

ÉCOLE DOCTORALE N°104 SCIENCES DE LA MATIÈRE DU RAYONNEMENT ET
DE L'ENVIRONNEMENT

Numéro d'ordre : **42555**

UNITÉ DE SERVICE ET RECHERCHE CNRS USR 3290
MINIATURISATION POUR LA SYNTHÈSE, L'ANALYSE ET LA PROTÉOMIQUE

THÈSE

Présentée par : **Fabrice BRAY**

Pour l'obtention du titre de
DOCTEUR DE L'UNIVERSITÉ DE LILLE 1

Titre de la Thèse

La spectrométrie de masse haute résolution :
Application à la FT-ICR bidimensionnelle et à la
protéomique dans les domaines de l'archéologie et la
paléontologie

Thèse soutenue le 12 décembre 2017 devant la commission d'examen composée de

Rapporteurs	Pr. Carlos AFONSO, laboratoire de chimie organique et analytique, UMR 6014 COBRA, Université de Rouen Dr. Emmanuelle LEIZE-WAGNER, Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 UNISTRA-CNRS, Strasbourg
Examineurs	Dr. Mathieu BOUDIN, Institut Royal du Patrimoine Artistique (IRPA), Bruxelles Pr. Valéry HEDOUIN, EA 7367 Unité de Taphonomie Médico-Légale (UTML), Université de Lille 2
Directrice de thèse	Pr Caroline TOKARSKI, laboratoire Miniaturisation pour la Synthèse, l'Analyse et la Protéomique (MSAP), Université de Lille 1
Co- Directeur de thèse	Dr Christian ROLANDO, laboratoire Miniaturisation pour la Synthèse, l'Analyse et la Protéomique (MSAP), Université de Lille 1

Remerciements,

Je remercie, en premier lieu, Madame la Professeur Caroline Tokarski pour avoir accepté que je réalise ma thèse au sein du laboratoire.

Merci de m'avoir suivi pendant ces trois ans et d'avoir obtenus des ossements de plus en plus anciens.

Je tiens à remercier Monsieur le Docteur Christian Rolando pour m'avoir initié à la 2D FT-ICR mais également pour toutes les choses apprises grâce à vous.

Je tiens également à remercier le Professeur Carlos AFONSO et le Docteur. Emmanuelle LEIZE-WAGNER pour avoir accepté d'examiner ce travail.

Je remercie le Docteur Mathieu BOUDIN et le professeur Valéry HEDOUIN d'examiner ce travail de thèse.

Merci à toute l'équipe du laboratoire MSAP : Stéphanie, Christophe, Amandine, Laëtitia, Maël, Ahmed, Karine, Nicolas, Stanislas, Maud...

Bien sûr, merci à ma femme Soizic qui m'a aidé pour les corrections de mon manuscrit mais aussi pour son soutien. Merci Soizic et mon fils Gaël d'être venus, de temps en temps, avec moi au laboratoire le week-end.

Je remercie aussi mes parents et mes frères de m'avoir écouté raconter mes projets.

-Sommaire-

Chapitre I : Généralités sur la spectrométrie de masse

I.	La spectrométrie de masse	2
II.	La résolution	3
III.	La précision en masse	6
IV.	La limite de masse	8
V.	Les sources d'ionisation	9
VI.	Source électronébulisation (ESI)	11
	1. L'étape de nébulisation	12
	2. La désolvatation et l'ionisation	13
	3. Les ions multichargés	14
VII.	Les analyseurs	16
	1. Les analyseurs de type FT-ICR	17
	a) Théorie de la FT-ICR	19
	b) Piégeage des ions	20
	c) Mouvement des ions dans la cellule	20
	d) Mesure du signal FT-ICR	21
	e) La cellule harmonisée	25
	2. L'analyseur Orbitrap	26
	a) Orbitrap, principe et utilisations	26
	b) Principe de l'Orbitrap	27
	c) La C-trap	28
	d) L'injection des ions	29
	e) Détection des ions	30
VIII.	La spectrométrie de masse en tandem ou MS/MS	31
	1. HCD : Higher –energy Collisional Dissociation	33
	2. IRMPD : dissociation Multiphotonique par infrarouge	34

CHAPITRE II : L'analyse bidimensionnelle par spectromètre de masse à transformée de Fourier (FT-ICR)

Partie 1 : Généralités sur la FT-ICR 2D

I.	Les premiers pas de la FT-ICR 2D	41
1.	L'expérience de Marshall	41
2.	Première expérience de 2D-FT	42
a)	Théorie	42
b)	FT-ICR 2D MS sur les ions méthane	45
3.	Les variantes de la 2D	49
II.	Récents développements en 2D-FT-ICR	50
1.	Ré-introduction de la 2D FT-ICR	51
2.	Optimisations de la méthode	51
3.	Traitement des données obtenue en 2D FT	53
4.	Le débruitage des données	55
III.	L'avenir de la FT-ICR 2D	58

Partie 2 : Analyse des triacylglycérols par FT-ICR 2D

I.	Les triacylglycérols	64
1.	Définition	64
2.	Métabolisme	64
3.	Maladies métabolique	65
II.	Les méthodes d'analyses classiques	65
III.	Préparation des lipides	65
1.	Méthodes d'extraction des lipides	65
2.	Méthodes de séparation et fractionnement	67
IV.	Analyse par spectrométrie de masse	68
1.	Chromatographie en phase gazeuse (GC)	69
2.	Analyse par bombardement atomique rapide	69
3.	La MALDI	69
4.	Les techniques : ESI, APCI, APPI, DESI, DART	70
5.	Le Couplage	71
V.	L'analyse des TAGs par fragmentation	71

VI.	Logiciels informatiques	72
VII.	Objectifs	74
VIII.	Matériels et méthodes	75
	1. Extraction des lipides du plasma	75
	2. Purification de familles lipidiques	75
	3. Analyse par spectrométrie de masse	76
	4. Traitement de l'information	77
IX.	Résultats	79
	1. Extraction des TAGs	79
	2. Analyse FT-ICR MS et MS/MS	80
	3. Cartes 2D FT-ICR des TAGs humain	84
	4. Identification des TAGs avec l'analyse 2D FT-ICR	86
	5. Analyse par pertes de neutres	87
X.	Conclusion	90

Partie 3 : FT-ICR 2D: Analyse haute résolution

I.	FT-ICR 2D : Analyse haute résolution	102
	1. Introduction	102
	2. Objectifs	105
II.	Matériels et méthodes	106
	1. Préparation des échantillons	106
	2. Analyse FT-ICR de la substance P	106
	3. Analyse FT-ICR de l'huile de sésame	107
	4. Retraitement informatique	107
III.	Résultats	108
	1. Intervalle de temps t_1	108
	2. Analyse NUS de la substance P	109
	3. Acquisition NUS pour l'analyse de l'huile de sésame.	116
IV.	Conclusion	121

Partie 4 : Empreinte lipidique par analyse FT-ICR 2D

I.	Analyse des triacylglycérols archéologiques	126
1.	Analyse par chromatographie en phase gazeuse	128
2.	Analyse par chromatographie en phase liquide	128
3.	Analyse par injection directe	128
II.	Objectifs	129
III.	Matériels et méthodes	130
1.	Echantillons	130
2.	Extraction lipides et purification des TAG	131
3.	Analyse 1D FT-ICR des TAGs	131
4.	Analyse 2D FT-ICR des TAGs	132
5.	Analyse informatique	132
IV.	Analyse FT-ICR MS et MS/MS	133
1.	Profils MS des échantillons	133
2.	Identification structurale des TAGs par fragmentation IRMPD	136
3.	Analyse d'échantillon archéologique	136
V.	Analyse 2D FT-ICR MS	138
1.	Empreinte lipidique	138
2.	Analyse 2D de l'échantillon archéologique	140
3.	Analyse haute résolution : identification structurale des TAGs	141
VI.	Conclusion	144

CHAPITRE III : L'analyse protéomique d'échantillons archéologiques et paléontologiques

Partie 1 : Généralités sur l'analyse des échantillons archéologiques et paléontologiques

I.	Introduction	149
II.	L'analyse d'échantillons anciens	149
III.	La détection des protéines dans les échantillons archéologiques	151
1.	Colorations des protéines	151
2.	Spectrométries vibrationnelles	151
a)	Spectroscopie Infra-rouge (IR)	152

b) Spectroscopie Raman	152
IV. Identification des résidus protéiques contenus dans les échantillons archéologiques ou paléontologiques	153
1. Analyse par TOF-SIMS	153
2. Méthodes d'analyses des acides aminés	154
3. Chromatographie en phase gazeuse	155
4. Méthodes immunologiques	156
V. Conclusion	157

Partie 2 : Analyse bottom-up des échantillons archéologiques et paléontologiques

I. La protéomique	162
II. L'approche bottom up	163
1. La préparation des échantillons	164
2. La digestion des protéines	165
3. Séparation des peptides par couplage chromatographique	166
4. La MS/MS ou spectrométrie de masse en tandem	166
5. Bio-informatique	167
III. Bottom up – archéologie & paléontologie	169
IV. Objectifs	172
V. Matériels et méthodes	173
1. Echantillons d'ossements analysés	173
2. Optimisation de la composition du tampon de déminéralisation	174
3. Analyse protéomique des ossements	174
4. Digestion des échantillons (méthode eFASP)	175
5. Couplage LC-MS/MS Orbitrap.	176
6. Analyse Bio-informatique des extractions	177
7. Analyse bio-informatique des ossements archéologique	178
8. Phylogénie des protéines	178
VI. Analyse des résultats sur les ossements	180
1. Mise au point du tampon de déminéralisation	180
2. Déminéralisation des ossements	172
3. Analyse protéomique	173

a)	Etude des tampons de déminéralisation sur les protéines identifiées dans la solution de déminéralisation et dans la poudre d'ossement.	174
b)	Analyse quantitative des protéines après déminéralisation.	176
4.	Analyse des échantillons anciens	180
a)	Analyse bio-informatique	180
b)	Protéines présentes et contaminants	181
c)	Le collagène	183
d)	Analyse de la déamidation des peptides du collagène	186
e)	Identification du genre de l'ossement	188
f)	Etude des peptides spécifiques	191
VII.	Conclusion	196

Partie 3 : Analyse sans digestion des échantillons archéologiques et paléontologiques

I.	L'approche top-down	206
II.	Les premières étapes de l'analyse TDP : le fractionnement des protéines	208
III.	Instruments et techniques de fragmentation	210
IV.	Applications de l'analyse top-down à des échantillons complexes.	213
V.	La quantification des protéines entières	214
VI.	L'analyse bio-informatique	216
VII.	L'analyse TPD pour l'archéologie et la paléontologie	217
VIII.	Objectifs	218
IX.	Matériels et méthodes	219
1.	Echantillons utilisés pour l'optimisation du protocole	219
2.	Echantillon archéologique	219
3.	Digestion des protéines avec la méthode eFASP	220
4.	Méthode de préparation des protéines	220
5.	Analyse des protéines par MALDI TOF	221
6.	Analyse des protéines par nanoLC-MS/MS	221
7.	Analyse Bio-informatique	222
X.	Résultats et discussions	224
1.	Méthodes de préparation des protéines intactes	224
2.	Analyse de l'échantillon Zubarch 55 par nanoLC-MSMS eFASP	206

3.	Analyse top down de l'échantillon Zurzach 55	206
a)	Analyse top down de lait de vache	206
b)	Analyse top down de l'échantillon archéologique Zurzach 55	207
4.	Analyse bio-informatique	208
XI.	Conclusion	213

ANNEXES

I.	Séquences des protéines utilisées pour l'analyse phylogénique du collagène I alpha 1	222
II.	Séquences des protéines utilisées pour l'analyse phylogénique du collagène I alpha 2	229
III.	Graphique de transition de phase de Donoho-Tanner	236
IV.	Publication de recherche	237
IV.	Publication de service	237
1.	Publication de service 1	238
2.	Publication de service 2	239

Abbréviations

ABC : Ammonium bicarbonate

ACN : Acétonitrile

AGC target: Automatic gain control

Ag-LC : Chromatographie liquide avec argent

BCA : BiCinchoninic acid Assay

BUP : Bottom-up

CH₃OH : Méthanol

CHCl₃ :Chloroforme

CID : Collision induit par dissociation

Da : Dalton

DCA : Acide déoxycholate

DTT : Dithiothréitol

ECD : Dissociation par capture d'électrons

eFASP : Enhanced Filter Aide Sample Preparation

ESI : Électronébulisation

ETD : Dissociation par transfert d'électrons

ev : Électrons volts

FT-ICR 2D : Spectrométrie de masse à résonance cyclotronique ionique à transformée de Fourier bidimensionnelle

FT-ICR : Spectrométrie de masse à résonance cyclotronique ionique à transformée de Fourier

FWMH : Full width at half-maximum

g : Gravité, unité d'accélération

g/L : Gramme par litre

GC : Chromatographie gazeuze

HCCA : α -Cyano-4-hydroxycinnamic acid

HCl: Acide chlorhydrique

HILIC : Chromatographie d'interaction hydrophile

IAA : Iodoacétamide

IRMPD : Multiple photo-dissociation par infra rouge

k : Kilo

kDa: Kilo Dalton

kHz : Kilo hertz
LC : Chromatographie liquide
LC-MS/MS : Chromatographie liquide avec
⁶Li : Chlorure de lithium 6
m : mètre
M : Molaire
m/z : Rapport masse / charge
MALDI : Matrix Assisted Laser Desorption Ionization
Maximum IT: Temps maximum d'injection
mg/dL : Milligramme par décilitre
mg/L : milligramme par litre
MHz : Mega hertz
min : Minute
mL : Millilitre
mM : Milli molaire
mm : Millimètre
mmol/L : Millimole par litre
mQ: MilliQ
MS : Spectrométrie de masse
MS/MS: Spectrométrie de masse / spectrométrie de masse
MTBE: Méthyl tert-butyl éther
N : Normalité
nanoESI : Nano electronébulisation
NaOH: Hydroxyde de sodium
NARP-LC : Chromatographie liquide sur phase inverse non aqueux
nL/min: Nanolitre par minute
NP-LC : Chromatographie liquide sur phase normale
NUS : L'échantillonnage non-uniforme
pH: Potentiel hydrogène
pmol/μL: Picomole par microliter
ppm : Partie par million
PSM: The number of peptide spectrum matches
PTM: Modifications post-traductionnelles
Q : Quadripôle

Q-TOF : Spectromètre de masse quadripôle avec temps de vol

RMN : Résonance magnétique nucléaire

RP-LC : Chromatographie liquide sur phase inverse

s : Seconde

SDS : Sodium dodécylsulfate

SFC : Chromatographie en phase supercritique

SNP : single nucleotide polymorphism

SPE : Extraction en phase solide

SVD : La décomposition en valeur singulière

TAG : Triacylglycérols

TCEP : Tris(2-carboxyethyl)phosphine

TDP : Top-down

TFA: Acide trifluoroacétique

TLC : Chromatographie sur couche mince

TOF : Temps de vol

Tris HCl : Tris hydrochloride

μL : Micro litre

μm : Micro mètre

μs : Micro seconde

UV : Ultra-violet

UVPD : Photodissociation par ultra-violet

V : Volt

v/v : Volume pour volume

Vpp : Peak-to-Peak Voltage

W : Watts

Chapitre I :

Généralités sur la spectrométrie de
masse

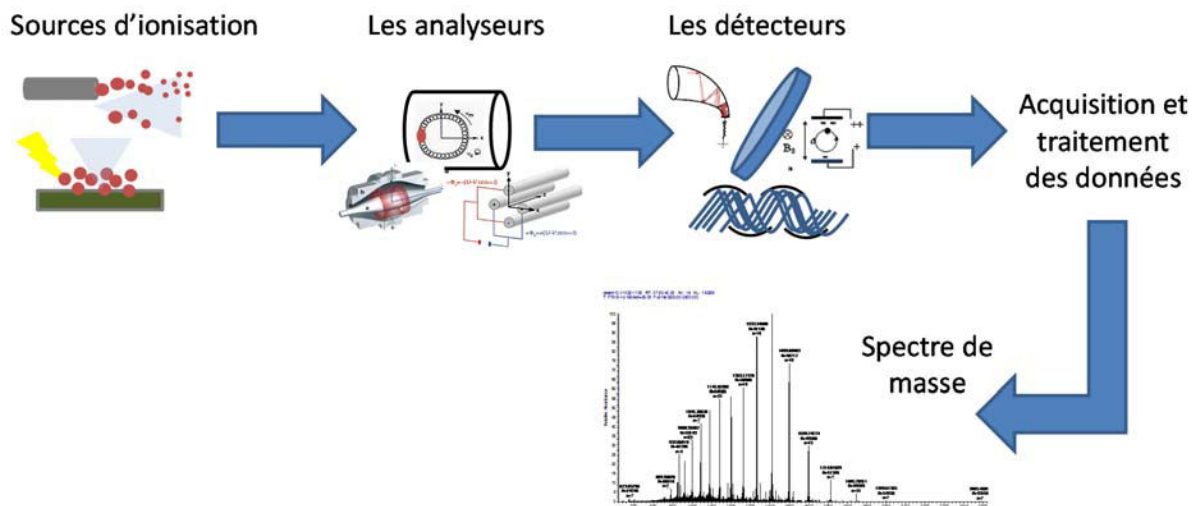
I. La spectrométrie de masse

C'est en 1898 que Joseph John Thomson effectua les premières mesures de rapport masse/charge (m/z) d'électrons. Ses recherches théoriques et expérimentales sur la conductivité électrique dans les gaz lui ont valu un prix Nobel en 1906.

La spectrométrie de masse est une technique analytique puissante utilisée pour quantifier des matériaux connus, pour identifier des composés inconnus dans un échantillon et pour élucider la structure et les propriétés chimiques de différentes molécules. Le processus implique la transformation de l'échantillon en ions gazeux, avec ou sans fragmentation, qui sont ensuite caractérisés par leurs rapports m/z et leurs abondances relatives.

Le spectromètre de masse est constitué de 3 parties principales (figure 1) :

- La source d'ionisation, qui permet de faire passer les molécules en phase gazeuse et de les ioniser,
- L'analyseur, qui sépare les ions en fonction de leur rapport m/z ,
- Le détecteur, couplé à un système informatique, qui assure le traitement des données,



Il existe une grande diversité d'appareils de spectrométrie de masse qui est due à la grande variété de sources, d'analyseurs, de types de couplage et de fragmentation. Bien sûr, toutes les configurations ne sont pas réalisables.

Ces configurations variables d'appareil permettent l'analyse des échantillons de natures différentes et permettent d'obtenir des informations variées. La résolution et la sensibilité seront différentes d'un appareil à un autre à cause de leur configuration.

Les deux caractéristiques instrumentales essentielles pour la mesure des ions par spectrométrie de masse sont le pouvoir résolutif et la précision de mesure en masse. Une bonne résolution spectrale va permettre de déterminer l'état de charge des ions, d'avoir accès à la distribution isotopique et de discriminer des ions isobares. Cette résolution est également liée à la précision de mesure de masse, qui correspond à la finesse du pic observée.

II. La résolution

Il existe plusieurs définitions de la résolution en spectrométrie de masse, qui permettent de comprendre les distinctions entre les différentes définitions et pour comprendre les caractéristiques des différents spectromètres de masse

La première définition est la résolution unitaire. Elle signifie qu'il est possible de séparer chaque masse de la prochaine masse intégrée. C'est-à-dire qu'il est possible de distinguer la masse 50 de la masse 51 ou la masse 700 de la masse 701. Cette résolution est utilisée pour les analyseurs de type quadripolaire ou piège ionique. Les formes des pics dans les spectromètres de masse quadripolaires sont habituellement caractérisées par un sommet plat.

Par exemple pour les secteurs magnétiques, il est possible de définir une puissance de résolution par : $R = m/\Delta m$, c'est-à-dire, la masse d'une molécule m_1 divisée par la différence entre la masse de la molécule m_1-m_2 qui peuvent être séparées. Les spectromètres de masse de type secteur magnétique ont des formes de pic triangulaires ou gaussiennes. Dans ces spectromètres de masse, les pics sont généralement définis pour être séparés à 10 % de la vallée, c'est-à-dire, un point qui est 1/10 de la hauteur du plus haut des deux pics.

Pour les spectromètres de masse à transformée de Fourier, la même définition de la résolution est utilisée, mais en utilisant la largeur des pics à mi-hauteur (figure 2), ceci en raison de l'élargissement à proximité de la ligne de base, dû à la forme des pics non Lorentzien. Le pouvoir de résolution est inversement proportionnel à la masse dans les spectromètres de masse FT-ICR.

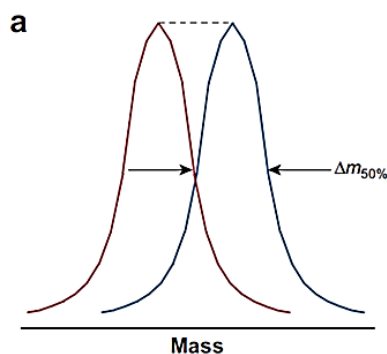


Figure 2: Pouvoir résolutif (ou résolution FWHM : « full width at half-maximum ») correspondant à $m/\Delta m_{50\%}$ pour deux masses (de Marshall et al, 2008 [1]).

La résolution d'un appareil est très importante lors de la détermination de la masse exacte d'une molécule car, grâce à celle-ci, il sera possible de séparer deux masses proches l'une de l'autre et de détecter les isotopes d'une molécule (par exemple : massif multichargé de protéine).

Il existe plusieurs niveaux de résolution spectrale qui peuvent être représentés sous forme de « marches » (figure 3). Avec une amélioration du pouvoir de résolution des instruments, il est possible d'observer de nouvelles familles de pics. Sur la figure 3, l'axe des ordonnées définit la hauteur de chaque marche, c'est-à-dire l'ordre de grandeur de la distance en Dalton séparant deux pics consécutifs d'un spectre. L'axe des abscisses $1/\Delta m_{50\%}$ (où $\Delta m_{50\%}$ correspond à la largeur d'un pic à mi-hauteur) équivaut au pouvoir résolutif de l'instrument ($m/\Delta m_{50\%}$) rapporté à la masse m de l'espèce mesurée. Une représentation du pouvoir résolutif indépendante de la masse est ainsi obtenue.

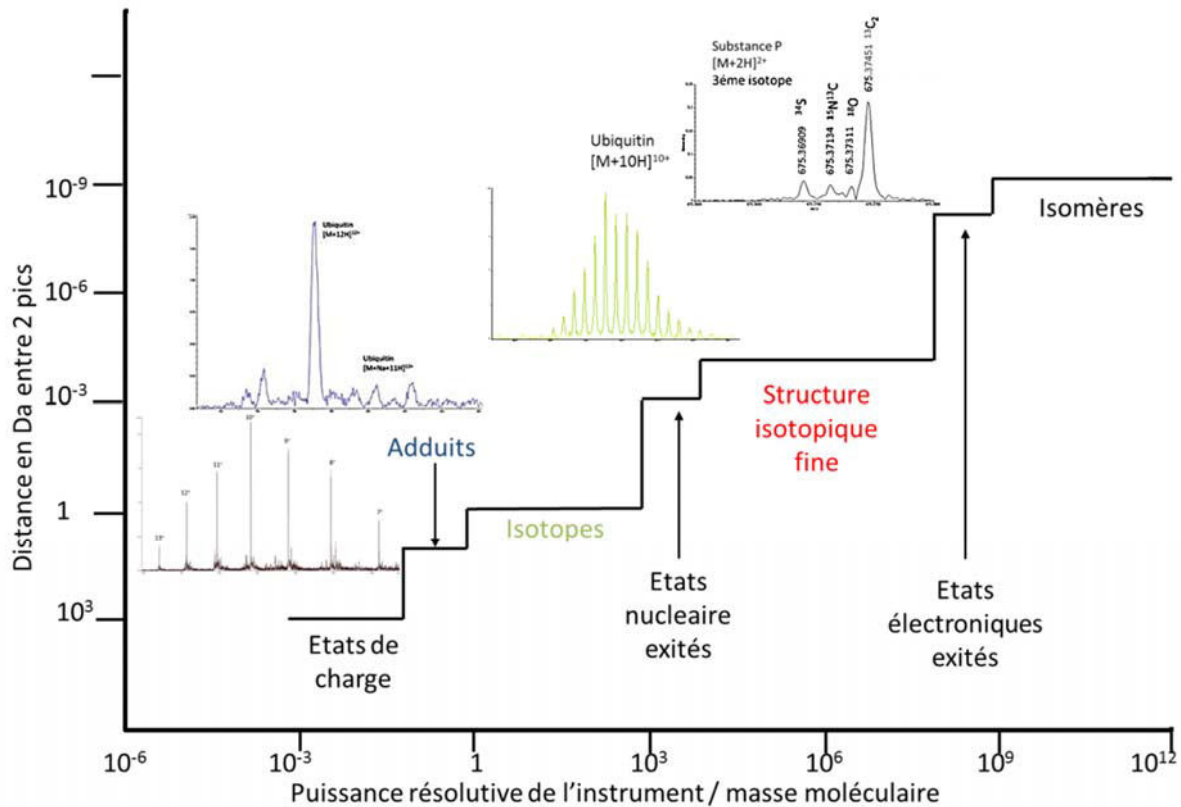


Figure 3: Résolution spectrale (distance en Da entre 2 pics à résoudre) en fonction du pouvoir résolutif de l'instrument rapporté à la masse moléculaire de l'espèce mesurée. En insert, spectres de masse de l'ubiquitine de masse moyenne 8,500 Da et structure isotopique du 3^{ème} isotope de la substance P. (Adapté de Marshall et al [2].)

- A $m/\Delta m_{50\%} > 100$, la séparation des charges des ions est possible $(M+H)^+$ versus $(M+2H)^{2+}$,
- A $m/\Delta m_{50\%} > 1\ 000$, il y a séparation des pics de différentes masses nominales (325 Da versus 326 Da),
- A $m/\Delta m_{50\%} > 10\ 000$, résolution des petits peptides (< 2 500 Da) de même masse nominale qui diffère d'un acide aminé (excepté pour les isomères de leucine et isoleucine),
- A $m/\Delta m_{50\%} > 100\ 000$, séparation des pics via leur espèce isobarique (molécules avec la même masse nominale mais une composition élémentaire variable, N₂ versus CO),

- A $m/\Delta m_{50\%} > 1\ 000\ 000$, séparation des pics avec une composition différentes, par exemple C₃ versus SH₄ qui correspond à une différence de 0,0034 Da.

Les différentes configurations d'instruments sur le marché possèdent différentes résolutions (tableau 1). Les spectromètres de masse quadripolaires ou avec une piège ionique possèdent le plus faible pouvoir de résolution. Ces appareils sont en mesure de détecter les états de charge d'une protéine ou la distribution isotopique des peptides.

Si ces spectromètres de masse quadripolaires sont couplés avec des analyseurs de type temps de vol, la résolution augmente très fortement passant de quelques daltons de résolution à environ 80 000 pour un appareil de type Q-TOF. Alors ces appareils peuvent, par exemple, mettre en évidence la distribution isotopique d'espèces multichargées.

Enfin, les analyseurs les plus résolutifs sont les FT-ICR qui possèdent une résolution pouvant atteindre 10 millions pour les dernières générations équipées de cellules harmonisées. Cette résolution permet de détecter l'enveloppe isotopique d'une protéine multichargée, ou séparer les pics via leurs isotopes.

Analyseur	Quadripôle	Triple quadripôles	Piège ionique	TOF/TOF	Q-TOF	FT-ICR	Orbitrap	Orbitrap
Instrument	ACQUITY Qda	Xevo TQ-XS	amaZon SL	ABI 4800 +	Maxis II	Solarix XR	Q-exactive	fusion Lumos
Résolution	0,7 Da	0,5 Da	0,6 Da	25 000	80 000	10 000 000	240 000	1 000 000
Gamme de masse m/z	1 500	2 000	4 000	200 000	20 000	6 000	6 000	6 000
Fournisseur	Waters	Waters	Bruker Daltonics	Absciex	Bruker Daltonics	Bruker Daltonics	Thermo scientific	Thermo scientific

Table 1 : Informations des différents instruments disponible sur le marché (données provenant des sites des constructeurs).

III. La précision en masse

La précision en masse est importante pour déterminer et identifier les molécules. Elle est exprimée en *ppm* (partie par million). Pour les appareils ayant une très grande précision en masse (de l'ordre de quelques *ppm*), il est possible de déterminer les

formules brutes de molécules (exemple tableau 2). La masse moléculaire exacte est calculée à partir de la combinaison des masses atomiques de la composition élémentaire d'un ion ou d'une molécule (habituellement basée sur l'isotope le plus abondant pour chaque atome), par exemple, $12,000000 + 4 \times (1,007825) = 16.0313$ Da qui correspond à CH₄.

Erreur calculée	Nombre de composition proposé	Composition élémentaire	Masse théorique
0 ppm	1	C ₅₃ H ₁₀₃ O ₆	835,77491
-1,6 ppm 0 ppm	3	C ₅₁ H ₁₀₁ N ₃ O ₅ C ₅₃ H ₁₀₃ O ₆	835,77625 835,77491
-4,8 -3,2 0 1,6 3,8	5	C ₄₇ H ₉₇ N ₉ O ₃ C ₄₉ H ₉₉ N ₆ O ₄ C ₅₃ H ₁₀₃ O ₆ C ₅₄ H ₉₉ N ₄ O ₂ C ₄₁ H ₉₇ N ₁₃ O ₄	835,77089 835,77223 835,77491 835,77625 835,77810

Tableau 2 : Tableau indiquant la composition élémentaire déterminée en fonction de l'erreur de masse pour un composé de type tryacylglycérol. La masse à m/z 835,77491 correspond au TG (16:0/18:0/16:0).

L'erreur de masse peut se calculer comme suivant :

- $\Delta m_i = (m_i - m_a)$ en Da
- $\Delta m_i = (m_i - m_a) \times 10^3$ en mDa
- $\Delta m_i = ((m_i - m_a) / m_a) \times 10^6$ en ppm (parts par million)

Avec m_i la mesure à un temps t en dalton (Da) et m_a la mesure théorique.

Le tableau 3 montre la précision de masse des différents types d'analyseurs. Les analyseurs de type FT-ICR possèdent la meilleure précision de masse car ils sont très stables et le nombre d'ions dans la cellule est constant.

Spectromètre de masse	Précision de masse en ppm ou dalton
FT-ICR-MS	0,1 - 1 ppm
Orbitrap	0,5 - 1 ppm
Secteur magnétique	1 - 2 ppm
TOF-MS	5 - 10 ppm
Q-TOF	3 - 5 ppm
Triple quadripôle	0,5 Da
Piège ionique linéaire	0,5 Da

Tableau 3 : précision de masse des différents types d'analyseurs. (données récupérés des sites internet des constructeurs).

Les appareils de type Orbitrap et FT-ICR peuvent être utilisés pour la détermination de masse exacte car ils possèdent le meilleur pouvoir résolutif, mais aussi la meilleure précision de masse. Grâce à ces deux niveaux de précision, le FT-ICR est souvent utilisé pour la pétrologie qui s'intéresse à l'identification de la totalité des constituants du pétrole [3]. Par exemple, les spectres de masse à haute résolution des pétroles bruts peuvent contenir plus de 100 000 pics. Le FT ICR-MS est capable de différencier deux molécules distinctes proches de seulement 0,0005 Dalton de distance en masse atomique.

IV. La limite de masse

Tout spectromètre permet de déterminer le rapport m/z jusqu'à une valeur maximale. La première limite sera fixée par l'analyseur lui-même, à savoir pour le temps de vol jusqu'à m/z 250 000, pour le Quadripôle 6 000 m/z . Mais la masse que pourra détecter le spectromètre prendra en compte également la charge portée par les molécules. Même avec un analyseur de type quadripolaire ayant une limite de masse de 6000 m/z , si une molécule de rapport m/z 10 000 qui possède deux charges celle-ci sera détectée par cette analyse car sa masse sera d'environ m/z 5 000.

V. Les sources d'ionisation

De nombreuses sources d'ions existent et chacune de ces sources repose sur un principe physique différent. La source sera sélectionnée en fonction des caractéristiques du composé à analyser. La volatilisation et l'ionisation des analytes se fait successivement ou simultanément selon le type de source.

Les critères de choix principaux sont :

- La volatilité et la stabilité thermique du composé à analyser,
- La labilité chimique,
- Les fonctions chimiques présentes et leur aptitude à induire une ionisation,
- La taille des molécules,
- Les quantités de produit disponibles,
- Le solvant de solubilisation du composé,
- La méthode d'injection de l'échantillon : directe ou couplage chromatographique.

Il existe de nombreux types de sources d'ionisation qui permettent d'analyser de nombreux types d'échantillons. Chaque source possède des caractéristiques qui lui sont propres, comme la source d'électronébulisation qui utilise des échantillons sous forme liquide ou encore comme la source MALDI qui emploie des échantillons solides ou cristallisés. Les sources sont présentées sur une échelle de temps dans la figure 4.

Généralités sur la spectrométrie de masse

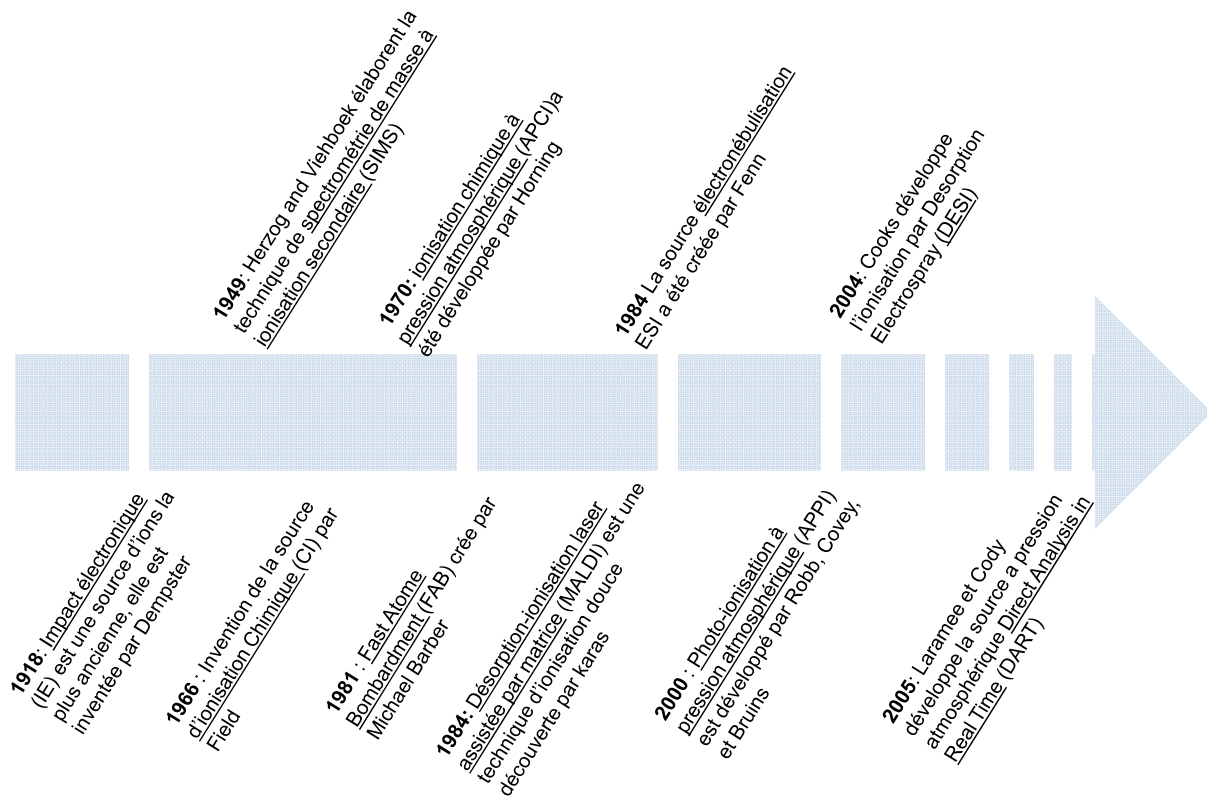


Figure 4 : Les sources de spectrométrie de masse placées sur une échelle de temps, avec les dates de création et leurs concepteurs.

Les sources d'ionisation sont choisies en fonction des molécules analysées. Comme l'illustre la figure 5, l'ESI est mieux adaptée aux composés polaires et aux masses moléculaires élevées, tandis que l'APCI convient mieux aux composés de polarité faible à moyenne. L'impact électronique est généralement utilisé en GC/MS pour les petites molécules volatiles. La source MALDI (qui n'est pas indiquée dans la figure 5) est adaptée aux molécules de petite et grande masse, non volatiles et aux composés thermiquement labiles (comme les protéines, peptides...)

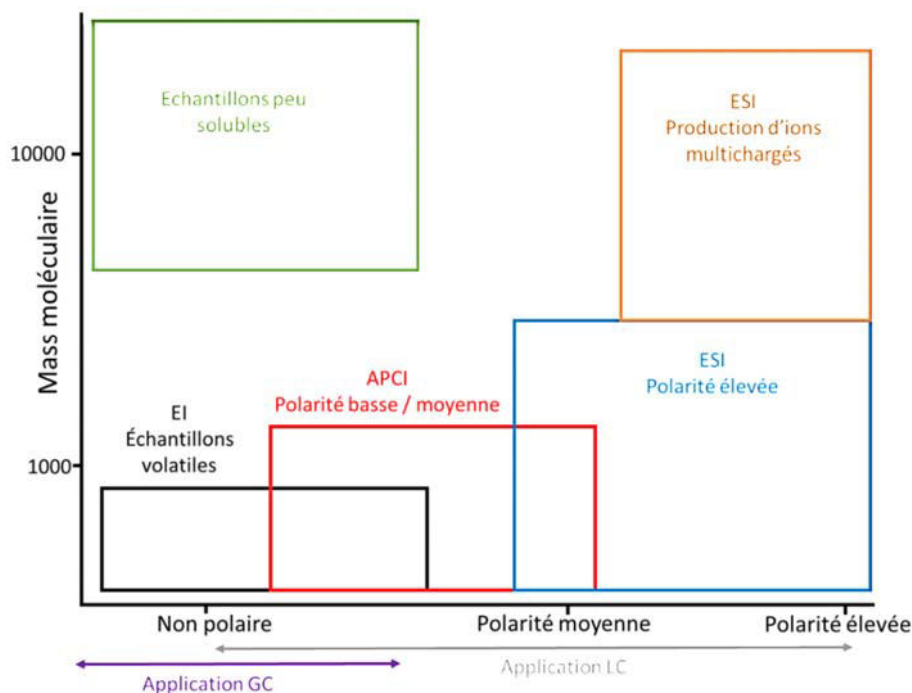


Figure 5 : Graphique montrant les différentes utilisations des sources d'ionisations en fonction de la polarité des molécules et de leurs masses moléculaires

L'étape d'ionisation est importante car elle assure la volatilisation, l'ionisation proprement dite et le guidage vers l'analyseur, mais elle détermine aussi l'énergie apportée aux ions parents.

Le développement de méthodes d'ionisation douces telles que l'ESI (mais aussi la MALDI) a permis à la spectrométrie de masse, depuis les années 80, de se développer dans le monde de la biologie. Dans la suite de cette introduction ne sera développée que la source d'électronébulisation qui a été utilisée pour les analyses de hautes résolutions.

VI. Source électronébulisation (ESI)

Mise au point par le groupe de Fenn (prix Nobel de Chimie en 2002), afin d'étudier l'ionisation des macromolécules biologiques, l'électrospray ou électronébulisation est la technique d'ionisation la plus utilisée qui rapidement élargi son utilisation à d'autres analyses. Elle permet l'étude des échantillons liquides avec une

grande sensibilité ainsi que l'analyse de la majeure partie des réactions chimiques et biologiques qui se déroulent en phase liquide. Le couplage à des techniques séparatives (chromatographie liquide, électrophorèse capillaire) est possible car l'échantillon de départ est liquide. Le point fort de cette technique repose sur le fait que lors du processus d'électronébulisation, l'énergie interne ajoutée aux ions peut être contrôlée afin de ne pas entraîner de fragmentation moléculaire.

L'ionisation de type ESI peut être divisée en trois étapes [4] (figure 6) :

- La nébulisation : formation de microgouttelettes chargées par différence de potentiel entre deux électrodes et par l'application d'un gaz de nébulisation,
- La désolvatation : le solvant des microgouttelettes s'évapore et leur diamètre diminue. La densité de charge au sein de chaque gouttelette augmente jusqu'à ce qu'elle devienne trop importante et qu'il se produise une explosion coulombienne, provoquant une libération des ions en phase gazeuse,
- Le transport des ions depuis la source, qui est à pression atmosphérique, jusqu'au vide poussé de l'analyseur,

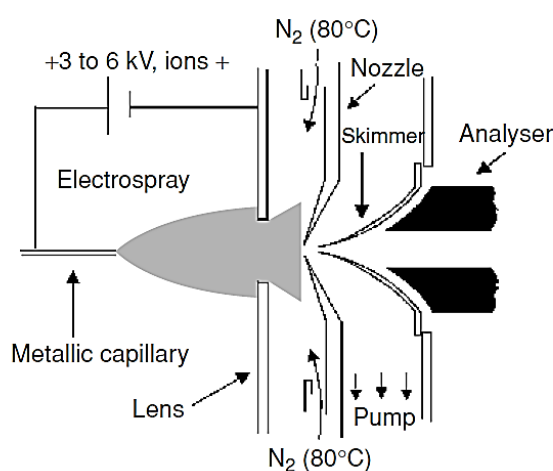


Figure 6 : Principe de l'électronébulisation avec l'utilisation de « skimmers » pour focaliser les ions et de gaz (ici du nitrogène) chauffé pour la désolvatation (extrait de De Hoffman et al. [5]).

1. L'étape de nébulisation

La nébulisation correspond à la dispersion d'un liquide en gouttelettes électriquement chargées grâce à deux processus :

- La formation des gouttelettes,
- L'acquisition de charges,

Ces mécanismes, qui s'effectuent à pression atmosphérique, sont la conséquence de l'application d'un potentiel élevé sur un capillaire au travers duquel passe la solution de l'échantillon à analyser. Un gaz de nébulisation (le plus souvent de l'azote) assure la nébulisation de la solution ou l'action de projeter un liquide en fines gouttelettes. Le champ électrique est créé à l'aide d'une différence de potentiel de quelque kilo Volt (pour la source ESI) entre le capillaire et la contre électrode, séparés de quelques centimètres. Une accumulation de charges à la surface du solvant est provoquée par le champ électrique, au niveau de l'extrémité du capillaire. Le solvant se déforme en prenant l'aspect d'un cône, appelé «cône de Taylor» [6] qui se casse et forme des gouttelettes ayant une densité de charge importante [7],[4]. Lorsque les forces de répulsions coulombiennes deviennent plus fortes que la tension de surface du solvant, de petites gouttelettes électriquement chargées quittent l'extrémité du capillaire [4]. A partir du cône de Taylor, ces gouttelettes primaires, de rayon de 0,5 à 1,5 μm sont émises (figure 7). Ce processus peut être réalisé en mode positif ou négatif pour permettre la formation de d'ions chargés positivement ou négativement. Le changement de polarité des électrodes permet de passer du mode positif au mode négatif.

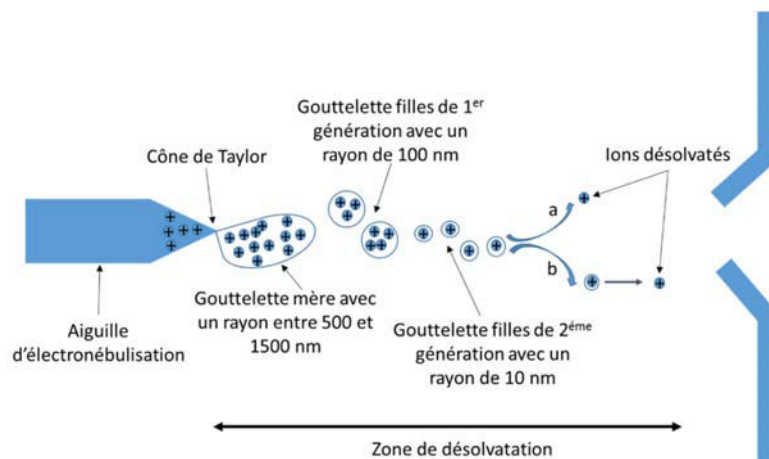


Figure 7 : Libération des ions à partir des gouttelettes émises lors de la nébulisation. a. libération directe des ions à partir de la surface des gouttelettes, b. éjection d'un ion encore solvaté et dont le solvant résiduel sera évaporé.

2. La désolvatation et l'ionisation

Au cours de leur trajet au niveau de la source, des forces de friction et le champ électrique intense déforment les gouttelettes. Les charges de surface augmentent

localement à cause de la combinaison de ces deux phénomènes. Chaque gouttelette se déforme localement en un cône de Taylor. Une vingtaine de nouvelles gouttelettes sont émises à partir de chacune des gouttelettes mères. Ces gouttelettes filles de première génération ont un rayon d'environ 100 nm. Le phénomène de fission coulombienne semble être atteint avant limite de Rayleigh qui est le point où les répulsions coulombiennes entre les charges égalent les forces cohésives de la tension de surface du liquide [8], [5].

Pour ce dernier point, deux théories sont proposées :

- Un ion encore solvaté peut aussi se séparer de la gouttelette (nano gouttelette). Le solvant restant s'évapore durant la fin du vol, ce modèle est appelé « Charge Residue Model » (CRM) [9].
- Le second modèle appelé IEM « mécanisme d'évaporation d'ion » consiste en une succession de fissions des gouttelettes et de nouvelles générations de gouttelettes filles sont ainsi formées [10]. Le phénomène se poursuit jusqu'à ce que la formation de gouttelettes ayant un rayon de l'ordre de 10 nm. La désorption d'ions est provoquée lorsque les répulsions de charges au niveau de ces dernières deviennent suffisantes [11], [12]. Il semblerait que la majorité des ions soit formée selon le modèle IEM.

3. Les ions multichargés

Les molécules de grande taille possédant plusieurs sites ionisables vont produire par ESI des molécules multichargées. La figure 8 présente un exemple de spectre multichargé de « Odorant binding protein » (OBP) obtenus par ionisation nano ESI sur un analyseur de type FT-ICR.

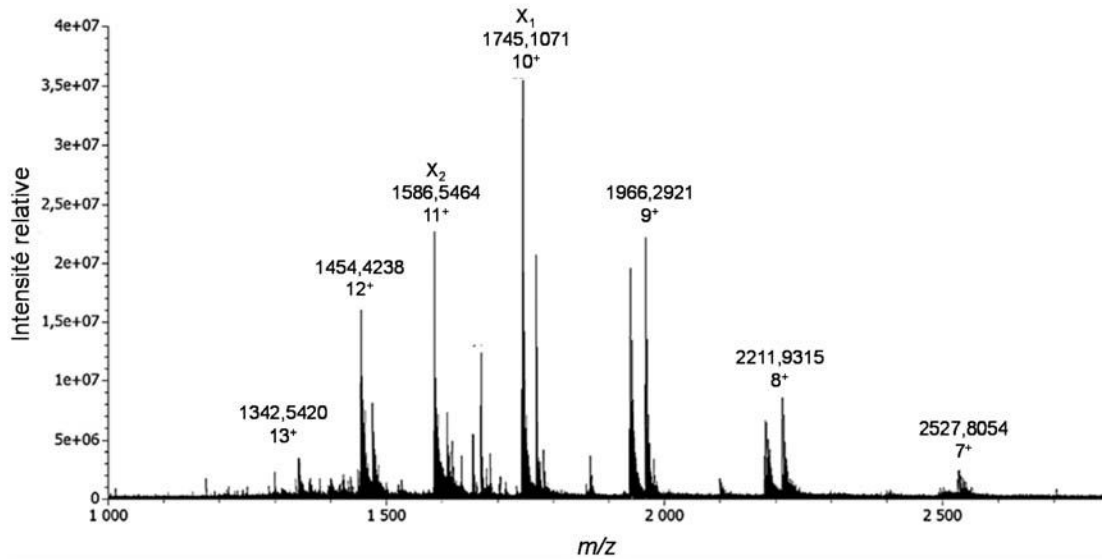


Figure 8 : Spectre multichargé obtenu en nanoESI Apex Qe FT-ICR 9,4T, mode positif, de la protéine OBP (Odorant Binding Protein) à 10 pmol/μL. La masse moyenne calculée est de 17 kDa.

Sur cette figure, la protéine OBP est multichargée et la détermination de la charge des pics se fait grâce à la formule [7] :

$$z_1 = \frac{x_2 - m_H}{x_1 - x_2}$$

Le nombre de charges z du signal de l'ion à $m/z = x_1$, permet de déterminer la masse moléculaire M de la protéine OBP, grâce à la formule :

$$M_1 = (z + 1) \times x_2 - (z + 1) \times m_H$$

L'avantage d'obtenir de molécules multichargées est l'augmentation de la sensibilité qui permet de détecter des molécules de hautes masses moléculaires avec un analyseur avec une limite de masse faible.

Les ions chargés générés par ESI ne reflètent pas l'état de charge des molécules dans la solution analysée, mais ils sont le résultat de l'accumulation des charges dans la gouttelette et des modifications des charges par le processus d'électrochimie à la pointe du cône [13].

L'ionisation ESI est également sensible à la concentration et non pas à la quantité totale de l'échantillon injectée dans la source. C'est la raison pour laquelle des modifications ont été réalisées sur les sources ESI aboutissant à des sources microelectrospray et nanoelectrospray utilisables à faibles concentrations et à faibles

débit lors du couplage avec des chromatographies liquides. Ceci a permis de diminuer la limite de détection et d'atteindre des quantités détectées de l'ordre de 10^{-15} moles [14,15].

VII. Les analyseurs

Les ions produits sont séparés d'après leur rapport m/z et leur masse est simultanément déterminée au moyen de l'analyseur.

De nombreux analyseurs sont disponibles (figure 9) mais, comme pour les sources d'ionisation, l'analyseur parfait n'existe pas. Chaque analyseur présente des avantages et des inconvénients.

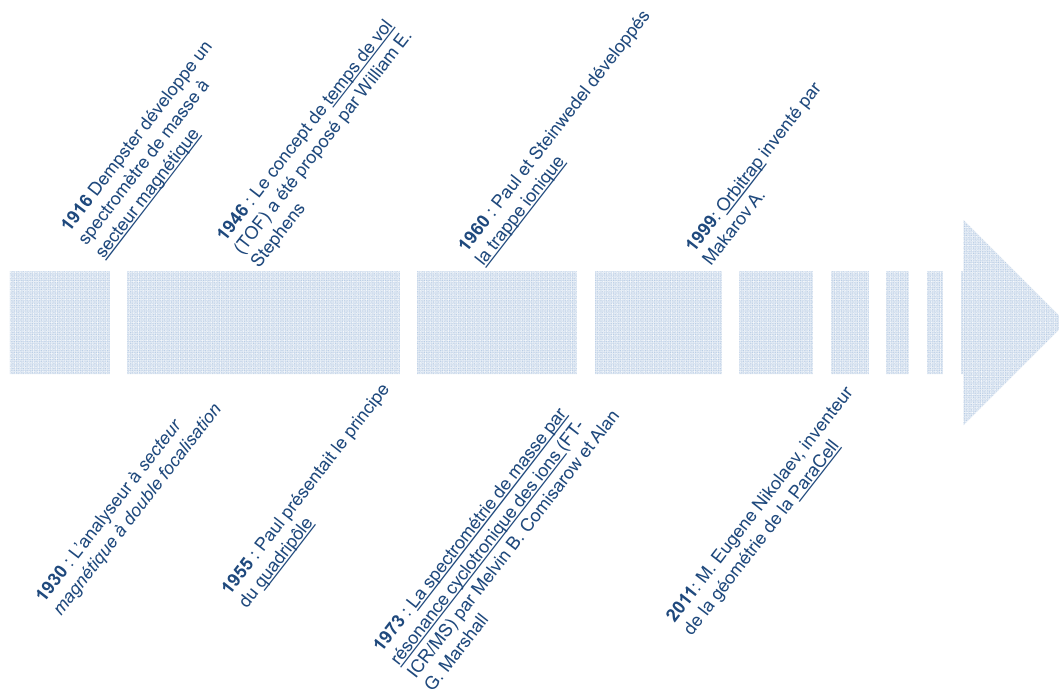


Figure 9 : Analyseurs développés pour la spectrométrie de masse avec les dates de création et les concepteurs.

Les analyseurs peuvent être divisés en deux grandes classes :

- Les analyseurs qui transmettent les ions de différentes masses le long d'une échelle de temps (secteur magnétique avec TOF, quadripôle),
- Les analyseurs qui permettent la transmission simultanée de tous les ions (analyseur magnétique, TOF, piège ionique, FT-ICR et orbitrap),

En fonction du type d'analyse à réaliser et du type de source utilisée, le choix de l'analyseur va dépendre principalement de :

- Sa gamme de masse (valeur limite du rapport m/z mesurable et gamme dynamique),
- Sa précision dans la mesure de la masse,
- Sa résolution, qui correspond à sa capacité à différencier des ions de rapport m/z voisins,
- Sa capacité à focaliser les ions.

L'analyseur FT-ICR et Orbitrap qui ont été utilisés dans cette thèse seront détaillés dans les paragraphes suivant.

1. Les analyseurs de type FT-ICR

La spectrométrie de masse par résonance cyclotronique des ions (FT-ICR, Fourier Transform Ion Cyclotron Resonance) a été développée par Comisarow et Marshall en 1973. Cette méthode permet de mesurer le mouvement cyclotron des ions dans un champ magnétique, de calculer leurs fréquences et de convertir ces informations en spectre de masse. Depuis sa création, cette technique a permis d'obtenir la meilleure résolution et de diminuer l'erreur de masse. Elle est devenue un outil puissant pour l'analyse de mélanges complexes. Pour optimiser la polyvalence de la FT-ICR pour l'analyse d'une large gamme d'échantillon, elle a été couplée avec de nombreuses sources : impact électronique (IE), chimique (IC), atmosphérique, fast atom bombardment (FAB), electrospray (ESI), desorption, Désorption-ionisation laser assistée par matrice (MALDI), photo ionisation à pression atmosphérique (API), bombardement par atomes métastables.

Les optiques de transfert ont été développées et optimisées pour permettre le pompage différentiel entre les sources à la pression atmosphérique et la cellule ICR pour permettre d'accumuler des ions et donc d'améliorer la sensibilité.

Pour obtenir de plus amples informations structurales, de nombreux modes de fragmentation ont également été mis en œuvre :

- La dissociation induite par collision (CID),
- La dissociation par infrarouge multi photon (IRMPD),

- La dissociation par capture d'électron (ECD),
- La dissociation par transfert d'électron (ETD),
- L'irradiation soutenue sans résonance (SORI),
- La dissociation radiative infrarouge du corps noir (BIRD),
- La photodissociation par ultra-violet (UVPD),

Ce type de spectromètre peut réaliser des expériences de fragmentations, soit dans une cellule de collision avant la cellule FT-ICR, soit dans la cellule ICR elle-même. Ces techniques de fragmentation ont pour objectif d'obtenir des informations détaillées sur la structure moléculaire des ions.

Les spectromètres de masse de type FT-ICR sont des analyseurs permettant d'obtenir des spectres MS d'une très grande résolution (jusque 10 000 000) et d'une très grande précision (quelques *ppm* d'erreurs). Un autre avantage est leur capacité à piéger les ions au sein de la cellule ICR.

Les FT-ICR sont constitués d'une cellule composée d'une paire de plaques excitatrices et d'une paire de plaques réceptrices (figure 10). Cette cellule est placée au sein d'un aimant qui génère un champ magnétique uniforme B . Les ions sont piégés au sein de la cellule ICR grâce à la combinaison du champ magnétique B et d'un champ électrique quadripolaire constant. Le champ magnétique piège les ions sur une trajectoire circulaire : c'est le principe du cyclotron. La fréquence de ce mouvement cyclotronique et sa vitesse angulaire dépendent à la fois du champ magnétique et du rapport m/z des ions piégés. [16,17]

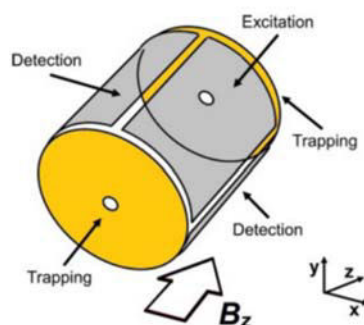


Figure 10 : Schéma d'une cellule ICR composée de deux plaques de détection, deux plaques d'excitation et deux plaques de piégeages devant et derrière permettant de garder les ions dans la cellule. Un champ magnétique B est généré par un aimant (extrait de Yulin Qi and Peter B. O'Connor [3]).

a) Théorie de la FT-ICR

Si la vitesse d'un ion est faible et si le champ magnétique est intense, le rayon de courbure de la trajectoire de l'ion devient faible. Cet ion peut alors être piégé sur une trajectoire circulaire dans le champ magnétique. Il subit une force de Lorentz qui induit un mouvement circulaire de l'ion, perpendiculaire à la direction du champ magnétique : c'est le principe du cyclotron et ce mouvement de rotation est appelé mouvement cyclotron. (figure 11).

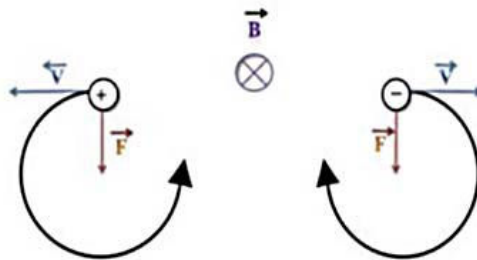


Figure 11 : Illustration de la force de Lorentz ($F = q v B$). Différence de rotation en fonction de la charge positive ou négative de l'ion dans un champ magnétique B

Si un ion est injecté dans un champ magnétique B avec une vitesse v , les équations permettent de calculer :

- La force de Lorentz grâce à l'équation : $F = q v B$
- La force centripète grâce à l'équation : $F' = \frac{m v^2}{r}$

Comme F et F' sont égaux dans cette situation :

$$q v B = \frac{m v^2}{r} \text{ ou } q B = \frac{m v}{r}$$

Avec q la charge de l'ion ($q = z e$), v sa vitesse, m sa masse, r le rayon de sa trajectoire circulaire et B l'intensité du champ magnétique.

L'ion boucle un trajet circulaire de $2\pi r$ à une fréquence ν_c :

$$\nu_c = \frac{v}{2 \pi r} = \frac{q B}{2 \pi m}$$

Sa vitesse angulaire ω s'exprime par l'équation :

$$\omega = 2 \pi \nu_c = \frac{v}{r} = \frac{q B}{m} = \frac{e B}{m/z}$$

La fréquence et la vitesse angulaire de l'ion ne dépendent que du champ magnétique B et de son rapport m/z , suivant la relation ($q B/m$). La vitesse de l'ion n'influence pas la fréquence et la vitesse angulaire de l'ion. La fréquence ν_c qui peut être mesurée permet de déterminer le rapport m/z de l'ion considéré.

b) Piégeage des ions

Lorsque les ions arrivent dans la cellule, ils passent à travers un orifice présent dans la cellule, puis ils sont piégés à l'aide de deux plaques perpendiculaires à l'axe z , appelées « trapping plate » (figure 11). Pour garder les ions dans la cellule sans qu'ils se déchargent sur les plaques, la tension des plaques varie.

Ce principe est appelé piège de Penning ou également résonance d'ion en cyclotron. Dans un cas idéal où des plaques de longueurs infinies sont utilisées et qu'un champ électrique homogène parallèle au champ magnétique est appliqué, le mouvement d'un ion serait hélicoïdal. Les ions auraient des oscillations périodiques suivant l'axe z dans le puits de potentiel électrique.

Une seconde fréquence est formée qui est appelée fréquence de piégeage ν_r

$$\nu_r = \frac{1}{2\pi} \sqrt{\frac{2qV_{trap}\alpha}{ma^2}}$$

V_{trap} : potentiel de piégeage ; α : facteur géométrique ; a : distance entre les plaques. Cette fréquence ne dépend pas de l'énergie cinétique initiale des ions.

c) Mouvement des ions dans la cellule

Le champ électrique généré n'est pas parallèle au champ magnétique car les électrodes ont une dimension finie.

Les particules piégées dans les cellules subissent trois types d'oscillations (figure 12):

1) une oscillation axiale ν_t , et deux oscillations dans le plan radial (similaires au mouvement d'une toupie sur le sol),

2) une oscillation cyclotron réduite ν_+ ou ν_c correspondant à l'orbite circulaire de la particule au sein du piège,

3) une oscillation magnétron ν_- ou ν_m correspondant à la rotation de la particule « sur elle-même »,

La fréquence cyclotron réduite est décrite via l'équation :

$$v_+ = \frac{v_c}{2} + \sqrt{\left(\frac{v_c}{2}\right)^2 - \frac{v_r^2}{2}}$$

Celle de la fréquence magnétron est :

$$v_- = \frac{v_c}{2} - \sqrt{\left(\frac{v_c}{2}\right)^2 - \frac{v_r^2}{2}}$$

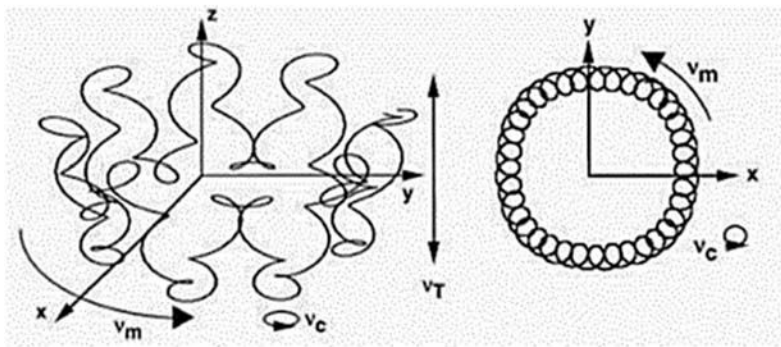


Figure 12 : Les différentes trajectoires d'un ion dans un champ électromagnétique, v_t fréquence d'oscillation axiale, v_m fréquence d'oscillation magnétron, v_c fréquence d'oscillation cyclotron (extrait de Marshall et al [16]).

Pour des rapports de m/z faibles, la fréquence cyclotronique réduite v_+ est très proche de la fréquence cyclotronique v_c .

Pour les ions de rapports m/z élevés (ou des fréquences basses), la situation est différente : il existe donc une région pour laquelle cette approximation n'est plus valable. Dans ce cas la fréquence de piégeage devient plus importante.

d) Mesure du signal FT-ICR

Les ions piégés dans la cellule ICR possèdent des rayons d'orbite et des positions différentes. Cette variation ne permet pas la mesure de leur fréquence cyclotronique.

Donc pour réaliser cette mesure, deux étapes doivent être effectuées :

Première étape : excitation résonante des ions

Cette étape consiste à appliquer un courant électrique entre les deux plaques d'excitation opposées. Ceci a pour effet de modifier les orbites des ions. Cette excitation a lieu à la fréquence cyclotronique d'un ion de rapport m/z avec une tension V_0 appliquée pendant un temps T_{exc} , le rayon de l'orbite de l'ion est donnée par l'équation :

$$r = \frac{V_0 T_{exc}}{B_0}$$

Le rayon de l'ion dépend de la tension et de la durée de l'excitation. Ces deux paramètres permettent d'amener les ions sur une orbite avec un rayon qui les fera passer proche des deux plaques de détection. Cette impulsion qui excite les ions balaye une grande gamme de fréquence. Ceci a pour effet d'amener les ions présents dans la cellule sur un même rayon mais avec des fréquences propres à chaque rapport m/z . Ce phasage des ions est réalisé avec une impulsion rapide.

Il existe différentes méthodes d'excitation (figure 13) :

- L'excitation par impulsion rectangulaire, qui consiste à appliquer une fréquence d'excitation correspondant à une population d'ions de même rapport m/z ,
- L'excitation "broadband", qui permet l'excitation de toute la gamme de fréquences d'intérêt. Deux approches existe : la méthode dite "R.F. chirp" [18] (Frequency-sweep) ou la méthode "SWIFT" [19] (Stored waveform inverse Fourier transform waveforms).

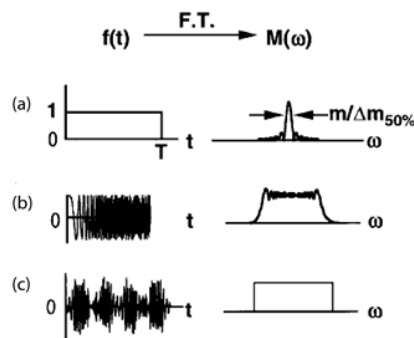


Figure 13 : Méthodes d'excitations avec, les formes d'onde d'excitation du signal transitoire à gauche et le domaine de fréquence à droite. (a) Excitation par impulsion rectangulaire. (b) Excitation "broadband", avec la fréquence-balayage («chirp»). (c) Excitation "broadband", en utilisant la méthode SWIFT (extrait de Marshall et al [16]).

Deuxième étape : Mesure de l'ensemble des fréquences de rotation des ions

La rotation cohérente de l'ensemble des ions (mouvement cyclotron) induit un courant qui est détecté par les deux plaques de détection. Ce courant est ensuite mesuré et digitalisé (interférogramme). Le signal détecté est appelé transient et correspond à la superposition des différents courants induits. Le signal est en fait proportionnel à ce courant induit.

$$\text{L'équation du signal enregistré est décrit par : } \Delta Q = -\frac{2zev_y}{d}$$

Avec ΔQ le courant induit, v la vitesse des ions et d la distance entre les plaques de détection.

La transformée de Fourier va permettre la conversion de l'intensité du signal en fonction du temps, puis l'intensité en fonction du rapport m/z (figure 14).

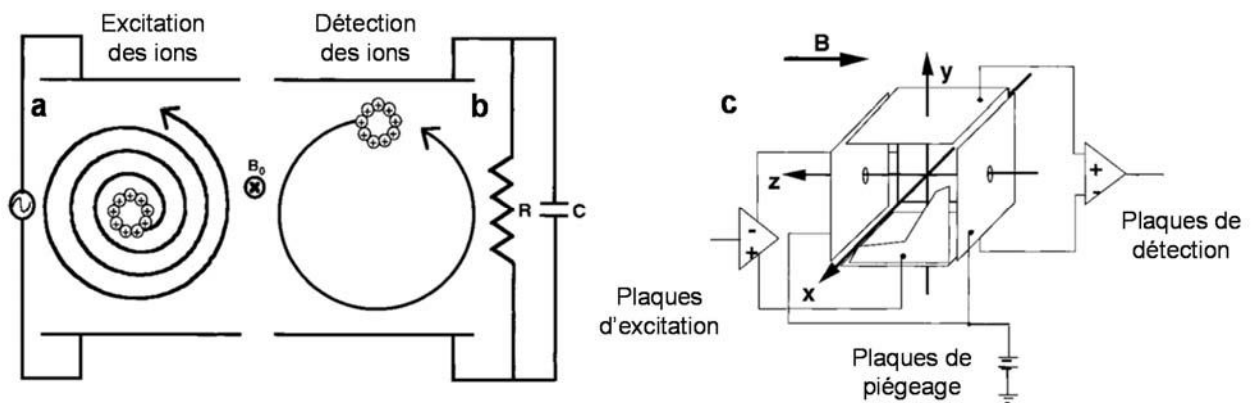


Figure 14 : Le mouvement cyclotron des ions incohérents (a) est converti en un mouvement cohérent (et donc détectable) (b) par l'application d'un champ électrique qui tourne dans le même sens et à la fréquence ICR des ions d'une valeur m/z donnée. (c) Circuits électroniques d'une cellule ICR pour les paires de plaques de détection, excitation et piégeage (extrait de Marshall et al [16]).

Le courant induit apparaît comme étant indépendant de la fréquence des ions et de l'intensité du champ magnétique ou électrostatique. Cependant, le courant induit augmente linéairement avec la charge des ions, donc le signal est plus intense pour un ion multichargé.

Pour les techniques utilisant la transformée de Fourier, la résolution dépend du temps d'acquisition et de la décroissance du signal observé (temps de relaxation). Cette décroissance résulte des collisions ions-molécules dans la cellule qui provoque une diminution de la vitesse des ions et donc de leur rayon.

Pour obtenir des spectres avec une haute résolution, le temps d'acquisition doit augmenter (figure 15). C'est pour cela que ce type d'appareil nécessite un vide poussé de l'ordre de 10^{-10} mbar. La limite principale de ce type d'appareil concerne leur gamme dynamique. En effet, le nombre d'ions présents dans la cellule ICR ne peut excéder 10^5 afin d'éviter un effet de charge trop important.

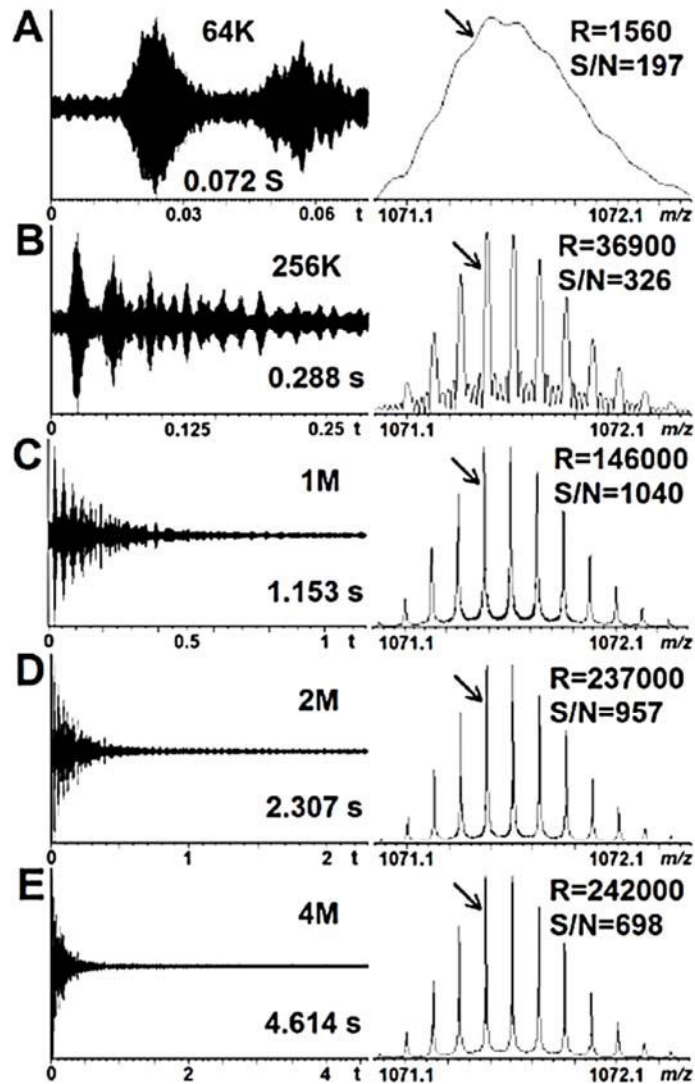


Figure 15 : Spectres de masse de l'ubiquitine acquis avec différents temps de transient ou résolution d'acquisition (64 kpoints jusque 4 mégapoints). Plus le temps du transient est long, plus la résolution spectrale augmente (extrait de Yulin Qi et al. [3]).

La résolution dépend également du champ magnétique de l'aimant. En effet, plus le champ magnétique B augmente, plus la résolution augmente (figure 16).

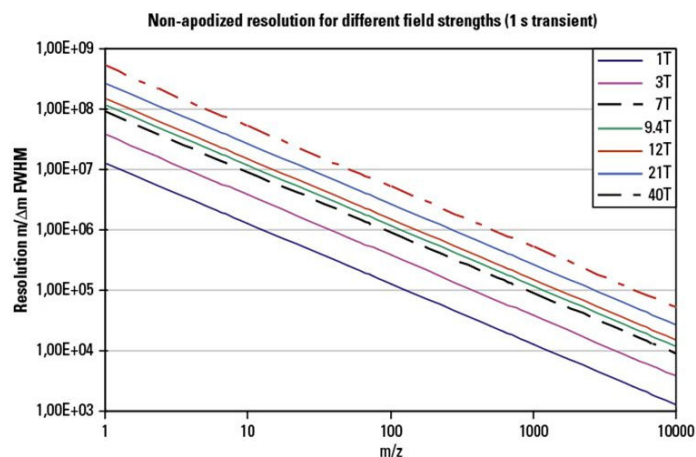


Figure 16 : Effet du champ magnétique de l'aimant sur la résolution du spectromètre de masse FT-ICR (extrait de Scigelova et al. [20]).

e) La cellule harmonisée

La mise au point de la cellule FT-ICR harmonisée a montré que pour des masses de 500-1000 Da (typiquement des peptides), ainsi que pour ions de masses très élevées allant jusqu'à 200 kDa (protéines) [21], la résolution était plus importante [22].

De tels résultats ont été obtenus pour les aimants supraconducteurs d'une homogénéité très élevée du champ magnétique. Pour les aimants avec une plus faible homogénéité, le temps de durée transitoire serait plus petit. Dans les aimants supraconducteurs utilisés dans la spectrométrie de masse FT-ICR, l'inhomogénéité du champ magnétique dans sa direction axiale prévaut sur l'inhomogénéité dans d'autres directions et devrait être considérée comme le principal facteur influençant le mouvement synchrone du nuage d'ions. L'inhomogénéité conduit à une dépendance de la fréquence du cyclotron à partir de l'amplitude de l'oscillation axiale dans le puits potentiel du piège à ions. En conséquence, les ions dans un nuage d'ions deviennent déphasés, ce qui entraîne une atténuation du signal et une diminution du pouvoir de résolution. La fréquence cyclotron de l'ion est également affectée par la composante radiale du champ électrique. Par conséquent, en ajustant de manière appropriée le champ électrique, il est possible de compenser l'inhomogénéité du champ magnétique et aligner la fréquence cyclotron dans toute la gamme des amplitudes des oscillations z. La cellule harmonisée compense l'inhomogénéité du champ magnétique par l'ajout d'électrodes supplémentaires dans la cellule de telle sorte que le champ électrique

moyen créé par ces électrodes produit une contre-force aux forces provoquées par le champ magnétique inhomogène (figure 17)[23]

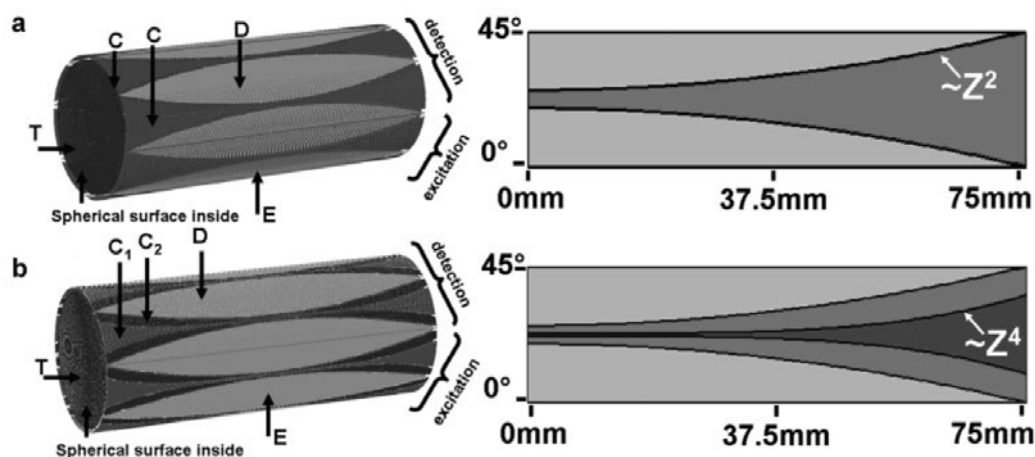


Figure 17: Schéma de la cellule FT-ICR harmonisée. E : plaques d'excitation, D : plaques de détection, T : plaques de piégeages, C électrodes de compensation. a. cellule avec harmonisation dynamique du champ électrostatique, b. cellule avec harmonisation dynamique du champ électrostatique et compensation de l'inhomogénéité du champ magnétique par l'addition de segments en forme de courbe (extrait de Boldin et al [23]).

2. L'analyseur Orbitrap

a) Orbitrap, principe et utilisations

Kindon [24] en 1923 a présenté le principe de l'Orbitrap dans lequel les ions sont piégés par des champs électrostatiques. Ce système n'était pas destiné à la spectrométrie de masse mais aux études de piégeage des ions. Alexander A. Makarov [25,26], à partir des années 2000, a largement contribué au développement de cette technique, notamment dans la façon d'injecter les ions dans la cellule Orbitrap, ainsi que dans la géométrie de l'Orbitrap.

C'est dans les années 2005 que le premier spectromètre de masse Orbitrap commercial, présenté par la société Thermo Fischer, est apparu. De nombreuses versions de spectromètre de masse Orbitrap ont été mises sur le marché : elles offrent de meilleures performances, de meilleures résolutions, des vitesses d'analyse très rapides...

b) Principe de l'Orbitrap

L'analyseur de type Orbitrap est composé d'une électrode creuse externe en forme de cylindre et d'une électrode centrale (Figure 18).

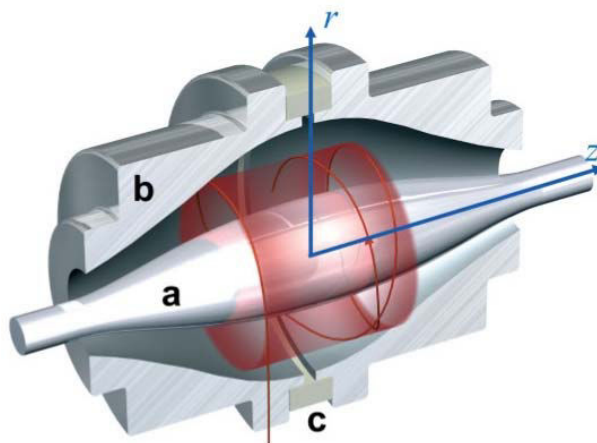


Figure 18 : Coupe d'une Orbitrap avec simulation des ions en rouge. (a) électrode centrale et (b) électrode annulaire (extrait de Scigelova et al [27]).

Les ions piégés subissent une rotation autour de l'électrode centrale et des oscillations le long de sa longueur. Les valeurs m/q des ions piégés sont liées aux fréquences de leurs oscillations harmoniques. Ceci est similaire à un analyseur de masse FT-ICR où les fréquences ioniques sont obtenues à partir du courant d'image induit entre les moitiés axiales de l'électrode externe et sont converties en spectres de masse en utilisant une transformée de Fourier.

Comme l'Orbitrap est également un piège ionique, la stabilité des ions est gouvernée par la solution de l'équation de Mathieu. Le champ électrostatique dans la cellule Orbitrap est décrit par l'équation suivante :

$$U(r, z) = \frac{k}{2} \left(z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \ln \left[\frac{r}{R_m} \right] + C$$

Avec r et z les coordonnées cylindriques, R_m le rayon caractéristique, C une constante, k un paramètre déterminé par la forme de l'électrode centrale et la tension appliquée.

La trajectoire des ions piégés est une spirale complexe qui est une combinaison de la rotation autour de l'électrode centrale et des oscillations le long de cet axe central. Ainsi, trois séries de fréquences peuvent être identifiées le long des trois coordonnées polaires r , φ et z : c'est-à-dire la fréquence des oscillations radiales ω_r , la fréquence de rotation ω_φ , et la fréquence des oscillations axiales ω . Cette fréquence axiale est décrite par l'équation :

$$\omega = \left(\frac{kq}{m} \right)^{1/2}$$

La fréquence axiale des ions ne dépend que du rapport m/z . Ce piège est peu sensible aux effets de charges d'espace et présente une importante capacité en nombre d'ions [28].

c) La C-trap

Les ions produits par la source d'électronébulisation sont guidés à l'aide de funnel et de quadripôle jusqu'à la C-trap. La C-trap est un piège à ions courbé en forme de C (Figure 19) qui permet de compacter les ions pour les envoyer dans l'Orbitrap.

Les ions sont piégés à l'aide d'un potentiel appliqué au niveau de la porte et de la Trap (figure 19). Une fois que les ions se sont accumulés suffisamment dans le piège, l'amplitude RF diminue rapidement et les ions sont rapidement éjectés du piège afin qu'ils ne heurtent pas les électrodes du piège. Ils sont ensuite accélérés et concentrés dans l'Orbitrap [29]. Les ions entrent dans l'Orbitrap après un temps très court.

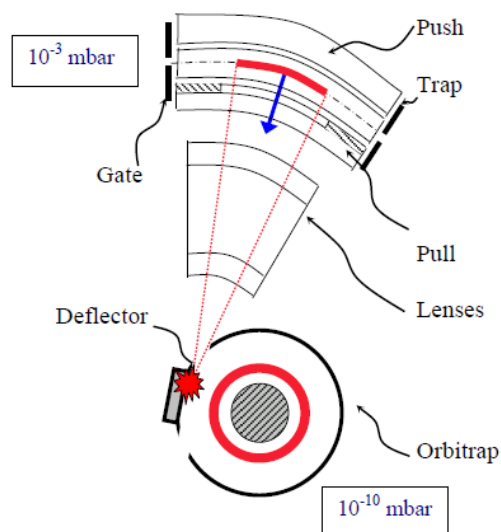


Figure 19 : Schéma de la C-trap et de l'analyseur Orbitrap (de Thermo Fisher Scientific).

d) L'injection des ions

Les ions sont typiquement injectés en un point décalé par rapport au plan x, y central, plus proche de l'une des extrémités étroites du piège. Ils entrent également dans le piège avec des trajectoires tangentes à l'axe central.

Pour des ions positifs, une tension négative relativement importante est appliquée sur l'électrode centrale. La partie radiale logarithmique du champ provoque le pliage des trajectoires d'ions vers l'axe central.

De plus, le potentiel quadratique dans l'axe z fait que les ions commencent leur oscillation axiale. Si le champ électrique restait fixe, les ions frapperaient probablement l'électrode extérieure à l'extrémité opposée du piège. Afin d'éviter ce problème, le courant continu sur l'électrode centrale est modifié très rapidement à une tension plus négative, provoquant le changement du champ électrique avant que les ions puissent se déplacer vers l'autre extrémité du piège.

Une grande tension négative empêchant les ions de se heurter à l'électrode extérieure en provoquant une contraction du mouvement axial des ions. Cette grande tension négative provoque également la contraction du mouvement radial des ions, ce qui provoque une diminution de l'orbite des ions. L'utilisation de ce changement de tension pour permettre une injection et un piégeage efficace des ions est appelée «compression électrodynamique» ou « electrodynamic squeezing » [24] (figure 20).

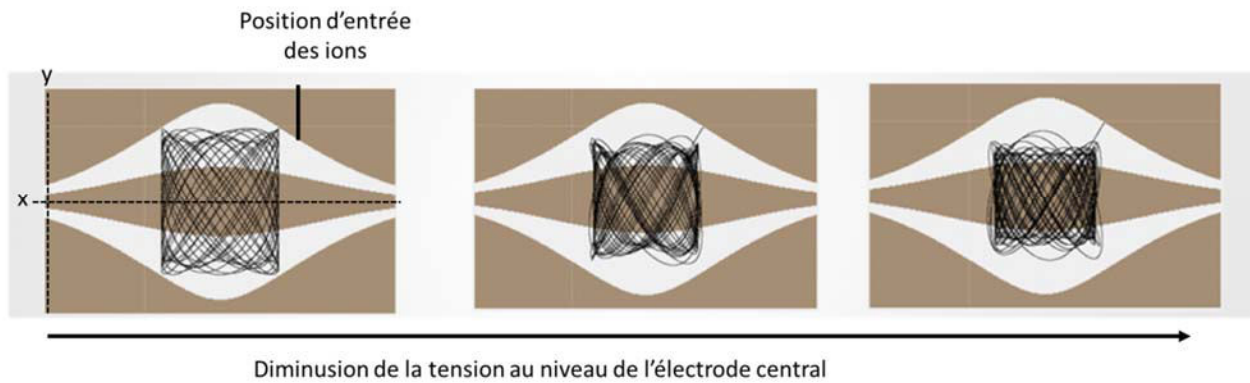


Figure 20 : Principe de la compression électrodynamique des ions au sein de l'Orbitrap.

e) Détection des ions

Par analogie avec le FT-ICR-MS, la détection des ions dans une Orbitrap est basée sur l'excitation à large bande, sur la détection de courant d'image et sur la conversion en spectre de masse par transformée de Fourier [24] (figure 21).

Bien qu'il ne soit pas utilisé pour l'analyse de masse, le mouvement rotationnel (r , φ) des ions dans l'Orbitrap est important pour leur piégeage. Par analogie avec un analyseur électrostatique, dans lequel une correspondance entre l'énergie cinétique ionique et le champ électrique entraînera une trajectoire ionique stable (indépendamment des rapports m/z), un mouvement orbital stable dans l'Orbitrap ne sera obtenu que si l'énergie cinétique ionique avant l'injection est convenablement adaptée à la composante radiale du champ électrique.

La fréquence d'oscillation axiale est indépendante de l'énergie et de la propagation spatiale des ions, ce qui a pour conséquence une grande précision de masse de l'ordre de 1 à 2 ppm et des pouvoirs de résolution jusqu'à 240 000 voire 1 000 000 pour les dernier modèles.

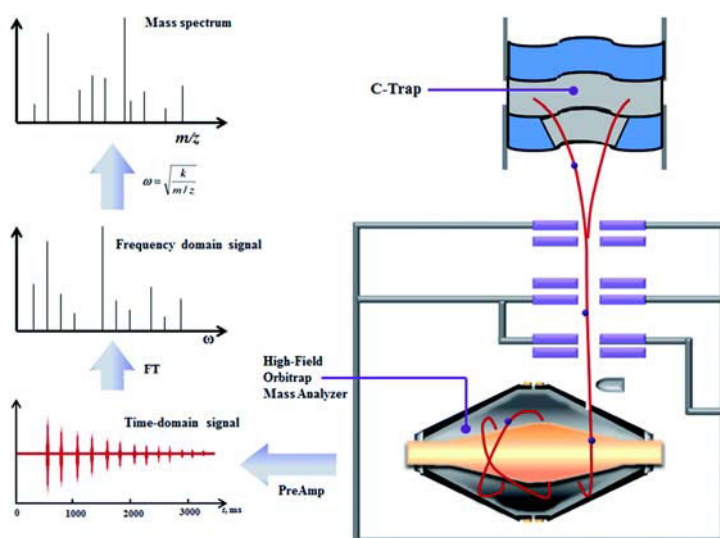


Figure 21: Détection du signal par un analyseur Orbitrap et transformation du signal en spectre de masse (extrait de Longfei Lin et al [30]).

VIII. La spectrométrie de masse en tandem ou MS/MS

La spectrométrie de masse en tandem (MS/MS) permet la détermination de la structure d'un ion sur la base de son profil de fragmentation. L'objectif est d'isoler un ion puis de le fragmenter. Ensuite, l'analyse des fragments permet de déterminer la structure de l'ion parent. Le développement des techniques de MS/MS a permis d'accroître ses domaines d'applications, en particulier dans les sciences du vivant [31,32]

La spectrométrie de masse en tandem porte ce nom, grâce à la conception des premiers instruments ayant permis son application, constitués de deux analyseurs de masse en tandem séparés par une zone de réaction (figure 22).

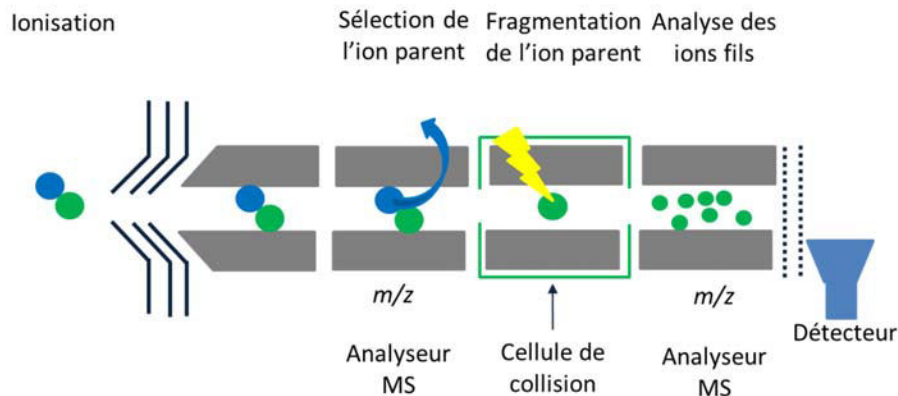


Figure 22 : Principe de la spectrométrie de masse en tandem (MS/MS).

Les analyses de type MS/MS peuvent être décomposées en trois étapes :

- Au cours de la première étape, le premier analyseur sépare les ions formés dans la source d'ionisation selon leur rapport m/z ,
- Durant la deuxième étape, ces ions sont fragmentés dans la zone de réaction,
- Dans la troisième étape, les fragments résultants sont séparés en fonction de leur rapport m/z dans le second analyseur de masse,

Enfin, les ions sont détectés par différentes méthodes suivant le spectromètre de masse (cylindres de Faraday, multiplicateurs d'électrons ou de photons, détecteurs à micro canaux).

Cette technique est très utile dans le domaine de la protéomique car elle permet de déterminer les séquences complètes de peptides ou de protéines à l'issue des étapes de fragmentations [33].

Il existe différents types d'appareils capables de réaliser la spectrométrie de masse en tandem qui sont séparés en trois catégories :

- La MS/MS dans l'espace ou spatiale (deux analyseurs séparés par une cellule de collision, c'est-à-dire un triple quadripôle, secteurs magnétiques et spectromètres de masse hybrides dont les plus courants sont actuellement les Q-q-Q, Q-TOF et QIT),
- La MS/MS dans le temps ou temporelle (un seul analyseur, par exemple, IT, FT-ICR),

- Le cas particulier du PSD (Post-Source Decay sur les instruments de type MALDI-TOF) qui est un mélange des deux précédents,

Il y a beaucoup de manières différentes d'augmenter l'énergie interne des ions afin que les liaisons chimiques se brisent et que des fragments se forment (figure 23). Les méthodes d'activation des ions varient en fonction de la distribution de l'énergie dans les ions activés. Ces facteurs affectent la sélectivité, l'efficacité et la reproductibilité des spectres de masse, et ils peuvent avoir des effets dramatiques sur les ions produits qui sont formés à partir d'un ion activé.



Figure 23 : Méthodes de fragmentation en spectrométrie de masse placées sur une échelle de temps avec les dates de création et les concepteurs.

1. HCD : Higher –energy Collisional Dissociation

Le HCD (Higher –energy Collisional Dissociation) est un mode de fragmentation initialement décrit par Olsen sur un appareil LTQ Orbitrap [34]. Ce mode de fragmentation est très proche du CID (Collision-Induced Dissociation). En effet, la dissociation des ions parents se fait par collision avec un gaz inerte et permet d'obtenir des ions fragments *b* et *y*. Cependant, contrairement au CID qui a lieu dans un piège à

ions linéaire, la fragmentation HCD se déroule dans une cellule de collision octopolaire spécialement conçue, de façon à augmenter l'énergie de fragmentation (figure 24). Cela entraîne deux conséquences principales. Premièrement, par rapport au CID, moins d'ions fragments sont observés sur le spectre. Deuxièmement, la limite du CID, appelée le «low-mass cutoff», se caractérise par l'absence des fragments de basse masse (inférieurs à un tiers de l'ion précurseur) sur les spectres CID classiques. Contrairement au CID, le HCD permet d'observer ces fragments.

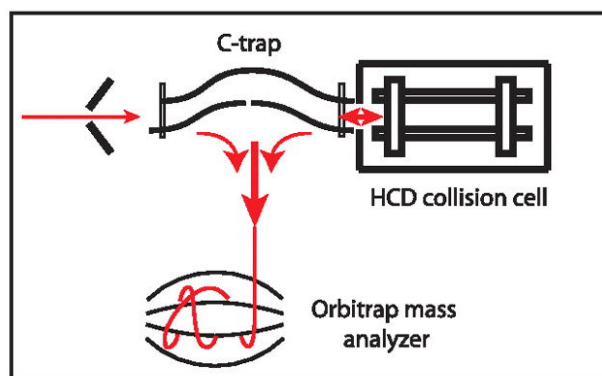


Figure 24 : Schéma de la cellule HCD, C-trap et Orbitrap.

2. IRMPD : dissociation Multiphotonique par infrarouge

La dissociation Multiphotonique par infrarouge (IRMPD) est une technique utilisée en spectrométrie de masse pour fragmenter des molécules en phase gazeuse [35]. Un laser infrarouge est dirigé à travers une fenêtre dans le vide du spectromètre de masse, où les ions sont présents. Le mécanisme de fragmentation comprend l'absorption par un ion de plusieurs photons infrarouges donné par le faisceau laser. L'ion parent devient excité dans des états vibratoires plus énergétiques jusqu'à ce qu'une ou plusieurs liaisons soient fragmentées en phase gazeuse. Dans le cas de puissantes impulsions laser, la dissociation se déroule par inner-valence de l'ionisation des électrons [36,37]. La fragmentation IRMPD est le plus souvent utilisée dans de spectromètres de masse FT-ICR [38] et elle permet la fragmentation de peptides ou protéines en formant des ions fragments de type b ou y comme la fragmentation CID [39]. Depuis peu cette méthode a été implanté dans un spectromètre de masse type Orbitrap [40]. Lisa A. Vasicek et al ont modifié une Orbitrap avec une fragmentation

Généralités sur la spectrométrie de masse

IRMPD au sein de la cellule HCD pour permettre la fragmentation de peptides ou protéines.

Bibliographie

- (1) Marshall, A. G.; Hendrickson, C. L.: High-resolution mass spectrometers. *Annu Rev Anal Chem (Palo Alto Calif)* **2008**, *1*, 579-99.
- (2) Alan G.; Hendrickson, C. L. S., Stone D.-H: Scaling, MS plateaus with high-resolution FT-ICRMS Marshall. *Anal. Chem.* **2002**, *74*.
- (3) Rodgers, R. P.; Marshall, A. G.: Petroleomics: Advanced characterization of petroleum-derived materials by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). *Asphaltenes, Heavy Oils, and Petroleomics* **2007**, 63-93.
- (4) Bruins, A. P.: Mechanistic aspects of electrospray ionization. *Journal of Chromatography A*, **1998**, *794*, 345–357.
- (5) Hoffmann, E. d.: Mass Spectrometry, third edition. *wiley* **2007**.
- (6) Taylor, G.: Disintegration of Water Drops in an Electric Field. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1964**, *280*, 383-397.
- (7) John B Fenn, M. M., Chin Kai Meng, Shek Fu Wong, Craig M. Whitehouse: Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64-71.
- (8) Gomez, A.; Tang, K.: Charge and fission of droplets in electrostatic sprays. *Physics of Fluids* **1994**, *6*, 404-414.
- (9) Malcolm Dole, L. L. M., R. L. Hines, R. C. Mobley , L. D. Ferguson , and M. B. Alice: Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49*, 2240-2249.
- (10) Kebarle, P.: On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Anal. Chim. Acta* **2000**, *406*, 11–35.
- (11) Iribarne, J. V., B. A. Thomson: On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **1976**, *64*, 2287.
- (12) Thomson, B. A., Iribarne, J. V.: Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J. Chem. Phys.* **1979**, *71*, 4451-4463.
- (13) Kelly, M. A.; Vestling, M. M.; Fenselau, C. C.; Smith, P. B.: Electrospray analysis of proteins: A comparison of positive-ion and negative-ion mass spectra at high and low pH. *J. Mass Spectrom.* **1992**, *27*, 1143-1147.
- (14) Emmett, M. R.; Caprioli, R. M.: Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 605-613.
- (15) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.: Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379*, 466.
- (16) Marshall, A. G.: Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.
- (17) Nikolaev, E. N.: Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom. Rev.* **2016**, *35*, 219-58.

- (18) Marshall, A. G.; Roe, D. C.: Theory of Fourier transform ion cyclotron resonance mass spectroscopy: Response to frequency-sweep excitation. *The Journal of Chemical Physics* **1980**, *73*, 1581-1590.
- (19) Marshall, A. G.; Wang, T. C. L.; Ricca, T. L.: Tailored excitation for Fourier transform ion cyclotron mass spectrometry. *Journal of the American Chemical Society* **1985**, *107*, 7893-7897.
- (20) Scigelova, M.: Fourier Transform Mass Spectrometry. *MCP* **2011**, *10*.
- (21) Popov, I. A.; Nagornov, K.; Vladimirov, G. N.; Kostyukevich, Y. I.; Nikolaev, E. N.: Twelve million resolving power on 4.7 T Fourier transform ion cyclotron resonance instrument with dynamically harmonized cell--observation of fine structure in peptide mass spectra. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 790-9.
- (22) Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A.: Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2014**, *86*, 317-20.
- (23) Boldin, I. A.; Nikolaev, E. N.: Fourier transform ion cyclotron resonance cell with dynamic harmonization of the electric field in the whole volume by shaping of the excitation and detection electrode assembly. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 122-6.
- (24) Perry, R. H.; Cooks, R. G.; Noll, R. J.: Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **2008**, *27*, 661-99.
- (25) Gillig, K. J., Bluhm, B. K., & Russell, D. H.: Ion motion in a Fourier transform ion cyclotron resonance wire ion guide cell. *International journal of mass spectrometry and ion processes* **1996**, *157*, 129-147.
- (26) Makarov, A.: Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72*, 1156-1162.
- (27) Scigelova, M.; Makarov, A.: Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006**, *6 Suppl 2*, 16-21.
- (28) Kharchenko, A.; Vladimirov, G.; Heeren, R. M.; Nikolaev, E. N.: Performance of Orbitrap mass analyzer at various space charge and non-ideal field conditions: simulation approach. *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 977-987.
- (29) Perry, R. H.; Hu, Q.; Salazar, G. A.; Cooks, R. G.; Noll, R. J.: Rephasing ion packets in the Orbitrap mass analyzer to improve resolution and peak shape. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1397-404.
- (30) Lin, L.; Lin, H.; Zhang, M.; Dong, X.; Yin, X.; Qu, C.; Ni, J.: Types, principle, and characteristics of tandem high-resolution mass spectrometry and its applications. *RSC Advances* **2015**, *5*, 107623-107636.
- (31) McLafferty, F. W.: Tandem mass spectrometry. *Science* **1981**, *214*, 280-287.
- (32) McLafferty, F. W.: Mass spectrometry across the sciences. *Proc Natl Acad Sci U S A* **2008**, *105*, 18088-9.

- (33) Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., & Hauer, C. R.: Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences* **1986**, *83*, 6233-6237.
- (34) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4*, 709-12.
- (35) Daniel P. Little, J. P. S., Michael W. Senko, Peter B. O'Connor, and Fred W. McLafferty: Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**, *66*, 2809-2815.
- (36) Talebpour, A.: Multiphoton ionization of inner-valence electrons and fragmentation of ethylene in an intense Ti:sapphire laser pulse. *hemical Physics Letters* **1999**, *313*, 789–794.
- (37) Talebpou, A.: Dissociative ionization of benzene in intense ultra-fast laser pulses. *J. Phys. B: At., Mol. Opt. Phys.* **2000**, *33*, 4615–4626.
- (38) Laskin, J.; Futrell, J. H.: Activation of large ions in FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 135-67.
- (39) Crowe, M. C.; Brodbelt, J. S.: Infrared multiphoton dissociation (IRMPD) and collisionally activated dissociation of peptides in a quadrupole ion trap with selective IRMPD of phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1581-92.
- (40) Vasicek, L. A.; Ledvina, A. R.; Shaw, J.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S.: Implementing photodissociation in an Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1105-8.

CHAPITRE II

L'analyse bidimensionnelle par
spectromètre de masse à
transformée de Fourier (FT-ICR)

Partie 1 :

Généralités sur la FT-ICR 2D

I. Les premiers pas de la FT-ICR 2D

1. L'expérience de Marshall

En 1984, Marshall et al. [1] ont mis en place la première expérience qui a permis de montrer que suite à l'excitation des ions présent dans une cellule ICR, il est possible de les désexciter.

Cette expérience a été réalisée en deux parties. La première expérience a consisté à exciter des ions avec une impulsion résonnante de durée croissante et de mesurer l'amplitude du signal dans la cellule ICR.

Dans une seconde expérience, la phase de la tension d'excitation a été inversée après 1,0 ms. La figure 1 montre l'amplitude du signal ICR mesuré dans les deux expériences en fonction de la durée de l'impulsion d'excitation. Pour les ions qui sont excités avec une phase constante, l'amplitude du signal ICR augmente avec le rayon cyclotron jusqu'à ce que les ions soient délocalisés ou atteignent les électrodes de la cellule ICR. Au contraire, dans le cas de l'inversion de phase le rayon de l'ion décroît et donc l'amplitude relative du signal relatif ICR diminue. Le signe du potentiel sur les plaques d'excitation est changé et les ions, au lieu d'être attirés par la plaque d'excitation sont repoussés par celle-ci : la force n'est plus centrifuge mais devient centripète.

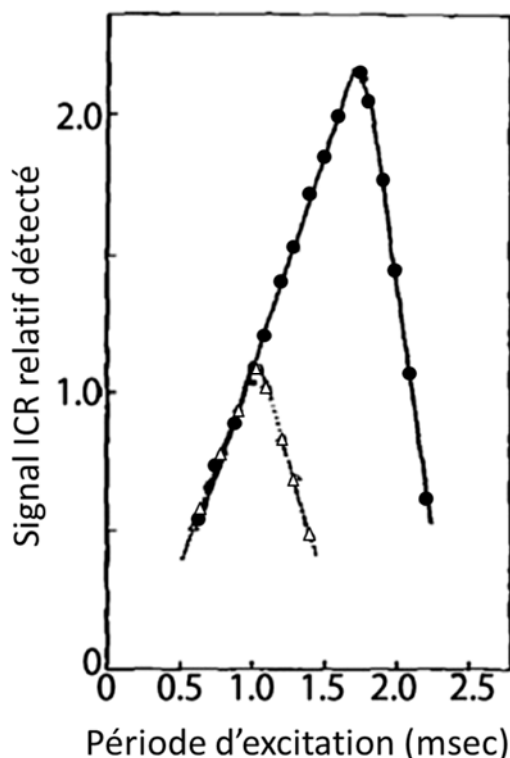


Figure 1 : Signal cyclotron de l'ion $C_9F_{20}N^+$ de la fluorobutylamine à m/z 501,971 après excitation avec un pulse de phase constante et croissant (cercles plein). Après une excitation où la phase est passée à 180° après 1 ms (triangle). D'après Marshall et al [1].

Cette observation a prouvé que l'excitation des ions est fortement linéaire et additive dans la cellule ICR, donc la « de-excitation » des ions est possible dans une cellule ICR aussi longtemps que le paquet d'ions reste cohérent.

2. Première expérience de 2D-FT

a) Théorie

En 1987, le groupe de RMN de Geoffrey Bodenhausen à l'Université de Lausanne et le groupe FT-ICR MS de Tino Gäumann à l'Ecole Polytechnique Fédérale de Lausanne en Suisse ont réalisé les premières expériences de FT-ICR 2D MS [2-4]. Cette expérience est similaire à la spectroscopie nucléaire à effet Overhauser (NOESY) [5] et rappelle la désexcitation sélectives d'ions dans une cellule ICR. La séquence de pulse de cette expérience est présentés sur la figure 2.

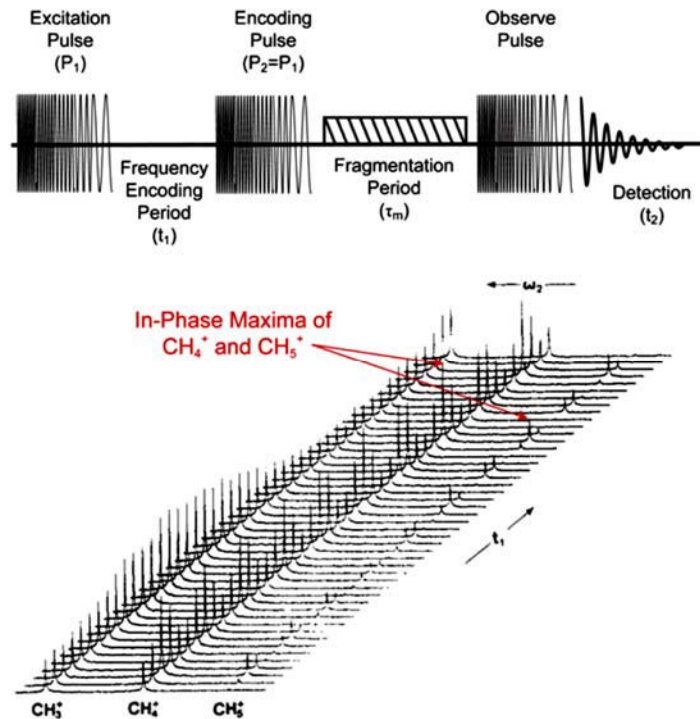


Figure 2 : Séquence du pulse 2D-FT-ICR proposée par Pfandler et al [3]. L'évolution du pic produit par l'ionisation du méthane par ionisation électronique est observée après une évolution suivant l'intervalle t_1 .

Après ionisation près du centre de la cellule ICR, deux impulsions radio fréquence égale P_1 , P_2 sont appliqués aux plaques d'excitation, séparés par un intervalle de temps t_1 d'encodage. Puis durant un temps τ_m les ions évoluent par réaction ion-molécule et une troisième impulsion P_3 est appliquée pour détecter tous les ions.

Le rayon d'un l'ion après le premier pulse $P1$ est régi par l'équation :

$$r_{ion} = \frac{V \times t_{excitation}}{2B \times r_{cellule}}$$

Avec B le champ magnétique, V la vitesse de l'ion, $r_{cellule}$ le rayon de la cellule ICR et $t_{excitation}$ le temps d'excitation de l'ion.

Au départ, les ions sont au centre de la cellule et possèdent une énergie cinétique faible. La fréquence cyclotron de l'ion est exprimé par :

$$\omega_{ion} = \frac{qB}{m} \text{ avec } m \text{ la masse de l'ion, } B \text{ le champ magnétique et } q \text{ la charge de}$$

l'ion. C'est ions possèdent une phase $\varphi = 0$.

La première impulsion excite les ions de manière cohérente sur un grand rayon cyclotron avec une énergie cinétique plus importante que précédemment. Cette impulsion est appelée « impulsion d'excitation ». Puis, pendant la période t_1 , les ions tournent en paquet en fonction de leur fréquence cyclotron. Une deuxième impulsion radio fréquence identique à la première est appliquée. Celle-ci n'a pas le même effet, étant donné que les ions au début de la deuxième impulsion ont un grand rayon et leur phase relative est proportionnelle au produit de leur fréquence cyclotron et du temps t_1 (cette impulsion est appelée impulsion de codage). La phase de l'ion s'écrit $\varphi = \omega_{ion} t_1$. Le rayon cyclotron de l'ion à la fin de la deuxième impulsion peut s'écrire :

$$r_{ionP_2} = r_0 \sqrt{2 \times (1 + \cos \omega_0(t_1 - T))}$$

Avec r_0 le rayon cyclotron après la première impulsion P_1 de durée T , ω_0 la fréquence de l'ion et t_1 le temps entre les deux impulsions P_1 et P_2 . Cette relation est identique pour les impulsions monochromatiques et pour les impulsions balayées en fréquences (ou « chirp »).

Le rayon cyclotron des ions à la fin de la seconde impulsion dépend de la fréquence cyclotron et du temps t_1 . A la fin de la seconde impulsion, les ions sont donc répartis sur différents rayons cyclotron en fonction de leur fréquences cyclotron. Lorsque la phase relative est égale à $2\pi n$ (avec $n = 0, 1, 2, \dots$) le rayon final est deux fois le rayon après le pulse P_1 . Au contraire lorsque $\omega_{ion} t_1 = \pi (2n + 1)$, le paquet d'ions est dé-excité et revient à son état initial (figure 3). Le résultat est que le rayon du paquet d'ion primaire est alors modulé avec la fréquence ω_{ion} [2,6,7].

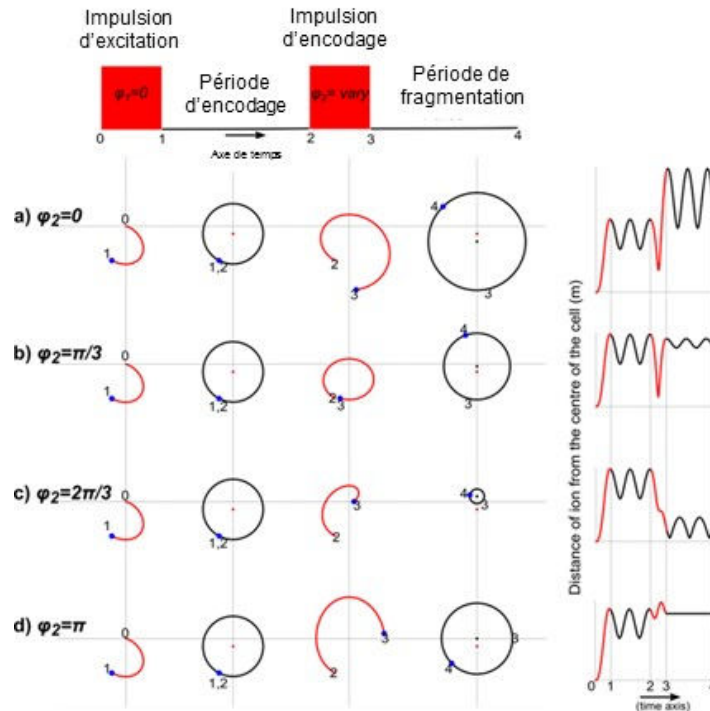


Figure 3 : Trajectoire de l'ion doublement chargé de la substance P dans un champ de 8,78 T avec une fréquence de $\omega_{ion}=200$ kHz. L'ion subit une 1^{er} impulsion d'une durée $T_{P1}=5 \mu s$ puis il évolue en fonction de t_1 . Il subit ensuite une seconde impulsion $P_2=5 \mu s$. Les deux impulsions P_1 et P_2 sont égales. La fréquence de l'ion est $\omega_{ion}T = 2\pi$ et la phase de l'ion varie de $0 < \varphi < 2\pi$ le point bleu représente l'ion, en rouge la trajectoire de l'ion lors des impulsions et en noir la trajectoire de l'ion après les impulsions. À droite est représentée la distance des ions du centre de la cellule après l'impulsion P_1 , le temps t_1 et l'impulsion P_2 (d'après Akansha et al [7]).

Lorsque les ions sont désexcités et reviennent au centre avec $\omega_{ion}t_1 = 2\pi n + \pi$, ces ions peuvent être fragmentés ou réagir durant un temps t_m (figure 3a). Ensuite, une impulsion P_3 est appliquée pour exciter tous les ions et permettre leur détection par les plaques de détection de la cellule. Le signal de l'ion parent peut être exprimé comme le produit de l'abondance de l'ion parent et du rayon cyclotron.

b) FT-ICR 2D MS sur les ions méthane

Dans la première expérience de FT-ICR 2D MS [3], des ions du méthane ont été produits par ionisation d'électrons et excités à différents rayons cyclotron par les deux premières impulsions. Ensuite, ils sont amenés à entrer en collision réactive lors d'un intervalle de mélange t_m avec du méthane, qui a été injecté dans la cellule ICR. Par conséquent, l'abondance des ions ayant réagi est une fonction périodique du rayon gyromagnétique des ions précurseurs et par conséquent de l'intervalle t_1 .

Après la réaction ou la dissociation métastable, les ions sont re-excités en utilisant une troisième impulsion d'excitation (ou impulsion de détection) afin de détecter les ions précurseurs et fragments. L'amplitude du signal ICR des ions précurseurs est proportionnelle à leur abondance. CH_5^+ est un ion qui est créé par les réactions molécule-ions entre le radical cation $\text{CH}_4^{+\cdot}$ et le méthane neutre [8]. La fréquence du signal ICR de CH_5^+ est donc égale à la fréquence cyclotron de son précurseur $\text{CH}_4^{+\cdot}$. Ceci est évident sur la figure 4 où l'intensité du pic de CH_5^+ est maximale lorsque le pic de $\text{CH}_4^{+\cdot}$ est également maximale. Le cation CH_3^+ est à la fois un précurseur qui est présent au début de la séquence d'impulsion et un produit de collisions impliquant $\text{CH}_4^{+\cdot}$. Son signal ICR comporte donc deux fréquences dans l'intervalle t_1 . Sa propre fréquence cyclotron et la fréquence cyclotron de $\text{CH}_4^{+\cdot}$ figure 4.

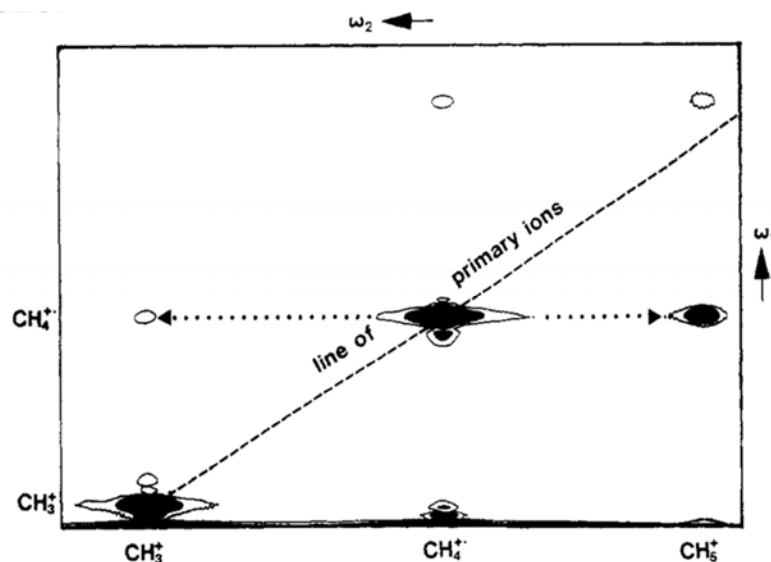


Figure 4 : Spectre FT-ICR 2D du méthane. Observation de l'ion parent CH_3^+ et $\text{CH}_4^{+\cdot}$ avec une modulation selon leur propre fréquence cyclotron et observation de l'ion fragment CH_3^+ provenant de l'ion précurseur $\text{CH}_4^{+\cdot}$ avec une modulation selon la fréquence cyclotron de l'ion parent [3].

En enregistrant les spectres de masse FT-ICR, avec des incréments de temps de $t_1 = n_1 \times \Delta t_1$, il est possible d'observer la modulation du rapport de l'amplitude du signal ICR de tous les ions dans un échantillon et de les mettre en corrélation avec l'amplitude du signal ICR de leurs fragments.

Δt_1 l'incrément de temps donne la fréquence d'échantillonnage qui fixe la fréquence de Nyquist (la fréquence maximale que doit contenir un signal pour permettre

sa description non ambiguë par un échantillonnage à intervalles réguliers qui vaut $f_{Nyquist} = \frac{f_{\text{échantillonnage}}}{2}$, donc le plus petit rapport m/z dans la dimension verticale. Sachant que n_{1max} est le nombre maximum d'incrément, n_{1max} donnant la résolution dans la dimension verticale.

Il suffit de calculer la transformée de Fourier de chaque transitoire enregistré en tant que fonction de l'intervalle t_2 de détection et de calculer une autre transformée de Fourier en fonction de l'évolution de t_1 . La figure 5 montre le cheminement du traitement du signal 2D-FT.

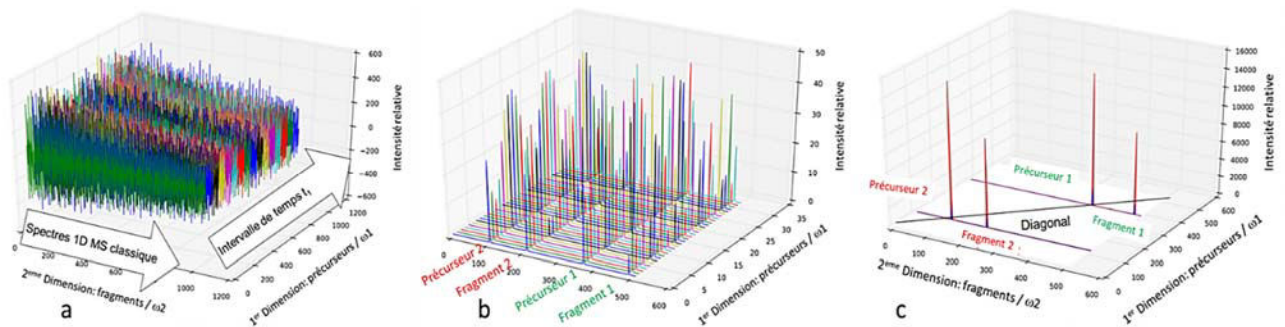


Figure 5 : Cheminement du traitement du signal 2D-FT transformée de Fourier. (a) FID enregistrée pour chaque temps de l'intervalle t_1 . (b) Application de la transformée de Fourier dans l'axe F2 et obtention des spectres de masse pour chaque intervalle t_1 , les parents et les fragments sont modulés à la même fréquence (c) Application de la transformée de Fourier dans l'axe F1 et obtention du spectre 2D-FT, corrélation entre le signal des précurseurs et des fragments.

Dans le spectre 2D de fréquence cyclotron résultant, tous les pics fragments des ions apparaissent le long de l'axe ω_2 (généralement tracés horizontalement) et de leurs précurseurs apparaissent le long de l'axe de ω_1 (qui est par convention tracés verticalement). Le schéma du spectre FT-ICR 2D (figure 6) montre une corrélation entre le signal des précurseurs et des fragments.

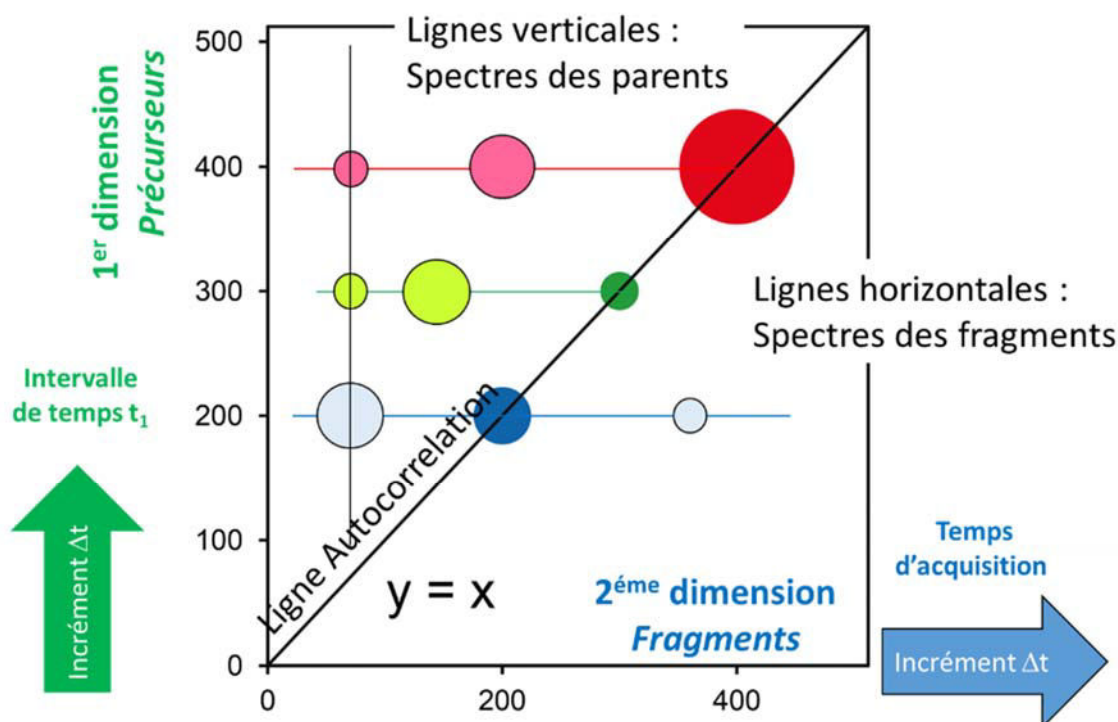


Figure 6 : Schéma représentant le spectre 2D-FT-ICR. Avec les lignes horizontales correspondant aux spectres de fragmentation, les lignes verticales correspondant aux spectres des précurseurs et la diagonale d'autocorrélation.

Cette séquence d'impulsions (figure 2) est une alternative à la FT-ICR MS/MS. En effet, dans les expériences classiques de MS/MS, les ions d'intérêt doivent d'abord être identifiés par l'utilisateur avant le réglage des paramètres d'isolement et de fragmentation. En FT-ICR 2D MS tous les ions de l'échantillon peuvent être fragmentés en même temps. Enfin, en raison des propriétés de la transformée de Fourier, tous les signaux qui sont acquis contribuent au rapport du signal sur bruit de toutes les espèces d'ions dans l'échantillon, les ions précurseurs et leurs fragments. Cette situation est différente pour la MS/MS classique où les transitoires de temps qui ont été accumulées pour chaque spectre servent seulement à améliorer le rapport signal sur bruit de l'ion parent sélectionné et de ses fragments [9].

Les fragmentations CID et les réactions ion-molécule ne sont pas les seules méthodes de fragmentation qui peuvent être utilisées en 2D-FT-ICR MS. Bensimon et al [4] ont montré qu'il était possible de réaliser une expérience de FT-ICR 2D en utilisant une fragmentation avec un laser infra-rouge sur les ions du CF₃I. Comme le faisceau laser est très focalisé au niveau du centre de la cellule ICR, seuls les ions qui ont été

déexcités vers le centre de la cellule peuvent être fragmentés. En conséquence, l'abondance des ions fragments dépend de t_1 et la fréquence cyclotron du précurseur tout comme dans l'expérience originale.

La plus grande expérience 2D FTICR a été réalisé en 1989. Elle comportait 512 points dans la première dimension et 8192 points dans la seconde dimension [4]. Cette expérience ne permettait pas d'arriver à une haute résolution. De plus, la capacité des ordinateurs au niveau de la taille de la mémoire et de la vitesse était limitée. C'est pour cela que les recherches en FT-ICR 2D MS ont été abandonnées.

3. Les variantes de la 2D

Une alternative à la FT-ICR 2D ont été développée par Ross en 1993 appelé méthode SWIM (Stored-Waveform ion radius modulation) [10]. Grâce à l'additivité des pulses une série d'excitation SWITF (Stored-Waveform Inverse Fourier Transform) peut être calculée, en utilisant le même encodage que les pulses P_1 et P_2 . La série d'excitation SWITF est aussi une sinusoïde dans le domaine des fréquences qui va créer une modulation sinusoïdale du rayon cyclotron des ions pour chaque fréquence des ions précurseurs. Les séries d'excitation d'ions permettent de moduler le rayon cyclotron des ions précurseurs de différentes masses à différentes fréquences.

En 2001, Van der Rest et Marshall [11] ont utilisés la 2D SWIM CID FT-ICR MS pour étudier les hétérodimères d'acides aminés avec un pont proton ou sodium. En 2002, Ross et al. [12] ont appliqués la méthode SWIM aux produits de synthèse combinatoire ayant un intérêt pharmaceutique et à un échantillon de revêtement pour voiture. Ce fut la première véritable étude analytique sur des échantillons complexe en utilisant la FT-ICR 2D MS.

D'autres méthodes permettant de corrélés les ions précurseurs et leur fragment existent :

- La méthode FT-ICR 2D de Hadamard ou HT/FT/ICR MS/MS [13,14], consiste à sélectionner la moitié des N précurseurs en utilisant des excitations SWITF et de les dissocier simultanément, puis le spectre des fragments combinés est mesuré. Une transformée d'Hadamard est appliquée pour obtenir un spectre de masse. Le processus est répété N fois, en sélectionnant une moitié différente des précurseurs à chaque fois.

Cette méthode permet d'améliorer d'un facteur $N/2$ la vitesse ou $\sqrt{\frac{N}{2}}$ pour le signal sur bruit, comparé à N analyses 1D FT-ICR MS/MS.

- PAcIFIC (Precursor Acquisition Independent From Ion Count), réalise des acquisitions de spectre MS/MS à chaque valeur de m/z sans se soucier de la présence d'ions précurseurs. L'échantillon est analysé par LC-MS/MS de façon répétée. A chaque injection un précurseur différent est sélectionné pour réaliser l'analyse PAcIFIC. Pour chaque fenêtre de masse m/z , tous les ions précurseurs sont isolés dans le piège à ions et soumis à une fragmentation par CID jusqu'à ce que la plage m/z souhaitée soit atteinte. Chaque spectre est ensuite recherché dans une base de données en utilisant le centre de la fenêtre comme masse de précurseur avec une tolérance correspondant à l'état de charge maximum autorisé multiplié par la largeur d'isolation divisée par deux. Cette méthode a permis d'identifier 756 protéines dans le plasma humain en réalisant 67 expériences de LC-MS/MS.

- La méthode SWATH (Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra) est une méthode d'acquisition indépendante des données, qui permet un enregistrement complet et permanent de tous les ions fragments des précurseurs peptidiques, isolés à l'aide d'une fenêtre de masse de quelques daltons (typiquement 25 Da), présents dans un échantillon complexe [15]. Les transitions des ions obtenus sont identifiées contre une bibliothèque spectrale, et des résultats d'identification et de quantification de peptides / protéines sont générés.

II. Récents développements en 2D-FT-ICR

Depuis 2000, des progrès considérables ont été réalisés en grâce à l'augmentation des performances des ordinateurs (mémoire, processeur, taille disque dur). L'électronique du système d'acquisition du FT-ICR MS a été entièrement numérisé, ce qui rend les impulsions d'excitation très stables et permet des changements faciles du protocole expérimental.

Grâce à ces avancées, la FT-ICR 2D MS peut être utilisée comme technique à part entière et à haute résolution. Les techniques de fragmentation sans gaz tels que l'IRMPD et l'ECD sont devenus disponibles sur les instruments FT-ICR commerciaux

[16]. Comme il n'y a pas de gaz injecté dans la cellule ICR, les ions ne subissent pas de collision avec les atomes et les paquets d'ions restent cohérents. L'utilisation de l'ECD ou IRMPD améliore donc à la fois la sensibilité et la résolution de la MS/MS et 2D FT-ICR MS.

1. Ré-introduction de la 2D FT-ICR

En 2010, l'expérience de Pfändler a été réalisée sur un FT-ICR Apex QE 9,4 T de Bruker (Brême, Allemagne) avec une source nano ESI positive et une fragmentation de type IRMPD [17]. Une expérience FT-ICR 2D a été réalisée en utilisant une gamme de masse de m/z 87,67 à 2000 dans le domaine ω_2 (l'axe horizontal) et un incrément de temps de $\Delta t_1 = 0,3 \mu s$ qui correspond à une fréquence maximale mesurée de 1667 MHz et une gamme de masse m/z 87,67 à 2000 dans le domaine ω_1 (vertical).

Lors de cette expérience le spectre de masse 2D enregistré a été constitué de 2048 spectres comprenant chacun 32 768 points, conduisant à une taille de fichier de 256 Mo (un point correspond à 4 octets de stockage). Le traitement des données utilisé a été réalisé avec le logiciel RMN Kernel, qui avait été développé pour la spectroscopie RMN [18].

Cette étude a montré que la séquence de pulse de Gäumann pouvait être réalisé sur un spectromètre de masse FT-ICR commercial et obtenir des spectres 2D informatif. Cette étude a également montré la présence des harmoniques pour chaque pic dans la dimension verticale car les rayons cyclotron des ions ne sont pas modulés sinusoïdalement, comme prédit par Guan et Jones [19].

Enfin, comme dans les expériences précédentes des bruits de scintillation ont été observés, ce qui a conduit à des raies verticales dans le spectre 2D aux fréquences des pics les plus intenses.

2. Optimisations de la méthode

Plusieurs problèmes doivent être résolus afin d'obtenir analytiquement des spectres de masse 2D utiles :

- Les harmoniques de pics précurseur d'ions (et, dans une moindre mesure, des pics d'ions de fragments) qui apparaissent dans la dimension verticale.

- La taille des données qui peuvent être traitées et la puissance de calcul nécessaire pour le traitement des données.

Dans les précédente FT-ICR 2D MS, la séquence de pulse utilise trois impulsions de durée et d'amplitude égales. Cependant, l'intensité du signal est faible, car il a été réparti sur une série d'harmoniques intenses. En optimisant l'amplitude et la durée des impulsions de codage afin de maintenir les rayons d'ions dans la zone de fragmentation, il est possible d'obtenir des spectres de masse 2D avec des harmoniques les plus faibles, ce qui améliore à la fois le rapport signal sur bruit et la lisibilité (figure 7).

Van Agthoven et al [20] ont montré l'importance de la forme et de la taille de la zone de fragmentation dans la cellule ICR pour les expériences de 2D-FT. En gardant le rayon des ions proche du faisceau laser de l'IRMPD, il est possible d'obtenir une modulation de l'efficacité de fragmentation qui ressemble à une fonction sinusoïdale de période t_1 et qui produit peut d'harmoniques.

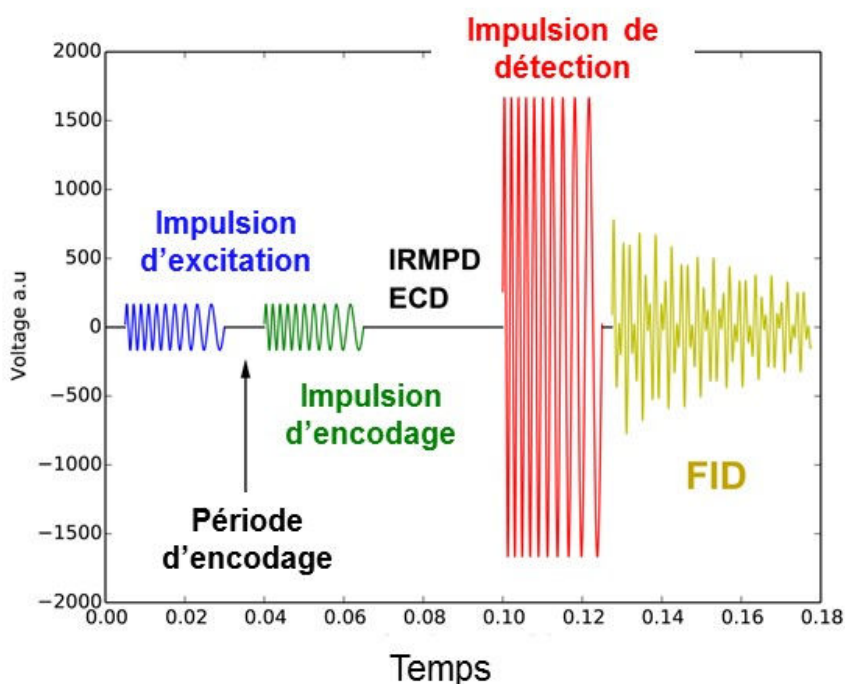


Figure 7 : Séquence FT-ICR 2D optimisée. Le premier et le second pulse sont identique. Le pulse de détection des ions possède un voltage plus important [20].

L'effet de l'augmentation de l'intensité du pulse d'observation est différente pour les ions fragmentés et les ions précurseurs. En effet les ions fragments voient leur intensité du signal et leur signal sur bruit augmenter lors de l'augmentation du pulse

d'observation. Le comportement des ions précurseurs, cependant, dépend de l'efficacité de la fragmentation au niveau du centre de la cellule ICR. L'impulsion d'observation qui est plus élevée que les impulsions de codage écrasent la modulation du rayon d'ions. Par conséquent, la ligne d'autocorrélation n'est pas équivalente au spectre de masse des ions précurseurs. Enfin, le rapport signal sur bruit des pics des ions précurseurs ne s'améliorent pas avec l'impulsion d'observation. Ceci est dû à la prévalence du bruit de scintillation, qui est proportionnelle à l'intensité du signal des pics précurseurs.

En théorie, la différence de phase entre l'ion et la seconde impulsion varie seulement avec la fréquence cyclotron ω_{ion} et l'intervalle de temps t_1 de sorte que si les ions sont en phase avec le second pulse P_2 , ils seront envoyés sur un rayon plus grand.

Si l'ion est en opposition de phase par rapport à P_2 , il doit revenir au centre de la cellule ICR. Cependant, cela ne se produit que lorsque les ions sont excités par des impulsions qui ont des fréquences RF qui sont exactement égalé avec la fréquence cyclotron de l'ion. Durant le 1^{er} pulse P_1 l'ion acquière une énergie cinétique, et se déplace en formant un spirale loin du centre de la cellule. Le mouvement de l'ion durant l'intervalle de temps t_1 suit une orbite qui est hors du centre de la cellule sauf si $T_{1\omega_{\text{ion}}}$ avec T_1 le temps de l'impulsion P_1 et la fréquence de l'ion ω_{ion} est un multiple de 2π .

Par conséquent, la seconde impulsion, même si elle est parfaitement déphasée avec l'ion, ne provoque pas exactement le retour au centre de la cellule. Il a été montré que le centre des trajectoires des ions présente un déplacement par rapport au centre de la cellule qui varie périodiquement avec les durées des deux premières impulsions. Si les impulsions sont de courte durée, le chevauchement du faisceau laser et des ions précurseurs ne sont pas compromis. Un autre avantage de l'utilisation des impulsions de courte durée est la diminution des intensités des signaux harmoniques qui apparaissent à plusieurs fréquences de base [7].

3. Traitement des données obtenue en 2D FT

Le traitement des analyses 2D-FT a été réalisé initialement en utilisant le programme Kermel utilisé pour la RMN codé sur 32 bits. Ce programme ne permettait pas un traitement des spectres 2D de grande dimension et un débruitage des spectres adéquat. Chiron et al. (SPIKE: Spectrometry Processing Innovative Kernel. Available at: <https://bitbucket.org/delsuc/spike>. Accessed May 2015) ont développé un programme

appelé SPIKE pour la RMN et la FT-ICR codé sur 64 octets. Ce logiciel a été développé pour l'analyse de données de grande dimension comme celles de 2D-FT-ICR et leur visualisation. Il a été écrit dans le langage Python pour bénéficier de la syntaxe claire, de la maintenance facile et des améliorations rapides. La vitesse de traitement est assurée par les modules scientifiques NumPy et Scipy. Le langage peut être utilisé pour l'écriture de scripts ou dans un contexte de programmation interactive tel que l'interface web Notebook IPython. SPIKE est également livré avec un mécanisme de plug-in qui est pratique pour l'ajout de nouvelles fonctionnalités à la bibliothèque native, sans la modification du code interne. Les données sont stockées au format h5 sous plusieurs niveaux pour permettre un affichage rapide comme dans Google Map. Spike propose des fonctionnalités diverses : analyse des ensembles de données de FT-ICR 1D (apodisation, phasing, module...), analyse des ensembles de données 2D (modulation de phase ou d'amplitude, algèbre complexe), un traitement robuste des données (aucune limite de taille de données, parallélisation des processus), caractéristiques de haut niveau (filtrage, prédiction linéaire, Cadzow, urQRd, sane, correction de ligne de base, 1D et 2D sélection de pics), l'architecture des Plugin et un affichage spectral complet en utilisant matplotlib (zoom, impression en format png ou pdf) (figure 8)

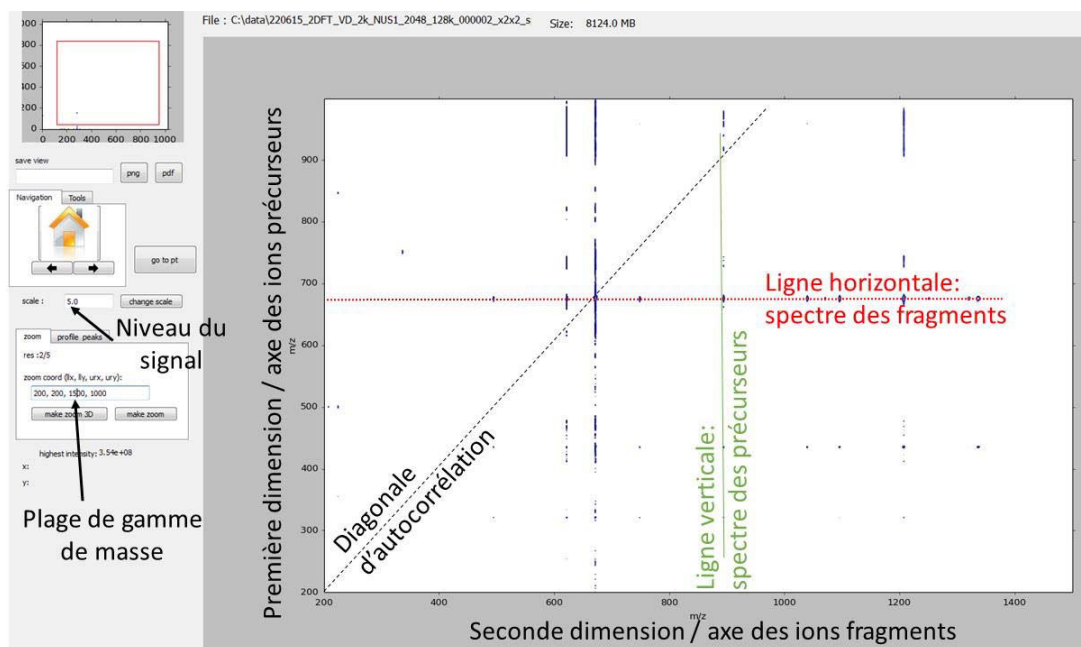


Figure 8 : Logiciel de visualisation des spectres FT-ICR 2D [21]. Au centre se trouve le spectre 2D-FT-ICR, sur la gauche la gamme de masse du spectre, le niveau du signal, la possibilité d'enregistrer le spectre en pdf ou png et le changement des axes en point ou m/z .

4. Le débruitage des données

Les groupes de Marshall et de Ross ont observés que le signal est composé de grande quantité de bruit de scintillation (comme en l'astronomie [22], ou au bruit t_1 en spectroscopie RMN [23]). Ce bruit conduit à des stries verticales dans les spectres 2D aux fréquences de ω_2 des pics les plus intenses. Le bruit de scintillation découle de fluctuations de l'amplitude ou de la fréquence d'un signal et ne peut pas être réduite en accumulant plusieurs scan, car il est proportionnel au signal. Une expérience FT-ICR 2D MS utilise un nouveau paquet d'ions pour chaque incrément de temps dans la première dimension. Donc, en 2D FTICR MS, la cause la plus probable du bruit de scintillation est la fluctuation du nombre d'ions créés dans la source d'ions, ce qui explique pourquoi la terminologie astronomique semble appropriée.

Les stries verticales causée par la fluctuation du nombre d'ions dans la cellule ICR peuvent masquer les signaux de faible intensité et provoquer des erreurs dans les données interprétation (figure 8, observation de stries verticale sur le spectre 2D). Le bruit de scintillation s'avère être un problème important dans les spectres FT-ICR 2D parce que ces pics parasites peuvent conduire à des erreurs dans la détermination de la fragmentation.

Pour supprimer le bruit de scintillation de la 2D, un algorithme développé et basé sur la valeur de décomposition singulière a été appliqué en 1987 par Cadzow et Wu [24]. Cet algorithme est basé sur le postulat selon lequel les signaux peuvent être décomposés en sommes de sinusoïdes exponentiellement amorties, ce qui est la base de toute l'analyse de Fourier, de sorte qu'après un certain nombre de points, le point suivant peut être prédit. Aucune hypothèse au sujet de la nature du bruit n'est nécessaire. Les données sont disposées dans une matrice de Toeplitz (une matrice dont les coefficients sur une diagonale descendant de gauche à droite sont les mêmes) qui est décomposée en vecteurs propres qui peuvent être disposés dans un ordre décroissant de leur valeurs propres en fonction d'une estimation du niveau de bruit, seuls les vecteurs propres avec les valeurs propres les plus élevés sont conservés et les données peuvent être reconstruites. Cet algorithme peut être transposé de la spectroscopie RMN, où il a souvent été appliqué avec succès [25]. Elle peut être

appliquée à chaque section verticale (parallèle à l'axe ω_1) des spectres FT-ICR 2D MS (figure 9) [26].

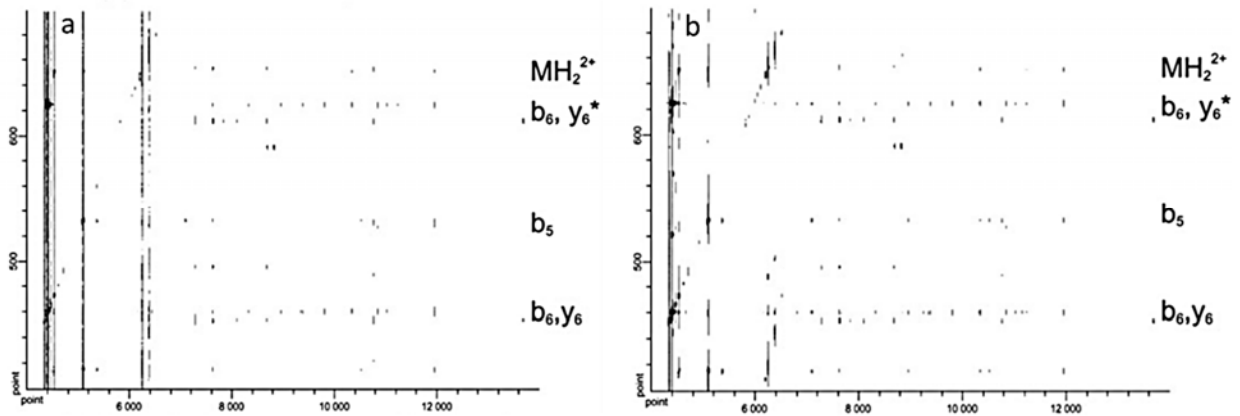


Figure 9 : Spectre de masse 2D FT de la bradykinine avec fragmentation IRMPD (a) sans application de l'algorithme Cadzow et (b) avec l'application de l'algorithme Cadzow en utilisant 30 lignes, 10 itérations et un ordre de 500. Les fréquences cyclotron sont représentées horizontalement et les fréquences de corrélation sont représentées verticalement. (de Van Agthoven et al. [26]).

L'algorithme Cadzow nous a permis de diminuer le bruit des spectres 2D de masse car elle ne présuppose pas de propriété particulière à la nature du bruit. Cependant, comme les causes exactes du bruit de scintillation ne sont pas encore connues, l'application de l'algorithme de Cadzow n'est pas entièrement satisfaisant car nous ne pouvons pas définir le nombre de vecteurs propres qui doivent être conservés et combien peuvent être mis au rebut. La recherche scientifique moderne produit des ensembles de données de taille et de complexité croissantes qui nécessitent des méthodes numériques dédiées au traitement. Dans de nombreux cas, l'analyse des données spectroscopiques implique le débruitage des données brutes avant tout traitement ultérieur. Les algorithmes de débruitage actuels exigent la décomposition en valeurs singulières d'une matrice avec une taille qui évolue comme le carré de la longueur des données, ce qui empêche leur utilisation sur de très grands ensembles de données. Profitant des progrès récents sur la projection aléatoire et algorithmes probabilistes, Chiron et al [27] ont développés une méthode simple et efficace pour le débruitage de très grands ensembles de données. Sur la base de la décomposition QR d'une matrice échantillonnés au hasard à partir des données, cette approche permet un gain de près de trois ordres de grandeur dans le temps de traitement par rapport aux décompositions classiques en valeurs singulières. Cette procédure, appelée urQRd

(uncoiled random QR denoising), réduit fortement l'encombrement de la mémoire de l'ordinateur et permet à l'algorithme de débruitage à appliquer à la taille des données quasi-illimitée (figure 10).

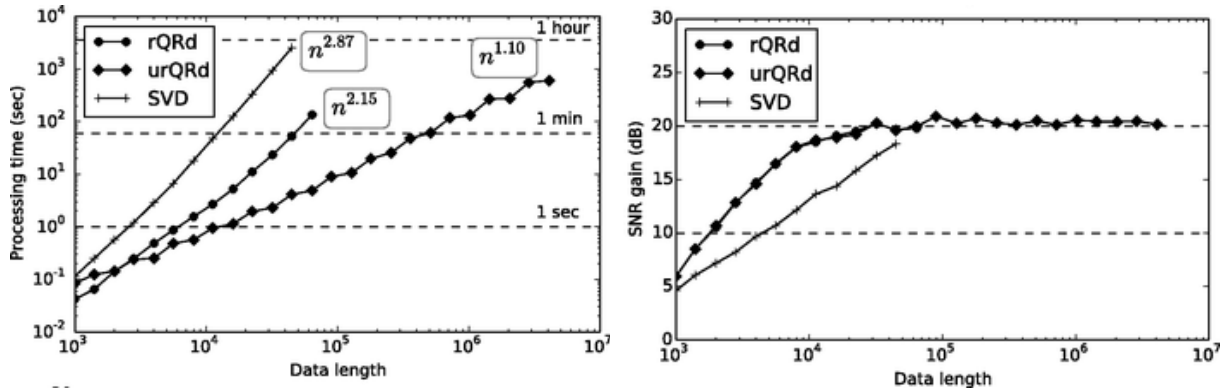


Figure 10 : Comparaison de l'algorithme de débruitage rQRd (rond), urQRd (losange) et SVD (croix). Sur le graphique de gauche, l'algorithme urQRd a un temps de processing plus rapide. L'efficacité de débruitage, exprimée comme le gain SNR généré par la méthode de valorisation; rQRd et urQRd sont presque indiscernables sur le graphique de droite (d'après Chiron et al [27])

Cette étude a montré que le débruitage urQRd est meilleur que les débruitage de type SVD et rQRd, qui sont limité par la mémoire de l'ordinateur, et s'arrête pour des données de 32 000 et 64 000 points. Cette algorithme a été appliqué sur une 2D-FT ICR 2k x 128 ponts de TAG extrait de plasma humain et a permis de diminuer le bruit de scintillation et d'observer les massifs isotopiques des triacylglycérols.

