus après marquage par échange H/D avec Ru/C

Ces exemples mettent en évidence la co-élution du standard non-marqué et du composé marqué. Cette propriété est souvent remise en cause dans le cas de composés deutérés. En effet, pour des composés avec une incorporation d'un grand nombre de deutériums, des effets isotopiques impliquant un décalage des temps de rétention peuvent être constatés. Dans notre cas, nous avons étudié des composés de faible masse moléculaire avec assez peu de deutériums incorporés ce qui, d'après nos observations, n'a pas eu d'impact sur les temps de rétention chromatographiques. L'analyse du produit de marquage de l'acide 2-aminobutyrique a également mis en évidence une perte partielle de régiospécificité avec Ru/C pour ce composé puisque, même si l'isotopologue majoritaire reste l'isotopologue M1 (correspondant probablement au marquage en alpha de l'amine), on observe la formation non négligeable d'isotopologues plus « lourds » (de M2 à M6).

Les paramètres de durée de réaction et nombre de cycles réactionnels ont été évalués pour les différents composés. En effet, chaque prélèvement à 24h et/ou 48h imposent un dégazage du CAT24. Les produits ont été remis en réaction par simple rechargement de l'autoclave en D₂ et remise à température du système. Globalement, il s'est avéré que, dans les conditions opératoires utilisées, le temps de réaction n'avait

pas un impact majeur sur les profils isotopiques obtenus pour les produits marqués. Néanmoins, nous avons pu constater qu'une recharge en D₂ (c'est-à-dire, un 2^{ème} cycle catalytique) pouvait conduire à une sensible augmentation du taux d'incorporation de deutérium et ainsi favoriser l'obtention d'une distribution isotopique plus homogène comme présenté en Figure 55 pour l'ornithine et le sorbitol.

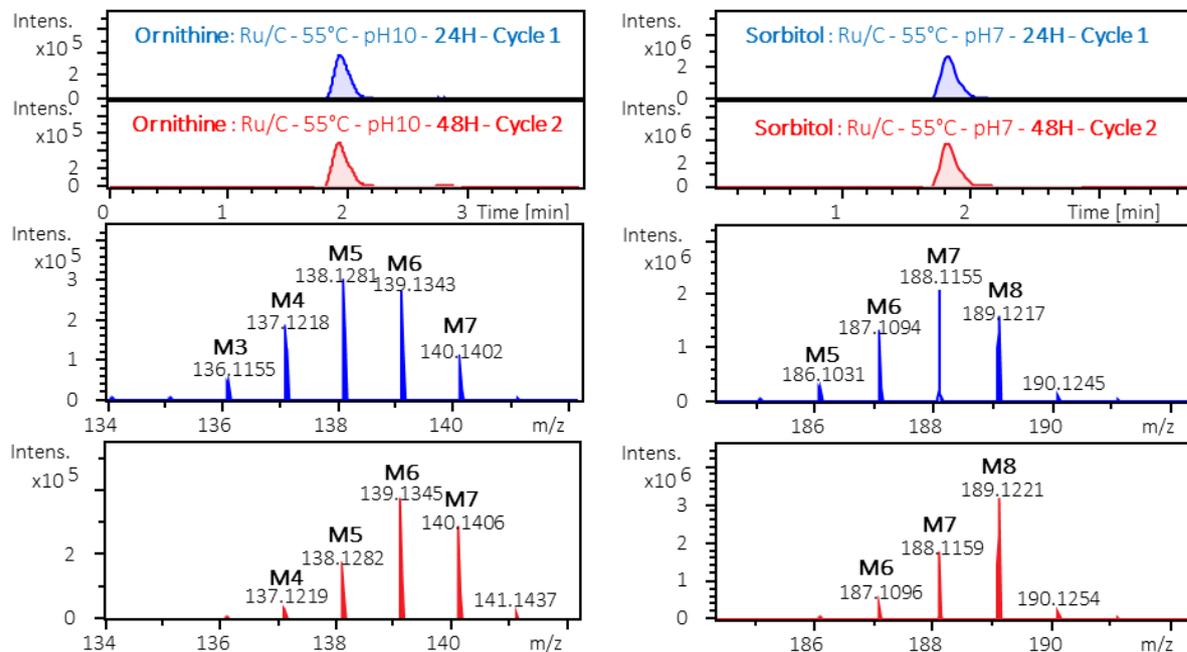


Figure 55 : Influence du temps de réaction (et du nombre de rechargement en D₂) pour la réaction d'échange H/D mono-étape métallo-catalysée.

Des tests complémentaires, réalisés sur un système de chimie en flux continu nommé H-Cube® (ThalesNano Inc.)²⁷³ disponible au SCBM/LMT (dispositif et protocole non détaillé ici), ont permis de suivre la cinétique d'incorporation des atomes de deutérium sur certains acides aminés. En bref, une solution d'acide aminé dans D₂O est mise à circuler en boucle dans le système H-Cube comportant une cartouche de Ru/C et intégrant un générateur de D₂ (produit par électrolyse de D₂O). L'exemple de suivi cinétique du marquage de la thréonine est présenté Figure 56 où des prélèvements et analyses LC-HRMS ont été réalisés à des intervalles de temps réguliers.

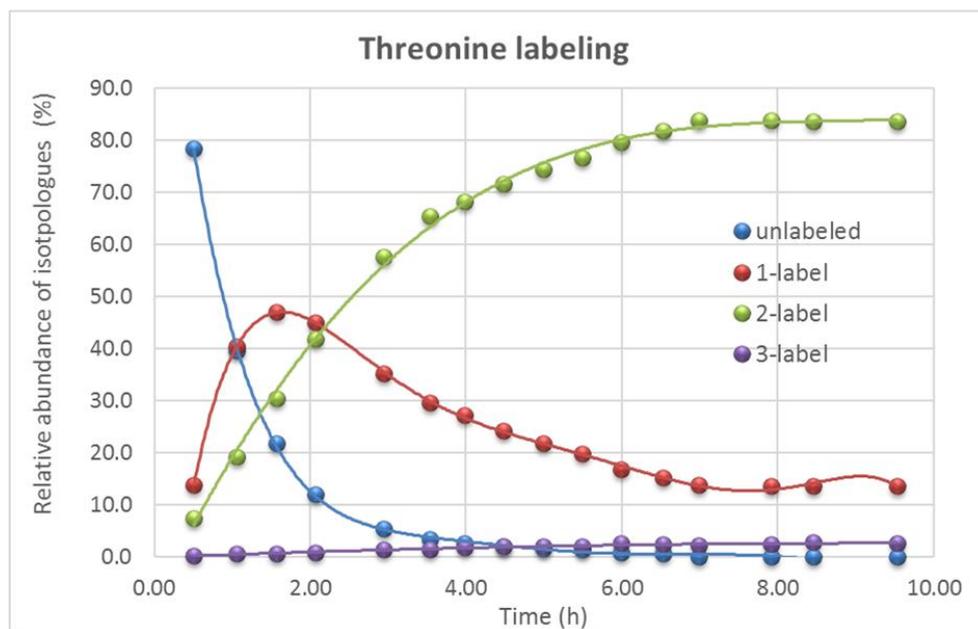
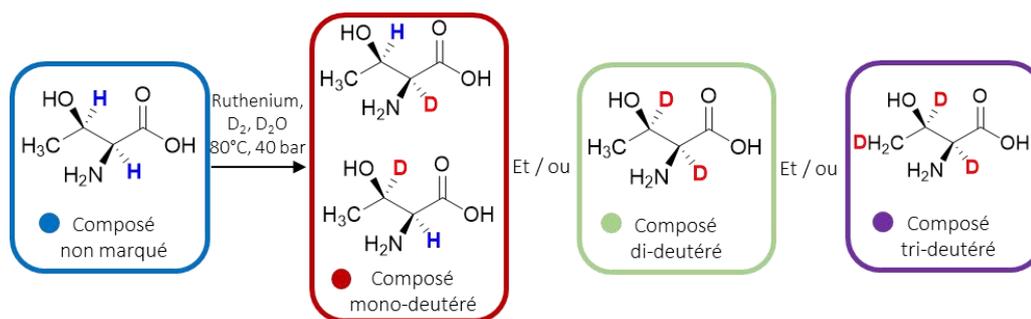


Figure 56 : Cinétique d'incorporation des atomes de deutériums par chimie en flux continu.

Cet exemple montre que l'incorporation des atomes de deutérium est assez rapide aux positions de marquage attendues avec Ru/C (en alpha des groupements amino et hydroxyl) jusqu'à atteindre un plateau au bout de sept heures de réaction. Avec le réacteur CAT 24, cette étude cinétique avec des intervalles de temps aussi rapprochés n'est, techniquement, pas possible. En effet, pour chaque prélèvement le système doit être ouvert. Il faut donc remettre le système sous D_2 gaz et de nouveau chauffer le système pour reprendre la réaction. Des prélèvements sont donc généralement effectués au bout de 24h, 48h et 72h ou 96h. Globalement, l'augmentation du temps de réaction sur CAT24 permet soit d'améliorer le taux d'incorporation de deutérium soit d'affiner le massif isotopique (Figure 55) probablement du fait de la recharge en D_2 gaz.

Même si les conditions de marquage avaient été préalablement optimisées par Taglang et al.²⁴³ spécifiquement pour le marquage de acides aminés avec Ru@NpPVP, avec notamment une température optimale de réaction de $55^\circ C$, nous avons évalué l'impact d'une augmentation de la température de la réaction à $80^\circ C$ dans le but d'augmenter les taux d'incorporation de deutérium dans les structures. Cependant, ces tests n'ont pas permis de mettre en évidence une amélioration de l'efficacité de marquage.

Il est intéressant de noter que pour quelques composés ayant été bien marqués avec Ru/C et pour lesquels nous disposions de suffisamment de matière (> 10 mg), des analyses par spectroscopie RMN ^1H ont été réalisées afin de confirmer les positions de marquage. C'est le cas de la 4-hydroxy-proline dont les spectres RMN ^1H avant et après marquage au deutérium sont présentés et comparés en Figure 57. Dans ce cas, seuls les hydrogènes portés par les carbones en alpha de la fonction amine ont été substitués par des atomes de deutérium (disparition des signaux RMN ^1H correspondant), alors qu'aucune incorporation de deutérium détectable n'est constatée sur le carbone qui porte la fonction alcool probablement en raison de facteurs liés à l'encombrement stérique. L'acquisition de spectres RMN proton et deutérium sera systématiquement envisagée à court terme et ce, en complément des analyses HRMS pour caractériser les lots marqués dans le cas où on passerait d'une phase d'investigation de marquage à une phase de production à plus grande échelle de produits marqués. La RMN du deutérium sera particulièrement utile pour la caractérisation des composés dont la totalité ou la quasi-totalité des hydrogènes non labiles auront été remplacés par des deutériums.

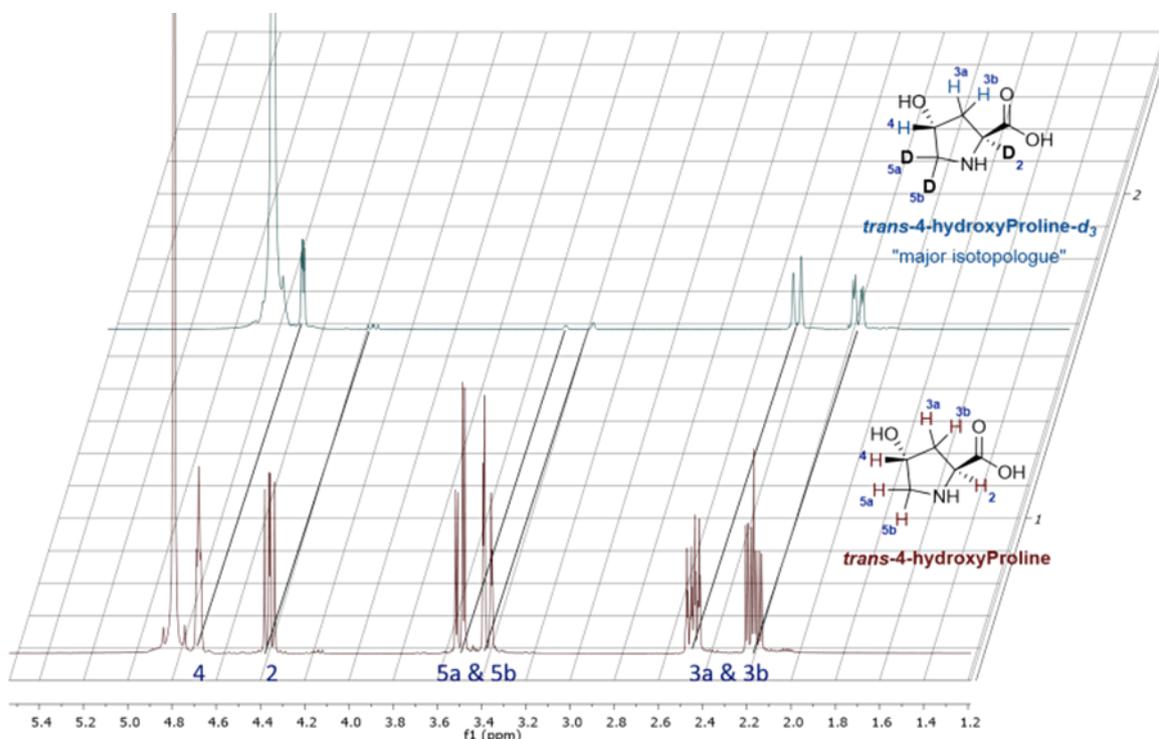


Figure 57 : Spectres RMN ^1H de la 4-hydroxy-proline (en bas) et de son homologue marqué au deutérium (en haut).

En résumé, 23 composés (sur les 41 composés comportant les fonctions adaptées au marquage avec le ruthénium) ont pu être marqués convenablement par échange H/D en présence de Ru/C dans des conditions proches de celles décrites par Taglang et al. et Pieters et al.^{237,243}. Pour les autres composés, n'ayant soit pas obtenus un taux d'incorporation de deutérium suffisant dans ces conditions, soit n'ayant pas été testés car ne présentant pas les groupements fonctionnels adaptés à l'introduction de deutérium avec Ru/C, l'évaluation du changement de catalyseur a été envisagée.

b - Adaptation du procédé de marquage avec remplacement de Ru/C par Pd/C

Pour les composés ne comportant pas (ou pas suffisamment) de fonctions amines ou hydroxyles et pour lesquels le procédé avec Ru/C n'est pas performant, l'utilisation de Pd/C, connu comme étant un catalyseur métallique beaucoup moins chimio- et régio-sélectif pour les réactions d'échanges H/D sur des structures carbonées, a été évaluée²⁴⁰. La Figure 58 montre les exemples de marquage de l'acide adipique et de l'acide citrique obtenus avec Pd/C dans des conditions équivalentes (concentration du substrat, charge catalytique, etc) à celles utilisées avec Ru/C mais en travaillant à une température plus élevée (minimum 80°C) comme souvent décrit dans la littérature avec ce catalyseur^{274,275}.

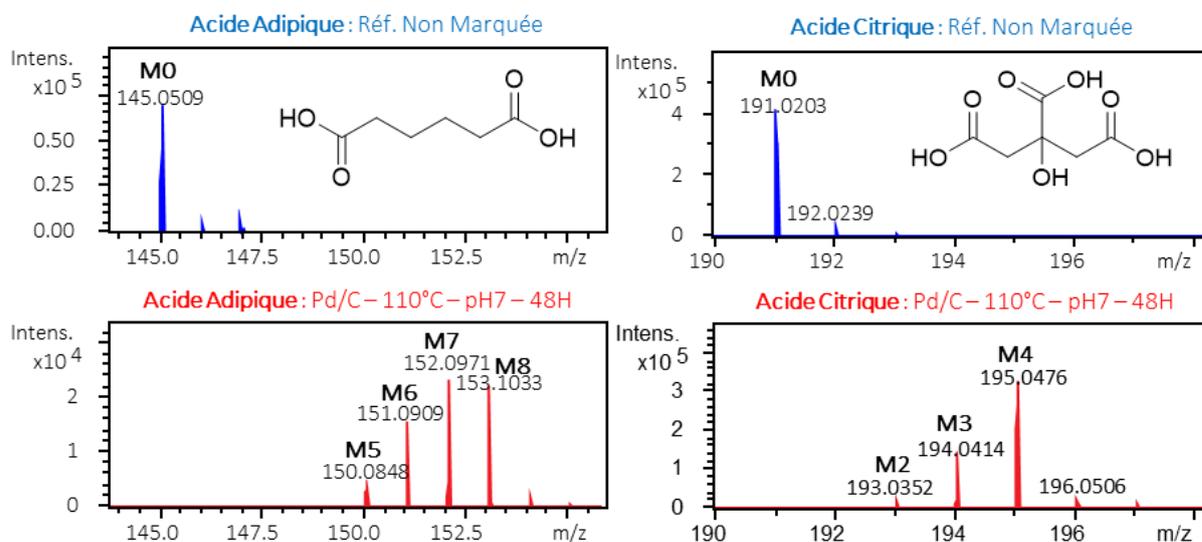


Figure 58 : Exemples de marquages obtenus par échange H/D métallo-catalysé par Pd/C.

Ces deux exemples de produits marqués mettent en évidence le fait que la réaction d'échange H/D catalysée par le palladium est très efficace pour le marquage au deutérium des acides organiques. Les massifs isotopiques obtenus sont nettement décalés par rapport à ceux de leur standard non marqué respectif. La distribution isotopique reste cependant plus au moins hétérogène selon les molécules : celle de l'acide citrique étant, à titre d'exemple, beaucoup plus homogène que celle de l'acide adipique (Figure 58).

Pour les composés n'étant toujours pas marqué efficacement dans ces conditions, l'ajout d'une base inorganique (NaOD ou Na₂CO₃) au milieu réactionnel a été testé dans le but de favoriser l'incorporation du deutérium. Une illustration de l'influence de la modification du pH de la solution réactionnelle est présentée en Figure 59 avec les exemples de l'acide diglycolique et de l'acide 3-hydroxy-3-méthylglutarique. On constate dans les deux cas que les conditions basiques favorisent l'incorporation de deutérium dans les structures puisque pour les deux exemples la disparition complète de l'isotopologue M0 est observée ainsi qu'un net déplacement du massif isotopique vers les plus hauts rapports *m/z*. Ces comparaisons ont pu être faites en engageant

simultanément, sur le système CAT24, les deux réactions (avec et sans base) dans les mêmes conditions de pression, température et durée, pour chacun des substrats.

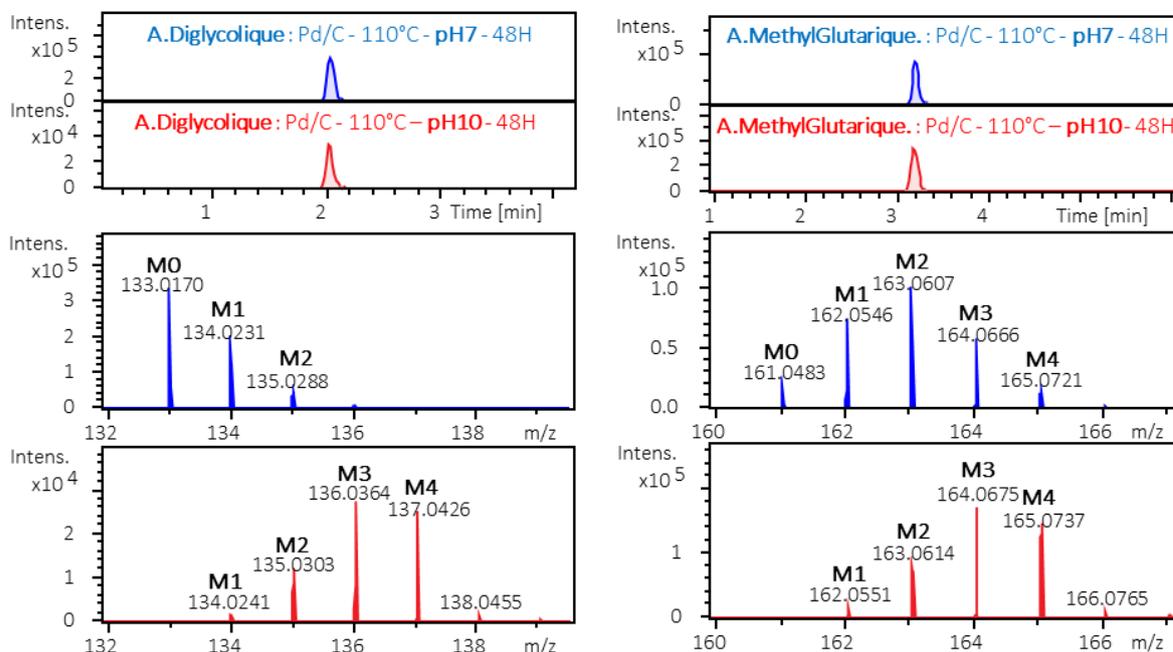


Figure 59 : Influence du pH pour la réaction mono-étape métallo-catalysée.

A.Diglycolique : Acide Diglycolique ; A.Methylglutarique : Acide-3-Hydroxy-3-Méthylglutarique

Ainsi, l'utilisation de Pd/C en tant que catalyseur semble bien adaptée au marquage des acides organiques et l'ajout d'une base a prouvé son efficacité pour améliorer significativement les taux d'incorporation de deutérium dans certaines structures. Des mélanges de catalyseurs (Ru/C et Pd/C) ont également été testé afin de diminuer la contribution de l'isotopologue monoisotopique et d'augmenter le taux d'incorporation des deutériums.

Les résultats globaux obtenus à la suite des différentes optimisations des protocoles d'échange H/D métallo-catalysés sont présentés composé par composé dans la partie qui suit. Globalement, l'ensemble de ces optimisations met en évidence la difficulté de définir un protocole unique pour un marquage optimal de l'ensemble du jeu de molécules d'intérêt. Au total, 13 composés présentent des massifs isotopiques sans MO avec l'utilisation de Pd/C.

6 – Bilan des marquages métallo-catalysés

L'ensemble des massifs isotopiques obtenus ont été comparés au cahier des charges initialement décrit. L'un des paramètres indispensables à l'utilisation de composés marqués comme standards internes, est la disparition complète du composé monoisotopique (c'est-à-dire non-marqué noté M0) ou une présence à une intensité relative inférieure à 1 % de l'isotopologue deutéré majoritaire. Pour 32 des 52 composés testés cet objectif a été atteint (Tableau 16). Après marquage, le décalage des massifs isotopiques de deux Da ou plus, par rapport au standard non-marqué, a été plus délicate à atteindre pour un grand nombre de composés mais nous verrons un peu plus loin que nous pouvons facilement nous affranchir de ce paramètre en travaillant à très haute résolution (avec l'Orbitrap Tribid Fusion par exemple) pour les applications en analyse quantitative. Un bilan des meilleures conditions de marquage (Tableau 16) utilisées pour chacun des composés nous permet, au final, de mettre en avant quatre conditions réactionnelles efficaces distinctes selon la nature du composé à marquer : Ru/C, 55°C, pH7, 24h ; Ru/C, 55°C, pH10, 24h ; Pd/C, 80°C, pH7, 24h ; et Pd/C, 80°C, pH10, 24h.

Notons cependant quelques particularités liées à ce type d'échange H/D métallo-catalysé. Tout d'abord, les conditions appliquées ne permettent pas de marquer ni l'O-acetylserine ni l'acide panthoténique qui se dégradent complètement lors de la réaction de marquage. De même, en raison du caractère réducteur des conditions de marquage utilisées (combinaison de Ru/C ou Pd/C et de D₂), l'acide maléique ne peut pas être marqué sans subir une réduction qui le convertit en acide succinique deutéré (Figure 60). Ces dernières conditions permettent d'avoir accès de façon simple à une version marquée de l'acide succinique à minima di-deutérée de façon plus efficace que directement à partir de l'acide succinique



Figure 60 : Conversion de l'acide maléique en acide succinique deutéré dans les conditions réductrices de réaction d'échange H/D métallo-catalysée..

D'ailleurs, en présence de palladium, l'acide succinique marqué obtenu répond assez bien au cahier des charges initial avec l'absence du composé monoisotopique et un massif isotopique homogène (Tableau 16). La réduction reste néanmoins un problème pour les métabolites sensibles à ces conditions et dont la conversion n'est pas toujours totale et/ou ne conduit pas nécessairement à des formes deutérées de métabolites d'intérêt.

Un des critères importants du cahier des charges est l'absence d'échange en retour (« back exchange »). Ce phénomène peut être rencontré pour les composés comportant des hydrogènes acides bien que non-portés par des hétéroatomes. C'est le cas de l'acide méthylmalonique pour lequel l'atome d'hydrogène porté par le carbone entre les deux fonctions acides carboxyliques est acide. En présence de D₂O cet hydrogène va naturellement être échangé avec un deutérium et *vice versa*, en présence d'H₂O, ce deutérium est de nouveau

échangé par un hydrogène (Figure 61). Bien qu'observable dans les conditions d'analyse LC-HRMS classiques (solvants non deutérés), cette version deutérée éphémère de l'acide méthylmalonique n'est en aucun cas utilisable en tant que standard interne marquée. En effet, le « back exchange » est complet en moins de 24h dans l'eau.

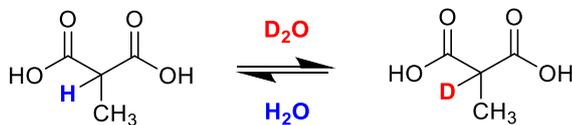


Figure 61 : Exemple d'échange en retour ("back exchange") observé avec l'acide méthylmalonique.

Partie B : Etudes Expérimentales – Chapitre 2

Tableau 16 : Bilan des marquages mono-étape métallo-catalysés.

Compounds	N°	MO	M1	M2	M3	M4	M5	M6	M7	M8	Selected labeling conditions
α-Aminoadipic Acid	1	3%	100%	12%	-	-	-	-	-	-	Ru/C - 55°C - 24h
2-Aminobutyric Acid	2	< 1%	100%	54%	50%	17%	9%	6%	-	-	Ru/C - 55°C - 24h
3-Amino-isobutyric Acid	3	-	< 1%	12%	100%	91%	52%	45%	< 1%	-	Ru/C - 55°C - 24h
Guanidinoacetic Acid	4	30%	76%	100%	-	-	-	-	-	-	Ru/C - 55°C - 24h
Guanidinosuccinic Acid	5	55%	100%	28%	2%	-	-	-	-	-	Ru/C - 55°C - 24h
3-Hydroxy-3-Methylglutaric acid	6	< 1%	10%	48%	100%	79%	6%	1%	-	-	Pd/C - NaOD - 80°C - 48h
3-Methylhistamine	7	-	< 1%	14%	100%	70%	21%	3%	< 1%	-	Ru/C - 55°C - 24h
3-Methylhistidine	8	-	-	< 1%	3%	26%	94%	100%	76%	41%	Ru/C + Pd/C - NaOD - 80°C - 48h
7-Methylguanine	9	62%	100%	-	-	-	-	-	-	-	Ru/C - NaOD - 55°C - 24h
Adipic Acid	10	-	-	-	-	18%	28%	59%	100%	84%	Pd/C - 80°C - 48h
Alanine	11	2%	100%	51%	17%	35%	< 1%	-	-	-	Ru/C + Pd/C - NaOD - 80°C - 48h
Arginine	12	5%	100%	56%	27%	7%	1%	-	-	-	Ru/C - 55°C - 24h
Asparagine	13	< 1%	6%	100%	49%	9%	< 1%	-	-	-	Ru/C - 55°C - 24h
Aspartic Acid	14	-	2.0%	30.0%	100.0%	-	-	-	-	-	Ru/C - 55°C - 24h
Citric Acid	15	-	< 1%	5%	41%	100%	7%	-	-	-	Pd/C - 80°C - 48h
Citrulline	16	3.7%	100.0%	3.6%	< 1%	-	-	-	-	-	Ru/C - 55°C - 72h
Diglycolic Acid	17	-	< 1%	28%	100%	90%	-	-	-	-	Pd/C - NaOD - 80°C - 48h
Glutaconic Acid	18	-	-	2.2%	13.3%	100.0%	-	-	-	-	NA - "back exchange"
Glutamic Acid	19	< 1%	8%	42%	100%	25%	1%	-	-	-	Ru/C + Pd/C - NaOD - 80°C - 48h
Glutamine	20	5.0%	100.0%	37.0%	10.0%	-	-	-	-	-	Ru/C - 55°C - 24h
Glycerol-3-Phosphate	21	< 1%	9%	54%	100%	67%	-	-	-	-	Ru/C - NaOD - 55°C - 24h
Glycine	22	-	4%	100%	1%	-	-	-	-	-	Ru/C - NaOD - 55°C - 24h
Histidine	23	-	-	< 1%	4%	31%	100%	7%	-	-	Ru/C + Pd/C - NaOD - 80°C - 48h
Hydroxyglutaric Acid	24	10%	94%	100%	42%	12%	2%	-	-	-	Ru/C - 55°C - 48h
Imidazolelactic Acid	25	-	< 1%	5%	35%	88%	100%	6%	-	-	Ru/C + Pd/C - NaOD - 80°C - 48h
Isocitric Acid	26	< 1%	4%	53%	100%	7%	-	-	-	-	Pd/C - NaOD - 80°C - 48h
Lactic Acid	27	100%	48%	4%	-	-	-	-	-	-	Ru/C - 55°C - 24h
Leucine	28	3%	100%	5%	3%	-	-	-	-	-	Ru/C + Pd/C - 80°C - 48h
Lysine	29	4%	< 1%	14%	100%	5%	< 1%	-	-	-	Ru/C - 55°C - 48h
Maleic Acid	30	11.0%	20.0%	100.0%	80.0%	-	-	-	-	-	Pd/C - 80°C - 48h
Malic Acid	31	-	3%	100%	18%	< 1%	-	-	-	-	Pd/C - NaOD - 80°C - 48h
Methylimidazoleacetic Acid	32	-	-	3%	35%	100%	18%	3%	1%	-	Ru/C + Pd/C - NaOD - 80°C - 48h
Methylmalonic Acid	33	-	-	-	Back exchange			-	-	-	Pd/C - NaOD - 80°C - 48h
N-Methyl-D-Aspartic Acid	34	-	-	-	< 1%	12%	69%	100%	5%	-	Ru/C + Pd/C - NaOD - 80°C - 48h
O-Acetylserine	35	-	-	-	Complete degradation			-	-	-	Ru/C - 55°C - 24h
Ornithine	36	-	-	-	-	-	2%	23%	100%	5%	Ru/C - NaOD - 55°C - 72h
Pantothenic Acid	37	-	-	-	Complete degradation			-	-	-	Pd/C - NaOD - 80°C - 24h
Phosphoethanolamine	38	< 1%	6%	100%	26%	2%	-	-	-	-	Ru/C - 55°C - 48h
Pipecolic Acid	39	-	< 1%	7%	100%	83%	28%	8%	1%	-	Ru/C - NaOD - 55°C - 24h
Proline	40	-	-	9%	100%	63%	15%	2%	-	-	Ru/C - 55°C - 48h
Pyroglutamic Acid	41	100%	29%	3%	< 1%	-	-	-	-	-	Ru/C - 55°C - 48h
Sarcosine	42	-	-	< 1%	3%	20%	100%	-	-	-	Ru/C - NaOD - 55°C - 24h
Serine	43	< 1%	3%	47%	100%	3%	-	-	-	-	Ru/C - 55°C - 48h
Sorbitol	44	-	-	-	-	-	2%	19%	83%	100%	Ru/C - 55°C - 48h
Succinic Acid	45	-	1%	10%	45%	100%	-	-	-	-	Pd/C - NaOD - 80°C - 48h
Taurine	46	-	3%	55%	100%	49%	-	-	-	-	Ru/C - 55°C - 72h
Taurocholic Acid	47	28%	100%	52%	19%	5%	-	-	-	-	Ru/C - 55°C - 72h
Threonic Acid	48	7%	34%	100%	86%	30%	-	-	-	-	Ru/C - 55°C - 72h
Threonine	49	1%	4%	100%	56%	43%	73%	2%	-	-	Ru/C - NaOD - 55°C - 72h
4-Hydroxyproline	50	-	< 1%	15%	100%	37%	4%	< 1%	-	-	Ru/C - NaOD - 55°C - 72h
Tyrosine	51	51%	100%	65%	24%	10%	3%	-	-	-	Ru/C - 55°C - 24h
Valine	52	-	-	-	-	-	2%	17%	66%	100%	Ru/C + Pd/C - NaOD - 80°C - 48h

Le troisième critère du cahier des charges, correspondant à l'obtention d'un isotopologue majoritaire avec un massif isotopique le plus homogène possible, est le plus difficile à valider pour la très grande majorité des composés marqués. En effet, pour moins de 6 % des composés marqués, l'isotopologue majoritaire représente plus de 90 % du massif isotopique. Pour 19 % et 31 % des composés, la contribution de l'isotopologue majoritaire est supérieure à 70 % et 50 % du massif isotopique, respectivement. Enfin, pour 39 % des composés marqués, l'isotopologue majoritaire représente moins de 50 % du massif isotopique total (Figure 62).

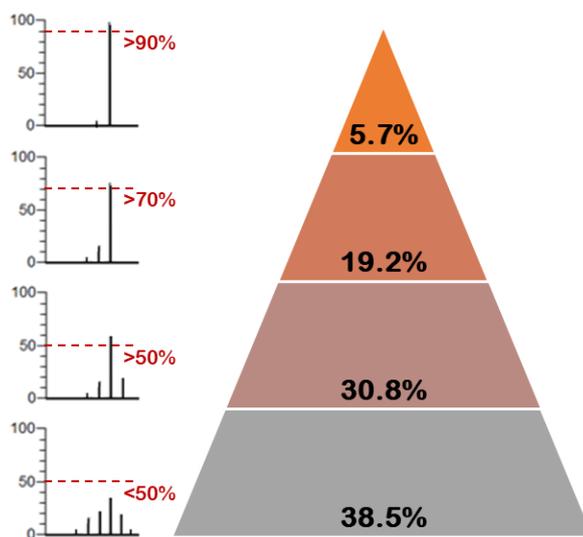


Figure 62 : Répartition des composés marqués en fonction de l'homogénéité de leur marquage.

Ces observations nous ont conduit à imaginer et évaluer une méthode innovante de quantification basée sur la construction de gamme de calibration interne en exploitant l'hétérogénéité des marquages et donc des massifs isotopiques obtenus. Cette méthode que l'on pourra nommer « multi-point internal calibration » consiste à considérer chaque isotopologue, d'un lot de métabolite marqué, comme un standard interne marqué indépendant. Cette méthode est détaillée dans la partie qui suit.

B – Développement d'une nouvelle méthode de quantification : « multi-point internal calibration »

1 – Contexte et objectif

Comme montré dans la partie précédente, les standards marqués, synthétisés par la méthode d'échange H/D mono-étape métallo-catalysée, présentent le plus souvent de multiples degrés de marquage (plusieurs isotopologues coexistent pour un même lot de produit marqué). Les massifs isotopiques ainsi obtenus nous ont permis d'imaginer et de développer une méthode innovante de quantification nommée « multi-point internal calibration ». Le principe de cette méthode est de considérer chaque isotopologue d'un composé

marqué comme un standard interne marqué indépendant pour la construction de gamme de calibration interne. En effet, une fois la concentration de chaque isotopologue déterminée, il est possible d'utiliser chacun d'entre eux pour construire une gamme d'étalonnage. Supplémenté à l'échantillon biologique d'intérêt, le mélange d'isotopologues constitutifs du métabolite marqué (IsoMix) doit permettre de quantifier le composé endogène correspondant. La Figure 63 illustre le principe de cette méthode avec l'exemple du sorbitol deutéré obtenu par la méthode de marquage précédemment décrite.

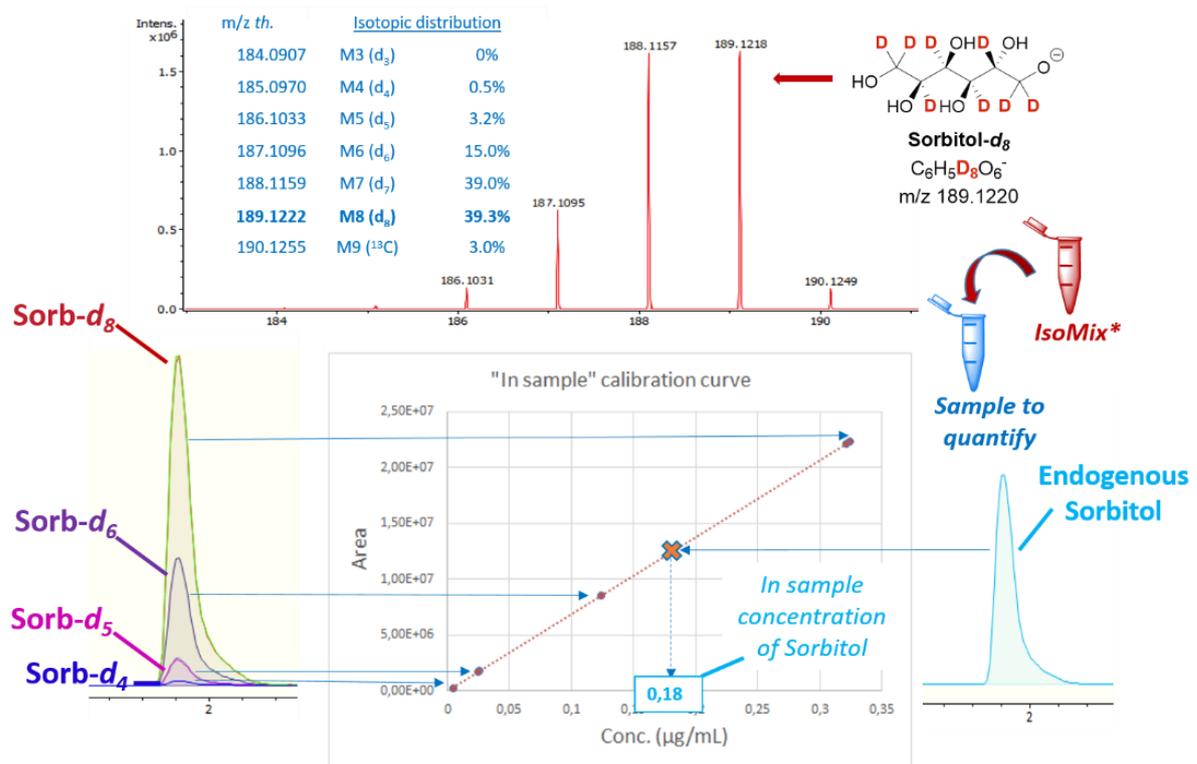


Figure 63 : Principe de la méthode de "multi-point internal calibration".

L'objectif de cette étude a donc été d'évaluer cette nouvelle approche de quantification et de la comparer aux méthodes de quantification plus traditionnelles basées sur l'utilisation de gammes externes de calibration.

2 – Méthodologie

La sélection des composés utilisés pour évaluer la méthode de quantification a principalement été réalisée en ne gardant que ceux pour lesquels l'isotopologue non-marqué (M0) n'était plus présent (ou avec une intensité relative inférieure à 1 %). Le critère du cahier des charges indiquant que deux deutérium doivent, idéalement, au minimum être incorporés dans la structure du composé marqué s'avère ne pas être obligatoire

pour des analyses LC-HRMS réalisées avec l'Orbitrap Fusion. En effet, à une résolution de 240 000 (à m/z 200) l'isotopologue C-13 du composé non marqué à quantifier et l'isotopologue mono-deutééré du lot marqué, sont bien résolus comme illustré sur la Figure 64 avec les exemples de la citrulline et de l'histidine (ce qui n'est pas le cas lorsque les acquisitions de spectres de masse sont réalisées sur le Q-TOF Impact HD ou sur un Orbitrap à une résolution inférieure à 100 000). Notons qu'avec une telle résolution, une tolérance d'écart de mesure de masse de cinq ppm a pu être appliquée pour la détermination des réponses analytiques des isotopologues deutérés, lors du traitement semi-manuel des données à l'aide de méthodes de traitement (« processing method ») mises en place sur QuantAnalysis (Xcalibur).

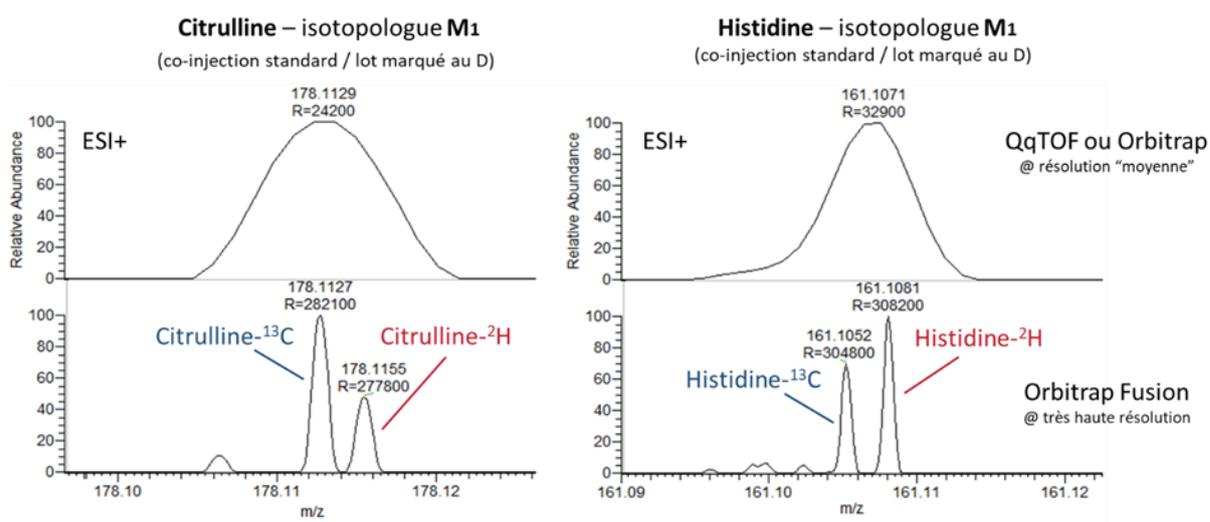


Figure 64 : Influence de la résolution sur la séparation des composés mono-C-13 du standard non marqué et du mono-deutééré du lot marqué. Illustration avec la citrulline et l'histidine

Dans un premier temps, les lots de composés marqués produits par échange H/D ont de nouveau été analysés par LC-HRMS sur l'Orbitrap Fusion pour vérifier la stabilité du massif isotopique par rapport aux premières analyses réalisées sur le Q-TOF (plusieurs mois auparavant) et pour quantifier les isotopologues de chaque lot à l'aide d'une gamme externe de standards non marqués correspondants. Ayant choisi de développer la nouvelle méthode de quantification dans un premier temps en mode d'ionisation positif, 21 composés ont été retenus.

La méthode de quantification « multi-point internal calibration » a tout d'abord été appliquée, en tant que preuve de concept, à la quantification de six composés non détectés dans le plasma NIST dans nos conditions d'analyse. Pour cela, les six standards non marqués ont été supplémentés à des concentrations connues au plasma NIST. Le but étant de vérifier que les concentrations obtenues par la nouvelle méthode de quantification sont bien similaires aux concentrations ajoutées. Les concentrations ainsi obtenues ont également été comparées aux valeurs obtenues par application de méthodes de quantification plus usuelles reposant sur des gammes externes de composés non-marqués.

Partie B : Etudes Expérimentales – Chapitre 2

Dans un second temps, quinze composés endogènes au plasma NIST ont été quantifiés par la méthode « multi-point internal calibration ». De la même façon que pour la preuve de concept, les concentrations obtenues ont été comparées à des valeurs de références trouvées dans des bases de données publiques (HMDB) et également avec les valeurs obtenues par la méthode de quantification par gamme externe.

Les résultats obtenus avec la production à façon des standards internes marqués au deutérium ainsi que l'évaluation de leur utilisation pour de la quantification de métabolites dans le plasma de référence NIST (SRM1950) seront rapportés dans une publication (à soumettre prochainement) dont une version non définitive est présentée ci-dessous.

3 – Publication (en cours de rédaction)

High-throughput metal-catalyzed deuterium labeling of small molecules for MS-based quantitative metabolomics and evaluation of an innovative multipoint calibration approach.

Kathleen Rousseau,¹ Jessica Michieletto,¹ Yu Min Kiw,¹ Sophie Feuillastre,² Grégory Pieters,² Christophe Junot,¹ François Fenaille,¹ Annelaure Damont.¹

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

²Service de chimie bio-organique et de marquage, Laboratoire de Marquage au Tritium, CEA, Université Paris Saclay, 91191 Gif-sur-Yvette, France

ABSTRACT: Large-scale absolute quantification is a major focus in the metabolomics field to accurately correlate metabolic changes with the development of a disease or in association with genetic or environmental factors. High-resolution mass spectrometry-based quantitative metabolomics essentially relies on the use of stable isotope labeled internal standards (SIL-IS) to account for matrix effects during the analysis. An attractive approach when using SIL-IS for quantification purposes, is to generate multipoint calibration curves with different isotopologues of the targeted metabolite. Here, we report (1) the production in one shot and on-demand of a panel of multi-deuterated analogs of more than 50 target metabolites and (2) the evaluation of their use as stable isotope internal standards (SIL-IS) for the absolute quantification by isotopic dilution mass spectrometry. The evaluated labeling protocol relies on the use of a one-step metal-catalyzed H/D reaction exchange. Overall, 21 molecules were selected to quantify compounds spiked in plasma as well as endogenous metabolites in blank plasma. A proof of concept of the use of such SIL-IS by a multi-point internal calibration approach was proposed and compared to the more traditional IDMS strategy used in targeted quantitative metabolomics studies. *In fine*, this innovative quantification approach proved to be highly reliable and gave similar quantitative results to the IDMS method (error < 15%). By permitting the calculation of quantitative data with a single injection per sample to quantify, the multi-point internal calibration strategy offers the substantial advantage of limiting the number of injection in mass-spectrometry based quantitative metabolomics studies.

▪ INTRODUCTION

Metabolomics is a relatively new science that studies small molecules with molecular weight below 1.5 kDa present in biological samples.¹ Large-scale absolute quantification is a major focus in the metabolomics field to accurately correlate metabolic changes with the development of a disease or in association with genetic or environmental factors. Absolute quantification is crucial to ensure analytical robustness when comparing analysis performed at different periods of time or when carried out with different analytical equipment. Metabolome analysis requires the use of advanced analytical techniques. High-resolution mass spectrometry (HRMS)² is the most widespread analytical technique for metabolomics studies essentially due to its high sensitivity and wide dynamic range coverage. Even though a variety of methodologies have been developed for MS-based quantitative analysis,³ strategies relying on isotopic dilution with the use of stable isotope labeled internal standards (SIL-IS) remain the gold standard methods when dealing with absolute quantification in metabolomics.^{5,6} Despite the increasing expansion of chemical catalogs proposing a variety of stable isotope labeled molecules in recent years, we are far from an entire coverage of the known metabolome. Additionally, the cost of some labeled versions of metabolites may be prohibitive.^{7,8} Due to these limitations, it is tempting for any metabolomics laboratory

interested in large-scale targeted quantitative analysis to develop strategies for the production of its own labeled compound library that could be used as SIL-IS. The most commonly used isotope for tagging molecule are carbone-13 (^{13}C), nitrogen-15 (^{15}N) and deuterium (D or ^2H) because of their common, even systematic, occurrence in organic molecules constituting the metabolome. To get access to pure standards labeled with carbone-13 or nitrogen-15, the elaboration of multi-step chemical reactions is usually mandatory. In contrast, a number of recent developments have been reported that allow the direct introduction of deuterium in replacement of hydrogen atoms at stable position in chemical structures. Such an approach, also named stable H/D exchange reaction, is attractive since it offers the possibility to generate, in one reaction step, SIL-IS of interest starting from unlabeled authentic standards. In particular, the one-step metal-catalyzed H/D exchange methodology seemed appealing for it allows the production of permanently deuterated compounds following a one-step reaction and starting from authentic standards of interest (target metabolites) as substrates. This process requires the use of a metal as catalyst, a deuterated solvent (*e.g.* D_2O) and deuterium gas as source of deuterium for the reaction. The presence of a metal (*e.g.* Ruthenium or Palladium) activates C-H bonds, allowing the exchange between hydrogen and deuterium. Among the list of metal that can be used as H/D exchange catalyst, ruthenium and palladium are well-suited for the labeling of common metabolites: the first one is reported to selectively favor deuterium incorporation on carbon at alpha-position of nitrogen^{9,10,11} and oxygen^{12,13} containing functions, such as amino and hydroxy groups, that are commonly encountered in the structure of polar metabolites (*e.g.* amino acids), while the latter is known for its reduced labeling site regioselectivity¹⁴ rendering it more suited for the labeling of metabolites lacking amine or alcohol functions (*e.g.* adipic acid). Whatever the labeling strategy employed, the resulting deuterated compounds should fulfill two major criteria to be used as SIL-IS for IDMS-based absolute quantification of metabolites: the presence of remaining unlabeled metabolite (noted M_0) should be avoided, or at least limited and the formation of one major isotopologue (with an isotopic purity > 95%) is preferred. Chemical and labeling stability (no H/D back-exchange) should also be guaranteed for the deuterium-labeled compounds to be used as SIL-IS. In this study, in a first part, we demonstrate the straightforward production of SIL-IS following a one-step metal catalyzed H/D exchange reaction protocol and using a high throughput reaction screening system. Then, we report the evaluation of the use of the in-house produced deuterium-labeled molecules for the quantification of metabolites in spiked and blank human plasma NIST® SRM® 1950 samples using IDMS-based strategies.

▪ EXPERIMENTAL SECTION

Reagents and Materials. Metabolite standards, reagent (NaOD solution, 40 wt. % in D_2O) and catalysts (Ru/C 5 wt. % loading, Pd/C 10 wt. % loading) were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France), D_2O was purchased from Eurisotop (Saint-Aubin, France), D_2 (Messer??). Acrodisc® syringe filter (PVDF membrane, diameter: 25 mm, pore size: 0.45 μm) for reaction mixture filtration and catalyst removal, were purchased from VWR International (Lutterworth, U.K.). All extraction (HPLC grade methanol and water) and UHPLC solvents (Optima grade water and acetonitrile) as well as formic acid (≥ 98.0 % purity) and ammonium carbonate as mobile phase additive, were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). NIST® SRM® 1950 reference plasma was ordered from the National Institute of Standards and Technology (NIST, Gaithersburg, Maryland, USA).

Instrumentation and protocol for the production of deuterium-labeled compounds. The CAT24 screening reaction system HEL (purchased from HEL Limited, Borehamwood, UK) was used to conduct all H/D exchange reactions. The CAT24 synthesis system consists in a single unit featuring 24 individual vials that can be pressurized, stirred and heated by a hot-plate, under manual operation. The design of the system prevents any cross-contamination between samples. It allows twenty-four reactions to be performed simultaneously at the same temperature and pressure conditions. Several reaction conditions were tested depending on the nature of the substrate (*i.e.* the target metabolite to be labeled) following the labeling investigation workflow depicted in **Figure**

1A. Typical H/D exchange reactions were carried out as follows: each vial was charged with 0.03 to 0.09 mmol of substrate in D₂O (0.5 mL), a magnetic stirrer and 15 mol% of metal catalyst (Ru/C 5 wt. % or Pd/C 10 wt. %). The system was tightly closed with the help of spanners to isolate it before being placed under D₂ atmosphere. Three cycles of pressurization and liberation of deuterium gas were repeated before being charged with 2 bars of D₂ gas. The system was then placed on a hot plate and heated to 55 °C (with Ru/C) or 80°C (with Pd/C) for at least 24h. After that, the system was cooled down to room temperature and the internal D₂ pressure was released under a fume hood. Small amounts (10 to 20 µL) of solution were collected from each vial to check (after filtration and dilution) for deuterium incorporation with LC-HRMS monitoring (cf sample preparation). The system was refilled with D₂ and heating was prolonged for an additional 24h or 48h for the compounds with insufficient deuterium incorporation. An additive, a base (NaOD 40% in D₂O, 20 µL) was also added if necessary to improve deuterium incorporation. Eventually, when the labeling was insufficient with Ru/C alone, Pd/C (15 mol%, 10 wt. %) was added to the reaction mixture.

Sample preparation for LC-HRMS deuterium-labeling monitoring. For each reaction mixture of the 24 vials of the CAT24 reactor, 10 to 20 µL were collected with a syringe and diluted in 500 µL of water before being filtered with an Acrodisc® syringe filter (PVDF membrane, d: 25 mm, pore size: 0.45 µm) to remove the metal catalyst particles. Each resulting mixture was then diluted to a range of 1 to 5 µg/mL (based on the initial starting material concentration) in the HPLC mobile phase (aqueous ammonium carbonate 10 mM pH 10.5 in water/acetonitrile 40/60 for the ZIC-p-HILIC column analysis, or 0.1% acid formic in water/0.1% acid formic in acetonitrile 95/5 for the PFPP column) before being injected into the LC-HRMS system. Authentic standard solutions were prepared in a similar range of concentration using the same HPLC mobile phase before being injected as reference compound into the LC-HRMS system. Deuterium incorporation was inspected by comparing the isotopic pattern of the reference compound with the one of the reaction product. Chromatogram were also inspected to control possible retention shift or shape modification.

High performance liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) analyses for deuterium incorporation evaluation. All purchased authentic standards as well as labeling reaction products were analyzed by LC-HRMS (isotopic pattern inspection) after sample preparation as previously described. Chromatographic separation was carried out on a Dionex Ultimate 3000 system (ThermoFisher Scientific) equipped either with a Discovery® HS-F5 PFPP column 2.1 x 250 mm, 5 µm (Supelco, Merck, Darmstadt, Germany) kept at 30°C applying elution condition (1) or a SeQuant® Zic-p-Hilic column 5 µm, 2.1 x 150 mm (Merck, Darmstadt, Germany) kept at 15°C applying elution condition (2). Elution condition were as follows: (1) mobile phase A: 0.1 % formic acid in water; mobile phase B: 0.1 % formic acid in acetonitrile. The following gradient was applied at a flow rate of 250 µL/min: t = 0 min, 2% B; t = 2 min, 2% B; t = 10 min, 5% B; t = 16 min, 35% B; t = 20 min, 100% B; t = 24 min, 100% B; t = 24.1 min, 2% B; t = 30 min, 2% B. (2) Mobile phase A: 10 mM ammonium carbonate aqueous buffer adjusted with ammonium hydroxide to pH 10.5; mobile phase B: acetonitrile. The following gradient was applied at a flow rate of 200 µL/min: t = 0 min, 80% B; t = 2 min, 80% B; t = 12 min, 40% B; t = 12.1 min, 0% B; t = 17 min, 0% B, t = 17.1 min, 80% B; t = 42 min, 80% B. High resolution mass spectrometry analyses were performed either on a Q-TOF Impact HD (Bruker, Wissembourg, France) or an Orbitrap Tribrid Fusion (ThermoFisher Scientific, Courtaboeuf, France) equipped with a heated electrospray ionization source operated in the positive or negative mode. Parameters were set as follows: a) Q-TOF in positive ionization mode: capillary voltage +4000 V at a temperature of 200 °C; gas flow rate 10 L/min; nebulizer pressure 2.1 bar; quadrupole collision energy 4 eV and collision cell energy 2 eV. b) Q-TOF in negative ionization mode: capillary voltage -4500V at a temperature of 200°C; gas flow rate 8 L/min; nebulizer pressure 1.7 bar; quadrupole collision energy 8 eV and collision cell energy 2 eV. MS spectra were acquired from 50 to 400 m/z at a frequency of 1 Hz in both ionization mode. b) Orbitrap Tribrid Fusion: spray voltage + 4000 V, sheath gas 40 (arbitrary units), auxiliary gas 12 (arbitrary units), sweep gas 1 (arbitrary units), ion transfer tube temperature 325 °C and vaporizer temperature 100 °C. Ion transfer parameters were set as follows: mass range at "normal", lens RF level at 60 %. Scan parameters applied were as follows: acquisition time from 0 to 20 min, positive ionization mode, micro scan

at "1", data type at "profile", AGC target at 5e4. For acquisition in MS mode, the data were acquired using the Orbitrap analyzer, at a resolution of 240,000 at m/z 200 (FWHM); maximum injection time was set at 400 ms, and the scan range from m/z 70 to m/z 500.

Quantification of the in-house produced labeled compounds. Batches of labeled material produced with the H/D exchange strategy that have been validated on the basis of their isotopic profiles, were first quantified for further use as stable-isotope labeled internal standard (SIL-IS). A first concentration estimation was performed by direct comparison of LC-HRMS peak area of the labeled compounds (sum of the major isotopologues) with the area of the corresponding unlabeled compound injected at 1 µg/mL. Then, internal calibration curves with unlabeled standard were constructed to precisely quantify the labeled batches. For each metabolite, a 7 point internal calibration curve was generated (0.1X – 0.2X – 0.5X – 1X – 2X – 5X – 10X where X is the previously estimated concentration of the labeled metabolite) and the sum of the area of the deuterium-labeled isotopologue peaks was reported on the corresponding curve to determine the total concentration of labeled material. The concentrations of each isotopologue were then calculated based on the isotopic pattern distribution for each metabolite.

Plasma sample extraction. The human plasma NIST® SRM® 1950 (NIST, Gaithersburg, Maryland, USA) was selected as reference biological matrix for evaluating different metabolite quantification approaches. Metabolites were extracted following a dedicated protocol routinely used in the laboratory: 10 µL of plasma were diluted with 30 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 withdrawn and dried under nitrogen. Plasma metabolites were resuspended in 40 µL of 10 mM aqueous ammonium carbonate (pH 10.5) blank or spiked with a mixture of labeled and/or unlabeled standards before sonication for 5 minutes. Then, 60 µL of acetonitrile were added and the resulting samples were sonicated for 15 min and centrifuged at 20,000 g for 15 min at 4°C. The supernatants were transferred to vials for injection and LC-HRMS analyses.

Blank plasma analysis and endogenous metabolite concentration estimation. All plasma samples were analyzed by LC-HRMS using the SeQuant® Zic-p-Hilic under condition (2) for chromatographic separation and the Orbitrap Tribrid Fusion in the positive mode for MS analysis. Blank plasma samples were analyzed first to evaluate the detectability of the 52 metabolites that were tested for deuterium incorporation. Among those that were detected, a short list of 15 targeted metabolites was defined and their concentration were estimated to further adjust the quantification approaches. Thus, after LC-HRMS analysis of blank NIST plasma sample (in triplicate), the peak areas for each targeted compound were directly compared to the peak areas of the corresponding standard injected in buffer (aqueous ammonium carbonate 10 mM pH 10.5 in water/acetonitrile 40/60) at 1 µg/mL. Then, an external calibration was performed with 5 different concentrations (0.01X – 0.1X – 1X – 10X – 100X where X was the first estimated concentration) to refine the preliminary concentration measurement. Finally, to further refine the concentration estimations obtained by external calibration in buffer, standard addition was performed in NIST plasma. The added concentrations of standard ranged from 0.25X to 5X (0.25X – 0.5X – 1X – 2X – 5X where X was the metabolite concentration estimation based on external calibration).

Absolute quantification of metabolites in NIST plasma using different quantification strategies. The quantification of target metabolites was investigated using two different isotope dilution mass spectrometry approaches: (1) the "classical" isotope dilution method (based on the use of external calibration curves with ranges of standards spiked with a constant amount of labeled material) and (2) an innovative multi-point internal calibration (exploiting the presence of series of isotopologues for each labeled compounds to directly built "in-sample" calibration curve with the labeled material spiked in the biological sample to analysed). Bias caused by matrix effects was also evaluated by comparison of quantitative results obtained by external calibration with no use of SIL-IS. Multicomponent standard stock solutions for external calibration of defined nonlabeled metabolite mixtures were prepared in water at 1mg/mL. Labeled compound injection solution concentrations were adjusted based on known or estimated target metabolite concentration in NIST plasma, and ranged from 541 µg/mL to 8

µg/mL (in water). Quantification methods were evaluated on two case study: (a) The quantification of 6 exogenous metabolite in NIST plasma; (b) The quantification of 15 endogenous metabolites in NIST plasma.

(a) Quantification of 6 spiked plasma compounds. For this case study, 6 unlabeled standards (exogenous compounds of NIST plasma) were spiked to the NIST plasma samples. Based on the known concentrations of the spiked compounds in NIST plasma, metabolite-specific calibration ranges with eight levels were adjusted. Labeled compounds were added at the same concentration in all points of range (See Supp Data) and in the NIST plasma sample. The ratio between labeled and unlabeled was 1:1, v/v. The pools of compounds were prepared in H₂O/Acetonitrile (40/60). In NIST plasma, concentrations tested were the same as for the external range. One mix with labeled and unlabeled compounds was prepared in water. The external range was analyzed in triplicate before and after the analyses of plasma NIST also analyzed in triplicate.

(b) Quantification of 15 endogenous plasma compounds. For this case study, 15 endogenous metabolite in plasma NIST were selected for quantification. Based on their previously estimated concentrations (obtained by spiking experiments in the end), metabolite-specific calibration ranges with heigh levels of concentration were adjusted. Labeled compounds were added at the same concentration in all points of range (See Supp Data). The ratio between labeled and unlabeled was 1:1, v/v. The pools of compounds were prepared in H₂O/Acetonitrile (40/60). In NIST plasma, concentrations tested were the same as for the external range. One mix with labeled compounds was prepared in water. The external range was analyzed in triplicate before and after the analyses of plasma NIST also analyzed in triplicate.

▪ RESULTS AND DISCUSSION

1 – Evaluation of high-throughput and on-demand production of deuterium labeled internal standards.

Considering the great importance of stable isotope labeled internal standards for metabolites absolute quantification and the current limitations, mainly their cost and availability, in getting access to a wide range of labeled compounds, our first objective was to apply a simple, fast and broad spectrum synthetic method to produce our own labeled internal standard library. Thus, we rapidly focused our attention on synthetic protocols allowing the rapid introduction of stable isotope(s) into a chemical structure such as metal-catalyzed H/D exchange reactions already described in the literature for the labeling of amino and hydroxyl groups containing compounds or more generally to any alkyl or aryl containing molecules.⁹⁻¹⁴ Traditionally, the metal-catalyzed H/D exchange labeling protocol requires the use of a special apparatus that can maintain a pressurized D₂ atmosphere during the reaction. Such apparatus usually allow the investigation of one single labeling reaction at a time. In view of the potentially large number of metabolites to label, we explored the use of the CAT24[®] equipment that permit to run and investigate 24 parallel labeling reactions in controlled temperature and pressure conditions (**Figure 1B**). The great advantage of this apparatus lies in the possibility to conduct deuterium labeling reaction on 24 substrates simultaneously, allowing efficient screening of various labeling conditions with a wide range of target metabolites. The labeling strategy investigation was carried out following the workflow depicted in **Figure 1A**. First, the metal catalyst used was selected depending on the chemical groups present in the chemical structure of each metabolite to be label: For compounds containing amino and/ or hydroxy group(s) (e.g. amino acids or polyhydroxylated compounds), Ru/C was chosen as catalyst and the reaction was run at 55°C,⁹⁻¹⁰ while Pd/C was selected to perform deuterium incorporation on any other metabolites (e.g. organic acids) with a reaction temperature of 80°C. Deuterium oxide (D₂O) was selected as deuterated solvent and was compatible with the solubility of all tested compounds (rather polar metabolites).

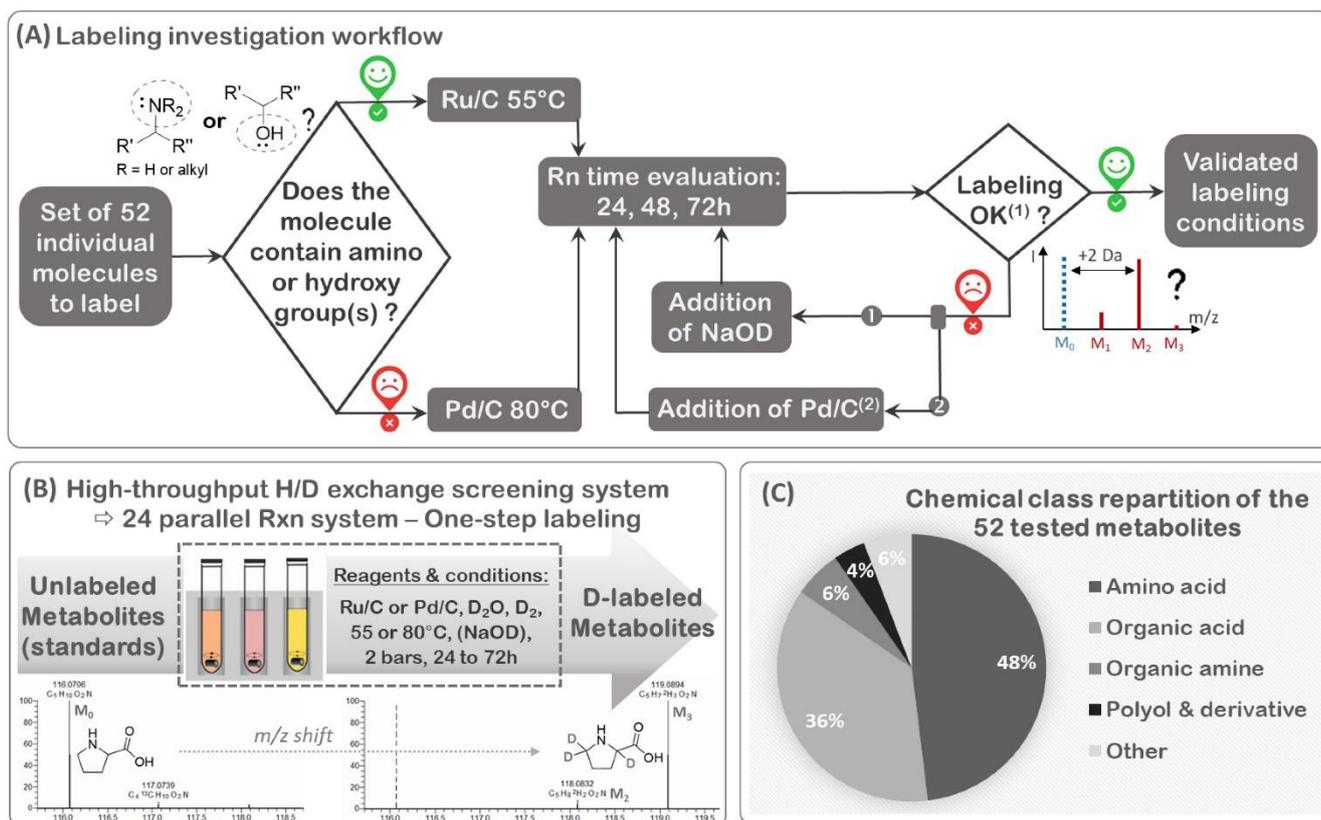


Figure 1. (A) Labeling investigation workflow followed for the metal-catalyzed H/D exchange reaction on a selection of 52 metabolites. (B) Schematic illustration of the labeling process applied for the production of in-house SIL-IS. (C) Chemical class repartition of the 52 metabolites tested for deuterium incorporation.

Another advantage of this protocol is the use of a heterogeneous metal catalyst, such as ruthenium or palladium on carbon, allowing easy recovery of the reaction product by simple filtration. In this study, the H/D exchange labeling reaction was evaluated on 52 compounds (see **Supplementary data S1** for names, smiles and formula) belonging to 4 chemical classes of metabolites (**Figure 1C**): amino acids (25), organic acid (19), organic amine (3), polyol and derivatives (2) and others (3). Reaction parameters such as the nature of the metal catalyst, the temperature, the reaction time and the use of an additive such as an inorganic base were adapted to maximize deuterium incorporation into the structures. For all compounds and reaction conditions, the labeling efficiency was evaluated by LC-HRMS. Chromatographic peak shape of all isotopologues was assessed to ensure no isotopic effects on HPLC retention time were observed. The possible presence of degradation product was also checked. For instance, it was observed that the labeling of glutamine was not ideal when working at pH 10 in D₂O since it was rapidly converted into glutamic acid. Similarly, the labeling of fumaric acid and maleic acid did not provided deuterated-fumaric acid or deuterated maleic acid but rather deuterated-succinic acid due to the reductive character of the reaction conditions. The table displayed in **Supplementary data S2** presents, for each tested metabolite, the isotopologue distribution obtained after deuterium labeling using their respective validated reaction conditions. During the labeling investigations, it was observed that a basic pH was frequently in favor of an increase in deuterium incorporation, to the detriment of the labeling regioselectivity that was not of prime importance for the production of an adequate labeled internal standard. Indeed, one predominant criteria in producing valuable in-house deuterated-standards was to avoid the presence of remaining unlabeled metabolite (noted M₀) in batch product that could hamper the quantification process afterwards. Thus, regioselectivity was not privileged because the larger the number of incorporated deuterium, the less the overlap of isotopic patterns between the labeled compound and the corresponding metabolite to quantify.

Illustrative examples of the LC-HRMS analysis and comparison of unlabeled and labeled compounds are presented in **Figure 2**. It is worth noting that the stable replacement of a hydrogen atom with a deuterium in a chemical structure results in a 1.0063 m/z shift of the mass spectral peak. As an example, the LC-HRMS analysis of citric acid and its tetra-deuterated version give two mass spectral peaks (M_0 and M_4 respectively) that are theoretically 4.0252 m/z (i.e. 4 x 1.0063) apart.

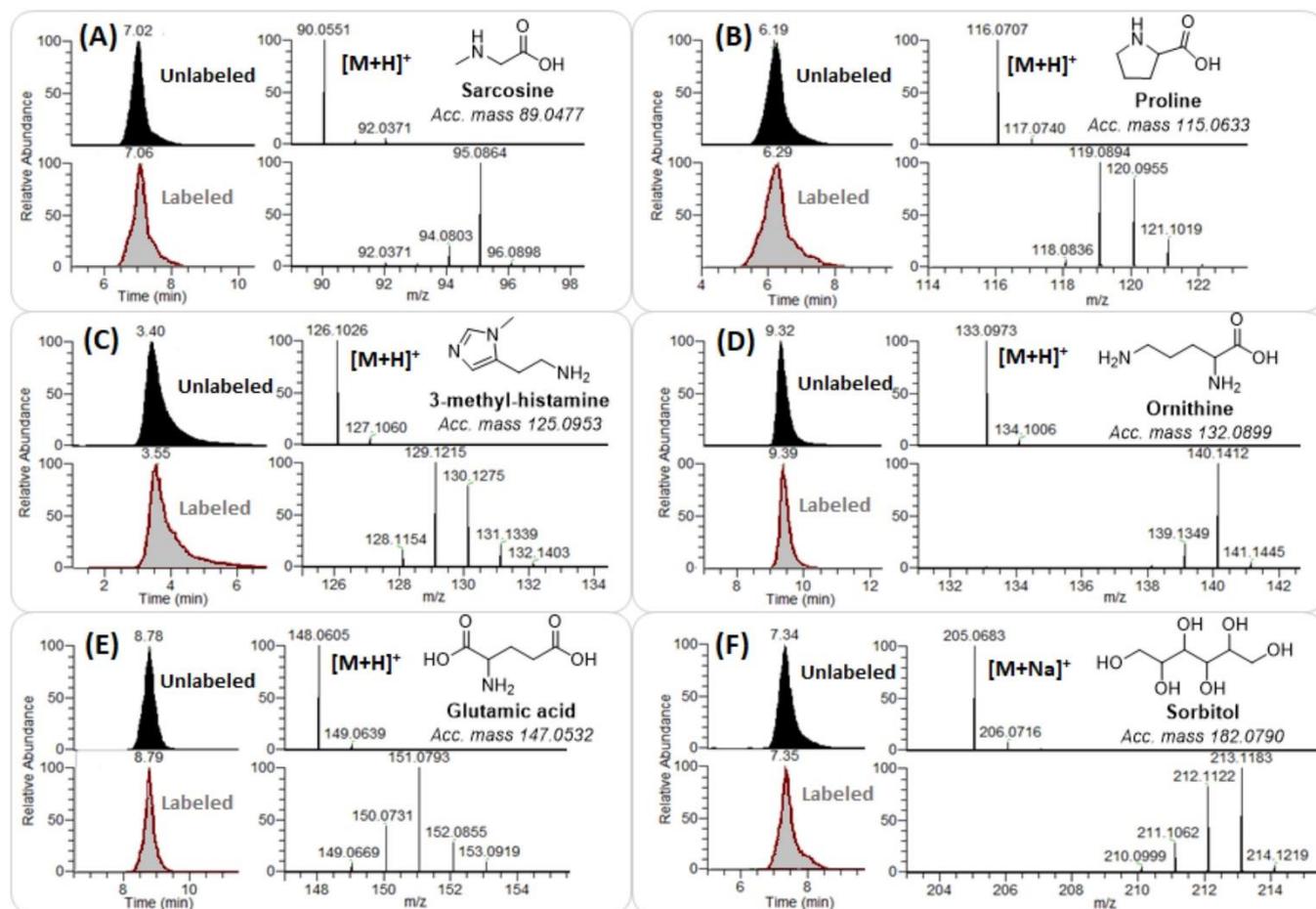
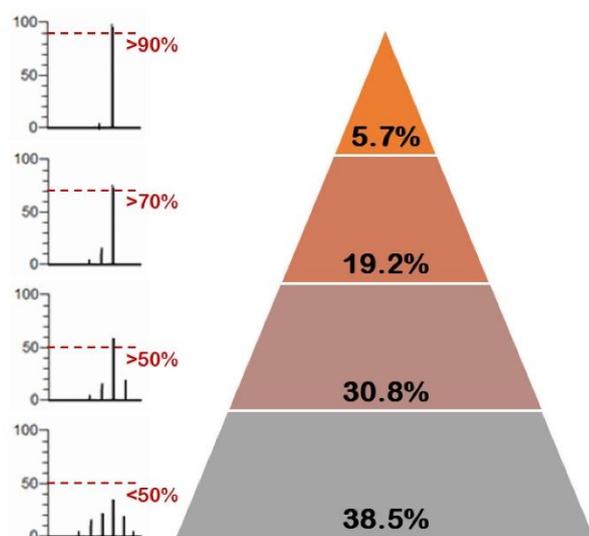


Figure 2: LC-HRMS analyses obtained with a sub-set of metabolites showing the H/D exchange reaction efficiency. Extracted ion chromatogram and mass spectra of unlabeled and labeled Sarcosine (A), Proline (B), 3-methyl-histamine (C), Ornithine (D), Glutamic acid (E) and Sorbitol (F) are presented.



As shown **Figure 2** and summarized in **Supplementary data S2**, isotopic pattern greatly differ from one metabolite to another. Only a few compounds (less than 6 %) could be deuterated with an isotopic purity > 90 % (i.e. with an absolute abundance of the major isotopologue above 90 %). For 19 % of the generated labeled metabolites, the absolute abundance of the major isotopologue was superior to 70 %, and almost 40% of the deuterated compounds displayed mass spectra with a major isotopologue counting for less than 50 % of the global isotopic pattern (**Figure 3**). Based on these observations, the evaluation of a novel isotope dilution mass spectrometry (IDMS) methodology was envisaged: the use of the deuterated-metabolite specific isotopic patterns to generate multi-point internal calibration curve for the one-

shot direct quantification of target metabolites in biofluid seemed appealing. The principle of this innovative approach is illustrated in **Supplementary data S3**. It consists in considering each isotopologue of a metabolite as an independent SIL-IS. Once each isotopologue concentration has been determined using unlabeled standard curves, it's possible to perform direct in-sample quantification of the corresponding endogenous metabolites in any biofluid or biological extract. The evaluation of this strategy is discussed in the next section.

3 – Absolute Quantification in a biological matrix

The usefulness as SIL-IS of the produced labeled compounds was evaluated for the quantification of metabolites in biological matrices using different approaches. The biofluid SRM®1950 (also called NIST® plasma) was chosen as reference matrix to investigate the absolute quantification of selected metabolites. First of all, the presence of the 52 metabolites was checked in NIST plasma using the selected LC-HRMS acquisition method (*i.e.* HILIC column, ammonium carbonate 10 mM pH 10.5 and acetonitrile as mobile phase, and ESI in positive ionization mode). Among these 52 metabolites, x were not detected in NIST plasma among which a subset of 6 molecules were selected to evaluate the quantification approaches by spiking them at known concentrations in blank plasma (see next section 3.2). The x other metabolites, which identities were fully validated by standard addition (unlabeled and labeled) and MS/MS experiment, were detected in NIST plasma with our analytical method. A representative set of fifteen molecules was selected for further absolute quantification and their concentrations were first estimated as described in the experimental part by external calibration followed by standard addition to adjust the amount of SIL-IS to be used for the quantification strategies.

B – Evaluation of the quantitative accuracy: measurement of 6 metabolites spiked in NIST® plasma.

To estimate the performance of the multi-point internal calibration, the novel quantitative approach was compared with more classical methods relying on external calibration curves with or without the use of SIL-IS (See Sup data 4 for method illustration). The main differences between this 3 approaches are: 1) the number of injections to take into account for the quantification of the target metabolites in a single sample and, 2) the correction for matrix effect with the IDMS-based approaches. Indeed, with the conventional quantitative methods (external range with or without SIL-IS), the number of injections increases drastically due to the need for a sufficient number of points to construct the external calibration curve, whereas the in-sample calibration approach requires only a single injection to measure the target metabolite concentration. Table 1 displays the measured concentrations obtained with the 3 different methodologies (See experimental part for details) for 6 metabolites that were spiked in NIST® plasma at known concentrations.

Table 1: Measured concentrations obtained with 3 different quantification methods for 6 compounds spiked at known concentrations in NIST SRM 1950 plasma.

	Spiked C° (µg/mL)	In-sample calibration		Ext. calibration (SIL-IS)		Ext. calibration (no SIL-IS)	
		Measured C° (µg/mL)	Accuracy	Measured C° (µg/mL)	Accuracy	Measured C° (µg/mL)	Accuracy
3-Methylhistamine	0.481	0.487	1.3%	0.480	-0.2%	0.295	-38.7%
2-Aminobutyric Acid	0.481	0.546	13.6%	0.501	4.2%	0.259	-46.2%
N-Methyl-D-Aspartic Acid	0.481	0.452	-6.0%	0.463	3.7%	0.363	-24.4%
Phosphoethanolamine	0.481	0.398	-17.3%	0.437	-9.2%	0.372	-22.8%
Sorbitol	1.042	1.070	2.7%	1.108	6.4%	0.396	-62.0%
Diglycolic acid	1.042	1.030	-1.1%	1.025	-1.6%	0.914	-12.3%

For all the compounds except for the 3-Methylhistamine, a linear range was obtained for the external calibration with and without SIL-IS ($r^2 > 0.997$). To compare, with the multi-point internal calibration, the curve obtained have a coefficient of correlation high than 0.9991 except for the diglycolic acid which has a low coefficient of correlation (0.9859). For the 3-MethylHistamine, a beginning of saturation was visible at the high points of the range. It was therefore necessary to adapt the range to a quadratic curve. For the sorbitol, in plasma NIST, only two isotopologues were detected. To obtain a straight, an artificial point corresponding to the sum of the concentration of all the isotopologues has been added. Moreover, the non-detection of the weakest sorbitol isotopologue implies that endogenous present in NIST plasma is no longer in calibration curve. That can explain the difference between the obtained and the theoretical concentration.

C – Quantification of 15 endogenous metabolites in NIST plasma

Absolute quantification of 15 endogenous metabolites in NIST plasma was evaluated using the 2 IDMS-based approaches and compared to external calibration with no SIL-IS. Table 3 summarizes the resulting concentration values obtained for each strategy.

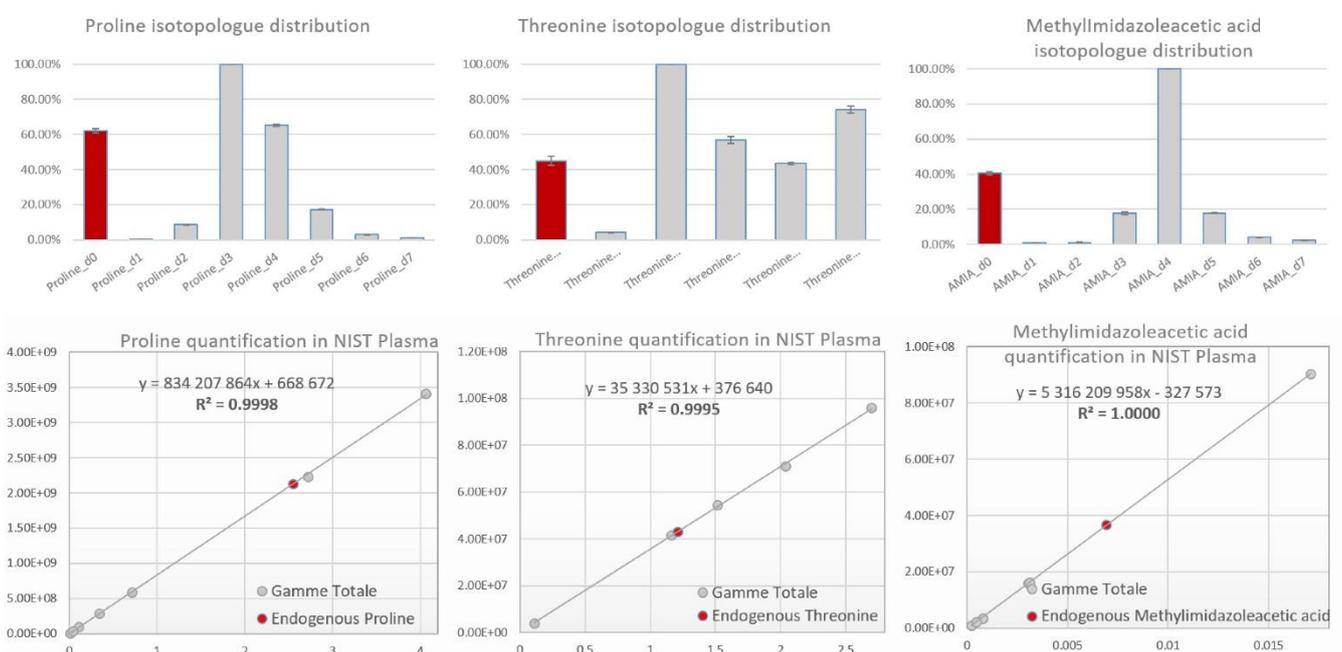


Figure 4: Isotopologue distribution of unlabeled and labeled Proline, Threonine and methylimidazoleacetic acid in NIST plasma and the corresponding multi-point internal calibration curves used for absolute quantification.

Table 3: Absolute concentrations measured with the isotope dilution approaches (multi-point internal calibration & external range + SIL-IS) for the 15 selected endogenous compounds. Comparison with reference values in plasma from SRM® 1950 certificate of analysis and HMDB.

Metabolite	Multi-point internal calibration	External range + SIL-IS		Ref values in plasma ^b	
	Measured C° (µg/mL)	Measured C° (µg/mL)	Relative difference ^a	Reported C° for SRM1950 (µg/mL)	Reported C° in HMDB (µg/mL)
Leucine	20.7	21.6	+4.4%	13.2	9.18 – 23.8
Methylimidazoleacetic Acid	0.069	0.071	+3.8%	ND	0.010 – 0.138
Imidazolelactic Acid	0.030	0.033	9.6%	ND	NQ
3-Methylhistidine	1.22	1.16	-5.2%	ND	0 – 0.998
Pipecolic Acid	0.40	0.38	-5.0%	ND	0.141 – 0.480
Proline	25.0	22.4	-11.4%	20.4	12.1 – 41.4
Threonine	11.9	9.78	-17.9%	14.2	9.53 – 26.21
Sarcosine	0.91	0.94	3.3%	ND	0 – 0.178
Taurine	NA	4.42	NA	ND	7.17 – 16.11
Serine	5.73	4.95	-13.6%	10.1	5.26 – 21.0
trans-4-Hydroxyproline	1.05	1.03	-1.5%	ND	0.525 – 3.93
Citrulline	4.77	4.54	-4.7%	ND	2.63 – 7.88
Glutamic Acid	8.63	8.09	-6.2%	ND	2.43 – 16.8
Ornithine	9.53	7.90	-17.1%	ND	3.97 – 14.5
Arginine	13.9	11.2	-19.8%	14.2	6.97 – 22.6

All quantified compounds (except sarcosine) were measured in the concentration range reported in the literature or in public databases (Ref HMDB). A number of metabolites, were already quantified in NIST plasma and their absolute concentration are reported in Table 3. Differences in absolute concentrations measured with our methods compared to values reported in the literature (publi NIST 2013 + analysis certificate) may be due to sample preparation methods and analytical methods being used. As an example, concentration determined by NMR are very accurate because they result from direct measurement of the metabolite in a sample obtained with minimum preparation.

Globally, the classical isotopic dilution strategy and the multi-point internal calibration gave comparable results with relative differences below 10% for the majority of the quantified metabolites: 9 out of 15. For these 9 compounds, linear regression curves were obtained with all quantification methods with $R^2 > 0.999$. One metabolite, Taurine, could not be quantified by multi-point internal calibration, because the calibration curve obtained with the different isotopologues was not linear. The 5 other compounds (names) were out of the linear range for the multi-point calibration due to signal saturation on high concentration (often 10X but also 5X) and quadratic curves were designed to quantify the metabolites. For the external calibration with SIL-IS, all curve were linear ($r^2 > 0.997$). For the multi-point internal calibration, a linear curve was obtained for 12 compounds ($r^2 > 0.997$). For Imidazolelactic acid and the 3-Methylhistidine, the calibration curve was no linear and the endogenous

was not comprise in the range. For serine, with the multi-point internal calibration, only 2 isotopologues can be used for the curve. So the artificial point corresponding to the sum of the area of all isotopologues was added.

▪ CONCLUSION

A fast and convenient way of producing deuterium-labeled compounds using a one-step metal-catalyzed reaction is presented. For 29 compounds out of 52, the labeling protocol afforded a deuterated version of the metabolite without remaining unlabeled material (no M0 isotopologue). In most of the cases, we observed that the labeling was quite heterogeneous, resulting in an isotopologue distribution, despite the expected regioselectivity usually encountered for the labeling of amine and/or alcohol containing molecule when using ruthenium as catalyst. Nevertheless, such isotope distribution could be exploited to develop a new methodology of quantification, called multi-point internal calibration. This method was evaluated for the quantification of 21 metabolites (6 exogenous + 15 endogenous) and compared with more traditional quantification methods. The different quantitative approaches gave similar results. The comparison to reference values reported for plasma (NIST SRM 1950 plasma analysis certificate and HMDB values) showed that the measured value are globally in the expected range of concentrations. Thus, the new quantitative method developed proves reliable and can be used to save analytical time since no external range are required to get calibration curves. In conclusion, we can produce on-demand SIL-IS for the quantification of metabolites and the new method of quantification, the multi-point internal calibration, allows to save a lot of time from an analytical point of view while keeping the same performances as the more classical methods. The multi-point internal calibration will make it possible to compare metabolomics analyses carried out at more or less long time and thus to make the analyses more robust.

▪ ASSOCIATED CONTENT

Supporting information

Additional information as noted in text. The Supporting Information is available free of charge on the ACS Publications website at DOI: XXX

▪ AUTHOR INFORMATION

Corresponding author

*E-mail: Annelaure.damont@cea.fr. Phone : +33-186-087350.

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) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics* **2001**, *2* (3), 155–168. <https://doi.org/10.1002/cfg.82>.
- (2) Junot, C.; Fenaille, F.; Colsch, B.; Bécher, F. High Resolution Mass Spectrometry Based Techniques at the Crossroads of Metabolic Pathways: HIGH RESOLUTION MASS SPECTROMETRY FOR METABOLOMICS. *Mass Spec Rev* **2014**, *33* (6), 471–500. <https://doi.org/10.1002/mas.21401>.
- (3) Zhou, J.; Yin, Y. Strategies for Large-Scale Targeted Metabolomics Quantification by Liquid Chromatography-Mass Spectrometry. *The Analyst* **2016**, *141* (23), 6362–6373. <https://doi.org/10.1039/C6AN01753C>.
- (4) Matuszewski, B. K. Standard Line Slopes as a Measure of a Relative Matrix Effect in Quantitative HPLC–MS Bioanalysis. *Journal of Chromatography B* **2006**, *830* (2), 293–300. <https://doi.org/10.1016/j.jchromb.2005.11.009>.
- (5) Nilsson, L. B.; Eklund, G. Direct Quantification in Bioanalytical LC-MS/MS Using Internal Calibration via Analyte/Stable Isotope Ratio. *J. Pharm. Biomed. Anal.* **2007**, *43* (3), 1094–1099. <https://doi.org/10.1016/j.jpba.2006.09.030>.
- (6) Joyce, R.; Kuziene, V.; Zou, X.; Wang, X.; Pullen, F.; Loo, R. L. Development and Validation of an Ultra-Performance Liquid Chromatography Quadrupole Time of Flight Mass Spectrometry Method for Rapid Quantification of Free Amino Acids in Human Urine. *Amino Acids* **2016**, *48* (1), 219–234. <https://doi.org/10.1007/s00726-015-2076-0>.
- (7) Khamis, M. M.; Adamko, D. J.; El-Aneed, A. Strategies and challenges in method development and validation for the absolute quantification of endogenous biomarker metabolites using liquid chromatography-tandem mass spectrometry. *Mass Spectrometry Reviews* **2019**, *00*, 1–22. <https://doi.org/10.1002/mas.21607>.
- (8) Liu, G.; Ji, Q. C.; Arnold, M. E. Identifying, Evaluating, and Controlling Bioanalytical Risks Resulting from Nonuniform Matrix Ion Suppression/Enhancement and Nonlinear Liquid Chromatography–Mass Spectrometry Assay Response. *Anal. Chem.* **2010**, *82* (23), 9671–9677. <https://doi.org/10.1021/ac1013018>.
- (9) Pieters, G.; Taglang, C.; Bonnefille, E.; Gutmann, T.; Puente, C.; Berthet, J.-C.; Dugave, C.; Chaudret, B.; Rousseau, B. Regioselective and Stereospecific Deuteration of Bioactive Aza Compounds by the Use of Ruthenium Nanoparticles. *Angew. Chem., Int. Ed.* **2014**, *53* (1), 230–234. <https://doi.org/10.1002/anie.201307930>.
- (10) Taglang C, Martínez-Prieto L, Del Rosal I, Maron L, Poteau R, Philippot K, Chaudret B, Perato S, Sam Lone A, Puente C, Dugave C, Rousseau B, Pieters G. Enantiospecific C-H Activation Using Ruthenium Nanocatalysts. *Angew. Chem.* **2015**, *127*, 10519-10519.
- (11) Michelotti A, Rodrigues F, Roche M. Development and scale-up of stereoretentive α -deuteration of amines. *Org. Process Res. Dev.* **2017**, *21*(11), 1741-1744.
- (12) Maegawa T, Fujiwara Y, Inagaki Y, Monguchi Y, Sajiki H. A convenient and effective method for the regioselective deuteration of alcohols. *Adv. Synth. Catal.* **2008**, *350*(14-15), 2215-2218.
- (13) Fujiwara Y, Iwata H, Sawama Y, Monguchi Y, Sajiki H. Method for regio-, chemo- and stereoselective deuterium labeling of sugars based on ruthenium-catalyzed C–H bond activation. *Chem. Commun.* **2010**;46(27):4977-4979.
- (14) Atzrodt, J.; Derdau, V. Pd- and Pt-Catalyzed H/D Exchange Methods and Their Application for Internal MS Standard Preparation from a Sanofi-Aventis Perspective. *Journal of Labelled Compounds and Radiopharmaceuticals* **2010**, *53* (11–12), 674–685. <https://doi.org/10.1002/jlcr.1818>.

Les données supplémentaires sont présentées en Annexe 5.

4 – Principaux résultats et conclusion

Dans un premier temps, une preuve de concept basée sur six composés a été réalisée en comparant les concentrations obtenues par « multi-point internal calibration », gamme externe avec ou sans standards marqués et également comparées par rapport à la concentration réelle supplémentée dans le plasma NIST. Les concentrations obtenues par la « multi-point internal calibration » et la gamme externe avec SIL-IS donne des valeurs très proches avec une bonne précision de mesure par rapport à la concentration réellement supplémentée dans l'échantillon (écart < 20 %). La quantification avec une gamme externe sans standard interne marqué révèle le caractère prononcé de l'effet matrice pour ces composés dans le plasma avec des écarts allant jusqu'à 64 % par rapport à la concentration théorique, d'où l'importance de l'utilisation des standards internes marqués (Figure 65).

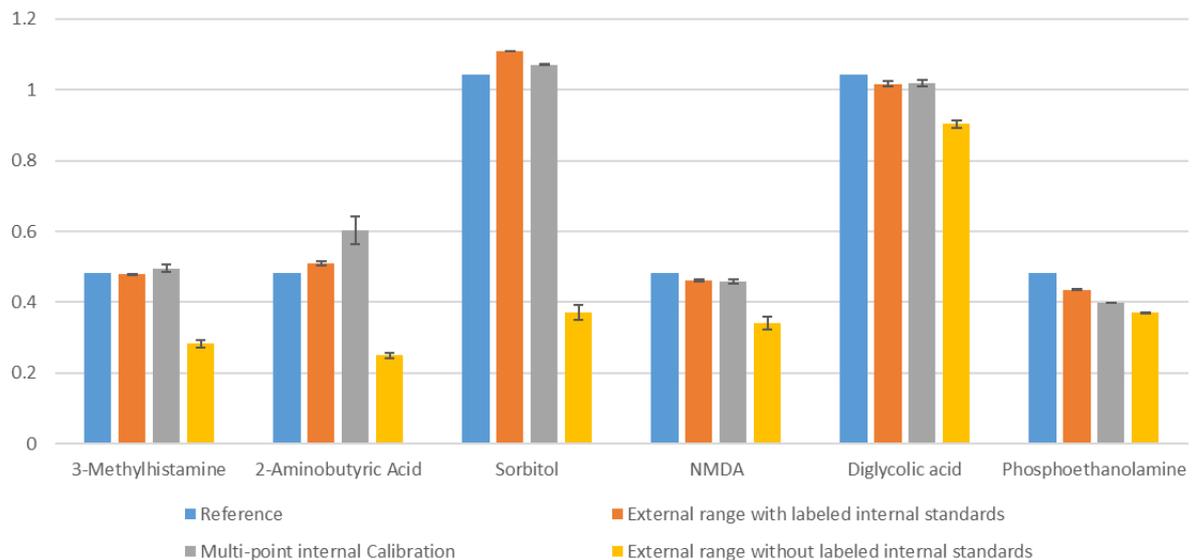


Figure 65 : Comparaison des mesures de concentrations obtenues par trois méthodes de quantification pour six composés supplémentés au plasma NIST..

Ensuite, les méthodes de quantification « multi-point internal calibration » et par gamme externe avec SIL-IS ont été appliquées à la quantification de 15 composés endogènes au plasma NIST. Exceptée la taurine qui n'a pas pu être quantifiée par la méthode « multi-point internal calibration » (gamme réalisée avec les isotopologues non linéaires certainement dû à un interférent), pour neuf composés sur les quatorze quantifiés, l'écart de concentration obtenu entre la quantification par gamme externe avec SIL-IS et par cette méthode innovante de quantification est inférieur à 10 %. Excepté pour la sarcosine, l'ensemble des concentrations obtenues par gamme externe avec SIL-IS ou par « multi-point internal calibration » sont comprises dans les gammes de concentration référencées dans HMDB.

La « multi-point internal calibration » montre donc de très bonnes performances pour la quantification de métabolites présents dans des matrices biologiques. En effet, les résultats obtenus par cette méthode sont très similaires à ceux obtenus par quantification par gamme externe avec standards internes marqués, qui a déjà montré son efficacité dans de nombreuses applications. L'avantage principal de la « multi-point internal calibration » par rapport à la gamme externe avec SIL-IS est le gain de temps. En effet, avec la méthode développée il n'est plus nécessaire de réaliser des gammes de calibration externe avant et après les analyses des échantillons d'intérêt.

En conclusion, la nouvelle approche de quantification par calibration interne s'avère tout aussi performante que la méthode plus usuelle de quantification par gamme externe avec standards internes marqués. Les deux méthodes peuvent s'avérer très complémentaires selon le type de composés marqués disponibles et la présence ou non d'interférents dans les analyses. On notera que les SIL-IS produits à façon au laboratoire peuvent être utilisés pour l'une ou l'autre des approches quantitatives (gamme interne ou gamme externe) et on peut également envisager leur utilisation en quantification par spectrométrie de masse en tandem. Les performances de quantification obtenues avec nos SIL-IS produits par échange H/D métallo-catalysé, nous ont conduits à développer d'autres protocoles de marquage complémentaires pour élargir la gamme de produits marqués disponibles au laboratoire : cela fait l'objet de la suite du manuscrit.

C – Evaluation de l'efficacité de méthodes alternatives de deutération pour l'obtention de versions marquées de métabolites d'intérêt

1 - Méthodologie

D'autres méthodes de deutération que celle préalablement présentée, ont été investiguées et en particulier, leurs performances en terme d'efficacité d'incorporation de deutérium dans les structures chimiques ont été évaluées. Pour des raisons de propriétés industrielles liées au dépôt d'un brevet, les méthodologies de marquage utilisées ne pourront pas être détaillées dans ce manuscrit.

Au total, 110 composés ont été sélectionnés dans notre chimiothèque principalement constituée de métabolites endogènes. La sélection s'est principalement orientée vers des composés comportant des fonctions chimiques pas ou peu compatibles avec les conditions de marquages réductrices présentées au chapitre précédent. Finalement, 103 composés ont pu être engagés dans les tests de deutération (entre quatre et huit méthodes ont pu être évaluées selon les molécules). Le marquage des sept composés restants n'a pu être évalué faute de solubilité de ceux-ci dans les solvants utilisés pour le marquage (Annexe 6).

L'ensemble de ces composés, avant et après tentative de marquage au deutérium, ont été analysés par LC-HRMS/MS en mode « Full Scan » et en mode DDA. La méthode chromatographique utilisée a été la même que celle décrite pour l'analyse des standards marqués par la réaction mono-étape métallo-catalysée avec la colonne HILIC (Partie B – Chapitre 2 – A – 4). Pour la partie MS, l'ensemble des échantillons ont été analysés avec

l'Orbitrap Fusion (Thermo Fisher Scientific, Courtaboeuf, France). Concernant les paramètres de source, une source d'ionisation ESI est installée avec les paramètres suivants : une tension de + 4000 V ou – 2800 V est appliquée au capillaire métallique de nébulisation qui est chauffé à 100°C, le débit pour les « sheath gaz » est de 40 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 1 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C. Les acquisitions sont réalisées en mode DDA. Les « Full Scan » sont acquis avec l'Orbitrap à une résolution de 120 000 (à m/z 200) avec une fenêtre de balayage de 75 à 550 m/z . L'« AGC target », est fixé à $5 \cdot 10^4$ ions et le temps maximal d'injection à 50 ms. Les spectres MS/MS sont également acquis avec l'Orbitrap à une résolution de 15 000 (à m/z 200). L'« AGC target » est fixé à $5 \cdot 10^4$ avec un temps maximal d'injection de 22 ms. L'isolement des ions se fait grâce au quadripôle avec une fenêtre de +/- 0.8 Da. Les fenêtres d'acquisition des spectres débutent à 75 m/z . Afin d'obtenir le plus d'informations sur les composés, notamment sur la position des atomes de deutériums dans la molécule, chaque ion sélectionné a été fragmenté à trois niveaux d'énergie en mode « stepped collision energy » : 20 +/- 10 % ; 50 +/- 10 % et 80 +/- 10 %. La sélection des ions est réalisé grâce au filtre d'intensité fixé à $1 \cdot 10^5$ et à l'exclusion dynamique des ions cinq secondes après trois sélections en quatre secondes (Figure 66).

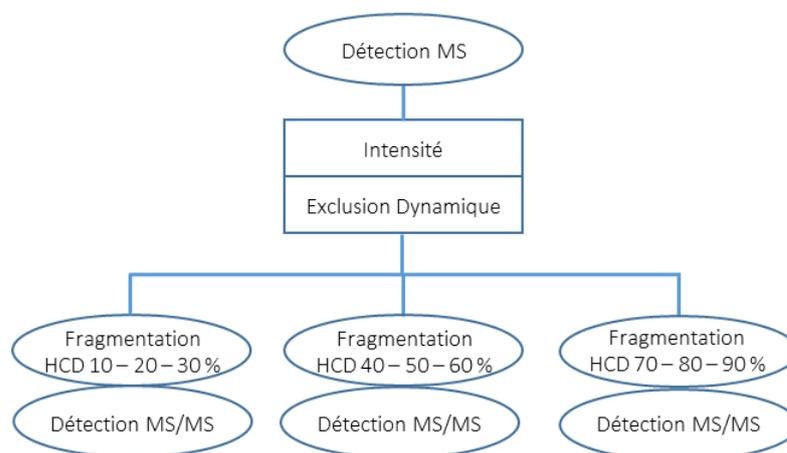


Figure 66 : Méthode DDA pour les analyses LC-HRMS/MS de composés deutérés.

Les données ont été traitées manuellement avec Xcalibur v.2.2.44 (ThermoFisher Scientific) pour voir, dans un premier temps, quels étaient les taux d'incorporation de deutérium dans les structures.

2 – Résultats

Différentes conditions de marquage ont été testées. Pour chacune des conditions, les spectres de masse ont été acquis en mode d'ionisation positif et négatif pour évaluer l'efficacité de marquage. Les résultats peuvent être résumés pour chaque composé en format tabulaire où, pour chaque condition, l'abondance relative de chaque isotopologue deutéré du massif isotopique est rapportée (Figure 67). Les analyses « Full Scan » ayant été réalisées sur l'Orbitrap Fusion à très haute résolution, les isotopologues deutérés et les isotopologues C-13, correspondant à la contribution naturelle de ce dernier isotope, ont pu être séparés. Ainsi, dans les tableaux de résultats, seules figurent les intensités relatives des isotopologues deutérés. Les distributions isotopiques obtenues dans différentes conditions de marquage pour l'acide xanthurénique sont présentées, à titre d'exemple, dans la Figure 67. On constate que selon les conditions de marquage appliquées, l'efficacité de la deutération est plus ou moins marquée et l'isotopologue majoritaire est variable. Pour une utilisation en tant que SIL-IS, la condition 5 sera retenue pour produire une version très majoritairement tri-deutérée de l'acide xanthurénique. Les positions de marquage restent, pour l'instant, à confirmer par RMN ^1H et RMN D dans la mesure où quatre positions de deutération sont possibles sur le bicyclic hétéroaromatique.

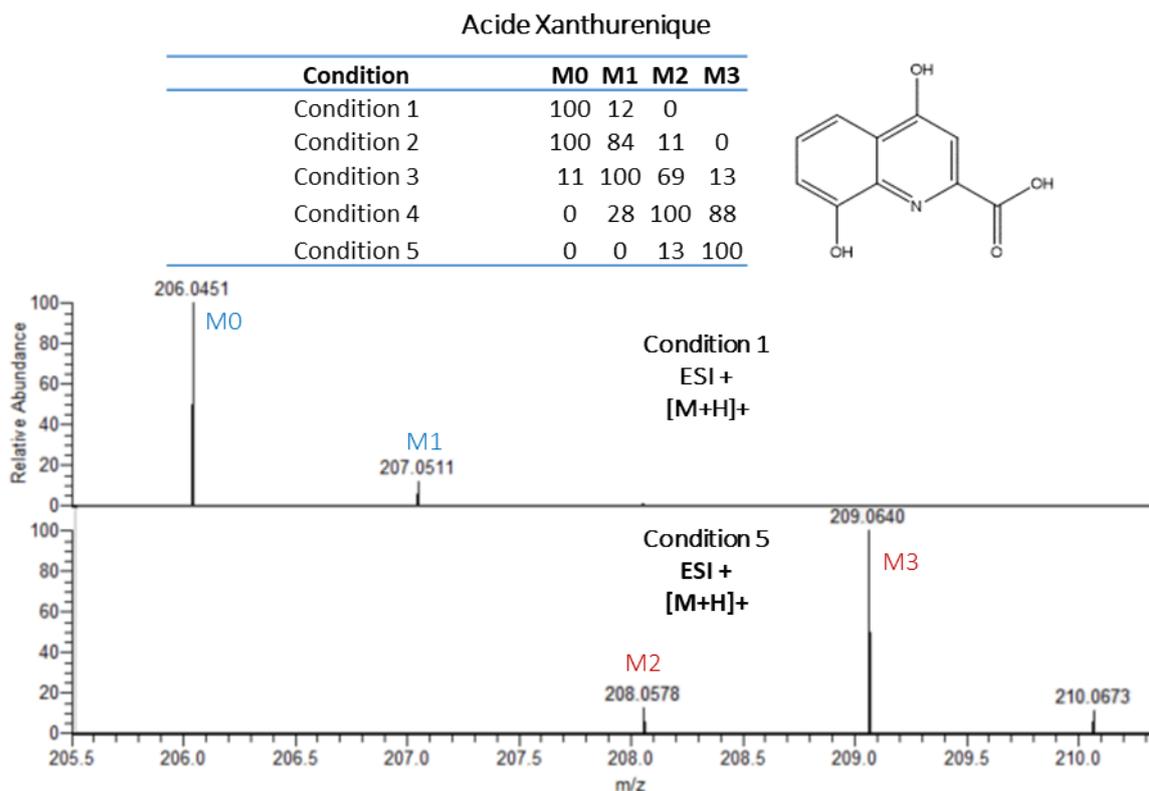


Figure 67 : Exemple des massifs isotopiques obtenus pour le marquage de l'acide xanthurénique analysé en mode d'ionisation positif.

Le Tableau 17 résume les massifs isotopiques obtenus dans les meilleures conditions de deutération pour chacun des 103 composés

Tableau 17 : Bilan des massifs isotopiques obtenus par la méthode alternative de deutération.

Composés	Famille	Formule Brute	Masse précise	d0	d1	d2	d3	d4	d5	d6
Thiamine pyrophosphate	Vitamine & dérivé	C12H18N4O7P2S	424.0371	0	100	0	0	0	0	0
Deoxyadenosine	Nucléoside & dérivé	C10H13N5O3	251.1018	0	100	0	0	0	0	0
7-Methylxanthine	Hétéroaromatique	C6H6N4O2	166.0491	0	100	0	0	0	0	0
4-Hydroxyindole	Hétéroaromatique	C8H7NO	133.0528	0	0	0	0	33	100	0
5-Hydroxyindole	Hétéroaromatique	C8H7NO	133.0528	0	0	100	0	0	0	0
5-Aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5'-monophosphate	Nucléoside & dérivé	C9H15N4O8P	338.0627	0	100	0	0	0	0	0
AICAR	Nucléoside & dérivé	C9H14N4O5	258.0964	0	100	0	0	0	0	0
Imidazoleacetic acid	Acide organique	C5H6N2O2	126.0429	0	0	0	23	100	0	0
Methylimidazoleacetic acid	Acide organique	C6H8N2O2	140.0586	0	0	0	28	100	0	0
α-Methylhistamine	Hétéroaromatique	C6H11N3	125.0953	0	100	12	0	0	0	0
L-Histidine methyl ester	Acide aminé & dérivé	C7H11N3O2	169.0851	0	100	84	1	0	0	0
Xanthurenic acid	Hétéroaromatique	C10H7NO4	205.0375	0	0	11	100	0	0	0
4-Acetamidobutanoic acid	Acide organique	C6H11NO3	145.0739	0	1	23	100	0	0	0
Dihydrouracil	Nucléobase	C4H6N2O2	114.0429	0	8	100	0	0	0	0
DL-Dihydroorotic acid	Acide organique	C5H6N2O4	158.0328	0	50	100	0	0	0	0
Gly-Pro	Peptide	C7H12N2O3	172.0848	0	0	9	100	0	0	0
N acetyl L glutamine	Acide aminé & dérivé	C7H12N2O4	188.0797	0	0	0	1	16	100	28
N-Acetyl-D-glucosamine	Hexose	C8H15NO6	221.0899	0	100	0	0	0	0	0
N-acetyl-DL-norvaline	Acide aminé & dérivé	C7H13NO3	159.0895	0	1	100	31	0	0	0
N-Acetyl-DL-serine	Acide aminé & dérivé	C5H9NO4	147.0532	0	0	1	15	100	0	0
N-Acetylglycine	Acide aminé & dérivé	C4H7NO3	117.0426	0	0	0	2	28	100	0
N-Acetyl-L-alanine	Acide aminé & dérivé	C5H9NO3	131.0582	0	0	7	100	11	0	0
N-Acetyl-L-aspartic acid	Acide aminé & dérivé	C6H9NO5	175.0481	0	0	11	100	77	15	1
N-Acetyl-L-glutamic acid	Acide aminé & dérivé	C7H11NO5	189.0637	0	0	9	100	17	0	0
N-Acetyl-L-leucine	Acide aminé & dérivé	C8H15NO3	173.1052	0	0	10	100	0	0	0
N-Acetyl-ornithine	Acide aminé & dérivé	C7H14N2O3	174.1004	0	0	2	33	100	0	0
N-Methyl-gamma-oxo-3-pyridinebutanamide	Hétéroaromatique	C10H12N2O2	192.0899	0	0	8	47	100	0	0
Phenylacetyl-L-glutamine	Acide aminé & dérivé	C13H16N2O4	264.1110	0	0	0	11	100	19	0
Betaine	Acide aminé & dérivé	C5H11NO2	117.0790	0	0	100	0	0	0	0
cis-Aconitic acid	Acide organique	C6H6O6	174.0164	0	0	0	100	0	0	0
Isocitric acid	Acide organique	C6H8O7	192.0270	0	33	100	36	11	0	0
L-Aspartic acid	Acide aminé & dérivé	C4H7NO4	133.0375	0	0	11	100	0	0	0
α-Ketoglutaric acid	Acide organique	C5H6O5	146.0215	0	0	3	34	100	0	0
N acetyl DL glutamic acid	Acide aminé & dérivé	C7H11NO5	189.0637	0	1	22	100	12	0	0
N-Acetyl-D-allo-isoleucine	Acide aminé & dérivé	C8H15NO3	173.1052	0	2	29	100	0	0	0
N-Acetyl-L-phenylalanine	Acide aminé & dérivé	C11H13NO3	207.0895	0	4	35	100	0	0	0
Cotinine	Hétéroaromatique	C10H12N2O	176.0950	1	22	100	0	0	0	0
paraxanthine	Hétéroaromatique	C7H8N4O2	180.0647	1	100	0	0	0	0	0
Kynurenic acid	Hétéroaromatique	C10H7NO3	189.0426	1	100	0	0	0	0	0
2'-O-Methylguanosine	Nucléoside & dérivé	C11H15N5O5	297.1073	1	100	0	0	0	0	0
Citric acid	Acide organique	C6H8O7	192.0270	1	4	26	69	100	0	0
2'-O-Methyladenosine	Nucléoside & dérivé	C11H15N5O4	281.1124	1	100	0	0	0	0	0
7-methylguanine	Nucléoside & dérivé	C6H7N5O	165.0651	1	100	0	0	0	0	0
Indolelactic acid	Acide organique	C11H11NO3	205.0739	1	100	89	24	3	0	0
L-Histidinol	Acide aminé & dérivé	C6H11N3O	141.0902	1	87	100	0	0	0	0
Carnosine	Acide aminé & dérivé	C9H14N4O3	226.1066	1	82	100	23	0	0	0
5-Aminoimidazole-4-carboxamide	Nucléobase	C4H6N4O	126.0542	1	100	1	0	0	0	0
Allantoin	Hétérocyclique	C4H6N4O3	158.0440	2	100	0	0	0	0	0
L-Anserine	Peptide	C10H16N4O3	240.1222	2	100	45	5	0	0	0
2'-O-Methylinosine	Nucléoside & dérivé	C11H14N4O5	282.0964	2	100	0	0	0	0	0
Imidazolelactic acid	Acide organique	C6H8N2O3	156.0535	2	100	47	0	0	0	0

Partie B : Etudes Expérimentales – Chapitre 2

Composés	Famille	Formule Brute	Masse précise	d0	d1	d2	d3	d4	d5	d6
N6-Acetyl-L-lysine	Acide aminé & dérivé	C8H16N2O3	188.1161	2	28	88	100	0	0	0
3-Methylhistidine	Acide aminé & dérivé	C7H11N3O2	169.0851	3	100	26	1	0	0	0
N-Acetylhistamine	Hétéroaromatique	C7H11N3O	153.0902	3	100	26	1	0	0	0
trans-Zeatin glucoside	Nucléoside & dérivé	C16H23N5O6	381.1648	3	100	0	0	0	0	0
4-Imidazoleacrylic acid (urocanic acid)	Acide organique	C6H6N2O2	138.0429	3	100	16	0	0	0	0
Histamine	Hétéroaromatique	C5H9N3	111.0797	3	100	23	0	0	0	0
Dihydrothymine	Nucléobase	C5H8N2O2	128.0588	3	100	0	0	0	0	0
D-Glutamine	Acide aminé & dérivé	C5H10N2O3	146.0691	3	10	100	0	0	0	0
1-Methylhistamine	Hétéroaromatique	C6H11N3	125.0953	3	100	86	0	0	0	0
o-Tyrosine	Acide aminé & dérivé	C9H11NO3	181.0739	3	35	100	93	0	0	0
Stachydrine	Acide aminé & dérivé	C7H13NO2	143.0946	4	100	0	0	0	0	0
Glycocholic acid	Acide biliaire	C26H43NO6	465.3091	4	33	83	100	75	8	0
Adenosine	Nucléoside & dérivé	C10H13N5O4	267.0967	6	100	0	0	0	0	0
5'-Deoxy-5'-(methylthio)adenosine	Nucléoside & dérivé	C11H15N5O3S	297.0896	9	100	0	0	0	0	0
Purine	Nucléobase	C5H4N4	120.0436	12	100	0	0	0	0	0
Thiamine	Nucléobase	C12H16N4OS	264.1045	20	96	100	17	5	0	0
Adenine	Nucléobase	C5H5N5	135.0545	21	100	0	0	0	0	0
N-Isovaleroylglycine	Acide aminé & dérivé	C7H13NO3	159.0895	22	85	100	35	3	0	0
Ergothioneine	Acide aminé & dérivé	C9H15N3O2S	229.0885	26	100	32	0	0	0	0
Indole	Hétéroaromatique	C8H7N	117.0578	29	100	0	0	0	0	0
Hexanoyl glycine	Acide aminé & dérivé	C8H15NO3	173.1052	30	90	100	43	5	0	0
N6-Methyladenine	Nucléobase	C6H7N5	149.0701	36	100	0	0	0	0	0
Trans-3-Hydroxy cotinine	Hétéroaromatique	C10H12N2O2	192.0899	73	100	0	0	0	0	0
Indole-2-carboxylic acid	Acide organique	C9H7NO2	161.0477	81	100	0	0	0	0	0
Succinic acid	Acide organique	C4H6O4	118.0266	93	100	40	6	0	0	0
Hypoxanthine	Hétéroaromatique	C5H4N4O	136.0385	97	100	0	0	0	0	0
Thiamine-PO4	Hétéroaromatique	C12H17N4O4PS	344.0708	100	30	17	24	3	0	0
1,7-Dimethyluracil	Hétéroaromatique	C7H8N4O3	196.0596	100	0	0	0	0	0	0
1-methyl guanine	Nucléoside & dérivé	C6H7N5O	165.0651	100	51	0	0	0	0	0
1-Methyladenosine	Nucléoside & dérivé	C11H15N5O4	281.1124	100	6	0	0	0	0	0
1-Methylxanthine	Hétéroaromatique	C6H6N4O2	166.0491	100	25	0	0	0	0	0
7-Methylguanosine	Nucléobase	C11H15N5O5	297.1073	100	0	0	0	0	0	0
9-Methyluracil	Hétéroaromatique	C6H6N4O3	182.0440	100	0	0	0	0	0	0
Caffeine	Hétéroaromatique	C8H10N4O2	194.0804	100	17	0	0	0	0	0
Cyclic AMP	Nucléoside & dérivé	C10H12N5O6P	329.0525	100	14	0	0	0	0	0
Guanine	Nucléobase	C5H5N5O	151.0494	100	24	0	0	0	0	0
S-Adenosyl homocysteine	Nucléoside & dérivé	C14H20N6O5S	384.1216	100	20	0	0	0	0	0
S-adenosylmethionine	Nucléoside & dérivé	C15H22N6O5S	398.1372	100	46	0	0	0	0	0
Theophylline	Hétéroaromatique	C7H8N4O2	180.0647	100	11	0	0	0	0	0
Xanthine	Hétéroaromatique	C5H4N4O2	152.0334	100	20	0	0	0	0	0
Xanthosine	Hétéroaromatique	C10H12N4O6	284.0757	100	7	0	0	0	0	0
3,7-Dimethyluracil	Nucléobase	C7H8N4O3	196.0596	100	0	0	0	0	0	0
Indole-3-acetyl-L-aspartic acid	Acide aminé & dérivé	C14H14N2O5	290.0903	100	51	7	0	0	0	0
Indoleacrylic acid	Acide organique	C11H9NO2	187.0633	100	1	0	0	0	0	0
2,4-Quinolinediol	Hétéroaromatique	C9H7NO2	161.0477	100	1	0	3	5	0	0
Isoxanthopterin	Hétéroaromatique	C6H5N5O2	179.0443	100	1	0	0	0	0	0
Adipic acid	Acide organique	C6H10O4	146.0579	100	13	1	0	0	0	0
Citraconic acid	Acide organique	C5H6O4	130.0266	100	1	4	0	0	0	0
Hydrocinnamic acid	Acide organique	C9H10O2	150.0681	100	1	0	0	0	0	0
L-Glutamic acid	Acide aminé & dérivé	C5H9NO4	147.0532	100	12	0	0	0	0	0
Maleic acid	Acide organique	C4H4O4	116.0110	100	1	0	0	0	0	0
Phenyllactic acid	Acide organique	C9H10O3	166.0630	100	0	0	0	0	0	0

L'analyse de l'ensemble des produits de marquage permet de dire que sur l'ensemble des 103 composés testés, 17 % ont montré un marquage totalement inefficace (comme dans le cas de la 7-méthylguanosine) ou un marquage trop peu efficace avec un isotopologue monoisotopique M0 (non-marqué) restant très majoritaire (les autres isotopologues étant présents à une intensité relative inférieure à 20 %). Des développements supplémentaires seront alors nécessaires spécifiquement pour améliorer le marquage de ces composés.

Par opposition, 41 % des composés testés présentent, après marquage, un massif isotopique ne comportant plus l'isotopologue monoisotopique M0 (ou à une abondance relative inférieure à 1 %). C'est le cas, par exemple de la *N*-acétylglycine et de l'acide *N*-acétylaspartique dont les spectres de masses obtenus sont présentés Figure 68. Globalement, pour ces composés pour lesquels l'incorporation de deutérium a été efficace, les massifs isotopiques obtenus sont beaucoup plus étroits qu'avec la méthode d'échange H/D mono-étape métal-catalysée. En effet, pour ces 43 composés ne comportant plus l'isotopologue monoisotopique, les massifs isotopiques présentent le plus souvent seulement deux isotopologues. Le nombre d'isotopologues varie néanmoins entre un et cinq.

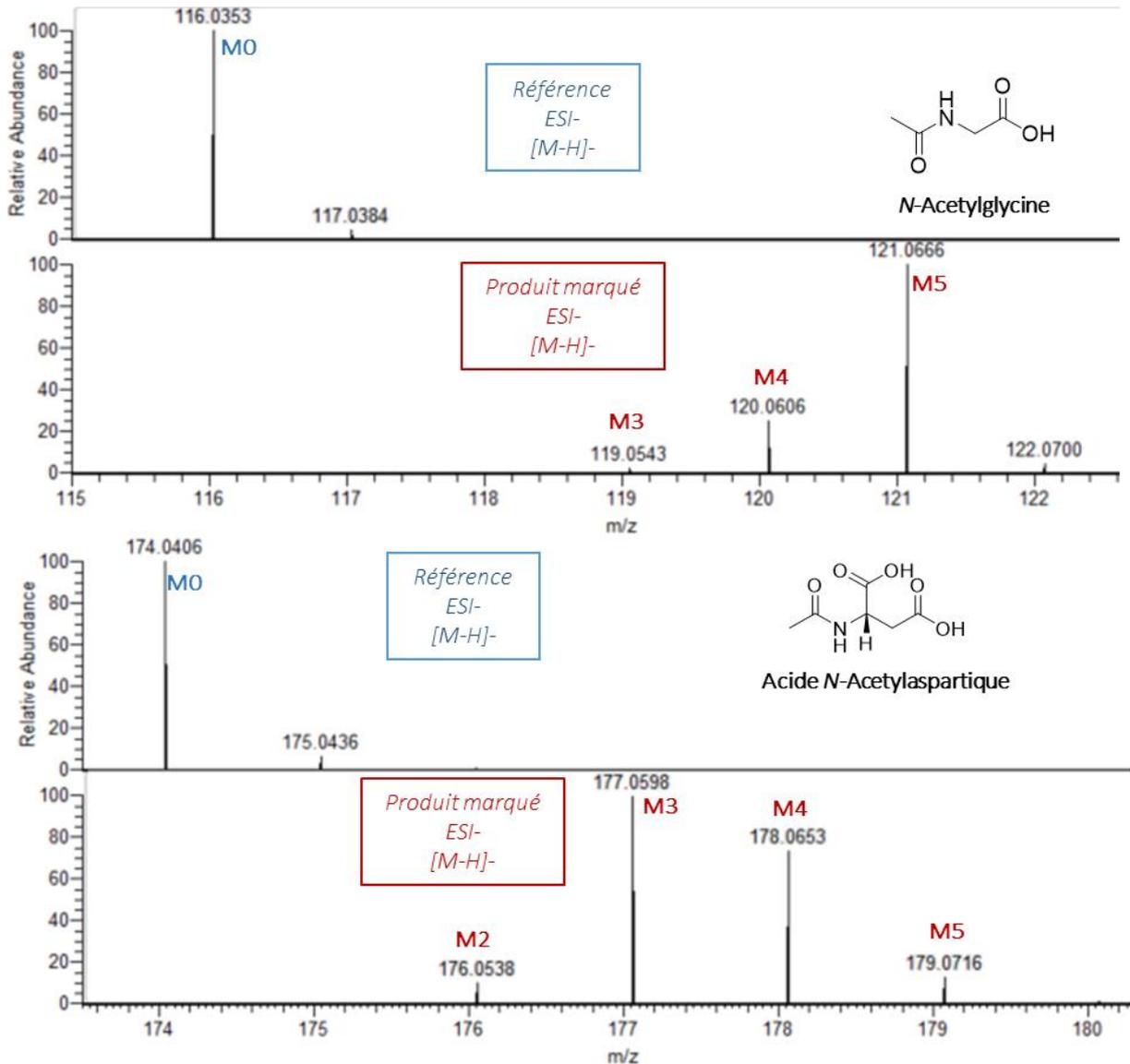


Figure 68 : Exemples de marquage efficace obtenus pour la *N*-acétylglycine et l'acide *N*-acétylaspartique

On notera que pour 11 de ces 43 composés, un seul isotopologue est présent et qu'il s'agit pour 13 d'entre eux du composé monodeutééré, pour deux d'entre eux du composé dideutééré (comme illustré ci-dessous dans la Figure 69 avec l'exemple de la bétaine) et pour un composé il s'agit de la forme trideutéérée (exemple de l'acide cis-aconitique dans la Figure 69).