Toutes ces améliorations ont permis d'analyser des échantillons de plus en plus complexes et de développer le potentiel de la technologie de 2D-FT-ICR. En 2015, Van Agthoven et al [28] a mis en œuvre la photo-ionisation à pression atmosphérique avec la 2D-FT pour l'étude de la fragmentation du cholestérol. Il a été montré que le motif de fragmentation du radical et de l'espèce protonnée du cholestérol est différent. L'utilisation de la ligne des fragments, des précurseurs et des pertes de neutres dans le spectre de 2D permet de déterminer le mécanisme de fragmentation de molécules connues et d'obtenir des informations sur des espèces ioniques inconnues dans le spectre. Van Agthoven et al [29] ont réalisé l'analyse 2D-FT-ICR d'une digestion de cytochrome C en utilisant la fragmentation ECD et IRMPD. Cette même technique a aussi été utilisé pour analyser une digestion de collagène et pour la comparer à une analyse simple de 1D MS/MS [30].

III. L'avenir de la FT-ICR 2D

L'un des objectifs est d'augmenter la résolution dans la première dimension, mais en évitant de prolonger les temps d'expériences et la consommation d'échantillon. Heureusement, il existe de nombreuses techniques d'échantillonnage partielles qui peuvent être adaptés de la 2D RMN à la FT-ICR 2D MS afin de réduire la taille des données, sans perte de résolution. Ces techniques peuvent conduire à des expériences plus courtes et moins consommatrice en échantillon [31]. D'autres séquences d'impulsions alternatives peuvent être utilisé en particulier en adaptant des idées telles que l'expérience de l'accordéon [32] de la RMN bidimensionnelle pour avoir accès à la durée de vie des ions.

Le maintien de la cohérence des paquets d'ions est très important dans les expériences FT-ICR 2D. Si les paquets d'ions perdent leur cohérence au cours de la séquence d'impulsions, la différence de phase entre le mouvement des ions et l'impulsion radio fréquence est perdue, ainsi que la diminution de la superposition entre le faisceau laser ou d'électrons et les ions. Cela provoque une diminution de l'efficacité de fragmentation et affecte l'abondance de fragments et la qualité des données. La géométrie de la cellule ICR a un grand impact sur la cohérence des paquets d'ions et peut donc grandement affecter la qualité des spectres FT-ICR 2D MS. La cellule ICR compensée telles que celles développées par Brustkern et al. [33] et la cellule harmonisée développée par Nikolaev et al. [34] sont très prometteuse.

D'autres développements peuvent être envisagés comme la modification des séquences d'impulsion pour créer des analyses 3D comme en RMN [35]. Cette expérience pourrait permettre l'utilisation de deux types de fragmentations (IRMPD, ECD) et d'obtenir des informations structurelles approfondies comme la localisation des instauration sur les triacylglycérols.

La 2D-FT-ICR montre un grand potentiel en tant qu'outil d'analyse. Étant donné que l'isolation des ions n'est pas nécessaire. La perte de ceci au cours du transfert de la source d'ions à la cellule ICR est minimisée. L'accumulation des signaux transitoires pour toutes les espèces d'ions constitue un avantage important par rapport MS/MS ou il y a sélection des ions dans la cellule ICR ou le quadripôle.

En conséquence, FT-ICR 2D MS pourrait bien devenir une partie essentielle de la recherche en protéomique, lipidomique ou métabolomique.

Bibliographie

- (1) Marshall, A. G., Wang, T. C. L., & Ricca, T. L.: Ion cyclotron resonance excitatio/de-excitation: A basis for Stochastic fourier transform ion cyclotron mass spectrometry. *Chem. Phys. Lett.* **1984**, *105*, 233-236.
- (2) PeterPfändler, G., JacquesRapin, RaymondHouriet, TinoGäumann: Two-dimensional fourier transform ion cyclotron resonance mass spectrometry. *Chem. Phys. Lett.* **1987**, *138*, 195-200.
- (3) Pfaendler, P., Bodenhausen, G., Rapin, J., Walser, M. E., & Gaeumann, T.: Broad-band two-dimensional Fourier transform ion cyclotron resonance. *Journal of the American Chemical Society*, **1988**, *110*, 5625-5628.
- (4) Bensimon, M.; Zhao, G.; Gäumann, T.: A method to generate phase continuity in two-dimensional Fourier transform ion cyclotron resonance mass spectrometry. *Chem. Phys. Lett.* **1989**, *157*, 97.
- (5) Kumar, A., Ernst, R. R., & Wüthrich, K.: A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1-6.
- (6) Jones, S. G. a. P. R.: A theory for two-dimensional Fourier-transform ion cyclotron resonance mass spectrometry *J. Chem. Phys.* **1988**, *91*, 5291.
- (7) Sehgal, A. A., Pelupessy, P., Rolando, C., & Bodenhausen, G.: Theory for spiralling ions for 2D FT-ICR and comparison with precessing magnetization vectors in 2D NMR. *Phys. Chem. Chem. Phys.* **2016**, *18*, 9167-9175.
- (8) Buttrill, S. E.: Temperature dependence of the rates of ion-molecule collisions. *J. Chem. Phys.* **1973**, *58*, 656-659.
- (9) Aue, W.; Bartholdi, E.; Ernst, R. R.: Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **1976**, *64*, 2229.
- (10) chules W. Ross, S. G., Peter B. Grcwsbaas, Tom L. Ricca, Alan G. Marshall: Two-Dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry/Mass Spectrometry with Stored- Waveform Ion Radius Modulation. *Am. Chem. Soc* **1993**, *115*, 7854-7861
- (11) Guillaume van der Rest, A. G. M.: Noise analysis for 2D tandem Fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom.* **2001**, *210/211*, 101-11.
- (12) Ross, C. W., Simonsick, W. J., & Aaserud, D. J.: Application of stored waveform ion modulation 2D-FTICR MS/MS to the analysis of complex mixtures. *Anal. Chem.* **2002**, *74*, 4625-4633.
- (13) Fred W. McLafferty, D. B. S., Stanton Y. Loh, and Evan R. Williams: Hadamard transform and "no-peak" enhancement in measurement of tandem Fourier transform mass spectra. *Anal. Chem.* **1987**, *59*, 2212-2213.

- (14) Williams, E. R.; Loh, S. Y.; McLafferty, F. W.; Cody, R. B.: Hadamard transform measurement of tandem Fourier-transform mass spectra. *Anal. Chem.* **1990**, *62*, 698-703.
- (15) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R.: Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.
- (16) Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Håkansson, P.: Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1759-1768.
- (17) van Agthoven, M. A.; Delsuc, M.-A.; Rolando, C.: Two-dimensional FT-ICR/MS with IRMPD as fragmentation mode. *Int. J. Mass Spectrom.* **2011**, *306*, 196-203.
- (18) Tramesel, D.; Catherinot, V.; Delsuc, M. A.: Modeling of NMR processing, toward efficient unattended processing of NMR experiments. *J. Magn. Reson.* **2007**, *188*, 56-67.
- (19) Guan, S.; Jones, P. R.: A theory for two-dimensional Fourier-transform ion cyclotron resonance mass spectrometry. *J. Chem. Phys.* **1989**, *91*, 5291-5295.
- (20) van Agthoven, M. A.; Chiron, L.; Coutouly, M.-A.; Sehgal, A. A.; Pelupessy, P.; Delsuc, M.-A.; Rolando, C.: Optimization of the discrete pulse sequence for two-dimensional FT-ICR mass spectrometry using infrared multiphoton dissociation. *Int. J. Mass Spectrom.* **2014**, *370*, 114-124.
- (21) Lionel Chiron, M.-A. C., Jean-Philippe Starck, Christian; Rolando, M.-A. D.: SPIKE a Processing Software dedicated to Fourier Spectroscopies. *Computational Physics* **2016**.
- (22) Reiger, S. H.: Starlight scintillation and atmospheric turbulence. *The Astronomical Journal* **1963**.
- (23) Mehlkopf, A. F., Korbee, D., Tiggeleman, T. A., & Freeman, R.: Sources of t1 noise in two-dimensional NMR. *Journal of Magnetic Resonance* **1984**, *58*, 315-323.
- (24) Cadzow, J.; Wu, M. M.: Analysis of transient data in noise. *IEE Proc. F* **1987**, *134*, 69.
- (25) Brissac C, M. T., Delsuc MA.: Use of the Cadzow procedure in 2D NMR for the reduction of t(1) noise. *J Biomol NMR* **1995**, *6*, 361-5.
- (26) van Agthoven, M. A.; Coutouly, M. A.; Rolando, C.; Delsuc, M. A.: Two-dimensional Fourier transform ion cyclotron resonance mass spectrometry: reduction of scintillation noise using Cadzow data processing. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1609-16.
- (27) Chiron, L.; van Agthoven, M. A.; Kieffer, B.; Rolando, C.; Delsuc, M. A.: Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry. *Proc Natl Acad Sci U S A* **2014**, *111*, 1385-90.
- (28) van Agthoven MA, B. M., Chiron L, Coutouly MA, Kilgour D, Wootton CA, Wei J, Soulby A, Delsuc MA, Rolando C, O'Connor PB: Differentiating Fragmentation Pathways of Cholesterol 2D FT-ICR. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 2105-10.

- (29) van Agthoven, M. A.; Wootton, C. A.; Chiron, L.; Coutouly, M. A.; Soulby, A.; Wei, J.; Barrow, M. P.; Delsuc, M. A.; Rolando, C.; O'Connor, P. B.: Two-Dimensional Mass Spectrometry for Proteomics, a Comparative Study with Cytochrome c. *Anal. Chem.* **2016**, *88*, 4409-17.
- (30) Simon, H. J.; van Agthoven, M. A.; Lam, P. Y.; Floris, F.; Chiron, L.; Delsuc, M. A.; Rolando, C.; Barrow, M. P.; O'Connor, P. B.: Uncoiling collagen: a multidimensional mass spectrometry study. *Analyst* **2016**, *141*, 157-65.
- (31) Delsuc, M.-A.; Tramesel, D.: Application du traitement par entropie maximale aux données RMN multidimensionnelles ; cas de l'échantillonnage partiel. *Comptes Rendus Chimie* **2006**, *9*, 364-373.
- (32) Bodenhausen, G.: Direct determination of rate constants of slow dynamic processes by two-dimensional "accordion" spectroscopy in nuclear magnetic resonance. *J. Am. Chem. Soc.* **1982**, *104*, 1304-1309.
- (33) Brustkern, A. M.; Rempel, D. L.; Gross, M. L.: An electrically compensated trap designed to eighth order for FT-ICR mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1281-5.
- (34) Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G.: Initial experimental characterization of a new ultra-high resolution FTICR cell with dynamic harmonization. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1125-33.
- (35) Ikura, M.; Kay, L. E.; Bax, A.: A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **1990**, *29*, 4659-4667.

Partie 2 :

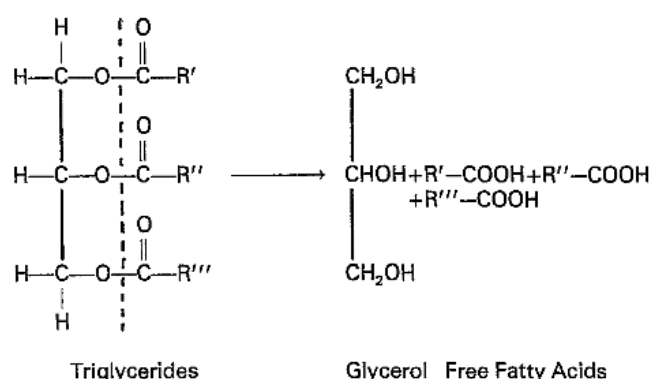
Analyse des triacylglycérols par FT-
ICR 2D

I. Les triacylglycérols

1. Définition

Les triglycérides ou triacylglycérols (abréviation TAGs) sont des composés lipidiques de notre organisme. Les triglycérides sont composés d'acides gras et sont stockés dans les tissus adipeux (la graisse). Ils sont fabriqués par notre organisme au niveau de l'intestin grêle et lors de la dégradation des sucres rapides par le foie.

Les triglycérides sont formés de l'estérification d'un glycérol par trois acides gras. Les triglycérides sont donc des esters d'acides gras et du glycérol.



où R'_1 , R''_2 et R'''_3 sont des chaînes carbonées d'acides gras. Les trois acides gras ne sont pas nécessairement identiques.

Dans les triglycérides, les acides gras peuvent comporter de 4 à 22 atomes de carbone, mais ceux comportant de 16 à 18 atomes sont les plus courantes. Les triglycérides peuvent être saturés, monoinsaturés ou polyinsaturés. Ces termes désignent la saturation en hydrogène des acides gras.

2. Métabolisme

Le métabolisme des acides gras comprend deux voies métaboliques principales :

- La lipogenèse, ou synthèse *de novo*, consiste à synthétiser un acide gras par condensation de molécules d'acide acétique. Les unités acétyle sont activées sous la forme d'acétyl-coenzyme A.
- La β -oxydation, consiste à dégrader des acides gras en unités acétyle afin de libérer de l'énergie métabolique et du pouvoir réducteur.

3. Maladies métabolique

Un taux anormalement élevé de triglycérides, appelé hypertriglycéridémie (au-delà de 2 g/L), augmente le risque d'athérome et d'accident cardiovasculaire. Ce risque cardiovasculaire est augmenté en cas d'excès de cholestérol, mais également d'une inflammation du pancréas.

Enfin, l'hypertriglycéridémie est l'un des 5 facteurs de risque du syndrome métabolique, aux côtés : du diabète, de l'obésité abdominale, du cholestérol, de l'hypertension [1-4]. Ce syndrome double le risque de décéder d'une complication cardiovasculaire.

II. Les méthodes d'analyses classiques

Les triglycérides sont dosés par des méthodes indirectes du glycérol contenu dans le plasma. Il s'agit d'un dosage colorimétrique où l'intensité de la couleur formée est proportionnelle à la concentration de triglycérides présents dans l'échantillon testé [5]. La limite de détection de 10 mg/dL jusqu'à la limite de linéarité de 2200 mg/dL.

L'application de la spectrométrie de masse pour la détection, l'identification et la quantification des TAGs est en plein essor. Les techniques chromatographiques sont utilisées pour séparer les différentes espèces moléculaires de triacylglycérols soit selon leur taille et / ou selon le degré d'insaturation. Le principal avantage de la spectrométrie de masse est de générer des ions utiles permettant l'élucidation de la structure, c'est-à-dire des ions fournissant des informations sur le poids moléculaire, les résidus d'acyle gras et leurs associations moléculaires, et la distribution de résidus d'acyle gras entre les positions primaires (*sn*-1/3) et le secondaire (*sn*-2) du glycérol.

Avant la détection et l'identification des TAGs, la première étape à réaliser est l'extraction des lipides.

III. Préparation des lipides

1. Méthodes d'extraction des lipides

Indépendamment de l'analyse ultérieure, la première étape de l'analyse des lipides est l'extraction des lipides. Cela semble plutôt évident, mais c'est une étape

majeure car de nombreux lipides ne peuvent pas être extraits si aisément puisqu'ils sont associés dans le matériau biologique d'origine avec d'autres composés. Par exemple, les fragments aliphatiques hydrophobes des lipides interagissent avec les régions non polaires des protéines, en particulier avec les acides aminés valine, leucine ou isoleucine. En outre, les groupes phosphate acide des lipides interagissent fortement avec des ions métalliques qui sont normalement liés aux protéines. Ces deux exemples fournissent déjà la preuve que l'extraction des lipides est une tâche sophistiquée mais largement négligée [6].

Les lipides neutres ou généralement les lipides de stockage sont extraits avec des solvants relativement non polaires tels que l'éther diéthylique ou le chloroforme, mais les lipides associés à la membrane sont plus polaires et nécessitent des solvants polaires tels que l'éthanol ou le méthanol pour perturber les liaisons hydrogène ou les forces électrostatiques.

Il existe de nombreuses méthodes d'extraction des lipides qui sont adaptées aux types d'échantillons, et aux lipides désirés [7]. Une liste de méthodes les plus utilisées pour extraire les lipides est présentée dans le tableau 1.

Solvants	Utilisation	Remarques et références
CHCl ₃ /CH ₃ OH (2:1, v/v), méthode de Folch	Lipides provenant de plantes, animaux et bactéries	L'eau des tissus est la composante ternaire et sa quantité est très importante pour éviter la perte de lipides [8]
CHCl ₃ /CH ₃ OH (1:1, v/v), méthode de Bligh and Dyer	Utile pour les systèmes riches en eau, en particulier fluides corporels	La perte partielle de lipides apolaires tels que les TAGs peut se produire [9]
Butanol saturé avec de l'eau	Les lipides végétaux, c'est-à-dire les lipides piégés dans l'amidon et les lipides plutôt polaires	très bonne extraction des phospholipides [10]
Hexane/2-propanol (3:2, v/v)	Faible teneur en lipides dans l'extrait parce que le mélange de solvants utilisé est hautement apolaire	Par rapport au CHCl ₃ , l'hexane et l'isopropanol sont des solvants de faible toxicité [11]

Chloroforme/isopropanol (7:11, v/v).	Particulièrement adapté aux érythrocytes avec une teneur élevée en lipides	Indiqué pour fournir des rendements lipidiques élevés [12]
Chloroforme/méthanol/12N HCl (2:4:0.1, v/v/v)	Les phospholipides acides tels que la phosphatidylsérine PS et en particulier les phosphoinositides	L'addition de HCl conduit à la détection neutralisation de la charge et à l'amélioration des rendements d'extraction des lipides acides [13] Cette méthode, cependant, conduit à une hydrolyse complète des plasmalogènes [14].
MTBE/méthanol/water (10:3:2.5, v/v/v)	Récupération des espèces lipidiques de presque toutes les classes principales	Procédure d'extraction rapide et permet d'extraire la majorité des classes de lipides [15].

Tableau 1 : Méthodes d'extraction des lipides les plus fréquemment utilisées.

2. Méthodes de séparation et fractionnement

Il existe de nombreuses méthodes permettant de séparer les grandes classes de lipides. Les procédés les plus largement utilisés pour isoler les triglycérides à partir de mélanges lipidiques sont la chromatographie sur colonne Florisil [16-18] ou la chromatographie sur couche mince (TLC) sur l'acide silicique [19]. Ces procédures sont relativement rapides, nécessitent des équipements simples, produisent une bonne résolution des triglycérides provenant d'autres classes de lipides et (lorsqu'elles sont correctement utilisées) ne produisent aucune altération chimique de l'échantillon. La chromatographie en colonne gère des échantillons de plus grande taille, mais la TLC donne une meilleure résolution des classes de lipides. Dans le tableau 2 ci-dessous, sont référencées les différentes méthodes de séparation des triacylglycerols. Les méthodes de séparation permettent de séparer les TAGs en fonction du nombre total de carbone, mais aussi en fonction de leurs oxydations.

Méthodes	Echantillons	phase	Références
SPE	Plasma humain, farine de blé, bactéries	Aminopropyl	[20-22]
SPE	Échantillons archéologique,	Diol	[23]
GC	Beurre de cacao, gras, huile coco	Silice	[24,25]
LC	Huiles, gras, plasma	Phase normale	[26-28]
LC	Lipides de lait, graines, cire	Phase inverse C18	[29-31]
LC	Huile végétale,	Argent	[32,33]
TLC	Huile végétale	Phase normale	[34,35]
TLC	Huile végétale	Phase inverse C18	[36]
TLC	Huile végétale	Argent	[33,37,38]
TLC	Huile végétale	Cuivre	[39]
SFC	Lait, huile, gras		[40,41]
TLC	Lait, huile, gras	2D	[42]
SFC	Huile végétale et poissons	Argent	[43,44]

Tableau 2 : Méthodes utilisées pour séparer et fractionner les lipides. LC : chromatographie liquide, SPE : Extraction en phase solide, TLC : Chromatographie sur couche mince, SFC : Chromatographie en phase supercritique.

IV. Analyse par spectrométrie de masse

La spectrométrie de masse est très utilisée dans l'analyse des lipides grâce au développement des technologies d'ionisation, de fragmentation, de détection. L'analyse des lipides utilise de nombreuses techniques d'ionisation pour permettre de détecter une large gamme de lipides dont les TAGs [45]. L'analyse des lipides à l'aide de la spectrométrie de masse est de plus en plus présente et se développe rapidement [46-49].

1. Chromatographie en phase gazeuse (GC)

La GC/MS avec les sources d'ionisation électronique ou chimique a été utilisée pour évaluer la complexité des espèces moléculaires des TAGs naturel dès les années 1970 [50-52]. Les TAG étaient estérifiés et dérivés pour obtenir des composés volatiles et analyser les acides gras [53]. Ou les TAGs sans les modifier, en fonction de leur nombre de carbones sur colonne à haute température [54].

2. Analyse par bombardement atomique rapide

Le bombardement atomique rapide (FAB, Fast Atom Bombardement) a été largement utilisé pour identifier la structure des lipides non volatiles, y compris les acides gras [55], les monoacylglycérols [56], les glycérophospholipides [57] et les sphingolipides [58,59]. Cependant, compte tenu de la complexité des extraits de lipides et des difficultés du couplage de la chromatographie et de la FAB MS, des efforts importants devaient être effectués pour analyser efficacement les lipides par FAB MS.

3. La MALDI

La MALDI (Matrix Assisted Laser Desorption Ionization) est utilisé pour l'analyse des TAGs car c'est une méthode simples, sensible et avec des possibilités d'automatisation [60,61]. De nombreuses matrices peuvent être utilisée pour leur analyse comme la DHB (acide 2,5-Dihydroxybenzoïque), HCCA (acide α -Cyano-4-hydroxycinnamique), [60], THAP (2',4',6'-Trihydroxyacetophenone) [62], 9-aminoacridine [63] et dithranol [60]. Lors des analyses MALDI, les TAG sont principalement détectés avec un adduit sodium [64] mais il existe d'autres cations qui peuvent être utilisé comme le lithium [65], ou le césium [66]. La présence quasi exclusive d'adduits comme le sodium simplifie considérablement l'analyse des mélanges et c'est probablement la raison pour laquelle, par exemple, les échantillons d'huiles végétales sont souvent caractérisés par MALDI-TOF MS [60].

Il a été montré que l'analyse MS/MS par MALDI des TAGs peut permettre d'obtenir des informations structurales [67]. De nouveaux développements MALDI ont permis d'analyser les TAGs à partir de séparation sur plaque TLC [68] mais aussi directement sur des tissus avec les avancés en imagerie MALDI [69].

D'autre part, les mesures de MALDI sont généralement affectées par des problèmes intrinsèques, c'est-à-dire des complications liées à l'interférence des ions de la matrice dans la région de faible masse et à l'absence de reproductibilité lors des analyses en raison du manque d'homogénéité dans les cristaux matriciels. Un autre point négatif est la fragmentation dans la source des TAGs lors de leurs analyses conduisant entre autre à la présence de diacylglycérols artéfactuels.

4. Les techniques : ESI, APCI, APPI, DESI, DART

L'électronébulisation (ESI, ElectroSpray Ionization) est une technique d'ionisation douce qui est mieux adaptée aux molécules polaires, sans avoir besoin de dérivatisation chimique ou d'extraction à partir de solutions polaires. La lipidomique "shotgun" a d'abord été proposée par Han et Gross en 2003 [70], dans laquelle la technique ESI-MS était utilisée pour l'analyse directe des lipides sans séparation par chromatographie liquide [71,72]. En ajustant la valeur du pH, à un pH neutre pour l'analyse en mode négatif [73], par ajout de certains agents d'ionisation spécifiques en solution pour l'analyse en mode positif comme le chlorure de lithium [74] ou en utilisant les ions sodium [75], ammoniacque [76], ou chlorure [77]. Cependant, le phénomène de la conversion mutuelle et de la suppression des ions parmi les différents lipides peuvent entraîner une erreur systématique lors de la détection directe d'extraits lipidiques complexes. Par exemple, lors de l'ionisation de solutions contenant des lysophosphatidylcholine, ceux-ci peuvent perdre le groupe de la choline et devenir un acide lysophosphatidique artificiel [78].

Par rapport à l'ionisation ESI qui utilise uniquement des champs électriques pour générer des gouttelettes chargées et des ions subséquents par évaporation ionique, L'ionisation chimique à pression atmosphérique ou APCI (Atmospheric Pressure Chemical Ionization) et La photo-ionisation à pression atmosphérique ou APPI peuvent fournir des mécanismes supplémentaires pour ioniser les analytes. En outre, ces méthodes d'ionisations sont moins sensibles à l'effet de la suppression de l'ionisation et des effets de tampon de sel que ESI [79-81].

D'autres techniques d'ionisation peuvent être utilisées mais sont moins courantes. C'est le cas de la désorption-ionisation par électronébulisation (Desorption electrospray

ionization, DESI) [82,83] ou l'ionisation à pression atmosphérique (Direct Analysis in Real Time, DART) [84]

5. Le Couplage

Les triacylglycérols peuvent être analysés avec différents types d'ionisation, par chromatographie gazeuse couplé à la spectrométrie de masse ou en analyse directe. Il existe de nombreux développements d'analyse des TAGs par couplage liquide (tableau 4).

Type	Classe lipides	phase colonne	Eluant	Références
HILIC	Lipides polaires	Si, NH ₂ , Diol	Acétonitrile, méthanol, 2-propanol, hexane et $\geq 2.5\%$ eau	[85]
NP-LC	Lipides non polaires	Si, NH ₂	Hexane, 2-propanol, chloroforme, heptane	[86]
SFC	Tous les lipides	C18, Si	CO ₂ super critique et méthanol	[87]
RP-LC	Tous les lipides	C18, C8	Acétonitrile, méthanol, 2-propanol, tampon volatile	[88,89]
NARP-LC	Lipides non polaires	C18	2-propanol, acétonitrile, acétone, dichlorméthane	[90]
Ag-LC	Lipides non polaires	IEX avec Argent	Hexane – acétonitrile (+2-propanol) ou dichlorméthane – acétonitrile	[91]
Chiral	Lipides non polaires	Normal	Hexane et 2-propanol	[92]

Tableau 4 : Méthodes de séparation des lipides avec un couplage avec un spectromètre de masse.

La nature de la séparation est assez connue, c'est-à-dire que le temps de rétention dépend du nombre total d'atomes de carbone dans les trois chaînes d'acides gras, chaque double liaison réduisant la longueur de chaîne efficace de l'espèce par l'équivalent d'environ deux carbones des atomes.

V. L'analyse des TAGs par fragmentation

La spectrométrie de masse permet l'élucidation de la structure des TAGs grâce à la fragmentation de ces molécules.

Marzilli et al démontré que l'analyse par spectrométrie de masse en piège ionique, combinée à l'ionisation par électronébulisation fournit une caractérisation simple et détaillée de la structure des TAGs. Ils ont montré que les spectres de masse obtenus par électronébulisation sont caractérisés par des pics molécules moléculaires intenses

générés par adduits et que la disposition du groupe acyle sur le squelette des TAGs peut être déterminée sur la base de différences quantitatives relativement bien définies d'ions fragments produits à partir de données de MS/MS [93]. Hsu et al ont caractérisés la structure des TAGs en utilisant la fragmentation CAD avec l'utilisation d'adduit lithium [74]. Pitternauer et al ont analysés les TAGs avec différents cations en utilisant un spectromètre de masse de type MALDI-TOF/R-TOF (R pour mode réflectron) et une fragmentation de type CID. Ils ont réussis à déterminer les positions des acides gras sur les TAGs et monter que l'adduit sodium été le mieux adapté pour la fragmentation en MALDI [64]. Les TAGs peuvent se fragmenter aussi en utilisant la fragmentation IRMPD implémentée sur un spectromètre FT-ICR [23]. Garnier et al ont montré que la fragmentation IRMPD permet d'obtenir des spectres de fragmentation plus simples et que des échantillons de type archéologiques peuvent être analysés. Plus récemment, la photo dissociation par ultra-violet (UVPD) implémentée sur un spectromètre de masse de type Orbitrap permet d'analyser les TAGs [94]

VI. Logiciels informatiques

Il existe quelques programmes et / ou progiciels qui exécutent plusieurs étapes de traitement de données telles que le filtrage spectral, la détection de pointe, l'alignement, la normalisation, l'analyse et la visualisation des données exploratoires pour les méthodes LC-MS. Ces programmes ont été développés et utilisent la plupart du temps les temps d'élution et les masses des d'ions observés. Les logiciels disponibles sur le marché et leur utilisation sont répertoriés (tableau 5).

Analyse des triacylglycérols par FT-ICR 2D

Nom	Type	Utilisation	Référence
XCMS	open-source	Traitement de données métabolique basé en utilisant un "cloud", Plate-forme qui fournit une analyse métabolique de haute qualité	[95]
MZmine 2	open-source	Outils utilisant les données de LC-MS. Présence de module de visualisation, identification pics.	[96]
LipidFinder	open-source	LipidFinder est un logiciel open source, développé initialement pour être utilisé avec SIEVE pour traiter les données MS à partir d'expériences lipidomiques	[97]
LipidView	Absciex	Le logiciel est un outil de traitement de données pour la caractérisation et la quantification des lipides espèces issues de l'analyse ESI-MS et LC	[98]
LipidSearch	Thermo Fisher Scientific	Un nouvel outil puissant pour l'identification automatique et la quantification relative des espèces de lipides cellulaires à partir d'une grande quantité de données de spectrométrie de masse obtenues à la fois par la LC-MS et les approches lipidomiques « shot gun »	[99]
SimLipid	PREMIER Biosoft	Logiciel qui permet l'identification et la quantification des lipides à haut débit. Il analyse les données lipidiques MS, MS/MS et MSn.	[100]
MultiQuant	Applied Biosystems	La quantification des espèces lipidiques à l'aide de tests MRM et permet la construction de scénarios courants standard hautement personnalisables	[101]
LIMS A	open-source	Cherche les pics dans les spectres de masse et recherche les masses dans une base de données.	[102]
LipidProfiler	open-source	Logiciel permettant l'identification et la quantification des lipides en utilisant la sélection des ions fragments.	[103]
AMDMS-SL	open-source	Identification et quantification des espèces lipidiques individuelles à partir des données obtenu à partir d'analyse « shotgun lipidomics »	[104]
LipidXplorer	open-source	Logiciel open source qui permet la caractérisation quantitative des lipides complexes en interprétant de grands ensembles de données de spectres de masse	[105]
ALEX	open-source	Traite les données MS et MSMS obtenus à partir de spectromètres de masse haute résolution	[106]
LIQUID	open-source	Analyse haut débit des données de lipidomique et contient une base de donnée personnalisable	[107]
LipidMiner	open-source	Interface graphique pour la détection et quantification des lipides obtenues par analyse LC-MS/MS	[108]
MS-LAMP	open-source	Logiciel d'aide d'interprétation des données de ESI et MALDI pour la lipidomique	[109]
LipidQA	open-source	Plate-forme logicielle pour identifier et quantifier les espèces moléculaires lipidiques complexes dans les mélanges biologiques	[110]
LipidInspector	open-source	Identification des lipides à partir de spectres MS/MS et en utilisant une base de donnée.	[111]

Tableau 5 : Liste des logiciels libre d'accès ou de commerciaux pour l'analyse, la quantification et l'identification des lipides.

Il existe quelques bases de données utilisées comme la plus connue qui est LipidsMaps [112] ou très récemment [113] LipidHome qui fournissent des molécules lipidiques théoriquement générées à partir des métadonnées utiles [114]. Tous ces outils permettent d'identifier et mettre en lien les lipides identifiés avec les aspects de la biologie.

VII. Objectifs

L'analyse des TAGs par spectrométrie de masse est très développée et de nombreuses méthodes d'analyses existent, avec ou sans séparation comme montré précédemment. Les TAGs sont des composants majeurs du corps humain et interviennent dans de nombreux processus biologiques et dans le développement de maladies métaboliques. Dans cette partie, une méthode d'analyse « shot gun » est développée pour identifier les lipides présents dans le plasma humain. La 2D-FT-ICR est appliquée sur ce type d'échantillons, comme vu dans le chapitre précédent cette méthode permet d'avoir les informations des MS et MS/MS en une fois sans réaliser de séparation par couplage. À partir de un millilitre de plasma humain, les TAGs présents dans le plasma sont profilés et identifiés. La 2D-FT-ICR MS permettra d'obtenir la masse de chacun des lipides et leur fragmentation IRMPD pour déterminer leur composition en acide gras.

VIII. Matériels et méthodes

1. Extraction des lipides du plasma

Cent microlitres de DMSO et 3 mL d'heptane ont été additionnés à 1 mL de plasma humain (concentration protéique estimée à 60 mg/mL). La solution a été agitée pendant 30 min à température ambiante pour séparer les lipides des protéines. La phase supérieure a été récupérée puis stockée.

Trois millilitres d'une solution chloroforme / méthanol (2/1) a été ajoutée au sous-nageant précédent. La solution a été agitée une heure à température ambiante puis la phase inférieure a été récupérée.

Les phases ont été réunies et le solvant a été évaporé à 37°C. L'extrait évaporé à sec, a été repris dans 2 mL de chloroforme pour la purification avec deux colonnes aminopropyl (figure 1).

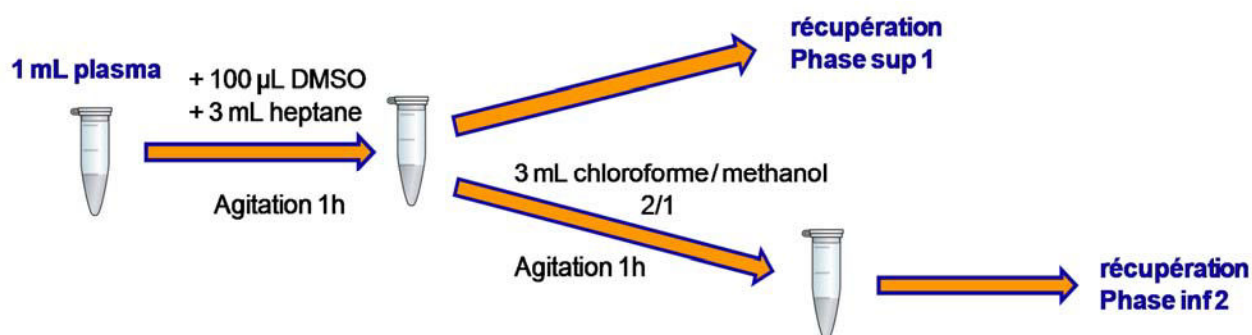


Figure 1 : Schéma des différentes étapes de l'extraction des lipides.

2. Purification de familles lipidiques

Deux millilitres d'hexane ont été déposés sur une première colonne aminopropyl (Supelco) avec 50 mg de phase afin de l'équilibrer. Un millilitre d'extrait total de lipides ainsi que 2 mL de chloroforme ont été déposés sur la colonne et sont récupérés en tant que fraction non retenue. Ensuite commence les éluions sur la première colonne (figure 2).

Deux millilitres d'acétone ont été déposés sur la colonne afin d'éluer les acides gras libres puis 2 mL de méthanol pour éluer les phospholipides.

Le solvant de la fraction non retenue (chloroforme) a ensuite évaporé. L'extrait évaporé à sec a été repris dans 2 mL d'hexane afin de commencer les éluions sur la deuxième colonne aminopropyl.

Le mélange a été déposé sur la colonne afin d'éluer les stérols ester. Puis 2 mL de hexane / éthyl acétate / CH₂CL₂ (85/13/2) ont été déposés pour éluer les triglycérides. Ensuite 2 mL d' hexane / éthyl acétate (5/95) été déposés pour éluer les Stérols, 2 mL hexane / éthyl acétate (15/85) été déposés pour éluer les diglycérides et enfin 2 mL chloroforme / isopropanol (2/1) été déposés pour éluer les monoglycérides. La fraction contenant les TAG a été évaporée à sec.

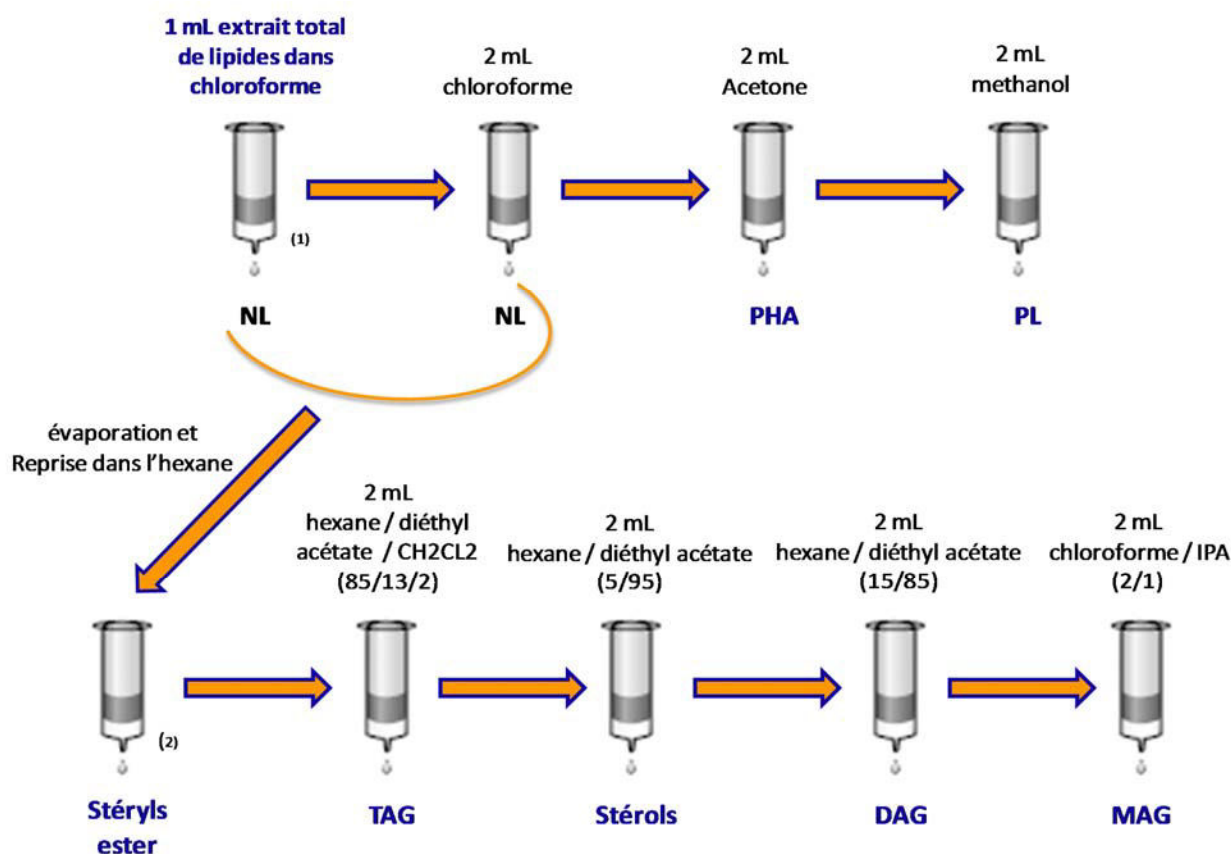


Figure 2 : Schéma des différentes étapes de la purification de grandes familles de lipides avec 2 colonnes aminopropyl.

3. Analyse par spectrométrie de masse

La fraction contenant les TAG a été re-suspendue dans 100 µL de chloroforme puis 10 µL de cette solution a été diluée dans 489 µL de méthanol / chloroforme, 5 :1,

v/v et 1 μL de chlorure de lithium isotopiquement pur (^6Li), à 10 mg/ml dans le méthanol a été ajouté. Toutes les expériences ont été effectuées sur un spectromètre de masse 9,4 T nanoESI-hQh-ApexQE FT-ICR (Bruker Daltonics, Bremen, Allemagne) en mode ionisation positive. L'échantillon a été injecté à un débit de 10 $\mu\text{L}/\text{heure}$ à l'aide d'un pousse seringue (Cole Parmer®, US) et une seringue de 250 μL (Gastight, Hamilton). L'acquisition a été réalisée avec 1M points et une gamme de masse de m/z 144,38 – 2500.

La séquence d'impulsion pour la 2D FT-ICR est représentée dans la figure 3. Toutes les impulsions ont été construites à l'aide du générateur d'impulsions ApexControl (Bruker Daltonics, Brême, Allemagne).

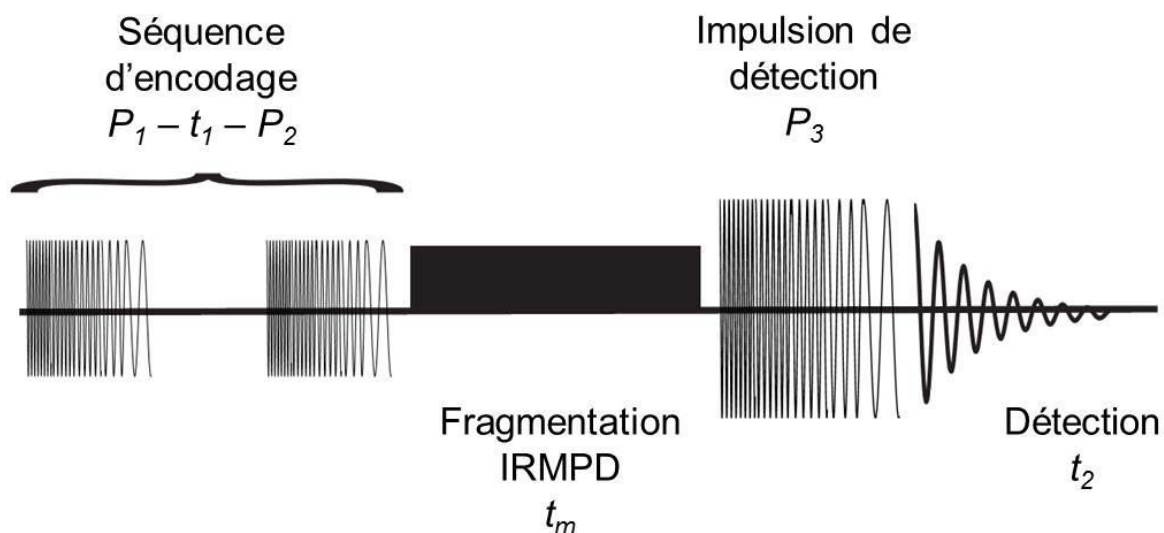


Figure 3: Séquence 2D FT-ICR utilisée pour réaliser les analyses sur triacylglycérols humains

Les impulsions de codage (P_1 , P_2) ont une amplitude de 70 Vpp et une durée de 0,5 μs par fréquence. L'augmentation en t_1 est de 0,5 μs . L'impulsion d'observation (P_3) a une amplitude de 100 Vpp et une durée de 20 μs par fréquence. La gamme de masse horizontale est m/z 144,39-3000 (1 MHz à 48 kHz). La fragmentation a été réalisée avec l'IRMPD avec un laser CO₂. La puissance du laser a été réglée à 12,5 W et la longueur d'onde de 10,4 μm . Les ions ont été irradiés pendant 0,4 s. Seize mille trois cent quatre-vingt-quatre points ont été enregistrés avec une résolution d'acquisition de 512k en réalisant quatre acquisitions de 16k et en compilant les données.

4. Traitement de l'information

Les spectres MS ont été traités et visualisés à l'aide du logiciel Data Analysis (Bruker Daltonics, Bremen, Allemagne). Le traitement informatique du spectre 2D de 64 k points a été réalisé à partir d'un programme écrit sous Python (SPIKE adapté en langage python 64 bits Enthought Python Distribution, Wavre, Belgique). Le spectre 2D de masse a été débruité à l'aide de l'algorithme de débruitage urQRd.

IX. Résultats

1. Extraction des TAGs

L'extraction des lipides comme décrit dans la publication de Folch's [8] fonctionne parfaitement. Sur le spectre présenté dans la figure 4, les différentes familles de lipides du plasma humain comme les TAG (triacylglycérols), PL (phosphoglycérides), DG (diacylglycérols), MG (monoacylglycérols) sont détectées. La méthode de séparation des TAG sur SPE aminopropyl est basé sur la publication de Gutiérrez [115]. Sur les spectres présentés dans les figures 5 à 7 on peut voir que les différentes fractions contiennent les lipides des différentes familles.

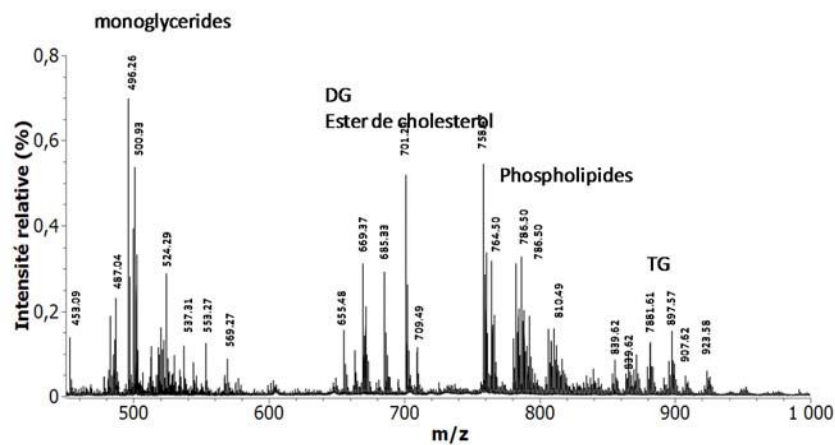


Figure 4 : Spectre nanoESI FT-ICR MS de l'extraction des lipides de plasma humain.

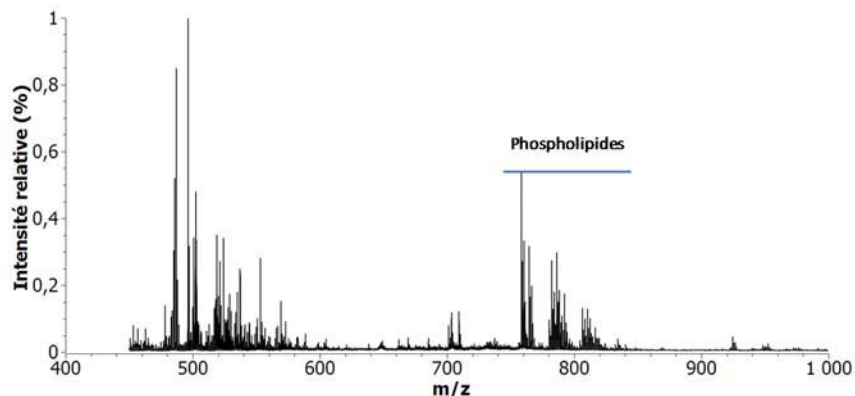


Figure 5 : Spectre nanoESI FT-ICR MS de la fraction PL.

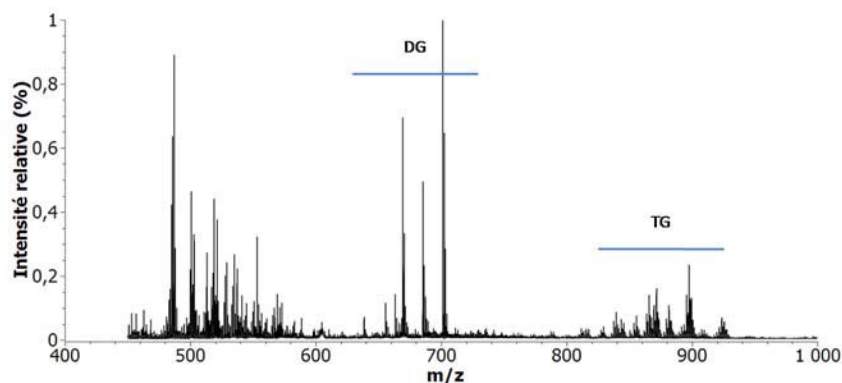


Figure 6 : Spectre nanoESI FT-ICR MS de la fraction NL (lipides neutres), contenant les DG, TG.

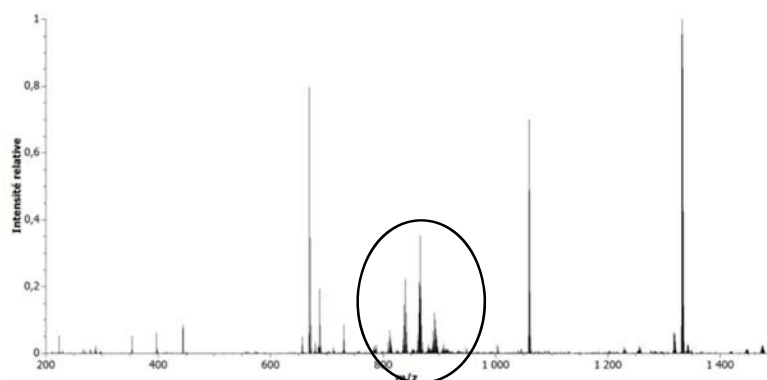


Figure 7 : Spectre nanoESI FT-ICR MS de la fraction purifiée de TAG du plasma humain (entourés d'un cercle noir).

La méthode de séparation des différentes familles de lipides fonctionne et elle permet d'obtenir une fraction de TAG propre pour l'analyse en nanoESI MS/MS. Les pics d'intensité élevés dans la figure 7 correspondent à des contaminants présents dans la solution de chloroforme.

2. Analyse FT-ICR MS et MS/MS

Les TAGs sont le produit de la triple estérification du glycérol avec des acides gras. Leur nomenclature est C:D, dans laquelle C représente le nombre de carbones dans les acides gras et D le nombre de doubles liaisons (ou insaturations) dans les acides gras. L'analyse des TAGs par nanoESI FT-ICR avec le chlorure de lithium utilisé comme adduit permet une amélioration du signal et l'obtention de spectres de fragmentation plus simple [23,74,93]. Hsu et Turk ont montré que les TAGs peuvent être

ionisés en ESI avec un bon rendement lorsqu'ils sont lithiés. Cette méthode a été utilisée avec succès dans de nombreuses applications afin d'analyser les lipides par spectrométrie de masse ESI [65,74,116]. Pour simplifier d'avantage l'analyse, du chlorure de lithium 6 isotopiquement pur, a été utilisée pour limiter les contributions isotopiques. La figure 8 montre le triacylglycérols C18 :1, C16 :1, C18 :0 ou TG C54 :2 avec son adduit lithium.

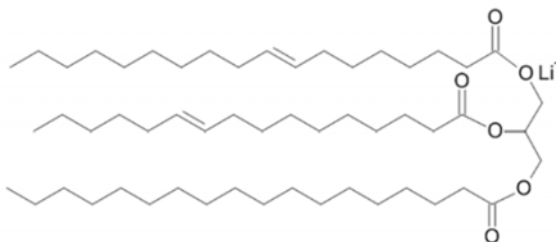


Figure 8 : Triacylglycérols C18 :1, C16 :1, C18 :0 ou TG C54 :2 avec son adduit lithium.

La figure 9 représente le spectre de masse FT-ICR des TAG lithiés extraits d'un échantillon de plasma humain. Ils sont présents dans le spectre de masse dans la gamme m/z 750-950. Les TAGs montrent des nombres de carbone compris entre 46 et 56. Le nombre de doubles liaisons des acides gras est compris entre 0 et 8, avec un nombre maximal de doubles liaisons augmentant avec le carbone. D'après la littérature, il existerait 289 espèces possible [117] et leur concentration dans le plasma est de 150 mg/L ou 1,7 mmol/L.

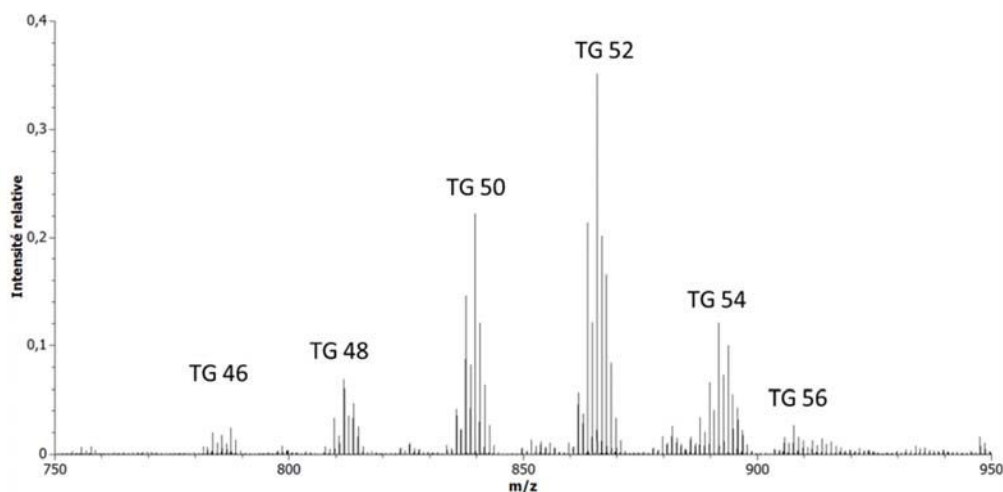


Figure 9 : Spectre nanoESI FT-ICR MS, de la région des TAGs de plasma humain de m/z 750-950.

Avec l'utilisation d'un spectromètre de masse haute résolution comme le FT-ICR, il est possible de déterminer les formules chimiques des TAGs. Le tableau 6 montre une liste des TAGs les plus intense présent dans le spectre présenté figure 9. La formule chimique de ces TAGs a été déterminée ce qui permet de connaître le nombre de carbone et d'insaturation. L'erreur de masse de masse calculée en ppm pour chacun des TAGs montre une variation de 7 à -5 ppm. Les TAG dans le tableau 6 ont des nombres de carbone compris entre 48 et 56 et un nombre de doubles liaisons de de 0 à 7.

Nombre de carbone	Masse observée [M+ ⁶ Li] ⁺	Masse exacte [M+ ⁶ Li] ⁺	erreur ppm	Nom	Formule
48	808,71709	808,72019	3,83	TG(48:2)	C ₅₁ H ₉₄ O ₆ Li
	810,73273	810,73584	3,84	TG(48:1)	C ₅₁ H ₉₆ O ₆ Li
50	834,73235	834,73584	4,18	TG(50:3)	C ₅₃ H ₉₆ O ₆ Li
	836,75559	836,75149	-4,90	TG(50:2)	C ₅₃ H ₉₈ O ₆ Li
	838,76927	838,76714	-2,54	TG(50:1)	C ₅₃ H ₁₀₀ O ₆ Li
	840,78329	840,78279	-0,59	TG(50:0)	C ₅₃ H ₁₀₂ O ₆ Li
52	860,74926	860,75149	2,59	TG(52:4)	C ₅₅ H ₉₈ O ₆ Li
	862,76757	862,76714	-0,50	TG(52:3)	C ₅₅ H ₁₀₀ O ₆ Li
	864,78347	864,78279	-0,79	TG(52:2)	C ₅₅ H ₁₀₂ O ₆ Li
	866,79926	866,79844	-0,95	TG(52:1)	C ₅₅ H ₁₀₄ O ₆ Li
	868,80987	868,81409	4,86	TG(52:0)	C ₅₅ H ₁₀₆ O ₆ Li
54	886,76076	886,76714	7,19	TG(54:5)	C ₅₇ H ₁₀₀ O ₆ Li
	888,77886	888,78279	4,42	TG(54:4)	C ₅₇ H ₁₀₂ O ₆ Li
	890,79787	890,79844	0,64	TG(54:3)	C ₅₇ H ₁₀₄ O ₆ Li
	892,81628	892,81409	-2,45	TG(54:2)	C ₅₇ H ₁₀₆ O ₆ Li
	894,83031	894,82974	-0,64	TG(54:1)	C ₅₇ H ₁₀₈ O ₆ Li
56	910,76207	910,76714	5,57	TG(56:7)	C ₅₉ H ₁₀₀ O ₆ Li
	912,77615	912,78279	7,27	TG(56:6)	C ₅₉ H ₁₀₂ O ₆ Li

Tableau 6 : Liste de 18 TAGs détectés par nanoESI FT-ICR dans le plasma humain. Les TAGs ont été regroupés par nombre de carbones. L'erreur de masse en ppm a été calculée. La formule chimique ainsi que la famille des TAG ont été déterminées à l'aide de la masse des TAGs.

La fragmentation des TAGs par IRMPD permet de les identifier. Hsu et Turk ont décrit la fragmentation des TAG en électronébulisation avec une collision CID [74]. La figure 10 montre un TAG lithié, dans laquelle R'₁, R''₂ et R'''₃ sont des chaînes

hydrocarbonées. Les positions de R'_1 , R''_2 et R'''_3 sont appelées *sn*-1, *sn*-2 et *sn*-3 respectivement

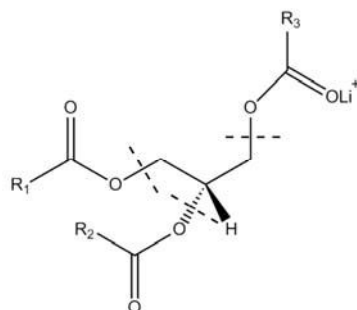


Figure 10 : Structure d'un TAG avec les positions correspondant aux acides gras remplacées par R_1 , R_2 , R_3 . L'adduit lithium est présente sur la fonction carboxyle d'un acide gras. En pointillé sont représentés les sites de fragmentation du TAG.

Les spectres MS2 permettent d'observer la perte des acides gras constituant les TAGs. La différence entre l'ion parent et l'ion fragment permet d'identifier les acides gras libères. Avec la fragmentation IRMPD seule l'ion fragment avec un adduit lithium est observés [23] contrairement à la fragmentation CID où les ions fragments avec et sans lithium est détectées. La figure 11 montre la fragmentation de l'ion parent à m/z 890,7978 $[M+Li]^+$. Un seul ion fragment est produit à m/z 608,5325 $[M+Li-RCH_2CO_2H]^+$ ce qui indique que le TAG est composée d'un seul type d'acides gras. La différence de masse entre l'ion parent et l'ion fragment donne une masse de 282,2653 ce qui permet de déterminer que l'acide gras libéré est un acide oléique. Un ion de faible intensité à m/z 289,2717 est détecté qui correspondant à l'ion $[RCH_2CO_2H+Li]^+$.

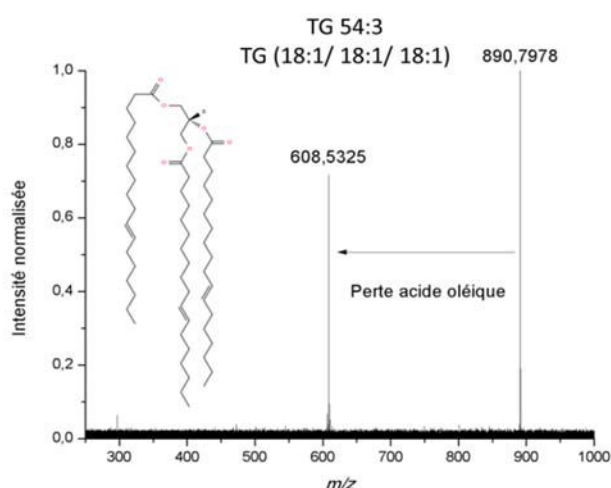


Figure 11 : Spectre de fragmentation IRMPD de l'ion à m/z 890,7. Deux pics sont observés, m/z 890,7 qui correspond à l'ion parent et m/z 608,5 correspond à l'ion fragment, ici la perte de l'acide oléique.

Le TAG à m/z 890,7978 possède la structure C18 :1/ C18 :1/ C18 :1. Ce TAG est aussi appelé trioléine. Ceci peut être réalisé sur chaque pic mais le temps d'expérience sera très élevé. C'est la raison pour laquelle la 2D FT-ICR peut être utilisée car cette méthode apportera des informations de MS et MS/MS en même temps.

3. Cartes 2D FT-ICR des TAGs humain

La figure 12a montre la carte 2D des TAGs humain. L'axe horizontal montre les rapports m/z des ions fragments de m/z 500-950. L'axe vertical montre les rapports m/z des ions précurseurs de m/z 700-950. Pour les TAGs, les ions précurseurs se situent dans la gamme m/z 775-925 (voir la figure 8). Les ions fragments résultant de la perte d'un AG sont dans la gamme m/z 500-650 (figure 11b). Deux lignes de fragments sont observées qui correspondent à une différence C_2H_4 qui est la différence par exemple entre la perte de C18 et C16.

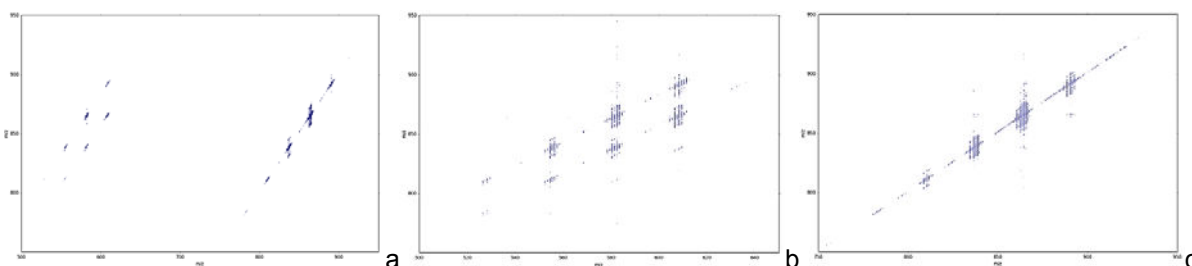


Figure 12 : Spectre 2D-FT-ICR des TAGs du plasma humain avec fragmentation IRMPD. (a) spectre complet 2D-FT des TAG. L'axe horizontal correspond à la gamme de masse des ions fragments, l'axe vertical correspond à la gamme de masse des ions parents, (b) zoom de la région des ions fragments des TAG avec observation de la perte de différent AG, (c) zoom de la région des ions parents des TAG, correspondant à la ligne d'autocorrélation.

L'équation de la ligne d'autocorrélation est $y=x$ et les pics correspondent à la modulation de l'intensité du signal des ions précurseurs selon leur propre fréquence cyclotron [118].

La figure 13 montre la ligne d'autocorrélation extraite du spectre de masse 2D. Les mêmes espèces ioniques sont présentes dans la ligne d'autocorrélation que dans le spectre de masse de l'échantillon TAG (voir la figure 9). Les intensités des signaux des pics dans la ligne d'autocorrélation peuvent être différentes des intensités du spectre de masse 1D car, alors que les intensités des signaux dans les spectres de masse unidimensionnels dépendent principalement des abondances dans la cellule ICR, dans

la ligne d'autocorrélation d'un spectre de masse 2D, elles dépendent à la fois de l'abondance d'ions dans la cellule ICR et de leur efficacité de fragmentation [118].

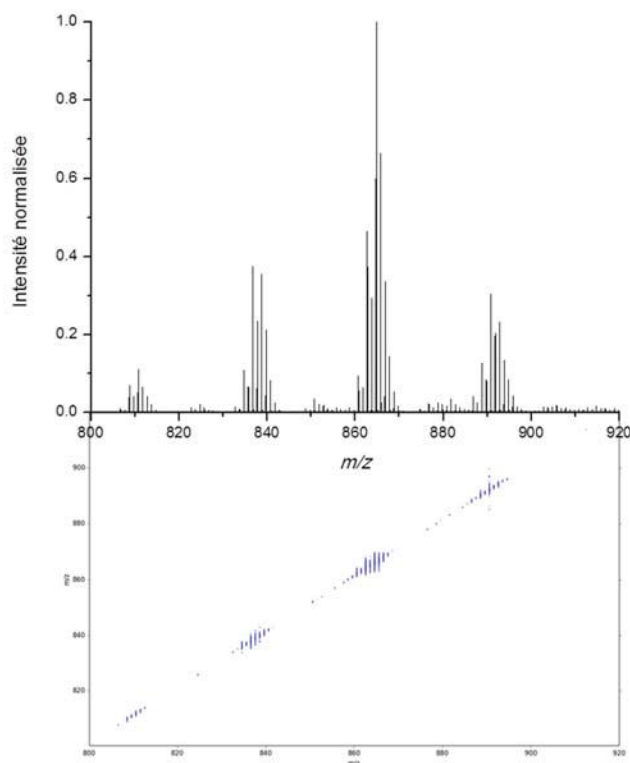


Figure 13 : Spectre 2D FT IRMPD des triacylglycérols extrait du plasma humain de m/z 800-922 sur l'axe des précurseurs et m/z 800-920 sur l'axe des fragments. En haut, le spectre de la ligne d'auto-corrélation de la carte 2D.

La ligne d'autocorrélation extraite du spectre de masse 2D et du spectre de masse 1D ont également des pouvoirs de résolution différents. En 1D MS, à m/z 864,7834 le spectre de masse, obtenu en utilisant une acquisition avec 1 mégapoints et démarrant à m/z 250, à un pouvoir de résolution de 55 000.

Le pic à m/z 864,7842 sur la ligne d'autocorrélation a une résolution horizontale de 29 000 (les paramètres d'acquisition en F2 sont 512 kpoints, avec une masse de départ à m/z 144,39) et une puissance de résolution verticale de 2 200 (à partir de 64 kpoints dans la première dimension en utilisant une 2D carrée : même temps entre les points de mesure en F2 que l'incrément de temps Δt_1 entre les points en F1.).

4. Identification des TAGs avec l'analyse 2D FT-ICR

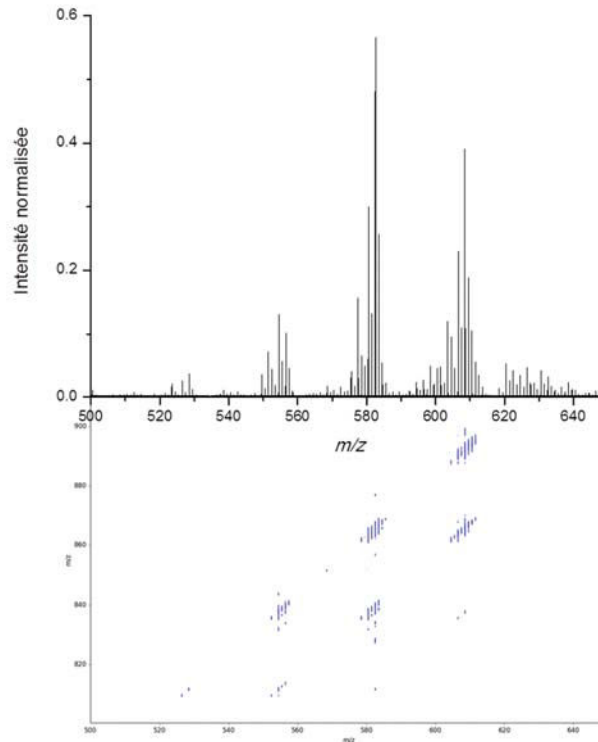


Figure 14 : Spectre 2D FT IRMPD des triacylglycérols extrait du plasma humain de m/z 800-922 sur l'axe des précurseurs et m/z 500-650 sur l'axe des fragments.

La figure 14 montre la zone du spectre de masse 2D correspondant aux pertes neutres des acides gras. Chaque pic correspond à une seule dissociation. Les coordonnées indiquent le rapport m/z de l'ion précurseur (vertical) et le rapport m/z de l'ion fragment (horizontal). Les spectres de fragments comparables aux spectres MS/MS peuvent être extraits selon des lignes horizontales du spectre de masse 2D.

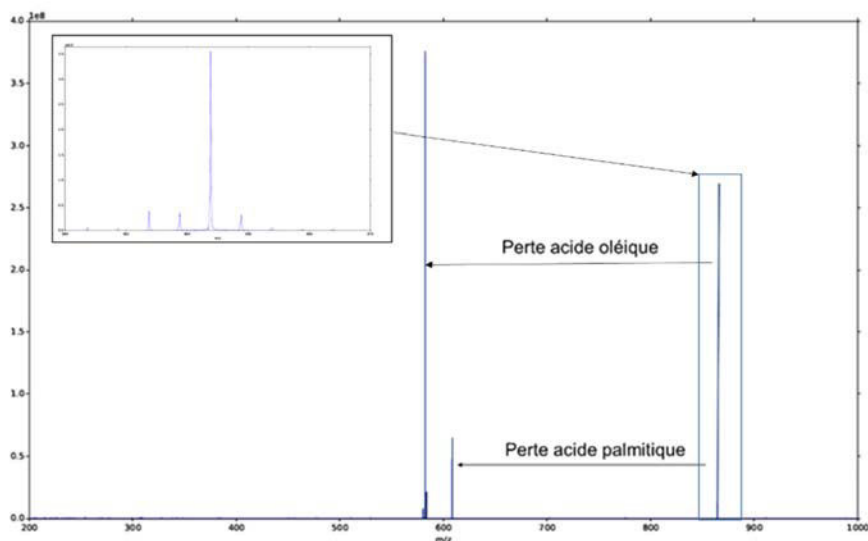


Figure 15 : Spectre de fragmentation de l'ion à m/z 864,7842 extrait du spectre de masse 2D avec le zoom du pic précurseur en haut à gauche.

La figure 15 correspond au profil de l'ion fragment à m/z 864,7842, correspondant à un TAG lithié de 52:2 de formule $C_{55}H_{102}O_6Li$. La masse de l'ion précurseur est mesurée à m/z 864,7842 (-1.63 ppm). Deux fragments sont clairement visibles: le premier à m/z 608,5441 de formule $C_{39}H_{70}O_4Li$ qui peut être affecté à une perte de neutre de 16:0 (-1,64 ppm) et le second à m/z 582,5299 de formule $C_{37}H_{68}O_4Li$ ce qui peut être attribué à une perte de 18:1 (-0,34 ppm).

Une structure possible du TAG à m/z 864,7842 peut donc être attribuée à 16:0/18:1/18:1 dans laquelle le FA 18:1 est en position Sn2, sinon l'ion fragment à m/z 609,5441 de la perte de 16:0 ne serait pas observée [74]. De chaque côté de l'ion précurseur à m/z 864,7842, des pics peuvent être observés : ils correspondent aux isotopes des TAGs qui l'entoure. Ce «dépassement» provient de la largeur du pic dans la dimension verticale, qui est d'environ 2 Da (FWHM, Full Width at Half Maximum).

5. Analyse par pertes de neutres

Une autre façon d'extraire des informations structurales du spectre de masse 2D est de rechercher des lignes de dissociation. Par exemple, les pertes neutres dans un spectre de masse 2D suivent la ligne: $y = x + m_{neutral}$ (équation 1) [71,72,119]

La figure 16 montre des lignes de perte neutres pour chaque nombre de carbones dans la chaîne AG, mais aussi pour chaque nombre d'insaturations, comme on peut le voir dans le zoom représenté sur la figure 16, dans lequel des lignes de perte de neutres pour 284 Da (18:0), 282 Da (18:1) et 280 Da (18:2) peuvent être clairement distingués.

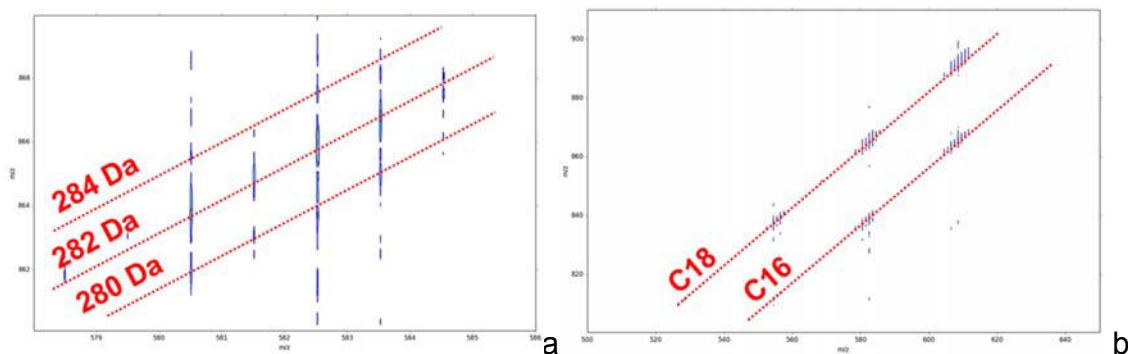


Figure 16: Lignes de perte de neutres pour 284 Da (18:0), 282 Da (18:1) et 280 Da (18:2) extraite du spectre 2D FT-ICR (a). En b la ligne de perte de neutre pour les acides gras C18 et C16 du spectre 2D-FTICR.

A partir de toutes ces informations, il est possible de déterminer les masses des TAGs présents dans le spectre 2D et les fragments correspondants. Les compositions de tous les TAGs les plus intenses sont listées dans le tableau 7.

Nombre de carbones	$[M+{}^6\text{Li}]^+$ observé	$[M+{}^6\text{Li}]^+$ théorique	Erreur en ppm	TG	Formule	composition
48	808,72019	808,72019	-3,70	TG(48:2)	$\text{C}_{51}\text{H}_{94}\text{O}_6\text{Li}$	C16:1/C14:0/C18:1
	810,73584	810,73584	-4,61	TG(48:1)	$\text{C}_{51}\text{H}_{96}\text{O}_6\text{Li}$	C16:0/C14:0/C18:1
50	834,73584	834,73584	-4,60	TG(50:3)	$\text{C}_{53}\text{H}_{96}\text{O}_6\text{Li}$	C16:1/C16:1/C18:1
	836,75149	836,75149	4,31	TG(50:2)	$\text{C}_{53}\text{H}_{98}\text{O}_6\text{Li}$	C16:0/C16:0/C18:2 C16:0/C16:1/C18:1
	838,76714	838,76714	2,58	TG(50:1)	$\text{C}_{53}\text{H}_{100}\text{O}_6\text{Li}$	C16:0/C16:0/C18:1
	840,78279	840,78279	0,73	TG(50:0)	$\text{C}_{53}\text{H}_{102}\text{O}_6\text{Li}$	C16:0/C16:0/C18:0
52	860,75149	860,75149	-1,85	TG(52:4)	$\text{C}_{55}\text{H}_{98}\text{O}_6\text{Li}$	C16:1/C16:1/C18:2
	862,76714	862,76714	0,65	TG(52:3)	$\text{C}_{55}\text{H}_{100}\text{O}_6\text{Li}$	C16:1/C18:1/C18:1
	864,78279	864,78279	1,63	TG(52:2)	$\text{C}_{55}\text{H}_{102}\text{O}_6\text{Li}$	C16:0/C18:1/C18:1
	866,79844	866,79844	-1,55	TG(52:1)	$\text{C}_{55}\text{H}_{104}\text{O}_6\text{Li}$	C16:0/C18:0/C18:1

54	886,76714	886,76714	-8,16	TG(54:5)	$C_{57}H_{100}O_6Li$	C18:2/C18:2/C18:1
	888,78279	888,78279	-4,38	TG(54:4)	$C_{57}H_{102}O_6Li$	C18:2/C18:1/C18:1 C18:0/C18:2/C18:2
	890,79844	890,79844	-1,28	TG(54:3)	$C_{57}H_{104}O_6Li$	C18:1/C18:1/C18:1
	892,81409	892,81409	2,70	TG(54:2)	$C_{57}H_{106}O_6Li$	C18:0/C18:1/C18:1
56	910,76714	910,76714	-5,42	TG(56:7)	$C_{59}H_{100}O_6Li$	C18:1/C18:2/C20:4
	912,78279	912,78279	-8,42	TG(56:6)	$C_{59}H_{102}O_6Li$	C16:1/C20:1/C20:4

Tableau 7 : Liste des TAGs identifiés dans le spectre de masse 2D

La position des insaturations des AG n'a pas pu être déterminée avec la fragmentation IRMPD. Afin d'obtenir cette information, des techniques à base d'électrons comme la dissociation induite par les électrons (EID, Electron Induced Dissociation) [120,121] ou l'excitation par impact électronique des ions à partir des composés organiques (EIEIO, Electron Impact Excitation of Ions from Organics) [122] qui permettent d'obtenir des fragmentations sur des espèces mono ou multichargées en utilisant une haute énergie >10eV ou égale à 10eV. Les diastéréoisomères trans ou cis peuvent être différenciés à l'aide de la dissociation induite par l'ozone (OzID) [123] car cette méthode oxyde les doubles liaisons dans les ions lipidiques et cette oxydation permet de localiser la position de la double liaison.

Une autre solution serait de réaliser des expériences 3D FT-ICR pour permettre d'obtenir une information de MS3.

Par exemple en réalisant une fragmentation CID dans la cellule de collision pour obtenir les ions parents des TAGs et les ions fragments (figure 17).

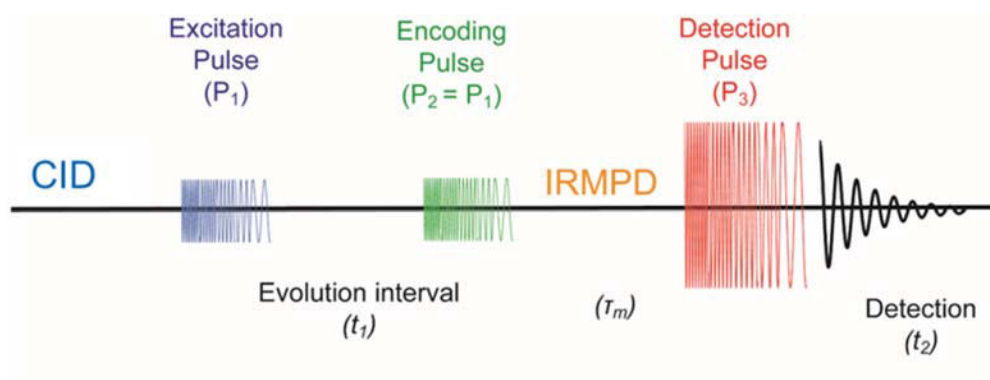


Figure 17 : Exemple de séquence pseudo 3D FT-ICR avec une fragmentation au niveau de cellule de collision.

Ensuite, la séquence 2D FT-ICR pourrait être appliquée pour obtenir une carte 2D avec les informations des fragments des TAGs et aussi la fragmentation des fragments produit en CID.

Il est possible de réaliser une expérience de 3D FT-ICR en réalisant deux séquences 2D FT-ICR l'une après l'autre (figure 18).

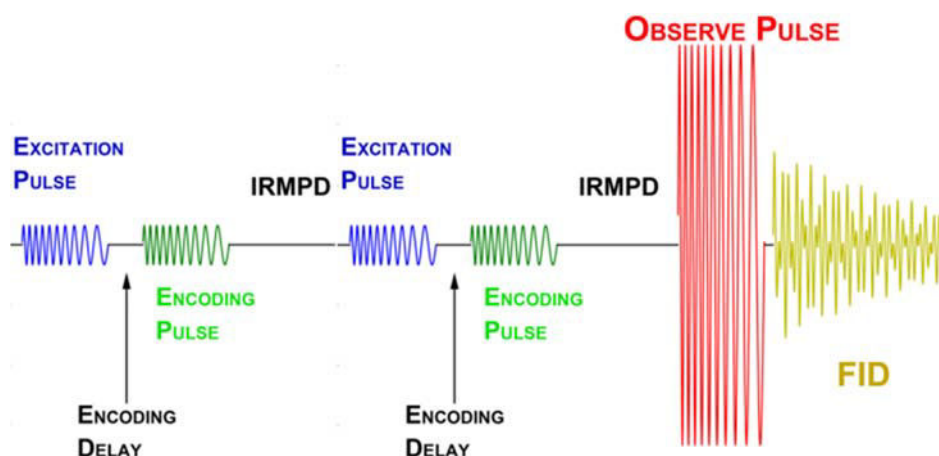


Figure 18 : Séquence d'impulsion 3D FT-ICR.

Cette séquence permettrait dans un premier temps d'avoir la modulation des fragments des TAGs puis avec la seconde séquence d'obtenir la modulation des fragments des premiers fragments. Cette expérience permettrait d'avoir une localisation des insaturations ou des oxidations en atteignant le niveau MS3.

X. Conclusion

L'analyse par spectrométrie de masse FT-ICR unidimensionnelle d'un extrait de TAG d'un échantillon de plasma sanguin humain montre la présence de nombreux TAG. Pour un mélange complexe de TAG, des analyseurs de masse haute résolution comme les spectromètres de masse à temps de vol, Orbitrap ou FT-ICR sont nécessaires afin d'attribuer correctement leur composition élémentaire [124].

Le spectre de masse 2D IRMPD des TAGs montre les pertes neutres des acides gras par les TAG. Nous avons pu déterminer ainsi la composition des acides gras des TAG. Les spectres des fragments ont montré qu'il y avait souvent des structures multiples pour chaque rapport m/z .

Cette étude montre que la 2D FT-ICR MS des extraits de TAG du plasma sanguin peut être utilisée pour les patients et les sujets sains afin de mesurer la composition de acides gras des TAG dans leur plasma sanguin et de corrélérer les résultats avec des informations sur leur alimentation et sur leur santé. Des études sur l'analyse quantitative de la composition de acides gras dans les TAG peuvent être entreprises pour déterminer si les intensités de pic du fragment dans les lignes de perte neutre dans les spectres de masse 2D reflètent l'abondance relative des AG dans les échantillons.

Dans cette étude, l'identité des acides gras constituant les TAGs dans l'échantillon a été déterminée à partir de leur masse moléculaire, qui informe sur le nombre de carbones et le nombre d'insaturations. Afin d'obtenir des informations structurelles sur la localisation des insaturations sur la chaîne AG, des études 2D EID FT-ICR MS pourraient être effectuées. L'utilisation de l'OzID peut être utilisé pour déterminer les conformations cis / trans des insaturations. Cependant, cette technique nécessite l'utilisation d'un gaz interagissant avec les ions, ce qui induirait une perte de résolution en 2D FT-ICR MS.

La spectrométrie de masse 2D peut également être utilisée pour analyser la structure d'autres lipides et peut être utilisée également pour l'empreinte digitale des classes lipidiques d'échantillons biologiques simplement en adaptant la préparation de l'échantillon. L'analyse des TAG en utilisant la spectrométrie de masse 2D peut également être appliquée à des échantillons provenant d'autres origines que le plasma sanguin, comme les échantillons archéologiques, les huiles animales, les membranes végétales, ou phéromones d'insectes.

Bibliographie

- (1) Talayero, B. G.; Sacks, F. M.: The Role of Triglycerides in Atherosclerosis. *Current cardiology reports* **2011**, *13*, 544-552.
- (2) Vergès, B.: Lipid Disorders in Type 1 Diabetes. In *Type 1 Diabetes - Complications, Pathogenesis, and Alternative Treatments*; Liu, C.-P., Ed.; InTech: Rijeka, 2011; pp Ch. 03.
- (3) Nordestgaard, B. G.; Varbo, A.: Triglycerides and cardiovascular disease. *The Lancet* **2014**, *384*, 626-635.
- (4) Sanders, F.; McNally, B.; Griffin, Julian L.: Blood triacylglycerols: a lipidomic window on diet and disease. *Biochem. Soc. Trans.* **2016**, *44*, 638.
- (5) Bucolo, G., & David, H.: Quantitative determination of serum triglycerides by the use of enzymes. *Clinical chemistry* **1973**, *19*, 476-482.
- (6) Carrapiso, A. I., & García, C.: Development in lipid analysis: some new extraction techniques and in situ transesterification. *Lipids* **2000**, *35*, 1167-1177.
- (7) William W.C, X.: Lipids Analysis: Isolation, Separation, Identification and Lipidomics Analysis. *The oily press* **2010**.
- (8) Folch J, L. M., Sloane Stanley GH.: A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem.* **1957**, *226*, 497-509.
- (9) Bligh, E. G., & Dyer, W. J.: A rapid method of total lipid extraction and purification. *anadian journal of biochemistry and physiology* **1959**, *37*, 911-917.
- (10) Colborne, A. J., & Laidman, D. L.: The extraction and analysis of wheat phospholipids. *Phytochem.* **1975**, *14*, 2639-2645.
- (11) Hara, A., & Radin, N. S.: Lipid extraction of tissues with a low-toxicity solvent. *Anal. Biochem.* **1978**, *90*, 420-426.
- (12) Rose, H. G.; Oklander, M.: Improved procedure for the extraction of lipids from human erythrocytes. *J. Lipid Res.* **1965**, *6*, 428-431.
- (13) Honeyman, T. W., Strohsnitter, W., Scheid, C. R., & Schimmel, R. J.: Phosphatidic acid and phosphatidylinositol labelling in adipose tissue. Relationship to the metabolic effects of insulin and insulin-like agents. *Biochem. J.* **1983**, *212*, 489-498.
- (14) Lessig, J., & Fuchs, B.: Plasmalogens in biological systems: their role in oxidative processes in biological membranes, their contribution to pathological processes and aging and plasmalogen analysis. *Current medicinal chemistry* **2009**, *16*, 2021-2041.
- (15) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D.: Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **2008**, *49*, 1137-46.
- (16) Carroll, K. K.: Separation of lipid classes by chromatography on florisil. *Journal of Lipid Research*, **1961**, *2*, 135-141.

- (17) Laakso, P., & Christie, W. W.: Chromatographic resolution of chiral diacylglycerol derivatives: potential in the stereospecific analysis of triacyl-sn-glycerols. *Lipids* **1990**, *25*, 349-353.
- (18) Christie, W. W., Nikolova-Damyanova, B., Laakso, P., & Herslof, B. : Stereospecific analysis of triacyl-sn-glycerols via resolution of diastereomeric diacylglycerol derivatives by high-performance liquid chromatography on silica. *J. Am. Oil Chem. Soc.* **1991**, *68*, 695-701.
- (19) Rouser, G., Kritchevsky, G., Simon, G., & Nelson, G. J.: Quantitative analysis of brain and spinach leaf lipids employing silicic acid column chromatography and acetone for elution of glycolipids. *Lipids* **1967**, *2*, 37-40.
- (20) Prieto, J. A., Ebri, A., & Collar, C.: Optimized separation of nonpolar and polar lipid classes from wheat flour by solid-phase extraction. *J. Am. Oil Chem. Soc.* **1992**, *69*, 387-391.
- (21) Pinkart, H. C., Devereux, R., & Chapman, P. J.: Rapid separation of microbial lipids using solid phase extraction columns. *Journal of Microbiological Methods* **1998**, *34*, 9-10.
- (22) Burdge, G. C.; Wright, P.; Jones, A. E.; Wootton, S. A.: A method for separation of phosphatidylcholine, triacylglycerol, non-esterified fatty acids and cholesterol esters from plasma by solid-phase extraction. *Br. J. Nutr.* **2007**, *84*, 781-787.
- (23) Garnier, N.; Rolando, C.; Høtje, J. M.; Tokarski, C.: Analysis of archaeological triacylglycerols by high resolution nanoESI, FT-ICR MS and IRMPD MS/MS: Application to 5th century BC–4th century AD oil lamps from Olbia (Ukraine). *Int. J. Mass Spectrom.* **2009**, *284*, 47-56.
- (24) Angers, P., & Arul, J.: A simple method for regiospecific analysis of triacylglycerols by gas chromatography. *J. Am. Oil Chem. Soc.* **1999**, *76*, 481-484.
- (25) Buchgraber, M.; Ulberth, F.; Anklam, E.: Capillary GLC: a robust method to characterise the triglyceride profile of cocoa butter— results of an intercomparison study. *European Journal of Lipid Science and Technology* **2003**, *105*, 754-760.
- (26) Hamilton, J. G., & Comai, K.: Separation of neutral lipid, free fatty acid and phospholipid classes by normal phase HPLC. *Lipids* **1988**, *23*, 1150-1153.
- (27) Kalo, P., Kempainen, A., Ollilainen, V., & Kuksis, A.: Regiospecific determination of short-chain triacylglycerols in butterfat by normal-phase HPLC with on-line electrospray-tandem mass spectrometry. *Lipids* **2004**, *39*, 915-928.
- (28) Kalo, P.; Kempainen, A.; Ollilainen, V.: Determination of triacylglycerols in butterfat by normal-phase HPLC and electrospray-tandem mass spectrometry. *Lipids* **2009**, *44*, 169-95.
- (29) Plattner, R. D., Spencer, G. F., & Kleiman, R.: Triglyceride separation by reverse phase high performance liquid chromatography. *J. Am. Oil Chem. Soc.* **1977**, *54*, 511-515.
- (30) Podlaha, O., & Töregård, B.: A system for identification of triglycerides in reversed phase HPLC chromatograms based on equivalent carbon numbers. *J. Sep. Sci.* **1982**, *5*, 553-558.
- (31) Aitzetmüller, K., & Grönheim, M.: Separation of highly unsaturated triacylglycerols by reversed phase HPLC with short wavelength UV detection. *J. Sep. Sci.* **1992**, *15*, 219-226.
- (32) Adlof, R. O.: Normal-phase separation effects with lipids on a silver ion high-performance liquid chromatography column. *J. Chromatogr. A* **1997**, *764*, 337-340.

- (33) Macher, M. B., & Holmqvist, A. (2001). . : Triacylglycerol analysis of partially hydrogenated vegetable oils by silver ion HPLC. *J. Sep. Sci.* **2001**, *24*, 179-185.
- (34) Skipski, V. P., Smolowe, A. F., Sullivan, R. C., & Barclay, M.: Separation of lipid classes by thin-layer chromatography. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* **1965**, *106*, 386-396.
- (35) Tiffany, J. M.: Separation of lipid classes by thin-layer chromatography on urea-silica gel plates. *J. Chromatogr. A* **1982**, *243*, 329-338.
- (36) Bhat, H. K., & Ansari, G. A. S.: Improved separation of lipid esters by thin-layer chromatography. *J. Chromatogr. A* **1989**, *483*, 369-378.
- (37) Momchilova, S., & Nikolova-Damyanova, B.: Stationary phases for silver ion chromatography of lipids: Preparation and properties. *Journal of separation science*, **2003**, *26*, 261-270.
- (38) Yan, Y.; Wang, X.; Liu, Y.; Xiang, J.; Wang, X.; Zhang, H.; Yao, Y.; Liu, R.; Zou, X.; Huang, J.; Jin, Q.: Combined urea-thin layer chromatography and silver nitrate-thin layer chromatography for micro separation and determination of hard-to-detect branched chain fatty acids in natural lipids. *J. Chromatogr. A* **2015**, *1425*, 293-301.
- (39) Fewster, M. E., Burns, B. J., & Mead, J. F.: Quantitative densitometric thin-layer chromatography of lipids using copper acetate reagent. *J. Chromatogr. A* **1969**, *43*, 120-126.
- (40) Manninen, P., Laakso, P., & Kallio, H.: Method for characterization of triacylglycerols and fat-soluble vitamins in edible oils and fats by supercritical fluid chromatography. *J. Am. Oil Chem. Soc.* **1995**, *72*, 1001-1008.
- (41) Manninen, P., & Laakso, P.: Capillary supercritical fluid chromatography-atmospheric pressure chemical ionization mass spectrometry of triacylglycerols in berry oils. *J. Am. Oil Chem. Soc.* **1997**, *75*, 1089-1098.
- (42) Lepage, M.: The separation and identification of plant phospholipids and glycolipids by two-dimensional thin-layer chromatography. *J. Chromatogr. A* **1964**, *13*, 99-103.
- (43) Demirbüker, M., & Blomberg, L. G.: Group separation of triacylglycerols on micropacked argentation columns using supercritical media as mobile phases. *J. Chromatogr. Sci.* **1990**, *28*, 67-72.
- (44) Demirbüker, M., & Blomberg, L. G. : Separation of triacylglycerols by supercritical-fluid argentation chromatography. *J. Chromatogr. A* **1991**, *550*, 765-774.
- (45) Murphy, R. C.; Axelsen, P. H.: Mass spectrometric analysis of long-chain lipids. *Mass Spectrom. Rev.* **2011**, *30*, 579-99.
- (46) Kohlwein, S. D.: Opinion articles on lipidomics — A critical assessment of the state-of-the-art. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862*, 729-730.
- (47) Kopczynski, D.; Coman, C.; Zahedi, R. P.; Lorenz, K.; Sickmann, A.; Ahrends, R.: Multi-OMICS: a critical technical perspective on integrative lipidomics approaches. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862*, 808-811.

- (48) Liebisch, G.; Ekroos, K.; Hermansson, M.; Ejsing, C. S.: Reporting of lipidomics data should be standardized. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862*, 747-751.
- (49) Triebel, A.; Hartler, J.; Trötz Müller, M.; Köfeler, H.: Lipidomics: Prospects from a technological perspective. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862*, 740-746.
- (50) Murata, T., & Takahashi, S.: Analysis of triglyceride mixtures by gas chromatography-mass spectrometry. *Analytical chemistry*, **1973**, *45*, 1816-1823.
- (51) Murata, T., & Takahashi, S.: Qualitative and quantitative chemical ionization mass spectrometry of triglycerides. *Anal. Chem.* **1977**, *49*, 728-731.
- (52) Murata, T.: Analysis of triglycerides by gas chromatography/chemical ionization mass spectrometry. *Analytical chemistry*, **1977**, *49*, 2209-2213.
- (53) Quehenberger, O.; Armando Am Fau - Dennis, E. A.; Dennis, E. A.: High sensitivity quantitative lipidomics analysis of fatty acids in biological samples by gas chromatography-mass spectrometry. **2011**.
- (54) RP, E.: High-resolution triacylglycerol mixture analysis using high-temperature gas chromatography / mass spectrometry with a polarizable stationary phase, negative ion chemical ionization, and mass-resolved chromatography. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 350-361.
- (55) Ji, H., Voinov, V. G., Deinzer, M. L., & Barofsky, D. F. : Distinguishing between cis/trans isomers of monounsaturated fatty acids by FAB MS. *Analytical chemistry*, **2007**, *79*, 1519-1522.
- (56) Gil, J. H.; Hong, J. Y.; Jung, J. H.; Kim, K. J.; Hong, J.: Structural determination of monoacylglycerols extracted from marine sponge by fast atom bombardment tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1264-70.
- (57) Hong, J., Kim, Y. H., Gil, J. H., Cho, K., Jung, J. H., & Han, S. Y.: Structural determination of hexadecanoic lysophosphatidylcholine regioisomers by fast atom bombardment tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2089-2093.
- (58) Ohashi, Y., Tanaka, T., Akashi, S., Morimoto, S., Kishimoto, Y., & Nagai, Y.: Squid nerve sphingomyelin containing an unusual sphingoid base. *Journal of Lipid Research*, **2000**, *41*, 1118-1124.
- (59) Ahn, Y. M.; Lee, W. W.; Jung, J. H.; Lee, S. G.; Hong, J.: Structural determination of glucosylceramides isolated from marine sponge by fast atom bombardment collision-induced dissociation linked scan at constant B/E. *J. Mass Spectrom.* **2009**, *44*, 1698-708.
- (60) Asbury, G. R., Al-Saad, K., Siems, W. F., Hannan, R. M., & Hill, H. H.: Analysis of triacylglycerols and whole oils by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 983-991.
- (61) Petkovic, M.; Schiller, J.; Muller, M.; Benard, S.; Reichl, S.; Arnold, K.; Arnhold, J.: Detection of individual phospholipids in lipid mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: phosphatidylcholine prevents the detection of further species. *Anal. Biochem.* **2001**, *289*, 202-16.

- (62) Guyon, F.; Absalon, C.; Eloy, A.; Salagoity, M. H.; Esclapez, M.; Medina, B.: Comparative study of matrix-assisted laser desorption/ionization and gas chromatography for quantitative determination of cocoa butter and cocoa butter equivalent triacylglycerol composition. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2317-22.
- (63) Cerruti, C. D.; Benabdellah, F.; Laprevote, O.; Touboul, D.; Brunelle, A.: MALDI imaging and structural analysis of rat brain lipid negative ions with 9-aminoacridine matrix. *Anal. Chem.* **2012**, *84*, 2164-71.
- (64) Pittenauer, E.; Allmaier, G.: The renaissance of high-energy CID for structural elucidation of complex lipids: MALDI-TOF/RTOF-MS of alkali cationized triacylglycerols. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1037-47.
- (65) Cvacka, J.; Svatos, A.: Matrix-assisted laser desorption/ionization analysis of lipids and high molecular weight hydrocarbons with lithium 2,5-dihydroxybenzoate matrix. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2203-7.
- (66) Popkova, Y.; Schiller, J.: Addition of CsCl reduces ion suppression effects in the matrix-assisted laser desorption/ionization mass spectra of triacylglycerol/phosphatidylcholine mixtures and adipose tissue extracts. *Rapid Commun. Mass Spectrom.* **2017**, *31*, 411-418.
- (67) Kubo, A.; Satoh, T.; Itoh, Y.; Hashimoto, M.; Tamura, J.; Cody, R. B.: Structural analysis of triacylglycerols by using a MALDI-TOF/TOF system with monoisotopic precursor selection. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 684-9.
- (68) Fuchs, B.; Schiller, J.; Suss, R.; Schurenberg, M.; Suckau, D.: A direct and simple method of coupling matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF MS) to thin-layer chromatography (TLC) for the analysis of phospholipids from egg yolk. *Anal. Bioanal. Chem.* **2007**, *389*, 827-34.
- (69) Murphy, R. C.; Hankin, J. A.; Barkley, R. M.: Imaging of lipid species by MALDI mass spectrometry. *J. Lipid Res.* **2009**, *50 Suppl*, S317-22.
- (70) Han, X.; Gross, R. W.: Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *J. Lipid Res.* **2003**, *44*, 1071-9.
- (71) Han, X.; Gross, R. W.: Quantitative analysis and molecular species fingerprinting of triacylglyceride molecular species directly from lipid extracts of biological samples by electrospray ionization tandem mass spectrometry. *Anal. Biochem.* **2001**, *295*, 88-100.
- (72) McAnoy, A. M.; Wu, C. C.; Murphy, R. C.: Direct qualitative analysis of triacylglycerols by electrospray mass spectrometry using a linear ion trap. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1498-509.
- (73) Currie, G. J., & Kallio, H.: Triacylglycerols of human milk: rapid analysis by ammonia negative ion tandem mass spectrometry. *Lipids* **1993**, *28*, 217-222.
- (74) Hsu, F. F., & Turk, J.: Structural characterization of triacylglycerols as lithiated adducts by electrospray ionization mass spectrometry using low-energy collisionally activated dissociation on a triple stage quadrupole instrument. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 587-599.

- (75) Herrera, L. C.; Potvin, M. A.; Melanson, J. E.: Quantitative analysis of positional isomers of triacylglycerols via electrospray ionization tandem mass spectrometry of sodiated adducts. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2745-52.
- (76) Cheng, C., Gross, M. L., & Pittenauer, E.: Complete structural elucidation of triacylglycerols by tandem sector mass spectrometry. *Anal. Chem.* **1998**, *70*, 4417-4426.
- (77) Kuksis, A., Marai, L., & Myher, J. J.: Reversed-phase liquid chromatography-mass spectrometry of complex mixtures of natural triacylglycerols with chloride-attachment negative chemical ionization. *J. Chromatogr. A* **1991**, *588*, 73-87.
- (78) Zhao, Z.; Xu, Y.: Measurement of endogenous lysophosphatidic acid by ESI-MS/MS in plasma samples requires pre-separation of lysophosphatidylcholine. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 3739-42.
- (79) Mottram, H. R., Woodbury, S. E., & Evershed, R. P.: Identification of triacylglycerol positional isomers present in vegetable oils by high performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1240-1252.
- (80) Mu, H., & Høy, C. E.: Application of atmospheric pressure chemical ionization liquid chromatography-mass spectrometry in identification of lymph triacylglycerols. *Journal of Chromatography B: Biomedical Sciences and Applications* **2000**, *748*, 425-437.
- (81) Wahid, H.: Changes in the triacylglycerol content of flaxseeds during development using liquid chromatography- atmospheric pressure photoionization-mass spectrometry (LC-APPI-MS). *African Journal of Biotechnology* **2012**, *11*.
- (82) Gerbig, S.; Takats, Z.: Analysis of triglycerides in food items by desorption electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2186-92.
- (83) Eberlin, L. S.; Ferreira, C. R.; Dill, A. L.; Ifa, D. R.; Cooks, R. G.: Desorption electrospray ionization mass spectrometry for lipid characterization and biological tissue imaging. *Biochim. Biophys. Acta* **2011**, *1811*, 946-60.
- (84) Vaclavik, L.; Cajka, T.; Hrbek, V.; Hajslova, J.: Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Anal. Chim. Acta* **2009**, *645*, 56-63.
- (85) Lisa, M.; Cifkova, E.; Holcapek, M.: Lipidomic profiling of biological tissues using off-line two-dimensional high-performance liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2011**, *1218*, 5146-56.
- (86) Hutchins, P. M.; Barkley, R. M.; Murphy, R. C.: Separation of cellular nonpolar neutral lipids by normal-phase chromatography and analysis by electrospray ionization mass spectrometry. *J. Lipid Res.* **2008**, *49*, 804-13.
- (87) Laakso, P., & Manninen, P. I.: Identification of milk fat triacylglycerols by capillary supercritical fluid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Lipids* **1997**, *32*, 1285-1295.

- (88) Marai, L., Myher, J. J., & Kuksis, A.: Analysis of triacylglycerols by reversed-phase high pressure liquid chromatography with direct liquid inlet mass spectrometry. *Canadian journal of biochemistry and cell biology* **1983**, *31*, 840-849.
- (89) Fauconnot, L.; Hau, J.; Aeschlimann, J. M.; Fay, L. B.; Dionisi, F.: Quantitative analysis of triacylglycerol regioisomers in fats and oils using reversed-phase high-performance liquid chromatography and atmospheric pressure chemical ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 218-24.
- (90) Leskinen, H. M.; Suomela, J. P.; Kallio, H. P.: Quantification of triacylglycerol regioisomers by ultra-high-performance liquid chromatography and ammonia negative ion atmospheric pressure chemical ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1-5.
- (91) Dugo, P.; Favoino, O.; Tranchida, P. Q.; Dugo, G.; Mondello, L.: Off-line coupling of non-aqueous reversed-phase and silver ion high-performance liquid chromatography–mass spectrometry for the characterization of rice oil triacylglycerol positional isomers. *J. Chromatogr. A* **2004**, *1041*, 135-142.
- (92) Nagai, T.; Mizobe, H.; Otake, I.; Ichioka, K.; Kojima, K.; Matsumoto, Y.; Gotoh, N.; Kuroda, I.; Wada, S.: Enantiomeric separation of asymmetric triacylglycerol by recycle high-performance liquid chromatography with chiral column. *J. Chromatogr. A* **2011**, *1218*, 2880-6.
- (93) Marzilli, L. A., Fay, L. B., Dionisi, F., & Vouros, P.: Structural characterization of triacylglycerols using electrospray ionization-MS_n ion-trap MS. *J. Am. Oil Chem. Soc.* **2003**, *80*, 195-202.
- (94) Brodbelt, J. S.: Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* **2014**, *43*, 2757-83.
- (95) 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. *Analytical chemistry*, **2006**, *78*, 779-787.
- (96) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M.: MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010**, *11*, 395.
- (97) O'Connor, A., Brasher, C. J., Slatter, D. A., Meckelmann, S. W., Hawksworth, J. I., Allen, S. M., & O'Donnell, V. B.: LipidFinder: A computational workflow for discovery of lipids identifies eicosanoid-phosphoinositides in platelets. *JCI insight* **2017**, *2*.
- (98) <https://sciex.com/products/software/lipidview-software>
- (99) <https://www.thermofisher.com/blog/proteomics/processing-a-complex-lipid-data-set-with-lipidsearch-software/>
- (100) <http://premierbiosoft.com/lipid/index.html>
- (101) <https://sciex.com/products/software/multiquant-software>
- (102) Haimi, P., Uphoff, A., Hermansson, M., & Somerharju, P.: Software tools for analysis of mass spectrometric lipidome data. *Anal. Chem.* **2006**, *78*, 8324-8331.

(103) Ejsing, C. S., Duchoslav, E., Sampaio, J., Simons, K., Bonner, R., Thiele, C. & Shevchenko, A. (2006). : Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. *Anal. Chem.* **2006**, *78*, 6202-6214.

(104) Yang, K., Cheng, H., Gross, R. W., & Han, X.: Automated lipid identification and quantification by multidimensional mass spectrometry-based shotgun lipidomics. *Anal. Chem.* **2009**, *81*, 4356-4368.

(105) Herzog, R., Schwudke, D., Schuhmann, K., Sampaio, J. L., Bornstein, S. R., Schroeder, M., & Shevchenko, A.: A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome biology* **2011**, *12*.

(106) Husen, P.; Tarasov, K.; Katafiasz, M.; Sokol, E.; Vogt, J.; Baumgart, J.; Nitsch, R.; Ekroos, K.; Ejsing, C. S.: Analysis of lipid experiments (ALEX): a software framework for analysis of high-resolution shotgun lipidomics data. *PLoS One* **2013**, *8*, e79736.

(107) Kyle, J. E.; Crowell, K. L.; Casey, C. P.; Fujimoto, G. M.; Kim, S.; Dautel, S. E.; Smith, R. D.; Payne, S. H.; Metz, T. O.: LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics* **2017**, *33*, 1744-1746.

(108) Meng, D.; Zhang, Q.; Gao, X.; Wu, S.; Lin, G.: LipidMiner: a software for automated identification and quantification of lipids from multiple liquid chromatography/mass spectrometry data files. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 981-5.

(109) Sabareesh, V.; Singh, G.: Mass spectrometry based lipid(ome) analyzer and molecular platform: a new software to interpret and analyze electrospray and/or matrix-assisted laser desorption/ionization mass spectrometric data of lipids: a case study from Mycobacterium tuberculosis. *J. Mass Spectrom.* **2013**, *48*, 465-77.

(110) Song, H.; Hsu, F. F.; Ladenson, J.; Turk, J.: Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1848-58.

(111) Schwudke, D., Oegema, J., Burton, L., Entchev, E., Hannich, J. T., Ejsing, C. S. & Shevchenko, A.: Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. *Anal. Chem.* **2006**, *78*, 585-595.

(112) Sud, M.; Fahy E Fau - Cotter, D.; Cotter D Fau - Brown, A.; Brown A Fau - Dennis, E. A.; Dennis Ea Fau - Glass, C. K.; Glass Ck Fau - Merrill, A. H., Jr.; Merrill Ah Jr Fau - Murphy, R. C.; Murphy Rc Fau - Raetz, C. R. H.; Raetz Cr Fau - Russell, D. W.; Russell Dw Fau - Subramaniam, S.; Subramaniam, S.: LMSD: LIPID MAPS structure database. **2006**.

(113) Tsubugawa, H.; Ikeda, K.; Arita, M.: The importance of bioinformatics for connecting data-driven lipidomics and biological insights. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862*, 762-765.

(114) Foster, J. M.; Moreno P Fau - Fabregat, A.; Fabregat A Fau - Hermjakob, H.; Hermjakob H Fau - Steinbeck, C.; Steinbeck C Fau - Apweiler, R.; Apweiler R Fau - Wakelam, M. J. O.; Wakelam Mj

Fau - Vizcaino, J. A.; Vizcaino, J. A.: LipidHome: a database of theoretical lipids optimized for high throughput mass spectrometry lipidomics. **2013**.

(115) Ruiz-Gutierrez, V., & Perez-Camino, M. C: Update on solid-phase extraction for the analysis of lipid classes and related compounds. *J. Chromatogr. A* **2000**, *885*, 321-341.

(116) Lin, J. T., & Arcinas, A. : Analysis of Regiospecific Triacylglycerols by Electrospray Ionization– Mass Spectrometry³ of Lithiated Adducts. *J. Agric. Food Chem.* **2008**, *56*, 4909-4915.

(117) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S.: The Human Serum Metabolome. *Plos One* **2011**, *6*, e16957.

(118) van Agthoven, M. A.; Chiron, L.; Coutouly, M.-A.; Sehgal, A. A.; Pelupessy, P.; Delsuc, M.-A.; Rolando, C.: Optimization of the discrete pulse sequence for two-dimensional FT-ICR mass spectrometry using infrared multiphoton dissociation. *Int. J. Mass Spectrom.* **2014**, *370*, 114-124.

(119) ReynoldHoman, M. K.: Rapid separation and quantitation of combined neutral and polar lipid classes by high-performance liquid chromatography and evaporative light-scattering mass detection. *Journal of Chromatography B: Biomedical Sciences and Applications* **1998**, *708*, 21-26.

(120) Wei, J.; Li, H.; Barrow, M. P.; O'Connor, P. B.: Structural characterization of chlorophyll-a by high resolution tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 753-760.

(121) Wills, R. H.; O'Connor, P. B.: Structural characterization of actinomycin D using multiple ion isolation and electron induced dissociation. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 186-195.

(122) Baba, T.; Campbell, J. L.; Le Blanc, J. C.; Baker, P. R.: Structural identification of triacylglycerol isomers using electron impact excitation of ions from organics (EIEIO). *J. Lipid Res.* **2016**, *57*, 2015-2027.

(123) Pham, H. T.; Maccarone, A. T.; Campbell, J. L.; Mitchell, T. W.; Blanksby, S. J.: Ozone-induced dissociation of conjugated lipids reveals significant reaction rate enhancements and characteristic odd-electron product ions. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 286-296.

(124) Marshall, A. G.; Hendrickson, C. L.: High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579-599.

Partie 3 :

FT-ICR 2D: Analyse haute résolution

I. FT-ICR 2D : Analyse haute résolution

1. Introduction

A l'heure actuelle la FT-ICR 2D est limitée par la résolution sur l'ion précurseur de l'ordre de plusieurs unités de masse. Dans la première dimension qui définit la résolution de l'ion précurseur, pour chaque intervalle de temps t_1 , un nouveau paquet d'ion est réinjecté car l'étape de fragmentation est irréversible à la différence de l'excitation en RMN. L'intervalle de temps t_1 correspond à la moitié de l'inverse de la fréquence de la masse la plus basse pour correspondre au critère de Nyquist. La résolution est proportionnelle au maximum du temps t_1 qui est un intervalle de temps multiplié par le nombre de point d'acquisition [1]. Le temps d'analyse d'un scan est de 1 seconde, donc pour une analyse avec un t_1 de 2048 points en FT-ICR 2D, le temps d'analyse sera de 30 minutes avec un nombre de point d'acquisition de 128k à 512k pour la seconde dimension qui définit la résolution des ions fragments. Ainsi pour obtenir une haute résolution sur le précurseur, il faut augmenter le nombre de points dans la première dimension pour atteindre la même résolution que dans la seconde dimension mais cette opération est difficile car coûteuse en temps.

Une des méthodes utilisée en spectrométrie RMN multidimensionnelle pour accélérer les acquisitions, est l'échantillonnage non-uniforme (NUS) [2]. L'acquisition NUS consiste en sauter des points lors de l'acquisition, ce qui permet de diminuer le temps d'acquisition. Un exemple est présenté en figure 1.

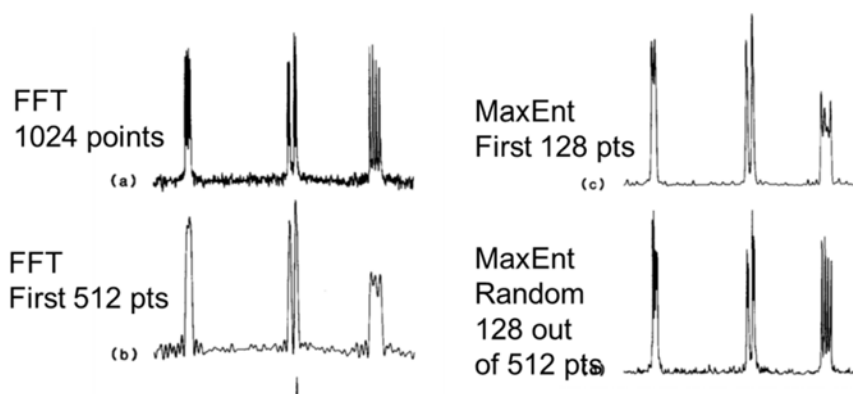


Figure 1 : Spectres proton de l'acide 2-furoïque dans le diméthyle sulfoxyde d_6 . (a) Transformée de Fourier (FT) des 512 premiers points d'un signal de 1024 points. (b) FT des 512 premiers points. (c) Reconstruction avec l'algorithme MaxEntropy en utilisant les 128 premiers points et (d) reconstruction avec l'algorithme MaxEntropy à partir de 128 point échantillonnés aléatoirement parmi les 512 points du signal mesuré (extrait de J.C.J Barna et al [3]).

Afin de récupérer l'information structurale, les points manquants doivent être reconstruits par un traitement de données. Dans la RMN multi-dimensions, cela se fait généralement en utilisant un traitement basé sur l'algorithme d'entropie maximum introduit par J. Skilling [4,5]. Plus récemment, des algorithmes avancés basés sur la détection comprimée (compress sensing) et sur les reconstructions à faible rang (low-rank reconstruction) ont été introduits. Malheureusement, la dépendance de taille de ces algorithmes empêche leur utilisation dans la spectrométrie de masse FTICR car leur temps d'exécution est proportionnel au nombre de points au carré dans la première dimension. De même, l'algorithme doit être répété pour tous les points de la deuxième dimension.

En effet, la taille des données 2D FTICR MS est beaucoup plus grande que celle de RMN 2D. Une petite FT-ICR 2D de $8k \times 128k = 1024$ Mega points permet une résolution sur les précurseurs dans la 1^{er} dimension similaire à un quadripôle (3 Da), comme celle utilisée en protéomique. Ce qui peut être comparé à une analyse 2D RMN de $2k \times 2k = 4$ Mega points. La taille d'une analyse 2D FTICR carrée est beaucoup plus élevée, par exemple $128k \times 128k = 16$ Giga points jusqu'à $4 M \times 4 M = 16$ Tera points.

Dans le chapitre II, partie I, l'algorithme de débruitage urQRd [6] qui est basé sur la décomposition QR d'une matrice échantillonnée au hasard à partir des données, a été présenté. Cet algorithme de débruitage urQRd évite la décomposition en valeur singulière (SVD) d'une matrice qui varie en temps d'exécution comme le cube de la longueur de données. La décomposition SVD d'une matrice est un outil important de factorisation des matrices rectangulaires réelles ou complexes. Cette décomposition cependant ne peut pas être utilisée sur des ensembles de données très volumineux. Donc l'utilisation de l'algorithme urQRd permet un gain d'au moins 3 ordres de grandeur en temps de traitement comparé à la décomposition SVD classique, réduit fortement la mémoire nécessaire de l'ordinateur et permet d'appliquer l'algorithme de débruitage à une taille de données pratiquement illimitée (chapitre II, partie I). Cet algorithme ne dépend que deux paramètres, le niveau de reconstruction qui est le nombre de pics à détecter et le nombre d'itérations qui correspond au nombre de fois que l'algorithme est appliqué. Contrairement à l'algorithme classique de Cadzov, urQRd n'est pas sensible au rang et au nombre d'itérations estimées. L'algorithme urQRd peut être utilisé pour

reconstituer les valeurs manquantes, en les considérant comme des valeurs corrompues par une sorte de bruit.

Les données NUS acquises sont incluses dans la totalité du signal enregistré dans lequel les points manquants sont initialement mis à zéro. Une version adaptée de l'algorithme urQRd optimisée pour les signaux faibles est appliquée au signal complet, après quoi les points de données expérimentales sont réinjectés. Les étapes combinées sont ensuite itérées. Une transformée de Fourier est ensuite appliquée sur la reconstruction. Les données sont traitées d'abord le long de F2 (seconde dimension, spectres MS classique pour les fragments) avec remplissage par zéros (zero-filling) de la taille des données et apodisation qui permettent de réduire les pics parasites de chaque côté des pics [7], [8]. Ensuite, une seconde transformée de Fourier est appliquée selon l'axe F1 (première dimension, spectres des précurseurs en fonction de l'intervalle de temps t_1) est traité avec l'algorithme urQRd avec réinjection des données, puis, en utilisant une représentation de données hypercomplexes, et en visualisant après remplissage par zéros (zero-filling) de la taille des données, les données en tant que module [9].

Pour la RMN bidimensionnelle, le meilleur échantillonnage non uniforme correspond à une distribution exponentielle, avec une densité d'échantillonnage inversement proportionnelle à la vitesse de relaxation. Pour la 2D FTICR MS, la situation est différente car pour chaque point une nouvelle injection d'ions doit être faite dans la dimension indirecte et avec l'utilisation d'une cellule harmonisée, le signal dans la dimension directe n'est pas presque amorti. Dans ces conditions, l'échantillonnage NUS le plus simple est une distribution uniforme. Le ratio NUS pour une taille d'acquisition physique donne un intervalle moyen. Les intervalles entre zéro et deux intervalles moyens sont générés aléatoirement ce qui conduit à des intervalles maximaux plus petits que l'échantillonnage de type Poisson [10].

2. Objectifs

L'analyse FT-ICR 2D a montré qu'il était possible d'analyser de nombreux types d'échantillons. L'implémentation sur des spectromètres de masses FT-ICR commerciaux est possible. Mais, pour l'instant la résolution obtenue dans la première dimension (F1 = dimension des précurseurs) est loin d'égaliser la résolution du FT-ICR. C'est pour cela que l'application d'acquisition avec un échantillonnage non uniforme pour le temps t_1 est intéressante car cette méthode permettrait d'augmenter la résolution dans la première dimension et des analyses 2D-FT-ICR plus fines.

II. Matériels et méthodes

1. Préparation des échantillons

Une solution de substance P à 1 pmol/ μ L a été préparée dans une solution méthanol/eau v/v avec 0,1% acide formique. Cette solution a été utilisée pour les mises au point de l'acquisition NUS.

Une solution d'huile de sésame a été préparée avec 1 μ L d'huile mélangé à 99 μ L de chloroforme. Puis 1 μ L de cette solution a été mélangé à 998 μ L 90% méthanol, 10% chloroforme et 1 μ L de chlorure de lithium (1 mg/mL dans le méthanol). Cette solution d'huile de sésame a été préparée pour montrer l'efficacité de l'acquisition NUS sur un échantillon plus complexe que le peptide de la substance P.

2. Analyse FT-ICR de la substance P

L'analyse a été réalisé sur un SolariX XR 12 T FTICR MS équipé d'une cellule ParaCell harmonisée de Bruker Daltonik GmbH (Bremen, Allemagne) en injection directe avec une seringue (Hamilton Gasthigh 250 μ L) à un débit de 10 μ L/h (pousse-seringue, Cole-Parmer®, US). Le voltage appliqué sur l'aiguille était de 1650 V. La fragmentation a été réalisée par ECD avec un ampérage de 1,6 A et 0,05 seconde de temps d'irradiation, la lentille ECD (ECD lens) à 10 volt et le potentiel relatif d'ECD (ECD bias) à 1 volt. L'ion parent de la substance P chargé 2 fois à m/z 675,37 a été sélectionnée avec le quadripôle avec une résolution de \pm 10 Da. La gamme de masse est de m/z 184,308 à m/z 1500 (1 MHz à 122 kHz).

Les paramètres pour réaliser l'expérience 2D FT-ICT sont 128k pour le nombre de point d'acquisition du FT-ICR, le pulse de détection a une amplitude de $P_3 = 100$ Vpp et une durée de 20 μ s pour la fréquence des pas. Pour les deux impulsions P_1 (pulse d'excitation), P_2 (pulse d'encodage), l'amplitude est de 70 Vpp et la fréquence des pas de 0,5 μ s. Les impulsions P_1 et P_2 ont une durée totale chacun de 723 μ s. La liste de temps t_1 s'appelle VD_list= VD_2D (fichier de temps), le nombre de point correspond au temps t_1 est la valeur L_20 = 2048.

Toutes les acquisitions ont été réalisées avec une taille physique de 2k \times 128k points correspondant à une acquisition de 30 minutes. A partir d'une acquisition

classique aux acquisitions NUS de taille 4k, 8k, 16k, 32k et 64k correspondant aux ratios NUS 1/2, 1/4, 1/8, 1/16 et 1/32.

3. Analyse FT-ICR de l'huile de sésame

L'analyse a été réalisé sur un Apex 9.4 T FTICR MS équipé d'une cellule Infinity de Bruker Daltonik GmbH (Bremen, Allemagne) en injection direct avec une seringue (Hamilton Gasthigh 250 μ L) à un débit de 10 μ L/h (pousse seringue, Cole-Parmer®, US). Le voltage appliqué sur l'aiguille était de 1700 V. La fragmentation a été réalisée avec un laser infrarouge CO₂ IRMPD avec une puissance de 50% et un temps d'irradiation de 0,4 sec.

Les paramètres pour réaliser l'expérience 2D FT-ICT sont 512k pour le nombre de point d'acquisition du FT-ICR, le pulse de détection à une amplitude de 100 Vpp et une durée de 20 μ s pour la fréquence des pas. Pour les deux pulse P_1 (pulse d'excitation), P_2 (pulse d'encodage), l'amplitude est de 70 Vpp et la fréquence des pas est de 0,5 μ s. Les pulses P_1 et P_2 ont une durée totale chacun de 723 μ s. La liste de temps t_1 s'appelle VD_list= VD_2D (fichier de temps), le nombre de point correspond à L_20 = 8 192. La gamme de masse utilisée est de m/z 144,38 à m/z 3 000 (1 MHz à 48 kHz).

4. Retraitement informatique

Le retraitement des données a été réalisé sur un cluster Dell R930 équipé de 4 processeurs Xeon comprenant 48 cœurs avec la suite Spike open source écrite en Python et parallélisée par MPI (Message Passing Interface) permettant un traitement rapide. Les listes de temps t_1 pour les acquisitions classiques ou NUS ont été réalisées à partir d'un programme écrit sous Python.

III. Résultats

Pour déterminer la portée et la limitation de l'acquisition 2D FTICR, nous avons utilisé la substance P, ionisée par nanoESI à une concentration analytique de 1 pmol/ μ L. L'ion doublement chargé à m/z 675,37 a été sélectionné puis fragmenté par ECD. Nous avons enregistré des spectres de masse 2D FTICR avec la séquence d'impulsions améliorée conçue précédemment et décrite dans la figure 7 [11]. Le pulse d'observation final utilise les paramètres analytiques par défaut pour l'excitation des ions. Les deux impulsions de codages initiaux avant la fragmentation ECD sont identiques et ont une durée la plus courte possible (1 μ s) et une tension plus faible correspondant à environ un rayon ionique après excitation deux fois plus grand que le faisceau de dissociation.

1. Intervalle de temps t_1

La liste de temps t_1 est réalisée à l'aide d'un programme réalisé sous Python. Le tableau 1 montre une partie de la liste de temps pour une analyse non NUS, NUS 2 et NUS 32 pour une acquisition de 2 048 points dans la première dimension et une 128 k points dans la seconde dimension.

	2 048	4 096	65 636
taille initiale	2 048	4 096	65 636
taille de l'acquisition	2 048	2 048	2 048
Rapport NUS	1	0,5	0,03125
Dwell time	0,5	0,5	0,5
1	0,500u	0,500u	0,500u
2	1,000u	1,500u	9,500u
3	1,500u	2,000u	27,500u
4	2,000u	3,500u	32,000u
5	2,500u	5,000u	33,500u
6	3,000u	6,000u	63,500u
7	3,500u	7,000u	77,000u
8	4,000u	8,500u	100,000u
9	4,500u	10,000u	125,000u
10	5,000u	11,000u	126,500u
11	5,500u	12,000u	134,000u
12	6,000u	13,500u	152,000u
13	6,500u	15,000u	179,000u
14	7,000u	15,500u	204,000u
15	7,500u	16,000u	231,500u

16	8,000u	16,500u	255,500u
17	8,500u	17,500u	265,000u
18	9,000u	18,500u	267,500u
19	9,500u	19,500u	298,000u
20	10,000u	20,000u	326,000u

Table 1 : Liste des 20 premiers temps de l'intervalle t_i pour une analyse de 2048 points avec une acquisition non NUS, NUS 2, NUS 32. L'évolution du temps avec le ratio NUS est clairement visible.

Pour le vingtième temps par exemple, le temps t_i pour le facteur NUS 2 est de 20 μ s tandis que le temps pour un facteur NUS 1 est de 10 μ s. Un facteur 2 est appliqué entre les deux séries. Pour le ratio NUS 32 le second temps est de 326 μ s. Pour ce ratio 31 des 32 points sont omis. Avec cette méthode d'acquisition, le nombre de points dans la première dimension reste identique entre les différents ratios mais le temps de chaque point d'acquisition est différent. Ceci permet, pour un ratio NUS 2, de réaliser une acquisition physique de 2048 points et d'avoir un nombre de point qui correspond à une acquisition de 4096 points, et pour le NUS 32, d'obtenir une acquisition de 65 635 points. L'algorithme urQRd est utilisé pour reconstituer les valeurs manquantes lors du traitement de données.

2. Analyse NUS de la substance P

L'échantillon utilisé pour les premières expériences est la substance P qui est un peptide dont la formule chimique de la substance P est $C_{63}H_{97}N_{17}O_{14}S_1$, ce qui correspond à une masse $M = 1347,7121$ g/mol. L'ion monoisotopique $[M+H]^+$ est à $m/z = 1348,7194$ et l'ion di chargée $[M+H]^{2+}$ est à $m/z = 674,8633$. La séquence en acide aminé est [Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met]. Ce peptide est facilement ionisable et la fragmentation est connue en ECD ce qui permettra une analyse plus facile des spectres [12].

La figure 2 montre la 2D de la substance P $2k \times 128k$ non NUS. Sur cette carte 2D, l'ion parent de la substance P est visible ainsi que la ligne de fragment.

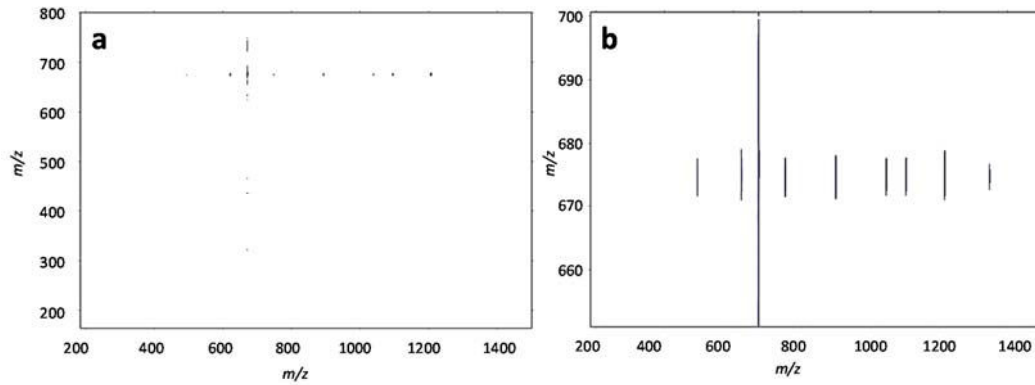


Figure 2 : Carte 2D de l'analyse de la substance P avec fragmentation ECD pour une analyse 2k × 128k non NUS. (a) carte 2D complète, (b) zoom de la ligne de fragmentation de la substance P.

La figure 3 (a) montre le zoom de l'ion parent de la substance P avec ses isotopes. Avec une acquisition 2k × 128k non NUS, les isotopes de l'ion parent se chevauche ce qui ne permet pas d'obtenir un spectre de fragmentation pour chaque isotope de la substance P. Le spectre de fragmentation en figure 3 (b) est propre. La fragmentation ECD de l'ion à m/z 674,471 est facilement observable et l'identification de la séquence est rapide.

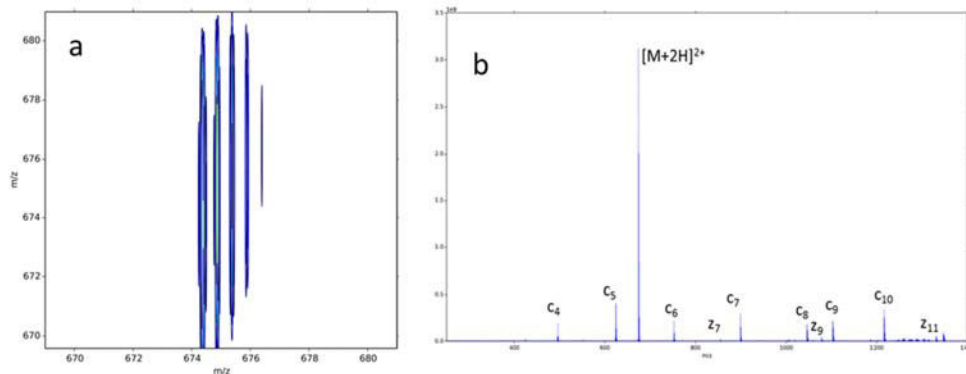


Figure 3 : (a) Zoom de l'ion parent de la substance P à m/z 675,38 avec les annotations du logiciel Spike, (b) Profile de la fragmentation de la substance P avec les annotations des fragments z et c.

Dans les spectres 2D de la figure 4, les spectres de fragmentation clairs sont visibles à tous les rapports NUS avec un bruit de scintillation légèrement croissant lorsque le rapport NUS diminue.

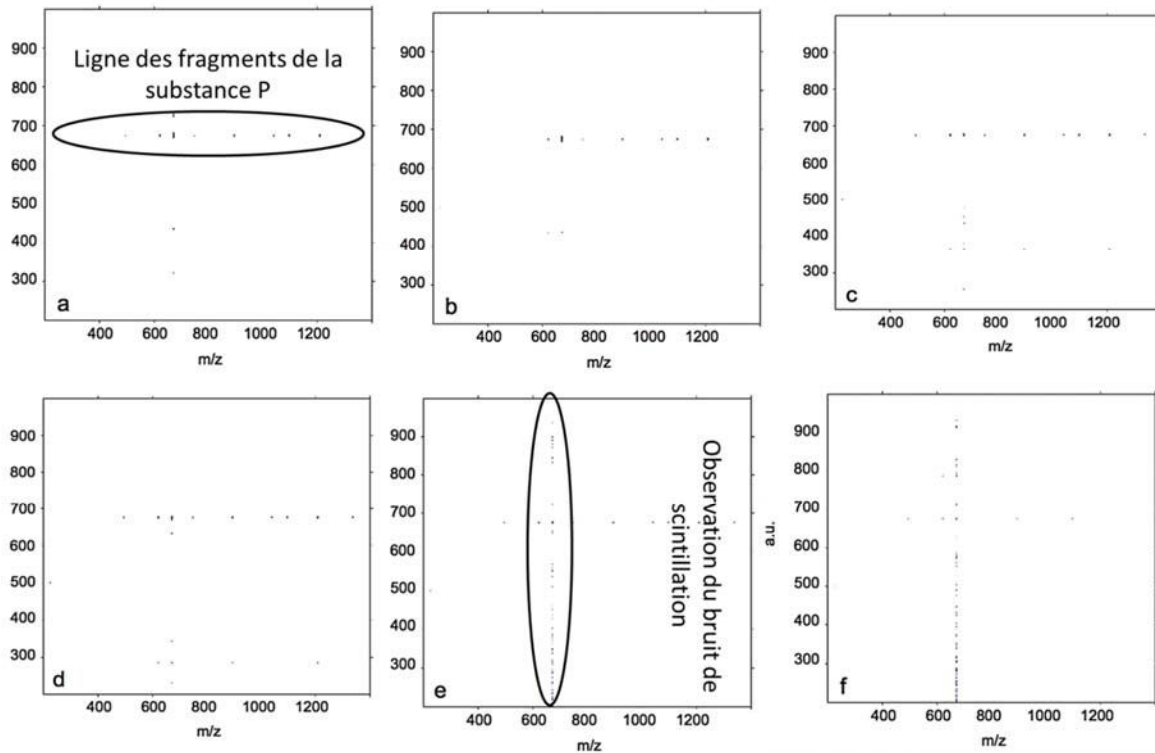


Figure 4 : Spectres 2D complet pour les différents ratios de NUS. a) NUS 2, a) NUS 4, a) NUS 8, a) NUS 16, a) NUS 32.

Pour obtenir une vue plus précise, la figure 5 affiche les spectres des fragments du pic monoisotopique normalisé sur le fragment c_5 (m/z 312,7007) à tous les rapports NUS. L'aspect du spectre des fragments est bien conservé. La variation des intensités de pointe est causée par une résolution progressive du pic monoisotopique t_1 lorsque la taille d'acquisition varie de 2k à 64k. La table 2 montre les masses des fragments de type c et z de la substance P avec l'ECD.

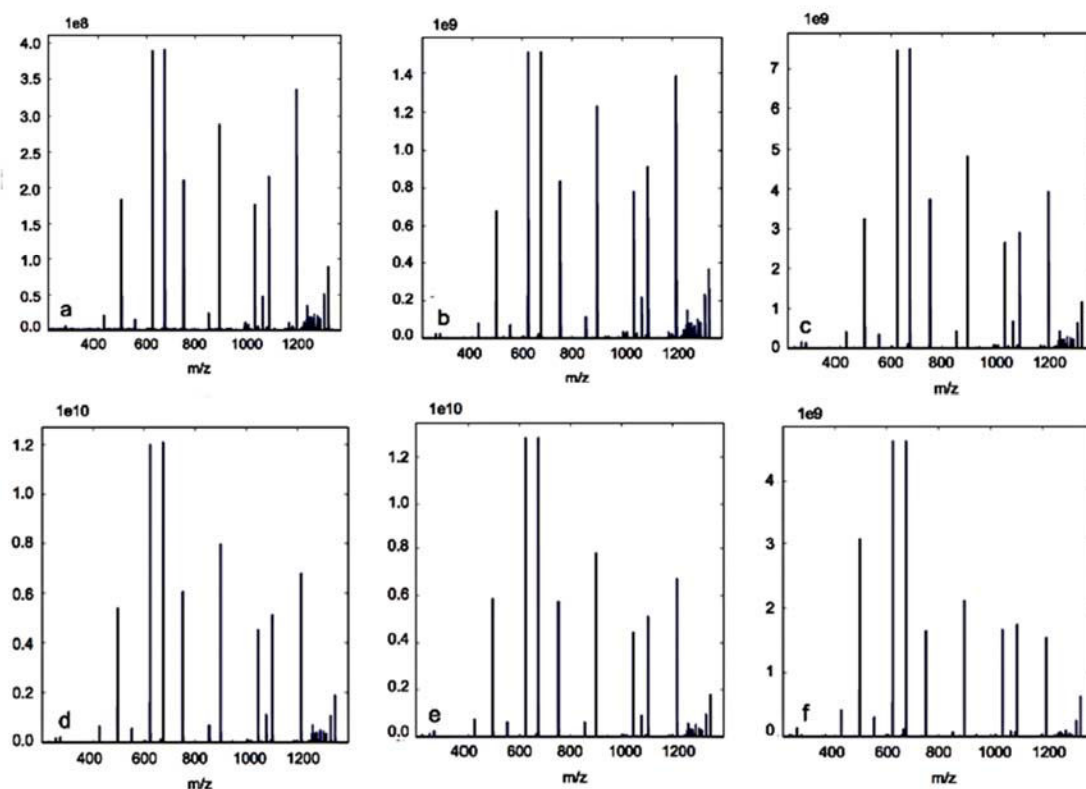


Figure 5 : Spectres des fragments du pic monoisotopique normalisé sur le fragment c5 (m/z 312,7007) à tous les rapports NUS. a) NUS 2, a) NUS 4, a) NUS 8, a) NUS 16, a) NUS 32.

AA	#	c	z	#
R	1	174,13498	1331,69403	11
P	2	271,18774	1175,59292	10
K	3	399,28271	1078,54015	9
P	4	496,33547	950,44519	8
Q	5	624,39405	853,39243	7
Q	6	752,45263	725,33385	6
F	7	899,52104	597,27527	5
F	8	1046,58945	450,20686	4
G	9	1103,61092	303,13845	3
L	10	1216,69498	246,11698	2
M	11	-	133,03292	1

Tableau 2 : Liste des masses des fragments c et z théorique de la substance P.

Le tableau 2 décrit le nombre de fragments attribués qui passe de 13 fragments pour l'acquisition simple de 2k à 30 fragments pour l'acquisition NUS 1/16 32k. L'évolution de la résolution FWHM du pic du précurseur dans la première dimension

(Figure 7) montre que, comme prévu, cette résolution augmente linéairement comme l'inverse du rapport NUS jusqu'à NUS 1/16 (Figure 7) et est légèrement inférieure à NUS 1/32.

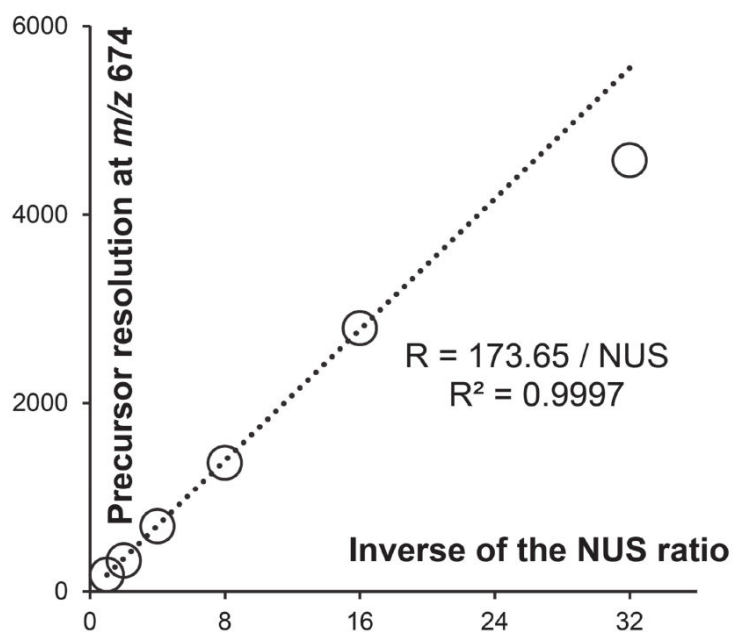


Figure 6 : Résolution à mi-hauteur de l'ion précurseur de la substance P en fonction de l'inverse du rapport de NUS.

La résolution pour NUS 1/16 atteint la valeur de 2 785 à m/z 674, ce qui correspond à une résolution normalisée d'environ 10 000 à m/z 200. Il est à noter que même au ratio NUS 1/32 (c'est-à-dire en sautant 31 points sur 32), la forme du pic est préservée (figure 7, f).

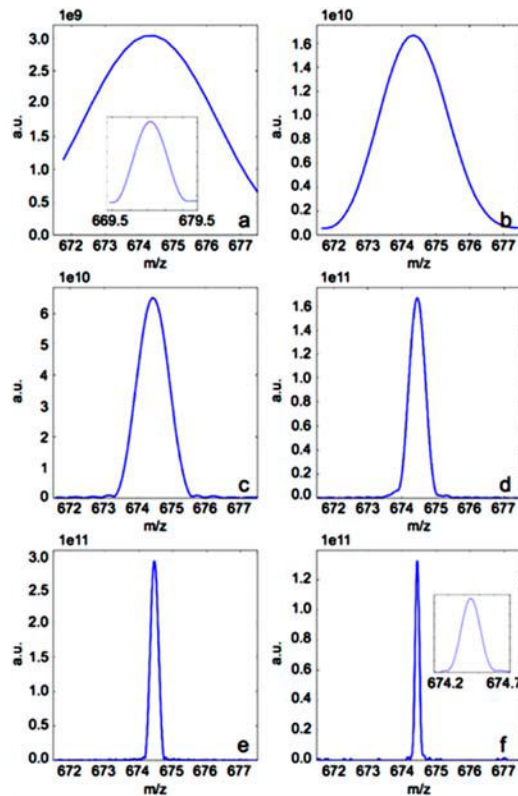


Figure 7 : Spectre de masse 2D FTICR du pic monoisotopique de la Substance P doublement chargé. Zoom le long de la dimension F1 (précurseur) (de A, 2k Acquisition classique; B, 4k, NUS 1/2; C, 8k, NUS 1/4; D, 16k, NUS 1 / 8; E, 32k, NUS 1/16; F, 64k, NUS 1/32). Tous les pics sont définis avec 24 points. Ce nombre est constant car la résolution augmente proportionnellement à la taille d'acquisition.

Comme prévu, les fragments à masse faible ou élevée ont la même résolution que le précurseur dans la 1^{ère} dimension, comme le montre le tableau 3.

taille (k)	NUS ratio	temps de traitement (h)	Résolution en F1			fragments identifiés	ratio intensité des fragments / précurseur
			Précurseur m/z 674,371	Fragment α m/z 496,335	Fragment m/z 1347,735		
2		1h18	176	180	178	13	65
4	1/2	1h34	325	348	335	20	65
8	1/4	2h14	687	626	626	24	75
16	1/8	3h42	1365	1258	1326	24	66
32	1/16	6h39	2795	2799	2830	30	63
64	1/32	12h17	4576	4528	51119	17	23

Table 3 : Temps de traitement avec l'algorithme urQRd pour chaque ratio de NUS avec les informations de résolution sur le précurseur et deux fragments ainsi que le nombre de fragments identifiés pour chaque ratio NUS et le rapport des intensités des fragments sur celle du précurseur.

À partir de la taille 32k (NUS 1/16), les isotopes du précurseur sont complètement séparés et le spectre des fragments de chaque isotope est extrait du spectre 2D. Dans la figure 8a, on peut voir le zoom du spectre 2D sur le pic parent avec la détection du 4^{ème} isotope à m/z 676,37. Comme son intensité est seulement de 0,84%, la plage dynamique est supérieure à 100. La figure 8b affiche le zoom du spectre 2D de l'espèce mono chargé réduite. Deux processus chimiques sont visibles: la capture d'un électron et la capture d'un électron suivie d'une perte d'atome d'hydrogène (figure 8b). Si les isotopes ne sont pas séparés, l'interprétation des données est compliquée alors qu'elle est plus aisée dans le cas présent [13].

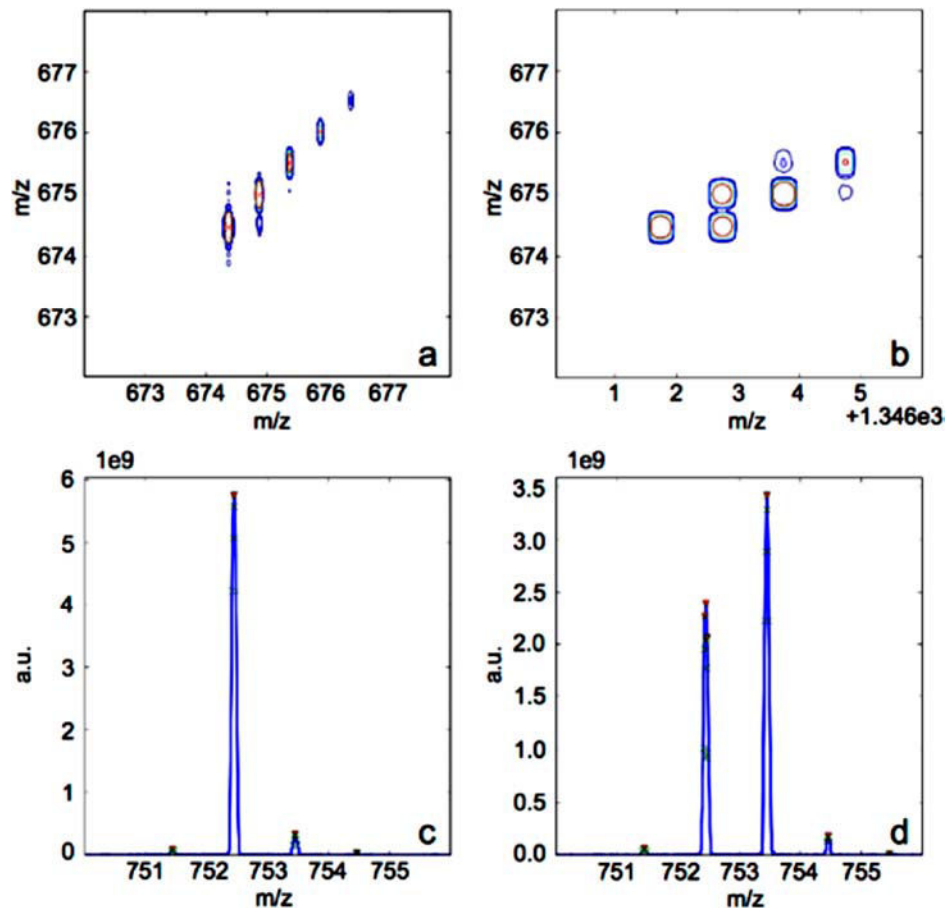


Figure 8 : Spectre de masse 2D FTICR du pic monoisotopique de la Substance P doublement chargé avec une acquisition 32k, NUS 1/16. (a) zoom sur le précurseur avec deux charge à m/z 675,3791, (b) zoom sur le pic monochargé réduit, (c) zoom sur le fragment c_6 à m/z 752,45263 correspondant au pic monoisotopique, (d) zoom sur le fragment c_6 correspondant au pic du 1^{er} isotope.