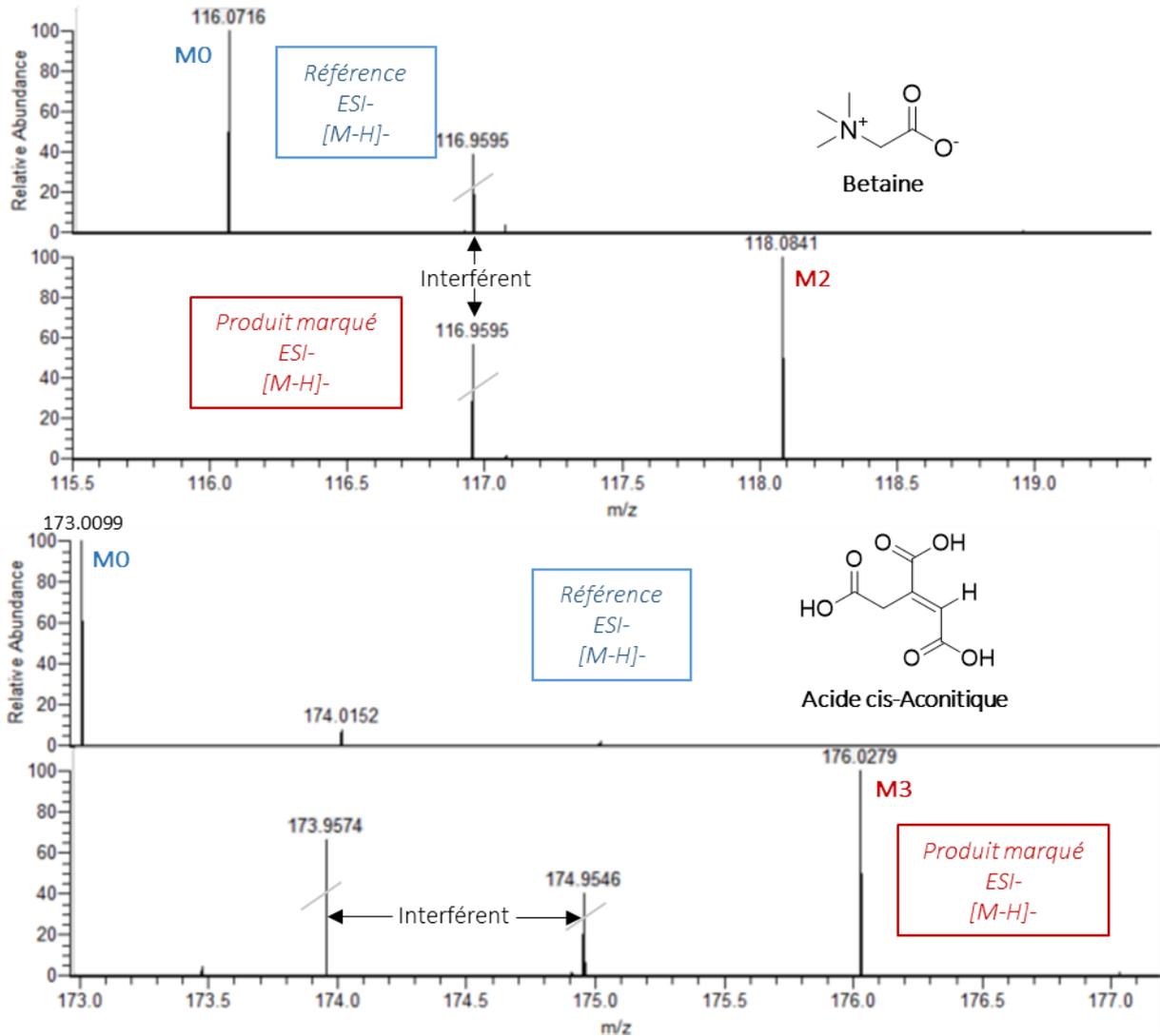


Figure 69 : Exemples de composés marqués comportant un isotopologue deutéré unique.

Pour les 42 composés restant, le marquage est incomplet, c'est-à-dire que le composé monoisotopique est présent en quantité non négligeable (Figure 70). Typiquement, la production d'acide succinique deutéré *via* la méthode d'échange H/D métallo-catalysée à partir de l'acide maléique (comme présenté dans la Chapitre 2 – A de la Partie études expérimentales) sera privilégiée en comparaison de l'approche présentée ici qui ne permet qu'un marquage très partiel du composé.

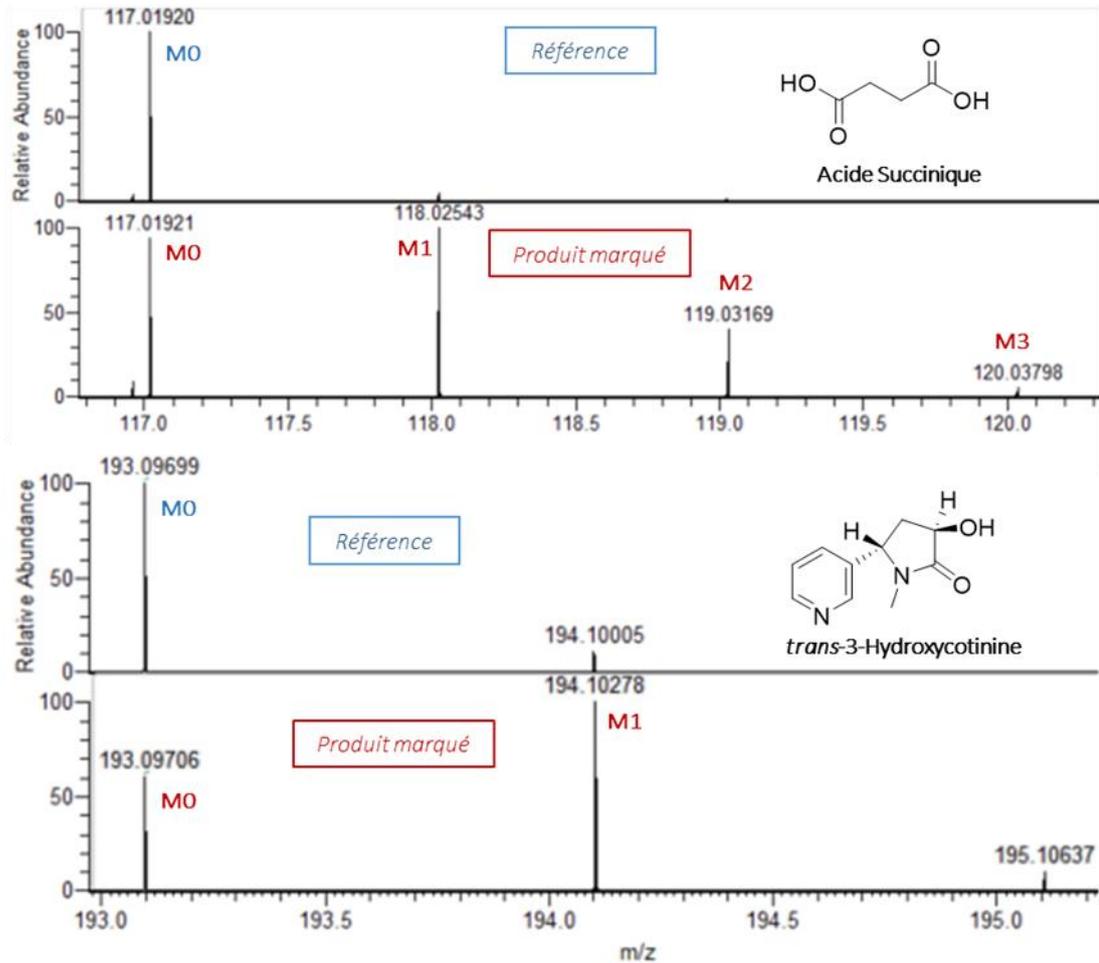


Figure 70 : Exemples de marquage incomplet de l'acide succinique et de la *trans*-3-hydroxycotinine.

Finalement, hormis les composés n'ayant pas ou peu été marqués (M0 majoritaire), les composés marqués ont pu être regroupés en 5 catégories basées sur la proportion ou abondance de l'isotopologue majoritaire au sein du massif isotopique global : ceux pour lesquels l'isotopologue majoritaire représente plus de 90% du massif isotopique ; entre 70 et 90% ; entre 50 et 70% ; moins de 50% ; et enfin les « non marqués » (ceux présentant majoritairement du M0). La répartition des composés dans ces 5 catégories est résumé dans la Figure 71.

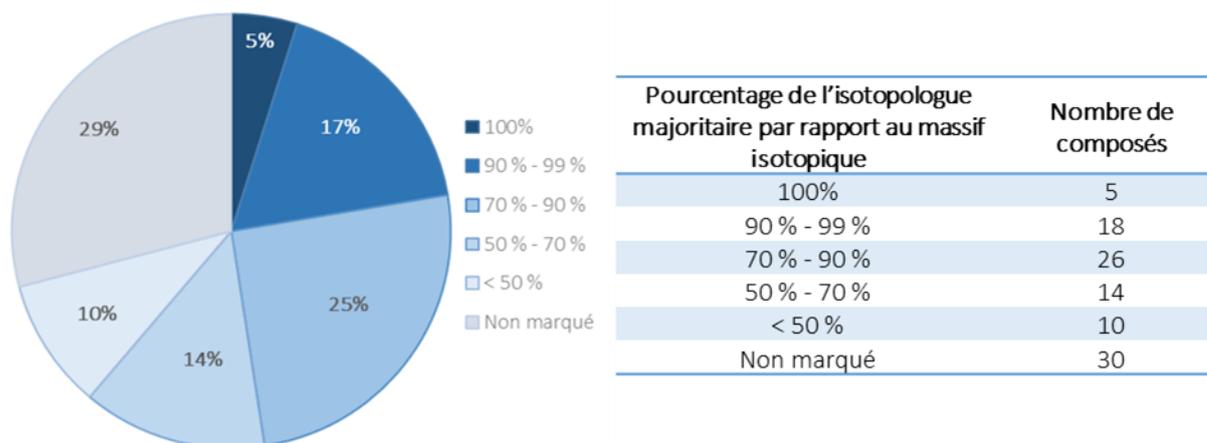


Figure 71 : Répartition des 103 composés investigués par la méthode de deutération alternative selon la proportion de l'isotopologue majoritaire par rapport à l'ensemble du massif isotopique.

Ces développements ont également permis de mettre en évidence certains effets isotopiques liés à la présence de deutérium dans les structures, par comparaison des spectres de masse acquis en mode d'ionisation positif et négatif. En effet, dans de rares cas, l'allure des massifs isotopiques d'un même produit marqué peut différer selon le mode d'ionisation. Par exemple, avec l'acide hippurique, lorsqu'on compare les massifs isotopiques des ions $[M+H]^+$ et $[M-H]^-$, on constate qu'en mode d'ionisation positif l'isotopologue majoritaire est le composé di-deutééré tandis qu'en mode d'ionisation négatif, l'isotopologue majoritaire est le composé mono-deutééré. Cela semble indiquer qu'en mode d'ionisation négatif, la déprotonation ne se fait pas uniquement sur la position acide carboxylique mais qu'il y a également une autre position, sur laquelle un deutérium a été incorporé, où se produit la déprotonation, expliquant le décalage du massif isotopique (**Error! Reference source not found.**). On peut ainsi supposer que les deutériums ont été incorporés dans la structure en alpha de la fonction acide carboxylique et qu'à cette position les protons ou deutériums sont suffisamment acides en phase gazeuse pour être arrachés et conduire à l'ion $[M-H]^-$, au même titre que le proton porté par la fonction acide carboxylique.

Acide Hippurique deutéré

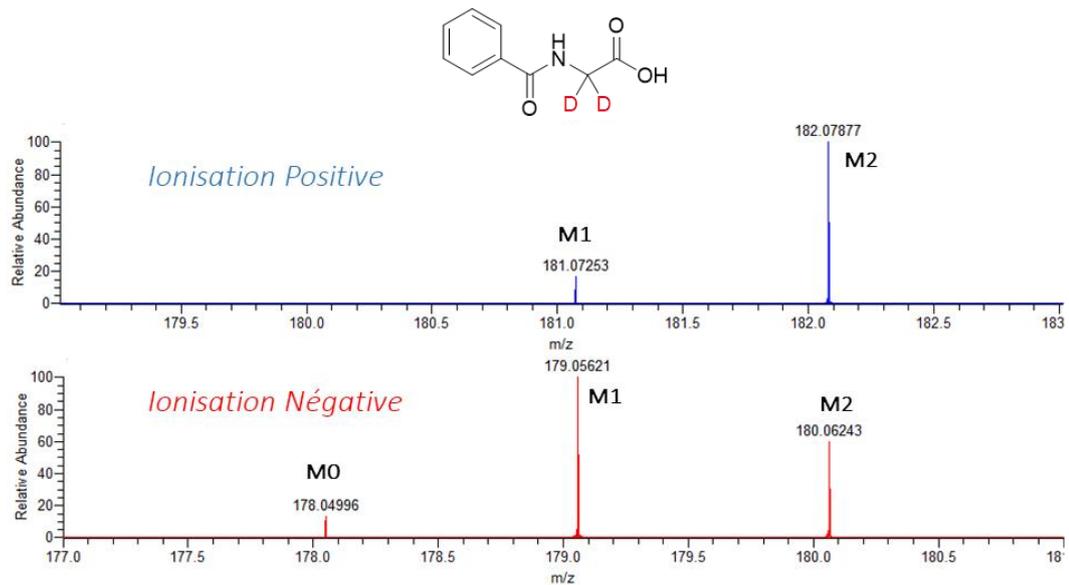


Figure 72 : Exemple d'effet isotopique lors de l'ionisation observé pour l'acide hippurique marqué.

Ces effets isotopiques complexifient l'interprétation des spectres mais permettent de mieux comprendre les processus d'ionisation mis en jeu lors des analyses par spectrométrie de masse. Ces aspects particulièrement intéressants seront approfondis lors de projet à venir.

L'analyse des spectres MS/MS des différents isotopologues deutérés, qui ont été acquis mais n'ont pour l'instant pas pu être pleinement exploités, permettront très probablement d'estimer les positions d'incorporation des deutériums. La Figure 73 présente les spectres de fragmentation de l'hexanoylglycine non marquée et des isotopologues deutérés majoritaires dans le produit marqué. Les ions fragments observés permettent d'annoter les spectres avec les différentes pertes caractéristiques de ce composé.

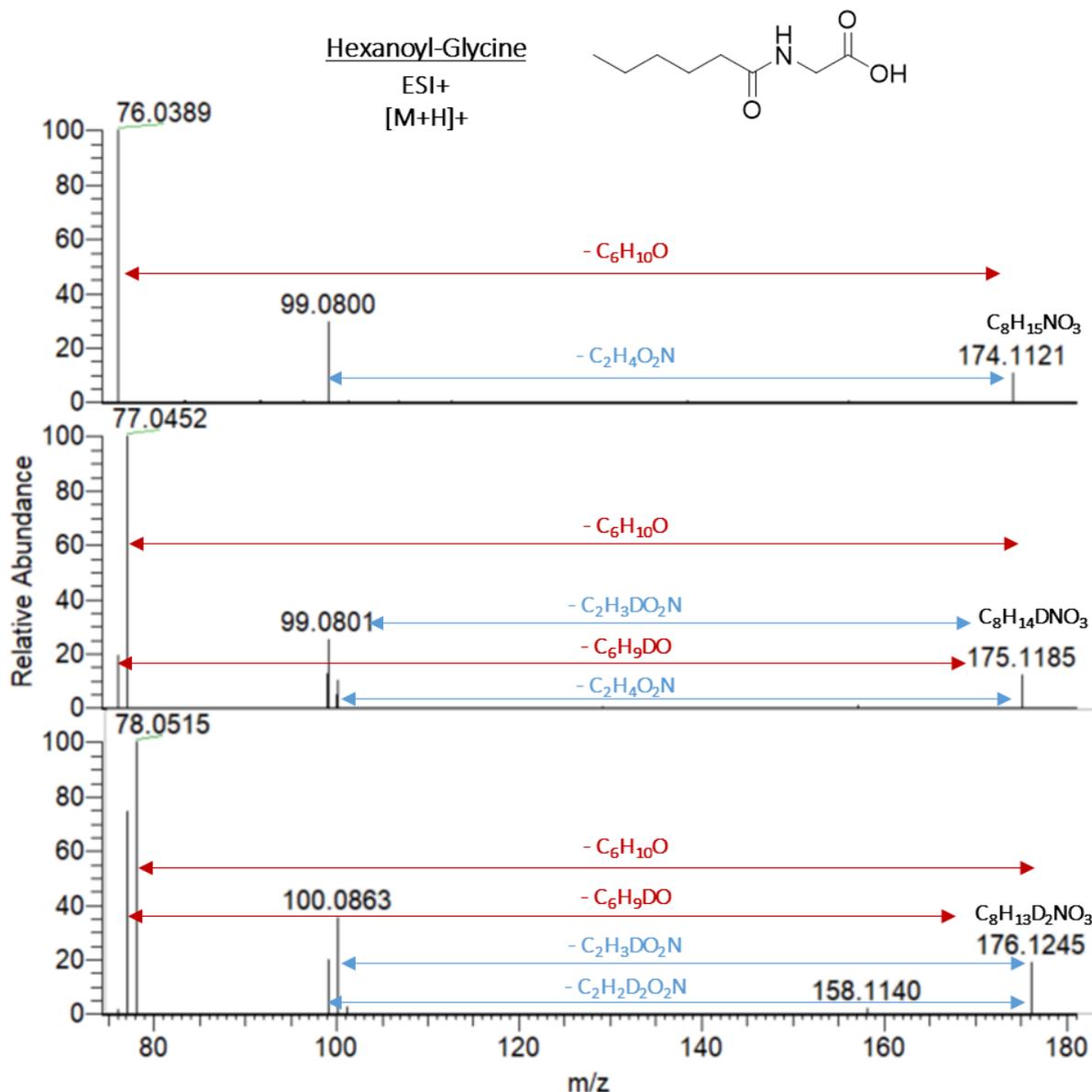


Figure 73 : Comparaison des spectres MS/MS indépendants des ions non-deuté, mono-deuté et di-deuté.

Notons que comme dans le cas de la production de composés deutérés par la méthode d'échange H/D métallo-catalysée, chaque composé « efficacement » marqué fera l'objet d'une caractérisation plus précise par RMN ¹H (et potentiellement par RMN D) en particulier pour valider les positions d'incorporation de deutérium. En résumé, les nouvelles conditions de production de produits marqués au deutérium sont très encourageantes et ont permis d'étoffer la gamme de produits marqués déjà disponibles par la voie de production métallo-catalysée. Des développements restent néanmoins à réaliser, en particulier pour les composés présentant des marquages incomplets et pour lesquels il reste de l'isotopologue non-marqué M0.

3 – Conclusions et perspectives

Les massifs isotopiques obtenus avec cette méthode alternative de production de SIL-IS par deutération sont très prometteurs. Des développements supplémentaires sont nécessaires pour affiner le marquage et répondre au mieux aux cahiers des charges initial.

Ces analyses ont également permis de mettre en évidence l'existence de certains effets isotopiques lors de l'ionisation. Cela peut également permettre de mieux appréhender les processus de fragmentation.

Cette méthode de production de standards deutérés est complémentaire à la réaction d'échange H/D mono-étape métallo-catalysée. Ces deux méthodes permettent d'obtenir un large panel de standards internes marqués.

Des expériences de RMN doivent être réalisées pour confirmer la position des deutériums dans la structure de la molécule.

D - Utilisation du marquage *In Vivo* pour la quantification et l'identification de métabolites

1 – Contexte et objectifs

Comme décrit précédemment dans la partie introductive du manuscrit, les signatures ou profils métaboliques peuvent conduire à une meilleure compréhension des phénomènes physiopathologiques et/ou conduire à la découverte de nouveaux biomarqueurs caractéristiques de certaines pathologies. Pour extraire les informations les plus pertinentes possibles lors des investigations métabolomiques il est nécessaire de pouvoir compter sur une fiabilité et une robustesse analytique maximale. La standardisation ou normalisation des données est d'une importance cruciale pour assurer la robustesse des résultats issus d'analyses métabolomiques par LC-HRMS. De la même façon, les techniques de normalisation des données peuvent permettre d'assurer une certaine comparabilité inter-équipement et inter-laboratoires. L'une des solutions permettant la normalisation des profils métaboliques est le recours à l'utilisation de mélanges de standards internes marqués (plusieurs centaines) et à ce titre l'utilisation d'extrait marqué métaboliquement pour ce type d'application prend de l'essor.

Le marquage *in vivo* ou biologique a originellement été initié à l'aide de petits organismes vivants tels que les plantes ou les levures. Comme expliqué précédemment, le marquage métabolique consiste à mettre en culture ces organismes en présence de nutriments marqués (sous atmosphère $^{13}\text{CO}_2$ pour les plantes et dans des milieux de culture marqués au ^{13}C pour les levures)^{216,217,276}. Cette technique de marquage permet de générer un nombre important de métabolites marqués dans un seul extrait puisque l'intégralité du métabolome de l'organisme a vocation à être marqué moyennant des durées de mise en culture suffisante et adaptée à l'utilisation exclusive de nutriments contenant des versions « lourdes » de l'isotope à introduire dans les

métabolites. Les limitations actuelles de ce type d'approche, restent la couverture métabolomique et la gamme dynamique des lots de métabolites marqués produits qui ne sont pas toujours bien adaptées à l'étude des échantillons humains. La répétabilité du marquage est également un questionnement majeur dans ce genre de procédé.

Le marquage d'organisme dont le métabolome se rapproche le plus possible de celui de l'homme est donc recommandé pour optimiser l'utilisation de ces échantillons dans un contexte de standardisation. C'est pourquoi, nous nous sommes intéressés au marquage ^{13}C *in vivo* de l'organisme entier de la souris.

Néanmoins, avant de pouvoir exploiter des extraits de souris C-13 pour de la normalisation en métabolomique, la caractérisation fine de ces extraits est un prérequis. Ainsi, l'objectif premier de cette étude préliminaire a été d'évaluer l'efficacité des marquages des métabolites urinaires de ce petit mammifère au cours de son alimentation en nutriments intégralement marqués au carbone-13. A plus long terme, l'étude approfondie des échantillons marqués de souris permettront d'avoir une meilleure connaissance du métabolome de la souris et par comparaison, probablement de celui de l'homme. La pertinence de l'utilisation de tels échantillons pour normaliser les signaux obtenus lors d'étude de cohorte sera également évaluée.

2 – Matériels et méthodes

- Echantillons

Au total, six souris ont été « produites » : trois souris témoins (non-marquées) et trois souris « marquées » au ^{13}C . Pour se faire, l'ensemble des souris ont été nourries exclusivement avec du blé et de la spiruline. Dans le cas des souris marquées au ^{13}C , il s'agissait de blé marqué au ^{13}C (produit par le CEA de Cadarache) et de spiruline ^{13}C commercialisée par Eurisotop ; alors que pour les souris témoins, les mêmes nutriments dans leur version non marquée ont été utilisés. Au total, 25 échantillons d'urine ont été prélevés pour chacune des six souris et ceci sur une durée de sept semaines (Figure 74). Ces souris ont été élevées au CEA de Marcoule avant le début de ma thèse et les échantillons urinaires ont pu être reçus et analysés durant ma thèse.

Partie B : Etudes Expérimentales – Chapitre 2

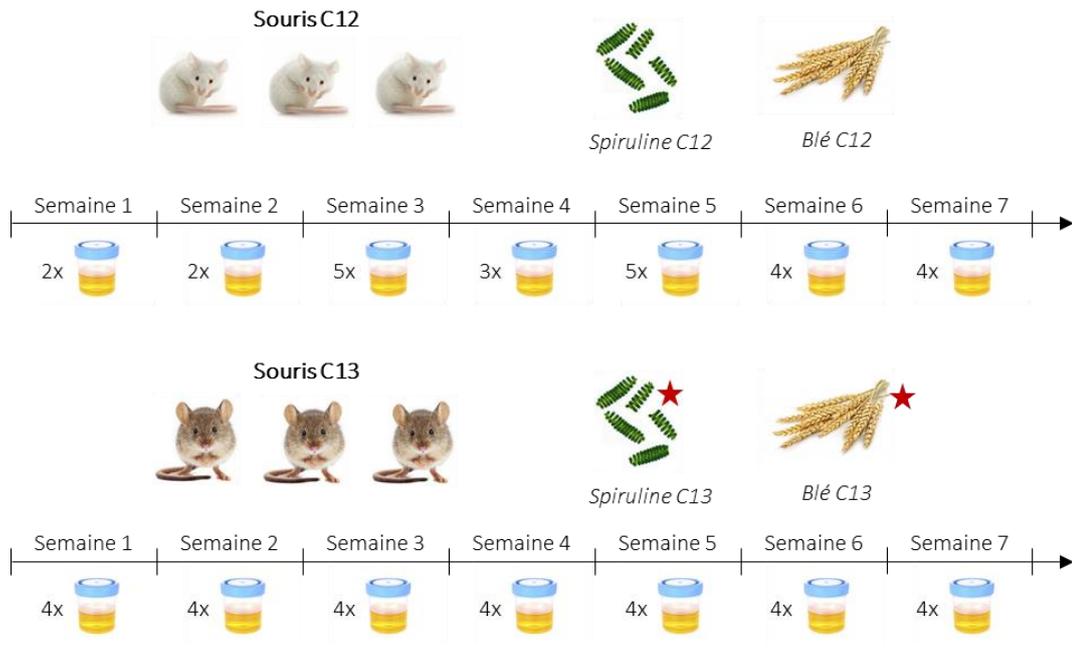


Figure 74 : Détails du prélèvement des échantillons d'urine de souris.

- Préparation des échantillons

La préparation des échantillons diffère selon le système chromatographique utilisé. Pour les analyses C18 et PFPP, 5 μ L d'urine sont prélevés et ajoutés à 45 μ L de blanc (95/5 phase A/phase B) en présence d'étalon interne. L'échantillon est centrifugé à 20 000 g pendant 15 minutes à 4°C. Le surnageant est ensuite prélevé et mis en vial. Pour les analyses HILIC, 5 μ L d'urine sont ajoutés à 18 μ L de phase A en présence d'étalon interne. L'échantillon est mis dans un bain à ultrasons pendant 15 minutes et centrifugé 15 minutes à 20 000 g à 4°C. Le surnageant est prélevé et ajouté à 27 μ L de phase B. L'échantillon est de nouveau centrifugé à 10 000 g pendant 5 minutes à 4°C. Le nouveau surnageant est prélevé et mis en vial. Des QCs sont réalisés en prélevant dans un cas 5 μ L de chacun des 75 échantillons d'urine de souris non marquées (appelées ^{12}C dans la suite de ce manuscrit) et dans un deuxième cas, 5 μ L de chacun des 75 échantillons d'urine de souris ^{13}C sont prélevés. Les QC obtenus sont ensuite dilués par 2, 4 et 8 avec du blanc de phase mobile.

- Méthodes LC/HRMS

L'ensemble des échantillons ont été analysés par LC/HRMS avec le système chromatographique UHPLC Dionex Ultimate 3000 (Thermo Fisher Scientific, Courtaboeuf, France) et le spectromètre de masse Orbitrap Fusion (Thermo Fisher Scientific, Courtaboeuf, France) équipé d'une source ESI en mode d'ionisation positif et négatif. Trois colonnes chromatographiques ont été utilisées (HILIC, C18 et PFPP). Pour la méthode HILIC, les paramètres sont détaillés dans le « matériels et méthodes » du Chapitre 2 de la Partie 2. Pour la méthode C18 se référer à la publication « Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics » (Partie 2 – Chapitre 1). Une

colonne PFPP a également été utilisée. Les phases mobiles, la température du four et du passeur d'échantillons ainsi que le volume d'injection sont les mêmes que pour la méthode C18. Le débit est fixé à 250 $\mu\text{L}/\text{min}$ et le gradient de 30 minutes commence par une période isocratique de 2 minutes à 5 % de phase B. Suit un gradient linéaire de 18 minutes pour arriver à 100 % de phase B. Continue une période isocratique de 4 minutes à 100 % de phase B. Le gradient fini par une période d'équilibration de 6 minutes à 5 % de phase B.

Les paramètres MS varient également selon la méthode chromatographique utilisée. Pour les analyses LC C18, les paramètres de source sont les suivants : une tension de + 3500 V ou – 2500 V est appliquée au capillaire métallique de nébulisation qui est chauffé à 350°C, le débit pour les « sheath gaz » est de 60 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 3 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C. Pour les analyses LC PFPP, les paramètres de source sont les suivants : une tension de + 3500 V ou – 2500 V est appliquée au capillaire métallique de nébulisation qui est chauffé à 200°C, le débit pour les « sheath gaz » est de 45 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 1 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C. Pour les analyses LC HILIC, les paramètres de source sont les suivants : une tension de + 4000 V ou – 2800 V est appliquée au capillaire métallique de nébulisation qui est chauffé à 100°C, le débit pour les « sheath gaz » est de 40 unités arbitraires, celui des gaz auxiliaires de 12 unités et de 1 unités pour les « sweep gaz ». Le tube de transfert des ions vers l'intérieur du spectromètre est chauffé à 325 °C.

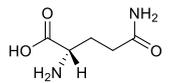
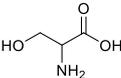
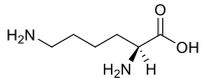
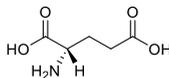
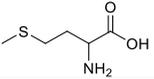
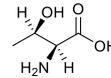
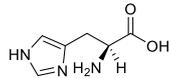
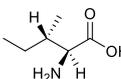
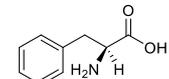
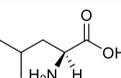
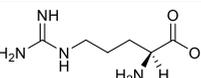
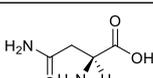
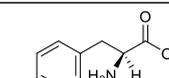
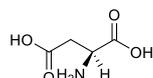
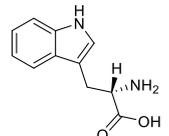
Les acquisitions ont été réalisées en mode « Full Scan » avec l'Orbitrap à une résolution de 240 000 (à m/z 200) avec une fenêtre de balayage de 85 à 1000 m/z . L'« AGC target » est fixé à $5 \cdot 10^4$ ions et le temps maximal d'injection à 400 ms.

Les données ont été traitées avec Xcalibur v.4.1 (ThermoFisher Scientific).

3 – Résultats et perspectives

Dans un premier temps, étant donné que nous disposons d'un panel d'échantillons urinaires recueillis sur une durée de sept semaines, nous avons cherché à évaluer les cinétiques d'incorporation du carbone-13 dans les métabolites urinaires de la souris, c'est-à-dire le temps nécessaire pour arriver à un plateau d'incorporation de C-13. Pour cela, 18 acides aminés, dont les structures sont présentées dans le Tableau 18 ont été suivis de façon ciblée.

Tableau 18 : Liste des 18 acides aminés d'intérêts.

Composé	Composition	Structure	Nature	Composé	Composition	Structure	Nature
Alanine	C3H7NO2		Non-essentielle	Glutamine	C5H10N2O3		Non-essentielle
Sérine	C3H7NO3		Non-essentielle	Lysine	C6H14N2O2		Essentielle
Proline	C2H9NO2		Non-essentielle	Acide Glutamique	C5H9NO4		Non-essentielle
Valine	C5H11NO2		Essentielle	Méthionine	C5H11NO2S		Essentielle
Thréonine	C4H9NO3		Essentielle	Histidine	C6H9N3O2		Essentielle
Isoleucine	C6H13NO2		Essentielle	Phénylalanine	C9H11NO2		Essentielle
Leucine	C6H13NO2		Essentielle	Arginine	C6H14N4O2		Non-essentielle
Asparagine	C4H8N2O3		Non-essentielle	Tyrosine	C9H11NO3		Non-essentielle
Acide Aspartique	C4H7NO4		Non-essentielle	Tryptophane	C11H12N2O2		Essentielle

Les données obtenues pour ces 18 acides aminés ont été traitées semi-manuellement à l'aide de méthodes de traitement (« processing method ») mises en place sur QuantAnalysis (Xcalibur). Chaque isotopologue C-13 de chaque composé est suivi indépendamment et l'ensemble des aires brutes est relevé. Lors de ce traitement des données, nous avons pu nous rendre compte que peu d'intermédiaires de biosynthèse de ces composés étaient observés. Généralement, seuls les composés monoisotopiques (appelés « full C-12 » dans la suite de ce manuscrit), les composés entièrement marqués au carbone-13 (appelés « full C-13 » dans la suite de ce manuscrit) et les composés entièrement marqués au carbone-13 moins un C-13 sont observés. La Figure 75 montre l'exemple de l'asparagine avec l'évolution des aires de ces principaux isotopologues qui sont l'asparagine « full C-12 », « full C-13 » et « full C-13 » moins un C-13(Asp_03).

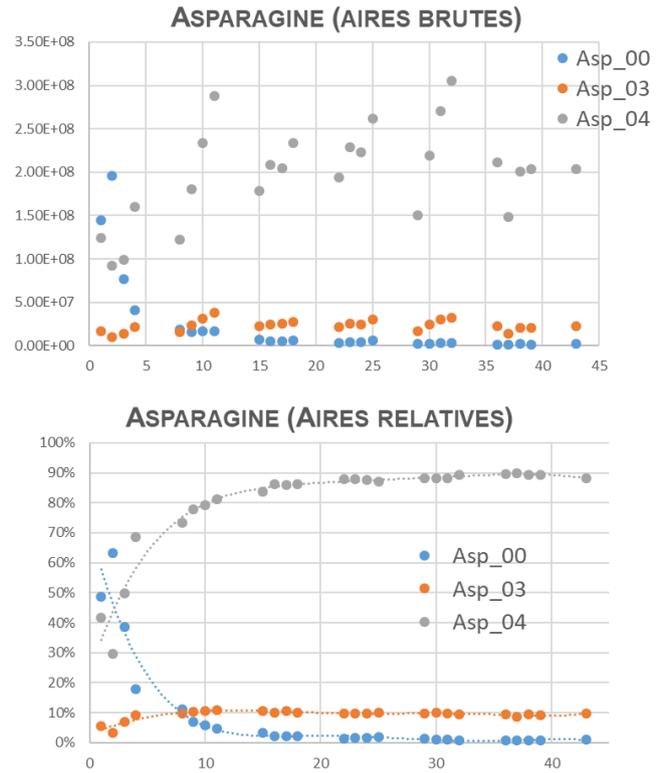
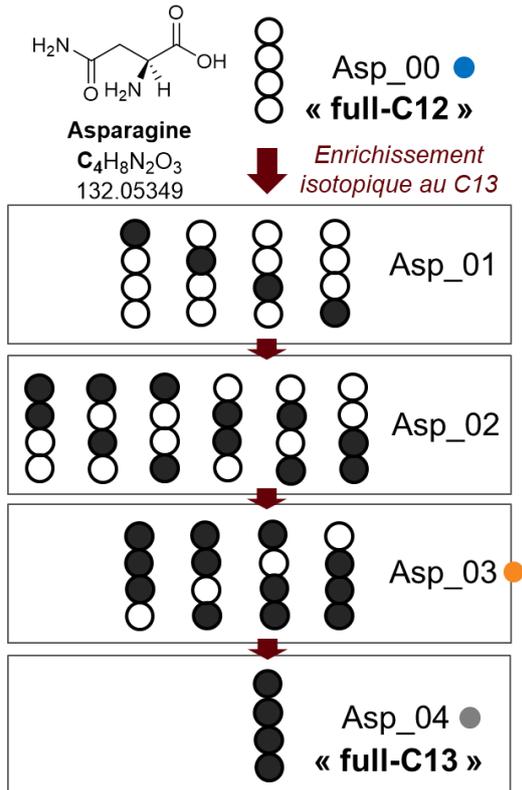


Figure 75 : Exemple de marquage *in vivo* de l'asparagine.

Ensuite, pour réaliser les courbes de cinétiques d'incorporation du carbone-13 dans les différents acides aminés, les signaux ont dû être normalisés. En effet, il est prouvé que les urines sont des échantillons dont la concentration varie selon les individus et le temps de prélèvement²⁷⁷⁻²⁸⁰. Cette normalisation avait pour but de comparer les massifs isotopiques obtenus pour chaque souris. Pour cela, les signaux ont été normalisés par rapport à l'ensemble du massif isotopique. Cela a permis de comparer le marquage des trois souris ¹³C. La cinétique et l'efficacité de ce marquage sont très similaires quelle que soit la souris étudiée. La Figure 76 illustre ce phénomène avec l'exemple de l'asparagine. Trois conditions analytiques ont été comparées : celles traditionnellement utilisées au laboratoire (« C18 Pos » et « HILIC Neg ») ainsi que « PFPP Pos » (en colonne dans la Figure 76). Pour chacune de ces trois conditions, les réponses analytiques obtenues pour chacune des trois souris ont été comparées (en ligne dans la Figure 76). Après normalisation des signaux par rapport au signal analytique de l'ensemble du massif isotopique de l'asparagine, les cinétiques d'incorporation du C-13 est comparable entre les souris et entre les méthodes analytique afin d'atteindre un plateau à 90 % de C-13 au bout de 10 jours.

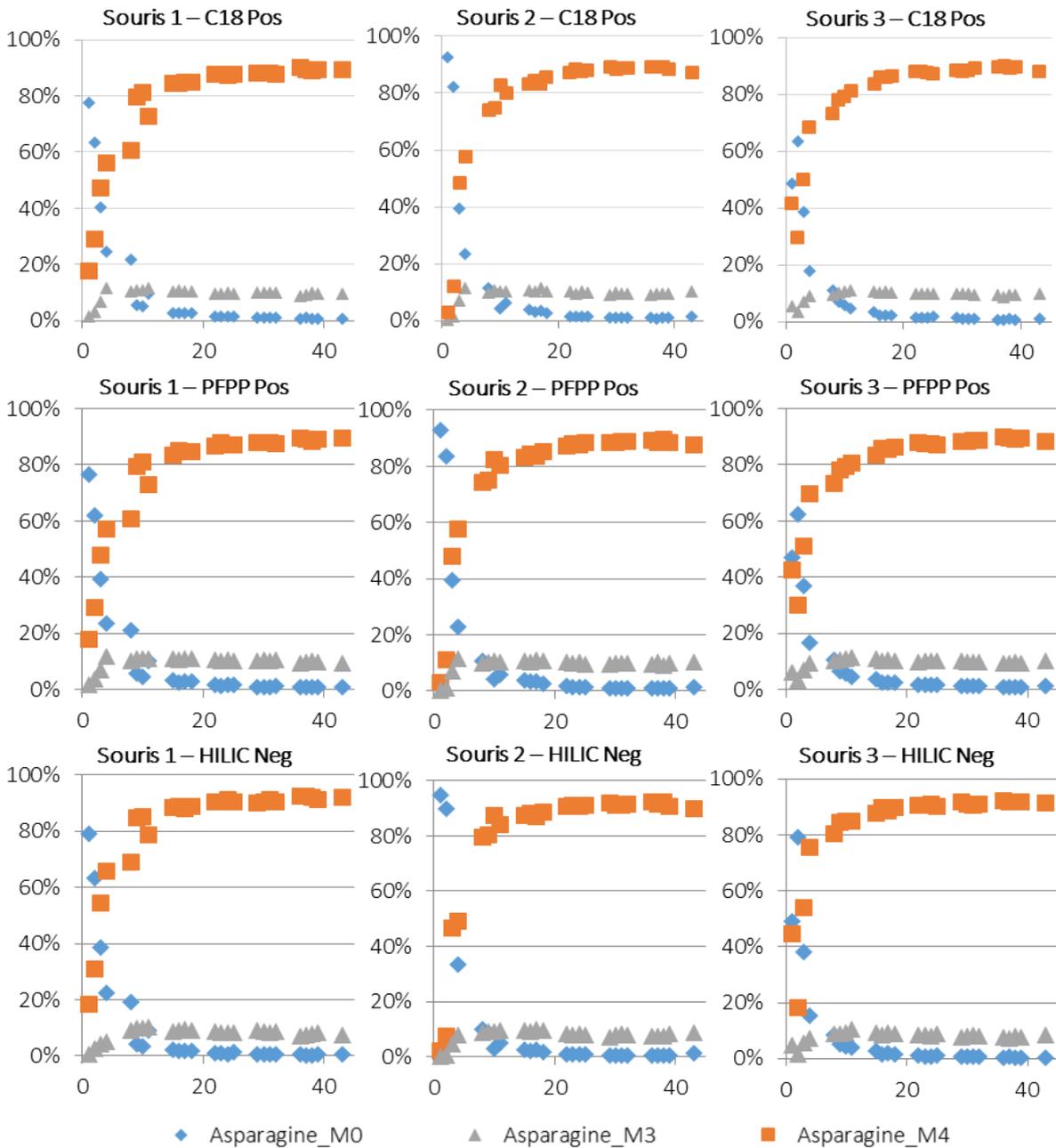


Figure 76 : Comparaison des cinétiques de marquage de l'asparagine pour les trois souris sur les trois systèmes chromatographiques dans les deux modes d'ionisation

Le traitement des données a débuté avec les spectres obtenus après séparation sur colonne C18 en mode d'ionisation négatif. Quatorze acides aminés ont montré une cinétique de marquage rapide avec un taux de métabolite entièrement marqué au C-13 (« full C-13 ») supérieur à 85 % au bout de 10 jours, comme dans le cas du tryptophane présenté Figure 77. Par contre, dans le cas de l'acide aspartique, la formation de l'isotopologue M4 (« full C-13 ») plafonne à 40 % et l'isotopologue majoritaire s'avère être de façon surprenante l'isotopologue M2 (Figure 77). Ceci est dû aux voies métaboliques dans lesquelles l'acide aspartique est impliqué

conduisant à un état d'équilibre particulier. Par ailleurs, l'histidine présente un marquage complet dès le premier échantillon. Ceci révèle peut-être une erreur dans le traitement des données.

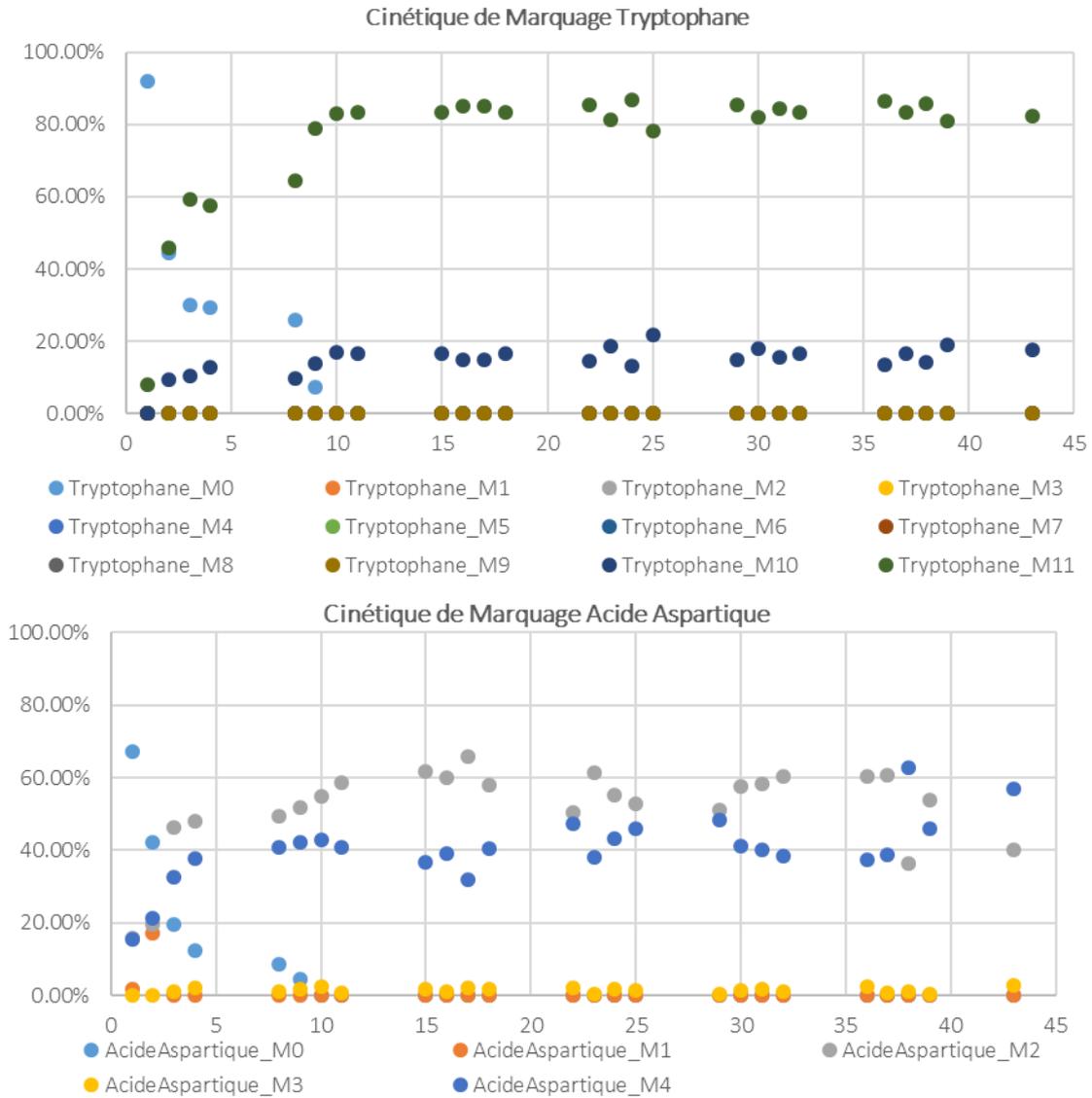
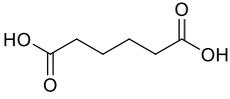
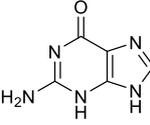
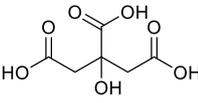
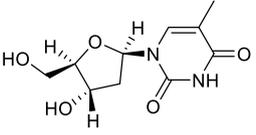
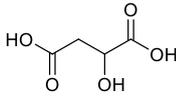
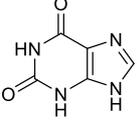
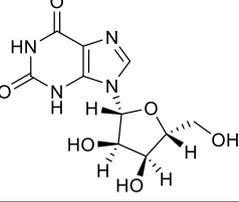


Figure 77 : Exemples de cinétiques de marquage d'acides aminés

Quelques autres composés, appartenant à d'autres familles chimiques, ont également été étudiés comme les acides organiques, les nucléosides ou les nucléobases (Tableau 19).

Tableau 19 : Autres composés d'intérêts pour l'étude du marquage des urines de souris.

Composé	Composition	Structure	Composé	Composition	Structure
Acide Adipique	C ₆ H ₁₀ O ₄		Guanine	C ₅ H ₅ N ₅ O	
Acide Citrique	C ₆ H ₈ O ₇		Thymidine	C ₁₀ H ₁₄ N ₂ O ₅	
Acide Malique	C ₄ H ₆ O ₅		Xanthine	C ₅ H ₄ N ₄ O ₂	
			Xanthosine	C ₁₀ H ₁₂ N ₄ O ₆	

Les trois acides organiques étudiés présentent une cinétique de marquage légèrement plus lente que celle obtenue avec les acides aminés avec un taux de métabolite « full C-13 » supérieur à 80 %. Pour trois des quatre autres composés d'intérêt, le taux de métabolite « full C-13 » est réduit à 70 % (Figure 78). La guanine n'est pas détectée dans ces conditions analytiques.

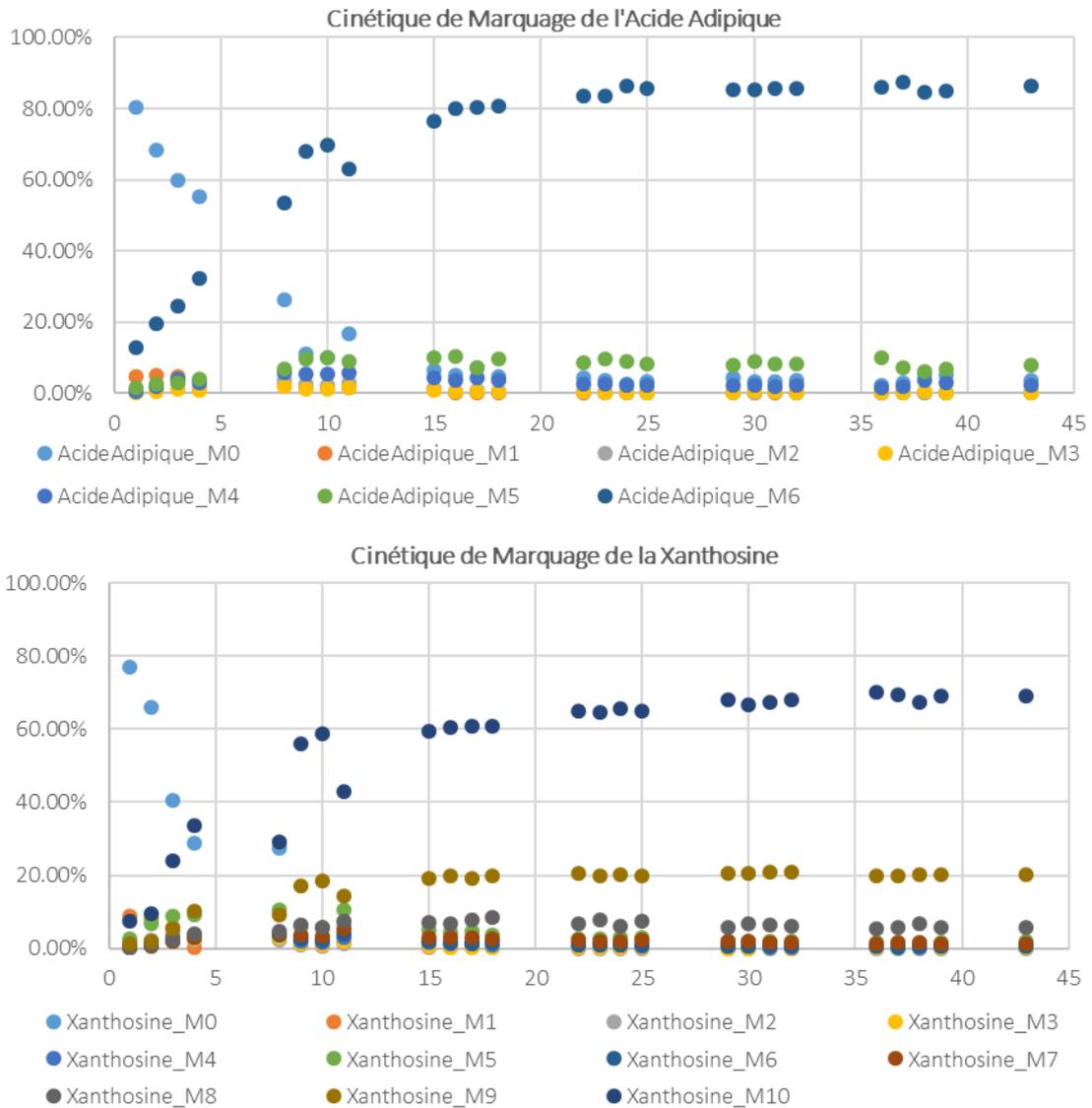


Figure 78 : Exemple de cinétique de marquage pour l'Acide Adipique et la Xanthosine.

Nous avons également cherché, dans un second temps, à faire une annotation des urines de souris marquées et non marquées avec les méthodologies de pré-traitement des données établis au laboratoire. Ainsi, les séquences analytiques acquises en mode positif sur la colonne C18 et celles acquises en mode négatif sur la colonne HILIC pour les souris C12 ont été traitées avec la plateforme Workflow4Metabolomics (W4M)²⁹ avec la méthodologie traditionnellement utilisée au laboratoire. La méthodologie de pré-traitement des signaux appliquée permet de détecter automatiquement les signaux LC-HRMS dans chaque échantillon, de regrouper les signaux similaires présents dans les différents échantillons, d'aligner les temps de rétention et d'intégrer les pics (utilisation des outils XCMS et CAMERA). Pour réduire le nombre de variables et s'affranchir des signaux artefactuels, des tris sont effectués selon trois critères. Tout d'abord les variables ayant une valeur d'aire dans le blanc supérieur à 33 % de celle dans l'échantillon sont supprimées. Les variables ne suivant pas les facteurs de dilution des QCs sont également écartées. Et enfin les variables dont les aires dans les QCs présentent une trop

grande variabilité (> 30 %) sont également enlevées. Ensuite, une fois la liste des variables triées, celles-ci peuvent être annotées automatiquement par comparaison des variables (m/z , RT) avec ceux de la base de données interne du laboratoire. Cette annotation a été réalisée pour les échantillons de souris non marqués avec les colonnes chromatographiques C18 et HILIC (l'annotation n'a pas pu être réalisée sur la colonne PFPP car le laboratoire ne possède pas de base de données avec ce système chromatographique). Le bilan des extractions des variables obtenus avec W4M avant et après tri, ainsi que le nombre d'annotations obtenus sur la base de notre chimiothèque, sont résumés dans le Tableau 20. Notons que les annotations n'ont pour l'instant pas pu être confirmées avec des analyses MS/MS faute de temps

Tableau 20 : Bilan du nombre de variables obtenues après traitement des données sur W4M

Nombre de variables	C18 Pos	C18 Neg	HILIC Neg
Détectées	9100	11660	9310
Triées	6110	7390	4630
Annotées	200	230	200

Le même pré-traitement des données opéré sur W4M a également été réalisé avec les échantillons d'urine marqués au C-13 pour les mêmes conditions analytiques (Tableau 21).

Tableau 21 : Nombre de variables détectées dans les échantillons d'urine de souris C-13.

Nombre de variables	C18 Pos	C18 Neg	HILIC Neg
Détectées	11370	11770	10120
Triées	5833	6546	5539

Les tableaux 20 et 21 montre que de nombreuses variables sont détectées et plus de la moitié d'entre elles passent l'étape du tri. Finalement, moins de cinq pourcent des variables triées sont annotées. On notera que le nombre de variables obtenues après tri sur les échantillons d'urine non-marqués et marqués est comparable et que la prochaine étape consistera à identifier chaque couple « full C-12 / full-C-13 » au sein de ces variables.

Les résultats de cette étude restent pour l'instant assez préliminaires. L'annotation la plus complète possible des échantillons d'urines est une étape très longue et fait l'objet, au moment de la rédaction du manuscrit, d'un travail de stage de Master 2. Cette annotation du métabolome urinaire murin doit être facilité par la comparaison des spectres obtenus dans les mêmes conditions pour les urines de souris C-12 et C-13. En

particulier, les formules brutes de métabolites inconnus pourront être validées sans ambiguïté grâce à la comparaison de leur spectres « full C-12 » et « full C-13 ». En effet, dans un premier temps, des dizaines de formules brutes sont proposées sur la base du m/z de l'ion en « full C-12 ». Le signal « full C-13 » correspondant (co-élution) nous permet de savoir combien d'atomes de carbones sont présents dans la structure et donc de limiter le nombre de compositions élémentaires à, le plus souvent, une unique proposition pour chaque composé (Figure 79).

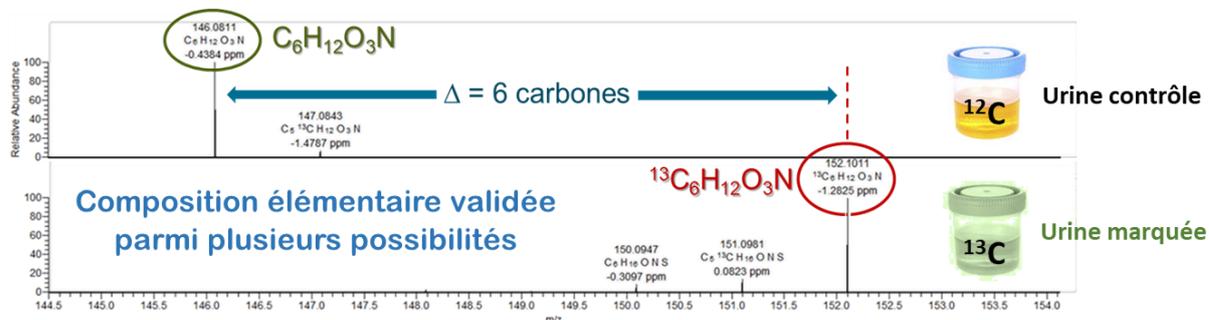


Figure 79 : Exemple de détermination de formule brute par comparaison signal "full C-12" et "full C-13".

L'exemple présenté Figure 80 illustre la possible identification de la decenoylcarnitine. En effet, les écarts de m/z entre la version marquée et non marquée du même métabolite (temps de rétention identique) montrent la présence de 17 carbones dans la structure, ce qui valide la formule brute $C_{17}H_{33}NO_4$ pour ce métabolite. Ce composé a été annoté comme étant la decenoylcarnitine. On remarquera dans ce cas, la présence d'un intermédiaire de biosynthèse qui renforce l'hypothèse d'identification du composé puisqu'il fait apparaître clairement les deux blocs constitutifs de la molécule : un premier bloc correspondant à la chaîne aliphatique de dix carbones (partie « decenoyl ») et le deuxième bloc correspondant à la structure de la carnitine à sept carbones.

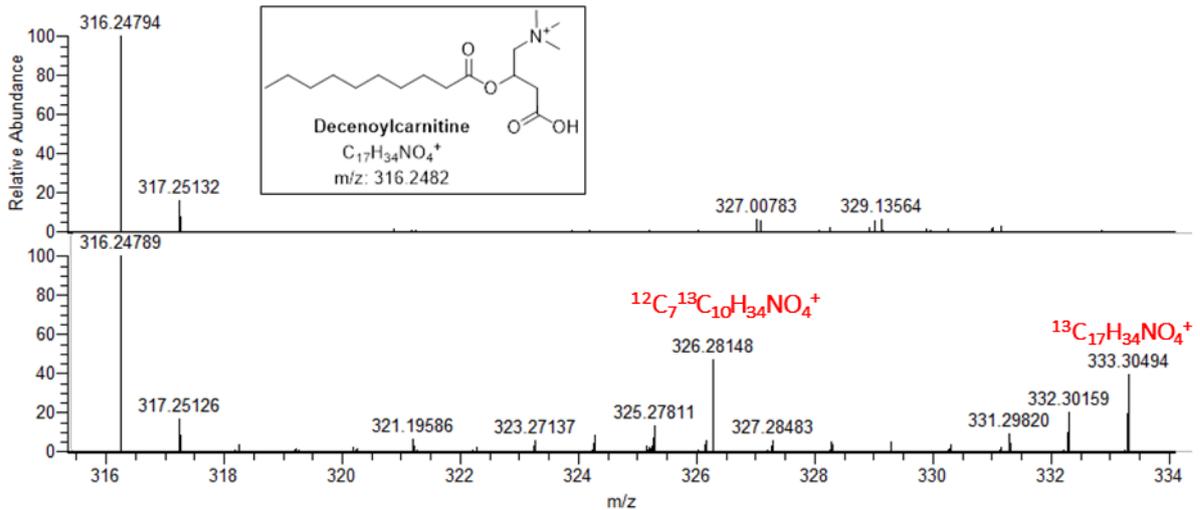


Figure 80 : Comparaison des versions non-marquée et marquée d'un même métabolite et hypothèse d'identification de la decenoylcarnitine.

Spectre MS du haut : souris Full C-12 ; Spectre MS du bas : mélange à quantité égale souris Full C-12/FullC-13

E – Conclusion et perspectives

L'utilisation de standards internes marqués (SIL-IS) est primordiale pour l'identification et la quantification de métabolites. Les SIL-IS n'étant pas toujours disponibles commercialement ou à des coûts prohibitifs, la synthèse chimique est aujourd'hui une méthode très utilisée pour obtenir ces standards. Cependant, les réactions mises en jeu sont souvent lentes avec le passage par plusieurs intermédiaires réactionnels. De plus, ces réactions dépendent fortement des composés ce qui limite le nombre de SIL-IS disponibles.

Le développement d'une méthode d'échange H/D mono-étape métallo-catalysée a permis d'obtenir des standards marqués en peu de temps et à faible coût. Cette méthode a l'avantage de pouvoir produire 24 composés simultanément. L'ensemble des conditions testées a montré que l'application d'une méthode « universelle » basée sur la définition d'un protocole unique avec des conditions réactionnelles fixées (une température, un catalyseur à une charge catalytique donnée, un temps de réaction, etc) était utopique pour parvenir à marquer efficacement les 52 composés d'intérêts. En effet, nous avons montré que les conditions de marquage à appliquer dépendent de la nature chimique de chaque composé. Les massifs isotopiques obtenus contiennent plusieurs isotopologues à des intensités variées contrairement aux SIL-IS disponibles commercialement. Afin d'optimiser les massifs isotopiques obtenus, une méthode de quantification « multi-point internal calibration » a été développée. Cette méthode consiste à considérer chaque isotopologue comme un SIL-IS indépendant. Cette méthode a été comparée à d'autres méthodes de quantification couramment utilisées. La quantification de métabolites endogènes au plasma NIST a montré des résultats très semblables entre ceux obtenus avec la méthode développée et ceux obtenus par gamme externe avec SIL-IS. La

quantification par gamme externe avec SIL-IS étant considéré comme la méthode de choix pour la quantification de métabolites, la « multi-point internal calibration » peut également être utilisée dans ce but. La principale différence est alors le temps d'analyse. En effet avec la quantification par gamme externe avec SIL-IS nécessite la réalisation de gamme avant et après l'analyse de l'échantillon d'intérêt ce qui n'est pas le cas avec la nouvelle méthode de quantification « multi-point internal calibration ». Cela diminue donc fortement le nombre d'échantillons à analyser.

Finalement, une autre méthode de production de SIL-IS a été développée et a montré des résultats très prometteurs qui feront potentiellement l'objet d'un brevet. Ces marquages ont permis de mettre en évidence des effets isotopiques lors des analyses par spectrométrie de masse. L'analyse des spectres de masses obtenus vont permettre de mieux comprendre les processus d'ionisation et de fragmentation mis en jeu lors d'analyse par spectrométrie de masse.

Les méthodes de production de standards marqués par échanges H/D couvre un large spectre de composés. Cependant certains métabolites ne peuvent pas être marqués dans ces conditions. Le marquage *in vivo* est alors une solution alternative pour accéder à ces composés marqués. L'étude préliminaire sur des échantillons d'urine de souris montrent une grande efficacité de marquage au ^{13}C avec une cinétique rapide. Ces échantillons représentent une mine d'information très importante pour l'étude du métabolome de la souris. En effet, ces échantillon, marqués et non marqués, permettent d'avoir un accès complet au métabolome de la souris et d'identifier un grand nombre de métabolites. Cette étape d'annotation est très longue. Une fois cette annotation réalisée, le métabolome de la souris sera comparé à celui de l'homme afin d'évaluer l'utilité des échantillons d'urine de souris pour la normalisation et la quantification des métabolites présents dans les échantillons humains. Il est également envisagé d'exploiter ces échantillons pour la mise en évidence des vrais métabolites par opposition aux contaminants et artefacts présents dans les échantillons biologiques (Figure 81). Ces travaux se rapprocheraient de ceux récemment publiés par Mahieu et al.¹⁶⁴ et Wang et al.²⁸¹ mais s'appliqueraient à des extraits de petits mammifères.

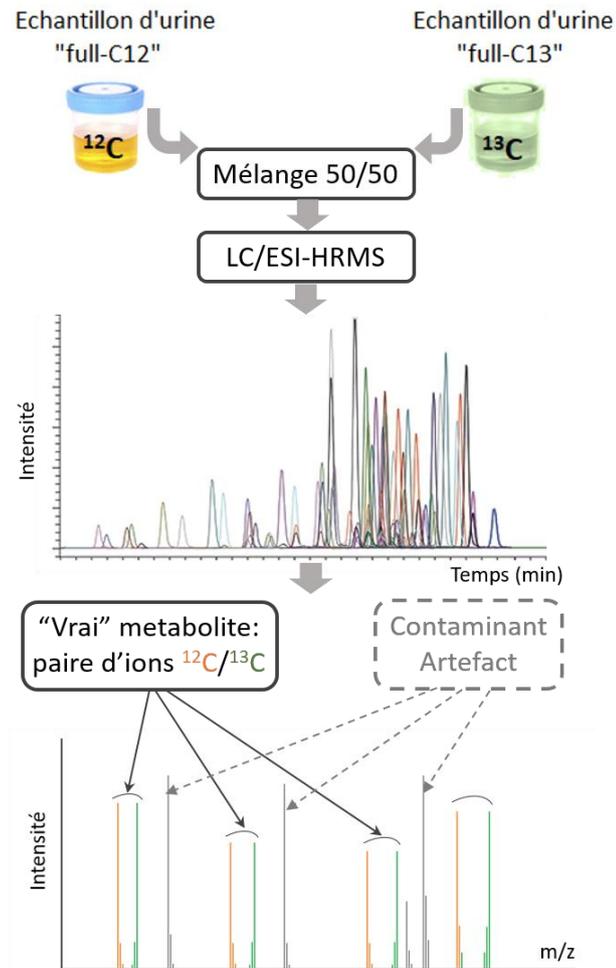


Figure 81 : Protocole envisageable pour la différenciation par LC-HRMS des vrais métabolites par rapport aux contaminants et ions du bruit de fond dans des extraits de petits mammifères.

Conclusions et discussions générales

La métabolomique est une science permettant d'étudier l'ultime réponse d'un organisme à un facteur modifiant son fonctionnement. Les études métabolomiques comportent de nombreuses étapes qui peuvent induire différents biais. En effet, aujourd'hui, les analyses métabolomiques réalisées en routine au laboratoire suivent une méthodologie comportant la préparation des échantillons, l'analyse LC-HRMS permettant le profilage métabolique de l'échantillon, le traitement des données et les analyses statistiques permettant de détecter les métabolites discriminants, les analyses LC-HRMS/MS permettant d'identifier ces métabolites, et l'interprétation biologique. L'ensemble de ces étapes peut introduire de nombreux biais qui peuvent nuire à l'identification des métabolites d'intérêt par rapport à la problématique étudiée. De plus, les analyses métabolomiques étant réalisées sur des cohortes de plus en plus importantes, comportant potentiellement plusieurs centaines d'échantillons devant pour des raisons pratiques être analysés en plusieurs séries distinctes, il devient difficile de comparer efficacement tous ces échantillons acquis dans des conditions différentes.

L'objectif de mon doctorat était donc de répondre à deux des challenges majeurs rencontrés en métabolomique : l'identification des métabolites et la normalisation des signaux voire la quantification absolue de métabolites à l'aide de standards internes marqués aux isotopes stables. Pour cela deux axes ont principalement été étudiés : le développement de méthodes d'acquisition simultanée de spectres MS et MS/MS, ainsi que la production et l'utilisation de standards internes marqués à grande échelle pour la quantification de métabolites.

Dans un premier temps, des méthodes permettant l'acquisition simultanée de spectres MS et MS/MS ont été développées afin de faciliter le processus d'annotation et d'identification des métabolites présents et absents des bases de données. L'objectif étant donc de pouvoir réaliser le profilage métabolique et l'identification des métabolites en une seule injection, nous avons cherché à exploiter au mieux les caractéristiques intrinsèques et la versatilité de l'Orbitrap Fusion. Les méthodes « Data Dependent Acquisition », permettant la fragmentation des composés répondant à certains critères analytiques comme l'intensité du signal, et « Data Independent Acquisition » de type « SWATH », permettant la fragmentation de tous les ions présents dans une gamme de masse donnée, ont montré de bonnes performances pour les analyses de plasma NIST. En effet, avec la méthode DDA, pour 80 % des composés fragmentés, les spectres MS/MS sont directement comparables à ceux présents dans notre base de données et ont donc pu être identifiés sans ambiguïté. De plus, la quantification réalisée à partir des spectres MS des méthodes DDA et DIA a montré des performances similaires à la quantification réalisée sur un « Full Scan » seul. Ces méthodes n'ont pour l'instant été développées qu'en mode d'ionisation positif et en utilisant une colonne chromatographique C-18, conditions les plus exigeantes en terme de vitesse d'acquisition de données. Pour être transférable en routine et être appliquées pour les analyses métabolomiques réalisées au sein du laboratoire, il faudrait également évaluer et optimiser les méthodes DDA

Conclusion et discussions générales

et DIA en utilisant une colonne chromatographique HILIC avec une détection dans le mode d'ionisation négatif. De plus, afin d'exploiter au maximum des données acquises en mode DIA, des logiciels alternatifs à MS-DIAL utilisés dans cette thèse et permettant la déconvolution des spectres MS/MS, pourraient être testés pour identifier au mieux les métabolites. Des bases de données les plus exhaustives possibles seront alors nécessaires. Ce qui impliquerait par exemple l'analyse d'un plus grand nombre de métabolites dans des conditions plus variées d'énergie de collision.

La construction d'une base de données spectrales de contaminants (pesticides, médicaments, retardateurs de flamme, etc) est l'un des objectifs du projet européen HBM4EU. En tant que contributeur de ce projet, nous avons participé à des études inter-laboratoires visant à harmoniser des méthodologies d'acquisition de spectres MS/MS. Nous avons ainsi pu appliquer et optimiser la méthode DDA développée précédemment à l'étude de polluants. Avec cette méthode DDA optimisée, plus de 90 % des composés détectés ont pu être identifiés avec succès par comparaison des spectres MS/MS obtenus par notre laboratoire avec ceux présents dans une base de données préexistante et générée à partir de données QTOF. Des acquisitions MS/MS ciblées ont également été réalisées avec la méthode couramment utilisée au laboratoire et tous les composés fragmentés ont pu être identifiés avec cette même base de données. Cette étude a mis en avant la bonne comparabilité des spectres quel que soit l'analyseur à haute résolution utilisé (Q-TOF vs Q-Orbitrap). Pour cela, des critères de contrôle qualité et d'assurance qualité ont été mis en place et ont fait l'objet d'une publication. Cette étude est très prometteuse pour la construction d'une base de données Européenne et interopérable que chaque laboratoire contribuerait à implémenter via l'acquisition de spectres MS/MS dans des conditions harmonisées. Ceci faciliterait énormément l'identification des métabolites pour l'ensemble des laboratoires du consortium comme pour ceux extérieurs à celui-ci.

Les méthodes DDA, bien que globalement très performantes en matière d'identification des métabolites, montrent cependant quelques limites. En effet, les spectres MS/MS obtenus sont très purs mais l'ensemble des métabolites d'intérêt ne sont pas fragmentés lors d'une seule analyse (en particulier les moins intenses). Les développements liés à l'optimisation de cette méthode par itérations successives permettraient de s'affranchir de cette limite et présenteraient un réel intérêt pour les futures analyses métabolomiques. Dans ce contexte, la méthode DDA « itérative » proposée par ThermoFisher Scientific, liée au développement d'un nouveau logiciel (Acquire X), présente des performances intéressantes pour l'identification d'un grand nombre de métabolites. Cette méthode permet par itérations successives de fragmenter un grand nombre de composés. Ainsi, plus de 400 métabolites ont pu être identifiés dans le plasma NIST contre moins de 200 avec la méthode DDA « simple » précédemment optimisée. Il serait maintenant intéressant d'appliquer les différentes méthodes DIA et DDA « itératives » à des études métabolomiques à grande échelle pour évaluer leurs performances dans des conditions plus réalistes. Pour cela, les échantillons d'intérêt pourraient être analysés en mode DIA et les échantillons de contrôle qualité pourraient être analysés avec la méthode DDA « itérative » (Figure 82). L'objectif serait alors d'identifier et quantifier les métabolites présents dans des bases de données avec les acquisitions DIA et d'identifier de nouveaux composés avec les acquisitions réalisées avec la méthode DDA « itérative ». Ces méthodes pourraient être utilisées de façon complémentaire pour la recherche de biomarqueurs, l'analyse MS

Conclusion et discussions générales

permettant de détecter les métabolites discriminants puis de les quantifier, et l'analyse MS/MS permettant de les identifier. La quantification de ces biomarqueurs est primordiale pour le développement de futures approches de diagnostic clinique.

N° d'injection	Echantillon	N° d'injection	Echantillon
1	Blanc-DIA	21	Echantillon_36-DIA
2	QC_dil8-DIA	22	Echantillon_50-DIA
3	QC_dil8-DIA	23	Echantillon_45-DIA
4	QC_dil8-DIA	24	Echantillon_2-DIA
5	QC_dil4-DIA	25	Echantillon_22-DIA
6	QC_dil4-DIA	26	Blanc-DIA
7	QC_dil4-DIA	27	QC_dil1-DIA
8	QC_dil2-DIA	28	Echantillon_12-DIA
9	QC_dil2-DIA	29-37	...
10	QC_dil2-DIA	38	Blanc-DIA
11	QC_dil1-DIA	39	QC_dil1-DIA
12	QC_dil1-DIA	40-47	...
13	QC_dil1-DIA	48	Blanc-DIA
14	Blanc-DIA	49	QC_dil1-DIA
15	QC_dil1-DIA	50	Blanc-Acquire X-LC-HRMS
16	Echantillon_31-DIA	51	QC_dil1-Acquire X-LC-HRMS
17	Echantillon_4-DIA	52	QC_dil1-Acquire X-DDA1
18	Echantillon_7-DIA	53	QC_dil1-Acquire X-DDA2
19	Echantillon_19-DIA	54	QC_dil1-Acquire X-DDA3
20	Echantillon_17-DIA		

dil1 ; dil2 ; dil4 ; dil8 =
facteurs de dilution des QC

Figure 82 : Exemple d'une possible future séquence analytique.

Les seconds développements se sont focalisés sur la production et l'utilisation de standards internes marqués. En effet, aujourd'hui, la méthode de choix pour la quantification de métabolites est l'utilisation de molécules marquées aux isotopes stables (SIL-IS). Afin de quantifier un maximum de métabolites présents dans les matrices biologiques, un grand nombre de SIL-IS est nécessaire. Dans un premier temps, la synthèse chimique a permis de produire (dans des conditions préalablement optimisées) par une réaction d'échange H/D mono-étape métal-catalysée en peu de temps de nombreux SIL-IS. Cependant, ces standards obtenus présentent le plus souvent des distributions isotopiques assez larges et donc des marquages hétérogènes par opposition aux puretés isotopiques des produits marqués commerciaux. L'exploitation de ces massifs isotopiques a cependant permis de développer une méthode de quantification originale appelée « multi-point internal calibration ». Cette méthode considère chaque isotopologue du massif isotopique comme un SIL-IS indépendant. Une gamme de calibration interne peut donc être réalisée pour chaque métabolite. Cette méthode a donné des résultats très similaires à la quantification par gamme externe avec des SIL-IS qui est aujourd'hui la méthode de référence pour

Conclusion et discussions générales

la quantification. La « multi-point internal calibration » permettrait donc de gagner du temps car permettrait de s'affranchir de l'acquisition de gammes externes avant et après l'échantillon d'intérêt.

L'utilisation de protocoles de deutération alternatifs à la méthode métallo-catalysée, dont les conditions réductrices peuvent en limiter l'utilisation, ont également été évalués afin d'augmenter le nombre de standards marqués disponibles au laboratoire. Les résultats obtenus avec cette méthode sont également très prometteurs. En effet, plus de 40 composés ont été marqués efficacement et pourront être utilisés comme SIL-IS. De plus, les distributions isotopiques sont plus étroites que celles obtenues par réaction d'échange H/D mono-étape métallo-catalysée. Ces développements pourraient faire l'objet d'un brevet. Dans ces conditions nous avons pu mettre en évidence des effets isotopiques assez importants, dont la compréhension pourrait permettre de mieux comprendre les processus d'ionisation et de fragmentation. Il serait donc très intéressant au niveau fondamental comme applicatif de pousser plus loin ces études d'effets isotopiques. Les méthodes de production de SIL-IS pourraient encore être optimisées et appliquées à d'autres familles de métabolites comme les lipides par exemple. En effet, la disponibilité de standards, marqués ou non-marqués, de lipides est encore un challenge aujourd'hui. En effet, seuls quelques standards de lipides sont généralement disponibles commercialement par classe (par exemple les glycérides, sphingolipides ou encore stérols). Des développements de marquage de lipides pourraient alors être réalisés sur ces différentes classes. De plus, afin d'avoir le plus de SIL-IS disponibles, il serait intéressant d'envisager le marquage direct de matrices biologiques complexes tels que le plasma ou les urines. De plus, l'évaluation à plus grande échelle de la méthode de quantification « multi-point internal calibration » apparaît comme particulièrement pertinente et novatrice, de même que son application à différentes matrices biologiques.

Comme nous l'avons vu, le marquage métabolique (*in vivo*) permet d'avoir accès à un grand nombre de métabolites marqués au carbone-13. L'étude du marquage *in vivo* d'urine de souris au ¹³C initié lors de ma thèse montre un très grand intérêt pour l'étude du métabolome de la souris et dans un contexte plus large d'identification de métabolites. Ces échantillons représentent en effet une ressource très importante d'informations pour l'identification de nouveaux métabolites. L'application de la méthode DDA « Acquire X » à ces échantillons (« full C-12 », « full C-13 » et mélange « full C-12/C-13 ») en ciblant par exemple des incréments en masse particuliers pourrait permettre d'avoir une connaissance approfondie de l'ensemble des métabolites présents dans le métabolome urinaire murin et potentiellement d'identifier de nouveaux métabolites. En effet, la comparaison des spectres MS et MS/MS obtenus pour ces échantillons pourrait fournir des informations clés pour l'élucidation structurale des composés. Dans le futur, il serait intéressant d'étudier des modèles animaux plus proches métaboliquement de l'homme tel que les modèles porcine ou simienne. Le marquage de ces animaux au carbone-13 pourrait nous donner accès à des quantités d'échantillons plus importantes et plus directement utilisables pour la quantification et/ou la normalisation de signaux d'échantillons humains.

La disponibilité de composés marqués, que ce soit par voie chimique ou métabolique, est aujourd'hui de plus en plus nécessaire pour les analyses métabolomiques. En effet, les cohortes médicales étant de plus en plus importantes, des outils de normalisation sont nécessaires pour assurer la robustesse des comparaisons « inter-batches » ou inter-plateformes. Les familles chimiques des métabolites étant très variées, de nombreux

Conclusion et discussions générales

standards sont nécessaires. Un lot bien défini de standards marqués obtenus par voie chimique pourrait par exemple être utilisé en routine pour améliorer la normalisation des signaux présents dans les plusieurs centaines d'échantillons analysés dans une cohorte médicale mais également afin de s'assurer de la qualité des conditions expérimentales. Une alternative, après avoir étudié le chevauchement des couvertures métaboliques des modèles animaux avec celle de l'homme, serait également d'utiliser les échantillons marqués *in vivo*. Supplémenté des échantillons marqués métaboliquement aux échantillons d'études pourrait permettre de normaliser simultanément un grand nombre de métabolites (connus et inconnus).

Globalement cette thèse avait pour objectif d'améliorer l'identification et la quantification des métabolites en rendant les analyses métabolomiques plus robustes, l'objectif à plus long terme étant de proposer des pistes pour le diagnostic clinique. Les développements réalisés, ou seulement initiés pour certains autres, ont permis d'améliorer l'identification et la quantification des métabolites. La combinaison des méthodes développées indépendamment, DDA et DIA pour l'identification et la quantification des métabolites, les productions chimique et métabolique de composés marqués ainsi que la méthode innovante pour la quantification, pourraient permettre d'améliorer à terme la recherche et la validation de biomarqueurs de pathologies voire le diagnostic clinique.

Références

- (1) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics* **2001**, *2* (3), 155–168. <https://doi.org/10.1002/cfg.82>.
- (2) Nicholson, J. K.; Lindon, J. C.; Holmes, E. “Metabonomics”: Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli via Multivariate Statistical Analysis of Biological NMR Spectroscopic Data. *Xenobiotica* **1999**, *29* (11), 1181–1189. <https://doi.org/10.1080/004982599238047>.
- (3) Yang, Q.; Zhang, A.; Miao, J.; Sun, H.; Han, Y.; Yan, G.; Wu, F.; Wang, X. Metabolomics Biotechnology, Applications, and Future Trends: A Systematic Review. *RSC Adv.* **2019**, *9* (64), 37245–37257. <https://doi.org/10.1039/C9RA06697G>.
- (4) Oliver, S. Systematic Functional Analysis of the Yeast Genome. *Trends in Biotechnology* **1998**, *16* (9), 373–378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1).
- (5) Junot, C.; Fenaille, F.; Colsch, B.; Bécher, F. High Resolution Mass Spectrometry Based Techniques at the Crossroads of Metabolic Pathways: High Resolution Mass Spectrometry for Metabolomics. *Mass Spec Rev* **2014**, *33* (6), 471–500. <https://doi.org/10.1002/mas.21401>.
- (6) Sanchez, S.; Demain, A. L. Metabolic Regulation and Overproduction of Primary Metabolites. *Microbial Biotech* **2008**, *1* (4), 283–319. <https://doi.org/10.1111/j.1751-7915.2007.00015.x>.
- (7) Carmichael, W. W. Cyanobacteria Secondary Metabolites—the Cyanotoxins. *Journal of Applied Bacteriology* **1992**, *72* (6), 445–459. <https://doi.org/10.1111/j.1365-2672.1992.tb01858.x>.
- (8) Aksenov, A. A.; da Silva, R.; Knight, R.; Lopes, N. P.; Dorrestein, P. C. Global Chemical Analysis of Biology by Mass Spectrometry. *Nat Rev Chem* **2017**, *1* (7), 0054. <https://doi.org/10.1038/s41570-017-0054>.
- (9) Ivanisevic, J.; Want, E. J. From Samples to Insights into Metabolism: Uncovering Biologically Relevant Information in LC-HRMS Metabolomics Data. *Metabolites* **2019**, *9* (12), 308. <https://doi.org/10.3390/metabo9120308>.
- (10) Klåvus, A.; Kokla, M.; Noerman, S.; Koistinen, V. M.; Tuomainen, M.; Zarei, I.; Meuronen, T.; Häkkinen, M. R.; Rummukainen, S.; Farizah Babu, A.; Sallinen, T.; Kärkkäinen, O.; Paananen, J.; Broadhurst, D.; Brunius, C.; Hanhineva, K. “Notame”: Workflow for Non-Targeted LC–MS Metabolic Profiling. *Metabolites* **2020**, *10* (4), 135. <https://doi.org/10.3390/metabo10040135>.
- (11) Lu, W.; Su, X.; Klein, M. S.; Lewis, I. A.; Fiehn, O.; Rabinowitz, J. D. Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annu. Rev. Biochem.* **2017**, *86* (1), 277–304. <https://doi.org/10.1146/annurev-biochem-061516-044952>.
- (12) Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O’Hagan, S.; Knowles, J. D.; Halsall, A.; HUSERMET Consortium; Wilson, I. D.; Kell, D. B. Development of a Robust and Repeatable UPLC–MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.* **2009**, *81* (4), 1357–1364. <https://doi.org/10.1021/ac8019366>.
- (13) The Human Serum Metabolome (HUSERMET) Consortium; 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. 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. <https://doi.org/10.1038/nprot.2011.335>.
- (14) A, J.; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklund, S. L.; Moritz, T. Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. *Anal. Chem.* **2005**, *77* (24), 8086–8094. <https://doi.org/10.1021/ac051211v>.
- (15) Want, E. J.; O’Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Anal. Chem.* **2006**, *78* (3), 743–752. <https://doi.org/10.1021/ac051312t>.
- (16) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Global Metabolic Profiling Procedures for Urine Using UPLC–MS. *Nat Protoc* **2010**, *5* (6), 1005–1018. <https://doi.org/10.1038/nprot.2010.50>.
- (17) Rosen Vollmar, A. K.; Rattray, N. J. W.; Cai, Y.; Santos-Neto, Á. J.; Deziel, N. C.; Jukic, A. M. Z.; Johnson, C. H. Normalizing Untargeted Periconceptional Urinary Metabolomics Data: A Comparison of Approaches. *Metabolites* **2019**, *9* (10), 198. <https://doi.org/10.3390/metabo9100198>.

Références

- (18) Roux, A.; Xu, Y.; Heilier, J.-F.; Olivier, M.-F.; Ezan, E.; Tabet, J.-C.; Junot, C. 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**, *84* (15), 6429–6437. <https://doi.org/10.1021/ac300829f>.
- (19) Vuckovic, D. Current Trends and Challenges in Sample Preparation for Global Metabolomics Using Liquid Chromatography–Mass Spectrometry. *Anal Bioanal Chem* **2012**, *403* (6), 1523–1548. <https://doi.org/10.1007/s00216-012-6039-y>.
- (20) Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K. Global Metabolic Profiling of Animal and Human Tissues via UPLC-MS. *Nat Protoc* **2013**, *8* (1), 17–32. <https://doi.org/10.1038/nprot.2012.135>.
- (21) Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. High-Throughput Tissue Extraction Protocol for NMR- and MS-Based Metabolomics. *Analytical Biochemistry* **2008**, *372* (2), 204–212. <https://doi.org/10.1016/j.ab.2007.10.002>.
- (22) Masson, P.; Alves, A. C.; Ebbels, T. M. D.; Nicholson, J. K.; Want, E. J. Optimization and Evaluation of Metabolite Extraction Protocols for Untargeted Metabolic Profiling of Liver Samples by UPLC-MS. *Anal. Chem.* **2010**, *82* (18), 7779–7786. <https://doi.org/10.1021/ac101722e>.
- (23) Surowiec, I.; Johansson, E.; Stenlund, H.; Rantapää-Dahlqvist, S.; Bergström, S.; Normark, J.; Trygg, J. Quantification of Run Order Effect on Chromatography - Mass Spectrometry Profiling Data. *Journal of Chromatography A* **2018**, *1568*, 229–234. <https://doi.org/10.1016/j.chroma.2018.07.019>.
- (24) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. A Pragmatic and Readily Implemented Quality Control Strategy for HPLC-MS and GC-MS-Based Metabonomic Analysis. *Analyst* **2006**, *131* (10), 1075–1078. <https://doi.org/10.1039/b604498k>.
- (25) Gika, H. G.; Theodoridis, G. A.; Wingate, J. E.; Wilson, I. D. Within-Day Reproducibility of an HPLC–MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J. Proteome Res.* **2007**, *6* (8), 3291–3303. <https://doi.org/10.1021/pr070183p>.
- (26) Smith, C. A.; Want, E. J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787. <https://doi.org/10.1021/ac051437y>.
- (27) Lommen, A. MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. *Anal. Chem.* **2009**, *81* (8), 3079–3086. <https://doi.org/10.1021/ac900036d>.
- (28) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11* (1), 395. <https://doi.org/10.1186/1471-2105-11-395>.
- (29) Giacomoni, F.; Le Corguille, G.; Monsoor, M.; Landi, M.; Pericard, P.; Petera, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; Goullitquer, S.; Thevenot, E. A.; Caron, C. Workflow4Metabolomics: A Collaborative Research Infrastructure for Computational Metabolomics. *Bioinformatics* **2015**, *31* (9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>.
- (30) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation. *Nucleic Acids Research* **2009**, *37* (Web Server), 652–660. <https://doi.org/10.1093/nar/gkp356>.
- (31) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84* (11), 5035–5039. <https://doi.org/10.1021/ac300698c>.
- (32) Croixmarie, V.; Umbdenstock, T.; Cloarec, O.; Moreau, A.; Pascussi, J.-M.; Boursier-Neyret, C.; Walther, B. Integrated Comparison of Drug-Related and Drug-Induced Ultra Performance Liquid Chromatography/Mass Spectrometry Metabonomic Profiles Using Human Hepatocyte Cultures. *Anal. Chem.* **2009**, *81* (15), 6061–6069. <https://doi.org/10.1021/ac900333e>.
- (33) Eliasson, M.; Rännar, S.; Madsen, R.; Donten, M. A.; Marsden-Edwards, E.; Moritz, T.; Shockcor, J. P.; Johansson, E.; Trygg, J. Strategy for Optimizing LC-MS Data Processing in Metabolomics: A Design of Experiments Approach. *Anal. Chem.* **2012**, *84* (15), 6869–6876. <https://doi.org/10.1021/ac301482k>.
- (34) 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, David. 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. HMDB: The Human Metabolome Database. *Nucleic Acids Research* **2007**, *35* (Database), D521–D526. <https://doi.org/10.1093/nar/gkl923>.

Références

- (35) 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. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Research* **2018**, *46* (D1), D608–D617. <https://doi.org/10.1093/nar/gkx1089>.
- (36) Smith, C. A.; Maille, G. O.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Therapeutic Drug Monitoring* **2005**, *27* (6), 747–751. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>.
- (37) Wu, Y.; Li, L. Sample Normalization Methods in Quantitative Metabolomics. *Journal of Chromatography A* **2016**, *1430*, 80–95. <https://doi.org/10.1016/j.chroma.2015.12.007>.
- (38) Webb-Robertson, B.-J. M.; Stratton, K. G.; Kyle, J. E.; Kim, Y.-M.; Bramer, L. M.; Waters, K. M.; Koeller, D. M.; Metz, T. O. Statistically Driven Metabolite and Lipid Profiling of Patients from the Undiagnosed Diseases Network. *Anal. Chem.* **2020**, *92* (2), 1796–1803. <https://doi.org/10.1021/acs.analchem.9b03522>.
- (39) Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Orešič, M. Normalization Method for Metabolomics Data Using Optimal Selection of Multiple Internal Standards. *BMC Bioinformatics* **2007**, *8* (1), 93. <https://doi.org/10.1186/1471-2105-8-93>.
- (40) Kamleh, M. A.; Ebbels, T. M. D.; Spagou, K.; Masson, P.; Want, E. J. Optimizing the Use of Quality Control Samples for Signal Drift Correction in Large-Scale Urine Metabolic Profiling Studies. *Anal. Chem.* **2012**, *84* (6), 2670–2677. <https://doi.org/10.1021/ac202733q>.
- (41) Wang, S.-Y.; Kuo, C.-H.; Tseng, Y. J. Batch Normalizer: A Fast Total Abundance Regression Calibration Method to Simultaneously Adjust Batch and Injection Order Effects in Liquid Chromatography/Time-of-Flight Mass Spectrometry-Based Metabolomics Data and Comparison with Current Calibration Methods. *Anal. Chem.* **2013**, *85* (2), 1037–1046. <https://doi.org/10.1021/ac302877x>.
- (42) Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B. Reflections on Univariate and Multivariate Analysis of Metabolomics Data. *Metabolomics* **2014**, *10* (3), 361–374. <https://doi.org/10.1007/s11306-013-0598-6>.
- (43) Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **2007**, *6* (2), 469–479. <https://doi.org/10.1021/pr060594q>.
- (44) Ren, S.; Hinzman, A. A.; Kang, E. L.; Szczesniak, R. D.; Lu, L. J. Computational and Statistical Analysis of Metabolomics Data. *Metabolomics* **2015**, *11* (6), 1492–1513. <https://doi.org/10.1007/s11306-015-0823-6>.
- (45) Clària, J.; Moreau, R.; Fenaille, F.; Amorós, A.; Junot, C.; Gronbaek, H.; Coenraad, M. J.; Pruvost, A.; Ghetas, A.; Chu-Van, E.; López-Vicario, C.; Oetti, K.; Caraceni, P.; Alessandria, C.; Trebicka, J.; Pavesi, M.; Deulofeu, C.; Albillos, A.; Gustot, T.; Welzel, T. M.; Fernández, J.; Stauber, R. E.; Saliba, F.; Butin, N.; Colsch, B.; Moreno, C.; Durand, F.; Nevens, F.; Bañares, R.; Benten, D.; Ginès, P.; Gerbes, A.; Jalan, R.; Angeli, P.; Bernardi, M.; Arroyo, V. Orchestration of Tryptophan-Kynurenine Pathway, Acute Decompensation, and Acute-on-Chronic Liver Failure in Cirrhosis. *Hepatology* **2018**, *69* (4), 1686–1701. <https://doi.org/10.1002/hep.30363>.
- (46) 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 in Analytical Chemistry* **2018**. <https://doi.org/10.1016/j.trac.2018.11.022>.
- (47) Kanehisa, M. KEGG Bioinformatics Resource for Plant Genomics and Metabolomics. In *Plant Bioinformatics*; Edwards, D., Ed.; Springer New York: New York, NY, 2016; Vol. 1374, pp 55–70. https://doi.org/10.1007/978-1-4939-3167-5_3.
- (48) Cottret, L.; Frainay, C.; Chazalviel, M.; Cabanettes, F.; Gloaguen, Y.; Camenen, E.; Merlet, B.; Heux, S.; Portais, J.-C.; Poupin, N.; Vinson, F.; Jourdan, F. MetExplore: Collaborative Edition and Exploration of Metabolic Networks. *Nucleic Acids Research* **2018**, *46* (W1), W495–W502. <https://doi.org/10.1093/nar/gky301>.
- (49) Peisl, B. Y. L.; Schymanski, E. L.; Wilmes, P. Dark Matter in Host-Microbiome Metabolomics: Tackling the Unknowns—A Review. *Analytica Chimica Acta* **2018**, *1037*, 13–27. <https://doi.org/10.1016/j.aca.2017.12.034>.
- (50) Sumner, L. W.; Mendes, P.; Dixon, R. A. Plant Metabolomics: Large-Scale Phytochemistry in the Functional Genomics Era. *Phytochemistry* **2003**, *62* (6), 817–836. [https://doi.org/10.1016/S0031-9422\(02\)00708-2](https://doi.org/10.1016/S0031-9422(02)00708-2).
- (51) Devaux, P. G.; Horning, M. G.; Horning, E. C. Benzylxime Derivatives of Steroids. A New Metabolic Profile Procedure for Human Urinary Steroids. *Analytical Letters* **1971**, *4* (3), 151–160. <https://doi.org/10.1080/00032717108059686>.
- (52) Horning, E. C.; Horning, M. G. Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites. *Clinical Chemistry* **1971**, *17* (8), 802–809. <https://doi.org/10.1093/clinchem/17.8.802>.

Références

- (53) Nicholson, J. K.; O'Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sönksen, P. H. Proton-Nuclear-Magnetic-Resonance Studies of Serum, Plasma and Urine from Fasting Normal and Diabetic Subjects. *Biochemical Journal* **1984**, *217* (2), 365–375. <https://doi.org/10.1042/bj2170365>.
- (54) Bales, J. R.; Higham, D. P.; Howe, I.; Nicholson, J. K.; Sadler, P. J. Use of High-Resolution Proton Nuclear Magnetic Resonance Spectroscopy for Rapid Multi-Component Analysis of Urine. *Clinical Chemistry* **1984**, *30* (3), 426–432. <https://doi.org/10.1093/clinchem/30.3.426>.
- (55) Oliver, S. G. Yeast as a Navigational Aid in Genome Analysis. *Microbiology* **1997**, *143* (5), 1483–1487. <https://doi.org/10.1099/00221287-143-5-1483>.
- (56) Tweeddale, H.; Notley-McRobb, L.; Ferenci, T. Effect of Slow Growth on Metabolism of Escherichia Coli, as Revealed by Global Metabolite Pool (“Metabolome”) Analysis. *Journal of Bacteriology* **1998**, *180* (19), 5109–5116. <https://doi.org/10.1128/JB.180.19.5109-5116.1998>.
- (57) Baggett, B. R.; Cooper, J. D.; Hogan, E. T.; Carper, J.; Paiva, N. L.; Smith, J. T. Profiling Isoflavonoids Found in Legume Root Extracts Using Capillary Electrophoresis. *Electrophoresis* **2002**, *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).
- (58) Arlt, K.; Brandt, S.; Kehr, J. Amino Acid Analysis in Five Pooled Single Plant Cell Samples Using Capillary Electrophoresis Coupled to Laser-Induced Fluorescence Detection. *Journal of Chromatography A* **2001**, *926* (2), 319–325. [https://doi.org/10.1016/S0021-9673\(01\)01052-4](https://doi.org/10.1016/S0021-9673(01)01052-4).
- (59) Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. Simultaneous Determination of Anionic Intermediates for *Bacillus s Ubtilis* Metabolic Pathways by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2002**, *74* (10), 2233–2239. <https://doi.org/10.1021/ac020064n>.
- (60) 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. Metabonomics: The Use of Electrospray Mass Spectrometry Coupled to Reversed-Phase Liquid Chromatography Shows Potential for the Screening of Rat Urine in Drug Development. *Rapid Commun. Mass Spectrom.* **2002**, *16* (20), 1991–1996. <https://doi.org/10.1002/rcm.813>.
- (61) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite Profiling for Plant Functional Genomics. *Nat Biotechnol* **2000**, *18* (11), 1157–1161. <https://doi.org/10.1038/81137>.
- (62) Stein, S. Mass Spectral Reference Libraries: An Ever-Expanding Resource for Chemical Identification. *Anal. Chem.* **2012**, *84* (17), 7274–7282. <https://doi.org/10.1021/ac301205z>.
- (63) Bhardwaj, C.; Hanley, L. Ion Sources for Mass Spectrometric Identification and Imaging of Molecular Species. *Nat. Prod. Rep.* **2014**, *31* (6), 756–767. <https://doi.org/10.1039/C3NP70094A>.
- (64) Horning, E. C.; Carroll, D. I.; Dzidic, I.; Haegele, K. D.; Horning, M. G.; Stiliweli, R. N. Atmospheric Pressure Ionization (API) Mass Spectrometry. Solvent-Mediated Ionization of Samples Introduced in Solution and in a Liquid Chromatograph Effluent Stream. *Journal Of Chromatographic Science* **12**, 725–729.
- (65) Huang, E. C.; Wachs, T.; Conboy, J. J.; Henion, J. D. Atmospheric Pressure Ionization Mass Spectrometry. Detection for the Separation Sciences. *Anal. Chem.* **1990**, *62* (13), 713A-725A. <https://doi.org/10.1021/ac00212a002>.
- (66) Robb, D. B.; Covey, T. R.; Bruins, A. P. Atmospheric Pressure Photoionization: An Ionization Method for Liquid Chromatography–Mass Spectrometry. *Anal. Chem.* **2000**, *72* (15), 3653–3659. <https://doi.org/10.1021/ac0001636>.
- (67) Kauppila, T. J.; Kuuranne, T.; Meurer, E. C.; Eberlin, M. N.; Kotiaho, T.; Kostianen, R. Atmospheric Pressure Photoionization Mass Spectrometry. Ionization Mechanism and the Effect of Solvent on the Ionization of Naphthalenes. *Anal. Chem.* **2002**, *74* (21), 5470–5479. <https://doi.org/10.1021/ac025659x>.
- (68) Marchi, I.; Rudaz, S.; Veuthey, J. Atmospheric Pressure Photoionization for Coupling Liquid-Chromatography to Mass Spectrometry: A Review. *Talanta* **2009**, *78* (1), 1–18. <https://doi.org/10.1016/j.talanta.2008.11.031>.
- (69) Yamashita, M.; Fenn, J. B. Negative Ion Production with the Electrospray Ion Source. *The Journal of Physical Chemistry* **1984**, *88* (20), 4671–4675. <https://doi.org/10.1021/j150664a046>.
- (70) Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **2001**, *20* (6), 362–387. <https://doi.org/10.1002/mas.10008>.
- (71) Syage, J. A.; Short, L. C.; Cai, S. S. Atmospheric Pressure Photoionization — The Second Source for LC-MS? *LCGC North America* **26** (3), 286–296.
- (72) Tian, H.; Bai, J.; An, Z.; Chen, Y.; Zhang, R.; He, J.; Bi, X.; Song, Y.; Abliz, Z. Plasma Metabolome Analysis by Integrated Ionization Rapid-resolution Liquid Chromatography/Tandem Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **2013**, *27*, 2071–2080. <https://doi.org/10.1002/rcm.6666>.
- (73) Vorkas, P. A.; Isaac, G.; Anwar, M. A.; Davies, A. H.; Want, E. J.; Nicholson, J. K.; Holmes, E. Untargeted UPLC-MS Profiling Pipeline to Expand Tissue Metabolome Coverage: Application to Cardiovascular Disease.

Références

- Anal. Chem.* **2015**, *87* (8), 4184–4193. <https://doi.org/10.1021/ac503775m>.
- (74) Paul, W. Electromagnetic Traps for Charged and Neutral Particles (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **1990**, *29* (7), 739–748. <https://doi.org/10.1002/anie.199007391>.
- (75) Introduction to mass analyzers : SHIMADZU (Shimadzu Corporation) https://www.shimadzu.com/an/lcms/support/fundamental/mass_analyzers.html (accessed Apr 2, 2020).
- (76) Londry, F. A.; Hager, J. W. Mass Selective Axial Ion Ejection from a Linear Quadrupole Ion Trap. *J Am Soc Mass Spectrom* **2003**, *14* (10), 1130–1147. [https://doi.org/10.1016/S1044-0305\(03\)00446-X](https://doi.org/10.1016/S1044-0305(03)00446-X).
- (77) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: A New Mass Spectrometer. *J. Mass Spectrom.* **2005**, *40* (4), 430–443. <https://doi.org/10.1002/jms.856>.
- (78) Guilhaus, M. Special Feature: Tutorial. Principles and Instrumentation in Time-of-Flight Mass Spectrometry. Physical and Instrumental Concepts. *J. Mass Spectrom.* **1995**, *30* (11), 1519–1532. <https://doi.org/10.1002/jms.1190301102>.
- (79) Vestal, M. L.; Juhasz, P.; Martin, S. A. Delayed Extraction Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1995**, *9* (11), 1044–1050. <https://doi.org/10.1002/rcm.1290091115>.
- (80) Mamyrin, B. A. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects). *International Journal of Mass Spectrometry* **2001**, *206* (3), 251–266. [https://doi.org/10.1016/S1387-3806\(00\)00392-4](https://doi.org/10.1016/S1387-3806(00)00392-4).
- (81) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162. <https://doi.org/10.1021/ac991131p>.
- (82) Makarov, A.; Scigelova, M. Coupling Liquid Chromatography to Orbitrap Mass Spectrometry. *Journal of Chromatography A* **2010**, *1217* (25), 3938–3945. <https://doi.org/10.1016/j.chroma.2010.02.022>.
- (83) Scigelova, M.; Makarov, A. Orbitrap Mass Analyzer – Overview and Applications in Proteomics. *Proteomics* **2006**, *6* (S2), 16–21. <https://doi.org/10.1002/pmic.200600528>.
- (84) Hardman, M.; Makarov, A. A. Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. *Anal. Chem.* **2003**, *75* (7), 1699–1705. <https://doi.org/10.1021/ac0258047>.
- (85) Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer <http://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFADBMBHQ> (accessed Apr 3, 2020).
- (86) Spectromètre de masse Orbitrap Fusion™ Tribrid™ <http://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFADBMBCX> (accessed Apr 3, 2020).
- (87) Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* **1974**, *25* (2), 282–283. [https://doi.org/10.1016/0009-2614\(74\)89137-2](https://doi.org/10.1016/0009-2614(74)89137-2).
- (88) Schrader, W.; Klein, H.-W. Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (LC-FTICR MS): An Early Overview. *Anal Bioanal Chem* **2004**, *379* (7–8). <https://doi.org/10.1007/s00216-004-2675-1>.
- (89) Marshall, A. G.; Hendrickson, C. L. Fourier Transform Ion Cyclotron Resonance Detection: Principles and Experimental Configurations. *International Journal of Mass Spectrometry* **2002**, *215* (1–3), 59–75. [https://doi.org/10.1016/S1387-3806\(01\)00588-7](https://doi.org/10.1016/S1387-3806(01)00588-7).
- (90) Himmelsbach, M. 10years of MS Instrumental Developments – Impact on LC–MS/MS in Clinical Chemistry. *Journal of Chromatography B* **2012**, *883–884*, 3–17. <https://doi.org/10.1016/j.jchromb.2011.11.038>.
- (91) Habchi, B.; Alves, S.; Jouan-Rimbaud Bouveresse, D.; Appenzeller, B.; Paris, A.; Rutledge, D. N.; Rathahao-Paris, E. Potential of Dynamically Harmonized Fourier Transform Ion Cyclotron Resonance Cell for High-Throughput Metabolomics Fingerprinting: Control of Data Quality. *Anal Bioanal Chem* **2018**, *410* (2), 483–490. <https://doi.org/10.1007/s00216-017-0738-3>.
- (92) Bennett, P. Future Developments for Triple Quadrupole Mass Spectrometers: Not Just Hardware. *Bioanalysis* **2011**, *3* (7), 709–711. <https://doi.org/10.4155/bio.11.41>.
- (93) Beck, S.; Michalski, A.; Raether, O.; Lubeck, M.; Kaspar, S.; Goedecke, N.; Baessmann, C.; Hornburg, D.; Meier, F.; Paron, I.; Kulak, N. A.; Cox, J.; Mann, M. The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Molecular & Cellular Proteomics* **2015**, *14* (7), 2014–2029. <https://doi.org/10.1074/mcp.M114.047407>.
- (94) Paxton, T. High Sensitivity Collisionally-Activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-Acceleration Time-of-Flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry* **1996**, *10*, 889–896. [https://doi.org/10.1002/\(SICI\)1097-0231\(19960610\)10:8<889::AID-RCM615>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0231(19960610)10:8<889::AID-RCM615>3.0.CO;2-F).
- (95) Ojanperä, I.; Kolmonen, M.; Pelander, A. Current Use of High-Resolution Mass Spectrometry in Drug Screening Relevant to Clinical and Forensic Toxicology and Doping Control. *Anal Bioanal Chem* **2012**, *403* (5), 1203–1220. <https://doi.org/10.1007/s00216-012-5726-z>.
- (96) Ramanathan, R.; Jemal, M.; Ramagiri, S.; Xia, Y.-Q.; Humphreys, W. G.; Olah, T.; Korfmacher, W. A. It Is Time

Références

- for a Paradigm Shift in Drug Discovery Bioanalysis: From SRM to HRMS. *J. Mass Spectrom.* **2011**, *46* (6), 595–601. <https://doi.org/10.1002/jms.1921>.
- (97) Stark, T.; Wollmann, N.; Lösch, S.; Hofmann, T. Quantitation of Resveratrol in Red Wines by Means of Stable Isotope Dilution Analysis–Ultra-Performance Liquid Chromatography–Quan-Time-of-Flight Mass Spectrometry and Cross Validation. *Anal. Chem.* **2011**, *83* (9), 3398–3405. <https://doi.org/10.1021/ac103305s>.
- (98) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass Spectrometry-Based Proteomics Using Q Exactive, a High-Performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular and Cellular Proteomics* **2011**, *10* (9), 12. <https://doi.org/10.1074/mcp.M111.011015-11>.
- (99) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-Filter, High-Performance Quadrupole and an Ultra-High-Field Orbitrap Analyzer. *Mol Cell Proteomics* **2014**, *13* (12), 3698–3708. <https://doi.org/10.1074/mcp.M114.043489>.
- (100) Zhang, N. R.; Yu, S.; Tiller, P.; Yeh, S.; Mahan, E.; Emary, W. B. Quantitation of Small Molecules Using High-Resolution Accurate Mass Spectrometers – a Different Approach for Analysis of Biological Samples. *Rapid Communications in Mass Spectrometry* **2009**, *23*, 1085–1094. <https://doi.org/10.1002/rcm.3975>.
- (101) Thermo Scientific Q Exactive HF-X Hybrid Quadrupole-Orbitrap MS System. Thermo Fisher Scientific 2017.
- (102) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *J Am Soc Mass Spectrom* **2006**, *17* (7), 977–982. <https://doi.org/10.1016/j.jasms.2006.03.006>.
- (103) 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. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* **2013**, *85* (24), 11710–11714. <https://doi.org/10.1021/ac403115c>.
- (104) Thermo Fisher :: Orbitrap :: Orbitrap Fusion <https://planetorbitrap.com/orbitrap-fusion#tab:schematic> (accessed Apr 2, 2020).
- (105) Stavrianidi, A. A Classification of Liquid Chromatography Mass Spectrometry Techniques for Evaluation of Chemical Composition and Quality Control of Traditional Medicines. *Journal of Chromatography A* **2020**, *1609*, 460501. <https://doi.org/10.1016/j.chroma.2019.460501>.
- (106) 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. 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>.
- (107) 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 Tribid, High-Field Orbitrap Instrument. *PLoS ONE* **2016**, *11* (7), e0160160. <https://doi.org/10.1371/journal.pone.0160160>.
- (108) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-Type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Mol Cell Proteomics* **2011**, *10* (12), 1–9. <https://doi.org/10.1074/mcp.M111.009910>.
- (109) Laphorn, C.; Pullen, F.; Chowdhry, B. Z. Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS) of Small Molecules: Separating and Assigning Structures to Ions. *Mass Spectrom. Rev.* **2013**, *32* (1), 43–71. <https://doi.org/10.1002/mas.21349>.
- (110) Kaplan, K. A.; Chiu, V. M.; Lukus, P. A.; Zhang, X.; Siems, W. F.; Schenk, J. O.; Hill, H. H. Neuronal Metabolomics by Ion Mobility Mass Spectrometry: Cocaine Effects on Glucose and Selected Biogenic Amine Metabolites in the Frontal Cortex, Striatum, and Thalamus of the Rat. *Anal Bioanal Chem* **2013**, *405* (6), 1959–1968. <https://doi.org/10.1007/s00216-012-6638-7>.
- (111) Turner, R. B.; Brokenshire, J. L. Hand-Held Ion Mobility Spectrometers. *TrAC Trends in Analytical Chemistry* **1994**, *13* (7), 275–280. [https://doi.org/10.1016/0165-9936\(94\)87064-0](https://doi.org/10.1016/0165-9936(94)87064-0).
- (112) Kaszycki, J. L.; La Rotta, A.; Colsch, B.; Fenaille, F.; Daully, C.; Kamleh, A.; Wu, C. Separation of Biologically Relevant Isomers on an Orbitrap Mass Spectrometer Using High-resolution Drift Tube Ion Mobility and Varied Drift Gas Mixtures. *Rapid Commun Mass Spectrom* **2019**, *33* (S2), 3–10. <https://doi.org/10.1002/rcm.8414>.
- (113) Zhang, X.; Quinn, K.; Cruickshank-Quinn, C.; Reisdorph, R.; Reisdorph, N. The Application of Ion Mobility Mass Spectrometry to Metabolomics. *Current Opinion in Chemical Biology* **2018**, *42*, 60–66. <https://doi.org/10.1016/j.cbpa.2017.11.001>.

Références

- (114) Jünger, M.; Bödeker, B.; Baumbach, J. I. Peak Assignment in Multi-Capillary Column–Ion Mobility Spectrometry Using Comparative Studies with Gas Chromatography–Mass Spectrometry for VOC Analysis. *Anal Bioanal Chem* **2010**, *396* (1), 471–482. <https://doi.org/10.1007/s00216-009-3168-z>.
- (115) Paglia, G.; Astarita, G. Metabolomics and Lipidomics Using Traveling-Wave Ion Mobility Mass Spectrometry. *Nat Protoc* **2017**, *12* (4), 797–813. <https://doi.org/10.1038/nprot.2017.013>.
- (116) May, J. C.; Morris, C. B.; McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Anal. Chem.* **2017**, *89* (2), 1032–1044. <https://doi.org/10.1021/acs.analchem.6b04905>.
- (117) Mairinger, T.; Causon, T. J.; Hann, S. The Potential of Ion Mobility–Mass Spectrometry for Non-Targeted Metabolomics. *Current Opinion in Chemical Biology* **2018**, *42*, 9–15. <https://doi.org/10.1016/j.cbpa.2017.10.015>.
- (118) Harry, E. L.; Weston, D. J.; Bristow, A. W. T.; Wilson, I. D.; Creaser, C. S. An Approach to Enhancing Coverage of the Urinary Metabonome Using Liquid Chromatography–Ion Mobility–Mass Spectrometry. *Journal of Chromatography B* **2008**, *871* (2), 357–361. <https://doi.org/10.1016/j.jchromb.2008.04.043>.
- (119) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G. An Optimized Protocol for Metabolome Analysis in Yeast Using Direct Infusion Electrospray Mass Spectrometry. *Phytochemistry* **2003**, *62* (6), 929–937. [https://doi.org/10.1016/S0031-9422\(02\)00713-6](https://doi.org/10.1016/S0031-9422(02)00713-6).
- (120) Madalinski, G.; Godat, E.; Alves, S.; Lesage, D.; Genin, E.; Levi, P.; Labarre, J.; Tabet, J.-C.; Ezan, E.; Junot, C. Direct Introduction of Biological Samples into a LTQ-Orbitrap Hybrid Mass Spectrometer as a Tool for Fast Metabolome Analysis. *Anal. Chem.* **2008**, *80* (9), 3291–3303. <https://doi.org/10.1021/ac7024915>.
- (121) Southam, A. D.; Weber, R. J. M.; Engel, J.; Jones, M. R.; Viant, M. R. A Complete Workflow for High-Resolution Spectral-Stitching Nanoelectrospray Direct-Infusion Mass-Spectrometry-Based Metabolomics and Lipidomics. *Nat Protoc* **2017**, *12* (2), 310–328. <https://doi.org/10.1038/nprot.2016.156>.
- (122) Han, J.; Datta, R.; Chan, S.; Borchers, C. H. Mass Spectrometry-Based Technologies for High-Throughput Metabolomics. *Bioanalysis* **2009**, *1* (9), 1665–1684. <https://doi.org/10.4155/bio.09.158>.
- (123) Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y. Gas Chromatography/Mass Spectrometry in Metabolic Profiling of Biological Fluids. *Journal of Chromatography B* **2008**, *871* (2), 202–211. <https://doi.org/10.1016/j.jchromb.2008.04.033>.
- (124) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M. Chemical Derivatization and Mass Spectral Libraries in Metabolic Profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany* **2005**, *56* (410), 219–243. <https://doi.org/10.1093/jxb/eri069>.
- (125) Birkemeyer, C.; Kolasa, A.; Kopka, J. Comprehensive Chemical Derivatization for Gas Chromatography–Mass Spectrometry-Based Multi-Targeted Profiling of the Major Phytohormones. *Journal of Chromatography A* **2003**, *993* (1–2), 89–102. [https://doi.org/10.1016/S0021-9673\(03\)00356-X](https://doi.org/10.1016/S0021-9673(03)00356-X).
- (126) Kawana, S.; Nakagawa, K.; Hasegawa, Y.; Kobayashi, H.; Yamaguchi, S. Improvement of Sample Throughput Using Fast Gas Chromatography Mass-Spectrometry for Biochemical Diagnosis of Organic Acid Disorders. *Clinica Chimica Acta* **2008**, *392* (1–2), 34–40. <https://doi.org/10.1016/j.cca.2008.02.025>.
- (127) Li, X.; Xu, Z.; Lu, X.; Yang, X.; Yin, P.; Kong, H.; Yu, Y.; Xu, G. Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry for Metabolomics: Biomarker Discovery for Diabetes Mellitus. *Analytica Chimica Acta* **2009**, *633* (2), 257–262. <https://doi.org/10.1016/j.aca.2008.11.058>.
- (128) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Pierce, K. M.; Young, E. T.; Synovec, R. E. Comprehensive Analysis of Yeast Metabolite GC×GC–TOFMS Data: Combining Discovery-Mode and Deconvolution Chemometric Software. *Analyst* **2007**, *132* (8), 756–767. <https://doi.org/10.1039/B700061H>.
- (129) Castillo, S.; Mattila, I.; Miettinen, J.; Orešič, M.; Hyötyläinen, T. Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2011**, *83* (8), 3058–3067. <https://doi.org/10.1021/ac103308x>.
- (130) Want, E. J.; Nordstro, A. From Exogenous to Endogenous: The Inevitable Imprint of Mass Spectrometry in Metabolomics. *Journal of Proteome Research* **2007**, *6*, 459–468. <https://doi.org/10.1021/pr060505+>.
- (131) Euerby, M. R.; Petersson, P. Chromatographic Classification and Comparison of Commercially Available Reversed-Phase Liquid Chromatographic Columns Using Principal Component Analysis. *Journal of Chromatography A* **2003**, *994* (1–2), 13–36. [https://doi.org/10.1016/S0021-9673\(03\)00393-5](https://doi.org/10.1016/S0021-9673(03)00393-5).
- (132) Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G. 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.* **2012**, *84* (4), 1994–2001. <https://doi.org/10.1021/ac2030738>.
- (133) Kirkland, J. Development of Some Stationary Phases for Reversed-Phase HPLC. *Journal of Chromatography A* **2004**, *1060* (1–2), 9–21. [https://doi.org/10.1016/S0021-9673\(04\)01892-8](https://doi.org/10.1016/S0021-9673(04)01892-8).
- (134) Vervoort, R. J. M.; Ruyter, E.; Debets, A. J. J.; Claessens, H. A.; Cramers, C. A.; de Jong, G. J. Characterisation

Références

- of Reversed-Phase Stationary Phases for the Liquid Chromatographic Analysis of Basic Pharmaceuticals by Thermodynamic Data. *Journal of Chromatography A* **2002**, *964* (1–2), 67–76. [https://doi.org/10.1016/S0021-9673\(02\)00589-7](https://doi.org/10.1016/S0021-9673(02)00589-7).
- (135) Stella, C.; Rudaz, S.; Gauvrit, J.-Y.; Lantéri, P.; Huteau, A.; Tchaplà, A.; Veuthey, J.-L. Characterization and Comparison of the Chromatographic Performance of Different Types of Reversed-Phase Stationary Phases. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43* (1), 89–98. <https://doi.org/10.1016/j.jpba.2006.06.018>.
- (136) Euerby, M. R.; Petersson, P.; Campbell, W.; Roe, W. Chromatographic Classification and Comparison of Commercially Available Reversed-Phase Liquid Chromatographic Columns Containing Phenyl Moieties Using Principal Component Analysis. *Journal of Chromatography A* **2007**, *1154* (1–2), 138–151. <https://doi.org/10.1016/j.chroma.2007.03.119>.
- (137) Chirita, R.-I.; West, C.; Zubrzycki, S.; Finaru, A.-L.; Elfakir, C. Investigations on the Chromatographic Behaviour of Zwitterionic Stationary Phases Used in Hydrophilic Interaction Chromatography. *Journal of Chromatography A* **2011**, *1218* (35), 5939–5963. <https://doi.org/10.1016/j.chroma.2011.04.002>.
- (138) Alpert, J. Hydrophilic-Interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds. *Journal of Chromatography* **1990**, *499*, 177–196. [https://doi.org/10.1016/S0021-9673\(00\)96972-3](https://doi.org/10.1016/S0021-9673(00)96972-3).
- (139) Boudah, S.; Olivier, M.-F.; Aros-Calt, S.; Oliveira, L.; Fenaille, F.; Tabet, J.-C.; Junot, C. Annotation of the Human Serum Metabolome by Coupling Three Liquid Chromatography Methods to High-Resolution Mass Spectrometry. *Journal of Chromatography B* **2014**, *966*, 34–47. <https://doi.org/10.1016/j.jchromb.2014.04.025>.
- (140) Moreau, R.; Clària, J.; Aguilar, F.; Fenaille, F.; Lozano, J. J.; Junot, C.; Colsch, B.; Caraceni, P.; Trebicka, J.; Pavesi, M.; Alessandria, C.; Nevens, F.; Saliba, F.; Welzel, T. M.; Albillos, A.; Gustot, T.; Fernández, J.; Moreno, C.; Baldassarre, M.; Zaccherini, G.; Piano, S.; Montagnese, S.; Vargas, V.; Genescà, J.; Solà, E.; Bernal, W.; Butin, N.; Hautbergue, T.; Cholet, S.; Castelli, F.; Jansen, C.; Steib, C.; Campion, D.; Mookerjee, R.; Rodríguez-Gandía, M.; Soriano, G.; Durand, F.; Benten, D.; Bañares, R.; Stauber, R. E.; Gronbaek, H.; Coenraad, M. J.; Ginès, P.; Gerbes, A.; Jalan, R.; Bernardi, M.; Arroyo, V.; Angeli, P. Blood Metabolomics Uncovers Inflammation-Associated Mitochondrial Dysfunction as a Potential Mechanism Underlying ACLF. *Journal of Hepatology* **2020**, *72* (4), 688–701. <https://doi.org/10.1016/j.jhep.2019.11.009>.
- (141) Leandro, C. C.; Hancock, P.; Fussell, R. J.; Keely, B. J. Comparison of Ultra-Performance Liquid Chromatography and High-Performance Liquid Chromatography for the Determination of Priority Pesticides in Baby Foods by Tandem Quadrupole Mass Spectrometry. *Journal of Chromatography A* **2006**, *1103* (1), 94–101. <https://doi.org/10.1016/j.chroma.2005.10.077>.
- (142) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. High Resolution “Ultra Performance” Liquid Chromatography Coupled to Oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies. *J. Proteome Res.* **2005**, *4* (2), 591–598. <https://doi.org/10.1021/pr049769r>.
- (143) The Nobel Prize in Chemistry 1948 <https://www.nobelprize.org/prizes/chemistry/1948/tiselius/facts/> (accessed Apr 21, 2020).
- (144) Hirayama, A.; Wakayama, M.; Soga, T. Metabolome Analysis Based on Capillary Electrophoresis-Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2014**, *61*, 215–222. <https://doi.org/10.1016/j.trac.2014.05.005>.
- (145) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS for Metabolomics: Developments and Applications in the Period 2010-2012. *ELECTROPHORESIS* **2013**, *34* (1), 86–98. <https://doi.org/10.1002/elps.201200390>.
- (146) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS for Metabolomics: Developments and Applications in the Period 2014-2016. *ELECTROPHORESIS* **2017**, *38* (1), 190–202. <https://doi.org/10.1002/elps.201600370>.
- (147) Ramautar, R.; Busnel, J.-M.; Deelder, A. M.; Mayboroda, O. A. Enhancing the Coverage of the Urinary Metabolome by Sheathless Capillary Electrophoresis-Mass Spectrometry. *Anal. Chem.* **2012**, *84* (2), 885–892. <https://doi.org/10.1021/ac202407v>.
- (148) Kuehnbaum, N. L.; Britz-McKibbin, P. New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era. *Chem. Rev.* **2013**, *113* (4), 2437–2468. <https://doi.org/10.1021/cr300484s>.
- (149) Kohler, I.; Verhoeven, A.; Derks, R. J.; Giera, M. Analytical Pitfalls and Challenges in Clinical Metabolomics. *Bioanalysis* **2016**, *8* (14), 1509–1532. <https://doi.org/10.4155/bio-2016-0090>.
- (150) Ramautar, R.; Mayboroda, O. A.; Somsen, G. W.; de Jong, G. J. CE-MS for Metabolomics: Developments and Applications in the Period 2008-2010. *ELECTROPHORESIS* **2011**, *32* (1), 52–65.

Références

- <https://doi.org/10.1002/elps.201000378>.
- (151) Ptolemy, A. S.; Britz-McKibbin, P. New Advances in On-Line Sample Preconcentration by Capillary Electrophoresis Using Dynamic PH Junction. *Analyst* **2008**, *133* (12), 1643. <https://doi.org/10.1039/b808236g>.
- (152) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS for Metabolomics: Developments and Applications in the Period 2016-2018. *ELECTROPHORESIS* **2019**, *40* (1), 165–179. <https://doi.org/10.1002/elps.201800323>.
- (153) Büscher, J. M.; Czernik, D.; Ewald, J. C.; Sauer, U.; Zamboni, N. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Anal. Chem.* **2009**, *81* (6), 2135–2143. <https://doi.org/10.1021/ac8022857>.
- (154) Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D. Current Practice of Liquid Chromatography–Mass Spectrometry in Metabolomics and Metabonomics. *Journal of Pharmaceutical and Biomedical Analysis* **2014**, *87*, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>.
- (155) Viant, M. R.; Kurland, I. J.; Jones, M. R.; Dunn, W. B. How Close Are We to Complete Annotation of Metabolomes? *Current Opinion in Chemical Biology* **2017**, *36*, 64–69. <https://doi.org/10.1016/j.cbpa.2017.01.001>.
- (156) DeBerardinis, R. J.; Thompson, C. B. Cellular Metabolism and Disease: What Do Metabolic Outliers Teach Us? *Cell* **2012**, *148* (6), 1132–1144. <https://doi.org/10.1016/j.cell.2012.02.032>.
- (157) Demain, A. L.; Fang, A. The Natural Functions of Secondary Metabolites. In *History of Modern Biotechnology I*; Fiechter, A., Ed.; Scheper, T., Babel, W., Blanch, H. W., Cooney, C. L., Endo, I., Enfors, S.-O., Eriksson, K.-E. L., Fiechter, A., Hoare, M., Mattiasson, B., Rehm, H. J., Rogers, P. L., Sahm, H., Schügerl, K., Stephanopoulos, G., Tsao, G. T., Venkat, K., Villadsen, J., von Stockar, U., Wandrey, C., Series Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2000; Vol. 69, pp 1–39. https://doi.org/10.1007/3-540-44964-7_1.
- (158) Cajka, T.; Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **2016**, *88* (1), 524–545. <https://doi.org/10.1021/acs.analchem.5b04491>.
- (159) Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C. E. Challenges, Progress and Promises of Metabolite Annotation for LC–MS-Based Metabolomics. *Current Opinion in Biotechnology* **2019**, *55*, 44–50. <https://doi.org/10.1016/j.copbio.2018.07.010>.
- (160) Sindelar, M.; Patti, G. J. Chemical Discovery in the Era of Metabolomics. *J. Am. Chem. Soc.* **2020**, *jacs.9b13198*. <https://doi.org/10.1021/jacs.9b13198>.
- (161) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. *Anal. Chem.* **2018**, *90* (1), 480–489. <https://doi.org/10.1021/acs.analchem.7b03929>.
- (162) Damont, A.; Olivier, M.-F.; Warnet, A.; Lyan, B.; Pujos-Guillot, E.; Jamin, E. L.; Debrauwer, L.; Bernillon, S.; Junot, C.; Tabet, J.-C.; Fenaille, F. Proposal for a Chemically Consistent Way to Annotate Ions Arising from the Analysis of Reference Compounds under ESI Conditions: A Prerequisite to Proper Mass Spectral Database Constitution in Metabolomics. *J Mass Spectrom* **2019**, *54* (6), 567–582. <https://doi.org/10.1002/jms.4372>.
- (163) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. Expanding Coverage of the Metabolome for Global Metabolite Profiling. *Anal. Chem.* **2011**, *83* (6), 2152–2161. <https://doi.org/10.1021/ac102981k>.
- (164) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89* (19), 10397–10406. <https://doi.org/10.1021/acs.analchem.7b02380>.
- (165) Vinaixa, M.; Schymanski, E. L.; Neumann, S.; Navarro, M.; Salek, R. M.; Yanes, O. Mass Spectral Databases for LC/MS- and GC/MS-Based Metabolomics: State of the Field and Future Prospects. *TrAC Trends in Analytical Chemistry* **2016**, *78*, 23–35. <https://doi.org/10.1016/j.trac.2015.09.005>.
- (166) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), 31. <https://doi.org/10.3390/metabo8020031>.
- (167) 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* (1), 234. <https://doi.org/10.1186/1471-2105-7-234>.
- (168) Kerber, A.; Laue, R.; Meringer, M.; Rucker, C. Molecules in Silico: The Generation of Structural Formulae and Its Applications. *J. Comput. Chem. Jpn.* **2004**, *3* (3), 85–96. <https://doi.org/10.2477/jccj.3.85>.
- (169) 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>.