Enfin, la figure 8c et d montre le zoom du fragment c6 correspondant au pic monoisotopique et au premier isotope. Comme prévu, le pic monoisotopique ne produit qu'un seul pic (le petit pic à M+1 est dû à la capture d'hydrogène) tandis que le premier isotope donne deux pics dans un rapport d'intensité proche de celui prévu par le nombre d'atomes de carbone restant dans le fragment (valeurs théoriques $m/z = \frac{m+1}{z}$; observées).

Dans les tableaux de la publication [14] qui sont en annexe (publication recherche), les valeurs observées pour tous les fragments du pic monoisotopique et les trois premiers isotopes sont données et comparées à leurs valeurs théoriques calculées avec le logiciel IsoPatrn [15]. Les tableaux de la publication montrent que deux phénomènes contraires se déroulent à mesure que le ratio NUS diminue:

- L'augmentation de la résolution apporte une meilleure séparation des isotopes requis pour obtenir des ratios isotopiques corrects,
- La qualité de la reconstruction, en particulier pour les pics avec une faible intensité diminue. Pour les spectres physiques 2k, le meilleur compromis est obtenu pour NUS 1/16.

3. Acquisition NUS pour l'analyse de l'huile de sésame.

Une application de la FT-ICR 2D haute résolution avec acquisition NUS a été réalisée sur un échantillon d'huile de sésame. La figure 9 montre l'analyse FT-ICR MS de l'huile de sésame. Les triacylglycérols (TAGs) observés sont cationisé avec de lithium pour améliorer l'ionisation et obtenir des spectres de fragmentations simple [16].

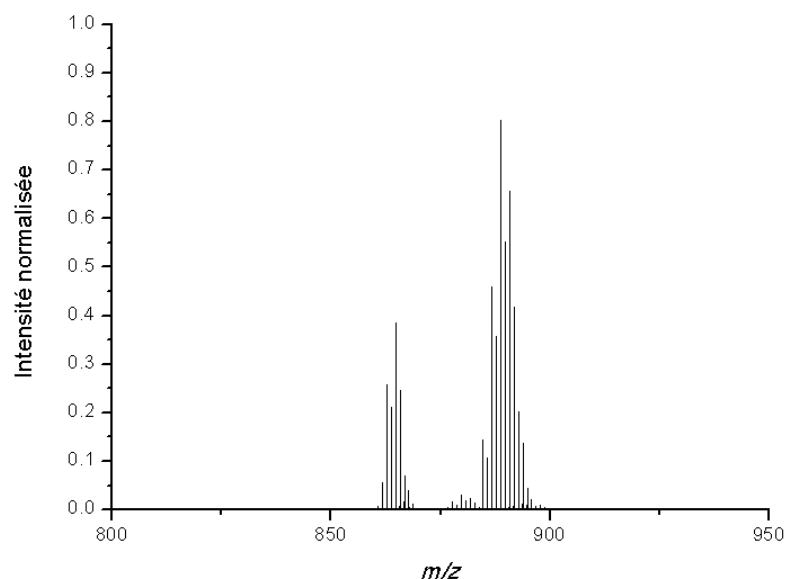


Figure 9 : Spectre MS des TAGs de l'huile de sésame.

A partir de ce spectre, il est possible de déterminer les TAGs présent avec la haute précision de mesure du FT-ICR et l'utilisation de base de données de lipides LIPIDS-MAPS. Dans le tableau 4, la liste des TAGs identifier avec LIPIDS MAPS est présentée. Les TAGs sont identifiés avec une erreur de masse entre 1 et 8 ppm. Cette identification permet d'obtenir le nombre de carbone et d'insaturation des TAGs mais pas les acides gras qui le composent.

$[M+{}^6\text{Li}]^+$	TAG	Formule	$[M+{}^6\text{Li}]^+$	erreur ppm
860,7495	TG(52:4)	$\text{C}_{55}\text{H}_{98}\text{O}_6\text{Li}$	860,7509	1,6
862,7592	TG(52:3)	$\text{C}_{55}\text{H}_{100}\text{O}_6\text{Li}$	862,7665	8,5
864,7796	TG(52:2)	$\text{C}_{55}\text{H}_{102}\text{O}_6\text{Li}$	864,7822	3
884,7503	TG(54:6)	$\text{C}_{57}\text{H}_{98}\text{O}_6\text{Li}$	884,7509	0,6
886,7614	TG(54:5)	$\text{C}_{57}\text{H}_{100}\text{O}_6\text{Li}$	886,7665	5,8
888,7780	TG(54:4)	$\text{C}_{57}\text{H}_{102}\text{O}_6\text{Li}$	888,7822	4,7
890,7966	TG(54:3)	$\text{C}_{57}\text{H}_{104}\text{O}_6\text{Li}$	890,7978	1,4
892,8084	TG(54:2)	$\text{C}_{57}\text{H}_{106}\text{O}_6\text{Li}$	892,8135	5,7

Tableau 4 : Liste des TAGs identifiés avec la base de données LIPIDS-MAPS.

Pour valider ces résultats, la sélection de chaque pic et sa fragmentation doit être réalisé pour obtenir son identification et les acides gras qui le composent. Un exemple de spectre de fragmentation avec l'IRMPD est présenté figure 10. Sur ce spectre, l'ion

parents est observé à m/z 890,7966 et l'ion fragment à m/z 608,5378. La différence entre ces deux masses est de 282 Da ce qui permet de déterminer l'acide gras qui le compose. Pour ce TAG, seul l'acide oléique (O) le compose donc il s'agit donc de la trioléine (OOO).

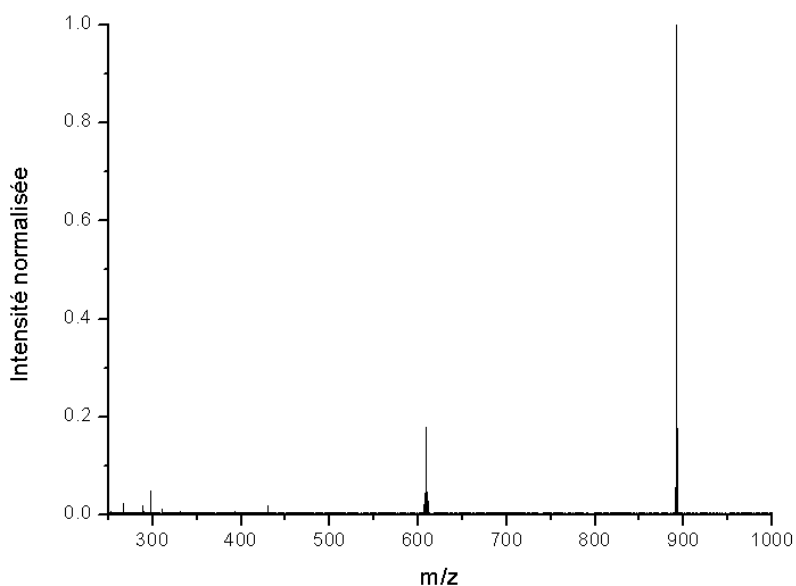


Figure 10 : Spectre de fragmentation IRMPD du TAG à m/z 890,7. L'ion fragment à m/z 608,5 correspond à la perte de l'acide oléique.

L'analyse classique en FT-ICR est longue et fastidieuse. C'est pour cela que l'application de la 2D-FT-ICR avec acquisition NUS permettra une analyse plus rapide. Avec l'acquisition NUS, il sera possible d'obtenir les spectres de fragmentations de chaque pic et d'identifier chaque TAG.

Le spectre 2D (64k × 512k à NUS 1/8) de triglycérides d'huile de sésame cationisés avec un adduit de lithium 6 isotopiquement pur et fragmentés par IRMPD est présenté figure 11. Les triglycérides sont constitués de trois acides gras et lors de la fragmentation, les fragments principaux observés correspondent à la perte d'acide gras comme vu précédemment. L'analyse FT-ICR 2D MS peut être effectuée dans n'importe quelle cellule ICR.

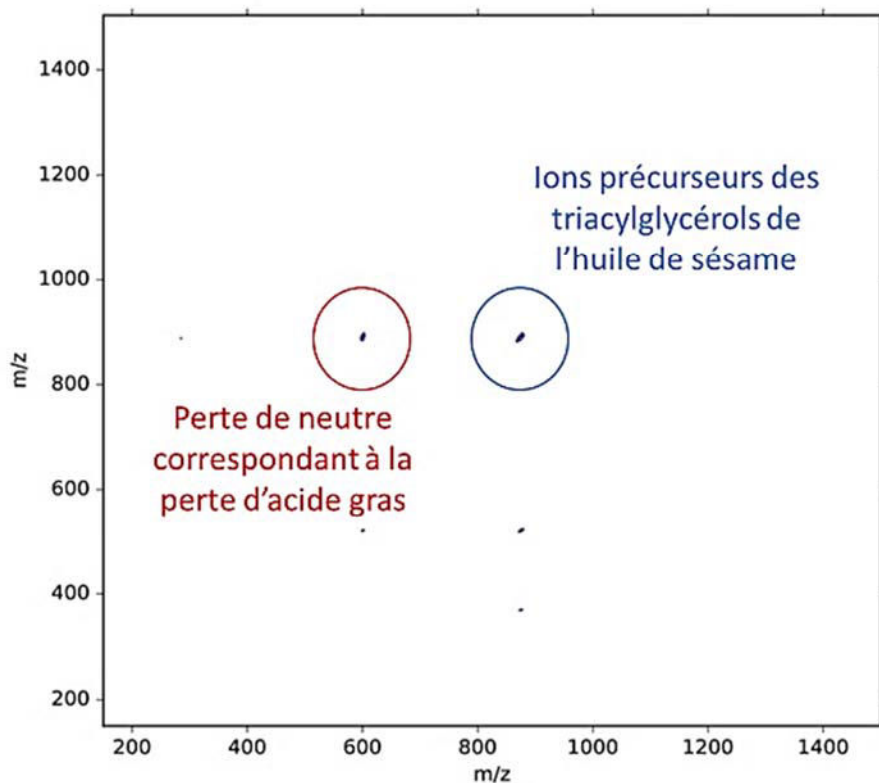


Figure 11 : Spectre 2D FT de l'huile de sésame cationisé avec un adduit lithium et une fragmentation IRMPD.

Dans la publication [14] présenté en annexe, publication recherche, la MS/MS des précurseurs des TAGs obtenue par isolement avec l'analyseur de quadripôle est donnée dans la Figure S5 a-e avec une table du m/z attendu (Tableau S11).

Sur le spectre 2D complet, la ligne de perte de neutre est clairement visible (Figure 11, cercle rouge) et, lorsqu'elle est agrandie, la perte d'acide linoléique et de l'acide oléique, qui diffère uniquement par 2 hydrogènes, se distingue aisément (figure 12a).

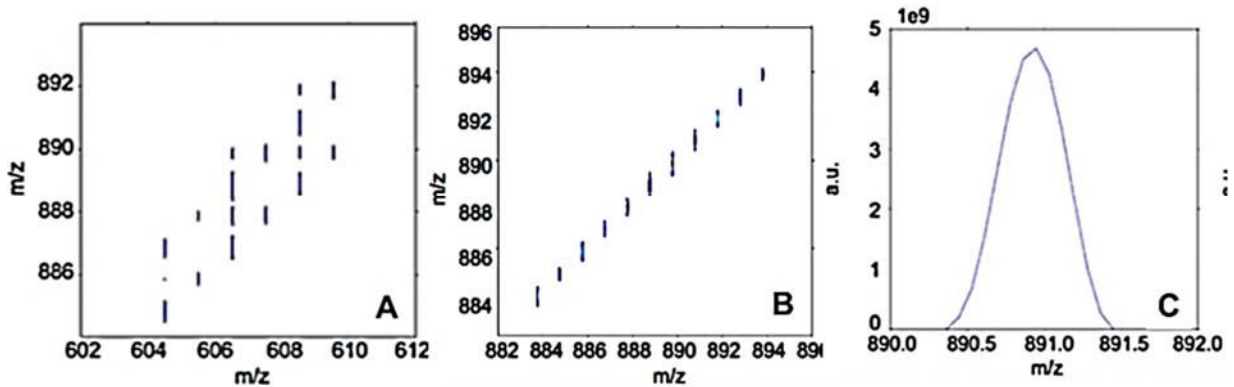


Figure 12 : (A) Zoom de la région des fragments sur le spectre 2D, (b) zoom de la région des précurseurs sur le spectre 2D et (C) zoom du précurseur à m/z 890,7 correspondant à la trioléine (OOO).

La figure 12b affiche un zoom sur la zone des précurseurs des TAGs de m/z 882 à 895 et la figure 12c correspond au profil d'un précurseur le long de la dimension F1. La résolution est 0,44 Da qui correspond à une résolution de masse d'environ 2000 à m/z 890 et donc à une résolution normalisée de 9000 à m/z 200.

À partir du précurseur à m/z 890,8, qui est la trioléine symétrique (OOO), un spectre des fragments présentant uniquement un pic à m/z 608,5 est obtenu (figure 13a). Les plus petits pics sont attribuables aux contributions isotopiques qui se chevauchent.

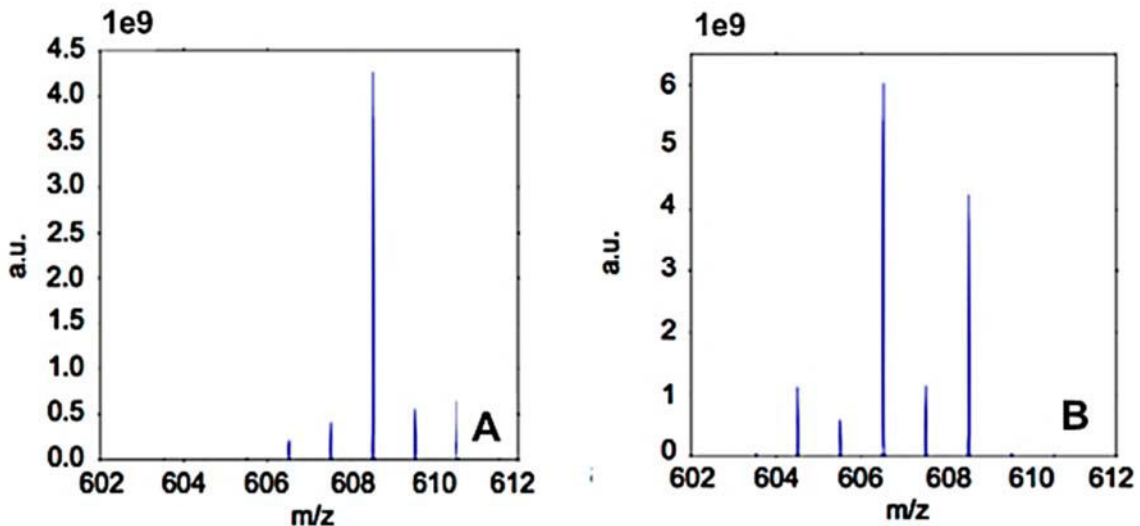


Figure 13 : Fragments des précurseurs à m/z 890,8 (a) et 888,8 (b) obtenues à partir des profils sur le spectre 2D NUS

En analysant le profil de fragmentation de l'ion à m/z 888,8, inférieur de 2 Da à la trioléine, on peut voir que le triglycéride possède une composition OOL ou OLO. Deux

fragments intenses à m/z 608,5 et m/z 606,5 correspondent à la perte de l'acide oléique O et l'acide linoléique L sont observés sur la Figure 13b.

Le même changement est observé pour le pic 4 Da inférieur à la trioléine qui correspond à un TAG composé de OLL ou LOL (Figure 14a). Pour le TAG à m/z 884,8 qui est à 6 Da en dessous de la trioléine, un seul fragment est observé à m/z 604,5 (Figure 14b). Ce TAG correspond à la trilinoléine composée de trois acides linoléiques L.

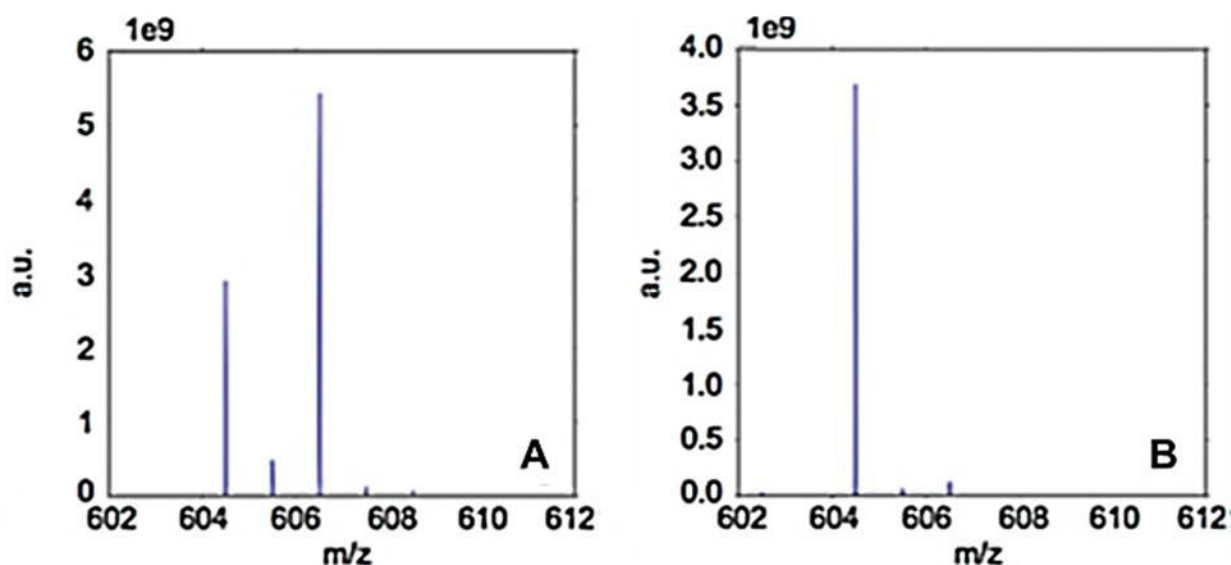


Figure 14 : Fragments des précurseurs à m/z 886,8 (g) et 884,8 (h) obtenues à partir des profils sur le spectre 2D NUS

IV. Conclusion

En conclusion, pour la première fois une acquisition 2D FTICR MS avec un échantillonnage non uniforme en utilisant un taux de d'échantillonnage jusqu'à $1/32$ a été mise en place. L'acquisition NUS ouvre la voie à la 2D-FT-ICR haute résolution pour les précurseurs et les fragments et dans l'analyse indépendante des données (DIA).

L'acquisition non uniforme en FT-ICR 2D haute résolution suit la transition de phase de Donoho-Tanner [17]. Cette transition de phase peut permettre de déterminer le rapport NUS pour une acquisition (graphique de transition de phase en annexe). Pour une acquisition physique de 2 k points, il est possible d'utiliser un rapport NUS de $1/16$ pour obtenir au final une acquisition de 32 k. Il est donc possible d'après la transition de

phase de Donoho-Tanner d'obtenir une taille d'acquisition de 1024 k points à partir d'une d'acquisition physique de 16 k points et un rapport NUS 1/64.

Cette méthode d'analyse pourrait permettre d'analyser des échantillons complexes comme les différentes familles de lipides : diacylglycérides, phosphoglycérides, sphingolipides ... mais aussi les métabolites présents dans des échantillons biologiques. Elle pourrait s'appliquer à des composés comme les tanins ou les phénols des différents vins.

Bibliographie

- (1) van Agthoven, M. A.; Delsuc, M. A.; Bodenhausen, G.; Rolando, C.: Towards analytically useful two-dimensional Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Bioanal. Chem.* **2013**, *405*, 51-61.
- (2) Marvasti, F.: *Nonuniform sampling: theory and practice*, 2012.
- (3) J.C.J Barna, E. D., M.RMayger, NJSkilling, S.J.PWorrall: Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *J. Magn. Reson.* **1987**, *73*, 69-77.
- (4) Gull, S. F.; Skilling, J.: The maximum entropy method. *Indirect Imaging, JA* **1984**.
- (5) Sibisi, S.; Skilling, J.; Brereton, R. G.; Laue, E. D.; Staunton, J.: Maximum entropy signal processing in practical NMR spectroscopy. *Nature* **1984**, *311*, 446-447.
- (6) Chiron, L.; van Agthoven, M. A.; Kieffer, B.; Rolando, C.; Delsuc, M. A.: Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry. *Proc Natl Acad Sci U S A* **2014**, *111*, 1385-90.
- (7) Aarstol, M.; Comisarow, M. B.: Apodization of FT-ICR spectra. *Int. J. Mass Spectrom. Ion Processes* **1987**, *76*, 287-297.
- (8) Ikura, M.; Kay, L. E.; Bax, A.: A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **1990**, *29*, 4659-4667.
- (9) Delsuc, M. A.: Spectral representation of 2D NMR spectra by hypercomplex numbers. *J. Magn. Reson.* **1988**, *77*, 119.
- (10) Hyberts, S. G.; Takeuchi, K.; Wagner, G.: Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **2010**, *132*, 2145.
- (11) van Agthoven, M. A.; Chiron, L.; Coutouly, M.-A.; Sehgal, A. A.; Pelupessy, P.; Delsuc, M.-A.; Rolando, C.: Optimization of the discrete pulse sequence for two-dimensional FT-ICR mass spectrometry using infrared multiphoton dissociation. *Int. J. Mass Spectrom.* **2014**, *370*, 114-124.
- (12) Jan Axelsson*, M. P., Kristina Hakansson, Per Hakansson: Electron Capture Dissociation of Substance P Using a Commercially Available Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 474-477.
- (13) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J.: Implementation of electron-transfer dissociation on a hybrid linear ion trap- orbitrap mass spectrometer. *Anal. Chem.* **2007**, *79*, 3525.
- (14) Bray, F.; Bouclon, J.; Chiron, L.; Witt, M.; Delsuc, M. A.; Rolando, C.: Non-Uniform Sampling (NUS) Acquisition of two-dimensional (2D) FT-ICR MS for increased mass resolution of MS/MS precursor ions. *Anal. Chem.* **2017**.
- (15) University, D.: IsoPatrn software. *Trace Analysis Research Centre* **2017**.

(16) Garnier, N.; Rolando, C.; Hotje, J. M.; Tokarski, C.: Analysis of archaeological triacylglycerols by high resolution nanoESI, FT-ICR MS and IRMPD MS/MS: application to 5th century BC–4th century AD oil lamps from Olbia (Ukraine). *Int. J. Mass Spectrom.* **2009**, *284*, 47.

(17) Donoho, D.; Tanner, J.: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc., A* **2009**, *367*, 4273.

Partie 4 :

Empreinte lipidique par analyse FT-
ICR 2D

I. Analyse des triacylglycérols archéologiques

Le matériel organique contenu dans des céramiques archéologiques permet de mettre en évidence des traces d'activités humaines. Ces céramiques sont des matériaux typologiquement bien connus (figure 1). L'identification des contenants des amphores peut être réalisée grâce à leur forme, aux inscriptions présentes sur le support et à l'analyse de la couche interne qui les protège (poix, cire).

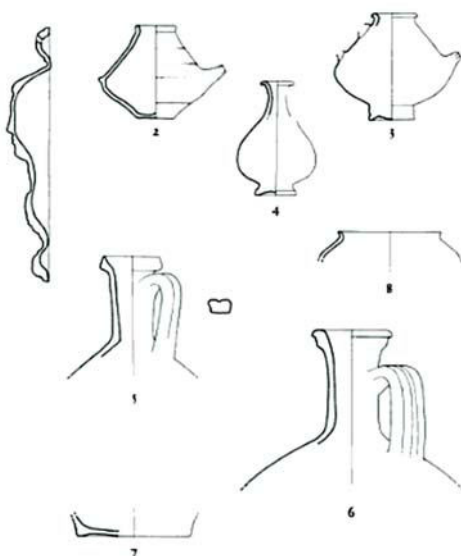


Figure 1 : Schémas typologiques de l'amphore biberon Zurzach 55 découvert sur le site de la nécropole de la cité antique de Tenedo du camp militaire établi à l'époque de l'empereur Claude (1^{er} siècle).

Cependant, la plupart des études chimiques sont consacrées à la composition de la céramique alors que peu de recherches sont consacrées à l'identification du contenu de la céramique [1].

Les matières grasses, notamment les huiles végétales ou les graisses animales ayant été présentes dans ces amphores, possèdent une constitution chimique simple et dominée par les triacylglycérols (TAGs). Ces derniers permettent d'identifier directement les acides gras conservés dans les tessons. L'identification de l'origine animale ou végétale est basée soit, sur la spectrométrie de masse isotopique par mesure du rapport isotopique $\delta^{13}\text{C}$ en chromatographie en phase gazeuse (GC-MS), et sur l'analyse des produits de dégradation oxygénés en utilisant la GC-MS [2].

En considérant plus particulièrement le contexte archéologique, il faut souligner que les TAGs sont sensibles à l'hydrolyse des fonctions ester ou à l'oxydation des doubles liaisons [3]. Les corps gras présents à l'intérieur de poteries ont été soumis à différents types de dégradation, à la fois induites par l'homme lors de la préparation du produit (chauffage, mélange avec d'autres substances, etc.), et faisant suite à l'enfouissement, dont les conditions (milieu humide ou sec, présence d'air ou non) jouent un rôle important sur les réactions (hydrolyse, oxydation). Toutes ces données ne sont pas connues de l'expérimentateur au moment de l'analyse, ce qui rend l'identification de ces corps gras délicate. De plus, les paramètres permettant une distinction aisée entre les graisses et les huiles fraîches ne sont plus valables en contexte archéologique.

Les huiles végétales étant riches en acides gras insaturés, elles s'oxydent rapidement et facilement. La disparition de composés insaturés au profit de composés saturés entraîne une modification de l'aspect de ces matériaux. Ainsi, les huiles végétales siccatives ou très oxydées peuvent devenir plus visqueuses, voire élastiques ou solides [4]. De manière générale, quelle que soit l'origine des corps gras retrouvés en contexte archéologique, ceux-ci sont solides, cireux et de couleur jaune/orange

Les analyses de corps gras archéologiques ont montré que les acides gras majoritaires qui les composent sont les acides palmitique et stéarique. Au début de l'archéologie moléculaire, le rapport acide palmitique/acide stéarique était utilisé pour déterminer l'origine biologique de ces corps gras. A l'heure actuelle, ce rapport n'est plus utilisé, car au cours de la dégradation des substances grasses, ce rapport varie. De plus, en présence de mélanges de corps gras, ces acides peuvent alors avoir plusieurs origines, dont ce rapport ne tient pas compte.

Les produits laitiers se distinguent des graisses animales par la présence d'acides gras à chaînes courtes. Cependant, comme précédemment mentionné, ces composés sont extrêmement sensibles au phénomène d'hydrolyse et ne sont jamais (ou très rarement) retrouvés en contexte archéologique. La perte de ces acides gras ne permet donc plus de faire la distinction entre un produit laitier et une graisse animale [5].

Malgré toutes ces difficultés, l'amélioration des outils analytiques, et une meilleure connaissance des corps gras ont permis de déterminer de nouveaux

biomarqueurs permettant de préciser une origine biologique. En effet, lorsque les conditions le permettent, les triglycérides (TAG) sont conservés, accompagnés ou non de produits d'hydrolyse comme les diglycérides (DAG) et/ou les monoglycérides (MAG). Ces composés témoignent de la présence initiale d'une graisse ou d'une huile. Cependant, les TAG ne subsistent pas toujours.

1. Analyse par chromatographie en phase gazeuse

Historiquement, dès 1976, la GC-MS a été utilisée pour identifier la présence de corps gras archéologiques, et tout particulièrement celle d'huile d'olive, à l'intérieur de tessons d'amphores servant à son transport [6,7]

Dans ce cas, l'analyse de la distribution des acides gras libres, des stérols et des produits d'oxydation des acides gras insaturés (diacides, hydroxy- et dihydroxyacides) par GC-MS, ainsi que l'étude des données isotopiques du $\delta^{13}\text{C}$ sur les acides gras par GC-C-IRMS permettent d'obtenir plus d'informations sur l'origine biologique du corps gras en question.

Depuis, de nombreuses études ont été réalisées sur les archéomatériaux, révélant l'utilisation de graisses animales et d'huiles végétales, ainsi que les dégradations qu'elles ont subies.

2. Analyse par chromatographie en phase liquide

L'analyse des TAGs par des méthodes basées sur la chromatographie liquide est complexe car leur labilité rend leur détection très difficile dans le matériel archéologique.

Quelques références exposent l'identification des graisses animales mélangées à de l'huile d'olive [7] et à la cire d'abeille [8] dans des échantillons archéologiques à l'aide de HPLC-APCI-MS. L'analyse des TAG par HPLC-APCI-MS ou HPLC-ESI-MS permet d'en déterminer la structure moléculaire et d'éventuellement d'en préciser l'origine biologique [9-11].

3. Analyse par injection directe

L'identification des TAGs archéologiques nécessite des protocoles d'extraction, de purification et d'analyse spécifiquement adaptés pour détecter des échantillons en état de dégradation avancé. Par conséquent, une procédure basée sur une étape préliminaire d'extraction des composants lipidiques peut être utilisée [12].

Une fois les TAGs purifiés, ils peuvent être analysés directement par spectrométrie de masse sans utiliser en amont une séparation. Cette méthode permet d'obtenir un profil des TAGs présent dans l'échantillon. La fragmentation par spectrométrie de masse, de ceux-ci, permet d'identifier les acides gras le composant (Comme expliqué dans le chapitre II, partie II, page 78-79). Garnier et al [12] a montré que la comparaison des profils MS de TAGs d'échantillon archéologique contre une base d'huile de référence peut permettre la caractérisation de ces échantillons.

II. Objectifs

La 2D FT-ICR est un outil pour l'analyse d'échantillon complexe comme nous l'avons vu dans les chapitres précédents. La possibilité d'améliorer la résolution dans la première dimension ouvre des portes pour l'analyse d'échantillons compliqués. L'analyse des TAGs par cette méthode rend possible l'obtention de leur masses mais aussi de leur structure en une seule acquisition.

C'est pour cela que la 2D-FT-ICR peut permettre la différenciation des TAGs venant d'espèces végétales ou animales différentes rapidement et, avec l'acquisition non uniforme, la caractérisation de ces triglycérides.

III. Matériels et méthodes

1. Echantillons

Les échantillons modernes constituant la base de données ont été choisis selon les espèces végétales cultivées et importées, et les animaux élevés dans les régions nord-méditerranéennes (tableau 1). Les graisses adipeuses ont été échantillonnées à partir de bœuf, de veau, d'agneau et de porc fraîchement abattus, à partir de tissus sous-cutanés et de tissus adipeux solides envahissant les reins. Les détails et les origines des huiles végétales, des produits laitiers et des graisses animales, sont détaillés dans le tableau 1.

	Espèces	Nom binominal	Provenance
Graisses animales	Boeuf, subcutaneous adipose tissue	<i>Bos taurus</i>	Auvergne, France
	Veau, subcutaneous adipose tissue	<i>Bos taurus</i>	Auvergne, France
	Agneau, subcutaneous adipose tissue	<i>Ovis aries</i>	Auvergne, France
	Porc, subcutaneous adipose tissue	<i>Sus scrofa</i>	Auvergne, France
	Canard, subcutaneous adipose tissue	<i>Anas platyrhynchos</i>	Auvergne, France
Huiles végétales	Huile d'avocat	<i>Persea americana</i>	Artisan Moulinier, Beaulieu, France
	Huile de soja	<i>Glycine max</i>	Artisan Moulinier, Beaulieu, France
	Huile de lin pressée et clarifiée	<i>Linum usitatissimum</i>	Kremer, Germany
	Huile de lin pressée	<i>Linum usitatissimum</i>	Kremer, Germany
	Huile de noix pressée	<i>Juglans regia</i>	Vic le Comte, France
	Huile d'olive pressée	<i>Olea europaea</i>	Nyons, France
	Huile de colza	<i>Brassica napus</i>	Artisan Moulinier, Beaulieu, France
	Huile de ricin	<i>Ricinus communis</i>	Artisan Moulinier, Beaulieu, France
	Huile de tournesol	<i>Helianthus annuus</i>	Vic le Comte, France
	Huile de sésame pressée	<i>Sesamum indicum</i>	Artisan Moulinier, Beaulieu, France
	Huile d'argan		Artisan Moulinier, Beaulieu, France
Produits laitiers	Lait de vache	<i>Bos taurus</i>	Auvergne, France
	Fromage au lait de vache (Saint-Félicien)	<i>Bos taurus</i>	Vic le Comte, France
	Fromage au lait de vache (Cantal)	<i>Bos taurus</i>	Vic le Comte, France
	Fromage au lait de brebis	<i>Capra hircus</i>	Vic le Comte, France
	Lait de chamelle	<i>Camelus dromedarius</i>	Petra, Jordan

Tableau 1 : Liste des échantillons analysés

L'échantillon archéologique provient d'un aryballe en céramique qui est un vase grec antique. Les aryballes sont très fréquemment utilisés comme réservoir d'huiles.

Cette échantillon provient du site de la nécropole de la cité antique de Tenedo du camp militaire établi à l'époque de l'empereur Claude.

2. Extraction lipides et purification des TAG

Le protocole d'extraction a été adapté à partir d'une méthodologie décrite par Folch et al. [13] Le tissu adipeux ou le produit laitier (~ 1 g) a été homogénéisé dans un mélange de dichlorométhane / méthanol 1: 1, v/v (20 ml) et placé dans un bain à ultrasons pendant 20 min. Le mélange a été filtré sur un filtre de diamètre 0,22 µm et une solution de chlorure de potassium (0,12 M) a été ajoutée au filtrat. Après extraction et séparation des deux phases par centrifugation, la phase organique a été séchée sur MgSO₄, évaporée à sec sous un courant d'azote à 40° C. Une aliquote de 2 mg du résidu a été dissoute dans du dichlorométhane (1 ml). Les TAG sont purifiés par fractionnement sur une cartouche Sep-Pak diol avant rinçage avec du cyclohexane (4 ml) [31,32]. L'extrait lipidique (100 µL) a été déposé et les TAG sont élués 3 fois avec 2 ml d'un mélange de cyclohexane / dichlorométhane / éthanol (89: 10: 1, v / v / v). La fraction contenant les TAGs a été évaporée à sec, le résidu a été dissous dans 100 µl de dichlorométhane / méthanol (1: 1, v / v)

Les huiles végétales liquides ont été diluées deux fois avec du dichlorométhane. Les TAGs ont été purifiés par fractionnement sur une cartouche Sep-Pak diol comme détaillé pour les tissus adipeux modernes et les produits laitiers

Un fragment d'environ 1,0 g de la céramique archéologique (TOUR 304) a été écrasé à l'aide d'un broyeur électrique (IKA A-11 basic) pendant 10 s. Les TAGs présent dans la poudre ont été extraits deux fois avec 5 ml de dichlorométhane / méthanol (2: 1, v / v), les phases organiques ont été séparées à chaque fois par centrifugation (10 min à 2000 g, puis 20 minutes à 4000 g). La solution limpide a été évaporée à sec sous un courant d'azote à 40 ° C, puis les TAG ont été purifiés selon la procédure décrite précédemment pour les tissus adipeux utilisant l'extraction SPE sur une cartouche diol.

3. Analyse 1D FT-ICR des TAGs

Les TAGs sont dilués dans 100 µL de chloroforme puis 10 µL de cette solution a été dilué dans 489 µL d'une solution méthanol : chloroforme 5 :1 v/v additionnée de 1 µL de chlorure de lithium 6 (10 mg/mL dans le méthanol). Toutes les expériences ont été

effectuées sur un spectromètre de masse FT-ICR 9,4 T nanoESI-hQh-ApexQE (Bruker Daltonics, Bremen, Allemagne) en mode ionisation positive par électronébulisation. L'échantillon a été injecté à un débit de 10 $\mu\text{L/h}$ à l'aide d'un pousse seringue (Cole-Parmer®, US) et une seringue de 250 μL (Gastight, Hamilton). Le voltage appliqué au niveau de la source nano ESI est de 1550 V, et 100 scans ont été accumulés pour chaque spectre. Le nombre de point d'acquisition est 1M point, avec une gamme de masse de m/z 144,38 à 2500. La fragmentation a été réalisée par l'IRMPD avec un laser CO_2 . La puissance du laser a été réglée à 12,5 W et la longueur d'onde de 10,4 μm . Les ions ont été irradiés pendant 0,4 s.

4. Analyse 2D FT-ICR des TAGs

La séquence d'impulsions pour MS 2D FT-ICR est représentée dans le schéma 1. Toutes les impulsions ont été construites à l'aide du générateur d'impulsions ApexControl (Bruker Daltonics, Brême, Allemagne). Les impulsions de codage (P_1 , P_2) ont une amplitude de 70 Vpp et une durée de 0,5 μs par fréquence. L'incrément de temps en t_1 est de 0,5 μs . L'impulsion d'observation (P_3) a une amplitude de 100 Vpp et une durée de 20 μs par fréquence. La gamme de masse horizontale est m/z 144,39-3000 (1 MHz à 48 kHz). La fragmentation a été réalisée avec l'IRMPD avec un laser CO_2 de longueur d'onde de 10,4 μm . La puissance du laser a été réglée à 12,5 W et les ions ont été irradiés pendant 0,4 s. Une première acquisition a été réalisée avec 2048 points dans la première dimension puis avec une résolution de 512k dans la deuxième dimension. Une seconde analyse a été réalisée en utilisant une acquisition non uniforme de 16k x 512k avec un ratio NUS de 1/16.

5. Analyse informatique

Les spectres MS ont été traités et visualisés à l'aide du logiciel Data Analysis (Bruker Daltonics, Bremen, Allemagne). Les traitements informatiques des expériences 2k et 16k avec des listes de temps t_1 pour des acquisitions classiques ou NUS ont été réalisées à partir d'un programme écrit sous Python (SPIKE adapté en langage python 64 bits Enthought Python Distribution, Wavre, Belgique).

IV. Analyse FT-ICR MS et MS/MS

1. Profils MS des échantillons

Les profils MS des TAGs purifiés provenant d'animaux, de produits laitiers et les graisses végétales dilués dans une solution de chlorure de lithium 6 isotopiquement pur, ont permis la création d'une base de données de TAG dépendante de l'espèce. Les spectres MS permettent de distinguer les composants en fonction de leur nombre d'équivalent carbone (ECN) et de leur degré d'insaturation les profils MS/MS, permettent l'élucidation de la structure des molécules. Il faut souligner que l'analyseur FT-ICR permet une haute précision de masse fournissant des données plus faciles interprétation et permettant l'identification des modifications des TAGs comme par exemple les changements oxydatifs [14,15].

Considérant plus précisément les huiles végétales (figure 2), on constate que la plupart des références modernes présentent une distribution étroite similaire de TAGs avec des ECN 50 à ECN 54 avec deux ou trois motifs, chacun séparé par 26-28 amu.

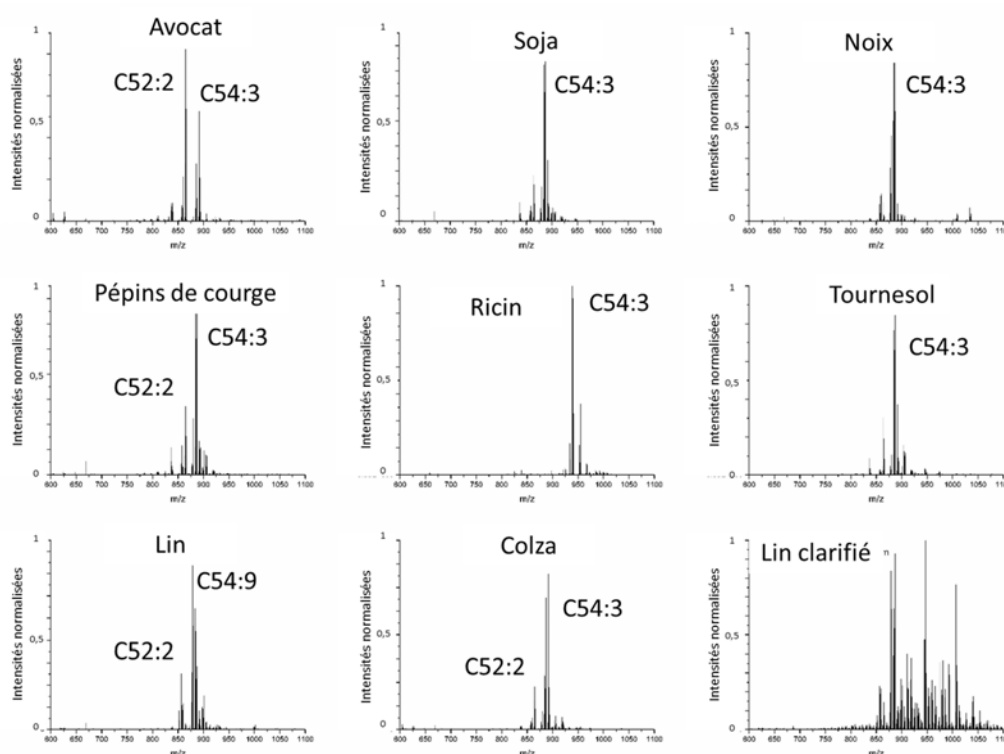


Figure 2 : Spectres MS obtenues en nanoESI Qh FT-ICR 9,4T en mode positif des TAGs des huiles végétales avec du chlorure de lithium (^6Li).

Cette distribution des TAGs est une conséquence directe de la composition des acides gras constitutifs dans les huiles, et de la prédominance d'un acide gras dans la plupart d'entre eux, à savoir l'acide oléique [16]. Cependant, chaque espèce d'huile végétale présente un modèle singulier qui permet de les discriminer et de les authentifier. Par exemple, le spectre de l'huile de soja (figure 2) est principalement dominé par des TAGs avec des ECN 54. Le pic majoritaire est le C54:3 qui peut être identifiés comme la trioléine OOO (O : acide oléique). Un autre exemple, l'huile de ricin, qui donne des motifs très distincts car il est composé d'acide ricinoléique qui possède un hydroxyle en position 12.

Le profil de masse des graisses animales (figure 3) couvre une large gamme de masse correspondant aux TAG avec ECN de 46 à 54.

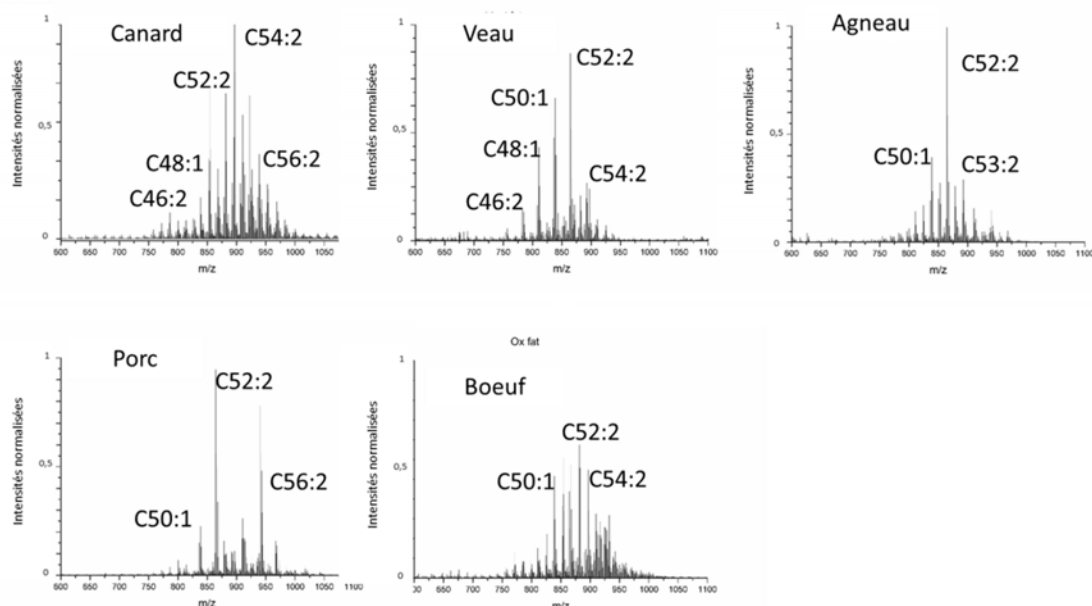


Figure 3 : Spectres MS obtenues en nanoESI Qh FT-ICR 9,4T en mode positif des TAGs des graisses animales avec du chlorure de lithium (^6Li).

Le marqueur principal est le TAG 52: 2 pour bovins. Les profils bovins et ovins présentent l'alternance d'un TAG à nombre pair abondant et de TAG de nombre impair mineur.

Les spectres des produits laitiers (figure 4), montrent qu'ils sont riches en acides gras palmitique (abréviation P), oléique (abréviation O), stéarique (abréviation S) et d'autres acides gras saturés à chaîne courte en moindre quantités. Cette diversité induit

une grande distribution des TAG de la ECN 24 à 54. Le TAG 36:0 est le TAG majoritaire pour la vache, la chèvre et les fromages de lait cru (figure 4).

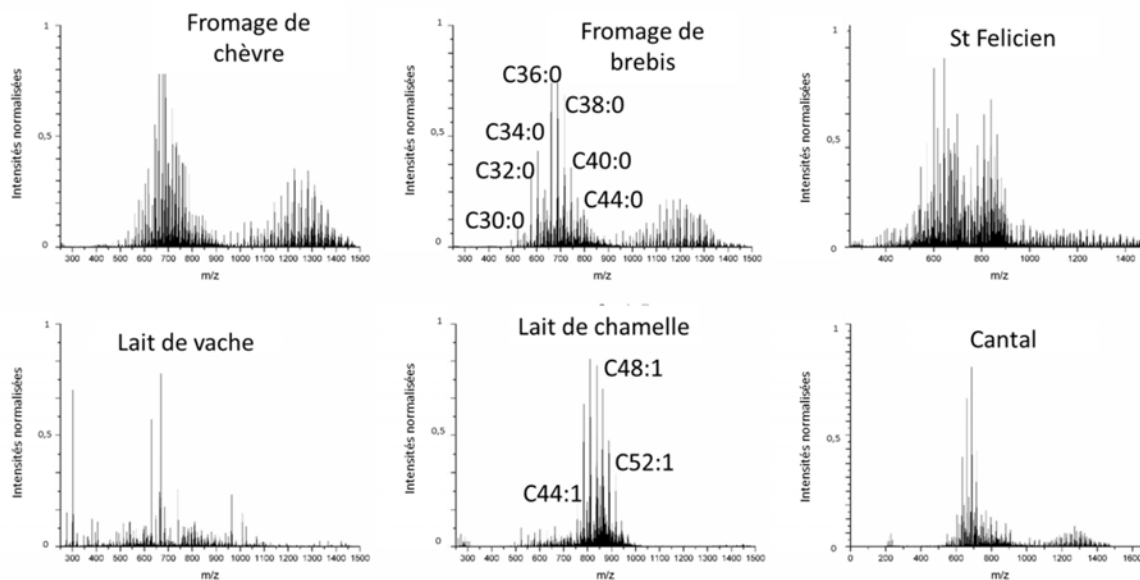


Figure 4 : Spectres MS obtenues en nanoESI Qh FT-ICR 9,4T en mode positif des TAGs des produits laitiers avec cationisé par du chlorure de lithium (^6Li).

Comme observé pour les produits laitiers, les acides gras avec un nombre impairs sont des marqueurs d'espèces de ruminants (bovin, vache, veau, agneau), principalement synthétisés par des bactéries présentes dans le rumen [17]. À l'exception de la distinction des TAGs présents dans les volailles et le porc, de ceux présents dans les graisses bovines et ovines, aucune autre authentification ne peut être suggérée par la seule analyse de MS et une étude MS/MS plus complète doit être étudiée pour obtenir des données pertinentes.

Globalement, on peut souligner que, par rapport aux méthodes utilisant d'autres sources d'ionisation, l'analyse des TAG par la spectrométrie de masse avec la nano électronébulisation conduit à des spectres simples et permet la détection des TAGs insaturés et saturés sans fragmentation.

Leurs profils de masse sont trop semblables pour permettre une distinction directe entre les produits laitiers bovins, ovins et caprins, et une méthodologie spécifique doit être utilisée pour discriminer l'origine de l'espèce [10].

2. Identification structurale des TAGs par fragmentation IRMPD

Pour élucider la structure des TAGs une étude par fragmentation a été réalisée. Lors de la fragmentation IRMPD du trioléate TAG homogène (OOO) à m/z 864,7978 (erreur de 0,33 ppm, figure 5), deux ions sont issu de la perte des acide gras libre de l'ion parent $[M+Li]^+$, à savoir les ions $[M+Li-RCH_2CO_2H]^+$ à m/z 608,5325 correspondant à la perte de l'acide oléique. L'identité des substituants d'acides gras dans les espèces TAGs peut également être déduite en observant le fragment $[RCH_2CO_2H+Li]^+$ dans la gamme de masse à m/z 300.

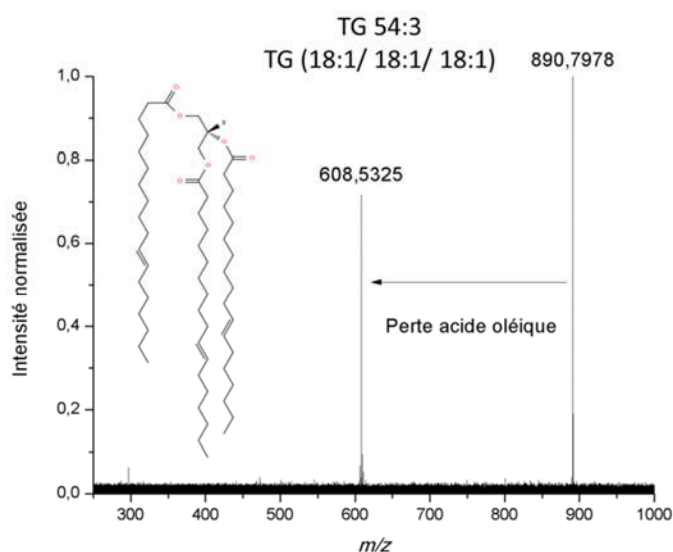


Figure 5 : Spectres de fragmentation IRMPD du TAG à m/z 890,7978. Les fragments des acides gras libres $[RCO_2+Li]^+$ sont observés dans la zone m/z 250-350. Les fragments de type $[M+Li-R_2CO_2H]^+$ correspondant à la perte d'un acide gras sont détectés dans la gamme de masse m/z 580-610.

Cette étude peut être réalisée sur toutes les masses des échantillons qu'il soit de provenance végétale, animale ou laitière. Cette opération est longue et fastidieuse, c'est pour cela que la 2D-FTICR peut être d'une grande aide.

3. Analyse d'échantillon archéologique

L'échantillon TOUR 304 présentait un dépôt organique visible sur sa face interne, la seule information chimique pouvait être obtenue à partir de la matière organique piégée et présente comme imprégnation de la matrice poreuse d'argile de la céramique.

Les fractions de TAGs purifiées ont été analysées dans les mêmes conditions que celles utilisées pour les graisses et huiles modernes. La figure 6 montre le profil MS des TAG natifs de l'échantillon TOUR 304.

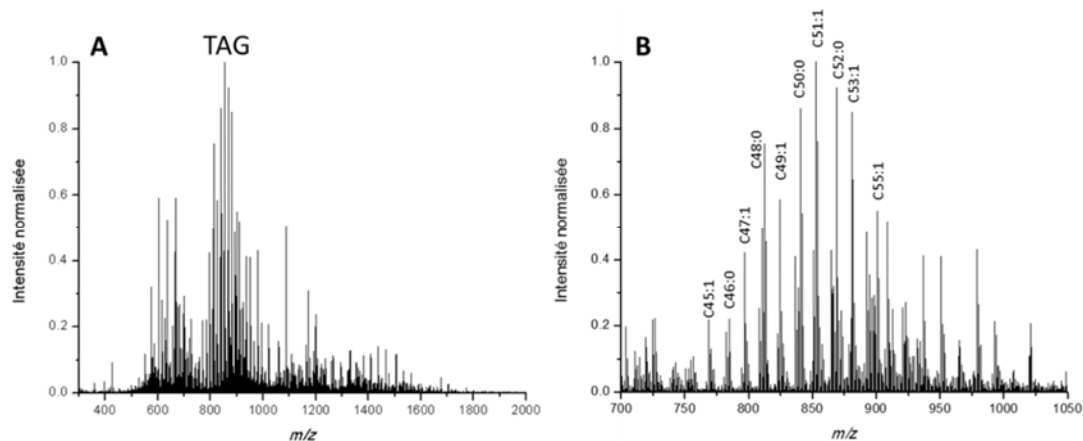


Figure 6 : Profil MS des TAGs présent dans l'échantillon TOUR 304 (A). Ionisation en mode positive en nanoESI sur le FT-ICR 9,4 T. l'ionisation est réalisé en présence de lithium. En B, zoom sur la région contenant les TAGs avec l'annotation des pics majoritaire.

Le profil est caractérisé par une large distribution de TAGs correspondant à des triacylglycérols avec ECN de 45 à 57. Compte tenu du peu de matière organique et de l'état de dégradation avancé de cet échantillon, on observe un très bon rapport signal sur bruit du spectre MS est observé.

Les intensités de pic les plus élevées sont des TAGs à nombre pair de carbone comme le TAG 48:0, 50:0, 52:0. Des TAGs à nombre impair de carbones sont détectés comme le TAG 47:1 et le TAG 53:1. Si les conditions environnementales ont garanti la conservation de certains TAG natifs de la dégradation par hydrolyse, le spectre révèle une forte dominance de TAGs saturés et une faible proportion de TAG monoinsaturés.

En effet, ces profils caractérisés par une forte proportion de TAG saturés peuvent s'expliquer par la forte sensibilité des acides gras insaturés aux processus d'oxydation. En conséquence, ces composés sont plus polaires et sont facilement oxydés par l'eau [18].

Les profils MS/MS de TAGs impairs révèlent la présence de groupements d'acides gras de nombres impairs pour les composés mineurs, par exemple le

groupements acide heptadécanoïque, caractéristiques des bactéries du rumen [17] (figure 7).

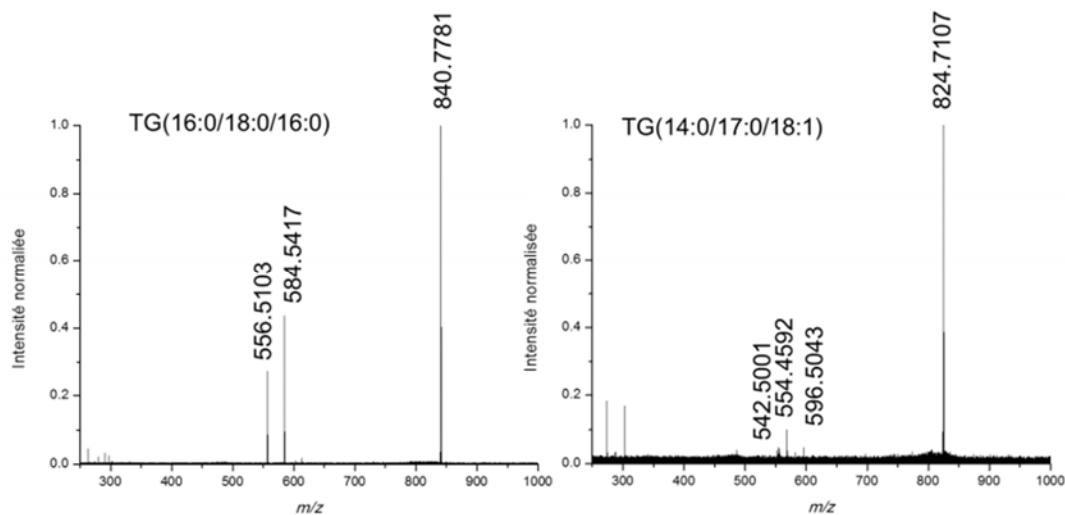


Figure 7 : Profil MS/MS à gauche du TAG 52 :0 (16:0/18:0/16:0) et à droite du TAG 51 :1 (14:0/17:0/18:1).

La comparaison des profils MS et MS/MS des graisses de référence extraites des tissus adipeux montre que seules les espèces bovines ou ovines présentent de telles séries impaires de carbone, par exemple des TAG contenant un fragment d'acide gras à nombre impair de carbone. L'identification de séries impaires de carbone dans les extraits archéologiques permet de conclure à la présence de graisses bovines ou ovines dans l'échantillon de TOUR 304.

V. Analyse 2D FT-ICR MS

1. Empreinte lipidique

Tout d'abord, une analyse 2D-FT-ICR de 2k x 512k a été réalisée. Le temps d'analyse est de 60 min et le temps de traitement informatique est de 30 min. Elle permet de cribler les échantillons de TAGs et de différencier les échantillons animaux, végétaux ou laitiers plus rapidement que la méthode précédente.

A partir des cartes 2D, il est facile de différencier les TAGs d'agneau ou de colza (figure 8). Les ions fragments sont spécifiques des espèces. La ligne des ions précurseurs et fragments est visualisable sur les spectres 2D.

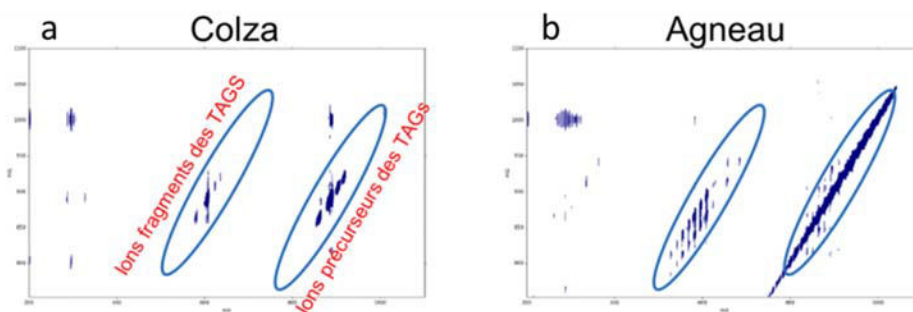


Figure 8 : Carte 2D des TAGs des huiles de colza (a) et Agneau (b)

La comparaison des échantillons se fait sur la base des ions précurseurs mais surtout sur la ligne des ions fragments ou perte de neutre (figure 9). En examinant la région des fragments, les acides gras libérés permettent de faire la différence entre l'huile de colza et d'avocat (figure 9a, b) ou la graisse de veau ou de canard (figure 9c, d).

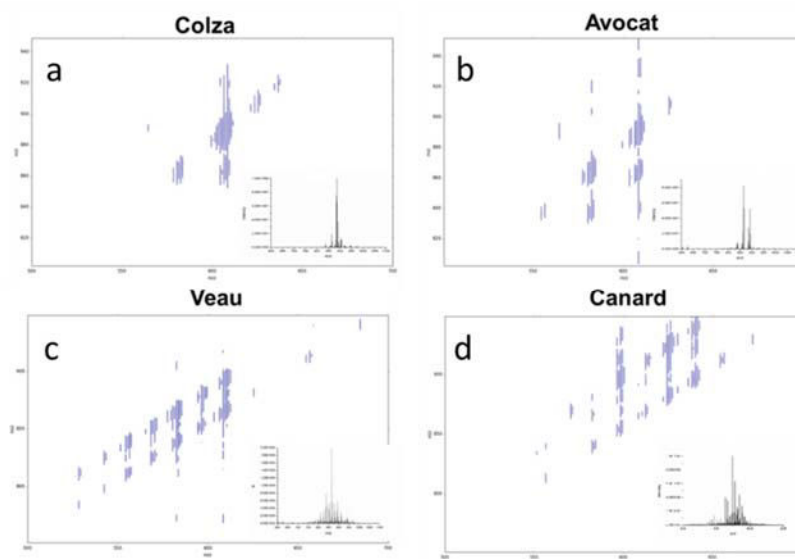


Figure 9 : Zoom sur la carte 2D des régions des fragmentations IRMPD des TAGs des huiles de colza (a), Avocat (b), Veau (c) et Canard (d). En bas à droite de chaque zoom, le spectres MS de l'échantillon.

La figure 10 montre les spectres 2D-FTICR des échantillons d'huiles végétales et animales. Il est clairement visible que de nombreuses différences sont présentes entre les échantillons aussi bien entre espèces qu'entre familles, mais certains restent tout de même difficile à différencier comme ceux de l'huile de colza et de tournesol (figure 10a) ou ceux de la graisse de bœuf et de veau figure (figure 10b).

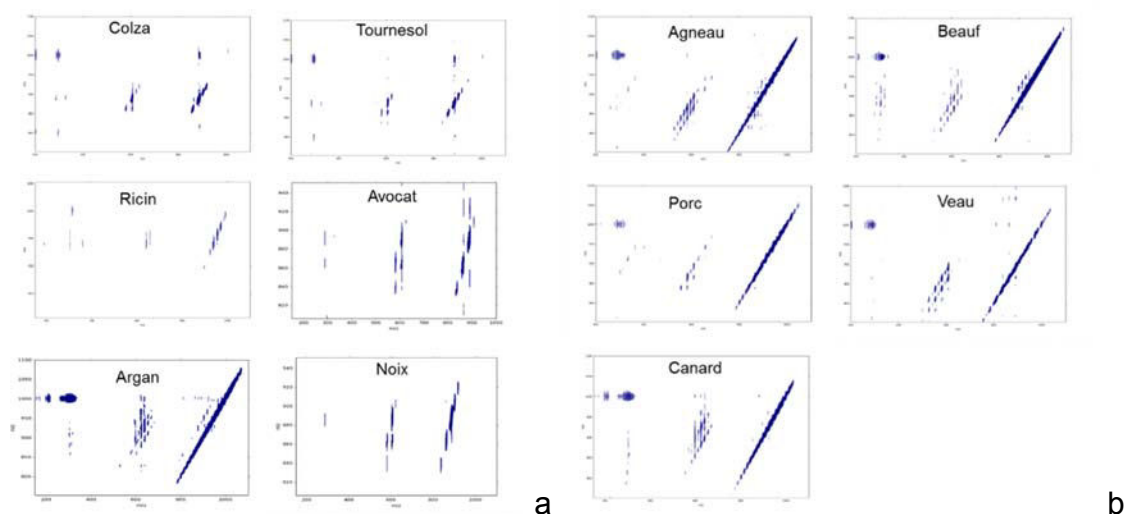


Figure 10 : Cartes 2D des échantillons d'huiles végétales (a) et de graisse animale (b).

Cette première expérience de 2D FT-ICR rapide, a permis de discriminer les TAGs de différentes espèces animales et végétales modernes mais la faible résolution en première dimension ne permet d'identifier les TAGs. Pour cela une expérience en haute résolution doit être mise en place.

2. Analyse 2D de l'échantillon archéologique

L'échantillon TOUR 304 a été analysé à l'aide d'une expérience de 2D FT-ICR 2k × 512 k comme pour les échantillons moderne. La figure 11 montre la carte 2D obtenue pour l'échantillon archéologique. La ligne des précurseurs et fragments sont très clairement visibles.

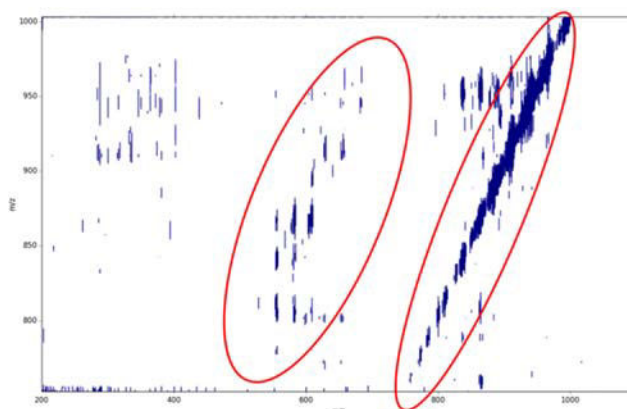


Figure 11 : Carte 2D FT-ICR de l'échantillon TOUR 304. En rouge sont entouré la ligne des précurseurs et fragments.

Cette expérience permet de voir que la carte 2D de l'échantillon archéologique ressemble aux cartes 2D des échantillons de graisses animales.

La figure 12 montre le zoom de la région des précurseurs (a) et des fragments (b). La région des fragments montre deux lignes de fragments qui correspondent à la perte d'acides gras de chaînes de taille différente.

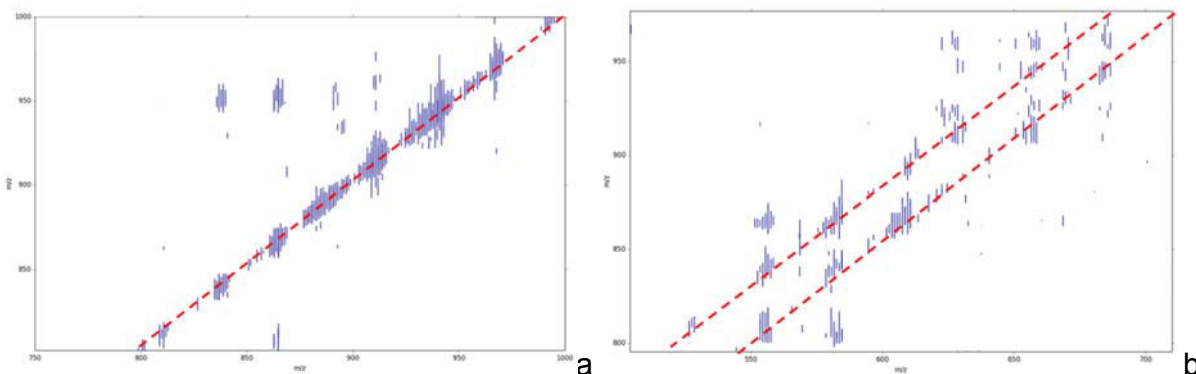


Figure 12 : Zoom de la région des précurseurs (a) et des fragments (a) de l'échantillon TOUR 304.

L'analyse par 2D FT-ICR de taille 2k points permet d'obtenir un résultat rapide. Elle permet de comparer un échantillon archéologique à une base de données d'huile ou graisse moderne pour déterminer l'origine des TAGs présents.

Mais cette analyse ne possède pas la résolution suffisante pour obtenir les informations de structure des TAGs présent. Pour réaliser ceci, une 2D FT-ICR haute résolution doit être réalisé pour obtenir ces informations.

3. Analyse haute résolution : identification structurale des TAGs

Pour permettre une identification plus précise, une analyse 2D-FTICR haute résolution a été réalisée. Cette méthode a été expliquée dans le chapitre II, partie 3 de ce manuscrit. Pour rappel, cette technique consiste à réaliser une acquisition 2D-FTICR avec un échantillonnage non uniforme dans la première dimension. Cette analyse a été faite avec une taille d'acquisition physique de 16 384 points dans la première dimension en utilisant un rapport NUS de 1/16 qui permet après traitement d'obtenir une analyse de 262 144 points et une résolution dans la seconde dimension de 512 k points. Cette acquisition permet d'avoir une résolution unitaire dans la première dimension et donc

d'obtenir des spectres de fragmentation pour chaque TAGs présents dans les échantillons.

Cette analyse a été menée sur l'échantillon de porc. La durée de l'expérience est de 16 heures et le traitement informatique de 24 heures. La figure 13 montre la carte 2D de l'échantillon de porc.

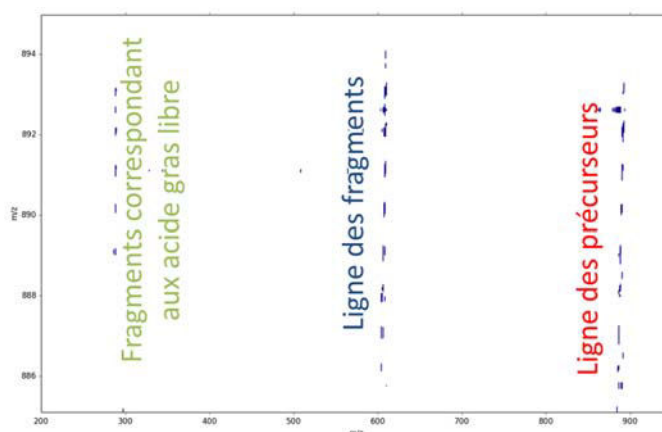


Figure 13 : Carte 2D FT-ICR NUS 1/16 de graisse de porc. Les lignes des précurseurs et fragments sont observés en rouge et bleu. En vert, il s'agit des fragments correspondants à la perte des acides gras.

Cette carte 2D montre les lignes des précurseurs $[M+Li]^+$ et fragments $[M+Li-CH_2CO_2H]^+$, mais aussi les fragments qui correspondent à la perte des acides gras $[RCH_2CO_2H+Li]^+$. Ces pertes d'acides gras présents dans la zone de masse de m/z 300 permettent de déterminer les acides gras présents dans les TAGs [12].

En réalisant un zoom sur la gamme de masse m/z 885 – 895 dans la dimension F1 (figure 14), il est possible de voir que les isotopes des TAGs ne se chevauchent pas. La résolution en F1 est d'environ 8 500. Sur la figure 14 b, trois lignes de fragmentation sont observées pour les TAG présent dans cette gamme de masse. Ces trois lignes correspondent à la perte d'acide gras avec des nombres d'insaturations différentes.

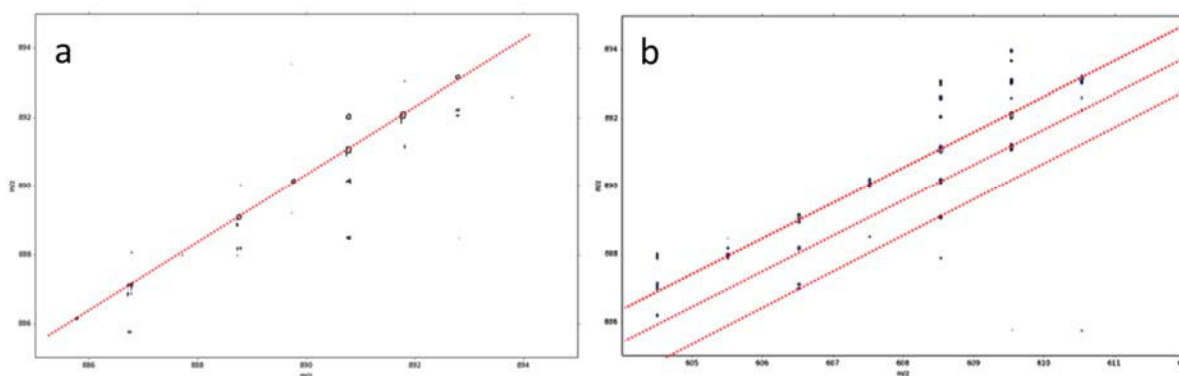


Figure 14 : Zoom de la région des précurseurs (a) et des fragments (b) de l'échantillon de porc sur la gamme de masse m/z 885 – 895. Trois lignes de fragmentation sont observées sur la figure b.

L'analyse en 2D FT-ICR haute résolution avec un rapport NUS 1/16 pour une acquisition de 16k points permet d'obtenir des profils de fragmentation propre et spécifique de l'ion sélectionnée (Figure 15).

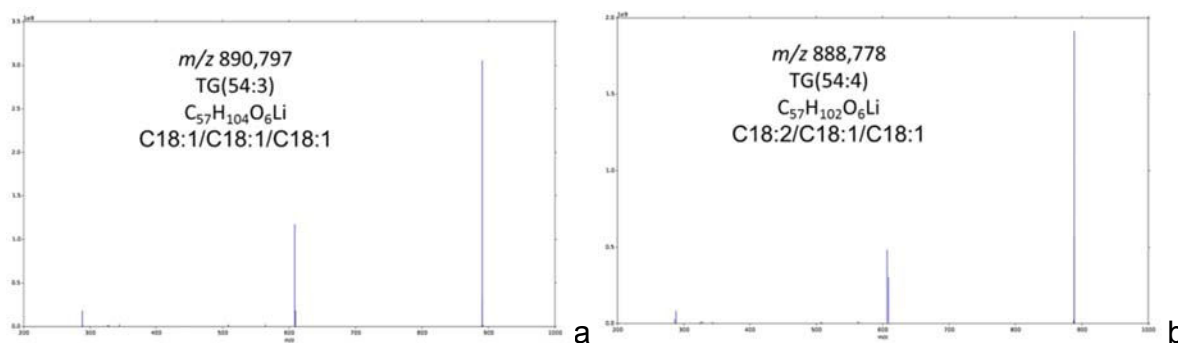


Figure 15 : Profil MS/MS du TAG 54:3, 18:1/18:1/18:1 à m/z 890,797 (a et (b) TAG 54:4, 18:2/18:1/18:1 à m/z 888,778.

Les spectres MS/MS des ions m/z 890,797 et m/z 888,778 permettent d'identifier le TAG. Le spectre MS/MS figure 15 a, correspond à la trioléine composé de 3 acides oléiques et le spectre MS/MS figure 15 b montre une composition de 2 acides oléiques et 1 acide linoléique. Malheureusement le spectre MS/MS de l'ion à m/z 888,778 ne permet de déterminer la position exacte des trois acides gras. Pour obtenir cette information, des expériences de MS3 seraient nécessaire comme la dissociation par les électrons (EID) [19], l'excitation par impact électronique des ions à partir des composés organiques (EIEIO) [20], la dissociation induite par l'ozone (OzID) [21] ou la dissociation ultra-violet (UVPD) [22].

L'utilisation de carte 2D permet une comparaison des échantillons rapidement. Il est envisageable de créer un score de ressemblance pour permettre une identification. La superposition de deux spectres et leur multiplication permettrait de voir les similitudes et les différences.

VI. Conclusion

La FT-ICR bidimensionnelle peut être utilisée comme une méthode rapide pour observer les profils des TAGs dans des échantillons. Cette méthode permet de détecter les TAGs et les fragments en une seule expérience et les représenter sur une carte 2D. Le temps d'analyse est réduit comparé à une analyse classique 1D FT-ICR où chaque masse doit être sélectionnée puis fragmentée.

Une base de données peut être créée pour comparer des échantillons inconnus et pour identifier la famille des TAGs présents (végétale, animale, produit laitier). La carte 2D permet d'avoir une visualisation complète de l'échantillon et donc de comparer rapidement les échantillons. L'échantillon archéologique analysé en 2D FT puis comparé à notre base de données a permis de déterminer que les TAGs présents sont d'origine animale.

Les analyses 2D non uniformes ont montré qu'il est possible d'obtenir une résolution suffisante pour séparer les isotopes des TAGs et d'obtenir des spectres MS/MS spécifiques. Bien que cette analyse reste longue, la qualité des données obtenues permet une identification structurale des TAGs.

A l'heure actuelle, l'échantillon archéologique n'a pas été analysé en 2D FT haute résolution car la technique est encore en phase d'amélioration mais il est possible que cette analyse permettra de corréler aux données obtenues avec les analyses 1D FT-ICR.

La mise en place de fragmentation MS3 permettrait d'avoir des informations de localisations des acides gras sur le squelette du TAG, mais aussi la position des insaturations des acides gras.

L'utilisation de la 2D FT-ICR dans l'analyse des lipides archéologiques peut permettre d'apporter des informations rapides aux archéologues sur leurs échantillons.

Bibliographie

- (1) Passi, S.; Rothschild-Boros, M. C.; Fasella, P.; Nazzaro-Porro, M.; Whitehouse, D.: An application of high performance liquid chromatography to analysis of lipids in archaeological samples. *J. Lipid Res.* **1981**, *22*, 778-784.
- (2) Copley, M.; Bland, H. A.; Rose, P.; Horton, M.; Evershed, R.: Gas chromatographic, mass spectrometric and stable carbon isotopic investigations of organic residues of plant oils and animal fats employed as illuminants in archaeological lamps from Egypt. *Analyst* **2005**, *130*, 860-871.
- (3) Stratakis, M.; Orfanopoulos, M.: Regioselectivity in the ene reaction of singlet oxygen with alkenes. *Tetrahedron* **2000**, *56*, 1595-1615.
- (4) Serpico, M.; White, R.: Oil, fat and wax. *Ancient Egyptian materials and technology* **2000**, 390-429.
- (5) Warinner, C.; Hendy, J.; Speller, C.; Cappellini, E.; Fischer, R.; Trachsel, C.; Arneborg, J.; Lynnerup, N.; Craig, O. E.; Swallow, D. M.: Direct evidence of milk consumption from ancient human dental calculus. *Scientific reports* **2014**, *4*.
- (6) Condamin, J.; Formenti, F.; Metais, M. O.; Michel, M.; Blond, P.: The application of gas chromatography to the tracing of oil in ancient amphorae. *Archaeometry* **1976**, *18*, 195-201.
- (7) Kimpe, K.; Jacobs, P.; Waelkens, M.: Analysis of oil used in late Roman oil lamps with different mass spectrometric techniques revealed the presence of predominantly olive oil together with traces of animal fat. *J. Chromatogr. A* **2001**, *937*, 87-95.
- (8) Kimpe, K.; Jacobs, P.; Waelkens, M.: Mass spectrometric methods prove the use of beeswax and ruminant fat in late Roman cooking pots. *J. Chromatogr. A* **2002**, *968*, 151-160.
- (9) Evershed, R. P.: Organic residue analysis in archaeology: the archaeological biomarker revolution. *Archaeometry* **2008**, *50*, 895-924.
- (10) Mirabaud, S.; Rolando, C.; Regert, M.: Molecular criteria for discriminating adipose fat and milk from different species by nanoESI MS and MS/MS of their triacylglycerols: application to archaeological remains. *Anal. Chem.* **2007**, *79*, 6182-6192.
- (11) Tuñón-López, J. A.; Beneito-Cambra, M.; Robles-Molina, J.; Parras-Guijarro, D. J.; Molina-Díaz, A.; Sánchez-Vizcaíno, A.; García-Reyes, J. F.: Multiclass profiling of lipids of archaeological interest by ultra-high pressure liquid chromatography-atmospheric pressure chemical ionization-high resolution mass spectrometry. *Microchem. J.* **2017**, *132*, 49-58.
- (12) Garnier, N.; Rolando, C.; Høtje, J. M.; Tokarski, C.: Analysis of archaeological triacylglycerols by high resolution nanoESI, FT-ICR MS and IRMPD MS/MS: application to 5th century BC–4th century AD oil lamps from Olbia (Ukraine). *Int. J. Mass Spectrom.* **2009**, *284*, 47-56.
- (13) Nie, S.; Pham, H. T.; Blanksby, S. J.; Reid, G. E.: Photoinduced intermolecular cross-linking of gas phase triacylglycerol lipid ions. *Eur. J. Mass Spectrom.* **2015**, *21*, 287-296.

- (14) van den Brink, O. F.; Boon, J. J.; O'Connor, P. B.; Duursma, M. C.; Heeren, R.: Matrix-assisted laser desorption/ionization Fourier transform mass spectrometric analysis of oxygenated triglycerides and phosphatidylcholines in egg tempera paint dosimeters used for environmental monitoring of museum display conditions. *J. Mass Spectrom.* **2001**, *36*, 479-492.
- (15) Wu, Z.; Rodgers, R. P.; Marshall, A. G.: Characterization of vegetable oils: detailed compositional fingerprints derived from electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J. Agric. Food Chem.* **2004**, *52*, 5322-5328.
- (16) Belitz, H.-D.; Grosch, W.; Schieberle, P.: Food Chemistry. 4. vyd. Berlin: Springer-Verlag, 2009.
- (17) Folch, J.; Lees, M.; Sloane-Stanley, G.: A simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem.* **1957**, *226*, 497-509.
- (18) Vlaeminck, B.; Fievez, V.; Cabrita, A.; Fonseca, A.; Dewhurst, R.: Factors affecting odd- and branched-chain fatty acids in milk: A review. *Anim. Feed Sci. Technol.* **2006**, *131*, 389-417.
- (19) Thomas, M. C.; Mitchell, T. W.; Harman, D. G.; Deeley, J. M.; Nealon, J. R.; Blanksby, S. J.: Ozone-induced dissociation: Elucidation of double bond position within mass-selected lipid ions. *Anal. Chem.* **2008**, *80*, 303-311.
- (20) Jones, J. W.; Thompson, C. J.; Carter, C. L.; Kane, M. A.: Electron-induced dissociation (EID) for structure characterization of glycerophosphatidylcholine: determination of double-bond positions and localization of acyl chains. *J. Mass Spectrom.* **2015**, *50*, 1327-1339.
- (21) Regert, M.; Bland, H. A.; Dudd, S. N.; Bergen, P. v.; Evershed, R. P.: Free and bound fatty acid oxidation products in archaeological ceramic vessels. *Proceedings of the Royal Society of London B: Biological Sciences* **1998**, *265*, 2027-2032.
- (22) Baba, T.; Campbell, J. L.; Le Blanc, J. Y.; Baker, P. R.: Structural identification of triacylglycerol isomers using electron impact excitation of ions from organics (EIEIO). *J. Lipid Res.* **2016**, *57*, 2015-2027.

CHAPITRE III

L'analyse protéomique
d'échantillons archéologiques et
paléontologiques

Partie 1 :

Généralités sur l'analyse des échantillons archéologiques et paléontologiques

I. Introduction

L'objectif principal de l'archéologie est l'étude des activités humaines dans le passé. La paléontologie, quant à elle, est la discipline scientifique qui étudie les restes fossiles des êtres vivants du passé et les implications évolutives ressortant de l'étude de ces restes. Pour étudier ces deux domaines, une large gamme de méthodes analytiques est utilisée pour permettre d'obtenir des informations à partir d'échantillons uniques et complexes à analyser.

De tels matériaux sont habituellement complexes au niveau de leur structure physique et chimique et peuvent s'altérer de diverses manières pendant les longues périodes d'enfouissement dans des environnements qui ne peuvent être totalement caractérisés. Ces matériaux se dégradent au cours du temps et sont exposés aux contaminants extérieurs, ce qui endommage l'intégrité chimique de la structure d'origine. Certains types de matériaux peuvent être ré-échantillonnés, alors que d'autres, par exemple, des fragments de pigments provenant de peintures rupestres préhistoriques ne peuvent fournir qu'une seule possibilité d'analyse.

II. L'analyse d'échantillons anciens

Les techniques d'analyse en archéologie et paléontologie sont nombreuses. Parmi ces techniques, la datation des échantillons est la plus utilisée [1,2]. Elle repose sur plusieurs techniques complémentaires qui vont utiliser les informations géologiques ou environnementales comme :

- Les méthodes de datation des couches de sédiments ou de l'environnement (stratigraphie). Ces méthodes permettent de fournir des informations sur le contexte dans lequel sont présent les objets ;
- L'étude du magnétisme terrestre qui est un bon outil de datation et de corrélation en archéologie (archéomagnétisme) ;
- La biochronologie, qui utilise les restes des espèces animales, ou la flore végétale pour obtenir une datation ;
- L'étude des objets présents sur un site, qui permettent d'apporter une information sur l'époque car les objets fabriqués évoluent dans le temps (typochronologie).

Bien évidemment, la datation peut être réalisée sur les objets archéologiques et paléontologiques comme :

- La datation radiométrique qui est une méthode de datation absolue utilisant la variation au cours du temps de la proportion de radio-isotopes (la plus connue étant la datation au carbone 14) ;
- La thermoluminescence, qui étudie l'émission de rayonnement à partir de minéraux qui ont été chauffés par le passé ;
- La résonance de spin électronique (ESR) qui mesure la dose de radiations d'un objet en utilisant les propriétés magnétiques ;
- La racémisation des acides aminés qui exploite leur chiralité dans des organismes fossiles pour permettre une datation ;
- La datation moléculaire qui étudie l'accumulation des changements dans les macromolécules biologiques en partant de l'hypothèse que ces changements ont une vitesse moyenne constante sur de longues périodes (étude de l'ADN mitochondrial).

Les paléontologues et les archéologues cherchent à obtenir une datation des objets découverts, mais également à identifier les échantillons trouvés sur les sites de fouilles. L'identification morphologique est, bien sûr, la première arme dont dispose les scientifiques mais d'autres méthodes existent.

Nous nous focaliserons sur les analyses des molécules organiques, avec deux objectifs :

- La détection des protéines dans les échantillons archéologiques en utilisant les techniques de colorations et de spectrométries ;
- Les méthodes d'identification des résidus protéiques contenus dans les échantillons archéologiques ou paléontologiques telles que le TOF-SIMS, qui permet de réaliser de l'imagerie, la chromatographie en phase gazeuse et les techniques immunologiques qui permettent de cibler les protéines avec des anticorps.

III. La détection des protéines dans les échantillons archéologiques

1. Colorations des protéines

La coloration des protéines est une technique classiquement utilisée pour détecter les protéines dans un échantillon en biologie. De nombreuses méthodes existent actuellement, de la coloration au bleu de coomassie à la fluorescence, et ceci pour des applications variées en recherche ou milieux hospitalier. La coloration des protéines dans les fossiles et les échantillons archéologiques a été décrite depuis le début des années 1980 [3]. Fosse et al, ont utilisé le mercurochrome sur des dents fossiles de mammifères du Crétacé, incluses dans une résine époxy. Ce colorant réagit avec les groupes sulfhydryle et les ponts disulfures des protéines. Dans cette étude, la coloration des dents fossiles était plus prononcée près de la cavité pulpaire, montrant la présence de protéines dans la dentine [3].

Des tests colorimétriques basés sur l'utilisation de la ninhydrine qui réagit avec les groupements amine des protéines donnant une coloration rose-violet ou d'OPA (o-Phthaldialdéhyde) qui réagit avec les groupes amines primaires de protéines pour former un produit fluorescent de couleur bleue, ont été utilisés pour détecter la présence de protéines sur des outils préhistoriques [4]. Des résidus de sang provenant de plusieurs espèces animales ont été découverts à la surface des outils préhistoriques de chert, de basalte et d'obsidienne (de 1 000 à 6 000 ans) dans des sites en plein air de la côte ouest et de la forêt boréale du Canada.

À la fin des années 1990, un test colorimétrique qui permet de détecter le sang dans les urines, Hemastix, a été appliqué pour détecter la présence de résidus de sang sur des artefacts de pierre provenant des fouilles archéologiques de Çayönü Tepasi en Turquie [5]. Le test est basé sur l'activité peroxydase de l'hémoglobine qui clive les molécules d'oxygène de l'hydroperoxyde et qui catalyse la transformation de la 3,3',5,5'-tétraméthylbenzidine (jaune clair incolore) en sa forme oxydée (orange à vert), pour de très fortes concentrations de sang.

2. Spectrométries vibrationnelles

Les techniques de spectroscopie vibrationnelle telles que la spectroscopie Infra Rouge (IR) et la spectroscopie Raman représentent des méthodes pertinentes

pour obtenir des informations moléculaires simultanées sur les phases minérales et organiques.

Ces deux techniques d'analyse utilisent les énergies associées aux mouvements vibrationnels entre les atomes d'une molécule. En spectroscopie IR, les fréquences absorbées par un échantillon positionné sur le trajet d'un faisceau IR sont mesurées (mode transmission ou réflexion), tandis que la technique Raman consiste à mesurer la lumière diffusée par un échantillon préalablement excité par un laser.

a) Spectroscopie Infra-rouge (IR)

Edwards et al [6] ont utilisés différentes méthodes pour étudier de la peau fossilisée de reptile. L'imagerie IR a permis pour observer la bande amide 1 dans des échantillons de peau de reptiles actuel et fossile (figure 2). Cette étude a permis de mettre en évidence la préservation de la structure de la peau de reptile fossilisé depuis 50 millions d'années (figure 2).

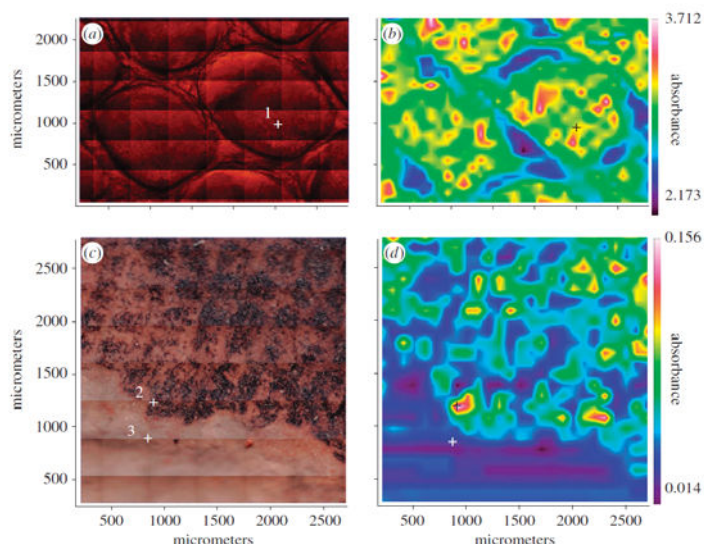


Figure 2 : (a) Image au microscope optique de la peau de reptile actuel avec la lumière transmise. (b) Carte de l'absorption à 1653 cm^{-1} de la peau de reptile actuel (mode de transmission). (c) Peau de reptile fossilisée en lumière réfléchie (zone encadrée sur la figure 1b). (d) Carte de l'absorption à 1653 cm^{-1} de la peau fossilisée (mode ATR) (extrait de Edwards et al. [6])

Cotte et al [7] ont montré à l'aide de microscopie IR l'état de dégradation des différentes couches de la peau (épiderme, derme) d'une momie Gréco-Romaine en étudiant les bandes amide A, B, I et II.

b) Spectroscopie Raman

L'analyse par spectrométrie Raman a été utilisée pour étudier des ossements notamment lors de la mise en évidence d'épisodes de minéralisation [8]. Cette étude s'est focalisée sur la présence de microsphères qui se forment durant la diagenèse

de l'os. L'analyse Raman a permis d'étudier les oxydes de fer qui composent les hématites et goethites présents dans ces microsphères. La présence des microsphères et des autres phases minérales dans l'étude des os a mis en évidence deux épisodes de minéralisation (formation d'oxyde de fer lors d'attaque microbienne et la cémentation tardive de calcite / barite / gypsum).

Un autre exemple d'étude par spectroscopie Raman est l'analyse d'ossements paléontologique et archéologique pour pré-sélectionner les os en fonction de la qualité du collagène. France et al [9] ont étudié des ossements et ont identifié avec succès le collagène conservé dans l'os. Ils ont aussi mis en évidence que les surfaces fraîches exposées offrent un maximum de succès pour déterminer la qualité du collagène. La figure 3 montre les spectres Raman d'échantillons d'ossement de bonne qualité en haut, et de mauvaise qualité en bas. Il est possible d'observer la diminution d'intensité des bandes amide III, I et de la liaison C-H dans les échantillons pour lesquels le collagène était mal préservé.

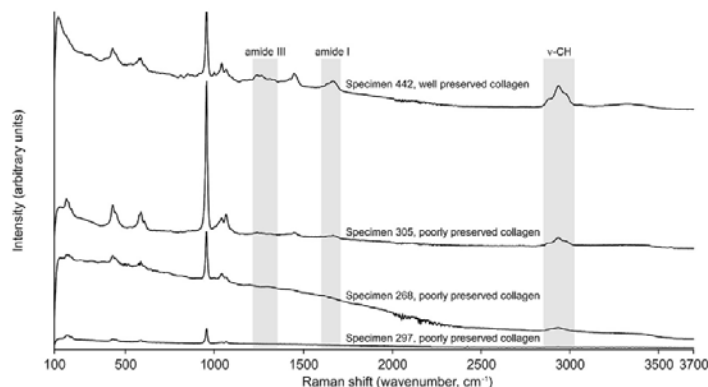


Figure 3 : Spectres Raman de 4 échantillons d'ossements. Le premier échantillon (442) en haut été composé de collagène bien conservé et les trois échantillons du bas (305, 258, 297), contenaient du collagène dégradé (extrait de France et al. [9]).

IV. Identification des résidus protéiques contenus dans les échantillons archéologiques ou paléontologiques

1. Analyse par TOF-SIMS

Le TOF-SIMS (Time of Flight Secondary Ion Mass Spectrometry) est une technique non destructive (plusieurs types de mesures possibles sur un échantillon unique) permettant l'identification simultanée de composés inorganiques et organiques. Le principe repose sur l'analyse des ions secondaires émis suite à une irradiation de la surface de l'échantillon à l'aide d'un faisceau d'ions primaires.

L'utilisation du TOF-SIMS a été grandement améliorée par le développement de sources qui utilisent des « clusters » d'or ou de bismuth, ce qui a eu pour conséquence une amélioration de la sensibilité, de la gamme de masse (jusqu'à m/z 1500) et de la résolution du faisceau d'ion. A titre d'exemple, les « clusters » de bismuth ont atteint une résolution inférieure à 400 nm [10].

Des analyses TOF-SIMS ont été réalisées sur de la peau de momie[11] (figure4) en étudiant les ions des acides aminés (leucine, valine...) et des acides gras (acide palmitique, acide oléique). L'analyse TOF-SIMS a permis de mettre en évidence une dégradation de la kératine et du collagène (en suivant les ions des acides aminés), et d'une redistribution des lipides présents sur la peau.

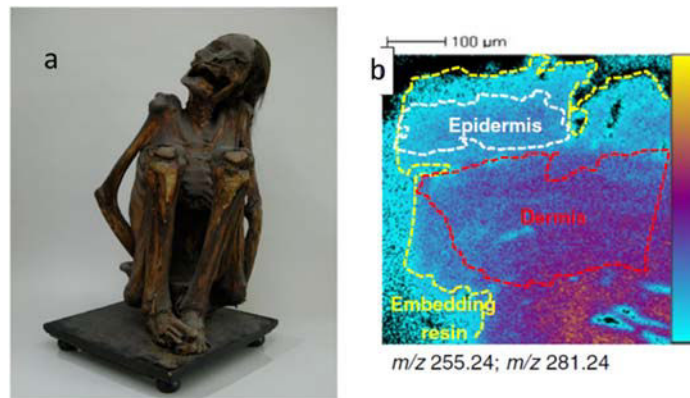


Figure 4 : (a) Photographie de la momie Pica-Tarapacá (1035–1217 AD) provenant du laboratoire anthropologie du Muséum National d'Histoire Naturel de Paris. (b) Image TOF-SIMS de l'ion palmitate à m/z 255,24 et de l'ion oléate à m/z 281,24 sur une partie de peau de momie (les pointillés blancs délimitent la partie de l'épiderme tandis que les pointillés rouge le derme). (Extrait de Cersoy et al. [11])

Le groupe de Surmik et al. [12] a étudié un échantillon fossilisé d'os du milieu du Triassique ($-252,2 \pm 0,5$ et $-201,3 \pm 0,2$ millions d'années) et, plus précisément, une structure qui pourrait être un vaisseau sanguin. L'analyse TOF-SIMS a permis de mettre en évidence la présence d'ions spécifiques de la glycine, alanine, proline, lysine, hydroxyproline et hydroxylysine.

2. Méthodes d'analyses des acides aminés

L'identification des protéines présentes dans les échantillons anciens est basée sur l'analyse de la composition des acides aminés.

Classiquement, cette méthode se réalise en appliquant successivement :

- Le nettoyage des échantillons,

- Le broyage de quelques grammes,
- L'hydrolyse qui permet de transformer les protéines en acides aminés à l'aide, par exemple, d'acide chlorhydrique,
- La dérivation des acides aminés pour les rendre volatiles.

Lors de l'analyse d'ossements, une étape de décalcification est réalisée à l'aide d'une solution acide (HCl 6N) ou basique (NaOH 1M).

L'étude des protéines présentes dans des fossiles ou des objets archéologiques a débuté dans les années 50.

La première méthode développée a permis l'analyse d'acides aminés par chromatographie sur papier, basée sur la séparation des acides aminés adsorbés sur un papier de cellulose grâce à un solvant s'élevant par capillarité. La détection des acides aminés est réalisée à l'aide de la ninhydrine. Des acides aminés présents dans des ossements du Miocène (il y a 23 à 5 millions d'années) [13] ou objets égyptiens [14] ont été ainsi analysés.

La seconde méthode est la chromatographie par échange de cations qui est basée sur l'absorbance mesurée des acides aminés dérivés avec la ninhydrine sur des ossements du Pléistocène [15] (2,58 millions d'années à 11 700 ans avant le présent) ou sur des pots en verre d'Ethiopie [16]. D'autres molécules ont été utilisées pour rendre plus sensible la détection des acides aminés comme le Phenylisothiocyanate, o-phthalaldehyde [17].

La troisième méthode est l'analyse des acides aminés en chromatographie liquide ont été réalisées en utilisant des séparations sur colonne C₁₈ [18-21].

3. Chromatographie en phase gazeuse

La chromatographie phase gazeuse (GC) qui est l'une des techniques les plus utilisées pour l'analyse d'échantillons anciens. Cette technique présente l'avantage de réduire la quantité d'échantillon utilisé.

Cette méthode se déroule de la façon suivante :

- La première étape est l'extraction puis l'hydrolyse des protéines,
- Ensuite, les acides aminés sont dérivés (à l'aide d'alkyle chloroformates ou trifluoroacetamide ou alkylsilyle acetamide) pour les rendre volatiles pour l'analyse GC
- Dans un dernier temps, la détection des acides aminés est réalisée à l'aide d'un détecteur à ionisation de flamme (FID).

Cette technique a permis d'identifier la composition en acides aminés de fossiles provenant de fossiles du Pléistocène jusqu'au Jurassique (201,3 à 145 millions d'années) [22,23]. Mais il a été montré que la quantité d'acides aminés diminue avec le temps et que certains acides aminés comme la cystéine ne sont pas détectés.

Le développement du couplage de ces méthodes avec la spectrométrie de masse pour l'analyse d'échantillons anciens dans les années 1980 a permis une diminution des quantités d'échantillons utilisées (quelques microgrammes) [1]. L'utilisation de la chromatographie gazeuse avec pyrolyse couplée à la MS, a été proposée pour l'étude d'échantillons anciens. Quelques microgrammes d'échantillons sont chauffés à très haute température (500-1000°C). La décomposition de l'échantillon génère des produits de pyrolyses qui sont analysés par MS. Cette technique est couramment utilisée en paléontologie [24,25] et en archéologie [26,27]. Par exemple, Manning et al. [52] ont montrés, la présence de composés organiques dérivés de la peau de façon endogène par spectrométrie de masse par chromatographie en phase gazeuse avec pyrolyse (Py-GCMS) sur un échantillon de *Edmontosaurus* (dinosauré vivant au Crétacé supérieur, il y a entre 73 et 66 millions d'années). Cette observation a permis de mettre en évidence la survie et la présence de macromolécules partiellement aliphatiques (figure 5).

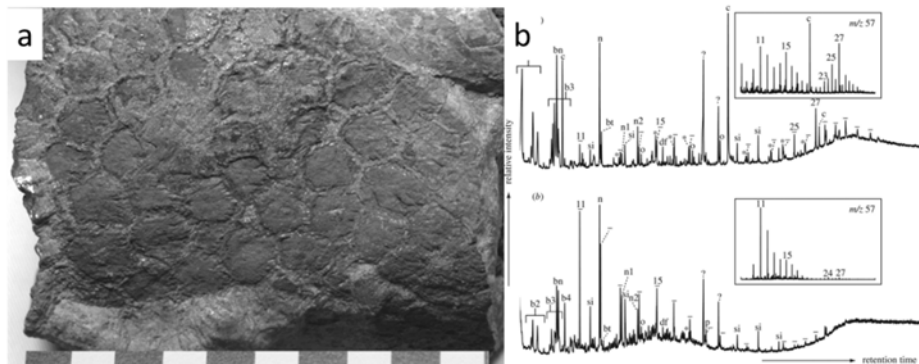


Figure 5 : (a) Surface dorsale d'une section de *Edmontosaurus*. (b) chromatogramme du courant ionique de l'analyse en Py-GCMS, en haut de l'enveloppe de la peau et en bas le Sédiments environnants associés à l'enveloppe de la peau. Le spectre encadré montre le spectre de l'ion à m/z 57 permettant de voir la distribution de n-alcane avec le nombre indiquant la longueur des chaînes (Extrait de Manning et al. [28]).

4. Méthodes immunologiques

Les méthodes immunologiques utilisant des anticorps spécifiques sont utilisées pour l'analyse des protéines contenues dans les échantillons anciens depuis le début des années 1960. Parmi ces techniques immunologiques, sont utilisées le

RIA (Radioimmunoassay), basé sur l'utilisation d'antigènes marqués radioactivement ou l'ELISA (enzyme-linked immunosorbent assay), technique immuno-enzymatique de détection qui permet de visualiser une réaction antigène-anticorps grâce à une réaction colorée produite par l'action sur un substrat d'une enzyme préalablement fixée à l'anticorps.

A titre d'exemple, de l'albumine a été détectée sur un fragment de crâne fossile daté d'il y a 1,6 millions d'années à l'aide de techniques immunologiques. Cette étude a montré que cette protéine se rapprochait plus de l'albumine présente chez l'homme que celle du cheval ou des bovins [29,30]. Ces techniques ont été également utilisées pour détecter des traces de sang sur des objets archéologique [31-34]. Ces méthodes immunologiques présentent l'inconvénient de ne cibler que les protéines à étudier. Bien que les quantités d'échantillons soient faibles pour ce genre de méthode, si l'anticorps ne détecte pas l'épitope car celui-ci a été dégradé ou sa conformation inexistante, il n'y aura pas détection.

V. Conclusion

Comme énoncé précédemment, de nombreuses techniques sont utilisées pour l'analyse d'échantillons archéologiques et paléontologiques. Outre les techniques de détection, des outils de caractérisation peuvent être utilisés pour détecter la présence des acides aminés ou des lipides. D'autres techniques permettent de détecter la présence de protéines en utilisant des anticorps spécifiques. Mais aucune d'entre elles ne permet l'identification de la séquence des protéines ou l'identification de l'espèce sans cibler la protéine. Récemment, une nouvelle technique, la protéomique, a été introduite pour l'étude des protéines anciennes. Cet outil est apparu dans les années 2000 avec le développement de méthode d'ionisation douce en spectrométrie de masse (MALDI et ESI). Elle permet d'obtenir l'identification des protéines d'un échantillon à partir du séquençage des peptides et de l'utilisation de base de données regroupant les protéines des différents organismes (cette méthode est détaillée dans le chapitre II, partie 2).

Dans la suite de ce manuscrit, deux méthodologies seront développées : l'analyse « bottom up » et « top down ». Ces deux techniques seront utilisées pour étudier deux types d'échantillons différents et seront optimisées pour permettre d'obtenir un grand nombre d'informations et aider les archéologues ou paléontologues dans leurs recherches.

Bibliographie

- (1) Colombini, M. P.; Modugno, F.: *Organic mass spectrometry in art and archaeology*; John Wiley & Sons, 2009.
- (2) Edwards, H.; Vandenabeele, P.: *Analytical archaeometry: selected topics*; Royal Society of Chemistry, 2016.
- (3) Fosse, G.; Justesen, N.-P.; Wesenberg, G.: Microstructure and chemical composition of fossil mammalian teeth. *Calcif. Tissue Int.* **1981**, *33*, 521.
- (4) Loy, T. H.: Prehistoric blood residues: detection on tool surfaces and identification of species of origin. *Science* **1983**, *220*, 1269-1271.
- (5) Loy, T. H.; Wood, A. R.: Blood residue analysis at Çayönü Tepesi, Turkey. *Journal of Field Archaeology* **1989**, *16*, 451-460.
- (6) Edwards, N. P.; Barden, H. E.; van Dongen, B. E.; Manning, P. L.; Larson, P. L.; Bergmann, U.; Sellers, W. I.; Wogelius, R. A.: Infrared mapping resolves soft tissue preservation in 50 million year-old reptile skin. *Proceedings of the Royal Society B: Biological Sciences* **2011**, *278*, 3209-3218.
- (7) Cotte, M.; Walter, P.; Tsoucaris, G.; Dumas, P.: Studying skin of an Egyptian mummy by infrared microscopy. *Vibrational Spectroscopy* **2005**, *38*, 159-167.
- (8) Kremer, B.; Owocki, K.; Królikowska, A.; Wrzosek, B.; Kazmierczak, J.: Mineral microbial structures in a bone of the Late Cretaceous dinosaur *Saurolophus angustirostris* from the Gobi Desert, Mongolia—a Raman spectroscopy study. *Palaeogeography, Palaeoclimatology, Palaeoecology* **2012**, *358*, 51-61.
- (9) France, C. A. M.; Thomas, D. B.; Doney, C. R.; Madden, O.: FT-Raman spectroscopy as a method for screening collagen diagenesis in bone. *J. Archaeo.Sci.* **2014**, *42*, 346-355.
- (10) Touboul, D.; Kollmer, F.; Niehuis, E.; Brunelle, A.; Laprévote, O.: Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1608-1618.
- (11) Cersoy, S.; Richardin, P.; Walter, P.; Brunelle, A.: Cluster TOF-SIMS imaging of human skin remains: analysis of a South-Andean mummy sample. *J. Mass Spectrom.* **2012**, *47*, 338-346.
- (12) Surmik, D.; Rothschild, B. M.; Pawlicki, R.: Unusual intraosseous fossilized soft tissues from the Middle Triassic *Nothosaurus* bone. *The Science of Nature* **2017**, *104*, 25.
- (13) Tuross, N.; Hare, P.: Collagen in fossil bone. *Carnegie Institution of Washington: Yearbook* **1978**, *77*, 891-895.
- (14) Nicholson, P. T.; Shaw, I.: *Ancient Egyptian materials and technology*; Cambridge University Press, 2000.
- (15) Wyckoff, R. W.; McCaughey, W. F.; Doberenz, A. R.: The amino acid composition of proteins from Pleistocene bones. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1964**, *93*, 374-377.

Généralités sur l'analyse des échantillons archéologiques et paléontologiques

- (16) Von Endt, D. W.: Amino-acid analysis of the contents of a vial excavated at Axum, Ethiopia. *J. Archaeo.Sci.* **1977**, *4*, 367-376.
- (17) Walton, D.; Curry, G. B.: Amino acids from fossils, facies and fingers. *Palaeontology* **1991**, *34*, 851-858.
- (18) Gurley, L. R., Valdez, J. G., Spall, W. D., Smith, B. F., & Gillette, D. D.: Proteins in the fossil bone of the dinosaur, *Seismosaurus*. *J. Protein Chem.* **1991**, *10*, 75-90.
- (19) Kaufman, D. S.; Manley, W. F.: A new procedure for determining DL amino acid ratios in fossils using reverse phase liquid chromatography. *Quaternary Science Reviews* **1998**, *17*, 987-1000.
- (20) O'Connell, T.; Hedges, R.: Isolation and Isotopic Analysis of Individual Amino Acids from Archaeological Bone Collagen: A New Method Using Rp-hplc. *Archaeometry* **2001**, *43*, 421-438.
- (21) Sroga, G. E.; Vashishth, D.: UPLC methodology for identification and quantitation of naturally fluorescent crosslinks in proteins: a study of bone collagen. *J. Chromatogr. B* **2011**, *879*, 379-385.
- (22) Akiyama, M.; Wyckoff, R. W.: The total amino acid content of fossil pecten shells. *Proceedings of the National Academy of Sciences* **1970**, *67*, 1097-1100.
- (23) Akiyama, M.; Davidson, F. D.; Matter, P.; Wyckoff, R. W.: Liquid and gas chromatographic analyses of several fossil proteins. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **1971**, *38*, 93-101.
- (24) Stankiewicz, B. A.; Briggs, D. E.; Evershed, R. P.: Chemical composition of Paleozoic and Mesozoic fossil invertebrate cuticles as revealed by pyrolysis- gas chromatography/mass spectrometry. *Energy Fuels* **1997**, *11*, 515-521.
- (25) Poinar, H. N.; Stankiewicz, B. A.: Protein preservation and DNA retrieval from ancient tissues. *Proceedings of the National Academy of Sciences* **1999**, *96*, 8426-8431.
- (26) Shedrinsky, A. M.; Stone, R. E.; Baer, N. S.: Pyrolysis gas chromatographic studies on Egyptian archaeological specimens: organic patinas on the "Three Princesses" gold vessels. *Journal of Analytical and Applied Pyrolysis* **1991**, *20*, 229-238.
- (27) Colombini, M. P.; Ribechini, E.; Rocchi, M.; Selleri, P.: Analytical pyrolysis with in-situ silylation, Py (HMDS)-GC/MS, for the chemical characterization of archaeological and historical amber objects. *Heritage Science* **2013**, *1*, 6.
- (28) Manning, P. L.; Morris, P. M.; McMahon, A.; Jones, E.; Gize, A.; Macquaker, J. H. S.; Wolff, G.; Thompson, A.; Marshall, J.; Taylor, K. G.; Lyson, T.; Gaskell, S.; Reamtong, O.; Sellers, W. I.; van Dongen, B. E.; Buckley, M.; Wogelius, R. A.: Mineralized soft-tissue structure and chemistry in a mummified hadrosaur from the Hell Creek Formation, North Dakota (USA). *Proceedings of the Royal Society B: Biological Sciences* **2009**, *276*, 3429-3437.
- (29) Borja, C.; García-Pacheco, M.; Olivares, E. G.; Scheuenstuhl, G.; Lowenstein, J. M.: Immunospecificity of albumin detected in 1.6 million-year-old fossils from Venta Micena in Orce, Granada, Spain. *American Journal of Physical Anthropology* **1997**, *103*, 433-441.
- (30) Lowenstein, J. M.; Borja, C.; García-Olivares, E.: Species-specific albumin in fossil bones from Orce, Granada, Spain. *Human Evolution* **1999**, *14*, 21-28.

Généralités sur l'analyse des échantillons archéologiques et paléontologiques

(31) Sensabaugh, G. F.; Wilson, A. C.; late P.L. Kirk, t.: Protein stability in preserved biological remains: II. Modification and aggregation of proteins in an 8-year-old sample of dried blood. *International Journal of Biochemistry* **1971**, *2*, 558-568.

(32) Sensabaugh, G. F.; Wilson, A. C.; late P.L. Kirk, t.: Protein stability in preserved biological remains: I. Survival of biologically active proteins in an 8-year-old sample of dried blood. *International Journal of Biochemistry* **1971**, *2*, 545-557.

(33) Downs, E. F.; Lowenstein, J. M.: Identification of archaeological blood proteins: A cautionarynote. *Journal of Archaeological Science* **1995**, *22*, 11-16.

(34) Newman, M. E.; Ceri, H.; Kooyman, B.: The use of immunological techniques in the analysis of archaeological materials—a response to Eisele; with report of studies at Head-Smashed-In Buffalo Jump. *Antiquity* **1996**, *70*, 677-682.

Partie 2 :

Analyse bottom-up des échantillons
archéologiques et paléontologiques

I. La protéomique

La protéomique désigne la science consacrée à l'étude des protéines, leur identification, leur quantification et l'étude de leur modification (post traductionnelle et chimique) à un moment précis dans un environnement donné. Le terme protéomique a vu son avènement dans les années 1990 [1].

Cette méthode est principalement basée sur l'utilisation de techniques complémentaires :

- La séparation de plusieurs centaines d'analytes par électrophorèse sur gel et / ou chromatographie (Mais aussi l'enrichissement des analytes) ;
- L'analyse par spectrométrie de masse des peptides et/ou des protéines en utilisant des techniques d'ionisation douce MALDI et ESI ;
- La bio-informatique, qui intègre des bases de données génomiques ou protéiques, permet l'identification des protéines et leur modification à l'aide des données de MS et MS/MS.

Ces trois étapes (séparation, analyse spectrométrique de masse et recherche de bases de données) permettent de mettre en place trois stratégies différentes pour l'identification des protéines : bottom up, middle down et top down (figure 1). Ces techniques se différencient par le traitement de l'échantillon en amont de l'analyse.

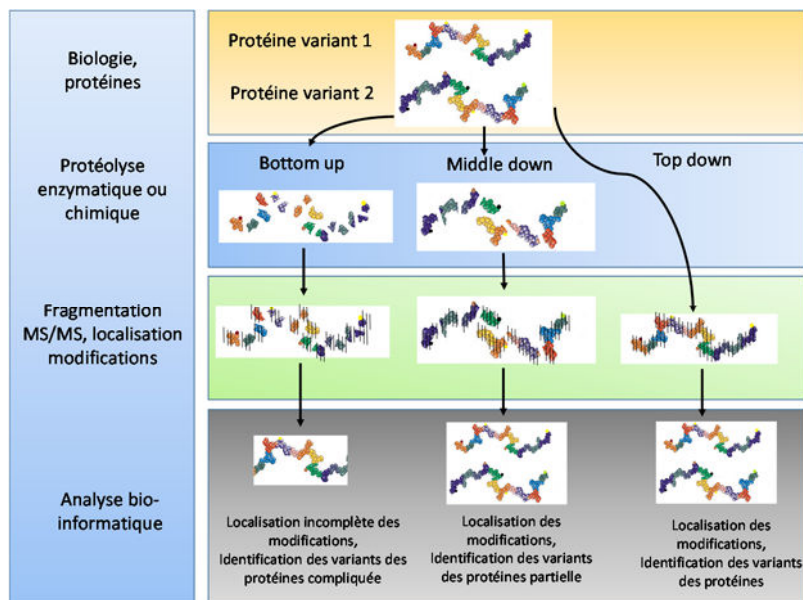


Figure 1: Stratégies bottom-up, middle-down et top-down. La stratégie top-down part de la protéine intacte pour arriver aux fragments peptidiques. La stratégie bottom-up et middle down partent quant à elles des peptides pour arriver aux protéines.

II. L'approche bottom up

L'approche bottom-up est illustrée dans la figure 2.

- Les protéines sont séparées sur gel SDS-PAGE 1D ou 2D (selon la complexité de l'échantillon) puis subissent une étape de digestion enzymatique ;
- Les protéines sont ensuite fractionnées en solution «off-gel» à l'aide de méthodes chromatographiques utilisant les propriétés des protéines (poids moléculaire, point isoélectrique, hydrophobicité). Ces protéines sont ensuite digérées en solution.
- Le mélange protéique peut être directement digéré et le mélange peptidiques analysé en spectrométrie de masse. Cette méthode est appelée « shotgun proteomics »,

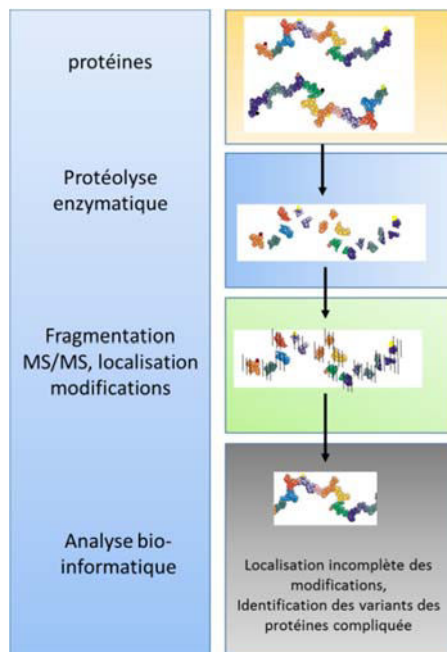


Figure 2 : Schéma de la stratégie bottom-up. L'échantillon de protéine est digéré à l'aide d'une enzyme de digestion (type trypsine), puis les peptides sont analysés par spectrométrie de masse en mode MS ou MS/MS. Les masses des peptides ou le séquençage des peptides, qui permet d'obtenir l'enchaînement en acides aminés, permettent l'identification de la protéine.

En général, l'enzyme utilisée est la trypsine car celle-ci présente une bonne spécificité et génère des peptides de taille relativement homogène. De plus, le site de coupure de la trypsine, situé après la lysine et après l'arginine au niveau C-terminal de la protéine, permet une meilleure ionisation des peptides produits.

Dans un premier cas, les peptides obtenus sont ensuite analysés par spectrométrie de masse en mode MS, comme la méthode PMF ou « Peptide Mass Fingerprint » réalisé à l'aide d'un spectromètre de masse MALDI-TOF. Cette méthode permet l'identification des protéines à partir des spectres de masse MS.

Dans un second cas, les peptides peuvent être analysés en mode MS/MS, ce qui va permettre d'obtenir l'enchaînement de la séquence en acides aminés des peptides après leur sélection.

Une analyse bio-informatique est réalisée pour identifier les protéines à partir des masses des peptides ou du séquençage des peptides qui permet d'obtenir l'enchaînement en acides aminés.

1. La préparation des échantillons

Avant de réaliser l'analyse bottom up, la préparation des échantillons est une étape primordiale qui déterminera la réussite de l'analyse. Brièvement, les protéines d'intérêt sont extraites via différentes méthodes qui seront spécifiques des échantillons [2,3]. Ces méthodes d'extraction peuvent être divisées en 5 groupes : l'homogénéisation mécanique, les ultrasons, la pression, un cycle de congélation, la lyse osmotique avec détergent.

Après l'extraction des protéines des échantillons, l'enjeu est de solubiliser les protéines en cassant les interactions telles que les interactions hydrophobes, cette étape reste une étape critique, si l'on considère la diversité des types de protéines et d'échantillons. Pour cela, il est possible d'utiliser des agents chaotropes, des détergents, des réducteurs, de modifier le pH des solutions ou la concentration en sels.

Ensuite, il est nécessaire d'éliminer les contaminants tels que :

- l'ADN ou l'ARN présent dans les échantillons pouvant être éliminés par une digestion enzymatique,
- Les lipides qui peuvent être extraits à l'aide d'une extraction liquide/liquide comme la méthode de Folch's ou Bligh & Dyer,
- Les sels et détergents qui peuvent être éliminés par dialyse,
- La précipitation des protéines (par l'acétone, le sulfate d'ammonium, l'acétate d'éthyle, l'acide trichloracétique) qui permet d'éliminer les sels, les détergents en récupérant les protéines sous forme de culot,
- La centrifugation,

Les protéines obtenues peuvent être séparées, soit par gel SDS-PAGE mono ou bidimensionnelle, soit par chromatographie liquide. Dans une dernière étape, les protéines sont digérées.

Dans un souci de clarté, l'analyse par empreinte peptidique ne sera pas expliquée mais, par contre, l'analyse par LC-MS/MS le sera.

2. La digestion des protéines

Avant de digérer les protéines avec une enzyme, les ponts disulfures entre les cystéines sont réduits par des agents réducteurs comme le Dithiothréitol (DTT) ou le tris(2-carboxyethyl)phosphine (TCEP). Ensuite, les fonctions thiols des cystéines sont protégées pour éviter la reformation des ponts disulfures pour permettre une digestion et une identification plus facile. De nombreux alkylants existent comme l'iodoacétamide, le 2-chloroacétamide, 4-vinylpyrolidine [2-4].

Après la réduction et l'alkylation des protéines, celles-ci peuvent être digérées pour former des peptides. L'enzyme la plus couramment utilisée pour générer des peptides protéolytiques est la trypsine. C'est une protéase stable, qui clive les protéines en peptides de manière très spécifique du côté C-terminal des résidus de lysine et d'arginine [5]. Cette enzyme permet d'obtenir des fragments de masse moyenne de 1500 Da, ce qui permet une bonne détection en MALDI ou lors du couplage avec une source ESI. De plus, les peptides générés sont facilement ionisables car les chaînes latérales des sites basiques ont des affinités protoniques élevées et le groupe amine N-terminal peut également se protonner [6]. La trypsine n'est pas la seule enzyme à être utilisée. De nombreuses autres enzymes peuvent être utilisées pour la digestion des protéines et l'analyse protéomique comme la LysC qui coupe les protéines en C-terminal après la lysine, l'AspN qui digère les protéines après les acides aspartique en N-terminal ... [4].

De nombreuses méthodes de digestion existent pour réaliser une analyse bottom up, allant de la simple digestion en gel ou liquide en passant par la digestion en utilisant des unités de filtration (FASP ou Filter Aide Sample Preparation) ou des cônes (iST ou In Stage Tips) [7-10].

Le séquençage du génome de divers organismes, comme celui de l'être humain, a permis de déterminer les séquences en acides aminés de la majeure partie des protéines de ces organismes. Les bases de données comme SwissProt (<http://www.uniprot.org/>) ou NCBI (<https://www.ncbi.nlm.nih.gov/>) regroupent ces

séquences de protéines. Ainsi, si la séquence d'une protéine est connue, il est possible de prédire le site de coupure de l'enzyme, et déterminer la séquence et la masse des peptides.

L'analyse bio-informatique qui suit permet de comparer les spectres de fragmentation des peptides de mélange de protéines obtenues par spectrométrie de masse aux spectres théoriques de fragmentation des peptides créés par les logiciels de bio-informatiques à partir des bases de données.

3. Séparation des peptides par couplage chromatographique

La séparation des peptides issus de la digestion enzymatique des protéines est réalisée le plus souvent en utilisant la chromatographie liquide. Les peptides sont séparés en règle générale sur une colonne contenant de la phase C₁₈, car cette phase est adaptée aux petites molécules hydrophobes comme les peptides. Ces molécules hydrophobes ou peptides présents dans la phase mobile polaire (eau) tendent à s'adsorber à la phase stationnaire hydrophobe (C₁₈), et les molécules hydrophiles dans la phase mobile traverseront la colonne et sont éluées en premier. L'éluion des peptides fixes sur la phase est réalisée à l'aide d'un gradient avec une phase mobile polaire (acétonitrile) de plus en plus élevé. Les peptides sont élués à des temps spécifique appelés temps de rétention (RT) ou temps d'éluion.

Les peptides élués sont ionisés dans une source électronébulisation, et injectés directement dans le spectromètre de masse. Celui-ci les détecte, les sélectionne et les fragmente. Classiquement en protéomique, la nanoLC-MS/MS est utilisée, car elle permet de travailler à des débits de l'ordre du nanolitre par minutes, d'utiliser des colonnes de longueur de 50 centimètres avec un diamètre de 75 µm et des particules de phase C₁₈ de 2 µm. Ce type de configuration améliore la résolution chromatographique et donc la séparation des peptides isobariques à des temps d'éluion différents et de les identifier. Comme les peptides sont séparés, le nombre d'ions introduits dans le spectromètre est réduit donc l'effet de suppression d'ion, où le signal d'un ion majoritaire supprime celui d'un autre ion minoritaire, est diminué [11], ce qui augmente la gamme dynamique observable.

4. La MS/MS ou spectrométrie de masse en tandem

Afin d'obtenir des informations sur la séquence en acides aminés d'un peptide, la spectrométrie de masse en tandem (ou MS/MS) peut être utilisée. Il s'agit d'un processus de fragmentation d'une ou plusieurs liaisons de la molécule. Les

processus de fragmentation des peptides varient en fonction de l'environnement et la nature des chaînes latérales. Une nomenclature a été établie en fonction de la localisation de la fragmentation et de la position de la charge sur la séquence peptidique (partie *N*-terminale ou *C*-terminale) [12,13].

La fragmentation peptidique intervient au niveau de la liaison amide entre les acides aminés (figure 3). Cette fragmentation permet d'obtenir des séries de fragments *N*-terminaux et *C*-terminaux (appelés respectivement ions *b* et *y* pour la fragmentation CID, HCD ou IRMPD).

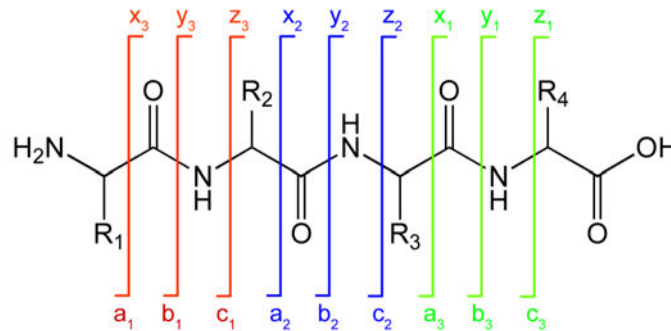


Figure 3 : Nomenclature des différentes séries d'ions peptidiques qui peuvent être générés lors de la fragmentation MS/MS. Proposée initialement par Roepstorff et Fohlman [14], elle fut modifiée par Johnson et al. [15] avant d'être définitivement adoptée. La charge portée par les fragments permettra leur détection. Si la charge du côté *N*-terminal alors l'ion est dit de série *a*, *b* ou *c*. Si la charge du côté *C*-terminal alors l'ion est dit de série *x*, *y* ou *z*.

Le spectre de fragmentation est composé de pic correspondant aux fragments des peptides. La différence entre ces pics permet de déterminer la masse d'un acide aminé et l'enchaînement de ces acides aminés dans le spectre de fragmentation rend possible l'identification des protéines. Ces spectres de fragmentation obtenus par spectrométrie de masse seront ensuite analysés avec des logiciels de bio-informatique pour identifier les peptides et les protéines.

5. Bio-informatique

L'identification des protéines après digestion est réalisée à l'aide de banques de données protéiques ou nucléiques. Les spectres MS/MS obtenus par spectrométrie de masse sont comparés aux spectres théoriques de chaque peptide, issu de la digestion *in silico* des protéines contenues dans les banques de données. Différents paramètres de recherche sont définis comme : la taxonomie, l'enzyme de digestion, le nombre de sites de coupure manqués, les modifications chimiques ou post-traductionnelles, le type de fragmentation, l'erreur tolérée en masse pour les précurseurs et les fragments.

Il existe de nombreuses bases de données comme UniProtKB (<http://www.uniprot.org/>) qui regroupe les bases de données Swiss-Prot (555 594 séquences) et TrEMBL (90 050 711 séquences). La base de données Swiss-Prot est composée de séquences « révisées » qui sont vérifiées manuellement avec la littérature et par informatique, mais également des séquences « non révisées », annotées par informatique uniquement.

La deuxième base de données la plus utilisée est NCBI avec 95 563 598 séquences (<https://www.ncbi.nlm.nih.gov/>), qui regroupe les séquences de plusieurs sources y compris des traductions de régions de codage annotées dans GenBank, RefSeq et TPA, ainsi que des enregistrements de SwissProt, PIR (proteins informatic resource), PRF (Protein Research Foundation) et PDB (protein data bank).

Il existe différents algorithmes de recherche qui permettent d'identifier les protéines à partir des spectres de fragmentations comme Mascot [16], Sequest [17] et Andromeda [18] qui ont été utilisés au cours de ce projet. Mais de nombreux autres algorithmes ont été développés comme Protein prospector [19], ProBlID [20], X!Tandem [21], OMSSA [22], ProSight [23] et Inspect [24]. Ces algorithmes comparent les spectres de fragmentations obtenus par spectrométrie de masse, issus des mélanges peptidiques de protéines digérées, aux spectres théoriques de fragmentation des peptides de toutes les protéines présentes dans les bases de données. Ces algorithmes intègrent différents paramètres de recherche comme les modifications post-traductionnelles, l'erreur de masse du spectromètre, les différentes bases de données, mais aussi les différents taxons des organismes vivants. Le résultat final est l'identification des protéines grâce aux peptides fragmentés par spectrométrie de masse et l'obtention d'un score qui permet de filtrer les données.

Chaque moteur de recherche calcule un score et une probabilité statistique. Le score obtenu est un score statistique indiquant dans quelle mesure les données expérimentales correspondent à la séquence de la base de données. Le plus souvent, ce score est accompagné d'une probabilité statistique appelé E-value pour Mascot qui est, le nombre de fois où il est possible d'obtenir un score X ou mieux par hasard. Une autre valeur statistique trouvée, est la P-value pour Byonic qui est la probabilité qu'un événement se produise par hasard.

Cette comparaison entre les spectres MS/MS expérimentaux et théoriques peut également se faire à l'aide d'une bibliothèque de spectres MS/MS de peptides

acquis et identifiés au préalable. Cet outil est intégré dans les logiciels comme MyProMS [25] ou sur Proteome discover 2.1 avec l'outil MSPepSearch [26]. Les logiciels attribuent un score qui mesure le degré de similitude entre le spectre expérimental et le spectre théorique. Ce score d'identification dépend de la banque utilisée et de la tolérance de masse qui permet aussi de filtrer les identifications.

Il existe d'autres méthodes pour identifier les protéines après digestion. Le séquençage *de novo* lors duquel les spectres MS/MS sont analysés et annotés pour déterminer la séquence en acides aminés. Cette méthode n'utilise pas de base de données de séquences. Des logiciels ont été conçus pour faciliter la tâche des analystes tels que PEAKS [27], PepNovo [28], Novor [29] ou DenoGui qui est une interface qui regroupe plusieurs algorithmes de *de novo* [30].

III. Bottom up – archéologie & paléontologie

La variation de la structure d'une protéine exprimée dans différentes espèces est l'aboutissement d'une longue histoire de sélection naturelle et de mutations tempérées par l'environnement au sein de laquelle les organismes ont évolué.

L'étude des séquences de base de l'ADN révèle cette variation montrant que certaines protéines sont bien conservées, avec une séquence similaire de paires de bases d'une espèce à l'autre (comme le collagène), tandis que d'autres présentent des variations importantes dues à la pression sélective.

La spectrométrie de masse (MS), et plus particulièrement la protéomique, peut permettre l'étude de la structure primaire des protéines qui sont réellement présentes dans les cellules. En plus de fournir une information sur la structure des protéines et leurs fonctions, la MS peut révéler des changements évolutifs de la séquence en acides aminés dans une protéine présentant un intérêt pour les archéologues et les paléontologues.

La protéomique est extrêmement sensible et ne nécessite aucune amplification pour séquencer la structure des protéines. De plus, les développements récents dans la miniaturisation des techniques d'ionisation douce et la chromatographie à l'échelle nanométrique ont permis de l'appliquer à de très petites quantités d'échantillons, en étendant les applications potentielles à un contexte archéologique [31,32].

L'analyse protéomique permet de répondre à différentes questions en archéologie ou en paléontologie comme :

- Fournir des méthodes pour identifier les protéines présentes dans les échantillons de différentes natures.

En 2014, Shevchenko et al [33] ont identifié des protéines dans des restes de pain retrouvé en Chine sur un site archéologique daté de 500 à 300 avant JC. Pour réaliser cette identification, ils ont extrait les protéines de l'échantillon puis séparé les protéines sur gel SDS-PAGE et analysé chaque bande par nanoLC-MS/MS. Cette analyse a mis en évidence que la nourriture retrouvée dans le tombeau de Subeixi était du pain au levain fabriqué à partir d'orge blanchie avec addition de millet et de sorgho en utilisant de la levure de boulanger et des bactéries lactiques.

- Distinguer les espèces en utilisant des protéines trouvées dans des restes archéologiques ou paléontologiques.

En Solazzo et al [34], ont pour la première fois montré que la protéomique permettait d'identifier des protéines et les espèces dont elles sont issues à partir de 250 mg d'échantillons archéologiques. Ils ont montré que des protéines musculaires de pinnipèdes et cétacés étaient présentes dans des échantillons archéologiques de 1200 – 1400 après Jésus Christ.

Une autre source clairement importante de protéines sont les restes osseux. Les principales protéines de l'os sont le collagène et l'ostéocalcine. Ces deux protéines peuvent être conservées longtemps après la disparition de l'ADN, en raison en partie, des effets protecteurs de la phase minérale de l'os. Les premières tentatives ont été faites sur l'extraction de l'ostéocalcine, une protéine relativement petite qui représente environ 3% des protéines osseuses. Sa petite taille et ses propriétés de liaison minérale suggèrent qu'elle serait appropriée pour l'analyse des peptides. Nielsen-Marsch et al ont analysé l'ostéocalcine par méthode bottom up avec une digestion trypsique pour déterminer la séquence de cette protéine chez Neandertal et des primates [35]. De nombreuses équipes ont travaillé sur la détection du collagène contenu dans les ossements anciens. Collins et al ont développé une méthode de détection des protéines des os par PMF (Peptide Mass Fingerprint) appelé ZooMS [36]. Elle consiste en une digestion des ossements puis purification des peptides et analyse MS et MS/MS par MALDI-TOF. Ensuite, une comparaison des peptides identifiés peut être réalisée entre les échantillons.

Cette méthode et l'analyse bottom up ont été appliquées à de nombreux échantillons de restes humains jusqu'au fossiles de dinosaures pour répondre à différentes questions :

- L'identification du collagène contenu dans des échantillons de mammoth et mastodonte.

Deux équipes Buckley et al, Cappellini et al, ont identifié des protéines provenant d'ossements de mammoth et mastodonte avec la méthode ZooMS à partir d'une centaine de milligrammes de poudre d'os. Après digestion de l'ossement, des peptides du collagène ont été identifiés et ont permis de différencier les deux espèces [37,38].

- La distinction entre des espèces à partir de la séquence de collagène.

Buckley et al. ont différencié à partir de peptides issus du collagène deux échantillons provenant de chèvre et de mouton [39] à l'aide d'une analyse PMF. Cette même équipe a montré qu'il est possible de distinguer les différentes espèces de rongeur à partir d'une analyse MS d'ossement [40]. Ceci a été possible avec l'identification du collagène 1 alpha 1 et alpha 2 à partir de 60-80 peptides qui représentent environ 50 à 60 % de couverture de séquence. L'identification de peptides spécifiques du collagène a permis de différencier les espèces de rongeurs.

- L'étude de l'évolution des espèces.

Welker et al a montré avec la méthode ZooMS que le collagène 1 (alpha 1 et alpha 2) contenu dans les os chez les mammifères placentaires possède suffisamment de variabilité pour fournir des informations systématiques fiables [41]. Bien sûr, une phylogénie basée sur deux protéines peut être sensible à différents facteurs tels que les taxons manquants, les taux de variations moléculaires aberrants, la sélection agissant sur les séquences codant pour les protéines.

- La détermination de marqueurs de vieillissement grâce aux sites de déamidation,

Van Dorm et al ont montré que la déamidation de la glutamine est un marqueur de la dégradation des ossements en utilisant la méthode ZooMS [42]. Ils ont montré que la déamidation était influencée par les conditions d'enfouissement et l'âge thermique.

- L'analyse d'ossements très anciens comme ceux du *Brachylophosaure* ou du *Tyrannosaure Rex*

Les avancées technologiques en spectrométrie de masse au niveau de la sensibilité de détection et les méthodes de préparation ont permis de détecter des peptides de collagène d'ossements de dinosaures datant de plusieurs millions d'années [43-45]. Pour ce faire, Schroeter et al.[45] ont extrait les protéines contenues dans 1 gramme d'os de *Brachylophosaurus canadensis*, puis séparé ces protéines sur gel SDS-PAGE, digéré les protéines et analysé par couplage LC-MS/MS. Cette étude a permis d'identifier 8 peptides uniques de collagène 1, et de compléter une étude antérieure réalisée par Schweitzer et al.[46].

- L'information sur le processus d'altération du matériel organique

Freinai et al [47] ont mis au point une approche protéomique pour la caractérisation des cheveux provenant de momies datées de 3880 ans. Cette méthode, nécessitant 500 µg d'échantillon, a permis d'identifier des kératines des cheveux de cette momie et d'évaluer, pour la première fois, la préservation moléculaire de cheveux par une approche protéomique. Par exemple certaines modifications chimiques telles que la déamidation des glutamines et asparagines ou l'oxydation des cystéines ont été étudiées et comparées aux résultats obtenus pour les cheveux modernes.

La protéomique bottom up est aussi utilisée pour l'identification des protéines archéologiques provenant d'autres types d'échantillons tels que les dents [48], ou encore la peau [49] de momies.

IV. Objectifs

Dans cette partie, l'objectif a été de mettre au point une nouvelle méthode robuste d'extraction, d'analyse et d'identification des échantillons archéologiques ou paléontologiques. Cette méthode doit intégrer de nombreux points, comme la faible quantité d'échantillon mis à disposition, la provenance de ces échantillons (enfouissement sous terre, immersions dans la mer, céramique...).

Cette méthode aura pour but d'améliorer la préparation des échantillons archéologiques en travaillant sur des quantités inférieures à 100 mg d'ossement.

Elle devra permettre d'obtenir une meilleure couverture de séquence des protéines identifiées (supérieur à 60 %), et un nombre de peptides identifiés important. L'un des objectifs est d'obtenir des informations sur les modifications présentes sur les séquences des protéines.

V. Matériels et méthodes

1. Echantillons d'ossements analysés

L'optimisation du protocole d'extraction a été réalisée à partir d'un échantillon d'ossement de crâne humain frais, appelé « Témoin1 » (figure 4), fourni dans le cadre d'une collaboration avec l'unité de Taphonomie Médico-légale (EA 7367, Lille 2) dirigée par le Professeur Valéry Hédouin.



Figure 4 : Échantillon de crâne humain utilisé pour les tests de mise au point

Les échantillons archéologiques étudiés sont des ossements qui nous ont été confiés par le Professeur Christine Keyser (laboratoire d'anthropologie moléculaire de l'Université de Strasbourg).

L'échantillon EG26 provient d'un os longs (fémurs) prélevé sur un individu datant de la période Xiongnu (3ème siècle av. J.-C. - 2ème siècle de notre ère) (figure 5 a) et exhumés du site d'Egyin Gol situé dans le Nord de la Mongolie [50,51]. Le second échantillon YAKa12, provenant également d'une diaphyse fémorale prélevé sur un individu exhumé de Yakoutie centrale (Sibérie) et datant du 15ème siècle (figure 5 b) [52,53].

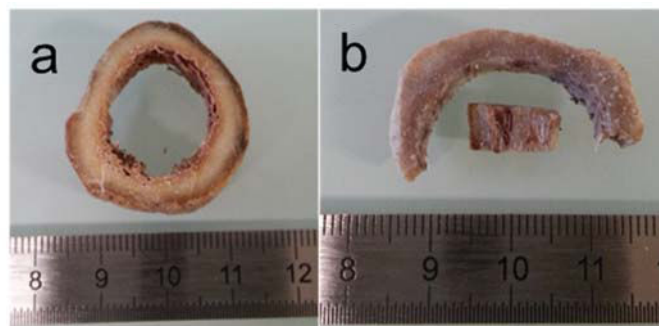


Figure 5 : Echantillon EG26 (a) et YAKa12 (b).

2. Optimisation de la composition du tampon de déminéralisation

La composition du tampon de déminéralisation a été optimisée. En particulier, les tampons suivants ont été utilisés : l'HCl (0,6 N), NaOH (1 N), des solutions contenant différents pourcentages de TFA (0,5, 1, 2,5, 5 et 10%). L'effet des tampons sur la composition chimique a été étudié à partir d'une protéine témoin, le lysozyme. Cette protéine a été préparée à une concentration de 1 mg/mL (69,88 pmol/ μ L) dans différents tampons.

Un volume de 200 μ L a été utilisé pour réaliser l'optimisation. Un témoin a été placé dans le même volume d'eau milliQ.

Les solutions ont été placées à 4 °C avec agitation pendant 24 heures, puis ces solutions ont été diluées avec une solution contenant 0,1% de TFA pour obtenir une solution à 10 pmol/ μ L. L'analyse a été réalisée avec le spectromètre de masse Apex Qe FT-ICR 9.4 T (Bruker Daltonics) en injection directe avec une seringue à un débit de 10 μ L/h et d'une aiguille d'infusion PicoTip Emitter de diamètre 10 +/- 1 μ m (MS wil GmbH).

Un voltage de 1600 V a été appliqué sur l'aiguille dans la source nanoESI. Une acquisition de 100 scans a été réalisée avec une résolution de 1 M points.

Ces tampons ont été utilisés sur une côte broyée à l'aide d'un mortier pour obtenir une fine poudre. Ensuite, la poudre d'os a été transférée dans un eppendorf de 2 mL puis différents tampons de déminéralisation ont été ajoutés comme précédemment. Des photos ont été réalisées pour comparer la déminéralisation sur la poudre d'ossement.

3. Analyse protéomique des ossements

Les ossements ont été broyés à l'aide d'un mortier pour obtenir une fine poudre. Ensuite, une quantité de 10 mg de cette poudre d'os a été transférée dans un eppendorf de 2 mL puis 1 mL de solution de déminéralisation (eau avec 5% TFA) a été ajouté.

La solution a été mise sous agitation 24 heures à 4 °C, puis centrifugée 10 minutes à 10 000 g à température ambiante. La solution de déminéralisation récupérée a été évaporée au Speed Vac à température ambiante.

La poudre a été lavée avec 1 mL d'eau milliQ puis placée sous agitation 10 minutes à 4°C. La solution a été centrifugée 5 minutes à 10 000 g, puis le surnageant a été éliminé et la poudre a été nettoyée une deuxième fois avec l'eau milliQ.

Ensuite, une étape de délipidation a été réalisée : 100 µL d'eau milliQ ont été ajoutés sur la poudre puis 900 µL d'une solution chloroforme / méthanol (2/1) ont été ajoutés. L'échantillon a été agité 1 heure à 4 °C puis centrifugé 10 min à 10 000 g. La phase inférieure (chloroforme) qui contient les lipides a été éliminée. La poudre a été lavée 2 fois à l'eau milliQ. Cette poudre a été ensuite évaporée au speed vac pour éliminer l'eau milliQ. La poudre d'os et la solution de déminéralisation évaporée ont été reprises dans 200 µL de tampon de lyse (8 M d'urée, 4% SDS, 0,2% DCA, 50 mM DTT, 100 mM ammonium bicarbonate pH 8,8) puis incubés avec agitation à 4°C pour l'étape de digestion eFASP.

4. Digestion des échantillons (méthode eFASP)

Les filtres Amicons[®], avec un seuil de coupure de 10 KDa, ont été incubés toute la nuit dans la solution de passivation contenant 5% (v/v) de TWEEN[®]-20. L'échantillon a été transféré dans le filtre Amicons[®] puis 100 µL de tampon d'échange (8 M d'urée, 0,2% DCA, 100 mM ammonium bicarbonate pH 8,8) ont été ajoutés dans l'Eppendorf et l'ensemble a été transféré dans l'Amicon[®] puis centrifugé 30 minutes à 10 000 g, le filtrat a été éliminé. 200 µL de tampon d'échange ont été de nouveau ajoutés dans l'Amicon[®] et centrifugés, l'opération a été répétée deux fois. Les protéines ont été alkylées pendant une heure à température dans le noir à l'aide de 100 µL de tampon d'alkylation (8 M urée, 50 mM iodoacétamide, et 100 mM ammonium bicarbonate, pH 8,8) puis centrifugé 30 minutes à 10 000 g, le filtrat a été éliminé. Après l'alkylation des protéines, 200 µL de tampon d'échange ont été ajoutés à l'Amicon[®], centrifugé, le filtrat a été éliminé. 200 µL de tampon de digestion (0,2% DCA, 50 mM ammonium bicarbonate pH 8,8) ont été ajoutés à l'Amicon[®] puis centrifugé. Cette étape a été répétée deux fois, en éliminant le filtrat à chaque étape. L'Amicon[®] a été transféré dans un nouveau tube de 2 mL. 100 µL de tampon de digestion et 40 µL de trypsine (trypsine : protéine 1/50 w : w) ont été ajoutés et incubés toute la nuit à 37 °C. Après cette étape, les peptides présents dans l'Amicon[®] ont été récupérés dans le tube grâce à une centrifugation de 15 min à 10 000 g, l'Amicon[®] a été lavé 2 fois avec 50 µL d'ammonium bicarbonate (50 mM pH 8,8). Le filtrat contenant tous les peptides a été transféré dans un Eppendorf de 1,5 mL. 200 µL d'acétate d'éthyle et 2,5 µL de TFA ont été ajoutés (un précipité blanc apparaît). 800 µL d'acétate d'éthyle ont été de nouveau ajoutés puis l'ensemble a été centrifugé 10 minutes à 10 000 g. La phase organique a été éliminée. Cette étape a

été répétée deux fois. Les Eppendorfs ont été placés 5 minutes à 60°C pour éliminer l'acétate d'éthyle restant. Les échantillons ont été évaporés à sec, puis 100 µL d'un mélange méthanol/eau (50/50) ont été ajoutés puis de nouveau évaporés à sec. Les échantillons ont été analysés par spectrométrie de masse (figure 6).

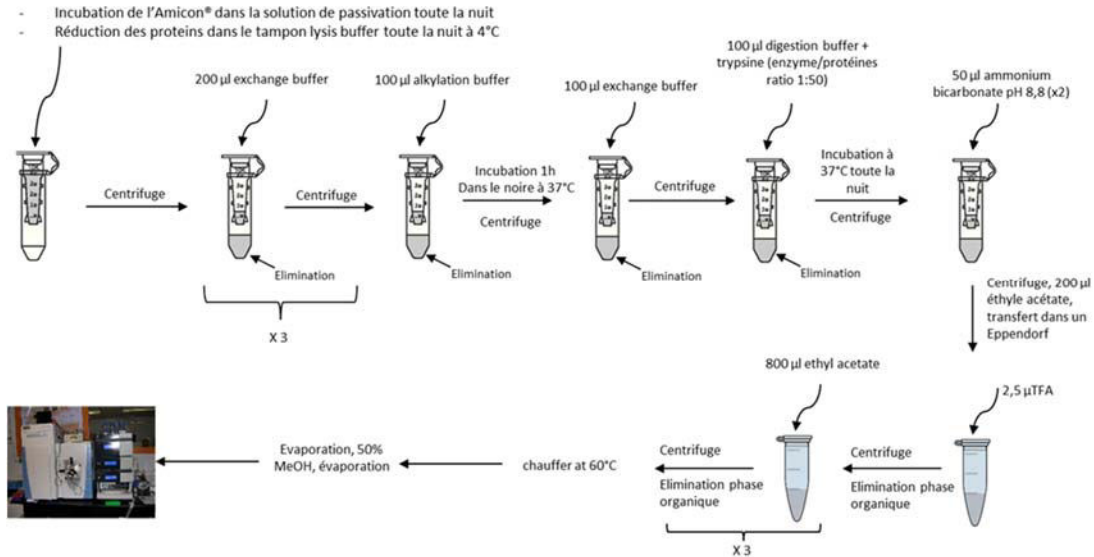


Figure 6 : Schéma de la méthode eFASP.

5. Couplage LC-MS/MS Orbitrap.

Les échantillons ont été repris dans 10 µl de solution H₂O contenant 0,1% d'acide formique. Le spectromètre de masse Orbitrap Q Exactive plus utilise un système microfluidique HPLC U3000 RSLC Thermo Fisher Scientific. Une pré-colonne Thermo Scientific, Acclaim PepMap100 C18 (5 µm, 300 µm i.d x 5 mm) a été utilisée, dans laquelle a été injecté 1 µl du mélange de peptides avec un solvant A : (5% acétonitrile et 0,1% acide formique) pendant 3 min à un débit de 10 µl.min⁻¹. Les peptides ont été ensuite séparés sur une colonne à phase inverse C18 Acclaim PepMap100 C18 (3 µm, 75 mm i.d. x 500 mm), grâce à un gradient linéaire (5-40%) de solution B (75% ACN et 0,1% acide formique) à un débit de 250 nL/min en 160 minutes puis 100% de solution B en 5 minutes. La colonne a été lavée durant 5 minutes avec le tampon B puis rééquilibré avec le tampon A. La colonne et la pré-colonne se trouvaient à une température de 45°C. La durée totale de l'analyse était de 180 minutes.

Les paramètres de l'Orbitrap Q-Exactive les plus utilisés pour l'analyse MS et MS sont décrits ci-dessous.

	MS	MS/MS
Résolution	70 000	35 000
AGC target	1e6	5e5
Maximum IT	80 ms	160 ms
Scan range	350 à 1600 m/z	

Fragmentation HCD à 28ev, exclusion des charges de 1, 8 et plus.

6. Analyse Bio-informatique des extractions

L'analyse de comparaison des extractions des protéines de la solution de déminéralisation et de la poudre d'os a été réalisée avec le logiciel Maxquant 1.5.5.0. Les données de l'Orbitrap ont été analysées avec le software contre la base de données SwissProt_Human_2017 (*Homo sapiens*, composée de 71 785 protéines) additionnée d'une liste de protéines contaminantes contenant 115 protéines (protéines contaminantes classiquement retrouvées lors des analyses par spectrométrie de masse). L'analyse a été réalisée en utilisant une erreur de masse en MS et MS/MS de 10 ppm. Les modifications fixes des peptides recherchés sont la carbamydométhylation des cystéines. Les modifications variables recherchées sur les peptides sont l'oxydation des méthionines et des prolines, la déamidation des arginines et glutamines. La trypsine a été utilisée comme enzyme de digestion avec 3 sites de coupure manqués. Seules les protéines identifiées avec un seuil de FDR de 0.01% ont été analysées. L'analyse statistique a été réalisée avec le logiciel Perseus 1.5.6.0. Classiquement, les protéines contaminantes ont été retirées puis les intensités ont été transformées en Log_2 . Les valeurs non attribués (les valeurs avec une intensité de 0 qui après transformation en Log_2 ne peuvent pas donner de chiffres) sont remplacées par des valeurs suivant une loi normale. La comparaison des groupes a été réalisée avec un t-test avec une correction de Benjamini Hochberg et un FDR (False Discovery Rate) de 0,05%. Un test ANOVA a été appliqué sur tous les groupes pour identifier les protéines ayant une expression significative. Les données ont été normalisées avec en utilisant le z-score (équation 1).

Equation 1 : $z - score = \frac{x - \mu}{\sigma}$, où σ = déviation standard de la protéine, μ = moyenne de la protéine, x = valeur d'une protéine pour un échantillon.

Une analyse en composant principale est réalisée sur les protéines ayant une expression différentielle significative à partir des résultats du test statistique ANOVA.

Puis une carte d'expression est réalisée en utilisant un regroupement hiérarchique euclidien.

L'analyse biologique des protéines identifiées a été réalisée avec le logiciel STRAP 2.0 qui permet d'obtenir les informations sur leurs localisations cellulaires, leurs fonctions, leurs activités.

7. Analyse bio-informatique des ossements archéologique

Pour l'analyse de l'ossement ancien, les fichiers obtenus à partir de l'Orbitrap ont été analysés avec le logiciel PEAKS 7.0 (Bioinformatics Solutions Inc.). Les fichiers ont été interrogés sur toute la base de données NCBI All_Entries (composée de 88 385 530 protéines à la date du 13 juin 2017) avec PEAKS DB. L'analyse utilise une erreur de masse de +/- 10 ppm pour la MS et 0,01 Da pour la MS/MS. Les mêmes modifications que précédemment ont été utilisées pour la recherche sur la base de données. La trypsine a été utilisée comme enzyme de digestion avec 3 sites de coupure manqués. Le logiciel PEAKS a réalisé une analyse *de novo* qui consiste à interpréter directement chaque spectre de fragmentation MS/MS et à le traduire en séquence d'acides aminés. Les séquences obtenues, bien qu'incomplètes, ont été utilisées pour identifier les protéines par homologie de séquence. Le logiciel PEAKS a utilisé, une analyse avec l'algorithme SPIDER, qui permet de faire correspondre les séquences des peptides, obtenues avec l'analyse *de novo*, avec les protéines de la base de données NCBI utilisée. Lorsqu'une similitude significative a été trouvée, l'algorithme a tenté d'utiliser à la fois les erreurs de séquençage *de novo* et les mutations de peptides homologues pour expliquer les différences. L'identification des protéines a été validée à partir de deux peptides uniques et d'un score PEAKS – $10 \log P > 20$.

8. Phylogénie des protéines

Pour observer les variations de séquence entre les espèces, une analyse de phylogénie a été réalisée sur les collagènes 1 de type alpha 1 et alpha 2 de 14 séquences. Le fichier fasta contenant les séquences des collagènes est en annexe. Les différentes espèces sont : *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Bos taurus*, *Gallus gallus*, *Canis lupus familiaris*, *Pan troglodytes*, *Pongo abelii*, *Ailuropoda melanoleuca*, *Macaca mulatta*, *Gorilla gorilla gorilla*, *Papio anubis*, *Equus caballus*, *Homo sapiens neanderthalensis*. Les outils en ligne du site <http://phylogeny.lirmm.fr/phylo.cgi/index.cgi> ont été utilisés dans l'ordre suivant :

Analyse bottom-up des échantillons archéologiques et paléontologiques

- Alignement de séquence avec MUSCLE 3.7
- Mesure des distances avec PhyML 3.0
- Création de l'arbre phylogénique à l'aide de TreeDyn 198.3.

VI. Analyse des résultats sur les ossements

Comme démontré précédemment, l'analyse protéomique est récemment devenue l'outil analytique de choix pour l'étude des protéines anciennes. Pour l'analyse archéologique et paléontologique, les ossements représentent des échantillons d'intérêt car ils se conservent des millions d'années et les protéines contenues se dégradent lentement car elles sont incluses dans des matrices minérales.

Notre objectif est de proposer une méthode d'analyse d'ossements anciens reproductible, sensible et précise dans le but de diminuer la quantité d'échantillons, d'améliorer l'identification des protéines et des espèces.

1. Mise au point du tampon de déminéralisation

La première étape a consisté en l'optimisation de la déminéralisation des ossements pour éviter l'apparition de modifications chimiques.

Classiquement, cette étape consiste à mettre en présence la poudre d'ossement (le plus souvent broyé à l'aide d'un mortier) en présence d'une solution acide comme l'acide chlorhydrique (HCl) ou basique comme l'hydroxyde de sodium (NaOH) [54,55]. Cette étape permet de libérer les protéines faiblement liées à l'os et d'éliminer les constituants inorganiques. Quand la solution d'HCl est utilisée, l'hydroxyapatite de calcium constituant l'os réagit avec l'acide formant du mono calcium de phosphate et du chlorure de calcium.

Dans le cadre de ce travail, nous avons développé une méthode de déminéralisation utilisant un autre acide, l'acide trifluoroacétique (TFA). Cette première étude a consisté à mettre en présence le lysozyme, protéine témoin, en présence de d'HCl (0,6 M), NaOH (1M), de l'eau mQ ou de TFA (0,1, 0,5, 1, 2,5, 5 et 10%) pendant 24 heures à 4 °C sous agitation. Ensuite, une analyse en protéines entière a été réalisée injection directe sur le FT-ICR (Bruker Daltonics) en nano ESI.

Un massif multichargé de 8 à 15 charges est détecté pour le lysozyme témoin dans l'eau mQ (figure 7). La masse du lysozyme issue de la déconvolution du spectre est de 14 304,80 Da et peu de modifications ont été détectées.

L'étape de déminéralisation a été réalisée à 4 °C pour éviter les modifications telles que des oxydations ou des déamidations. Les spectres du lysozyme dans le tampon TFA 0,1, 0,5, 1% en TFA sont similaires au spectre témoin dans l'eau mQ. Par contre, pour les spectres 2,5 et 5% de faibles différences sont observées

(intensité des modifications), comparé au lysozyme dans les solutions NaOH, HCl et TFA 10% qui présentent de nombreuses modifications avec des intensités élevées.

La figure 8 présente les spectres du massif multichargé 10^+ à m/z 1431,3748 du lysozyme. Cette masse correspond à la formule brute $C_{613}H_{951}N_{193}O_{185}S_{10}$. Sur le spectre du lysozyme dans l'eau mQ, différentes masses ont été détectées : la masse du lysozyme natif à m/z 1431,3748 ($\Delta m < 1$ ppm), la perte d'une molécule d'eau à m/z 1429,6374 ($C_{613}H_{949}N_{193}O_{184}S_{10}$; $\Delta m < 1$ ppm). Des massifs à m/z 1432,9745 ($C_{613}H_{949}N_{193}O_{185}S_{10}$) et m/z 1434,5748 ($C_{613}H_{949}N_{193}O_{186}S_{10}$), correspondent à l'addition d'un ou deux oxygènes sur le lysozyme.

Comme dit précédemment le lysozyme dans les solutions TFA 0,1, 0,5% sont peu différents du témoin. Pour la solution 1, 2,5 et 5% des massifs d'oxydations sont détectés mais leur intensité est faible.

Les spectres du lysozyme dans la solution HCl, NaOH et TFA 10% montrent la présence de nombreuses formes d'oxydations avec des intensités très élevées

Analyse bottom-up des échantillons archéologiques et paléontologiques

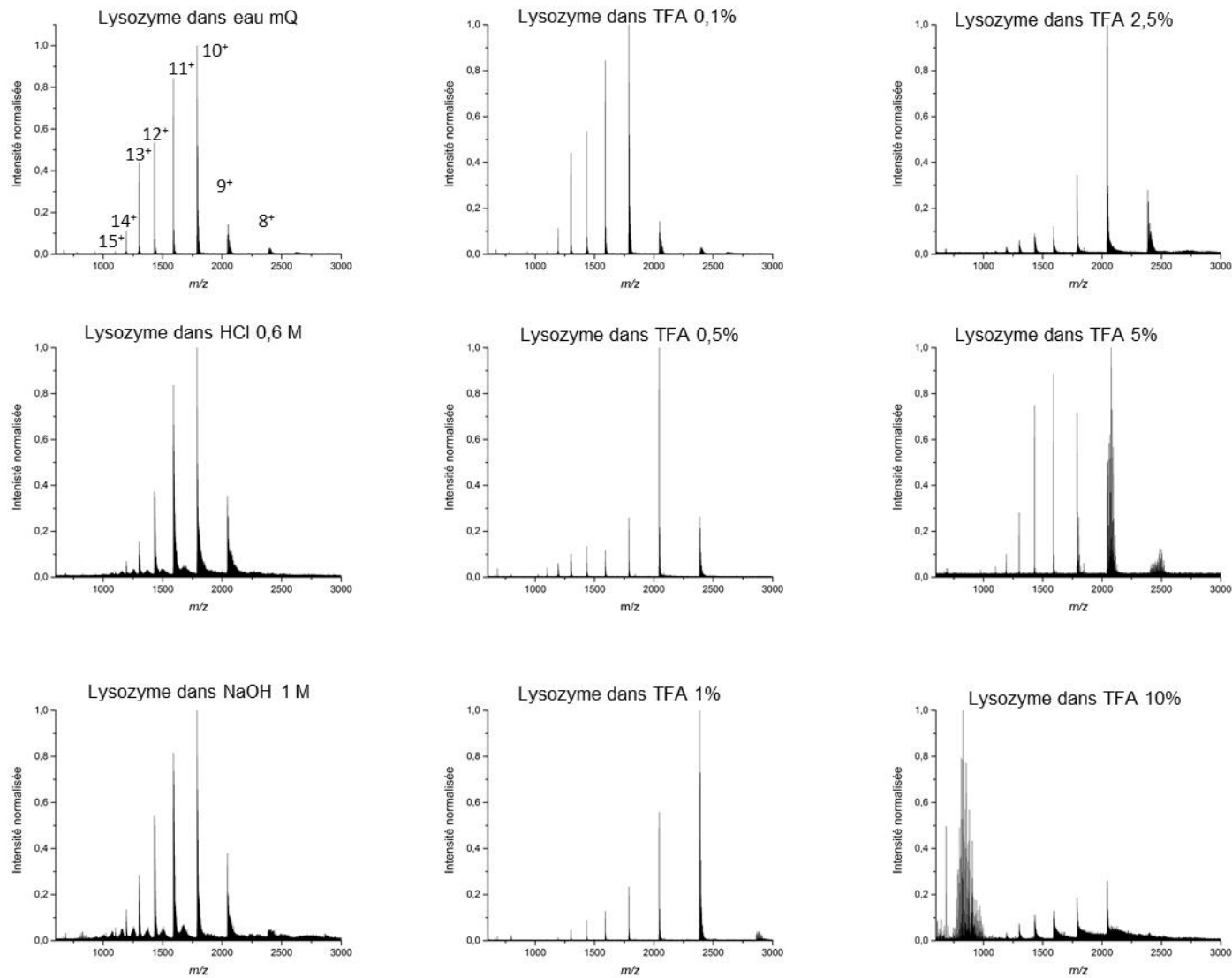


Figure 7 : Spectres du lysozyme à 10 pmol / μL . analyse par nanoESI Qh FTICR 9.4T dans les différentes solutions de déminéralisation. Sur le spectre du lysozyme témoin dans l'eau mQ, les états de charges sont annotés de 15 à 8.

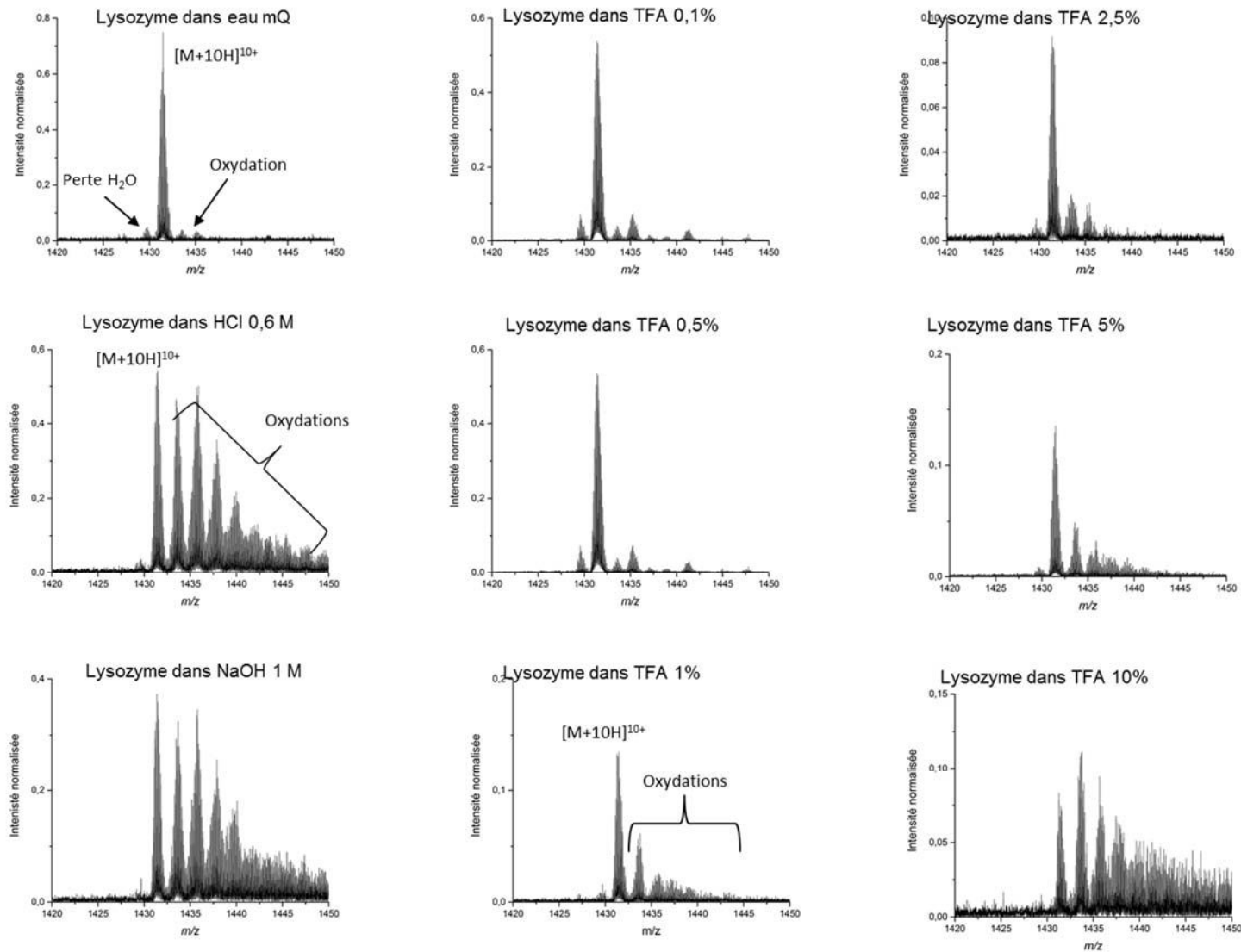


Figure 8 : Zoom sur le massif du lysozyme $[M+10H]^{10+}$ pour chaque solution de déminéralisation. Observation de perte d'eau et d'oxydations.

Cette première étape montre que la solution de TFA jusque 5% induit peu de modifications de type oxydations chimiques sur le lysozyme comparé aux solutions d'HCl et de NaOH qui sont les tampons de déminéralisation les plus utilisés dans les études de protéomique sur les ossements archéologiques [37,39].

2. Déminéralisation des ossements

Les solutions ont été testées sur 10 mg de poudre d'ossements dans les mêmes conditions (une nuit à 4 °C avec agitation).

La première observation est que la masse d'ossement présente dans le fond du tube diminue avec le pourcentage de TFA. Après récupération de cette poudre, élimination du surnageant et séchage, les poudres ont été pesées.

Les échantillons issus des déminéralisations par le TFA 5% et l'HCl 0,6 M possèdent la même quantité de poudre après déminéralisation (tableau 1) tandis que la déminéralisation de l'ossement avec la solution TFA 10% est plus faible. Cette mesure de masse résiduelle de poudre d'ossement après déminéralisation montre que la solution TFA 5% permet une diminution de la masse d'ossement similaire au protocole classique avec l'HCl 0,6 M.

	Eau mQ	0,6 M HCl	1 M NaOH	0,1% TFA	0,5% TFA	1% TFA	2,5% TFA	5% TFA	10% TFA
quantité en mg restante de poudre d'os	49,5	36	30	49,5	48	45	39	35	20

Tableau 1 : Quantité en milligramme de poudre d'os restant après déminéralisation.

La figure 9 montre l'effet des solutions de déminéralisation sur la poudre d'ossement. La solution TFA à 10% présente une couleur différente de celle des autres solutions de déminéralisation, de même que la solution NaOH. En utilisant la solution HCl 0,6 M comme référence, les solutions à base de TFA sont similaires à cette solution.



Figure 9 : Photographies de tubes contenant la poudre d'ossement avec différentes solutions de déminéralisation.

Habituellement, la déminéralisation est réalisée avec des solutions contenant de l'HCl, NaOH, EDTA [55]. La solution TFA à 5% permet d'éviter la formation de modifications chimiques sur les protéines d'ossement comme expliqué dans le paragraphe VI-1. La solution HCl 0,6 M et TFA à 5% permettent une diminution de la masse de la poudre d'ossement similaire. Ainsi, nous utiliserons donc la solution TFA à 5% comme solution de déminéralisation pour la suite de l'étude. Cette solution de déminéralisation est une alternative aux tampons actuels.

3. Analyse protéomique

Après avoir optimisé cette étape de déminéralisation, nous avons effectué une digestion eFASP [56] de la solution de déminéralisation et la poudre d'ossement provenant de l'ossement témoin « Témoin1 », après traitement au TFA 0,5%, 5%, HCl 0,6M et eau milliQ pour observer l'influence de la solution acide et observer la différence en les protéines détectées dans la solution de déminéralisation et dans la poudre.

La première étape de la méthode eFASP est la dénaturation et la réduction des protéines. Cette étape a été réalisée à 4 °C pour éviter la formation de modifications sur les protéines.

La digestion eFASP a été utilisée car elle possède différents avantages :

- L'utilisation de filtre pour permettre d'éliminer les sels, les détergents ou autres molécules. L'échantillon est ainsi solubilisé dans un tampon comportant des agents réducteurs, chaotropes et des détergents car ceux-ci seront éliminés avec les lavages.

- L'incubation des filtres Amicon® dans une solution de tween 5% pour permettre de diminuer la fixation aspécifique des protéines sur la membrane.
- L'incubation des échantillons dans une solution de lyse contenant du SDS, de l'urée, du déoxycholate permet une dénaturation optimale. L'addition de réducteur comme le dithiothréitol réduit les ponts disulfures des protéines.
- L'amélioration de la digestion due à l'utilisation de déoxycholate dans le tampon de digestion qui permet une meilleure solubilisation des protéines et peptides.
- Cette méthode permet d'utiliser une combinaison d'enzymes de digestion en une fois ou en série. Le filtre permet de récupérer de façon séquentielle les peptides après chaque digestion.

Les données de spectrométrie de masse obtenues à partir de toutes les fractions ont été analysées à en utilisant différentes informations :

- Le nombre de peptides identifiés à partir des spectres pour une protéine (PSM) et les peptides identifiés.
- L'analyse du collagène présent dans les ossements.
- L'analyse en composante principale.
- Le regroupement hiérarchique euclidien.
- L'analyse des réseaux protéines - protéines.

a) Etude des tampons de déminéralisation sur les protéines identifiées dans la solution de déminéralisation et dans la poudre d'ossement.

L'identification des protéines a été réalisée avec le logiciel Maxquant en utilisant la base de données *Homo sapiens*. Le logiciel Maxquant est un logiciel qui permet l'identification des protéines en utilisant des bases de données et aussi la quantification des protéines entre les échantillons [57]. Les données obtenues avec Maxquant ont ensuite été analysées avec le logiciel Perseus qui permet de filtrer les données (élimination contaminants, protéines sans valeurs), d'ajouter les informations de fonctions des protéines, de réaliser des tests statistiques (t-test, ANOVA), ainsi que des analyses ACP et des cartes d'expression [58].

La figure 10 expose le nombre de protéines, de peptides et de PSM (peptides identifiés à partir des spectres pour une protéine) identifiés avec le logiciel Maxquant, pour les 4 tampons de déminéralisation différents pour la fraction acide ou fraction

déminéralisation et la fraction de poudre d'ossement obtenus sur l'échantillon témoin frais « Témoin1 ».

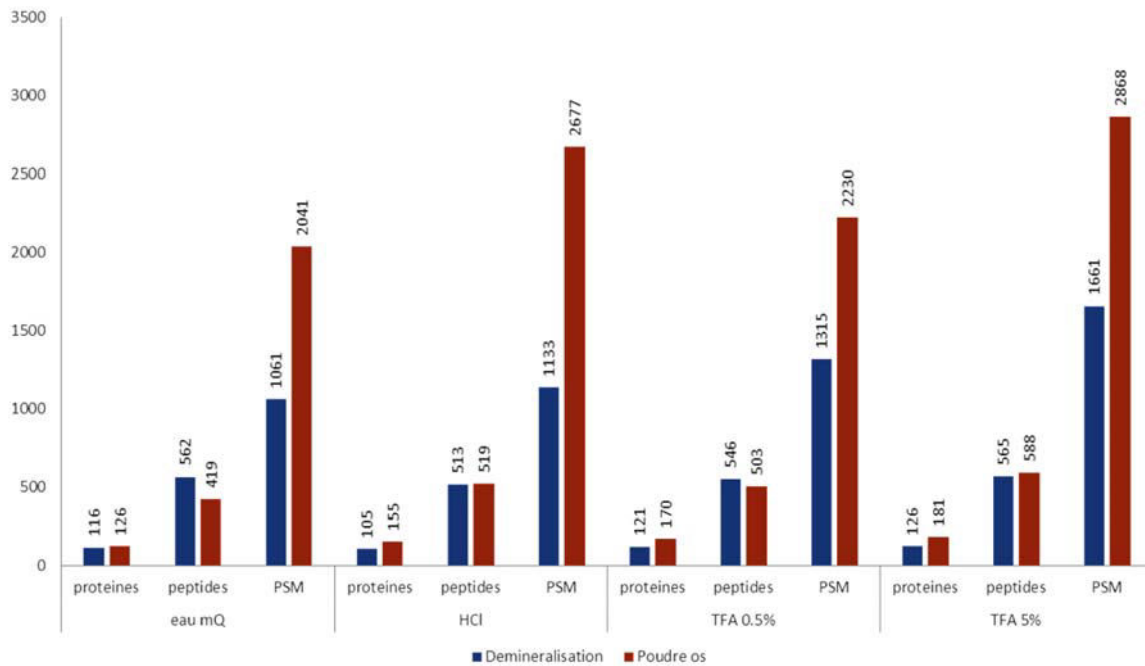


Figure 10 : Nombre de protéines, peptides et PSM identifiés avec les 4 tampons eau mQ, HCl, TFA 0,5% et TFA 5%. Pour la fraction acide ou déminéralisation (en bleu) et la fraction de poudre d'ossement (en orange).

Le tampon TFA à 5% permet l'identification du plus grand nombre de protéines, peptides et PSM pour la fraction déminéralisation et poudre d'os, par rapport aux autres tampons de déminéralisation. Par exemple, pour la solution TFA à 5%, 181 et 126 protéines ont été identifiées dans la fraction déminéralisation et poudre d'os comparé aux 116 et 126 protéines identifiées pour le tampon eau mQ.

Le tissu osseux est constitué d'eau (environ 1/4 du poids de l'os), de matières organiques (environ 1/3 du poids de l'os, dont la majeure partie est représentée par des protéines) et de sels minéraux inorganiques (le calcium, le phosphore, et le magnésium prédominant, bien que l'on trouve également du fer, du sodium, du potassium, du chlore, et du fluor en petites quantités). En se focalisant sur le collagène 1 chaîne A1, une protéine d'environ 300 kDa, des différences sont observées en fonction des tampons utilisés (figure 11).

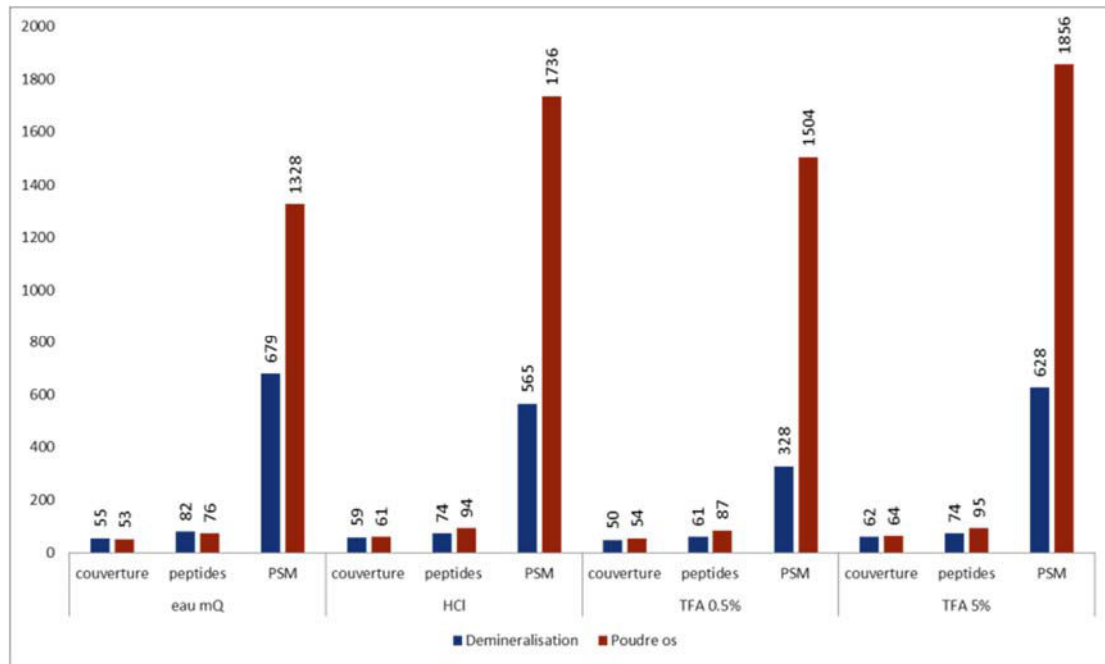


Figure 11 : Analyse du COL1A1 dans la fraction acide (bleu) et la fraction de poudre d'ossemment (orange) pour les quatre tampons de déminéralisation.

La figure 11 montre la couverture de séquence pour le collagène 1 alpha1, le nombre de peptides et le nombre de PSM (peptides identifiés à partir des spectres pour une protéine) identifiés, pour les 4 tampons de déminéralisation et pour la fraction acide (ou fraction de déminéralisation) et la fraction de poudre d'ossemment obtenus sur l'échantillon témoin frais « Témoin1 ».

Le tampon de déminéralisation au TFA à 5% permet une identification du collagène avec 60% de couverture, grâce à l'identification de 74 à 95 peptides et de 628 à 1856 PSM pour la fraction acide et poudre d'ossemment. Les autres tampons ont des nombres de peptides, de PSM inférieurs ainsi que des couvertures de séquences plus faibles.

Ces résultats permettent de montrer que la solution de déminéralisation avec la solution de TFA à 5% permet une meilleure identification des protéines.

b) Analyse quantitative des protéines après déminéralisation.

L'Analyse en Composantes Principales (ACP) est une méthode statistique exploratoire permettant une description essentiellement graphique de l'information contenue dans des grands tableaux de données [59]. L'ACP est une méthode très efficace pour représenter des données corrélées entre elles. Elle condense l'information contenue dans un jeu de données immense en un petit nombre de facteurs. Ces facteurs permettent d'expliquer la majorité des variances observées.

Une composante principale (CP) est une combinaison linéaire calculée en fonction des variables existantes (protéines). Concrètement, si des protéines possèdent des valeurs élevées dans une CP spécifique, ceci implique que ces protéines expliquent une part importante de la variance de l'échantillon.

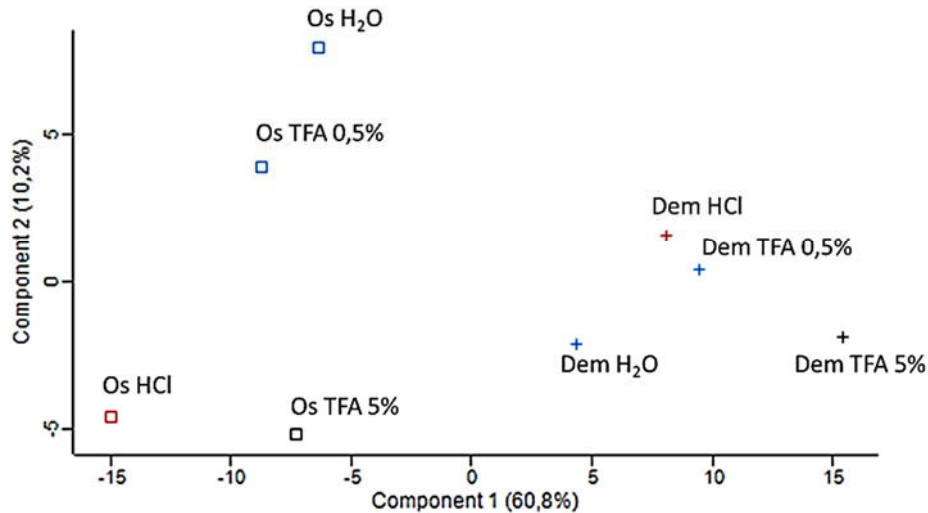


Figure 12 : Analyse en composante principale des différents tampons pour les protéines identifiées et quantifiées dans la solution de déminéralisation (Dem) et la poudre d'ossement (Os). Les croix bleues correspondent au tampon eau mQ, carré rouge : HCl, rond vert : TFA 0,5%, triangle noir : TFA 5%.

L'analyse ACP (figure 12) réalisée avec le logiciel Perseus, montre une séparation de la fraction déminéralisation et poudre d'ossement via l'axe de la composante 1 (60,8%). Seules les protéines ayant une différence d'expression significative (test ANOVA avec $p < 0,05$) entre les groupes sont utilisées pour réaliser l'ACP. L'analyse ACP sépare les types de tampons selon l'axe de la composant 2 (10,2%). Les protéines identifiées dans les différents tampons permettent de distinguer les deux types de fractions (poudre et déminéralisation).

La carte d'expression (figure 13) montre l'expression des protéines dans les échantillons HCl, eau, TFA 0,5% et 5% pour les fractions déminéralisation et poudre d'ossement. En rouge sont représentées les protéines les plus exprimées ou dont la quantité est la plus élevée dans la fraction, et en vert sont représentées les protéines les moins présentes. Le regroupement hiérarchique euclidien permet de regrouper les échantillons en fonction de leur ressemblance.

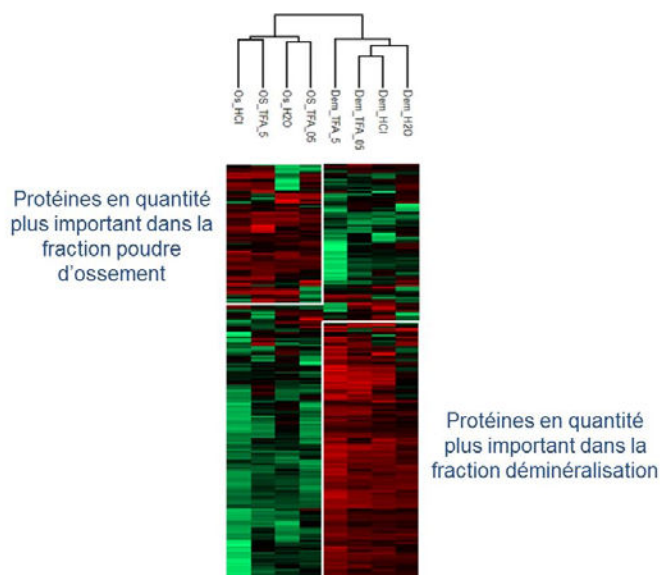


Figure 13 : Carte d'expression des protéines réalisées à partir d'un regroupement hiérarchique euclidien pour la fraction déminéralisation (Dem) et poudre d'ossement (Os). On distingue la formation de groupe de protéines spécifique de certains tampons. En rouge les protéines les plus présente et en vert celle qui sont en quantité moins élevé. Le regroupement hiérarchique euclidien sépare les deux fractions et une différence d'expression des protéines est observée pour les tampons utilisés.

Dans la fraction déminéralisation, 142 protéines possèdent une expression différentielle significative ($p > 0,05$). Ces protéines correspondent aux contaminants de type kératine, mais également à des protéines spécifiques des os comme l'ostéonectine, la lumican ou la sialoprotéine. D'autres protéines ont été identifiées, provenant du plasma, avec des fonctions dans le développement des os et d'autres avec une fonction de coagulation comme l'albumine, transferrine, fibrinogène ou impliqué dans le transport de lipoprotéines (APOA1, fetuin).

La carte d'expression pour la poudre d'ossement montre un groupe de 59 protéines ayant une expression différentielle significative de couleur rouge qui indique que celles-ci sont présentes en plus grande quantité dans cette fraction. Elles correspondent aux protéines collagéniques (différents types de collagène), à l'ostéocalcine.... Ce groupe est composé de protéines de l'os et avec une fonction de liaison.

Une analyse des réseaux de protéines fonctionnelles est réalisée avec le logiciel STRING [60] (« Search Tool for the Retrieval of Interacting Genes/Proteins », version 10.5). Le logiciel prédit les interactions protéines – protéines en utilisant les informations de génomique, les fonctions des protéines.

Les résultats de l'analyse STRING (figure 14) montrent que les différentes fractions possèdent des protéines spécifiques. La figure 14 a et b représentent les

protéines identifiées dans les différentes fractions (142 protéines pour la fraction déminéralisation et 59 protéines pour la fraction de poudre d'ossement) ayant une différence significative ($p > 0,05$). Les protéines ayant une ressemblance sont très proches les unes des autres. Ceci se voit très clairement sur la figure 14b où les protéines collagéniques sont regroupées au centre. Ces protéines correspondent aux différents types collagènes (exemple : collagène 1 alpha 1, 2, collagène 2 alpha 1...).

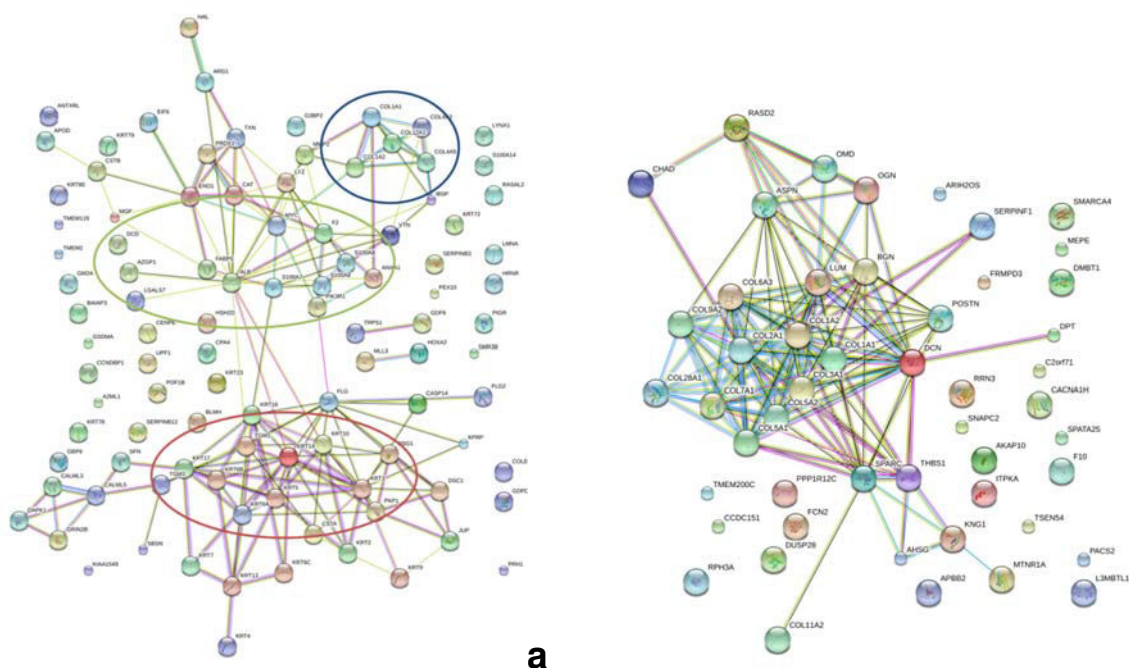


Figure 14 : Analyse avec le logiciel STRING des protéines présentes en plus grandes quantité dans la fraction déminéralisation (a) et la fraction de poudre d'ossement (b).

Pour la fraction déminéralisation, trois groupes peuvent être observés sur le diagramme de STRING (figure 14 a). En bleu, un premier groupe correspondant aux collagènes, en rouge celui des contaminants comme les kératines et en vert, les protéines plasmatiques comme l'albumine, la vimentine.

Pour conclure sur la méthode de préparation des ossements et leurs analyses, il a été montré que la solution TFA 5% permet d'éviter la formation de modifications chimiques sur les protéines.

La méthode de digestion eFASP permet une identification des protéines présentes dans les différentes fractions : déminéralisation et poudre d'os. La quantité analysée par digestion eFASP est de 10 mg, mais cette quantité peut être diminuée pour les ossements frais car les protéines ne sont pas encore dégradées comme c'est le cas dans des échantillons archéologiques.

Les protéines identifiées correspondent bien à un protéome d'ossement [61] car les protéines comme le collagène, l'ostéocalcine, le lumican ont été identifiées. Des protéines du plasma ont été identifiées dans la fraction déminéralisation car l'ossement analysé a été conservée à -80°C et que cet ossement n'était pas ancien. Les protéines contenues dans l'échantillon ont donc été correctement conservées.

Les analyses statistiques montrent une différence des protéines entre la fraction de poudre d'os et déminéralisation. Cette différence est aussi retrouvée dans de nombreuses études sur les ossements anciens ou actuels [55,61,62].

4. Analyse des échantillons anciens

La méthodologie mise au point précédemment a été appliquée à l'analyse des échantillons archéologiques qui nous ont été confié par le Professeur Christine Keyser (laboratoire d'anthropologie moléculaire de l'Université de Strasbourg). L'échantillon EG26 provenant d'un os longs (fémurs) prélevé sur un individu datant de la période Xiongnu (3ème siècle av. J.-C. - 2ème siècle de notre ère) et exhumés du site d'Egyin Gol situé dans le Nord de la Mongolie. Le second échantillon YAKa12, provenant également d'une diaphyse fémorale prélevée sur un individu exhumé de Yakoutie centrale (Sibérie) et datant du 15ème siècle. Ces ossements archéologiques seront comparés à un ossement témoin, d'*Homo sapiens* datant de moins de 50 ans, que nous appellerons « Témoin1 »

L'objectif est de savoir si les ossements archéologiques contiennent toujours des protéines et de valider qu'il s'agit d'ossement de l'espèce *Homo sapiens*.

Dans la bibliographie, il a été montré la possibilité d'identifier différentes espèces avec une approche protéomique [34,41,54,63].

a) Analyse bio-informatique

Dix milligrammes d'ossement ont été broyés, puis ont été déminéralisés avec la solution de TFA 5%. La partie interne de l'ossement a été broyée pour limiter la contamination par des protéines extérieures.

Ensuite, chaque fraction a été digérée avec la méthode eFASP puis analysée en nano LC-MS/MS. Les fichiers Orbitrap ont été analysés sur le logiciel PEAKS pour identifier les protéines. Le tableau 1 regroupe les informations de MS, MS/MS et résultat de l'analyse PEAKS sur les ossements.

	fraction déminéralisation			fraction poudre d'os		
	EG26_Dem	YAKa12_Dem	Témoin1_Dem	EG26_Os	YAKa12_Os	Témoin1_Os
Nombre de scans MS	19207	16722	38007	20824	21229	38515
Nombre de scans MS/MS	8450	10771	22238	6899	6520	21134
Peptide-Spectrum Matches	2756	3712	4 404	2686	2930	4 846
Groupe de protéines	23	56	162	18	32	165
Protéines	51	131	312	49	89	221
Site de coupure manqué = 0	965	1246	1845	1021	1336	1503
Site de coupure manqué = 1	557	552	686	556	960	712
Site de coupure manqué = 2	181	43	129	153	359	210
Site de coupure manqué = 3	71	12	58	59	147	77

Tableau 2 : Tableau récapitulant les données des analyses LC-MS/MS avec le logiciel PEAK 7.0. Dénomination : Dem = déminéralisation et Os = poudre d'ossement.

La compilation des résultats de PEAKS montre que le nombre de MS et MS/MS pour l'échantillon Témoin1 sont plus importants que pour les deux échantillons archéologiques. Ensuite, dans l'échantillon témoin « Témoin1 » un nombre plus élevé de protéines est détecté dans la fraction acide et dans l'ossement par rapport aux deux autres échantillons. Cette différence de nombre de MS, MS/MS, protéines est due au fait que l'échantillon « Témoin1 » est plus récent et a été conservé dans de meilleures conditions.

Il faut noter l'analyse protéomique utilisée, permet d'identifier des protéines dans les échantillons archéologiques. Le nombre de protéines identifiées est plus faible dans les échantillons archéologiques, car les protéines ont été dégradées. Cette dégradation des protéines peut être due à la présence de bactéries dans les ossements, dans l'environnement mais aussi aux conditions de conservation (température, sols, luminosité, conditions climatiques).

b) Protéines présentes et contaminants

L'interrogation a été réalisée sur la base de données NCBI « All entries » contenant plus de 88 385 530 protéines de toutes les espèces. Quarante-vingt-dix pourcent des protéines identifiées sont des protéines de mammifères (figure 15). L'analyse des spectres montre une faible identification de protéines provenant

d'espèces non vertébrées, qui correspondent aux contaminations de l'environnement.

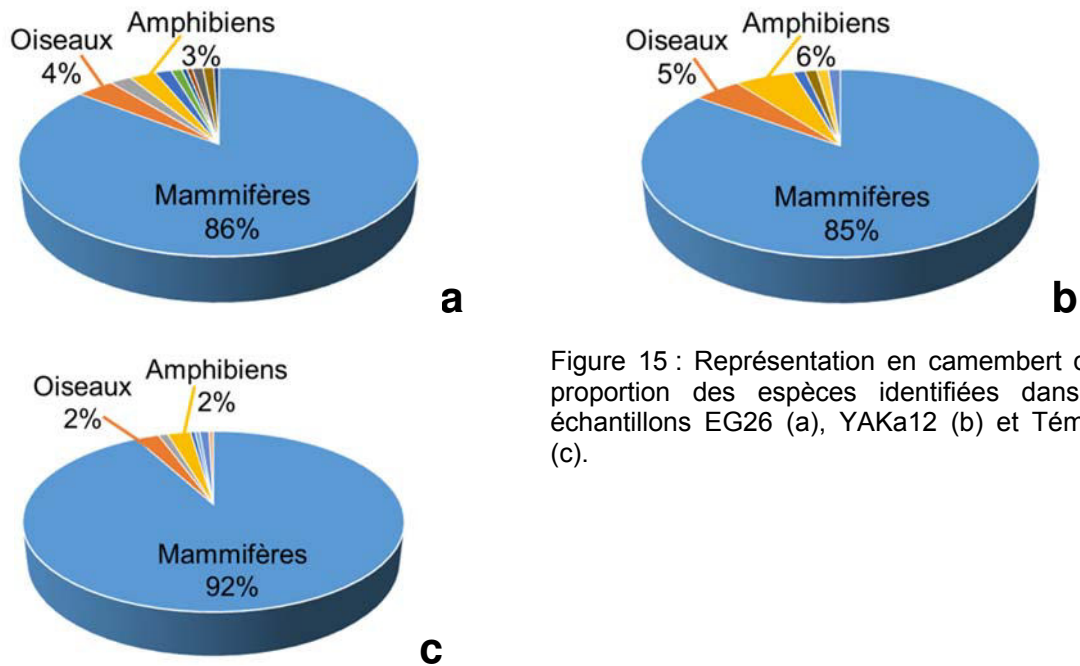


Figure 15 : Représentation en camembert de la proportion des espèces identifiées dans les échantillons EG26 (a), YAKa12 (b) et Témoin1 (c).

En examinant les protéines identifiées, plusieurs protéines contaminantes comme les protéines de la peau (les keratines, l'hornérine, les desmoplakines) et de la sueur (dermicine) sont détectées. Ces protéines sont présentes car les paléontologues manipulent les ossements sans moyen de protection adapté aux analyses de protéomique (figure 16).

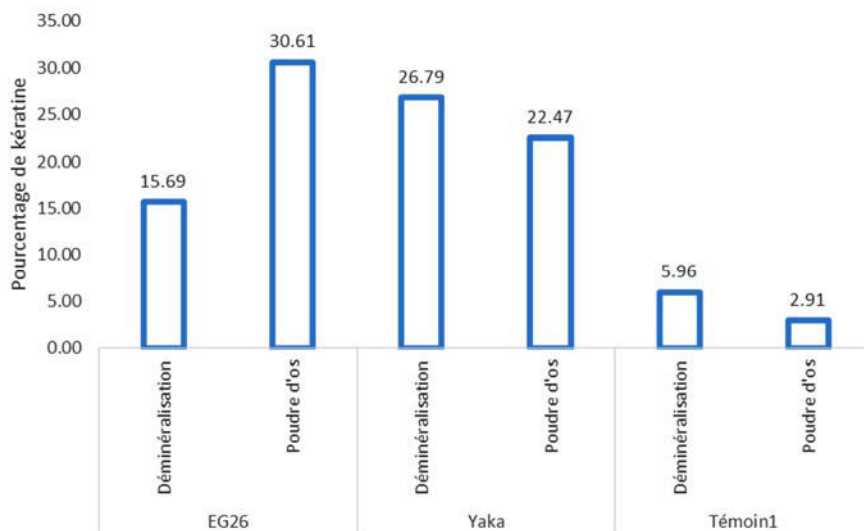


Figure 16 : Histogramme du pourcentage de kératine détectée dans les échantillons.

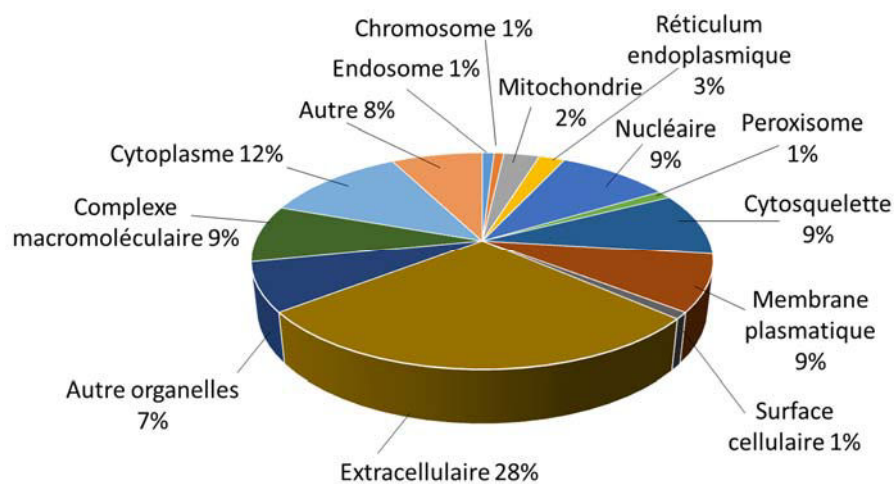
c) Le collagène

Dans les trois échantillons de nombreux types de collagènes ont été identifiées comme le COL1A1, COL2A1, COL3A1, COL1A5, COL5A1, COL12A1 ainsi que des protéines du cartilage comme l'ostéopontine, l'ostéomoduline, la chondroadhérine.

De nombreuses protéines non collagéniques ont également été identifiées comme : l'Alpha-2-HS-glycoprotéine, le biglycan, la vitronectine, la protéine du complément C9, la décorine, la bone sialoprotéine, la tetranectine et la prothrombine.

Malgré l'enfouissement et la dégradation subis par les ossements archéologiques, des protéines collagéniques et non collagéniques sont encore détectées. Des publications ont montré que l'os est une partie du corps qui permet la conservation des protéines pendant de très longues années [64-69].

Cinquante-cinq protéines uniques différentes ont été détectées dans l'os et la fraction acide de EG26 (protéines uniques des deux fractions, non redondante) et 138 protéines uniques différentes ont été détectées dans l'os et la fraction acide de YAKa12. Après une analyse des numéros Gene Ontology des protéines, leur fonction moléculaire principale identifiée est la fonction de liaison, ce qui correspond à la fonction du collagène. Leurs localisations sont principalement extracellulaires. Ces protéines sont réparties majoritairement dans deux processus biologiques : la régulation et les processus cellulaires (figure 17).



a

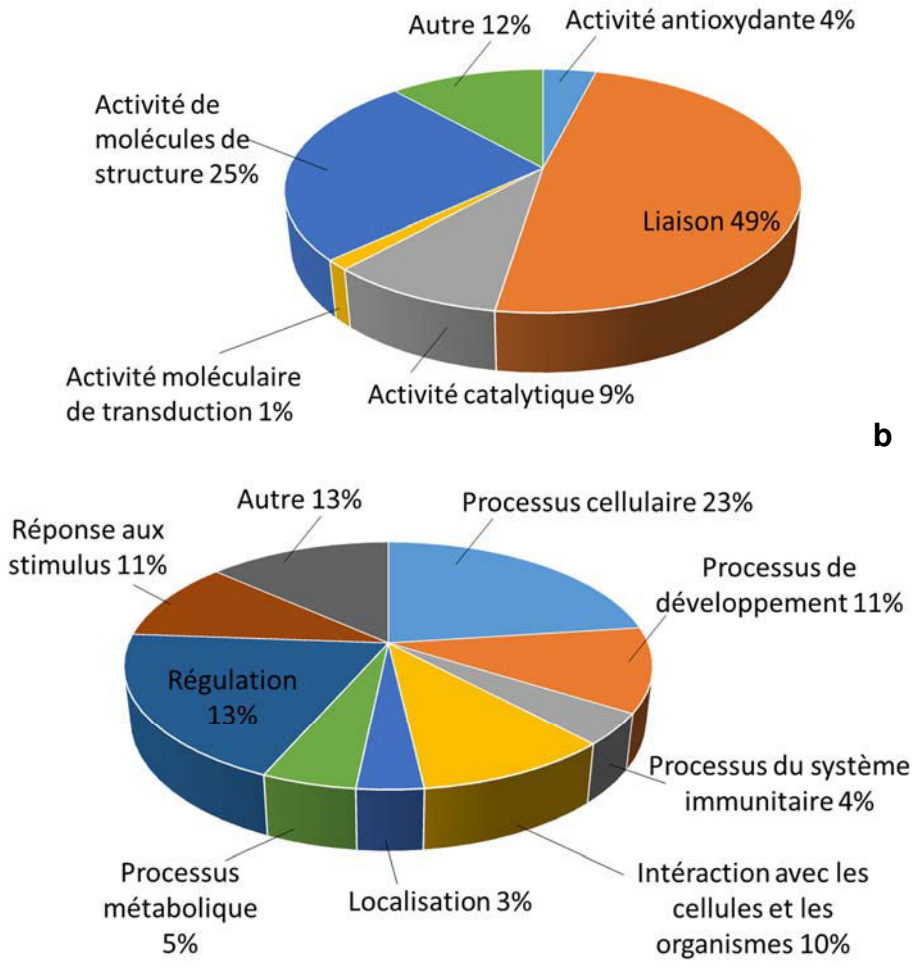


Figure 17 : Répartition des protéines identifiées au sein de l'os EG26 en fonction de leur localisation cellulaire (a), fonction moléculaire (b) et processus biologique (c).

Après l'élimination des protéines contaminantes dans les trois échantillons, les protéines présentes possèdent des similitudes au niveau de leurs fonctions, structures (exemple : nombreux type de collagènes identifiés). Ceci suggère que les ossements archéologiques possèdent des protéines d'os et que ces protéines identifiées ne proviennent pas de contaminant environnementaux.

En se focalisant uniquement sur la famille du collagène, de nombreux types ont été détectés provenant d'espèces différentes (tableau 3). Le collagène 1A1, 1A2 et 2A1 sont retrouvés dans chaque ossement.

	EG26		Yaka		Témoïn1	
	Deminéralisation	Poudre d'os	Deminéralisation	Poudre d'os	Deminéralisation	Poudre d'os
COL11A1	x	•	x	•	x	•
COL11a2	•	•	x	•	x	•
COL12A1	•	x	•	•	•	•
COL14A1	x	•	x	x	•	x
COL17A1	x	x	•	x	•	x
COL18A1	x	•	x	x	x	x
COL1A1	•	•	•	•	•	•
COL1A2	•	•	•	•	•	•
COL21A1	•	•	x	x	x	x
COL22A1	•	•	x	•	•	•
COL2A1	•	•	•	•	•	•
COL39	•	x	x	•	x	•
COL3A1	x	•	x	•	•	•
COL4A3	•	•	x	•	x	•
COL4A4	x	x	x	x	•	x
COL4A5	•	•	•	x	x	x
COL5A1	x	•	x	•	•	•
COL5A2	•	x	x	•	•	•
COL6A1	•	•	x	x	•	x
COL6A3	x	•	x	x	•	x

Tableau 3 : Collagènes identifiés dans la fraction acide et poudre d'ossement pour les trois échantillons. Le point noir indique la présence de la protéine et la croix l'absence. En gras et rouge sont indiqués les collagènes détectés dans les deux ossements.

Le collagène de type I représente 90% de la matière organique de l'os et la triple hélice de collagène est composée de deux chaînes alpha 1 et d'une chaîne alpha 2. Par conséquent, il est attendu que la proportion de collagène de type I soit dominante dans les os archéologiques. Dans nos échantillons, le collagène type I alpha 1 (COL1A1) a le plus grand nombre de peptides uniques suivis du collagène type I alpha 2 (COL1A2) (tableau 4). De plus, les peptides identifiés pour ces collagènes dans l'ossement archéologique sont principalement retrouvés sur la séquence du collagène d'*Homo sapiens* de la base de données NCBI. La composition protéique et le rapport entre le collagène A1 et A2 est similaire entre l'ossement de Témoïn1 et celui des deux échantillons archéologiques EG26 et YAKa12.

d) Analyse de la déamidation des peptides du collagène

La déamidation est un indicateur de la qualité de conservation, de nombreuses études ont montré que l'augmentation de la déamidation du collagène peut être corrélée avec l'âge et l'environnement dans lequel l'ossement a été conservé [42,70-74]. Bien sûr, il a été démontré par ailleurs que certaines procédures d'extraction des protéines peuvent induire une déamidation lors de la déminéralisation et/ou lors de leur solubilisation [75-77]. Ces procédures d'extraction en général diffèrent entre les sources publiées.

Cette étude sur les déamidations de la glutamine (Q) a été faite à partir des extractions des ions de deux des peptides étudiés par van Doorn et al. [42,72] (Figure 21 et 22). Les peptides étudiés, GVQGPPGPAGPR, COL1 α 1 de masse $[M+H]^{2+}$ m/z 553,2914 en position 686-697 et DGEAGAQQGPPGPAGPAGER, COL1 α 1 de masse $[M+H]^{2+}$ m/z 853,8905 en position 613-631 [70] (tableau 5).

Séquence peptides	Protéine	Charges	Position	Masse peptides	Intensité EG26	Intensité Yaka	Intensité Témoin1	Déamidation EG26	Déamidation Yaka	Déamidation Tem1
GVQGP <u>P</u> GPAGPR	COL1 α 1	2	669-687	553,2914	1,16 10 ⁺⁸	1,01 10 ⁺⁸	2,8 10 ⁺⁷	30%	36%	22%
GV <u>Q</u> GP <u>P</u> GPAGPR	COL1 α 1	2	669-687	553,7827	3,5 10 ⁺⁷	3,69 10 ⁺⁷	0,61 10 ⁺⁷			
DGEAGAQQGP <u>P</u> GPAGPAGER	COL1 α 1	2	613-631	853,8905	8,92 10 ⁺⁵	5.30 10 ⁺⁶	8,22 10 ⁺⁶	36%	33%	20%
DGEAGA <u>Q</u> GP <u>P</u> GPAGPAGER	COL1 α 1	2	613-631	854,3820	3,25 10 ⁺⁵	1,65 10 ⁺⁶	2,79 10 ⁺⁶			

Tableau 5 : Tableau récapitulatif des deux peptides étudiés pour l'analyse de déamidation avec la charge, la position sur la chaîne du COL1 α 1, la masse et intensité des peptides dans chacun des échantillons et le pourcentage des peptides déamidés présent dans les 3 échantillons.

Ces peptides ne contiennent qu'une seule position qui peut subir une déamidation et possèdent des taux de déamidation bas par rapport aux autres positions de glutamine dans le COL1 [42]. Le pourcentage de déamidation des deux peptides du collagène COL1A1 dans les ossements est calculé en réalisant le rapport des appels des ions des peptides natifs sur les peptides déamidés.

Une différence est observée entre les échantillons archéologiques et l'échantillon témoin Témoin1. Dans les échantillons archéologiques entre 30 et 36% des peptides sont déamidés contre 20-22% dans l'échantillon Témoin1. La figure 18 correspond à l'extraction des ions des peptides sans déamidation et avec dans l'échantillon Témoin1.

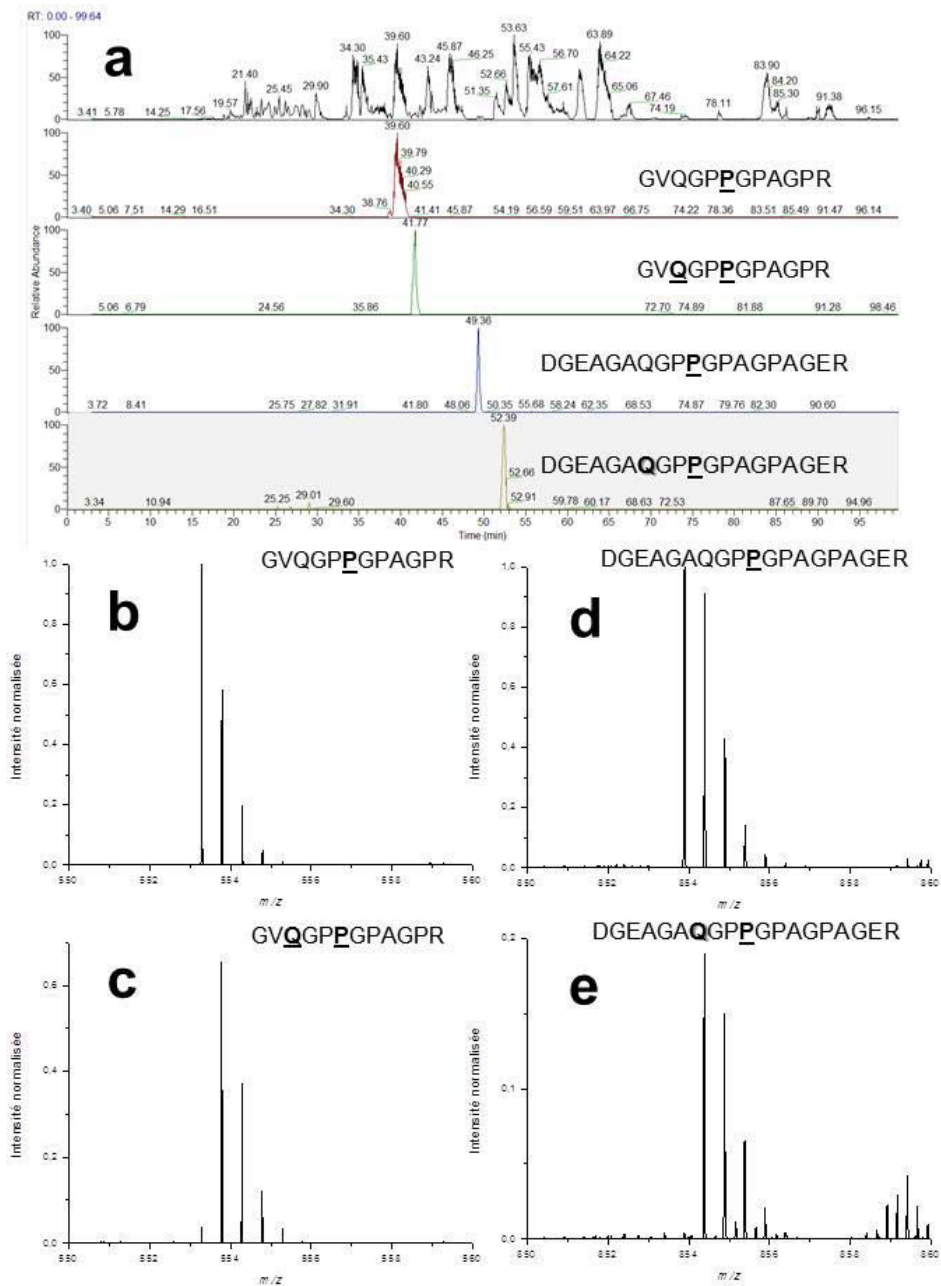


Figure 18 : Courant ionique total de la digestion eFASP de l'échantillon témoin « Témoin1 » avec l'extraction des courants ioniques des peptides utilisés pour l'analyse de la déamidation (a). Les spectres MS de ion du peptide GVQGPPGPAGP sans déamidation et avec déamidation sont observés en b et c. Pour le peptide DGEAGAQGPPGPAGPAGER, les spectres sont observés en d et e. La déamidation est détectée avec un décalage de masse de 0,5 Da.

La fragmentation de ces peptides est montrée dans la figure 19 et permet de vérifier la séquence des peptides et la position de la modification.

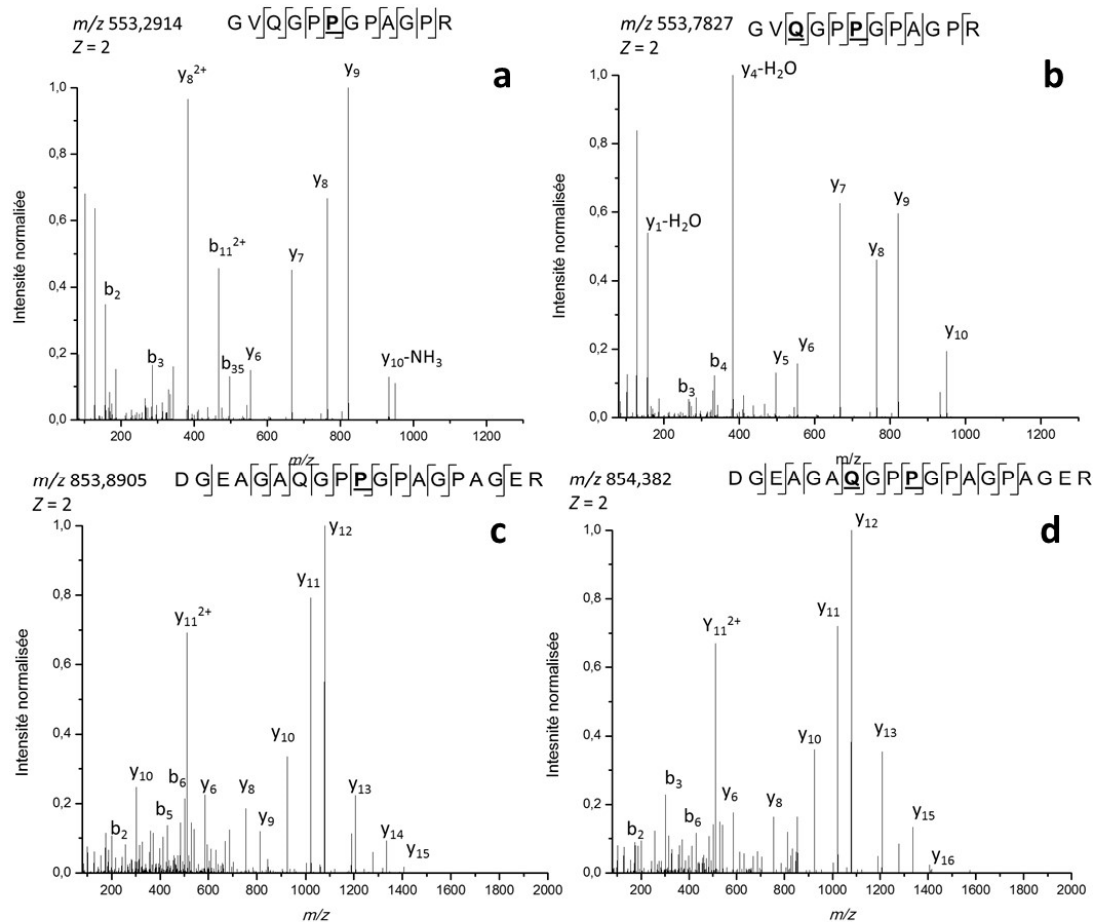


Figure 19 : Spectres MS/MS des peptides utilisés pour étudier le pourcentage de déamidation. Spectres MSMS du peptide GVQGGPPGPAGPR (a) et GVQGGPPGPAGPR (b). Les spectres de fragmentation du peptide DGEAGAQQGPPGPAGPAGER (c) et DGEAGAQQGPPGPAGPAGER (d).

La déamidation est un indicateur de la qualité de conservation, ce pourcentage élevé de déamidation indique que les ossements archéologiques sont anciens [42,73].

e) Identification du genre de l'ossement

L'analyse avec le logiciel PEAK sur la base de données NCBI « all entries » identifie des protéines d'*Homo sapiens* avec un grand nombre de peptides pour les deux échantillons. Ces protéines correspondent à COL1A2, K1C9, COL1A1. Dans le tableau 4, sont rassemblés les résultats sur les collagènes humains identifiés COL1A1 et A2.

Analyse bottom-up des échantillons archéologiques et paléontologiques

		Accession	ID	Gène	Organisme	Taille	-10lgP	Couverture (%)	Peptides	Peptides uniques
EG26	Poudre d'os	P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	476	72	363	86
		P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	513	69	275	30
	Démérisation	P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	545	70	499	51
		P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	448	73	223	53
YAKA12	Poudre d'os	P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	564	79	254	58
		P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	484	73	313	34
	Démérisation	P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	489	71	495	48
		P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	450	74	399	97
Témoin1	Poudre d'os	P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	328	73	219	54
		P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	336	70	239	30
	Démérisation	P02452	CO1A1_HUMAN	COL1A1	Homo sapiens	1464	351	69	473	46
		P08123	CO1A2_HUMAN	COL1A2	Homo sapiens	1366	338	72	452	98

Tableau 4 : Compilation des résultats de PEAKS 7.0 sur les 2 types de collagènes identifiés avec le plus grand nombre de peptides.

Les résultats montrent une identification des collagènes 1A1 et 1A2 avec un haut pourcentage de couverture (d'environ 70%), des nombres de peptides et peptides uniques très élevés pour les trois échantillons et dans les deux fractions. Environ 300 peptides sont identifiés sur le CO1A1 et CO1A2 dans les trois échantillons pour la fraction de poudre d'ossement.

Seul les parties *N* et *C*-terminal ne sont pas identifiées car elles contiennent le peptide signal et les pro-peptides qui sont clivés lors de la maturation de la protéine (figure 20). De nombreuses modifications ont été identifiées comme : les hydroxylations des prolines (P) qui sont des modifications naturelles sur le collagène. L'analyse protéomique montre que plus de 70% des prolines du collagène 1 alpha 1 (figure 20) et 1 alpha 2 sont hydroxylées.

Des oxydations sur les méthionines (M) et déamidation des glutamines (Q) et asparagine (N) sont détectées, et peuvent être induites lors de la digestion.

Analyse bottom-up des échantillons archéologiques et paléontologiques

```

1 MFSFVLDLRL LLLAATALLT HGQEEGQVEG QDEDIPPTIC VQNGRLYHDR DVWKPPECRI CVCNDGKVLG DDVICDETKN CPGAEPVEGE CCFVCPDGSE SPTDQETTGV
111 EGPKGDTGPR GPRGPAGPPP RDGIFGQPLG PGPFPFPPPP GPPGLGNFA PQLSYGYDEK STGGISVPGF MGPSGPRGLP GPPGAPGQG FQGGPEPGE PGASGMPGPR
221 GPPGPPGKNG DDGEAGKPR FGERGPPFQ GARGLPGTAG LPMGMKGRGF SGLDGAKDA GPAGPKGEG SPGENGAPQG MGRPGLPGER GRPGAPGAG ARGNDGATGA
331 AGPPGPTGPA GPPGFPAGV AKGEAGPQGF RGSEGPQVVR GEPGPPGAG AAGPAGNFGA DGQPGAKGAN GAPGIAGAPG FPGARGPSGF QGGGPPGPK GNSGEPGAPG
441 SKGDTGAKGE PGPVVGQPPP GPAGEEGKRG ARGEPGPTGL PGPFPGERGGP GSRGFPGADG VAGPKGPAGE RGSFGPAGPK GSPGEAGRPG EAGLPGAKGL TSPGSPGPD
551 KKTGPPGAG QDGRGPPFP PGARQAGVM GPPGPKAAG EPKAGGERV PGPFGAVPA GHDGEGAQG PPGPAGPAGE RGEQGPAGSP GFQGLPGPAG PPGEAGKPE
661 QVPGDLGAF GPSGARGERG FPGERGVQGF PGPAGFRGAN GAPGNDGAKG DAGAPGAPGS QGAPLQGMF GERGAAGLPG PKGDRGDAGP KGADGSPGKD GVRGLTGPIG
771 PPGPAGAPGD KGESGSPGPA GPTGARGAP DRGEPGPPGP AGFAGPPGAD GQPGAKGEPG DAGAKGDAGP PGPAGPAGPP GPIGNVAGP AKGARGSAGP PGATGPPGAA
881 GRVGPFGPSG NAGPPGPPGP AGKEGKGRP GETGPAGRPG EVGPPGPPGP AGEKSGPAD GPAGAGTTPG PQGIAGQGRV VGLPGQRGER GPPGLPGPSG EPGKQGPSGA
991 SGERGPPGPM GPPGLAGPFG ESGREGAPGA EGSPGRDSSP GAKGDRGETG PAGPPGAPGA PGAGPVGPA GKSGRDRETG PAGTGPVGF VGARGPAGFQ GPRGDKGETG
1101 EQGDRGIKGH RGFSGLQGGP GPPGSPGEGQ PSGASGPAGP RGPFGSAGAP GHDGLNLPLG FIGPFGPRGR TGDAGVPGFP GPPGPPGPPG PPSAGFDFSF LPQPPEKAKH
1211 DGGRYV RADD ANVVRDRDLE VDTTLKSLSQ QIENIRSPG SRKNPARTCR DLKMCMSDWK SGEYWDPNQ GCNLDAIKVF CNMETGETCV YPTQPSVAQK NWYISKNPKD
1321 KRHVWFGE SM TDGFQFEYGG QGSDPADVAI QLTFLRLMST EASQNTIYHC KNSVAYMDQQ TGNLKKALLL QGSNEIEIRA EGNRSFTYSV TVDGCRTSHTG AWGKLVIEYK
1431 TTKTSRLPII DVAPLDVAGP DQEFGFDVGP VCFL
  
```

Figure 20 : Séquence du collagène 1 alpha 1 de *Homo sapiens*, identifiée dans la fraction poudre d'os dans l'échantillon de EG26. 275 peptides ont été identifiés recouvrant 70 % de la séquence. En gris est indiqué où les peptides ont été identifiés avec les modifications h : hydroxylation proline, d : déamidation arginine ou acide aspartique, o : oxydation des méthionines.

Sur la base des résultats d'identification du collagène, une étude phylogénétique a été réalisée pour observer la divergence du collagène 1 alpha 1 et 2 dans différentes espèces (figure 21) à l'aide de l'outil en ligne « phylogeny » [78] et des séquences de 13 espèces différentes dont les séquences du collagène de *Homo neanderthalensis*. Dans un premier temps, les séquences protéiques des collagènes ont été alignées avec l'outil MUSCLE 3.7 (Multiple Sequence Comparison by Log-Expectation). Puis l'outil PhyML 3.0, qui est un logiciel de phylogénie basé sur le principe du maximum de vraisemblance, a été utilisé. Il calcule une ressemblance entre les séquences et évalue une hypothèse sur l'histoire de l'évolution en termes de probabilité. Puis l'arbre phylogénétique a été créé à l'aide de TreeDyn 198.3.

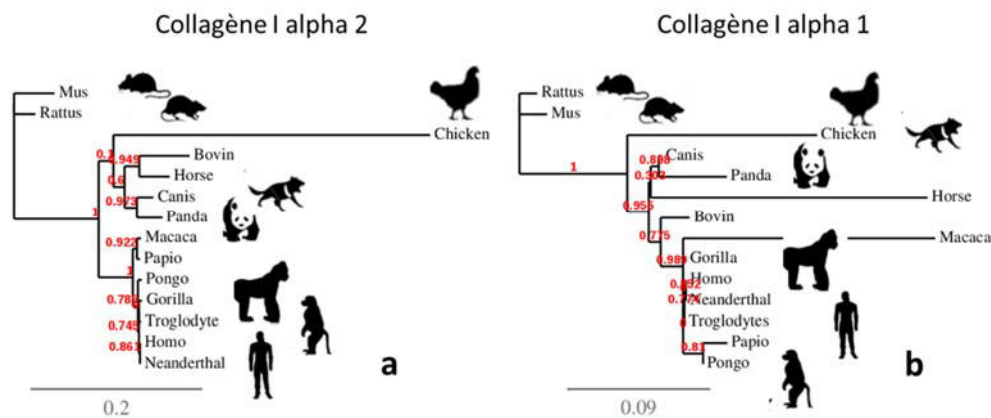


Figure 21 : Phylogénie des séquences des collagènes I alpha 1 (b) et alpha 2 (a) réalisé avec les outils sur le site <http://phylogeny.lirmm.fr/phylo.cgi/index.cgi>. La valeur en rouge ou en gris correspond à la distance « génétique » entre les séquences. Plus la valeur en rouge est élevée, plus les séquences sont proches.

Comme montré dans la figure 21, les séquences des collagènes sont très similaires (plus de 90% de similarité) pour la famille des *Hominidae* (*Gorilla*, *Homo*, *Papio*, *Pongo*, *Macaca*, *Pan troglodyte*, *neandethalis*). Les résultats de protéomique et de phylogénie montrent une forte ressemblance entre *Homo sapiens* et *Homo neanderthalensis*, avec une valeur de similarité des séquences de 99%. Des données de littératures ont montré des résultats similaires à partir d'analyse protéomique sur les collagènes de différentes espèces [64,79].

Nos résultats permettent de montrer que les ossements archéologiques EG26, YAKa12 sont de la famille des *Hominidae*, du genre *Homo*. Les informations archéologiques, stratigraphique, ADN combinées aux informations de protéomique tendent à prouver que ces ossements pourraient être de la famille *Homo sapiens* [50-53].

Pour vérifier ces données, une analyse comparative des collagènes 1 et 2 entre *Homo sapiens* et *Homo neanderthalensis* va être réalisée, puis les peptides spécifiques de la famille *Homo sapiens* vont être analysés avec l'outil BlastP pour confirmer que ces peptides soient uniques de cette famille.

f) Etude des peptides spécifiques

Il existe une similitude de 99% entre *Homo Sapiens* et *Homo neandertalensis* pour les séquences de collagènes 1A1 et 1A2. Entre ces deux espèces, il existe des modifications d'acides aminés spécifiques. Ces variations ont été identifiées sur la dans « The Neanderthal Genome » [78]. Pour le collagène 1 alpha 1, 41 variations ou single nucleotide polymorphism (SNP) ont été répertoriées sur le gène dans le génome de Neandertal. Sur ces 41 SNP, 20 sont des SNP codantes, ce qui signifie que la SNP provoque un changement d'acides aminés (tableau 5).

Position	SNP ID	Allèles	Acide aminé
120	rs35183627	C/T	R, Q
197	rs8179178	C/A ou C>T	G, C
476	rs57377812	C/G ou C/T	G, R
564	rs1800211	C/T	R, H
571	rs5028520	G/A	P, L
591	rs1135345	C/T	E, K
823	rs1800214	G/C ou G/T	P, A
993	rs62062995	C/A ou C/T	E, *
1019	rs1135348	C/G	G, A

1072	rs11556513	C/A ou C/T	A, S
1075	rs1800215	C/T	A, T
1141	rs41316713	C/T	R, Q
1177	rs41316719	C/T	V, I
1251	rs3205325	C/G	T, S
1379	rs3205509	G/T	K, Q
1391	rs2586486	G/T	K, Q
1431	rs1059454	T/G	P, T
1434	rs1800220	T/A	T, S
1438	rs17857117	G/C	P, R
1460	rs17853657	G/T	P, H

Tableau 5 : SNP codantes présents sur le gène du collagène 1 alpha 1 localisé sur le chromosome 17 du génome de Neandertal. La première colonne correspond à la position de l'acide aminé qui varie, la seconde colonne est la référence de la SNP, la colonne allèles correspond au nucléotide qui change C en T pour rs1800215, la dernière colonne correspond à la variation de l'acide aminé R en Q. * veut dire aucun acide aminé.

La SNP correspond à l'identifiant rs1800215 correspond à l'acide aminé en position 1075 sur le collagène 1A1 qui correspond à une substitution de l'alanine (A) en thréonine (T) en avec une fréquence qui varie entre le génome d'*Homo sapiens* et d'*Homo neandertalensis*. Pour confirmer que les ossements anciens correspondent à des ossement de l'espèce *Homo Sapiens*, cette substitution a été recherchée dans les trois échantillons. Les extractions des ions m/z 773,9022 et m/z 788,9075 correspondent aux peptides GETGPAGPAGPVGPVGAR et GETGPAGPTGPVGPVGAR ont été réalisées. Le rapport du signal extrait de l'ion du peptide deux fois chargé GETGPAGPAGPVGPVGAR (SNP A) à m/z 773,9022 ($C_{66}H_{107}N_{21}O_{22}$) sur l'ion du peptide GETGPAGPTGPVGPVGAR (SNP T) à m/z 788,9075 ($C_{61}H_{97}N_{17}O_{22}$) deux fois chargé (figure 22) nous permet d'obtenir une estimation de la fréquence de substitution (tableau 6).

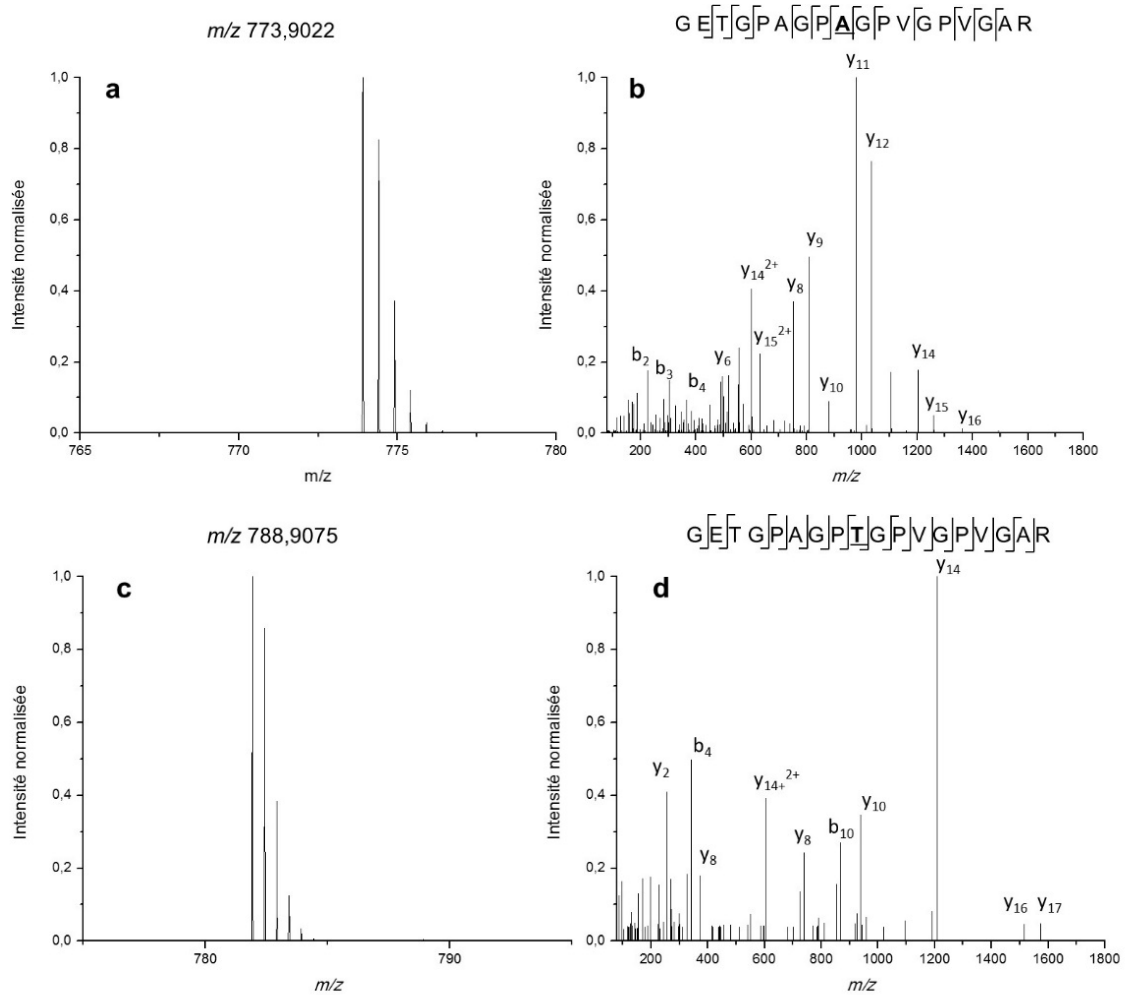


Figure 22 : Spectres MS (a) et MS/MS (b) du peptide GETGPAGPAGPVGPGAR contenant l'acide aminé A en position 1075. Spectres MS (c) et MS/MS (d) du peptide GETGPAGPTGPVGPVGAR.

Les spectres de fragmentation des deux peptides avec et sans la substitution de l'alanine (A) en thréonine (T) sont présentés figure 22. Les peptides sont identifiés avec plus de 90 % des ions *b* et *y*, ce qui permet une identification de la substitution.

La fréquence de substitution de l'acide aminé A en position 1075 du collagène 1A1 est d'environ 95% pour les échantillons et 5% pour l'acide aminé T, ce qui suggère que les ossements sont de la famille de *Homidae*, du genre *Homo*, de l'espèce *Sapiens*, car pour la famille *Homo neandertalensis* seul la substitution en thréonine T est présente et ceux d'après les informations obtenue sur le site « The Neanderthal Genome ».

	Echantillon	Nom du gène	Position de la mutation	dbSNP	Neanderthal	AA homo sapiens	Fréquence théorique (1000G global)	freq SNP A	freq SNP T
poudre d'os	EG26	COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,84	4,16
fraction acide		COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,63	4,37
poudre d'os	YAKa12	COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,84	4,16
fraction acide		COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,63	4,37
poudre d'os	Témoin1	COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,94	4,06
fraction acide		COL1A1	1075	rs1800215	T	A , T	97,9 , 2,1	95,89	4,11

Tableau 6 : Fréquence de la substitution en position 1075 présent du le collagène COL1A1 identifiée dans les trois échantillons pour les deux fractions. La colonne Neanderthal correspond à l'acide aminé présent dans le génome de *Neanderthalensis*, AA homo sapiens indique les acides aminés dans le génome de *Homo sapiens*, Fréquence théorique indiqué sur le site <http://www.internationalgenome.org/>, freq SNP A ou T correspond au pourcentage de présence de la SNP dans les échantillons.

Une analyse avec le logiciel Blast P est ensuite utilisée pour vérifier la présence de peptides spécifique de la famille *Homo Sapiens* (tableau 7). L'analyse Blast P consiste à rechercher une séquence d'acides aminés dans une base de données de protéines et de chercher les homologies. Le résultat est un score S correspondant à la somme des alignements des acides aminés de la séquence (exemple : L – L = +4, E·Q = +2) et une p-value qui est la probabilité d'obtenir par hasard un score x au moins égal au score S.

Cette analyse montre que le peptide GETGPAGPAGPVGPVGAR est spécifique du collagène 1 alpha 1 provenant de la famille *Homo Sapiens* car ce peptide a été trouvé dans la base de données de Blast P (protéine non redondante de GenBank) avec un p-value très faible (8,00E-12) et un pourcentage d'identité de 100% (tableau 7 A). La famille suivante identifiée avec ce peptide est *Macaca mulatta* avec une p-value de 0,32 et un pourcentage d'identité de 67% (tableau 7 B).

Le second peptide EQGPPGPPGFQGLPGPSGPAGEVGK est spécifique du collagène 1 alpha 2 provenant de la famille *Homo Sapiens* car ce peptide a été trouvé dans la base de données de Blast P (protéine non redondante de GenBank) avec un p-value très faible (3,00E-16) et un pourcentage d'identité de 100% (tableau 7 C). La famille suivante identifiée avec ce peptide est *Colobus angolensis palliatus* avec une p-value très faible de 3,00E-14, mais avec un pourcentage d'identité de 96% et une séquence peptidique de 25 acides aminés utilisée au lieu de 26 (tableau 7 D).

Score	p-value	Identité	A
67,7	8,00E-12	18/18 (100%)	
	Début	Séquence	Fin
peptide identifié	1	GETGPAGPAGPVGPVGAR	18
		GETGPAGPAGPVGPVGAR	
correspondance	1067	GETGPAGPAGPVGPVGAR	1084

Score	p-value	Identité	B
37,5	0,32	12/18 (67%)	
	Début	Séquence	Fin
peptide identifié	2	GETGPAGPAGPVGPVGAR	18
		GETGP G P G P GAR	
correspondance	1101	GETGPSGPAG- - PTGAR	1118

Score	p-value	Identité	C
81,2	3,00E-16	26/26 (100%)	
	Début	Séquence	Fin
peptide identifié	1	GEQGPPGPPGFQGLPGPSGPAGEVGK	26
		GEQGPPGPPGFQGLPGPSGPAGEVGK	
correspondance	544	GEQGPPGPPGFQGLPGPSGPAGEVGK	569

Score	p-value	Identité	D
74,4	3,00E-14	24/25 (96%)	
	Début	Séquence	Fin
peptide identifié	4	EQGPPGPPGFQGLPGPSGPAGEVGK	22
		EQGP GPPGFQGLPGPSGPAGEVGK	
correspondance	547	EQGPSGPPGFQGLPGPSGPAGEVGK	471

Tableau 7 : Analyse BlastP du peptide GETGPAGPAGPVGPVGAR du collagène 1A1. A) Résultat Blast P avec une identité de 100% sur le collagène 1A1 d'*Homo sapiens*, B) Blast P du même peptide mais avec une identité de 67% sur le collagène 1A1 de *Macaca mulatta*. Analyse BlastP du peptide GEQGPPGPPGFQGLPGPSGPAGEVGK du collagène 1A2, C) résultat Blast P avec une identité de 100% sur le collagène 1A2 d'*Homo sapiens*, D) Blast P du même peptide mais avec une identité de 96% sur le collagène 1A2 de *Colobus angolensis palliatus*.

VII. Conclusion

La méthode développée a été appliquée avec succès à l'étude d'un échantillon ancien. Parmi les faits marquants de notre méthodologie peuvent être cités :

- L'utilisation de la solution de TFA 5% pour réaliser la déminéralisation des ossements. Cette solution permet de limiter la formation d'oxydations au cours de l'étape de déminéralisation comparé aux méthodes classiques (HCl, NaOH).
- L'utilisation de la méthode eFASP permet d'obtenir des couvertures de séquence pour le collagène de plus de 70%, ainsi qu'un nombre de peptides identifiées de 300 peptides pour le collagène dans un échantillon témoin et ancien.
- La quantité utilisée sur l'échantillon témoin est de 10 mg pour l'étude, mais cette quantité peut être diminuée à plusieurs centaines de microgrammes, car les protéines présentes sont correctement conservées.
- L'analyse quantitative par spectrométrie de masse sur les fractions déminéralisation et poudre d'os a permis de mettre en évidence la présence de protéines différentes dans chaque fraction. Dans la fraction déminéralisation, les protéines du plasma et du sang comme l'albumine, la vimentine... y sont majoritaires, tandis que dans la poudre d'ossement, il s'agit des protéines constituant l'os.
- Pour les ossements anciens, 10 mg ont été utilisés et ont permis d'identifier de nombreuses protéines (contaminantes et de l'ossement). Notre méthode permet de diminuer les quantités d'échantillons (méthodes classique : 50 à 100 mg). Les couvertures de séquence sur le collagène 1 alpha 1 et 2 humain sont de 70% et plus 300 peptides ont été identifiées.
- L'étude de peptides natifs et déamidés sur le collagène 1 alpha 1, permet de monter une variation du pourcentage de déamidation entre l'échantillon témoin et anciens (20 et 35 % respectivement). Le pourcentage élevé de déamidation présent sur les peptides des ossements anciens du permet de valider qu'il s'agit d'échantillons ancien.
- L'analyse protéomique bottom-up montre que les ossements anciens correspondent à la l'espèce *Hominidae* et plus précisément à *Homo Sapiens*.

L'un des défis des paléontologues est d'obtenir des informations phylogéniques sur des espèces éteintes. La génomique peut répondre à cette question en étudiant l'ADN mitochondrial [80] mais cette analyse est limitée par la présence d'ADN dans l'ossement et des études ont montré que les ossements contiennent de moins en moins de matériel et la limite actuelle est de 430 000 ans pour un ossement d'*Hominidae* [81] et 700 000 pour celui d'un cheval [82]. La protéomique est capable de détecter des protéines jusqu'à plusieurs millions d'années [83]. L'utilisation de la protéomique bottom up et de la LC-MSMS avec des spectromètres de masse haute résolution pourrait permettre d'identifier des espèces inconnues. L'implémentation de différents logiciels *de novo* [84] et des moteurs de recherche [85] permettraient d'identifier d'avantage de peptides et de substitutions d'acides aminés pour permettre une meilleure identification des espèces. De plus, il est envisageable de mettre en place des algorithmes de classification [86-88] pour identifier les échantillons à partir de bases de données d'ossements déjà analysés et pour identifier des espèces inconnues en utilisant la ressemblance entre deux analyses LC-MSMS ou à partir des données de *de novo*.

Bibliographie

- (1) Anderson, N. L.; Anderson, N. G.: Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **1998**, *19*, 1853-1861.
- (2) Bodzon-Kulakowska, A.; Bierczynska-Krzysik, A.; Dylag, T.; Drabik, A.; Suder, P.; Noga, M.; Jarzebinska, J.; Silberring, J.: Methods for samples preparation in proteomic research. *J. Chromatogr. B* **2007**, *849*, 1-31.
- (3) Feist, P.; Hummon, A. B.: Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *International journal of molecular sciences* **2015**, *16*, 3537-3563.
- (4) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates III, J. R.: Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **2013**, *113*, 2343-2394.
- (5) Olsen, J. V.; Ong, S.-E.; Mann, M.: Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3*, 608-614.
- (6) Shi, J.; Wu, F.-X.: Peptide charge state determination of tandem mass spectra from low-resolution collision induced dissociation. *Proteome science* **2011**, *9*, S3.
- (7) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M.: In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2007**, *1*, 2856-2860.
- (8) Stevens, S. M.; Wolverton, S.; Venables, B.; Barker, A.; Seeley, K. W.; Adhikari, P.: Evaluation of microwave-assisted enzymatic digestion and tandem mass spectrometry for the identification of protein residues from an inorganic solid matrix: implications in archaeological research. *Anal. Bioanal. Chem.* **2010**, *396*, 1491-1499.
- (9) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M.: Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **2014**, *11*, 319-324.
- (10) Procopio, N.; Buckley, M.: Minimizing Laboratory-Induced Decay in Bone Proteomics. *J. Proteome Res.* **2017**, *16*, 447-458.
- (11) Annesley, T. M.: Ion suppression in mass spectrometry. *Clinical chemistry* **2003**, *49*, 1041-1044.
- (12) Touboul, D.; Kollmer, F.; Niehuis, E.; Brunelle, A.; Lapr evote, O.: Improvement of biological time-of-flight-secondary ion mass spectrometry imaging with a bismuth cluster ion source. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1608-1618.
- (13) Biemann, K.: Contributions of Mass Spectrometry to Peptide and Protein Structure. *Biomedical and Environmental Mass Spectrometry.* **1988**, *16*.
- (14) Roepstorff, P.; Fohlman, J.: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984**, *11*, 601.
- (15) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T.: Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, *59*, 2621-2625.

- (16) Cottrell, J. S.; London, U.: Probability-based protein identification by searching sequence databases using mass spectrometry data. *electrophoresis* **1999**, *20*, 3551-3567.
- (17) Eng, J. K.; McCormack, A. L.; Yates, J. R.: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.
- (18) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M.: Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794-1805.
- (19) Jimenez, C. R.; Huang, L.; Qiu, Y.; Burlingame, A. L.: Searching sequence databases over the internet: protein identification using MS-tag. *Curr Protoc Protein Sci* **2001**, *Chapter 16*, Unit 16.6.
- (20) Zhang, N.; Aebersold, R.; Schwikowski, B.: ProbiD: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2002**, *2*, 1406-1412.
- (21) Craig, R.; Beavis, R. C.: TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466-1467.
- (22) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H.: Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958-964.
- (23) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L.: ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007**, *35*, W701-W706.
- (24) Tanner, S.; Shu, H.; Frank, A.; Wang, L.-C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V.: InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **2005**, *77*, 4626-4639.
- (25) Pouillet, P.; Carpentier, S.; Barillot, E.: myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics* **2007**, *7*, 2553-2556.
- (26) Stein, S. E.; Scott, D. R.: Optimization and testing of mass spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 859-866.
- (27) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G.: PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337-2342.
- (28) Frank, A.; Pevzner, P.: PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77*, 964-973.
- (29) Ma, B.: Novor: real-time peptide de novo sequencing software. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1885-1894.
- (30) Muth, T.; Weilnböck, L.; Rapp, E.; Huber, C. G.; Martens, L.; Vaudel, M.; Barsnes, H.: DeNovoGUI: an open source graphical user interface for de novo sequencing of tandem mass spectra. *J. Proteome Res.* **2014**, *13*, 1143-1146.

- (31) Dallongeville, S.; Garnier, N.; Rolando, C.; Tokarski, C.: Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. *Chem. Rev.* **2016**, *116*, 2-79.
- (32) Vinciguerra, R.; De Chiaro, A.; Pucci, P.; Marino, G.; Birolo, L.: Proteomic strategies for cultural heritage: From bones to paintings. *Microchem. J.* **2016**, *126*, 341-348.
- (33) Shevchenko, A.; Yang, Y.; Knaust, A.; Thomas, H.; Jiang, H.; Lu, E.; Wang, C.; Shevchenko, A.: Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China. *J. Proteomics* **2014**, *105*, 363-371.
- (34) Solazzo, C., Fitzhugh, W. W., Rolando, C., & Tokarski, C. : Identification of protein remains in archaeological potsherds by proteomics. *Anal. Chem.* **2008**, *80*, 4590-4597.
- (35) Nielsen-Marsh, C. M.; Richards, M. P.; Hauschka, P. V.; Thomas-Oates, J. E.; Trinkaus, E.; Pettitt, P. B.; Karavanic, I.; Poinar, H.; Collins, M. J.: Osteocalcin protein sequences of Neanderthals and modern primates. *Proc Natl Acad Sci U S A* **2005**, *102*, 4409-13.
- (36) Collins, M.; Buckley, M.; Grundy, H. H.; Thomas-Oates, J.; Wilson, J.; van Doorn, N.: ZooMS: the collagen barcode and fingerprints. *Spectroscopy Europe* **2010**, *22*, 2-2.
- (37) Buckley, M.; Larkin, N.; Collins, M.: Mammoth and Mastodon collagen sequences; survival and utility. *Geochim. Cosmochim. Acta* **2011**, *75*, 2007-2016.
- (38) Cappellini, E.; Jensen, L. J.; Szklarczyk, D.; Ginolhac, A.; da Fonseca, R. A.; Stafford, T. W.; Holen, S. R.; Collins, M. J.; Orlando, L.; Willerslev, E.; Gilbert, M. T.; Olsen, J. V.: Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* **2012**, *11*, 917-26.
- (39) Buckley, M.; Whitcher Kansa, S.; Howard, S.; Campbell, S.; Thomas-Oates, J.; Collins, M.: Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeo.Sci.* **2010**, *37*, 13-20.
- (40) Buckley, M.; Gu, M.; Shameer, S.; Patel, S.; Chamberlain, A. T.: High-throughput collagen fingerprinting of intact microfaunal remains; a low-cost method for distinguishing between murine rodent bones. *Rapid Commun. Mass Spectrom.* **2016**, *30*, 805-812.
- (41) Welker, F.; Collins, M. J.; Thomas, J. A.; Wadsley, M.; Brace, S.; Cappellini, E.; Turvey, S. T.; Reguero, M.; Gelfo, J. N.; Kramarz, A.; Burger, J.; Thomas-Oates, J.; Ashford, D. A.; Ashton, P. D.; Rowsell, K.; Porter, D. M.; Kessler, B.; Fischer, R.; Baessmann, C.; Kaspar, S.; Olsen, J. V.; Kiley, P.; Elliott, J. A.; Kelstrup, C. D.; Mullin, V.; Hofreiter, M.; Willerslev, E.; Hublin, J. J.; Orlando, L.; Barnes, I.; MacPhee, R. D.: Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* **2015**, *522*, 81-4.
- (42) van Doorn, N. L.; Wilson, J.; Hollund, H.; Soressi, M.; Collins, M. J.: Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2319-27.
- (43) Asara, J. M.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C.: Protein sequences from mastodon and Tyrannosaurus rex revealed by mass spectrometry. *Science* **2007**, *316*, 280-285.

(44) James D. San Antonio , M. H. S., Shane T. Jensen,Raghu Kalluri,Michael Buckley,Joseph P. R. O. Orgel: Dinosaur Peptides Suggest Mechanisms of Protein Survival. *PLoS One* **2011**, *6*, e20381.

(45) Schroeter, E. R.; DeHart, C. J.; Cleland, T. P.; Zheng, W.; Thomas, P. M.; Kelleher, N. L.; Bern, M.; Schweitzer, M. H.: Expansion for the *Brachylophosaurus canadensis* Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *J. Proteome Res.* **2017**, *16*, 920-932.

(46) Fosse, G.; Justesen, N.-P.; Wesenberg, G.: Microstructure and chemical composition of fossil mammalian teeth. *Calcif. Tissue Int.* **1981**, *33*, 521.

(47) Fresnais, M.; Richardin, P.; Sepulveda, M.; Leize-Wagner, E.; Charrie-Duhaut, A.: Omics for Precious Rare Biosamples: Characterization of Ancient Human Hair by a Proteomic Approach. *OMICS* **2017**, *21*, 361-370.

(48) Stewart, N. A.; Molina, G. F.; Issa, J. P. M.; Yates, N. A.; Sosovicka, M.; Vieira, A. R.; Line, S. R. P.; Montgomery, J.; Gerlach, R. F.: The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC Adv.* **2016**, *6*, 61673-61679.

(49) Corthals, A.; Koller, A.; Martin, D. W.; Rieger, R.; Chen, E. I.; Bernaski, M.; Recagno, G.; Dávalos, L. M.: Detecting the immune system response of a 500 year-old Inca mummy. *PLoS One* **2012**, *7*, e41244.

(50) Keyser-Tracqui, C.; Crubezy, E.; Ludes, B.: Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *The American Journal of Human Genetics* **2003**, *73*, 247-260.

(51) Keyser-Tracqui, C.; Crubézy, E.; Pamzav, H.; Varga, T.; Ludes, B.: Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. *American journal of physical anthropology* **2006**, *131*, 272-281.

(52) Crubézy, E.; Amory, S.; Keyser, C.; Bouakaze, C.; Bodner, M.; Gibert, M.; Röck, A.; Parson, W.; Alexeev, A.; Ludes, B.: Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC evolutionary biology* **2010**, *10*, 25.

(53) Keyser, C.; Hollard, C.; Gonzalez, A.; Fausser, J.-L.; Rivals, E.; Alexeev, A. N.; Riberon, A.; Crubézy, E.; Ludes, B.: The ancient Yakuts: a population genetic enigma. *Phil. Trans. R. Soc. B* **2015**, *370*, 20130385.

(54) Buckley, M.; Collins, M.; Thomas-Oates, J.; Wilson, J. C.: Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3843-54.

(55) Schroeter, E. R.; DeHart, C. J.; Schweitzer, M. H.; Thomas, P. M.; Kelleher, N. L.: Bone protein "extractomics": comparing the efficiency of bone protein extractions of *Gallus gallus* in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ* **2016**, *4*, e2603.

(56) Erde, J.; Loo, R. R. O.; Loo, J. A.: Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **2014**, *13*, 1885-1895.

- (57) Cox, J.; Mann, M.: MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **2008**, *26*, 1367-1372.
- (58) Tyanova, S.; Temu, T.; Cox, J.: The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301-2319.
- (59) Kumar, C.; Mann, M.: Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS Lett.* **2009**, *583*, 1703-1712.
- (60) Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.: STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2014**, *43*, D447-D452.
- (61) Procopio, N.; Chamberlain, A. T.; Buckley, M.: Intra- and Interskeletal Proteome Variations in Fresh and Buried Bones. *J. Proteome Res.* **2017**, *16*, 2016-2029.
- (62) Sawafuji R, C. E., Nagaoka T, Fotakis AK, Jersie-Christensen RR, Olsen JV, Hirata K, Ueda S: Proteomic profiling of archaeological human bone. *R Soc Open Sci.* **2017**, *4*.
- (63) Brown, S.; Higham, T.; Slon, V.; Pääbo, S.; Meyer, M.; Douka, K.; Brock, F.; Comeskey, D.; Procopio, N.; Shunkov, M.: Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Scientific reports* **2016**, *6*, 23559.
- (64) Buckley, M.; Fariña, R. A.; Lawless, C.; Tambusso, P. S.; Varela, L.; Carlini, A. A.; Powell, J. E.; Martinez, J. G.: Collagen sequence analysis of the extinct giant ground sloths *Lestodon* and *Megatherium*. *PloS one* **2015**, *10*, e0139611.
- (65) Hill, R. C.; Wither, M. J.; Nemkov, T.; Barrett, A.; D'Alessandro, A.; Dzieciatkowska, M.; Hansen, K. C.: Preserved proteins from extinct bison *Latifrons* identified by tandem mass spectrometry; hydroxylysine glycosides are a common feature of ancient collagen. *Mol. Cell. Proteomics* **2015**, *14*, 1946-1958.
- (66) Brown, S.; Higham, T.; Slon, V.; Paabo, S.; Meyer, M.; Douka, K.; Brock, F.; Comeskey, D.; Procopio, N.; Shunkov, M.; Derevianko, A.; Buckley, M.: Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep* **2016**, *6*, 23559.
- (67) Cleland, T. P.; Schroeter, E. R.; Feranec, R. S.; Vashishth, D.: Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proc Biol Sci* **2016**, *283*.
- (68) Welker, F.; Hajdinjak, M.; Talamo, S.; Jaouen, K.; Dannemann, M.; David, F.; Julien, M.; Meyer, M.; Kelso, J.; Barnes, I.; Brace, S.; Kamminga, P.; Fischer, R.; Kessler, B. M.; Stewart, J. R.; Paabo, S.; Collins, M. J.; Hublin, J. J.: Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne. *Proc Natl Acad Sci U S A* **2016**, *113*, 11162-11167.
- (69) Welker, F.; Smith, G. M.; Hutson, J. M.; Kindler, L.; Garcia-Moreno, A.; Villaluenga, A.; Turner, E.; Gaudzinski-Windheuser, S.: Middle Pleistocene protein sequences from the rhinoceros

genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* **2017**, *5*, e3033.

(70) Robinson, N. E.; Robinson, A. B.: Prediction of protein deamidation rates from primary and three-dimensional structure. *Proceedings of the National Academy of Sciences* **2001**, *98*, 4367-4372.

(71) Hurtado, P. P.; O'Connor, P. B.: Deamidation of collagen. *Anal. Chem.* **2012**, *84*, 3017-25.

(72) Wilson, J.; van Doorn, N. L.; Collins, M. J.: Assessing the Extent of Bone Degradation Using Glutamine Deamidation in Collagen. *Anal. Chem.* **2012**, *84*, 9041-9048.

(73) Schroeter, E. R.; Cleland, T. P.: Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Commun. Mass Spectrom.* **2016**, *30*, 251-5.

(74) Welker, F.; Soressi, M. A.; Roussel, M.; van Riemsdijk, I.; Hublin, J.-J.; Collins, M. J.: Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. *STAR: Science & Technology of Archaeological Research* **2017**, *3*, 15-27.

(75) Ren, D.; Pipes, G. D.; Liu, D.; Shih, L.-Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V.: An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Anal. Biochem.* **2009**, *392*, 12-21.

(76) Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K.: Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol. Cell. Proteomics* **2011**, *10*, O111. 009381.

(77) Simpson, J. P.; Penkman, K. E. H.; Demarchi, B.; Koon, H.; Collins, M. J.; Thomas-Oates, J.; Shapiro, B.; Stark, M.; Wilson, J.: The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *J. Archaeo.Sci.* **2016**, *69*, 29-38.

(78) Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.: Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465-W469.

(79) Buckley, M.: Ancient collagen reveals evolutionary history of the endemic South American 'ungulates'. *Proc Biol Sci* **2015**, *282*, 20142671.

(80) Hofreiter, M.; Serre, D.; Poinar, H. N.; Kuch, M.; Paabo, S.: Ancient DNA. *Nat Rev Genet* **2001**, *2*, 353-359.

(81) Meyer, M.; Arsuaga, J.-L.; de Filippo, C.; Nagel, S.; Aximu-Petri, A.; Nickel, B.; Martínez, I.; Gracia, A.; de Castro, J. M. B.; Carbonell, E.; Viola, B.; Kelso, J.; Prüfer, K.; Pääbo, S.: Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **2016**, *531*, 504-507.

(82) Orlando, L.; Ginolhac, A.; Zhang, G.; Froese, D.; Albrechtsen, A.; Stiller, M.; Schubert, M.; Cappellini, E.; Petersen, B.; Moltke, I.; Johnson, P. L. F.; Fumagalli, M.; Vilstrup, J. T.; Raghavan, M.; Korneliussen, T.; Malaspinas, A.-S.; Vogt, J.; Szklarczyk, D.; Kelstrup, C. D.; Vinther, J.; Dolocan, A.; Stenderup, J.; Velazquez, A. M. V.; Cahill, J.; Rasmussen, M.; Wang, X.; Min, J.; Zazula, G. D.;

Seguin-Orlando, A.; Mortensen, C.; Magnussen, K.; Thompson, J. F.; Weinstock, J.; Gregersen, K.; Roed, K. H.; Eisenmann, V.; Rubin, C. J.; Miller, D. C.; Antczak, D. F.; Bertelsen, M. F.; Brunak, S.; Al-Rasheid, K. A. S.; Ryder, O.; Andersson, L.; Mundy, J.; Krogh, A.; Gilbert, M. T. P.; Kjaer, K.; Sicheritz-Ponten, T.; Jensen, L. J.; Olsen, J. V.; Hofreiter, M.; Nielsen, R.; Shapiro, B.; Wang, J.; Willerslev, E.: Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **2013**, *499*, 74-78.

(83) Demarchi, B.; Hall, S.; Roncal-Herrero, T.; Freeman, C. L.; Woolley, J.; Crisp, M. K.; Wilson, J.; Fotakis, A.; Fischer, R.; Kessler, B. M.; Rakownikow Jersie-Christensen, R.; Olsen, J. V.; Haile, J.; Thomas, J.; Marean, C. W.; Parkington, J.; Presslee, S.; Lee-Thorp, J.; Ditchfield, P.; Hamilton, J. F.; Ward, M. W.; Wang, C. M.; Shaw, M. D.; Harrison, T.; Dominguez-Rodrigo, M.; MacPhee, R.; Kwekason, A.; Ecker, M.; Kolska Horwitz, L.; Chazan, M.; Kroger, R.; Thomas-Oates, J.; Harding, J. H.; Cappellini, E.; Penkman, K.; Collins, M. J.: Protein sequences bound to mineral surfaces persist into deep time. *Elife* **2016**, *5*.

(84) Blank-Landeshammer, B.; Kollipara, L.; Biß, K.; Pfenninger, M.; Malchow, S.; Shuvaev, K.; Zahedi, R. P.; Sickmann, A.: Combining de novo peptide sequencing algorithms, a synergistic approach to boost both identifications and confidence in bottom-up proteomics. *J. Proteome Res.* **2017**, *16*, 3209-3218.

(85) Shteynberg, D.; Nesvizhskii, A. I.; Moritz, R. L.; Deutsch, E. W.: Combining results of multiple search engines in proteomics. *Mol. Cell. Proteomics* **2013**, *12*, 2383-2393.

(86) Palmblad, M.; Deelder, A. M.: Molecular phylogenetics by direct comparison of tandem mass spectra. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 728-32.

(87) Önder, Ö.; Shao, W.; Kemps, B. D.; Lam, H.; Brisson, D.: Identifying sources of tick blood meals using unidentified tandem mass spectral libraries. *Nature communications* **2013**, *4*, 1746.

(88) Rieder, V.; Blank-Landeshammer, B.; Stuhr, M.; Schell, T.; Biß, K.; Kollipara, L.; Meyer, A.; Pfenninger, M.; Westphal, H.; Sickmann, A.: DISMS2: A flexible algorithm for direct proteome-wide distance calculation of LC-MS/MS runs. *BMC bioinformatics* **2017**, *18*, 148.

Partie 3 :

Analyse sans digestion des
échantillons archéologiques et
paléontologiques

I. L'approche top-down

La protéomique top-down (TDP) est l'étude des protéoformes. Ce terme désigne l'ensemble des différentes formes moléculaires sous lesquelles le produit protéique d'un gène unique peut être trouvé. Il englobe toutes les formes de variation génétique, les épissages alternatifs des transcrits ARN et les modifications post-traductionnelles (PTM) [1].

Le terme protéoforme, désigne l'unité de base du TDP : une structure primaire spécifique d'une protéine traduite à partir d'un gène spécifique possède des combinaisons de PTM, de variant d'épissage, et de variation génétique (figure 1). À l'heure actuelle, il n'y a pas d'analogue universel du numéro d'accèsion de la protéine pour les protéoformes.

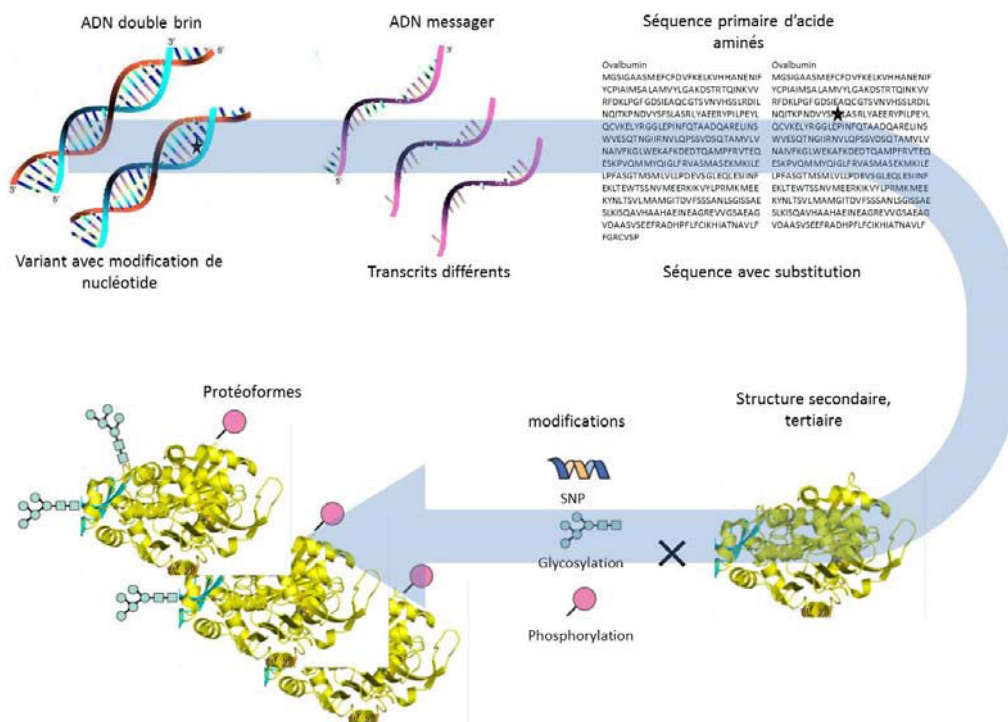


Figure 1 : Schéma de la variabilité protéique. A partir d'un gène spécifique, pouvant porter des de variations génétiques, plusieurs transcrits d'ARNm sont créés avec des variants d'épissages. Ces transcrits sont traduits en protéines portant les modifications du gène. Ces protéines traduites subissent des combinaisons de PTM. Le résultat final est la création de protéoformes.

Le TDP représente un changement de modèle majeur par rapport à la protéomique classique bottom-up (BUP), qui utilise des enzymes protéolytiques pour digérer les protéoformes et protéomes entiers en peptides avant l'analyse (figure 2). Il y

a de nombreuses raisons pour l'utilisation des techniques BUP. En général, il est plus aisé d'analyser les peptides car ils sont très solubles et se séparent facilement avant la MS. Les spectres de fragmentations sont plus facilement interprétables, ce qui permet une identification grâce aux différentes bases de données [2]. Lors de l'analyse BUP, l'étape de digestion enzymatique ne permet plus d'observer les interactions entre protéines ce qui complique l'interprétation des données biologique [3].

L'analyse TDP n'utilise pas d'enzyme pour digérer les protéines. Ceci entraîne des défis au niveau de la séparation, de la spectrométrie de masse et de l'analyse informatique. Tout d'abord, il est fondamental de noter que des informations importantes sont perdues lors de la digestion des protéoformes (figure 2).

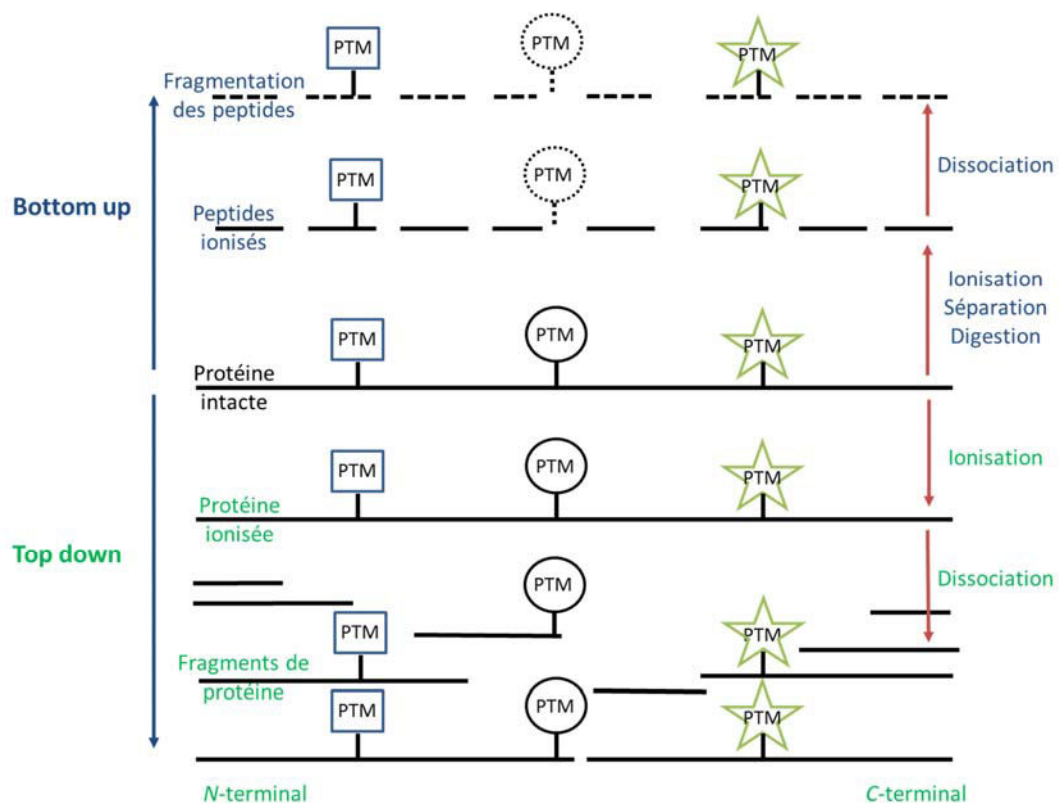


Figure 2 : Comparaison de la spectrométrie de masse top-down et bottom-up. Dans l'approche bottom-up traditionnelle, les protéines sont digérées pour obtenir des peptides avant l'introduction dans le spectromètre de masse où elles sont ensuite fragmentées et détectées. Dans la spectrométrie de masse top-down, la protéine est ionisée directement, puis fragmentée et détectée. Cette méthode qui permet d'améliorer la couverture des séquences et la détection des PTM

La découverte de l'ionisation ESI par Fenn et al. [4], [5] a permis le développement de la protéomique. Ensuite, les études fondamentales du laboratoire de McLafferty au début et au milieu des années 1990 ont contribué à formaliser l'analyse

TDP et d'établir l'utilisation de spectromètre de masse à haute résolution de type FT-ICR avec fragmentation pour l'analyse de grands polypeptides [6]. Le groupe de Kelleher et al [7].[8] ont montré que l'utilisation de la spectrométrie de masse à haute résolution et la fragmentation en source permettait d'étudier des protéines de haut poids moléculaire (la thiaminase I 42 kDa, l'anhydrase carbonique 29 kDa) et d'obtenir des couvertures de séquence proche de 100%.

La TDP est une technique plus difficile à mettre en œuvre que le BUP. De nombreuses limitations existent, comme l'augmentation de l'effet de charge due à la taille des protéines, la complexité des spectres de fragmentation (présences de fragments multichargés), la faible résolution de séparation des protéines en LC-MS.

L'analyse TPD peut être séparée en trois parties:

- Le fractionnement de mélanges complexes,
- L'acquisition de spectre de masse MS et MS/MS à haute résolution,
- Le traitement des données des bases de recherche et d'annotation [9].

II. Les premières étapes de l'analyse TDP : le fractionnement des protéines

Le protéome est extraordinairement complexe. Si nous supposons qu'une variété de PTM et des variants de séquences peuvent se produire, dans une combinaison quelconque, à travers les produits des 20 300 gènes identifiés dans le génome humain, des millions, voire des milliards de protéoformes possibles peuvent exister. Cette variabilité peut évoluer car les protéomes évoluent au cours du temps. La gamme dynamique du protéome est vaste et celle des spectromètres de masse limitée de telle sorte que seules les protéoformes les plus abondantes peuvent être identifiées si la complexité de l'échantillon n'est pas réduite avant l'analyse MS.

La première étape de l'analyse TPD a été largement étudiée pour permettre de réduire la complexité de l'échantillon avant l'analyse MS [10]. En 1998, Kelleher et al ont montré que l'utilisation de mini colonne contenant de la résine de phase inverse, et en réalisant des éluions séquentielles pouvait permettre une réduction de la complexité des échantillons avant l'analyse par spectrométrie de masse [11].

Analyse sans digestion des échantillons archéologiques et paléontologiques

Peu après, d'autres groupes de recherche ont montré que l'analyse à haut débit en spectrométrie de masse de protéines digérées et intactes était possible en utilisant la chromatographie capillaire liquide en phase inverse [12,13] ou une focalisation isoélectrique capillaire (CIEF) [14] couplée à un FTICR qui pouvaient faciliter l'étude des mélanges complexes de protéines.

Une autre méthode, l'isoélectrofocalisation, est une stratégie de fractionnement efficace avant le TDP, bien que la récupération des protéines reste un problème majeur de la technique [10,15]. L'électrophorèse capillaire de zone a également prouvé son efficacité, en raison de ses faibles débits compatibles avec l'ESI, ce qui permet de détecter des quantités de protéines de quelques attomoles [16].

Le développement de l'électrophorèse sur gel avec élution liquide de fractions (Gelfree) par Tran et Doucette [17] a grandement facilité la séparation multidimensionnelle pour l'analyse top-down en fonction du poids moléculaire des protéines à partir de protéomes complexes. De multiples études ont combiné la séparation Gelfree avec d'autres techniques de fractionnement pour développer des workflows puissants à grande échelle permettant l'analyse de protéines intactes dans des protéomes [18-23]. Une limitation importante de la Gelfree, est l'utilisation du SDS (sodium dodécylsulfate), qui est un tensioactif qui interfère fortement avec l'ESI et qui doit donc être éliminé des échantillons.

La méthode pour éliminer ce SDS est le plus souvent effectuée par une précipitation au méthanol/chloroforme, mais il est possible d'utiliser des spins colonnes, qui permettent une analyse reproductible ou automatisable. Kim et al [24] ont développé une méthode pour permettre l'élimination du SDS en quelques minutes est l'utilisation du fractionnement par flux d'écoulement asymétrique entre la Gelfree et la MS partir d'échantillons en seulement 5 minutes.

Très récemment une équipe a réalisé des analyses top-down sur des protéines extraites de gel SDS-PAGE hydrolysable [25], et d'autre ont mis en place

Un autre développement récent et intéressant dans le domaine des séparations, intègre la séparation des ions avec la mobilité ionique des fragments des protéoformes, par la dissociation induite par collision (CID) [26].

Analyse sans digestion des échantillons archéologiques et paléontologiques

La chromatographie d'interaction hydrophobe est aussi utilisée comme technique de préfractionnement pour la séparation à haute résolution des protéines intactes, grâce à l'adoption de tartrate d'ammonium en tant que sel compatible avec la MS et qui est complémentaire à celle de la phase inverse [27]. La majorité des progrès récents dans le préfractionnement des protéoformes avant la MS ont servi à augmenter la dynamique d'identification du protéome réalisé par TDP à grande échelle. La figure 3 montre des méthodologies d'analyse pour les expériences top-down, en utilisant plusieurs dimensions de séparation.

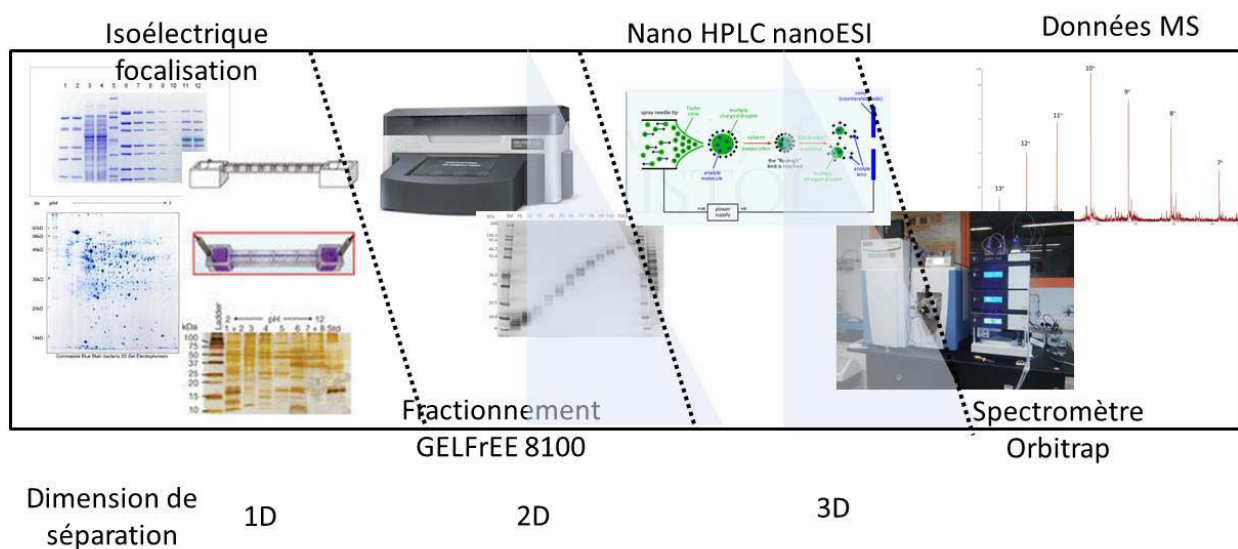


Figure 3 : Méthode d'analyse pour les expériences top-down. La première étape consiste à réaliser un préfractionnement des protéines de l'échantillon puis séparation des protéines contenues dans les fraction par nanoHPLC et analyse par spectrométrie de masse [28].

III. Instruments et techniques de fragmentation

En général, la caractérisation par spectrométrie de masse des polypeptides nécessite trois étapes principales:

- L'ionisation, qui permet le passage du polypeptide à partir du liquide à la phase gazeuse, en tant que particule chargée qui peuvent être manipulées à l'intérieur du spectromètre de masse [4],
- La détermination de la masse intacte du polypeptide par MS,

Analyse sans digestion des échantillons archéologiques et paléontologiques

- L'acquisition d'informations des séquences peptidiques par l'activation et la fragmentation des ions précurseurs sélectionnés ou MS/MS,

Dans le cas des protéoformes, ces opérations sont compliquées à cause de la taille de la chaîne polypeptidique. Bien que des exemples de protéines intactes analysées par MALDI existent [29], l'analyse TDP emploie traditionnellement l'ionisation ESI comme la technique d'ionisation de choix.

Les spectromètres de masse de haute résolution sont fortement préférés pour l'analyse TDP, en particulier lors de l'analyse des ions fragments créés lors d'une expérience en tandem. Pour les données MS sur les protéoformes intactes, la résolution isotopique des clusters d'ions est utile mais pas nécessaire pour déterminer leur état de charge. Cependant, les spectres de fragmentation des protéoformes intactes sont compliqués et contiennent de nombreux ions avec différents états de charge qui peuvent se chevaucher. Un pouvoir de résolution supérieur à 50 000 est généralement nécessaire pour déconvoluer correctement les spectres et interpréter les fragmentations obtenues. Le pouvoir de résolution nécessaire pour l'analyse TPD peut être obtenu par de nombreux appareils hybrides à temps de vol (Q-TOF) et analyseurs à transformée de Fourier (FT-ICR, Orbitrap).

Le spectromètre de masse à transformée de Fourier FT-ICR était l'instrument dédié à l'analyse TDP grâce aux avancées technologiques clés comme l'amélioration du piège de Penning, l'accumulation sélective d'ions avec la résolution du quadripôle et l'augmentation du signal sur bruit grâce à l'application d'un gradient de tension continue sur l'octopole au cours d'accumulation durant le transfert des ions [30]. D'autres améliorations au niveau de la cellule de piégeage d'ions ont permis de diminuer le vide pour tenir compte de l'augmentation de l'énergie cinétique des ions [31].

Récemment, l'Orbitrap est devenu une alternative au FT-ICR car cet instrument peut être utilisé dans de nombreux champs de recherche [32-34]. La résolution élevée (1 000 000 pour l'Orbitrap Fusion Lumos) et la sensibilité (quelques femtomole de protéine) de l'Orbitrap ont été rapidement mises en évidence pour faciliter les mesures des protéines dont le poids moléculaire est inférieur à 30 kDa [35].

Ces appareils ne sont pas les seuls à être utilisés pour l'analyse top-down, l'utilisation d'instrumentation à temps de vol pour l'analyse des protéines intactes est

possible. Par exemple, l'utilisation d'un spectromètre de masse de type Q-TOF (quadripolaire et temps de vol) a permis l'analyse de protéines de moyenne et grande taille [36], [37].

Les spectromètres de masse, généralement utilisés pour l'analyse TDP combinent souvent un analyseur haute résolution dans des architectures hybrides. La présence de multiples analyseurs de masse en série permet non seulement des manipulations d'ions plus complexes, mais, est aussi avantageuse pour la réalisation d'activation d'ions / fragmentation pour de gros ions biomoléculaires.

La technique de fragmentation « higher energy collisional dissociation » (HCD) ou « Collision-Induced Dissociation » qui sont réalisées dans des dispositifs quadripolaires ou multipolaires, permettent des identifications de protéines mais avec des pourcentages de couverture de séquences de l'ordre de 50 %.

Il est possible d'augmenter ce pourcentage, par des techniques d'activation d'ions radicaux, à savoir la dissociation par capture d'électron (ECD) et la dissociation par transfert d'électron (ETD) [38], mais la faible efficacité de fragmentation limite l'utilisation de ces méthodes. L'ECD et ETD sont généralement appliquées à la caractérisation des protéines de petite et grande taille [39], [40].

Les photons infrarouges à faible énergie ont été utilisés pour augmenter la l'efficacité de l'ETD, dans un processus appelé activation d'ion ETD. Sur des protéines modèles, son utilisation permet d'obtenir un rapport signal sur bruit élevé au niveau des ions fragments [41]. De plus, l'utilisation de photons ultra-violet à 193 nm a été utilisée pour caractériser les protéines ribosomiques en LC-MS/MS sur des échantillons complexes. Cette fragmentation est appelée photodissociation par ultra-violet (UVPD) [42]. Cette méthode utilise l'absorption du rayonnement UV par la chaîne d'acides aminés de la protéine qui initie un clivage de la protéine et permet d'obtenir un haut pourcentage de couverture [43].

Une autre étude réalisée au synchrotron SOLEIL a permis à partir de la fragmentation appelée vacuum UV (VUV) avec des photons de 193 nm et 157 nm, de déterminer le site de fixation de tanin procyanidine sur la protéine humaine IDP (intrinsically disordered protein) [44].

Par exemple, Shaw et al. [45] ont utilisé l'UVPD sur un instrument Orbitrap pour caractériser une protéoforme de Pin1 humain di-oxydé. Cette protéine est impliquée dans la maladie d'Alzheimer et dans différents cancers.

La combinaison de ces différents types de fragmentation ETD et HCD (ou CID) permet d'obtenir des spectres MS/MS plus informatifs par rapport à l'utilisation d'un seul type de fragmentation [46]. La combinaison de ces types de fragmentation est appelé EThcD et ETciD. L'utilisation de ces méthodes de fragmentations a permis d'identifier les protéoformes de l'histones H3 et de localiser jusqu'à huit sites de méthylations sur cette protéine [47]. Dans le domaine de la biopharmaceutique, une autre étude a montré qu'il est possible d'obtenir une couverture de séquence de 95% sur la protéine GCSF (granulocytic colony stimulating factor) par le biais de la combinaison de ETD, HCD et CID [48]

IV. Applications de l'analyse top-down à des échantillons complexes.

Les premières mises en place des méthodes TDP entre 2000 et 2010 ont porté sur des bactéries, des archées et des levures car elles permettaient de développer toutes les étapes (préparation de l'échantillon, acquisition de données, et analyse informatique). Par exemple, Bungert et al. [49] ont appliqué les récents développements liés à ETD et aux fractionnements multidimensionnels pour identifier 322 protéoformes correspondant à 174 protéines dans *Escherichia coli* en utilisant des bases de données personnalisées et des algorithmes de score. Un grand nombre d'analyses TDP ont permis d'approfondir le protéome de *Saccharomyces cerevisiae* et ont permis de développer de nouvelles approches de préfractionnement en deux dimensions pour identifier les protéines non annotées qui ne sont pas identifiées par l'analyse bottom-up [50].

Pour transférer les progrès de l'analyse TDP développée à l'aide des systèmes microbiens aux cellules humaines, Roth et al. [51] ont réalisé une étude protéomique axée sur les leucocytes humains primaires. Ils ont utilisé une stratégie de séparation bidimensionnelle pour la caractérisation des protéines par top-down. Ce travail a conduit à l'identification de 133 protéoformes identifiées à partir de 67 gènes uniques.

Tran et al. [18] ont utilisé plusieurs avancées pour améliorer le préfractionnement en introduisant une séparation à quatre dimensions : une focalisation isoélectrique suivie d'une séparation par Gelfree, puis une LC-MSMS pour les bas et hautes masses moléculaires à partir de lysat de cellules humaines.

Ces récentes études du protéome humain par l'analyse TPD ont montré la capacité de l'analyse TDP pour caractériser des protéoformes en haut débit et pour permettre leur application dans le domaine de la biologie.

V. La quantification des protéines entières

La stratégie de quantification des protéines dans les études protéomiques représente un autre thème dans la progression de l'analyse TDP (figure 4). La capacité de quantifier les abondances des protéoformes à travers des échantillons va ouvrir la voie à une augmentation d'analyse TDP dans les domaines cliniques et en particulier dans le domaine de la découverte de biomarqueurs.

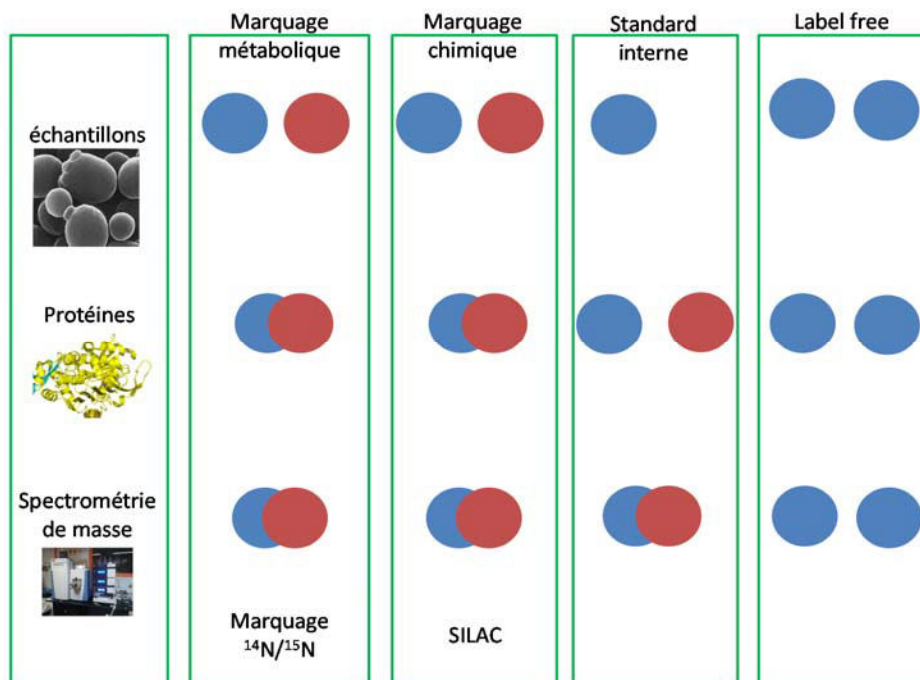


Figure 4 : Aperçu de la protéomique quantitative par spectrométrie de masse. Deux types de méthodes de quantification existent : la quantification relative qui permet d'avoir une estimation de la quantité des protéines dans les échantillons et la quantification absolue qui donne une concentration des protéines ciblées.

Analyse sans digestion des échantillons archéologiques et paléontologiques

De nombreuses stratégies de marquage *in vitro* et *in vivo* ont été explorées pour quantifier les protéoformes à travers des protéomes entiers. Le marquage des cystéines et le marquage métabolique avec de l'azote $^{14}\text{N}/^{15}\text{N}$ ont été utilisés avec succès chez la levure. Du et al. ont quantifié 50 protéines et ont ainsi identifié 14 protéines uniques [52]. Cette méthode est limitée à l'utilisation d'isotopes stables incorporés dans les cellules levures ou bactéries pendant leur croissance. Ces études ont permis d'établir une preuve de concept d'étude quantitative de protéines intactes par l'analyse top-down.

Peu après, Collier et al. [53] ont testé la faisabilité de quantification par TDP à l'aide d'un marquage grâce à des acides aminés marqués dans des cultures cellulaires (SILAC), apparentés aux méthodes de quantification en bottom-up [54]. La limitation de cette technique est le pourcentage d'intégration d'arginines marquées dans les protéines de haute masse moléculaire [53].

John F. Kellie et al [55] ont développé une méthode TDP robuste, pour profiler des espèces d'alpha-synucléine provenant d'autopsies de tissus cérébraux du cortex frontal d'humain provenant de témoins et de patients atteint de la maladie de Parkinson. La quantification de alpha-synucléine a été réalisé dans des échantillons de cerveaux en ajoutant alpha-synucléine ^{15}N et avec une courbe de calibration de cette protéine. Dans cette étude, tous les états de charges de la protéine ont été extraits et intégrés pour la quantification.

La méthode de quantification sans marquage, dite « label free », pour quantifier les protéomes peut être appliquée pour l'analyse TPD. Cette méthode est facile à mettre en œuvre, car la préparation des échantillons ne nécessite aucun marquage. Cette technique peut être automatisée pour permettre une quantification des protéines d'un protéome avec des analyses LC-MS/MS. Il faut alors obtenir les rapports masse sur charge, les temps de rétention, les d'intensités des pics d'intérêt et réaliser l'analyse en répliquat [56]. En 2010, Mazur et al. [57] ont publié une preuve du concept d'analyse quantitative sans marquage différentiel sur les protéoformes de l'apolipoprotéine dans les HDL de patients présentant un bas ou un haut niveau de cholestérol. En 2014, Ntai I et al [23] ont réalisés une quantification label free sur une souche de levure de *Saccharomyces cerevisiae* sauvage versus un mutant. Ils ont quantifiés 100

proteoformes qui possèdent une expression différentielle et une hyperacétylation des histones.

Même avec l'avènement des études de quantification sans marquage pour quantifier les protéoformes dans l'ensemble des protéomes, des limitations existent toujours. Par exemple, les analyses statistiques dans la plupart de ces études ont été limitées à un *t* test de Student, ce qui est insuffisant pour répondre aux nombreuses sources de variations techniques. Cette méthode de quantification sans marquage représente un outil pour l'avenir de la recherche car elle peut être appliquée à divers types d'échantillons qui ne peuvent être métaboliquement marqués. L'amélioration des méthodes statistiques pourra permettre de traiter un grand nombre de données.

VI. L'analyse bio-informatique

Bien que les dispositifs de séparation soient performants et que les spectromètres de masse soient de plus en plus puissants pour générer des données impressionnantes en qualité et en quantité, ces données doivent être traitées de manière adéquate afin d'identifier et de caractériser les protéoformes.

La première étape d'interprétation des spectres top-down est généralement de transformer le fichier obtenu par le spectromètre en fichiers lisibles par les logiciels. Le plus souvent ces fichiers sont transformés en fichier mzXML qui est un fichier standardisé à l'aide du logiciel MSconvert [58]. Le spectre est ensuite déconvolué à l'aide du logiciel MSdeconv [59] qui permet de transformer les masses des fragments multichargés en monochargés, puis cette liste de masse peut être interrogée contre une base de données contenant des protéines pour permettre d'obtenir une identification avec un score appelé PrSM (Protein-Spectrum-Match).

De nombreux logiciels ont été mis au point pour permettre l'identification des protéines intactes. Une liste des principaux logiciels est indiquée dans le tableau 1.

Logiciel	fonction	Références
ProSight PTM 2.0	Identification des protéines à partir des données de fragmentation, après déconvolution des spectres	[60] [61]
ProsightPC	Logiciel de la société Thermo Scientific qui utilise différents modes de recherche	[62]
BIG Mascot MascotTD	Utilise le logiciel Mascot et permet d'analyser des spectres de fragmentations de protéines de 16 kDa à 110 kDa.	[63,64]
PIITA	Logiciel libre d'accès pour l'identification des protéines pas analyse TDP	[64]
YADA		[65]
pTOP		[66]
MS Align+		[67]
TopPIC		[68]
MASH		[69]
Byonic	Logiciel de la société Protein Metric pour l'analyse TDP ou BUP.	

Tableau 1 : Liste des logiciels utilisés pour l'analyse des données TDP.

VII. L'analyse TPD pour l'archéologie et la paléontologie

La possibilité d'analyser les protéines anciennes dans leur forme intacte est l'un des défis actuels pour permettre d'obtenir des informations sur leur dégradation, sur les modifications chimiques et sur les substitutions d'acides aminés.

Les premières expériences fournissant des informations sur les poids moléculaires des protéines intactes extraites d'échantillons anciens (fossiles) ont été réalisées à l'aide de techniques comme le gel électrophorétique à la fin des années 1980 [70,71]. Cette technique a permis de vérifier la teneur en protéines après une extraction. Avec cette méthode, il a été possible d'observer sur le gel la présence de protéines coupées suite à des mécanismes de dégradation.

Les premières expériences pour analyser des protéines intactes en utilisant la spectrométrie de masse résultant de l'extraction de fossiles ont été effectuées à l'aide de l'analyse MALDI-TOF dans les années 2000 [72,73]. Les spectres ont permis de

déterminer des poids moléculaires moyens de protéines (erreur de mesure liée à l'hétérogénéité de la protéine ciblée et à la résolution de l'instrument).

L'analyse des protéines natives (pour étudier les modifications des protéines et les dégradations de la séquence des protéines) et l'ajout des expériences de TDP (pour obtenir la séquence d'acides aminés et la localisation potentielle des modifications) représentent des outils puissants pour l'étude des échantillons du patrimoine culturel. Ces techniques sont capables de fournir de nouvelles informations mais elles restent un challenge analytique.

VIII. Objectifs

L'analyse BUP sur les échantillons archéologique ou paléontologique est très développée comme nous l'avons vu dans le chapitre III, partie 2. Cette technique permet d'identifier les protéines et les modifications présentes sur la protéine, mais certains points ne peuvent être étudiés comme la dégradation *in situ*.