Références

- (170) 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. Identification of Small Molecules Using Accurate Mass MS/MS Search. *Mass Spec Rev* **2018**, *37* (4), 513–532. <https://doi.org/10.1002/mas.21535>.
- (171) Oberacher, H.; Reinstadler, V.; Kreidl, M.; Stravs, M.; Hollender, J.; Schymanski, E. Annotating Nontargeted LC-HRMS/MS Data with Two Complementary Tandem Mass Spectral Libraries. *Metabolites* **2018**, *9* (1), 3. <https://doi.org/10.3390/metabo9010003>.
- (172) Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S. S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; Jahn, D.; Fiehn, O. Comprehensive Comparison of in Silico MS/MS Fragmentation Tools of the CASMI Contest: Database Boosting Is Needed to Achieve 93% Accuracy. *J Cheminform* **2017**, *9* (1), 32. <https://doi.org/10.1186/s13321-017-0219-x>.
- (173) 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. Proposed Minimum Reporting Standards for Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3* (3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>.
- (174) Metabolomics Society News. *MetaboNews* **2020**, *10* (2), 1–19.
- (175) Yan, Z.; Yan, R. 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. *Analytica Chimica Acta* **2015**, *894*, 65–75. <https://doi.org/10.1016/j.aca.2015.08.056>.
- (176) Wei, R.; Li, G.; Seymour, A. B. High-Throughput and Multiplexed LC/MS/MRM Method for Targeted Metabolomics. *Anal. Chem.* **2010**, *82* (13), 5527–5533. <https://doi.org/10.1021/ac100331b>.
- (177) 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. Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling. *Anal. Chem.* **2015**, *87* (2), 884–891. <https://doi.org/10.1021/ac5025649>.
- (178) Lim, H.-K.; Chen, J.; Cook, K.; Sensenhauser, C.; Silva, J.; Evans, D. C. A Generic Method to Detect Electrophilic Intermediates Using Isotopic Pattern Triggered Data-Dependent High-Resolution Accurate Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22* (8), 1295–1311. <https://doi.org/10.1002/rcm.3504>.
- (179) Zhang, H.; Zhang, D.; Ray, K.; Zhu, M. Mass Defect Filter Technique and Its Applications to Drug Metabolite Identification by High-Resolution Mass Spectrometry. *J. Mass Spectrom.* **2009**, *44* (7), 999–1016. <https://doi.org/10.1002/jms.1610>.
- (180) Zhu, X.; Chen, Y.; Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MS/MS 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>.
- (181) 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? *Journal of Chromatography A* **2017**, *1526*, 1–12. <https://doi.org/10.1016/j.chroma.2017.10.043>.
- (182) Perry, S. J.; Nász, S.; Saeed, M. 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). *Rapid Commun. Mass Spectrom.* **2015**, *29* (17), 1545–1555. <https://doi.org/10.1002/rcm.7250>.
- (183) Hilaire, P. B. S.; Rousseau, K.; Seyer, A.; Dechaumet, S.; Damont, A.; Junot, C.; Fenaille, F. Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics. *Metabolites* **2020**, *10* (4), 158. <https://doi.org/10.3390/metabo10040158>.
- (184) Hopfgartner, G. Can MS Fully Exploit the Benefits of Fast Chromatography? *Bioanalysis* **2011**, *3* (2), 121–123. <https://doi.org/10.4155/bio.10.191>.
- (185) Hopfgartner, G.; Tonoli, D.; Varesio, E. High-Resolution Mass Spectrometry for Integrated Qualitative and Quantitative Analysis of Pharmaceuticals in Biological Matrices. *Anal Bioanal Chem* **2012**, *402* (8), 2587–2596. <https://doi.org/10.1007/s00216-011-5641-8>.
- (186) Chapman, J. D.; Goodlett, D. R.; Masselon, C. D. Multiplexed and Data-Independent Tandem Mass Spectrometry for Global Proteome Profiling: MULTIPLEXED AND DATA-INDEPENDENT TANDEM MS. *Mass Spec Rev* **2014**, *33* (6), 452–470. <https://doi.org/10.1002/mas.21400>.
- (187) Anjo, S. I.; Santa, C.; Manadas, B. SWATH-MS as a Tool for Biomarker Discovery: From Basic Research to Clinical Applications. *Proteomics* **2017**, *17* (3–4), 1600278. <https://doi.org/10.1002/pmic.201600278>.

Références

- (188) 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>.
- (189) 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*, 1989–1994. <https://doi.org/10.1002/rcm.2550>.
- (190) Naz, S.; Gallart-Ayala, H.; Reinke, S. N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C. E. Development of a Liquid Chromatography–High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal. Chem.* **2017**, *89* (15), 7933–7942. <https://doi.org/10.1021/acs.analchem.7b00925>.
- (191) Bonner, R.; Hopfgartner, G. SWATH Data Independent Acquisition Mass Spectrometry for Metabolomics. *TrAC Trends in Analytical Chemistry* **2019**, *120*, 115278. <https://doi.org/10.1016/j.trac.2018.10.014>.
- (192) Gillet, L. C.; Navarro, P.; Tate, S.; 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. *Molecular & Cellular Proteomics* **2012**, *11* (6), 1–17. <https://doi.org/10.1074/mcp.O111.016717>.
- (193) Arnhard, K.; Gottschall, A.; Pittler, F.; Oberacher, H. 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* **2015**, *407* (2), 405–414. <https://doi.org/10.1007/s00216-014-8262-1>.
- (194) 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>.
- (195) Bruderer, T.; Varesio, E.; Hopfgartner, G. The Use of LC Predicted Retention Times to Extend Metabolites Identification with SWATH Data Acquisition. *Journal of Chromatography B* **2017**, *1071*, 3–10. <https://doi.org/10.1016/j.jchromb.2017.07.016>.
- (196) 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>.
- (197) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nature Methods* **2015**, *12* (6), 523–526. <https://doi.org/10.1038/nmeth.3393>.
- (198) Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing Strategies and Software Solutions for Data-Independent Acquisition in Mass Spectrometry. *Proteomics* **2015**, *15* (5–6), 964–980. <https://doi.org/10.1002/pmic.201400323>.
- (199) Yin, Y.; Wang, R.; Cai, Y.; Wang, Z.; Zhu, Z.-J. DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics. *Anal. Chem.* **2019**, *91* (18), 11897–11904. <https://doi.org/10.1021/acs.analchem.9b02655>.
- (200) Raetz, M.; Bonner, R.; Hopfgartner, G. SWATH-MS for Metabolomics and Lipidomics: Critical Aspects of Qualitative and Quantitative Analysis. *Metabolomics* **2020**, *16*:71. <https://doi.org/10.1007/s11306-020-01692-0>.
- (201) Cuevas-Delgado, P.; Dudzik, D.; Miguel, V.; Lamas, S.; Barbas, C. Data-Dependent Normalization Strategies for Untargeted Metabolomics—a Case Study. *Anal Bioanal Chem* **2020**. <https://doi.org/10.1007/s00216-020-02594-9>.
- (202) Wang, S.; Yang, H. PseudoQC: A Regression-Based Simulation Software for Correction and Normalization of Complex Metabolomics and Proteomics Datasets. *Proteomics* **2019**, *19* (19), 1900264. <https://doi.org/10.1002/pmic.201900264>.
- (203) Misra, B. B. Data Normalization Strategies in Metabolomics: Current Challenges, Approaches, and Tools. *Eur J Mass Spectrom (Chichester)* **2020**, *26* (3), 165–174. <https://doi.org/10.1177/1469066720918446>.
- (204) Nilsson, L. B.; Eklund, G. Direct Quantification in Bioanalytical LC-MS/MS Using Internal Calibration via Analyte/Stable Isotope Ratio. *J. Pharm. Biomed. Anal.* **2007**, *43* (3), 1094–1099. <https://doi.org/10.1016/j.jpba.2006.09.030>.
- (205) Zhou, J.; Yin, Y. Strategies for Large-Scale Targeted Metabolomics Quantification by Liquid Chromatography-Mass Spectrometry. *The Analyst* **2016**, *141* (23), 6362–6373. <https://doi.org/10.1039/C6AN01753C>.
- (206) Nemkov, T.; Hansen, K. C.; D'Alessandro, A. A Three-Minute Method for High-Throughput Quantitative Metabolomics and Quantitative Tracing Experiments of Central Carbon and Nitrogen Pathways: Three-

Références

- Minute Semi-Targeted Metabolomics. *Rapid Commun. Mass Spectrom.* **2017**, *31* (8), 663–673. <https://doi.org/10.1002/rcm.7834>.
- (207) Rochat, B. From Targeted Quantification to Untargeted Metabolomics: Why LC-High-Resolution-MS Will Become a Key Instrument in Clinical Labs. *TrAC Trends in Analytical Chemistry* **2016**, *84*, 151–164. <https://doi.org/10.1016/j.trac.2016.02.009>.
- (208) Chen, S.; Kong, H.; Lu, X.; Li, Y.; Yin, P.; Zeng, Z.; Xu, G. 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.* **2013**, *85* (17), 8326–8333. <https://doi.org/10.1021/ac4016787>.
- (209) Kitteringham, N. R.; Jenkins, R. E.; Lane, C. S.; Elliott, V. L.; Park, B. K. Multiple Reaction Monitoring for Quantitative Biomarker Analysis in Proteomics and Metabolomics☆. *Journal of Chromatography B* **2009**, *877* (13), 1229–1239. <https://doi.org/10.1016/j.jchromb.2008.11.013>.
- (210) Burnum-Johnson, K. E.; Nie, S.; Casey, C. P.; Monroe, M. E.; Orton, D. J.; Ibrahim, Y. M.; Gritsenko, M. A.; Clauss, T. R. W.; Shukla, A. K.; Moore, R. J.; Purvine, S. O.; Shi, T.; Qian, W.; Liu, T.; Baker, E. S.; Smith, R. D. Simultaneous Proteomic Discovery and Targeted Monitoring Using Liquid Chromatography, Ion Mobility Spectrometry, and Mass Spectrometry. *Mol Cell Proteomics* **2016**, *15* (12), 3694–3705. <https://doi.org/10.1074/mcp.M116.061143>.
- (211) Lu, W.; Bennett, B. D.; Rabinowitz, J. D. Analytical Strategies for LC–MS–Based Targeted Metabolomics. *Journal of Chromatography B* **2008**, *871* (2), 236–242. <https://doi.org/10.1016/j.jchromb.2008.04.031>.
- (212) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics* **2012**, *11* (11), 14. <https://doi.org/10.1074/mcp.O112.020131>.
- (213) Colgrave, M. L.; Byrne, K.; Blundell, M.; Heidelberger, S.; Lane, C. S.; Tanner, G. J.; Howitt, C. A. 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.* **2016**, *88* (18), 9127–9135. <https://doi.org/10.1021/acs.analchem.6b02108>.
- (214) Lehmann, W. D. A Timeline of Stable Isotopes and Mass Spectrometry in the Life Sciences. *Mass Spectrometry Reviews* **2017**, *36* (1), 58–85. <https://doi.org/10.1002/mas.21497>.
- (215) Freund, D. M.; Hegeman, A. D. Recent Advances in Stable Isotope-Enabled Mass Spectrometry-Based Plant Metabolomics. *Current Opinion in Biotechnology* **2017**, *43*, 41–48. <https://doi.org/10.1016/j.copbio.2016.08.002>.
- (216) Acién Fernández, F. G.; Fernández Sevilla, J. M.; Egorova-Zachernyuk, T. A.; Molina Grima, E. Cost-Effective Production of ¹³C, ¹⁵N Stable Isotope-Labelled Biomass from Phototrophic Microalgae for Various Biotechnological Applications. *Biomolecular Engineering* **2005**, *22* (5–6), 193–200. <https://doi.org/10.1016/j.bioeng.2005.09.002>.
- (217) Bueschl, C.; Kluger, B.; Lemmens, M.; Adam, G.; Wiesenberger, G.; Maschietto, V.; Marocco, A.; Strauss, J.; Bödi, S.; Thallinger, G. G.; Krska, R.; Schuhmacher, R. A Novel Stable Isotope Labelling Assisted Workflow for Improved Untargeted LC–HRMS Based Metabolomics Research. *Metabolomics* **2014**, *10* (4), 754–769. <https://doi.org/10.1007/s11306-013-0611-0>.
- (218) Hermann, G.; Schwaiger, M.; Volejnik, P.; Koellensperger, G. ¹³C-Labelled Yeast as Internal Standard for LC–MS/MS and LC High Resolution MS Based Amino Acid Quantification in Human Plasma. *Journal of Pharmaceutical and Biomedical Analysis* **2018**, *155*, 329–334. <https://doi.org/10.1016/j.jpba.2018.03.050>.
- (219) Rampler, E.; Coman, C.; Hermann, G.; Sickmann, A.; Ahrends, R.; Koellensperger, G. LILY-Lipidome Isotope Labeling of Yeast: In Vivo Synthesis of ¹³C Labeled Reference Lipids for Quantification by Mass Spectrometry. *Analyst* **2017**, *142* (11), 1891–1899. <https://doi.org/10.1039/C7AN00107J>.
- (220) Giavalisco, P.; Hummel, J.; Lisek, J.; Inostroza, A. C.; Catchpole, G.; Willmitzer, L. High-Resolution Direct Infusion-Based Mass Spectrometry in Combination with Whole ¹³C Metabolome Isotope Labeling Allows Unambiguous Assignment of Chemical Sum Formulas. *Anal. Chem.* **2008**, *80* (24), 9417–9425. <https://doi.org/10.1021/ac8014627>.
- (221) Heuillet, M.; Bellvert, F.; Cahoreau, E.; Letisse, F.; Millard, P.; Portais, J.-C. Methodology for the Validation of Isotopic Analyses by Mass Spectrometry in Stable-Isotope Labeling Experiments. *Anal. Chem.* **2018**, *90* (3), 1852–1860. <https://doi.org/10.1021/acs.analchem.7b03886>.
- (222) Heuillet, M.; Millard, P.; Cissé, M. Y.; Linares, L. K.; Létisse, F.; Manié, S.; Le Cam, L.; Portais, J.-C.; Bellvert, F. Simultaneous Measurement of Metabolite Concentration and Isotope Incorporation by Mass Spectrometry. *Anal. Chem.* **2020**, *92* (8), 5890–5896. <https://doi.org/10.1021/acs.analchem.9b05709>.
- (223) Cui, J.; Zhang, J.; Zhu, X.; Bai, F.; Feng, Y.; Guan, W.; Cui, Q. Separation and Quantification of Water-Soluble Cellular Metabolites in *Clostridium Thermocellum* Using Liquid Chromatography-Isotope Dilution Tandem

Références

- Mass Spectrometry. *Analytical Letters* **2013**, *46* (17), 2767–2786. <https://doi.org/10.1080/00032719.2013.811680>.
- (224) Dethloff, F.; Bueschl, C.; Heumann, H.; Schuhmacher, R.; Turck, C. W. Partially ¹³C-Labeled Mouse Tissue as Reference for LC-MS Based Untargeted Metabolomics. *Analytical Biochemistry* **2018**, *556*, 63–69. <https://doi.org/10.1016/j.ab.2018.06.023>.
- (225) Mayers, M. D.; Moon, C.; Stupp, G. S.; Su, A. I.; Wolan, D. W. Quantitative Metaproteomics and Activity-Based Probe Enrichment Reveals Significant Alterations in Protein Expression from a Mouse Model of Inflammatory Bowel Disease. *J. Proteome Res.* **2017**, *16* (2), 1014–1026. <https://doi.org/10.1021/acs.jproteome.6b00938>.
- (226) Khamis, M. M.; Adamko, D. J.; El-Aneed, A. STRATEGIES AND CHALLENGES IN METHOD DEVELOPMENT AND VALIDATION FOR THE ABSOLUTE QUANTIFICATION OF ENDOGENOUS BIOMARKER METABOLITES USING LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY. *Mass Spectrometry Reviews* **2019**. <https://doi.org/10.1002/mas.21607>.
- (227) Higashi, T.; Ogawa, S. Chemical Derivatization for Enhancing Sensitivity during LC/ESI-MS/MS Quantification of Steroids in Biological Samples: A Review. *The Journal of Steroid Biochemistry and Molecular Biology* **2016**, *162*, 57–69. <https://doi.org/10.1016/j.jsbmb.2015.10.003>.
- (228) Li, H.; Han, J.; Pan, J.; Liu, T.; Parker, C. E.; Borchers, C. H. Current Trends in Quantitative Proteomics - an Update: Current Trends in Quantitative Proteomics. *J. Mass Spectrom.* **2017**, *52* (5), 319–341. <https://doi.org/10.1002/jms.3932>.
- (229) Higashi, T.; Ogawa, S. Isotope-Coded ESI-Enhancing Derivatization Reagents for Differential Analysis, Quantification and Profiling of Metabolites in Biological Samples by LC/MS: A Review. *Journal of Pharmaceutical and Biomedical Analysis* **2016**, *130*, 181–193. <https://doi.org/10.1016/j.jpba.2016.04.033>.
- (230) Zhao, S.; Li, H.; Han, W.; Chan, W.; Li, L. Metabolomic Coverage of Chemical-Group-Submetabolome Analysis: Group Classification and Four-Channel Chemical Isotope Labeling LC-MS. *Anal. Chem.* **2019**, *91* (18), 12108–12115. <https://doi.org/10.1021/acs.analchem.9b03431>.
- (231) Zhao, S.; Li, L. Chemical Derivatization in LC-MS Based Metabolomics Study. *TrAC Trends in Analytical Chemistry* **2020**, 115988. <https://doi.org/10.1016/j.trac.2020.115988>.
- (232) Leitner, A. A Review of the Role of Chemical Modification Methods in Contemporary Mass Spectrometry-Based Proteomics Research. *Analytica Chimica Acta* **2018**, *1000*, 2–19. <https://doi.org/10.1016/j.aca.2017.08.026>.
- (233) Tortajada, A.; Duan, Y.; Sahoo, B.; Cong, F.; Toupalas, G.; Sallustrau, A.; Loreau, O.; Audisio, D.; Martin, R. Catalytic Decarboxylation/Carboxylation Platform for Accessing Isotopically Labeled Carboxylic Acids. *ACS Catal.* **2019**, *9* (7), 5897–5901. <https://doi.org/10.1021/acscatal.9b01921>.
- (234) Zheng, S.-J.; Zheng, J.; Xiong, C.-F.; Xiao, H.-M.; Liu, S.-J.; Feng, Y.-Q. Hydrogen–Deuterium Scrambling Based on Chemical Isotope Labeling Coupled with LC–MS: Application to Amine Metabolite Identification in Untargeted Metabolomics. *Anal. Chem.* **2020**, *92* (2), 2043–2051. <https://doi.org/10.1021/acs.analchem.9b04512>.
- (235) Hale, L. V. A.; Szymczak, N. Stereoretentive Deuteration of A-Chiral Amines with D₂O. *Journal of the American Chemical Society* **2016**, *138*, 13489–13492. <https://doi.org/10.1021/jacs.6b07879>.
- (236) Atzrodt, J.; Derdau, V. Pd- and Pt-Catalyzed H/D Exchange Methods and Their Application for Internal MS Standard Preparation from a Sanofi-Aventis Perspective. *Journal of Labelled Compounds and Radiopharmaceuticals* **2010**, *53* (11–12), 674–685. <https://doi.org/10.1002/jlcr.1818>.
- (237) Pieters, G.; Taglang, C.; Bonnefille, E.; Gutmann, T.; Puente, C.; Berthet, J.-C.; Dugave, C.; Chaudret, B.; Rousseau, B. Regioselective and Stereospecific Deuteration of Bioactive Aza Compounds by the Use of Ruthenium Nanoparticles. *Angewandte Chemie International Edition* **2014**, *53* (1), 230–234. <https://doi.org/10.1002/anie.201307930>.
- (238) Junk, T.; Catallo, W. J. Hydrogen Isotope Exchange Reactions Involving C–H (D, T) Bonds. *Chem. Soc. Rev.* **1997**, *26* (5), 401–406. <https://doi.org/10.1039/CS9972600401>.
- (239) Lee, J. H.; Yoo, K. S.; Park, C. P.; Olsen, J. M.; Sakaguchi, S.; Prakash, S.; Mathew, T.; Jung, K. W. Air/Water-Stable Tridentate NHC-PdII Complex; Catalytic C–H Activation of Hydrocarbons via H/D Exchange Process in D₂O. *Advanced Synthesis & Catalysis* **2010**, *341* (4), 563–568. <https://doi.org/10.1002/adsc.200800698>.
- (240) Sajiki, H.; Ito, N.; Esaki, H.; Maesawa, T.; Maegawa, T.; Hirota, K. Aromatic Ring Favorable and Efficient H–D Exchange Reaction Catalyzed by Pt/C. *Tetrahedron Letters* **2005**, *46* (41), 6995–6998. <https://doi.org/10.1016/j.tetlet.2005.08.067>.
- (241) Fujiwara, Y.; Iwata, H.; Sawama, Y.; Monguchi, Y.; Sajiki, H. Method for Regio-, Chemo- and Stereoselective Deuterium Labeling of Sugars Based on Ruthenium-Catalyzed C–H Bond Activation. *Chem. Commun.* **2010**, *46* (27), 4977. <https://doi.org/10.1039/c0cc01197e>.

Références

- (242) Maegawa, T.; Fujiwara, Y.; Inagaki, Y.; Monguchi, Y.; Sajiki, H. A Convenient and Effective Method for the Regioselective Deuteration of Alcohols. *Adv. Synth. Catal.* **2008**, *350* (14–15), 2215–2218. <https://doi.org/10.1002/adsc.200800407>.
- (243) Taglang, C.; Martínez-Prieto, L. M.; del Rosal, I.; Maron, L.; Poteau, R.; Philippot, K.; Chaudret, B.; Perato, S.; Sam Lone, A.; Puente, C.; Dugave, C.; Rousseau, B.; Pieters, G. Enantiospecific C-H Activation Using Ruthenium Nanocatalysts. *Angewandte Chemie International Edition* **2015**, *54* (36), 10474–10477. <https://doi.org/10.1002/anie.201504554>.
- (244) Berg, T.; Strand, D. H. ¹³C Labelled Internal Standards—A Solution to Minimize Ion Suppression Effects in Liquid Chromatography–Tandem Mass Spectrometry Analyses of Drugs in Biological Samples? *Journal of Chromatography A* **2011**, *1218* (52), 9366–9374. <https://doi.org/10.1016/j.chroma.2011.10.081>.
- (245) Lee, J. W.; Weiner, R. S.; Sailstad, J. M.; Bowsher, R. R.; Knuth, D. W.; O'Brien, P. J.; Fourcroy, J. L.; Dixit, R.; Pandite, L.; Pietrusko, R. G.; Soares, H. D.; Quarmby, V.; Vesterqvist, O. L.; Potter, D. M.; Witliff, J. L.; Fritche, H. A.; O'Leary, T.; Perlee, L.; Kadam, S.; Wagner, J. A. Method Validation and Measurement of Biomarkers in Nonclinical and Clinical Samples in Drug Development: A Conference Report. *Pharm Res* **2005**, *22* (4), 499–511. <https://doi.org/10.1007/s11095-005-2495-9>.
- (246) Schatschneider, S.; Abdelrazig, S.; Safo, L.; Henstra, A. M.; Millat, T.; Kim, D.-H.; Winzer, K.; Minton, N. P.; Barrett, D. A. Quantitative Isotope-Dilution High-Resolution-Mass-Spectrometry Analysis of Multiple Intracellular Metabolites in *Clostridium Autoethanogenum* with Uniformly ¹³C-Labeled Standards Derived from *Spirulina*. *Anal. Chem.* **2018**, *90* (7), 4470–4477. <https://doi.org/10.1021/acs.analchem.7b04758>.
- (247) Matuszewski, B. K. Standard Line Slopes as a Measure of a Relative Matrix Effect in Quantitative HPLC–MS Bioanalysis. *Journal of Chromatography B* **2006**, *830* (2), 293–300. <https://doi.org/10.1016/j.jchromb.2005.11.009>.
- (248) Rule, G. S.; Clark, Z. D.; Yue, B.; Rockwood, A. L. Correction for Isotopic Interferences between Analyte and Internal Standard in Quantitative Mass Spectrometry by a Nonlinear Calibration Function. *Anal. Chem.* **2013**, *85* (8), 3879–3885. <https://doi.org/10.1021/ac303096w>.
- (249) Bakhtiar, R.; Majumdar, T. K. Tracking Problems and Possible Solutions in the Quantitative Determination of Small Molecule Drugs and Metabolites in Biological Fluids Using Liquid Chromatography–Mass Spectrometry. *Journal of Pharmacological and Toxicological Methods* **2007**, *55* (3), 227–243. <https://doi.org/10.1016/j.vascn.2006.10.002>.
- (250) Tang, K.; Page, J. S.; Smith, R. D. Charge Competition and the Linear Dynamic Range of Detection in Electrospray Ionization Mass Spectrometry. *J Am Soc Mass Spectrom* **2004**, *15* (10), 1416–1423. <https://doi.org/10.1016/j.jasms.2004.04.034>.
- (251) Commissioner, O. of the. U.S. Food and Drug Administration <http://www.fda.gov/home> (accessed Mar 30, 2020).
- (252) Liu, Q.; Jiang, F.; Zhu, J.; Zhong, G.; Huang, M. Development, Validation, and Application of a New Method To Correct the Nonlinearity Problem in LC-MS/MS Quantification Using Stable Isotope-Labeled Internal Standards. *Analytical Chemistry* **2019**, *91* (15), 9616–9622. <https://doi.org/10.1021/acs.analchem.9b00947>.
- (253) Meija, J.; Mester, Z. Paradigms in Isotope Dilution Mass Spectrometry for Elemental Speciation Analysis. *Analytica Chimica Acta* **2008**, *607* (2), 115–125. <https://doi.org/10.1016/j.aca.2007.11.050>.
- (254) Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS). In *Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS)*; Royal Society of Chemistry: Cambridge, 2007; pp 1–34. <https://doi.org/10.1039/9781847559302-00001>.
- (255) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72. <https://doi.org/10.1007/s11306-018-1367-3>.
- (256) Blaženović, I.; Kind, T.; Sa, M. R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B. S.; Torbašinović, H.; Lee, T.; Mehta, S. S.; Showalter, M. R.; Song, H.; Kwok, J.; Jahn, D.; Kim, J.; Fiehn, O. Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **2019**, *91* (3), 2155–2162. <https://doi.org/10.1021/acs.analchem.8b04698>.
- (257) Chen, G.; Walmsley, S.; Cheung, G. C. M.; Chen, L.; Cheng, C.-Y.; Beuerman, R. W.; Wong, T. Y.; Zhou, L.; Choi, H. Customized Consensus Spectral Library Building for Untargeted Quantitative Metabolomics Analysis with Data Independent Acquisition Mass Spectrometry and MetaboDIA Workflow. *Anal. Chem.* **2017**, *89* (9), 4897–4906. <https://doi.org/10.1021/acs.analchem.6b05006>.
- (258) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Reiter, L. High-Precision IRT Prediction in the Targeted Analysis of Data-Independent Acquisition and Its Impact on Identification and Quantitation. *PROTEOMICS* **2016**, *16*

Références

- (15–16), 2246–2256. <https://doi.org/10.1002/pmic.201500488>.
- (259) Tada, I.; Chaleckis, R.; Tsugawa, H.; Meister, I.; Zhang, P.; Lazarinis, N.; Dahlén, B.; Wheelock, C. E.; Arita, M. Correlation-Based Deconvolution (CorrDec) to Generate High-Quality MS2 Spectra from Data-Independent Acquisition in Multi-Sample Studies. *Anal. Chem.* **2020**, *acs.analchem.0c01980*. <https://doi.org/10.1021/acs.analchem.0c01980>.
- (260) HBM4EU - science and policy for a healthy future <https://www.hbm4eu.eu/> (accessed May 28, 2020).
- (261) Ganzleben, C.; Antignac, J.-P.; Barouki, R.; Castaño, A.; Fiddicke, U.; Klánová, J.; Lebret, E.; Olea, N.; Sarigiannis, D.; Schoeters, G. R.; Sepai, O.; Tolonen, H.; Kolossa-Gehring, M. Human Biomonitoring as a Tool to Support Chemicals Regulation in the European Union. *International Journal of Hygiene and Environmental Health* **2017**, *220* (2, Part A), 94–97. <https://doi.org/10.1016/j.ijheh.2017.01.007>.
- (262) Vekey, K. Internal Energy Effects in Mass Spectrometry. *Journal of Mass Spectrometry* **1996**, *31*, 445–463. <https://doi.org/Internal Energy Effects in Mass Spectrometry>.
- (263) Pourchet, M.; Debrauwer, L.; Klanova, J.; Price, E. J.; Covaci, A.; Caballero-Casero, N.; Oberacher, H.; Lamoree, M.; Damont, A.; Fenaille, F.; Vlaanderen, J.; Meijer, J.; Krauss, M.; Sarigiannis, D.; Barouki, R.; Le Bizec, B.; Antignac, J.-P. Suspect and Non-Targeted Screening of Chemicals of Emerging Concern for Human Biomonitoring, Environmental Health Studies and Support to Risk Assessment: From Promises to Challenges and Harmonisation Issues. *Environment International* **2020**, *139*, 105545. <https://doi.org/10.1016/j.envint.2020.105545>.
- (264) Oberacher, H.; Pitterl, F.; Siapi, E.; Steele, B. R.; Letzel, T.; Grosse, S.; Poschner, B.; Tagliaro, F.; Gottardo, R.; Chacko, S. A.; Josephs, J. L. On the Inter-instrument and the Inter-laboratory Transferability of a Tandem Mass Spectral Reference Library. 3. Focus on Ion Trap and Upfront CID.Pdf. *Journal of Mass Spectrometry* **2012**, *47*, 263–270. <https://doi.org/10.1002/jms.2961>.
- (265) Oberacher, H.; Sasse, M.; Antignac, J.-P.; Guitton, Y.; Debrauwer, L.; Jamin, E. L.; Schulze, T.; Krauss, M.; Covaci, A.; Caballero-Casero, N.; Rousseau, K.; Damont, A.; Fenaille, F.; Lamoree, M.; Schymanski, E. L. A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries. *Environ Sci Eur* **2020**, *32* (1), 43. <https://doi.org/10.1186/s12302-020-00314-9>.
- (266) Compounding Insights Thermo Scientific Compound Discoverer Software. Thermo Fisher Scientific 2018.
- (267) Cerrato, A.; Cannazza, G.; Capriotti, A. L.; Citti, C.; La Barbera, G.; Laganà, A.; Montone, C. M.; Piovesana, S.; Cavaliere, C. A New Software-Assisted Analytical Workflow Based on High-Resolution Mass Spectrometry for the Systematic Study of Phenolic Compounds in Complex Matrices. *Talanta* **2020**, *209*, 120573. <https://doi.org/10.1016/j.talanta.2019.120573>.
- (268) Huang, T.; Bruderer, R.; Muntel, J.; Xuan, Y.; Vitek, O.; Reiter, L. Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition. *Mol Cell Proteomics* **2020**, *19* (2), 421–430. <https://doi.org/10.1074/mcp.RA119.001705>.
- (269) Derdau, V.; Atzrodt, J.; Zimmermann, J.; Kroll, C.; Brückner, F. Hydrogen–Deuterium Exchange Reactions of Aromatic Compounds and Heterocycles by NaBD4-Activated Rhodium, Platinum and Palladium Catalysts.Pdf. *Chemistry-A European Journal* **2009**, *15*, 10397–10404. <https://doi.org/10.1002/chem.200901107>.
- (270) Alexakis, E.; Jones, J. R.; Lockley, W. J. S. One-Step Exchange-Labeling of Pyridines and Other N-Heteroaromatics Using Deuterium Gas: Catalysis by Heterogeneous Rhodium and Ruthenium Catalysts. *Tetrahedron Letters* **2006**, *47* (29), 5025–5028. <https://doi.org/10.1016/j.tetlet.2006.05.106>.
- (271) Gröll, B.; Schnürch, M.; Mihovilovic, M. D. Selective Ru(0)-Catalyzed Deuteration of Electron-Rich and Electron-Poor Nitrogen-Containing Heterocycles. *J. Org. Chem.* **2012**, *77* (9), 4432–4437. <https://doi.org/10.1021/jo300219v>.
- (272) Guy, K. A.; Shapley, J. R. H–D Exchange between N-Heterocyclic Compounds and D2O with a Pd/PVP Colloid Catalyst. *Organometallics* **2009**, *28* (14), 4020–4027. <https://doi.org/10.1021/om9001796>.
- (273) Ötvös, S.; Hsieh, C.-T.; Wu, Y.-C.; Li, J.-H.; Chang, F.-R.; Fülöp, F. Continuous-Flow Synthesis of Deuterium-Labeled Antidiabetic Chalcones: Studies towards the Selective Deuteration of the Alkynone Core. *Molecules* **2016**, *21* (3), 318. <https://doi.org/10.3390/molecules21030318>.
- (274) Modutlwa, N.; Maegawa, T.; Monguchi, Y.; Sajiki, H. Synthesis of Deuterium-labelled Drugs by Hydrogen–Deuterium (H–D) Exchange Using Heterogeneous Catalysis.Pdf. *Journal of Labelled Compounds and Radiopharmaceuticals* **2010**, *53*, 686–692. <https://doi.org/10.1002/jlcr.1848>.
- (275) Ito, N.; Watahiki, T.; Maesawa, T.; Maegawa, T.; Sajiki, H. Synergistic Effect of a Palladium-on-Carbon/Platinum-on-Carbon Mixed Catalyst in Hydrogen/Deuterium Exchange Reactions of Alkyl-Substituted Aromatic Compounds. *Adv. Synth. Catal.* **2006**, *348* (9), 1025–1028. <https://doi.org/10.1002/adsc.200606053>.
- (276) Sandberg, T. E.; Long, C. P.; Gonzalez, J. E.; Feist, A. M.; Antoniewicz, M. R.; Palsson, B. O. Evolution of E.

Références

- Coli on [U-13C]Glucose Reveals a Negligible Isotopic Influence on Metabolism and Physiology. *PLoS ONE* **2016**, *11* (3), e0151130. <https://doi.org/10.1371/journal.pone.0151130>.
- (277) Chen, Y.; Shen, G.; Zhang, R.; He, J.; Zhang, Y.; Xu, J.; Yang, W.; Chen, X.; Song, Y.; Abliz, Z. Combination of Injection Volume Calibration by Creatinine and MS Signals' Normalization to Overcome Urine Variability in LC-MS-Based Metabolomics Studies. *Anal. Chem.* **2013**, *85* (16), 7659–7665. <https://doi.org/10.1021/ac401400b>.
- (278) Warrack, B. M.; Hnatyshyn, S.; Ott, K.-H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. Normalization Strategies for Metabonomic Analysis of Urine Samples. *Journal of Chromatography B* **2009**, *877* (5–6), 547–552. <https://doi.org/10.1016/j.jchromb.2009.01.007>.
- (279) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczar, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. *Anal. Chem.* **2011**, *83* (15), 5864–5872. <https://doi.org/10.1021/ac201065j>.
- (280) Chetwynd, A. J.; Abdul-Sada, A.; Holt, S. G.; Hill, E. M. Use of a Pre-Analysis Osmolality Normalisation Method to Correct for Variable Urine Concentrations and for Improved Metabolomic Analyses. *Journal of Chromatography A* **2016**, *1431*, 103–110. <https://doi.org/10.1016/j.chroma.2015.12.056>.
- (281) Wang, L.; Xing, X.; Yang, L.; Su, X.; Rabitz, H.; Lu, W.; Rabinowitz, J. D. Peak Annotation and Verification Engine (PAVE) for Untargeted LC-MS Metabolomics. *Analytical Chemistry* **2019**, *91* (3), 1838–1846. <https://doi.org/10.1021/acs.analchem.8b03132>.

Partie C : Annexes

*Annexe 1 : Revue sur les modes
d'acquisitions en spectrométrie de masse*

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



Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

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 12ème 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-

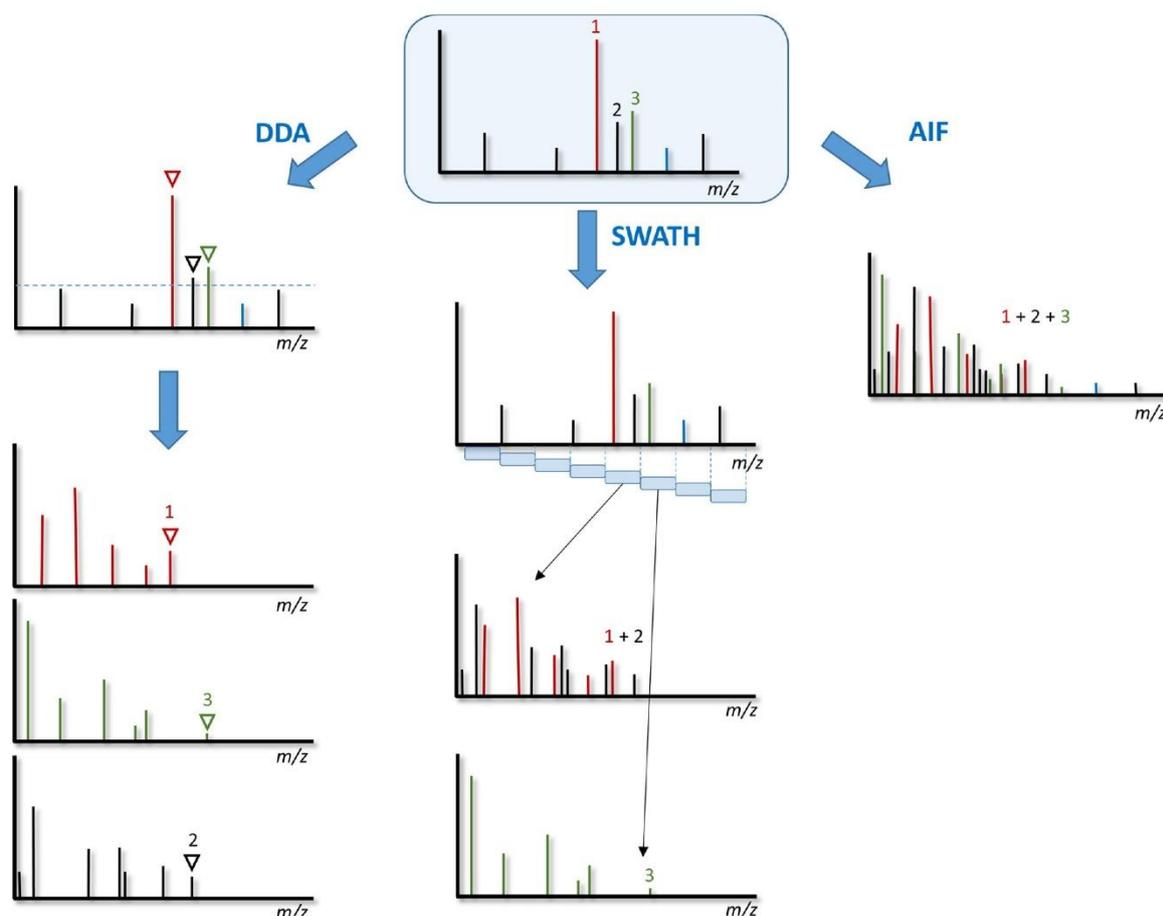


Fig. 1. Scheme illustrating the DDA, SWATH and AIF acquisition workflows.

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.

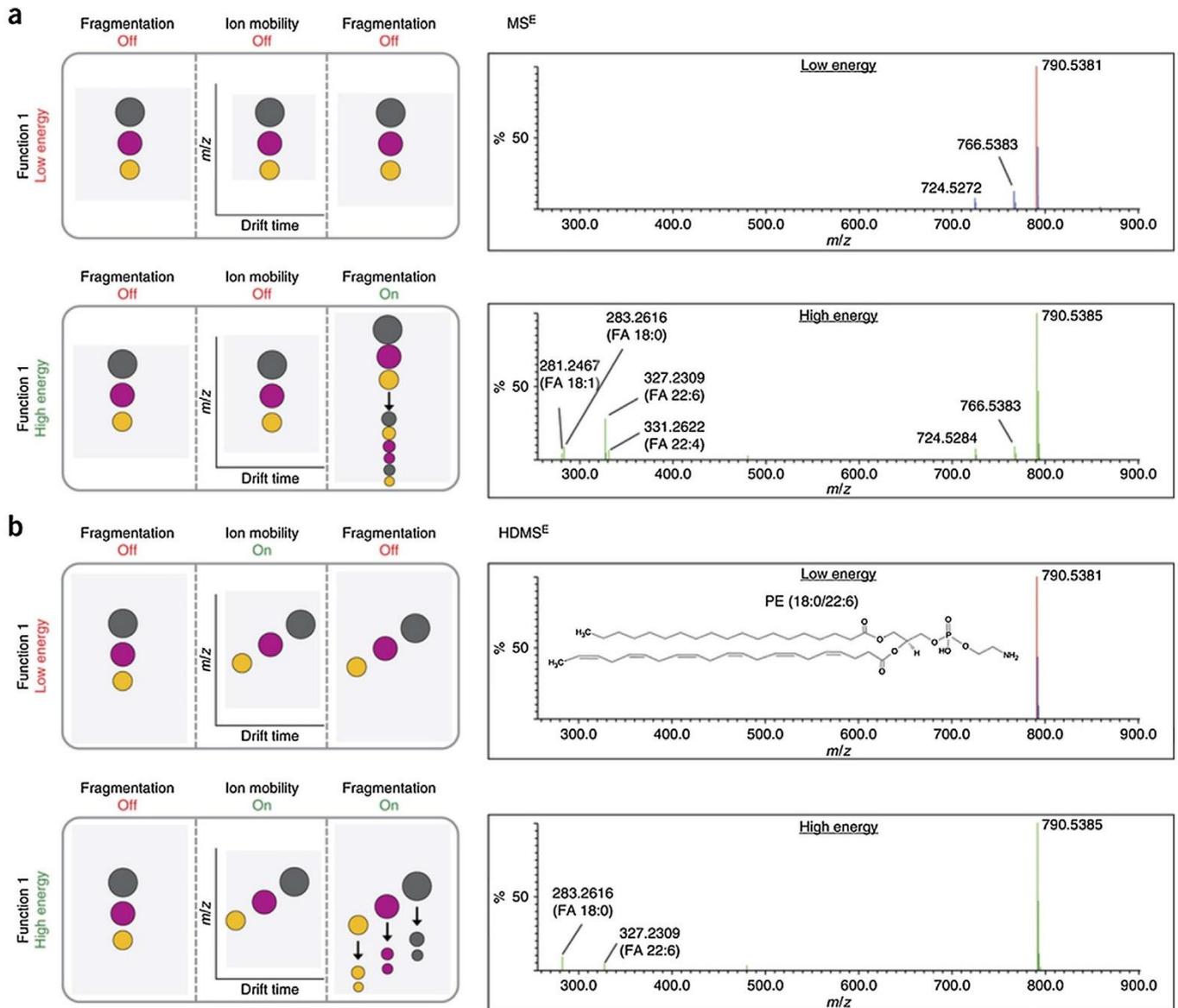


Fig. 2. Characterization of lipid species by combining ion mobility and MS^E. (a) MS^E-only data at low and high energies, providing complex mass spectra resulting from coelution of different phosphatidylethanolamine (PE) species. (b) Ion mobility efficiently separates the coeluting PE species, thus 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^E 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

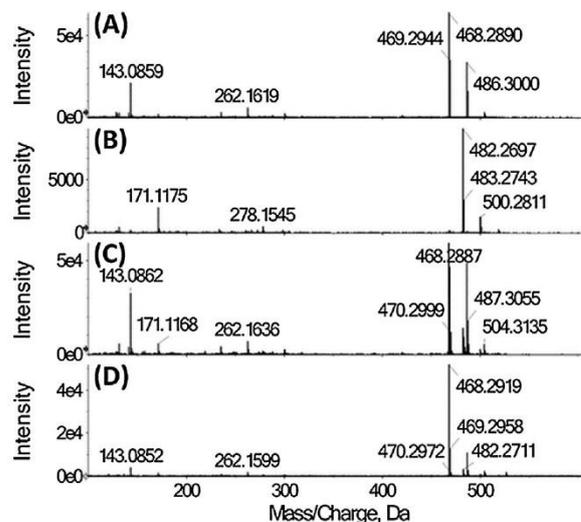


Fig. 3. Tandem mass spectra obtained from various manners (A) DDA of a terfenadine metabolite at *m/z* 504.31, (B) DDA of another terfenadine metabolite at *m/z* 518.29, (C) SWATH with a Q1 isolation window at *m/z* 499–525, and (D) AIF, high-energy TOF MS scan at *m/z* 100–1000.

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

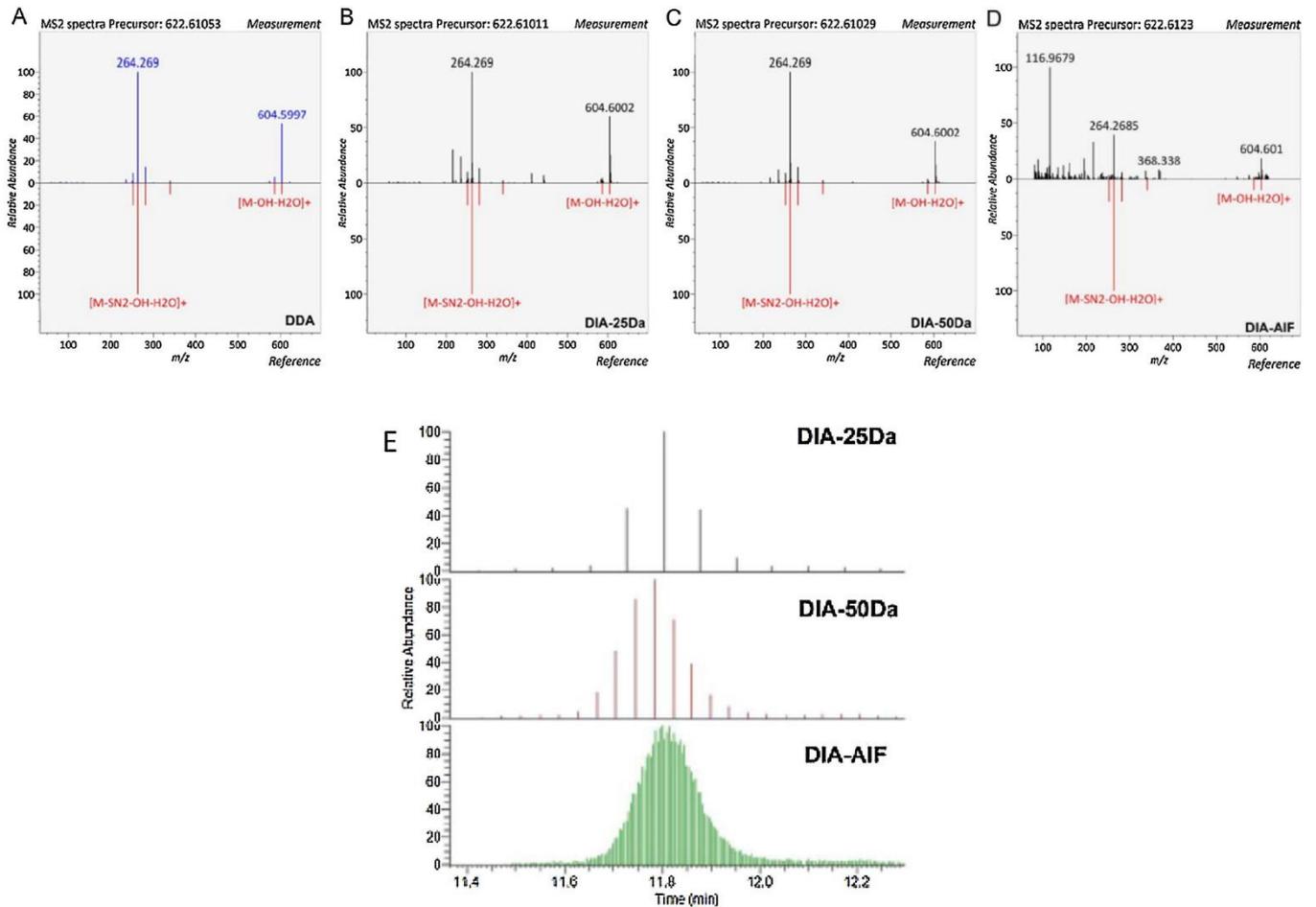


Fig. 4. Tandem mass spectra of ceramide 22:0 obtained (A) by the DDA strategy, and by the DIA strategy with (B) a 25 Da or (C) a 50 Da isolation window, and (D) using an AIF strategy. (E) Chromatographic traces obtained using the DIA–25 Da, DIA–50 Da, and AIF strategies.

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. Wohlgenuth, 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. Breiting, 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. Krokshin, 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 Hoof, 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_MassFrontier7.4-115.L.pdf.
- [21] K. Dührkop, H. Shen, M. Meusel, J. Rous, 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. Yukihira, 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 Hoof, 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. Mohny, 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>.
- [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.

- [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-¹⁵N-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.O112.020131>.
- [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.O112.019802>.
- [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.aca.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>.

*Annexe 2 : Données Supplémentaires de la
Publication sur les Modes d'Acquisitions
DDA et DIA*

« Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics »

Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics

Pierre BARBIER SAINT HILAIRE^{1,#}, Kathleen ROUSSEAU^{1,#}, Alexandre SEYER², Sylvain
DECHAUMET², Annelaure DAMONT¹, Christophe JUNOT¹, François FENAILLE^{*,1}

¹ Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), MetaboHUB, F-91191 Gif sur Yvette, France.

² MedDay Pharmaceuticals SA, 24 Rue de la Pépinière, F-75008 Paris, France.

* E-mail: francois.fenaille@cea.fr

Text S1. Optimization of a first “HCD-only” DDA method

We first optimized precursor selection filters such as intensity threshold, target mass exclusion, dynamic exclusion, apex detection, and monoisotopic precursor selection (Table S2, Supporting Information). In particular, special attention was paid to excluding frequently observed irrelevant background ions (“target mass exclusion” filter), not to waste time collecting MS/MS data on meaningless signals originating from various sources of contaminants (e.g., mobile phases, solvents, formic acid clusters). A solvent blank sample was analyzed prior to each acquisition batch to (manually) generate an exclusion list from detected background signals present at rather high intensities above 1×10^5 (to avoid exclusion of any minor metabolite signals resulting from slight but significant sample-to-blank memory or carry-over effects) with an m/z tolerance of 5 ppm. As mobile phase composition changes according to the gradient, contaminant detection was accomplished throughout the whole chromatographic run. An example of such an exclusion list is given in Table S6 (Supporting Information), including more than 50 distinct recurrently observed rather abundant contaminant ions. Target ions (produced from relevant metabolites) already selected for MS/MS were dynamically excluded after two selections for 4 seconds (“dynamic exclusion” filter), which corresponds to the typical chromatographic peak full width at half maximum. Similarly, the “apex detection” filter proved efficient for selecting an m/z signal at its highest intensity. Last, the best choice for the “monoisotopic precursor selection” (MIPS) filter to avoid selection of ^{13}C -isotopes was rather unexpectedly demonstrated to be the “peptides” setting and not the “small molecules” one (Table S2, Supporting Information). Of note, this filter should remain unselected when working on chemical compounds exhibiting particular isotopic patterns such as those incorporating chlorine or bromine atoms.

Table S1. List of the 47 compounds included in the test sample as analyzed under ESI+ conditions. Under our experimental conditions, a few compounds did not yield $[M+H]^+$ as most intense species but rather fragment ions resulting from in-source loss(es) of water.

Compound	Composition	$[M+H]^+$	Retention time (min)	Approximative "optimal NCE" (%)	3 Most Intense Fragment ions	Estimated limit of detection in solvent (ng/mL)
2-Aminophenol	C6H7NO	110.0600	1.12	40	92.0496/65.0387/82.0649	1
Nicotinic acid	C6H5NO2	124.0393	0.93	100	96.0445/112.0392/80.0497	1
Isoleucine	C6H13NO2	132.1019	1.26	15	86.0966	0.1
4-Pyridylacetic acid	C7H7NO2	138.0550	0.83	80	93.0575/120.0445/94.0652	0.1
Triethanolamine	C6H15NO3	150.1125	0.75	40	132.1019/88.0758/114.0914	0.1
p-Coumaric acid	C9H8O3	165.0546	5.77	10	147.0441/164.9300/141.9789	>100
4-Methylumbelliferone	C10H8O3	177.0546	6.60	50	103.0544/121.0649/91.0544	10
6,7-Dihydroxy-4-methylcoumarin	C10H8O4	193.0495	5.66	40	147.044/119.0492/103.0544	1
Capryloylglycine	C10H19NO3	202.1438	7.78	15	184.1332/127.1118/109.1013	10
Pantothenic acid	C9H17NO5	220.1179	2.42	20	90.0551/202.1074/184.0968	1
Flavone	C15H10O2	223.0754	9.35	80	121.0286/103.0547/129.0338	1
Dodecanedioic acid	C12H22O4	231.1591	8.34	15	213.1482/167.1429/149.1323	>100
6-Hydroxymelatonin	C13H16N2O3	249.1234	5.40	10	232.0969/190.0862/158.0602	>100
Ala-Tyr	C12H16N2O4	253.1183	1.68	10	182.0811/136.0757/165.0546	0.1
Dextrorphan	C17H23NO	258.1852	5.89	65	199.1124/133.0653/201.1277	0.1
(±)-Propranolol	C16H21NO2	260.1645	7.10	30	116.1071/183.0803/155.0855	0.1
Formononetin	C16H12O4	269.0808	8.33	50	254.0570/213.0913/253.0499	0.1
Dextromethorphan	C18H25NO	272.2009	7.20	50	215.1433/213.1279/147.0807	0.1
19-Nortestosterone	C18H26O2	275.2006	8.65	40	109.065/257.1901/105.0701	1
Testosterone	C19H28O2	289.2162	9.01	30	109.0649/271.2058/253.1953	1
Atropine	C17H23NO3	290.1751	5.60	40	124.1122/93.0701/91.0546	0.1
D-Sphingosine (-H2O)	C18H37NO2 (C18H35NO)	300.2897 (282.2791)	10.63	10	282.2789/211.2057/252.2683	100
(-)-Scopolamine	C17H21NO4	304.1543	5.00	40	138.0914/103.0546/121.0649	0.1
7α-Hydroxytestosterone	C19H28O3	305.2111	6.86	30	287.2005/269.1900/145.1011	1
Stanozolol	C21H32N2O	329.2587	9.22	80	121.1013/107.0857/105.0701	0.1
21-Deoxycortisol	C21H30O4	347.2217	7.99	30	311.2004/121.0648/147.0805	1
Prednisone	C21H26O5	359.1853	7.24	20	341.1748/147.0806/171.0807	1
Curcumin	C21H20O6	369.1333	9.39	20	177.0545/245.0809/285.1117	1
Cholic acid (-2H2O)	C24H40O5 (C24H36O3)	409.2948 (373.2737)	8.89	20	355.2627/159.1167/145.1011	10
Finasteride	C23H36N2O2	373.2850	9.26	60	305.259/317.2226/121.1013	0.1
Riboflavin	C17H20N4O6	377.1456	5.19	30	243.0874/172.0869/216.0767	1
trans-Zeatin glucoside	C16H23N5O6	382.1721	2.93	30	220.1191/202.1091/136.0618	0.1
Ochratoxin A	C20H18ClNO6	404.08954	9.39	20	239.0104/257.0209/358.0836	10
Lincomycin	C18H34N2O6S	407.2210	4.88	30	126.1278/359.2177/389.2101	0.1
Folic acid	C19H19N7O6	442.1470	4.64	10	295.0934/176.0566/313.1041	1
Glycodeoxycholate	C26H43NO5	450.3214	9.19	10	414.3001/432.31/339.2679	10
Psychosine	C24H47NO7	462.3425	9.84	20	444.3321/282.2792/264.2686	10
Glycocholic acid	C26H43NO6	466.3163	8.12	15	430.2948/412.2843/337.2524	1
Deoxycorticosterone 21-glucoside	C27H40O8	493.2796	7.57	20	331.2264/313.2162/145.0496	1
Taurodeoxycholic acid	C26H45NO6S	500.3040	8.28	10	464.2824/126.0220/482.2929	10
Taurocholic acid (-H2O)	C26H45NO7S (C26H43NO6S)	516.2990 (498.2867)	7.46	10	462.2673/480.2778/337.2522	10
D-Pantethine	C22H42N4O8S2	555.2517	5.55	10	425.1887/295.1254/147.0587	1
a-Ergocryptine	C32H41N5O5	576.3181	7.59	30	223.1233/268.1447/558.3087	1
Apigenin 7-O-neohesperidoside	C27H30O14	579.1708	6.18	15	271.0598/433.1129/129.0547	1

Naringin	C ₂₇ H ₃₂ O ₁₄	581.1865	6.24	10	273.0757/419.1331/435.1272	10
Rutin	C ₂₇ H ₃₀ O ₁₆	611.1607	5.73	10	465.1026/303.0498/449.1078	10
3'-Dephosphocoenzyme A	C ₂₁ H ₃₅ N ₇ O ₁₃ P ₂ S	688.1562	2.05	20	428.036/348.0699/261.1264	>100

Table S2. Acquisition parameters of the “reference” DDA workflow

MS1 acquisition	
Scan range	m/z 85-1000
Resolution	240,000 at m/z 200 (FWHM)
Maximum injection time	400 ms
AGC target	5×10^4
MS/MS acquisition	
Activation	HCD
Resolution	30,000 at m/z 200 (FWHM)
AGC target	5×10^4
Maximum injection time	54 ms
Isolation width	0.8 Da
Stepped collision energy	10, 30, 50%
<i>Filters</i>	
Intensity threshold	2.5×10^4
Exclusion list	Intensity $> 1 \times 10^5$, 5 ppm (see Table S6 as an example)
Dynamic exclusion	4 s, after 2 repetitions in 1 s
Apex detection	3.5 s, 65% maximal peak height
Monoisotopic precursor selection (MIPS)	Peptide

Table S3. Acquisition parameters of the DDA workflow combining low- and high-mass resolution acquisitions.