Dans cette partie, est présenté le développement d'une méthode de préparation des échantillons pour l'analyse TDP et son application à un échantillon archéologique dans le but d'identifier les protéines présentes dans l'échantillon et les modifications présentes sur ces protéines. Il faut noter que cette étude représente la première application de l'analyse top-down à des échantillons du patrimoine culturel.

IX. Matériels et méthodes

1. Echantillons utilisés pour l'optimisation du protocole

L'optimisation du protocole d'extraction a été réalisée à partir d'une solution d'alpha lactalbumine (Sigma Aldrich) préparée à une concentration de 1 mg/mL dans de l'eau mQ. Cette solution va permettre l'optimisation de la méthode de préparation des protéines en utilisant des filtres Amicons®.

Le protocole mis en place a ensuite été testé sur deux échantillons complexes pour valider notre méthodologie. 1 mg de bactérie *Escherichia coli* et 1 mg de muscle d'autruche ont été mis en présence de 500 µL de tampon de lyse (4 M urée, 0,2% DCA, 50 mM Tris HCl, pH 7,5), puis soniqué 30 secondes avec une sonde ultrasons à 50% de puissance dans la glace. La solution a été incubée à 4°C pendant 1 heure sous agitation puis centrifugée 30 minutes à 10 000 g à température ambiante. Le surnageant a été récupéré puis dosé avec la méthode BCA à 575 nm. Les échantillons ont ensuite été traités avec la méthode de préparation des protéines en utilisant des filtres Amicons®.

Un échantillon de lait de vache demi-écrémé provenant du magasin Auchan V2 (Villeneuve d'Ascq) a été traité avec notre méthodologie pour obtenir un témoin pour notre échantillon archéologique. 100 µL de lait de vache ont été additionnés à 500 µL de tampon de lyse (4 M urée, 0,2% DCA, 50 mM Tris HCl, pH 7,5). La solution a été incubée à 4°C pendant 1 heure sous agitation puis centrifugée 30 minutes à 10 000 g à température ambiante. L'échantillon a ensuite été traité avec la méthode de préparation des protéines en utilisant des filtres Amicons®.

2. Echantillon archéologique

L'échantillon archéologique utilisé pour cette étude nous a été fourni par le Dr Nicolas Garnier (S.A.S, LNG, Vic Le Comte). Cet échantillon provient d'une céramique retrouvée dans la commune de Bad Zurzach en Suisse. Cette amphore a été prélevée sur le site de la nécropole du camp militaire établie à l'époque claudienne (de 41 à 54 après Jésus Christ) dans la cité antique de Tenedo.

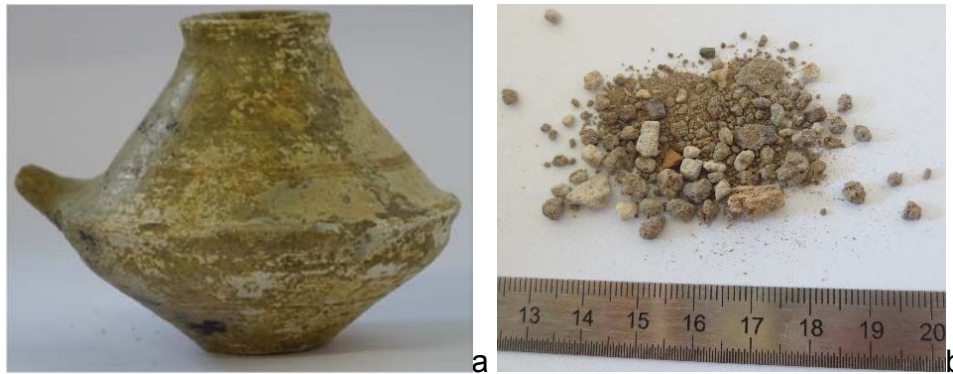


Figure 5 : Amphore Zurarch 55 (a) et prélèvement intérieur de céramique de l'amphore (b).

110 mg de l'échantillon de Zurzarch 55 (figure 5) ont été broyés dans un mortier en verre dépoli.

Dix milligrammes de la poudre obtenues ont été mis en présence de 100 μ L du tampon de lyse eFASP (8 M d'urée, 4% SDS, 0,2% DCA, 50 mM DTT, 100 mM ammonium bicarbonate pH 8,8) pour réaliser l'analyse bottom-up, puis agité à 4°C toute la nuit.

Cent milligrammes de la poudre obtenues ont été mis en présence de 500 μ L de tampon d'extraction (4 M urée, 0,2% DCA, 0,2% SDS, 50 mM DTT, 50 mM Tris HCl pH 7,5), pour l'analyse top-down, puis agité à 4°C toute la nuit.

3. Digestion des protéines avec la méthode eFASP

La méthode a été détaillée dans le chapitre III partie 2. La digestion eFASP a été réalisée avec des filtres Amicon® avec un seuil de coupure de 10 kDa et en utilisant une digestion par la trypsine.

4. Méthode de préparation des protéines

Les filtres Amicons® avec un seuil de coupure de 3 KDa ont été incubés toute la nuit dans la solution de passivation contenant 5% (v/v) de TWEEN®-20. 100 μ L de tampon d'échange (4 M d'urée; 0,2% DCA, 50 mM tris HCl pH 7,5) ont été ajoutés dans l'Eppendorf et transférés dans un Amicon® puis centrifugés 30 minutes à 10 000 g, le filtrat a été éliminé. 200 μ L de tampon d'échange ont été à nouveau ajoutés dans l'Amicon® et centrifugés, l'opération a été répétée deux fois.

Analyse sans digestion des échantillons archéologiques et paléontologiques

Les protéines ont été alkylées pendant une heure à température ambiante dans le noir à l'aide de 100 µL de tampon d'alkylation (4 M urée, 50 mM iodoacétamide, et 50 mM Tris HCl, pH 7,5) puis centrifugés 30 minutes à 10 000 g, le filtrat a été éliminé. Après l'alkylation des protéines, 200 µL de tampon d'échange ont été ajoutés dans l'Amicon®, centrifugé, puis le filtrat a été éliminé. 200 µL de tampon de lavage (50 mM Tris HCl pH 7,5) ont été ajoutés à l'Amicon® puis centrifugé, le filtrat a été éliminé. 200 µL d'eau milliQ ont été de nouveaux ajoutés dans l'Amicon® et centrifugés, l'opération a été répétée deux fois. Le surnageant contenu dans le filtre Amicon® a été récupéré, puis a été analysé par spectrométrie de masse (figure 6).

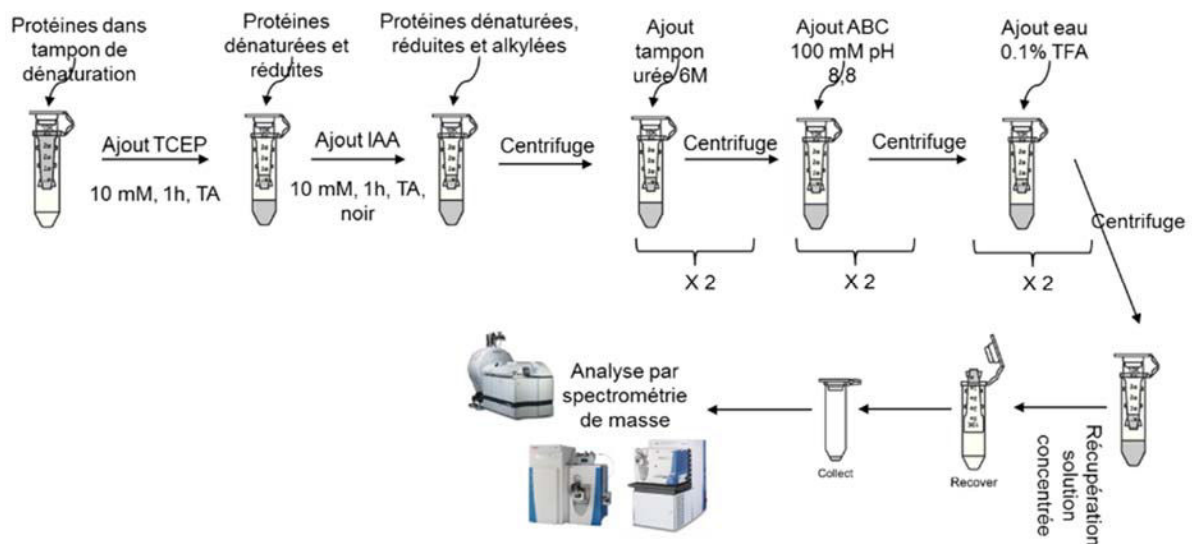


Figure 6 : Schéma de la méthode de préparation des protéines natives pour une analyse TDP.

5. Analyse des protéines par MALDI TOF

Les protéines ont été analysées par MALDI TOF en mélangeant 1 µL d'échantillon avec 2 µL de matrice HCCA à 10 mg/mL (30% H₂O, 70% ACN, 0,1% TFA). 1 µL a été déposé sur la plaque MALDI. L'analyse a été réalisée en mode linéaire, positif et 1000 scans ont été enregistrés.

6. Analyse des protéines par nanoLC-MS/MS

Les échantillons ont été repris dans 10 µl de solution H₂O contenant 0,1% d'acide formique. Le spectromètre de masse Orbitrap Q Exactive plus utilise un système microfluidique HPLC U3000 RSLC Thermo Fisher Scientific. Une pré-colonne Thermo

Analyse sans digestion des échantillons archéologiques et paléontologiques

Scientific, Acclaim PepMap100 C18 (5 μm , 300 μm i.d x 5 mm) a été utilisée, dans laquelle a été injecté 1 μl du mélange de peptides et protéines avec un solvant A (95% H_2O , 5% acétonitrile, 0,1% acide formique) pendant 3 minutes à un débit de 10 $\mu\text{l}\cdot\text{min}^{-1}$. Les peptides ont été ensuite séparés sur une colonne à phase inverse C18 (Acclaim PepMap100 C18, 3 μm , 75 mm i.d. x 500 mm), grâce à un gradient linéaire (5-40%) de solution B (25% H_2O , 75% ACN, 0,1% acide formique) à un débit de 250 nL/min en 160 minutes puis 100% de solution B en 5 minutes. La colonne a été lavée durant 5 minutes avec le tampon B puis rééquilibrée avec le tampon A. La colonne et la pré-colonne étaient placées à une température de 45 °C. La durée totale de l'analyse a été de 180 minutes. Les paramètres utilisés pour l'analyse bottom-up et top-down sont regroupés dans le tableau 2.

Paramètres analyse peptides			Paramètres analyse protéines		
	MS	MS/MS		MS	MS/MS
Résolution	70 000	35 000	Résolution	70 000	70 000
AGC target	1e ⁶	5e ⁵	AGC target	1e ⁶	1e ⁶
Maximum IT	80 ms	120 ms	Maximum IT	100 ms	200 ms
Scan range	350 à 1600 m/z		Scan range	350 à 2000 m/z	
Scan	1 scan / spectre		Scan	4 scans / spectre	
HCD	28 ev		HCD	30 ev	
isolation	1,6 m/z		isolation	5 m/z	
exclusion	1, 8 et plus		exclusion	1, 2, 3	

Tableau 2 : paramètres de l'Orbitrap Q-Exactive plus pour l'analyse BUP et TDP.

Pour l'analyse des protéines, l'Orbitrap Q-Exactive était en mode protéine (optimisation de pression au sein de la C-trap).

7. Analyse Bio-informatique

L'analyse du fichier obtenu par l'Orbitrap de la digestion eFASP a été réalisée avec le logiciel Proteome discover 2.1. Le fichier a été interrogé sur toute la base de données NCBI avec le moteur de recherche SEQUEST. Une erreur de masse de +/- 10 ppm pour la MS et 0,01 Da pour la MS/MS a été appliquée. Les modifications fixes des peptides recherchées sont la carbamydométhylation des cystéines. Les modifications variables recherchées sur les peptides sont l'oxydation des méthionines et des prolines, la déamidation des arginines et glutamines. La trypsine a été utilisée comme enzyme de digestion avec la possibilité d'avoir 3 sites de coupure manqués.

Analyse sans digestion des échantillons archéologiques et paléontologiques

Le fichier obtenu par l'Orbitrap pour l'analyse top-down a été analysé avec le logiciel Byonic (Protein Metrics). Le fichier a été interrogé sur toute la base de données *Bos Taurus* (Swissprot, 2017, 32 202 séquences). Une erreur de masse de +/- 20 ppm pour la MS et MS/MS a été appliquée. Les modifications variables utilisées pour la recherche dans les banques de données sont les oxydations des méthionines et prolines, les déamidations des arginines et glutamines.

X. Résultats et discussions

1. Méthodes de préparation des protéines intactes

La méthode de préparation des protéines intactes a été mise au point en se basant sur la méthode eFASP de Erde J. et al [74]. Les différences majeures avec la méthode eFASP sont liées à la composition des tampons et l'absence d'utilisation d'enzymes de digestion.

Dans la méthode eFASP, la première étape est l'incubation des filtres Amicon® dans le tween 20 à 0,5%. Cette étape a été conservée car elle permet une diminution des fixations des protéines sur la membrane.

Ensuite, l'échantillon est dénaturé dans une solution différente de la méthode eFASP. Le SDS a été éliminé pour éviter les contaminations mais il est possible de réaliser une extraction chloroforme/méthanol pour l'éliminer mais une perte d'échantillon est observée [75].

Il est possible d'ajouter du dithiotreitol (DTT) pour réduire les ponts disulfures des protéines. Ensuite, ces protéines ont été lavées avec une solution 4 M d'urée, 50 mM tris HCl pH 7,5. Puis une étape d'alkylation a été réalisée pour protéger les fonctions thiol des cystéines. Cette étape se réalise à température ambiante car, à 37°C, il est possible de créer des modifications de types carbamylation sur la protéine à cause de la présence de l'urée [76,77]. Les échantillons ont été une nouvelle fois lavés pour éliminer l'excès du réactif d'alkylation.

L'étape suivante consiste à réaliser un échange de tampons, dans un premier temps, avec un tampon Tris HCl (50 mM pH 7,5), puis dans un second avec de l'eau mQ. Les protéines peuvent être ensuite analysées en spectrométrie de masse MALDI-TOF ou en couplage nano LC-MS/MS.

La figure 7 montre les spectres de l'alpha lactalbumine préparée avec cette méthode. Six tampons ont été testés :

- Eau,
- 100 mM ammonium bicarbonate (ABC) pH 8,8,

Analyse sans digestion des échantillons archéologiques et paléontologiques

- 6 M urée avec 100 mM ABC pH 8,8,
- 6 M urée avec 0,2% déoxycholate (DCA),
- 100 mM ABC pH 8,8, 4 M urée, 0,2% déoxycholate (DCA), 50 mM tris HCl pH 7,5,
- 6 M urée avec 0,2% déoxycholate (DCA), 0,2% sodium dodécylsulfate (SDS) et 100 mM ABC pH 8,8.
-

L'alpha lactalbumine a été détecté dans tous les cas. Le signal de la protéine est plus faible pour le tampon contenant du SDS. Ceci signifie que le détergent n'est pas totalement éliminé et provoque une perte de signal. Bien sûr, il est possible de réaliser une purification pour éliminer le SDS restant en utilisant des spin cartouche « detergent removal » (Thermo Scientific) ou une extraction chloroforme méthanol mais cela ajoute une étape et augmente la probabilité de perdre de l'échantillon.

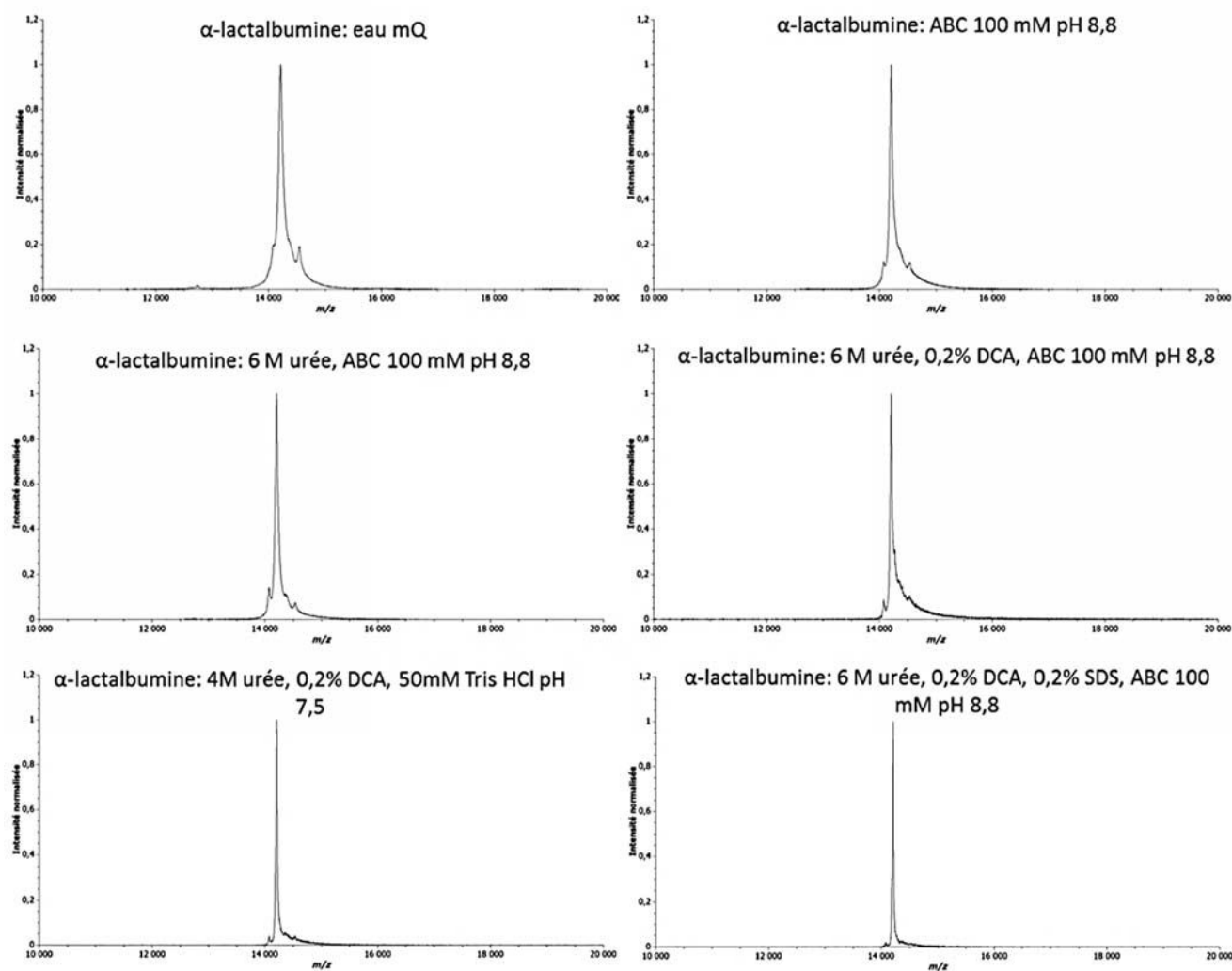


Figure 7 : Spectres MALDI de l'alpha lactalbumine traitée avec notre protocole dans différents tampons. Les spectres ont été réalisés en mode positif, linéaire avec la matrice HCCA. En (A) la protéine témoin dans l'eau, (B) 100 mM ammonium bicarbonate (ABC) pH 8,8, (C) 6 M urée, 100 mM ammonium bicarbonate (ABC) pH 8,8, (D) 6 M urée, 0,2% déoxycholate (DCA), 100 mM ammonium bicarbonate (ABC) pH 8,8, (E) 4 M urée, 0,2% déoxycholate (DCA), 50 mM Tris HCl pH 7,5, (F) 6 M urée, 0,2% déoxycholate (DCA), 0,2% sodium dodécylsulfate (SDS), 100 mM ammonium bicarbonate (ABC) pH 8,8.

Cette méthode a été appliquée sur un échantillon de protéines de lysat de bactérie *Escherichia coli* et d'extraction de protéines provenant de muscle d'autruche dans le tampon 4 M urée, 0,2% déoxycholate (DCA), 50 mM Tris HCl pH 7,5. Cent microgrammes ont été traités avec notre méthode sans réaliser de réduction et alkylation.

Des protéines de poids moléculaire de 10 à 20 kDa ont été observées sur le spectre MALDI TOF de lysat d'*Escherichia coli* et de 10 à 70 kDa pour le lysat de muscle d'autruche (figure 8). La méthode de préparation permet de traiter des protéines seules ou en mélange. L'utilisation du filtre permet l'utilisation de détergent (avec des concentrations limites) dans la solution d'extraction, car ils seront éliminés par les centrifugations.

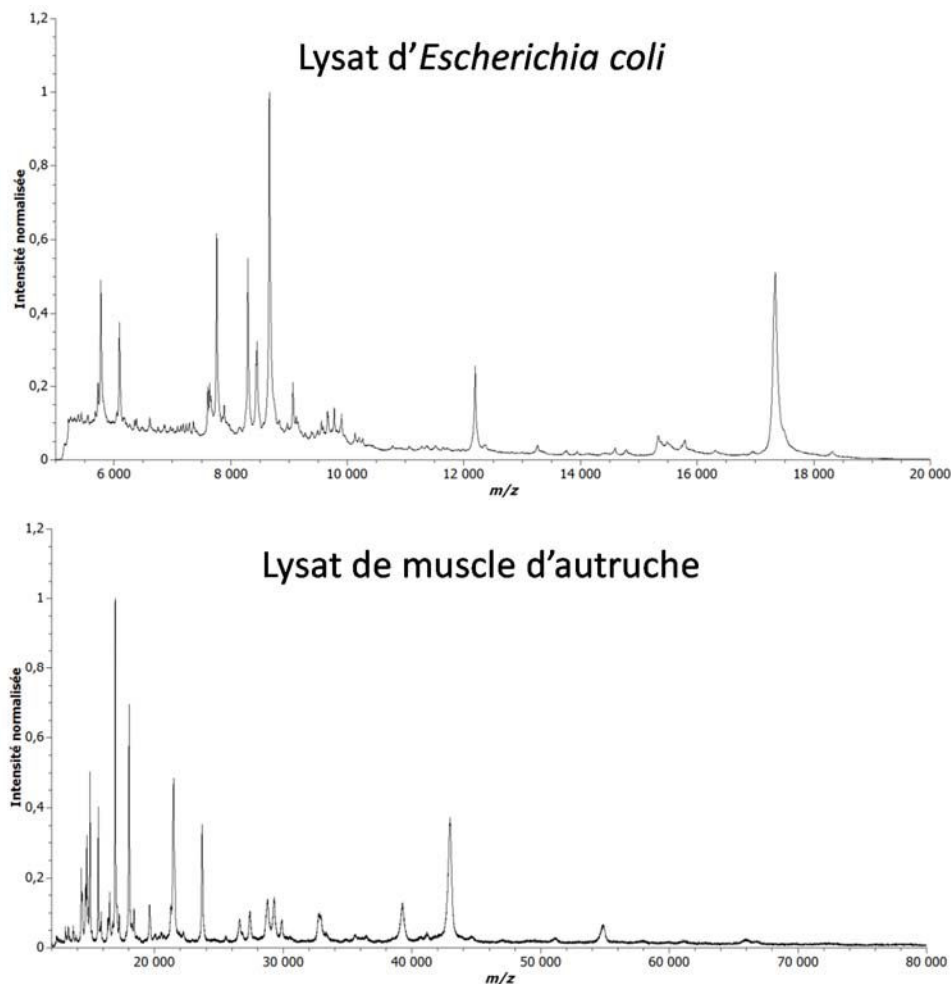


Figure 8 : Spectres MALDI-TOF de l'extrait de bactérie et de muscle traité avec la méthode eFASP modifié dans le tampon 4 M urée, 0,2% déoxycholate (DCA), 50 mM Tris HCl pH 7,5.

Comme décrit dans l'introduction, de nombreuses méthodes de préparation d'échantillons en vue de l'analyse sans digestion existent mais contrairement à ces méthodes, celle mise au point dans le cadre de ce travail est rapide et permet de traiter l'échantillon de différentes façons (réalisation ou non des étapes de réduction et alkylation).

2. Analyse de l'échantillon Zubarch 55 par nanoLC-MSMS eFASP

10 mg de l'échantillon Zurzarch 55 ont été digérés avec la méthode eFASP à l'aide d'un Amicon® avec un seuil de coupure de 10 kDa puis analysés par nanoLC Orbitrap Q-Exactive plus. L'analyse bio-informatique a été réalisée avec le logiciel Proteome Discoverer 2.1 en utilisant la base de données Swissprot All Entries.

161 protéines ont été identifiées à partir de la base de données Swissprot All Entries. Parmi ces 161 protéines, 21% ont été identifiées comme étant des kératines, 28,5% sont des protéines identifiées chez le bovin et 35% chez l'humain. Connaissant le contexte archéologique de l'échantillon (amphore biberon), notre analyse s'est focalisée sur les protéines provenant de l'espèce *Bos taurus*.

Les peptides identifiés dans l'échantillon après digestion par la trypsine ont une masse moyenne de 2000 Da (jusque 5000 Da) et l'erreur de masse de ces peptides est centrée autour de 0,5 +/- 3 ppm (figure 9). Ce résultat indique que la digestion a fonctionné et qu'il est possible d'obtenir une haute précision de mesure avec l'Orbitrap.

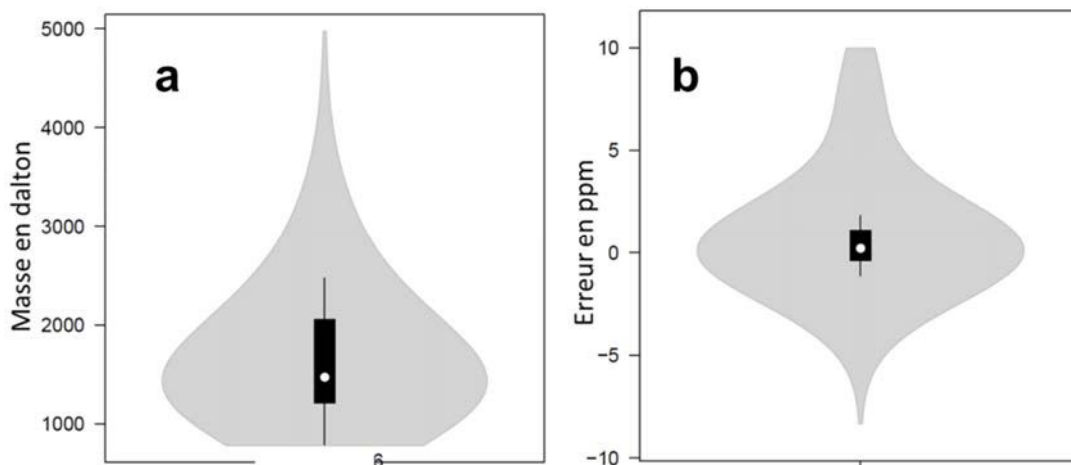


Figure 9 : Violin plot représentant la répartition des masses des peptides détectés dans l'échantillon après digestion (a) et un violin plot représentant l'erreur en ppm des peptides identifiés (b).

Analyse sans digestion des échantillons archéologiques et paléontologiques

Dans le tableau 3 sont regroupés les protéines identifiées dans l'échantillon en éliminant les contaminants de type kératine présents en grande quantité et en réalisant l'interrogation des données sur la base de données de Swissprot *Bos Taurus* contenant 32 204 séquences protéiques.

Accession	Description	Couverture %	Peptides	PSMs	Unique Peptides	MW [kDa]	calc. pi	emPAI	Score Sequest HT
P24627	Lactotransferrin OS= <i>Bos taurus</i> GN=LTF PE=1 SV=2	58,76	42	141	29	78,0	8,32	44,61	538,98
P02769	Serum albumin OS= <i>Bos taurus</i> GN=ALB PE=1 SV=4	63,26	34	140	32	69,2	6,18	18,85	478,06
P02663	Alpha-S2-casein OS= <i>Bos taurus</i> GN=CSN1S2 PE=1 SV=2	61,26	23	670	23	26,0	8,43	1493,63	2005,32
P02662	Alpha-S1-casein OS= <i>Bos taurus</i> GN=CSN1S1 PE=1 SV=2	78,04	17	1111	16	24,5	5,02	4923,17	3803,18
Q95114	Lactadherin OS= <i>Bos taurus</i> GN=MFGE8 PE=1 SV=2	39,11	14	52	14	47,4	7,15	5,62	157,36
P02754	Beta-lactoglobulin OS= <i>Bos taurus</i> GN=LGB PE=1 SV=3	70,22	13	218	5	19,9	5,02	463,15	785,85
P00711	Alpha-lactalbumin OS= <i>Bos taurus</i> GN=LALBA PE=1 SV=2	45,77	12	64	12	16,2	5,14	681,69	198,57
P80025	Lactoperoxidase OS= <i>Bos taurus</i> GN=LPO PE=1 SV=1	18,82	11	20	11	80,6	8,54	1,11	52,27

Tableau 3 : Protéines identifiées dans l'échantillon Zurzach 55 après digestion avec la méthode eFASP et analyse par nanoLC Orbitrap Q-exactive plus.

De nombreuses protéines du lait ont été détectées dans l'échantillon Zurzach 55 provenant de la céramique de l'amphore. La méthode eFASP et l'utilisation de l'Orbitrap appliquées à un échantillon archéologique permettent une identification des protéines avec un haut pourcentage de couverture et un grand nombre de peptides (tableau 3). La caséine alpha S1 a été identifiée avec 78 % de couverture, 17 peptides et 16 peptides uniques identifiés sur la séquence. Le score emPAI qui permet d'avoir une information sur la quantité relative d'une protéine dans un échantillon [78] indique que l'alpha caséine S1 et S2 sont les protéines de lait les plus présentes. Dans la littérature, il est décrit que les caséines représentent 80% des protéines du lait [79], ce qui montre que du lait était présent dans l'amphore biberon Zurzach 55.

L'analyse avec le logiciel Blast P a permis de mettre en évidence des peptides spécifiques de la caséine alpha 1 de l'espèce *Bos taurus* (tableau 4). L'analyse Blast P consiste à rechercher une séquence d'acides aminés dans une base de données de protéines et de chercher les homologues. Le résultat est un score S correspondant à la somme des alignements des acides aminés de la séquence (exemple : L – L = +4, E · Q = +2) et une p-value qui est la probabilité d'obtenir par hasard un score x au moins égal au score S.

Cette analyse montre que le peptide DIGSESTEDQAMEDIK est spécifique de l'alpha-S1-caseine provenant de *Bos taurus* car ce peptide a été trouvé dans la base de données de Blast P (protéine non redondante de GenBank) avec un p-value de 1,00E-07 et un pourcentage d'identité de 100%. Le peptide correspondant à la position 58 à 73 sur la séquence de l'alpha-S1-caseine.

Score	p-value	Identité
54,5	1,00E-07	16/16 (100%)

	Début peptide	Séquence	Fin peptide
peptide identifié	1	DIGSESTEDQAMEDIK	16
	1	DIGSESTEDQAMEDIK	16
correspondance	58	DIGSESTEDQAMEDIK	73

Tableau 4 : Résultats de l'analyse Blast P sur le peptide DIGSESTEDQAMEDIK. L'analyse permet de montrer que le peptide est de l'espèce *Bos taurus* et provient de la protéine alpha caséine S1. Les colonnes correspondent au score, à la p-value, à l'identité (pourcentage de ressemblance).

L'analyse protéomique a permis de mettre en évidence la présence de lait de vache dans ces amphores biberon. Dans la littérature, des publications ont mis en évidence la présence de lait dans des échantillons archéologiques à l'aide de la protéomique [80-82]. La protéomique est une méthode très utilisée pour l'identification de protéines contenues dans des objets archéologiques [82-84].

A l'heure actuelle, aucun travail de recherche n'a permis de démontrer l'utilisation de l'analyse sans digestion ou top down sur un échantillon archéologique.

3. Analyse top down de l'échantillon Zurzach 55

a) Analyse top down de lait de vache

Avant de traiter et d'analyser l'échantillon archéologique Zurzach 55 avec la méthode de préparation des protéines mises au point au cours de ce travail de thèse, un échantillon de lait de vache a été analysé pour vérifier que les caséines peuvent être analysées et détectées par spectrométrie de masse.

Un volume de 100 μ L de lait de vache a été traité avec notre méthode de préparation des protéines puis ces dernières ont été séparés à l'aide de la chromatographie liquide C18 couplée à l'obitrap Q-exactive plus. La figure 10 montre l'analyse du lait de vache dilué cent fois sur une colonne C18 couplée avec un spectromètre de masse de type Orbitrap Q-Exactive plus. Des protéines ont été détectées comme le montre la figure 10b, dans lequel on observe un massif composé de 27 à 13 charges a été observé. La figure 10c montre le massif de la protéine avec 10 charges à m/z 1411,7771 correspondant à une masse de $M = 24493,1$ Da. La masse observée peut correspondre à l'alpha-S2-caséine bovine, avec le numéro d'accèsion P02663 et la séquence en acides aminés de 16 – 222. La masse moyenne théorique de cette protéine est de 24348,2 (formule brute : $C_{1083}H_{1707}N_{287}O_{338}S_6$). La différence de masse observée est due à la présence de modifications post-traductionnelles. Dix sites de phosphorylation de sérine ont été identifiés sur l'alpha-S2 caséine-bovine [85].

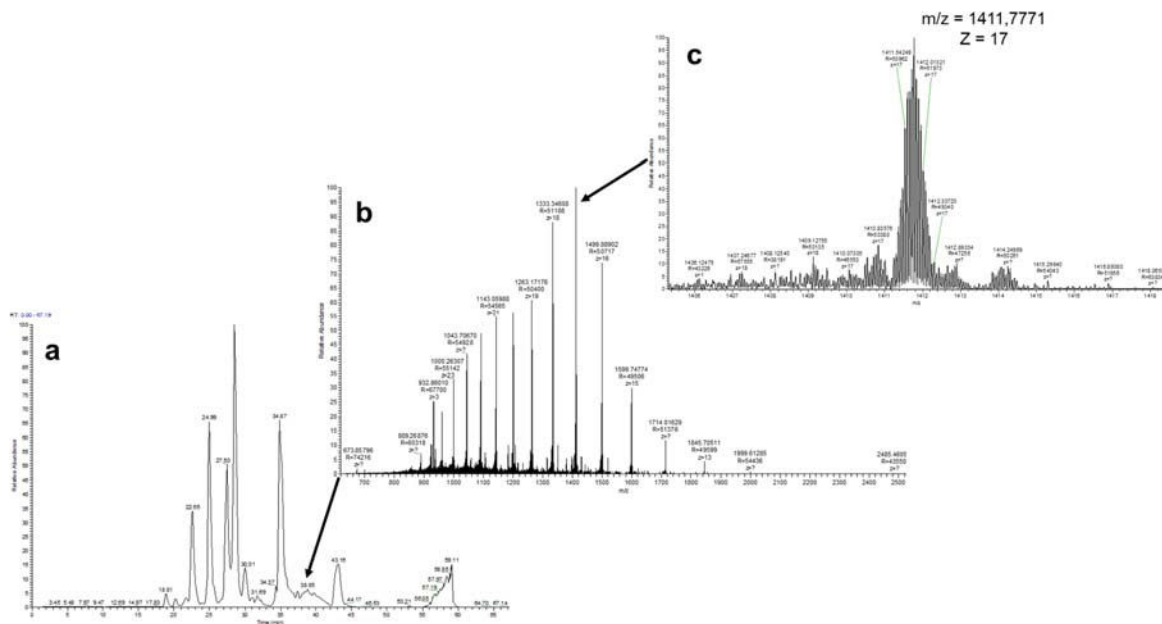


Figure 10 : Solution de lait diluée 100 fois et les protéines ont été séparées sur une colonne C18 couplée avec un spectromètre de masse de type Orbitrap Q-Exactive plus. Le courant total des ions est observé sur le chromatogramme (a), le spectre MS de la caséine alpha en b et en c le zoom sur le pic à m/z 1411,7771, $z=17$.

Cette analyse a permis de valider la méthode et l'adéquation de la méthodologie analytique pour l'analyse des protéines de lait. L'utilisation de notre méthode de préparation de protéines, la séparation sur une colonne C18 et la détection avec une Orbitrap Q-Exactive plus permettent de séparer les protéines du lait mais aussi de les détecter.

b) Analyse top down de l'échantillon archéologique Zurzach 55

L'analyse précédente a montré que les protéines de lait pouvaient être séparées en appliquant la méthodologie analytique que nous avons mise au point. Ainsi, l'échantillon archéologique a été préparé de la même façon. Une quantité plus importante d'échantillons a été utilisée pour l'analyse top down. En effet, 100 mg d'échantillon ont été nécessaires pour obtenir des spectres de bonne qualité. La figure 11 montre le chromatogramme du courant ionique total de l'analyse LC-MSMS de l'échantillon Zurzach 55. Le signal MS est présent sur tout le long du chromatogramme avec une intensité élevée ce qui va permettre d'obtenir des spectres MS et MS/MS de bonne qualité.

Analyse sans digestion des échantillons archéologiques et paléontologiques

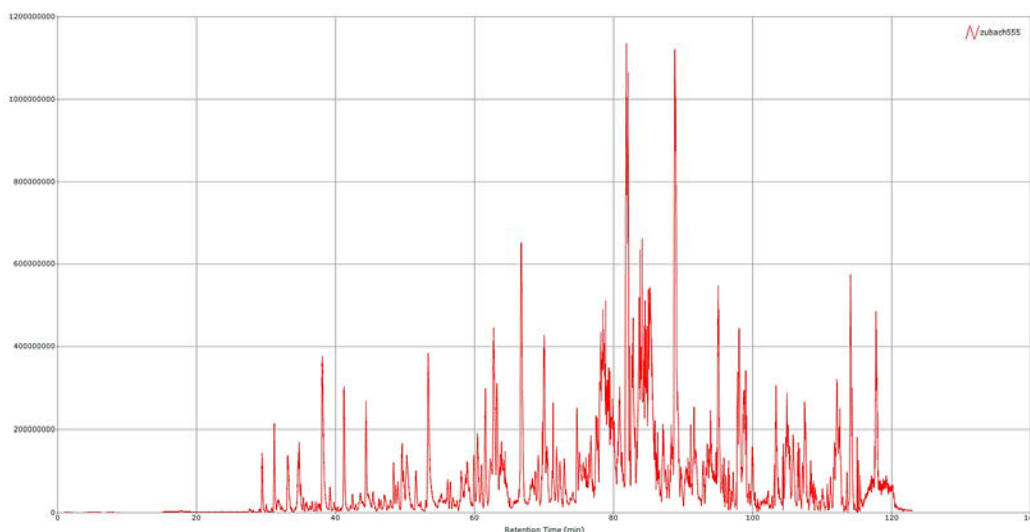


Figure 11 : Chromatogramme du courant des ions totaux (TIC) de l'analyse LC-MSMS top down de l'échantillon archéologique.

La figure 12 montre l'histogramme des charges des masses dans l'analyse LC-MS/MS. Les masses détectées dans l'analyse possèdent de 2 à 10 charges. Ceci permet de dire que les protéines de l'échantillon ne sont pas entières car elles ont subi une dégradation au cours du temps. De plus, la majorité des molécules détectées possèdent 2 ou 3 charges.

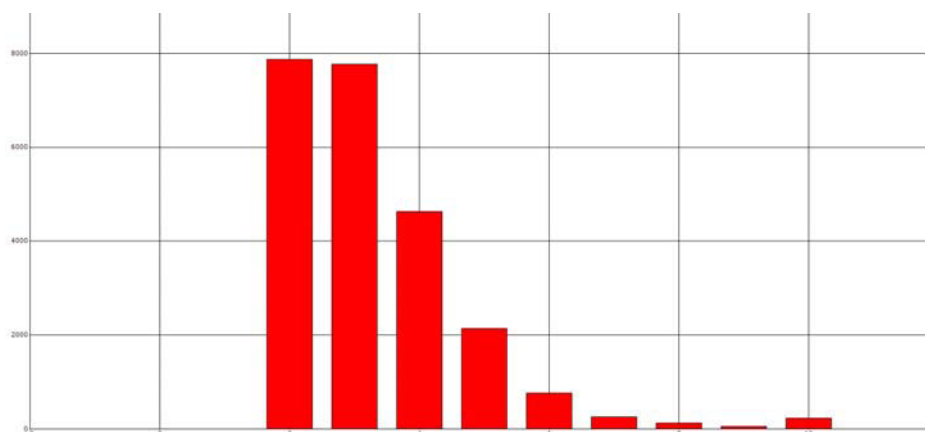


Figure 12 : Histogramme de la répartition des charges des masses détectées dans le chromatogramme.

4. Analyse bio-informatique

A partir de l'analyse bio-informatique réalisée sur le logiciel Byonic, les peptides identifiés possèdent des poids moléculaire de 800 Da à 10 000 Da. L'analyse permet de

montrer que les peptides identifiés possèdent une erreur de masse centrée autour de 0,8 +/- 3 ppm (figure 13). En comparant les résultats de l'analyse top-down et bottom-up, l'analyse top-down permet de détecter des peptides ou des petites protéines de plus grande taille.

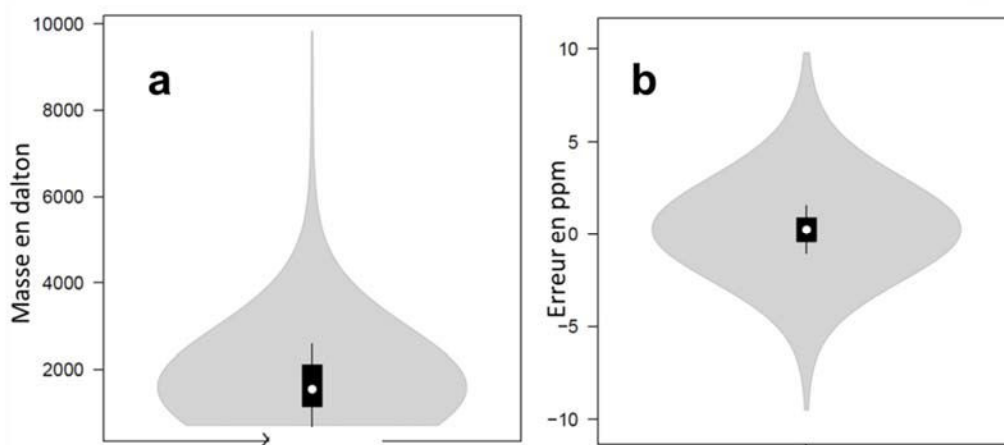


Figure 13 : Violin plot représentant la répartition des masses des peptides détectés dans l'échantillon après extraction et analyse top down (a) et un violin plot représentant l'erreur en ppm des peptides identifiés (b).

Le tableau 5 montre les protéines identifiées par analyse top-down dans l'échantillon en utilisant une interrogation sur la base de données de l'espèce *Bos taurus*. Les protéines identifiées correspondent à celle de l'analyse bottom-up. Le nombre de peptides identifiés et la couverture de séquence sont élevés. L'analyse TDP a fonctionné sur l'échantillon Zurzach 55 et elle permet d'identifier l'alpha caséine 1 avec 90 peptides uniques et 74% de couverture de séquence.

Numéro accésion	Protéines	Log P	Peptides uniques	Couverture %	Acides aminés
P02666	CASB_BOVIN Beta-casein	3401,88	56	73,3	224
P02662	CASA1_BOVIN Alpha-S1-casein	2096,84	90	74,9	214
P02663	CASA2_BOVIN Alpha-S2-casein	1804,2	76	71,9	222
P02668	CASK_BOVIN Kappa-casein	129,74	49	40	190
P31096	OSTP_BOVIN Osteopontin	94,54	35	37,1	278
P24627	TRFL_BOVIN Lactotransferrin	107,57	6	28	708

Tableau 5 : Protéines identifiées à partir de l'analyse TDP dans l'échantillon archéologique avec le logiciel Byonic en utilisant la banque de donnée Swissprot *Bos Taurus*. Les colonnes correspondent aux protéines, au Log P (Log base 10 de la p-value de la protéine), aux peptides uniques, au pourcentage de couverture et au nombre d'acides aminés.

Analyse sans digestion des échantillons archéologiques et paléontologiques

Des modifications ont été identifiées sur les séquences de protéines comme des oxydations des méthionines (M), proline (P), la conversion de Glutamate en Pyroglutamate en *N*-terminal (Gln->pyro-Glu),

Lors de l'approche bottom-up, un Amicon® avec un seuil de coupure de 10 kDa a été utilisé. Tandis que lors de l'analyse top-down un Amicon® avec un seuil de coupure de 3 kDa a été utilisé, ce qui permet d'expliquer l'identification d'un plus grand nombre de peptides à l'issue de l'analyse top-down. Lors de l'analyse bottom-up, une digestion à la trypsine est réalisée, ce qui diminue la taille des peptides. Certains peptides ne seront pas détectés lors de l'analyse par spectrométrie de masse car leur masse est trop faible.

En examinant la caséine beta, des peptides de masse de 800 à 9000 Da ont été identifiés avec des charges de $z = 2$ à $z = 7$. En analysant les spectres MS/MS, les peptides de grande taille (m/z 5000 à 9000) possèdent des ions de fragmentation de la série *b* et *y* (figure 14). Le peptide de masse $M = 9821,1484$ est identifié avec une série d'ion *b* et *y* sur les extrémités *N* et *C*-terminales et possède une oxydation sur la proline. Ceci peut être dû à l'utilisation de la fragmentation HCD qui n'est pas parfaitement adaptée aux peptides de grande masse. L'utilisation de fragmentation mixte comme EtHCD ou la fragmentation avec l'utilisation d'ultra-violet [41,46,86].

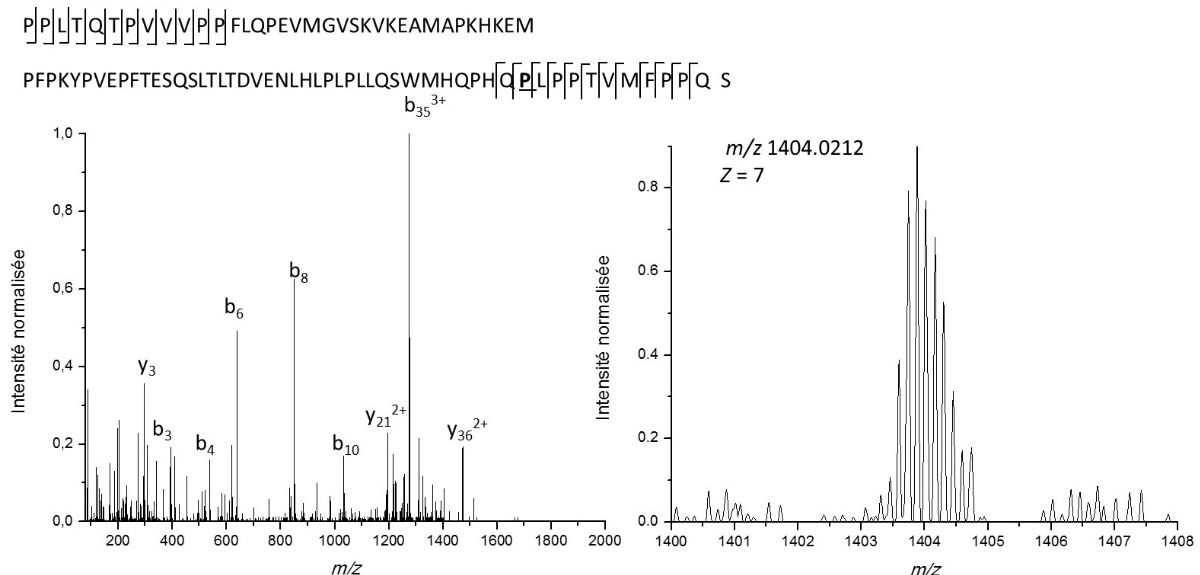


Figure 14 : Spectre de fragmentation HCD du peptide de la beta caséine à m/z 1404,1602 $z=7$ à gauche. A droite spectre MS du peptide multichargé détecté par l'Orbitrap.

Analyse sans digestion des échantillons archéologiques et paléontologiques

Un autre exemple de peptide de masse à m/z 1214,9034, $z = 4$ (peptide de masse $M= 4855,6136$) possèdent des spectres MS/MS informatif avec des séries d'ions fragments b et y qui couvrent les parties N et C -terminales de la séquence peptidique (figure 15).

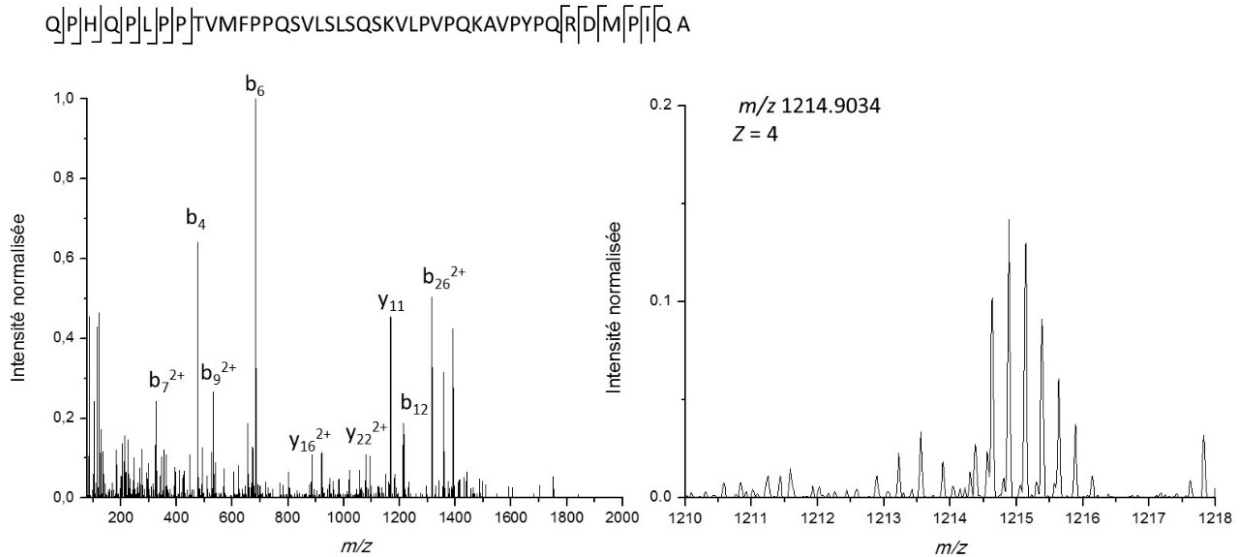


Figure 15 : Spectre de fragmentation HCD du peptide de la beta caséine à m/z 1214,4014 $z=4$ à gauche. A droite spectre MS du peptide multichargé détecté par l'Orbitrap

Des peptides de masse similaire ont été identifiés sur la protéine alpha caséine S1. La figure 16 montre le spectre de fragmentation d'un peptide possédant une masse de $M= 5485,9378$.

Analyse sans digestion des échantillons archéologiques et paléontologiques

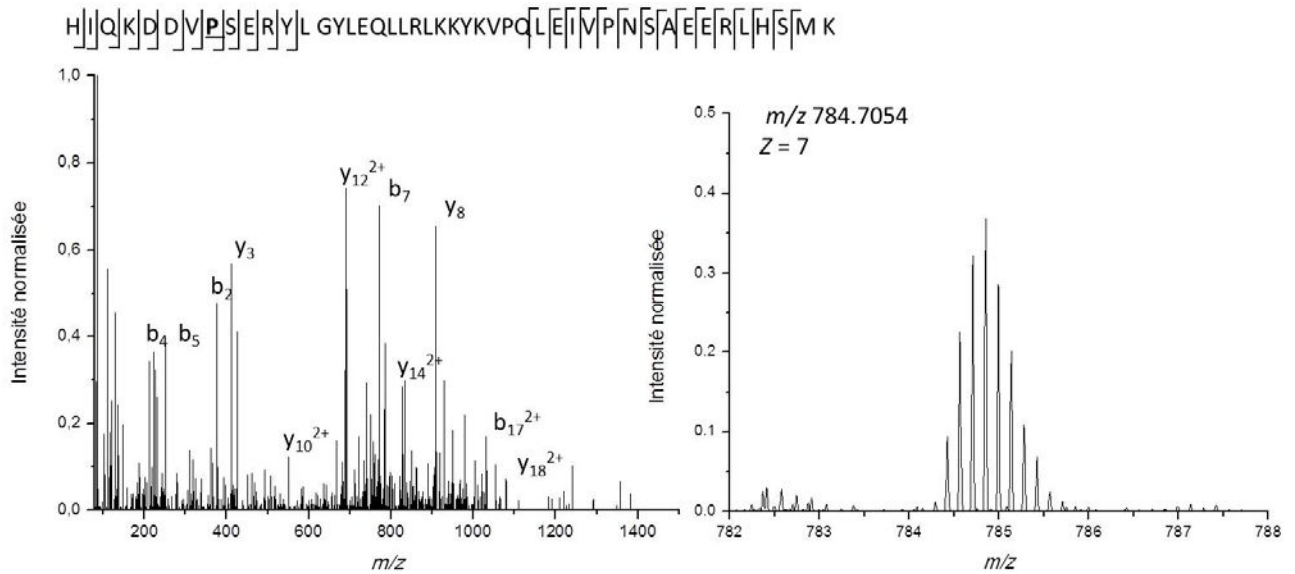


Figure 16 : Spectre de fragmentation HCD du peptide à m/z 1037,5099 $z=4$ de l'alpha caséine ainsi que le spectre MS du peptide.

L'analyse top down sur l'échantillon Zurzach 55 a permis d'identifier de la caséine sous la forme de fragments. Il s'agit de la première étude en top down sur un échantillon archéologique. Malheureusement, aucune protéine intacte n'a été détectée dans l'échantillon, ce qui montre que ces protéines ont été dégradées au cours du temps et par les conditions environnementales.

Cette première analyse top-down d'un échantillon archéologique ouvre les portes sur de nouvelles possibilités d'informations. Cette technique pourrait permettre d'étudier les protéines contenues dans des amphores et d'observer si ces protéines ont subi des modifications lors de leur préparation (chauffage, fermentation).

L'utilisation de la fragmentation UVPD ou EtHCD pourrait permettre une meilleure identification des protéines, des fragmentations des protéines et de couverture de séquence des peptides et des protéines. Ce type de fragmentation est implémenté sur les nouveaux appareils de type Orbitrap fusion de thermo Scientifique.

XI. Conclusion

Les analyses bottom-up et top-down ont permis de mettre en évidence la présence de protéines de lait dans l'échantillon Zurzarch 55. Les protéines identifiées, majoritairement de la caséine, l'ont été avec une couverture de séquence importante, ce qui prouve que, malgré les années et les conditions d'enfouissement, les protéines imprégnées dans les céramiques peuvent encore être identifiées avec succès.

Après l'interrogation des bases de données et l'élimination des protéines contaminantes, ces deux méthodes ont permis d'identifier les protéines comme étant des protéines de l'espèce *Bos taurus*.

L'analyse top down a permis d'identifier de nombreux peptides avec des masses élevées mais aucune protéine intacte n'a été détectée. L'utilisation de la spectrométrie de masse haute résolution a permis de montrer qu'il était possible de détecter des protéines sans effectuer de digestion au préalable pour observer la dégradation de celle-ci. Une analyse approfondie des modifications chimiques sur les peptides identifiés permettrait de connaître l'utilisation de ces protéines. Ce travail est actuellement en cours au laboratoire.

Il serait intéressant de déterminer si les phosphorylations des caséines sont encore présentes ou si des modifications types lactosylations peuvent être détectées.

L'étude d'échantillons archéologiques par spectrométrie de masse haute résolution sans digestion ouvre une nouvelle voie dans l'analyse. Les archéologues, à partir de l'objet et des données de spectrométrie de masse, pourraient obtenir de nouvelles informations sur l'utilisation des objets découverts sur les sites archéologiques, mais aussi sur les modes de vie sociétaux comme la préparation de l'alimentation à cette époque.

Bibliographie

- (1) Smith, L. M.; Kelleher, N. L.: Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186-187.
- (2) Chait, B. T.: Mass spectrometry: bottom-up or top-down? *Science* **2006**, *314*, 65-66.
- (3) Nesvizhskii, A. I.; Aebersold, R.: Interpretation of shotgun proteomic data the protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419-1440.
- (4) Fenn J, M. M., Meng C, Wong S, Whitehorse C: Electrospray ionization for mass-spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.
- (5) Henry, I.; Grandjouan, S.; Couillin, P.; Barichard, F.; Huerre-Jeanpierre, C.; Glaser, T.; Philip, T.; Lenoir, G.; Chaussain, J.; Junien, C.: Tumor-specific loss of 11p15. 5 alleles in del11p13 Wilms tumor and in familial adrenocortical carcinoma. *Proceedings of the National Academy of Sciences* **1989**, *86*, 3247-3251.
- (6) McLafferty, F. W.: High-resolution tandem FT mass spectrometry above 10 kDa. *Acc. Chem. Res.* **1994**, *27*, 379-386.
- (7) Kelleher, N. L.; Costello, C. A.; Begley, T. P.; McLafferty, F. W.: Thiaminase I (42 kDa) heterogeneity, sequence refinement, and active site location from high-resolution tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 981-984.
- (8) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W.: Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.
- (9) Siuti, N.; Kelleher, N. L.: Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4*, 817.
- (10) Tran, J. C.; Doucette, A. A.: Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. *J. Proteome Res.* **2008**, *7*, 1761-1766.
- (11) Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; Mclafferty, F. W.: Efficient sequence analysis of the six gene products (7-74 kDA) from the escherichia coli thiamin biosynthetic operon by tandem high-resolution mass spectrometry. *Protein Science* **1998**, *7*, 1796-1801.
- (12) Shen, Y.; Tolić, N.; Zhao, R.; Paša-Tolić, L.; Li, L.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D.: High-throughput proteomics using high-efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. *Anal. Chem.* **2001**, *73*, 3011-3021.
- (13) Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y.; Zhao, R.; Smith, R. D.: Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc Natl Acad Sci U S A* **2002**, *99*, 5942-7.

Analyse sans digestion des échantillons archéologiques et paléontologiques

- (14) Martinović, S.; Berger, S. J.; Paša-Tolić, L.; Smith, R. D.: Separation and detection of intact noncovalent protein complexes from mixtures by on-line capillary isoelectric focusing-mass spectrometry. *Anal. Chem.* **2000**, *72*, 5356-5360.
- (15) Ouvry-Patat, S. A.; Torres, M. P.; Quek, H. H.; Gelfand, C. A.; O'Mullan, P.; Nissum, M.; Schroeder, G. K.; Han, J.; Elliott, M.; Dryhurst, D.: Free-flow electrophoresis for top-down proteomics by Fourier transform ion cyclotron resonance mass spectrometry. *Proteomics* **2008**, *8*, 2798-2808.
- (16) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W.: Attomole protein characterization by capillary electrophoresis--mass spectrometry. *Science* **1996**, *273*, 1199.
- (17) Tran, J. C.; Doucette, A. A.: Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008**, *80*, 1568-1573.
- (18) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L.: Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480*, 254-8.
- (19) Ahlf, D. R.; Compton, P. D.; Tran, J. C.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.: Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J. Proteome Res.* **2012**, *11*, 4308-14.
- (20) Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L.: Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol. Cell. Proteomics* **2013**, *12*, 3465-3473.
- (21) Catherman, A. D.; Li, M.; Tran, J. C.; Durbin, K. R.; Compton, P. D.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.: Top down proteomics of human membrane proteins from enriched mitochondrial fractions. *Anal. Chem.* **2013**, *85*, 1880-8.
- (22) Li, Y.; Compton, P. D.; Tran, J. C.; Ntai, I.; Kelleher, N. L.: Optimizing capillary electrophoresis for top-down proteomics of 30-80 kDa proteins. *Proteomics* **2014**, *14*, 1158-64.
- (23) Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D. t.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L.: Applying label-free quantitation to top down proteomics. *Anal. Chem.* **2014**, *86*, 4961-8.
- (24) Kim, K. H.; Compton, P. D.; Tran, J. C.; Kelleher, N. L.: Online matrix removal platform for coupling gel-based separations to whole protein electrospray ionization mass spectrometry. *J. Proteome Res.* **2015**, *14*, 2199-206.
- (25) Takemori, N.; Takemori, A.; Wongkongkathep, P.; Nshanian, M.; Loo, R. R. O.; Lermyte, F.; Loo, J. A.: Top-down/bottom-up mass spectrometry workflow using dissolvable polyacrylamide gels. *Anal. Chem.* **2017**.

Analyse sans digestion des échantillons archéologiques et paléontologiques

- (26) Zinnel, N. F.; Pai, P. J.; Russell, D. H.: Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage. *Anal. Chem.* **2012**, *84*, 3390-7.
- (27) Xiu, L.; Valeja, S. G.; Alpert, A. J.; Jin, S.; Ge, Y.: Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal. Chem.* **2014**, *86*, 7899-906.
- (28) Toby, T. K.; Fornelli, L.; Kelleher, N. L.: Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual review of analytical chemistry (Palo Alto, Calif.)* **2016**, *9*, 499-519.
- (29) Katta, V.; Chow, D. T.; Rohde, M. F.: Applications of In-Source Fragmentation of Protein Ions for Direct Sequence Analysis by Delayed Extraction MALDI-TOF Mass Spectrometry. *Anal. Chem.* **1998**, *70*, 4410-4416.
- (30) Patrie, S. M.; Charlebois, J. P.; Whipple, D.; Kelleher, N. L.; Hendrickson, C. L.; Quinn, J. P.; Marshall, A. G.; Mukhopadhyay, B.: Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1099-108.
- (31) Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Pasa-Tolic, L.; Smith, R. D.: FT-ICR MS optimization for the analysis of intact proteins. *Int. J. Mass Spectrom.* **2009**, *281*, 32-38.
- (32) 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*, 430-43.
- (33) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S.: Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2113-2120.
- (34) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wiegand, 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. *Mol. Cell. Proteomics* **2011**, *10*, M111. 011015.
- (35) Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M.: Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol. Cell. Proteomics* **2006**, *5*, 949-958.
- (36) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O.: Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* **2013**, *405*, 8505-14.
- (37) Chamot-Rooke, J.; Mikaty, G.; Malosse, C.; Soyer, M.; Dumont, A.; Gault, J.; Imhaus, A.-F.; Martin, P.; Trellet, M.; Clary, G.: Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science* **2011**, *331*, 778-782.
- (38) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, U. A.; Tsybin, Y. O.: Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **2013**, *42*, 5014-30.

Analyse sans digestion des échantillons archéologiques et paléontologiques

- (39) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J.: Implementation of electron-transfer dissociation on a hybrid linear ion trap– orbitrap mass spectrometer. *Anal. Chem.* **2007**, *79*, 3525-3534.
- (40) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O.: Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **2012**, *11*, 1758-67.
- (41) Riley, N. M.; Westphall, M. S.; Coon, J. J.: Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87*, 7109-16.
- (42) Warnke, S.; von Helden, G.; Pagel, K.: Analyzing the higher order structure of proteins with conformer-selective ultraviolet photodissociation. *Proteomics* **2015**, *15*, 2804-12.
- (43) Dang, X.; Young, N. L.: Ultraviolet photodissociation enhances top-down mass spectrometry as demonstrated on green fluorescent protein variants. *Proteomics* **2014**, *14*, 1128-9.
- (44) Milosavljević, A.; Nicolas, C.; Gil, J.-F.; Canon, F.; Réfrégiers, M.; Nahon, L.; Giuliani, A.: VUV synchrotron radiation: a new activation technique for tandem mass spectrometry. *Journal of synchrotron radiation* **2012**, *19*, 174-178.
- (45) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S.: Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646-51.
- (46) Brunner, A. M.; Lossl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F.; Heck, A. J.: Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal. Chem.* **2015**, *87*, 4152-8.
- (47) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.: Unabridged analysis of human histone H3 by differential top-down mass spectrometry reveals hypermethylated proteoforms from MMSET/NSD2 overexpression. *Mol. Cell. Proteomics* **2016**, *15*, 776-790.
- (48) Levy, M. J.; Gucinski, A. C.; Boyne, M. T., 2nd: Primary Sequence Confirmation of a Protein Therapeutic Using Top Down MS/MS and MS(3). *Anal. Chem.* **2015**, *87*, 6995-9.
- (49) Bungler, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L.: Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal. Chem.* **2008**, *80*, 1459-1467.
- (50) Meng, F.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L.: Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal. Chem.* **2002**, *74*, 2923-2929.
- (51) Roth, M. J.; Parks, B. A.; Ferguson, J. T.; Boyne, M. T.; Kelleher, N. L.: "Proteotyping": population proteomics of human leukocytes using top down mass spectrometry. *Anal. Chem.* **2008**, *80*, 2857-2866.

Analyse sans digestion des échantillons archéologiques et paléontologiques

- (52) Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L.: Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal. Chem.* **2006**, *78*, 686-694.
- (53) Collier, T. S.; Hawkrige, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C.: Top-down identification and quantification of stable isotope labeled proteins from *Aspergillus flavus* using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal. Chem.* **2008**, *80*, 4994-5001.
- (54) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M.: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.
- (55) Kellie, J. F.; Higgs, R. E.; Ryder, J. W.; Major, A.; Beach, T. G.; Adler, C. H.; Merchant, K.; Knierman, M. D.: Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. *Sci Rep* **2014**, *4*, 5797.
- (56) Wiener, M. C.; Sachs, J. R.; Deyanova, E. G.; Yates, N. A.: Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.* **2004**, *76*, 6085-6096.
- (57) Mazur, M. T.; Cardasis, H. L.; Spellman, D. S.; Liaw, A.; Yates, N. A.; Hendrickson, R. C.: Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc Natl Acad Sci U S A* **2010**, *107*, 7728-33.
- (58) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P.: ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534-2536.
- (59) Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A.: Deconvolution and database search of complex tandem mass spectra of intact proteins a combinatorial approach. *Molecular & Cellular Proteomics* **2010**, *9*, 2772-2782.
- (60) Zamborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L.: ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007**, *35*, W701-6.
- (61) Durbin, K. R.; Fellers, R. T.; Ntai, I.; Kelleher, N. L.; Compton, P. D.: Autopilot: an online data acquisition control system for the enhanced high-throughput characterization of intact proteins. *Anal. Chem.* **2014**, *86*, 1485-92.
- (62) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W.: Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320-332.
- (63) Karabacak, N. M.; Li, L.; Tiwari, A.; Hayward, L. J.; Hong, P.; Easterling, M. L.; Agar, J. N.: Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 846-56.

Analyse sans digestion des échantillons archéologiques et paléontologiques

- (64) Tsai, Y. S.; Scherl, A.; Shaw, J. L.; MacKay, C. L.; Shaffer, S. A.; Langridge-Smith, P. R.; Goodlett, D. R.: Precursor ion independent algorithm for top-down shotgun proteomics. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2154-66.
- (65) Ayaz-Guner, S.; Zhang, J.; Li, L.; Walker, J. W.; Ge, Y.: In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry* **2009**, *48*, 8161-70.
- (66) Sun, R. X.; Luo, L.; Wu, L.; Wang, R. M.; Zeng, W. F.; Chi, H.; Liu, C.; He, S. M.: pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. *Anal. Chem.* **2016**, *88*, 3082-90.
- (67) Frank, A. M.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L.; Pevzner, P. A.: Interpreting top-down mass spectra using spectral alignment. *Anal. Chem.* **2008**, *80*, 2499-2505.
- (68) Kou, Q.; Xun, L.; Liu, X.: TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32*, 3495-3497.
- (69) Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y.: MASH Suite Pro: A comprehensive software tool for top-down proteomics. *Mol. Cell. Proteomics* **2016**, *15*, 703-714.
- (70) Tuross, N.: Albumin preservation in the Taima-taima mastodon skeleton. *Applied Geochemistry* **1989**, *4*, 255-259.
- (71) Grupe, G.: Preservation of collagen in bone from dry, sandy soil. *J. Archaeo.Sci.* **1995**, *22*, 193-199.
- (72) Ostrom, P. H.; Schall, M.; Gandhi, H.; Shen, T.-L.; Hauschka, P. V.; Strahler, J. R.; Gage, D. A.: New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochim. Cosmochim. Acta* **2000**, *64*, 1043-1050.
- (73) Nielsen-Marsh, C. M.; Ostrom, P. H.; Gandhi, H.; Shapiro, B.; Cooper, A.; Hauschka, P. V.; Collins, M. J.: Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. *Geology* **2002**, *30*, 1099-1102.
- (74) Erde, J.; Loo, R. R. O.; Loo, J. A.: Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **2014**, *13*, 1885-1895.
- (75) Feist, P.; Hummon, A. B.: Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015**, *16*, 3537-63.
- (76) Kollipara, L.; Zahedi, R. P.: Protein carbamylation: In vivo modification or in vitro artefact? *Proteomics* **2013**, *13*, 941-944.
- (77) Sun, S.; Zhou, J.-Y.; Yang, W.; Zhang, H.: Inhibition of Protein Carbamylation in Urea Solution Using Ammonium Containing Buffers. *Analytical biochemistry* **2014**, *446*, 76-81.
- (78) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M.: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in

Analyse sans digestion des échantillons archéologiques et paléontologiques

proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* **2005**, 4, 1265-72.

(79) Meurant, G.: *Handbook of milk composition*; Academic press, 1995.

(80) Hong, C.; Jiang, H.; Lü, E.; Wu, Y.; Guo, L.; Xie, Y.; Wang, C.; Yang, Y.: Identification of milk component in ancient food residue by proteomics. *PLoS one* **2012**, 7, e37053.

(81) Leger, T.; Garcia, C.; Videlier, M.; Camadro, J. M.: Label-Free Quantitative Proteomics in Yeast. *Methods Mol. Biol.* **2016**, 1361, 289-307.

(82) Léger, T.; Garcia, C.; Collomb, L.; Camadro, J.-M.: A Simple Light Isotope Metabolic labeling (SLIM-labeling) strategy: a powerful tool to address the dynamics of proteome variations in vivo. *Mol. Cell. Proteomics* **2017**, 16.

(83) Solazzo, C.; Fitzhugh, W. W.; Rolando, C.; Tokarski, C.: Identification of Protein Remains in Archaeological Potsherds by Proteomics. *Analytical chemistry* **2008**, 80, 4590-4597.

(84) Dallongeville, S.; Garnier, N.; Casasola, D. B.; Bonifay, M.; Rolando, C.; Tokarski, C.: Dealing with the identification of protein species in ancient amphorae. *Analytical and bioanalytical chemistry* **2011**, 399, 3053-3063.

(85) Imanishi, S. Y.; Kochin, V.; Ferraris, S. E.; de Thonel, A.; Pallari, H. M.; Corthals, G. L.; Eriksson, J. E.: Reference-facilitated phosphoproteomics: fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS. *Mol. Cell. Proteomics* **2007**, 6, 1380-91.

(86) Halim, M. A.; Girod, M.; MacAleese, L.; Lemoine, J.; Antoine, R.; Dugourd, P.: Combined Infrared Multiphoton Dissociation with Ultraviolet Photodissociation for Ubiquitin Characterization. *J. Am. Soc. Mass Spectrom.* **2016**, 27, 1435-42.

Annexes

I. Séquences des protéines utilisées pour l'analyse phylogénétique du collagène I alpha 1

>sp|P02452|CO1A1_HUMAN Collagen alpha-1(I) chain OS=Homo sapiens GN=COL1A1 PE=1 SV=5
MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGRLYHHRD VVKPEPCRI
CVC DNGKVLCD DVICDETKNCPGAEVPEGECCPVC PDGSE SPTDQETT GVEGPKGDTGPR
GPRGPAGPPGRDGI PGQPGLPGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGP
MGPSGPRGLPGPPGAPG PQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPRG
PGERGPPGPQGARGLP GTAGLPGMKGHRGFSGLDGA KG DAGPAGPKGEPGSPGENGAPGQ
MGPRGLPGERGRPGAPG PAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGP
RGSEG PQGVRGEPGPPG PAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGP
QGPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPTGL
PGPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGL
TGSPGSPGPDGKTGPPG PAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
PGPPGAVGPAGKDG EAGAQQPPG PAGPAGERGEQGPAGSPGFQGLPGAGPPGEAGKPGE
QGVPDGLGAPGSPGARGERGFPGERGVQGPPG PAGPRGANGAPGNDGAKGDAGAPGAPGS
QGAPGLQGM PGERGAAGLP GPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGAPAGPD
KGESGSPGAPGTGARGAPDRGEPGPPG PAGFAGPPGADGQPGAKGEPGDAGAKGDAGP
PGPAGPAGPPGPIGNV GAPGAKGARGSA GPPGATGFPGAAGR VGP GPPGSGNAGPPGPPGP
AGKEGGKGRGETG PAGRPEVGP PPGPPG PAGEKGS PGADGPAGAPGTPGQGIAGQRGV
VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
EGSPGRD GSPGAKGDRGETGPAGPPGAPGAPGAPGVGPAGKSGDRGETGPAGPTGPVGP
VGARGPAGPQGP RGDKGETGEQDGRG IKGHRGFSGLQGPPGPPGSPGEEQGPSGASGAPG
RGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSAGFDFSF
LPQPPQEKAHDGG RYYRADDANVVRDRDLEVDTTLKSLSQQIENIRSPEGSRKNPARTCR
DLKMCHSDWKS GEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNP KD
KRHVWFGESMTDGFQFEYGGQGS DPA DVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ
TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKT VIEYKTTKTSRLPII
DVAPLDVGAPDQEF GFDVGPVCF L

>sp|P02454|CO1A1_RAT Collagen alpha-1(I) chain OS=Rattus norvegicus GN=Col1a1 PE=1 SV=5
MFSFVDLRLLLLLL GATALLTHGQEDIPEVSCIHNGLRVPNGETWKP DVCLICICHNGTAV
CDGVLCKEDLDCPNPQKREG ECCPFCE EYVSPDAE VIGVEGPKGDPGPQGP RGPVGP
QDGI PGQPGLPGPPGPPGPPGPPGPPGLGGNFASQMSYGYDEKSAGVSVPGPMGPSGPRGLPG
PPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPRGPRGERGPPGPQG
ARGLPGTAGLPGMKGHRGFSGLDGA KGDTGPAGPKGEPGSPGENGAPGQMGP RGLPGERG
RPGPPGSAGARGNDGAVGAAGPPGPTGPTGPPGFPGAAGAKGEAGPQGARGSEG PQGVRG
EPGPPG PAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPSGAPGPKG
NSGEPGAPGNK GDTGAKGEPGAGVQGPPGPAGEEGKRGARGEPPSGLPGPPGERGGPG
SRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDG
KTGPPG PAGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGEPGKAGERGVPGPPGAVGPAG
KDGEAGAQQAGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPG
PSGARGERGFPGERGVQGPPG PAGPRGNNGAPGNDGAKGDTGAPGAPGSQQGAPGLQGM PG
ERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPG PAGAPGDKGETGPSGPAG
PTGARGAPGDRGEPGPPG PAGFAGPPGADGQPGAKGEPGDTGVKGDAGPPG PAGPAGPPG
PIGNV GAPGPKGSRGAAGPPGATGFPGAAGR VGP GPPGSGNAGPPGPPGPPV GKEGKGRG
ETGPAGRPEVGP PPGPPG PAGEKGS PGADGPAGSPGTPGPGQGIAGQRGVVGLPGQRGERG
FPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGSPGAEGSPGRD GAPG
AKGDRGETGPAGPPGAPGAPGAPGVGPAGKNGDRGETGPAGPAGPIGPARGPAGPQG
PRGDKGETGEQDGRG IKGHRGFSGLQGPPGSPGSPGEEQGPSGASGAPRGP PPSAGSPG
KDGLNGLPGPIGPPGPRGRTGDSG PAGPPGPPGPPGPPGPPSGGYDFSFLPQPPQEKSQD
GGRYRADDANVVRDRDLEVDTTLKSLSQQIENIRSPEGSRKNPARTCRDLKMCHSDWKS
GEYWIDPNQGCNLDAIKVYCNMETGQTCVFPTQPSVPQKNWYISP NPKKHHVWFGESMT
DGFQFEYGSEGS DPA DVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ TGNLKKSLLLQ
GSNEIELRGEGNSRFTYSTLV DGCTSHTGTWGKT VIEYKTTKTSRLPIIDVAPLDIGAPD
QEFGM DIGPACFV

>sp|P11087|CO1A1_MOUSE Collagen alpha-1(I) chain OS=Mus musculus GN=Col1a1 PE=1 SV=4
MFSFVDLRLLLLLGATALLTHGQEDIPEVSCIHNGLRVPNGETWKPEVCLICIHNGTAV
CDDVQCNEELDCPNPQRREGECCAFCEPEEYVSPNSEDVGVVEGPKGDPGPGQGRGPPVGGPPG
RDGIPGQPLGPPGPPGPPGPPGLGGNFASQMSYGYDEKSAGVSVPGMGPSPGRGLPG
PPGAPGPGQFQGGPPGEPGEGGSGPMGPRGPPGPPGKNGDDGEAGKPRPGERGPPGPPGQ
ARGLPGTAGLPGMKGHRGFSLDGAKGADGAPGPKGEPGSPGENGAPGQMGRPLPGERG
RPGPPGTAGARGNDGAVGAAGPPGPTGPTGPPGFPGAVGAKGEAGPQGARGSEGPQGVGRG
EPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPSGPPGPKG
NSGEPGAPGNKGDGTAKGEPGATGVQQPPGPAGEEGKRGARGEPGPSGLPGPPGERGGPG
SRGFFGADGVAGPKGSPGERGAPGAPGKSPGEGRPEAGLPGAKGLTGSPPGSPGPDG
KTGPPGAGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGEPGKAGERGLPGPPGAVGPAG
KDGEAGAQQGAPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEGVPGDLGAPG
PSGARGERGFPGERGVQGGPPGAPGRGNGAPGNDGAKGDTGAPGAPGSQQGAPGLQGMPPG
ERGAAGLPGPKGDRGDAGPKGADGSPGKDGARGLTGPIGPPGPAGAPGDKGEAGPSGPPG
PTGARGAPGDRGEAGPPGPAGFAGPPGADGQPGAKGEPGDTGVKGDAGPPGPAGPAGPPG
PIGNVAGPAGPKGPRGAAGPPGATGFPGAAGRVPVGGPSGNAGPPGPPGPPVKGEGGKGRG
ETGPAGRPGEVGGPPGPPGAGEKGSPPGADGAPGSPGTPGPPQGIAGQRGVGLPGQRGERG
FPGLPGPSGEPGKQGPSGSSGERGPPGPMGPPGLAGPPGESGREGSPGAEGSPGRDGAPG
AKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGPAGPAGPIGPAGARGPAGPQG
PRGDKGETGEQDGRGKGRGFSLQGGPPGSPGSPGEGQGPSGASGPAGPRGPPGSAGSPG
KDGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPPSSGGYDFSFLPQPPQEKSD
GGRYRADDANVVRDRDLEVDTTLSLSQIENIRSPEGSRKNPARTCRDLKMCHSDWKS
GEYWIDPNQGCNLDAIKVYCNMETGQTCVFPTQPSVPQKNWYISPNPKEKKHVWFGESMT
DGFPFEYGSEGSDDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQTGNLKKALLLQ
GSNEIELRGEGNSRFTYSTLVDGCTSHTGTWGKTVIEYKTTKTSRLLPIIDVAPLDIGAPD
QEFGLDIGPACFV

>sp|P02453|CO1A1_BOVIN Collagen alpha-1(I) chain OS=Bos taurus GN=COL1A1 PE=1 SV=3
MFSFVDLRLLLLLAATALLTHGQEEGQEEGQEEEDIPPVTCVQNGRLRYHHRDVKVPVPCQI
CVCNDGNVLCDDVICDELKDCPNAKVPTDECCPVCPEGQESPTDQETTGVVEGPKGDTGPR
GPRGPAGPPGRDGIPIGQPLGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGISVPGPM
GPSGPRGLPGPPGAPGPGQFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPRP
GERGPPGPGQARGLPGTAGLPGMKGHRGFSLDGAKGADGAPGPKGEPGSPGENGAPGQM
GPRGLPGERGRPGAPGARGNDGATGAAGPPGPTGAPPPGFPGAVGAKGEGGPPQGRG
GSEGPQGVRRGEPGPPGAPGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ
GPSGPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGIQGGPPGPAGEEGKRGARGEPGAPGLP
GPPGERGGPSRGFFGADGVAGPKGPAGERGAPGAPGPKGSPGEGRPEAGLPGAKGLT
GSPGSPGPDGKTGPPGPAGQDGRGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
GPPGAVGPAGKDGEAGAQQGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ
GVPGDLGAPGPSGARGERGFPGERGVQGGPPGAPGRGANGAPGNDGAKGDAGAPGAPGSQ
GAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK
GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP
GPAGPAGPPGPIGNVAGPAGKARGSPGATGFPGAAGRVPVGGPSGNAGPPGPPGPA
GKEGSKGPRGETGPAGRPGEVGGPPGPPGAGEKGPAGADGAPGAPGTPGPPQGIAGQRGV
GLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAE
GSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGPAGPAGPIGPV
GARGPAGPQGRGDKGETGEQDGRGKGRGFSLQGGPPGPPGSPGEGQGPSGASGPAGPR
GPPGSAGSPGKDGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPGPPSSGGYDLSFL
PQPPQEKADGGRYRADDANVVRDRDLEVDTTLSLSQIENIRSPEGSRKNPARTCRD
LKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPKEK
RHVWYGESMTGGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQT
GNLKKALLLQGSNEIEIRAEGNSRFTYSVTYDGTGCTSHTGAWGKTVIEYKTTKTSRLLPIID
VAPLDVGPADQEFDFVGPACFL

>sp|P02457|CO1A1_CHICK Collagen alpha-1(I) chain OS=Gallus gallus GN=COL1A1 PE=1 SV=3
MFSFVDSRLLLLIAATVLLTRGEEEDIQTGSCVQDGLTYNDKDVWKPEPCQICVCDGSGN
ILCDEVICEDTSDCPNAEIPFGECCPICPDVADSPVYVESAGVEGPKGDTGPRGDRGLPG
PPGRDGIPIGQPLGPPGPPGPPGPPGLGGNFAPQMSYGYDEKSAGVAVPGMGPAGPRGLPG
PPGAPGPGQFQGGPPGEPGEPGASGPMGPRGAPPPGKNGDDGEAGKPRPGRGPGRGPPGQ
ARGLPGTAGLPGMKGHRGFSLDGAKGQPGAPGPKGEPGSPGENGAPGQMGRPLPGERG

RPGPSGPAGARGNDGAPGAAGPPGPTGPAGPPGFPGAAGAKGETGPQGARGSEGPOGSRG
 EPGPPGAGAAGPAGNPGADGQPGAKGATGAPGIAGAPGFPGARGPSGPQGPSGAPGPKG
 NSGEPGAPGNKGDGTGAKGEPGPAGVQGPFPAGEEGKRGARGEPGAPGLPGAGERGAPG
 SRGFPGADGIAGPKGPPGERGSPGAVGPKGSPGEAGRPGEAGLPAGKGLTGSPGSPGPDG
 KTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGAAGEPGKPGERGAPGPPGAVGAAG
 KDGEAGAQQGPPGPTGPAGERGEQGPAGAPGFQGLPGPAGPPGEAGKPGEQGVPGNAGAPG
 PAGARGERGFPGERGVQGPFGPPGQGPARGANGAPGNDGAKGDAGAPGAPGNEGPPGLEGMPG
 ERGAAGLPAGAKGDRGDPGPKGADGAPGKDGLRGLTGPIGPPGPAGAPGDKGEAGPPGPAG
 PTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGETGDAGAKGDAGPPGPAGPTGAPG
 PAGZVGAPGPKGARGGAGPPGATGFPGAAGRVPFPGPSNIGLPGPPGPAGKZGSKGPRG
 ETGPAGRPGEPGAPGPPGPPGEGKSGADGPIGAGTGPQGIAGQRGVVGLPGQRGERG
 FPGLPGSPGEPGKQGPSASGERGPPGPMGPPGLAGPPGEAGREGAPGAEGAPGRDGAAG
 PKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGPAGPAGPPGPAGARGPAGPQG
 PRGDKGETGEQDGRGMKGHRGFSGLQGPFPAGGEPGQGPSGASGPAGPRGPPGSAGAAG
 KDGLNGLPGPIGPPGPRGRTGEVGPVGPFPGPPGPPGPPGPPSGGFDLSFLPQPPQEKHAH
 GGYYRADDANVMRDRDLEVDTTLSLSQIENIRSPEGTRKNPARTCRDLKMCHGDWKS
 GEYWIDPNQGCNLDIAKVYCNMETGETCVYPTQATIAQKNWYLSKNPKEKHHVWFGETMS
 DGFQFEYGGEGSNPADVAIQLTFLRLMSTEATQNVTYHCKNSVAYMDHDTGNLKKALLLQ
 GANEIEIRAEGNSRFTYGVTEGCTSHTGAWGKTVIEYKTTKTSRPLIIDLAPMDVGAPD
 QEFGIDIGPVCFL

>sp|Q9XSJ7|CO1A1_CANLF Collagen alpha-1(I) chain OS=Canis lupus familiaris GN=COL1A1 PE=1 SV=1

MFSFVDLRLLLLLAATALLTHGQEEGQEEDIPPVTCVQNGLRYYDRDVKPEACRICVCD
 NGNVLCDDVICDETKNCPGAQVPPGECPCVCPDGEASPTDQETTGVGPKGDTGPRGPRG
 PAGPPGRDGIPIGQPLPGPPGPPGPPGPPGLGGNFAPQMSYGYDEKSTGGISVPGMGPS
 GPRGLPGPPGAPGPPGQGFQGPPEGEPGASGPMGPRGPPGPPGKNGDDGEAGKPRGPRG
 GPPGPPGQARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGR
 GLPGERGRPGAPGARGNDGATGAAGPPGPTGPAGPPGFPGAAGVAKGEAGPQGARGSE
 GPQGVREPGPPGAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPS
 GPPGPKGNSGEPGAPGNKGDGTGAKGEPGPTGIQGPFPAGEEGKRGARGEPGPTGLPGPP
 GERGGPGRGFPADGVAGPKGPAGERGSPGAPGPKGSPGEAGRPGEAGLPAGKGLTGSP
 GSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVGPP
 GAVGAPAGKGEAGAQQGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGV
 GDLGAPGSPGARGERGFPGERGVQGPFPAGPRGANGAPGNDGAKGDAGAPGAPGSPQGP
 GLQGMPPERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDKGEA
 GPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPA
 GPTGPPGPIGNVAPGPKGARGGAGPPGATGFPGAAGRVPFPGPSNAGPPGPPGAPGKE
 GGKGARGETGPAGRPGEVGPFPGPPGAGEKSPGADGPAGAPGTPGPQGIAGQRGVVGLP
 GQRGERGFPLPGSPGEPGKQGPSGTSGERGPPGPMGPPGLAGPPGESGREGSPGAEGSP
 GRDGSPGPKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKNGDRGETGPAGPAGPIGPVGAR
 GPAGPQGPGRGDKGETGEQDGRGIKHRGFSGLQGPFPSPGEPGQGPSGASGPAGPRGPP
 GSAGSPGKDGLNGLPGPIGPPGPRGRTGDAGVGPFPGPPGPPGPPGPPSGGFDLSFLPQP
 PQEKADGGYYRADDANVVRDRDLEVDTTLSLSQIENIRSPEGSRKNPARTCRDLKM
 CHSDWKSGEYWIDPNQGCNLDIAKVFENMETGETCVYPTQPQVAQKNWYISKNPKEKRHV
 WYGESMTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQQTGNL
 KKALLLQGSNEIEIRAEGNSRFTYSVTYDGCTSHTGAWGKTVIEYKTTKTSRPLIIDVAP
 LDVGAPDQEFGMIDIGPVCFL

>tr|H2QDE6|H2QDE6_PANTR Collagen type I alpha 1 OS=Pan troglodytes GN=COL1A1 PE=2 SV=1

MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEEDIPPITCVQNGLRHYHRRDVKPEPCRI
 CVCDNGKVLCDVICDETKNCPGAEVPEGECCPCVCPDGESESPTDQETTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLPGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGP
 MGPSGPRGLPGPPGAPGPPGQGFQGPPEGEPGASGPMGPRGPPGPPGKNGDDGEAGKPR
 PGERGPPGPPGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQ
 MGPRGLPGERGRPGAPGARGNDGATGAAGPPGPTGPAGPPGFPGAAGVAKGEAGPQGP
 RGSEGPQGVREPGPPGAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGP
 QGPPGPPGPKGNSGEPGAPGSKGDTGAKGEPGVPVQGPFPAGEEGKRGARGEPGPTGL
 PGPGERGGPGRGFPADGVAGPKGPAGERGSPGAPGPKGSPGEAGRPGEAGLPAGKGL
 TGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV

PGPPGAVGPAGKDGEGAQGGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGE
 QGVPDGLGAPGSPGARGERGFPPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGS
 QGAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPAGPD
 KGESGSPGAPGTGARGAPDRGEPGPPGAPGAFAGPPGADGQPGAKGEPGDAGAKGDAGP
 PGPAGPAGPPGPIGNVGAAPGAKGARGASAGPPGATGFPGAAGRVPVPPGSPGNAGPPGPPGP
 AGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGPAGAPGTPGPQGIAGQRGV
 VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
 EGSPGRDGSAGKGDGRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGPAGPAGVGP
 VGARGPAGPQGPGRDKGETGEQDGRGIKGRHGFSGLQGGPPGPPGSPGEEQGPSGASGAPG
 RGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGVPVPPGPPGPPGPPGPPSAGFDFS
 LPQPPQEKAHDGGRYRADDANVVRDRDLEVDTTLSLQQIENIRSEGSRKNPARTCR
 LKMHSDWKSGEYWIDPNQGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK
 KRHVWFGESMTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ
 TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTTKTSRLPII
 DVAPLDVGAPDQEFDFVGPVCF

>tr|H2NVM9|H2NVM9_PONAB Collagen type I alpha 1 chain OS=Pongo abelii GN=COL1A1 PE=4
 SV=1

MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEIDIPPITCVQNGRLRYHDRDVKPEPCRI
 CVCDNGKVLCDVICDETKNCPGAEVPEGECCPVCDDGSESPDQETTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGP
 MGPSGPRGLPGPPGAPGPPQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGR
 PGERGPPGPPGARGPLGTAGLPGMKGHRGFSGLDGAAGKGDAGPAGPKGEPGSPGENGAPGQ
 MGPRGLPGERGRPGAPGAPGARGNDGATGAAGPPGPTGPAGPPGPPGAVGAKGEAGPQGP
 RGSEGPQGVRRGEPGPPGAPGAGPAGNPAGDQPGAKGANGAPGIAGAPGFPARGPSGP
 QGPPGPPGPKGNSGEPGAPGSKGDTGAKGEPGPPVGVQGGPPGAGEEGKRGARGEPPGPTGL
 PGPPGERGGPGRGFPGADGVAGPKGPAGERGSPGAPGKGSPPGEAGRPGEAGLPAGKGL
 TGSPGSPGPDGKTGPPGAPGQDGRGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
 PGPPGAVGPAGKDGEGAQGGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGE
 QGVPDGLGAPGSPGARGERGFPPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGS
 QGAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPAGPD
 KGETGPSGPAGPTGARGAPDRGEPGPPGAPGAFAGPPGADGQPGAKGEPGDAGAKGDAGP
 PGPAGPAGPPGPIGNVGAAPGAKGARGASAGPPGATGFPGAAGRVPVPPGSPGNAGPPGPPGP
 AGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGPAGAPGTPGPQGIAGQRGV
 VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
 EGSPGRDGSAGKGDGRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGPAGPAGVGP
 VGARGPAGPQGPGRDKGETGEQDGRGIKGRHGFSGLQGGPPGPPVSMSPSPVMPFPALQ
 GPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGVPVPPGPPGPPGPPGPPSGGDFDFSFL
 PQQPQEKAHDGGRYRADDANVVRDRDLEVDTTLSLQQIENIRSEGSRKNPARTCRD
 LKMHSDWKSGEYWIDPNQGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK
 RHVWFGESMTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQT
 GNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTTKTSRLPII
 VAPLDVGAPDQEFDFVGPVCF

>tr|G1M024|G1M024_AILME Collagen type I alpha 1 chain OS=Ailuropoda melanoleuca
 GN=COL1A1 PE=4 SV=1

MFSFVDLRLLLLLVAATALLTHGQEEGQEEEDSKSQIPPVTCVQNGRLRYDRDVKPEACRI
 CVCDNGNVLCDVICDETKNCPGAQVPPGECCPVCDDGEGTALGAWGWGGAGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLGPPGPPGPPGPPGPPGLGGNFAPQMSYGYDEKSTGGISVPGP
 MGPSGPRGLPGPPGAPGPPQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGR
 PGERGPPGPPGARGPLGTAGLPGMKGHRGFSGLDGAAGKGDAGPAGPKGEPGSPGENGAPGQ
 MGPRGLPGERGRPGAPGAVSTPGPLGDTFFHFSQGPAGPPGPPGAVGAKGEAGPQG
 ARGSEGPQGVRRGEPGPPGAPGAGPAGNPAGDQPGAKGANGAPGIAGAPGFPARGPSG
 PQGPSGPPGPKGNSGDGSPHCRPSPGLCLQGGPTGIQGGPPGAGEEGKRGARGEPPGPTGL
 PGPPGERGGPGRGFPGADGVAGPKGPAGERGSPGAPGKGSPPGEAGRPGEAGLPAGKGL
 TGSPGSPGPDGKTGPPGAPGQDGRGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
 PGPPGAVGPAGKDGEGAQGGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGE
 QGVPDGLGAPGSPGARGERGFPPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGS
 QGAPGLQGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPAGPD
 KGEAGPSGPAGPTGARGAPDRGEPGPPGAPGAFAGPPGADGQPGAKGEPGDAGAKGDAGP

PGPAGPTGPPGPIGNVAGPGBKARGSAGPPGATGFPGAAGRVPVPPGPPSGNAGPPGPPGP
 AGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGPAGAPGTPGPPQGIAGQRGV
 VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGSPGA
 EGSPGRDGSPPGPKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGPAGPAGPIGP
 VGARGPAGPQGRGDKGETGEQGDRIKGRHGFSGLQGGPPGPPGSPGEEQGPSGASGPAGP
 RGPPGSAGSPGKDLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSGGDFDFSF
 LPQPPQEKAHDGGRYRADDANVVRDRDLEVDTTLSLSQQIENIRSPEGSRKNPARTCR
 DLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPHVAQKNWYISKNPKE
 KRHVWYGESMTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ
 TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTYDGTSTHTGAWGKTVIEYKTTKTSRLPII
 DVAPLDVGAPDQEFGMDVGPVCF

>tr|F7BJV7|F7BJV7_MACMU Collagen type I alpha 1 chain OS=Macaca mulatta GN=COL1A1 PE=4
 SV=2

MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGRLRYHDRDVWKPEPCRI
 CVCDNGKVLCDDEVICDETKNCPGAEVPEGECCVCPDGSAGALGAWGLGPGGWGWGWGL
 GSGISKPCVALSGLWGAGPMGPRGPPGPPGKNGDDVSIPSKKIPSDPMGARGLPGTAGLP
 GMKGHRGFSGLDGAKGDAGPAGARWRPGRGTRLSPIFDLWFGCQGRGLPGERGRPGAP
 GPAVSTPSPLGEPWVGFPRVASDMSFLLIQGEAGPQGRGSEGPQGVREGEPPGPGAGAA
 GPACLAGAHVLSNLRLPGGCPCCSLGIVFLFWPMAKSHKLETLTGPAGERGSPGPAGP
 KGSPGEAGRPEAGLPKAKVRPQSFLAWPGPDHTGLTGSPPGSPGPDGKTGPPGPAGQDG
 RPPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVPPGAVGPAGKDGEGAGAQGPPG
 PAVMAPGGFCSLSVPPSLPPQGVPGDLGAPGSPGARVSRPLSLHPSRCLVRGPVSRTGE
 RGFPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGSQAGPLQGMPPGERGAAGL
 PGPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPGAPGDKGETGPPGAPGPTGARGA
 PGDRGEPGPPGAPFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGAPPPGPIGNAGP
 PGPPGAPGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGPAVSRAGEGSLCC
 CPGAPGTPGPPQGIAGQRGVVGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMG
 PPGLAGPPGESGREGAPGAEGSPGRDGSPPGPKGDRGETGPAGPPGAPGAPGAPGVPVGPAG
 KSGDRGETGPAGPAGPVGVPVARGPAVNADSAPLDGPQGRGDKGETGEQGDRIKGRHGF
 FSGLQGGPPGPPGSPGEEQGPSGASGPAGPRGPPGSAGTPGKDLNGLPGPIGPPGPRGRTG
 DAGPVGPPGPPGPPGPPGPPGPPSGGDFDFSLPQPPQEKAHDGGRYRADDANVVRDRDLEVD
 TTLKLSQQIENIRSPEGSRKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCN
 METGETCVYPTQPSVAQKNWYISKNPDKRHWVWFGESMTDGFQFEYGGEGSDPADVAIQL
 TFLRLMSTEASQNITYHCKNSVAYMDQGTGNLKKALLLQGSNEIEIRAEGNSRFTYSVTV
 DGCTSTHTGAWGKTVIEYKTTKTSRLPIIDVAPLDVGAPDQEFGFDVGPVCF

>tr|G3RBN8|G3RBN8_GORGO Collagen type I alpha 1 chain OS=Gorilla gorilla GN=COL1A1
 PE=4 SV=1

MFSFVDLRLLLLLLAATALLTHGQEDGQVEGQDEDIPPITCVQNGRLRYHDRDVWKPEPCRI
 CVCDNGKVLCDDEVICDETKNCPGAEVPEGECCVCPDGSSEPTDQETTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSAGGISVPGP
 MGPSGPRGLPGPPGAPGPPGQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPR
 PGERGPPGPPGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQ
 MGPRGLPGERGRPGAPGAPARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGP
 RGSEGPQGVREGEPPGPPGAGAAGPAGNPAGDQPGAKGANGAPGIAGAPGFPGARGPSGP
 QGPPGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGGPPGAGEEGKRGARGEPGPTGL
 PGPPGERGGPPSRGFPAGDGVAGPKGPAGERGSPGAPGPKGSPGEAGRPEAGLPKAKGL
 TGSPGSPGPDGKTGPPGAPGQDGRGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
 PGPPGAVGPAGKDGEGAGAQGPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPE
 QGVPDGLGAPGSPGARGERGFPPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGS
 QGAPGLQGMPPGERGAAGLPKPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPGPD
 KGESGSPGAPGPTGARGAPDRGEPGPPGAPGAGPPGADGQPGAKGEPGDAGAKGDAGP
 PGPAGPAGPPGPIGNVAGPAGKARGSAGPPGATGFPGAAGRVPVPPGPPSGNAGPPGPPGP
 AGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGPAGAPGTPGPPQGIAGQRGV
 VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
 EGSPGRDGSPPGPKGDRGETGPAGPPGAPGAPGAPGVPVGPAGKSGDRGETGPAGPAGPVG
 VGARGPAGPQGRGDKGETGEQGDRIKGRHGFSGLQGGPPGPPGSPGEEQGPSGASGPAGP
 RGPPGSAGAPGKDLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSGGDFDFSF
 LPQPPQEKAHDGGRYRADDANVVRDRDLEVDTTLSLSQQIENIRSPEGSRKNPARTCR

DLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK
 KRHVWFGESMTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ
 TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTTKTSRPLII
 DVAPLDVGAPDQEFQFDVGPVCF

>tr|A0A096P4E5|A0A096P4E5_PAPAN Collagen type I alpha 1 chain OS=Papio anubis GN=COL1A1
 PE=4 SV=1

MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGRLRYHDRDVWKPEPCRI
 CVCNDNGKVLCDVICDETKNCPGAEVPEGECCPVC PDGSESPDQETTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGP
 MGPSGPRGLPGPPGAPGPQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGR
 PGERGPPGQGARGLPGTAGLPGMKGHRGFSGLDGAAGDAGPAGPKGEPGSPGANGAPGQ
 MGPRGLPGERGRPGAPGARGNDGATGAAGPPGPTGPAGPPGFPAGVAVGAKGEAGPQGA
 RGSEGPGQVRGEPGPPGAPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPARGPSGP
 QGPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPGVGVQGGPPGAGEEGKRGARGEPGPTGL
 PGPPGERGGPGSRGFPAGDGVAGPKGPAGERGSPGAPGPKGSPGEAGRPEAGLPAGKGL
 TGSPGSPGPDGKTGPPGAPQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
 PGPPGAVGPAGKDGEAGAQQPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGE
 QGVPGDLGAPGSPGARGERGFPPGERGVQGGPPGAPPRGANGAPGNDGAKGDAGAPGAPGS
 QGAPGLQGMPPGERGAAGLPKPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPAGPD
 KGETGPSGPAGPTGARGAPDRGEPGPPGAPGAGPPGADGQPGAKGEPGDAGAKGDAGP
 PGPAGPAGPPGPIGNGAPGPKGARGGAPGATGFPGAAGRVPVGGPSGNAGPPGPPGP
 AGKEGGKGRGETGPAGRPGEVGGPPGPPGAGEKGSPPGADGAPGAPGTPGPGQGIAGQRGV
 VGLPGQRGERGFPLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
 EGSPGRDGSPPGKDRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETVSSWAPVPCVW
 SGQGLSRPP

>tr|F7A3F7|F7A3F7_HORSE Collagen type I alpha 1 chain OS=Equus caballus GN=COL1A1
 PE=4 SV=1

MFSFVDLRLLLLLLAATALLTHGQEEGQEEGQEEEDIPAVTICIQDGLRYHDRAVWKPEPCRV
 CICDNGNVLCDDVICEDTKNCPGASVPKDECCPVCPEGQVSPTDDQTTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLPGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSAGISVPGPM
 GPSGPRGLPGPPGAPGPQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGR
 GERGPPGQGARGLPGTAGLPGMKGHRGFSGLDGAAGDAGPAGPKGEPGSPGANGAPGQ
 GPRGLPGERGRPGAPGARGNDGATGAAGPPGPTGPAGPPGFPAGVAVGAKGEAGPQGA
 GSEGPGQVRGEPGPPGAPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPARGPSGP
 GPSGPPGPKGNSGEPGAPGNKGDGTGAKGEPGPTGIQGGPPGAGEEGKRGARGEPGPTGLP
 GPPGERGGPGARGFPAGDGVAGPKGPAGERGAPGAPGPKGSPGEAGRPEAGLPAGKGLT
 GSPGSPGPDGKTGPPGAPQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
 GPPGAVGPAGKDGEAGAQQPPGAPGAGERGEQGPAGSPGFQGLPGPAGPPGESGKPGEQ
 GVPGDGAPGSPGARGERGFPPGERGVQGGPPGAPPRGNSGAPGNDGAKGDAGAPGAPGSQ
 GAPGLQGMPPGERGAAGLPKPKGDRGDAGPKGADGSPGKDGVRGLTGPPIGPPGAPAGPD
 GETGPSGPAGPTGARGAPDRGEPGPPGAPGAGPPGADGQPGAKGEPGDAGAKGDAGPP
 GPAGPAGPPGPIGNSVAGPAPGPKGARGGAPGATGFPGAAGRVPVGGPSVSIKGGDATS LG
 VGAQAVRRVVGPTASVVLQNPLESDFSPRSGKCNWPLDPPGPVVKEGGKGRVRLGP
 LTGDRGEAVSSWVPACCFCSSRGLSRPPQGPVDGHLTSPRLGGEDILQDVQAFGLPENIE
 PQISFKAPVQQYVPSEEWLSPGASPGSHDAKEEGLGLGGPAGPAGPIGPVARGPAGP
 QGPRGDKGETGEQGDRIKGRGFSGLQGGPPGPPGSPGEGQGPSGASGPAGPRGPPGSAGA
 PGKDGLNGLPGPIGPPGPRGRTGDAGVGGPPGPPGPPGPPGPPSGGDFDFLQPPQEK
 HDGGRYRADDANVVRDRDLEVDTTLSLSQIENIRSPESRKNPARTCRDLKMCHSDW
 KSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRHVWYGES
 MTDGFQFEYGGQSDPADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQTGNLKKALL
 LQGSNEIEIRAEGNSRFTYSVTYDGTSTHTGAWGKTVIEYKTTKTSRPLIIDVAPLDIGA
 PDQEFIDIGPVCFL

>ENST00000225964 peptide:ENSP00000225964 pep:KNOWN_protein_coding
 MFSFVDLRLLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGRLRYHDRDVWKPEPCRI
 CVCNDNGKVLCDVICDETKNCPGAEVPEGECCPVC PDGSESPDQETTGVGPKGDTGPR
 GPRGPAGPPGRDGIPIGQPLPGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGP
 MGPSGPRGLPGPPGAPGPQGFQGGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGR

PGERGGPPGQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQ
MGPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFFGAVGAKGEAGPQGP
RGSEGPQGVRRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGP
QGPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPAGPAGEEGKRGARGEPTGL
PGPPGERGGPSRGGFPADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPEAGLPGAKGL
TGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGV
PGPPGAVGPAGKDGEGAQPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGE
QGVPGDLGAPGSPGARGERGFPGERGVQGPAGPRGANGAPGNDGAKGDAGAPGAPGS
QGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPD
KGESGSPGAPPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGP
PGPAGPAGPPGPIGNVAGPAGAKGARGGAGPPGATGFPGAAGRVPAGPSGNAGPPGPPGP
AGKEGGKGRGETGPAGRPGEVGPAGPPGPPGAGEKGSPPGADGPAGAPGTPGQGIAGQRGV
VGLPGQRGERGFPLPGPSGEPGKQPSGASGERGPPGPMGPPGLAGPPGESGREGAPGA
EGSPGRDGSPPGAKGDRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGPAGPTGPVGP
VGARGPAGPQGPARGDKGETGEQDGRGKGRGHRGFSGLQGPPGPPGSPGEEQPSGASGAPG
RGPPGSAGAPGKDLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSAGFDFSF
LPQPPQEKAGHDGGRYRADDANVVRDRDLEVDTTLSLSQQIENIRSPEGSRKNPARTCR
DLKMCHSDWKSGEYWIDPNQGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK
KRHVWFGESMTDGFQFEYGGQGSADVAIQLTFLRLMSTEASQNITYHCKNSVAYMDQQ
TGNLKKALLLQGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTTKTSRLPII
DVAPLDVGAPDQEFGFDVGPVCF

II. Séquences des protéines utilisées pour l'analyse phylogénétique du collagène I alpha 2

>sp|P08123|CO1A2_HUMAN Collagen alpha-2(I) chain OS=Homo sapiens GN=COL1A2 PE=1 SV=7
 MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
 PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQGFQGPAGEP
 GEPGQTGPAGARGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLPGFKGIR
 GHNGLDGLKGQPGAPGVKGEPEGAPGENGTPGQTGARGLPGERGRVGAAPGARGSDGSV
 GPVGPAGPIGSAGPPGFPAGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNP
 GANGLTGAKGAAGLPVAGAPGLPGRGIPGPVGAAGATGARGLVGEPPAGSKGESGNK
 GEPGSAGPQGGPPGSGEKGKRGPNGEAGSAGPPGPPGLRGSPPSRGLPGADGRAGVMGPP
 GSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIPAGKEGPVGLPGIDGRPGPI
 GPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPDPGNGAQQGPPGPQGVQ
 GGKGEQGGPPGPPGFQGLPGPSGPAGEVVKPGERGLHGEFGLPGPAGPRGERGPPGESGAA
 GPTGPIGSRGSPGPPGPDGNKGEPPVAVGTAGPESGSPGLPGERGAAGIPGGKGEKGEK
 GLRGEIGNPGRDARGAPGAVGAPGATGDRGEAGAAGPAGPAGPRGSPGERGEVGA
 GPNGFAGPAGAAGQPGAKGERGAKGPKGENGVGPTGPVGAAGPAGPNGPPGPAGSRGDG
 GPPGMTGFPGAAGRTGPPGSPGISGPPGPPGPAKKEGLRGRGDQGPVGRTEVAVGPP
 GFAGEKGPSGEAGTAGPPGTPGQGLLAPGILGLPGRGERGLPGVAVAVGEPGLGIA
 GPPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDQPGHKGERGYPGNIGPVGAA
 GAPGPHGPVGPAGKHGHRGETGPSGPVGPAGAVGPRGSPGQGIRGDKGEPGEKGRGLP
 GLKHGHLQGLPGIAGHHGDQAGPSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIR
 GPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGFYRADQPRAPSRLPKDYEVDATLK
 SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG
 ETCIRAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLL
 ANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKK
 TNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

>sp|Q01149|CO1A2_MOUSE Collagen alpha-2(I) chain OS=Mus musculus GN=Col1a2 PE=1 SV=2
 MLSFVDTRTLLLLAVTSCLATCQYLQSGSVRKGPAGDRGPRGRGQRPAGPRGRDGVDPGMG
 PPGPPGSPGPPGSPAPPGLTGNFAAQYSDKGVSSGPGPMGLMGPRGPPGAVGAPGPQGFQ
 GPAGEPGEPEGQTGPAGPRGPAGSPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLP
 GFKGVKGHSGMDGLKGQPGAQGVKGEPEGAPGENGTPGQAGARGLPGERGRVGAAPGAGAR
 GSDGSVGPVGPAGPIGSAGPPGFPAGPKGELGPVGNPAGPAGPRGEVGLPGLSGPV
 GPPGNPNTGLTGAKGATGLPGVAGAPGLPGRGIPGPAGAAGATGARGLVGEPPAGSK
 GESGNKGEPSVGAQQGPPGSPGEEGKRGSPGEAGSAGPAGPPGLRGSPPSRGLPGADGRA
 GVMGPPGNRGSTGPAGIRGPNGDAGRPGEPGLMGPRGLPGSPGNVGPSGKEGPVGLPGID
 GRPQPIGAPPRGEAGNIGFPGPKGSPGDPGKPRGERGHPGLAGARGAPDPGNGAQQGPP
 GPQGVQGGKGEQGPAGPPGFQGLPGPSGTEAGTTEVVKPGERGLPGEFGLPGPAGPRGERGTP
 GESGAAGSPGPIGSRGSPGAPDPGNKGEAGAVGAPGSAGASGPGGLPGERGAAGIPGGK
 GEKGETGLRGDTGNTGRDARGIPGAVGAPGAGASGDRGEAGAAGPSGPAGPRGSPGER
 GEVGPAGPNGFAGPAGAAGQPGAKGEKGTGPKGENGIVGPTGSVGAAGPSGPNGPPGPV
 GSRGDGPPGMTGFPGAAGRTGPPGSPGIAGPPGPPGAAGKEGIRGPRGDQGPVGRTEGET
 GASGPPGFVGEKGPSGEPGTAGAPGTAGPQGLLAPGILGLPGRGERGLPGIAGALGEP
 GPLGISGPPGARGPPGAVGSPGVNGAPGEAGRDGNPDSGPPGRDQPGHKGERGYPGSI
 GPTGAAGAPGPHGSVGPAGKHGHRGEPGAGSVGPVAVGPRGSPGQGIRGDKGEPGDK
 GHRGLPGLKGYSLQGLPGLAGLHGDQAGPVPVGPAGPRGPAGPSGPVGDGRSGQPGPV
 GPAGVRGSQGSQGPAGPPGPPGPPGPPGVSGGGYDFGFEGDFYRADQPRSQPSLRPKDYE
 VDATLKSLLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWNSDYYWIDPNQGCTMDAIKVY
 CDFSTGETCIQAQPVNTPAKNSYSRAQANKHVWLGETINGGSQFEYNVEGVSSKEMATQL
 AFMRLLANRASQNITYHCKNSIAYLDEETGSLNKAVLLQGSNDVELVAEGNSRFTYSVLV
 DGCSKKTNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFRVEVGPVCFK

>sp|P02465|CO1A2_BOVIN Collagen alpha-2(I) chain OS=Bos taurus GN=COL1A2 PE=1 SV=2
 MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPSGDRGPRGERGPPGPPGRDGGDDGIPG
 PPGPPGPPGPPGLGGNFAAQFDKGGGPGPMGLMGPRGPPGASGAPGPQGFQGGPPGEPGE
 PGQTGPAGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLPGFKGIRGH

NGLDGLKGQPGAPGVKGEPEGAPGENGTGPGQTGARGLPGERGRVVGAPGPAGARGSDGSVGP
 VGPAGPIGSAGPPGFPGAPGPKGELGPVGNPGPAGPAGPRGEVGLPGLSGPVGPPGNPGA
 NGLPGAAGKGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGLVGEPGPAGSKGESGNKGE
 PGAVGQPGPPGSPGEEGKRGSTGEIGPAGPPGPPGLRGNPGRSLPGADGRAGVMGPAGS
 RGATGPAGVRGPNNGDSGRPGEPGLMGRGFPSPGNIGPAGKEGPVGLPGIDGRPGPIGP
 AGARGEPGNIGFPGPKGSPGDPGKAGEKGHAGLAGARGAPGPDGNNGAQQGPPGLQGVQGG
 KGEQGPAGPPGFQGLPGPAGTAGEAGKPGERGIPEFGLPGPAGARGERGPPEGESGAAGP
 TGPIGSRGSPGPPGPDGNKGEPEGVVGAPGTAGPSGSPGLPGERGAAGIPGGKGEKGETGL
 RGDIGSPGRDGARGAPGAIGAPGPAGANGDRGEAGPAGPAGPAGPRGSPGERGEVGPAGP
 NGFAGPAGAAGQPGAKGERGTGKPKGENGVPPTGPVGAAGPSGPNGPPGPAGSRGDGGP
 PGATGFPGAAGRTGPPGSPGSGPPGAGKGLRGRGPRGDQGPVGRSGETGASGPPGF
 VGEKGPSGEPGTAGPGTGPQGLLGPAGFLGLPGRGRERGLPGVAGSVGEPGLGIAGP
 PGARGPPGNVGNPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNAGPVGAAGA
 PGQQGPVGPVGVKHGNRGEPGPAGAVGPAGAVGPRGSPGQGIRGDKGEPGDKGPRGLPGL
 KGHNLQGLPGLAGHHGDQGAPGAVGPAGPRGPAGPSGPAGKDRIGQPGAVGPAGIRGS
 QGSQGPAGPPGPPGPPGPPGSPGGGYEFGFDGDFYRADQPRSPTSLRPKDYEVDTLKSL
 NNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGET
 CIRAQPEDIPVKNWYRNSKAKKHVVWVGETINGGTQFEYNVEGVTTKEMATQLAFMRLLAN
 HASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTN
 EWQKTIIEYKTNKPSRLPILDIAPLDIGGADQEIRLNIGPVCFK

>sp|P02466|CO1A2_RAT Collagen alpha-2(I) chain OS=Rattus norvegicus GN=Col1a2 PE=1 SV=3
 MLSFVDTRLLLLAVTSCLATCQSLQMGSVRKGPTGDRGPRGQRGPAGPRGRDGDVDPVGP
 PPGPPGAPGPPGPPGPPGLTGNFAAQYSDKGVSAAGPMPGLMGRGPPGAVGAPGPQGFQ
 GPAGEPEGPGQTGPAGSRGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTGGLP
 GFKGIRGHNLGLDGLKGQPGAQGVKGEPEGAPGENGTGQAGARGLPGERGRVVGAPGPAGAR
 GSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGELGPVGNPGPAGPAGPRGEAGLPLSGPV
 GPPGNPGANGLTGAKGATGLPGVAGAPGLPGPRGIPGPVGAAGATGPRGLVGEPEGPAGSK
 GETGNKGEPEGSAQAQGGPPGSPGEEGKRGSPGEPGSAGPAGPPGLRGSPPSRGLPGADGRA
 GVMGPPGNRSTGPAGVRGPNGDAGRPGEPGLMGRGLPGSPGNVGPAGKEGPVGLPGID
 GRPGPIGPAGPRGEAGNIGFPGPKGSPGDPGKPEKGHPLAGARGAPGPDGNNGAQQGPP
 GPQGVQGGKGEQGPAGPPGFQGLPGPSGTAGEVGKPGERGLPEFGLPGPAGPRGERGPP
 GESGAAGPSGPIGIRGSPGAPGPDGNKGEAGAVGAPGSAGASGPGGLPGERGAAGIPGGK
 GEKGETGLRGEIGNPGRDGARGAPGAIGAPGPAGASGDRGEAGAAGPSGAPGRGSPGER
 GEVGPAGPNFAGPAGSAGQPGAKGEKGTGPKGENGIVGPTGPVGAAGPSGPNGPPGPA
 GSRGDGPPGMTGFPGAAGRTGPPGSPGITGPPGPPGAAGKEGIRGPRGDQGPVGRTEI
 GASGPPGFAGEKGPSGEPGTTGPPGTAGPQGLLGPAGILGLPGRGRERGQPIAGALGEP
 GPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGSDGPPGRDGQPGHKGERGYPGNI
 GPTGAAGAPGPHGSVGPAGKHGNRGEPGPAGSVGPVAVGPRGSPGQGIRGDKGEPGDK
 GARGLPGLKGHNLQGLPGLAGLHGDQGAPGPVGPAGPRGPAGPSGPIKDRSGHPGPV
 GPAGVRGSQGSQGPAGPPGPPGPPGPPGVSGGGYDFGFEGGFYRADQPRSQPSLRPKDYE
 VDATLKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWKSYYWIDPNQGCTMDAIKVY
 CDFSTGETCIRAQPVNTPAKNAYSRAQANKHVWLGETINGGSQFEYNAEGVSSKEMATQL
 AFMRLLANRASQNITYHCKNSIAYLDEETGRLNKAVILQGSNDVELVAEGNSRFTYTVLV
 DGCSSKKTNEWDKTVIEYKTNKPSRLPFLDIAPLDIGGTNQEFRVEVGPVCFK

>sp|P02467|CO1A2_CHICK Collagen alpha-2(I) chain OS=Gallus gallus GN=COL1A2 PE=1 SV=3
 MLSFVDTRILLLLAVTSYLATSQHLFQASAGRKGRGDKGPPQGERGPPGPPGRDGEDGPP
 GPPGPPGPPGLGGNFAAQYDPSKAADFGPGPMGLMGRGPPGASGPPGPPGFQGVPEGP
 EPQGTGPQGRGPPGPPGKAGEDGHPGKPRPGERGVAGPQGARGFPPTGGLPGFKGIRG
 HNLGDGQKQPGTPTGKGEPEGAPGENGTGQPGARGLPGERGRIGAPGPAGARGSDGSAG
 PTGPAGPIGAAGPPGFPGAPGAKGEIGPAGNVGPTGPAGPRGEIGLPGSSGPVGPNNPG
 ANGLPGAAGLPGVAGAPGLPGPRGIPGPPGPAGPSGARGLVGEPGPAGAKGESGNKGE
 EPGAAGPPGPPGSPGEEGKRGSNGEPEGSAAGPPGAGLRGVPGSRGLPGADGRAGVMGPAG
 NRGASGPVGAAGPNGDAGRPGEPGLMGRGLPGQGPSPGAPGKEGPVGFPGADGRVGPPIG
 PAGNRGEPGNIGFPGPKGPTGEPGKPEKGNVLAGPRGAPGPEGNNGAQQGPPGVTGNQG
 AKGETGPAGPPGFQGLPGSPGAGEAGKPGERGLHGEFVGPAGPRGERGLPGESGAVG
 PGPAGSRGSPGPPGPDGNKGEPEGVNPAGAPGAGPGGIPGERGVAGVPVGGKGEKAGP
 LRGDTGATGRDGARGLPGAIGAPGAGGAGDRGEGGPAGPAGPAGARGIPGERGEPGPV
 PSGFAGPPGAAGQPGAKGERGPKGPKGETGPTGAIGPIGASGPPGPVGAAGPAGPRGDAG

PPGMTGFPGAAGRVPVGGPAGITGPPGPPGAGKDGPRGLRGDVGVPVGRTEEQGIAGPPG
 FAGEKGPSGEAGAAGPPGTPGPQGILGAPGILGLPGSRGERGLPGIAGATGEPGPLGVSG
 PPGARGPSGPVGSPPNGAPGEAGRDGNPNDGPPGRDGPAGFKGERGAPGNPSPGALG
 APGPHGQVGPSPKPNRGGDPVGPVGPAGAFGPRGLAGPQGPGEKGEKGDGHRGLPG
 LKGHNGLQGLPGLAGQHGDQGGPPNNGPAGPRGPPGSPGPKDGRNGLPGPIGPAGVRG
 SHGSQGPAGPPGPPGPPGPPGPPGNGGGYEVGFDAEYRADQPSLRPKDYEV DATLKLNNQ
 IETLLTPEGSKKNPARTCRDLRLSHPEWSSGFYWIDPNQGCTADAIRAYCDFATGETCIH
 ASLEDIPTKTWYVSKNPKDKKHIWFGETINGGTQFEYNVEGVTTKDMATQLAFMRLANH
 ASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELRAEGNSRFTFSVLVDGCSKKNK
 WGKTIIERTNKPSRPLILDIAPLDIGGADQEFGLHIGPVCFK

>sp|O46392|CO1A2_CANLF Collagen alpha-2(I) chain OS=Canis lupus familiaris GN=COL1A2 PE=2
 SV=2

MLSFVDTRTLLLLAVTSCLATCQSLQEATARKGPTGDRGPRGERGPPGPPGRDGDGIPG
 PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGMGLMGPRGPPGASGAPGQGFQGPAGEP
 GEPGQTGPAGARGPPGPPGKAGEDGHPGKPRGERGVVGPQGARGFPGTPLPGFKGIR
 GHNGLDGLKGQPGAPGVKGEPPGAPGENGTPGQTGARGLPGERGRVGPAGARGSDGSV
 GPVGPAGPIGSAGPPGFPAGPKGEIGPVGNPAGPAGPRGEVGLPGVSGPVGPPGNP
 GANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGIVGEPGAGSKGESGNK
 GEPGSAGAQQPPGSPGEEGKRGPNGEAGSAGSPGPPGLRGSPPSRGLPGADGPAGVMGPP
 GPRGATGPAGVRGPNDSRPGEPGLMGPRGFPAGPNVGPAGKEGPMGLPGIDGRPGPI
 GPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQQPPGPQGVQ
 GKGGEQGPAGPPGFQGLPGPAGTAGEVVKPGERGLPGEFGLPGPAGPRGERGPPGESGAA
 GPSGPIGSRGSPGPPGPDGNKGEPPVGLGAPGTAGASGPGGLPGERGAAGIPGGKGEKGET
 GLRGEIGNPGRDARGAPGAMGAPGAGATGDRGEAGPAGPAGPAGPRGTPGERGEVGA
 GPNGFAGPAGAAGQPAGKGERGTPGPKGENGPVGPPTGPIGSAGSPGPNPAGSRGDG
 GPPGATGFPGAAGRTGPPGPGSITGPPGPPGAAGKEGLRGRDQGPVGRTEGASGPP
 GFTGEKGPSGEPGTAGPPGTPGPQGLLAPGILGLPGSRGERGLPGVAGSVGEPGLGIA
 GPPGARGPPGAVGAPGVNGAPGEAGRDGNPNDGPPGRDGGQAGHKGERGYPNIGPVGAV
 GAPGPHGVPVGTGKHGHRGEPGAGSVGPVAVGPRGSPGQGIRGDKGEPGEKGRGLP
 GLKGHNGLQGLPGLAGQHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGQPGTVGPAGIR
 GSQGSQGPAGPPGPPGPPGPPGPPGSPGGYDFGYEGDFYRADQPRSPSLRPKDYEV DATLK
 SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG
 ETCIRAQPENIPAKNWYRNSKVKKHIWLGETINGGTQFEYNVEGVTTKEMATQLAFMRL
 ANHASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSK
 TNEWRTIIIEYKTNKPSRPLILDIAPLDIGDADQEFRVDVGPVCFK

>tr|H2QUY2|H2QUY2_PANTR Uncharacterized protein OS=Pan troglodytes GN=COL1A2 PE=4
 SV=1

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
 PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGMGLMGPRGPPGAAGAPGQGFQGPAGEP
 GEPGQTGPAGARGPAGPPGKAGEDGHPGKPRGERGVVGPQGARGFPGTPLPGFKGIR
 GHNGLDGLKGQPGAPGVKGEPPGAPGENGTPGQTGARGLPGERGRVGPAGARGSDGSV
 GPVGPAGPIGSAGPPGFPAGPKGEIGAVNAGPAGPAGPRGEVGLPGLSGPVGPPGNP
 GANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPGAGSKGESGNK
 GEPGSAGPQQPPGSPGEEGKRGPNGEAGSAGPPGPPGLRGSPPSRGLPGADGRAGVMGPA
 GSRGASGPAGVRGPNGDAGRPGEPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPI
 GPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQQPPGPQGVQ
 GKGGEQGPAGPPGFQGLPGPSGPTGEVVKPGERGLHGEFGLPGPAGPRGERGPPGESGAA
 GPTGPIGSRGSPGPPGPDGNKGEPPVGVAVGTAGPSGSPGLPGERGAAGIPGGKGEKGEK
 GLRGEIGNPGRDARGAPGAVGAPGAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGA
 GPNGFAGPAGAAGQPAGKGERGAKGPKGENGVVGPVGAAGPAGPNPAGSRGDG
 GPPGMTGFPGAAGRTGPPGPGSISGPPGPPGAGKEGLRGRDQGPVGRTEVAVGPP
 GFAGEKGPSGEAGTAGPPGTPGPQGLLAPGILGLPGSRGERGLPGVAVGAVGEPGLGIA
 GPPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDGGQGHKGERGYPNIGPVGAA
 GAPGPHGVPVGPAGKHGHRGEPGAGSVGPAGAVGPRGSPGQGIRGDKGEPGEKGRGLP
 GLKGHNGLQGLPGLAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIR
 GPQGHGQGPAGPPGPPGPPGPPGPPGSPGGYDFGYDGFYRADQPRSAPSLRPKDYEV DATLK
 SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG
 ETCIRAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRL

ANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKK
TNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK
>tr|H9Z2D1|H9Z2D1_MACMU Collagen alpha-2(I) chain OS=Macaca mulatta GN=COL1A2 PE=2
SV=1
MLSFVDTRTLLLLAVTSCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGRGPPGAAGAPGPQGFQGPAGEP
GEPGQTGPAGSRGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPLPGFKGIR
GHNGLDGLKGQPGAPGVKGEPEGAPGENGTPGQTGARGLPGERGRVGPAGPARGSDGSV
GPVGPAGPIGSAGPPGFPAPGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNP
GANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPPGAGSKGESGNK
GEPGSAGPQGPVPGSGEEGKRGPNGEVGSAGPPGGLRGGPGSRGLPGADGRAGVMGPP
SGRSGSPAGVRGPGSDAGRPGEPGLMGRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPI
GPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQQPPGPQGVQ
GKGEQGPAGPPGFQGLPGPSGPAGEVVKPGERGLPGEFGLPGPAGARGERPPEGESGAA
GPTGPIGSRGSPGPPGPDGNKGEPPGVVGAAGTAGPSGSPGLPGERGAAGIPGGKGEKGE
GLRGEIGNPGRDGARGAPGAVGAPGAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGA
GPNGFAGPAGAAGQPAGKGERGAKGPKGENGVVGTGPVGAAGPSGPNPAGPAGSRGDG
GPPGMTGFPGAAGRTGPPGSPGISGPPGPPGSPGKEGLRGRGDQGPVGRTEVAVGPP
GFAGEKGPSGEAGTAGPPGTPGPQGLLAPGILGLPSRGERGLPGVAVVGEPPGLGIA
GPPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDQPGHKGERGYPGNNGPVGAA
GAPGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGSPGQGIRGDKGEPGDKGPRGLP
GLKGHNGLQGLPGLAGHHGDQAGPSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIR
GPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGFYRADQPRAPSRLPKDYEV DATLK
SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG
ETCIRAQENIPAKNWRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLL
ANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKK
TNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGANQEFFVDIGPVCFK

>tr|H2PMW7|H2PMW7_PONAB Collagen type I alpha 2 chain OS=Pongo abelii GN=COL1A2 PE=4
SV=1
MLSFVDTRTLLLLAVTSYLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGRGPPGAAGAPGPQGFQGPAGEPG
EPGQTGPAGARGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPLPGFKGIRG
HNGLDGLKGQPGAPGVKGEPEGAPGENGTPGQTGARGLPGERGRVGPAGPARGSDGSV
PVGPAGPIGSAGPPGFPAPGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNP
ANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPPGAGSKGESGNK
EPGSAGPQGPVPGSGEEGKRGPNGEAGSAGPPGPPGLRGGSPSRGLPGADGRAGVMGPPG
SRGASGPAGVRGPGSDAGRPGEPGLMGRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIG
PAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQQPPGPQGVQ
GKGEQGPAGPPGFQGLPGPSGPAGEVVKPGERGLHGEFGLPGPAGPRGERGPPGESGAAG
PTGPIGSRGSPGPPGPDGNKGEPPGVVGAAGTAGPSGSPGLPGERGAAGIPGGKGEKGE
LRGEIGNPGRDGARGAPGAVGAPGAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAG
PNGFAGPAGAAGQPAGKGERGTKGPKGENGVVGTGPVGAAGPAGPAGPAGSRGDG
PPGMTGFPGAAGRTGPPGSPGISGPPGPPGAGKEGLRGRGDQGPVGRTEVAVGPPG
FAGEKGPSGEAGTAGPPGTPGPQGLLAPGILGLPSRGERGLPGVAVVGEPPGLGIAG
PPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDQPGHKGERGYPGNIGPVGAAG
APGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGSPGQGIRGDKGEPGDKGPRGLPG
LKGHNGLQGLPGLIAGHHGDQAGPSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIRG
PQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGFYRADQPRAPSRLPKDYEV DATLKS
LNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGE
TCIRAQENIPAKNWRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLA
NYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKT
NEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

>tr|F6RUA6|F6RUA6_HORSE Collagen type I alpha 2 chain OS=Equus caballus GN=COL1A2 PE=4
SV=1
MLSFVDTRTLLLLAVTSCLATCQSLQEATAGKGPAGDRGPRGERGPPGPPGRDGEDDGI
PPGPPGPPGPPGLGGNFAAQFDAQGGGPGPMGLMGRGPPGASGAPGPQGFQGPAGEPGE
PGQTGPAGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPLPGFKGIRGH
NGLDGLKGQPGAPGVKGEPEGAPGENGTPGQAGARGLPGERGRVGPAGPARGSDGSVGP

VGPAGPIGSAGPPGFPAGPCKGELGPVGNPAGPAGPRGEVGLPGLSGPVGPPGNPGA
 NGLTGAKGAAGLPGVAGAPGLPGRGIPGPAGAAGATGARGLVGEPGPAGSKGESGNKGE
 PGAAGPQGGPPGSGEEGKRGNPGEPESTGPAGPPGLRGSPPSRGLPGADGRAGVMGPAGS
 RGASGPAGVRGPNGDSGRPGEPGLMGRGFPSPGNIGPAGKEGPVGLPGIDGRPGPIGP
 AGARGEPGNIGFPGPKGSPGEPGKPGDKGHAGLAGARGAPGPDGNNGAQGGPPGQGVQGG
 KGEQGPAGPPGFQGLPGPAGTAGEVKGPERGLPGEFGLPGPAGARGERGPPGESGAAGP
 AGPIGSRGPPSGPPGPDGNKGEVGLGAPGTAGPSGSPGLPGERGAAGIPGGKGEKGETGL
 RGEIGNPGRDGARGAPGAVGAPGPAGANGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGP
 NGFAGPAGAAGQPGAKGERGKPKGKENGVPVGTGPVGAAGPSGPNGPPGPAGSRGDGGP
 PGVTGFPGAAGRTGPPGSPGISGPPGPPGAAGKEGLRGRGDQGPVGRAGETGASGPPGF
 AGEKGPSGEPGTAGPPGTPGQGLLGAAGILGLPGRGERGLPGVAGSLGEPGLGIAGP
 PGARGPPGAVGAPGAVGAPGEAGRDGNPNSDGGPPGRDQPGHKGGERGYPGNAGPVAVGA
 PGPHGVPVGTGKHGHRGEPGVPVGVGAVGPRGSPGQGVGRGDKGEPGDKGPRGLPGI
 KGHNGLQGLPGLAGQHGDQAGPSVGPAGPRGPAGPTGPVVKDGRSGQPGTVGPAGVRS
 QGSQGPAGPPGPPGPPGPPGSGGGYDFGYDGFYRADQPRSPSLRPKDYEV DATLKSL
 NNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGET
 CIRAQPENIPAKNWYRSSKAKKHIWLGETINGGTQFEYNVEGVTTKEMATQLAFMRLLAN
 HASQNITYHCKNSIAYLDEETGNLKKAVTLQGSNDVELVAEGNSRFTYTVLVDGCSRKTN
 EWGKTIIYEYKTNKPSRLPILDIALLDIGGADQEFGLDIGPVCFK

>tr|A0A096NVZ9|A0A096NVZ9_PAPAN Collagen type I alpha 2 chain OS=Papio anubis
 GN=COL1A2 PE=4 SV=1

MLSFVDTRTLLLLAVTSCLATCQCGPAGDRGPRGERGPPGPPGRDGEDGPTGPPGPPGPP
 GPPGLGGNFAAQYDGKGVGLGPGPMGLMGRGPPGAAGAPGPQGFQGPAGEPEPEGQTGP
 AGSRGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTPLPGFKGIRGHNGLDGL
 KGQPGAPGVKGEPEGAPGENGTPGQTGARGLPGERGRVGAAGPAGARGSDGSVGPVGPAGP
 IGSAGPPGFPGAPGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNPANGLTGA
 KGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPGPAGSKGESGNKGEPSAGP
 QGPPGSGEEGKRGNPGEAGSAGPPGPPGLRGGPGRGLPGADGRAGVMGPPGSRGASGP
 AGVRGPNGDAGRPGEPGLMGRGFPSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGE
 PGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGGPPGQGVQGGKGEQGP
 AGPPGFQGLPGPSGPAGEVVKPERGLPGEFGLPGPAGARGERGPPGESGAAGPTGPIGS
 RGPSGPPGPDGNKGEVGVGAAGTAGPSGSPGLPGERGAAGIPGGKGEKGEPLRGEIGN
 PGRDGARGAPGAVGAPGAGTADRGEAGAAGPAGPAGPRGSPGERGEVGPAGPAGPAGP
 AGAAGQPGAKGERGAKGPKGKENGVPVGTGPVGAAGPSGPNGPPGPAGSRGDGGPPGMTGF
 PGAAGRTGPPGSPGISGPPGPPGAGKEGLRGRGDQGPVGRTEVAVGPPGPFAGEKGP
 SGEAGTAGPPGTPGQGLLGAAGILGLPGRGERGLPGVAGALGEPGLGIAGPPGARGP
 PGAVGSPGVNAPGEAGRDGNPNDGPPGRDQPGHKGGERGYPGNNGPVGAAGAPGPHGP
 VGPAGKHGNRGETGSPGPVGPAGAVGPRGSPGQGIRGDKGEPGDKGPRGLPGLKGHNGL
 QGLPGLAGHHGDQAGPSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIRGPQGHQGP
 AGPPGPPGPPGPPGVSGGGYDFGYDGFYRADQPRSPSLRPKDYEV DATLKSLNNQIET
 LLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQ
 ENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYSQNI
 TYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTI
 IEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

>tr|G3QT97|G3QT97_GORGO Collagen type I alpha 2 chain OS=Gorilla gorilla gorilla
 GN=COL1A2 PE=4 SV=1

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
 PPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGRGPPGAAGAPGPQGFQGPAGEP
 GEPGQTGPAGARGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTPLPGFKGIR
 GHNGLDGLKGQPGAPGVKGEPEGAPGENGTPGQTGARGLPGERGRVGAAGPAGARGSDGSV
 GPVGPAGPIGSAGPPGFPAGPCKGELGPVGNPAGPAGPRGEVGLPGLSGPVGPPGNP
 GANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPGPAGSKGESGNK
 GEPSAGPQGGPPGSGEEGKRGNPGEAGSAGPPGPPGLRGSPPSRGLPGADGRAGVMGPP
 GSRGASGPAGVRGPNGDAGRPGEPGLMGRGFPSPGNIGPAGKEGPVGLPGIDGRPGPI
 GPAGARGEPEGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPGPDGNNGAQGGPPGQGVQ
 GGKGEQGPAGPPGFQGLPGPSGPAGEVVKPERGLPGEFGLPGPAGARGERGPPGESGA
 GPTGPIGSRGSPGPPGPDGNKGEVGVGAAGTAGPSGSPGLPGERGAAGIPGGKGEKGE
 GLRGEIGNPGRDGARGAPGAVGAPGPAGTADRGEAGAAGPAGPAGPRGSPGERGEVGP

GPNGFAGPAGAAGQPAGKGERGAKGPKGENGVVGTGPVGAAGPAGPNGPPGPAGSRGDG
 GPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTEVAVGPP
 GFAGEKGPSGEAGTAGPPGTPGPQGLLAPGILGLPGRGERGLPGVAVAVGESGPLGIA
 GPPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDQGPQGHKGERGYPGNIGPVGAA
 GAPGPHGPVGPAGKHGHRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGRGLP
 GLKGHNGLQGLPGIAGHHGDQGASGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIR
 GPQGHQGPAGPPGPPGPPGPPGVSSGGYDFGYDGDYFADQPRAPSRLPKDYEVDTLTK
 SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG
 ETCIRAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLL
 ANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKK
 TNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

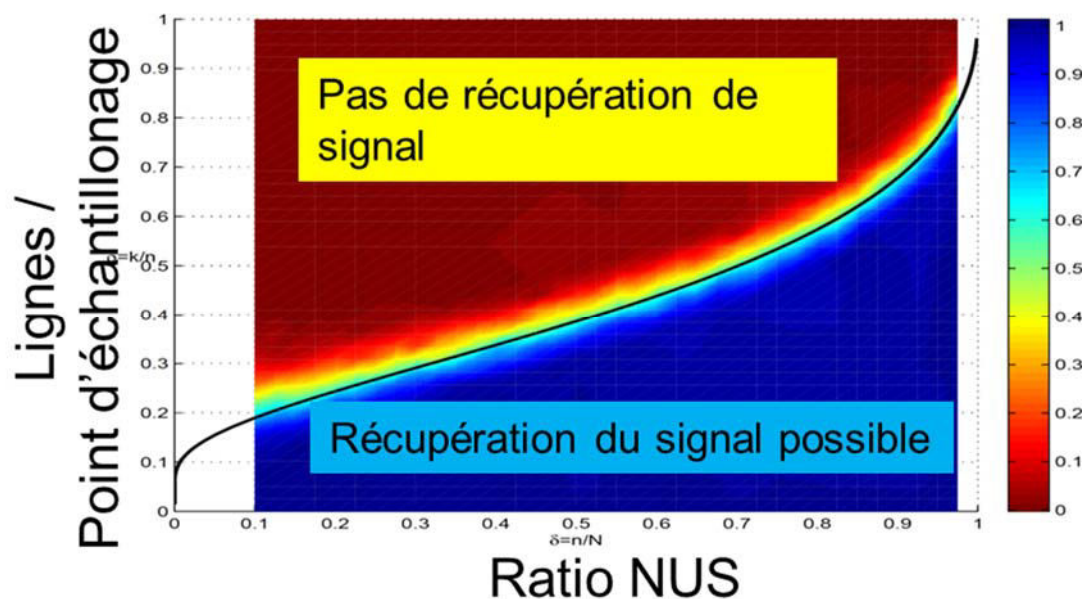
>tr|G1MH95|G1MH95_AILME Collagen type I alpha 2 chain OS=Ailuropoda melanoleuca
 GN=COL1A2 PE=4 SV=1

MLSFVDTRTLLLLAVTSCLATCQFYTLWILRSNCIEQIHITNFVFMRSSEGGKLPKPKGM
 KTDLGKSEDINAGCGPPSRIVTIATDVLTPNDTLLGHLIKGPAGDRGPRGERGPPGPPGR
 DGDDGIPGPPGPPGPPGPPGLGGNFAAQYDPGKGVGLGPGMGLMGPGRGPPGASGAPGPQ
 GFQGPAGEPGEPPQTGPAGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTP
 GLPGFKGIRGHNGLDGLKGQPGAPGVKGEPPGAPGENTPGQTGARGLPGERGRVGPAGPA
 GARGSDGSVGPVGPAGPIGSAGPPGFPAGPKGELGPVGNPGPAGPAGPRGEVGLPGVS
 GPVGPNGPANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPGPA
 GSKGESGNKGEPPKQGSVGPQGPSPGEEGKRGPNGEAGSAGPSGPPGLRGSPPSRGLPG
 ADGRAGVMGPPGPRGSTGPAGVRGPNDSGRPGEPGLMGPGRGFPAGPNVGPAGKEGPMG
 LPGIDGRPGPIGPAGARGEPGNIGFPGPKGPSGEPGKAGEKGHAGLAGARGAPDPGNNG
 AQQPPGPQGVQGGKGEQGPAGPPGFQGLPGPAGTAGEVGKPGERGLPGEFGLPGPAGPRG
 ERGPPGESGAAGPSGPIGSRGPSGPPGPDGNKGEPPVLAGPAGTAGPSGPPGLPGERGAAG
 VPGGKGEKGETGLRGEVGNPGRDARGAPGAVGAPGAGATGDRGEAGPAGPAGPAGPRG
 SPGERGEVGPAGPNGFAGPAGAAGQPAGKGERGKGPKEGNGPVGTPGPVGSAGPSGPNP
 PPGPAGSRDGGPPGATGFPGAAGRTGPPGPSGITGPPGPPGAAGKEGLRGPRGDQGPV
 RTGETGAHGGPPGFAGEKGPSGEPGTAGPPGTAGPQGLLAPGILGLPGRGERGLPGVSG
 SVGEPGLGIAGPPGARGPPGAVGAPGVNGAPGEAGRDGNPNDGPPGRDQGPQGHKGERG
 YPGNIGPVGTVPAGPHGPVGTGKHGHRGEPGPAGAVGPVAVGPRGPSGPQGVGRDQK
 EPGDKGPRGLPGLKGHNGLQGLPGLAGQHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTG
 HPGTVGPAGVRGSQGSQGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP
 PKDYEVDTLTKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMD
 AIKVYCDFSTGETCIRAQPENIPAKNWYRNSKVKKHIWLGETINGGTQFEYNVEGVTTKE
 MATQLAFMRLLANHASQNITYHCKNSIAYMDEETGNLKKAVMLQGSNDVELVAEGNSRFT
 YTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPILDIAPLDIGGADQEFRVDVGPVCFK

>ENST00000297268 peptide:ENSP00000297268 pep:KNOWN protein coding
 MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGEDGPTG
 PPGPPGPPGPPGLGGNFAAQYDGGKGVGLGPGMGLMGPGRGPPGAAGAPGPQGFQGPAGEP
 GEPGQTGPAGARGPAGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPPTPGLPGFKGIR
 GHNGLDGLKGQPGAPGVKGEPPGAPGENTPGQTGARGLPGERGRVGPAGARGSDGSV
 GPVGPAGPIGSAGPPGFPAGPKGEIGAVGNAGPAGPAGPRGEVGLPGLSGPVGPPGNP
 GANGLTGAKGAAGLPGVAGAPGLPGRGIPGPVGAAGATGARGLVGEPGAGSKGESGNK
 GEPGSAGPQGPSPGEEGKRGPNGEAGSAGPPGPPGLRGSPPSRGLPGADGRAGVMGPP
 GSRGASGPAGVRGPNGDAGRPGEPGLMGPGRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPI
 GPAGARGEPGNIGFPGPKGPTGDPGKNGDKGHAGLAGARGAPDPGNNGAQQPPGPQGVQ
 GKGGEQGPPPGPPGFQGLPGPSGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAA
 GPTGPIGSRGPSGPPGPDGNKGEPPGVVAVGTAGPSGSPGLPGERGAAGIPGGKGEKGEPP
 GLRGEIGNPGRDARGAPGAVGAPGAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGP
 GPNGFAGPAGAAGQPAGKGERGAKGPKGENGVVGTGPVGAAGPAGPNGPPGPAGSRGDG
 GPPGMTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTEVAVGPP
 GFAGEKGPSGEAGTAGPPGTPGPQGLLAPGILGLPGRGERGLPGVAVAVGEPGPLGIA
 GPPGARGPPGAVGSPGVNGAPGEAGRDGNPNDGPPGRDQGPQGHKGERGYPGNIGPVGAA
 GAPGPHGPVGPAGKHGHRGETGPSGPVGPAGAVGPRGPSGPQGIRGDKGEPGEKGRGLP
 GLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIR
 GPQGHQGPAGPPGPPGPPGPPGPPGVSSGGYDFGYDGDYFADQPRAPSRLPKDYEVDTLTK
 SLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTG

ETCIRAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRL
ANYASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSK
TNEWGKTIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK

III. Graphique de transition de phase de Donoho-Tanner



Graphique de transition de phase extrait de Donoho, D., & Tanner, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367, 4273-4293.).

Taille d'acquisition physique (k)	Temps d'acquisition	Ration NUS minimum	Taille d'acquisition NUS (k)
0.5	8 min	1/8	4
1	15 min	1/16	32
2	30 min	1/32	64
4	1 h	1/64	256
8	2 h	1/64	512
16	8 h	1/128	2048
32	16 h	1/256	8192
64	32 h	1/512	32778

Tableau montrant pour une acquisition physique donnée, le ratio NUS utilisable avec la taille d'acquisition final en utilisant le graphique de transition de phase.

IV. Publication de recherche

Bray, F.; Bouclon, J.; Chiron, L.; Witt, M.; Delsuc, M.-A.; Rolando, C. Nonuniform Sampling Acquisition of Two-Dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Increased Mass Resolution of Tandem Mass Spectrometry Precursor Ions. *Analytical Chemistry* 2017, **89**, 8589-8593.

1 Nonuniform Sampling Acquisition of Two-Dimensional Fourier 2 Transform Ion Cyclotron Resonance Mass Spectrometry for Increased 3 Mass Resolution of Tandem Mass Spectrometry Precursor Ions

4 Fabrice Bray,[†] Julien Bouclon,^{†,‡} Lionel Chiron,[§] Matthias Witt,^{||} Marc-André Delsuc,[⊥]
5 and Christian Rolando^{*,†,ⓧ}

6 [†]Univ. Lille, CNRS, MSAP USR 3290, FR 3688 FRABIO, FR 2638 Institut Eugène-Michel Chevreul, F-59000 Lille, France

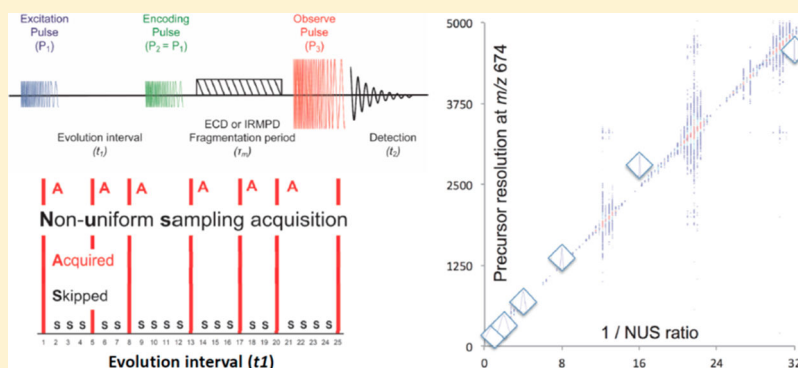
7 [‡]École Normale Supérieure, PSL Research University, Département de Chimie, 24, Rue Lhomond, F-75005 Paris, France

8 [§]CASC4DE, Le Lodge, 20, Avenue du Neuhof, F-67100 Strasbourg, France

9 ^{||}Bruker Daltonik, FTMS Applications, Fahrenheitstrasse 4, D-28359 Bremen, Germany

10 [⊥]Univ. Strasbourg, INSERM U596, CNRS UMR 7104, F-67404 Illkirch-Graffenstaden, France

11 **S** Supporting Information



12 **ABSTRACT:** Obtaining the full MS/MS map for fragments and precursors of complex mixtures without hyphenation with
13 chromatographic separation by a data-independent acquisition is a challenge in mass spectrometry which is solved by two-
14 dimensional (2D) Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS). Until now 2D FTICR MS
15 afforded only a moderate resolution for precursor ion since this resolution is limited by the number of evolution interval steps to
16 which the number of scans, the acquisition time, and the sample consumption are proportional. An overnight acquisition is
17 already required to reach a quadrupole mass filter-like unit mass resolution. Here, we report that 2D FTICR MS using
18 nonuniform sampling (NUS) obtained by randomly skipping points in the first dimension corresponding to the precursor
19 selection gives access, after data processing, to the same structural information contained in a complex mixture. The resolution
20 increases roughly as the inverse of the NUS ratio, up to 26 times at NUS 1/32, leading to an acquisition time reduced in the same
21 ratio compared to a classical acquisition at the same resolution. As an example, the analysis of a natural oil is presented.

22 **I**n mass spectrometry although mass/charge information on
23 all compounds is acquired at the same time in the MS
24 spectrum leading to high-throughput parallel acquisition,
25 structural information can only be acquired through fragmen-
26 tation of one compound after the other leading to a slow serial
27 acquisition. Recently data-independent analysis (DIA)
28 methods have been introduced which partly solve this issue
29 for a sample comprising a homogeneous class of molecules such
30 as peptides. Briefly, the full mass spectrometry (MS) spectrum
31 chopped up in tenths of a continuous m/z range, and MS/MS
32 are systematically recorded on a chromatographic time scale.
33 The mixed MS/MS spectra are identified thanks to a
34 comprehensive library search acquired of a set of data-
35 dependent acquisitions (DDA).^{1,2} Obviously, this strategy
36 cannot be applied to complex mixtures that are intractable by

chromatography such as petroleum or environmental samples. 37
2D Fourier transform techniques have revolutionized nuclear 38
magnetic resonance (NMR) since their introduction by R. R. 39
Ernst in 1974.³ They paved the way for the analysis of complex 40
samples such as purified proteins but also of complex mixtures 41
such as plasma or urine. In the same way, the principle of two- 42
dimensional 2D Fourier transform ion cyclotron resonance 43
(FTICR) MS was established in the late eighties inspired by the 44
nuclear Overhauser effect spectroscopy (NOESY) sequence.^{4–6} 45
Unfortunately 2D FTICR MS did not follow the rapid 46

Received: May 16, 2017

Accepted: August 8, 2017

Published: August 8, 2017



development of 2D NMR as its initial version suffered from three main drawbacks: (i) loss of resolution caused by poor vacuum induced by collision induced dissociation in-cell fragmentation, (ii) difficulty in data treatment at full FTICR resolution, and (iii) intense scintillation noise.⁷ During the last years we revisited 2D FTICR MS with modern instrumentation using gas-free fragmentation methods such as infrared multi-photon dissociation (IRMPD)⁸ and electron capture dissociation (ECD)⁹ and introduced optimized pulse sequence^{10,11} and advanced data processing.¹² With these developments, 2D FTICR MS has been able to provide structural information on samples of increasing complexity ranging from small molecules,¹³ peptides, and modified peptides⁹ to cytochrome c and collagen digests^{14,15} and small proteins.¹⁶

Nevertheless, until now 2D FTICR was hampered by the limited resolution for precursor ion selection on the order of several mass units. Indeed, in the first dimension (using NMR convention) which defines precursor ion resolution, a new bunch of ions has to be reinjected for each step of the delay time t_1 (Scheme S1) as the fragmentation is an irreversible process in contrast to the excitation of transverse magnetization in NMR. The delay time step is fixed to half of the low mass frequency inverse for satisfying the Nyquist criterion. The resolution is proportional to the maximum delay time t_1 which is equal to delay time step times the number of acquisition points (for a review, see ref 7). The duty time of such a scan is typically around 1 s leading to approximately an acquisition of 74 30 min for a 2k 2D, of 2 h for a 8k, and of 8 h for a 32k, using a moderate number of points (128k to 512k) in the second dimension which defines fragment ion resolution and may reach several million points. So increasing the number of data points in the first indirect dimension to the same value as for second classical dimension for reaching high resolution on precursors ion is impractical.

A powerful approach for accelerating acquisition in two-dimensional spectrometries based on Fourier transform like NMR is nonuniform sampling (NUS).¹⁷ NUS consists in skipping points in the dimension with the highest cost in terms of acquisition time, which in 2D FTICR MS is the first indirect dimension which determine precursors' resolution. In order to recover the structural information, the missing points need to be reconstructed by a data treatment. In n D NMR this is usually done using a processing based on the maximum entropy algorithm introduced by J. Skilling.^{18–20} More recently, advanced algorithms based on compressed sensing and low-rank reconstruction have been introduced.^{21–23} Unfortunately the size dependence of these algorithms precludes their use in FTICR mass spectrometry as their execution time is proportional at least to the number of points squared in the first dimension.¹⁹ Also the algorithm has to be repeated for all the points in the second dimension. Indeed, the size of 2D FTICR MS data are considerably larger than 2D NMR: a small 2D FTICR is $8k \times 128k = 1024$ Mega points affording only the resolution of a typical quadrupole mass filter in the first dimension compared to $2k \times 2k = 4$ Mega for a typical 2D NMR. The size of a square 2D FTICR measurement is much higher starting at $128k \times 128k = 16$ Giga points up to $8M \times 8M = 64$ Tera points. We described previously the urQRd denoising algorithm which is based on the QR decomposition of a matrix randomly sampled from the data.¹² The urQRd denoising algorithm avoids the singular value decomposition (SVD) of a matrix which scales in size as the cube of the data length, preventing its use on very large data sets. This approach

allows a gain of nearly 3 orders of magnitude in processing time compared to the classical SVD denoising and reduces strongly the computer memory footprint and allows the denoising algorithm to be applied to virtually unlimited data size. The urQRd algorithm has only two parameters, the reconstruction rank which is the number of peaks to detect and the number of iteration. Contrary to the classical Cadzov algorithm,²⁴ urQRd is not sensitive to the rank and to an overestimated number of iteration. The urQRd algorithm can be used to reconstruct missing values, by considering them as values corrupted by some kind of noise. The acquired NUS data are inserted into a full length signal in which missing points are initially set to zero. An adapted version of the urQRd algorithm is applied to the full length signal, after which experimental data points are reinjected. The combined steps are then iterated. The adapted urQRd algorithm has been optimized for weak signals by using the original signal for the projective space determination. FFT is then applied on the resulting reconstruction. Data are processed first along F2 by FFT with zerofilling of the data size apodization and, then along F1 by urQRd with reinjection as described using a hypercomplex data representation, and displayed after zerofilling of the data size as modulus.^{25,26}

In 2D NMR the best NUS sampling distribution is usually a exponentially biased sampling protocol, with a sampling density proportional to the inverse of the relaxation rate.²⁷ The situation is quite different for 2D FTICR MS as a new bunch of ions has to be introduced for each point in the indirect dimension and as using a harmonized cell the signal in the direct dimension is almost not damped. In those conditions, the simplest NUS sampling distribution is the uniform distribution. The ratio of NUS to physical acquisition size afford the average gap. Gaps between zero and 2 average gaps are randomly generated which generates smaller gaps than a Poisson scheme.²⁸

For determining the scope and limitation of the 2D FTICR acquisition, we used a simple peptide substance P ionized by nanoESI at an analytical concentration of $1 \text{ pmol } \mu\text{L}^{-1}$. The selected doubly charged ion was fragmented by ECD which has been comprehensively studied.^{9,29–31} Experiments were performed on a SolariX XR 12 T FTICR MS fitted with a harmonized ParaCell³² from Bruker Daltonik GmbH (Bremen, Germany). We recorded 2D FTICR mass spectra with the improved pulse sequence designed previously and described in Scheme S1.¹⁰ The final observed pulse used the default analytical settings for ion excitation. The two initial encoding pulses before ECD are identical with the shorter duration available and a lower voltage corresponding to approximately an ion radius two times bigger than the dissociating beam. Data treatments were done on a Dell R930 cluster fitted with 4 Xeon processors encompassing 48 cores with the open source Spike suite³³ written in Python and which is parallelized allowing fast processing. All the acquisition were performed with a physical size of $2k \times 128k$ points corresponding to a 30 min acquisition, starting from a classical acquisition to NUS acquisitions of size 4k, 8k, 16k, 32k, and 64k corresponding to NUS ratios 1/2, 1/4, 1/8, 1/16, and 1/32, respectively. In 2D spectra displayed in Figure S1, clear fragmentation spectra are visible at all NUS ratios with scintillation noise slightly increasing when the NUS ratio is decreasing.

To get a more precise view, Figure S2 displays the fragment spectrum of the monoisotopic peak normalized on the c5 fragment at all NUS ratios. The aspect of the fragment spectrum is well preserved. The change in peak intensities is

Table 1. Substance P Doubly Charged Monoisotopic Peak 2D FTICR Spectrum: Quantitative Data

size (k)	NUS ratio	processing time ^a (hour)	F1 resolution ^b			identified fragments ^c	fragments/precursor %
			Prec.	Frag. 1	Frag. 2		
2		1 h 18 m	176	180	178	13	65
4	1/2	1 h 34 m	325	348	335	20	65
8	1/4	2 h 14 m	687	626	626	24	75
16	1/8	3 h 42 m	1365	1258	1326	24	66
32	1/16	6 h 39 m	2795	2799	2830	30	63
64	1/32	12 h 17 m	4576	4528	5119	17	23

^aProcessing time with the following parameters: urQRd rank 49 and iteration 1, reinjection 2401 times on column containing signal; urQRd rank 7 and iteration 1, reinjection 7 times elsewhere. ^bPrecursor m/z 674.371; fragment 1, m/z 496.335; fragment 2, m/z 1347.735. ^cThe full list of identified fragments is given in Tables S1–S6.

caused by progressive resolution of the monoisotopic peak when the acquisition size varies from 2k to 64k. Table 1 describes the number of assigned fragments which increases from 13 for the plain 2k acquisition to 30 for the NUS 1/16 32k acquisition. The evolution of the precursor peak fwhm (Figure 1) shows that as expected the resolution increases linearly as the inverse of the NUS ratio until NUS 1/16 (Figure S3) and is slightly below at NUS 1/

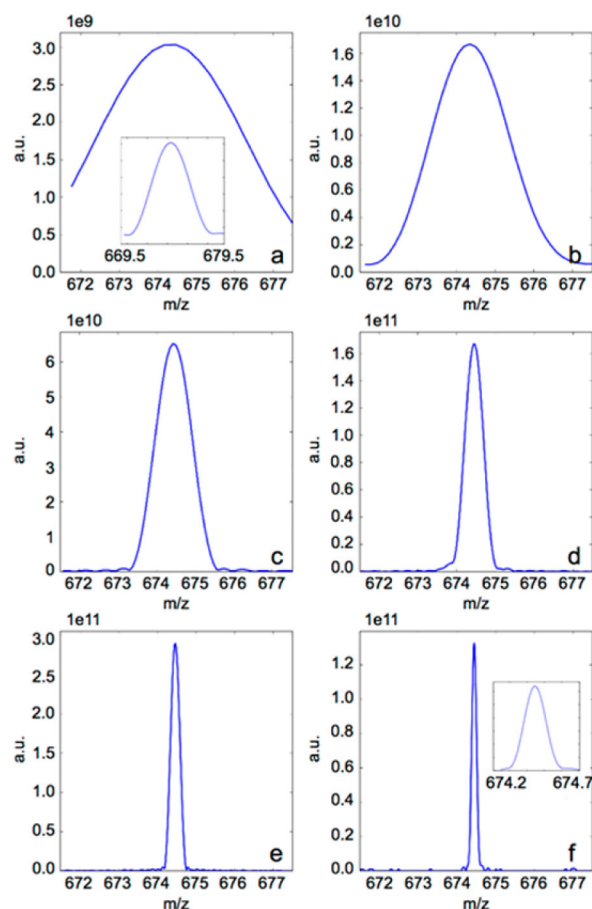


Figure 1. Substance P doubly charged monoisotopic peak 2D FTICR mass spectrum. Zoom along the F1 (precursor) dimension (from a, 2k classical acquisition; b, 4k, NUS 1/2; c, 8k, NUS 1/4; d, 16k, NUS 1/8; e, 32k, NUS 1/16; f, 64k, NUS 1/32). All peaks are defined by approximately 24 points which is constant as resolution increases proportionally to the acquisition size.

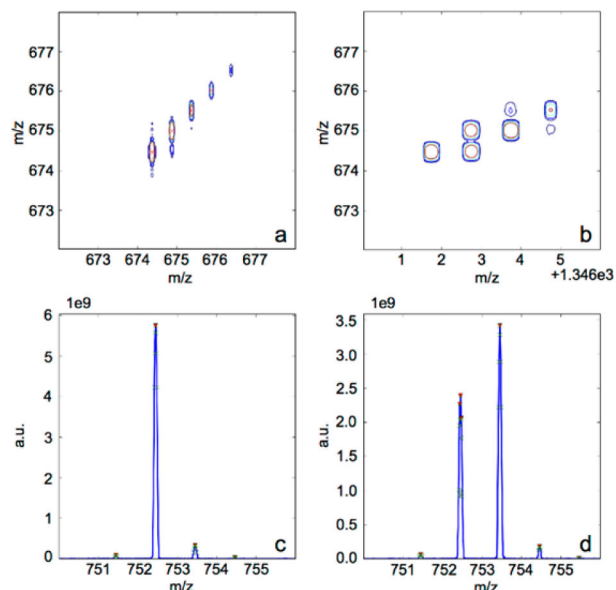


Figure 2. Substance P doubly charged monoisotopic peak 2D FTICR spectrum 32 k acquisition, NUS 1/16: (a) zoom on the doubly charged precursor, (b) zoom on the reduced monocharged species, (c) zoom on the c6 fragment corresponding to the monoisotopic, (d) zoom on the first isotope (theoretical ratio m/z , $(m + 1)/z$: 48, 100).

noteworthy that even at NUS ratio 1/32, which means skipping 31 points out of 32, the peak shape is preserved. In 2D NMR not so small NUS factor are generally used. As expected, the fragments at low or high mass have the same resolution as the precursor as shown in Table 1.

Starting at size 32k (NUS 1/16), the isotope precursors are fully separated and the fragment spectrum of each isotope is extracted from the 2D spectrum. In Figure 2a, we can see in the zoom of the 2D spectrum on the parent peak that the fourth isotope can be detected at m/z 676.37. As its intensity is only 0.84%, the dynamic range is higher than 100. Figure 2b displays the 2D spectrum zoom of the reduced monocharged species. Two chemical processes are visible: the capture of one electron and the capture of one electron followed by hydrogen atom loss. If isotopes are not separated, data interpretation is blurred whereas it is obvious here.³¹ Finally in Figure 2c,d the zoom of the c6 fragment corresponding to the monoisotopic and first isotope are displayed. As expected the monoisotopic peak affords only one peak (the small peak at $M + 1$ is due to hydrogen capture discussed above) whereas the first isotope gives two peaks in an intensity ratio close to the predicted by the number of carbon atoms remaining in the fragment (theoretical m/z , $(m + 1)/z$; intensities, 48, 100; observed, 70, 205

16 reaches the value of 2 785 at m/z 674, which corresponds to a normalized resolution of around 10 000 at m/z 200. It is

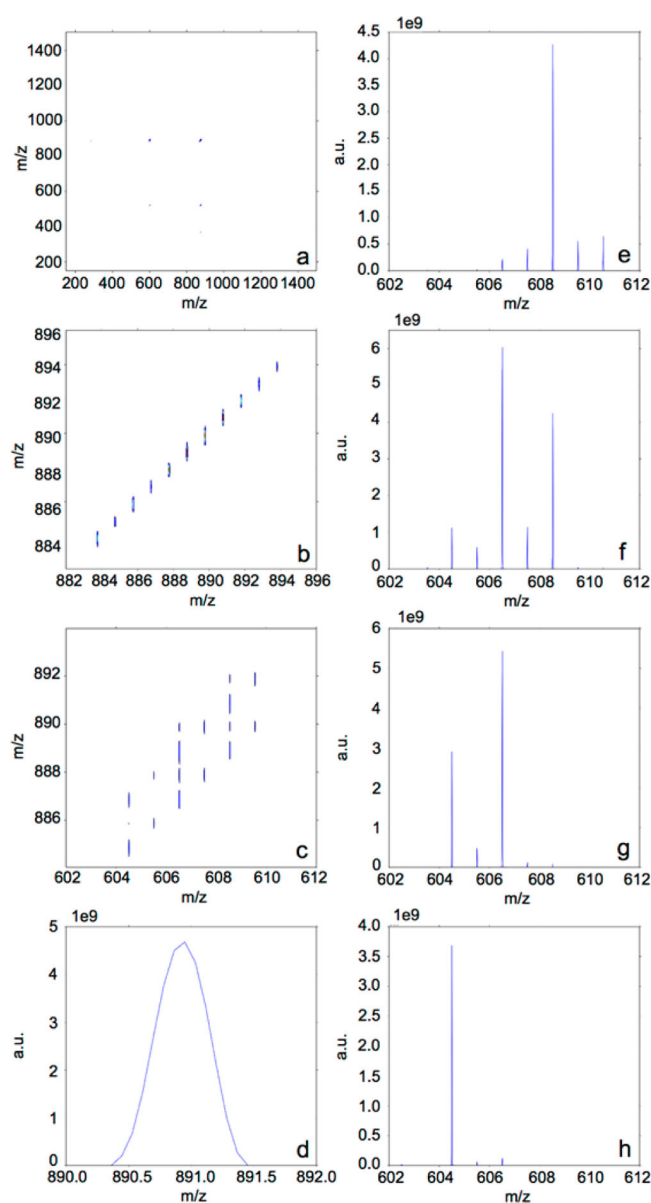


Figure 3. 2D NUS FTICR spectra of sesame oil cationized by Li^+ and fragmented by IRMPD ($64\text{k} \times 512\text{k}$ at NUS 1/8): (a) full 2D spectrum (a larger view is presented in Figure S4), (b) zoom on the precursors, (c) zoom on the fatty acid neutral loss, (d) zoom on the m/z 890.8 (OOO) profile along the F1 (precursor) dimension. Zoom on profile along the F2 (fragment) dimension corresponding to the loss of a fatty acid from (e) m/z 890.8 (OOO), (f) m/z 888.8 (OLO or OOL) profile, (g) m/z 886.8 (LOL or OLL), and (h) m/z 884.8 (LLL).

the physical acquisition size and to use a higher NUS ratio as shown below.

Finally, as an example, we present the 2D spectra ($64\text{k} \times 512\text{k}$ at NUS 1/8) of sesame oil triglycerides cationized by lithium cation and fragmented by IRMPD which affords as main fragment the loss of a fatty acid.³⁷ Sesame oil experiment was performed on an Apex 9.4 T FTICR MS fitted with an Infinity cell from Bruker Daltonik GmbH (Bremen, Germany) showing that 2D NUS acquisition may be performed in any ICR cell. For comparison the MS/MS obtained by isolation of the same precursors with the front quadrupole analyzer are given in Figure S5a–e along with a table of the expected m/z (Table S11). On the full 2D spectrum, the neutral loss line is clearly apparent (Figure 3a) and when zoomed the loss of linolenic acid and oleic acid, which differs only by 2 hydrogens, are easily distinguished (Figure 3b). Figure 3c displays a zoom on the precursor area and Figure 3d the profile of a precursor along the F1 precursor dimension. Its fwhm is 0.44 Da which corresponds to a mass resolution of about 2000 at m/z 890 and so to a normalized resolution of 9000 at m/z 200. Starting from precursor at m/z 890.8, which is the symmetrical triolein (OOO), a fragment spectrum containing only one peak is observed (Figure 3e). The smaller peaks are due to overlapping isotope contributions. When moving 2 Da lower, the triglyceride has the composition OOL or OLO and two peaks corresponding to the loss of O and L are observed (Figure 3f). The same shift is observed 4 Da lower from triolein for OLL or LOL (Figure 3g). Finally a fragment spectrum containing only one peak is observed for the precursor at m/z 884.8, 6 Da lower from triolein corresponding again to a symmetrical TAG (LLL) (Figure 3h).

In conclusion we successively implemented for the first time a 2D FTICR MS acquisition with nonuniform sampling using under-sampling ratio as low as 1/32. The higher sparsity of the FTICR data and the larger acquisition size compared to NMR accounts for this increase in the useful NUS ratio range before reaching the Donoho-Tanner phase transition.³⁸ NUS acquisition paves the way to high-resolution 2D FTICR MS for both precursor and fragments and open a new area in data independent analysis (DIA).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b01850.

Experimental details and detailed results; 2D FTICR MS pulse sequence; full 2D spectrum at different NUS ratios; fragment spectrum of the monoisotopic peak, normalized on the c_5 fragment, for different NUS ratios; fwhm of the precursor peak as a function of the inverse of the NUS ratio; 2D FTICR spectra of sesame oil cationized by $^6\text{Li}^+$ and fragmented by IRMPD ($64\text{k} \times 512\text{k}$ at NUS 1/8); classical MS/MS FTICR spectra of sesame oil cationized by Li^+ and fragmented by IRMPD; interpretation of fragment spectrum of the monoisotopic peak for NUS ratios; observed isotopic distribution of 10 major fragments from the mono, 2nd, 3rd and 4th isotopic peak for compared to the theoretical distribution calculated with the IsoPatrn software; and theoretical fragmentation of the sesame oil major triglycerides cationized by ^6Li and fragmented by IRMPD (PDF)

100).^{34,35} In Tables S7–S10, the observed values for all fragments of the monoisotopic peak and the first three isotopes are given and compared to their theoretical values calculated with the IsoPatrn software.³⁶ Tables S7–S10 show that two contrary phenomena are taking place as the NUS ratio decreases: (i) the resolution increase brings a better separation of the isotopes required for obtaining correct isotopic ratios and (ii) the quality of the reconstruction especially for peaks with a low intensity decreases. For 2k physical spectra, the best trade-off is obtained for NUS 1/16. A simple way to improve the isotopic ratio accuracy is for a given acquisition size to increase

278 ■ AUTHOR INFORMATION

279 Corresponding Author

280 *E-mail: Christian.Rolando@univ-lille1.fr.281 ORCID 

282 Christian Rolando: 0000-0002-3266-8860

283 Notes

284 The authors declare no competing financial interest.

285 ■ ACKNOWLEDGMENTS

286 The authors thank Professor Geoffrey Bodenhausen for seminal
287 discussions on NMR and FTICR. The authors thank the
288 Agence Nationale de la Recherche (ANR, France; call, Défi de
289 tous le savoirs 2014; Grant, ONE_SHOT_FT-ICR_MS_2D),
290 the Défi CNRS MASTODONS (Grant, TABASCO), and the
291 French FT-ICR MS network (FR CNRS 3624, Réseau National
292 de Spectrométrie de Masse FT-ICR à très haut champ) for
293 their financial support.

294 ■ REFERENCES

- 295 (1) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter,
296 L.; Bonner, R.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11*,
297 O111.016717.
- 298 (2) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.;
299 MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.;
300 Kellmann, M.; Zabrouskov, V. *Nat. Methods* **2013**, *10*, 744–746.
- 301 (3) Aue, W.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*,
302 2229–2246.
- 303 (4) Pfändler, P.; Bodenhausen, G.; Rapin, J.; Houriet, R.; Gäumann,
304 T. *Chem. Phys. Lett.* **1987**, *138*, 195–200.
- 305 (5) Pfaendler, P.; Bodenhausen, G.; Rapin, J.; Walser, M. E.;
306 Gäumann, T. *J. Am. Chem. Soc.* **1988**, *110*, 5625–5628.
- 307 (6) Bensimon, M.; Zhao, G.; Gäumann, T. *Chem. Phys. Lett.* **1989**,
308 *157*, 97–100.
- 309 (7) van Agthoven, M. A.; Delsuc, M.-A.; Bodenhausen, G.; Rolando,
310 C. *Anal. Bioanal. Chem.* **2013**, *405*, 51–61.
- 311 (8) van Agthoven, M. A.; Delsuc, M.-A.; Rolando, C. *Int. J. Mass*
312 *Spectrom.* **2011**, *306*, 196–203.
- 313 (9) van Agthoven, M. A.; Chiron, L.; Coutouly, M.-A.; Delsuc, M.-A.;
314 Rolando, C. *Anal. Chem.* **2012**, *84*, 5589–5595.
- 315 (10) van Agthoven, M. A.; Chiron, L.; Coutouly, M.-A.; Sehgal, A. A.;
316 Pelupessy, P.; Delsuc, M.-A.; Rolando, C. *Int. J. Mass Spectrom.* **2014**,
317 *370*, 114–124.
- 318 (11) Sehgal, A. A.; Pelupessy, P.; Rolando, C.; Bodenhausen, G. *Phys.*
319 *Chem. Chem. Phys.* **2016**, *18*, 9167–9175.
- 320 (12) Chiron, L.; van Agthoven, M. A.; Kieffer, B.; Rolando, C.;
321 Delsuc, M.-A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1385–1390.
- 322 (13) van Agthoven, M. A.; Barrow, M. P.; Chiron, L.; Coutouly, M.-
323 A.; Kilgour, D.; Wootton, C. A.; Wei, J.; Soulby, A.; Delsuc, M.-A.;
324 Rolando, C. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 2105–2114.
- 325 (14) van Agthoven, M. A.; Wootton, C. A.; Chiron, L.; Coutouly, M.-
326 A.; Soulby, A.; Wei, J.; Barrow, M. P.; Delsuc, M.-A.; Rolando, C.;
327 O'Connor, P. B. *Anal. Chem.* **2016**, *88*, 4409–4417.
- 328 (15) Simon, H. J.; van Agthoven, M. A.; Lam, P. Y.; Floris, F.;
329 Chiron, L.; Delsuc, M. A.; Rolando, C.; Barrow, M. P.; O'Connor, P.
330 B. *Analyst* **2016**, *141*, 157–165.
- 331 (16) Floris, F.; van Agthoven, M.; Chiron, L.; Soulby, A. J.; Wootton,
332 C. A.; Lam, Y. P.; Barrow, M. P.; Delsuc, M.-A.; O'Connor, P. B. *J. Am.*
333 *Soc. Mass Spectrom.* **2016**, *27*, 1531–1538.
- 334 (17) Marvasti, F. *Nonuniform sampling: theory and practice*; Springer
335 Science & Business Media: New York, 2012.
- 336 (18) Barna, J. C.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J.
337 *Biochem. Soc. Trans.* **1986**, *14*, 1262–1263.
- 338 (19) Hyberts, S. G.; Arthanari, H.; Wagner, G. In *Novel Sampling*
339 *Approaches in Higher Dimensional NMR*; Billeter, M., Orekhov, V.,
340 Eds.; Springer, Berlin Heidelberg, 2012; p 125–148.

- (20) Hoch, J. C.; Maciejewski, M. W.; Mobli, M.; Schuyler, A. D.;
Stern, A. S. *Acc. Chem. Res.* **2014**, *47*, 708–717. 342
- (21) Kazimierczuk, K.; Orekhov, V. Y. *Angew. Chem., Int. Ed.* **2011**,
343 *50*, 5556–5559. 344
- (22) Stern, A. S.; Hoch, J. C. *Magn. Reson. Chem.* **2015**, *53*, 908–912. 345
- (23) Qu, X.; Mayzel, M.; Cai, J. F.; Chen, Z.; Orekhov, V. *Angew.*
346 *Chem., Int. Ed.* **2015**, *54*, 852–854. 347
- (24) Cadzow, J.; Wu, M.-M. *IEE Proc. F* **1987**, *134*, 69–78. 348
- (25) Delsuc, M.-A. *J. Magn. Reson.* **1988**, *77*, 119–124. 349
- (26) Schuyler, A. D.; Maciejewski, M. W.; Stern, A. S.; Hoch, J. C. *J.*
350 *Magn. Reson.* **2013**, *227*, 20–24. 351
- (27) Rovnyak, D.; Hoch, J.; Stern, A.; Wagner, G. *J. Biomol. NMR*
352 **2004**, *30*, 1–10. 353
- (28) Hyberts, S. G.; Takeuchi, K.; Wagner, G. *J. Am. Chem. Soc.* **2010**,
354 *132*, 2145–2147. 355
- (29) Axelsson, J.; Palmblad, M.; Håkansson, K.; Håkansson, P. *Rapid*
356 *Commun. Mass Spectrom.* **1999**, *13*, 474–477. 357
- (30) Belyayev, M. A.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B. *J. Am.*
358 *Soc. Mass Spectrom.* **2006**, *17*, 1429–1437. 359
- (31) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.;
360 Coon, J. *J. Anal. Chem.* **2007**, *79*, 3525–3534. 361
- (32) Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G. *J. Am. Soc.*
362 *Mass Spectrom.* **2011**, *22*, 1125–1133. 363
- (33) Delsuc, M.-A. *SPIKE*. <http://www.bitbucket.org/delsuc/spike>
364 (accessed July 4 2017). 365
- (34) Singleton, K. E.; Cooks, R. G.; Wood, K. V. *Anal. Chem.* **1983**,
366 *55*, 762–764. 367
- (35) Ramaley, L.; Herrera, L. C. *Rapid Commun. Mass Spectrom.* **2008**,
368 *22*, 2707–2714. 369
- (36) Dalhousie University, Trace Analysis Research Centre. [http://](http://tarc.chemistry.dal.ca/soft_down.htm)
370 tarc.chemistry.dal.ca/soft_down.htm (accessed July 4 2017). 371
- (37) Garnier, N.; Rolando, C.; Hotje, J. M.; Tokarski, C. *Int. J. Mass*
372 *Spectrom.* **2009**, *284*, 47–56. 373
- (38) Donoho, D.; Tanner, J. *Philos. Trans. R. Soc., A* **2009**, *367*,
374 4273–4293. 375

Supporting Information

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS for increased mass resolution of MS/MS precursor ions

Author names

Fabrice Bray,[†] Julien Bouclon,^{†,‡} Lionel Chiron,^{||} Matthias Witt,[§] Marc-André Delsuc[#] and Christian Rolando^{*,†}

1

Author affiliations

[†]Univ. Lille, CNRS, MSAP USR 3290, FR 3688 FRABIO, FR 2638 Institut Eugène-Michel Chevreul, F-59000 Lille, France

[‡]École Normale Supérieure, PSL Research University, Département de Chimie, 24, rue Lhomond, F-75005 Paris, France

^{||}CASC4DE, Le Lodge, 20, Avenue du Neuhof, F-67100 Strasbourg, France

[§]Bruker Daltonik, FTMS Applications, Fahrenheitstrasse 4, D-28359 Bremen, Germany

[#]Univ. Strasbourg, INSERM U596, CNRS UMR 7104, F-67404 Illkirch-Graffenstaden, France

Corresponding author

Email: christian.rolando@univ-lille1.fr

Table of contents

Supporting Information Manuscript. This manuscript gives detailed results of experiments on substance P.

Scheme S1. 2D FT-ICR MS optimized pulse sequence.

Figure S1. Full 2D spectrum at different NUS ratios. a) NUS 1, b) NUS 1/2, c) NUS 1/4, d) NUS 1/8, e) NUS 1/16, f) NUS 1/32.

Figure S2. Fragment spectrum of the mono-isotopic peak normalized for the c_5 fragment for different NUS ratios. a) NUS 1, b) NUS 1/2, c) NUS 1/4, d) NUS 1/8, e) NUS 1/16, f) NUS 1/32.

Figure S3. FWHM of the precursor peak as a function of the inverse of the NUS ratio.

Figure S4. 2D FT-ICR spectra of sesame oil cationized by $^6\text{Li}^+$ and fragmented by IRMPD.

Figure S5. Classical MS/MS FT-ICR spectra of sesame oil cationized by $^6\text{Li}^+$ and fragmented by IRMPD a) unselected precursors and with front quadrupole selection of ions at m/z b) 890.8, c) 888.8 d) 886.8 and e) 884.8.

Table S1. Interpretation of fragment spectrum of the mono-isotopic peak for a classical not-NUS (ie NUS 1) acquisition.

Table S2. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/2 acquisition.

Table S3. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/4 acquisition.

Table S4. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/8 acquisition.

Table S5. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/16 acquisition.

Table S6. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/32 acquisition.

Table S7. Observed isotopic distribution of 10 major fragments from the mono-isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Table S8. Observed isotopic distribution of 10 major fragments from the 2nd isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

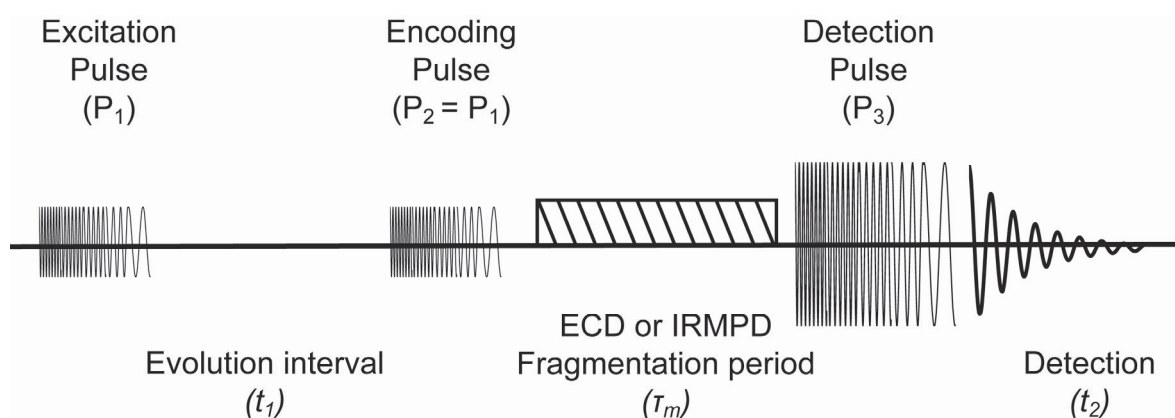
Table S9. Observed isotopic distribution of 10 major fragments from the 3rd isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Table S10. Observed isotopic distribution of 10 major fragments produced from the 4th isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Table S11. Theoretical fragmentation of the sesame oil triglycerides cationized by ⁶Li⁺ and fragmented by IRMPD.

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Scheme S1. 2D FT-ICR MS pulse sequence.



The excitation and encoding pulses P_1 and P_2 are identical and have an overall duration of 723 μs each with 0.5 μs per frequency step. The P_1 pulse amplitude was tuned to decrease the fragment intensity to about half of the maximum. The detection pulse P_3 used the default settings for ion excitation in Apex/SolariX Control (100 V_{pp} amplitude and 20 μs per frequency step).

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

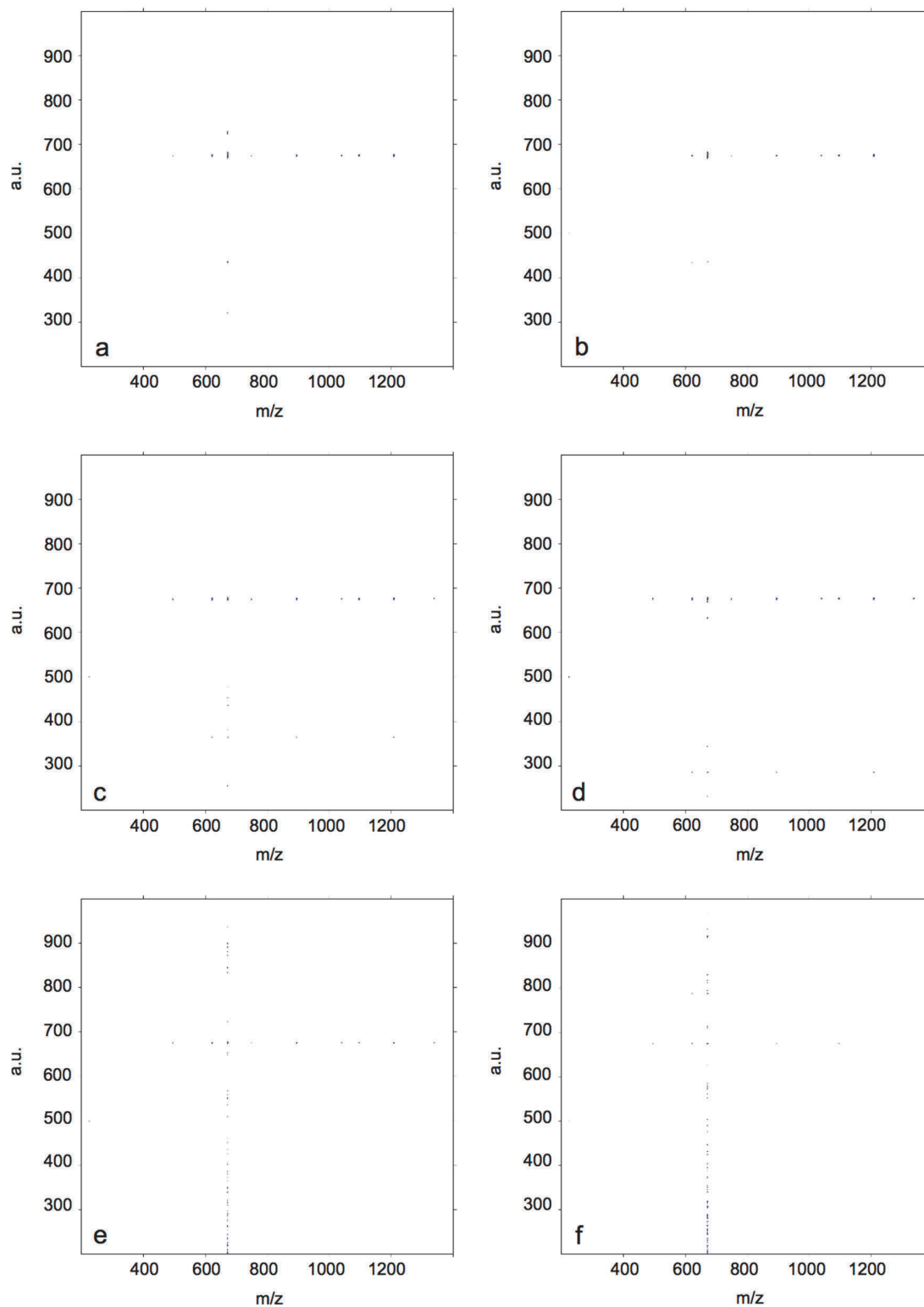


Figure S1. Full 2D spectrum at different NUS ratios. a) NUS 1, b) NUS 1/2, c) NUS 1/4, d) NUS 1/8, e) NUS 1/16, f) NUS 1/32.

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

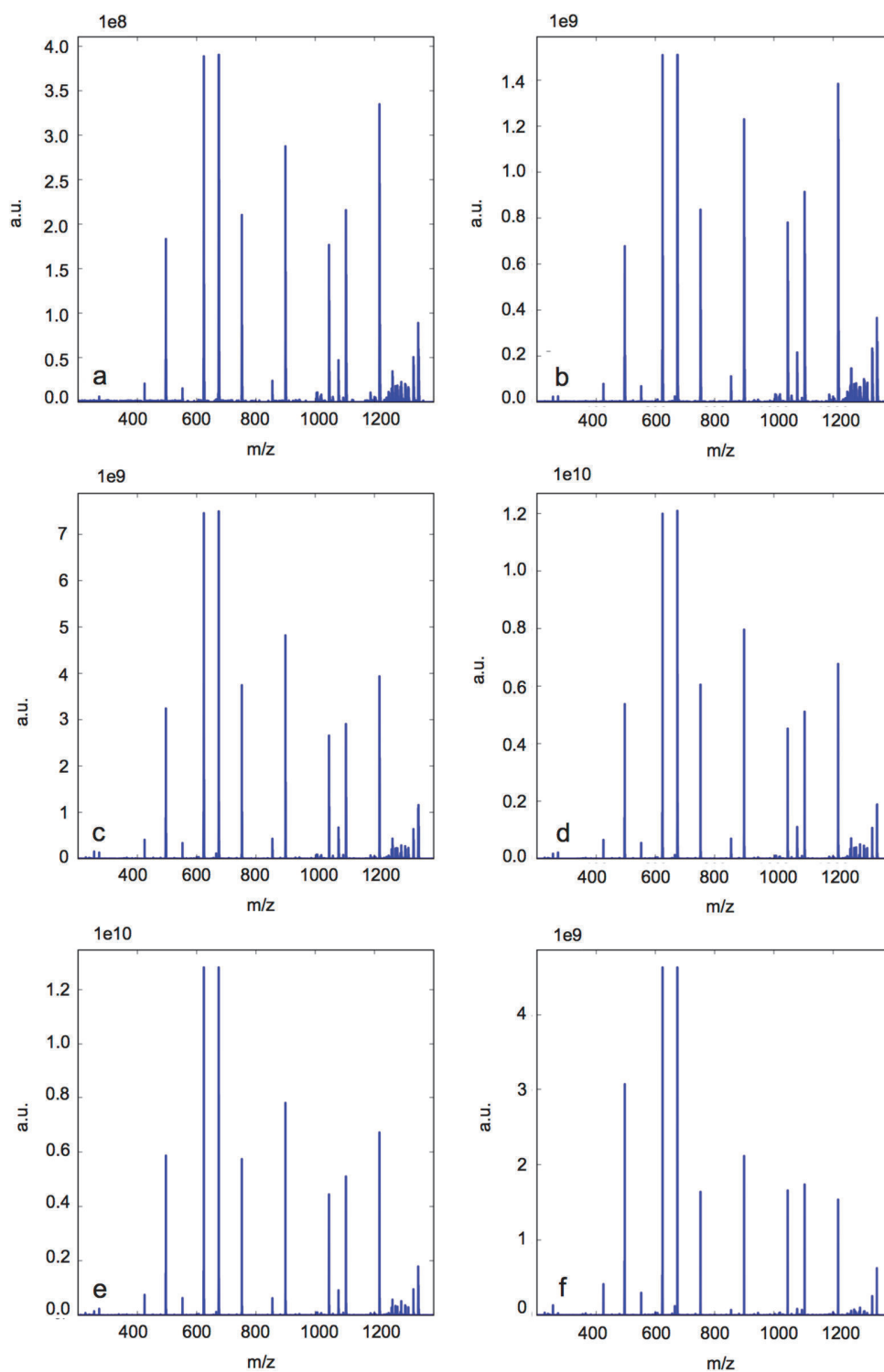


Figure S2. Fragment spectrum of the normalized mono-isotopic peak of the c5 fragment at different NUS ratios. a) NUS 1, b) NUS 1/2, c) NUS 1/4, d) NUS 1/8, e) NUS 1/16, f) NUS 1/32.

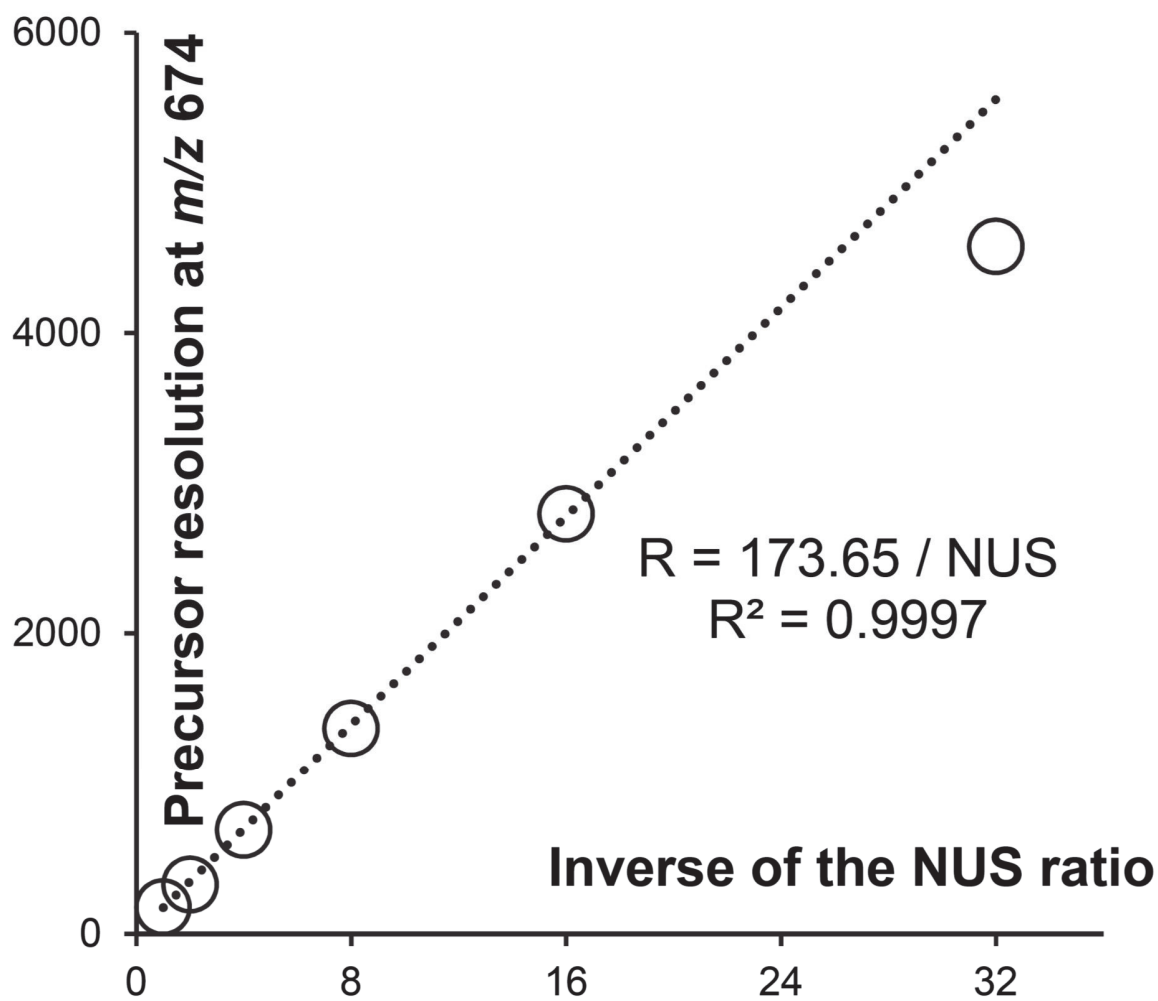


Figure S3. FWHM of the precursor peak as a function of the inverse of the NUS ratio.

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

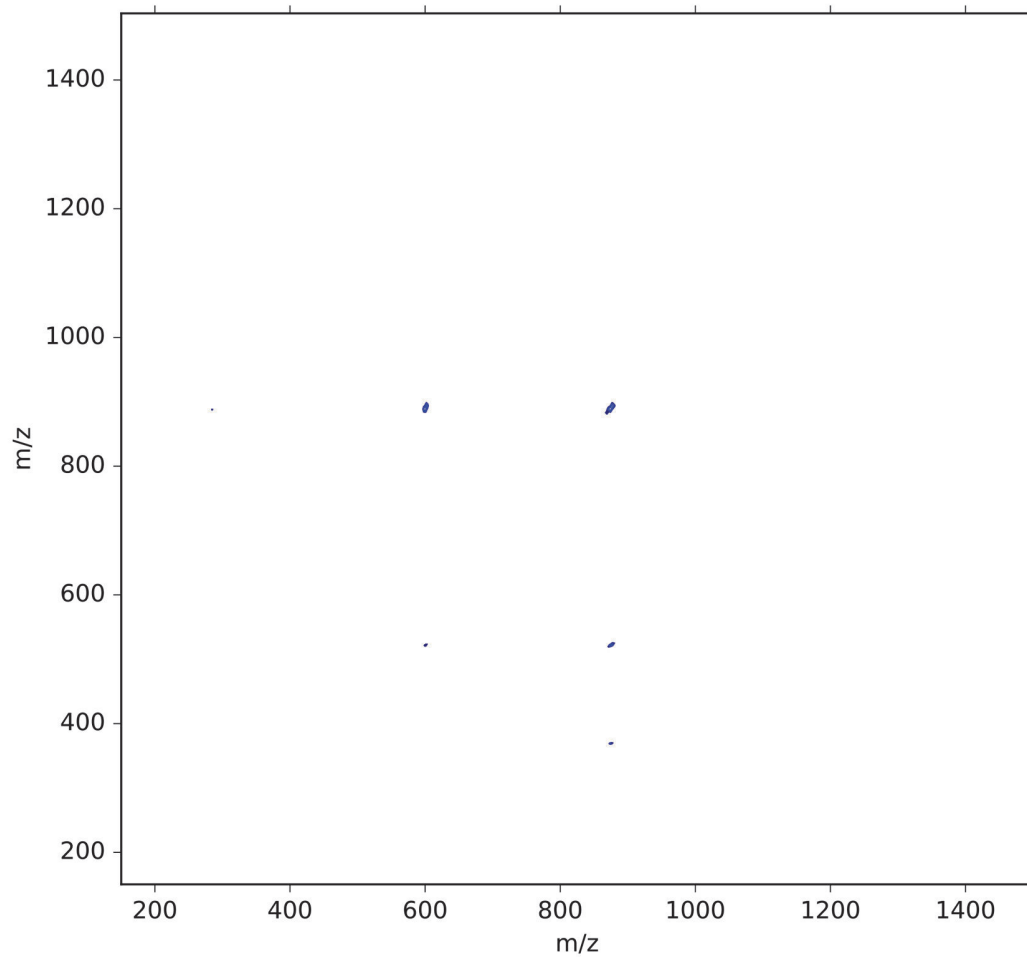


Figure S4. 2D FT-ICR spectra of sesame oil cationized by Li^+ and fragmented by IRMPD.

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

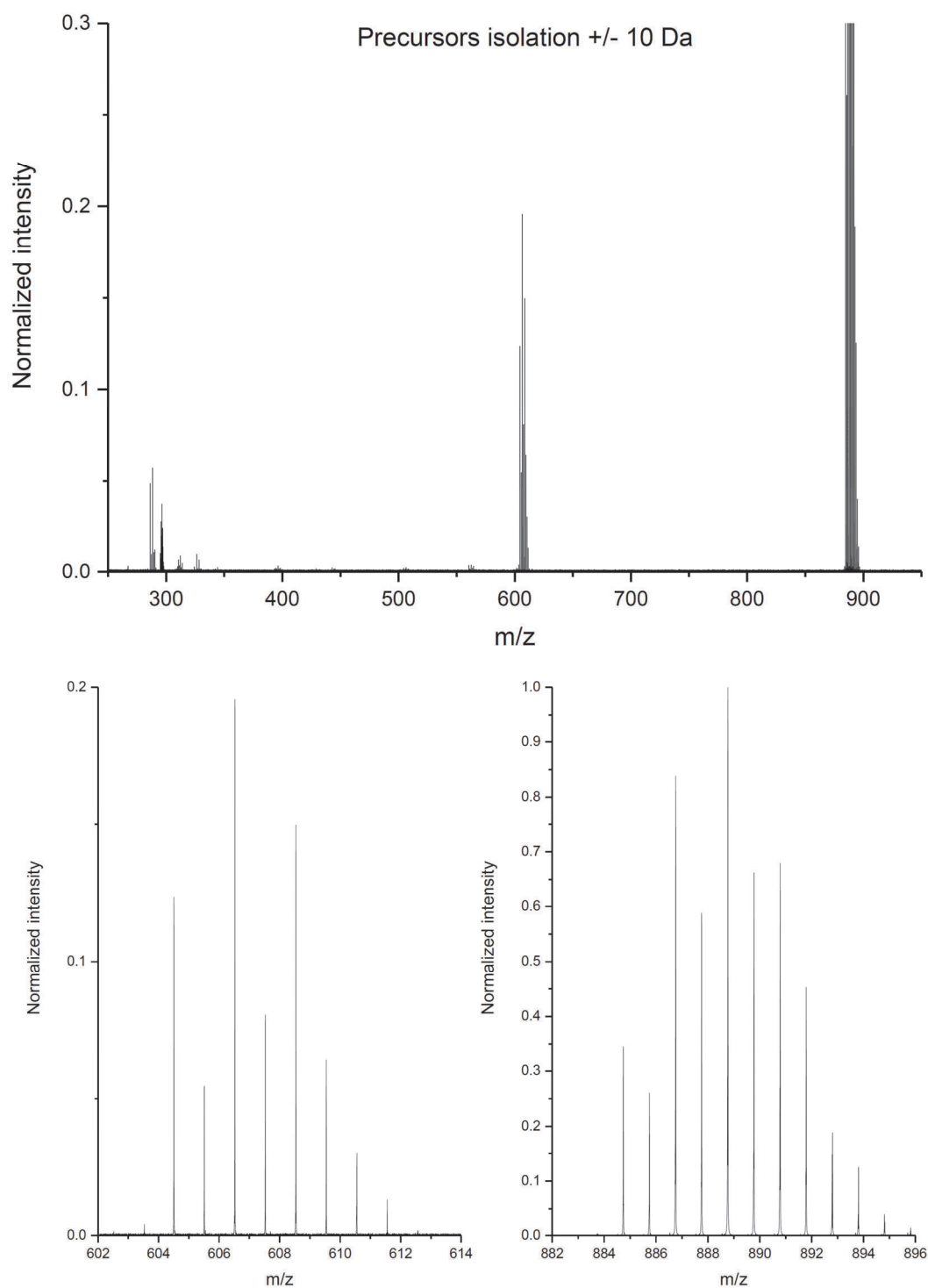


Figure S5a. Classical MS/MS FT-ICR spectra of sesame oil cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD: upper panel full MS spectrum, right lower panel zoom on the precursor, left lower panel, zoom on fragments corresponding to the loss of a neutral fatty acid a) unselected precursors.

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

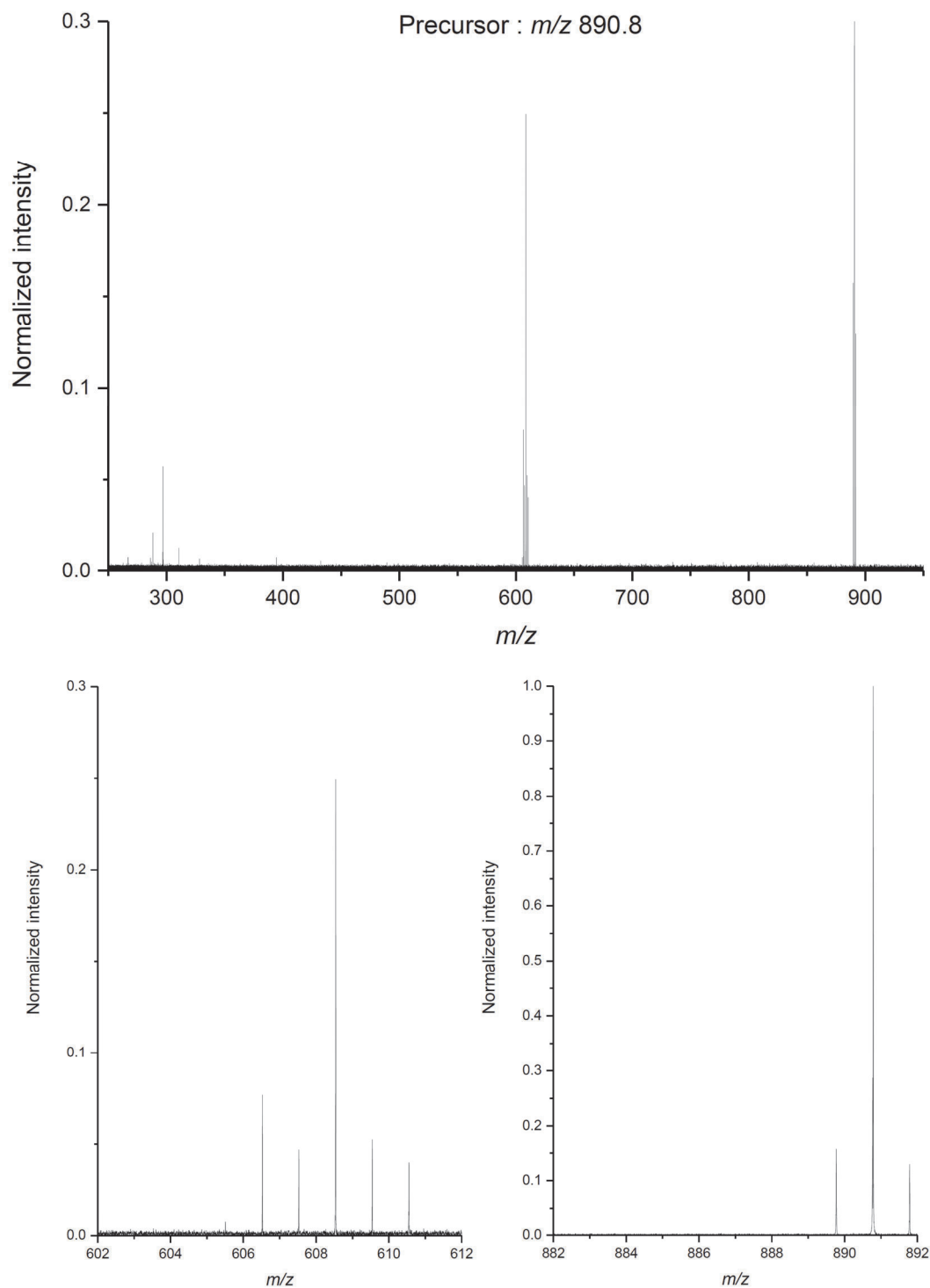


Figure S5b. Classical MS/MS FT-ICR spectra of sesame oil cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD: upper panel full MS spectrum, right lower panel zoom on the precursor, left lower panel, zoom on fragments corresponding to the loss of a neutral fatty acid b) precursor m/z 890.8 (quadrupole selection 0.5 m/z).

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

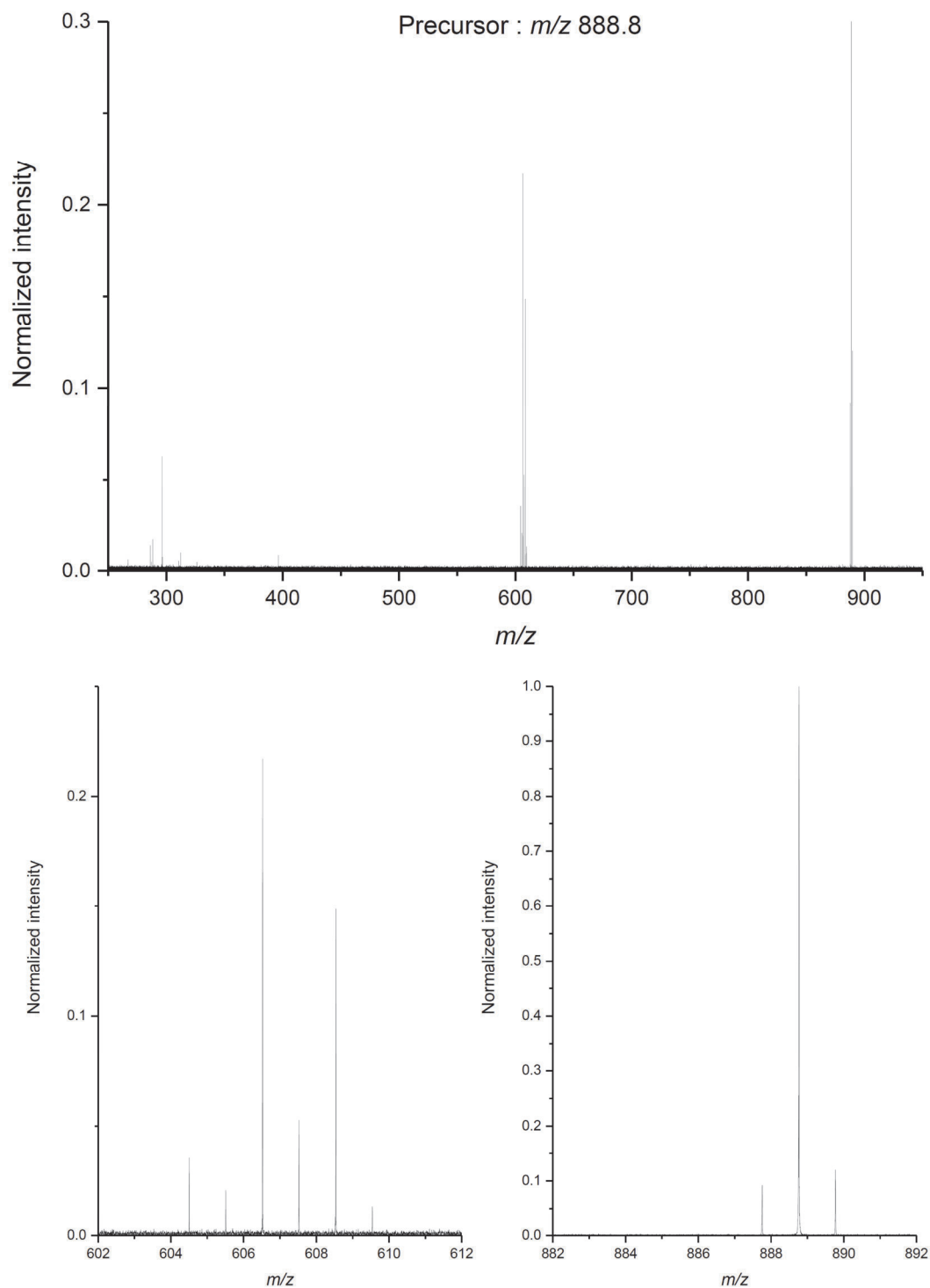


Figure S5c. Classical MS/MS FT-ICR spectra of sesame oil cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD: upper panel full MS spectrum, right lower panel zoom on the precursor, left lower panel, zoom on fragments corresponding to the loss of a neutral fatty acid b) precursor m/z 888.8 (quadrupole selection $\pm 0.5 m/z$).

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

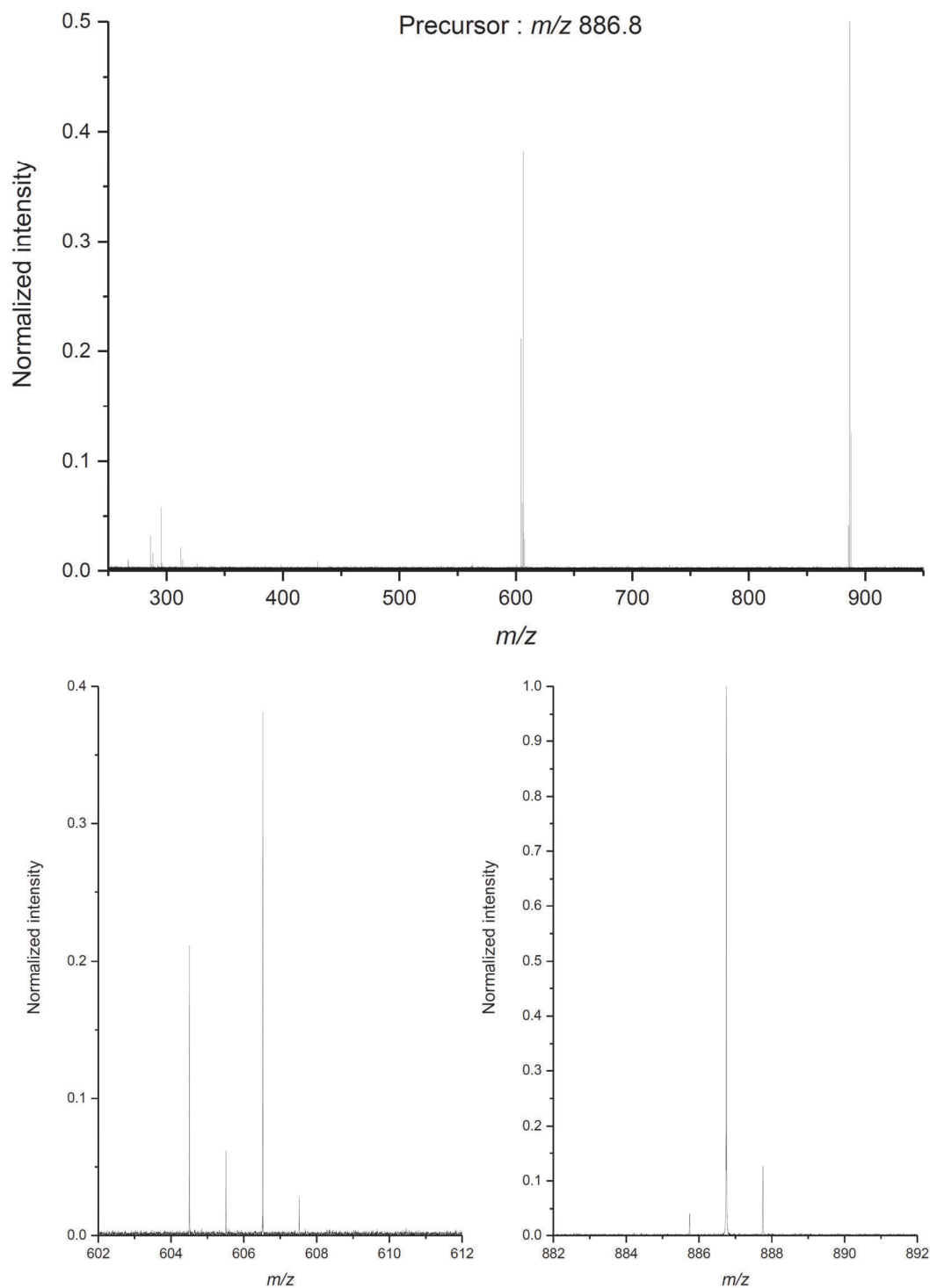


Figure S5d. Classical MS/MS FT-ICR spectra of sesame oil cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD: upper panel full MS spectrum, right lower panel zoom on the precursor, left lower panel, zoom on fragments corresponding to the loss of a neutral fatty acid b) precursor m/z 886.8 (quadrupole selection $\pm 0.5 m/z$).

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

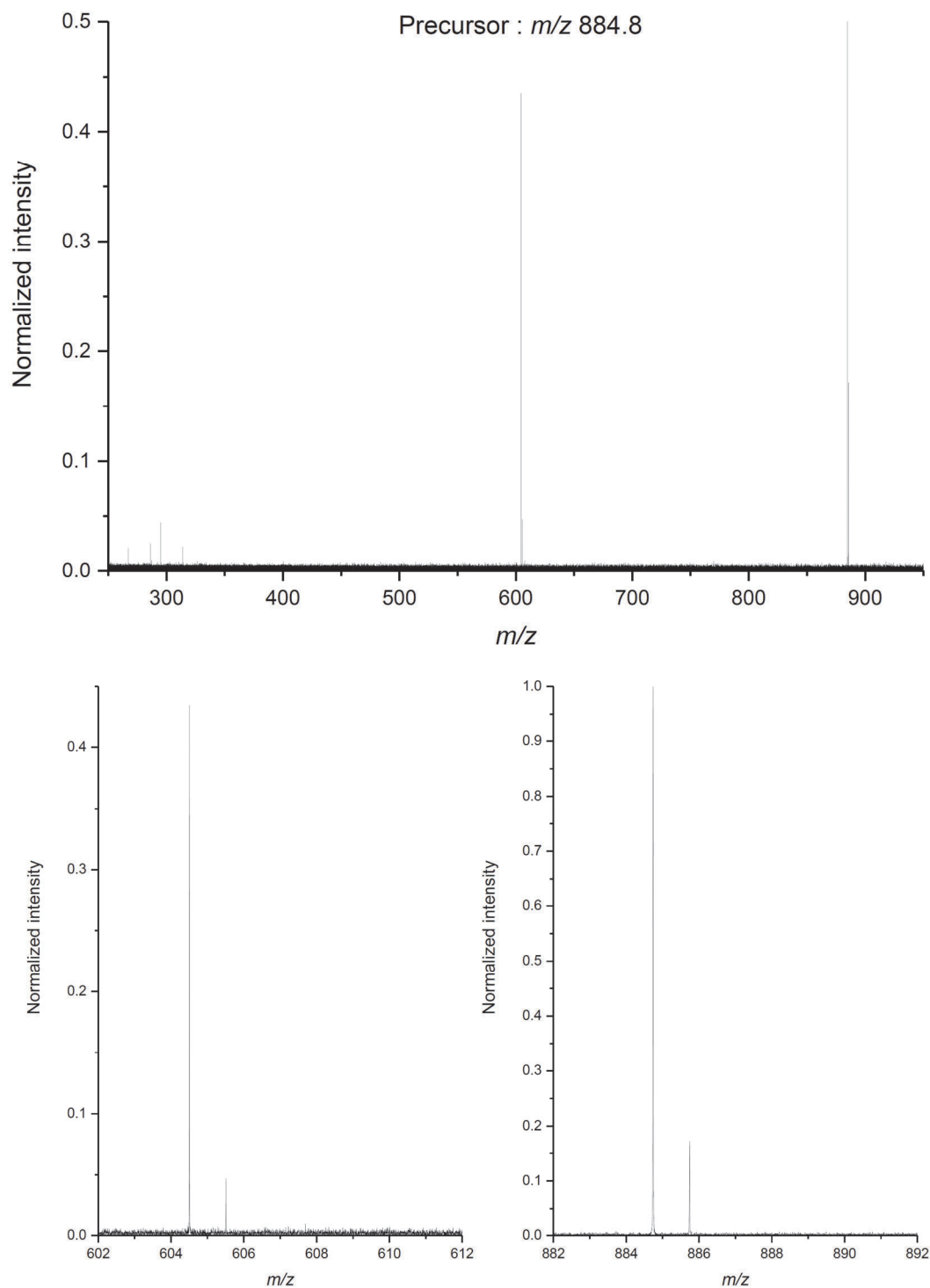


Figure S5e. Classical MS/MS FT-ICR spectra of sesame oil cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD: upper panel full MS spectrum, right lower panel zoom on the precursor, left lower panel, zoom on fragments corresponding to the loss of a neutral fatty acid b) precursor m/z 884.7 (quadrupole selection $\pm 0.5 m/z$).

Table S1. Interpretation of fragment spectrum of the mono-isotopic peak for a classical not-NUS (ie NUS 1) acquisition.

Plain acquisition (NUS 1)									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
1350.75922	MH ⁺ (+3)	1350.7557	-0.00352	-2.6		2.38E+07	6.08	26.67%	5624
1349.75128	MH ⁺ (+2)	1349.7562	0.00492	3.6		5.13E+07	13.11	57.50%	5581
1348.74334	MH ⁺ (+1)	1348.7466	0.00326	2.4		5.97E+07	15.27	66.98%	5548
1347.7354	MH ⁺	1347.7421	0.0067	5.0		8.91E+07	22.80	100.00%	5595
1333.73272	MH ⁺ -NH3 (+3)	1333.7286	-0.00412	-3.1		1.73E+07	4.43		5823
1332.72478	MH ⁺ -NH3 (+2)	1332.7249	0.00012	0.1		4.17E+07	10.66		5686
1331.71684	MH ⁺ -NH3 (+1)	1331.7236	0.00676	5.1		5.05E+07	12.93		5676
					1315.6942	1.50E+07	3.84		5805
					1314.6936	1.67E+07	4.27		6039
					1306.7248	1.43E+07	3.65		5914
					1305.7246	1.93E+07	4.94		6010
					1304.7313	1.66E+07	4.24		6118
1291.67294	m10 (+1)	1291.7237	0.05076	39.3		1.71E+07	4.37		6201
1290.665	m10	1290.7197	0.0547	42.4		2.29E+07	5.85		5931
					1289.6829	1.07E+07	2.75		5866
					1288.7291	1.33E+07	3.41		5689

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Plain acquisition (NUS 1)									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
				1287.7329	1.74E+07	4.45		5920	
				1278.704	1.38E+07	3.53		6228	
				1277.7109	1.87E+07	4.79		5944	
1275.6905	m5 / m6	1275.7063	0.0158 12.4	1274.7214	1.72E+07	4.40		5905	
				1273.689	1.39E+07	3.54		6479	
				1271.7236	1.24E+07	3.17		6309	
				1270.7166	1.81E+07	4.63		6166	
				1262.6656	1.06E+07	2.72		5869	
				1261.659	2.86E+07	7.31		5831	
				1260.6585	3.46E+07	8.86		6094	
				1258.7077	1.21E+07	3.09		6158	
				1257.7092	1.54E+07	3.94		6151	
				1247.662	1.16E+07	2.96		6183	
1219.71872	c10 (+3)	1219.7077	-0.01102 -9.0		1.99E+07	5.10	5.94%	6355	
1218.71078	c10 (+2)	1218.7051	-0.00568 -4.7		8.81E+07	22.52	26.27%	6101	
1217.70284	c10 (+1)	1217.7021	-0.00074 -0.6		2.46E+08	62.84	73.28%	6172	
1216.6949	c10	1216.6996	0.0047 3.9		3.35E+08	85.75	100.00%	6174	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Plain acquisition (NUS 1)									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
				1186.6354	1.07E+07	2.74		6490	
1105.62678	c9 (+2)	1105.6195	-0.00728		4.40E+07	11.25	20.37%	6798	
1104.61884	c9 (+1)	1104.6166	-0.00224		1.43E+08	36.63	66.31%	6811	
1103.6109	c9	1103.6143	0.0034		2.16E+08	55.24	100.00%	6819	
1079.57074	z9 (+1)	1079.5652	-0.00554		2.42E+07	6.19	50.94%	7021	
1078.5628	z9	1078.5659	0.0031		4.75E+07	12.14	100.00%	7009	
1048.60528	c8 (+2)	1048.5957	-0.00958		3.42E+07	8.75	19.36%	7365	
1047.59734	c8 (+1)	1047.594	-0.00334		1.14E+08	29.09	64.34%	7209	
1046.5894	c8	1046.5924	0.003		1.77E+08	45.22	100.00%	7226	
				1007.5232	1.09E+07	2.78		8016	
				1004.5563	1.05E+07	2.68		8002	
901.53688	c7 (+2)	901.5272	-0.00968		3.78E+07	9.67	13.15%	8438	
900.52894	c7 (+1)	900.5251	-0.00384		1.46E+08	37.43	50.88%	8378	
899.521	c7	899.5219	0.0009		2.88E+08	73.57	100.00%	8372	
				855.5069	2.41E+07	6.16		8901	
754.46848	c6 (+2)	754.458	-0.01048		1.74E+07	4.46	8.27%	10296	
753.46054	c6 (+1)	753.454	-0.00654		8.66E+07	22.14	41.09%	9994	
752.4526	c6	752.4522	-0.0004		2.11E+08	53.88	100.00%	10006	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Plain acquisition (NUS 1)									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
676.38718	MH2+ (+4)	676.3739	-0.01328	-19.6	5.72E+07	14.63		10383	
675.88321	MH2+ (+3)	675.8727	-0.01051	-15.6	2.53E+08	64.67		10840	
675.37924	MH2+ (+2)	675.3719	-0.00734	-10.9	8.44E+08	215.77		10935	
674.87527	MH2+ (+1)	674.8711	-0.00417	-6.2	2.00E+09	511.66		10998	
674.3713	MH2+	674.3699	-0.0014	-2.1	3.09E+09	789.17		11120	
626.40988	c5 (+2)	626.3981	-0.01178	-18.8	4.26E7	10.90		17233	
625.40194	c5 (+1)	625.3958	-0.00614	-9.8	1.07E+07	2.75		15794	
624.394	c5	624.3929	-0.0011	-1.8	2.28E+07	5.84	5.84%	12439	
623.3862	c5 (-1)	623.3854	-0.0008	-1.3	1.33E+08	33.90	33.90%	12014	
552.3091	c9 ²⁺	552.3133	0.0042	7.6	3.91E+08	100.00	100.00%	11976	
497.34334	c4 (+1)	497.3372	-0.00614	-12.3	4.94E+07	12.64		11915	
496.3354	c4	496.3338	-0.0016	-3.2	1.55E+07	3.96		13142	
495.3276	c4 (-1)	495.3259	-0.0017	-3.4	4.83E+07	12.36	26.35%	15032	
					1.83E+08	46.91	100.00%	15191	
					4.75E+07	12.15		15255	
					2.08E+07	5.32		17995	
				424.2523					

Table S2. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/2 acquisition.

NUS 1/2										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1351.74322	MH ⁺ (+4)	1351.7581	0.01488	11.0		2.58E+07	1.70	7.03%	6279	
1350.73528	MH ⁺ (+3)	1350.7575	0.02222	16.5		1.00E+08	6.60	27.30%	5648	
1349.72734	MH ⁺ (+2)	1349.7545	0.02716	20.1		2.63E+08	17.32	71.64%	5592	
1348.7194	MH ⁺ (+1)	1348.7626	0.0432	32.0		9.74E+07	6.42	26.55%	5418	
1347.7116	MH ⁺	1347.7411	0.0295	21.9		3.67E+08	24.18	100.00%	5562	
1334.71672	MH+ -NH3 (+4)	1334.7255	0.00878	6.6		2.22E+07	1.46		9651	
1333.70878	MH+ -NH3 (+3)	1333.728	0.01922	14.4		7.92E+07	5.22		5997	
1332.70084	MH ⁺ -NH3 (+2)	1332.7248	0.02396	18.0		2.18E+08	14.39		5657	
1331.6929	MH ⁺ -NH3 (+1)	1331.7237	0.0308	23.1		2.33E+08	15.37		5700	
					1316.7246	1.36E+07	0.89		10832	
					1316.6158	1.46E+07	0.96		7740	
					1315.7047	8.48E+07	5.59		7147	
					1314.698	8.10E+07	5.34		5764	
					1307.7076	1.35E+07	0.89		7779	
					1306.7385	6.97E+07	4.60		8669	
					1305.7215	9.31E+07	6.14		5746	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2									
m/z theoretical	Assignment (isotope)	m/z assigned	Mass shift	ppm	m/z unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
					1304.7709	9.14E+07	6.02		8382
					1303.7257	1.00E+08	6.61		5765
					1301.7556	1.46E+07	0.96		10765
1293.68882	m10 (+3)	1293.6811	-0.00772	-6.0		2.28E+07	1.50		10299
1292.68088	m10 (+2)	1292.6975	0.01662	12.9		4.17E+07	2.75		8513
1291.67294	m10 (+1)	1291.723	0.05006	38.8		6.60E+07	4.35		5918
1290.665	m10	1290.7237	0.0587	45.5		5.69E+07	3.75		5848
					1289.7971	2.83E+07	1.87		11205
					1289.7046	2.68E+07	1.76		5397
					1289.6371	3.11E+07	2.05		11241
					1288.7439	6.53E+07	4.30		10021
					1285.6974	1.20E+07	0.79		9354
1279.72226	m5 / m6 (+4)	1279.7019	-0.02036	-15.9		2.25E+07	1.48		9152
1278.71432	m5 / m6 (+3)	1278.7084	-0.00592	-4.6		5.77E+07	3.80		3537
1277.70638	m5 / m6 (+2)	1277.6931	-0.01328	-10.4		8.21E+07	5.41		5842
1276.69844	m5 / m6 (+1)	1276.7129	0.01446	11.3		2.24E+07	1.47		7522
1275.6905	m5 / m6	1275.726	0.0355	27.8		7.66E+07	5.05		5560
1274.72448	m11 (+2)	1274.7359	0.01142	9.0		6.00E+07	3.95		5812

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1273.71654	m11 (+1)	1273.6908	-0.02574	-20.2		4.94E+07	3.26		4693	
1272.7086	m11	1272.7151	0.0065	5.1	1271.7174	1.31E+07	0.87		8175	
					1270.7111	6.14E+07	4.05		9343	
					1262.6357	7.87E+07	5.19		6925	
					1261.6619	5.22E+07	3.44		11306	
					1260.6585	1.47E+08	9.70		5884	
					1258.69	1.47E+08	9.66		5995	
					1257.7071	7.33E+07	4.83		10007	
					1249.6948	6.60E+07	4.35		6032	
					1248.6352	4.17E+07	2.75		10055	
					1241.6911	3.97E+07	2.61		10510	
					1240.638	1.76E+07	1.16		10841	
					1233.678	1.58E+07	1.04		10431	
						1.20E+07	0.79		8864	
1219.71872	c10 (+3)	1219.7057	-0.01302	-10.7		6.07E+07	4.00	4.38%	5852	
1218.71078	c10 (+2)	1218.7069	-0.00388	-3.2		3.12E+08	20.55	22.51%	6064	
					1217.9406	1.87E+07	1.23		10208	
1217.70284	c10 (+1)	1217.7026	-0.00024	-0.2		1.14E+09	75.28	82.43%	6154	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2									
m/z theoretical	Assignment (isotope)	m/z assigned	Mass shift	ppm	m/z unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
1216.6949	c10	1216.6996	0.0047	3.9		1.39E+09	91.31	100.00%	6231
					1216.4428	1.27E+07	0.84		8905
1200.67634	b10 (+1)			-7.6	1204.6508	1.52E+07	1.00		9810
		1200.6672	-0.00914			1.38E+07	0.91	56.88%	10063
1199.6684	b10	1199.6905	0.0221	18.4		2.43E+07	1.60	100.00%	8019
1191.6343	y10	1191.6095	-0.0248	-20.8		1.04E+07	0.68		11666
					1187.6884	2.74E+07	1.81		9018
1106.63472	c9 (+3)			-21.5	1186.6212	3.37E+07	2.22		11162
		1106.6109	-0.02382			3.15E+07	2.08	3.45%	11935
1105.62678	c9 (+2)	1105.6203	-0.00648	-5.9		1.56E+08	10.26	17.01%	6678
					1104.8139	1.18E+07	0.78		12921
1104.61884	c9 (+1)	1104.6168	-0.00204	-1.8		6.44E+08	42.45	70.40%	6804
1103.6109	c9	1103.6149	0.004	3.6		9.15E+08	60.30	100.00%	6709
1094.5815	y9	1094.5753	-0.0062	-5.7		1.88E+07	1.24		9444
1080.55474	z9 (+1)	1080.5557	0.00096	0.9		1.23E+07	0.81	9.71%	6211
1079.5468	z9	1079.5658	0.019	17.6		1.27E+08	8.34	100.00%	6785
1078.539	y9 -NH3 (+1)	1078.5656	0.0266	24.7		2.17E+08	14.33		6932
					1060.5735	1.47E+07	0.97		12752

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
					1059.5873	2.81E+07	1.85		13769	
1049.61322	c8 (+3)	1049.6193	0.00608	5.8		1.89E+07	1.25	2.42%	9455	
1048.60528	c8 (+2)	1048.596	-0.00928	-8.8		1.27E+08	8.35	16.22%	7281	
1047.59734	c8 (+1)	1047.5941	-0.00324	-3.1		5.26E+08	34.67	67.33%	7242	
1046.5894	c8	1046.5924	0.003	2.9		7.81E+08	51.48	100.00%	7079	
					1021.53	1.64E+07	1.08		7155	
					1020.5428	3.53E+07	2.32		14511	
					1018.5504	1.07E+07	0.71		8611	
					1008.5569	2.11E+07	1.39		11590	
					1007.5164	2.50E+07	1.64		9734	
					1007.4699	2.23E+07	1.47		11519	
					1005.5475	2.37E+07	1.56		4584	
					1004.5826	3.44E+07	2.27		10111	
					1004.4933	2.61E+07	1.72		12541	
					946.5182	1.07E+07	0.70		8889	
902.54482	c7 (+3)	902.5243	-0.02052	-22.7		1.88E+07	1.24	1.53%	12735	
901.53688	c7 (+2)	901.5294	-0.00748	-8.3		1.42E+08	9.36	11.54%	8479	
900.52894	c7 (+1)	900.5242	-0.00474	-5.3		7.16E+08	47.21	58.21%	8320	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2									
m/z theoretical	Assignment (isotope)	m/z assigned	Mass shift	ppm	m/z unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
899.521	c7	899.5219	0.0009	1.0		1.23E+09	81.11	100.00%	8413
754.46848	c6 (+2)	754.4585	-0.00998	-13.2	856.5093	4.87E+07	3.21		9408
753.46054	c6 (+1)	753.4545	-0.00604	-8.0	855.5079	1.13E+08	7.42		8574
752.4526	c6	752.4525	-1E-04	-0.1		6.49E+07	4.27	7.75%	10067
751.4448	c6 (-1)	751.4238	-0.021	-27.9		4.02E+08	26.50	48.01%	9933
676.38718	MH ²⁺ (+4)	676.3713	-0.01588	-23.5		8.37E+08	55.19	100.00%	9976
675.88321	MH ²⁺ (+3)	675.8712	-0.01201	-17.8		2.45E+07	1.62		13194
					675.6994	7.03E+07	4.63		10838
					675.6646	5.16E+08	34.02		11099
					675.636	1.89E+07	1.24		16604
					675.5551	2.17E+07	1.43		12371
					675.3721	1.48E+07	0.98		22276
675.37924	MH ²⁺ (+2)	675.3721	-0.00714	-10.6		1.35E+07	0.89		19818
					675.2299	2.59E+09	170.71		10892
					675.1908	3.08E+07	2.03		17706
					675.1592	2.22E+07	1.47		20636
					675.1224	2.08E+07	1.37		20501
						2.29E+07	1.51		17941

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
		674.871	-0.00427	-6.3	675.0871	2.75E+07	1.81		19322
674.87527	MH ²⁺ (+1)				675.0517	3.15E+07	2.08		20482
674.3713	MH ²⁺	674.3698	-0.0015	-2.2		8.32E+09	548.37		10999
						1.65E+10	1086.59		11186
					674.2937	2.30E+08	15.17		17226
					673.8217	2.07E+07	1.36		22124
					673.6689	1.16E+07	0.77		20532
					673.5557	2.42E+07	1.60		17517
					673.4858	1.58E+07	1.04		21134
					673.4136	1.43E+07	0.94		19065
					665.86	2.50E+07	1.65		14273
626.40988	c5 (+2)	626.3982	-0.01168	-18.6		7.90E+07	5.20	5.20%	12452
625.40194	c5 (+1)	625.3953	-0.00664	-10.6		6.38E+08	42.02	42.02%	11768
624.394	c5	624.3929	-0.0011	-1.8		1.52E+09	100.00	100.00%	11892
623.3862	c5 (-1)	623.3851	-0.0011	-1.8		2.28E+08	15.02		12083
607.3675	b5	607.3685	0.001	1.6	606.3847	1.18E+07	0.78		16895
					553.3202	1.02E+07	0.67		13898
						1.72E+07	1.13		9261

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/2									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
552.3091	c9 ²⁺	552.3124				6.91E+07	4.55		12798
497.34334	c4 (+1)	497.3366				2.03E+08	13.39	29.97%	14964
496.3354	c4	496.3338				6.78E+08	44.69	100.00%	15045
495.3276	c4 (-1)	495.3258				1.96E+08	12.90		15126
					425.2555	1.19E+07	0.79		17691
					425.2504	1.27E+07	0.84		27752
					424.2523	7.94E+07	5.23		18031
254.1612	b2	254.158				2.36E+07	1.56		43673

Table S3. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/4 acquisition.

NUS 1/4										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1350.75922	MH ⁺ (+3)	1350.7823	0.02308	17.1		4.25E+07	0.57	3.67%	6602	
1349.75128	MH ⁺ (+2)	1349.7531	0.00182	1.3		6.98E+08	9.29	60.21%	5216	
1348.74334	MH ⁺ (+1)	1348.752	0.00866	6.4		1.16E+09	15.43	100.00%	5549	
1347.7354	MH ⁺	1347.7399	0.0045	3.3		1.09E+09	14.52	94.07%	5642	
1333.70878	MH ⁺ -NH3 (+2)	1333.7235	0.01472	11.0		2.71E+07	0.36		9496	
1332.70084	MH ⁺ -NH3 (+1)	1332.7256	0.02476	18.6		4.04E+08	5.38		5880	
1331.6929	MH ⁺ -NH3 +1)	1331.7234	0.0305	22.9		6.40E+08	8.52		5600	
					1315.7342	8.74E+07	1.16		9077	
					1314.6967	2.34E+08	3.11		6011	
					1306.7347	4.48E+07	0.60		5641	
					1305.7159	2.08E+08	2.77		5674	
					1304.8106	7.42E+07	0.99		10477	
					1304.7242	7.16E+07	0.95		3936	
					1304.655	7.38E+07	0.98		8940	
					1303.7287	2.71E+08	3.60		6193	
					1302.7097	1.22E+07	0.16		5591	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
					1301.76	2.65E+07	0.35		11136
1291.67294	m10 (+1)	1291.719	0.04606	35.7	1293.6652	1.39E+07	0.18		5437
1290.665	m10	1290.7158	0.0508	39.4		1.54E+08	2.06		5917
					1288.7777	6.05E+07	0.81		6075
					1287.72	1.21E+08	1.61		9934
					1285.7249	1.79E+07	0.24		7454
					1284.785	1.90E+07	0.25		7575
					1284.7399	1.95E+07	0.26		10800
					1277.7069	2.25E+08	3.00		9155
					1275.7257	1.49E+08	1.99		5172
					1274.7228	2.31E+08	3.07		5277
					1273.7233	1.10E+08	1.47		5061
					1271.7506	5.98E+07	0.80		10303
					1270.7075	2.22E+08	2.96		8867
					1261.6633	2.84E+08	3.78		8390
					1260.6609	4.34E+08	5.78		5897
					1258.6982	8.76E+07	1.17		5850
									9052

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1257.6647	m7 / m8	1257.7056	0.0409	32.5	2.05E+08	2.73		6337	
				1256.6642	3.06E+07	0.41		3754	
				1253.6746	2.24E+07	0.30		8697	
				1249.7288	4.60E+07	0.61		10544	
				1249.6575	4.35E+07	0.58		6747	
				1248.6563	4.56E+07	0.61		6277	
				1247.7053	6.77E+07	0.90		8264	
				1244.6608	1.05E+07	0.14		5974	
				1243.6635	1.36E+07	0.18		8010	
				1240.6756	4.08E+07	0.54		11159	
				1232.6385	2.10E+07	0.28		8529	
1219.71872	c10 (+3)	1219.7192	0.00048	0.4	2.67E+07	0.36	0.68%	8116	
1218.71078	c10 (+2)	1218.7243	0.01352	11.1	6.05E+07	0.81	1.53%	6680	
1217.70284	c10 (+1)	1217.7036	0.00076	0.6	2.17E+09	28.83	54.93%	6263	
1216.6949	c10	1216.7004	0.0055	4.5	3.94E+09	52.49	100.00%	6346	
				1204.6614	1.58E+07	0.21		5045	
				1203.6667	3.28E+07	0.44		9223	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution		
1199.6684	b10	1199.6611	-6.1	1201.5976	2.12E+07	0.28		12527		
					5.20E+07	0.69	100.00%	7388		
1191.6343	y10	1191.6314	-2.4	1199.6062	5.14E+07	0.68		11753		
1189.6187	Y10	1189.5942	-20.6		2.39E+07	0.32		11079		
					2.02E+07	0.27		12043		
				1187.6551	2.83E+07	0.38		7021		
				1187.592	2.86E+07	0.38		6917		
				1186.6438	7.15E+07	0.95		6128		
1105.62678	c9 (+2)	1105.6286	1.6		4.41E+07	0.59	1.52%	12330		
1104.61884	c9 (+1)	1104.6182	-0.6		1.32E+09	17.60	45.42%	6848		
1103.6109	c9	1103.6155	4.2		2.91E+09	38.76	100.00%	6761		
				1094.5723	8.48E+07	1.13		11108		
1086.5843	b9	1086.6056	19.6		1.21E+07	0.16		11373		
1080.55474	z9 (+1)	1080.5792	22.6		1.01E+08	1.34	23.94%	7990		
1079.5468	z9	1079.5719	23.2		4.22E+08	5.61	100.00%	7126		
1078.539	y9 -NH3	1078.567	26.0		6.81E+08	9.07		7325		
				1077.5488	2.81E+07	0.37		8645		
				1060.5639	1.91E+07	0.25		10567		

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1048.60528	c8 (+2)	1048.6195	0.01422	13.6	1059.6058	6.34E+07	0.84		10408	
1047.59734	c8 (+1)	1047.5961	-0.00124	-1.2		3.63E+07	0.48	1.36%	13822	
1046.5894	c8	1046.5928	0.0034	3.2		1.12E+09	14.90	42.10%	6845	
						2.66E+09	35.38	100.00%	7082	
					1045.5602	1.71E+07	0.23		13182	
					1031.5934	1.08E+07	0.14		11273	
					1031.5673	1.04E+07	0.14		8226	
1029.5629	b8	1029.5628	-1E-04	-0.1		1.28E+07	0.17		9330	
					1022.4944	2.09E+07	0.28		12924	
					1021.558	2.36E+07	0.31		11830	
					1021.5293	2.54E+07	0.34		14780	
					1020.5387	7.89E+07	1.05		12957	
					1018.5376	1.12E+07	0.15		14941	
					1017.5768	4.90E+07	0.65		11215	
					1008.5495	2.98E+07	0.40		12031	
					1008.4955	4.00E+07	0.53		14737	
					1007.4909	8.87E+07	1.18		13064	
					1005.5373	3.26E+07	0.43		12994	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
					1004.5154	8.79E+07	1.17		12983
					1003.5646	2.33E+07	0.31		10688
					1002.5885	1.17E+07	0.16		9756
					986.5795	1.57E+07	0.21		15189
					946.5152	2.84E+07	0.38		9160
					933.5082	2.28E+07	0.30		12497
					933.4852	2.37E+07	0.32		14957
901.53688	c7 (+2)	901.5363	-0.00058	-0.6		5.69E+07	0.76	1.18%	10434
900.52894	c7 (+1)	900.5243	-0.00464	-5.2		1.65E+09	21.93	34.16%	8059
899.521	c7	899.5222	0.0012	1.3		4.82E+09	64.20	100.00%	8673
882.4944	b7	882.497	0.0026	2.9	898.5481	2.66E+07	0.35		16299
					857.5131	1.98E+07	0.26		12773
					856.5111	2.02E+07	0.27		17775
					855.5066	1.88E+08	2.50		8789
					839.4859	4.33E+08	5.77		7993
						1.15E+07	0.15		14241
754.46848	c6 (+2)	754.4635	-0.00498	-6.6		3.55E+07	0.47	0.95%	14141
753.46054	c6 (+1)	753.4547	-0.00584	-7.8		1.01E+09	13.43	26.90%	9388

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
752.4526	c6	752.4523	-0.0003		3.75E+09	49.92	100.00%	10309	
735.426	b6	735.4223	-0.0037	751.4332	9.42E+07	1.25		18598	
			-5.0	734.4389	1.60E+07	0.21		11465	
				708.4366	1.12E+07	0.15		19195	
				675.9396	1.53E+07	0.20		10885	
					1.77E+07	0.24		22212	
675.88321	MH ²⁺ (+3)	675.883	-0.00021		3.98E+08	5.29		12605	
675.37924	MH ²⁺ (+2)	675.3747	-0.00454		1.23E+09	16.39		12050	
			-6.7	675.2992	5.62E+07	0.75		14473	
				675.2242	5.96E+07	0.79		22180	
				675.0785	3.83E+07	0.51		19150	
				675.0439	4.31E+07	0.57		22097	
				675.0153	9.90E+07	1.32		17928	
674.87527	MH ²⁺ (+1)	674.8711	-0.00417		1.55E+10	206.69		10934	
			-6.2	674.7019	8.41E+07	1.12		18827	
674.3713	MH ²⁺	674.3697	-0.0016		5.48E+10	729.56		11305	
			-2.4	673.5938	2.64E+07	0.35		22483	
				665.8588	1.16E+08	1.54		16486	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
					656.852	1.89E+07	0.25		10169	
626.40988	c5 (+2)	626.3998	-0.01008	-16.1		4.54E+07	0.60	0.60%	14813	
625.40194	c5 (+1)	625.3952	-0.00674	-10.8		1.71E+09	22.74	22.74%	11525	
624.394	c5	624.3925	-0.0015	-2.4		7.51E+09	100.00	100.00%	12048	
623.3862	c5 (-1)	623.3841	-0.0021	-3.4		1.02E+09	13.51		12400	
607.3675	b5	607.3701	0.0026	4.3		3.36E+07	0.45		12786	
					606.3779	3.43E+07	0.46		12604	
					600.337	3.26E+07	0.43		21645	
					553.323	3.34E+07	0.44		19855	
552.3091	c9 ²⁺	552.3125	0.0034	6.2		3.38E+08	4.50		12399	
					524.3035	1.64E+07	0.22		18518	
497.34334	c4 (+1)	497.3365	-0.00684	-13.8		5.42E+08	7.21	16.71%	14872	
496.3354	c4	496.3334	-0.002	-4.0		3.24E+09	43.19	100.00%	15255	
495.3276	c4 (-1)	495.3256	-0.002	-4.0		8.82E+08	11.74		15291	
479.3089	b4	479.3065	-0.0024	-5.0		1.20E+07	0.16		20506	
					478.3209	2.02E+07	0.27		20610	
					453.3268	2.16E+07	0.29		30506	
					425.2538	3.07E+07	0.41		28387	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/4									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
382.2561	b3	382.2523	-0.0038	-9.9	424.2521	4.09E+08	5.45		17663
					364.2445	1.17E+07	0.16		32853
					354.1759	2.20E+07	0.29		24399
					271.1864	1.01E+07	0.13		13291
					270.1801	1.35E+08	1.79		28639
					254.159	1.57E+07	0.21		55436
					243.1451	1.58E+08	2.10		46992
					243.1433	1.17E+07	0.16		56376
					237.1329	1.22E+07	0.16		53572
					226.153	2.04E+07	0.27		28226
					226.1177	1.28E+07	0.17		42999
						3.02E+07	0.40		55868

Table S4. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/8 acquisition.

NUS 1/8										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1350.73528	MH ⁺ (+3)	1350.7586	0.02332	17.3		5.53E+07	0.46	2.93%	8950	
1349.72734	MH ⁺ (+2)	1349.7526	0.02526	18.7		1.93E+08	1.60	10.23%	6044	
1348.7194	MH ⁺ (+1)	1348.7468	0.0274	20.3		1.89E+09	15.63	100.00%	6405	
1347.7116	MH ⁺	1347.7353	0.0237	17.6		1.88E+09	15.51	99.27%	5432	
1332.70084	MH ⁺ -NH3 (+2)	1332.7177	0.01686	12.7		9.83E+07	0.81		5781	
1331.6929	MH ⁺ -NH3 (+1)	1331.7191	0.0262	19.7		1.07E+09	8.88		5097	
					1314.6827	3.71E+08	3.07		5290	
					1305.7188	3.46E+08	2.86		3458	
					1303.7219	4.55E+08	3.77		6133	
					1301.7091	2.56E+07	0.21		7279	
					1292.6791	4.01E+07	0.33		5703	
					1292.6191	3.86E+07	0.32		8043	
					1291.7399	4.65E+07	0.38		5893	
1290.665	m10	1290.7155	0.0505	39.1		5.03E+08	4.16		5864	
					1289.7092	7.47E+07	0.62		7230	
					1287.7523	1.31E+08	1.08		7475	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolutio n	
				1284.7133	2.65E+07	0.22		9156	
				1277.6968	3.95E+08	3.26		5570	
				1275.8314	2.33E+07	0.19		10992	
				1274.7155	3.76E+08	3.11		7731	
				1273.6571	1.03E+08	0.85		6408	
				1270.8003	1.43E+08	1.18		8836	
				1261.6745	6.51E+07	0.54		6734	
				1260.6532	7.12E+08	5.89		5624	
1257.68854	m7 / m8 (+1)	1257.6984	0.00986		3.45E+08	2.86		6517	
1256.6806	m7 / m8	1256.6752	-0.0054		3.49E+07	0.29		5937	
				1253.6683	3.59E+07	0.30		11462	
				1249.6591	6.51E+07	0.54		7063	
				1247.6373	9.27E+07	0.77		9836	
				1245.6728	1.14E+07	0.09		8823	
				1243.6301	2.11E+07	0.17		7504	
				1240.6966	3.71E+07	0.31		7600	
				1234.6647	1.03E+07	0.09		11611	
				1232.6586	2.55E+07	0.21		9607	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
					1221.6584	1.12E+07	0.09		9984	
1219.71872	c10 (+3)	1219.7007	-0.01802	-14.8		1.88E+07	0.16	0.28%	5436	
1218.71078	c10 (+2)	1218.6971	-0.01368	-11.2		1.03E+08	0.85	1.52%	10800	
1217.70284	c10 (+1)	1217.7046	0.00176	1.4		4.63E+08	3.83	6.83%	7139	
1216.6949	c10	1216.6969	0.002	1.6		6.77E+09	56.03	100.00%	6005	
1215.6871	c10 (-1)	1215.7119	0.0248	20.4		1.13E+07	0.09		7608	
					1204.5967	1.76E+07	0.15		10107	
					1203.6956	4.04E+07	0.33		11413	
					1202.6042	1.30E+07	0.11		11495	
					1201.6269	1.96E+07	0.16		9581	
1199.6684	b10	1199.6606	-0.0078	-6.5		8.18E+07	0.68	100.00%	6849	
1191.6343	y10	1191.648	0.0137	11.5		2.30E+07	0.19		9841	
					1189.5943	2.81E+07	0.23		11076	
					1186.6805	7.20E+07	0.60		10487	
					1130.5868	1.49E+07	0.12		8685	
					1115.595	1.36E+07	0.11		9200	
1105.62678	c9 (+2)	1105.6082	-0.01858	-16.8		6.01E+07	0.50	1.18%	10851	
1104.61884	c9 (+1)	1104.6186	-0.00024	-0.2		3.09E+08	2.56	6.04%	7212	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
1103.6109	c9	1103.6116	0.0007	0.6		5.11E+09	42.28	100.00%	7059
1094.5815	y9	1094.5972	0.0157	14.3		9.62E+07	0.80		4389
1086.5843	b9	1086.5921	0.0078	7.2		1.99E+07	0.16		7163
1080.55474	z9 (+1)	1080.5743	0.01956	18.1		4.41E+07	0.37	12.54%	7246
1079.5468	z9	1079.5726	0.0258	23.9		3.52E+08	2.91	100.00%	8070
1078.539	y9 -NH3	1078.5639	0.0249	23.1		1.11E+09	9.17		6758
					1077.5503	4.54E+07	0.38		10778
					1059.6222	5.44E+07	0.45		9163
1048.60528	c8 (+2)	1048.6207	0.01542	14.7		3.65E+07	0.30	0.80%	7250
1047.59734	c8 (+1)	1047.5975	0.00016	0.2		2.72E+08	2.25	6.00%	7987
1046.5894	c8	1046.5891	-0.0003	-0.3		4.54E+09	37.52	100.00%	7230
					1045.5734	1.67E+07	0.14		8728
					1031.5601	1.23E+07	0.10		6914
1029.5629	b8	1029.5535	-0.0094	-9.1		1.77E+07	0.15		5447
					1022.5303	2.11E+07	0.17		10126
					1020.534	7.58E+07	0.63		7679
					1017.5276	6.29E+07	0.52		11825
					1007.523	1.02E+08	0.84		7343

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
					1004.5405	1.05E+08	0.86		10189
					1003.591	3.05E+07	0.25		6904
					1002.5843	1.58E+07	0.13		9168
					1002.5623	1.62E+07	0.13		13765
					986.5592	1.91E+07	0.16		12036
					948.4899	1.03E+07	0.09		6812
					946.5148	4.74E+07	0.39		12802
					933.5014	3.14E+07	0.26		11781
901.53688	c7 (+2)	901.5465	0.00962	10.7		4.73E+07	0.39	0.59%	14269
900.52894	c7 (+1)	900.5236	-0.00534	-5.9		4.21E+08	3.49	5.29%	9091
899.521	c7	899.5189	-0.0021	-2.3		7.97E+09	65.91	100.00%	8819
898.5132	c7 (-1)	898.5151	0.0019	2.1		3.53E+07	0.29		12954
882.4944	b7	882.4897	-0.0047	-5.3		2.60E+07	0.22		11123
		881.4993			1002.5702	1.21E+07	0.10		8408
		856.5139			986.5576	1.58E+08	1.31		10173
		855.505			953.474	7.07E+08	5.85		9532
		839.4869			949.4714	1.62E+07	0.13		12990
		756.423			948.4947	1.74E+07	0.14		16047

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
753.46054	c6 (+1)	753.4533	-0.00724	-9.6		2.44E+08	2.02		10976
752.4526	c6	752.4491	-0.0035	-4.7	752.5488	4.73E+07	0.39		13158
751.4448	c6 (-1)	751.4458	0.001	1.3		6.10E+09	50.46		9503
735.426	b6	735.4233	-0.0027	-3.7		9.21E+07	0.76		12462
						2.59E+07	0.21		12265
					734.4453	1.49E+07	0.12		10881
675.88321	MH ²⁺ (+3)	675.8779	-0.00531	-7.9	708.4362	1.45E+07	0.12		17089
675.37924	MH ²⁺ (+2)	675.3719	-0.00734	-10.9	676.0145	4.22E+07	0.35		21322
						4.30E+08	3.55		17571
674.87527	MH ²⁺ (+1)	674.8719	-0.00337	-5.0		1.95E+09	16.16		11135
674.3713	MH ²⁺	674.3685	-0.0028	-4.2	675.2866	3.26E+07	0.27		19683
						6.48E+09	53.63		11097
						9.66E+10	799.28		11100
625.40194	c5 (+1)	625.3958	-0.00614	-9.8	665.8543	1.31E+08	1.09		10274
624.394	c5	624.3911	-0.0029	-4.6	656.8515	2.77E+07	0.23		11928
						3.78E+08	3.13	3.13%	12287
623.3862	c5 (-1)	623.3831	-0.0031	-5.0		1.21E+10	100.00	100.00%	11133
						1.60E+09	13.27		12028

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolutio n	
607.3675	b5	607.3585	-0.009		4.79E+07	0.40		18771	
				606.3813	4.76E+07	0.39		17155	
600.3378	b10 ²⁺	600.3361	-0.0017		3.98E+07	0.33		12591	
			-2.8	591.3329	1.17E+07	0.10		17672	
				581.388	1.25E+07	0.10		16613	
				565.3529	1.46E+07	0.12		9542	
552.3091	c9 ²⁺	552.3113	0.0022		5.49E+08	4.54		13308	
			4.0		2.14E+07	0.18		25938	
497.34334	c4 (+1)	497.3368	-0.00654		1.51E+08	1.25	2.79%	15367	
			-13.2		5.40E+09	44.63	100.00%	15176	
496.3354	c4	496.3327	-0.0027		1.43E+09	11.85		14865	
			-5.4		1.42E+07	0.12		16520	
495.3276	c4 (-1)	495.3247	-0.0029		1.42E+07	0.12		16520	
			-5.9		2.74E+07	0.23		23433	
479.3089	b4	479.3045	-0.0044		2.15E+07	0.18		24574	
			-9.2		6.58E+08	5.44		17498	
				478.3238	1.25E+07	0.10		14865	
			-4.4		2.99E+07	0.25		21816	
382.2561	b3	382.2544	-0.0017		1.57E+07	0.13		29498	
				364.2426	2.99E+07	0.25		21816	
				356.2756	1.57E+07	0.13		29498	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/8								
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolutio n
		354.1753		1.37E+07	0.11			27820
		254.1598		1.76E+08	1.45			38773
		243.1436		1.34E+07	0.11			34591
		237.1329		2.95E+07	0.24			36946
		226.1536		1.61E+07	0.13			26690
		226.1169		3.59E+07	0.30			22630

Table S5. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/16 acquisition.

NUS 1/16										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1350.75922	MH ⁺ (+3)	1350.7732	0.01398	10.3		6.95E+07	0.53	3.88%	8726	
1349.75128	MH ⁺ (+2)	1349.7442	-0.00708	-5.2		2.26E+08	1.74	12.64%	5959	
1348.74334	MH ⁺ (+1)	1348.7251	-0.01824	-13.5		1.10E+09	8.44	61.27%	7561	
1347.7354	MH ⁺	1347.7285	-0.0069	-5.1		1.79E+09	13.78	100.00%	5930	
					1332.7259	1.06E+08	0.81		6567	
					1331.6975	9.63E+08	7.40		8392	
					1314.6991	2.92E+08	2.24		4059	
					1305.6985	3.29E+08	2.53		9032	
					1303.7216	3.66E+08	2.82		7426	
					1301.7689	2.84E+07	0.22		6812	
1292.68088	m10 (+2)	1292.7054	0.02452	19.0		5.63E+07	0.43		7582	
1291.67294	m10 (+1)	1291.7178	0.04486	34.7		6.22E+07	0.48		7433	
1290.665	m10	1290.7165	0.0515	39.9		5.22E+08	4.01		6673	
					1289.6997	7.37E+07	0.57		8829	
					1288.6217	2.46E+07	0.19		10769	
					1287.7334	1.26E+08	0.97		5805	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
		1284.7216			1284.7216	2.62E+07	0.20		7464
		1277.6856			1277.6856	3.19E+08	2.45		5414
		1273.6741			1273.6741	1.05E+08	0.81		7738
		1267.6417			1267.6417	1.01E+07	0.08		4692
		1261.6548			1261.6548	7.88E+07	0.61		7983
		1260.6905			1260.6905	5.33E+08	4.10		7126
		1260.6249			1260.6249	5.74E+08	4.42		7702
1256.6806	m7 / m8	1256.6645	-0.0161	-12.8		3.88E+07	0.30		10208
					1253.6656	3.13E+07	0.24		6038
					1249.6546	6.29E+07	0.48		4681
1247.648	m1	1247.6341	-0.0139	-11.1		7.99E+07	0.61		4000
					1245.6803	1.23E+07	0.09		3308
					1243.6456	2.36E+07	0.18		6069
1242.7106	da11	1242.6789	-0.0317	-25.5		1.13E+07	0.09		11495
					1240.6667	4.00E+07	0.31		7393
					1234.6392	1.49E+07	0.11		8307
					1233.6611	1.05E+07	0.08		11171
					1232.6602	2.53E+07	0.19		6864

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	Resolution
1218.71078	c10 (+2)	1218.7157	0.00492	4.0	1221.6463	1.07E+07	0.08		6815	6815
1217.70284	c10 (+1)	1217.7054	0.00256	2.1	1218.6737	1.46E+08	1.12	2.18%	7739	7739
1216.6949	c10	1216.6959	0.001	0.8		1.42E+08	1.09		8872	8872
1215.6871	c10 (-1)	1215.7121	0.025	20.6		6.19E+08	4.76	9.21%	6576	6576
						6.72E+09	51.63	100.00%	7351	7351
						1.31E+07	0.10		9039	9039
					1214.5869	1.18E+07	0.09		12089	12089
					1204.6139	1.86E+07	0.14		5466	5466
					1203.6833	3.59E+07	0.28		5962	5962
					1202.6111	1.83E+07	0.14		7238	7238
					1201.6415	2.00E+07	0.15		8033	8033
1199.6684	b10	1199.6534	-0.015	-12.5		7.72E+07	0.59	100.00%	6393	6393
1191.6343	y10	1191.6341	-0.0002	-0.2		2.58E+07	0.20		8167	8167
1189.6187	Y10	1189.6169	-0.0018	-1.5	1190.6127	1.12E+07	0.09		10371	10371
1182.6418	b10 -NH3	1182.6463	0.0045	3.8	1186.6432	2.46E+07	0.19		7323	7323
						7.28E+07	0.56		7359	7359
						1.15E+07	0.09		8418	8418
					1175.6546	1.05E+07	0.08		7483	7483

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
					1130.5706	1.21E+07	0.09		6584
1105.62678	c9 (+2)	1105.6236	-0.00318	-2.9	1115.5856	1.27E+07	0.10		6489
1104.61884	c9 (+1)	1104.6057	-0.01314	-11.9		8.01E+07	0.62	1.57%	9852
1103.6109	c9	1103.6189	0.008	7.2		4.37E+08	3.36	8.56%	8750
1094.5815	y9	1094.5842	0.0027	2.5		5.11E+09	39.29	100.00%	4334
1086.5843	b9	1086.5743	-0.01	-9.2		9.38E+07	0.72		8602
1080.55474	z9 (+1)	1080.5669	0.01216	11.3		2.04E+07	0.16		6743
1079.5468	z9	1079.5548	0.008	7.4		5.58E+07	0.43	22.71%	7209
1078.539	y9 -NH3	1078.5744	0.0354	32.8		2.46E+08	1.89	100.00%	11808
					1078.541	9.10E+08	7.00		9628
					1077.5539	4.04E+07	0.31		7199
					1076.5687	1.05E+07	0.08		11154
					1059.6037	5.40E+07	0.41		8847
1048.60528	c8 (+2)	1048.5967	-0.00858	-8.2		6.57E+07	0.51	1.47%	6417
1047.59734	c8 (+1)	1047.5926	-0.00474	-4.5		3.64E+08	2.80	8.14%	6494
1046.5894	c8	1046.5847	-0.0047	-4.5		4.47E+09	34.38	100.00%	6060
1045.5816	c8 (-1)	1045.5655	-0.0161	-15.4		1.79E+07	0.14		9950

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
		1031.5697				1.38E+07	0.11		8239
1029.5629	b8	1029.5469	-0.016	-15.5		1.65E+07	0.13		5966
					1022.508	2.28E+07	0.18		9829
					1020.5408	7.93E+07	0.61		6350
					1018.5474	1.02E+07	0.08		6943
					1017.5569	4.98E+07	0.38		6906
					1008.5141	3.63E+07	0.28		13427
					1007.5142	1.02E+08	0.78		8556
					1004.5365	1.02E+08	0.78		8814
					1003.5787	3.02E+07	0.23		8145
					1002.5745	1.46E+07	0.11		12134
					986.5629	1.67E+07	0.13		4578
					948.4683	1.30E+07	0.10		9513
					946.5087	4.24E+07	0.33		6732
					933.5037	2.86E+07	0.22		8970
901.53688	c7 (+2)	901.5276	-0.00928	-10.3		7.17E+07	0.55	0.91%	7079
900.52894	c7 (+1)	900.5219	-0.00704	-7.8		5.27E+08	4.05	6.72%	6798
899.521	c7	899.514	-0.007	-7.8		7.85E+09	60.35	100.00%	7901

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
898.5132	c7 (-1)	898.5212	0.008	8.9		2.98E+07	0.23		7965	
882.4944	b7	882.4995	0.0051	5.8		2.52E+07	0.19		8027	
					881.5042	1.40E+07	0.11		8401	
					856.5054	1.29E+08	0.99		11618	
					855.508	6.28E+08	4.83		7647	
					839.4769	1.77E+07	0.14		11158	
					756.4249	1.55E+07	0.12		12427	
754.46848	c6 (+2)	754.4619	-0.00658	-8.7		3.23E+07	0.25	0.56%	14269	
753.46054	c6 (+1)	753.4509	-0.00964	-12.8		3.30E+08	2.54	5.74%	11078	
					752.5445	2.25E+07	0.17		18925	
752.4526	c6	752.4512	-0.0014	-1.9		5.75E+09	44.19	100.00%	7787	
751.4448	c6 (-1)	751.4433	-0.0015	-2.0		8.68E+07	0.67		13408	
735.426	b6	735.423	-0.003	-4.1		2.70E+07	0.21		13692	
					734.4391	1.53E+07	0.12		6477	
					708.4339	1.65E+07	0.13		9577	
676.38718	MH ²⁺ (+4)	676.3765	-0.01068	-15.8		5.69E+07	0.44		13396	
675.88321	MH ²⁺ (+3)	675.883	-0.00021	-0.3		2.37E+08	1.82		18422	
675.37924	MH ²⁺ (+2)	675.3728	-0.00644	-9.5		1.37E+09	10.54		14366	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
674.87527	MH ²⁺ (+1)	674.8711	-0.00417	-6.2		8.17E+09	62.78		11027	
674.3713	MH ²⁺	674.3681	-0.0032	-4.7		1.03E+11	788.56		11235	
626.40988	c5 (+2)	626.3964	-0.01348	-21.5	665.8561	1.21E+08	0.93		9610	
625.40194	c5 (+1)	625.3945	-0.00744	-11.9	656.8522	2.21E+07	0.17		12971	
624.394	c5	624.3919	-0.0021	-3.4		4.12E+07	0.32	0.32%	12769	
623.3862	c5 (-1)	623.3877	0.0015	2.4		5.85E+08	4.50	4.50%	13415	
607.3675	b5	607.3631	-0.0044	-7.2		1.30E+10	100.00	100.00%	12121	
600.3378	b10 ²⁺	600.335	-0.0028	-4.7	606.3811	1.63E+09	12.53		16889	
						4.50E+07	0.35		14039	
						4.16E+07	0.32		12603	
						3.73E+07	0.29		13169	
						1.02E+07	0.08		7666	
						1.20E+07	0.09		12289	
						1.51E+07	0.12		14300	
						1.42E+07	0.11		18529	
552.3091	c9 ²⁺	552.3118	0.0027	4.9		6.30E+08	4.84		15475	
497.34334	c4 (+1)	497.3358	-0.00754	-15.2	524.3027	1.93E+07	0.15		10975	
						2.17E+08	1.67	3.68%	15860	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/16									
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution
496.3354	c4	496.3323	-0.0031	-6.2		5.89E+09	45.30	100.00%	13462
495.3276	c4 (-1)	495.3243	-0.0033	-6.7		1.55E+09	11.88		14581
479.3089	b4	479.3064	-0.0025	-5.2		1.57E+07	0.12		20530
					478.3212	2.43E+07	0.19		15080
					453.3274	1.99E+07	0.15		24174
					424.2513	7.54E+08	5.80		17256
382.2561	b3	382.2543	-0.0018	-4.7		1.24E+07	0.10		27383
					364.2428	2.74E+07	0.21		19818
					356.2749	1.17E+07	0.09		29183
					354.1741	1.43E+07	0.11		24575
254.1612	b2	254.1592	-0.002	-7.9		1.45E+08	1.12		31040
					243.1436	1.43E+07	0.11		29220
237.1346	b2 -NH3	237.1326	-0.002	-8.4		2.73E+07	0.21		37061
					226.1528	1.53E+07	0.12		36350
					226.1169	3.40E+07	0.26		34814
					224.7898	8.19E+07	0.63		31193

Table S6. Interpretation of fragment spectrum of the mono-isotopic peak for a NUS 1/32 acquisition.

NUS 1/32										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
		1314.6906			3.08E+07	0.7			9444	
		1303.689			4.19E+07	0.9			9128	
		1301.7581			1.45E+07	0.3			6521	
		1287.7194			4.70E+07	1.0			6914	
		1273.6736			4.62E+07	1.0			7159	
		1260.7161			5.33E+07	1.2			10760	
		1260.6594			5.97E+07	1.3			9347	
		1256.6544			1.63E+07	0.4			6237	
		1253.6858			1.14E+07	0.2			2774	
		1203.6823			1.57E+07	0.3			7045	
1199.6684	b10	1199.6668	-0.0016	-1.3	4.01E+07	0.9	100.00%		6561	
					1.22E+07	0.3			7703	
					1.02E+07	0.2			8583	
1103.6109	c9	1103.6289	0.018	16.3	1.75E+09	37.8	100.00%		10093	
1094.5815	y9	1094.5715	-0.01	-9.1	7.20E+07	1.6			8447	
1086.5843	b9	1086.5778	-0.0065	-6.0	1.16E+07	0.3			7073	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/32										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
1077.539	y9 -NH3	1077.5548	0.0158	14.7		3.46E+07	0.7		6490	
1046.5894	c8	1046.6326	0.0432	41.3		1.27E+09	27.5	100.00%	12186	
1029.5629	b8	1029.5549	-0.008	-7.8	1274.7263	1.28E+07	0.3		7373	
					1020.5309	3.54E+07	0.8		9253	
					1017.5446	2.79E+07	0.6		9293	
					1003.5788	1.63E+07	0.4		9726	
					986.5473	1.05E+07	0.2		11345	
					946.5158	3.14E+07	0.7		9048	
					933.5107	1.38E+07	0.3		9647	
899.521	c7	899.5228	0.0018	2.0		1.74E+09	37.5	100.00%	8171	
882.4944	b7	882.4895	-0.0049	-5.6		1.88E+07	0.4		9051	
					855.5016	7.09E+07	1.5		11205	
752.4526	c6	752.4437	-0.0089	-11.8		1.64E+09	35.4	100.00%	6466	
735.426	b6	735.4215	-0.0045	-6.1		1.73E+07	0.4		10080	
674.3713	MH ²⁺	674.3678	-0.0035	-5.2		7.41E+10	1600.1		10170	
					665.8546	1.21E+08	2.6		12197	
					656.8511	2.23E+07	0.5		8415	
624.394	c5	624.4026	0.0086	13.8		4.63E+09	100.0	100.00%	20318	

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

NUS 1/32										
<i>m/z</i> theoretical	Assignment (isotope)	<i>m/z</i> assigned	Mass shift	ppm	<i>m/z</i> unassigned	Relative intensity	Normalized	Isotopic ratio	Resolution	
623.3862	c5 (-1)	623.3809	-0.0053	-8.5		5.29E+08	11.4		21488	
607.3675	b5	607.3631	-0.0044	-7.2		3.07E+07	0.7		13984	
600.3378	b10 ²⁺	600.3353	-0.0025	-4.2	606.3798	2.65E+07	0.6		13143	
					552.3167	3.47E+07	0.7		14710	
496.3354	c4	496.3304	-0.005	-10.1		2.99E+08	6.5		19802	
495.3276	c4 (-1)	495.3246	-0.003	-6.1		3.09E+09	66.7	100.00%	21188	
						7.27E+08	15.7		20347	
					478.3232	1.60E+07	0.3		18950	
					424.2501	4.16E+08	9.0		23187	
					364.2433	2.44E+07	0.5		26919	
					354.1745	1.46E+07	0.3		19945	
254.1612	b2	254.1592	-0.002	-7.9		1.33E+08	2.9		30816	
					243.1432	1.02E+07	0.2		30073	
237.1346	b2 -NH3	237.1328	-0.0018	-7.6		2.02E+07	0.4		30311	
					226.1528	1.34E+07	0.3		41377	
					226.1169	3.35E+07	0.7		29131	

Table S7. Observed isotopic distribution of 10 major fragments from the mono-isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
MONO	Mono	8.91E+07	100.00%	3.67E+08	100.00%	1.09E+09	93.97%	1.88E+09	99.47%	1.79E+09	100.00%	0	0.00%	100
	1 st	5.97E+07	67.00%	9.74E+07	26.54%	1.16E+09	100.00%	1.89E+09	100.00%	1.10E+09	61.39%	0	0.00%	0
	2 nd	5.13E+07	57.58%	2.63E+08	71.66%	6.98E+08	60.17%	1.93E+08	10.21%	2.26E+08	12.61%	0	0.00%	0
	3 rd	2.38E+07	26.71%	1.00E+08	27.25%	4.25E+07	3.66%	5.53E+07	2.93%	6.95E+07	3.88%	0	0.00%	0
c10	Mono	3.35E+08	100.00%	1.39E+09	100.00%	3.94E+09	100.00%	6.77E+09	100.00%	6.72E+09	100.00%	0	0.00%	100
	1 st	2.46E+08	73.43%	1.14E+09	82.01%	2.17E+09	55.08%	4.63E+08	6.84%	6.19E+08	9.21%	0	0.00%	0
	2 nd	8.81E+07	26.30%	3.12E+08	22.45%	6.05E+07	1.54%	1.03E+08	1.52%	1.46E+08	2.17%	0	0.00%	0
	3 rd	1.99E+07	5.94%	6.07E+07	4.37%	2.67E+07	0.68%	1.88E+07	0.28%	0.00E+00	0.00%	0	0.00%	0
b10	Mono	0.00E+00	0.00%	2.43E+07	100.00%	5.20E+07	100.00%	8.18E+07	100.00%	7.72E+07	100.00%	4.01E+07	100.00%	100
	1 st	0.00E+00	0.00%	1.38E+07	56.79%	2.12E+07	0.00%	2.03E+07	24.82%	0.00E+00	0.00%	0	0.00%	0
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	2.12E+07	0.00%	1.95E+07	23.84%	0.00E+00	0.00%	0	0.00%	0

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
c9	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	1.30E+07	15.89%	0.00E+00	0.00%	0	0.00%	0
	Mono	2.16E+08	100.00%	9.15E+08	100.00%	2.91E+09	100.00%	5.11E+09	100.00%	5.11E+09	100.00%	1.75E+09	100.00%	100
	1 st	1.43E+08	66.20%	6.44E+08	70.38%	1.32E+09	45.36%	3.09E+08	6.05%	4.37E+08	8.55%	0	0.00%	0
	2 nd	4.40E+07	20.37%	1.56E+08	17.05%	4.41E+07	1.52%	6.01E+07	1.18%	8.01E+07	1.57%	0	0.00%	0
z9	3 rd	0.00E+00	0.00%	3.15E+07	3.44%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	4.75E+07	100.00%	1.27E+08	100.00%	4.22E+08	100.00%	3.52E+08	100.00%	2.46E+08	100.00%	0	0.00%	100
	1 st	2.42E+07	50.95%	1.23E+08	96.85%	1.01E+08	23.93%	4.41E+07	12.53%	5.58E+07	22.68%	0	0.00%	0
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
c8	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	1.77E+08	100.00%	7.81E+08	100.00%	2.66E+09	100.00%	4.54E+09	100.00%	4.47E+09	100.00%	1.27E+09	100.00%	100
	1 st	1.14E+08	64.41%	5.26E+08	67.35%	1.12E+09	42.11%	2.72E+08	5.99%	3.64E+08	8.14%	0	0.00%	0
	2 nd	3.42E+07	19.32%	1.27E+08	16.26%	3.63E+07	1.36%	3.65E+07	0.80%	6.57E+07	1.47%	0	0.00%	0
C7	3 rd	0.00E+00	0.00%	1.89E+07	2.42%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	2.88E+08	100.00%	1.23E+09	100.00%	4.82E+09	100.00%	7.97E+09	100.00%	7.85E+09	100.00%	1.74E+09	100.00%	100
	1 st	1.46E+08	50.69%	7.16E+08	58.21%	1.65E+09	34.23%	4.21E+08	5.28%	5.27E+08	6.71%	0	0.00%	0
	2 nd	3.78E+07	13.13%	1.42E+08	11.54%	5.69E+07	1.18%	4.73E+07	0.59%	7.17E+07	0.91%	0	0.00%	0

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
c6	3 rd	0.00E+00	0.00%	1.88E+07	1.53%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	2.11E+08	100.00%	8.37E+08	100.00%	3.75E+09	100.00%	6.10E+09	100.00%	5.75E+09	100.00%	1.64E+09	100.00%	100
	1 st	8.66E+07	41.04%	4.02E+08	48.03%	1.01E+09	26.93%	2.43E+08	3.98%	3.30E+08	5.74%	0	0.00%	0
	2 nd	1.74E+07	8.25%	6.49E+07	7.75%	3.55E+07	0.95%	0.00E+00	0.00%	3.23E+07	0.56%	0	0.00%	0
c5	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	3.91E+08	100.00%	1.52E+09	100.00%	7.51E+09	100.00%	1.21E+10	100.00%	1.30E+10	100.00%	4.63E+09	100.00%	100
	1 st	1.33E+08	34.02%	6.38E+08	41.97%	1.71E+09	22.77%	3.78E+08	3.12%	5.85E+08	4.50%	0	0.00%	0
	2 nd	2.28E+07	5.83%	7.90E+07	5.20%	4.54E+07	0.60%	0.00E+00	0.00%	4.12E+07	0.32%	0	0.00%	0
c4	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
	Mono	1.83E+08	100.00%	6.78E+08	100.00%	3.24E+09	100.00%	5.40E+09	100.00%	5.89E+09	100.00%	3.09E+09	100.00%	100
	1 st	4.83E+07	26.39%	2.03E+08	29.94%	5.42E+08	16.73%	1.51E+08	2.80%	2.17E+08	3.68%	0	0.00%	0
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0

Table S8. Observed isotopic distribution of 10 major fragments from the 1st isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
MH ⁺	Mono	8.31E+07	100.00%	2.79E+08	100.00%	5.16E+08	43.00%	3.79E+08	16.92%	1.69E+08	6.60%	0	0.00%	0.015
	1 st	5.67E+07	68.23%	5.96E+07	21.36%	1.04E+09	86.67%	2.06E+09	91.96%	1.92E+09	75.00%	1.31E+08	100.00%	100
	2 nd	5.05E+07	60.77%	2.34E+08	83.87%	1.20E+09	100.00%	2.24E+09	100.00%	2.56E+09	100.00%	0	0.00%	0
	3 rd	2.58E+07	31.05%	1.28E+08	45.88%	2.60E+08	21.67%	2.21E+08	9.87%	3.85E+08	15.04%	0	0.00%	0
c10	Mono	3.10E+08	32.33%	1.06E+09	89.83%	1.76E+09	48.22%	1.83E+09	27.94%	9.42E+08	12.96%	0	0.00%	9.535
	1 st	2.47E+08	25.76%	1.18E+09	100.00%	3.65E+09	100.00%	6.55E+09	100.00%	7.27E+09	100.00%	8.60E+08	100.00%	100
	2 nd	9.59E+08	100.00%	4.27E+08	36.19%	6.46E+08	17.70%	3.68E+08	5.62%	7.61E+08	10.47%	0	0.00%	0
	3 rd	2.39E+07	2.49%	1.09E+08	9.24%	2.14E+07	0.59%	1.02E+08	1.56%	0.00E+00	0.00%	0	0.00%	0
b10	Mono	0.00E+00	0.00%	1.96E+07	100.00%	2.53E+07	87.85%	1.32E+07	36.26%	1.84E+07	48.81%	0	0.00%	10.16
	1 st	0.00E+00	0.00%	1.42E+07	72.45%	2.88E+07	100.00%	3.64E+07	100.00%	3.77E+07	100.00%	1.22E+07	100.00%	100
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	1.47E+07	38.99%	0	0.00%	0
	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	2.40E+07	63.66%	0	0.00%	0
c9	Mono	1.99E+08	100.00%	6.85E+08	100.00%	1.33E+09	60.18%	1.69E+09	42.46%	8.50E+08	18.93%	0	0.00%	21.659
	1 st	1.42E+08	71.36%	6.58E+08	96.06%	2.21E+09	100.00%	3.98E+09	100.00%	4.49E+09	100.00%	3.27E+08	100.00%	100
	2 nd	4.79E+07	24.07%	2.14E+08	31.24%	3.35E+08	15.16%	1.63E+08	4.10%	3.99E+08	8.89%	0	0.00%	0

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent FIRST	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
	3 rd	1.00E+07	5.03%	5.33E+07	7.78%	0.00E+00	0.00%	0.00E+00	0.00%	3.23E+07	0.72%	0	0.00%	0
z9	Mono	4.36E+07	100.00%	1.58E+08	100.00%	3.48E+08	52.65%	4.70E+08	37.30%	1.70E+08	12.59%	0	0.00%	22.984
	1 st	2.47E+07	56.65%	1.40E+08	88.61%	6.61E+08	100.00%	1.26E+09	100.00%	1.35E+09	100.00%	0	0.00%	100
	2 nd	0.00E+00	0.00%	2.87E+07	18.16%	2.09E+08	31.62%	3.23E+08	25.63%	4.44E+08	32.89%	0	0.00%	0
	3 rd	0.00E+00	0.00%	1.09E+07	6.90%	0.00E+00	0.00%	0.00E+00	0.00%	2.93E+07	2.17%	0	0.00%	0
c8	Mono	1.62E+08	100.00%	5.82E+08	100.00%	1.21E+09	65.05%	1.64E+09	48.24%	7.67E+08	20.18%	0	0.00%	26.863
	1 st	1.13E+08	69.75%	5.33E+08	91.58%	1.86E+09	100.00%	3.40E+09	100.00%	3.80E+09	100.00%	2.67E+08	100.00%	100
	2 nd	3.73E+07	23.02%	1.72E+08	29.55%	2.71E+08	14.57%	1.25E+08	3.68%	3.19E+08	8.39%	0	0.00%	0
	3 rd	0.00E+00	0.00%	3.15E+07	5.41%	0.00E+00	0.00%	0.00E+00	0.00%	2.34E+07	0.62%	0	0.00%	0
C7	Mono	2.64E+08	100.00%	8.76E+08	100.00%	2.40E+09	87.27%	3.77E+09	76.01%	2.01E+09	36.09%	0	0.00%	52.553
	1 st	1.44E+08	54.55%	7.09E+08	80.94%	2.75E+09	100.00%	4.96E+09	100.00%	5.57E+09	100.00%	3.05E+08	100.00%	100
	2 nd	4.12E+07	15.61%	1.89E+08	21.58%	3.26E+08	11.85%	1.29E+08	2.60%	3.82E+08	6.86%	0	0.00%	0
	3 rd	0.00E+00	0.00%	3.61E+07	4.12%	0.00E+00	0.00%	0.00E+00	0.00%	3.16E+07	0.57%	0	0.00%	0
c6	Mono	1.91E+08	100.00%	5.65E+08	100.00%	2.07E+09	100.00%	3.52E+09	100.00%	2.39E+09	70.09%	0	0.00%	91.29
	1 st	8.51E+07	44.55%	3.89E+08	68.85%	1.67E+09	80.68%	3.03E+09	86.08%	3.41E+09	100.00%	1.63E+08	100.00%	100
	2 nd	1.91E+07	10.00%	8.34E+07	14.76%	1.51E+08	7.29%	6.72E+07	1.91%	1.79E+08	5.25%	0	0.00%	0

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
	3 rd	0.00E+00	0.00%	1.16E+07	2.05%	0.00E+00	0.00%	0.00E+00	0.00%	1.27E+07	0.37%	0	0.00%	0
c5	Mono	3.53E+08	100.00%	9.94E+08	100.00%	4.62E+09	100.00%	8.42E+09	100.00%	7.44E+09	100.00%	0	0.00%	100
	1 st	1.29E+08	36.54%	6.03E+08	60.66%	2.87E+09	62.12%	5.07E+09	60.21%	5.72E+09	76.88%	4.57E+08	100.00%	78.898
	2 nd	2.42E+07	6.86%	1.03E+08	10.36%	2.09E+08	4.52%	1.02E+08	1.21%	2.39E+08	3.21%	0	0.00%	0
	3 rd	0.00E+00	0.00%	1.12E+07	1.13%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
c4	Mono	1.65E+08	100.00%	4.39E+08	100.00%	2.12E+09	100.00%	3.87E+09	100.00%	3.90E+09	100.00%	0	0.00%	100
	1 st	4.68E+07	28.36%	1.87E+08	42.60%	9.25E+08	43.63%	1.60E+09	41.34%	1.76E+09	45.13%	5.08E+08	100.00%	56.073
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	3.43E+07	1.62%	0.00E+00	0.00%	2.11E+07	0.54%	0	0.00%	0
	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0

Table S9. Observed isotopic distribution of 10 major fragments from the 2nd isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
SECOND	Mono	7.71E+07	100.00%	2.13E+08	100.00%	3.44E+07	5.21%	1.57E+08	31.78%	1.63E+08	18.38%	0	0.00%	0
	1 st	5.35E+07	69.39%	7.11E+07	33.38%	3.58E+08	54.24%	3.36E+08	68.02%	4.57E+08	51.52%	0	0.00%	0
	2 nd	4.86E+07	63.04%	1.89E+08	88.73%	6.60E+08	100.00%	2.48E+08	50.20%	8.87E+08	100.00%	0	0.00%	0
	3 rd	2.59E+07	33.59%	1.26E+08	59.15%	5.19E+08	78.64%	4.94E+08	100.00%	8.22E+08	92.67%	0	0.00%	0
c10	Mono	2.87E+08	100.00%	8.09E+08	76.32%	9.83E+08	65.97%	4.97E+08	39.76%	7.87E+08	39.75%	0	0.00%	0
	1 st	2.38E+08	82.93%	1.06E+09	100.00%	1.49E+09	100.00%	5.04E+08	40.32%	1.24E+09	62.63%	0	0.00%	0
	2 nd	9.63E+07	33.55%	4.41E+08	41.60%	1.22E+09	81.88%	1.25E+09	100.00%	1.98E+09	100.00%	0	0.00%	0
	3 rd	2.51E+07	8.75%	1.28E+08	12.08%	2.27E+08	15.23%	9.59E+07	7.67%	9.38E+07	4.74%	0	0.00%	0
b10	Mono	0.00E+00	0.00%	1.55E+07	100.00%	0.00E+00	0.00%	0.00E+00	0.00%	1.43E+07	76.88%	0	0.00%	0
	1 st	0.00E+00	0.00%	1.28E+07	82.58%	0.00E+00	0.00%	0.00E+00	0.00%	1.63E+07	87.63%	0	0.00%	0
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	1.86E+07	100.00%	0	0.00%	0
	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0
c9	Mono	1.84E+08	100.00%	5.19E+08	88.87%	1.07E+08	11.92%	3.61E+08	55.45%	5.61E+08	50.54%	0	0.00%	0
	1 st	1.37E+08	74.46%	5.84E+08	100.00%	8.98E+08	100.00%	3.42E+08	52.53%	9.33E+08	84.05%	0	0.00%	0
	2 nd	4.82E+07	26.20%	2.19E+08	37.50%	6.28E+08	69.93%	6.51E+08	100.00%	1.11E+09	100.00%	0	0.00%	0

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
SECOND	3 rd	1.07E+07	5.82%	6.07E+07	10.39%	7.12E+07	7.93%	0.00E+00	0.00%	3.18E+07	2.86%	0	0.00%	0
	Mono	4.00E+07	100.00%	1.19E+08	92.97%	4.63E+07	16.59%	1.18E+08	56.73%	1.28E+04	0.00%	0	0.00%	5.296
	1 st	2.42E+07	60.50%	1.28E+08	100.00%	2.79E+08	100.00%	1.21E+08	58.17%	3.02E+08	71.90%	0	0.00%	34.661
	2 nd	0.00E+00	0.00%	3.43E+07	26.80%	2.49E+08	89.25%	2.08E+08	100.00%	4.20E+08	100.00%	0	0.00%	100
z9	3 rd	0.00E+00	0.00%	1.40E+07	10.94%	4.88E+07	17.49%	0.00E+00	0.00%	4.97E+07	11.83%	0	0.00%	0
	Mono	1.50E+08	100.00%	4.43E+08	94.06%	1.23E+08	16.44%	3.13E+08	59.73%	5.07E+08	53.76%	0	0.00%	31.337
	1 st	1.09E+08	72.67%	4.71E+08	100.00%	7.48E+08	100.00%	3.03E+08	57.82%	8.51E+08	90.24%	0	0.00%	49.075
	2 nd	3.75E+07	25.00%	1.76E+08	37.37%	5.11E+08	68.32%	5.24E+08	100.00%	9.43E+08	100.00%	0	0.00%	100
c8	3 rd	0.00E+00	0.00%	4.10E+07	8.70%	3.76E+07	5.03%	0.00E+00	0.00%	2.85E+07	3.02%	0	0.00%	0
	Mono	2.42E+08	100.00%	6.44E+08	100.00%	4.18E+08	36.03%	5.64E+08	100.00%	1.03E+09	70.07%	0	0.00%	60.421
	1 st	1.38E+08	57.02%	6.16E+08	95.65%	1.16E+09	100.00%	5.12E+08	90.78%	1.47E+09	100.00%	0	0.00%	93.416
	2 nd	4.13E+07	17.07%	1.91E+08	29.66%	6.16E+08	53.10%	5.54E+08	98.23%	1.19E+09	80.95%	0	0.00%	100
C7	3 rd	0.00E+00	0.00%	4.21E+07	6.54%	6.50E+07	5.60%	0.00E+00	0.00%	2.85E+07	1.94%	0	0.00%	0
	Mono	1.74E+08	100.00%	4.00E+08	100.00%	5.15E+08	71.63%	2.69E+08	97.46%	9.24E+08	98.30%	0	0.00%	81.43
	1 st	8.15E+07	46.84%	3.34E+08	83.50%	7.19E+08	100.00%	1.76E+08	63.77%	9.40E+08	100.00%	0	0.00%	100
	2 nd	1.95E+07	11.21%	8.20E+07	20.50%	2.93E+08	40.75%	2.76E+08	100.00%	5.61E+08	59.68%	0	0.00%	64.522

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
SECOND	3 rd	2.00E+07	11.49%	1.38E+07	3.45%	4.09E+08	56.88%	0.00E+00	0.00%	1.64E+07	1.74%	0	0.00%	0
	Mono	3.22E+08	100.00%	6.89E+08	100.00%	1.50E+09	100.00%	4.08E+08	100.00%	2.70E+09	100.00%	0	0.00%	100
	1 st	1.24E+08	38.51%	5.11E+08	74.17%	1.32E+09	88.00%	0.00E+00	0.00%	1.95E+09	72.22%	0	0.00%	97.866
	2 nd	2.40E+07	7.45%	1.03E+08	14.95%	3.97E+08	26.47%	0.00E+00	0.00%	7.72E+08	28.59%	0	0.00%	45.728
c5	3 rd	0.00E+00	0.00%	1.19E+07	1.73%	2.30E+07	1.53%	0.00E+00	0.00%	1.63E+07	0.60%	0	0.00%	0
	Mono	1.50E+08	100.00%	3.01E+08	100.00%	8.08E+08	100.00%	6.32E+08	100.00%	1.41E+09	100.00%	0	0.00%	100
	1 st	4.45E+07	29.67%	1.56E+08	51.83%	4.60E+08	56.93%	3.67E+08	58.07%	6.38E+08	45.25%	0	0.00%	74.904
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	6.10E+07	7.55%	3.89E+07	6.16%	5.48E+07	3.89%	0	0.00%	24.758
c4	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	0

Table S10. Observed isotopic distribution of 10 major fragments produced from the 3rd isotopic peak compared to the theoretical distribution calculated with the IsoPatrn software (the closest are in green, incorrect patterns in red).