MS1 acquisition	
Scan range	<i>m/z</i> 85-1000
Resolution	120,000 at <i>m/z</i> 200 (FWHM)
AGC target	4 x 10 ⁵
MS/MS acquisition for “present” molecules (“S1 subset”)	
Activation	HCD
Detection	Ion Trap
Resolution	Normal
AGC target	1 x 10 ⁴
Maximum injection time	35 ms
Isolation width	0.8 Da
Stepped collision energy	15±10% and 45±20%
MS/MS acquisition for “absent” molecules (“S2 subset”)	
Activation	HCD and CID
Detection	Orbitrap
Resolution	15,000 at <i>m/z</i> 200 (FWHM)
AGC target	5 x 10 ⁴
Maximum injection time	22 ms
Isolation width	0.8 Da
Collision energy	HCD@30±20%, CID@22%
<i>Common filters for “S1 and S2 subsets”</i>	
Intensity threshold	2.5 x 10 ⁴
Apex detection	3.5 s, 65% maximal peak height
Monoisotopic precursor selection (MIPS)	Peptide
<i>Filters for “S1 subset”</i>	
Targeted Mass	Compounds of interest (<i>m/z</i> +/- 5 ppm, RT range)
Targeted Mass trigger	Compounds of interest (<i>m/z</i> +/-5 ppm)
Dynamic exclusion	8 s, after 3 repetitions in 3 s
<i>Filters for “S2 subset”</i>	
Exclusion list	All compounds from the “S1 subset” and all compounds with an intensity > 1 x 10 ⁵ in blank samples (5 ppm)
Dynamic exclusion	4 s, after 2 repetitions in 1.5 s

Table S4. Acquisition parameters of the DIA workflow

MS1 acquisition	
Scan range	<i>m/z</i> 85-1000
Resolution	120,000 at <i>m/z</i> 200 (FWHM)
Maximum injection time	200 ms
AGC target	5×10^4
MS/MS acquisition	
Activation	HCD
Resolution	15,000 at <i>m/z</i> 200 (FWHM)
AGC target	5×10^4
Maximum injection time	22 ms
Stepped collision energy	30±20%
<i>m/z Windows</i>	
1	98 – 132
2	129.5 – 145.5
3	144.5 – 160.5
4	159.5 – 175.5
5	174.5 – 190.5
6	189 – 211
7	209 – 262
8	259 – 301
9	300 – 400
10	399 - 602

Table S5. List of the 72 metabolites annotated in plasma NIST in C18 LC condition and ESI+ ionization

Name	Formula	Mass	Retention Time (min)
Uracil	C4H4N2O2	112.0273	1.06
Creatinine	C4H7N3O	113.0589	0.83
Proline	C5H9NO2	115.0633	0.85
Betaine	C5H11NO2	117.0790	0.82
Valine	C5H11NO2	117.0790	0.93
Threonine / D-allo-Threonine	C4H9NO3	119.0582	0.84
Nicotinamide	C6H6N2O	122.0480	0.99
Pyroglutamic-acid	C5H7NO3	129.0426	1.08
Pipecolinic-acid	C6H11NO2	129.0790	0.94
Creatine	C4H9N3O2	131.0695	0.84
Isoleucine	C6H13NO2	131.0946	1.28
Leucine	C6H13NO2	131.0946	1.38
5-Hydroxyindole	C8H7NO	133.0528	4.92
Hypoxanthine	C5H4N4O	136.0385	0.99
Trigonelline	C7H7NO2	137.0477	0.84
4-Imidazoleacrylic acid	C6H6N2O2	138.0429	0.93
Methylimidazoleacetic-acid	C6H8N2O2	140.0586	0.86
Stachydrine	C7H13NO2	143.0946	0.85
4-Guanidinobutyric-acid	C5H11N3O2	145.0851	0.93
Glutamine	C5H10N2O3	146.0691	0.81
Lysine	C6H14N2O2	146.1055	0.72
L-Glutamic-acid	C5H9NO4	147.0532	0.82
Methionine	C5H11NO2S	149.0511	1.04
Acetaminophen-(4-Acetamidophenol)	C8H9NO2	151.0633	2.88
Histidine	C6H9N3O2	155.0695	0.76
Carnitine	C7H15NO3	161.1052	0.82
1-methyl-guanine / 7-methylguanine	C6H7N5O	165.0651	0.97
Phenylalanine	C9H11NO2	165.0790	2.10
7-Methylxanthine	C6H6N4O2	166.0491	1.45
3-Methylxanthine	C6H6N4O2	166.0491	1.60
1-Methylxanthine	C6H6N4O2	166.0491	1.74
Uric-acid	C5H4N4O3	168.0283	0.94
1-Methylhistidine	C7H11N3O2	169.0851	0.78
Arginine	C6H14N4O2	174.1117	0.77
L-Citrulline	C6H13N3O3	175.0957	0.81
Cotinine	C10H12N2O	176.0950	1.08
Hippuric-acid	C9H9NO3	179.0582	4.93
paraxanthine / Theophylline	C7H8N4O2	180.0647	3.95
Theobromine	C7H8N4O2	180.0647	2.52
Tyrosine	C9H11NO3	181.0739	1.16
3-Amino-3-(4-hydroxyphenyl)propanoic	C9H11NO3	181.0739	1.17
1-Methyluric-acid	C6H6N4O3	182.0440	1.46
4-Pyridoxic-acid	C8H9NO4	183.0532	1.37
N-acetyl-L-glutamine	C7H12N2O4	188.0797	0.94
N6-Acetyl-L-lysine	C8H16N2O3	188.1161	0.90
N6,N6,N6-Trimethyl-L-lysine	C9H20N2O2	188.1525	0.76
Trans-3-Hydroxy-cotinine	C10H12N2O2	192.0899	0.91
(S)-Cotinine-N-oxide	C10H12N2O2	192.0899	1.52
Caffeine	C8H10N4O2	194.0804	4.89
Acetyl-L-carnitin	C9H17NO4	203.1158	1.01
Tryptophan	C11H12N2O2	204.0899	4.24
Panthenol	C9H19NO4	205.1314	2.48

L-Kynurenine	C10H12N2O3	208.0848	2.11
N-alpha-acetyl-L-arginine	C8H16N4O3	216.1222	0.90
Propionylcarnitine	C10H19NO4	217.1314	1.54
Pantothenic-acid	C9H17NO5	219.1107	2.45
(R)-Butyryl-carnitine	C11H21NO4	231.1471	3.48
L-Cystine	C6H12N2O4S2	240.0239	0.84
L-a-Glycerophosphorylcholine	C8H20NO6P	257.1028	0.77
Hexanoylcarnitine	C13H25NO4	259.1784	6.19
Phenylacetyl-L-glutamine	C13H16N2O4	264.1110	5.07
N-Acetyl-L-carnosine	C11H16N4O4	268.1172	0.88
1-Methyladenosine	C11H15N5O4	281.1124	0.92
Octanoylcarnitine	C15H29NO4	287.2097	7.63
5-Deoxy-5-(methylthio)adenosine	C11H15N5O3S	297.0896	4.16
Decanoylcarnitine	C17H33NO4	315.2410	8.79
Acetaminophen-glucuronide	C14H17NO8	327.0954	1.56
Sphingosine-1-phosphate	C18H38NO5P	379.2488	10.24
glycochenodeoxycholic-acid	C26H43NO5	449.3141	9.12
glycodeoxycholate	C26H43NO5	449.3141	9.29
Glycocholic-acid	C26H43NO6	465.3090	8.21
Stercobilin	C33H46N4O6	594.3417	7.16

Table S6. Example of an exclusion list generated from contaminant background ions

<i>m/z</i>	Start time (min)	End time (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

Table S7. Dot product scores obtained for the standard mixture by using the HCD-only DDA workflow (mean values of three replicate measurements).

Metabolite	Dot product
2-Aminophenol	999
Nicotinic acid	480
Isoleucine	687
4-Pyridylacetic acid	569
Triethanolamine	645
4-Methylumbelliferone	508
6,7-Dihydroxy-4-methylcoumarin	581
Capryloylglycine	296
Pantothenic acid	752
Flavone	727
Ala-Tyr	810
Dextrorphan	489
(±)-Propranolol	674
Formononetin	495
Dextromethorphan	551
19-Nortestosterone	685
Testosterone	646
Atropine	578
D-Sphingosine	720
(-)-Scopolamine	693
7 α -Hydroxytestosterone	676
Stanozolol	939
21-Deoxycortisol	761
Prednisone	839
Curcumin	773
Cholic acid	630
Finasteride	544
Riboflavin	620
trans-Zeatin glucoside	758
Ochratoxin A	779
Lincomycin	805
Folic acid	672
Glycodeoxycholate	729
Psychosine	852
Glycocholic acid	770
Deoxycorticosterone 21-glucoside	804
Taurodeoxycholic acid	733
Taurocholic acid (-H ₂ O)	718
D-Pantethine	617
α -Ergocryptine	791
Apigenin 7-O-neohesperidoside	874
Naringin	699
Rutin	807

Table S8. Dot product scores obtained by DDA and DIA for a set of 34 metabolites (mean of 3 replicates per extraction).

Metabolite	DDA		DIA	
	Mean Dot Product Extraction 1	Mean Dot Product Extraction 2	Mean Dot Product Extraction 1	Mean Dot Product Extraction 2
Proline	460	604	626	605
Betaine	750	641	656	657
Pyroglutamic-acid	579	588	465	478
Creatine	807	810	236	247
Leucine/Isoleucine	747	749	621	708
Hypoxanthine	596	615	570	552
Trigonelline	725	742	30	35
Stachydrine	868	870	570	596
Glutamine	741	769	467	467
Lysine	841	852	415	447
Methionine	861	931	585	630
Acetaminophen-(4-Acetamidophenol)	708	714	708	737
Histidine	783	748	51	51
Carnitine	489	495	643	694
Phenylalanine	656	662	534	535
Uric-acid	863	824	765	805
1-Methylhistidine	542	585	133	97
Arginine	778	807	688	685
Cotinine	810	805	831	830
Paraxanthine / Theophylline	755	690	764	763
Theobromine	762	693	947	948
Tyrosine	831	838	709	711
Trans-3-Hydroxy-cotinine	840	789	672	703
Tryptophan	835	841	517	515
Propionylcarnitine	853	848	919	913
Pantothenic-acid	505	479	338	402
(R)-Butyryl-carnitine	598	483	855	689
Hexanoylcarnitine	487	477	741	743
Phenylacetyl-L-glutamine	900	898	898	899
Decanoylcarnitine	566	660	873	870
Acetaminophen-glucuronide	557	590	452	449
Sphingosine-1-phosphate	764	721	193	163
Glycochenodeoxycholic-acid	822	742	865	866
Glycocholic-acid	664	513	431	376

Table S9. Comparison of performance characteristics of MS-Only, DDA-MS, DIA-MS, and DIA-MS/MS workflows

Metabolite	MS-Only					DDA-MS				
	Estimated LOD (ng/mL) ^a	Dynamic range (ng/mL)	r2	Accuracy		Estimated LOD (ng/mL) ^a	Dynamic range (ng/mL)	r2	Accuracy	
				@0.25 ng/mL	@3 ng/mL				@0.25 ng/mL	@3 ng/mL
Scopolamine	0.05	0.05-10	0.99	102	96	0.05	0.05-10	0.99	102	98
Flavone	0.25	0.25-10	0.95	108	93	0.1	0.1-3	0.93	81	102
Formononetin	0.5	0.5-10	0.96	103	85	0.25	0.25-10	0.95	107	86
Finasteride	0.1	0.1-3	0.92	83	104	0.1	0.1-3	0.92	86	103
Propranolol	0.05	0.05-10	0.98	98	99	0.05	0.05-10	0.99	99	99
trans-ZeatinGlucoside	0.25	0.25-10	0.97	103	94	0.25	0.25-10	0.97	105	95
Dextrorphan	0.05	0.05-10	0.98	97	99	0.05	0.05-10	0.98	97	98
Lincomycin	0.25	0.25-10	0.97	105	96	0.05	0.05-10	0.98	96	100
alpha-Ergocryptine	0.5	0.5-10	0.98	103	94	0.25	0.25-10	0.98	100	94

Metabolite	DIA-MS					DIA-MS/MS				
	Estimated LOD (ng/mL) ^a	Dynamic range (ng/mL)	r2	Accuracy		Estimated LOD (ng/mL) ^a	Dynamic range (ng/mL)	r2	Accuracy	
				@0.25 ng/mL	@3 ng/mL				@0.25 ng/mL	@3 ng/mL
Scopolamine	0.05	0.05-10	0.98	100	97	0.25	0.25-10	0.98	102	97
Flavone	0.1	0.1-3	0.92	84	102	0.25	0.25-10	0.96	107	94
Formononetin	3	3-10	NA	NA	NA	3	3-10	NA	NA	NA
Finasteride	0.1	0.1-3	0.92	85	102	0.25	0.25-10	0.96	109	98
Propranolol	0.05	0.05-10	0.98	99	98	0.1	0.1-10	0.98	100	99
trans-ZeatinGlucoside	0.25	0.25-10	0.97	103	94	3	3-10	NA	NA	NA
Dextrorphan	0.05	0.05-10	0.98	98	99	0.1	0.1-10	0.98	101	98
Lincomycin	0.25	0.25-10	0.97	105	95	3	3-10	NA	NA	NA
alpha-Ergocryptine	0.25	0.25-10	0.98	101	95	0.5	0.5-10	0.97	100	99

^a Determined as the lowest calibration point with a CV<20%

NA: Not Applicable

Highlighted with a yellow background are the accuracies measured at 0.5ng/mL

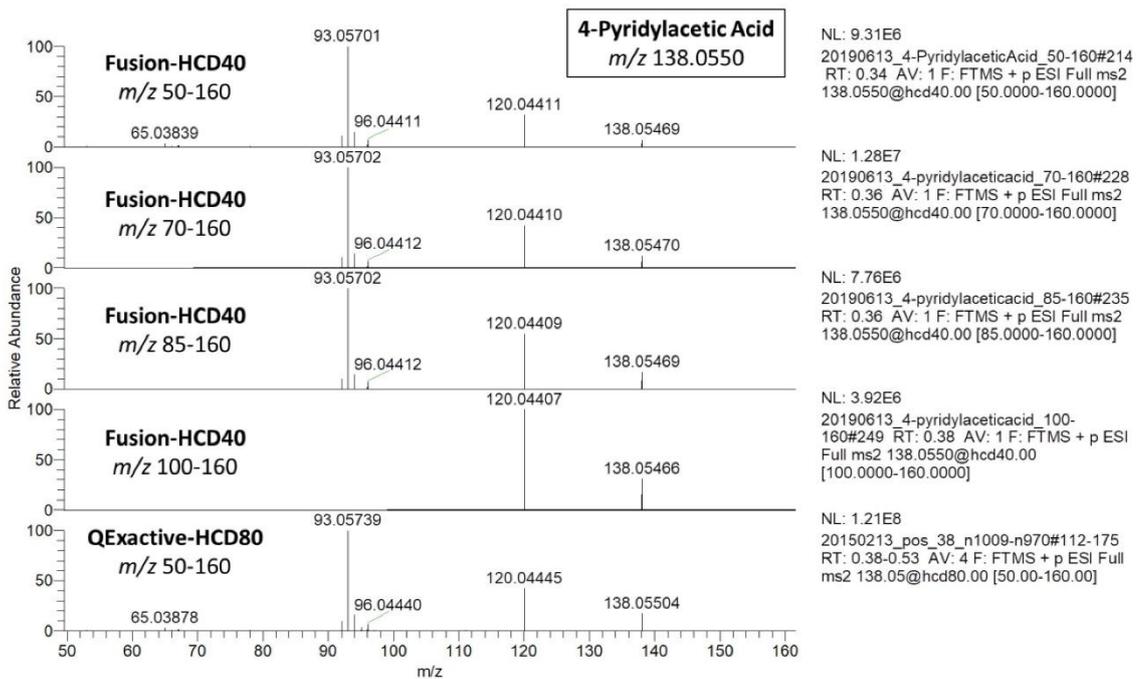


Figure S1. Comparison of HCD spectra acquired for 4-pyridylacetic acid on the Orbitrap FusionTM or on a Q-ExactiveTM using variable *m/z* windows for fragment ions.

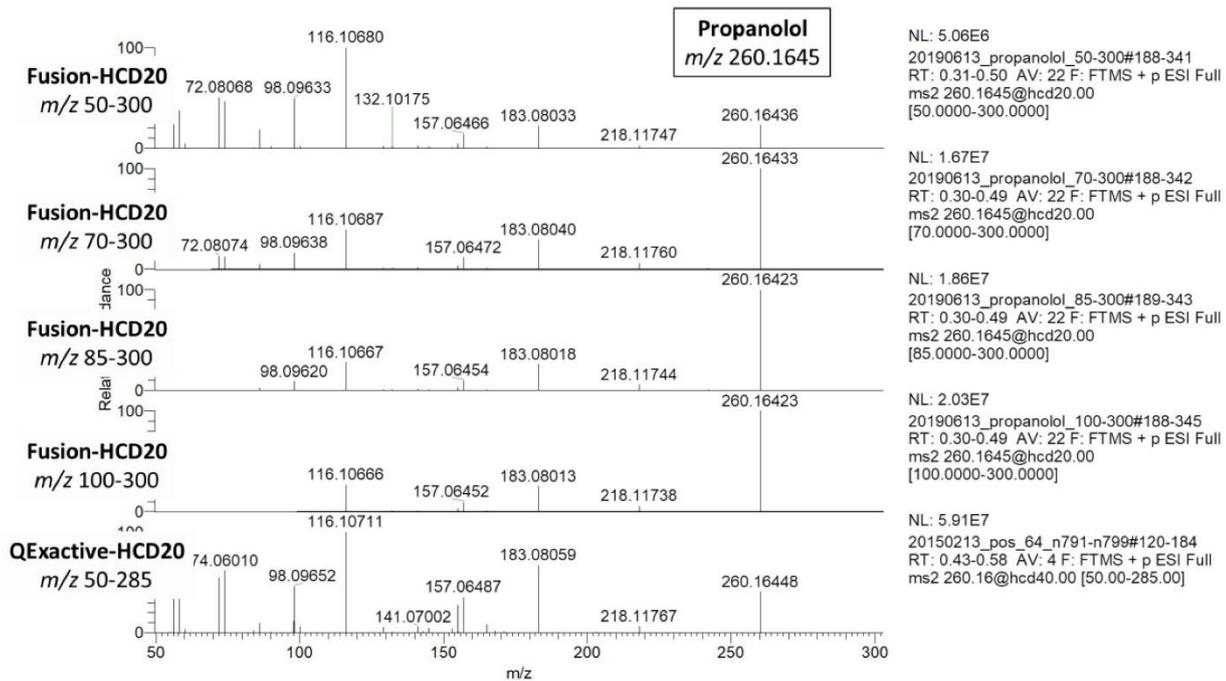


Figure S2. Comparison of HCD spectra acquired for propranolol on the Orbitrap Fusion™ or on a Q-Exactive™ using variable m/z windows for fragment ions.

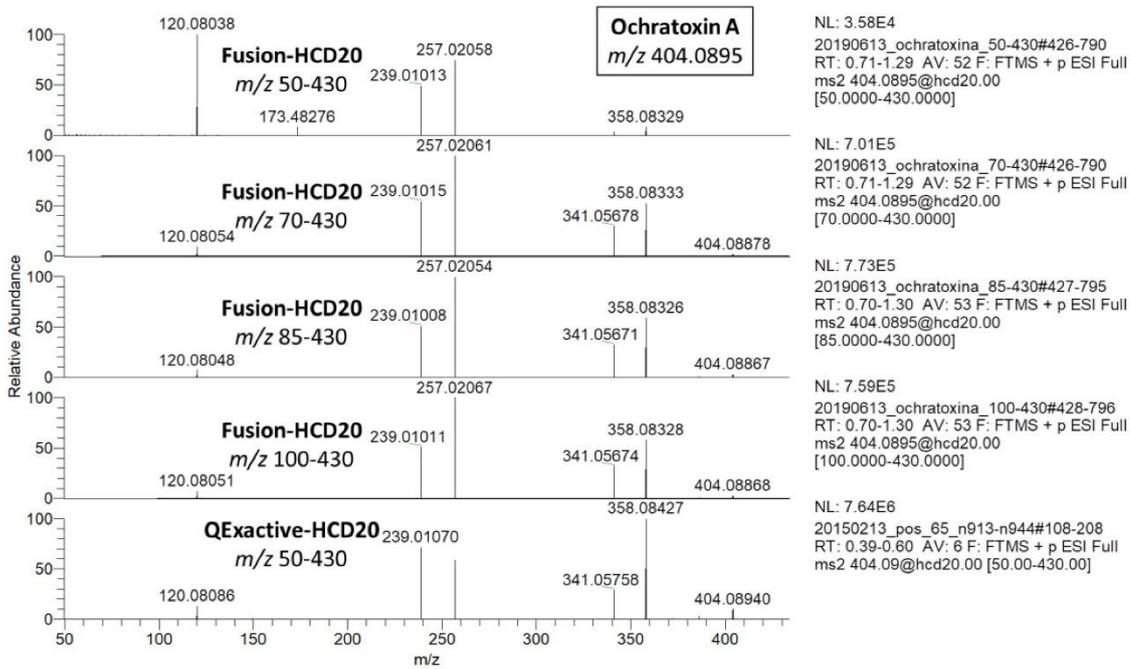


Figure S3. Comparison of HCD spectra acquired for ochratoxin A on the Orbitrap FusionTM or on a Q-ExactiveTM using variable *m/z* windows for fragment ions.

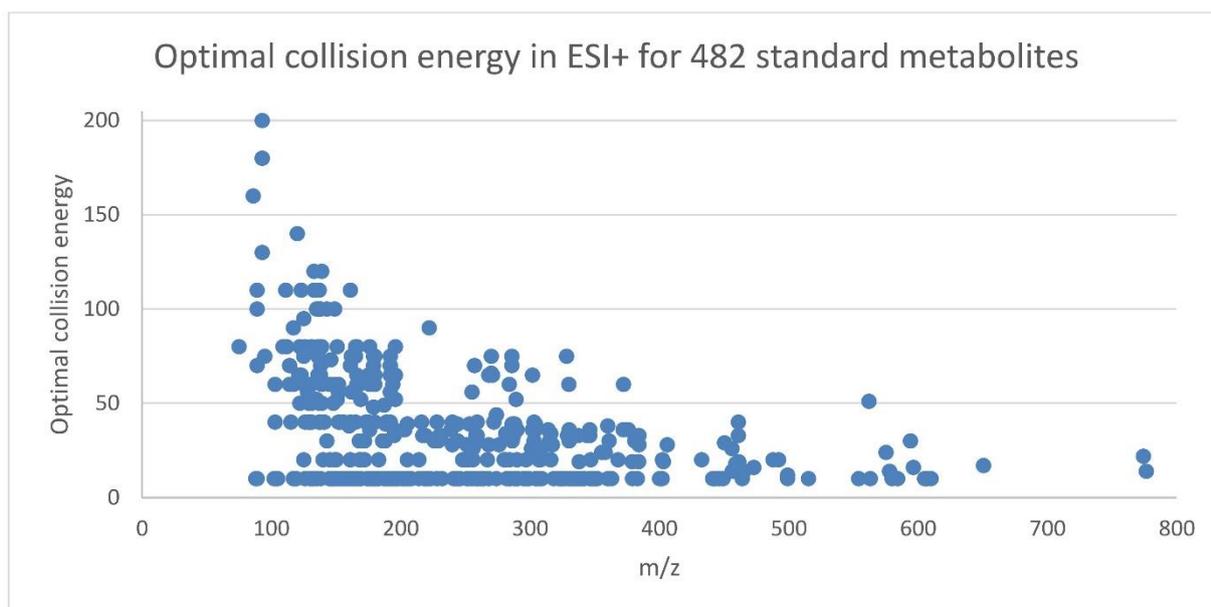


Figure S4. Distribution of optimal collision energy for 482 standard metabolites.

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 15 and 45%.

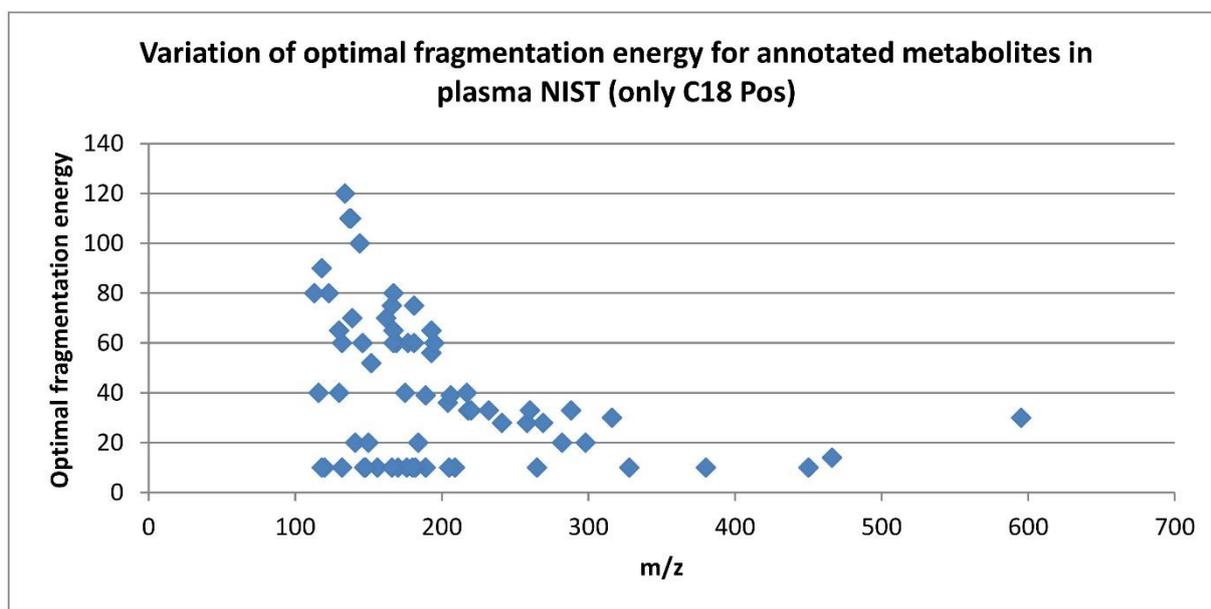


Figure S5. Distribution of optimal collision energy for 72 standard metabolites identified in NIST plasma using C₁₈ UHPLC system coupled to (ESI+)-MS/MS.

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 15 and 45%.

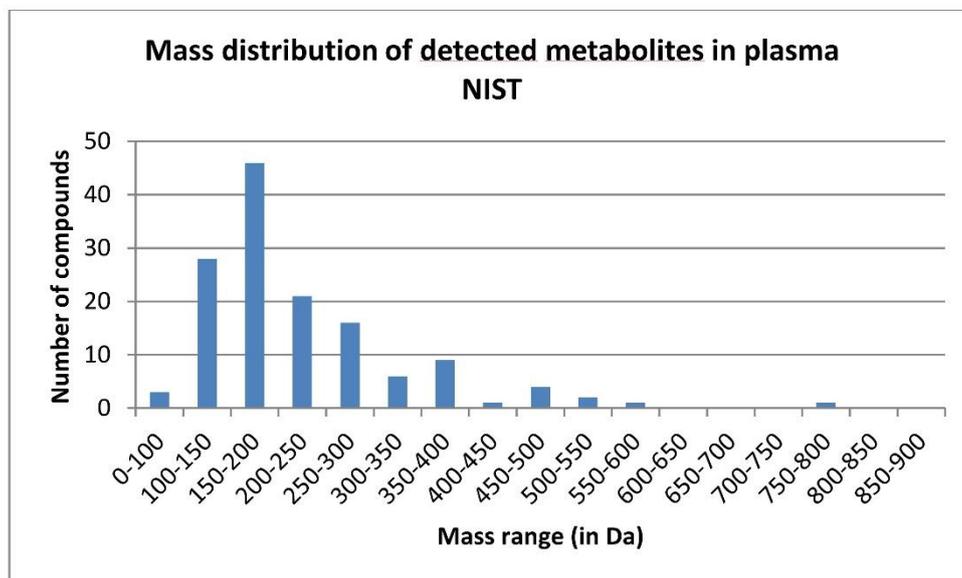
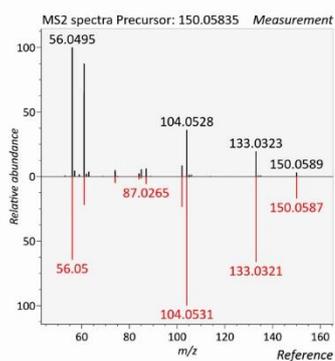
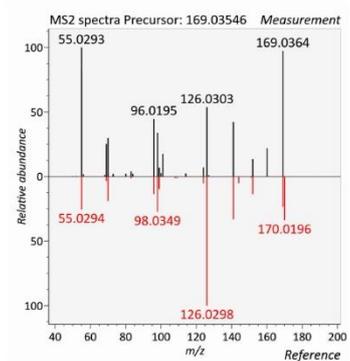


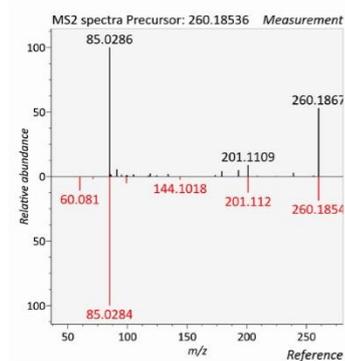
Figure S6. Mass distribution of metabolites identified in plasma.



Methionine (DP: 529)



Uric acid (DP: 690)



Hexanoylcarnitine (DP: 764)

Figure S7. Head-to-tail comparison of evaluated versus reference MS/MS spectra. Evaluated MS/MS spectra were obtained using the DIA acquisition workflow.

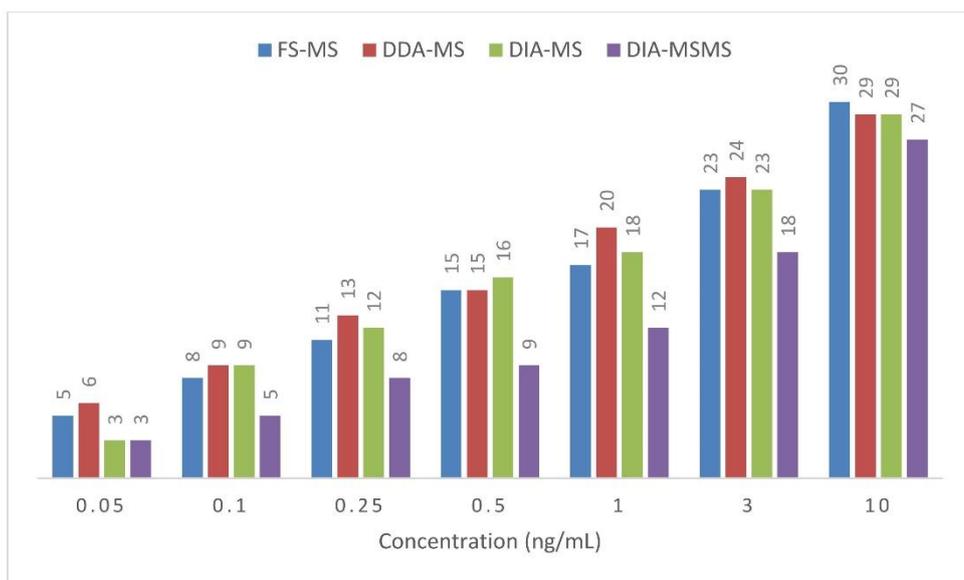


Figure S8. Number of metabolites reproducibly detected (with CV < 30%) using full-scan only (blue), DDA (red), DIA-MS (green) and DIA MS/MS workflows (purple). Each bar is the result of 6 independent measurements. Regarding the four molecules from the spiked test mixture also present endogenously in plasma, only signals exceeding more than 5 times the endogenous ones were considered.

Annexe 3 : Publication HBM4EU sur la proposition de critères qualités pour la confection d'une base de données spectrale MS/MS

« A European proposal for quality control and quality assurance of tandem mass spectral libraries »

RESEARCH

Open Access



A European proposal for quality control and quality assurance of tandem mass spectral libraries

Herbert Oberacher^{1*}, Michael Sasse², Jean-Philippe Antignac³, Yann Guitton³, Laurent Debrauwer^{4,5}, Emilien L. Jamin^{4,5}, Tobias Schulze⁶, Martin Krauss⁶, Adrian Covaci⁷, Noelia Caballero-Casero⁷, Kathleen Rousseau⁸, Annelaure Damont⁸, François Fenaille⁸, Marja Lamoree⁹ and Emma L. Schymanski^{10*} 

Abstract

Background: High resolution mass spectrometry (HRMS) is being used increasingly in the context of suspect and non-targeted screening for the identification of bioorganic molecules. There is correspondingly increasing awareness that higher confidence identification will require a systematic, group effort to increase the fraction of compounds with tandem mass spectra available in central, publicly available resources. While typical suspect screening efforts will only result in tentative annotations with a moderate level of confidence, library spectral matches will yield higher confidence or even full confirmation of the identity if the reference standards are available.

Results: This article first explores representative percent coverage of measured tandem mass spectra in selected major environmental suspect databases of interest in the context of human biomonitoring, demonstrating the current extensive gap between the number of potential substances of interest (up to hundreds of thousands) and measured spectra (0.57–3.6% of the total chemicals have spectral information available). Furthermore, certain datasets are benchmarked, based on previous efforts, to show the extent to which acquired experimental data were comparable between laboratories, even with HRMS instruments based on different technologies (i.e., quadrupole–quadrupole-time of flight versus ion trap/quadrupole-Orbitrap). Instruments and settings that are less comparable are also revealed, primarily linear ion trap instruments, which show distinctly lower comparability.

Conclusions: Based on these efforts, harmonization guidelines for the acquisition and processing of tandem mass spectrometry data are proposed to enable European (and ideally worldwide) laboratories to contribute to common resources, without requiring extensive changes to their current *in house* methods.

Keywords: Compound identification, Tandem mass spectral library, Liquid chromatography mass spectrometry, Non-targeted analysis, Quality control, Quality assurance, Exposomics, Environmental science, Human biomonitoring, HRMS

*Correspondence: herbert.oberacher@i-med.ac.at; emma.schymanski@uni.lu

¹ Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Muellerstrasse 44, 6020 Innsbruck, Austria

¹⁰ Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg
Full list of author information is available at the end of the article

Background

Detection, annotation and identification

The goal of suspect and non-targeted analysis is to provide extensive qualitative information on the chemical composition of a sample. These analytical approaches are possible because of recent technologies and instrumentation, which are capable of generating large amounts of chemical information from low sample amounts. In particular, high-resolution mass spectrometry (HRMS) is one of these innovative technologies for large-scale and high-throughput profiling of complex samples [1]. However, assigning chemical identities to a set of mass spectrometric signals is not trivial and requires strongly consolidated data processing as well as appropriate quality assurance (QA) and quality control (QC) procedures. Particularly the confidence of the assigned chemical identities (annotations) is a crucial issue [2]. As the aim is typically to confirm the presence of as many compounds as possible, adequate strategies have to be used in reporting the results, both for research and in the context of regulatory use. While QA/QC aspects are well established in the field of conventional targeted methods, these are currently less developed for non-targeted analyses, although these are discussed actively in both the metabolomics and environmental communities (e.g., [2, 3]).

This article explores possible strategies for QA/QC of tandem mass spectral libraries and databases, especially suitable for the annotation of chemicals of emerging concern in the context of human biomonitoring (specifically within the Human Biomonitoring for Europe project, HBM4EU, <http://www.hbm4eu.eu>) and environmental monitoring (initiatives originating from the NORMAN Network, <http://www.norman-network.net>). It reflects on existing initiatives and approaches that can be used to assign well-defined confidence levels to annotated biomarkers (either of exposure or effect), to check the quality of existing sets of tandem mass spectra data, and for acquiring new experimental data, as well as any gaps that may need to be addressed.

For this article, a few definitions are clarified here, with further definitions given in the Glossary below. *Detection* refers to the collection of compound-specific data by instrumental analysis. For chromatography coupled to mass spectrometry, this collected data may include retention times, mass-to-charge ratios (m/z) of molecular ions, adducts and possibly fragment ions as well as the presence and relative abundances of fragment ions and isotopologues. *Annotation* is the act of linking a detected mass spectrometric feature with a chemical identity, taking into account the detected chromatographic and spectrometric characteristics. *Identification* is the process of proving or verifying that the annotated compound is indeed the proposed chemical (i.e., the annotation can be confirmed).

Annotation is generally performed using analytical evidence of the measured dataset alone [1–3], along with additional supporting evidence (e.g., experimental context [2] and metadata [4–6]). In contrast, *identification* is generally accomplished by comparing measured data sets (e.g., using reference standards), where one set of features is obtained from the analysis of an unknown compound; the other from a reference standard of known identity. In this context, defining objective metrics for confirmation of identity is a challenging task. Ideal metrics should minimize the number of false positive and false negative identifications (see below).

Sufficient analytical data must be available to enable definitive identification. Like a fingerprint, this evidence should be a unique set of information capable of excluding all other chemical entities from consideration (unequivocal identification). In the case of a mass spectrometry (MS)-based characterization, such a chemical fingerprint can partly be created in silico (e.g., m/z -values of molecular ions, relative abundances of isotopologues given the molecular formula and, to some extent, fragment ions) and/or by analyzing reference standards (e.g., m/z -values of fragment ions, retention times in chromatographic systems). These “fingerprints” (often specific for distinct instrumental settings) are generally stored in databases. While predictive (in silico) methods exist for both fragment (e.g., [7, 8]) and retention time information (e.g., [9–12]), these are not yet sufficiently accurate for unequivocal identification, although these are constantly improving in accuracy. Thus, complete “standards-free” identification is not yet possible for HRMS, although it is becoming increasingly possible to get reasonable annotations using “standards-free” approaches.

From an analytical point of view and depending on the available information for annotation, chemicals can be divided into three categories, which we define as follows for this article (see Fig. 1).

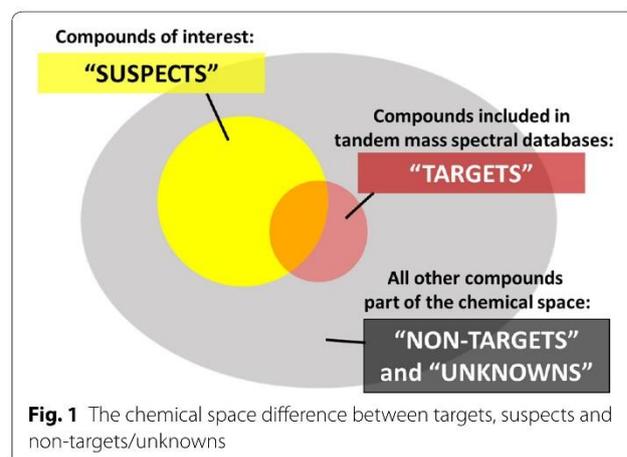


Fig. 1 The chemical space difference between targets, suspects and non-targets/unknowns

Targets are compounds (“knowns”) that are preselected for analysis in a sample and for which full mass spectrometric reference data, including MS/MS fragmentation and retention time, is available for annotation. The reference data are usually acquired with certified reference standards in house; the reference mass spectra and, in some cases retention times (depending on the database), are stored in mass spectral databases.

Suspects are known compounds (“known unknowns” [13]) that are expected (“suspected”) to be present in a sample, but for which either no reference standard (in house) or incomplete mass spectrometric reference data are available, such that unequivocal annotation is not always directly possible. In suspect screening, this could be either missing or measured with alternative methods, or predicted with computational tools, and as such not of sufficient accuracy in many cases to allow reliable annotation—although it may provide supporting evidence. For instance, while MS information can be computed reliably from the structure, the *in silico* prediction of MS/MS fragmentation and retention time information still needs to be improved.

Both targets and suspects represent subsets of the entire chemical space in a sample (Fig. 1). Suspects can be “converted” into targets by collecting comprehensive mass spectrometric reference data that enables unequivocal identification of the suspect compound (usually reliant on the availability of reference standard compounds). The remaining signals in the sample are generally termed *non-targets* or *unknowns*—for which no target or suspect identity can be assigned readily. These require full elucidation and are beyond the scope of this article.

Inspired by European regulatory documents and a classification system originally proposed by Sumner et al. [14] for metabolomics, a classification system tailored for HRMS primarily for the environmental context was proposed by Schymanski et al. [2] in 2014. Level 1 (confirmed structure) described identification that has been verified via the appropriate measurement of a reference standard with MS, MS/MS and retention time matching, matching the definition of targets above. A “probable structure” (Level 2) is obtained by unambiguous matching literature or library data (Level 2a) or via diagnostic evidence (Level 2b), where the diagnostic evidence must clearly rule out all other candidate structures. “Tentative candidates” (Level 3) describe the case where the available data provide evidence for possible or likely structure(s), but insufficient information exists for one exact structure only (e.g., positional isomers). Level 4 or 5 identifications are typically “unknowns”, where only the molecular formula (Level 4) or exact mass (Level 5) are known. Since initiation of the level system in 2014, several practical cases have evolved within each level and

these are, for instance, now encoded into the mass spectral processing software RMassBank [15].

Tandem mass spectral databases: current status

Tandem mass spectral databases are indispensable tools for compound annotation in non-targeted HRMS workflows based on soft-ionization mass spectrometry (typically liquid chromatography (LC)-HRMS) and good matches can yield Level 2a annotations in many cases. Several reviews are available describing the development and application of tandem mass spectral databases [3, 16–22]. Typically, a tandem mass spectral database represents an organized collection of tandem mass spectral data within a management system. The database management system enables the user, or other applications, to interact with data within the database itself. Tandem mass spectral databases are acquired by the analysis of reference standards. Since a fragmentation spectrum can look different depending on the excitation process (e.g., resonant vs. non-resonant) as well as the collision energy applied to the parent ion, state-of-the-art databases include sets of compound-specific spectra that were acquired by applying different collisions energy settings, as well as different instruments [23, 24]. Fragmentation is typically accomplished by collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD). Usually, the spectral information is processed prior to storage in a library. Curation efforts may include manual inspection of mass spectra by experienced mass spectrometrists, noise and artifact removal, recalibration of spectra and peak annotations, as well as inter-library comparisons [15, 23, 25–27]. In some databases, such as the Human Metabolome Database (HMDB [28]) and MassBank of North America (MoNA, [29]), experimental data are now complemented with *in silico*-generated spectra.

In 2016, the overlap of compounds with tandem mass spectra from authentic reference standards in most public and commercial databases was evaluated by Vinaixa et al. [16]. A total of 27,622 unique compounds were present across all databases. Among the 7127 compounds in the four open databases HMDB 3.0 [30], MassBank [24], the Global Natural Product Social Molecular Networking library (GNPS) [31], and the RIKEN MSⁿ spectral database for phytochemicals (ReSpect) [32], only 18 compounds (<1%) had at that time at least one form of spectral data in all databases. When comparing all combined open databases versus four commercial ones, only 225 compounds out of 27,622 (<1%) had at least one form of spectral data in all databases. The ratio of compounds in each database with any type of spectral data in two or more databases was generally >50%, with the exception of METLIN and GNPS, which only overlapped

approximately 35% with other databases in terms of compounds. As there is a relatively low overlap of compounds among existing spectral databases, most scientists currently use multiple databases. Since the 2016 review, many of the databases have expanded their compound coverage immensely and many of the open libraries have cross-imported their spectral records. However, the issue of overlap and coverage of relevant substances in chemical space remains [17].

Conceptually, the premise of spectral library searching is very simple: the fragmentation pattern of a molecule is a reproducible fingerprint of that molecule under a given set of fixed conditions, such that unknown spectra acquired under similar conditions can be identified via spectral matching [33]. Automated spectral library searching involves software with tailor-made search algorithms for tandem mass spectral databases [34–37]. The search score obtained following the database search represents the likelihood that the searched spectrum corresponds to a given reference spectrum in the mass spectral database. A low score indicates that the experimental fragmentation pattern has low similarity to any stored reference spectrum. A high score indicates significant spectral overlap and, consequently, that the analyte is likely either structurally similar or even identical to the reference compound. Library search should be both sensitive and specific, producing as few false negative and false positive results as possible. Ideally, the scores obtained should be able to distinguish true and false positive matches [18]. To compare with historical targeted methods applied in a regulatory context (e.g., forensic toxicology and food safety), the primary objective of a screening method is to limit the risk of false negatives (e.g., 1%) and to keep to an acceptable risk of false positives (e.g., 5%). The latter should be further reduced by confirmatory analyses using reference standard compounds. Non-targeted methods with consolidated QA/QC thus have to consider and document these issues during the development and evaluation of method performances, as well as in reporting of results. False discovery rates, applied successfully in proteomics for many years [38], are now being developed for small molecule MS/MS [39], but are not yet widely integrated into tandem library software.

There is also extensive discussion about the robustness and transferability of tandem mass spectral libraries. For a long time, the predominant opinion was that libraries would only be useful on the instrument used to acquire reference spectra, due to the limited reproducibility of tandem mass spectra. This situation has changed thanks to both progress in instrument technologies and informatics tools. Databases combining advanced library

designs with tailor-made search algorithms have been shown to enable reliable compound identification with spectra acquired in different laboratories with various instruments and different instrument settings [18, 27]. While pre-acquisition harmonization of analytical procedures was researched, the participating laboratories encountered a number of difficulties [40]. Thus, the current trend is rather to look for a post-acquisition flexibility of the MS/MS reference library and associated matching algorithms to deal (as much as possible) with the diversity of imported experimental data without sacrificing the ambitioned confidence level in terms of correct annotation.

Over the last 10 years, there has been substantial progress in the quality of tandem mass spectral databases. Today, spectral acquisition of reference spectra is accomplished regularly on high-resolution instrumentation (i.e., quadrupole–quadrupole-time of flight (QqTOF), Orbitrap) employing multiple collision energies for fragmentation to comprehensively cover the breakdown curves of reference compounds. Besides the protonated and deprotonated molecular ions, adduct ions, in-source fragments as well as isotopologues are commonly selected as precursor ions [18, 25, 26]. Furthermore, to improve spectral quality, generally only curated spectra are stored in databases, which come bundled with improved search algorithms. As knowledge and understanding of mass spectra increase, automated curation procedures are being implemented and constantly improved to reduce the manual curation load associated with mass spectral database creation [15, 25]. Overall, the ambition for a “universal tandem mass spectral database” is closer to a reality. However, this ambition requires definition and implementation of some common procedures to ensure the reliability and robustness of the generated data, both stored in the desired tandem mass spectral reference library and generated from each experimental sample.

A number of reference tandem mass spectral databases exist that have to be considered in the frame of identification of chemicals of emerging concern, in terms of structuration and/or content. These existing resources should serve as a basis to avoid unnecessary time spent in re-implementing existing and reliable elements, and to ensure a coherence of the ambitioned outputs with potential established standards. However, an obvious lack of high-level QA/QC consolidation appears within many of these existing databases (e.g., percentage of erroneous information, insufficiently or non-adequately curated spectra), together with some necessary adjustments for specific applications (e.g., human metabolites of contaminants are not well represented as compared to parent compounds).

Methods—proposed QA/QC framework

Generic strategy for converting suspects into targets

The long-term goal in terms of developing screening capabilities would be to progressively include a large part, ideally all, of the compounds listed in “suspect lists of interest” (e.g., [41]) into corresponding entries in tandem mass spectral libraries. This would allow the conversion of suspects (generally Level 3) into targets (Level 1, if the retention time information has been measured in house or on an identical chromatographic regime elsewhere) or higher confidence tentative matches (Level 2a, spectral library match). The strategy for accomplishing this aim

involves (1) QC of already acquired tandem mass spectral data to determine how many suitably comparable mass spectral records exist and (2) QA-guided acquisition of new reference spectra, shown in Fig. 2.

To determine a baseline for the status of environmentally and toxicologically relevant compounds and their presence in various resources, a mapping exercise was performed. Compound numbers (number of entries) were obtained for the CompTox Chemicals Dashboard [42], NORMAN SusDat [43], HMDB [28], DrugBank [44], the Toxic Exposome Database (TEDB) [45] and Exposome Explorer [46] from their respective websites



Fig. 2 The proposed workflow to convert suspects into targets

or download files on March 15, 2019. CompTox numbers mapping to mzCloud [47], MassBank [48] and WRTMD [49] were obtained via downloading the respective list files (list codes on https://comptox.epa.gov/dashboard/chemical_lists are: MZCLOUD, HDXNOEX, MASSBANKREF, MYCOTOXINS, WRTMD), also on March 15, 2019 and counting/merging by InChIKey first block (thus ignoring stereochemistry). HMDB MS/MS numbers were obtained from download files (March 15, 2019) and cross-checked with InChIKey mappings still on record from a previous study [16]; also counting by InChIKey first block. SusDat mappings to MassBank were obtained by extracting list S1 results from the download file. To provide a global, up-to-date overview, all compounds with MS/MS annotation that were listed in the PubChem [50] Table of Contents Browser (<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72>) were exported as SMILES [51], converted to InChIKeys using Open Babel [52], and counted by unique InChIKey first block. While both DrugBank and T3DB contain MS/MS records, this information is not available in their export files, and these contain high overlap with HMDB where the information is mapped extensively in the download files.

The authors note that while many more resources are available, these are open resources with pre-mapped information to the highest quality and relevant MS/MS records to form a sufficient information basis for the outcome of this article.

Quality control and benchmarking of tandem mass spectral libraries

The library of the Helmholtz Centre for Environmental Research (UFZ) being part of MassBank was used as test set to demonstrate the usefulness of the proposed strategy. The UFZ library (at that stage) contained 636 MS/MS spectra corresponding to 167 compounds. Reference spectra were recorded on a LTQ-Orbitrap XL (Thermo Fisher Scientific, Waltham, MA, USA). HCD product-ion spectra were acquired at three different collision energy levels (HCD 35, 55, 80) at a nominal resolving power of 30,000. The R package RMassBank was used to perform recalibration and clean-up of acquired spectra [15]. The curated spectra are available at <https://github.com/MassBank/MassBank-data/tree/master/UFZ>.

The spectra of the UFZ library were searched against “The Wiley Registry of Tandem Mass Spectral Data” (WRTMD) [53]. Library search was accomplished using ‘MSforID Search’ [34, 35] as described in the Additional file 1. The spectra of compounds covered in both UFZ and WRTMD served as positive controls. The number of positive identifications obtained with the positive controls were counted and used to calculate the statistical parameter sensitivity (= true positive rate).

Interlaboratory study to validate acquired reference spectra

An interlaboratory study was organized to verify that participating laboratories were generating new reference data with experimental settings and workflows that were

Table 1 Fifteen test compounds used to assess the interlaboratory comparability of tandem mass spectral data

Name	DTXSID	SMILES	Molecular formula	Monoisotopic mass
Amiloride	DTXSID9043853	<chem>NC(=N)NC(=O)C1=NC(CI)=C(N)N=C1N</chem>	C ₆ H ₈ ClN ₇ O	229.04789
Phenazone	DTXSID6021117	<chem>CN1N(C(=O)C=C1C)C1=CC=CC=C1</chem>	C ₁₁ H ₁₂ N ₂ O	188.09496
Cyclizine	DTXSID4022864	<chem>CN1CCN(CC1)C(C1=CC=CC=C1)C1=CC=CC=C1</chem>	C ₁₈ H ₂₂ N ₂	266.17830
Desipramine	DTXSID6022896	<chem>CNCCCNC1C2=C(CCC3=C1C=CC=C3)C=CC=C2</chem>	C ₁₈ H ₂₂ N ₂	266.17830
Dibucaïne	DTXSID3045271	<chem>CCCCOC1=CC(C(=O)NCCN(CC)CC)=C2C=CC=CC2=N1</chem>	C ₂₀ H ₂₉ N ₃ O ₂	343.22598
Dixyrazine	DTXSID10947639	<chem>CC(CN1CCN(CC1)CCOCCO)CN2C3=CC=CC=C3SC4=CC=CC=C42</chem>	C ₂₄ H ₃₃ N ₃ O ₂ S	427.22935
Dothiepin	DTXSID2022961	<chem>CN(C)CCC=C1C2=CC=CC=C2CSC2=C1C=CC=C2</chem>	C ₁₉ H ₂₁ NS	295.13947
Ethambutol	DTXSID8023006	<chem>CC[C@@H](CO)NCCN[C@H](CC)CO</chem>	C ₁₀ H ₂₄ N ₂ O ₂	204.18378
Etofylline	DTXSID5023031	<chem>CN1C2=C(N(CCO)C=N2)C(=O)N(C)C1=O</chem>	C ₉ H ₁₂ N ₄ O ₃	224.09094
Mefruside	DTXSID0048844	<chem>CN(CC1(C)CCCCO1)S(=O)(=O)C1=CC=C(CI)C(=C1)S(N)(=O)=O</chem>	C ₁₃ H ₁₉ ClN ₂ O ₅ S ₂	382.04239
Metoclopramide	DTXSID6045169	<chem>CCN(CC)CCNC(=O)C1=CC(CI)=C(N)C=C1OC</chem>	C ₁₄ H ₂₂ ClN ₃ O ₂	299.14005
Nylidrin	DTXSID4023387	<chem>CC(CCC1=CC=CC=C1)NC(C)C(O)C1=CC=C(O)C=C1</chem>	C ₁₉ H ₂₅ NO ₂	299.18853
Phentermine	DTXSID9023461	<chem>CC(C)N)CC1=CC=CC=C1</chem>	C ₁₀ H ₁₅ N	149.12045
Sulfamethoxazole	DTXSID8026064	<chem>CC1=CC(NS(=O)(=O)C2=CC=C(N)C=C2)=NO1</chem>	C ₁₀ H ₁₁ N ₃ O ₃ S	253.05211
Sulfamoxole	DTXSID5023617	<chem>CC1=C(C)N=C(NS(=O)(=O)C2=CC=C(N)C=C2)O1</chem>	C ₁₁ H ₁₃ N ₃ O ₃ S	267.06776

Further information (InChI, SMILES, InChIKeys, etc.) is provided in Additional file 2

Table 2 Overview of instrumentation used by the laboratories participating in the interlaboratory study dedicated to evaluating the degree of comparability and transferability of acquired reference spectra

Laboratory	Instrument type	Brand	Fragmentation technique
1	Q-Orbitrap-LIT	Orbitrap Fusion Tribrid, Thermo	HCD
2	Q-Orbitrap	Q Exactive, Thermo	HCD
3	Q-Orbitrap	Q Exactive Plus, Thermo	HCD
4	LIT-Orbitrap	LTQ Orbitrap XL, Thermo	HCD
5	LIT-Orbitrap	LTQ Orbitrap XL, Thermo	HCD and CID
6	QqTOF	6530 QTOF, Agilent	CID
7	QqTOF	TripleTOF 5600+, Sciex	CID

compatible with existing reference spectra collections. Each participating lab was asked to use fifteen reference compounds for producing tandem mass spectral libraries. Table 1 includes the set of compounds used in this study. They have already been applied to test the transferability of the WRTMD in different laboratories with various available instruments and procedures (e.g., [18, 34, 36]). Seven laboratories involved in the NORMAN Network and/or HBM4EU participated in the interlaboratory study. An overview of the applied instrumentation as well as the applied fragmentation technique is provided in Table 2 as well as in the Additional file 1.

The seven collections of centroided, averaged and curated tandem mass spectra were benchmarked against the WRTMD. Benchmarking included two sets of experiments: (1) matching the test spectra to the WRTMD, and (2) matching the spectra of the 15 test compounds included in the WRTMD to modified libraries derived from the WRTMD by substituting the original reference spectra with the newly generated libraries. For statistical evaluation of library search performance, all test sets were grouped according to the collision energy settings used to acquire the individual spectra.

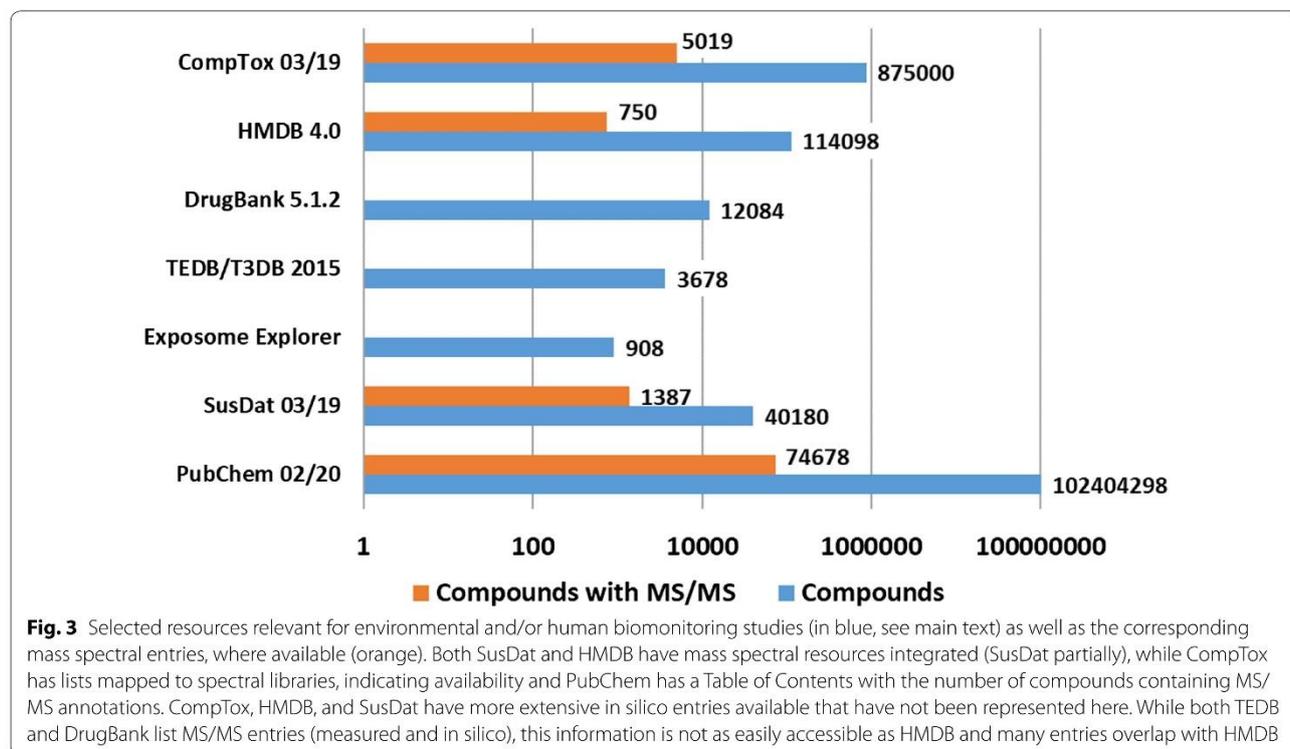
Results and discussion

Starting point—overview on existing tandem mass spectral data for chemicals of potential concern

Various initiatives and/or sources of information have documented proposed lists of chemicals of emerging concern in various contexts, mainly in the field of environment and toxicology. Examples are the many lists on the NORMAN Suspect List Exchange [41] and the corresponding merged database SusDat [43], the CompTox Chemicals Dashboard [42] and a series of topical databases from the Wishart laboratory and collaborators (e.g., HMDB [28, 30], DrugBank [44], the Toxin and Toxin Target Database (T3DB) or Toxic Exposome Database (TEDB) [45] and the Exposome Explorer database [46]). While these sources overlap to some extent, they also

provide a lot of complementary information and functions. As several reviews have covered the number of overlapping substances in tandem mass spectral libraries in various contexts (especially metabolomics) recently [16, 17, 54], the focus here will be on substances of interest for the environmental and biomonitoring contexts, using the resources mentioned. Looking into these collections, tandem mass spectral fragmentation information is already available for a considerable number of suspects in MassBank, HMDB, WRTMD or mzCloud, but this represents only a fraction of the substances actually present in the respective resources.

An overview of the number of compounds in the respective resources, as well as the number of entries that map to mass spectral data within that resource is given in Fig. 3. The CompTox Dashboard (875,000 compounds) includes 3997 compounds in mzCloud, 2377 in MassBank and 1429 in WRTMD, corresponding with 5019 unique compounds (ignoring stereochemistry differences), thus 0.57% of the resource. HMDB (144,098 compounds) contains MS/MS data corresponding to 750 unique compounds (ignoring stereochemistry), or 0.66% of the resource. NORMAN SusDat contains 40,180 entries, of which 1387 are in MassBank (3.6% of SusDat). This overview shows that tandem mass spectral data is available only for a rather low number of compounds. A further complicating factor is that these tandem mass spectral data are spread among several spectral collections. For the vast majority of interesting suspects, no public mass spectral data exists and measured mass spectral data will have to be newly generated, if possible. While METLIN now claims MS/MS spectra of over 500,000 chemicals (<https://metlin.scripps.edu/>, accessed 8 Dec. 2019), information on the coverage is not available, nor are the spectra openly available. However, as the PubChem database [50] aggregates information from a number of sources, the 74,678 compounds with MS/MS annotations (ignoring stereochemistry; 89,726 with stereochemistry), of 102,404,298 compounds (~ 0.073%



of PubChem) give a reasonable indication of the total number of compounds with MS/MS information available to some extent, although some of these are in silico and many of these are not directly relevant for human biomonitoring or environmental studies.

The generation of new reference spectra is considered to represent an important element for the successful and long-term establishment of non-targeted LC–MS. As a single laboratory will not have the necessary resources available to handle the huge number of suspects ahead, this challenge must be addressed as a group. For successful realization of multi-partner generation of reference spectra, harmonization of acquisition and processing strategies is essential. Related QA actions imply application of generally agreed best practice procedures and participation in interlaboratory studies (see below). However, even by joining forces with respect to manpower and instrumentation, there will be further challenges ahead, and these are related to prioritization of suspect lists and availability of the corresponding reference standards.

Existing techniques for prioritizing chemicals are generally based on risk assessment [55, 56], which involves assessment of exposure and hazard. Other useful criteria might represent detectability by analytical techniques (e.g., LC–MS with ESI in positive or negative ion mode), legal status, importance for a defined research project [1], or simply the availability of reference information [4, 5, 57].

As things are now, over the next years a steady increase of the number of chemicals of emerging concern included in tandem mass spectral libraries is expected. Already available spectral collections are considered to represent nuclei for even larger collections. Therefore, much effort should also be put into the QC of already acquired tandem mass spectral data to determine how many suitably comparable mass spectral records already exist (see below).

Quality control of MassBank collections

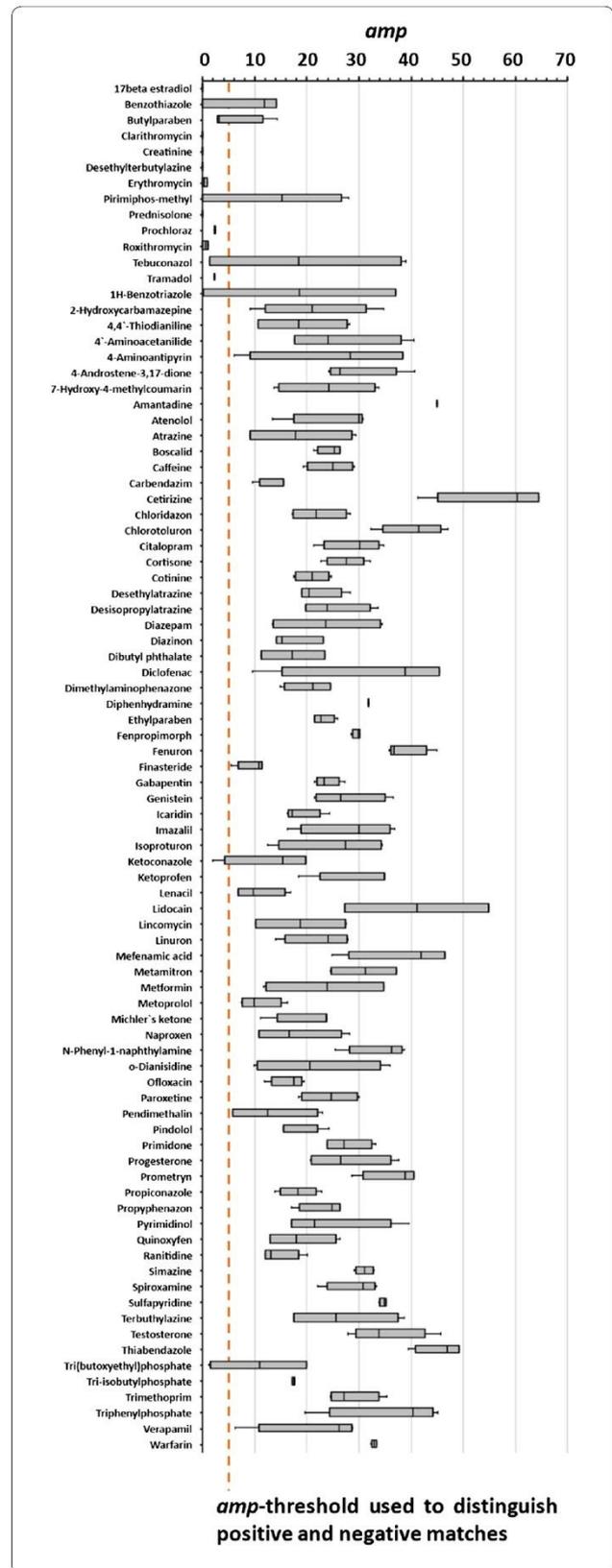
MassBank is an important collection of reference tandem mass spectra [24]. Currently, 45 collections are available on MassBank (<https://massbank.eu/MassBank/RecordIndex>) with more than 55,075 tandem mass spectra (of 76,037 spectra total) representing 14,297 compounds (15,988 stereoisomers) total (over all spectral types). In terms of compound coverage, there is significant overlap between MassBank and the WRTMD that can be used to create sets of positive controls for testing the libraries.