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical Isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
THIRD	Mono	6.10E+07	100.00%	1.20E+08	100.00%	2.45E+07	7.78%	9.50E+07	20.56%	9.49E+07	40.04%	0	0.00%	0
	1 st	4.48E+07	73.44%	7.13E+07	59.42%	5.86E+07	18.60%	3.08E+08	66.67%	1.79E+08	75.53%	0	0.00%	0
	2 nd	4.21E+07	69.02%	1.08E+08	90.00%	2.13E+08	67.62%	3.15E+08	68.18%	2.37E+08	100.00%	0	0.00%	0.033
	3 rd	2.40E+07	39.34%	1.02E+08	85.00%	3.15E+08	100.00%	4.62E+08	100.00%	1.35E+08	56.96%	0	0.00%	100
c10	Mono	2.25E+08	100.00%	4.63E+08	61.00%	6.32E+07	10.97%	5.09E+08	60.17%	3.90E+08	96.53%	0	0.00%	3.854
	1 st	2.08E+08	92.44%	7.59E+08	100.00%	2.60E+08	45.14%	8.46E+08	100.00%	4.04E+08	100.00%	0	0.00%	46.912
	2 nd	9.02E+07	40.09%	3.96E+08	52.17%	5.76E+08	100.00%	6.28E+08	74.23%	2.71E+08	67.08%	0	0.00%	24.715
	3 rd	2.57E+07	11.42%	1.39E+08	18.31%	3.92E+08	68.06%	5.95E+08	70.33%	2.84E+08	70.30%	0	0.00%	100
b10	Mono	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	4.182
	1 st	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	47.598
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	26.297
	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	100
c9	Mono	1.43E+08	100.00%	2.92E+08	70.87%	8.39E+07	28.63%	3.65E+08	69.92%	2.66E+08	65.84%	0	0.00%	12.209
	1 st	1.18E+08	82.52%	4.12E+08	100.00%	1.95E+08	66.55%	5.22E+08	100.00%	2.91E+08	72.03%	0	0.00%	66.585
	2 nd	4.50E+07	31.47%	1.94E+08	47.09%	2.93E+08	100.00%	3.85E+08	73.75%	1.47E+08	36.39%	0	0.00%	55.219

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical Isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
z9	3 rd	1.10E+07	7.69%	6.32E+07	15.34%	1.10E+08	37.54%	9.64E+07	18.47%	4.04E+08	100.00%	0	0.00%	100
	Mono	3.07E+07	100.00%	6.65E+07	72.92%	2.13E+07	17.90%	1.06E+08	62.72%	7.72E+07	88.33%	0	0.00%	1.229
	1 st	2.16E+07	70.36%	9.12E+07	100.00%	3.86E+07	32.44%	1.69E+08	100.00%	8.74E+07	100.00%	0	0.00%	10.694
	2 nd	0.00E+00	0.00%	3.67E+07	40.24%	1.19E+08	100.00%	1.49E+08	88.17%	7.57E+07	86.61%	0	0.00%	46.41
c8	3 rd	0.00E+00	0.00%	1.66E+07	18.20%	5.84E+07	49.08%	6.59E+07	38.99%	2.41E+07	27.57%	0	0.00%	100
	Mono	1.16E+08	100.00%	2.54E+08	77.20%	7.60E+07	31.28%	3.59E+08	84.27%	2.60E+08	100.00%	0	0.00%	17.777
	1 st	9.48E+07	81.72%	3.29E+08	100.00%	1.86E+08	76.54%	4.26E+08	100.00%	2.44E+08	93.85%	0	0.00%	80.115
	2 nd	3.52E+07	30.34%	1.55E+08	47.11%	2.43E+08	100.00%	3.22E+08	75.59%	1.30E+08	50.00%	0	0.00%	68.676
C7	3 rd	0.00E+00	0.00%	4.38E+07	13.31%	6.40E+07	26.34%	9.09E+07	21.34%	4.17E+07	16.04%	0	0.00%	100
	Mono	1.87E+08	100.00%	3.46E+08	83.37%	1.95E+08	58.38%	8.14E+08	100.00%	5.05E+08	100.00%	0	0.00%	37.531
	1 st	1.19E+08	63.64%	4.15E+08	100.00%	3.34E+08	100.00%	7.48E+08	91.89%	3.83E+08	75.84%	0	0.00%	100
	2 nd	3.84E+07	20.53%	1.65E+08	39.76%	3.05E+08	91.32%	4.66E+08	57.25%	1.90E+08	37.62%	0	0.00%	86.978
c6	3 rd	0.00E+00	0.00%	4.42E+07	10.65%	9.54E+07	28.56%	8.24E+07	10.12%	4.67E+07	9.25%	0	0.00%	67.464
	Mono	1.32E+08	100.00%	1.98E+08	90.00%	2.10E+08	78.07%	6.85E+08	100.00%	3.86E+08	100.00%	0	0.00%	56.47
	1 st	6.96E+07	52.73%	2.20E+08	100.00%	2.69E+08	100.00%	5.07E+08	74.01%	2.48E+08	64.25%	0	0.00%	100
	2 nd	1.77E+07	13.41%	6.85E+07	31.14%	1.59E+08	59.11%	2.92E+08	42.63%	9.71E+07	25.16%	0	0.00%	72.335

Non-Uniform Sampling Acquisition of two-dimensional (2D) FT-ICR MS
for increased mass resolution of MS/MS precursor ions

Parent	frag isotope	NUS1		NUS2		NUS4		NUS8		NUS16		NUS32		Theoretical Isotopic ratio %
		Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	Intensity	Isotopic ratio %	
c5	3 rd	0.00E+00	0.00%	1.49E+08	67.73%	0.00E+00	0.00%	3.80E+07	5.55%	2.12E+07	5.49%	0	0.00%	34.411
	Mono	2.43E+08	100.00%	3.28E+08	100.00%	5.02E+08	94.72%	1.56E+09	100.00%	7.95E+08	100.00%	0	0.00%	73.667
	1 st	1.05E+08	43.21%	3.28E+08	100.00%	5.30E+08	100.00%	9.95E+08	63.78%	4.58E+08	57.61%	0	0.00%	100
	2 nd	2.19E+07	9.01%	8.68E+07	26.46%	2.30E+08	43.40%	4.53E+08	29.04%	1.54E+08	19.37%	0	0.00%	57.959
c4	3 rd	0.00E+00	0.00%	1.09E+07	3.32%	3.86E+07	7.28%	4.33E+07	2.78%	2.45E+07	3.08%	0	0.00%	20.003
	Mono	1.14E+08	100.00%	1.38E+08	100.00%	2.66E+08	100.00%	6.84E+08	100.00%	3.09E+08	100.00%	0	0.00%	98.253
	1 st	3.74E+07	32.81%	9.67E+07	70.07%	1.99E+08	74.81%	4.07E+08	59.50%	1.13E+08	36.57%	0	0.00%	100
	2 nd	0.00E+00	0.00%	0.00E+00	0.00%	3.54E+07	13.31%	4.33E+07	6.33%	2.22E+07	7.18%	0	0.00%	44.152
c4	3 rd	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0.00E+00	0.00%	0	0.00%	10.77

Table S11. Theoretical fragmentation of the major sesame oil triglycerides cationized by ${}^6\text{Li}^+$ and fragmented by IRMPD.

Common name	Abbreviated formula	Molecular formula	m/z cationized by ${}^6\text{Li}^+$	Loss of oleic acid acid (-O) ^a		Loss of linolenic acid acid (-L) ^b	
				Molecular formula	m/z cationized by ${}^6\text{Li}^+$	Molecular formula	m/z cationized by ${}^6\text{Li}^+$
Triolein	OOO	$\text{C}_{57}\text{H}_{104}\text{O}_6$	890.798	$\text{C}_{39}\text{H}_{70}\text{O}_4$	608.542	-	-
	OOL or OLO	$\text{C}_{57}\text{H}_{102}\text{O}_6$	888.782	$\text{C}_{39}\text{H}_{68}\text{O}_4$	606.526	$\text{C}_{39}\text{H}_{70}\text{O}_4$	608.542
	OLL or LOL	$\text{C}_{57}\text{H}_{100}\text{O}_6$	886.767	$\text{C}_{39}\text{H}_{66}\text{O}_4$	604.511	$\text{C}_{39}\text{H}_{68}\text{O}_4$	606.526
Trilinolein	LLL	$\text{C}_{57}\text{H}_{98}\text{O}_6$	884.751	-	-	$\text{C}_{39}\text{H}_{66}\text{O}_4$	604.511

^aO: oleic acid ($\text{C}_{18}\text{H}_{34}\text{O}_2$)

^bL: Linolenic acid ($\text{C}_{18}\text{H}_{32}\text{O}_2$)

V. Publication de service

1. Publication de service 1

Barbe, C.; Bray, F.; Gueugneau, M.; Devassine, S.; Lause, P.; Tokarski, C.; Rolando, C.; Thissen, J.-P. Comparative Proteomic and Transcriptomic Analysis of Follistatin-Induced Skeletal Muscle Hypertrophy. *Journal of Proteome Research* 2017, **16**(10), 3477–3490.

Comparative Proteomic and Transcriptomic Analysis of Follistatin-Induced Skeletal Muscle Hypertrophy

Caroline Barbé,^{*,†} Fabrice Bray,[‡] Marine Gueugneau,[†] Stéphanie Devassine,[‡] Pascale Lause,[†] Caroline Tokarski,[‡] Christian Rolando,[‡] and Jean-Paul Thissen[†]

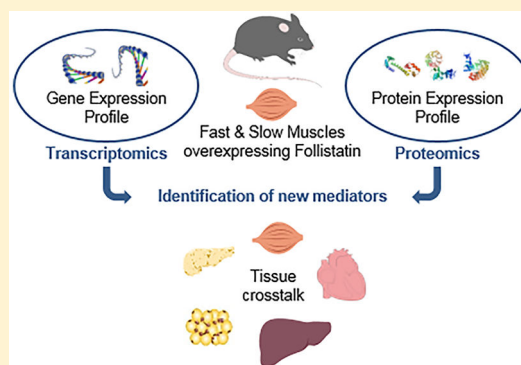
[†]Pole of Endocrinology, Diabetes and Nutrition, Institute of Experimental and Clinical Research, Université Catholique de Louvain, 1200 Brussels, Belgium

[‡]Miniaturisation pour la Synthèse, l'Analyse & la Protéomique (MSAP), CNRS, USR 3290, Université de Lille; Biochimie Structurale & Fonctionnelle des Assemblages Biomoléculaires, CNRS, FR 3688, FRABIO, Université de Lille and Institut Eugène-Michel Chevreul, CNRS, FR 2638, Université de Lille, 59000 Lille, France

S Supporting Information

ABSTRACT: Skeletal muscle, the most abundant body tissue, plays vital roles in locomotion and metabolism. Myostatin is a negative regulator of skeletal muscle mass. In addition to increasing muscle mass, Myostatin inhibition impacts muscle contractility and energy metabolism. To decipher the mechanisms of action of the Myostatin inhibitors, we used proteomic and transcriptomic approaches to investigate the changes induced in skeletal muscles of transgenic mice overexpressing Follistatin, a physiological Myostatin inhibitor. Our proteomic workflow included a fractionation step to identify weakly expressed proteins and a comparison of fast versus slow muscles. Functional annotation of altered proteins supports the phenotypic changes induced by Myostatin inhibition, including modifications in energy metabolism, fiber type, insulin and calcium signaling, as well as membrane repair and regeneration. Less than 10% of the differentially expressed proteins were found to be also regulated at the mRNA level but the Biological Process annotation, and the KEGG pathways analysis of transcriptomic results shows a great concordance with the proteomic data. Thus this study describes the most extensive omics analysis of muscle overexpressing Follistatin, providing molecular-level insights to explain the observed muscle phenotypic changes.

KEYWORDS: *follistatin, myostatin, skeletal muscle hypertrophy, microarray, mass spectrometry*



■ INTRODUCTION

Myostatin (Mstn), a member of the TGF- β family primarily expressed in skeletal muscle, is a negative regulator of skeletal muscle mass, as shown by the marked increase in skeletal muscle mass following Mstn inhibition or gene deletion.^{1,2} Increasing size and strength of skeletal muscle represents a promising therapy for muscular disorders. To date, however, pharmacologic strategies to treat muscle disorders have produced modest clinical benefits. Therefore, inhibition of the Mstn pathway represents a hopeful strategy to improve functional outcomes in mice^{3–5} and, above all, in humans.^{6–8} The muscle mass increase induced by Mstn inhibition is associated with several positive physiological changes, such as increased insulin sensitivity,⁹ but also greater muscle fatigability with reduced exercise tolerance.^{10–12} In this regard, Mstn inhibitors could be considered as deleterious. However, several works have demonstrated the beneficial effect of Mstn inhibition in muscle dystrophies,^{3,8,13,14} cancer cachexia,^{4,15} sarcopenia,^{5–7,16} and states of insulin resistance.^{9,17} Considering these results, we may speculate about the potential benefits that can be expected from the use of Mstn inhibitors in

different pathological conditions. Although the physiological and metabolic features induced by Mstn inhibition have been well investigated, the molecular mechanisms responsible for these phenotypic changes are still poorly understood, impeding the development of more targeted therapeutic strategies.

Among Mstn inhibitors, Follistatin (FS) has been shown to bind and antagonize Mstn, leading to a dramatic increase in muscle mass.^{18,19} The muscle fiber hypertrophy induced by FS results from increased protein synthesis and requires the activation of Smad1/5 and the type I IGF receptor (IGF-IR)/Akt/mTOR pathway.^{20,21} Therefore, FS itself or other Mstn inhibitors may represent therapeutic strategies as treatments for reversing muscle wasting (cancer cachexia, sarcopenia) and for treating obesity and the metabolic syndrome.

To gain new insights into the molecular mechanisms responsible for the phenotype caused by FS, we combined transcriptomic and proteomic analyses of skeletal muscles from muscle-specific FS transgenic (mTrFS) mice. In this study, we

Received: February 8, 2017

Published: August 15, 2017

used a deep label-free proteomic approach to quantitate the relative proteomic changes in muscles from mTrFS mice. Indeed, our proteomic workflow included a fractionation step to facilitate identification of weakly expressed proteins. Moreover, we compared the proteomic profile from two phenotypically different muscles, namely, the tibialis anterior (TA) and the soleus (SOL) muscles, overexpressing FS. In addition, a microarray analysis was used to investigate transcriptional changes occurring in TA muscle from mTrFS mice.

■ EXPERIMENTAL PROCEDURES

Animals

The study was conducted in accordance with the directives of and approved by the Animal Ethics Committee of the Catholic University of Louvain (Brussels, Belgium). Analyses were carried out on 6-week-old C57BL/6 transgenic male mice (mTrFS) overexpressing a human FS288 (hFS288) short form specifically in skeletal muscle¹⁹ and their control wild-type littermates (WT), kindly provided by S. J. Lee (John Hopkins University). Mice were housed individually under controlled conditions of lighting (12 h light, 12 h dark cycle) and temperature (22 ± 2 °C). They were allowed free access to chow and water. Animal were anesthetized with a mixture of 100 mg/kg ketamine (Nimatek; Eurovet Animal Health) and 15 mg/kg xylazine hydrochloride (Rompun; Bayer) administered by ip injection after an overnight fasting. They were killed by decapitation, and blood was collected from the trunk vessels. Hind limb muscles were dissected and quickly frozen in liquid nitrogen, followed by storage at -80 °C.

Experimental Design

For the proteomic analyses, we used three animals per genotype (mTrFS and WT), from which two different muscles (TA and SOL) were harvested. Proteins from each of these muscles were extracted into a soluble fraction (SF) and a solubilized fraction (sF), and each fraction was digested in peptides with trypsin. Finally, each peptides fraction ($n = 24$) was injected three times in the LC-MS/MS instrument. In other words, we performed LC-MS/MS mass spectrometry analysis in triplicate on three biological replicates (Figure 2). For transcriptomic analyses, we used three animals per genotype (mTrFS and WT), from which TA muscles were harvested. Total RNA from each of these muscles was extracted and analyzed by microarray in simplicite.

Protein Extraction

Frozen muscles, previously pestled in liquid nitrogen, were homogenized (50 mg/mL) with Ultraturrax (IKA Labortechnik) in ice-cold buffer containing 40 mM Tris (pH 7.0), 2 mM EDTA, and 10% protease inhibitor cocktail (Roche Applied Science), according to Sayd et al.²² After centrifugation at 4 °C for 10 min at 10 000g, the supernatants, referred to as SF, were immediately aliquoted and stored at -80 °C. The insoluble pellets were resuspended in solubilization buffer containing 50 mM Tris (pH 7.0) and 8 M urea, regularly shaken for 90 min on ice, and centrifuged at 4 °C for 5 min at 5000g. The supernatants, referred to as sF, were immediately aliquoted and stored at -80 °C.

Protein Digestion

The quantity of proteins was measured with BCAssay method. Fifty micrograms of proteins from each fraction was processed in the 10 kDa filtration units (Millipore) using eFASP protocol

(enhanced Filter-Aided Samples Preparation) protocol.²³ Trypsin was used for digestion of proteins with a ratio 1/50. Peptides were extracted and evaporated before being dissolved in 10 μ L of water Milli-Q. The complete digestion eFASP procedure is described in the [Supplemental Methods](#). Samples were then diluted 10-fold in buffer A of nano-HPLC (5% acetonitrile and 0.1% formic acid), and each sample ($n = 24$) was injected three times in HPLC instrument to be analyzed in triplicate.

Nano-LC-MS/MS Analysis

Peptide mixtures were analyzed using a nanoflow HPLC instrument (U3000 RSLC Thermo Fisher Scientific) coupled online to a quadrupole-Orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) with a nanoelectrospray ion source. One μ L of peptide mixture (corresponding to 500 ng of proteins) was loaded onto the preconcentration trap (Thermo Scientific, Acclaim PepMap100 C18, 5 μ m, 300 μ m i.d. \times 5 mm) using partial loop injection for 5 min at a 10 μ L/min flow rate with buffer A. Peptides were separated on analytical column (Acclaim PepMap100 C18, 3 μ m, 75 mm i.d. \times 500 mm) with a linear gradient of 5–50% buffer B (75% acetonitrile and 0.1% formic acid) at a 250 nL/min flow rate and a 45 °C controlled temperature. The total time for a LC-MS/MS run was \sim 240 min. MS data were acquired in a top-20 data-dependent method dynamically choosing the most abundant precursor ions from the survey scan for HCD fragmentation.²⁴ Dynamic exclusion duration was 60 s. Isolation of precursors was performed with a 1.6 m/z window, and MS/MS scans were acquired with a starting mass of 80 m/z . Survey scans were acquired in the Orbitrap analyzer with mass range of 400–1600 m/z and 70 000 resolutions at m/z 400 (AGC set to 10^6 ions with a maximum fill time of 100 ms). Resolution for HCD spectra was set to 35 500 at m/z 200 (AGC set to 10^5 ions with a maximum fill time of 120 ms) and 28 eV normalized collision energy. The underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.4%. The instrument was run with peptide recognition mode (i.e., from two to eight charges), exclusion of singly charged and of unassigned precursor ions enabled.

Data Analysis

The acquired raw files were analyzed with MaxQuant software (version 1.5.3.30) using the Andromeda search engine. Peptides and proteins were identified by using the mouse subset of the UniProt database (June 2016, 49 870 entries) supplemented with 245 frequently observed contaminants and forward/reverse sequences. The “target-decoy” search strategy was used for estimating the frequencies of incorrect protein identifications (FDR) based on a reverse database generated automatically in MaxQuant. The precursor mass and fragment mass were identified with an initial mass tolerance of 10 and 20 ppm, respectively. The search included variable modifications of methionine and proline oxidation, asparagine and glutamine deamidation, tyrosine, serine and threonine phosphorylation and N-terminal acetylation and glutamine to pyroglutamate conversion, and fixed modifications of carbamidomethyl cysteine. Minimal peptide length was set to six amino acids, and a maximum of three miscleavages was allowed. The FDR was set to 0.05 for peptide and protein identifications. To maximize the number of quantification events across samples, MS runs from skeletal muscle were analyzed with the “match between runs” option in the MaxQuant software, which allowed

Table 1. Sequences of Primers (5'-3')

primers ^a	forward	reverse
EGF	TGGACTGGATTGGCCGGAG	CCTTGGATGCACAGCTATTCCCTCG
FGF6	GTAGAACGGGTGTGGTGGAG	GGAAGTGCATTTCGTCGTGG
IGF1	GCTATGGCTCCAGCATTCCG	TCCGGAAGCAACACTCATCC
IGFBP5	TCCGAACAAGGCCCTGCCG	GCTGTGCAAGCGGTGGCACT
IL12a	CAAGGGCATGCTGGTGGCCA	GTCACGACGCGGGTGTGAA
LGALS1	GCTTCAATCATGGCCTGTGG	CACAATGGTGTGGCGTCTC
NOS1	CCCCTGGGCGGGGAGAATGAT	CCGAGGCGGGAGACTGTTCG
OGN	AATGATGAAATGCCACATGCCTGT	TGGCATGTCTGCAAAATCCTTGGCA
PAK1	TCCGCCAGATGCTTTGACCCG	AATGGCCACTCTGCCCTGT
SLN	TGTGCCCTGCTCCTCTTCAGGAAG	GGGCACACAGCAGTCACTCCC
WNT5a	GCTTAACCCGACGCTTCGCT	AATGGGCTTCTTCATGGCGAGG
GAPDH	TGCACCACCAACTGCTTA	GGATGCAGGGATGATGTTT

^aEGF: Epidermal Growth Factor; FGF6: Fibroblast Growth Factor 6; IGF1: Insulin-like Growth Factor 1; IGFBP5: Insulin-like Growth Factor-Binding Protein 5; IL12a: Interleukin-12 subunit alpha; LGALS1: Galectin-1; NOS1: Nitric Oxide Synthase 1; OGN: Osteoglycin; PAK1: p21-activated Kinase 1; SLN: Sarcoplin; WNT5a: Wnt family member 5A; GAPDH: Glyceraldehyde-3-phosphate Dehydrogenase.

the quantification of high-resolution MS1 features that were not identified in each single measurement. This algorithm was enabled using a 120 s retention time window for individual matching and a 30 min retention time window for complete alignment of the spectrum. In the case of identified peptides that are all shared between two proteins, these were combined and reported as one protein group. Moreover, protein contaminants, proteins identified only based on variable modifications sites and proteins matching to the reverse decoy database, were filtered out. For the proteins that were identified with single peptide and found as differentially expressed between mTrFS and WT mice, each single peptide was validated by BlastP interrogation provided in [Supplemental Table S1](#). Moreover, for all of these proteins, detailed information about the MS/MS spectrum, peptide sequence, and precursor *m/z* is provided in [Supplemental Figure S1](#). LFQ intensities for respective protein groups were uploaded in Perseus (version 1.5.6.0) and analyzed.²⁵ Raw LFQ intensities were logarithmized by Log₂. At least three LFQ values per protein group needed to be present for the analysis. To replace nonquantified values with low intensities, data imputation was performed based on normal distribution of LFQ intensities. Significant interactors were determined using a two-sample analysis *t* test and multiple sample tests with Benjamini-Hochberg FDR at 0.05. Normalization with Z-score was determined, and hierarchical clustering was constructed with Euclidean distance.

Western Blots Analysis

Equal amounts of proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred to PVDF membranes. Membranes were probed with the following primary antibodies: anti-ACADM (1:1000; Genetex), anti-eEF1A2 (1:1000; Genetex), anti-LGALS1 (1:1000; Genetex), and anti-GAPDH (1:20 000; Cell Signaling Technology). Then, membranes were incubated with a horseradish peroxidase-coupled secondary antibody (Cell Signaling Technology) and developed using Enhanced Chemiluminescence (ECL) Western blotting Detection System Plus (GE Healthcare). Developed films were scanned and analyzed as previously described.²⁶ All results were normalized for the signal to GAPDH protein.

RNA Extraction and Transcriptomic Analysis

Total RNA was isolated from the frozen TA muscle using TRIzol reagent, as described by the manufacturer. RNA was quantified with UV absorption at 260 nm with a NanoDrop Spectrophotometer, and its integrity was assessed with the RNA 6000 Nano Chip on the Agilent 2100 Bioanalyzer (Agilent). RNA target for microarray analysis was prepared with The GeneChip 3' IVT Express Kit according to the manufacturer's instructions. GeneChip Mouse Genome 430A 2.0 (Affymetrix) was used for the hybridization. Results were analyzed using the GCOS suite software and normalized using the control condition as baseline for the comparison. Using MeV, the Significant Analysis of Microarray (SAM) algorithm was applied to detect the significant probe sets differentially expressed between the two conditions ($FC > 1.4$ and P value < 0.05).

mRNA Analysis by Real-Time Quantitative (RTQ)-PCR

cDNA for quantitative real-time polymerase chain reaction (RT-qPCR) analysis was synthesized from 1 μ g of total RNA, as previously described.²⁷ Accession numbers for the sequences and primers used were: EGF (NM_010113), FGF6 (NM_010204.1), IGF1 (NM_178866), IGFBP5 (NM_010518.2), IL12a (NM_008351.2), LGALS1 (NM_008495.2), NOS1 (NM_008712.2), OGN (NM_008760.4), PAK1 (NM_011035.2), SLN (NM_025540.2), WNT5a (NM_022631.1), and GAPDH (NM_001289726.1) used as the reporter gene. Primers sequences are reported in [Table 1](#). Primers were tested to avoid primer dimers, self-priming formation, or unspecific amplification. The primers were designed to have standardized optimal PCR conditions. We have confirmed that the expression values of GAPDH normalized to RNA were not affected by FS overexpression.

Functional Annotation

The differentially expressed proteins were classified using the Gene Ontology categories "Cellular Component" and "Biological Process" extracted from the Uniprot database. The Gene Annotation Co-occurrence Discovery (GeneCodis) web-tool was used to analyze the differentially expressed genes. GeneCodis is a function analysis tool which integrates different information resources (GO, Panther pathways, SwissProt, etc.).^{28–30} Genes of interest, defined as at least 1.4-fold differentially expressed according to microarray analysis, were

uploaded to examine the biological processes and KEGG pathways significantly regulated.

Statistical Analysis

For statistical analysis not related to the proteomic and transcriptomic data sets, results were expressed relatively to WT mice and are presented as means \pm standard error of the mean (SEM), biological replicates being stated in Figure legends. Unpaired Student's *t* test was used to assess statistical differences among means with a significant threshold set for a *P*-value < 0.05 .

RESULTS AND DISCUSSION

FS Overexpression Causes Hypertrophy of Fast Skeletal Muscles

Without any change in the total body weight (Figure 1A), the mTrFS mice are characterized by marked skeletal muscle

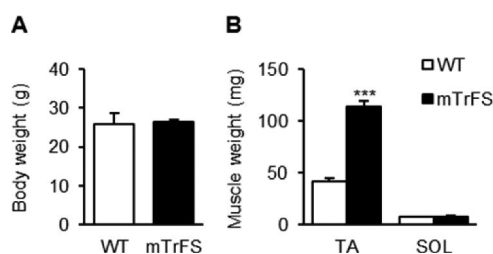


Figure 1. Effect of muscle-specific FS overexpression on mice skeletal muscles. Body weight (A) and TA and SOL muscles weight (B) of 6-week-old C57Bl/6 mTr-FS versus WT male mice ($n = 3$ /group). Results are expressed as means \pm SEM. Statistical analysis was performed using unpaired *t* test (***, $P < 0.001$).

hypertrophy of TA muscle, while the SOL muscle mass does not change (TA: +176%, $P < 0.001$ and SOL: +6%, $P > 0.05$, Figure 1B). Contrary to the TA muscle, which is a predominantly fast contracting muscle with a large number of type IIB fibers (60%), the SOL muscle is a slow contracting muscle containing mainly type I fibers (40–60%).³¹ It is worth noting that while hypertrophy of fast muscles has been well documented in mice treated with Mstn inhibitors such as FS or soluble Activin type IIB Receptor (sActRIIB),^{21,32} most studies reported a smaller hypertrophy for SOL muscle.^{33–36} An intriguing possibility is that the fast muscles would be more responsive to the effects of Mstn inhibitors due to higher basal expression levels of Mstn^{37,38} and its receptor ActRIIB.^{34,36} Alternatively, it is possible that the SOL hypertrophy appears only at a later age, as suggested by some studies.^{35,34} Thereby, these factors could explain the absence of hypertrophy of the SOL in this early stage of age.

Fractionation and Proteomic Workflow

A major challenge in the proteomic analysis of skeletal muscle lies in the broad dynamic range of protein abundance. To date, only three studies have examined the muscle proteomic changes in Mstn loss-of-function models,^{39–41} all of them having been performed on the whole proteome. However, the skeletal muscle expresses a relatively small number of myofibrillar proteins representing $>50\%$ of its whole protein content.⁴² Because of their high abundance, myofibrillar proteins may therefore hamper the detection of low abundance proteins. Recent reports on proteomic analysis emphasize that fractionation methods help to detect and identify a larger number of weakly expressed proteins.^{43–45} For these reasons,

we separated whole muscle into SF, mainly composed of sarcoplasmic proteins, and sF, more concentrated in myofibrillar proteins, the two fractions being separately submitted to the proteomic analyses (Figure 2).

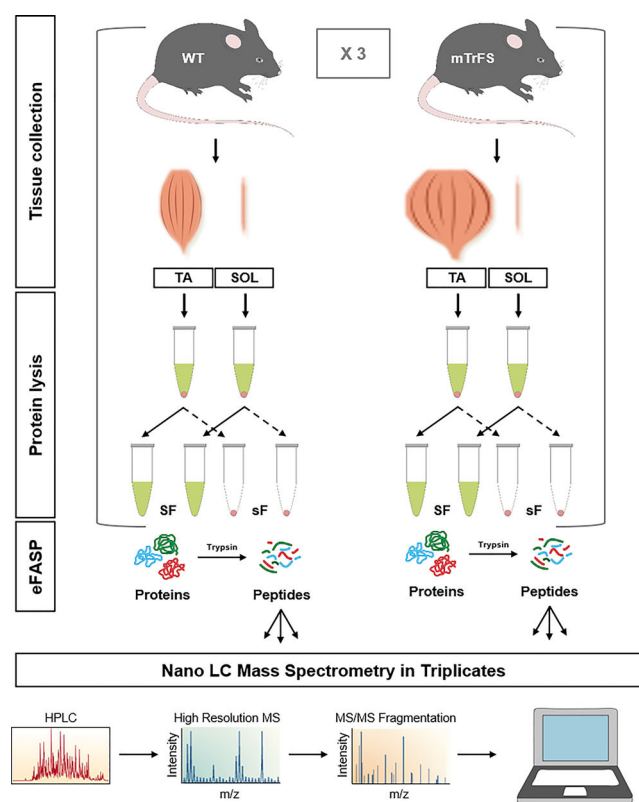


Figure 2. Simplified scheme of the mass-spectrometry-based proteomic workflow. Protein homogenates from TA and SOL muscles were separated into SF and sF as described in the “Experimental Procedures”. All fractions ($n = 24$) were digested on FASP filters, and peptides were analyzed in triplicates by LC–MS/MS on a Q Exactive mass spectrometer. HPLC, MS, and MS/MS diagrams courtesy of A. Heck and R. Aebersold.

The amount of proteins extracted was similar between the 3 mTrFS and 3 WT mice in the two fractions (SF: 59.7 ± 3.5 mg/g and 61.1 ± 0.9 mg/g, respectively; sF: 107.2 ± 9.4 mg/g and 113.2 ± 5.4 mg/g, respectively, in the TA muscle), indicating that extractability of SF and sF fractions was not affected by FS overexpression. As reported in the schematic workflow (Figure 2), we performed a proteomic analysis on one predominantly fast muscle (TA) and one slow muscle (SOL). Therefore, our study constitutes the first proteomic analysis investigating the proteome profile changes induced by FS in the presence of hypertrophy (TA) and in its absence (SOL).

For each fraction (SF and sF) of each muscle (TA and SOL) of the three animals per genotype, we performed LC–MS/MS mass spectrometry analysis in triplicate. The quality of our analysis is highlighted by the high technical reproducibility of our experimental approach and the low biological variability of our muscle samples. Indeed, the coefficients of variation (CV) are $<1\%$ for mTrFS and WT mice in the two fractions both for technical (SF: 0.6 and 0.8%, respectively; sF: 0.3 and 0.7%, respectively, in the TA muscle) and biological (SF: 0.4 and 0.7%, respectively; sF: 0.4 and 0.4%, respectively, in the TA

muscle) replicates (Figure S2A). Moreover, as expected, a Principal Component Analysis (PCA) revealed clear separation in the proteomic profile of slow SOL muscle compared with the fast TA muscle (Figure S2B).

We identified on average 1021 nonredundant proteins in TA and SOL muscles. The total number of identified proteins was not affected by the FS overexpression or by the muscle nature (Figure 3A). Among the proteins detected, ~14% were found

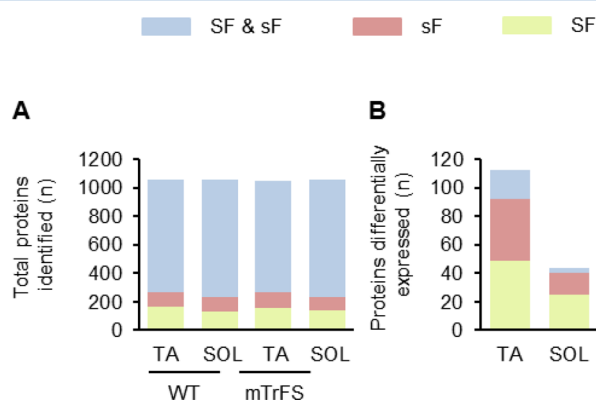


Figure 3. Global and comparative proteomic analysis of skeletal muscles from mTrFS and WT Mice. (A) Total number of proteins identified for each muscle in WT and mTrFS mice and (B) number of proteins differentially expressed in SF and sF for TA and SOL muscles. Proteins identified in the SF are shown in green, those in the sF are shown in red, and those found both in SF and sF are in blue.

only in the SF, 10% were found only in the sF, and 76% were identified in both fractions. It should be noted that although the separated analysis did not increase the total number of proteins identified in comparison with analysis performed on the whole muscle (data not shown), fractionation significantly increased the number of proteins identified as differentially expressed between mTrFS and WT mice.

Proteomic Signature of Slow and Fast Muscle Overexpressing FS

The results of the LFQ analyses performed on the SF and sF were compared and gathered, avoiding duplication. It is worth noting that directional changes were concordant between the two fractions for >95% of the proteins detected in both fractions and reported as differentially expressed between mTrFS and WT mice. In case of discrepancy, the amplitude of the changes is reported for the fraction where the protein was the most expected, based on the LFQ intensity, the MaxQuant score, and the known protein localization. Thereby, the comparative proteomic analysis indicated that 115 (Table S2) and 44 (Table S3) proteins were significantly differentially expressed in TA and SOL muscles, respectively, between mTrFS and WT mice (Figures 3B and 4).

Interestingly, the number of FS-regulated proteins was not correlated to the total number of proteins identified. Indeed, despite a similar number of identified proteins (Figure 3A), fewer proteins were found to be significantly differentially expressed (Figure 3B) in the nonhypertrophied SOL muscle compared with the hypertrophied TA muscle. This observation indicates that the effect of Mstn inhibition on muscle proteome is muscle-type-dependent, like changes in muscle mass, and that Mstn may affect the muscle proteome independently of its effect on muscle mass. Comparing analyses performed on the two different muscles, we observed that only 13 proteins were

identified, as regulated by FS overexpression in both TA and SOL muscles. From these proteins, β -enolase 3 (ENO3), Obscurin (OBSCN), and the 60 kDa heat shock protein (HSPD1) were upregulated in TA muscle, but downregulated in SOL muscle. Interestingly, it has been reported that the glycolytic enzyme ENO3 might have an auxiliary role in transcription and DNA replication or repair and appears to have a function in striated muscle development and regeneration.⁴⁶ Moreover, several works have reported a decrease of its expression in denervated⁴⁷ and in aging muscle in rat.^{48,49} However, whether these protein changes explain the differential hypertrophic response of TA and SOL muscles to Mstn inhibition is unknown.

Proteomic Changes Related to Muscle Phenotype Induced by Mstn Inhibition

GO annotations, namely, the Biological Processes (BP) and Cellular Components (CC) of the proteins that we found regulated by FS (Tables S2 and S3), have been used to unravel the altered proteins regulating the most relevant BP, getting a potential impact on phenotypic changes previously described in response to FS. The correlations between proteome and physiological changes are discussed thereafter.

Changes in Energy Metabolism. The energy metabolism appears to be the most affected biological process in response to FS overexpression in skeletal muscle, with 39% of the FS-regulated proteins in both TA and SOL muscles. Our analysis shows that FS overexpression in TA muscle leads to a strong increase in components of the glycolytic metabolism but a reduction in the abundance of proteins involved in mitochondrial lipid and oxidative metabolisms (Table 2 and Figure 5). Among the proteins involved in fatty acid β -oxidation, we confirmed, by Western blotting analysis, the decreased abundance of ACADM, responsible for the oxidation of medium-chain fatty acids (Figure 6A,B).

Altogether, these proteome changes may contribute to the increased muscle glucose metabolism observed in genetic mouse models of Mstn inhibition. Indeed, it has been demonstrated that Mstn inhibition increases muscle insulin-stimulated glucose uptake and glycogen storage,^{50,51} with the energy balance being equilibrated by a reduction in lipid metabolism.^{9,52,53} Moreover, the shift from fatty acid toward glucose consumption is reflected by an increase in the respiratory exchange ratio also reported in Mstn KO mice.¹¹ Interestingly, the reduction in the oxidative metabolism is consistent with mitochondria depletion, reduced aerobic metabolism, and decrease in PGC1 α levels that we and others have observed in Mstn KO muscles.^{40,54–57} Moreover, this metabolic deficit agrees with the increased fatigability, illustrated by the lower endurance run capacity of the Mstn KO mice.⁵⁸ Finally, because blockade of mitochondrial respiration in skeletal muscle results in decreased tetanic force generation,^{59,60} the mitochondrial impairment that we observed may explain the decrease in specific force induced by Mstn inhibition.^{10,54,61}

The SOL muscle from mTrFS mice is characterized by similar changes in its energetic homeostasis (Table 2). However, in contrast with fast muscles, the SOL muscle shows a surprising lower abundance in glycogenolytic and glycolytic enzymes. This observation raises an interesting question about the coverage of energy requirement in slow muscles following Mstn inhibition.

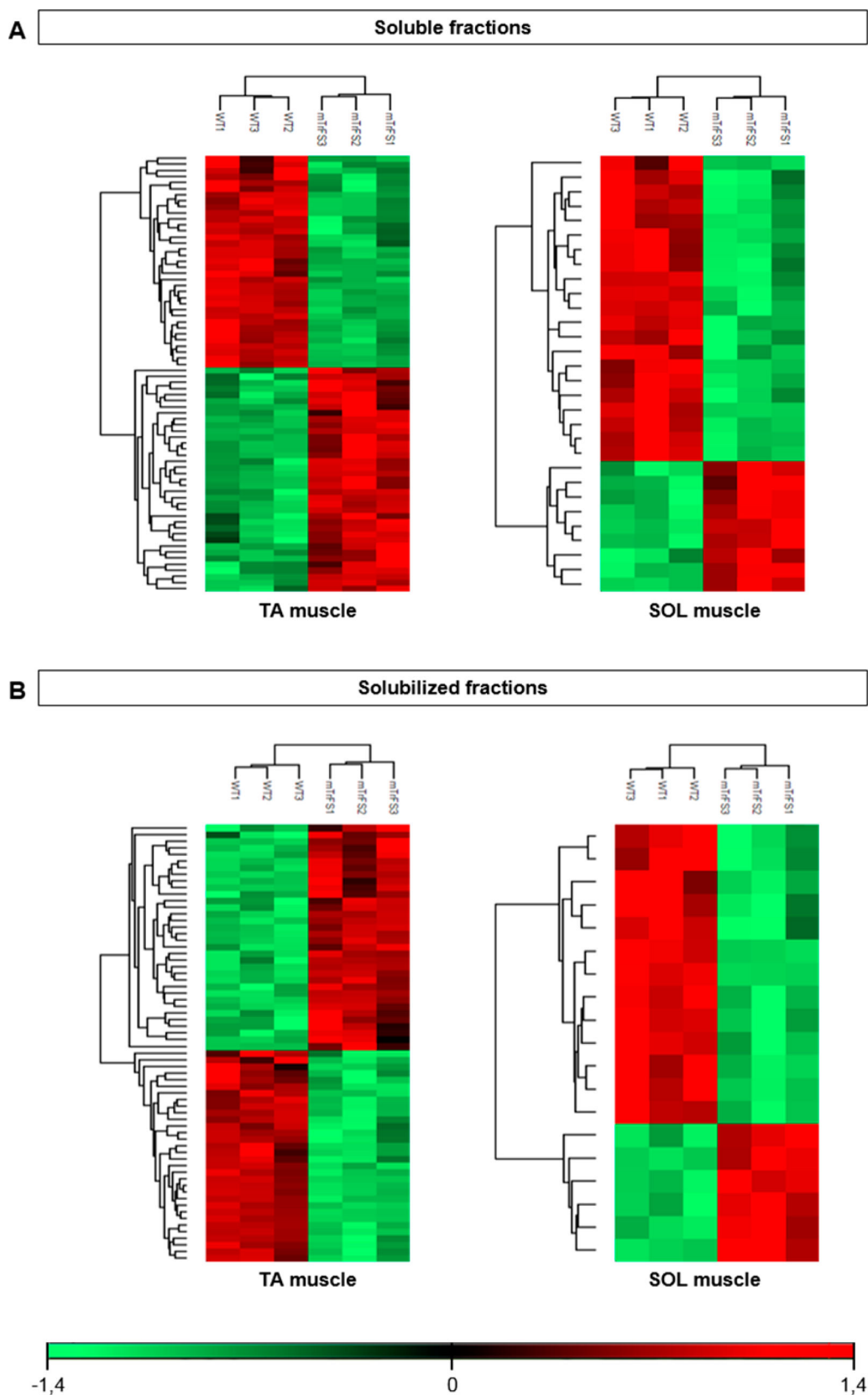


Figure 4. Comparative proteomic profile of skeletal muscles from mTrFS and WT Mice. Hierarchical clustering of significantly changing proteins in (A) SF and (B) sF from TA and SOL muscles of mTrFS mice compared with their WT control littermates with upregulated proteins in red and downregulated proteins in green. The quality of the heat maps highlights the very low dispersion between the biological replicates as well as the clear separation between the two genotypes, namely, mTrFS and WT mice.

Table 2. Main FS-Induced Changes in Energy Metabolism

BP	proteins
TA Muscle	
↑ Glycogenolysis	STBD1; PHKB
↑ Glycolysis	LDHA; PGK1; GPI; PFKM; ENO3; SLC16A3
↑ G3P shuttle	GPD1; GPD2
↓ Malate shuttle	GOT1; GOT2; MDH1; MDH2
↓ Lipid oxidation	ACADL; ACAT1; ACADM; ACADS
↓ Krebs cycle	PDK2; ACO2; CS; IDH3A; OGDH; PDHA1; IDH2
↓ Oxidative phosphorylation	NDUFA8/9/10; NDUFS1; MTNDS; NDUFB11; NDUFV1/2; NDUFC2; MTCO2; UQCRRS1; SDHA
SOL Muscle	
↓ Glycogenolysis	PGM1/2
↓ Glycolysis	ENO3; PKM; PGAM2
↓ Malate shuttle	MDH1; MDH2
↓ Lipid oxidation	ACAT1
↓ Krebs cycle	IDH2; PDHB
↓ Oxidative phosphorylation	ATP5B; COX6B1; COX7A1; NDUFA13; NDUFS5; SDHA

Changes in Myosin Isoforms. Skeletal muscle is composed of four types of muscle fibers based on the expression of different isoforms of myosin heavy chain (MHC) proteins, which are directly responsible for the contractile performance.³¹ While MYH4 is the predominant MHC in fast type IIB fibers, MYH7 is more abundant in slow type I muscle fibers, MYH2 and MYH1 being associated with

intermediate fast oxidative type IIA and IIX fibers, respectively. On the basis of the expression of these MHC proteins, it has been demonstrated that loss of Mstn signaling alters the fiber-type distribution, promoting a larger proportion of type II B fibers and a reduced proportion of type I fibers.^{11,62} Moreover, this slow-to-fast switch in the MHC, concomitant with the glycolytic characteristics, is consistent with the reduction in overall endurance of Mstn KO mice.⁵⁸

Although Mstn regulates oxidative metabolism independently of muscle fiber composition,¹¹ our proteomic analysis demonstrates a slow-to-fast switch of the contractile apparatus of TA muscle (Table 3), which is perfectly in line with the metabolic changes just described above. Furthermore, this switch is supported by a decrease in the abundance of the muscle specific protein myoglobin (MB), the main oxygen carrier into the muscle fiber. In SOL muscle, there was no clear evidence of a slow-to-fast switch on the basis of myosin isoforms.

Alterations in the Muscle Proteostasis. Muscle mass is tightly controlled by a balance between protein synthesis and degradation. A wealth of data strongly supports that Mstn affects this equilibrium to increase muscle mass.⁶³ An increased protein synthesis rate in skeletal muscle is well established in response to Mstn inhibition.⁶⁴ Moreover, several genes encoding translation initiation and elongation factors are known to be upregulated in muscle of Mstn KO mice.^{39,65} In contrast, a reduced protein degradation, namely, a reduced activity of the ubiquitin–proteasome system, has been

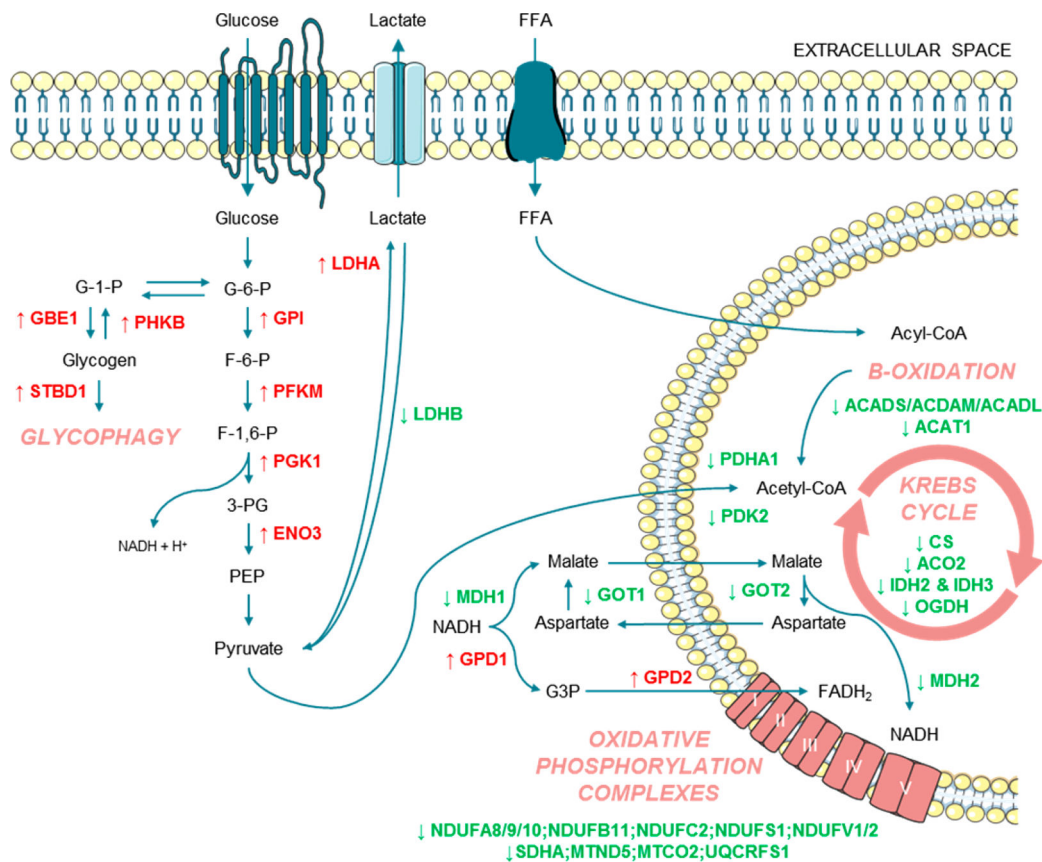


Figure 5. Regulation induced by FS in the energy metabolism of fast muscles. All of the proteins discussed in the **Changes in Energy Metabolism** section from the description of proteomic results are reported, with increased proteins in red and decreased proteins in green.

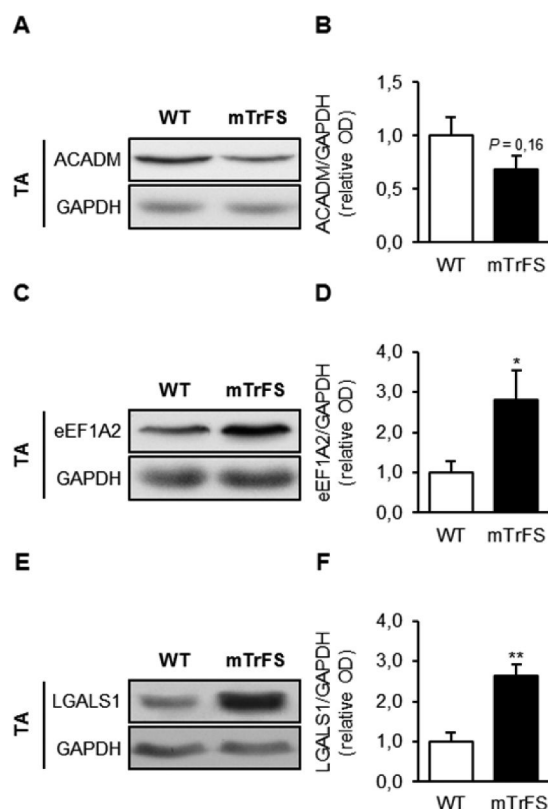


Figure 6. Immunoblot verification of selected muscle proteins identified as regulated by FS. FS overexpression decreases muscle abundance of (A,B) the β -oxidation enzyme ACADM and increases the abundance of (C,D) the translation factor eEF1A2 and (E,F) the lectin LGALS1. Western blot quantifications were assessed on TA muscle ($n = 6$ /group). Results are expressed as means \pm SEM. Statistical analysis was performed using unpaired t test (*, $P < 0.05$; **, $P < 0.01$).

Table 3. Main FS-Induced Changes in Myosin Isoforms

BP	proteins
TA Muscle	
↑ Fast and glycolytic	MYBPC2; MYH4
↓ Slow and oxidative	MYBPC1; MYH1; MYH2; MB

suggested to be one of the mechanisms contributing to muscle growth induced by Mstn inhibition.^{66,67}

Our proteomic analysis shows changes in the abundance of proteins involved in both protein synthesis and degradation pathways (Table 4). In TA muscle, we observed that FS overexpression induced an increase in the abundance of

Table 4. Main FS-Induced Changes in Muscle Proteostasis

BP	proteins
TA Muscle	
↑ Transcription	PBXIP1; ZFP692
↑ Translation	MRPS36; RPS14; eEF1G; eEF1A2
↑ UPS component	UBXN1
↑ Calpain activator	DBI
SOL Muscle	
↑ Transcription	HNRNPH1/2; HNRNPF
↓ Translation	EEF2; RPLP2
↓ Proteolysis	SERPINB1A

transcription and translation factors, among which we confirmed the higher abundance of the elongation factor eEF1A2 (Figure 6C,D). Surprisingly, we also observed an increase in the levels of mediators of both the ubiquitin-proteasome catabolic system (UPS) and the calcium-dependent proteolytic system (Calpains). However, these last observations are not sufficient to conclude to an activation of global proteolysis.

In SOL muscle, we also observed an increase in transcription factors. However, in contrast with TA muscle, we observed a lower abundance of proteins implicated both in protein synthesis and proteolysis. These last results partly contrast with those observed in TA muscle, which develops hypertrophy. However, whether these changes explain the absence of SOL hypertrophy in response to Mstn inhibition is unknown.

Perturbations in Signal Transduction. Skeletal muscle constitutes one of the main targets of the insulin action. There is some evidence to indicate that Mstn targets the muscle insulin transduction pathway.³² Indeed, enhanced muscle insulin-stimulated glucose uptake in response to Mstn inhibition is well established, but the mechanisms are not yet clearly described, although it does not appear to be only due to increased muscle mass by itself.^{9,50,68,69} Consistent with these observations, our proteomic analysis on TA muscle shows an increase in the abundance of two proteins, YWHAE (14–3–3 protein) and FABP3 (Fatty Acid Binding Protein 3), regulating the insulin-mediated glucose uptake (Table 5). While YWHAE

Table 5. Main FS-Induced Changes in Signal Transduction

BP	proteins
TA Muscle	
↑/↓ Calcium signaling	↓ CAV3; ↓ RYR1; ↓ TNNC2; ↓ ATP2A1; ↑ SRL
↑ Insulin signaling	YWHAE; FABP3
SOL Muscle	
↓ Calcium signaling	RYR1

interacts with the RabGAP TBC1D4 and TBC1D1 to allow the insulin-stimulated GLUT4 translocation,^{70–72} FABP3 stimulates TBC1D4 phosphorylation and glucose uptake via AMPK and AKT activation.⁷³ However, to date, no study has investigated the role of these proteins in the enhanced insulin sensitivity observed in response to Mstn inhibition.

Calcium regulates numerous signaling pathways in skeletal muscle such as excitation–contraction coupling process (ECC). ECC begins when depolarization of the transverse tubular membrane (T-tubules) initiates the release of Ca^{2+} from its intracellular store, the sarcoplasmic reticulum (SR), via the ryanodine receptors (RyRs). RyR activation is achieved by the depolarization-evoked conformational change of the dihydropyridine receptors (DHPRs) localized in the T-tubules.⁷⁴ Ca^{2+} ions then bind to the regulatory sites on troponin-C (TNNC) to initiate a conformational change, which, in turn, enables the binding of myosin to actin and, ultimately, results in muscle contraction.⁷⁵ Caveolin-3 (CAV3), mainly expressed in striated muscle tissues,⁷⁶ has been suggested to be an important regulator of ECC in skeletal muscle. Indeed, on one hand, the expression levels of CAV3 in skeletal muscle modulate Ca^{2+} currents through DHPR,⁷⁷ and, on the other hand, there is a direct interaction between CAV3 and RYR1.⁷⁸ After initiation of muscle contraction by the increase in cytosolic Ca^{2+} concentration, Ca^{2+} ions are pumped back to the SR by SR- Ca^{2+} -ATPase (SERCA), whose ATP2A1 is the major adult fast-

twitch specific isoform in skeletal muscle, leading to relaxation. Moreover, the sarcalumenin (SRL) is a major luminal glycoprotein that codistributes with SERCA and accelerates Ca^{2+} transport and sequestration.^{79–81} It has been demonstrated that despite the larger muscle mass relative to age-matched WT mice, Mstn KO mice show a decreased specific force production and a shortening of contraction and relaxation time during isometric twitch contraction.⁵⁴ Later, it has been shown that mutations in the Mstn gene are accompanied by a reduction in the calcium release from the SR.⁸² In line with these data, our proteomic analysis reveals changes in the level of proteins involved in SR calcium flux, which might be involved in the defective mechanical function observed in response to Mstn inhibition (Table 5). Interestingly, SRL KO mice are characterized by an enhanced fatigue resistance,⁸³ suggesting that this protein is involved in muscle fatigue. Therefore, the increase that we show in the abundance of SRL, concomitant with the mitochondrial functional depletion, could contribute to the impaired exercise tolerance and fatigability reported for Mstn KO mice.

Muscle Remodeling and Repair. Repair of damage to the plasma membrane is an important aspect of cellular physiology, and disruption of this process can contribute to pathophysiology in a number of human diseases, including muscular dystrophies. It has been demonstrated that Mstn inhibition accelerates skeletal muscle healing after injury through increased muscle regeneration and decreased fibrosis.^{84,85}

Fibrosis is characterized by an excessive accumulation of extracellular matrix (ECM) components, particularly collagen. A number of studies have reported a decreased fibrosis in mdx mice with Mstn deletion or postnatal inhibition.^{86–88} Our proteomic study points to a reduced muscle collagen content, as previously demonstrated,^{34,64} concomitant with a lower level in other ECM components in TA muscle overexpressing FS (Table 6), which may contribute to fibrosis attenuation. These

Table 6. Main FS-Induced Changes in Muscle Remodeling and Repair

BP	Proteins
TA Muscle	
↑ Repair patch formation	TRIM72
↑ Stem cells fusion	LGALS1
↓ ECM components	COL1A1; COL1A2; FMOD; LUM; DCN

alterations in the connective tissue may lead to a more rapid uptake of the series elastic components⁸⁹ and act as an additional contributing factor to the more rapid twitch kinetics observed in response to Mstn deficiency.⁵⁴

The cell membrane repair response involves translocation of intracellular vesicles to the injury site to form a membrane repair patch composed of several proteins, including the tripartite motif-containing protein 72 (TRIM72, also called MG53). TRIM72, which is found increased in TA muscle from mTrFS mice, contributes to intracellular vesicle trafficking and is an essential component of the membrane repair machinery in striated muscle.⁹⁰ Galectin-1 (LGALS1) is a small 14 kDa lectin found in many tissues, including skeletal, smooth, and cardiac muscle. In skeletal muscle, LGALS1 aids in the fusion and differentiation of myoblasts during muscle repair and regeneration.^{91–93} Interestingly, LGALS1 KO mice present a delay in muscle fiber development in the neonatal stage and a reduced muscle fiber diameter in the adult stage.⁹² Its

abundance also appears to be increased in TA muscle from mTrFS mice, as reported by the proteomic data and the Western blotting confirmation (Figure 6E,F); however, the role of these two proteins, TRIM72 and LGALS1, in the improved muscle repair after Mstn inhibition remains unexplored.

Transcriptomic Signature of Hypertrophied Muscle Overexpressing FS

In an attempt to ascertain the origin of the proteomic changes, a microarray analysis was used to investigate transcriptional changes occurring in TA muscle from mTrFS mice. The expression of 315 genes appears significantly altered in mTrFS mice compared to WT mice ($\text{FC} > 1.4$, $P < 0.05$), including 132 upregulated and 183 downregulated genes (Figure 7A and Table S4). To confirm the validity of our microarray results, expression levels of 10 randomly selected genes were validated with RT-qPCR. A nice correlation of 0.9417 was observed between the fold changes from the microarray and the RT-qPCR analysis (Figure 7B).

Consistent with the proteomic data, examination of the GO: Biological Process annotation reveals that metabolism has the highest number of gene hits (25% of the regulated genes). Moreover, the KEGG pathways analysis highlights the regulation in ECM components (KEGG: 04512), as revealed by the reduced expression of several collagen genes (Col1a1, Col1a2, Col3a1, Col5a2, and Col11a1). Moreover, the web-based tool also pinpoints changes in insulin signaling pathway (KEGG: 0491), TGF β signaling pathway (KEGG: 04350), and calcium signaling pathway (KEGG: 04020). Like Rahimov et al., we find a decrease in the expression of Nos1, involved in nitric oxide (NO) synthesis.⁶⁷ Interestingly, reduced NO production has been correlated to a decreased muscle twitch force⁹⁴ and increased fatigability,^{95,96} two characteristics already described above of Mstn-deficient mice.

Moreover, we performed a global comparison of the omic data from TA muscles overexpressing FS to provide insight into the program of transcription and translation events governing the FS-mediated skeletal muscle phenotype. As recently reported in a similar work by Davey et al.,⁴¹ a large number of significant transcriptome changes (315 genes) were not accompanied by a significant change in the proteome. Indeed, we revealed a redundancy of <10% between proteomic and transcriptomic data. However, on the basis of our experimental configuration, we cannot rule out that a subset of differentially expressed proteins observed in our proteomic data may be correlated with a nonsignificant transcriptomics change and vice versa. Nevertheless, all eight proteins significantly regulated both at the transcriptomic and proteomic levels in muscles from mTrFS mice show the same directional changes in the two analysis sets (Figure 7C). Among these, five proteins are upregulated (MYBPH, STBD1, BIN1, LGALS1, and GSTM1) while the three others are downregulated (MYH1, COL1A2, COL1A1) by FS overexpression. The Myosin-binding protein H (MYBPH) interacts with Rho kinase 2 (ROCK2) to affect the interaction of actin-myosin.^{97,98} Interestingly, Miao et al. has previously suggested the role of the Ky/MYBPH/ACTINS pathway as a potential player in the muscle hypertrophy of the MSTN propeptide transgenic mice.⁹⁹

CONCLUSIONS

Our work describes the most extensive omics study of skeletal muscle overexpressing FS to explain the observed muscle phenotypic changes in both fast and slow muscle from mTrFS

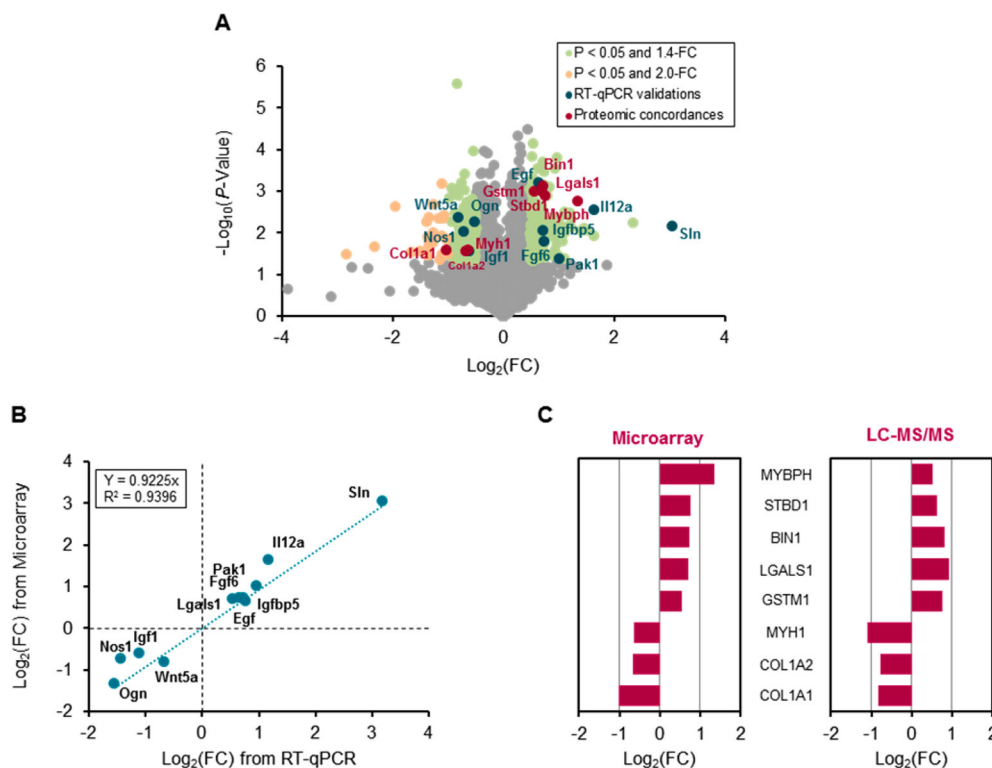


Figure 7. Changes in the TA muscle transcriptomic profile induced by FS overexpression. (A) Volcano plot of the 315 genes significantly altered in response to FS ($FC > 1.4$, $P < 0.05$), with 132 up- and 183 downregulated genes. (B) RT-qPCR validations show a nice correlation with the transcriptomic results in terms of FC. (C) Omics data comparison revealed eight proteins similarly regulated at the transcriptomic and proteomic levels in TA muscle from mTrFS mice compared with WT mice.

mice. We provide molecular-level insights into protein modifications potentially contributing to changes observed in muscle mass (changes in protein synthesis and degradation mediators), in muscle performance such as fatigability and exercise intolerance (decreases in mitochondrial proteins and changes in MHC isoforms) and ECC process (alterations in SR calcium flux players), in muscle remodeling (changes in ECM component and regeneration mediators), and in energy metabolic function (changes in glycolytic metabolism enzymes and regulators of insulin signaling). Our results highlight genes and proteins not previously associated with the FS-mediated skeletal muscle phenotype. In conclusion, this work generates new hypotheses to deepen the intracellular processes mediating the skeletal muscle hypertrophy and metabolic adaptations in response to FS to allow the development of targeted therapies for muscle wasting and insulin resistance.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00069.

Supplemental Methods – Experimental procedure details. Protein digestion eFASP with passivated ultrafiltration unit. Supplemental Table S1 – BlastP validations for one single peptide-based protein identifications. BlastP interrogations for each single peptide of proteins identified with one single peptide and found as differentially expressed between mTrFS and WT mice. Supplemental Figure 1 – Detailed information for one single peptide-based protein identification. Detailed

information about the MS/MS spectrum, peptide sequence, and precursor m/z for each single peptide of proteins identified with one single peptide and found as differentially expressed between mTrFS and WT mice. Supplemental Figure S2 – Quality of the proteomic analysis. Box plots representing Log_2 signal intensities for each replicate from TA muscle of WT mice indicating high technical reproducibility and very low biological variance and PCA of TA and SOL muscles from WT mice. (PDF)

Supplemental Table S2 – Proteomic profile of TA muscle in response to FS. List of the 115 proteins significantly regulated in TA muscle of mTrFS mice compared with WT mice. (PDF)

Supplemental Table S3 – Proteomic profile of SOL muscle in response to FS. List of the 44 proteins significantly regulated in SOL muscle of mTrFS mice compared with WT mice. (PDF)

Supplemental Table S4 – Transcriptomic profile of TA muscle in response to FS. List of the 315 genes significantly regulated in TA muscle of mTrFS mice compared with WT mice. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 32 (0) 2 764 54 24. Fax: 32 (0) 2 764 54 79. E-mail: caroline.barbe@uclouvain.be.

ORCID

Caroline Barbé: 0000-0001-7372-6302

Christian Rolando: 0000-0002-3266-8860

Notes

The authors declare no competing financial interest. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the Prida partner repository with the data set identifier PXD005899

ACKNOWLEDGMENTS

We thank Dr. S. J. Lee (Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, USA) for the genetically modified mice and Claire Ploquin for her contribution in starting this work on the transcriptomic analysis. The proteomic platform was supported by the IBISA network. The mass spectrometers were funded by the Université Lille 1, Sciences et Technologie, the CNRS, the Région Hauts-de-France, and the European Regional Development Fund. The support of the high field FT-ICR MS network (FR 3624 CNRS) is also gratefully acknowledge. Finally, this work was funded by grants from the French Anti-Doping Agency (AFLD, France), Fonds de la Recherche Scientifique Médicale (FRSM-FNRS, Belgium), and Fonds spéciaux de Recherche (FSR, Université Catholique de Louvain, Belgium).

REFERENCES

- (1) Grobet, L.; Pirotin, D.; Farnir, F.; Poncelet, D.; Royo, L. J.; Brouwers, B.; Christians, E.; Desmecht, D.; Coignoul, F.; Kahn, R.; Georges, M. Modulating skeletal muscle mass by postnatal, muscle-specific inactivation of the myostatin gene. *Genesis (Hoboken, NJ, U. S.)* **2003**, *35* (4), 227–38.
- (2) McPherron, A. C.; Lawler, A. M.; Lee, S. J. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* **1997**, *387* (6628), 83–90.
- (3) Rodino-Klapac, L. R.; Janssen, P. M.; Shontz, K. M.; Canan, B.; Montgomery, C. L.; Griffin, D.; Heller, K.; Schmelzer, L.; Handy, C.; Clark, K. R.; Sahenk, Z.; Mendell, J. R.; Kaspar, B. K. Micro-dystrophin and follistatin co-delivery restores muscle function in aged DMD model. *Hum. Mol. Genet.* **2013**, *22* (24), 4929–37.
- (4) Gallot, Y. S.; Durieux, A. C.; Castells, J.; Desgeorges, M. M.; Vernus, B.; Plantureux, L.; Remond, D.; Jahnke, V. E.; Lefai, E.; Dardevet, D.; Nemoz, G.; Schaeffer, L.; Bonniou, A.; Freyssen, D. G. Myostatin gene inactivation prevents skeletal muscle wasting in cancer. *Cancer Res.* **2014**, *74* (24), 7344–56.
- (5) Pauly, M.; Chabi, B.; Favier, F. B.; Vanterpool, F.; Matecki, S.; Fouret, G.; Bonafos, B.; Vernus, B.; Feillet-Coudray, C.; Coudray, C.; Bonniou, A.; Ramonaxo, C. Combined Strategies for Maintaining Skeletal Muscle Mass and Function in Aging: Myostatin Inactivation and AICAR-Associated Oxidative Metabolism Induction. *J. Gerontol., Ser. A* **2015**, *70* (9), 1077–87.
- (6) Attie, K. M.; Borgstein, N. G.; Yang, Y.; Condon, C. H.; Wilson, D. M.; Pearsall, A. E.; Kumar, R.; Willins, D. A.; Seehra, J. S.; Sherman, M. L. A single ascending-dose study of muscle regulator ACE-031 in healthy volunteers. *Muscle Nerve* **2013**, *47* (3), 416–23.
- (7) Cohen, S.; Nathan, J. A.; Goldberg, A. L. Muscle wasting in disease: molecular mechanisms and promising therapies. *Nat. Rev. Drug Discovery* **2014**, *14* (1), 58–74.
- (8) Mendell, J. R.; Sahenk, Z.; Malik, V.; Gomez, A. M.; Flanigan, K. M.; Lowes, L. P.; Alfano, L. N.; Berry, K.; Meadows, E.; Lewis, S.; Braun, L.; Shontz, K.; Rouhana, M.; Clark, K. R.; Rosales, X. Q.; Al-Zaidy, S.; Govoni, A.; Rodino-Klapac, L. R.; Hogan, M. J.; Kaspar, B. K. A phase 1/2a follistatin gene therapy trial for becker muscular dystrophy. *Mol. Ther.* **2015**, *23* (1), 192–201.
- (9) Guo, T.; Jou, W.; Chanturiya, T.; Portas, J.; Gavrilo, O.; McPherron, A. C. Myostatin inhibition in muscle, but not adipose tissue, decreases fat mass and improves insulin sensitivity. *PLoS One* **2009**, *4* (3), e4937.
- (10) Ploquin, C.; Chabi, B.; Fouret, G.; Vernus, B.; Feillet-Coudray, C.; Coudray, C.; Bonniou, A.; Ramonaxo, C. Lack of myostatin alters intermyofibrillar mitochondria activity, unbalances redox status, and impairs tolerance to chronic repetitive contractions in muscle. *American journal of physiology. Endocrinology and metabolism* **2012**, *302* (8), E1000–8.
- (11) Mouisel, E.; Relizani, K.; Mille-Hamard, L.; Denis, R.; Hourde, C.; Agbulut, O.; Patel, K.; Arandel, L.; Morales-Gonzalez, S.; Vignaud, A.; Garcia, L.; Ferry, A.; Luquet, S.; Billat, V.; Ventura-Clapier, R.; Schuelke, M.; Amthor, H. Myostatin is a key mediator between energy metabolism and endurance capacity of skeletal muscle. *American journal of physiology. Regulatory, integrative and comparative physiology* **2014**, *307* (4), R444–54.
- (12) Rossiter, H. B.; Abraham, A.; Kovanecz, I.; Ali, H.; Porszasz, J.; Reisz-Porszasz, S. The Regulatory Role of Myostatin on Skeletal Muscle Metabolism and Fatigue Resistance. *FASEB J.* **2016**, *30* (1 Supplement), 1245.26.
- (13) Wagner, K. R.; Fleckenstein, J. L.; Amato, A. A.; Barohn, R. J.; Bushby, K.; Escolar, D. M.; Flanigan, K. M.; Pestronk, A.; Tawil, R.; Wolfe, G. I.; Eagle, M.; Florence, J. M.; King, W. M.; Pandya, S.; Straub, V.; Juneau, P.; Meyers, K.; Csimm, C.; Araujo, T.; Allen, R.; Parsons, S. A.; Wozney, J. M.; Lavallie, E. R.; Mendell, J. R. A phase I/II trial of MYO-029 in adult subjects with muscular dystrophy. *Ann. Neurol.* **2008**, *63* (5), 561–71.
- (14) Foley, J. W.; Bercury, S. D.; Finn, P.; Cheng, S. H.; Scheule, R. K.; Ziegler, R. J. Evaluation of systemic follistatin as an adjuvant to stimulate muscle repair and improve motor function in Pompe mice. *Mol. Ther.* **2010**, *18* (9), 1584–91.
- (15) Smith, R. C.; Cramer, M. S.; Mitchell, P. J.; Capen, A.; Huber, L.; Wang, R.; Myers, L.; Jones, B. E.; Eastwood, B. J.; Ballard, D.; Hanson, J.; Credille, K. M.; Wroblewski, V. J.; Lin, B. K.; Heuer, J. G. Myostatin Neutralization Results in Preservation of Muscle Mass and Strength in Preclinical Models of Tumor-Induced Muscle Wasting. *Mol. Cancer Ther.* **2015**, *14* (7), 1661–70.
- (16) Collins-Hooper, H.; Sartori, R.; Macharia, R.; Visanuvimol, K.; Foster, K.; Matsakas, A.; Flasskamp, H.; Ray, S.; Dash, P. R.; Sandri, M.; Patel, K. Propeptide-mediated inhibition of myostatin increases muscle mass through inhibiting proteolytic pathways in aged mice. *J. Gerontol., Ser. A* **2014**, *69* (9), 1049–59.
- (17) Dong, J.; Dong, Y.; Dong, Y.; Chen, F.; Mitch, W. E.; Zhang, L. Inhibition of myostatin in mice improves insulin sensitivity via irisin-mediated cross talk between muscle and adipose tissues. *Int. J. Obes.* **2016**, *40* (3), 434–42.
- (18) Amthor, H.; Nicholas, G.; McKinnell, I.; Kemp, C. F.; Sharma, M.; Kambadur, R.; Patel, K. Follistatin complexes Myostatin and antagonises Myostatin-mediated inhibition of myogenesis. *Dev. Biol.* **2004**, *270* (1), 19–30.
- (19) Lee, S. J.; McPherron, A. C. Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (16), 9306–11.
- (20) Winbanks, C. E.; Weeks, K. L.; Thomson, R. E.; Sepulveda, P. V.; Beyer, C.; Qian, H.; Chen, J. L.; Allen, J. M.; Lancaster, G. I.; Febbraio, M. A.; Harrison, C. A.; McMullen, J. R.; Chamberlain, J. S.; Gregorevic, P. Follistatin-mediated skeletal muscle hypertrophy is regulated by Smad3 and mTOR independently of myostatin. *J. Cell Biol.* **2012**, *197* (7), 997–1008.
- (21) Kalista, S.; Schakman, O.; Gilson, H.; Lause, P.; Demeulder, B.; Bertrand, L.; Pende, M.; Thissen, J. P. The type 1 insulin-like growth factor receptor (IGF-IR) pathway is mandatory for the follistatin-induced skeletal muscle hypertrophy. *Endocrinology* **2012**, *153* (1), 241–53.
- (22) Sayd, T.; Morzel, M.; Chambon, C.; Franck, M.; Figwer, P.; Larzul, C.; Le Roy, P.; Monin, G.; Cherel, P.; Laville, E. Proteome analysis of the sarcoplasmic fraction of pig semimembranosus muscle: implications on meat color development. *J. Agric. Food Chem.* **2006**, *54* (7), 2732–7.
- (23) Erde, J.; Loo, R. R.; Loo, J. A. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **2014**, *13* (4), 1885–95.

- (24) Sun, L.; Zhu, G.; Dovichi, N. J. Comparison of the LTQ-Orbitrap Velos and the Q-Exactive for proteomic analysis of 1–1000 ng RAW 264.7 cell lysate digests. *Rapid Commun. Mass Spectrom.* **2013**, *27* (1), 157–62.
- (25) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13* (9), 731–40.
- (26) Schakman, O.; Kalista, S.; Bertrand, L.; Lause, P.; Verniers, J.; Ketelslegers, J. M.; Thissen, J. P. Role of Akt/GSK-3beta/beta-catenin transduction pathway in the muscle anti-atrophy action of insulin-like growth factor-I in glucocorticoid-treated rats. *Endocrinology* **2008**, *149* (8), 3900–8.
- (27) Dehoux, M.; Van Beneden, R.; Pasko, N.; Lause, P.; Verniers, J.; Underwood, L.; Ketelslegers, J. M.; Thissen, J. P. Role of the insulin-like growth factor I decline in the induction of atrogen-1/MAFbx during fasting and diabetes. *Endocrinology* **2004**, *145* (11), 4806–12.
- (28) Carmona-Saez, P.; Chagoyen, M.; Tirado, F.; Carazo, J. M.; Pascual-Montano, A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome biology* **2007**, *8* (1), R3.
- (29) Nogales-Cadenas, R.; Carmona-Saez, P.; Vazquez, M.; Vicente, C.; Yang, X.; Tirado, F.; Carazo, J. M.; Pascual-Montano, A. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W317–22.
- (30) Tabas-Madrid, D.; Nogales-Cadenas, R.; Pascual-Montano, A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res.* **2012**, *40* (Web Server issue), W478–83.
- (31) Schiaffino, S.; Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol. Rev.* **2011**, *91* (4), 1447–531.
- (32) Barbe, C.; Kalista, S.; Loumaye, A.; Ritvos, O.; Lause, P.; Ferracin, B.; Thissen, J. P. Role of IGF-I in follistatin-induced skeletal muscle hypertrophy. *American journal of physiology. Endocrinology and metabolism* **2015**, *309* (6), E557–67.
- (33) Parsons, S. A.; Millay, D. P.; Sargent, M. A.; McNally, E. M.; Molkentin, J. D. Age-dependent effect of myostatin blockade on disease severity in a murine model of limb-girdle muscular dystrophy. *Am. J. Pathol.* **2006**, *168* (6), 1975–85.
- (34) Mendias, C. L.; Marcin, J. E.; Calerdon, D. R.; Faulkner, J. A. Contractile properties of EDL and soleus muscles of myostatin-deficient mice. *J. Appl. Physiol.* **2006**, *101* (3), 898–905.
- (35) Matsakas, A.; Foster, K.; Otto, A.; Macharia, R.; Elashry, M. I.; Feist, S.; Graham, L.; Foster, H.; Yaworsky, P.; Walsh, F.; Dickson, G.; Patel, K. Molecular, cellular and physiological investigation of myostatin propeptide-mediated muscle growth in adult mice. *Neuromuscular disorders: NMD* **2009**, *19* (7), 489–99.
- (36) Cadena, S. M.; Tomkinson, K. N.; Monnell, T. E.; Spait, M. S.; Kumar, R.; Underwood, K. W.; Pearsall, R. S.; Lachey, J. L. Administration of a soluble activin type IIB receptor promotes skeletal muscle growth independent of fiber type. *J. Appl. Physiol.* **2010**, *109* (3), 635–42.
- (37) Carlson, C. J.; Booth, F. W.; Gordon, S. E. Skeletal muscle myostatin mRNA expression is fiber-type specific and increases during hindlimb unloading. *Am. J. Physiol.* **1999**, *277* (2 Pt 2), R601–6.
- (38) Allen, D. L.; Unterman, T. G. Regulation of myostatin expression and myoblast differentiation by FoxO and SMAD transcription factors. *American journal of physiology. Cell physiology* **2006**, *292* (1), C188–99.
- (39) Chelh, I.; Meunier, B.; Picard, B.; Reecy, M. J.; Chevalier, C.; Hocquette, J. F.; Cassar-Malek, I. Molecular profiles of Quadriceps muscle in myostatin-null mice reveal PI3K and apoptotic pathways as myostatin targets. *BMC Genomics* **2009**, *10*, 196.
- (40) Salzler, R. R.; Shah, D.; Dore, A.; Bauerlein, R.; Miloscio, L.; Latres, E.; Papadopoulos, N. J.; Olson, W. C.; MacDonald, D.; Duan, X. Myostatin deficiency but not anti-myostatin blockade induces marked proteomic changes in mouse skeletal muscle. *Proteomics* **2016**, *16* (14), 2019–27.
- (41) Davey, J. R.; Watt, K. I.; Parker, B. L.; Chaudhuri, R.; Ryall, J. G.; Cunningham, L.; Qian, H.; Sartorelli, V.; Sandri, M.; Chamberlain, J.; James, D. E.; Gregorevic, P. Integrated expression analysis of muscle hypertrophy identifies Asb2 as a negative regulator of muscle mass. *JCI insight* **2016**, *1* (5), 85477 DOI: 10.1172/jci.insight.85477.
- (42) Deshmukh, A. S.; Murgia, M.; Nagaraj, N.; Treebak, J. T.; Cox, J.; Mann, M. Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol. Cell. Proteomics* **2015**, *14* (4), 841–53.
- (43) Hamelin, M.; Sayd, T.; Chambon, C.; Bouix, J.; Bibe, B.; Milenkovic, D.; Leveziel, H.; Georges, M.; Clop, A.; Marinova, P.; Laville, E. Differential expression of sarcoplasmic proteins in four heterogeneous ovine skeletal muscles. *Proteomics* **2007**, *7* (2), 271–80.
- (44) Vitorino, R.; Ferreira, R.; Neuparth, M.; Guedes, S.; Williams, J.; Tomer, K. B.; Domingues, P. M.; Appell, H. J.; Duarte, J. A.; Amado, F. M. Subcellular proteomics of mice gastrocnemius and soleus muscles. *Anal. Biochem.* **2007**, *366* (2), 156–69.
- (45) Sato, Y.; Shimizu, M.; Mizunoya, W.; Wariishi, H.; Tatsumi, R.; Buchman, V. L.; Ikeuchi, Y. Differential expression of sarcoplasmic and myofibrillar proteins of rat soleus muscle during denervation atrophy. *Biosci., Biotechnol., Biochem.* **2009**, *73* (8), 1748–56.
- (46) Keller, A.; Peltzer, J.; Carpentier, G.; Horvath, I.; Olah, J.; Duchesnay, A.; Orosz, F.; Ovadi, J. Interactions of enolase isoforms with tubulin and microtubules during myogenesis. *Biochim. Biophys. Acta, Gen. Subj.* **2007**, *1770* (6), 919–26.
- (47) Sun, H.; Qiu, J.; Chen, Y.; Yu, M.; Ding, F.; Gu, X. Proteomic and bioinformatic analysis of differentially expressed proteins in denervated skeletal muscle. *Int. J. Mol. Med.* **2014**, *33* (6), 1586–96.
- (48) Gannon, J.; Staunton, L.; O'Connell, K.; Doran, P.; Ohlndieck, K. Phosphoproteomic analysis of aged skeletal muscle. *Int. J. Mol. Med.* **2008**, *22* (1), 33–42.
- (49) Capitanio, D.; Vasso, M.; Fania, C.; Moriggi, M.; Viganò, A.; Procacci, P.; Magnaghi, V.; Gelfi, C. Comparative proteomic profile of rat sciatic nerve and gastrocnemius muscle tissues in ageing by 2-D DIGE. *Proteomics* **2009**, *9* (7), 2004–20.
- (50) Cleasby, M. E.; Jarmin, S.; Eilers, W.; Elashry, M.; Andersen, D. K.; Dickson, G.; Foster, K. Local overexpression of the myostatin propeptide increases glucose transporter expression and enhances skeletal muscle glucose disposal. *American journal of physiology. Endocrinology and metabolism* **2014**, *306* (7), E814–23.
- (51) Kocsis, T.; Trencsenyi, G.; Szabo, K.; Baan, J. A.; Müller, G.; Mendl, L.; Garai, I.; Reinauer, H.; Deak, F.; Dux, L.; Keller-Pinter, A. Myostatin propeptide mutation of the hypermuscular Compact mice decreases the formation of myostatin and improves insulin sensitivity. *American journal of physiology. Endocrinology and metabolism* **2017**, *312*, E150.
- (52) Gangopadhyay, S. S. Systemic administration of follistatin288 increases muscle mass and reduces fat accumulation in mice. *Sci. Rep.* **2013**, *3*, 2441.
- (53) Brandt, C.; Hansen, R. H.; Hansen, J. B.; Olsen, C. H.; Galle, P.; Mandrup-Poulsen, T.; Gehl, J.; Pedersen, B. K.; Hojman, P. Overexpression of Follistatin-like 3 attenuates fat accumulation and improves insulin sensitivity in mice. *Metab., Clin. Exp.* **2015**, *64* (2), 283–95.
- (54) Amthor, H.; Macharia, R.; Navarrete, R.; Schuelke, M.; Brown, S. C.; Otto, A.; Voit, T.; Munttoni, F.; Vrbova, G.; Partridge, T.; Zammit, P.; Bunker, L.; Patel, K. Lack of myostatin results in excessive muscle growth but impaired force generation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (6), 1835–40.
- (55) Baligand, C.; Gilson, H.; Menard, J. C.; Schakman, O.; Wary, C.; Thissen, J. P.; Carlier, P. G. Functional assessment of skeletal muscle in intact mice lacking myostatin by concurrent NMR imaging and spectroscopy. *Gene Ther.* **2010**, *17* (3), 328–37.
- (56) Lipina, C.; Kendall, H.; McPherron, A. C.; Taylor, P. M.; Hundal, H. S. Mechanisms involved in the enhancement of mammalian target of rapamycin signalling and hypertrophy in skeletal muscle of myostatin-deficient mice. *FEBS Lett.* **2010**, *584* (11), 2403–8.

- (57) Matsakas, A.; Narkar, V. A. Endurance exercise mimetics in skeletal muscle. *Current sports medicine reports* **2010**, *9* (4), 227–32.
- (58) Savage, K. J.; McPherron, A. C. Endurance exercise training in myostatin null mice. *Muscle Nerve* **2010**, *42* (3), 355–62.
- (59) Mancuso, M.; Salviati, L.; Sacconi, S.; Otaegui, D.; Camano, P.; Marina, A.; Bacman, S.; Moraes, C. T.; Carlo, J. R.; Garcia, M.; Garcia-Alvarez, M.; Monzon, L.; Naini, A. B.; Hirano, M.; Bonilla, E.; Taratuto, A. L.; DiMauro, S.; Vu, T. H. Mitochondrial DNA depletion: mutations in thymidine kinase gene with myopathy and SMA. *Neurology* **2002**, *59* (8), 1197–202.
- (60) Zhang, S. J.; Bruton, J. D.; Katz, A.; Westerblad, H. Limited oxygen diffusion accelerates fatigue development in mouse skeletal muscle. *J. Physiol.* **2006**, *572* (Pt 2), 551–9.
- (61) Gentry, B. A.; Ferreira, J. A.; Phillips, C. L.; Brown, M. Hindlimb skeletal muscle function in myostatin-deficient mice. *Muscle Nerve* **2011**, *43* (1), 49–57.
- (62) Girgenrath, S.; Song, K.; Whittemore, L. A. Loss of myostatin expression alters fiber-type distribution and expression of myosin heavy chain isoforms in slow- and fast-type skeletal muscle. *Muscle Nerve* **2005**, *31* (1), 34–40.
- (63) Rodriguez, J.; Vernus, B.; Chelch, I.; Cassar-Malek, I.; Gabillard, J. C.; Hadj Sassi, A.; Seiliez, I.; Picard, B.; Bonnieu, A. Myostatin and the skeletal muscle atrophy and hypertrophy signaling pathways. *Cell. Mol. Life Sci.* **2014**, *71* (22), 4361–71.
- (64) Welle, S.; Burgess, K.; Mehta, S. Stimulation of skeletal muscle myofibrillar protein synthesis, p70 S6 kinase phosphorylation, and ribosomal protein S6 phosphorylation by inhibition of myostatin in mature mice. *American journal of physiology. Endocrinology and metabolism* **2008**, *296* (3), E567–72.
- (65) Steelman, C. A.; Recknor, J. C.; Nettleton, D.; Reecy, J. M. Transcriptional profiling of myostatin-knockout mice implicates Wnt signaling in postnatal skeletal muscle growth and hypertrophy. *FASEB J.* **2006**, *20* (3), 580–2.
- (66) Zhou, X.; Wang, J. L.; Lu, J.; Song, Y.; Kwak, K. S.; Jiao, Q.; Rosenfeld, R.; Chen, Q.; Boone, T.; Simonet, W. S.; Lacey, D. L.; Goldberg, A. L.; Han, H. Q. Reversal of cancer cachexia and muscle wasting by ActRIIB antagonism leads to prolonged survival. *Cell* **2010**, *142* (4), 531–43.
- (67) Rahimov, F.; King, O. D.; Warsing, L. C.; Powell, R. E.; Emerson, C. P., Jr.; Kunkel, L. M.; Wagner, K. R. Gene expression profiling of skeletal muscles treated with a soluble activin type IIB receptor. *Physiol. Genomics* **2011**, *43* (8), 398–407.
- (68) Zhao, B.; Wall, R. J.; Yang, J. Transgenic expression of myostatin propeptide prevents diet-induced obesity and insulin resistance. *Biochem. Biophys. Res. Commun.* **2005**, *337* (1), 248–55.
- (69) Morissette, M. R.; Cook, S. A.; Buranasombati, C.; Rosenberg, M. A.; Rosenzweig, A. Myostatin inhibits IGF-I-induced myotube hypertrophy through Akt. *American journal of physiology. Cell physiology* **2009**, *297* (5), 1124–32.
- (70) Ramm, G.; Larance, M.; Guilhaus, M.; James, D. E. A role for 14–3-3 in insulin-stimulated GLUT4 translocation through its interaction with the RabGAP AS160. *J. Biol. Chem.* **2006**, *281* (39), 29174–80.
- (71) Pehmoller, C.; Treebak, J. T.; Birk, J. B.; Chen, S.; Mackintosh, C.; Hardie, D. G.; Richter, E. A.; Wojtaszewski, J. F. Genetic disruption of AMPK signaling abolishes both contraction- and insulin-stimulated TBC1D1 phosphorylation and 14–3-3 binding in mouse skeletal muscle. *American journal of physiology. Endocrinology and metabolism* **2009**, *297* (3), E665–75.
- (72) Chen, Q.; Quan, C.; Xie, B.; Chen, L.; Zhou, S.; Toth, R.; Campbell, D. G.; Lu, S.; Shirakawa, R.; Horiuchi, H.; Li, C.; Yang, Z.; MacKintosh, C.; Wang, H. Y.; Chen, S. GARNL1, a major RalGAP alpha subunit in skeletal muscle, regulates insulin-stimulated RalA activation and GLUT4 trafficking via interaction with 14–3-3 proteins. *Cell. Signalling* **2014**, *26* (8), 1636–48.
- (73) Kusudo, T.; Kontani, Y.; Kataoka, N.; Ando, F.; Shimokata, H.; Yamashita, H. Fatty acid-binding protein 3 stimulates glucose uptake by facilitating AS160 phosphorylation in mouse muscle cells. *Genes to cells: devoted to molecular & cellular mechanisms* **2011**, *16* (6), 681–91.
- (74) Lanner, J. T.; Georgiou, D. K.; Joshi, A. D.; Hamilton, S. L. Ryanodine receptors: structure, expression, molecular details, and function in calcium release. *Cold Spring Harbor Perspect. Biol.* **2010**, *2* (11), a003996.
- (75) Baylor, S. M.; Hollingworth, S. Intracellular calcium movements during excitation-contraction coupling in mammalian slow-twitch and fast-twitch muscle fibers. *J. Gen. Physiol.* **2012**, *139* (4), 261–72.
- (76) Song, K. S.; Scherer, P. E.; Tang, Z.; Okamoto, T.; Li, S.; Chafel, M.; Chu, C.; Kohtz, D. S.; Lisanti, M. P. Expression of caveolin-3 in skeletal, cardiac, and smooth muscle cells. Caveolin-3 is a component of the sarcolemma and co-fractionates with dystrophin and dystrophin-associated glycoproteins. *J. Biol. Chem.* **1996**, *271* (25), 15160–5.
- (77) Weiss, N.; Couchoux, H.; Legrand, C.; Berthier, C.; Allard, B.; Jacquemond, V. Expression of the muscular dystrophy-associated caveolin-3(P104L) mutant in adult mouse skeletal muscle specifically alters the Ca(2+) channel function of the dihydropyridine receptor. *Pflugers Arch.* **2008**, *457* (2), 361–75.
- (78) Whiteley, G.; Collins, R. F.; Kitmitto, A. Characterization of the molecular architecture of human caveolin-3 and interaction with the skeletal muscle ryanodine receptor. *J. Biol. Chem.* **2012**, *287* (48), 40302–16.
- (79) Leberer, E.; Timms, B. G.; Campbell, K. P.; MacLennan, D. H. Purification, calcium binding properties, and ultrastructural localization of the 53,000- and 160,000 (sarcalumenin)-dalton glycoproteins of the sarcoplasmic reticulum. *J. Biol. Chem.* **1990**, *265* (17), 10118–24.
- (80) Mahaney, J. E.; Weis, C. P.; Grisham, C. M.; Kutchai, H. Antibodies against the 53 kDa glycoprotein inhibit the rotational dynamics of both the 53 kDa glycoprotein and the Ca(2+)-ATPase in the sarcoplasmic reticulum membrane. *Biochim. Biophys. Acta, Biomembr.* **1991**, *1064* (1), 55–68.
- (81) Yoshida, M.; Minamisawa, S.; Shimura, M.; Komazaki, S.; Kume, H.; Zhang, M.; Matsumura, K.; Nishi, M.; Saito, M.; Saeki, Y.; Ishikawa, Y.; Yanagisawa, T.; Takeshima, H. Impaired Ca²⁺ store functions in skeletal and cardiac muscle cells from sarcalumenin-deficient mice. *J. Biol. Chem.* **2005**, *280* (5), 3500–6.
- (82) Bodnar, D.; Geyer, N.; Ruzsnavszky, O.; Olah, T.; Hegyi, B.; Sztretye, M.; Fodor, J.; Dienes, B.; Balogh, A.; Papp, Z.; Szabo, L.; Muller, G.; Csernoch, L.; Szentesi, P. Hypermuscular mice with mutation in the myostatin gene display altered calcium signalling. *J. Physiol.* **2014**, *592* (6), 1353–65.
- (83) Zhao, X.; Yoshida, M.; Brotto, L.; Takeshima, H.; Weisleder, N.; Hirata, Y.; Nosek, T. M.; Ma, J.; Brotto, M. Enhanced resistance to fatigue and altered calcium handling properties of sarcalumenin knockout mice. *Physiol. Genomics* **2005**, *23* (1), 72–8.
- (84) Zhu, J.; Li, Y.; Shen, W.; Qiao, C.; Ambrosio, F.; Lavasani, M.; Nozaki, M.; Branca, M. F.; Huard, J. Relationships between transforming growth factor-beta1, myostatin, and decorin: implications for skeletal muscle fibrosis. *J. Biol. Chem.* **2007**, *282* (35), 25852–63.
- (85) Zhu, J.; Li, Y.; Lu, A.; Gharaibeh, B.; Ma, J.; Kobayashi, T.; Quintero, A. J.; Huard, J. Follistatin improves skeletal muscle healing after injury and disease through an interaction with muscle regeneration, angiogenesis, and fibrosis. *Am. J. Pathol.* **2011**, *179* (2), 915–30.
- (86) Bogdanovich, S.; Krag, T. O.; Barton, E. R.; Morris, L. D.; Whittemore, L. A.; Ahima, R. S.; Khurana, T. S. Functional improvement of dystrophic muscle by myostatin blockade. *Nature* **2002**, *420* (6914), 418–21.
- (87) Wagner, K. R.; McPherron, A. C.; Winik, N.; Lee, S. J. Loss of myostatin attenuates severity of muscular dystrophy in mdx mice. *Ann. Neurol.* **2002**, *52* (6), 832–6.
- (88) Qiao, C.; Li, J.; Jiang, J.; Zhu, X.; Wang, B.; Li, J.; Xiao, X. Myostatin propeptide gene delivery by adeno-associated virus serotype 8 vectors enhances muscle growth and ameliorates dystrophic phenotypes in mdx mice. *Hum. Gene Ther.* **2008**, *19* (3), 241–54.
- (89) Kovanen, V. Effects of ageing and physical training on rat skeletal muscle. An experimental study on the properties of collagen, laminin, and fibre types in muscles serving different functions. *Acta Physiol. Scand., Suppl.* **1989**, *577*, 1–56.

(90) Cai, C.; Masumiya, H.; Weisleder, N.; Matsuda, N.; Nishi, M.; Hwang, M.; Ko, J. K.; Lin, P.; Thornton, A.; Zhao, X.; Pan, Z.; Komazaki, S.; Brotto, M.; Takeshima, H.; Ma, J. MG53 nucleates assembly of cell membrane repair machinery. *Nat. Cell Biol.* **2009**, *11* (1), 56–64.

(91) Goldring, K.; Jones, G. E.; Thiagarajah, R.; Watt, D. J. The effect of galectin-1 on the differentiation of fibroblasts and myoblasts in vitro. *J. Cell Sci.* **2002**, *115* (Pt 2), 355–66.

(92) Georgiadis, V.; Stewart, H. J.; Pollard, H. J.; Tavsanoglu, Y.; Prasad, R.; Horwood, J.; Deltour, L.; Goldring, K.; Poirier, F.; Lawrence-Watt, D. J. Lack of galectin-1 results in defects in myoblast fusion and muscle regeneration. *Dev. Dyn.* **2007**, *236* (4), 1014–24.

(93) Cerri, D. G.; Rodrigues, L. C.; Stowell, S. R.; Araujo, D. D.; Coelho, M. C.; Oliveira, S. R.; Bizario, J. C.; Cummings, R. D.; Dias-Baruffi, M.; Costa, M. C. Degeneration of dystrophic or injured skeletal muscles induces high expression of Galectin-1. *Glycobiology* **2008**, *18* (11), 842–50.

(94) Eu, J. P.; Hare, J. M.; Hess, D. T.; Skaf, M.; Sun, J.; Cardenas-Navina, I.; Sun, Q.-A.; Dewhirst, M.; Meissner, G.; Stamler, J. S. Concerted regulation of skeletal muscle contractility by oxygen tension and endogenous nitric oxide. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (25), 15229–15234.

(95) Kobayashi, Y. M.; Rader, E. P.; Crawford, R. W.; Iyengar, N. K.; Thedens, D. R.; Faulkner, J. A.; Parikh, S. V.; Weiss, R. M.; Chamberlain, J. S.; Moore, S. A.; Campbell, K. P. Sarcolemma-localized nNOS is required to maintain activity after mild exercise. *Nature* **2008**, *456* (7221), 511–5.

(96) Percival, J. M.; Anderson, K. N.; Gregorevic, P.; Chamberlain, J. S.; Froehner, S. C. Functional deficits in nNOS^{mu}-deficient skeletal muscle: myopathy in nNOS knockout mice. *PLoS One* **2008**, *3* (10), e3387.

(97) Lin, Y.; Ishikawa, R.; Okagaki, T.; Ye, L. H.; Kohama, K. Stimulation of the ATP-dependent interaction between actin and myosin by a myosin-binding fragment of smooth muscle caldesmon. *Cell Motil. Cytoskeleton* **1994**, *29* (3), 250–8.

(98) Conti, A.; Riva, N.; Pesca, M.; Iannaccone, S.; Cannistraci, C. V.; Corbo, M.; Previtali, S. C.; Quattrini, A.; Alessio, M. Increased expression of Myosin binding protein H in the skeletal muscle of amyotrophic lateral sclerosis patients. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2014**, *1842* (1), 99–106.

(99) Miao, Y.; Yang, J.; Xu, Z.; Jing, L.; Zhao, S.; Li, X. RNA sequencing identifies upregulated kyphoscoliosis peptidase and phosphatidic acid signaling pathways in muscle hypertrophy generated by transgenic expression of myostatin propeptide. *Int. J. Mol. Sci.* **2015**, *16* (4), 7976–94.

Comparative Proteomic and Transcriptomic Analysis of Follistatin-induced Skeletal Muscle Hypertrophy

Caroline Barbé†, Marine Gueugneau†, Stéphanie Devassine‡, Pascale Lause†, Olga Pléchakova‡, Fabrice Bray‡, Caroline Tokarski‡, Christian Rolando‡ and Jean-Paul Thissen†*

†Pole of Endocrinology, Diabetes and Nutrition, Institute of Experimental and Clinical Research,
Université Catholique de Louvain, 1200 Brussels, Belgium

‡Miniaturisation pour la Synthèse, l'Analyse & la Protéomique (MSAP), USR CNRS 3290,
Université de Lille 1 Sciences et Technologies, 59655 Villeneuve d'Ascq, France

SUPPORTING INFORMATION

TABLE OF CONTENTS

Supplemental Methods – *Experimental Procedures Details*. Protein digestion eFASP with passivated ultra-filtration unit. **(PDF 1 page S3)**

Supplemental Table S1 – *BlastP validations for one Single Peptide-Based Protein Identifications*. BlastP interrogations for each single peptide of proteins identified with one single peptide and found as differentially expressed between mTrFS and WT mice. **(PDF 1 page S4)**

Supplemental Table S2 – *Proteomic Profile of TA Muscle in Response to FS*. List of the 115 proteins significantly regulated in TA muscle of mTrFS mice compared to WT mice. **(PDF 2)**

Supplemental Table S3 – *Proteomic Profile of SOL Muscle in Response to FS*. List of the 44 proteins significantly regulated in SOL muscle of mTrFS mice compared to WT mice. **(PDF 3)**

Supplemental Table S4 – *Transcriptomic Profile of Fast TA Muscle in Response to FS*. List of the 315 genes significantly regulated in TA muscle of mTrFS mice compared to WT mice. **(PDF 4)**

Supplemental Figure S1 – *Detailed information for one Single Peptide-Based Protein Identifications*. Detailed information about the MS/MS spectrum, peptide sequence and precursor m/z for each single peptide of proteins identified with one single peptide and found as differentially expressed between mTrFS and WT mice. **(PDF 1 page S5-S17)**

Supplemental Figure S2 – *Quality of the Proteomic Analysis*. Box plots representing Log₂ signal intensities for each replicate from TA muscle of WT mice indicating high technical reproducibility and very low biological variance and PCA of TA and SOL muscles from WT mice. **(PDF 1 page S18)**

SUPPLEMENTAL METHODS

Digestion eFASP With Passivated Ultra filtration Unit

UF filters from Amicon® units (10 kDa cutoff limit; Millipore) were incubated overnight 5% v/v in TWEEN-20 and then rinsed thoroughly by three immersions in MS-grade water. Proteins samples (50 µg) were solubilized and reduced in buffer A (4% SDS, 0.2 % deoxycholic acid [DCA], 50 mM DTT, 200 mM ammonium bicarbonate [ABC]) overnight at 4°C. After centrifugation, supernatants were mixed with buffer exchange B (8 M urea, 0.2 % DCA, 100 mM ABC, pH 8), transferred on the Amicon® units and centrifuged. The filtrates were discarded and an additional volume of buffer B was deposited in each unit so as to repeat the centrifugation step three times. Then, the reduced proteins were alkylated with buffer C (8 M urea, 50 mM iodoacetamide, and 100 mM ABC, pH 8). After centrifugation, the filtrates were discarded and a volume of buffer B was deposited in each unit so as to repeat the centrifugation step two times. Three more exchange steps were repeated with eFASP digestion buffer D (50 mM ABC, 0.2% DCA pH 8). Finally, buffer C was added to each Amicon® units with 1 µg of trypsin (1:50 w/w) and digestion of proteins proceeded overnight at 37°C. Peptides were recovered by rinsing with a solution of 50 mM ABC and centrifugation. Finally, the peptides-containing filtrates were cleared from organic residues and volatile salts and peptides were extracted and evaporated before being dissolved in 10 µl of water milliQ. Samples were then diluted tenfold in buffer A of nano-HPLC (5% acetonitrile and 0.1% formic acid) and each sample, namely each biological replicate, was injected three times in HPLC instrument to be analyzed in triplicate.

SUPPLEMENTAL TABLES

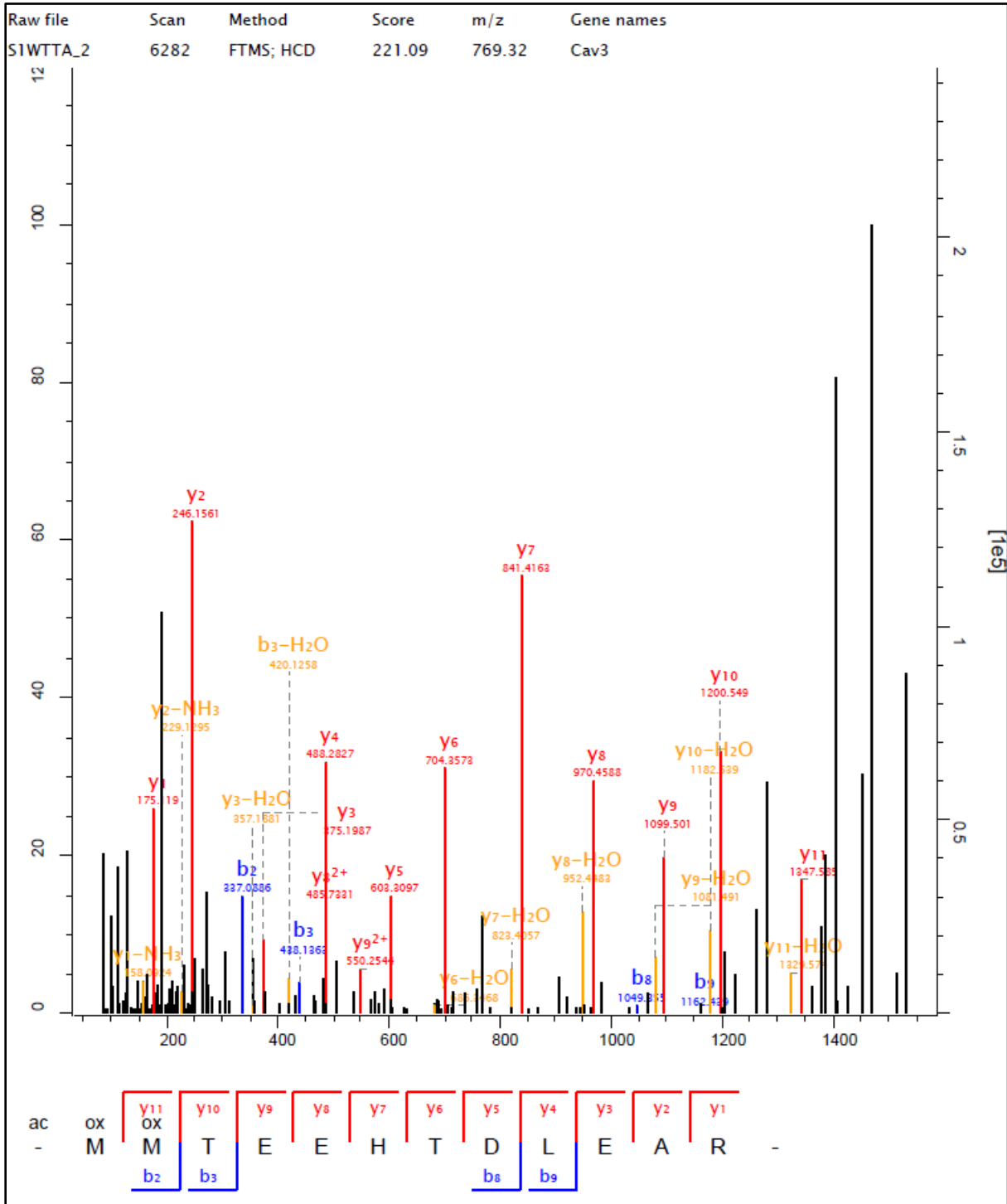
Table S1. BlastP validations for one Single Peptide-Based Protein Identified Differentially Expressed Between mTrFS and WT mice

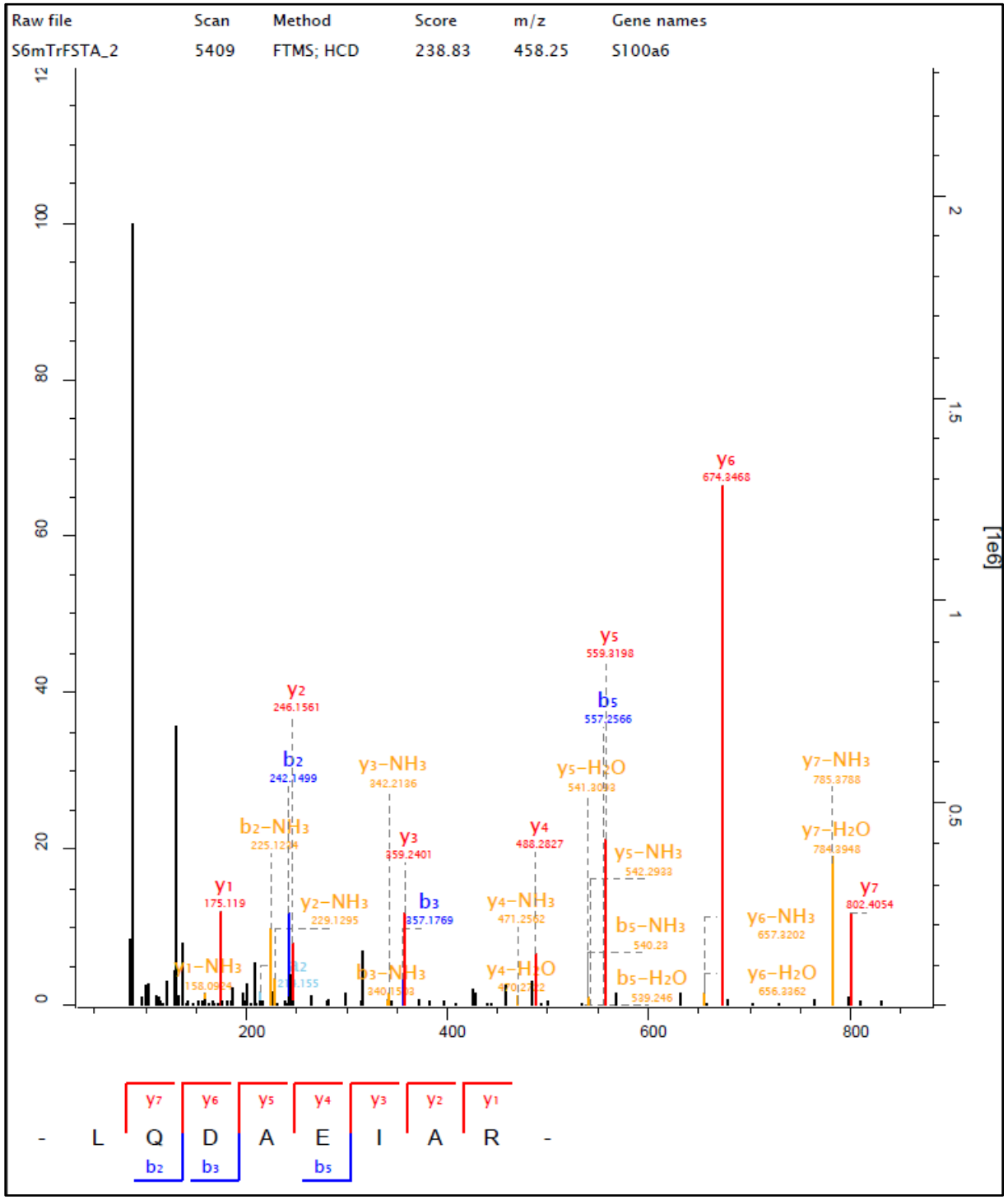
Accession	Gene	Score	m/z	Peptide	BlastP				
					Max score	Total score	Query cover	E value	Ident.
TA Muscle									
Soluble Fraction									
P51637	Cav3	221.09	769.32	MMTEEHTDLEAR	44.8	44.8	100%	0.000	100%
P14069	S100a6	238.83	458.25	LQDAEIAR	28.2	28.2	100%	0.045	100%
P70168	Kpnb1	130.56	613.34	VLANPGNSQVAR	39.2	39.2	100%	0.000	100%
Q922Y1	Ubxn1	44.85	773.05	SSPPATDPGPVPSSPSQEPPTKR	73.6	73.6	100%	0.000	100%
Q9CQX8	Mrps36	192.15	614.63	RKPMSEQEMEFIQR	52.4	52.4	100%	0.000	100%
Q9JK42	Pdk2	163.32	489.25	ALSTDSVER	29.9	29.9	100%	0.013	100%
Solubilized Fraction									
P03921	Mtnd5	203.99	947.96	STSTLHTNM TTL TTNQK	56.6	99.7	100%	0.000	100%
Q3U381	Zfp692	293.73	470.25	HAETAAALR	29.9	29.9	100%	0.013	100%
P62264	Rps14	160.18	527.79	TPGPGAQSALR	35.4	35.4	100%	0.000	100%
SOL Muscle									
Soluble Fraction									
P99027	Rplp2	66.27	621.84	NIEDVIAQGVGK	40.1	40.1	100%	0.000	100%
Q8C1A5	Thop1	112.13	566.82	FKQEGVLSPK	34.1	34.1	100%	0.001	100%
Q8QZS1	Hibch	243.13	614.13	AVLIDKDQTPK	37.5	37.5	100%	0.000	100%
Q99LY9	Ndufs5	158.58	586.29	YTPPPHHSGREPRP	53.7	53.7	100%	0.000	100%
Solubilized Fraction									
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

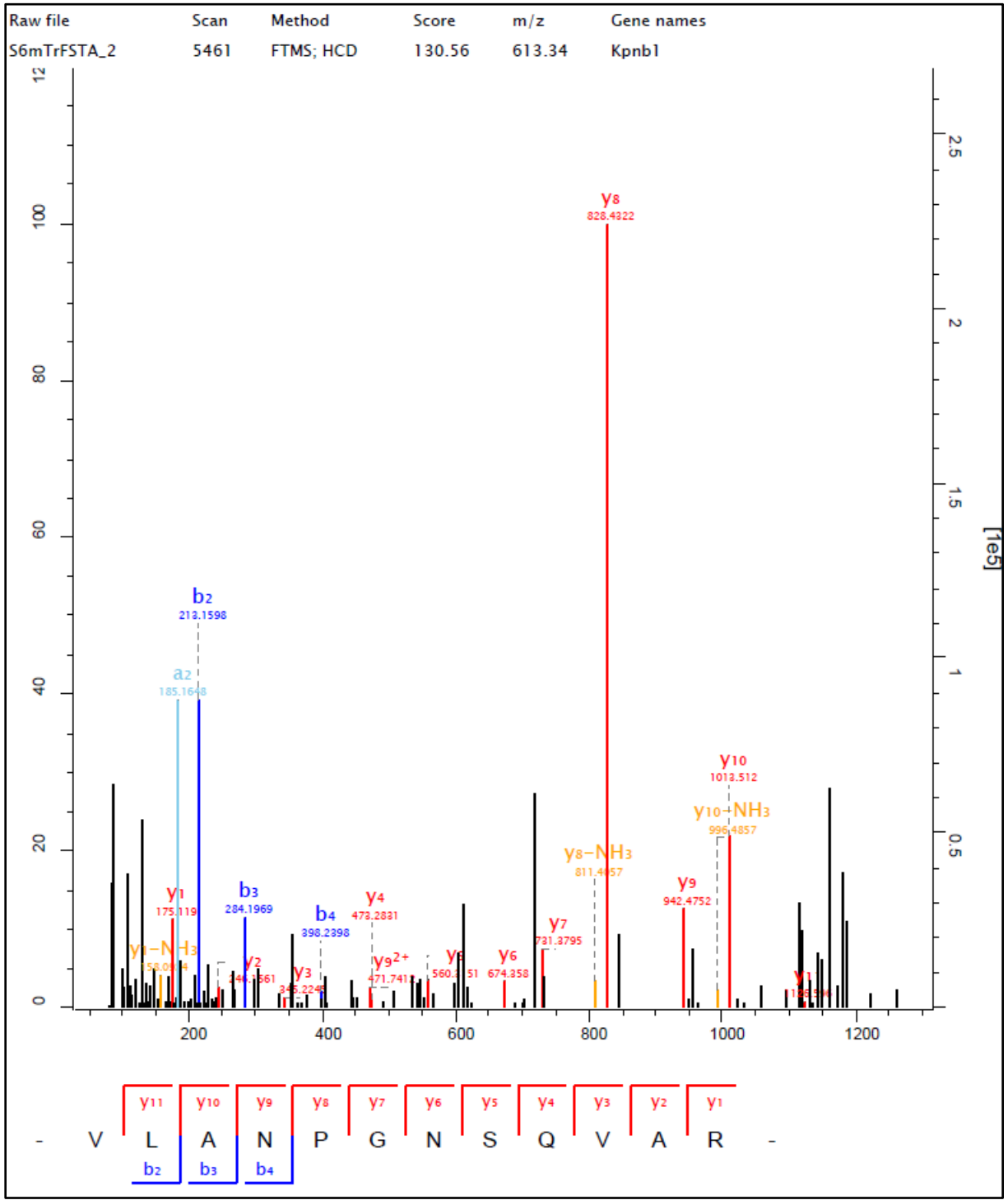
Table S1. BlastP validations for one Single Peptide-Based Protein Identifications. For all the proteins that were identified on the basis of one single peptide and that were found as differentially regulated by FS, each single peptide was validated by BlastP interrogation, as reported in the table above.

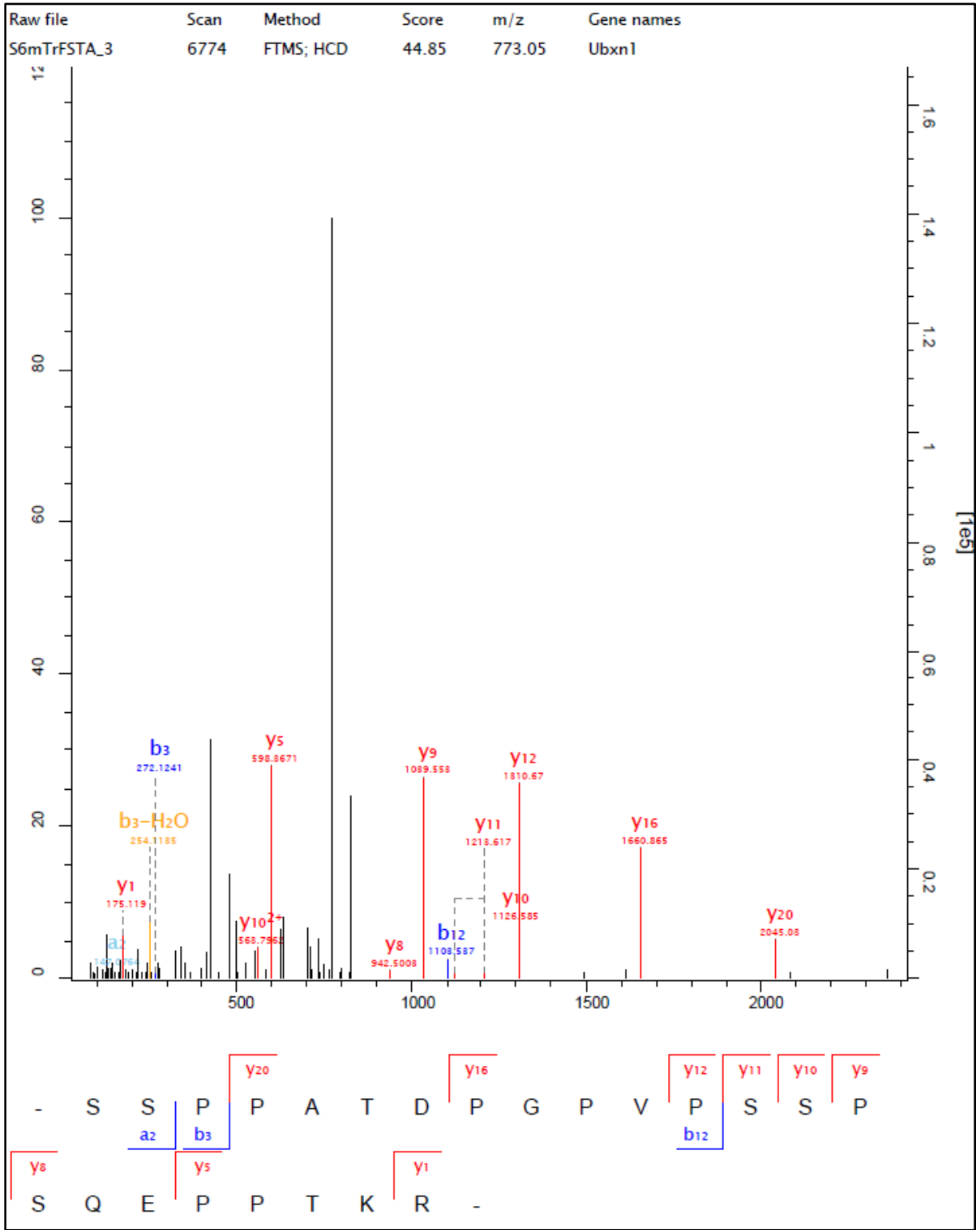
SUPPLEMENTAL FIGURES

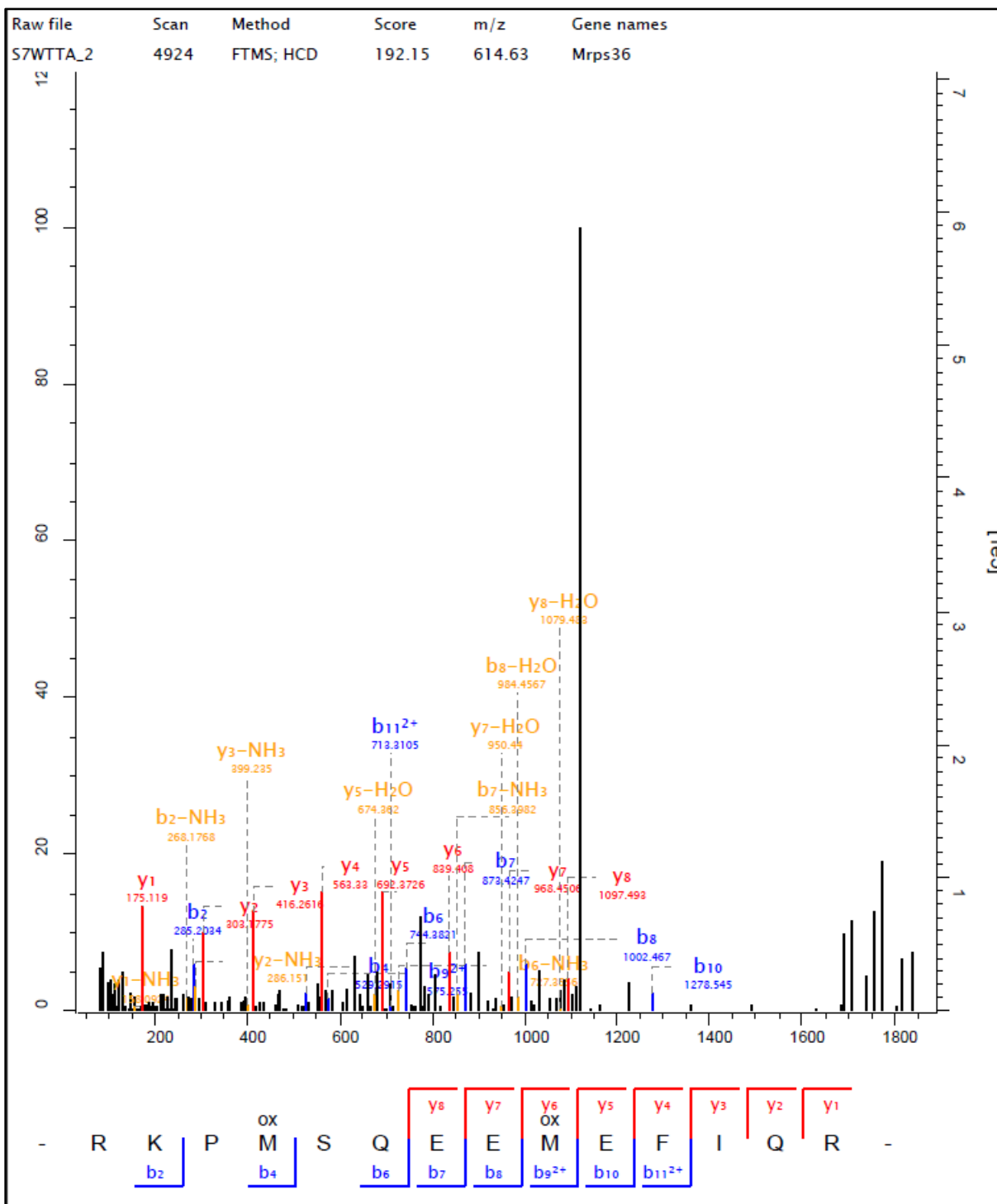
Supplemental Figure S1

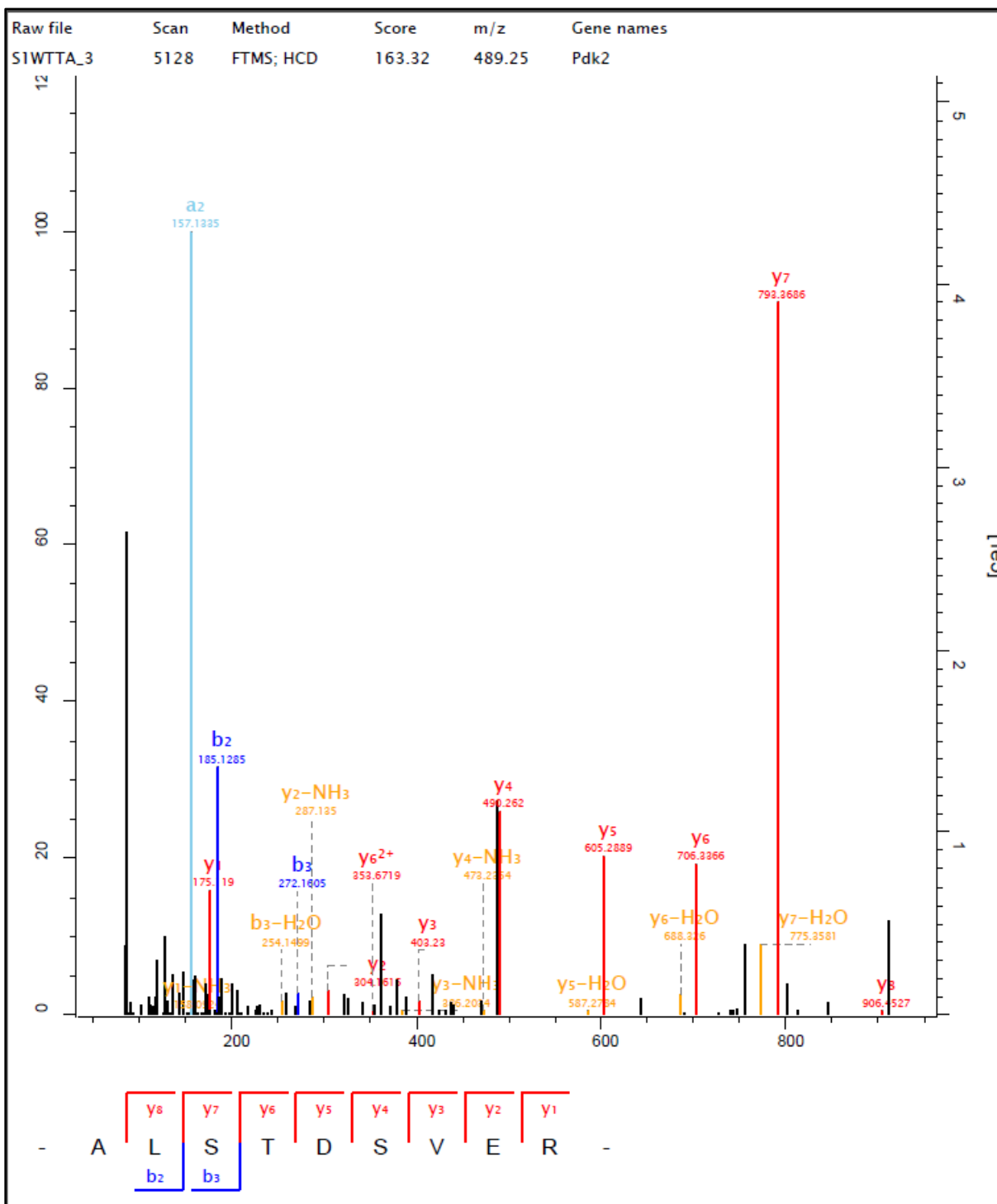


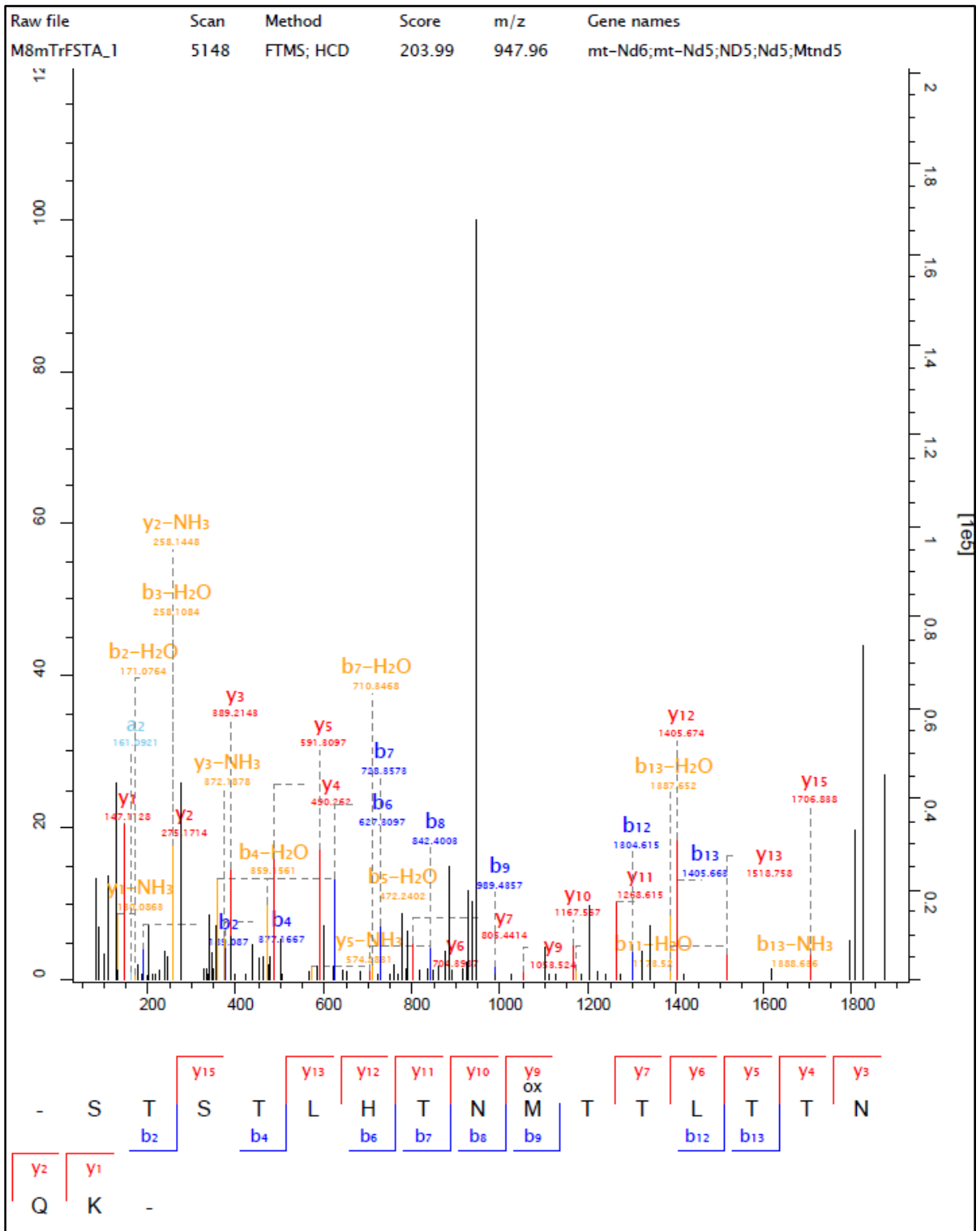


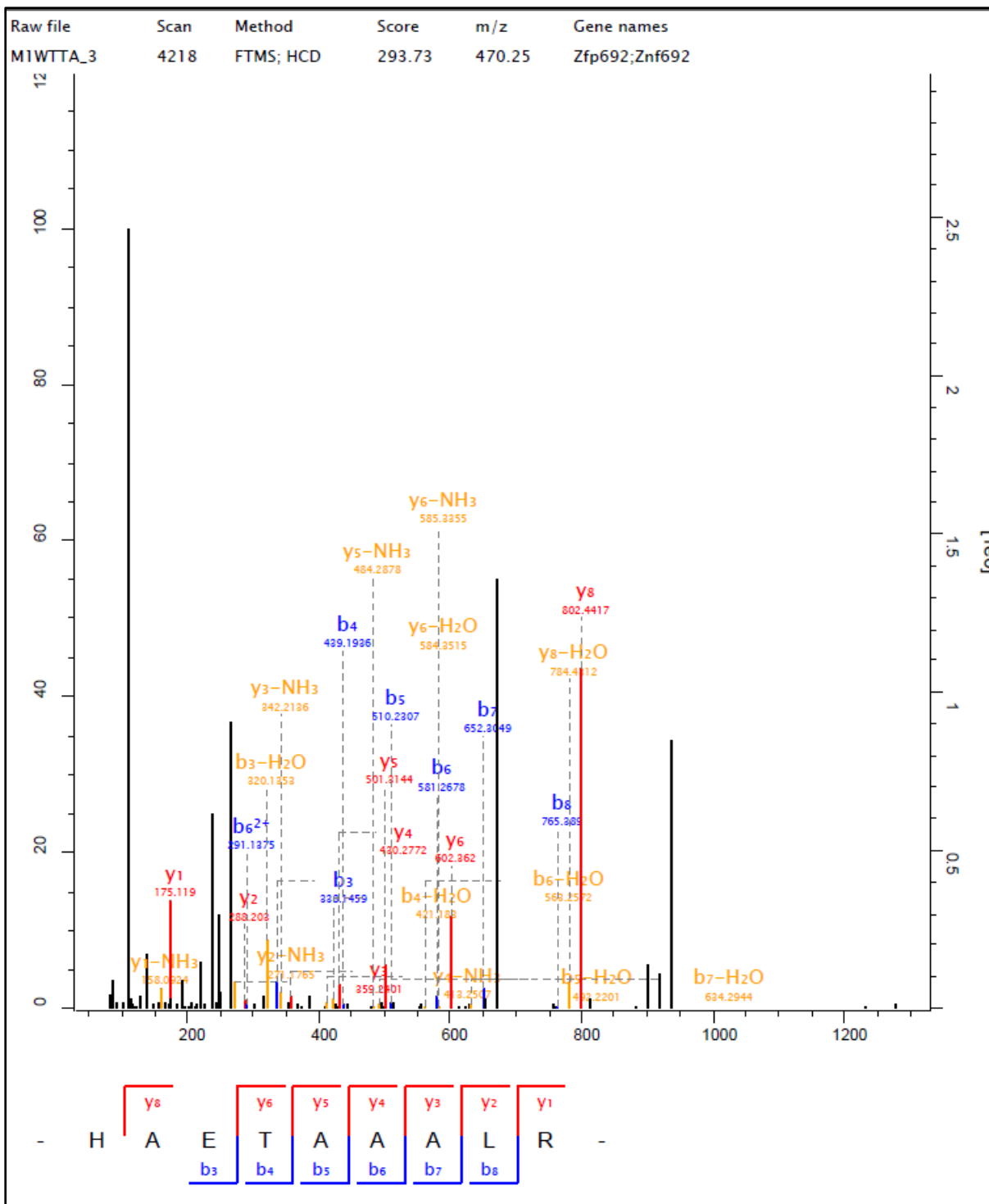


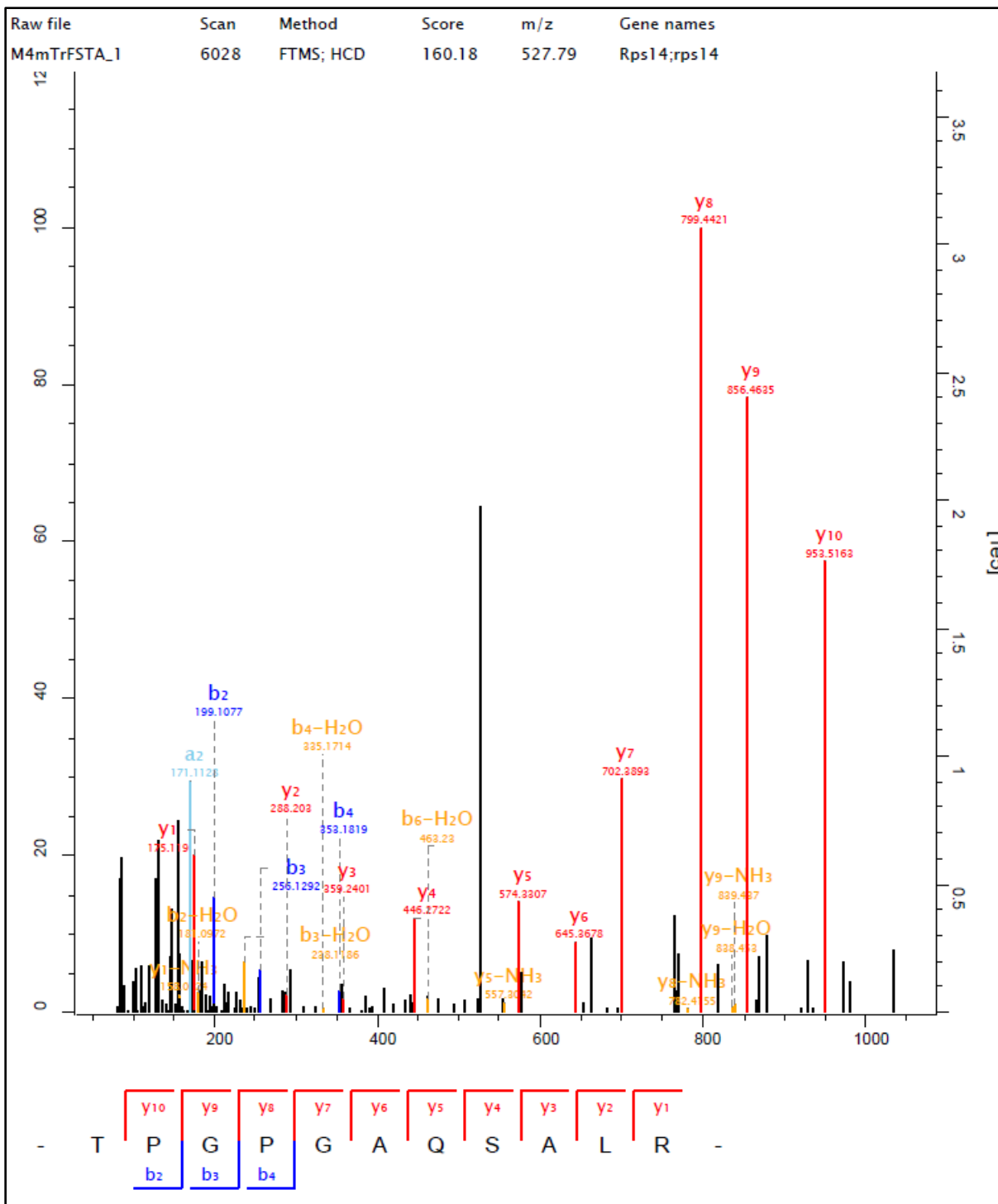


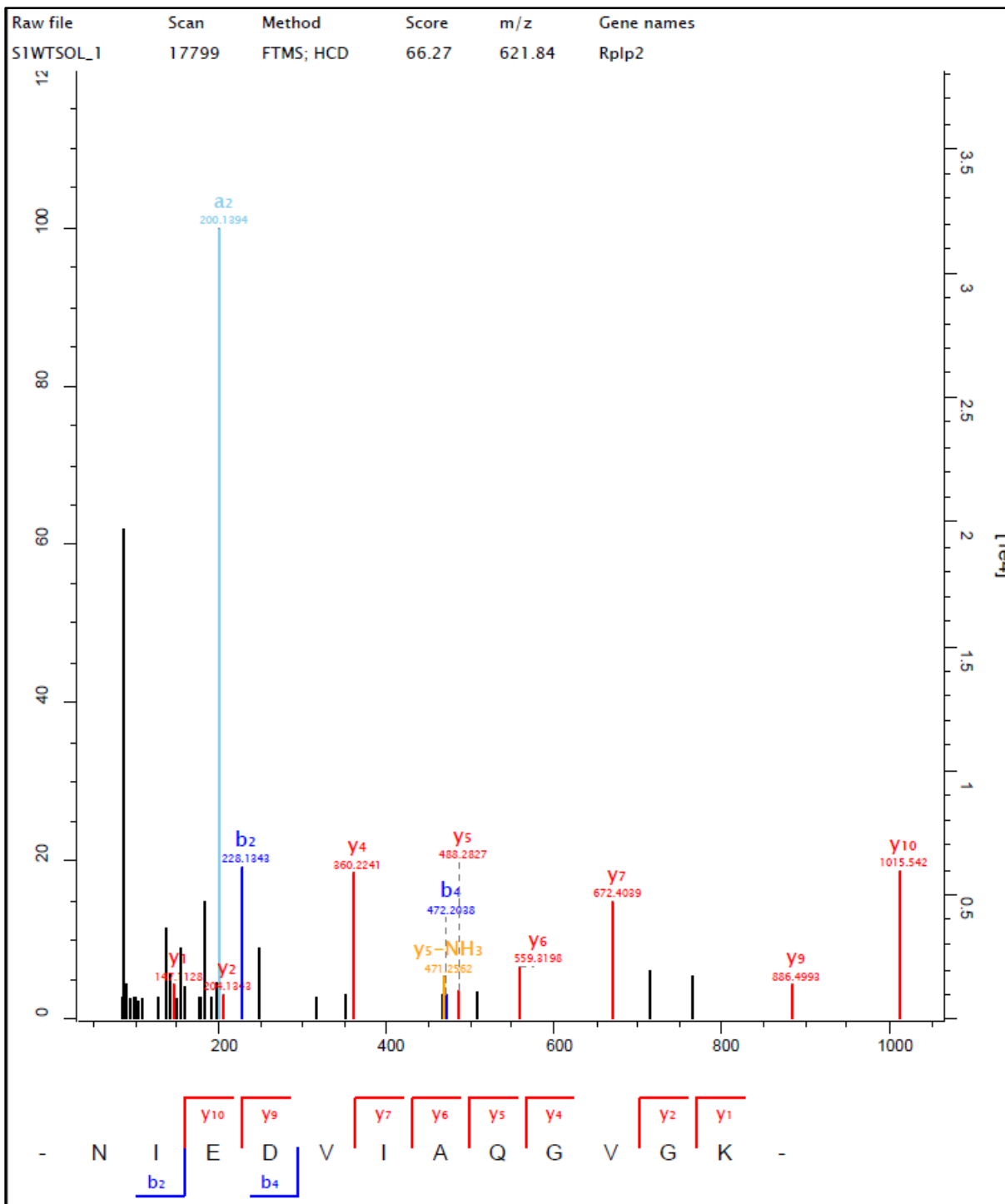


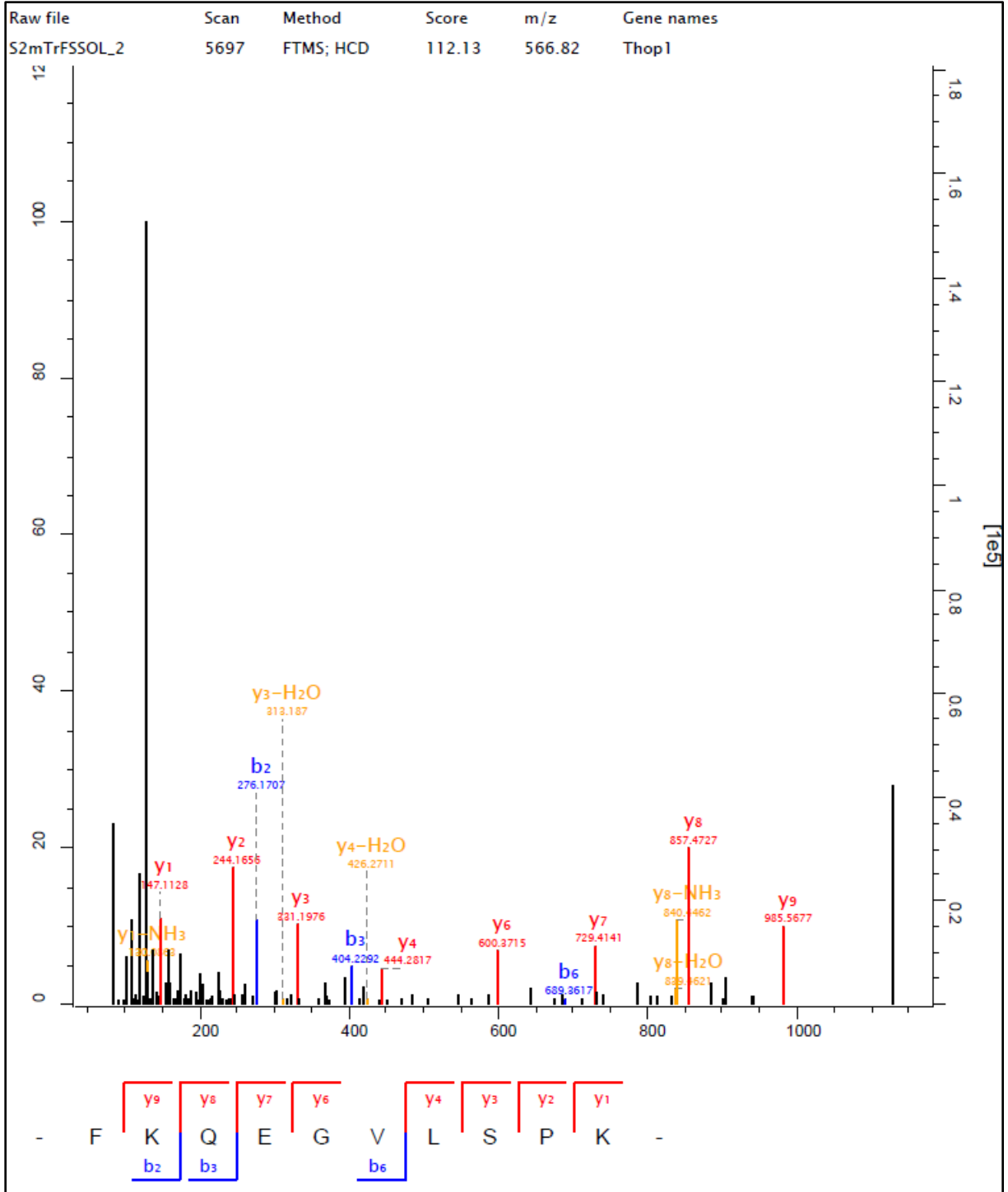


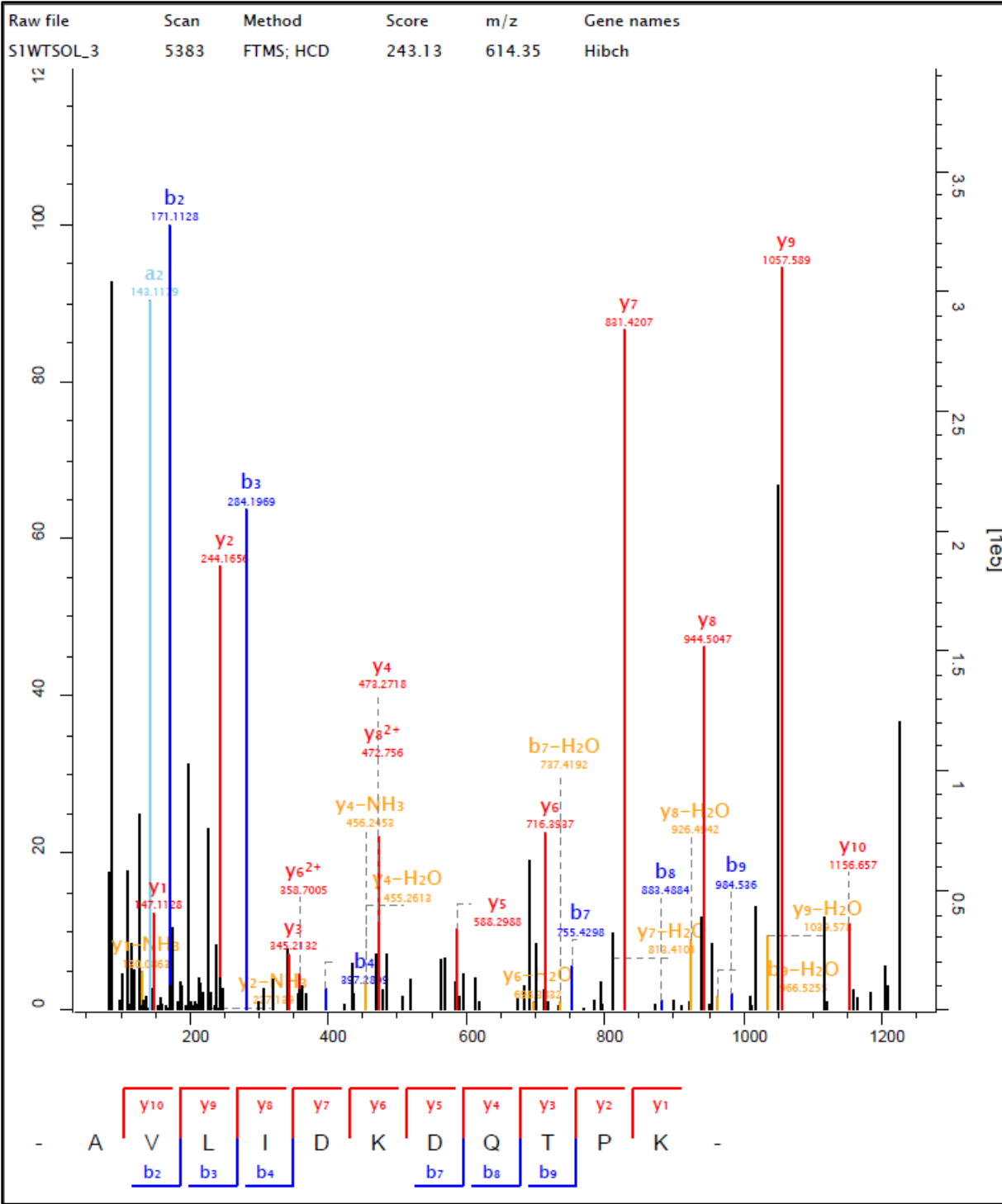


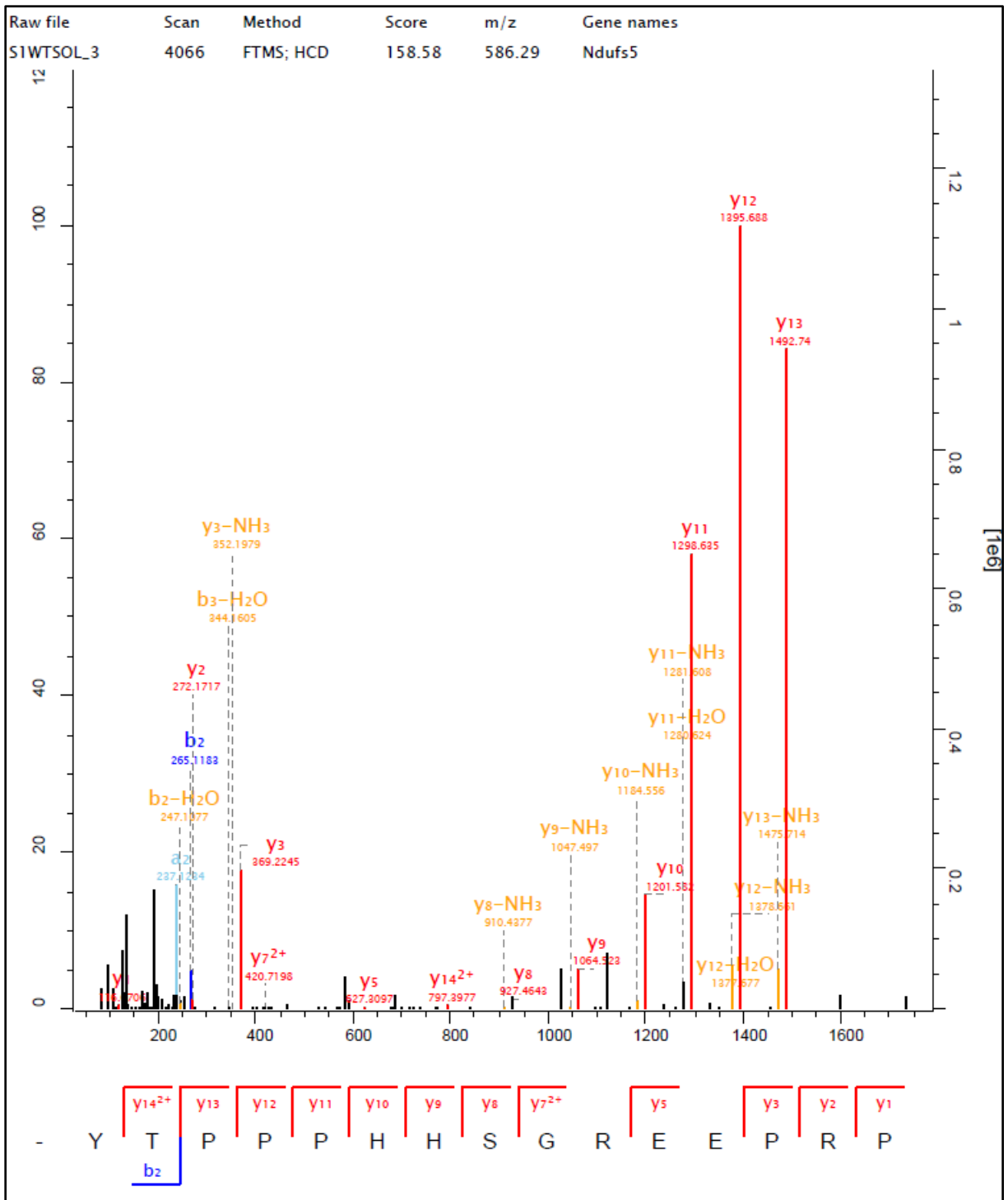












Supplemental Figure S1. Detailed information for one Single Peptide-Based Protein Identifications. Detailed information about the MS/MS spectrum, peptide sequence and precursor m/z for each single peptide of proteins identified with one single peptide and found as differentially expressed between mTrFS and WT mice.