Positive controls are particularly suitable for testing the quality and comparability of databases. Matching positive controls is used to determine the sensitivity (= true positive rates) of a database. Ideally, the obtained sensitivity values should be close to 100%. Negative controls are used to test the specificity (= true negative rate) of a database.

Fig. 4 Overview on the average match probability (*amp*) distribution obtained from matching 352 positive controls representing 87 compounds part of the UFZ library to the WRTMD. An *amp*-threshold of 5.0 was used to distinguish between positive and negative matches. The overall sensitivity was 89.7%. For 70 compounds, all test spectra led to positive matches with $amp > 5.0$. For sixteen compounds at least one test spectrum retrieved an *amp*-value below the specified threshold of 5.0 indicating insufficient similarity between test and reference spectra

Initial benchmarking efforts between the Swiss Federal Institute of Aquatic Science and Technology (Eawag) MassBank collection and the WRTMD were published recently [27]. Spectra from the 233 overlapping substances between the two collections were used as positive controls. Of particular interest was the fact that the Eawag spectra were acquired with an Orbitrap instrument (HCD and CID), whereas the WRTMD spectra were acquired on a QqTOF. Spectra in the range of collision energy 20–50 eV on the QqTOF and 30–60% NCE on the Orbitrap provided optimal library matching results with sensitivity-values 95.1–98.4% [27]. Therefore, it was concluded that both collections enable reliable compound identifications, and that they are ready for use in suspect screening applications.

Another important spectral collection within MassBank is the UFZ library. The library contains tandem mass spectra of 167 compounds. The spectra were acquired on an Orbitrap with HCD. For each compound, reference spectra were acquired with three different collision energy settings. All spectra were curated and recalibrated before storing in MassBank. 87 reference compounds included in the UFZ library were also covered by the WRTMD. For each of these compounds two to eight spectra acquired at different collision energy settings (35, 55, 80%) were available. The corresponding 352 spectra represented positive controls suitable for QC of the UFZ library. The spectra were matched to the WRTMD and the number of positive matches was statistically evaluated (Fig. 4). The overall sensitivity was 89.7%. For 70 compounds, all test spectra performed well ($amp > 5.0$) and led to a positive match. There were, however, 16 compounds, of which at least one test spectra retrieved an *amp*-value below the specified threshold of 5.0 indicating insufficient similarity between test and reference spectra. Communicating the benchmarking results to the authors of the library initiated a fruitful discussion that also included reviewing of the raw data. This process identified reasons for the observed differences between test and reference spectra. In this way, the issues could be resolved and the corresponding entries in MassBank were updated.



Thorough quality control of already existing spectral collections is able to identify libraries (or subsets thereof) for immediate application to compound identification in suspect screening. Likewise, benchmarking of tandem mass spectral libraries is a suitable approach to identify specific errors like low signal-to-noise ratios, improper mass calibration or wrong compound labeling. In this way, low-quality spectra can be identified, corrected or even deprecated.

Recommendations for QC of existing tandem mass spectral libraries

On the basis of the above results, as well as the previously published and discussed benchmarking studies, the following two-step QC procedure could be drafted and adopted:

Firstly, tandem mass spectral data should meet the following quality criteria:

- QC1 Acquisition: High-resolution instrumentation (e.g., QqTOF, Orbitrap, Fourier Transform ion cyclotron resonance (FTICR)), typically with a minimum resolution of 10,000 in MS/MS mode, and m/z error lower than 10 ppm, to ensure contributions of broadly applicable spectra can be made from many laboratories.
- QC2 Ionization: Positive and/or negative mode with a specified ionization technique (ESI, atmospheric pressure chemical ionization or atmospheric pressure photo ionization), as these are the most common methods.
- QC3 Precursor ion isolation: The isolation window should be as narrow as possible to avoid fragmentation of multiple precursors, including isotopic peaks.
- QC4 Fragmentation: Either Orbitrap-HCD or QqTOF-CID should be used for generating MS/MS spectra, as such spectra can be searched with higher sensitivity (see discussion below and Fig. 5).
- QC5 Mass range: Ideally the mass range should start at $m/z \leq 50$ (instrumental limitations may preclude this, e.g., for instruments relying on ion trapping) wherever possible to include also small mass fragments. The acquired mass range should be given to avoid poor spectral matches due to the presence/absence of low m/z fragments in fragmentation spectra acquired with different scan ranges (see discussion below and Fig. 6).
- QC6 Collision energies: Multiple collision energies (minimum 3) should have been recorded wherever possible over a meaningful range (e.g., 5

to 60 eV CID; NCE 10–60% HCD [27]) to form compound-specific breakdown curves.

- QC7 Curation: Centroiding, filtering, noise removal, and recalibration should have been performed where possible to provide the best quality reference spectra.
- QC8 Expert review: This is necessary to identify issues, such as artifacts, improper noise removal, or truncated spectra, which cannot always be captured automatically.

Secondly, spectral collections satisfying these conditions will proceed to the benchmarking step, using procedures described in the section “Quality Control of Mass Bank Collections”.

Multi-partner acquisition of new tandem mass spectral data

Interlaboratory harmonization studies are useful to verify that a laboratory is generating new reference data with experimental settings and workflows that are compatible with existing reference spectra collections. One way to characterize this interlaboratory comparability is to introduce a number of predefined known compounds and accompanying QA/QC criteria, such that these compounds must be detected and successfully identified with a given instrumentation and related procedure to validate the method appropriateness and reliability.

Seven laboratories involved in the NORMAN Network and/or HBM4EU project participated in the first harmonization study. The study was aimed to demonstrate compatibility and transferability of newly acquired tandem mass spectral data among participating laboratories as well as with already available reference spectra collections. The study involved the measurement of 15 reference standards (Table 1) on three different Orbitrap configurations (at four locations) and two QqTOFs (Table 2). The WRTMD served as the reference library.

The eight collections of centroided, averaged and curated tandem mass spectra were benchmarked against the WRTMD (Table 3). In a first set of experiments, acquired tandem mass spectra were matched to the WRTMD. The number of positive identifications obtained with individual test sets ranged from 73.2 to 100%. To prove that even the sets that showed a significant number of negative matches (e.g., laboratories 3, 4, and 5) contained suitable collision energy windows, the eight test sets were further grouped according to the collision energy settings used to acquire the individual spectra. As expected, a considerable number of subgroups were identified that led to 100% correct positive identifications (Fig. 5). The collision energy windows

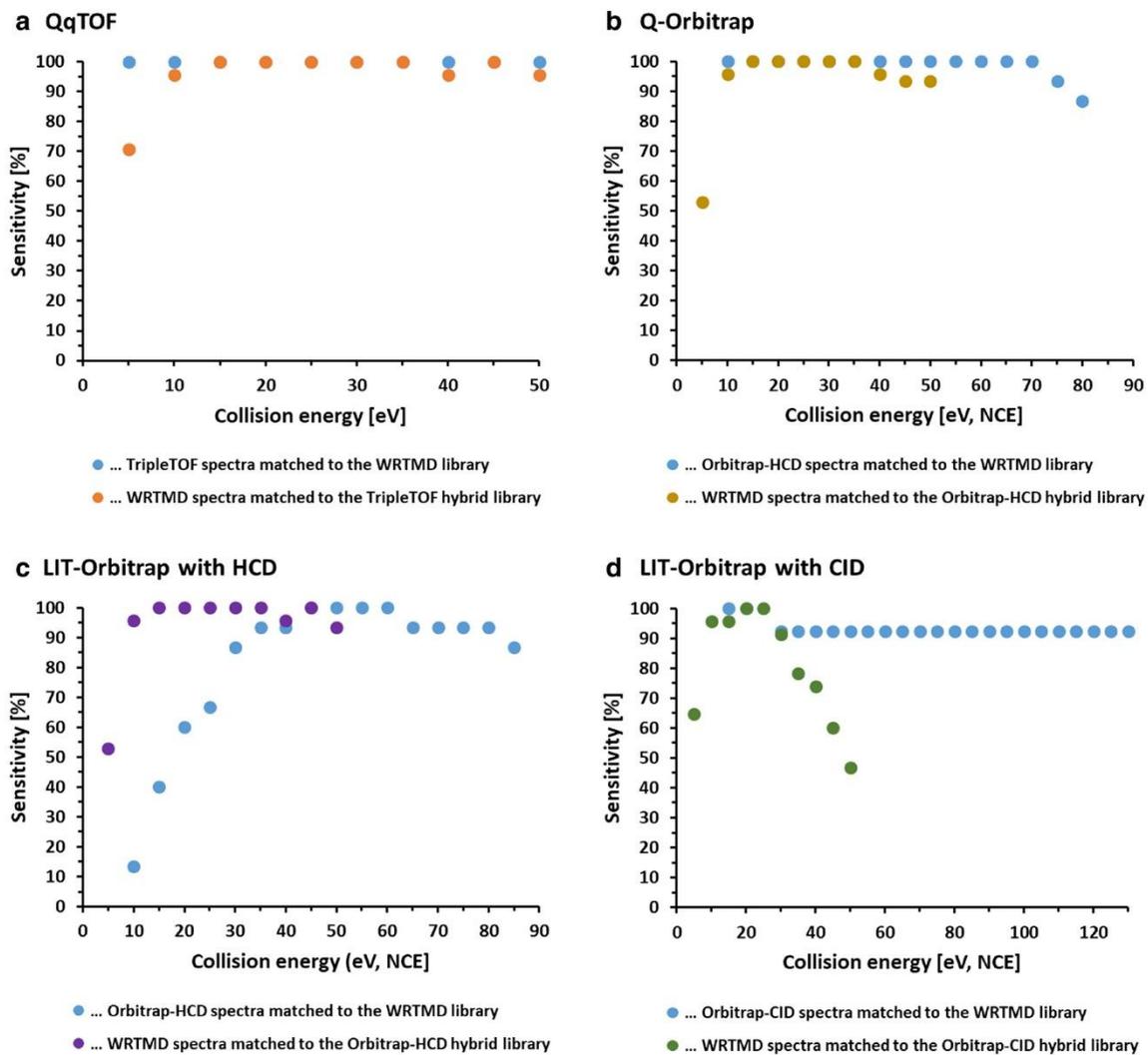


Fig. 5 The reliability of library matching with the WRTMD and libraries derived from the WRTMD by substituting the original reference spectra with instrument-specific library spectra is shown by sensitivity versus collision energy for **a** a QqTOF instrument, **b** a Q-Orbitrap instrument with HCD, **c** a LIT-Orbitrap instrument with HCD, and **d** a LIT-Orbitrap instrument with CID, respectively. The collision energies are given in eV for the QqTOFs and NCE for the Orbitraps. **a–c** Reliable matches in a wide CE range, while **d** shows that the optimal CE window is smaller for LIT-Orbitrap instrument with CID

that appeared to be suitable for acquiring test spectra spanned at least 15 units (eV or % NCE). In the second set of benchmarking experiments, the spectra of the 15 test compounds included in the WRTMD were matched to eight libraries derived from the WRTMD by substituting the original reference spectra with the newly generated sets of reference spectra. The number of positive identifications obtained with individual test sets ranged from 78.5 to 99.5%. Also in this case, the test sets were further grouped according to the collision energy settings used to acquire the individual spectra. Like in the other experiment, a considerable number of subgroups were

identified that led to 100% correct positive identifications (Fig. 5). The collision energy windows that appears to be suitable for acquiring library spectra spanned at least 10 units (eV or % NCE).

The interlaboratory study clearly demonstrated that the participating laboratories are able to acquire high-quality reference spectra for building libraries, while also providing further evidence that Orbitrap-HCD and QqTOF-CID introduce quite similar fragmentation reactions (Fig. 5). Thus, libraries produced on these types of instruments will offer complementary identification possibilities. Of utmost importance was the observation

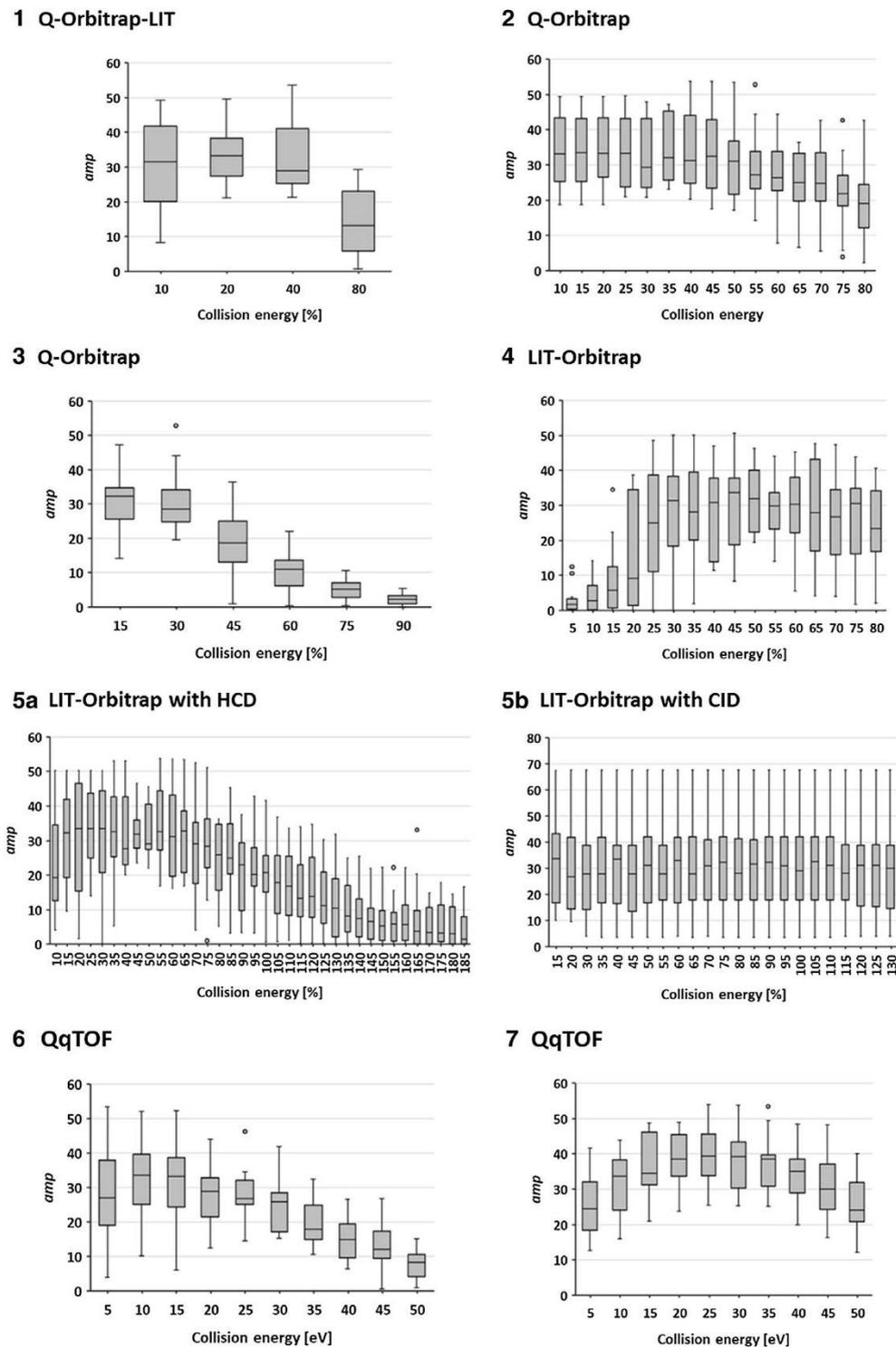


Fig. 6 Overview on the average match probability (*amp*) distribution obtained from library matching of the spectral collections acquired during the interlaboratory study on the following instrument configurations: (1) Q-Orbitrap-LIT, (2, 3) Q-Orbitrap, (4) LIT-Orbitrap, (5a) LIT-Orbitrap with HCD, (5b) LIT-Orbitrap with CID, and (6, 7) QqTOF. The numbering (1–7) decodes the laboratories. Fragmentation on the Orbitraps 1–4 was accomplished with HCD. The collision energies are given in eV for the QqTOFs and NCE for the Orbitraps. The WRTMD served as reference library. Even for instruments with identical configurations a considerable inter-instrument variability in the optimal collision energy range necessary for obtaining library matches with high *amp*-values was observed

Table 3 Results of the interlaboratory study involving 1972 spectra of 15 compounds acquired in seven labs using different Orbitrap configurations and QqTOFs

Laboratory	Instrument type	CE range	Number of spectra	Sensitivity [%] of the test set	Sensitivity [%] of the library
1	Q-Orbitrap-LIT	10–80%	60	95.0	92.8
2	Q-Orbitrap	10–80%	225	98.7	94.2
3	Q-Orbitrap	15–90%	82	73.2	78.5
4	LIT-Orbitrap with HCD	5–80%	236	81.8	94.7
5	LIT-Orbitrap with CID	10–130%	518	82.4	83.2
	LIT-Orbitrap with HCD	10–185%	343	93.4	93.8
6	QqTOF	5–50eV	150	94.0	79.9
7	QqTOF	5–50 eV	150	100	96.2

The eight spectral collections were benchmarked against the WRTMD. The numbers of correct positive identifications (= sensitivities) obtained for the individual sets are listed

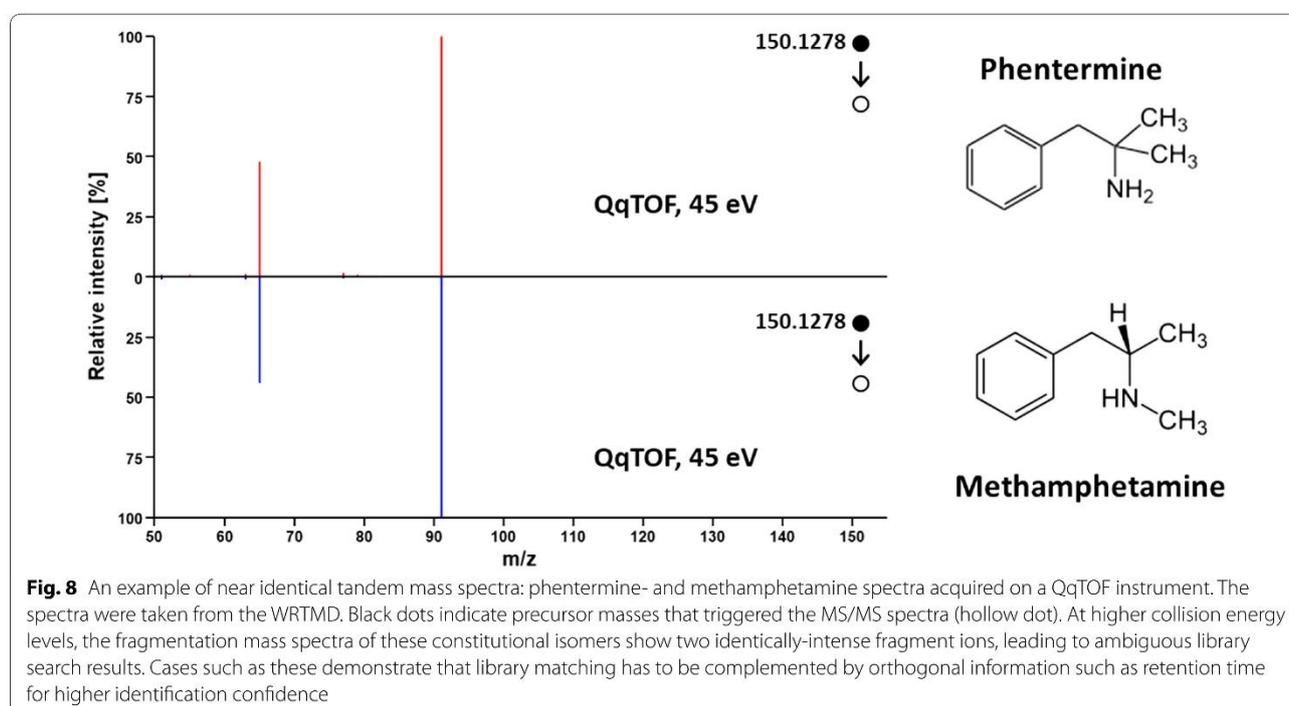
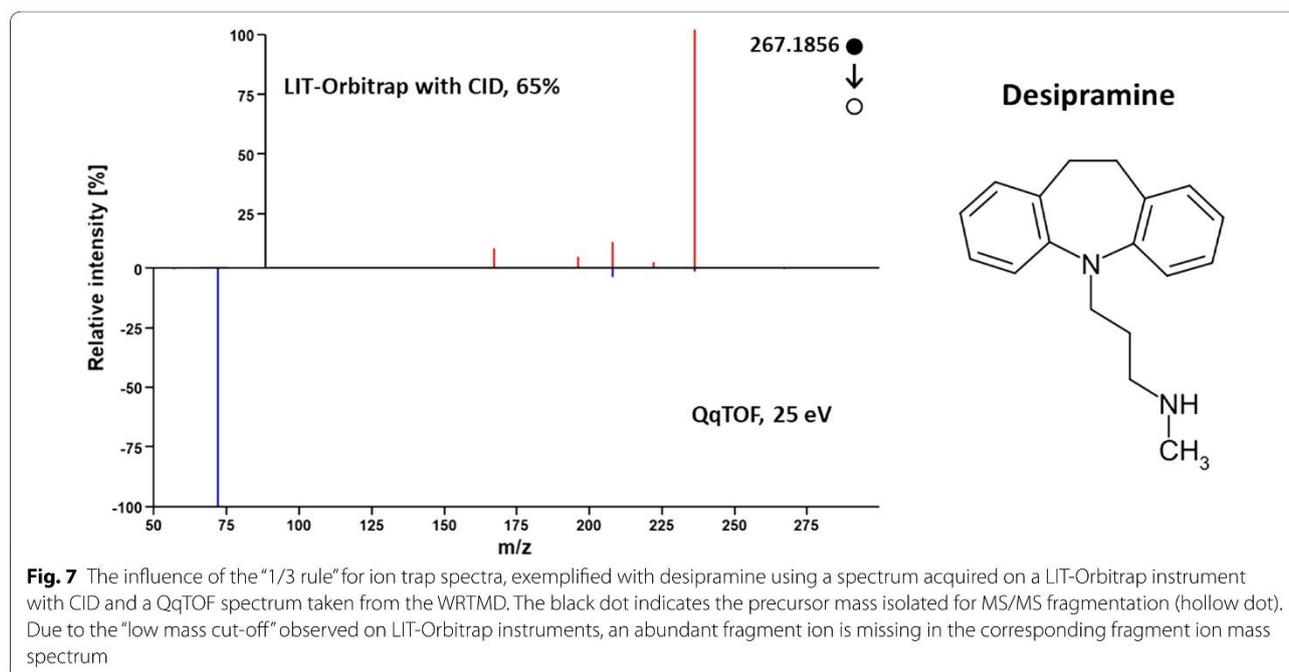
that there is a significant overlap of the compound-specific collision energy ranges between instruments (Fig. 5). Thus, databases that contain series of multiple spectra acquired on one instrument will enable reliable compound identifications when querying spectra from other instruments. Clearly, databases produced in different laboratories will offer complementary identification possibilities.

Another important result of the interlab study was the observation that even for instruments with identical configurations, a considerable inter-instrument variability in the optimal collision energy range necessary for obtaining library matches with high match probability was observed (Fig. 6). Taking into consideration that all Orbitrap technology-based instruments were provided by the same manufacturer, a higher degree of similarity between those instruments regarding compound-specific breakdown curves was expected. The observation suggests that even after years of instrument development and optimization, harmonization of collision energy values has hardly been accomplished yet. The good news is, however, that state-of-the-art tandem mass spectral databases can cope with spectral variability leading to high reliability of a library match. The analyst applying these libraries just needs to find the optimal collision energy corridor for acquiring test spectra. The herein presented interlaboratory study could represent an appropriate strategy for this purpose.

The interlaboratory study also highlighted some limitations. These are mainly connected with the use of tandem-in-time fragmentation in the ion trap of the linear ion trap (LIT)-Orbitrap instrument (Fig. 5d). In contrast to quadrupole collision or HCD cells that use non-resonant CID, ion traps use resonant CID (i.e., several low-energy collisions during a longer time than non-resonant), which enables to produce fragmentation trees

beyond MS^2 . Generally, these fragmentation trees cover the full range of possible fragmentation pathways, and are therefore specific identifiers for the corresponding molecules, which can be stored in databases (e.g., mzCloud [47]). With ion trap MS^2 , only parts of the entire range of possible fragmentation reactions are covered, even when applying higher collision energies [58, 59]. Such spectra match the lower energy part of spectral series acquired on tandem-in-space instruments (Orbitrap-HCD and QqTOF-CID) well. There is, however, limited overlap with spectra acquired at higher collision energies. Another problem of ion trap fragmentation is related to the “low mass cut-off”, or the so-called “1/3 rule” [59]. This means that fragment ions with an m/z -value below 1/3 of the m/z -value of the precursor ion are not trapped under normal operation conditions and are lost. Thus, a considerable part of fragment ions that are observed with higher collision energy of fragmentation on tandem-in-space instruments is not detectable with IT analysers. Thus, in comparison to Orbitrap-HCD and QqTOF-CID, Orbitrap-CID spectra are truncated. This truncation can hamper compound identification if abundant fragment ions are missing. One such example is desipramine (Fig. 7). At low collision energies, this compound has one abundant fragment ion at m/z 72.0808. This ion was not observed in the LIT-Orbitrap spectra since it displayed a m/z -ratio lower than 1/3 of the precursor ion (m/z 267 for $[M+H]^+$). Accordingly, spectral match gave a low score.

Another limitation of tandem mass spectral databases is highlighted in Fig. 8. It is well recognized that stereoisomers can hardly be distinguished from each other by tandem mass spectral fragmentation [13, 60]. But even the differentiation of constitutional isomers can be challenging. Fragmentation of such compounds may lead to identical products. In the worst-case scenario, tandem



mass spectra will be identical. One such example of a pair of constitutional isomers comprises phentermine and methamphetamine. At higher collision energy levels, the fragmentation mass spectra of these two compounds show two identically intense fragment ions (Fig. 8), leading to ambiguous library search results.

QA recommendations for acquisition and processing of tandem mass spectral reference data

The results of the interlaboratory study, as well as the available experience and knowhow in building tandem mass spectral libraries, formed the basis for drafting and adopting the following recommendations:

- QA1 Acquisition: High-resolution instrumentation (e.g., QqTOF, Orbitrap, FTICR) should be used for the acquisition of reference tandem mass spectra.
- QA2 Instrument performance: Instruments should be properly tuned and calibrated (ideally daily or before commencing a batch analysis). High mass accuracy should be maintained using a lock mass or similar. The instrument should be capable of a minimum resolution of 10,000 in MS/MS mode, and the m/z error should be lower than 10 ppm.
- QA3 Standards: Certified reference standards should be used to ensure that spectra will represent the linked structure.
- QA4 Sample introduction: Samples may be introduced by direct infusion, flow injection or chromatography. A special caution should be paid to the minimal number of acquisition points (related to dwell time values and scan speed capabilities) to ensure a sufficient number of spectra for averaging. The possible occurrence of background interferences should be checked by introducing blank samples.
- QA5 Separation of mixtures: If reference compounds are introduced in mixtures, proper separation of the individual precursor ions (either during sample introduction or the mass spectrometric analysis) must be ensured to avoid the acquisition of chimeric spectra.
- QA6 Dealing with isobars: If reference spectra are acquired in batches, isobaric compounds must not be processed consecutively, to avoid interferences due to carryover effects resulting in chimeric spectra.
- QA7 Alternative precursors: While the primary precursors of interest may be protonated or deprotonated molecules, for some molecules other abundant signals corresponding to in-source fragments, isotopic peaks or other related species might be considered as additional precursor ions.
- QA8 Isolation width: The precursor isolation window should be as narrow as possible to avoid fragmentation of multiple precursors, including isotopic peaks.
- QA9 Fragmentation: Fragmentation should be accomplished by tandem-in-space techniques (e.g., HCD for Orbitrap, CID for QqTOF). As shown in Fig. 7, CID fragmentation with tandem-in-time instruments may produce spectra with limited overlap to spectra acquired with tandem-in-space techniques at higher collision energies. Another problem of ion trap fragmentation is related to the “low mass cut-off”.
- QA10 Scan range: The lower limit of the applied scan range should ideally be $\leq m/z$ 50. The lower limit of the applied scan range must not exceed m/z 100 to avoid the production of truncated spectra (instrumental limitations may preclude this, e.g., for instruments relying on ion trapping) wherever possible to include also small mass fragments. The acquired mass range should be given to avoid poor spectral matches due to the presence/absence of low m/z fragments in fragmentation spectra acquired with different scan ranges (see Fig. 7).
- QA11 Collision energies: Compound-specific breakdown curves should be covered by spectra acquired at multiple collision energies. A spectral series should contain at least 3-5 fragment ion spectra acquired at sufficiently different collision energies within the defined range. With this strategy, libraries are produced that are robust against inter-instrumental collision energy variability (see Figs. 5 and 6). If ramped collision energies are used, these should be clearly labelled as such.
- QA12 Signal-to-noise: Sample concentration should be sufficiently high to produce fragment ion mass spectra with signal-to-noise ratios >100 , to enable reliable acquisition of low-abundant fragment ions.
- QA13 Saturation: Detector saturation must be avoided for fragment ions, and is only acceptable for precursor ions if the resulting artifacts are removed during curation.
- QA14 Centroid spectra: Fragment ion mass spectra should be acquired in centroid mode or centroided during export and curation.
- QA15 Curation: Spectra should be curated, which includes multiple steps of filtering, noise removal, and recalibration, to provide the best quality reference spectra.
- QA16 Expert review: Spectral series should be reviewed by an expert to identify issues like artifacts, improper noise removal, or truncated spectra, which cannot always be captured automatically.
- The curation of acquired tandem mass spectral data is of utmost importance to obtain a high-quality library.

Curation efforts may include noise and artifact removal, recalibration of spectra and peak annotations, manual inspection of mass spectra by experienced mass spectrometrists, as well as inter-library comparisons (e.g., [15, 25, 27, 61]). Removal of noise during data processing may lead to losses of spectral information of compounds. Accordingly, processed spectra should be reviewed by experienced mass spectrometrists to check the integrity of data with a special focus on the occurrence of artifacts and processing errors.

Conclusions

Acquisition of tandem mass spectral data requires the physical availability of reference substances and sufficient experimental capacities for acquiring fragment ion mass spectra. Even by joining forces, for instance within HBM4EU and NORMAN initiatives, the acquisition of reference data for ten thousands of compounds is a multi-annual project requiring significant resources. With this document, an outline for harmonized acquisition of suitable spectra for the expansion of public resources is proposed, which balances the consideration of individual instruments and methods at individual laboratories and the comparability of the resulting data. As a result, it is hoped that the next few years will see an increase in the number of environmentally relevant spectra in (open) mass spectral libraries. The question of how to prioritize the compounds for acquisition is being addressed in other activities currently in progress in the HBM4EU project and NORMAN Network and will be communicated separately.

Glossary

Annotation/identification

Capability to assign a signal detected by non-targeted or suspect screening (i.e., a spectrometric descriptor) to a chemical with a given confidence level, by means of a reference library and/or structural elucidation work. Annotation is the act of linking a detected mass spectrometric feature with a chemical. Identification is the act of proving to be the same.

Non-targeted LC–MS

Analytical process for gathering comprehensive information on the composition of a sample. Workflows involve different steps of sample collection, sample preparation, data acquisition and data mining. The fraction of compounds accessible by a certain workflow depends on the characteristics of the individual steps applied. Data-dependent or data-independent acquisition techniques are employed for data acquisition. Detected features are characterized by retention time, MS, and, where possible, MS/MS information to enable annotation.

Targeted LC–MS

Analytical process for gathering specific information on the composition of a sample. Workflows involve different steps of sample collection, sample preparation, data acquisition and data mining. The steps were optimized for a preselected number of molecules. Often selected reaction monitoring techniques are employed for data acquisition. Furthermore, target screening usually involves a reference standard measured in-house under the same analytical conditions such that retention time, MS, and, where possible, MS/MS information is available for identification and confirmation.

Tandem mass spectrometry

Tandem mass spectrometry, also known as MS/MS or MS², involves multiple steps of mass spectrometry selection, with some form of fragmentation occurring in between the stages. Multiple stages of mass analysis separation can be accomplished with individual mass spectrometer elements separated in space or using a single mass spectrometer with the MS steps separated in time.

Target

A compound that is expected to be included in a sample and of which full mass spectrometric reference data, including MS/MS fragmentation, is available to enable annotation. The reference data is usually acquired with certified reference standards, and is stored in tandem mass spectral databases. The mass spectrometric data is often accompanied by metadata.

Suspect

A compound that is expected to be included in a sample. Typically, the available mass spectrometric data is incomplete and does not allow unequivocal annotation. Often, information on MS/MS fragmentation and retention time is missing or has only been predicted with computational tools.

Known

A detected signal that was annotated to a suspect or target at a certain confidence level.

Unknown

An unannotated signal.

Identification level

An approach for communicating identification confidence. A commonly used classification system in environmental research [2] includes five levels: exact

mass—unequivocal molecular formula—tentative candidate—probable structure—confirmed structure.

Tandem mass spectral database

An organized collection of tandem mass spectral data which comes bundled with a management system. The database management system is a software application that interacts with the user, other applications, and the database itself to capture and analyze data. Tandem mass spectra are typically acquired from certified reference compounds. Spectral information is processed prior to storage in a library.

Tandem mass spectral library

A curated and annotated collection of mass spectra acquired from certified reference compounds. Curation efforts may include manual inspection of mass spectra by experienced mass spectrometrists, noise and artifact removal, recalibration of spectra and peak annotations, as well as inter-library comparisons. The mass spectrometric data is often accompanied by metadata.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12302-020-00314-9>.

Additional file 1. Extended description of experimental settings.

Additional file 2. Reference compounds for producing tandem mass spectral libraries.

Abbreviations

AGC: Automatic gain control; amp: Average match probability; CID: Collision-induced dissociation; Eawag: Swiss Federal Institute of Aquatic Science and Technology; ESI: Electrospray ionization; FTICR: Fourier transform ion cyclotron resonance; GNPS: Global Natural Product Social Molecular Networking; HBM4EU: Human Biomonitoring for Europe project; HCD: Higher-energy collisional dissociation; HMDB: Human Metabolome Database; HRMS: High-resolution mass spectrometry; IT: Ion trap; LC: Liquid chromatography; LIT: Linear ion trap; m/z: Mass-to-charge ratio; MoNA: MassBank of North America; mp: Reference spectrum-specific match probability; MS: Mass spectrometry; MS/MS, MS2: Tandem mass spectrometry; NCE: Normalized collision energy; NORMAN: Network of reference laboratories, research centers and related organizations for monitoring of emerging environmental substances; Q: Quadrupole; QA: Quality assurance; QC: Quality control; QqTOF: Quadrupole–quadrupole time of flight instrument; ramp: Relative average match probability; ReSpec: RIKEN MSn spectral database for phytochemicals; TEDB: Toxic Exposome Database; UFZ: Helmholtz Centre for Environmental Research; WRTMD: Wiley Registry of Tandem Mass Spectra Database.

Acknowledgements

LD thanks Alyssa Bouville for her assisting in acquiring tandem mass spectral data.

Authors' contributions

HO, JPA, LD, MK, AC, FF, ML, ELS elaborated conception and design of this study. HO, MS, JPA, YG, LD, ELJ, TS, MK, AC, NCC, KR, AD, FF, ELS were involved in data acquisition and analysis. HO and ELS wrote a first version of the manuscript. All authors were involved in drafting and revising the manuscript. All authors read and approved the final manuscript.

Funding

The authors acknowledge financial support by the HBM4EU project. HBM4EU has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 733032. This text has been modified and extended for publication from Deliverable AD16.4 as part of WP16 within the HBM4EU project. ELS acknowledges funding from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006. MS acknowledges financial support by the University of Innsbruck.

Availability of data and materials

Datasets generated and analyzed during the current study will be available in MassBank (<https://massbank.eu/MassBank/>) in a HBM4EU_INTERLAB_2019 folder and for direct download via GitHub (https://github.com/MassBank/MassBank-data/tree/master/HBM4EU_INTERLAB_2019) The WRTMD is available from John Wiley & Sons, Inc. [53].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Muellerstrasse 44, 6020 Innsbruck, Austria. ² Institute of Analytical Chemistry and Radiochemistry, University of Innsbruck, CCB-Center of Chemistry and Biomedicine, Innrain 80-82, 6020 Innsbruck, Austria. ³ Laboratoire d'Étude des Résidus et Contaminants dans les Aliments (LABERCA), Oniris, INRAE, UMR 1329, 44307 Nantes, France. ⁴ TOXALIM (Research Centre in Food Toxicology), Toulouse University, INRAE UMR 1331, ENVT, INP-Purpan, Paul Sabatier University, 31027 Toulouse, France. ⁵ Metatoul-AXIOM Platform, National Infrastructure for Metabolomics and Fluxomics, MetaboHUB, Toxalim, INRAE UMR 1331, 31027 Toulouse, France. ⁶ UFZ – Helmholtz Centre for Environmental Research, Permoserstrasse 15, 04318 Leipzig, Germany. ⁷ Toxicological Center, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium. ⁸ Université Paris-Saclay, CEA, INRAE, Médicaments et Technologies pour la Santé (MTS), MetaboHUB, 91191 Gif-sur-Yvette, France. ⁹ Department Environment & Health, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands. ¹⁰ Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg.

Received: 26 December 2019 Accepted: 22 February 2020

Published online: 13 March 2020

References

- Hollender J, Schymanski EL, Singer HP, Ferguson PL (2017) Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ Sci Technol* 51:11505–11512. <https://doi.org/10.1021/acs.est.7b02184>
- Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, 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>
- Blaženović I, Kind T, Ji J, Fiehn O (2018) Software tools and approaches for compound identification of LC–MS/MS data in metabolomics. *Metabolites* 8:31. <https://doi.org/10.3390/metabo8020031>
- Schymanski EL, Ruttikies C, Krauss M, Brouard C, Kind T, Dührkop K, Allen F, Vaniya A, Verdegem D, Böcker S, Rousou J, Shen H, Tsugawa H, Sajed T, Fiehn O, Ghesquière B, Neumann S (2017) Critical assessment of small molecule identification 2016: automated methods. *J Cheminform* 9:22. <https://doi.org/10.1186/s13321-017-0207-1>
- Ruttikies C, Schymanski EL, Wolf S, Hollender J, Neumann S (2016) MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform* 8:3. <https://doi.org/10.1186/s13321-016-0115-9>
- Blaženović I, Kind T, Torbašinović H, Obrenović S, Mehta SS, Tsugawa H, Wermuth T, Schauer N, Jahn M, Biedendieck R, Jahn D, Fiehn O (2017)

- Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. *J Cheminform* 9:32. <https://doi.org/10.1186/s13321-017-0219-x>
7. Allen F, Pon A, Wilson M, Greiner R, Wishart D (2014) CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res* 42:W94–W99. <https://doi.org/10.1093/nar/gku436>
 8. Djoumbou-Feunang Y, Pon A, Karu N, Zheng J, Li C, Arndt D, Gautam M, Allen F, Wishart DS (2019) CFM-ID 3.0: Significantly improved ESI-MS/MS prediction and compound identification. *Metabolites* 9:72. <https://doi.org/10.3390/metabo9040072>
 9. Bade R, Bijlsma L, Miller TH, Barron LP, Sancho JV, Hernández F (2015) Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis. *Sci Total Environ* 538:934–941. <https://doi.org/10.1016/j.scitotenv.2015.08.078>
 10. Creek DJ, Jankevics A, Breiting R, Watson DG, Barrett MP, Burgess KEV (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/ac202182z>
 11. Stanstrup J, Neumann S, Vrhovšek U (2015) PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. *Anal Chem* 87:9421–9428. <https://doi.org/10.1021/acs.analchem.5b02287>
 12. Goryński K, Bojko B, Nowaczyk A, Bucirski A, Pawliszyn J, Kalisz R (2013) Quantitative structure–retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds. *Anal Chim Acta* 797:13–19. <https://doi.org/10.1016/j.jaca.2013.08.025>
 13. Schymanski EL, Williams AJ (2017) Open science for identifying “known unknown” chemicals. *Environ Sci Technol* 51:5357–5359. <https://doi.org/10.1021/acs.est.7b01908>
 14. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reilly MD, Thaden JJ, Viant MR (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>
 15. Stravs MA, Schymanski EL, Singer HP, Hollender J (2013) Automatic recalibration and processing of tandem mass spectra using formula annotation: recalibration and processing of MS/MS spectra. *J Mass Spectrom* 48:89–99. <https://doi.org/10.1002/jms.3131>
 16. Vainixa M, Schymanski EL, Neumann S, Navarro M, Salek RM, Yanes O (2016) Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects. *TrAC Trends Anal Chem* 78:23–35. <https://doi.org/10.1016/j.trac.2015.09.005>
 17. Frainay C, Schymanski E, Neumann S, Merlet B, Salek R, Jourdan F, Yanes O (2016) Mind the gap: mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered areas. *Metabolites* 8:51. <https://doi.org/10.3390/metabo8030051>
 18. Oberacher H, Arnhard K (2016) Current status of non-targeted liquid chromatography-tandem mass spectrometry in forensic toxicology. *TrAC Trends Anal Chem* 84:94–105. <https://doi.org/10.1016/j.trac.2015.12.019>
 19. Oberacher H, Arnhard K (2015) Compound identification in forensic toxicological analysis with untargeted LC–MS-based techniques. *Bioanalysis* 7:2825–2840. <https://doi.org/10.4155/bio.15.193>
 20. Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, Wohlgemuth G, Barupal DK, Showalter MR, 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>
 21. Milman BL, Zhurkovich IK (2016) Mass spectral libraries: a statistical review of the visible use. *TrAC Trends Anal Chem* 80:636–640. <https://doi.org/10.1016/j.trac.2016.04.024>
 22. Cooper BT, Yan X, Simón-Manso Y, Tchekhovskoi DV, Mirokhin YA, Stein SE (2019) Hybrid search: a method for identifying metabolites absent from tandem mass spectrometry libraries. *Anal Chem* 91(21):13924–13932. <https://doi.org/10.1021/acs.analchem.9b03415>
 23. 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>
 24. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, 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>
 25. Wallace WE, Ji W, Tchekhovskoi DV, Phinney KW, Stein SE (2017) Mass spectral library quality assurance by inter-library comparison. *J Am Soc Mass Spectrom* 28:733–738. <https://doi.org/10.1007/s13361-016-1589-4>
 26. Yang X, Neta P, Stein SE (2014) Quality control for building libraries from electrospray ionization tandem mass spectra. *Anal Chem* 86:6393–6400. <https://doi.org/10.1021/ac500711m>
 27. Oberacher H, Reinstadler V, Kreidl M, Stravs M, Hollender J, Schymanski E (2018) Annotating nontargeted LC-HRMS/MS data with two complementary tandem mass spectral libraries. *Metabolites* 9:3. <https://doi.org/10.3390/metabo9010003>
 28. Wishart DS, Feunang YD, Marcu A, Guo AC, 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>
 29. FiehnLab (2019) MassBank of North America. <http://mona.fiehnlab.ucdavis.edu/>. Accessed 14 Mar 2019
 30. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorn Dahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0—the Human Metabolome Database in 2013. *Nucleic Acids Res* 41:D801–D807. <https://doi.org/10.1093/nar/gks1065>
 31. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavalan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O’Neill EC, Briand E, Helfrich EJN, Granatovsky EA, Glukhov E, Ryyffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>
 32. Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, Hirai MY, Saito K (2012) RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* 82:38–45. <https://doi.org/10.1016/j.phytochem.2012.07.007>
 33. Lam H (2011) Building and Searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics* 10(R111):008565. <https://doi.org/10.1074/mcp.R111.008565>
 34. Oberacher H, Pavlic M, Libiseller K, Schubert B, Sulyok M, Schuhmacher R, Csaszar E, Köfeler HC (2009) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm: about an advanced search algorithm for tandem mass spectral reference libraries. *J Mass Spectrom* 44:494–502. <https://doi.org/10.1002/jms.1525>
 35. Pavlic M, Libiseller K, Oberacher H (2006) Combined use of ESI–QqTOF–MS and ESI–QqTOF–MS/MS with mass-spectral library search for qualitative analysis of drugs. *Anal Bioanal Chem* 386:69–82. <https://doi.org/10.1007/s00216-006-0634-8>

36. Oberacher H, Whitley G, Berger B, Weinmann W (2013) Testing an alternative search algorithm for compound identification with the 'Wiley Registry of Tandem Mass Spectral Data, MSforID': an alternative search algorithm for the Wiley Registry MSMS. *J Mass Spectrom* 48:497–504. <https://doi.org/10.1002/jms.3185>
37. Mylonas R, Mauron Y, Masselot A, Binz P-A, Budin N, Fathi M, Viette V, Hochstrasser DF, 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>
38. Nesvizhskii AI, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 4:787–797. <https://doi.org/10.1038/nmeth1088>
39. Scheubert K, Hufsky F, Petras D, Wang M, Nothias L-F, Dührkop K, Bandeira N, Dorrestein PC, Böcker S (2017) Significance estimation for large scale metabolomics annotations by spectral matching. *Nat Commun* 8:1494. <https://doi.org/10.1038/s41467-017-01318-5>
40. Ichou F, Schwarzenberg A, Lesage D, Alves S, Junot C, Machuron-Mandard X, Tabet J-C (2014) Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes: theoretical comparison of CID, PQD and HCD. *J Mass Spectrom* 49:498–508. <https://doi.org/10.1002/jms.3365>
41. NORMAN Network NORMAN Suspect List Exchange. <https://www.norman-network.com/nds/SLE/>. Accessed 9 Jun 2019
42. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM (2017) The CompTox chemistry dashboard: a community data resource for environmental chemistry. *J Cheminform* 9:61. <https://doi.org/10.1186/s13321-017-0247-6>
43. NORMAN Network (2019) NORMAN suspect list exchange database SusDat. <https://www.norman-network.com/nds/susdat/>. Accessed 15 Mar 2019
44. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 46:D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
45. Wishart D, Arndt D, Pon A, Sajed T, Guo AC, Djoumbou Y, Knox C, Wilson M, Liang Y, Grant J, Liu Y, Goldansaz SA, Rappaport SM (2015) T3DB: the toxic exposome database. *Nucleic Acids Res* 43:D928–D934. <https://doi.org/10.1093/nar/gku1004>
46. Neveu V, Moussy A, Rouaix H, Wedekind R, Pon A, Knox C, Wishart DS, Scalbert A (2017) Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res* 45:D979–D984. <https://doi.org/10.1093/nar/gkw980>
47. HighChem LLC (2019) mzCloud advanced mass spectral database. <https://www.mzcloud.org/>. Accessed 14 Mar 2019
48. NORMAN Network, MassBank Consortium (2019) MassBank EU: European MassBank (NORMAN MassBank). <https://massbank.eu/MassBank/>. Accessed 15 Mar 2019
49. Oberacher HM (2019) WRTMD or MSforID: Tandem mass spectral identification of small molecules. <https://msforid.com/>. Accessed 20 Dec 2019
50. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 47:D1102–D1109. <https://doi.org/10.1093/nar/gky1033>
51. Daylight Chemical Information Systems, Inc. (2008) SMILES—a simplified chemical language. <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>. Accessed 13 Apr 2019
52. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. *J Cheminform* 3:33. <https://doi.org/10.1186/1758-2946-3-33>
53. Oberacher HM (2011) The Wiley registry of tandem mass spectral data, MSforID, 1st edn. John Wiley & Sons, Hoboken
54. Peisl BYL, Schymanski EL, Wilmes P (2018) Dark matter in host-microbiome metabolomics: tackling the unknowns—a review. *Anal Chim Acta* 1037:13–27. <https://doi.org/10.1016/j.aca.2017.12.034>
55. Gaston L, Lapworth DJ, Stuart M, Arnscheidt J (2019) Prioritization approaches for substances of emerging concern in groundwater: a critical review. *Environ Sci Technol* 53:6107–6122. <https://doi.org/10.1021/acs.est.8b04490>
56. Götz CW, Stamm C, Fenner K, Singer H, Schärer M, Hollender J (2010) Targeting aquatic microcontaminants for monitoring: exposure categorization and application to the Swiss situation. *Environ Sci Pollut Res* 17:341–354. <https://doi.org/10.1007/s11356-009-0167-8>
57. Little JL, Clevon CD, Brown SD (2011) Identification of “known unknowns” utilizing accurate mass data and chemical abstracts service databases. *J Am Soc Mass Spectrom* 22:348–359. <https://doi.org/10.1007/s13361-010-0034-3>
58. Oberacher H, Pitterl F, Siapi E, Steele BR, Letzel T, Grosse S, Poschner B, Tagliaro F, Gottardo R, Chacko SA, Josephs JL (2012) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library. 3. Focus on ion trap and upfront CID: on the transferability of a tandem mass spectral reference library. *J Mass Spectrom* 47:263–270. <https://doi.org/10.1002/jms.2961>
59. Boyd RK, Basic C, Bethem RA (2013) Trace quantitative analysis by mass spectrometry. Wiley, Hoboken
60. McEachran AD, Mansouri K, Grulke C, Schymanski EL, Ruttkies C, Williams AJ (2018) “MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies. *J Cheminform* 10:45. <https://doi.org/10.1186/s13321-018-0299-2>
61. Damont A, Olivier M-F, Warnet A, Lyan B, Pujos-Guillot E, Jamin EL, Debrauwer L, Bernillon S, Junot C, Tabet J-C, Fenaille F (2019) Proposal for a chemically consistent way to annotate ions arising from the analysis of reference compounds under ESI conditions: a prerequisite to proper mass spectral database constitution in metabolomics. *J Mass Spectrom* 54:567–582. <https://doi.org/10.1002/jms.4372>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries

Herbert Oberacher¹, Michael Sasse², Jean-Philippe Antignac³, Yann Guitton³, Laurent Debrauwer^{4,5}, Emilien L. Jamin^{4,5}, Tobias Schulze⁶, Martin Krauss⁶, Adrian Covaci⁷, Noelia Caballero-Casero⁷, Kathleen Rousseau⁸, Annelaure Damont⁸, François Fenaille⁸, Marja Lamoree⁹, Emma L. Schymanski¹⁰

¹ Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Mullerstrasse 44, 6020 Innsbruck, Austria

² Institute of Analytical Chemistry and Radiochemistry, University of Innsbruck, CCB-Center of Chemistry and Biomedicine, Innrain 80-82, 6020 Innsbruck, Austria

³ Laboratoire d'Étude des Résidus et Contaminants dans les Aliments (LABERCA), Oniris, INRAE, UMR 1329, 44307, Nantes, France

⁴ TOXALIM (Research Centre in Food Toxicology), Toulouse University, INRAE UMR 1331, ENVT, INP-Purpan, Paul Sabatier University, 31027 Toulouse, France

⁵ Metatoul-AXIOM platform, National Infrastructure for Metabolomics and Fluxomics, MetaboHUB, Toxalim, INRAE UMR 1331, 31027 Toulouse, France

⁶ UFZ – Helmholtz Centre for Environmental Research, Permoserstrasse 15, D-04318 Leipzig, Germany

⁷ Toxicological Center, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

⁸ Université Paris-Saclay, CEA, INRAE, Médicaments et Technologies pour la Santé (MTS), MetaboHUB, 91191 Gif-sur-Yvette, France

⁹ Vrije Universiteit, Department Environment & Health, De Boelelaan 1087, 1081 HV Amsterdam, the Netherlands

¹⁰ Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367 Belvaux, Luxembourg

Experimental settings used in the laboratories participating to the interlaboratory study

In Laboratory 1, tandem mass spectra were acquired on an Orbitrap Fusion Tribrid mass spectrometer coupled with a Dionex Ultimate 3000 liquid chromatography system (Thermo Fisher Scientific). Compounds of interest were solubilized at 1 $\mu\text{g/mL}$ and analyzed by flow injection analysis at a flow rate of 100 $\mu\text{L/min}$ using water containing 0.1% formic acid as solvent A and acetonitrile containing 0.1% formic acid as solvent B. A 95/5 (A/B) mixture was used as mobile phase during 3 minutes, while 10 μL of sample have been injected. ESI source parameters were as follows: spray voltage, +3500 V; sheath gas pressure, 60 arbitrary units; auxiliary gas pressure, 12 arbitrary units; sweep gas pressure, 3 arbitrary units. The temperatures of the ion transfer tube and the vaporizer were set at 325°C and at 350°C, respectively. For MS acquisition, the Orbitrap mass analyser was operated at a resolution of 240,000 at m/z 200 (full width at half maximum) following quadrupole isolation. The mass range was m/z 85-1000 and the automatic gain control (AGC) target fixed at 5×10^4 . MS/MS acquisition was performed in the targeted mode at a resolution of 30,000, with an isolation window of 0.8 Da and an AGC target of 5×10^4 . Precursor ions were fragmented under HCD conditions using normalized collision energy (NCE) set at 10, 20, 40 and 80%. To guarantee an efficient transmission of the whole ion range in the MS/MS mode, the scan range started at m/z 50-90 depending on the precursor ion mass ($\sim 1/5$ of the m/z of the precursor ion). Obviously, this may lead to the loss of some low-mass diagnostic fragment ions impairing good matching of MS/MS spectra. The instrument was controlled using the Orbitrap Fusion Tune Application version 3.1.2412.25, while raw data files were manually processed using Xcalibur version 4.1.50 (Thermo Fisher Scientific).

In Laboratory 2, tandem mass spectra were acquired as mixtures (3 mixes total) on a QExactive, equipped with a heated electrospray II source, and coupled with a Dionex Ultimate 3000 liquid chromatography system (Thermo Fisher Scientific). Separation was performed on a reversed-phase column (Hypersil GOLD 1.9 μm , 2.1x100 mm, from Thermo Scientific) with the program as follows (A/B): 95/5 from 0 to 2.40 min; 75/25 at 4.50 min; 30/70 at 11 min; 0/100 at 14 min to 16.5 min; 95/5 at 19 min to 25 min. The injected volume was 5 μL and the flow rate was 400 $\mu\text{L/min}$. Solvent A was water with 0.1 % acetic acid, Solvent B was acetonitrile with 0.1% acetic acid. MS was acquired in positive mode with source parameters set to: spray voltage ± 3 kV; capillary temperature 350°C; sheath gas flow rate 55 a.u. (arbitrary units), auxiliary gas flow rate 10 a.u., spare gas flow rate 0; max spray current ± 100 μA ; probe heater temperature 300°C; S-lens RF level 50. MS was acquired at a resolution of 70,000 (m/z range 65-975) and target MS/MS at a resolution of 17,500 with an isolation window of m/z 1

and an AGC target of 1×10^5 . The instrument was calibrated in advance according to the manufacturer's instructions. The mixtures of standards were injected several times to obtain data on multiple collision energies (HCD from 15% to 90% by 5% NCE steps; one NCE per injection). Data processing and MS/MS spectrum extraction was performed manually and spectra exported to tabular files.

In Laboratory 3, tandem mass spectra were acquired as mixtures on a QExactive Plus, coupled with a Dionex Ultimate 3000 liquid chromatography system (Thermo Fisher Scientific). Compounds of interest were solubilized at 500 ng/mL. Separation was performed on a reversed-phase column (Acquity UPLC BEH C₁₈ 1.7 μ m, 2.1x150 mm, from Waters) with the program as follows (A/B): 90/10 at 0 min, 90/10 at 2 min, 0/100 at 15 min, 0/100 at 20 min, 90/10 at 21 min, 90/10 at 30 min at a flow rate of 200 μ L/min. Solvent A was water with 0.1 % formic acid, Solvent B was methanol. MS was acquired at a resolution of 70,000 (m/z range 60-900), MS/MS as data-dependent acquisition at a resolution of 17,500 with an isolation window of m/z 1 and an AGC target of 5×10^5 . The instrument was calibrated in advance according to the manufacturer's instructions. The mixture was injected several times to obtain data on multiple collision energies (HCD 15%, 30%, 45%, 60%, 75% and 90% NCE; one per injection). Data processing was performed using an *in house* approach "shinysscreen", which is based on pre-screening methods from the ReSOLUTION package [1] and tandem mass spectral processing from the RMassBank package [2]. The spectra passing quality control were provided in MassBank record format.

In Laboratory 4, tandem mass spectra were acquired on an LTQ-Orbitrap XL hybrid mass spectrometer coupled with a Dionex RSLC 3000 liquid chromatography system (Thermo Fisher Scientific). The instrument was set-up and calibrated according to internal laboratory procedures. Compounds of interest were solubilized at 1 mg/mL in methanol and then diluted to 1 μ g/mL in mobile phase A for analysis. The chromatographic separation was achieved on a Hypersil GOLD C18 (100x2.1 mm, 1.9 μ m) column (Thermo Fisher Scientific), maintained at 40°C and operated at a flow rate of 0.3 mL/min with the following gradient program: 0 min 100% of A, linear gradient from 0 to 100% of B in 30 min, then 5 min 100% of B. Mobile phases were composed of (A) water/methanol/acetic acid 95/5/0.1 (v:v:v) and (B) methanol/acetic acid 100/0.1 (v:v). The electrospray ionization (ESI) source was used in the positive mode with the following parameters: spray voltage, 4.0 kV, sheath gas flow rate, 30 a.u., auxiliary gas flow rate, 10 a.u. and capillary temperature, 300°C. High resolution MS/MS spectra were acquired at a resolution of 7,500 from m/z 50 to 10 m/z over the precursor. Spectra were recorded in three separate injections at multiple collision energies (5% to 80% by 5% NCE steps; one NCE per injection) in the HCD cell of the LTQ-Orbitrap XL instrument. Data were acquired and

processed using Xcalibur (version 2.0.7) from Thermo Fisher.

In Laboratory 5, tandem mass spectra were acquired on a LTQ Orbitrap XL instrument with a heated electrospray source coupled to an Ultimate 3000 LC system (both from Thermo Fisher Scientific). Compounds were dissolved in water:methanol 95:5 at 10 µg/L and separated using a Kinetex C₁₈ EVO column (50 x 2.1 mm, 2.6 µm particle size, Phenomenex), equipped with a pre-column of the same type (5 x 2.1 mm) and an in-line filter. The injection volume was 25 µL. Gradient elution was conducted using water and methanol, both containing 0.1% (v/v) of formic acid. MS and MS/MS spectra were acquired at a nominal resolving power of 30,000 and 15,000, respectively. The mixture was injected altogether six times to obtain MS/MS spectra for HCD (10% to 180% by 10% NCE steps; one NCE per injection) and CID (15% to 125% by 10% steps; one energy per injection) fragmentation. Data-dependent acquisition was used with the precursor ion *m/z* in an inclusion list, an isolation width of *m/z* 1.3 and an AGC target of 5x10⁵. External mass calibration was conducted on each day prior to analysis using the standard mixture recommended by the vendor. The RMassbank package was used [2] to extract and generate MS/MS spectra.

In Laboratory 6, tandem mass spectra were acquired on a quadrupole time-of-flight instrument (6530 QTOF from Agilent). The mass spectrometer was operated in positive ESI mode using Agilent Jet-Stream ion source and 2GHz (extended dynamic range) mode, providing a full width at half maximum resolution of approximately 3600 at *m/z* 118.0862 and 12000 at *m/z* 922.0095, and automatically recalibrated using a calibration solution delivered via a calibration delivery system (Agilent). Drying gas temperature and flow rate were set at 300 °C and 10 L/min, respectively. Sheath gas temperature and flow rate were adjusted at 350 °C and 12 L/min. Nebulizer pressure was set at 45 psig. Capillary, nozzle and fragmentor voltages were set at 4000, 0 and 120 V, respectively. Test compounds were solubilized in methanol at 1.0 µg/mL and introduced into the mass spectrometer by using an Agilent 1290 Infinity UPLC chromatographic system without analytical column. Mobile phase composition was water and methanol 50:50 (v/v) and flow rate 0.3 mL/min. Volume of injection was set at 1 µL. The acquisition parameters were set for *m/z* range from 100 to 600 at a scan rate of 4 scans/s and 6 scans/s for MS and MS/MS spectra, respectively. Collision energies were applied at the 5-80 eV range (5 eV steps). The software used to perform the analysis was LC/MS Data Acquisition MassHunter B.05.01 (Agilent).

In Laboratory 7, tandem mass spectra were acquired on a TripleTOF 5600+ (Sciex). The mass spectrometer was operated in positive ESI mode using a DuoSpray ion source. The spray voltage was set to 5.5 kV. Gas flows of 40 arbitrary units for the nebulizer gas and 30 arbitrary

units for the turbo gas were employed. The temperature of the turbo gas was adjusted to 200 °C. The instrument was operated at a mass resolution of ~30,000 for MS and ~15,000 for MS/MS, and automatically recalibrated using a calibration solution delivered via a calibration delivery system (Sciex). Test compounds were dissolved in 0.1% aqueous acetic acid solution containing 50% (v/v) acetonitrile. Sample solutions of 1.0 µg/mL were directly introduced into the ESI source at a flow rate of 20 µL/min. The scan range was m/z 50-700 for the acquisition of MS/MS spectra. CID was accomplished at 5 eV to 50 eV by 5 eV steps (one energy per injection). The accumulation time for a single spectrum was set to 1.0 s. To generate an average spectrum, 60 spectra were used. The instrument was controlled by the Analyst TF 1.6 software (Sciex).

Tandem mass spectral library search

The library search and benchmarking was accomplished using the 'MSforID Search' [3,4]. The search algorithm determines the similarity between a sample spectrum and library spectra. The estimation of similarity starts with the identification of fragment ions that are present in both of the spectra being compared. These ions are called "matching fragments." The spectral information retrieved is used to calculate the "reference spectrum specific match probability" (*mp*). If the mass spectral library contains multiple spectra per reference compound, multiple *mp*-values per reference compound are obtained. To combine all these compound specific *mp*-values to one value that specifies the similarity between the unknown and the specific reference compound, the compound specific *mp*-values are averaged and normalized to yield the compound specific "average match probability" (*amp*) and "relative average match probability" (*ramp*), respectively. These values range between 0 and 100. High compound-specific *amp*- and *ramp*-values indicate high similarity between the unknown and the reference compound. The substance with the highest *amp*- and *ramp*-value is considered to be the best match to the unknown compound.

Automated MSforID Search was performed with a program written in Pascal using Delphi 6 for Windows (Borland Software Corporation, Scotts Valley, CA, USA; now Embarcadero Technologies, Inc., San Francisco, CA, USA) using the following search parameters: mass-to-charge ratio (*m/z*) tolerance of ± 0.01 , intensity cut-off factor of 0.01. The following criteria were used to classify obtained search results as tentatively correct positive results: precursor ion mass tolerance of ± 0.01 , *amp* > 5.0 and *ramp* > 30-50. The thresholds were determined using quality tests and represent a compromise between sensitivity and specificity [3, 4]. The tentative identifications were checked for accuracy by expert reviewing, which included visual inspection and comparison of tandem mass spectral data.

References

1. Schymanski EL (2019) ReSOLUTION: SOLUTIONS for High ReSOLUTION Mass Spectrometry. <https://github.com/schymane/ReSOLUTION>. Accessed 20 Dec 2019.
2. Stravs MA, Schymanski EL, Singer HP, Hollender J (2013) Automatic recalibration and processing of tandem mass spectra using formula annotation: Recalibration and processing of MS/MS spectra. *Journal of Mass Spectrometry* 48:89–99. <https://doi.org/10.1002/jms.3131>
3. Oberacher H, Pavlic M, Libiseller K, Schubert B, Sulyok M, Schuhmacher R, Csaszar E, Köfeler HC (2009) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm: About an advanced search algorithm for tandem mass spectral reference libraries. *Journal of Mass Spectrometry* 44:494–502. <https://doi.org/10.1002/jms.1525>
4. Pavlic M, Libiseller K, Oberacher H (2006) Combined use of ESI–QqTOF-MS and ESI–QqTOF-MS/MS with mass-spectral library search for qualitative analysis of drugs. *Analytical and Bioanalytical Chemistry* 386:69–82. <https://doi.org/10.1007/s00216-006-0634-8>

A European Proposal for Quality Control and Quality Assurance of Tandem Mass Spectral Libraries

Herbert Oberacher, Michael Sasse, Jean-Philippe Artignac, Yann Guitton, Laurent Debrauwer, Emilien L. Jamin, Tobias Schulte, Martin Krauss, Adrian Covaci, Noelia Caballero-Casero, Kathleen Rousseau, Annelaure Damont, François Fenaille, Marja Lamoree, Emma L. Schyns

Name	CompTox_DTXSID	CASRN	InChIkey
Amiloride	DTXSID9043853	2609-46-3	XSDOTBWBPRPKA-UHFFFAOYSA-N
Phenazone	DTXSID6021117	60-80-0	VEGOALMAJBPNY-UHFFFAOYSA-N
Cyclizine	DTXSID4022864	82-92-8	UVKZSORRUEBAZ-UHFFFAOYSA-N
Desipramine	DTXSID6022896	50-47-5	HCAVAFALSYZDH-UHFFFAOYSA-N
Dibucaine	DTXSID3045271	85-79-0	PURQVATATUTTEAL-UHFFFAOYSA-N
Dixyrazine	DTXSID10947639	2470-73-7	MISYUWPGNGDNTQ-UHFFFAOYSA-N
Dothiepin	DTXSID2022961	113-53-1	PHTUQLWOUWZIMZ-UHFFFAOYSA-N
Ethambutol	DTXSID8023006	74-55-5	AEUTYOVWVDBAKS-UWVGGRQHSAN
Etofliline	DTXSID5023031	519-37-9	NWPRCROWMGIBOT-UHFFFAOYSA-N
Metfruside	DTXSID0048844	7195-27-9	SMNDEBSLNYGGOU-UHFFFAOYSA-N
Metoclopramide	DTXSID6045169	364-62-5	TTWJBBZEZQIBI-UHFFFAOYSA-N
Nylidrin	DTXSID9023461	447-41-6	PTGXALUBQBSGPKF-UHFFFAOYSA-N
Phentermine	DTXSID9023461	122-09-8	DHVAAGZRIUJOJK-UHFFFAOYSA-N
Sulfamethoxazole	DTXSID8025064	723-46-6	ILKIGFTWXXRPMT-UHFFFAOYSA-N
Sulfamoxole	DTXSID5023617	729-99-7	CYELXLSBQBMFT-UHFFFAOYSA-N

Name	IUPAC_Name	SMILES
Amiloride	3,5-Diamino-N-carbamimidoyl-6-chloropyrazine-2-carboxamide	NC(=N)NC(=O)C1=NC(C)=C(N)N=C1N
Phenazone	1,5-Dimethyl-2-phenyl-1,2-dihydro-3H-pyrazol-3-one	CN1NC(C)C=C1O1C1=CC=CC=C1
Cyclizine	1-(Di(phenylmethyl)-4-methylpiperazine	CN1CCN(C)C1(C)C=C(C)C=C1C1=CC=CC=C1
Desipramine	3-(10,11-Dihydro-5H-dibenzol[b][f]azepin-5-yl)-N-methylpropan-1-amine	CN(C)CCN(C)C1(C)C=C(C)C=C1C1=CC=CC=C1
Dibucaine	2-Butoxy-N-[2-(diethylamino)ethyl]quinoline-4-carboxamide	CCCCOC1=CC(C)O1NCCN(C)C1(C)C=C(C)C=C1
Dixyrazine	2-[2-(4-(2-methyl-3-phenothiazin-10-ylpropyl)piperazin-1-yl)ethoxy]ethanol	CC(C)N1CCN(C)C1(C)C=C(C)C=C1C1=CC=CC=C1
Dothiepin	3-(Dibenzol[b,e]thiepin-11(6H)-ylidene)-N,N-dimethylpropan-1-amine	CN(C)C(=C)C1=CC=C(C)C=C1C1=CC=CC=C1
Ethambutol	(2S,2'S)-2,2'-(ethane-1,2-diyldiazanediyl)di(butan-1-ol)	CC(C)C@H(CO)NCCN(C)C@H(C)C(C)CO
Etofliline	7-(2-Hydroxyethyl)-1,3-dimethyl-3,7-dihydro-1H-purine-2,6-dione	CN1C2=C(N)C(CO)C=N2(C)C(=O)N1C(C)C=O
Metfruside	4-Chloro-N-(1'-methyl-N'-1'-[2-methyl-2-ylamino]propyl)phenol	CN(C)C(=O)C1=CC=C(C)C=C1C1=CC=CC=C1
Metoclopramide	4-Amino-5-chloro-N-[2-(diethylamino)ethyl]-2-methoxybenzamide	CC(C)C(=O)C1=CC=C(C)C=C1C1=CC=CC=C1
Nylidrin	4-[1-Hydroxy-2-(4-phenylbutan-2-ylamino)propyl]phenol	CC(C)C(=O)C1=CC=C(C)C=C1C1=CC=CC=C1
Phentermine	2-Methyl-1-phenylpropan-2-amine	CC(C)N(C)C1=CC=CC=C1
Sulfamethoxazole	4-Amino-N-(5-methyl-1,2-oxazol-3-yl)benzene-1-sulfonamide	CC1=CC(N)C(=O)C=C1C(=O)N1S(=O)(=O)N=C1N
Sulfamoxole	4-Amino-N-(4,5-dimethyl-1,3-oxazol-2-yl)benzene-1-sulfonamide	CC1=CC(N)C(=O)C=C1C(=O)N1S(=O)(=O)N=C1N

Name	InChI	Molecular_Formula	Monoisotopic_Mass
Amiloride	InChI=1S/C6H8ClN7O/c7-2-4(9)13-3(8)11(12-2)5(15)14-6(10)11/h(4,8,9,13)(H4,10,11,14,15)	C6H8ClN7O	229.04789
Phenazone	InChI=1S/C11H12N2O/c1-9-8-11(14)13(12)9)10-6-4-3-5-7-10/h3-8H,1-2H3	C11H12N2O	188.09496
Cyclizine	InChI=1S/C18H22N2/c1-19-12-14-20(15-13-19)18(16-8-4-2-5-9-16)17-10-6-3-7-11-17/h2-11,18H,12-15H2,1H3	C18H22N2	266.17830
Desipramine	InChI=1S/C18H22N2/c1-19-13-6-14-20-17-9-4-2-7-15(17)11-12-16-8-3-5-10-18(16)20/h2-5,7-10,19H,6,11-14H2,1H3	C18H22N2	266.17830
Dibucaine	InChI=1S/C20H29N3O2/c1-4-7-14-25-19-15-17(16-10-8-9-11-18(16)22-19)20(24)12-13-23(5-2)6-3/h8-11,15H,4-7,16H2,1-3H3,(H,17,19)	C20H29N3O2	343.22598
Dixyrazine	InChI=1S/C24H33N3O2S/c1-20(18-26-12-10-25(11-13-26)14-16-29-17-15-28)19-27-21-6-2-4-8-23(21)30-24-9-5-7-22(24)27/h2-9,20,28H,10-19H2,1H3	C24H33N3O2S	427.22935
Dothiepin	InChI=1S/C19H21NS/c1-20(2)13-7-11-17-16-9-4-3-8-15(16)14-21-19-12-6-5-10-18(17)19/h3-6,8-12H,7,13-14H2,1-2H3	C19H21NS	295.13947
Ethambutol	InChI=1S/C9H12N4O3/c1-3-9(7-13)11-5-6-12-10(4-2)8-14/h9-14H,3-8H2,1-2H3/9-10/mo/51	C9H12N4O3	204.18378
Etofliline	InChI=1S/C9H12N4O3/c1-11-7-6(8(15)12(2)9(11)16)13(8-4-14)5-10-7/h5,14H,3-4H2,1-2H3	C9H12N4O3	224.09094
Metfruside	InChI=1S/C14H22ClN2O2/c1-3-9(7-13)11-5-6-12-10(4-2)8-14/h9-14H,3-8H2,1-2H3	C14H22ClN2O2	282.04239
Metoclopramide	InChI=1S/C14H22ClN2O2/c1-3-9(7-13)11-5-6-12-10(4-2)8-14/h9-14H,3-8H2,1-2H3	C14H22ClN2O2	382.04239
Nylidrin	InChI=1S/C10H15N/c1-10(2-11)9(10-8-11(15)12(16)9-13(10)20-3)8-9H,4-7,16H2,1-3H3,(H,17,19)	C10H15N	149.12045
Phentermine	InChI=1S/C10H15N/c1-10(2-11)9(10-8-11(15)12(16)9-13(10)20-3)8-9H,4-7,16H2,1-3H3,(H,17,19)	C10H15N	149.12045
Sulfamethoxazole	InChI=1S/C10H11N3O3S/c1-7-6-10(12-16-7)13-17(14,15)9-4-2-8(11)3-5-9/h2-6H,11H2,1H3,(H,12,13)	C10H11N3O3S	253.05211
Sulfamoxole	InChI=1S/C11H13N3O3S/c1-7-8(2)17-11(13-7)14-18(15,16)10-5-3-9(12)4-6-10(10)3-6H,12H2,1-2H3,(H,13,14)	C11H13N3O3S	267.06776

*Annexe 4 : Détails conditions d'échange H/D
par la réaction mono-étape métallo-
catalysée*

Partie C : Annexes – Annexe 4

Condition de Synthèse	Nombre de Composés Testés
Pd/C - 110°C - pH10 - 48H - 10 mol%	10
Pd/C - 110°C - pH7 - 48H - 10 mol%	9
Ru/C - 40°C - pH7 - 24H - 5 mol%	18
Ru/C - 50°C - pH10 - 24H - 15 mol%	1
Ru/C - 50°C - pH7 - 24H - 15 mol%	7
Ru/C - 50°C - pH7 - 24H - 5 mol%	8
Ru/C - 50°C - pH7 - 48H - 5 mol%	8
Ru/C - 55°C - pH 10 - 24H - 15 mol%	1
Ru/C - 55°C - pH 10 - 96H - 15 mol%	1
Ru/C - 55°C - pH 10 - 96H - 20 mol%	3
Ru/C - 55°C - pH 7 - 96H - 15 mol%	1
Ru/C - 55°C - pH 7 - 96H - 20 mol %	1
Ru/C - 55°C - pH 7 - 96H - 5 mol%	18
Ru/C - 55°C - pH10 - 144H - 5 mol%	18
Ru/C - 55°C - pH10 - 24H - 10 mol%	3
Ru/C - 55°C - pH10 - 24H - 15 mol%	5
Ru/C - 55°C - pH10 - 24H - 20 mol%	11
Ru/C - 55°C - pH10 - 48H - 10 mol%	2
Ru/C - 55°C - pH10 - 48H - 15 mol%	1
Ru/C - 55°C - pH10 - 48H - 20 mol%	3
Ru/C - 55°C - pH10 - 48H - 5 mol%	3
Ru/C - 55°C - pH10 - 96H - 15 mol%	4
Ru/C - 55°C - pH10 - 96H - 20 mol%	7
Ru/C - 55°C - pH7 - 24H - 10 mol%	4
Ru/C - 55°C - pH7 - 24H - 15 mol%	19
Ru/C - 55°C - pH7 - 24H - 20 mol%	11
Ru/C - 55°C - pH7 - 24H - 5 mol%	9
Ru/C - 55°C - pH7 - 24H - 8 mol%	8
Ru/C - 55°C - pH7 - 48H - 10 mol%	2
Ru/C - 55°C - pH7 - 48H - 15 mol%	13
Ru/C - 55°C - pH7 - 48H - 20 mol%	5
Ru/C - 55°C - pH7 - 48H - 5 mol%	8
Ru/C - 55°C - pH7 - 96H - 15 mol%	4
Ru/C - 55°C - pH7 - 96H - 20 mol%	6
Ru/C - 55°C - pH 7 - 96H - 5 mol%	18
Ru/C - 60°C - pH10 - 96H - 5 mol%	1
Ru/C - 60°C - pH7 - 96H - 5 mol%	6
Ru/C - 80°C - pH10 - 96H - 15 mol%	1
Ru/C - 80°C - pH7 - 96H - 15 mol%	6
Ru/C - Pd/C - 110°C - pH10 - 144H - 10 mol%	4
Ru/C - Pd/C - 110°C - pH10 - 144H - 15 mol%	3
Ru/C - Pd/C - 110°C - pH10 - 144H - 20 mol%	1
Ru/C - Pd/C - 110°C - pH7 - 144H - 10 mol%	5
Ru/C - Pd/C - 110°C - pH7 - 144H - 15 mol%	3
Ru/C - Pd/C - 110°C - pH7 - 144H - 20 mol%	2
Ru/C - Pd/C - 80°C - pH10 - 192H - 10 mol%	17

*Annexe 5 : Données supplémentaires sur la
publication de la nouvelle méthode de
quantification*

« High-throughput metal-catalyzed deutérium labeling
of small molecules for MS-based quantitative
metabolomics and evaluation of an innovative
multipoint calibration approach »

Supplementary data

High-throughput metal-catalyzed deutérium labeling of small molecules for MS-based quantitative metabolomics and evaluation of an innovative multipoint calibration approach.

Kathleen Rousseau,¹ Jessica Michieletto,¹ Yu Min Kiw,¹ Sophie Feuillastre,² Grégory Pieters,² Christophe Junot,¹ François Fenaille,¹ Annelaure Damont.¹

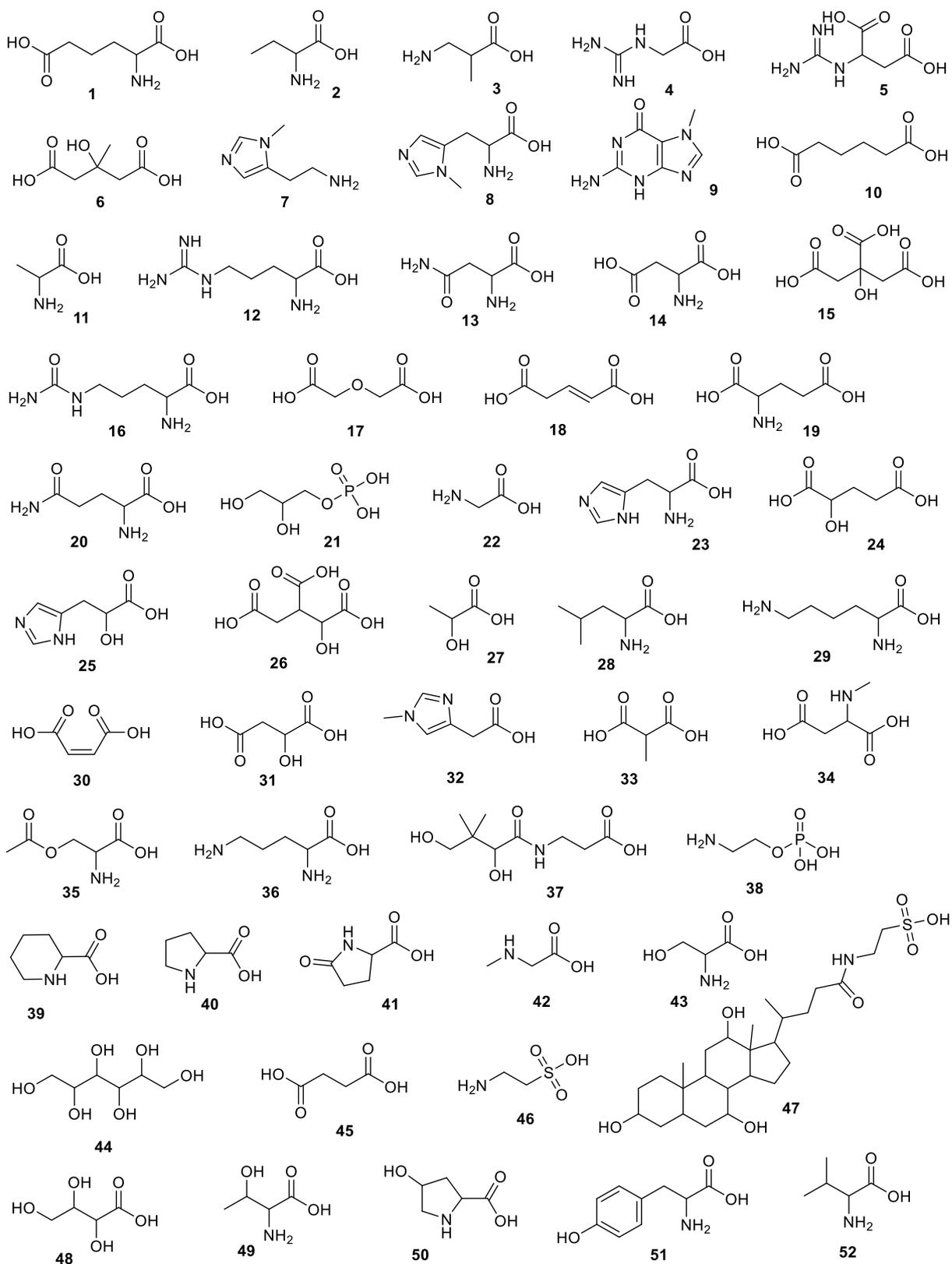
¹Service de Pharmacologie et d'Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, MetaboHUB, 91191 Gif-sur-Yvette, France

²Service de chimie bio-organique et de marquage, Laboratoire de Marquage au Tritium, CEA, Université Paris Saclay, 91191 Gif-sur-Yvette, France

Supplementary Data S1: Compounds tested and selected labeling conditions

Compound	N°	Formula	Accurate mas	Smiles	Chemical class	Selected labeling conditions
α-Aminoadipic Acid	1	C6H11NO4	161.0688	O=C(O)C(N)CCCC(O)=O	Amino acid	Ru/C - 55°C - 24h
2-Aminobutyric Acid	2	C4H9NO2	103.0633	CCC(C(O)=O)N	Amino acid	Ru/C - 55°C - 24h
3-Amino-isobutyric Acid	3	C4H9NO2	103.0633	CC(C(O)=O)CN	Amino acid	Ru/C - 55°C - 24h
Guanidinoacetic Acid	4	C3H7N3O2	117.0538	N=C(N)C(O)=O	Organic acid	Ru/C - 55°C - 24h
Guanidinosuccinic Acid	5	C5H9N3O4	175.0593	N=C(N)NC(C(O)=O)C(O)=O	Organic acid	Ru/C - 55°C - 24h
3-Hydroxy-3-Methylglutaric acid	6	C6H10O5	162.0528	CC(C(O)=O)(O)CC(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
3-Methylhistamine	7	C6H11N3	125.0953	CN1C=NC=C1CCN	Organic amine	Ru/C - 55°C - 24h
3-Methylhistidine	8	C7H11N3O2	169.0851	CN1C=NC=C1CC(C(O)=O)N	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h
7-Methylguanine	9	C6H7N5O	165.0651	CN1C=NC(NC(N)=N2)=C1C2=O	Heteroaromatic	Ru/C - NaOD - 55°C - 24h
Adipic Acid	10	C6H10O4	146.0579	O=C(O)CCCC(O)=O	Organic acid	Pd/C - 80°C - 48h
Alanine	11	C3H7NO2	89.0477	CC(C(O)=O)N	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h
Arginine	12	C6H14N4O2	174.1117	NC(CCCN(N)=N)C(O)=O	Amino acid	Ru/C - 55°C - 24h
Asparagine	13	C4H8N2O3	132.0535	NC(C(N)=O)C(O)=O	Amino acid	Ru/C - 55°C - 24h
Aspartic Acid	14	C4H7NO4	133.0375	NC(C(O)=O)C(O)=O	Amino acid	Ru/C - 55°C - 24h
Citric Acid	15	C6H8O7	192.0270	O=C(O)CC(C(O)=O)(O)CC(O)=O	Organic acid	Pd/C - 80°C - 48h
Citrulline	16	C6H13N3O3	175.0957	NC(CCCN(N)=O)C(O)=O	Amino acid	Ru/C - 55°C - 72h
Diglycolic Acid	17	C4H6O5	134.0215	O=C(O)COCC(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
Glutaconic Acid	18	C5H6O4	130.0266	O=C(O)C=C(C)C(O)=O	Organic acid	NA - "back exchange"
Glutamic Acid	19	C5H9NO4	147.0532	NC(CCC(O)=O)C(O)=O	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h
Glutamine	20	C5H10N2O3	146.0691	NC(C(O)=O)CCC(N)=O	Amino acid	Ru/C - 55°C - 24h
Glycerol-3-Phosphate	21	C3H9O6P	172.0137	OCC(COP(O)(O)=O)O	Polyol & derivative	Ru/C - NaOD - 55°C - 24h
Glycine	22	C2H5NO2	75.0320	NCC(O)=O	Amino acid	Ru/C - NaOD - 55°C - 24h
Histidine	23	C6H9N3O2	155.0695	NC(CC1=CN=CN1)C(O)=O	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h
Hydroxyglutaric Acid	24	C5H8O5	148.0372	C(C(=O)O)C(C(=O)O)O	Organic acid	Ru/C - 55°C - 48h
Imidazolelactic Acid	25	C6H8N2O3	156.0535	OC(C(O)=O)CC1=CN=CN1	Organic acid	Ru/C + Pd/C - NaOD - 80°C - 48h
Isoctric Acid	26	C6H8O7	192.0270	O=C(O)CC(C(O)=O)C(O)C(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
Lactic Acid	27	C3H6O3	90.0317	CC(O)C(O)=O	Organic acid	Ru/C - 55°C - 24h
Leucine	28	C6H13NO2	131.0946	CC(C)CC(C(O)=O)N	Amino acid	Ru/C + Pd/C - 80°C - 48h
Lysine	29	C6H14N2O2	146.1055	NCCCC(N)C(O)=O	Amino acid	Ru/C - 55°C - 48h
Maleic Acid	30	C4H4O4	116.0110	O=C(O)C=C(C)C(O)=O	Organic acid	Pd/C - 80°C - 48h
Malic Acid	31	C4H6O5	134.0215	OC(C(O)=O)C(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
Methylimidazoleacetic Acid	32	C6H8N2O2	140.0586	CN(C=N1)C=C1CC(O)=O	Organic acid	Ru/C + Pd/C - NaOD - 80°C - 48h
Methylmalonic Acid	33	C4H6O4	118.0266	CC(C(O)=O)C(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
N-Methyl-D-Aspartic Acid	34	C5H9NO4	147.0532	CNC(C(C(O)=O)C(O)=O)	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h
O-Acetyls erine	35	C5H9NO4	147.0532	CC(OC)C(C(O)=O)N=O	Amino acid derivative	Ru/C - 55°C - 24h
Ornithine	36	C5H12N2O2	132.0899	NCCCC(C(O)=O)N	Amino acid	Ru/C - NaOD - 55°C - 72h
Pantothenic Acid	37	C9H17NO5	219.1107	CC(CO)(C(O)C(NCCC(O)=O)=O)C	Organic acid	Pd/C - NaOD - 80°C - 24h
Phosphoethanolamine	38	C2H8NO4P	141.0191	NCCOP(O)(O)=O	Organic amine	Ru/C - 55°C - 48h
Pipecolic Acid	39	C6H11NO2	129.0790	O=C(O)C1CCCCN1	Amino acid	Ru/C - NaOD - 55°C - 24h
Proline	40	C5H9NO2	115.0633	O=C(C1CCCN1)O	Amino acid	Ru/C - 55°C - 48h
Pyroglutamic Acid	41	C5H7NO3	129.0426	O=C1CCC(N1)C(O)=O	Organic acid	Ru/C - 55°C - 48h
Sarcosine	42	C3H7NO2	89.0477	CNCC(O)=O	Amino acid	Ru/C - NaOD - 55°C - 24h
Serine	43	C3H7NO3	105.0426	NC(CO)C(O)=O	Amino acid	Ru/C - 55°C - 48h
Sorbitol	44	C6H14O6	182.0790	OCC(O)C(O)C(O)C(O)CO	Polyol & derivative	Ru/C - 55°C - 48h
Succinic Acid	45	C4H6O4	118.0266	O=C(O)CCC(O)=O	Organic acid	Pd/C - NaOD - 80°C - 48h
Taurine	46	C2H7NO3S	125.0147	NCCS(O)=O	Organic amine	Ru/C - 55°C - 72h
Taurocholic Acid	47	C26H45NO7S	515.2917	OC1C2C3CCG(C(CCC(NCCS(O)=O)=O)O)C)C3(C)C(O)CC2C4(C)CCC(O)CC4C1	Bile acid derivative	Ru/C - 55°C - 72h
Threonine	48	C4H8O5	136.0372	OCC(C(C(O)=O)O)O	Organic acid	Ru/C - 55°C - 72h
Threonine	49	C4H9NO3	119.0582	CC(C(C(O)=O)N)O	Amino acid	Ru/C - NaOD - 55°C - 72h
4-Hydroxyproline	50	C5H9NO3	131.0582	OCC1NC(C(O)=O)C1	Amino acid	Ru/C - NaOD - 55°C - 72h
Tyrosine	51	C9H11NO3	181.0739	NC(C(O)=O)CC1=CC=C(C=C1)O	Amino acid	Ru/C - 55°C - 24h
Valine	52	C5H11NO2	117.0790	CC(C)C(C(O)=O)N	Amino acid	Ru/C + Pd/C - NaOD - 80°C - 48h

Partie C : Annexes – Annexes 5



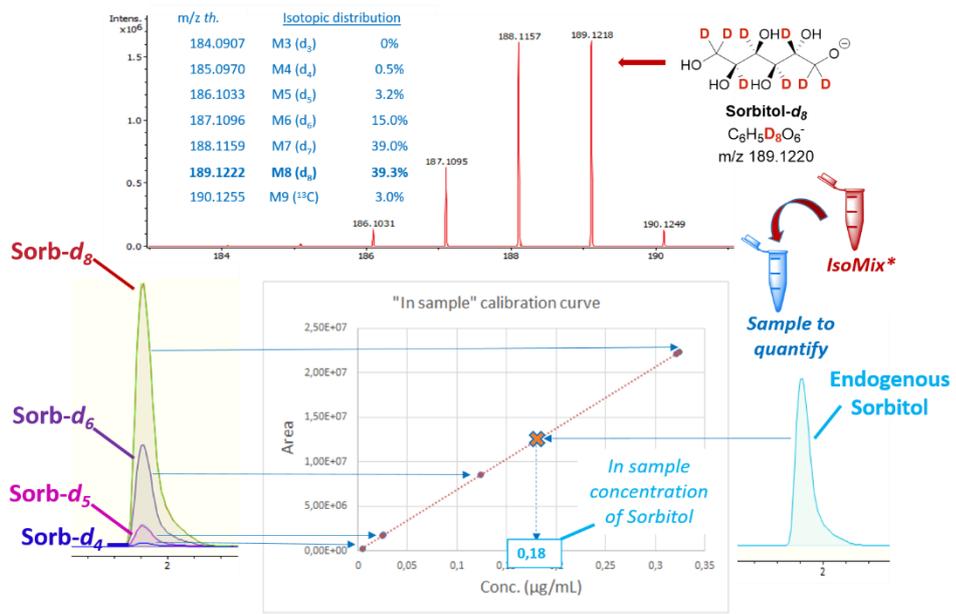
Supplementary Data S2 : Selected examples of isotopologue distribution for each labeled compounds

Partie C : Annexes – Annexes 5

Compounds	N°	Isotope distribution - Relative isotopologue abundances									Selected for quantification	
		M0	M1	M2	M3	M4	M5	M6	M7	M8		
α-Aminoadipic Acid	1	3%	100%	12%	-	-	-	-	-	-	-	
2-Aminobutyric Acid	2	< 1%	100%	54%	50%	17%	9%	6%	-	-	-	✓: Proof of concept
3-Amino-isobutyric Acid	3	-	< 1%	12%	100%	91%	52%	45%	< 1%	-	-	
Guanidinoacetic Acid	4	30%	76%	100%	-	-	-	-	-	-	-	
Guanidinossuccinic Acid	5	55%	100%	28%	2%	-	-	-	-	-	-	
3-Hydroxy-3-Methylglutaric acid	6	< 1%	10%	48%	100%	79%	6%	1%	-	-	-	
3-Methylhistamine	7	-	< 1%	14%	100%	70%	21%	3%	< 1%	-	-	✓: Proof of concept
3-Methylhistidine	8	-	-	< 1%	3%	26%	94%	100%	76%	41%	-	✓: Endogenous quantification
7-Methylguanidine	9	62%	100%	-	-	-	-	-	-	-	-	
Adipic Acid	10	-	-	-	-	18%	28%	59%	100%	84%	-	
Alanine	11	2%	100%	51%	17%	35%	< 1%	-	-	-	-	
Arginine	12	5%	100%	56%	27%	7%	1%	-	-	-	-	✓: Endogenous quantification
Asparagine	13	< 1%	6%	100%	49%	9%	< 1%	-	-	-	-	
Aspartic Acid	14	-	2.0%	30.0%	100.0%	-	-	-	-	-	-	
Citric Acid	15	-	< 1%	5%	41%	100%	7%	-	-	-	-	
Citrulline	16	3.7%	100.0%	3.6%	< 1%	-	-	-	-	-	-	✓: Endogenous quantification
Diglycolic Acid	17	-	< 1%	28%	100%	90%	-	-	-	-	-	✓: Proof of concept
Glutaconic Acid	18	-	-	2.2%	13.3%	100.0%	-	-	-	-	-	
Glutamic Acid	19	< 1%	8%	42%	100%	25%	1%	-	-	-	-	✓: Endogenous quantification
Glutamine	20	5.0%	100.0%	37.0%	10.0%	-	-	-	-	-	-	
Glycerol-3-Phosphate	21	< 1%	9%	54%	100%	67%	-	-	-	-	-	
Glycine	22	-	4%	100%	1%	-	-	-	-	-	-	
Histidine	23	-	-	< 1%	4%	31%	100%	7%	-	-	-	
Hydroxyglutaric Acid	24	10%	94%	100%	42%	12%	2%	-	-	-	-	
Imidazolelactic Acid	25	-	< 1%	5%	35%	88%	100%	6%	-	-	-	✓: Endogenous quantification
Isocitric Acid	26	< 1%	4%	53%	100%	7%	-	-	-	-	-	
Lactic Acid	27	100%	48%	4%	-	-	-	-	-	-	-	
Leucine	28	3%	100%	5%	3%	-	-	-	-	-	-	✓: Endogenous quantification
Lysine	29	4%	< 1%	14%	100%	5%	< 1%	-	-	-	-	
Maleic Acid	30	11.0%	20.0%	100.0%	80.0%	-	-	-	-	-	-	
Malic Acid	31	-	3%	100%	18%	< 1%	-	-	-	-	-	
Methylimidazoleacetic Acid	32	-	-	3%	35%	100%	18%	3%	1%	-	-	✓: Endogenous quantification
Methylmalonic Acid	33	Back exchange										
N-Methyl-D-Aspartic Acid	34	-	-	-	< 1%	12%	69%	100%	5%	-	-	✓: Proof of concept
O-Acetylserine	35	Complete degradation										
Ornithine	36	-	-	-	-	-	2%	23%	100%	5%	-	✓: Endogenous quantification
Pantothenic Acid	37	Complete degradation										
Phosphoethanolamine	38	< 1%	6%	100%	26%	2%	-	-	-	-	-	✓: Proof of concept
Pipelicolic Acid	39	-	< 1%	7%	100%	83%	28%	8%	1%	-	-	✓: Endogenous quantification
Proline	40	-	-	9%	100%	63%	15%	2%	-	-	-	✓: Endogenous quantification
Pyroglutamic Acid	41	100%	29%	3%	< 1%	-	-	-	-	-	-	
Sarcosine	42	-	-	< 1%	3%	20%	100%	-	-	-	-	✓: Endogenous quantification
Serine	43	< 1%	3%	47%	100%	3%	-	-	-	-	-	✓: Endogenous quantification
Sorbitol	44	-	-	-	-	-	2%	19%	83%	100%	-	✓: Proof of concept
Succinic Acid	45	-	1%	10%	45%	100%	-	-	-	-	-	
Taurine	46	-	3%	55%	100%	49%	-	-	-	-	-	✓: Endogenous quantification
Taurocholic Acid	47	28%	100%	52%	19%	5%	-	-	-	-	-	
Threonic Acid	48	7%	34%	100%	86%	30%	-	-	-	-	-	✓: Endogenous quantification
Threonine	49	1%	4%	100%	56%	43%	73%	2%	-	-	-	
4-Hydroxyproline	50	-	< 1%	15%	100%	37%	4%	< 1%	-	-	-	✓: Endogenous quantification
Tyrosine	51	51%	100%	65%	24%	10%	3%	-	-	-	-	
Valine	52	-	-	-	-	-	2%	17%	66%	100%	-	

Supplementary Data S3 : Principle of the in-sample multi-point internal calibration strategy. Illustration with for the absolute quantification of sorbitol in a biological matrix.

Partie C : Annexes – Annexes 5

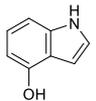
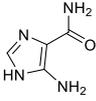
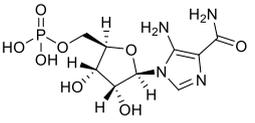
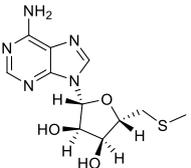
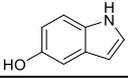
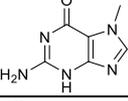
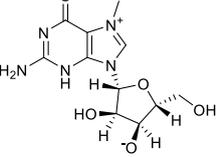
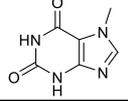
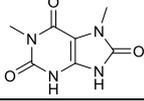
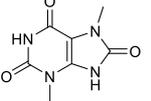


*Annexe 6 : Liste des composés investigués
avec la nouvelle méthode de production de
standards internes marqués (SIL-IS)*

Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
1-MethylGuanine	C6H7N5O	165.0651		<chem>CN(C(N)=N1)C(C2=C1NC=N2)=O</chem>
1-MethylAdenosine	C11H15N5O4	281.1124		<chem>CN1C=NC2=C(N=CN2[C@@]3([H])C[C@@]([O])([H])C[C@@]([O])([H])C[C@@]([O3])([H])CO)C1=N</chem>
1-MethylHistamine	C6H11N3	125.0953		<chem>CN(C=N1)C=C1CCN</chem>
1-MethylHistidine	C7H11N3O2	169.0851		<chem>CN(C=N1)C=C1C[C@@]([N])([H])C(O)=O</chem>
1-MethylXanthine	C6H6N4O2	166.0491		<chem>CN(C1=O)C(C2=C(N1)NC=N2)=O</chem>
2,4-Quinolinediol	C9H7NO2	161.0477		<chem>OC1=CC(O)=NC2=CC=CC=C21</chem>
2'-O-MethylAdenosine	C11H15N5O4	281.1124		<chem>CO[C@]([C@]1([H])N2C=NC(C2=NC=N3)=C3N)([H])C[C@@]([O])([H])C[C@@]([O1])([H])CO</chem>
2'-O-MethylGuanosine	C11H15N5O5	297.1073		<chem>CO[C@]([C@]1([H])N2C=NC3=C2NC(N)=NC3=O)([H])C[C@@]([O])([H])C[C@@]([O1])([H])CO</chem>
2'-O-MethylInosine	C11H14N4O5	282.0964		<chem>CO[C@]([C@]1([H])N2C=NC3=C2N=CNC3=O)([H])C[C@@]([O])([H])C[C@@]([O1])([H])CO</chem>
3-MethylHistidine	C7H11N3O2	169.0851		<chem>CN1C=NC=C1C[C@@]([N])([H])C(O)=O</chem>

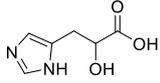
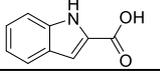
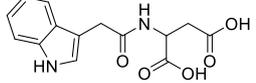
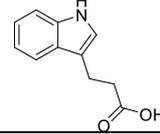
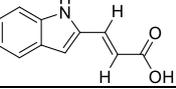
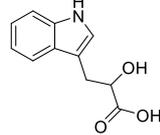
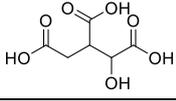
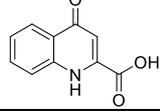
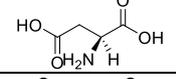
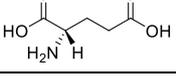
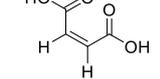
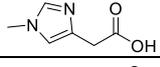
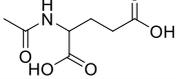
Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
4-Hydroxyindole	C8H7NO	133.0528		<chem>OC1=CC=CC2=C1C=CN2</chem>
5-Aminimidazole-4-Carboxamide	C4H6N4O	126.0542		<chem>NC1=C(N=CN1)C(N)=O</chem>
5-Aminimidazole-4-Carboxamide-1-β-D-Ribofuranosyl 5'-Monophosphate	C9H15N4O8P	338.0627		<chem>NC1=C(N=CN1[C@@]2([H])[C@@](O)([H])[C@@](O)([H])[C@@](O2)([H])COP(=O)(O)O)C(N)=O</chem>
5'-Deoxy-5'-(MethylThio)Adenosine	C11H15N5O3S	297.0896		<chem>CSC[C@](O1)([H])[C@](O)([H])[C@](O)([H])[C@]1([H])N2C=NC(C2=NC=N3)=C3N</chem>
5-Hydroxyindole	C8H7NO	133.0528		<chem>OC1=CC=C2C(C=CN2)=C1</chem>
7-MethylGuanine	C6H7N5O	165.0651		<chem>CN1C=NC(NC(N)=N2)=C1C2=O</chem>
7-MethylGuanosine	C11H15N5O5	297.1073		<chem>C[N+](=CN1([C@@]2([H])[C@@](O)([H])[C@@](O)([O-])[C@@]([H])C(=O)C(NC(N)=N3)=C1C3=O</chem>
7-MethylXanthine	C6H6N4O2	166.0491		<chem>CN1C=NC(NC2=O)=C1C(N2)=O</chem>
Acide 1,7-DimethylUrique	C7H8N4O3	196.0596		<chem>CN1C(C(N2C=O)=C(NC2=O)N1=O</chem>
Acide 3,7-DimethylUrique	C7H8N4O3	196.0596		<chem>CN1C(C(NC2=O)=O)=C(N2C)N1=O</chem>

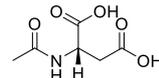
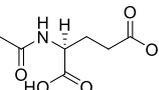
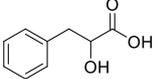
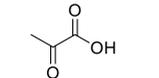
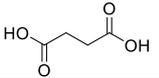
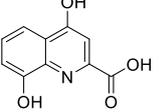
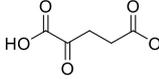
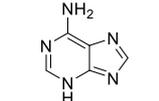
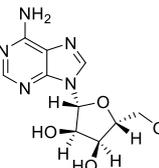
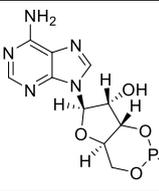
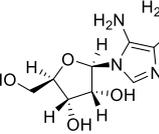
Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Acide 4-AcetamidoButanoïque	C6H11NO3	145.0739		<chem>CC(NCCCC(O)=O)=O</chem>
Acide 4-ImidazoleAcrylique	C6H6N2O2	138.0429		<chem>O=C(O)/C([H])=C([H])/C1=CN=CN1</chem>
Acide 9-MéthylUrique	C6H6N4O3	182.0440		<chem>CN(C1=O)C(NC2=O)=C(N1)C(N2)=O</chem>
Acide Adipique	C6H10O4	146.0579		<chem>O=C(O)CCCC(O)=O</chem>
Acide cis-Aconitique	C6H6O6	174.0164		<chem>O=C(O)/C([H])=C(C(O)=O)\CC(O)=O</chem>
Acide Citraconique	C5H6O4	130.0266		<chem>C/C(C(O)=O)=C([H])\C(O)=O</chem>
Acide Citrique	C6H8O7	192.0270		<chem>O=C(O)CC(C(O)=O)(O)CC(O)=O</chem>
Acide D-2-HydroxyGlutarique	C5H8O5	148.0372		<chem>O[C@](C(O)=O)([H])CCC(O)=O</chem>
Acide DL-DihydroOrotique	C5H6N2O4	158.0328		<chem>O=C(N1)C[C@@]([H])(NC1=O)C(O)=O</chem>
Acide Glutaconique	C5H6O4	130.0266		<chem>O=C(O)/C([H])=C([H])/CC(O)=O</chem>
Acide Glycocholique	C26H43NO6	465.3091		<chem>C[C@]([C@@]1([H])CC[C@@]2([H])(C@@)([C@@]([O]([H])CC3([H])C[C@@]([O])([H])CC[C@@]34C([H])(C@]4([H]))C[C@@]([O])([H])[C@]12C([H])CCC(NCC(O)=O)=O</chem>
Acide HydroCinnamique	C9H10O2	150.0681		<chem>O=C(O)CCC1=CC=CC=C1</chem>
Acide ImidazoleAcétique	C5H6N2O2	126.0429		<chem>O=C(O)CC1=CN=CN1</chem>

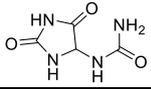
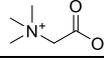
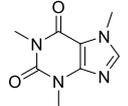
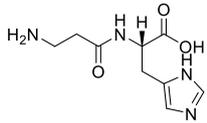
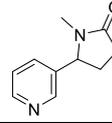
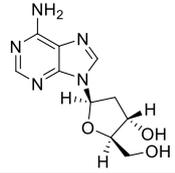
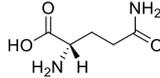
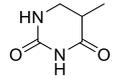
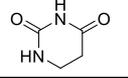
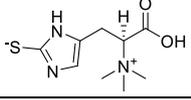
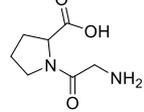
Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Acide ImidazoleLactique	C6H8N2O3	156.0535		<chem>OC(C(O)=O)CC1=CN=CN1</chem>
Acide Indole-2-Carboxylique	C9H7NO2	161.0477		<chem>O=C(O)C1=CC2=CC=CC=C2N1</chem>
Acide Indole-3-Acetyl-L-Aspartique	C14H14N2O5	290.0903		<chem>O=C(NC(C(O)=O)CC(O)=O)CC1=CNC2=CC=CC=C12</chem>
Acide Indole-3-Propionique	C11H11NO2	189.0790		<chem>O=C(O)CCC1=CNC2=CC=CC=C12</chem>
Acide IndoleAcrylique	C11H9NO2	187.0633		<chem>O=C(O)/C([H])=C([H])/C1=CC2=CC=CC=C2N1</chem>
Acide IndoleLactique	C11H11NO3	205.0739		<chem>OC(C(O)=O)CC1=CNC2=CC=CC=C12</chem>
Acide Isocitrique	C6H8O7	192.0270		<chem>O=C(O)CC(C(O)=O)C(O)C(O)=O</chem>
Acide Kynurenique	C10H7NO3	189.0426		<chem>O=C(C1=CC=CC=C1N2)C(=O)C2=O</chem>
Acide L-Aspartique	C4H7NO4	133.0375		<chem>N[C@@]([C(O)=O])([H])CC(O)=O</chem>
Acide L-Glutamique	C5H9NO4	147.0532		<chem>N[C@@]([C(O)=O])([H])CCC(O)=O</chem>
Acide Maleique	C4H4O4	116.0110		<chem>O=C(O)/C([H])=C([H])C(O)=O</chem>
Acide MethylimidazoleAcetique	C6H8N2O2	140.0586		<chem>CN(C=N1)C=C1CC(O)=O</chem>
Acide N-Acetyl-DL-Glutamique	C7H11NO5	189.0637		<chem>CC(NC(C(O)=O)CCC(O)=O)=O</chem>

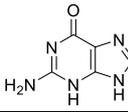
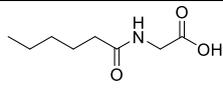
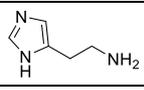
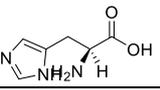
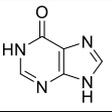
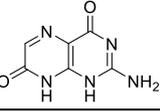
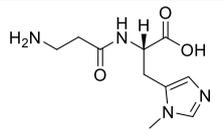
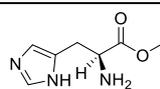
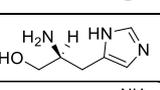
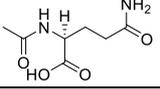
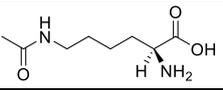
Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Acide N-Acetyl-L-Aspartique	C6H9NO5	175.0481		<chem>CC(N[C@@](C(=O)O)([H])CC(=O)O)=O</chem>
Acide N-Acetyl-L-Glutamique	C7H11NO5	189.0637		<chem>CC(N[C@@](C(=O)O)([H])CCC(=O)O)=O</chem>
Acide Phenyllactique	C9H10O3	166.0630		<chem>OC(C(=O)O)CC1=CC=CC=C1</chem>
Acide Pyruvique	C3H4O3	88.0160		<chem>CC(C(=O)O)=O</chem>
Acide Succinique	C4H6O4	118.0266		<chem>O=C(O)CCC(=O)O</chem>
Acide Xanthurenique	C10H7NO4	205.0375		<chem>OC1=CC=CC(C1=N2)=C(O)C=C2C(=O)O</chem>
Acide α-Ketoglutarique	C5H6O5	146.0215		<chem>O=C(C(=O)O)CCC(=O)O</chem>
Adenine	C5H5N5	135.0545		<chem>NC(N=CN1)=C2C1=NC=N2</chem>
Adenosine	C10H13N5O4	267.0967		<chem>NC1=C(N=CN2[C@@]3([H])[C@@]([O])([H])[C@@]([O])([H])[C@@]([O3])([H])CO)C2=NC=N1</chem>
Adenosine Monophosphate Cyclique	C10H12N5O6P	329.0525		<chem>NC1=C(N=CN2[C@@]3([H])[C@@]([O])([H])[C@@]([O4])([H])[C@@]([O3])([H])COP4(O)=O)C2=NC=N1</chem>
5-Aminimidazole-4-Carboxamide 1-β-D-Ribofuranoside	C9H14N4O5	258.0964		<chem>NC1=C(N=CN1[C@@]2([H])[C@@]([O])([H])[C@@]([O])([H])[C@@]([O2])([H])CO)C(N)=O</chem>

Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Allantoin	C4H6N4O3	158.0440		<chem>NC(NC(=O)C(N)=O)=O</chem>
Betaine	C5H11NO2	117.0790		<chem>C[N+](C)(C)C([O-])=O</chem>
Caffeine	C8H10N4O2	194.0804		<chem>CN1C=NC(N(C2=O)C)=C1C(N2C)=O</chem>
Carnosine	C9H14N4O3	226.1066		<chem>NCCC(N[C@@](C(O)=O)([H])CC1=CN=CN1)=O</chem>
Cotinine	C10H12N2O	176.0950		<chem>CN1C(C2=CN=CC=C2)CCC1=O</chem>
DeoxyAdenosine	C10H13N5O3	251.1018		<chem>NC1=C(N=CN2[C@@]3([H])C[C@@]([O])([H])[C@@]([O3])([H])CO)C2=NC=N1</chem>
D-Glutamine	C5H10N2O3	146.0691		<chem>N[C@@](C(O)=O)([H])CCC(N)=O</chem>
DihydroThymine	C5H8N2O2	128.0588		<chem>CC(C(N1=O)CNC1=O</chem>
DihydroUracil	C4H6N2O2	114.0429		<chem>CC(C(N1=O)CNC1=O</chem>
Ergothioneine	C9H15N3O2S	229.0885		<chem>C[N+](C)(C)[C@@](C(O)=O)([H])CC1=CN=C([S-])N1)C</chem>
Glycine-Proline	C7H12N2O3	172.0848		<chem>NCC(N1CCCC1C(O)=O)=O</chem>

Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Guanine	C5H5N5O	151.0494		<chem>NC1=NC(C2=C(N1)NC=N2)=O</chem>
Hexanoyl-Glycine	C8H15NO3	173.1052		<chem>CCCCCC(NCC(O)=O)=O</chem>
Histamine	C5H9N3	111.0797		<chem>NCCC1=CN=CN1</chem>
Histidine	C6H9N3O2	155.0695		<chem>N[C@@](C(O)=O)([H])CC1=CN=CN1</chem>
Hypoxanthine	C5H4N4O	136.0385		<chem>O=C1C2=C(N=CN1)NC=N2</chem>
Indole	C8H7N	117.0578		<chem>C12=CC=CC=C1NC2=C</chem>
Isoxanthopterin	C6H5N5O2	179.0443		<chem>NC1=NC(C2=C(N1)NC(=N2)=O)=O</chem>
L-Anserine	C10H16N4O3	240.1222		<chem>CN1C=NC=C1C[C@@](N(C(CCN)=O)([H])C(O)=O)</chem>
L-Histidine-Methyl-Ester	C7H11N3O2	169.0851		<chem>COC([C@](N)([H])CC1=CN=CN1)=O</chem>
L-Histidinol	C6H11N3O	141.0902		<chem>N[C@@](CO)([H])CC1=CN=CN1</chem>
N-Acetyl-L-Glutamine	C7H12N2O4	188.0797		<chem>CC(N[C@@](C(O)=O)([H])CCC(N)=O)=O</chem>
N6-Acetyl-L-Lysine	C8H16N2O3	188.1161		<chem>CC(NCCCC[C@@](N)([H])C(O)=O)=O</chem>

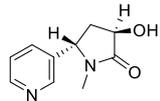
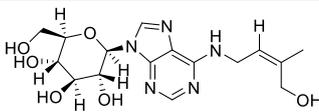
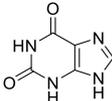
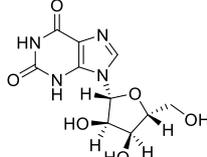
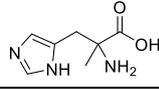
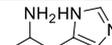
Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
N6-MethylAdenine	C6H7N5	149.0701		<chem>CNC(N=CN1)=C2C1=NC=N2</chem>
N-Acetyl-D-Allo-Isoleucine	C8H15NO3	173.1052		<chem>CC[C@]([C@](NC(C)=O)([H])C(O)=O)([H])C</chem>
N-Acetyl-D-Glucosamine	C8H15NO6	221.0899		<chem>CC(N[C@@H](C1O)[C@](O)([H])[C@@](O)([H])[C@@](O1)([H])CO)=O</chem>
N-Acetyl-DL-Norvaline	C7H13NO3	159.0895		<chem>CCCC(NC(C)=O)C(O)=O</chem>
N-Acetyl-DL-Serine	C5H9NO4	147.0532		<chem>CC(NC(C(O)=O)CO)=O</chem>
N-AcetylGlycine	C4H7NO3	117.0426		<chem>CC(NCC(O)=O)=O</chem>
N-AcetylHistamine	C7H11N3O	153.0902		<chem>CC(NCCC1=CN=CN1)=O</chem>
N-Acetyl-L-alanine	C5H9NO3	131.0582		<chem>C[C@]([NC(C)=O)([H])C(O)=O</chem>
N-Acetyl-L-Leucine	C8H15NO3	173.1052		<chem>CC[C@]([C@](NC(C)=O)([H])C(O)=O)C</chem>
N-Acetyl-L-Phenylalanine	C11H13NO3	207.0895		<chem>CC(N[C@@](C(O)=O)([H])CC1=CC=CC=C1)=O</chem>
N-Acetyl-Ornithine	C7H14N2O3	174.1004		<chem>CC(N[C@@](C(O)=O)([H])CCCN)=O</chem>
N-IsovaleroylGlycine	C7H13NO3	159.0895		<chem>CC(CC(NCC(O)=O)=O)C</chem>
N-Methyl-Gamma-Oxo-3-PyridineButanamide	C10H12N2O2	192.0899		<chem>CNC(CCC1=CN=CC=C1)=O</chem>

Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
o-Tyrosine	C9H11NO3	181.0739		<chem>NC(C(O)=O)CC1=CC=CC=C1O</chem>
Paraxanthine	C7H8N4O2	180.0647		<chem>NC(C(O)=O)CC1=CC=CC=C1O</chem>
Phenylacetyl-L-Glutamine	C13H16N2O4	264.1110		<chem>NC(CC[C@@H](NC(CC1=CC=CC=C1)=O)[H])C(O)=O</chem>
Purine	C5H4N4	120.0436		<chem>C12=CN=CN=C1NC=N2</chem>
S-Adenosyl-Homocysteine	C14H20N6O5S	384.1216		<chem>N[C@@H](C(O)=O)[C@H](O)CCSC[C@@H](O)[C@H](O)[C@H](O)[C@H]1N2C=NC(=C(N3)C=C(N3)=C3N</chem>
S-AdenosylMethionine	C15H22N6O5S	398.1372		<chem>C[S+](C[C@@H](O)[H])[C@H](O)[H][C@H](O)[C@H]1([H])N2C=NC(C2=NC=N3)=C3N)CC[C@@H](N)[C](O)=O</chem>
Stachydrine	C7H13NO2	143.0946		<chem>C[N+]1(CCC[C@@H]1([H])C([O-])=O)C</chem>
Theophylline	C7H8N4O2	180.0647		<chem>CN(C1=O)C2=C(N=CN2)C(N1C)=O</chem>
Thiamine	C12H16N4OS	264.1045		<chem>CC([N+](CC(C(N)=N1)=CN=C1C)=CS2)=C2CC([O-])</chem>
Thiamine Pyrophosphate	C12H18N4O7P2S	424.0371		<chem>CC([N+](CC(C(N)=N1)=CN=C1C)=CS2)=C2CCOP([O-])OP(O)(O)=O</chem>
Thiamine-PO4	C12H17N4O4PS	344.0708		<chem>CC([N+](CC(C1=N)=CN=C(N1)C)=CS2)=C2CCOP(O)([O-])=O</chem>

Partie C : Annexes – Annexe 6

Metabolite	Formule Brute	Masse Précise	Structure	Smile
Trans-3-HydroxyCotinine	C10H12N2O2	192.0899		<chem>CN1[C@@](C2=CN=CC=C2)([H])C[C@](O)([H])C1=O</chem>
trans-Zeatin Glucoside	C16H23N5O6	381.1648		<chem>C/C(CO)=C([H])/CNC1=C(N=CN2[C@@]3([H])[C@@]([O])([H])[C@@]([O])([H])[C@@]([O])([H])C@@]3([O3])([H])CO)C2=NC=N1</chem>
Xanthine	C5H4N4O2	152.0334		<chem>O=C(N1)C2=C(NC1=O)NC=N2</chem>
Xanthosine	C10H12N4O6	284.0757		<chem>OC[C@](O1)([H])[C@@](O)([H])[C@@](O)([H])[C@]1([H])N2C=NC(C(N3)=O)=C2N3=O</chem>
α-Methyl-DL-Histidine	C7H11N3O2	169.0851		<chem>CC(C(O)=O)(N)CC1=CN=CN1</chem>
α-MethylHistamine	C6H11N3	125.0953		<chem>CC(N)CC1=CN=CN1</chem>

Développement de nouveaux outils pour l'analyse métabolomique par spectrométrie de masse haute résolution : de l'acquisition de spectres MS/MS pour l'identification large-spectre de métabolites au marquage isotopique pour la quantification

Spectrométrie de masse à haute résolution, Métabolomique, Marquage isotopique, Quantification

La métabolomique est une science relativement récente qui étudie les molécules de masses moléculaires inférieures à 1.5 kDa, présentes dans une matrice biologique donnée. Elle est le dernier maillon des sciences « omiques » et représente l'ultime réponse d'un organisme aux facteurs perturbant son fonctionnement. Aujourd'hui, de nombreux développements en spectrométrie de masse à haute résolution couplée à la chromatographie liquide (LC-HRMS), visent à améliorer la détection, l'identification et la quantification des métabolites mais également à augmenter la robustesse des analyses. C'est dans ce contexte que mon projet de thèse s'est intégré. L'un des objectifs de cette thèse a donc été de développer des méthodes permettant l'acquisition simultanée de spectres MS et MS/MS via des analyses en mode d'acquisitions dits « données dépendantes » et « données indépendantes ». Ces méthodes ont permis d'avoir la même sensibilité que les analyses classiquement réalisées en spectrométrie de masse à haute résolution, tout en acquérant des informations importantes pour l'identification des métabolites grâce aux spectres MS/MS. Ces développements ont pu être appliqués à une étude inter-laboratoires dans le cadre d'un projet européen. L'autre

objectif majeur de ce doctorat a été la mise en place de stratégies pour la quantification de métabolites en matrice biologique. Pour ce faire, la production et la caractérisation de standards internes marqués aux isotopes stables ont été envisagées. Ainsi, des méthodes de synthèses chimiques par échanges H/D ont été évaluées et ont permis d'obtenir de nombreux composés marqués. L'exploitation des profils isotopiques obtenus pour ces composés a permis la mise en place d'une approche quantitative innovante par étalonnage interne multipoints. Cette méthode a pu être comparée aux approches classiques de dilution isotopique. Par ailleurs, l'étude de profils isotopiques a également pu être appliquée à des analyses préliminaires d'urines de souris métaboliquement marquées (*in vivo*) au carbone-13, à la fois pour du suivi cinétique d'incorporation du C-13 au sein de quelques métabolites ciblés mais aussi, à plus long terme, pour l'annotation et l'identification de nouveaux métabolites. In fine, l'ensemble des développements réalisés au cours de cette thèse ont permis d'atteindre un meilleur niveau d'identification des métabolites et d'améliorer leur quantification

Development of new tools for metabolomic analysis by high-resolution mass spectrometry: from the acquisition of MS/MS spectra for large-scale metabolite identification to isotope labeling for quantification

High Resolution Mass Spectrometry, Metabolomics, Isotopic Labeling, Quantification

Metabolomics is a relatively new science that studies molecules with a molecular weight below 1.5 kDa, present in a given biological matrix. It is the last link in the "omics" sciences and represents the ultimate response of an organism to factors that disrupt its functioning. Today, many developments in high-resolution mass spectrometry coupled to liquid chromatography (LC/HRMS) aim at improving the detection, identification and quantification of metabolites but also at increasing the robustness of the analyses. This PhD thesis took place in that context. One of its main objectives was to develop methods allowing the simultaneous acquisition of MS and MS/MS spectra via analyses in "Data Dependent Acquisition" and "Data Independent Acquisition" modes. These methods allowed the production of data with the same sensitivity as the traditional "Full Scan" analysis while providing more precise and relevant information for further metabolite identification. These methodological developments were applied to an inter-laboratory study as part of a European project. The other major objective of this

PhD was to set-up strategies for the large-scale quantification of metabolites in biological samples. To do so, the production and characterization of isotopically labeled internal standards were investigated. Thus, chemical synthesis methods relying on H/D exchange was evaluated and enabled the production of a large panel of labeled compounds. Their isotopic pattern exploitation allowed the development of an innovative quantitative approach by multi-point internal calibration. This method was compared to the conventional isotope dilution approach. At last, isotope profiling was applied to the preliminary analysis of mouse urine samples obtained from mice that were *in vivo* labeled with carbone-13. The aims were to study the C-13 incorporation kinetics in targeted metabolites, but also to annotate and identify new metabolites. Overall, the set of developments carried out during this thesis have allowed to reach a better level of identification of metabolites and to improve their quantification