Supplemental Figure S2

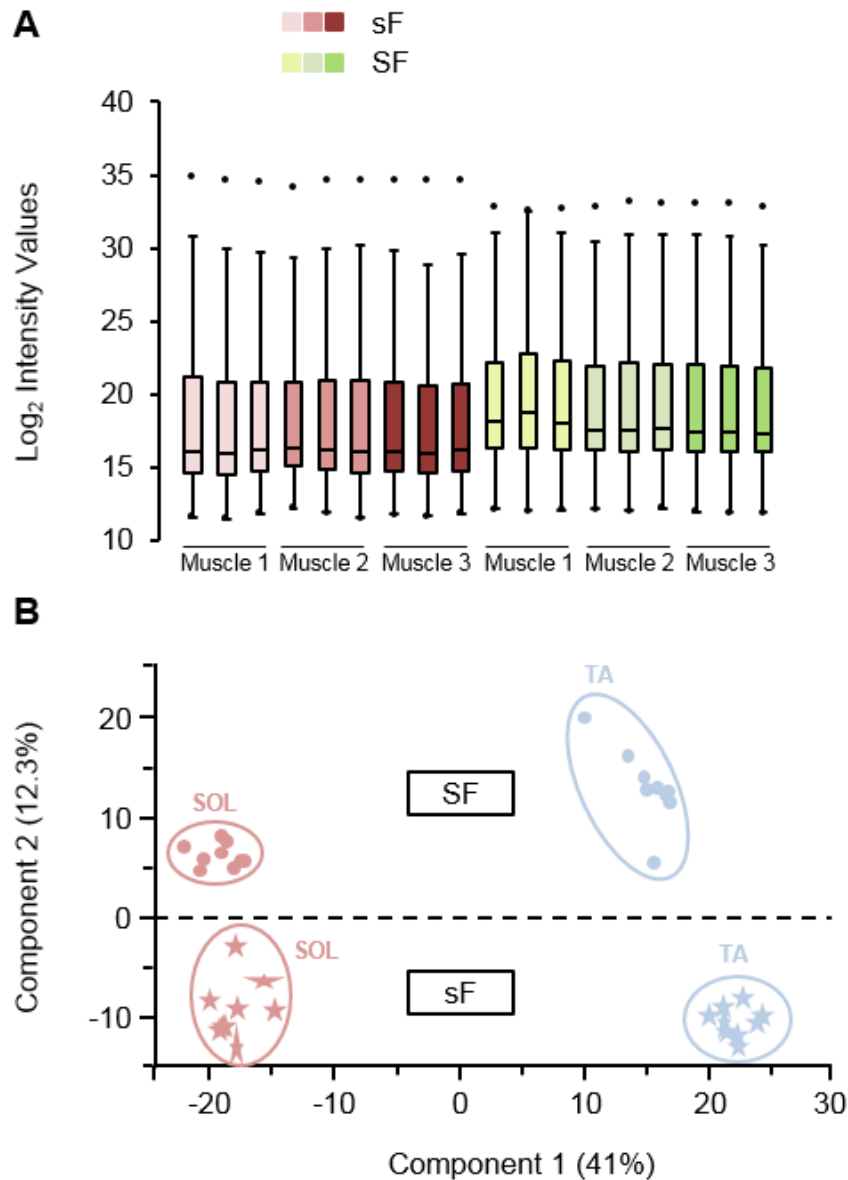


Figure S2. *Quality of the Proteomic Analysis.* (A) Box plots representing Log₂ signal intensities for each replicate from sF (in red) and SF (in green) of TA muscle of WT mice show high similarity among all runs, indicating high technical reproducibility and very low biological variance. The three muscle biological replicates (Muscle 1-3) are presented in different grades of the same color while the three technological replicates are reported exactly in the same color. (B) PCA of WT muscles clearly discriminates the proteomic profiles of the fast TA muscle (blue) from the slow SOL muscle (red) in sF (stars) and SF (circles) fractions.

Table S2 - Proteomic Profile of TA muscle in Response to FS

# Accession	Symbol	Name	Anova (P)	Fold change	Score	%Coverage	# Peptides	Main Biological Process	Main Cellular Component
O08539	BIN1	Myc box-dependent-interacting protein 1	0.013	1.76	323.3	19.7	8	Apoptosis, Endocytosis	Cytoplasm, Nucleus
Q9WTR5	CDH13	Cadherin-13	0.044	0.68	11.0	9.1	2	Cell adhesion	Membrane
P63038	HSPD1	60 kDa heat shock protein, mitochondrial	0.024	1.44	16.3	10.1	5	Detoxification, cytoprotection	Mitochondrion
P20029	HSPA5	78 kDa glucose-regulated protein	0.038	1.51	87.0	18.7	10	Detoxification, cytoprotection	Endoplasmic reticulum
Q61316	HSPA4	Heat shock 70 kDa protein 4	0.037	1.55	3.2	3.1	2	Detoxification, cytoprotection	Cytoplasm
P10649	GSTM1	Glutathione S-transferase Mu 1	0.014	1.68	39.0	45.8	9	Detoxification, cytoprotection	Cytoplasm
P11499	HSP90AB1	Heat shock protein HSP 90-beta	0.018	1.69	218.3	25.6	20	Detoxification, cytoprotection	Cytoplasm
Q5EBG6	HSPB6	Heat shock protein beta-6	0.005	1.70	41.7	29.0	4	Detoxification, cytoprotection	Cytoplasm
P14602	HSPB1	Heat shock protein beta-1	0.009	1.72	182.7	44.5	10	Detoxification, cytoprotection	Cytoplasm
P97447	FHL1	Four and a half LIM domains protein 1	0.009	0.64	11.1	12.5	4	Developmental process	Cytoplasm, Nucleus
P50247	AHCY	Adenosylhomocysteinase	0.007	1.54	20.1	17.9	5	Energy metabolism, Amino-acid biosynthesis	Cytoplasm
Q9DCW4	ETFB	Electron transfer flavoprotein subunit beta	0.044	1.46	17.7	25.1	5	Energy metabolism, Electron transfer	Mitochondrion
P13707	GPD1	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	0.004	2.08	323.3	53.7	20	Energy metabolism, G3P NADH Shuttle	Cytoplasm
Q64521	GPD2	Glycerol-3-phosphate dehydrogenase, mitochondrial	0.003	1.54	27.2	13.8	9	Energy metabolism, G3P NADH Shuttle	Mitochondrion
Q9D6Y9	GBE1	1,4-alpha-glucan-branching enzyme	0.019	1.54	27.7	7.8	5	Energy metabolism, Glycogen biosynthesis	Cytoplasm
Q8C7E7	STBD1	Starch-binding domain-containing protein 1	0.028	1.54	6.4	7.9	2	Energy metabolism, Glycogenolysis	Cytoplasm
Q7TSH2	PHKB	Phosphorylase b kinase regulatory subunit beta	0.046	1.68	202.7	9.4	10	Energy metabolism, Glycogenolysis	Cytoplasm
P16125	LDHB	L-lactate dehydrogenase B chain	0.004	0.64	132.9	34.4	13	Energy metabolism, Glycolysis	Cytoplasm
P09411	PGK1	Phosphoglycerate kinase 1	0.015	1.83	323.3	52.5	25	Energy metabolism, Glycolysis	Cytoplasm
P06745	GPI	Glucose-6-phosphate isomerase	0.004	1.89	323.3	41.8	23	Energy metabolism, Glycolysis	Cytoplasm
P47857	PFKM	ATP-dependent 6-phosphofructokinase, muscle type	0.014	1.93	323.3	45.1	34	Energy metabolism, Glycolysis	Cytoplasm
P06151	LDHA	L-lactate dehydrogenase A chain	0.004	2.06	323.3	54.2	27	Energy metabolism, Glycolysis	Cytoplasm
P21550	ENO3	Beta-enolase	0.005	2.50	323.3	69.1	42	Energy metabolism, Glycolysis	Cytoplasm
Q9JK42	PDK2	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 2, mitochondrial	0.013	0.68	2.3	2.2	1	Energy metabolism, Krebs cycle	Mitochondrion
Q99K10	ACO2	Aconitate hydratase, mitochondrial	0.004	0.53	323.3	38.2	31	Energy metabolism, Krebs cycle	Mitochondrion
Q9CZU6	CS	Citrate synthase, mitochondrial	0.007	0.54	253.4	29.5	13	Energy metabolism, Krebs cycle	Mitochondrion
Q9D6R2	IDH3A	Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	0.009	0.56	177.7	37.2	13	Energy metabolism, Krebs cycle	Mitochondrion
Q60597	OGDH	2-oxoglutarate dehydrogenase, mitochondrial	0.004	0.57	323.3	25.0	25	Energy metabolism, Krebs cycle	Mitochondrion
P35486	PDHA1	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	0.004	0.58	218.9	25.1	10	Energy metabolism, Krebs cycle	Mitochondrion
P54071	IDH2	Isocitrate dehydrogenase [NADP], mitochondrial	0.010	0.64	260.4	19.9	8	Energy metabolism, Krebs cycle	Mitochondrion
P97807	FH	Fumarate hydratase, mitochondrial	0.010	1.50	19.6	13.8	7	Energy metabolism, Krebs cycle	Mitochondrion
Q99JY0	HADHB	Trifunctional enzyme subunit beta, mitochondrial	0.005	1.46	38.1	29.9	12	Energy metabolism, Lipid biosynthesis	Mitochondrion
O88492	PLIN4	Perilipin-4	0.005	1.57	7.0	13.3	3	Energy metabolism, Lipid storage	Lipid droplet
P05202	GOT2	Aspartate aminotransferase, mitochondrial	0.012	0.55	175.0	39.1	18	Energy metabolism, Malate-aspartate NADH Shuttle	Mitochondrion
P14152	MDH1	Malate dehydrogenase, cytoplasmic	0.016	0.56	143.3	39.2	12	Energy metabolism, Malate-aspartate NADH Shuttle	Cytoplasm
P08249	MDH2	Malate dehydrogenase, mitochondrial	0.004	0.66	69.2	53.8	16	Energy metabolism, Malate-aspartate NADH Shuttle	Mitochondrion
P05201	GOT1	Aspartate aminotransferase, cytoplasmic	0.029	0.72	28.2	19.9	6	Energy metabolism, Malate-aspartate NADH Shuttle	Cytoplasm
Q9DC69	NDUFA9	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	0.007	0.58	89.3	40.6	11	Energy metabolism, Oxidative phosphorylation	Mitochondrion
P00405	MTCO2	Cytochrome c oxidase subunit 2	0.009	0.61	33.2	26.4	5	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q9CR68	UQCRCF1	Cytochrome b-c1 complex subunit Rieske, mitochondrial	0.010	0.61	36.6	33.9	11	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q8K2B3	SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	0.005	0.61	61.0	20.0	12	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q99LC3	NDUFA10	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10	0.012	0.64	27.2	28.2	8	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q91YT0	NDUFV1	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	0.013	0.65	52.1	24.4	10	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q9DCJ5	NDUFA8	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	0.020	0.66	36.9	40.7	6	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q91VD9	NDUFS1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	0.021	0.66	103.4	37.6	24	Energy metabolism, Oxidative phosphorylation	Mitochondrion
P03921	MTND5	NADH-ubiquinone oxidoreductase chain 5	0.010	0.69	16.0	21.2	1	Energy metabolism, Oxidative phosphorylation	Mitochondrion
O09111	NDUFB11	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	0.018	0.70	4.6	10.6	2	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q9D6J6	NDUFV2	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	0.005	0.70	16.8	27.0	6	Energy metabolism, Oxidative phosphorylation	Mitochondrion

Q9CQ54	NDUFC2	NADH dehydrogenase [ubiquinone] 1 subunit C2	0.030	0.74	4.8	15.0	2	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q9CQ75	NDUFA2	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2	0.020	1.30	9.0	31.3	3	Energy metabolism, Oxidative phosphorylation	Mitochondrion
P28650	ADSSL1	Adenylosuccinate synthetase isozyme 1	0.004	0.58	107.5	39.4	19	Energy metabolism, Purine nucleotide cycle	Cytoplasm
P51174	ACADL	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	0.032	0.58	279.2	29.3	13	Energy metabolism, β -oxydation	Mitochondrion
Q8QZT1	ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	0.017	0.60	133.5	20.8	7	Energy metabolism, β -oxydation	Mitochondrion
P45952	ACADM	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	0.014	0.68	323.3	20.7	9	Energy metabolism, β -oxydation	Mitochondrion
Q07417	ACADS	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	0.022	0.75	6.2	10.7	4	Energy metabolism, β -oxydation	Mitochondrion
P11087	COL1A1	Collagen alpha-1(I) chain	0.004	0.57	323.3	46.0	37	Extracellular structure organization	Extracellular matrix
Q01149	COL1A2	Collagen alpha-2(I) chain	0.004	0.58	323.3	37.5	29	Extracellular structure organization	Extracellular matrix
P50608	FMOD	Fibromodulin	0.006	0.63	57.0	4.8	2	Extracellular structure organization	Extracellular matrix
P51885	LUM	Lumican	0.010	0.64	3.4	8.3	3	Extracellular structure organization	Extracellular matrix
P28654	DCN	Decorin	0.004	0.73	5.0	5.4	2	Extracellular structure organization	Extracellular matrix, Serum
Q9WV35	APOBEC2	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 2	0.031	1.55	24.6	37.9	6	Gene regulation	Nucleus
Q1XH17	TRIM72	Tripartite motif-containing protein 72	0.020	1.65	64.5	16.1	8	Muscle repair and regeneration	Membrane
P16045	LGALS1	Galectin-1	0.004	1.90	198.4	42.2	6	Muscle repair and regeneration	Serum
Q5SX40	MYH1	Myosin-1	0.003	0.47	323.3	68.1	206	Myofilaments and cytoskeleton	Sarcomere
P05977	MYL1	Myosin light chain 1/3, skeletal muscle isoform	0.010	0.47	323.3	68.6	13	Myofilaments and cytoskeleton	Sarcomere
I6L963	MYH2	Myosin-2	0.004	0.47	323.3	66.7	193	Myofilaments and cytoskeleton	Sarcomere
Q9JI91	ACTN2	Alpha-actinin-2	0.003	0.52	323.3	65.4	57	Myofilaments and cytoskeleton	Sarcomere
P97457	MYL6P	Myosin regulatory light chain 2, skeletal muscle isoform	0.004	0.54	267.1	76.3	17	Myofilaments and cytoskeleton	Sarcomere
D3YU50	MYBPC1	Myosin-binding protein C, slow-type	0.003	0.56	323.3	42.0	48	Myofilaments and cytoskeleton	Sarcomere
P09542	MYL3	Myosin light chain 3	0.003	0.58	46.9	41.2	8	Myofilaments and cytoskeleton	Sarcomere
Q9JLH8	TMOD4	Tropomodulin-4	0.010	1.34	15.4	14.8	4	Myofilaments and cytoskeleton	Sarcomere
P70402	MYBPH	Myosin-binding protein H	0.004	1.43	17.0	15.3	5	Myofilaments and cytoskeleton	Sarcomere
Q8BWB1	SYNPO2L	Synaptopodin 2-like protein	0.029	1.44	16.0	14.4	8	Myofilaments and cytoskeleton	Sarcomere
P20152	VIM	Vimentin	0.027	1.47	29.4	22.3	11	Myofilaments and cytoskeleton	Sarcomere
A2AAJ9	OBSCN	Obscurin	0.020	1.53	13.4	4.2	8	Myofilaments and cytoskeleton	Sarcomere
P63028	TPT1	Translationally-controlled tumor protein	0.039	1.56	40.3	21.5	3	Myofilaments and cytoskeleton	Sarcomere
Q9QZ47	TNNT3	Troponin T, fast skeletal muscle	0.005	0.51	323.3	44.4	22	Myofilaments and cytoskeleton	Sarcomere
P68368;P05213; P68373;P05214; P68369;Q9JJZ2	TUBA4A; TUBA1B/C/A; TUBA3A;TUBA8	Tubulin alpha-chain	0.046	1.63	11.6	15.3	6	Myofilaments and cytoskeleton	Sarcomere
P68372;P99024; Q9CWF2;Q7TMM9;Q9 D6F9;Q9CVR0	TUBB4B/A; TUBB5; TUBB2B/A	Tubulin beta-chain	0.009	1.70	26.9	28.5	11	Myofilaments and cytoskeleton	Sarcomere
P20801	TNNC2	Troponin C, skeletal muscle	0.004	0.53	323.3	90.6	15	Myofilaments and cytoskeleton	Sarcomere
Q9JKS4	LDB3	LIM domain-binding protein 3	0.012	1.79	295.7	32.0	18	Myofilaments and cytoskeleton	Sarcomere
P31001	DES	Desmin	0.004	1.80	245.5	77.2	40	Myofilaments and cytoskeleton	Sarcomere
Q5XKE0	MYBPC2	Myosin-binding protein C, fast-type	0.003	1.82	323.3	53.6	60	Myofilaments and cytoskeleton	Sarcomere
P13412	TNNI2	Troponin I, fast skeletal muscle	0.018	0.55	83.3	34.2	12	Myofilaments and cytoskeleton	Sarcomere
A2ASS6	TTN	Titin	0.005	2.02	323.3	44.5	1310	Myofilaments and cytoskeleton	Sarcomere
Q5SX39	MYH4	Myosin-4	0.003	2.32	323.3	69.0	214	Myofilaments and cytoskeleton	Sarcomere
P68134	ACTA1	Actin, alpha skeletal muscle	0.004	0.49	323.3	62.1	28	Myofilaments and cytoskeleton	Sarcomere
F6X9B6;Q9JL90	NEB	Nebulin	0.011	1.44	3.8	57.0	41	Myofilaments and cytoskeleton	Sarcomere
P58771	TPM1	Tropomyosin alpha-1 chain	0.004	0.40	323.3	90.5	51	Myofilaments and cytoskeleton	Sarcomere
Q9D1G3	HHATL	Protein-cysteine N-palmitoyltransferase HHAT-like protein	0.013	0.68	8.2	6.0	3	Protein biosynthesis, Maturation	Endoplasmic reticulum
Q9DC61	PMPCA	Mitochondrial-processing peptidase subunit alpha	0.046	0.71	4.5	6.7	3	Protein biosynthesis, Maturation	Mitochondrion
Q3TVI8	PBXIP1	Pre-B-cell leukemia transcription factor-interacting protein 1	0.010	1.53	6.1	3.9	2	Protein biosynthesis, Transcription	Cytoplasm, Nucleus
Q3U381	ZFP692	Zinc finger protein 692	0.004	1.60	3.8	1.7	1	Protein biosynthesis, Transcription	Nucleus
Q9CQX8	MRPS36	28S ribosomal protein S36, mitochondrial	0.014	0.60	4.6	14.0	1	Protein biosynthesis, Translation	Mitochondrion
P62264	RPS14	40S ribosomal protein S14	0.047	1.27	3.3	9.2	1	Protein biosynthesis, Translation	Ribosome

Q9D8N0	EEFIG	Elongation factor 1-gamma	0.011	1.95	34.7	21.8	9	Protein biosynthesis, Translation	Cytoplasm
P62631	EEF1A2	Elongation factor 1-alpha 2	0.004	2.12	47.7	22.9	13	Protein biosynthesis, Translation	Cytoplasm
Q922Y1	UBXN1	UBX domain-containing protein 1	0.028	1.41	2.3	7.7	1	Proteolysis	Proteasome
P31786	DBI	Acyl-CoA-binding protein	0.019	1.78	59.5	58.7	4	Proteolysis	Endoplasmic reticulum
P14069	S100A6	Protein S100-A6	0.028	1.63	26.7	9.0	1	Signal transduction, Calcium signaling	Cytoplasm, Nucleus
Q9JK37	MYOZ1	Myozenin-1	0.026	1.85	323.3	88.2	18	Signal transduction, Calcium signaling	Sarcomere
Q91ZI0	CELSR3	Cadherin EGF LAG seven-pass G-type receptor 3	0.015	1.87	3.7	0.5	2	Signal transduction, Calcium signaling	Membrane
Q7TQ48	SRL	Sarcalumenin	0.010	2.01	323.3	28.1	25	Signal transduction, Calcium signaling	Endoplasmic reticulum
O55126	NIPSNAP	Protein NipSnap homolog 2	0.010	1.58	9.8	12.1	4	Signal transduction, Calcium signaling	Mitochondrion
P49817	CAV3	Caveolin-3	0.019	0.62	32.0	21.3	1	Signal Transduction, Calcium signaling	Membrane
P11404	FABP3	Fatty acid-binding protein, heart	0.005	1.98	323.3	49.6	8	Signal transduction, Insulin signaling	Cytoplasm
P62259	YWHAE	14-3-3 protein epsilon	0.005	1.83	323.3	40.0	11	Signal transduction, Insulin signaling	Cytoplasm
Q8R429	ATP2A1	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	0.005	0.48	323.3	40.2	43	Transport, Calcium transport	Endoplasmic reticulum
E9PZQ0	RYR1	Ryanodine receptor 1	0.003	0.65	177.0	9.5	40	Transport, Calcium transport	Endoplasmic reticulum
Q921I1	TF	Serotransferrin	0.036	1.63	323.3	20.4	13	Transport, Iron transport	Serum
P57787	SLC16A3	Monocarboxylate transporter 4	0.012	1.43	10.1	10.4	4	Transport, Lactate transport	Membrane
Q6PIE5	ATP1A2	Sodium/potassium-transporting ATPase subunit alpha-2	0.004	0.72	39.5	15.6	13	Transport, Nutrients transport	Membrane
P04247	MB	Myoglobin	0.003	0.40	323.3	85.7	19	Transport, Oxygen transport	Cytoplasm
P70168	KPNB1	Importin subunit beta-1	0.042	1.47	2.0	1.4	1	Transport	Cytoplasm, Nucleus
P62827	RAN	GTP-binding nuclear protein Ran	0.028	1.37	12.0	6.9	3	Transport	Nucleus

Accession: UniProt Ids; Symbol: Official Gene Symbol; Name: Musmusculus Protein name; Score: MaxQuant Score; %Coverage: Protein sequence covered by all identified peptides (%); # Peptides: Number of identified peptides

Table S3 - Proteomic Profile of SOL muscle in Response to FS

# Accession	Symbol	Name	Anova (P)	Fold change	Score	% Coverage	# Peptides	Main Biological Process	Main Cellular Component
P63038	HSPD1	60 kDa heat shock protein, mitochondrial	0.048	0.6	26.1	21.1	11	Detoxification, cytoprotection	Mitochondrion
P35700	PRDX1	Peroxiredoxin-1	0.013	0.7	16.6	37.1	7	Detoxification, cytoprotection	Cytoplasm
Q9QZ57	HSPB3	Heat shock protein beta-3	0.038	0.7	4.3	11.0	2	Detoxification, cytoprotection	Cytoplasm
P07310	CKM	Creatine kinase M-type	0.016	0.5	323.3	81.4	43	Energy metabolism	Cytoplasm
Q8QZS1	HIBCH	3-hydroxyisobutyryl-CoA hydrolase, mitochondrial	0.042	0.7	7.4	2.9	1	Energy metabolism, Amino-acid degradation	Mitochondrion
Q5RJV4;Q9D0F9	PGM1/2	Phosphoglucomutase-1/2	0.025	0.6	144.1	44.3	21	Energy metabolism, glycogenolysis	Cytoplasm
P21550	ENO3	Beta-enolase	0.018	0.5	323.3	63.1	33	Energy metabolism, Glycolysis	Cytoplasm
P52480	PKM	Pyruvate kinase PKM	0.019	0.6	323.3	58.6	31	Energy metabolism, Glycolysis	Cytoplasm
O70250	PGAM2	Phosphoglycerate mutase 2	0.044	0.6	54.6	28.1	9	Energy metabolism, Glycolysis	Cytoplasm
P54071	IDH2	Isocitrate dehydrogenase [NADP], mitochondrial	0.019	0.5	323.3	58.0	29	Energy metabolism, Krebs cycle	Mitochondrion
Q9D051	PDHB	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	0.018	0.6	190.4	28.1	8	Energy metabolism, Krebs cycle	Mitochondrion
P14152	MDH1	Malate dehydrogenase, cytoplasmic	0.018	0.6	323.3	47.9	17	Energy metabolism, Malate-aspartate NADH Shuttle	Cytoplasm
P08249	MDH2	Malate dehydrogenase, mitochondrial	0.023	0.6	87.0	58.3	17	Energy metabolism, Malate-aspartate NADH Shuttle	Mitochondrion
P56480	ATP5B	ATP synthase subunit beta, mitochondrial	0.019	0.6	323.3	53.5	21	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q9ERS2	NDUFA13	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	0.028	0.6	26.4	53.5	7	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q8K2B3	SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	0.035	0.6	309.4	19.4	15	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q99LY9	NDUFS5	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	0.022	0.6	8.8	14.2	1	Energy metabolism, Oxidative phosphorylation	Mitochondrion
P56391	COX6B1	Cytochrome c oxidase subunit 6B1	0.037	0.7	7.8	41.9	4	Energy metabolism, Oxidative phosphorylation	Mitochondrion
P56392	COX7A1	Cytochrome c oxidase subunit 7A1, mitochondrial	0.020	0.7	4.9	28.8	2	Energy metabolism, Oxidative phosphorylation	Mitochondrion
Q8QZT1	ACAT1	Acetyl-CoA acetyltransferase, mitochondrial	0.048	0.7	97.6	30.9	10	Energy metabolism, β -oxydation	Mitochondrion
Q8CGP2;Q8CGP1; Q64475;P10853; Q9D2U9;Q8CGP0; P70696;Q64524	HIST1H2BP/K/B/F/A;HI ST1H2BE; HIST3H2BA/B	Histone H2B type 1-P/K/B/F/A; Histone H2B type 2-E; Histone H2B type 3-A/B	0.015	1.9	66.0	52.0	7	Gene regulation	Nucleus
Q61838	PZP	Alpha-2-macroglobulin	0.038	1.7	17.9	8.2	11	Miscellaneous	Miscellaneous
P02088	HBB-B1	Hemoglobin subunit beta-1	0.020	2.0	323.3	84.4	15	Miscellaneous	Miscellaneous
P01942	HBA	Hemoglobin subunit alpha	0.040	2.0	108.9	53.3	8	Miscellaneous	Miscellaneous
Q8VHX6	FLNC	Filamin-C	0.048	0.5	323.3	37.6	88	Myofilaments and cytoskeleton	Sarcomere
Q9QXS1	PLEC	Plectin	0.023	0.5	323.3	34.5	149	Myofilaments and cytoskeleton	Sarcomere
A2AAJ9	OBSCN	Obscurin	0.001	0.5	323.3	19.0	91	Myofilaments and cytoskeleton	Sarcomere
P09542	MYL3	Myosin light chain 3	0.019	0.6	195.5	66.2	14	Myofilaments and cytoskeleton	Sarcomere
P97457	MYLPF	Myosin regulatory light chain 2, skeletal muscle isoform	0.028	0.6	224.3	78.7	15	Myofilaments and cytoskeleton	Sarcomere
P05213;P68373; P68369;P05214	TUBA1A/B/C; TUBA3A	Tubulin alpha-1A/B/C chain;Tubulin alpha-3 chain	0.047	1.6	11.8	21.3	8	Myofilaments and cytoskeleton	Sarcomere
E9Q616	AHNAK	Desmoyokin	0.019	1.6	2.1	1.5	2	Myofilaments and cytoskeleton	Sarcomere
P63260;P60710	ACTG1/ACTB	Actin, cytoplasmic 2/1	0.040	1.6	5.9	40.0	20	Myofilaments and cytoskeleton	Sarcomere
O88346	TNNT1	Troponin T, slow skeletal muscle	0.017	1.7	323.3	31.3	12	Myofilaments and cytoskeleton	Sarcomere
P17742	PPIA	Peptidyl-prolyl cis-trans isomerase A	0.023	1.4	12.5	23.2	4	Protein biosynthesis, Maturation	Cytoplasm
P70333;O35737; Q9Z2X1	HNRNPH1/2; HNRNPF	Heterogeneous nuclear ribonucleoprotein H2/H/F	0.019	1.3	12.9	5.8	2	Protein biosynthesis, Transcription	Nucleus
P58252	EEF2	Elongation factor 2	0.027	0.6	25.4	11.1	9	Protein biosynthesis, Translation	Cytoplasm
P99027	RPLP2	60S acidic ribosomal protein P2	0.038	0.7	1.6	10.4	1	Protein biosynthesis, Translation	Ribosome
Q8C1A5	THOP1	Thimet oligopeptidase	0.029	0.8	2.5	1.5	1	Miscellaneous	Cytoplasm
Q9D154	SERPINB1A	Leukocyte elastase inhibitor A	0.039	1.7	2.2	5.5	2	Proteolysis	Cytoplasm
P62880;P62874; Q61011	GNB2/1/3	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2/1/3	0.022	1.5	2.2	5.7	2	Signal transduction, Calcium signaling	Cytoplasm
E9PZQ0	RYR1	Ryanodine receptor 1	0.023	0.6	149.1	7.3	34	Transport, Calcium transport	Endoplasmic reticulum
Q02789	CACNA1S	Voltage-dependent L-type calcium channel subunit alpha-1S	0.033	0.8	4.6	7.9	2	Transport, Calcium transport	Membrane
P63254	CRIP1	Cysteine-rich protein 1	0.025	1.3	4.2	16.9	2	Transport, Zinc transport	Cytoplasm

Q6PIE5	ATP1A2	Sodium/potassium-transporting ATPase subunit alpha-2	0.042	0.6	119.2	30.0	25	Transport, Nutrients transport	Membrane
--------	--------	--	-------	-----	-------	------	----	--------------------------------	----------

Accession: UniProt Ids; **Symbol:** Official Gene Symbol; **Name:** Musmusculus Protein name; **Score:** MaxQuant Score; **%Coverage:** Protein sequence covered by all identified peptides (%); **# Peptides:** Number of identified peptides

Table S4. Transcriptomic Profile of TA Muscle in Response to FS

# Probeset	Symbol	Gene Title	Fold Change	P-Value
1433582_at	I190002N15Rik	RIKEN cDNA I190002N15 gene	0.69	0.027
1457964_at	I1810044D09Rik	RIKEN cDNA I1810044D09 gene	1.56	0.002
1450972_at	I3110040N11Rik	RIKEN cDNA I3110040N11 gene	1.46	0.012
1421840_at	Abca1	ATP-binding cassette, sub-family A (ABC1), member 1	0.55	0.016
1427053_at	Abi3bp	ABI gene family, member 3 (NESH) binding protein	0.57	0.041
1427174_at	Abrac1	ABRA C-terminal like	0.50	0.014
1439459_x_at	Acly	ATP citrate lyase	0.69	0.029
1428082_at	Acs15	Acyl-CoA synthetase long-chain family member 5	0.71	0.022
1448460_at	Acvr1	Activin A receptor, type 1	0.58	0.015
1451932_a_at	Adamts14	ADAMTS-like 4	0.61	0.009
1426574_a_at	Add3	Adducin 3 (gamma)	0.67	0.021
1424151_at	Aimp2	Aminoacyl tRNA synthetase complex-interacting multifunctional protein 2	1.59	0.004
1448450_at	Ak2	Adenylate kinase 2	0.69	0.020
1455151_at	Akap9	A kinase (PRKA) anchor protein (yotiao) 9	0.65	0.032
1450959_at	Akip1	A kinase (PRKA) interacting protein 1	1.51	0.021
1424328_s_at	Anapc15	Anaphase promoting complex C subunit 15	1.44	0.017
1419421_at	Ank1	Ankyrin 1, erythroid	1.60	0.006
1451446_at	Antxr1	Anthrax toxin receptor 1	0.67	0.028
1449396_at	Aoc3	Amine oxidase, copper containing 3	0.57	0.005
1420346_at	Asb12	Ankyrin repeat and SOCS box-containing protein 12	1.48	0.017
1416652_at	Aspn	Asporin	0.66	0.007
1433918_at	Atg4d	Autophagy-related 4D (yeast)	0.70	0.025
1434513_at	Atp13a3	ATPase type 13A3	1.77	0.001
1418453_a_at	Atp1b1	ATPase, Na ⁺ /K ⁺ transporting, beta 1 polypeptide	0.59	0.024
1422908_at	Atp1b4	ATPase, (Na ⁺)/K ⁺ transporting, beta 4 polypeptide	1.83	0.008
1452363_a_at	Atp2a2	ATPase, Ca ⁺⁺ transporting, cardiac muscle, slow twitch 2	0.59	0.007
1423597_at	Atp8a1	ATPase, aminophospholipid transporter (APLT), class I, type 8A, member 1	1.54	0.004
1426459_s_at	AW549877	Expressed sequence AW549877	0.62	0.031
1425394_at	BC023105	cDNA sequence BC023105	5.10	0.006
1439388_s_at	Bcar1	Breast cancer anti-estrogen resistance 1	1.43	0.015
1418421_at	Bcl6b	B-cell CLL/lymphoma 6, member B	0.57	0.026
1452257_at	Bdh1	3-hydroxybutyrate dehydrogenase, type 1	0.42	0.010
1448595_a_at	Bex1	Brain expressed gene 1	2.10	0.016
1425532_a_at	Bin1	Bridging integrator 1	1.65	0.001
1425814_a_at	Calcr1	Calcitonin receptor-like	0.64	0.016
1426710_at	Calm1	Calmodulin 1	1.61	0.017
1426710_at	Calm2	Calmodulin 2	1.61	0.017
1426710_at	Calm3	Calmodulin 3	1.61	0.017
1450725_s_at	Car14	Carbonic anhydrase 14	1.42	0.014
1448659_at	Casp7	Caspase 7	0.60	0.006
1418509_at	Cbr2	Carbonyl reductase 2	2.12	0.065
1452721_a_at	Ccdc53	Coiled-coil domain containing 53	1.44	0.001
1424186_at	Ccdc80	Coiled-coil domain containing 80	0.56	0.048
1448229_s_at	Ccnd2	Cyclin D2	0.67	0.006
1416072_at	Cd34	CD34 antigen	0.66	0.010
1416401_at	Cd82	CD82 antigen	1.55	0.008
1448193_at	Cdip1	Cell death inducing Trp53 target 1	0.65	0.004
1418579_at	Cetn2	Centrin 2	1.43	0.004
1425400_a_at	Cited4	Chp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4	1.74	0.001
1427279_at	Clip4	CAP-GLY domain containing linker protein family, member 4	0.67	0.011
1454849_x_at	Clu	Clusterin	1.51	0.003
1451322_at	Cmb1	Carboxymethylenebutenolidase-like (Pseudomonas)	1.55	0.003
1453015_at	Cmip	c-Maf inducing protein	0.57	0.008
1450981_at	Cnn2	Calponin 2	0.64	0.010
1437341_x_at	Cnp	2',3'-cyclic nucleotide 3' phosphodiesterase	0.64	0.005
1418599_at	Col11a1	Procollagen, type XI, alpha 1	0.41	0.002
1423669_at	Col1a1	Procollagen, type I, alpha 1	0.49	0.027
1423110_at	Col1a2	Procollagen, type I, alpha 2	0.63	0.027
1427884_at	Col3a1	Procollagen, type III, alpha 1	0.63	0.023
1420384_at	Col4a3bp	Procollagen, type IV, alpha 3 (Goodpasture antigen) binding protein	0.65	0.033
1450625_at	Col5a2	Procollagen, type V, alpha 2	0.60	0.017
1416776_at	Crym	Crystallin, mu	1.43	0.001
1460220_a_at	Csf1	Colony stimulating factor 1 (macrophage)	0.69	0.024
1425810_a_at	Csrp1	Cysteine and glycine-rich protein 1	0.70	0.003
1417567_at	Ctnnbip1	Catenin beta interacting protein 1	1.69	0.018
1418365_at	Ctsh	Cathepsin H	0.62	0.031

1417574_at	Cxcl12	Chemokine (C-X-C motif) ligand 12	0.63	0.022
1454268_a_at	Cyba	Cytochrome b-245, alpha polypeptide	0.50	0.062
1449406_at	Cyhr1	Cysteine and histidine rich 1	1.77	0.006
1438133_a_at	Cyr61	Cysteine rich protein 61	0.14	0.033
1426521_at	D230025D16Rik	RIKEN cDNA D230025D16 gene	1.42	0.007
1437917_at	D530037H12Rik	RIKEN cDNA D530037H12 gene	0.56	0.042
1426439_at	Ddx3y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	1.96	0.000
1416755_at	Dnajb1	DnaJ (Hsp40) homolog, subfamily B, member 1	0.55	0.026
1451177_at	Dnajb4	DnaJ (Hsp40) homolog, subfamily B, member 4	0.67	0.031
1423065_at	Dnmt3a	DNA methyltransferase 3A	0.69	0.024
1444028_s_at	Dock9	Dedicator of cytokinesis 9	0.67	0.018
1425292_at	Dtna	Dystrobrevin alpha	1.55	0.001
1431422_a_at	Dusp14	Dual specificity phosphatase 14	0.64	0.004
1425848_a_at	Dusp26	Dual specificity phosphatase 26 (putative)	0.47	0.005
1418371_at	Dynll2	Dynein light chain LC8-type 2	1.45	0.014
1417298_at	Ebpl	Emopamil binding protein-like	1.42	0.006
1418093_a_at	Egf	Epidermal growth factor	1.56	0.001
1417065_at	Egr1	Early growth response 1	0.15	0.070
1417713_at	Eif2s2	Eukaryotic translation initiation factor 2, subunit 2 (beta)	1.43	0.001
1417562_at	Eif4ebp1	Eukaryotic translation initiation factor 4E binding protein 1	1.49	0.002
1455994_x_at	Elov1	Elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 1	0.64	0.017
1415840_at	Elov5	ELOVL family member 5, elongation of long chain fatty acids (yeast)	0.69	0.002
1437712_x_at	Exosc4	Exosome component 4	1.72	0.007
1423680_at	Fads1	Fatty acid desaturase 1	0.68	0.002
1449258_at	Fam104a	Family with sequence similarity 104, member A	1.58	0.007
1416892_s_at	Fam107b	Family with sequence similarity 107, member B	0.65	0.018
1426812_a_at	Fam129b	Family with sequence similarity 129, member B	0.59	0.012
1426729_at	Fam131a	Family with sequence similarity 131, member A	1.61	0.010
1416805_at	Fam198b	Family with sequence similarity 198, member B	0.62	0.000
1423266_at	Fam213b	Family with sequence similarity 213, member B	1.48	0.010
1460555_at	Fam65b	Family with sequence similarity 65, member B	2.66	0.009
1435315_s_at	Far1	Fatty acyl CoA reductase 1	0.69	0.023
1423828_at	Fasn	Fatty acid synthase	0.56	0.017
1427582_at	Fgf6	Fibroblast growth factor 6	1.66	0.016
1426677_at	Flna	Filamin, alpha	0.68	0.016
1423100_at	Fos	Proto-oncogene c-Fos	0.07	0.226
1455242_at	Foxp1	Forkhead box P1	0.65	0.008
1437543_at	Fubp1	Far upstream element (FUSE) binding protein 1	0.56	0.005
1417343_at	Fxyd6	FXYP domain-containing ion transport regulator 6	0.53	0.031
1448700_at	G0s2	G0/G1 switch gene 2	0.57	0.066
1449519_at	Gadd45a	Growth arrest and DNA-damage-inducible 45 alpha	1.90	0.013
1453851_a_at	Gadd45g	Growth arrest and DNA-damage-inducible 45 gamma	2.22	0.007
1422558_at	Gamt	Guanidinoacetate methyltransferase	1.91	0.026
1423569_at	Gatm	Glycine amidinotransferase (L-arginine:glycine amidinotransferase)	1.96	0.054
1418627_at	Gclm	Glutamate-cysteine ligase, modifier subunit	1.46	0.003
1450940_at	Gdap1	Ganglioside-induced differentiation-associated-protein 1	2.33	0.008
1454959_s_at	Gnai1	Guanine nucleotide binding protein, alpha inhibiting 1	0.61	0.009
1415698_at	Golm1	Golgi membrane protein 1	2.32	0.005
1417673_at	Grb14	Growth factor receptor bound protein 14	0.66	0.011
1448148_at	Grn	Granulin	0.66	0.045
1415812_at	Gsn	Gelsolin	0.65	0.028
1416416_x_at	Gstm1	Glutathione S-transferase, mu 1	1.47	0.001
1419158_a_at	Hars2	Histidyl-tRNA synthetase 2, mitochondrial (putative)	0.66	0.007
1418102_at	Hes1	Hairy and enhancer of split 1 (Drosophila)	0.61	0.059
1424863_a_at	Hipk2	Homeodomain interacting protein kinase 2	1.45	0.002
1422155_at	Hist1h3b	Histone cluster 1, H3b	1.52	0.019
1422155_at	Hist1h3c	Histone cluster 1, H3c	1.52	0.019
1422155_at	Hist1h3d	Histone cluster 1, H3d	1.52	0.019
1422155_at	Hist1h3e	Histone cluster 1, H3e	1.52	0.019
1422155_at	Hist1h3f	Histone cluster 1, H3f	1.52	0.019
1422155_at	Hist2h3b	Histone cluster 2, H3b	1.52	0.019
1422155_at	Hist2h3c1	Histone cluster 2, H3c1	1.52	0.019
1422155_at	Hist2h3c2	Histone cluster 2, H3c2	1.52	0.019
1424639_a_at	Hmgcl	3-hydroxy-3-methylglutaryl-Coenzyme A lyase	1.41	0.007
1434875_a_at	Hmgn3	High mobility group nucleosomal binding domain 3	0.63	0.009
1424132_at	Hras	Harvey rat sarcoma virus oncogene	1.49	0.009
1450047_at	Hs6st2	Heparan sulfate 6-O-sulfotransferase 2	1.47	0.006
1453016_at	Hspb1	Heat shock protein family B (small), member 11	1.45	0.002

1417014_at	Hspb8	Heat shock protein 8	2.15	0.008
1416749_at	Htra1	HtrA serine peptidase 1	0.58	0.007
1448862_at	Icam2	Intercellular adhesion molecule 2	0.65	0.017
1416442_at	Ier2	Immediate early response 2	0.40	0.054
1454629_at	Iffo1	Intermediate filament family orphan 1	1.98	0.036
1452349_x_at	Ifi205	Interferon activated gene 205	0.63	0.034
1452349_x_at	Ifi211	Interferon activated gene 211	0.63	0.008
1427239_at	Ift122	Intraflagellar transport 122 homolog (Chlamydomonas)	1.56	0.012
1424648_at	Ift27	Intraflagellar transport 27	1.87	0.027
1437401_at	Igf1	Insulin-like growth factor 1	0.66	0.009
1452114_s_at	Igfbp5	Insulin-like growth factor binding protein 5	1.65	0.003
1425454_a_at	Il12a	Interleukin 12a	3.12	0.029
1418697_at	Inmt	Indolethylamine N-methyltransferase	0.32	0.002
1456195_x_at	Itgb5	Integrin beta 5	0.51	0.024
1423184_at	Itsn2	Intersectin 2	0.58	0.004
1450710_at	Jarid2	Jumonji, AT rich interactive domain 2	0.66	0.027
1426900_at	Jmjd1c	Jumonji domain containing 1C	0.59	0.047
1423559_at	Kenc1	Potassium voltage gated channel, Shaw-related subfamily, member 1	0.66	0.024
1449379_at	Kdr	Kinase insert domain protein receptor	0.69	0.013
1418063_at	Kera	Keratocan	0.39	0.008
1449439_at	Klf7	Kruppel-like factor 7 (ubiquitous)	0.63	0.018
1428081_at	Klhl21	Kelch-like 21 (Drosophila)	0.70	0.016
1426406_at	Kmt5a	Lysine methyltransferase 5A	1.45	0.004
1427011_a_at	Lancl1	LanC (bacterial lantibiotic synthetase component C)-like 1	1.48	0.000
1416148_at	Laptm4b	Lysosomal-associated protein transmembrane 4B	1.44	0.001
1419573_a_at	Lgals1	Lectin, galactose binding, soluble 1	1.62	0.007
1448883_at	Lgmn	Legumain	1.67	0.001
1450629_at	Lima1	LIM domain and actin binding 1	0.60	0.016
1452345_at	Lmod2	Leiomodin 2 (cardiac)	0.39	0.006
1448228_at	Lox	Lysyl oxidase	0.39	0.006
1460438_at	Lysmd1	LysM, putative peptidoglycan-binding, domain containing 1	1.56	0.005
1455928_x_at	Lztr1	Leucine-zipper-like transcriptional regulator, 1	0.63	0.005
1448647_at	Man2a1	Mannosidase 2, alpha 1	0.70	0.008
1423688_at	Man2c1	Mannosidase, alpha, class 2C, member 1	0.65	0.001
1451290_at	Map1lc3a	Microtubule-associated protein 1 light chain 3 alpha	1.60	0.010
1436893_a_at	March7	Membrane-associated ring finger (C3HC4) 7	0.68	0.057
1456028_x_at	Marcks	Myristoylated alanine rich protein kinase C substrate	0.62	0.011
1433532_a_at	Mbp	Myelin basic protein	0.58	0.008
1424234_s_at	Meox2	Mesenchyme homeobox 2	0.65	0.003
1416313_at	Mllt11	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11	1.56	0.034
1416136_at	Mmp2	Matrix metalloproteinase 2	0.67	0.034
1419077_at	Mpp3	Membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	0.54	0.014
1419253_at	Mthfd2	Methylenetetrahydrofolate dehydrogenase (NAD+ dependent), methylenetetrahydrofolate cyclohydrolase	2.81	0.009
1418522_at	Mtx1	Metaxin 1	0.68	0.002
1427201_at	Mustn1	Musculoskeletal, embryonic nuclear protein 1	1.44	0.017
1423321_at	Myadm	Myeloid-associated differentiation marker	0.69	0.011
1419487_at	Mybph	Myosin binding protein H	2.53	0.002
1452465_at	Myh1	Myosin, heavy polypeptide 1, skeletal muscle, adult	0.65	0.026
1420805_at	Myl10	Myosin, light chain 10, regulatory	0.45	0.045
1425505_at	Mylk	Myosin, light polypeptide kinase	0.65	0.002
1422671_s_at	Naalad2	N-acetylated alpha-linked acidic dipeptidase 2	0.71	0.004
1423269_a_at	Nedd4l	Neural precursor cell expressed, developmentally down-regulated gene 4-like	1.63	0.010
1416817_at	Nek7	NIMA (never in mitosis gene a)-related expressed kinase 7	0.69	0.027
1431936_a_at	Neu2	Neuraminidase 2	1.47	0.013
1438244_at	Nfib	Nuclear factor I/B	0.59	0.017
1419652_s_at	Nkain1	Na+/K+ transporting ATPase interacting 1	1.91	0.007
1423249_at	Nktr	Natural killer tumor recognition sequence	0.70	0.006
1451503_at	Nol3	Nucleolar protein 3 (apoptosis repressor with CARD domain)	1.52	0.007
1438483_at	Nos1	Nitric oxide synthase 1, neuronal	0.61	0.010
1436736_x_at	Nrep	Neuronal regeneration related protein	0.63	0.002
1418084_at	Nrp1	Neuropilin 1	0.69	0.004
1419662_at	Ogn	Osteoglycin	0.40	0.020
1422792_at	Pafah1b2	Platelet-activating factor acetylhydrolase, isoform 1b, alpha2 subunit	0.70	0.015
1420979_at	Pak1	p21 (CDKN1A)-activated kinase 1	2.01	0.042
1416379_at	Panx1	Pannexin 1	0.54	0.014
1452913_at	Pcp4l1	Purkinje cell protein 4-like 1	0.64	0.018
1422473_at	Pde4b	Phosphodiesterase 4B, cAMP specific	0.60	0.008
1436970_a_at	Pdgfrb	Platelet derived growth factor receptor, beta polypeptide	0.71	0.008

1449178_at	Pdlim3	PDZ and LIM domain 3	1.44	0.004
1419309_at	Pdpm	Podoplanin	1.46	0.015
1427038_at	Penk	Preproenkephalin	0.59	0.038
1416432_at	Pfkfb3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	3.68	0.059
1423451_at	Pgrmc1	Progesterone receptor membrane component 1	0.70	0.020
1428025_s_at	Pitpnc1	Phosphatidylinositol transfer protein, cytoplasmic 1	0.67	0.004
1452026_a_at	Pla2g12a	Phospholipase A2, group X1IA	0.67	0.020
1451611_at	Pla2g16	Phospholipase A2, group XVI	1.40	0.001
1430700_a_at	Pla2g7	Phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	2.15	0.010
1416424_at	Plin3	Perilipin 3	1.50	0.006
1427005_at	Plk2	Polo-like kinase 2 (Drosophila)	0.66	0.021
1453299_a_at	Pnp	Purine-nucleoside phosphorylase	0.62	0.001
1453299_a_at	Pnp2	Purine-nucleoside phosphorylase 2	0.62	0.001
1450914_at	Ppp1r14b	Protein phosphatase 1, regulatory (inhibitor) subunit 14B	1.60	0.007
1423831_at	Prkag2	Protein kinase, AMP-activated, gamma 2 non-catalytic subunit	2.31	0.003
1425528_at	Prrx1	Paired related homeobox 1	0.65	0.002
1421411_at	Pstpip2	Proline-serine-threonine phosphatase-interacting protein 2	1.61	0.003
1417667_a_at	Pter	Phosphotriesterase related	1.46	0.001
1448816_at	Ptgis	Prostaglandin I2 (prostacyclin) synthase	0.42	0.018
1455207_at	Pthr1	Peptidyl-tRNA hydrolase domain containing 1	1.44	0.019
1418666_at	Ptx3	Pentraxin related gene	1.65	0.005
1424193_at	Pwp2	PWP2 periodic tryptophan protein homolog (yeast)	1.46	0.002
1417073_a_at	Qk	Quaking	0.63	0.005
1424894_at	Rab13	RAB13, member RAS oncogene family	1.48	0.002
1434914_at	Rab6b	RAB6B, member RAS oncogene family	0.66	0.009
1435517_x_at	Ralb	V-ral simian leukemia viral oncogene homolog B (ras related)	0.64	0.055
1417481_at	Ramp1	Receptor (calcitonin) activity modifying protein 1	0.57	0.001
1456262_at	Rbm5	RNA binding motif protein 5	0.58	0.010
1424382_at	Rcn3	Reticulocalbin 3, EF-hand calcium binding domain	0.52	0.001
1416216_at	Reps1	RalBP1 associated Eps domain containing protein	1.50	0.018
1439200_x_at	Rhox4b	Reproductive homeobox 4B	3.13	0.012
1423134_at	Rilpl2	Rab interacting lysosomal protein-like 2	0.70	0.041
1416700_at	Rnd3	Rho family GTPase 3	0.60	0.025
1423615_at	Rnf115	Ring finger protein 115	1.47	0.011
1450423_s_at	Rpf2	Ribosome production factor 2 homolog (S. cerevisiae)	1.41	0.003
1422623_x_at	Rpl41	Ribosomal protein L41	0.69	0.013
1424095_at	Rtca	RNA 3'-terminal phosphate cyclase	1.45	0.001
1419814_s_at	S100a1	S100 calcium binding protein A1	1.42	0.002
1452885_at	Scaff11	SR-related CTD-associated factor 11	0.68	0.000
1451204_at	Scara5	Scavenger receptor class A, member 5 (putative)	0.70	0.016
1415964_at	Scd1	Stearoyl-Coenzyme A desaturase 1	0.35	0.028
1417012_at	Sdc2	Syndecan 2	1.90	0.000
1416779_at	Sdpr	Serum deprivation response	0.71	0.021
1416318_at	Serpnb1a	Serine (or cysteine) peptidase inhibitor, clade B, member 1a	2.17	0.043
1450138_a_at	Serpnb6a	Serine (or cysteine) peptidase inhibitor, clade B, member 6a	1.59	0.024
1456733_x_at	Serpinh1	Serine (or cysteine) peptidase inhibitor, clade H, member 1	0.66	0.010
1421077_at	Sertad3	SERTA domain containing 3	0.61	0.013
1436898_at	Sfpq	Splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	0.57	0.006
1428107_at	Sh3bgr1	SH3-binding domain glutamic acid-rich protein like	0.56	0.000
1449084_s_at	Sh3d19	SH3 domain protein D19	0.68	0.003
1437503_a_at	Shisa5	Shisa homolog 5 (Xenopus laevis)	0.68	0.007
1452214_at	Skil	SKI-like	0.44	0.005
1416629_at	Slc1a5	Solute carrier family 1 (neutral amino acid transporter), member 5	0.61	0.015
1438187_at	Slc25a29	Solute carrier family 25 (mitochondrial carrier, palmitoylcarnitine transporter), member 29	0.69	0.020
1448889_at	Slc38a4	Solute carrier family 38, member 4	2.02	0.005
1425440_x_at	Slc41a3	Solute carrier family 41, member 3	1.63	0.002
1420884_at	Sln	Sarcolipin	8.26	0.007
1415935_at	Smoc2	SPARC related modular calcium binding 2	1.49	0.003
1424268_at	Smox	Spermine oxidase	2.53	0.070
1421052_a_at	Sms	Spermine synthase	1.56	0.003
1449109_at	Socs2	Suppressor of cytokine signaling 2	0.33	0.057
1455899_x_at	Socs3	Suppressor of cytokine signaling 3	0.18	0.071
1419329_at	Sorbs3	Sorbin and SH3 domain containing 3	0.57	0.013
1423362_at	Sort1	Sortilin 1	1.66	0.001
1426018_a_at	Sox6	SRY-box containing gene 6	1.85	0.036
1418180_at	Sp1	Trans-acting transcription factor 1	0.71	0.018
1418798_s_at	Srpk3	Serine/arginine-rich protein specific kinase 3	1.53	0.000
1434972_x_at	Srsf1	Serine/arginine-rich splicing factor 1	0.70	0.008

1438215_at	Srsf3	Serine/arginine-rich splicing factor 3	0.56	0.001
1421892_at	St3gal2	ST3 beta-galactoside alpha-2,3-sialyltransferase 2	0.63	0.025
1425533_a_at	Stau2	Staufen (RNA binding protein) homolog 2 (Drosophila)	1.56	0.001
1434442_at	Stbd1	Starch binding domain 1	1.70	0.001
1419741_at	Supt16	Suppressor of Ty 16	0.70	0.003
1419447_s_at	Tbc1d1	TBC1 domain family, member 1	0.46	0.001
1450961_a_at	Tceal3	Transcription elongation factor A (SII)-like 3	2.76	0.008
1450961_a_at	Tceal5	Transcription elongation factor A (SII)-like 5	2.76	0.008
1450961_a_at	Tceal6	Transcription elongation factor A (SII)-like 6	2.76	0.008
1417455_at	Tgfb3	Transforming growth factor, beta 3	0.63	0.056
1456250_x_at	Tgfb1	Transforming growth factor, beta induced	0.49	0.014
1417500_a_at	Tgm2	Transglutaminase 2, C polypeptide	1.40	0.011
1450663_at	Thbs2	Thrombospondin 2	0.51	0.022
1424642_at	Thoc1	THO complex 1	0.71	0.018
1452125_at	Thrap3	Thyroid hormone receptor associated protein 3	0.67	0.032
1419089_at	Timp3	Tissue inhibitor of metalloproteinase 3	0.61	0.037
1434600_at	Tjp2	Tight junction protein 2	0.55	0.013
1450853_at	Tle4	Transducin-like enhancer of split 4, homolog of Drosophila E(spl)	0.53	0.005
1436042_at	Tln1	Talin 1	1.87	0.010
1424443_at	Tm6sf1	Transmembrane 6 superfamily member 1	1.83	0.011
1449533_at	Tmem100	Transmembrane protein 100	0.26	0.002
1416342_at	Tnc	Tenascin C	0.61	0.020
1418572_x_at	Tnfrsf12a	Tumor necrosis factor receptor superfamily, member 12a	2.30	0.096
1417979_at	Tnmd	Tenomodulin	0.47	0.033
1418412_at	Tpd5211	Tumor protein D52-like 1	1.66	0.000
1449996_a_at	Tpm3	Tropomyosin 3, gamma	0.50	0.034
1416799_at	Trpm7	Transient receptor potential cation channel, subfamily M, member 7	0.71	0.025
1424653_at	Tspan15	Tetraspanin 15	1.47	0.025
1424567_at	Tspan2	Tetraspanin 2	0.67	0.020
1448276_at	Tspan4	Tetraspanin 4	0.63	0.019
1416484_at	Ttc3	Tetratricopeptide repeat domain 3	0.69	0.008
1436833_x_at	Ttl1	Tubulin tyrosine ligase-like 1	1.53	0.015
1423846_x_at	Tuba1b	Tubulin, alpha 1B	0.64	0.024
1431292_a_at	Twf2	Twinfilin, actin-binding protein, homolog 2 (Drosophila)	1.44	0.003
1424056_at	Usp48	Ubiquitin specific peptidase 48	0.67	0.004
1450641_at	Vim	Vimentin	0.64	0.047
1436791_at	Wnt5a	Wingless-related MMTV integration site 5A	0.57	0.004
1451968_at	Xrcc5	X-ray repair complementing defective repair in Chinese hamster cells 5	1.86	0.004
1420830_x_at	Ywhaq	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide	0.69	0.007
1420865_at	Zbtb14	Zinc finger and BTB domain containing 14	0.65	0.011
1434901_at	Zbtb2	Zinc finger and BTB domain containing 2	0.69	0.011
1417778_at	Zfp35	Zinc finger protein 35	0.61	0.018
1452519_a_at	Zfp36	Zinc finger protein 36	0.39	0.035
1423835_at	Zfp503	Zinc finger protein 503	0.47	0.012
1418607_at	Zkscan14	Zinc finger with KRAB and SCAN domains 14	1.45	0.019

2. Publication de service 2

Bouclon, J.; Le Danvic, C.; Guettier, E.; Bray, F.; Tokarski, C. ; Rolando, C. ; Nagnan-Le Meillour, P. Identification of post-translational modifications on Odorant-Binding Protein isoforms from pig olfactory secretome by high-resolution mass spectrometry: O-beta-N-acetylglucosaminylation and phosphorylation. *Frontier in Ecology and Evolution*, **5**, 2017

Identification of post-translational modifications on Odorant-Binding Protein isoforms from pig olfactory secretome by high-resolution mass spectrometry: O-beta-N-acetylglucosaminylation and phosphorylation

Julien BOUCLON^{1, 2}, Chrystelle LE DANVIC³, Elodie GUETTIER⁴, Fabrice BRAY², Caroline TOKARSKI², Christian ROLANDO², Patricia NAGNAN-LE MEILLOUR^{1*}

¹Glycobiology of olfaction, UMR8576 Unité de glycobiologie structurale et fonctionnelle, France,

²USR3290 Miniaturisation pour la Synthèse, l'Analyse et la Protéomique (MSAP), France, ³Département R&D, ALLICE, France, ⁴UEPAO - Unité Expérimentale de Physiologie Animale de l'Orfrasière, INRA, France

Submitted to Journal:

Frontiers in Ecology and Evolution

Specialty Section:

Chemical Ecology

Article type:

Original Research Article

Manuscript ID:

303711

Received on:

18 Aug 2017

Revised on:

27 Oct 2017

Frontiers website link:

www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

PNLM, CLD, EG and JB conceived and performed experiments in 2D-E and western-blot. PNLM, CT and CR designed MS experiments. FB and JB acquired and analyzed MS data. All authors were involved in drafting and revising the manuscript and approve the final version submitted.

Keywords

Odorant-binding protein, O-GlcNAc, Phosphorylation, 2-Dimensional Gel Electrophoresis, nanoLC-nanoESI-HCD-MS/MS-Orbitrap

Abstract

Word count: 233

Odorant-Binding Proteins (OBP) are major players of perireceptor events in olfaction. Despite their importance, a molecular mechanism explaining their specificity for odors and pheromones has yet to be proposed. A new approach is provided by the analysis of the pig olfactory secretome that is mainly composed of OBP isoforms, generated from 3 gene products by two types of post-translational modifications (PTM): i) phosphorylation and ii) O-GlcNAc-N-acetylglucosaminylation (O-GlcNAcylation), which are unusual for secreted proteins. Although both types of PTM can be demonstrated on OBP isoforms by specific antibodies, they have to be identified by mass spectrometry (MS), as localizing PTM sites and identifying PTM patterns can help predict binding affinities. In this paper, we report the identification of phosphorylation and O-GlcNAcylation sites on peptides coming from trypsin digestion of only OBP (*sensu stricto*) by nanoLC-nanoESI-HCD-MS/MS-Orbitrap. These PTM were not present on VEG and SAL. PEAKS software analysis of raw MS data allowed selecting spectra that were analyzed manually to identify PTMs. Four peptides corresponding to two different portions of OBP sequence were modified either by a phosphate group or by a hexNAc moiety. Due to the high energy used in HCD, the data did not allow precise localization of the modified sites. The new findings contribute to a better understanding of the mechanisms by which OBP isoforms could extend the binding repertoire of the secreted OBPs. Data are available via ProteomeXchange with identifier PXD007955.

Funding statement

This work was funded by French INRA (Institut National de Recherche Agronomique) and CNRS (Centre National de Recherche Scientifique).

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: No

1 **Identification of post-translational modifications on Odorant-Binding**
2 **Protein isoforms from pig olfactory secretome by high-resolution mass**
3 **spectrometry: *O*- β -N-acetylglucosaminylation and phosphorylation**
4

5 **Julien Bouclon** ^{1,2,‡}, **Chrystelle Le Danvic** ^{1,3}, **Elodie Guettier** ⁴, **Fabrice Bray** ², **Caroline**
6 **Tokarski** ², **Christian Rolando** ², **Patricia Nagnan-Le Meillour** ^{1,*}

7 ¹Univ. Lille, CNRS UMR8576, INRA USC 1409, UGSF, Unité de Glycobiologie Structurale et
8 Fonctionnelle, F 59000, Lille, France

9 ²Univ. Lille, CNRS USR 3290, MSAP, Miniaturisation pour la Synthèse, l'Analyse et la
10 Protéomique, F 59000, Lille, France

11 ³ALLICE, 149 rue de Bercy F 75012 Paris, France

12 ⁴INRA, UEPAO, Unité Expérimentale de Physiologie Animale de l'Orfrasière, F 37380 Nouzilly,
13 France

14 [‡]Present address: LBM, UMR 7203 CNRS, Ecole Normale Supérieure, Paris, France

15

16 *** Correspondence:**

17 Patricia Nagnan-Le Meillour, Univ. Lille, CNRS UMR8576, INRA USC 1409, UGSF, Unité de
18 Glycobiologie Structurale et Fonctionnelle, F 59000, Lille, France

19 patricia.le-meillour@univ-lille1.fr

20 **Keywords: odorant-binding protein, nanoLC-nanoESI-HCD-MS/MS-Orbitrap, *O*-GlcNAc,**
21 **phosphorylation, 2-dimensional gel electrophoresis.**

spectrometry

22 **Abstract**

23 Odorant-Binding Proteins (OBP) are major players of perireceptor events in olfaction. Despite their
24 importance, a molecular mechanism explaining their specificity for odors and pheromones has yet to
25 be proposed. A new approach is provided by the analysis of the pig olfactory secretome that is
26 mainly composed of OBP isoforms, generated from 3 gene products by two types of post-
27 translational modifications (PTM): i) phosphorylation and ii) *O*- β -N-acetylglucosaminylation (*O*-
28 GlcNAcylation), which are unusual for secreted proteins. Although both types of PTM can be
29 demonstrated on OBP isoforms by specific antibodies, they have to be identified by mass
30 spectrometry (MS), as localizing PTM sites and identifying PTM patterns can help predict binding
31 affinities. In this paper, we report the identification of phosphorylation and *O*-GlcNAcylation sites on
32 peptides coming from trypsin digestion of only OBP (*sensu stricto*) by nanoLC-nanoESI-HCD-
33 MS/MS-Orbitrap. These PTM were not present on VEG and SAL. PEAKS software analysis of raw
34 MS data allowed selecting spectra that were analyzed manually to identify PTMs. Four peptides
35 corresponding to two different portions of OBP sequence were modified either by a phosphate group
36 or by a hexNAc moiety. Due to the high energy used in HCD, the data did not allow precise
37 localization of the modified sites. The new findings contribute to a better understanding of the
38 mechanisms by which OBP isoforms could extend the binding repertoire of the secreted OBPs.
39 Data are available via ProteomeXchange with identifier PXD007955.

40 **1 Introduction**

41 Odorant-binding proteins (OBPs) are one of the molecular players involved in perireceptor events of
42 olfaction (Pelosi, 1996). These globular proteins bind odorant molecules into a hydrophobic pocket
43 (Spinelli et al., 1998) and transport them to olfactory receptors, ultimately responsible for olfactory
44 transduction (Buck and Axel, 1991). Whether the odorant molecule or the complex OBP/odorant acts

45 as a receptor ligand is still an open question, as none of the two hypotheses have received direct
46 experimental evidence. The recent discovery of OBP isoform diversity in pig (*Sus scrofa*) nasal
47 mucus (Nagnan-Le Meillour et al., 2014) favors the hypothesis that OBPs could be more than passive
48 carriers of odorants, being more likely involved in the selection of chemical structures to be
49 presented to a given olfactory receptor. Proteomic analysis of pig nasal mucus revealed that the
50 olfactory secretome (Nagnan-Le Meillour et al., 2014) is mainly composed of around 30 OBP-,
51 VEG- (Von Ebner's Gland protein) and SAL- (Salivary Lipocalin) isoforms, resulting from post-
52 translational modifications (PTMs) of these three gene products and variants. The pig genome
53 contains three genes coding for odorant-binding proteins: i) the *OBP* gene *sensu stricto*, ii) a second
54 OBP transcript encoded by the porcine VEG (Von Ebner's gland protein) gene, *LCN1*, iii) a third
55 OBP, the salivary lipocalin (SAL) encoded by *SALI* gene. *OBP* and *LCN1* genes have two transcripts
56 (*OBPX1* and *OBPX2*, *VEG-VNO* and *VEG-RM*, respectively), whereas *SALI* leads to a single
57 transcript. In addition to their splicing variants, protein isoforms for these three *OBP* genes vary
58 between tissues, and in the nasal mucus, around 9 protein isoforms have been previously identified
59 for VEG, 7 for SAL and 12 for OBP (Nagnan- Le Meillour et al., 2014), due to post-translational
60 modifications targeting the secreted proteins. Immunodetection with specific antibodies indicated
61 that only OBP *sensu stricto* isoforms are modified by phosphorylation and *O*-linked- β -N-
62 acetylglucosaminylation (*O*-GlcNAcylation).

63 If phosphorylation of proteins following the secretion pathway has recently been documented
64 (Tagliabracci et al., 2015), few data report on *O*-GlcNAcylation of secreted proteins (Alfaro et al.,
65 2012). Indeed, typical glycosylation takes place in cellular compartments processing secreted
66 proteins, such as the Endoplasmic Reticulum (ER) and Golgi apparatus (Schwartz and Aebi, 2011). It
67 consists in the addition of complex glycan chains on asparagine (N-linked) or serine/threonine (*O*-
68 linked mucin type) residues, whereas *O*-GlcNAc moieties attached on nuclear and cytoplasmic
69 proteins are not elongated to form complex glycans. Since its discovery (Torres and Hart, 1984), *O*-

spectrometry

70 GlcNAcylation has been shown to regulate almost all intracellular functions by a dynamic interplay
71 together with phosphorylation on the same S/T residues or on adjacent sites (for a full review see
72 Hart and Akimoto, 2009). In the same way that nucleocytoplasmic glycosylation was absent from the
73 Glycobiology dogma until its discovery, many publications also claimed that *O*-GlcNAcylation
74 cannot occur within the secretion pathway. However, it was recently reported in *Drosophila* that EGF
75 repeats of extracellular domains of Notch1 are *O*-GlcNAcyated by a newly identified
76 glycosyltransferase, EOGT (EGF domain specific *O*-Linked N-acetylglucosamine transferase;
77 Sakaidani et al., 2011), ER-resident and totally unrelated to its intracellular counterpart, OGT (*O*-
78 linked N-acetylglucosamine transferase). The new enzyme has – for obvious reasons - been named
79 “EGF-domain specific *O*-GlcNAc transferase”, even if its substrate specificity has not been fully
80 investigated. Although the EOGT gene is well conserved throughout evolution (Nagnan-Le Meillour
81 et al., 2014), the role of extracellular *O*-GlcNAcylation in Mammals remains to be elucidated, despite
82 identification of an increasing number of modified proteins (Alfaro et al., 2012; Hoffman et al.,
83 2012). A first opening on possible roles of this process is the report of a mutation in the EOGT gene
84 associated with a phenotype of Adams-Oliver syndrome, a human congenital disorder (Shaheen et
85 al., 2013). In pigs, the odorant binding properties of OBP (*sensu stricto*) isoforms seem to be driven
86 at least by phosphorylation (Brimau et al., 2010) and we hypothesize that *O*-GlcNAcylation patterns
87 could have a similar function. Before initiating functional assays, we need to identify OBP *O*-
88 GlcNAcylation by high-resolution mass spectrometry. In this paper we report on the identification of
89 phosphorylation and *O*-GlcNAcylation of OBP (*sensu stricto*) peptides, by nanoLC-nanoESI-HCD-
90 MS/MS-Orbitrap from 2-DE spots of total pig olfactory secretome.

91 **2 Materials and methods**92 **2.1 Animals and Tissues**

93 Animals (Large White *Sus scrofa*) were bred at the experimental farm of INRA (UEPAO, Nouzilly,
94 France). Individuals coming from the same offspring (brothers) were slaughtered by a licensed
95 butcher at UEPAO slaughterhouse (authorization # A37801 E37-175-2) in agreement with EU
96 directives about animal welfare. Two adult males were used in this study (300014 - M14 and 300039
97 - M39). Respiratory mucosa was collected immediately after death from each animal and stored in
98 tubes at -80°C before protein extraction.

99 **2.2 Protein extraction**

100 Proteins were extracted from pig tissues by phase partition using chloroform/methanol (v/v, 2/1) on
101 ice. After soft grinding in this solution, samples were centrifuged (15,000 g for 15 min at 4 °C) and
102 the methanol phase was collected and evaporated in a Speed-Vac concentrator (Eppendorf). Dried
103 samples were stored at -20°C. Aliquots representing 1/250 and 1/500 of each tissue were tested by
104 native-PAGE as already described (Guiraudie et al., 2003) and compared with an abacus previously
105 obtained with recombinant porcine OBP (Brimau et al., 2010) in order to estimate the quantity of
106 proteins in each tissue for the standardization of sample loading in 2-DE.

107 **2.3 Two-dimensional gel electrophoresis (2-DE)**

108 All chemicals and reagents were from Sigma-Aldrich. Each sample (M14 and M39) was divided into
109 six identical aliquots for Coomassie-blue staining, immunodetection with anti-OBP, anti-VEG, anti-
110 SAL, anti-phosphoS/Y/T, and anti-*O*-GlcNAc CTD110.6. For the first separation in IEF (isoelectric
111 focalization), each aliquot of dried proteins (15 µg) representing 1/6 of each sample (M14 and M39)
112 was solubilized in 150 µL of rehydration buffer [8 M urea, 2 M thiourea, 2% (w/v) CHAPS, 10 mM
113 dithiothreitol (DTT), 1.2% (v/v) IPG (immobilized pH gradient) buffer (pH 3-5.6, GE Healthcare)
114 and bromophenol blue]. After vigorous shaking, proteins were loaded onto 7-cm IPG strip (non linear
115 pH 3-5.6, GE Healthcare) by overnight passive rehydration at room temperature (RT). The first

spectrometry

116 dimensional IEF was carried out on a PROTEAN IEF Cell (Bio-Rad) using the following program:
117 300 V for 30 min, 1000 V for 1 h, 5000 V for 2 h and 500 V for 3 h, with a current limited to 50
118 μ A/gel. All steps were performed at rapid ramped-voltage. When IEF was complete (10,000 V.h
119 final), strips were incubated twice for 15 min in the equilibration buffer [375 mM tris-HCl pH 8.8, 6
120 M urea, 2% (w/v) SDS (sodium dodecyl sulfate) and 30% (v/v) glycerol] containing first 1.5% (w/v)
121 DTT then 2% (w/v) iodoacetamide. The second dimension separation was performed using 16.8%
122 SDS-PAGE. The 12 gels were run at the same time in Criterion® Dodeca™ Cell (Bio-Rad) to
123 minimize differences in migration in order to superimpose the scanned images with ImageJ®
124 software.

2.4 Staining and western-blot

126 After 2-DE, gels were either stained with colloidal Coomassie blue R solution (12% trichloroacetic
127 acid, 5% ethanolic solution of 0.035% Serva blue R-250), or transferred onto nitrocellulose
128 membrane (Hybond C-Extra, GE Healthcare) for PTM characterization, or PVDF membranes
129 (Immobilon P, Millipore) for primary sequence characterization. For immunodetection of *O*-GlcNAc
130 modified proteins, we used the Thermo Scientific Pierce “*O*-GlcNAc western-blot detection kit”,
131 following manufacturer’s instructions. For other immunodetections, membranes were blocked
132 overnight at 4°C either in 5% (w/v) non-fat dry milk in tris-buffered saline-0.05% tween-20 (TBS-T)
133 for probing with polyclonal antibodies (anti-OBP, anti-SAL and anti-VEG) or in 3% bovine serum
134 albumin (BSA) for probing with anti-phosphoserine/threonine/tyrosine monoclonal antibodies
135 (SPM101 Mab, Fisher Scientific). Membranes were then incubated with antibodies in PBS-T
136 overnight at 4°C [CTD110.6 of the kit mentioned above, Thermo Scientific Pierce, 1:5,000] or TBS-
137 T 1 h at room temperature (RT) [SPM101, 1:500; Anti-OBP, 1:50,000; Anti-VEG, 1:5,000; Anti-
138 SAL, 1:10,000]. After six washes in PBS-T or TBS-T, membranes were incubated 1 h at RT with the
139 appropriate horseradish peroxidase-conjugated secondary antibody [anti-mouse IgM-HRP for

140 CTD110.6, 1:10,000; anti-rabbit IgG-HRP for polyclonal antibodies, 1:30,000; anti-mouse IgG-HRP
141 for SPM101 (all secondary antibodies were from Fisher Scientific)]. After 6 final washes in PBS-T or
142 TBS-T, blots were developed using enhanced chemiluminescence (ECL Plus and ECL Prime
143 reagents, GE Healthcare; Super-Signal™ West Dura, Fisher Scientific). Scanned images of stained
144 gels and blots were merged using Image J software.

145 **2.5 Digestion of protein spots**

146 Spots of interest were excised from the Coomassie-stained gels and rinsed with a mixture of ACN
147 50%/50 mM ammonium bicarbonate (v/v), then dehydrated with ACN (acetonitrile). The gel slices
148 were subjected to reduction of disulfide bonds by 10 mM DTT at 45 °C for 1 h. Alkylation step was
149 then performed with 50 mM iodoacetamide for 1 hour at room temperature in dark. Before trypsin
150 digestion, the gel slices were washed with 50 mM ammonium bicarbonate and dehydrated with ACN.
151 Gel slices were then incubated overnight with 100 ng of trypsin in 50 mM ammonium bicarbonate
152 (Trypsin Gold; Promega) at 37 °C. Peptides were extracted by two incubations in formic acid
153 10%/ACN at 30°C for 15 min, then in formic acid 5%/ACN at RT for 15 min. Supernatants were
154 pooled and dried in Speed-Vac concentrator (Eppendorf).

155 **2.6 Mass spectrometry (MS) analysis**

156 A nanoflow liquid chromatography (LC) instrument (nanoLC U3000 RSLC, Thermo Fisher
157 Scientific) was coupled on-line to a Q Exactive plus (Orbitrap, Thermo Scientific) with a
158 nanoelectrospray ion source (nanoLC-nanoESI). Peptides were resuspended in 10 µL of nano-HPLC
159 buffer A (5% ACN/0.1% formic acid) and 1 µL (corresponding to 200 ng of proteins) was loaded
160 onto the pre-concentration trap (Thermo Scientific, Acclaim PepMap100 C18, 5 µm, 300 µm i.d × 5
161 mm) using partial loop injection, for 5 min at a flow rate of 10 µL.min⁻¹ with buffer A, then
162 separated on column (Acclaim PepMap100 C18, 3 µm, 75 mm i.d. × 500 mm) with a linear gradient

spectrometry

163 of 5–50% buffer B (75% ACN/0.1% formic acid) at a flow rate of 250 nL.min⁻¹ and 45 °C. The total
164 time for a LC MS/MS (tandem mass spectrometry) run was about 240 min long and each sample was
165 injected three times.

166 MS data were acquired on Q Exactive plus using a data-dependent top 20 method dynamically
167 choosing the most abundant precursor ions from the survey scan (400–1600 m/z) for HCD
168 fragmentation (High-energy Collision Dissociation). Dynamic exclusion duration was 60 s. Isolation
169 of precursors was performed with a 1.6 m/z window and MS/MS scans were acquired with a starting
170 mass of 80 m/z. Survey scans were acquired at a resolution of 70,000 at m/z 400 (AGC set to 1 106
171 ions with a maximum fill time of 120 ms). Resolution for HCD spectra was set to 35,500 at m/z 200
172 (AGC set to 5.105 ions with a maximum fill time of 180 ms). Normalized collision energy was 28
173 eV. The underfill ratio, which specifies the minimum percentage of the target value likely to be
174 reached at maximum fill time, was defined as 0.3%. The instrument was run with peptide recognition
175 mode (i.e. from 2 to 8 charges), exclusion of singly charged and of unassigned precursor ions
176 enabled. The acquired raw files were analyzed with PEAKS 7 studio software (Bioinformatics
177 Solutions Inc., Waterloo, Canada) using a custom-made OBP database (Table 1). The full sequences
178 and details are provided in Supplementary data 1. The peptide mass tolerance was set to 10 ppm and
179 0.01 Da for MS/MS. Variable modifications included were as follows: Oxidation of M, Y, H,
180 deamidation of N, Q, carbamidomethylation of C, phosphorylation of Y, S, T, pyro-G, and *O*-N-
181 acetylhexosaminylation of S, T. For high-confidence peptide identification a FDR (False Discover
182 Rate) of 1% and a minimum of 2 ranked peptides were used for peptide filtering.

183 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
184 via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD007955.

185 3 Results

186 **3.1 Identification of OBP-, VEG-, and SAL- isoforms contained in 2-DE spots by western-blot**
187 **with specific antibodies**

188 With the extraction method used in this study, only secreted proteins (OBP *sensu stricto*, VEG, and
189 SAL) of the nasal mucus were extracted and, as revealed by 2-DE, compose the pig olfactory
190 secretome. Coomassie-blue staining revealed similar profiles for the two male secretomes (Figure 1
191 and Figure 2A, 2E). The distribution pattern shows that OBP (Figure 2B, 2F), VEG (Figure 2C, 2G),
192 and SAL (Figure 2D, 2H) isoforms segregate into 3 well-separated groups with no apparent
193 overlapping. SAL isoforms are not visible with Coomassie-blue staining (Figure 2A, 2E), but are
194 well labeled by specific antibodies (Figure 2D, 2H). Western-blot with anti-*O*-GlcNAc and anti-
195 phosphoS/Y/T labeling co-localized only with OBP *sensu stricto*-containing spots (Figure 3A-C and
196 3B-D, respectively) in both samples. From the Coomassie-stained gel of M14 sample, 42 spots were
197 cut (Figure 4) and analyzed by nanoLC-nanoESI-HCD-MS/MS-Orbitrap.

198 **3.2 Characterization of OBP-, VEG-, and SAL-isoforms from 2-DE spots by nanoLC-**
199 **nanoESI-HCD-MS/MS-Orbitrap**

200 Raw files were analyzed by PEAKS software using the custom-made OBP database (Table 1) and
201 resulting data are given in Supplementary Table S1. PEAKS software considers as relevant those
202 identifications with score higher than 67. A few identifications were below this score (in peaks 32,
203 34, 37, and 41) and correspond to spots of faint intensity. Nonetheless, accurate identification does
204 not depend on spot intensity. For example in spot 1, the OBP sequence is covered at 52% (PEAKS
205 score of 307.18, 13 peptides) although spot 38 displaying the same intensity gave only 8% of
206 sequence coverage (PEAKS score 67.14, 1 peptide). OBP was characterized in all spots, except spot
207 6 that contains only a VEG peptide. OBP is alone in spots 1-5, 7-12, 15-19, 21-23, and 38-42. OBP is
208 present with VEG in spots 13, 14, 20, 24-25, 27-37, and with SAL in spot 26. Other SAL isoforms,

spectrometry

209 labeled with anti-SAL antibodies were not visible with Coomassie blue staining and were not
210 analyzed. Data from spot 11 displayed the best PEAKS score (603.29) with 285 peptides and 93% of
211 sequence coverage for the four OBP isoforms (X1A, X1J, X2A, X2J). Peptides from OBPX2 were
212 only retrieved in spots 9 and 11. This analysis did not allow reconstruction of the isoform sequences
213 due to the complexity remaining in each spot, despite the resolution of 2-DE separation.

214 **3.3 PTM characterization by nanoLC-nanoESI-HCD-MS/MS-Orbitrap**

215 Analysis by PEAKS software of the 30617 peptide spectra of spot n°11 revealed PTMs along almost
216 all the OBP sequence. After a rapid spectrum screening, only spectra with a PEAKS score above 67
217 were kept in order to focus on relevant identifications. This allowed identification of 4 OBP peptides
218 (Table 2) modified by phosphorylation (+79.9797) or hexNAcylation (+203.0794). Spectrum
219 n°20288 also revealed an aspartate hydroxylation, a rarely described PTM known as β -
220 hydroxyaspartate, which remains intact and observable under precise conditions of hydrolysis
221 (Castellino et al., 2008). Its role appears to inhibit fucosylation of adjacent threonine (Harris et al.,
222 1992). In all cases, we were unable to precisely assign these PTMs to a particular amino acid, due to
223 the fragmentation method and the number of potential sites on each peptide. However,
224 phosphorylation and hexNAcylation revealed by spectra n°13801 and 28697, respectively, modify
225 peptides of OBPX1 isoform only (see isoform sequences in Supplementary data S1). Considering
226 that no PTMs were found on peptides in this region without the specific OBPX1 peptide sequence
227 NYAGNN, phosphorylation and hexNAcylation are potentially localized on the Y82 residue even if
228 tyrosine glycosylation is rarely described (Jank et al., 2015).

229 The raw data of these four MS/MS spectra (Figure 5) and their corresponding MS spectra
230 (Figure 6) were analyzed (Table 3) in order to confirm the conclusions of the PEAKS approach. The
231 manual assignment of each MS/MS spectrum peak led us to the same peptide identification. The
232 most intense peak in spectrum n°27502 ($m/z = 1012.002$, $\Delta m = 0.6$ ppm) corresponds to the presence

233 of the naked peptide, resulting from the first fragmentation of the parent ion, given that PTM are
234 readily removed by HCD. When the HCD energy is too high, the naked peptide is fragmented just
235 after losing PTM, leaving no time to detect it, which explains its absence in the other spectra.
236 However, neither PEAKS nor manual analysis were able to attribute all peaks. This could be
237 explained by ions of the same mass in mixtures as supported by the distributions of parent ions
238 isotopes (Figure 6). Indeed, experimental isotopic distributions of parent ions showed some
239 differences with respect to the theoretical ones. For each supposed PTM, the difference between the
240 parent ions experimental mass and the theoretical mass of the corresponding naked peptide was
241 calculated. Results showed mass differences close to the theoretical mass of PTM (Phosphorylation:
242 +79.9797; HexNAcylation: +203.0794). Furthermore, in each case the mass accuracy of the modified
243 parent ion signal is below 10 ppm, providing strong evidence of its supposed PTM. Finally, the
244 intensity of each modified parent ion was compared to the intensity of the corresponding unmodified
245 peptide to determine a PTM-modified peptides ratio. This ratio is very low in each case and far from
246 the predicted average of 10% for phosphorylation (Ma and Hart, 2014). Indeed, these PTMs are
247 extremely labile, making their investigation very hard in real-world samples, due to losses during
248 sample treatment and the limited quantities of material available. In the case of OBP, there is also
249 probably strong ion suppression for detecting modified peptides in the presence of naked peptides,
250 which are much more abundant (Wang et al., 2010).

251

252 4 Discussion

253 We have shown in a previous work (Nagnan-Le Meillour et al., 2014) that around 30 OBP-, VEG-
254 and SAL- isoforms compose the pig olfactory secretome. Among them, only OBP *sensu stricto*
255 isoforms were labeled with specific antibodies raised against phosphorylation and *O*-GlcNAcylation.
256 These two types of PTM, although unusual for secreted proteins, could be involved in odorant and

spectrometry

257 pheromone binding specificity of OBP isoforms. To better understand how PTM could drive OBP
258 binding properties, it is necessary to identify PTM sites and patterns. Indeed, identification by using
259 antibodies raised some concerns because CTD110.6, although specific to *O*-GlcNAc moieties,
260 recognize a large range of modified proteins (Tashima and Stanley, 2014). In particular, they can
261 bind to terminal β -GlcNAc on complex N-glycans, but with careful controls, they can be used to
262 detect proteins processed in the secretory pathway that are modified by *O*-GlcNAc. Several
263 arguments strongly support that OBP labeling obtained here is specific: i) The same OBP isoforms
264 were recognized by both RL2 and CTD110.6 (Nagnan-Le Meillour et al., 2014) and there is no
265 specificity concern for RL2. ii) In the pig olfactory secretome, only SAL bears N-glycan chains on
266 N53 (Loëbel et al., 2000; Scaloni et al., 2001), whereas OBP and VEG do not (Scaloni et al., 2001;
267 Paolini et al., 1998). In this study, only one SAL isoform was detectable with Coomassie blue
268 staining (spot 26), but was not labeled by CTD 110.6. iii) False positive would not be possible in the
269 case of OBP, which is neither N-glycosylated nor *O*-glycosylated (mucin type).

270 The MS analysis described here confirmed that only OBP *sensu stricto* is modified by both
271 phosphorylation and *O*-GlcNAcylation, on the same peptides. If BEMAD method already assessed
272 OBP phosphorylation (Brimau et al., 2010), this is the first time that OBP *O*-GlcNAcylation is
273 confirmed by mass spectrometry.

274 The identification of PTM was derived from PEAKS data, but assigned and controlled
275 manually on MS/MS spectra. Only spot 11 provided enough material to identify PTM without
276 ambiguity. For each peptide, several sites are potentially modified. For sequence 121-137, two
277 threonines (T122 and T136) could be phosphorylated and/or hexNAcylation (Table 2). These two
278 sites were previously identified for phosphorylation by using the BEMAD method (Brimau et al.,
279 2010), T122 in recombinant OBP produced by CHO cells, and T136 in recombinant OBP produced
280 by the yeast *Pichia pastoris*. The peptide of sequence 73-87 could be modified at three positions,

281 T77, Y78, Y82 (Table 2), all 3 potentially modified when using BEMAD (Brimau et al., 2010). In
282 this previous paper, the results of BEMAD were misinterpreted concerning the phosphorylation of
283 Y78 and Y82, because the cycle of tyrosine cannot be opened by beta-elimination. The MS analysis
284 conducted here does not bring any additional information about site assignment. Concerning
285 hexNAcylation, this is the first time that such a modification is assessed by mass spectrometry on
286 OBP peptides, confirming the data obtained by 2-DE and labeling with specific antibodies. In the
287 sequence 60-87, three amino acids, S67, T71 and T77 could bear a GlcNAc moiety (Table 2). We
288 cannot totally exclude that Y78 and Y82 could also be modified even if intracellular *O*-
289 GlcNAcylation is assumed to only occur on serine and threonine. Indeed, Jank et al. (2015) reported
290 the transfer of UDP-GlcNAc onto tyrosine by a new bacterial glycosyltransferase involved in the
291 pathogenicity of *Yersinia* species. The *O*-GlcNAcylation of secreted proteins is new and not well
292 understood, and the function of EOGT has not been investigated in mammals, but could differ from
293 that of OGT, the intracellular enzyme. Indeed, there is no consensus sequence described until now,
294 which could predict the nature of amino acid residues targeted by this PTM. The hypothesis of
295 tyrosine *O*-GlcNAcylation should be kept in mind for further studies.

296 It is remarkable that the same peptides were identified as modified by either phosphorylation
297 or hexNAcylation, and particularly on the sole peptide that differs between the products of the two
298 splicing variants (OBPX1⁸¹NYAGNNK⁸⁷ and OBPX2⁸¹NCNNK⁸⁵). Only the peptide specific to the
299 OBPX1 isoform was identified as modified by both PTM. The fact that the same peptides could be
300 modified by phosphorylation and *O*-GlcNAcylation raises questions about the possibility of a
301 phospho/GlcNAc balance on extracellular proteins, as well as on nuclear and cytoplasmic ones. But
302 to date, no ER-resident or extracellular counterpart of *O*-GlcNAcase, the enzyme that removes
303 GlcNAc residues from intracellular modified proteins (OGA), has been identified. There is another
304 hypothesis that does not involve an OGA-like enzyme. In this scenario, OBP could be modified in
305 the ER, secreted in the extracellular compartment, and internalized after binding with an odorant

spectrometry

306 and/or olfactory receptor, this latter event being evidenced by Strotmann and Breer (2011) in the
307 mouse sustentacular cells of olfactory epithelium. This scheme fits well with the rapid turnover of
308 OBPs described in the nasal mucus. If secreted together, phosphorylated and *O*-GlcNAcylated OBP
309 isoforms could have different binding properties towards ligands, extending their binding repertoire,
310 thereby supporting an active role of OBPs in odorant discrimination.

311 Functional studies (binding assays) are needed to address this question. However, before
312 realizing site-directed mutagenesis for this purpose, it is necessary to identify *O*-GlcNAc sites and
313 this remains a challenging task, especially when several amino acids of the same peptide could be
314 modified. Indeed, the study of native proteins in complex mixture could explain why we did not find
315 the oxonium ions (204.08) for HexNAcylation, typical of the HCD method (Zhao et al., 2011). These
316 ions are usually detected in glycoprotein samples enriched in *O*-GlcNAcylated peptides (Toghi
317 Eshghi et al., 2016) or exclusively composed of synthetic *O*-GlcNAcylated peptides (Chalkley and
318 Burlingame, 2001; Malaker et al., 2017). In this study, identification of modified peptides was made
319 by a manual assignment approach instead of using software as it is done in almost all PTM
320 identification/localization papers. PEAKS software is useful to decrease the spectra number to a few
321 spectra of interest, but should mostly be used as a first approach of investigation. When ones looks
322 carefully at most of the spectra attributed to phosphorylated/hexNAcylated peptides by PEAKS
323 software, there is no obvious proof of such modifications according to the MS/MS spectrum as the
324 modification is the first to be lost in the collision activated dissociation. We therefore recommend
325 that the software should be best used to limit the essential manual assignment work to a reduced
326 number of relevant spectra instead of using it blindly. ESI-ETD-MS/MS has been reported to be the
327 only method able to confidently localize the modification sites (Alfaro et al., 2012; Ma and Hart,
328 2014) because the energy of fragmentation is low and preserves the link between the protein and
329 labile PTMs. But ETD (Electron Transfer Dissociation) is less sensitive than HCD and less useful for

330 complex mixtures than for single proteins. In the absence of an ETD-MS/MS, we were able to
331 characterize four peptides modified by PTMs without any enrichment. The first step to have all
332 proofs of hexNAcylation is to enrich our sample in modified peptides by the use of specific columns
333 to observe oxonium ions with HCD fragmentation. However, HCD fragmentation removes PTMs in
334 priority making their localization possible only if one S/T/Y residue is present on the characterized
335 modified peptide. The second step could be the use of ETD fragmentation on modified peptides
336 characterized with HCD fragmentation. However, ETD requires more modified peptides overall
337 quantity than HCD, making enrichment a crucial step for our future studies.

338 In conclusion, this work is the first step in PTM localization on OBP isoforms from a
339 complex sample of whole secretome. It opens the way to molecular and cellular studies of this new
340 metabolic pathway involved in olfaction.

341 **5 Conflict of Interest**

342 The authors declare that the research was conducted in the absence of any commercial or financial
343 relationships that could be construed as a potential conflict of interest.

344 **6 Author Contributions**

345• PNLM, CT and CR conceptualized the project. 2-DE analyses were performed by JB and PNLM.
346 CLD and JB identified isoforms. High-resolution mass spectrometry was performed by FB. FB, JB
347 and CR analyzed Peaks data. All authors have participated in the preparation of the manuscript.

348 **7 Funding**

349 This work was funded by French INRA (Institut National de Recherche Agronomique) and CNRS
350 (Centre National de Recherche Scientifique).

351 **8 References**

spectrometry

- 352 Alfaro JF, Gong CX, Monroe ME, Aldrich JT, Clauss TR, Purvine SO, Wang Z, Camp DG 2nd,
353 Shabanowitz J, Stanley P, Hart GW, Hunt DF, Yang F, Smith RD (2012) Tandem mass spectrometry
354 identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc
355 transferase targets. Proc Natl Acad Sci USA 109:7280-7285. doi:10.1073/pnas.1200425109
- 356 Brimau F, Cornard JP, Le Danvic C, Lagant P, Vergoten G, Grebert D, Pajot E, Nagnan-Le Meillour
357 P (2010) Binding specificity of recombinant Odorant Binding Protein isoforms is driven by
358 phosphorylation. J Chem Ecol 36:801-813. doi:10.1007/s10886-010-9820-4
- 359 Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis
360 for odor recognition. Cell 65:175-187. doi:10.1016/0092-8674(91)90418-X
- 361 Castellino FJ, Ploplis VA, Zhang L (2008) Gamma-glutamate and beta-hydroxyaspartate in proteins.
362 Methods Mol Biol 446:85-94. doi:10.1038/msb.2012.31
- 363 Chalkley RJ, Burlingame AL (2001) Identification of GlcNAcylation sites of peptides and α -
364 crystallin using Q-TOF mass spectrometry. J Am Soc Mass Spectrom, 12: 1106-1113.
365 doi:10.1016/S1044-0305(01)00295-1
- 366 Guiraudie G, Pageat P, Cain AH, Madec I, Nagnan-Le Meillour P (2003) Functional characterization
367 of olfactory binding proteins for appeasing compounds and molecular cloning in the vomeronasal
368 organ of pre-pubertal pigs. Chem Senses 28:609-619. doi:10.1093/chemse/bjg052
- 369 Harris RJ, Ling VT, Spellman MW (1992) O-linked fucose is present in the first epidermal growth
370 factor domain of factor XII but not in protein C. J Biol Chem 267:5102-5107
- 371 Hart GW, Akimoto Y (2009) in Essentials of Glycobiology. 2nd edition. Chapter 18, The O-GlcNAc
372 modification. Copyright © 2009, The Consortium of Glycobiology Editors, La Jolla, California.
373 Bookshelf ID: NBK1954PMID: 20301273. Varki A, Cummings RD, Esko JD, et al., editors. Cold
374 Spring Harbor (NY): Cold Spring Harbor Laboratory Press

375 Hoffman BR, Liu Y, Mosher DF (2012) Modification of EGF-like module 1 of thrombospondin-1,
376 an animal extracellular protein, by O-linked N-acetylglucosamine. *PloS one* 7,e32762.
377 doi:10.1371/journal.pone.0032762

378 Jank T, Eckerle S, Steinemann M, Trillhaase C, Schimpl M, Wiese S, van Alten DMF, Driever W,
379 Aktories K (2015) Tyrosine glycosylation of Rho by Yersinia toxin impairs blastomere cell
380 behaviour in Zebrafish embryos. *Nat Commun* 6:7807. doi:10.1038/ncomms8807

381 Loëbel D, Scaloni A, Paolini S, Fini C, Ferrara L, Breer H, Pelosi P (2000) Cloning, post-
382 translational modifications, heterologous expression and ligand-binding of boar salivary lipocalin.
383 *Biochem J* 350:369-379. doi:10.1042/bj3500369

384 Ma J, Hart GW (2014) O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics* 11:8.
385 doi:10.1186/1559-0275-11-8

386 Malaker SA, Penny SA, Steadman LG, Myers PT, Loke JC, Raghavan M, Bai DL, Shabanowitz J,
387 Hunt DF, Cobbold M (2017) Identification of glycopeptides as posttranslationally modified
388 neoantigens in leukemia. *Cancer Immunol Res* 5(5): 376-384. doi:10.1158/2326-6066.CIR-16-0280

389 Nagnan-Le Meillour P, Vercoutter-Edouart AS, Hilliou F, Le Danvic C, Lévy F (2014) Proteomic
390 analysis of pig (*Sus scrofa*) olfactory soluble proteome reveals O-linked-N-acetylglucosaminylation
391 of secreted odorant-binding proteins. *Front Endocrinol* 5:202. doi:10.3389/fendo.2014.00202

392 Paolini S, Scaloni A, Amoresano A, Marchese S, Napolitano E, Pelosi P (1998) Amino acid
393 sequence, post-translational modifications, binding and labelling of porcine odorant-binding protein.
394 *Chem Senses* 23:689–698.

395 Pelosi P (1996) Perireceptor events in olfaction. *J Neurobiol* 30:3-19. doi:10.1093/chemse/23.6.689

396 Sakaidani Y, Nomura T, Matsuura A, Ito M, Suzuki E, Murakami K, Nadano D, Matsuda T,
397 Furukawa K, Okajima T (2011) O-linked-N-acetylglucosamine on extracellular protein domains
398 mediates epithelial cell-matrix interactions. *Nat Commun* 2:583. doi:10.1038/ncomms1591

spectrometry

- 399 Scaloni A, Paolini S, Brandazza A, Fantacci M, Bottiglieri C, Marchese S, Navarrini A, Fini C,
400 Ferrara L, Pelosi P (2001) Purification, cloning and characterisation of odorant-binding proteins from
401 pig nasal epithelium. *Cell Mol Life Sci* 58:823-834. doi:10.1007/PL00000903
- 402 Schwarz F, Aebi M (2011) Mechanisms and principles of N-linked protein glycosylation. *Curr Opin*
403 *Struct Biol* 21:576-582. doi:10.1016/j.sbi.2011.08.005
- 404 Shaheen R, Aglan M, Keppler-Noreuil K, Faqeih E, Ansari S, Horton K, Ashour A, Zaki MS, Al-
405 Zahrani F, Cueto-Gonzales AM, Abdel-Salam G, Temtamy S, Alkuraya FS (2013) Mutations in
406 EOGT confirm the genetic heterogeneity of autosomal-recessive Adams-Oliver syndrome. *Am J*
407 *Hum Genet* 92:598-604. doi:10.1016/j.ajhg.2013.02.012
- 408 Spinelli S, Ramoni R, Grolli S, Bonicel J, Cambillau C, Tegoni M (1998) The structure of the
409 monomeric porcine odorant binding protein sheds light on the domain swapping mechanism.
410 *Biochemistry* 37:7913-7918. doi:10.1021/bi980179e
- 411 Strotmann R, Breer H (2011) Internalization of odorant-binding proteins into the mouse olfactory
412 epithelium. *Histochem Cell Biol* 136:357-369. doi:10.1007/s00418-011-0850-y
- 413 Tagliabracci VS, Wiley SE, Guo X, Kinch LN, Durrant E, Wen J, Xiao J, Cui J, Nguyen KB, Engel
414 JL, Coon JJ, Grishin N, Pinna LA, Pagliarni DJ, Dixon JE (2015) A single kinase generates the
415 majority of the secreted phosphoproteome. *Cell* 161:1619-1632. doi:10.1016/j.cell.2015.05.028
- 416 Tashima Y, Stanley P (2014) Antibodies that detect *O*-GlcNAc on the extracellular domain of cell
417 surface glycoproteins. *J Biol Chem* 289:11132-42. doi:10.1074/jbc.M113.492512
- 418 Toghi Eshghi S, Yang W, Hu Y, Shah P, Sun S, Li X, Zhang H (2016) Classification of tandem mass
419 spectra for identification of N- and *O*-linked glycopeptides. *Sci Rep* 6:37189. doi:10.1038/srep37189
- 420 Torres CR, Hart GW (1984) Topography and polypeptide distribution of terminal N-
421 acetylglucosamine residues on the surface of intact lymphocytes. Evidence for *O*-linked GlcNAc. *J*
422 *Biol Chem* 259:3308-3317

423 Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y,
424 Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H (2016) 2016 update of the PRIDE database
425 and related tools. *Nucleic acid Res* 44(D1):D447-D456.

426 Wang Z, Hart GW (2008) Glycomic approaches to study *O*-GlcNAcylation: protein identification,
427 site-mapping, and site-specific *O*-GlcNAc quantitation. *Clin Proteomics* 4:5-13. doi:10.1007/s12014-
428 008-9008-x

429 Zhao P, Viner R, Teo CF, Boons GJ, Horn D, Wells L (2011) Combining high-energy C-trap
430 dissociation and electron transfer dissociation for protein *O*-GlcNAc modification site assignment. *J*
431 *Prot Research* 10:4099-4104/ doi:10.1021/pr2002726

432 **9 Figure legends**

433 Figure 1: Two-dimensional electrophoresis of soluble proteins extracted from two pubertal male pigs
434 M14 and M39. Merge of Coomassie-blue staining of M14 and M39 secretomes obtained by 2-DE. X-
435 axis: pI, Y-axis: molecular weight

436 Figure 2: Identification of OBP, VEG, and SAL with specific antibodies in M14 (A-D) and M39 (E-
437 H)) secretomes. A) and E) Coomassie-blue staining. B) and F) Merge of Coomassie-blue staining
438 (red) and western-blot with anti-OBP (green). C) and G) Merge of Coomassie-blue staining (red) and
439 western-blot with anti-VEG (green). D) and H) Merge of Coomassie-blue staining (red) and western-
440 blot with anti-SAL (green). X-axis: pI, Y-axis: molecular weight

441 Figure 3: Identification of potential phosphorylation and *O*-GlcNAcylation of M14 (A-B) and M39
442 (C-D) protein spots. M14: (A,C) represent the merging of Coomassie-blue staining (red) and
443 western-blot with anti-*O*-GlcNAc (CTD110.6, green) for individuals M14 and M39, respectively.
444 (B,D) represent the merging of Coomassie-blue staining (red) and western-blot with anti-
445 phosphoS/Y/T (green) for M14 and M39, respectively. X-axis: pI, Y-axis: molecular weight

spectrometry

446 Figure 4: Map of M14 sample spotting (Coomassie blue staining) for nanoLC-nanoESI-HCD-
 447 MS/MS-Orbitrap analysis. X-axis: pI, Y-axis: molecular weight. Numbering corresponds with data in
 448 Table S1.

449 Figure 5: MS/MS spectra of modified peptides and the relevant fragments for their attribution. A)
 450 Spectrum n°27502. B) Spectrum n°13801. C) Spectrum n°28697. D) Spectrum n°20288.

451 Figure 6: MS spectra of modified peptides and their theoretical isotopic distribution (in red). A)
 452 Spectrum n°27502. B) Spectrum n°13801. C) Spectrum n°28697. D) Spectrum n°20288.

453 **10 Tables**

Custom name	Protein name	Average mass (Da)	Comments	Accession number
P81245	OBPX1J	17836	PAK C-terminal	NCBI OBP X1 isoform GI/47523248
P81246	OBPX1A	17647	PA C-terminal	GI/3122574
P81247	OBPX2J	17707	PAK C-terminal	NCBI OBP X2 isoform GI/545883991
P81248	OBPX2A	17519	PA C-terminal	Derived from OBP X2 isoform (no ref.)
P53715	VEG VNO	17445	VNO variant	GI/1718160

P53716	VEG RM	17461	RM variant	GI/27657971
F1SN92	SAL B	19902	B variant: V45- I48-A73	GI/21465464
F1SN93	SAL A	19916	A variant: A45- V48-V73	GI/20178087

454 Table 1: Custom-made OBP database used for characterization of 2-DE spots from M14 sample

455

Spectrum number	Peptide	PEAKS score	PTM	Potential Sites (underlined)
27502	121-137	100.04	Phosphorylation	<u>G</u> <u>T</u> <u>D</u> <u>I</u> <u>E</u> <u>D</u> <u>Q</u> <u>D</u> <u>L</u> <u>E</u> <u>K</u> <u>F</u> <u>K</u> <u>E</u> <u>V</u> <u>T</u> <u>R</u> (T122, T136)
13801	73-87	72.53	Phosphorylation	<u>Q</u> <u>E</u> <u>G</u> <u>N</u> <u>T</u> <u>Y</u> <u>D</u> <u>V</u> <u>N</u> <u>Y</u> <u>A</u> <u>G</u> <u>N</u> <u>N</u> <u>K</u> (T77, Y78, Y82)
28697	60-87	80.45	HexNAcylation	<u>N</u> <u>G</u> <u>I</u> <u>C</u> <u>E</u> <u>E</u> <u>F</u> <u>S</u> <u>L</u> <u>I</u> <u>G</u> <u>T</u> <u>K</u> <u>Q</u> <u>E</u> <u>G</u> <u>N</u> <u>T</u> <u>Y</u> <u>D</u> <u>V</u> <u>N</u> <u>Y</u> AGNNK (S67, T71, T77, Y78? Y82?)

spectrometry

			HexNAcylation	<u>T</u> DIEDQDLEKFKEV <u>T</u> R
20288	122-137	71.84	Aspartate hydroxylation	(T122, T136)

456 Table 2: Identification of the peptides modified by phosphorylation and hexNAcylation: PEAKS

457 analysis of PTMs in spot 11 from M14 sample

458

Spectrum number	Peptide	Δm		PTM	Δm naked peptide (ppm)	% modified peptide
		Δm Parent (ppm)	(expm(parent)- thm(naked peptide)) m/z with z = 1			
27502	121-137	4.1	79.9576	Phosphorylation (+79.9797)	0.6	0.17
13801	73-87	2.3	79.9706	Phosphorylation (+79.9797)	/	0.17
28697	60-87	4.1	203.0658	HexNAcylation (+203.0794)	/	2.45

				HexNAcylation		
				Aspartate		
20288	122-137	7.8	219.0864	hydroxylation	/	0.81
				(+219,0788)		

459 Table 3: Manual analysis of PTMs raw data obtained for spot 11 from M14 sample

In review

Figure 1.TIFF



Figure 2.TIFF

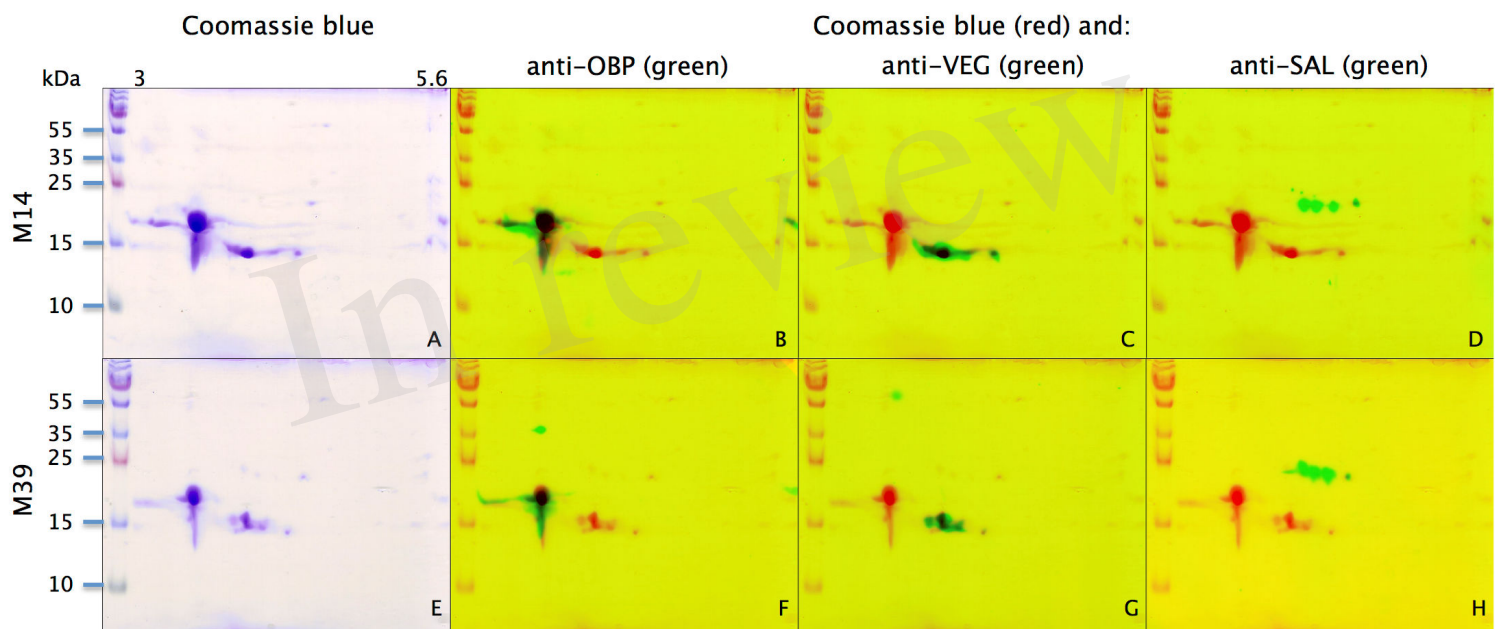


Figure 3.TIFF

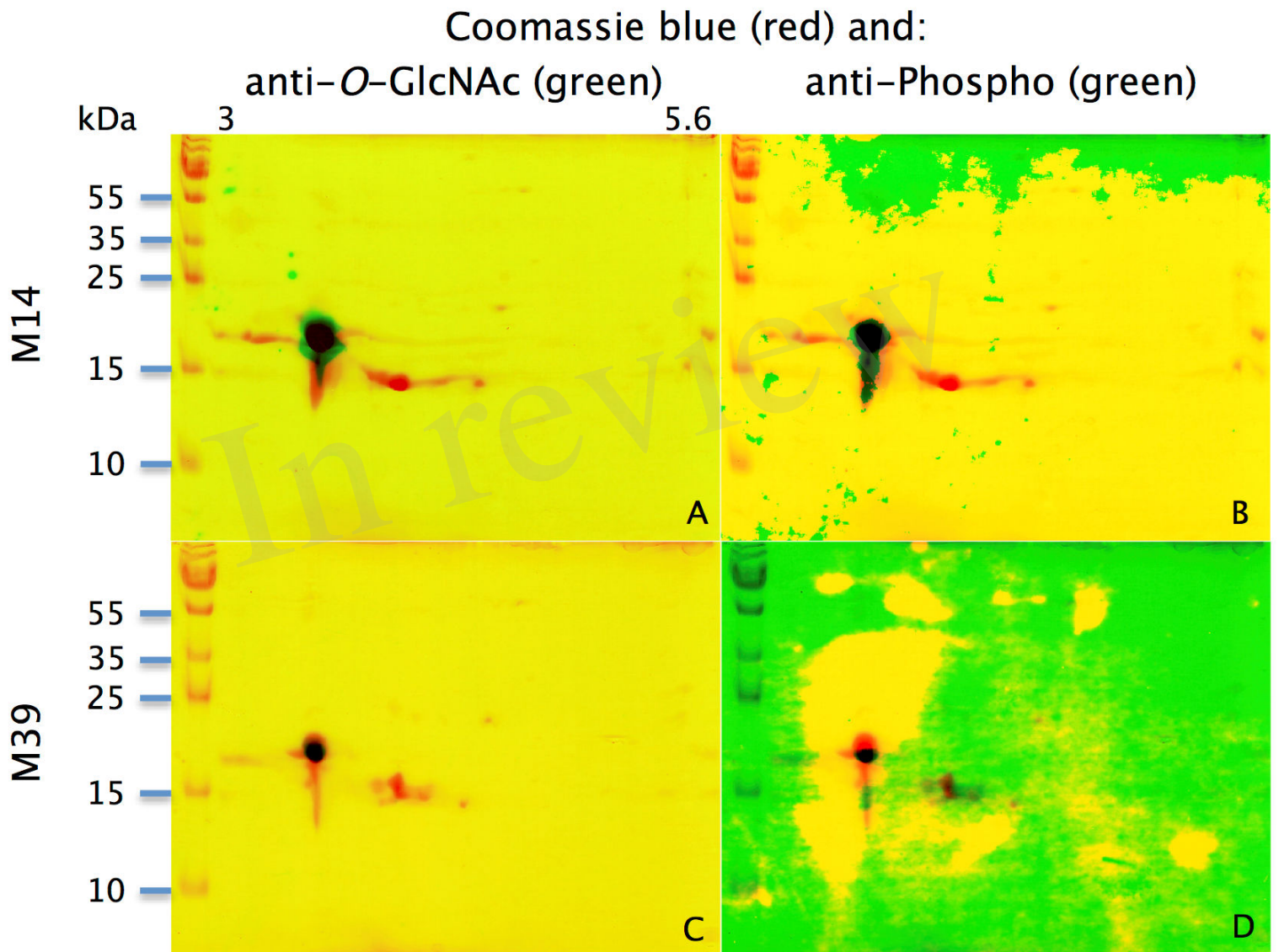


Figure 4.TIFF

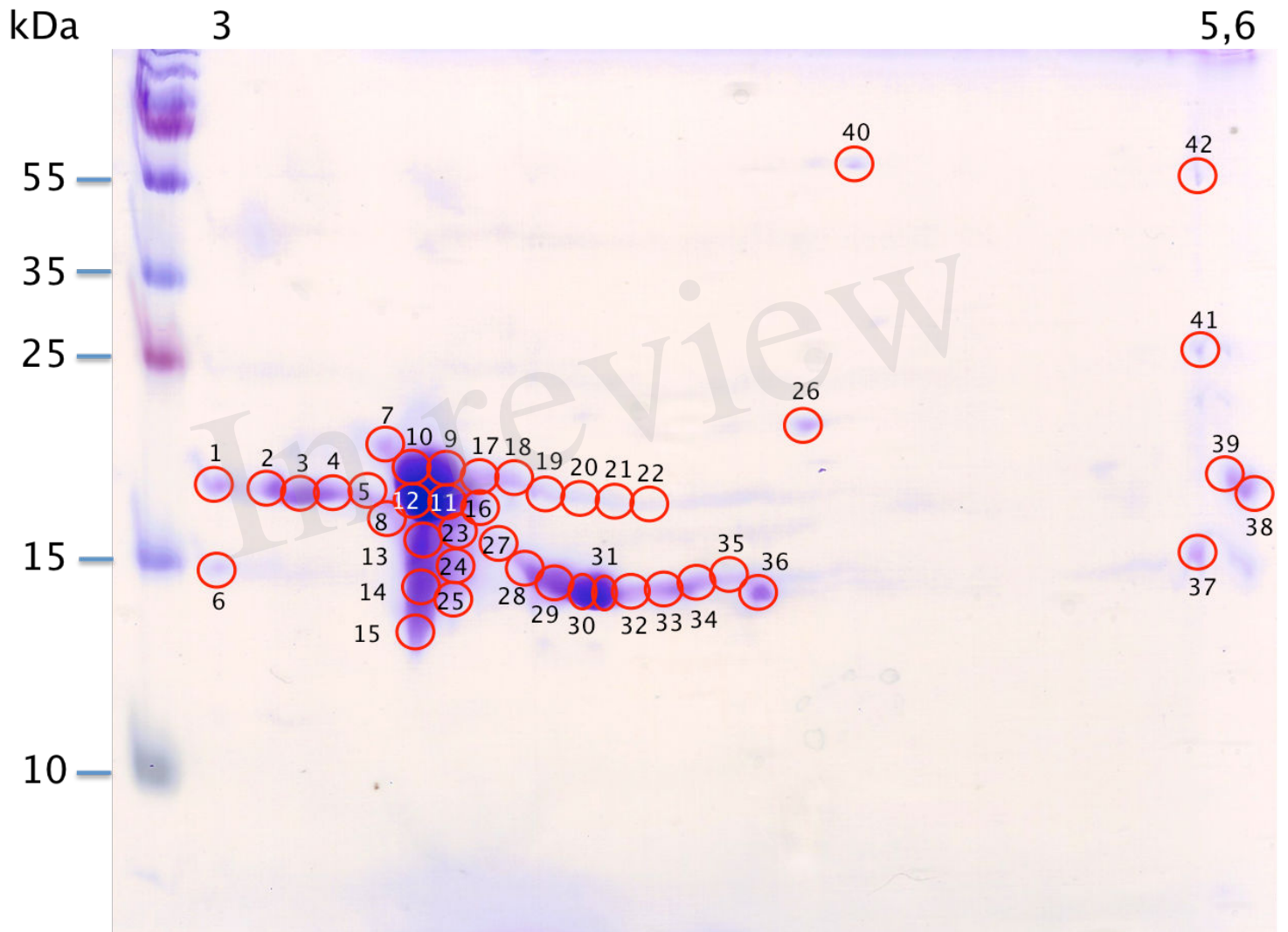
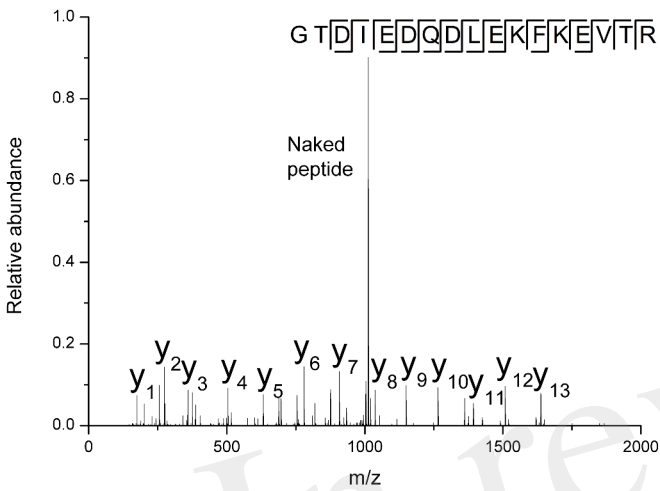
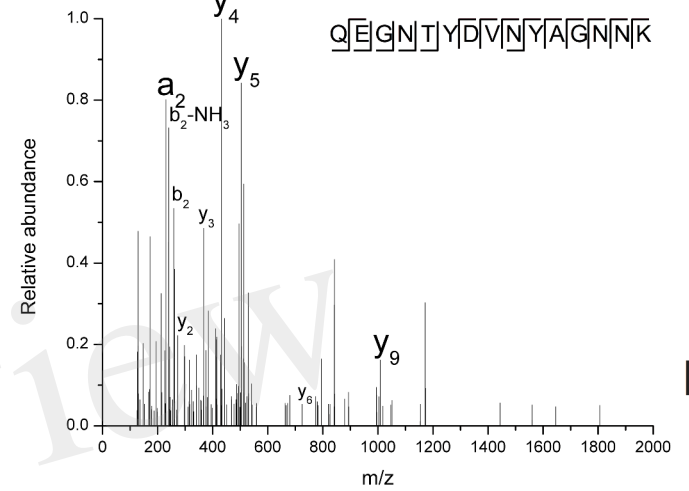


Figure 5.TIFF

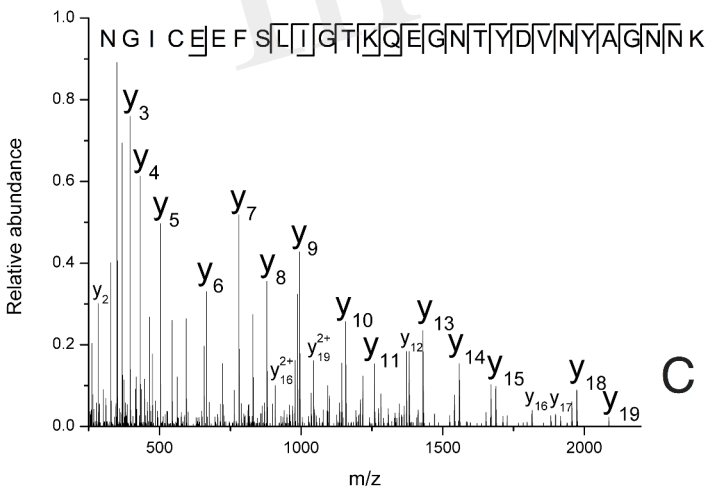
Spectrum 27502: m/z = 1052.9822, z = 2, RT = 108.78, -10lgP = 100.04



Spectrum 13801: m/z = 883.8569, z = 2, RT = 63.11, -10lgP = 72.53



Spectrum 28697: m/z = 1094.5024, z = 3, RT = 112.14, -10lgP = 80.45



Spectrum 20288: m/z = 547.0193, z = 4, RT = 86.55, -10lgP = 71.84

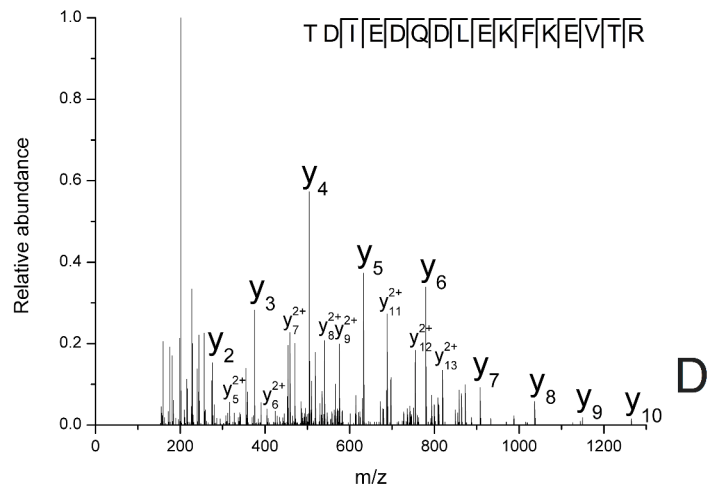
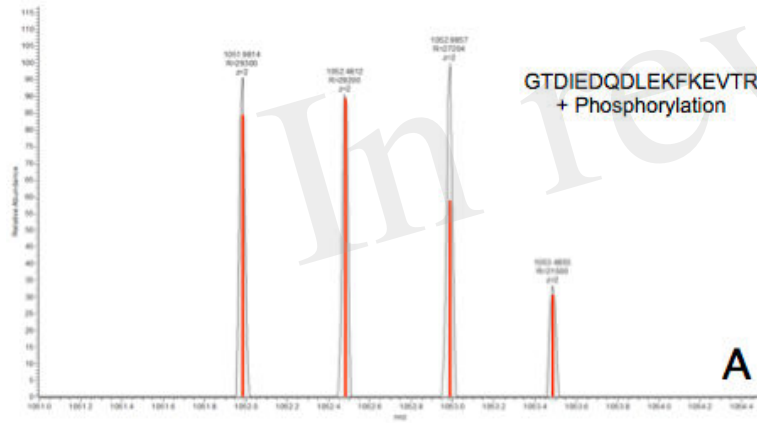
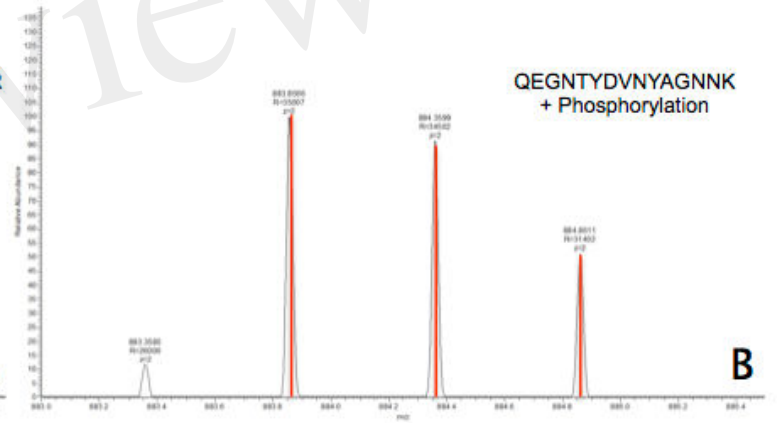


Figure 6.TIFF

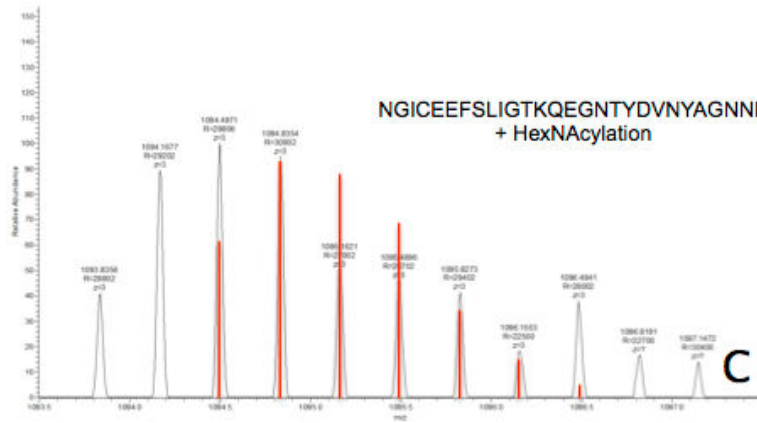
20_11421998_RT 130.16 AV 1 NL 3.76E5
T FTMS - y NS Full scan (500.00-2000.00)



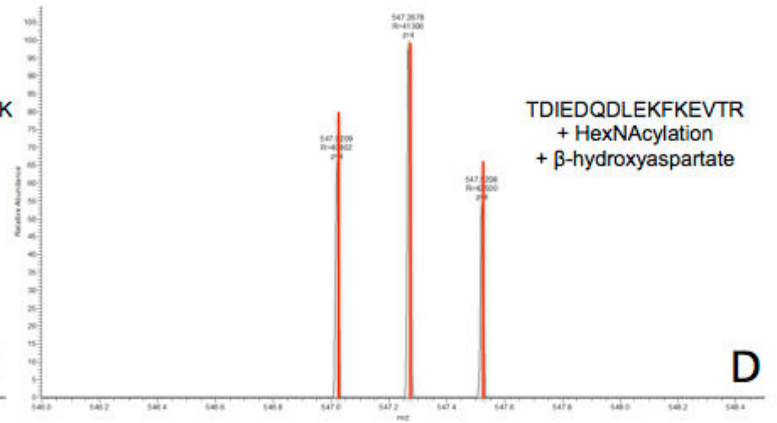
20_11413607_RT 133.11 AV 1 NL 1.82E5
T FTMS - y NS Full scan (500.00-2000.00)



20_11420988_RT 112.14 AV 1 NL 7.51E5
T FTMS - y NS Full scan (500.00-2000.00)



20_11420280_RT 60.52 AV 1 NL 1.52E6
T FTMS - y NS Full scan (500.00-2000.00)



Supplementary Material

Identification of post-translational modifications on Odorant-Binding Protein isoforms from pig olfactory secretome by high-resolution mass spectrometry: *O*- β -*N*-acetylglucosaminylation and phosphorylation

Julien Bouclon^{1,2‡}, Chrystelle Le Danvic^{1,3}, Elodie Guettier⁴, Fabrice Bray², Caroline Tokarski², Christian Rolando², Patricia Nagnan-Le Meillour^{1,*}

* **Correspondence:** Patricia Nagnan-Le Meillour: patricia.le-meillour@univ-lille1.fr

1 Supplementary Data

Sequences used for the constructed database

Porcine OBPs

GI/47523248 (OBP X1 J)

QEPQPEQDPF ELSGKWITSY IGSSDLEKIG ENAPFQVFMR SIEFDDKESK
VYLNFFSKEN GICEEFSLIG TKQEGNTYDV NYAGNNKFVV SYASETALII
SNINVDEEGD KTIMTGLLGK GTDIEDQDLE KFKEVTRENG IPEENIVNII
ERDDCPAK

GI/545883991 (OBP X2 J)

QEPQPEQDPF ELSGKWITSY IGSSDLEKIG ENAPFQVFMR SIEFDDKESK
VYLNFFSKEN GICEEFSLIG TKQEGNTYDV NCNNKFVVSY ASETALIISN
INVDEEGDKT IMTGLLGKGT DIEDQDLEKF KEVTRENGIP EENIVNIER
DDCPAK

GI/3122574 (OBP X1 A)

QEPQPEQDPF ELSGKWITSY IGSSDLEKIG ENAPFQVFMR SIEFDDKESK
VYLNFFSKEN GICEEFSLIG TKQEGNTYDV NYAGNNKFVV SYASETALII
SNINVDEEGD KTIMTGLLGK GTDIEDQDLE KFKEVTRENG IPEENIVNII
ERDDCPA

No reference, derived from GI/3122574 (OBP X2 A)

QEPQPEQDPF ELSGKWITSY IGSSDLEKIG ENAPFQVFMR SIEFDDKESK
VYLNFFSKEN GICEEFSLIG TKQEGNTYDV NCNNKFVVSY ASETALIISN
INVDEEGDKT IMTGLLGKGT DIEDQDLEKF KEVTRENGIP EENIVNIER
DDCPA

Porcine VEG

GI/1718160 Vomeronasal Organ (VNO) variant

QEFPAVGQPL QDLLGRWYLK AMTSDPEIPG KKPESVTPLI LKALEGGDLE
AQITFLIDGQ CQDVTLVLKK TNQPFTFTAY DGKRVVYILP SKVKDHYILY
CEGELDGQEV RMAKLVGRDP ENNPEALEEF KEVARAKGLN PDIVRPQQSE
TCSPGGN

GI/27657971 Respiratory Mucosa (RM) variant

QEFPAVGQPL QDLLGRWYLK AMTSDPEIPG KKPESVTPLI LKALEGGDLE
AQITFLIDGQ CQDVTLVLKK TNQPFTFTAY DGKRVVYILP SKVKDHYILY
CEGELDGQEV RMAKLVGRDP ENNPEALEEF KEVARAKGLN LDIVRPQQSE
TCSPGGN

Salivary lipocalin, porcine SAL

GI/20178087 A variant

HKEAGQDVVT SNFDASKIAG EWYSILLASD AKENIEENGs MRVFAEHVRV
LDNSSLAFKF QRKVNGECTD FYVVC DKVGD GVYTVAYYGE NKFRLLEVNY
SDYVILHLVN VNGDKTFQLM EfyGRKPDVE PKLKDKFVEI CQYGIKEN
IIDLTKIDRC FQLRSGGVQ ESSAE

GI/21465464 B variant

HKEAGQDVVT SNFDASKIAG EWYSILLASD AKENIEENGs MRVFVEHIRV
LDNSSLAFKF QRKVNGECTD FYAVC DKVGD GVYTVAYYGE NKFRLLEVNY

SDYVILHLVN VNGDKTFQLM EFGYGRKPDVE PKLKDKFVEI CQQYGIKEN
 IIDLTKIDRC FQLRSGSGVQ ESSAE

2 Supplementary Tables

2.1 Supplementary Table S1

Spot number	Primary Sequence	PEAKS score	N° Peptides	% of covered sequence
1	OBP X1	307.18	13	52
2	OBP X1	309.30	7	45
3	OBP X1	491.48	20	63
4	OBP X1	412.13	18	76
5	OBP X1	309.79	6	45
6	VEG	72.96	1	5
7	OBP X1	364.82	18	63
8	OBP X1	336.55	13	52
9	OBP X1, X2 (A & J)	407.73	96	91
11	OBP X1, X2 (A & J)	603.29	285	93
12	OBP X1 (A & J)	514.93	79	79
13	OBP X1 (A & J)	450.35	62	85
	VEG	146.03	4	32
14	OBP X1 (A & J)	568.41	66	85
	VEG RM	170.70	2	21
15	OBP X1 A	472.32	23	76
16	OBP X1	503.24	29	73
17	OBP X1	337.53	9	48
18	OBP X1	424.24	25	82
19	OBP	486.35	9	53
20	OBP X1	262.96	6	44
	VEG	100.26	2	17
21	OBP X1	280.58	5	38
23	OBP X1	531.07	32	82
24	OBP X1	433.10	34	82
	VEG	223.46	5	31
25	OBP X1	377.73	23	70
	VEG	206.18	4	29
26	OBP	235.24	6	43
	SAL	144.35	2	13
27	OBP X1	276.70	6	47
	VEG RM	532.43	24	24
28	OBP	194.33	3	25
	VEG RM	443.62	15	61
29	OBP	84.80	1	7
	VEG	195.30	3	20
30	OBP	201.65	4	24

	VEG RM	612.90	99	84
31	OBP	82.22	1	8
	VEG RM	694.63	73	76
32	OBP	46.25	1	8
	VEG	58.20	2	5
33	OBP	69.17	1	8
	VEG RM	345.47	18	61
34	OBP	61.52	1	8
	VEG RM	209.28	9	47
35	OBP	108.76	1	8
	VEG	75.02	1	5
36	OBP	78.64	1	8
	VEG RM	257.19	5	30
37	OBP	61.13	1	8
	VEG	64.52	1	5
38	OBP	67.14	1	8
39	OBP	287.86	9	40
40	OBP	124.58	2	13
41	OBP	65.39	1	8
42	OBP	82.52	1	8

Supplementary Table S1: Identification of primary sequences contained in 2DE spots from M14 sample by ESI-HCD-MS/MS

La spectrométrie de masse est une méthode d'analyse qui permet de travailler sur une large gamme d'échantillons. Elle est utilisée dans de nombreux domaines de recherche comme la chimie analytique, la protéomique, la lipidomique et la métabolomique...

Dans un premier temps, mon travail s'est focalisé sur le développement d'une méthode indépendante d'analyse de données par spectrométrie de masse à transformée de Fourier bidimensionnelle. Pour augmenter la résolution en première dimension, une analyse FT-ICR 2D avec un échantillonnage non uniforme (NUS) a été développée. L'augmentation de la résolution dans la première dimension a permis l'obtention d'une haute résolution pour les ions précurseurs. La FT-ICR 2D a été utilisée avec succès pour l'analyse de triacylglycérols contenus dans du plasma mais aussi pour l'analyse d'échantillons archéologiques.

Dans un second temps, une stratégie protéomique conjointe bottom-up et top-down a été appliquée à l'analyse d'échantillons archéologiques et paléontologiques à partir d'ossements ou de céramiques. Le développement d'une méthodologie bottom-up, a permis à partir d'ossements archéologiques d'espèces inconnues l'identification des protéines et leurs modifications chimiques. Ces ossements ont pu être attribués comme appartenant à *Homo sapiens* et l'absence de substitution d'acide aminé établie. Une analyse top-down a été utilisée pour l'analyse d'échantillons archéologiques. Pour la première fois, une protéine (la caséine de lait) a été identifiée dans un échantillon archéologique d'amphore de l'époque de l'empereur Claude (1er siècle de notre ère) grâce la détection de grands fragments de caséine.

Mots clés : 2D FT-ICR, archéologie, paléontologie, protéomique bottom-up, protéomique top-down

Mass spectrometry is a method of analysis which works on a wide range of sample types. It is used in many research fields such as analytical chemistry, proteomics, lipidomics and metabolomics...

Firstly, my work was dedicated to the development of an independent data analysis methodology based on two-dimensional Fourier transform mass spectrometry. For increasing the resolution on the first dimension a 2D FT-ICR analysis with non-uniform sampling was developed. The resolution increase in the first dimension leading to high resolution for the precursor ions. The 2D FT-ICR has been successfully applied for the analysis of triacylglycerol contained on plasma and also for archaeological samples. This methodology led to 2D maps allowing a rapid classification of plants or animals samples.

Secondly, a joint bottom-up and top-down proteomics strategy was applied for the analysis of archaeological and paleontological samples from bones or ceramics. The development of a bottom-up methodology, allowed the identification of proteins and their chemical modifications from archaeological bones. These bones have been attributed to *Homo sapiens* with no amino acid substitutions. The development of a top-down methodology was applied to the analysis of archaeological ceramic. For the first time, a protein (milk casein) was identified in an archaeological sample of an amphora from Claudius emperor era (1st century A.D) via the detection of large fragments of casein. This first application of the top down proteomics showed that new information can be provided such as the in situ molecular degradation.

Keywords: 2D FT-ICR, archeology, paleontology, bottom-up proteomics, top-down proteomics