



**HAL**  
open science

# Quantification multiplexe de biomarqueurs d'intérêt clinique et de leurs protéoformes par spectrométrie de masse. Application à l'analyse de cohortes médicales

Arthur Viodé

## ► To cite this version:

Arthur Viodé. Quantification multiplexe de biomarqueurs d'intérêt clinique et de leurs protéoformes par spectrométrie de masse. Application à l'analyse de cohortes médicales. Chimie analytique. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLS515 . tel-02887089

**HAL Id: tel-02887089**

**<https://theses.hal.science/tel-02887089>**

Submitted on 2 Jul 2020

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

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

# Quantification multiplexe de biomarqueurs d'intérêt clinique et de leurs protéoformes par spectrométrie de masse. Application à l'analyse de cohortes médicales

Thèse de doctorat de l'Université Paris-Saclay  
Préparée au CEA de Saclay et à l'Université Paris Sud

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

Thèse présentée et soutenue à Saclay, le 05/12/2018, par

**Arthur Viodé**

Composition du Jury :

Claire Smadja Professeur des universités, Institut Galien Paris-Sud	Président
Arnaud Salvador Professeur des universités, Université Lyon 1	Rapporteur
Christophe Hirtz Professeur des universités, Université Montpellier	Rapporteur
Virginie Redeker Chargé de recherche, Institut des Neurosciences CNRS	Examineur
Foudil Lamari Praticien hospitalier, UMPC, Sorbonne Université	Examineur
François Becher Ingénieur de recherche, CEA Saclay	Directeur de thèse
Isabelle Le Potier Maître de conférences, C2N Université Paris Sud	Invité



# Remerciements

Ce travail de thèse a été réalisé au CEA de Saclay, au Service de Pharmacologie et Immunoanalyse (SPI) dans le Laboratoire d'Etudes du Métabolisme des Médicaments (LEMM). Je tiens à remercier le Dr. Christophe Junot et le Dr. François Fenaille pour leur accueil au sein du SPI et du LEMM.

Ces recherches ont été réalisées sous la direction du Dr. François Becher que je remercie pour son investissement et sa pédagogie. J'ai énormément appris lors de nos échanges.

Je tiens à remercier les différents collaborateurs qui ont permis ce travail de thèse. A l'institut du Cerveau et de la Moelle Epinière (ICM), notamment le Dr. Isabelle Le Ber, le Dr. Clémence Fournier et le Dr. Vincent Anquetil. Au Dr. Foudil Lamari de l'Hôpital de la Pitié-Salpêtrière qui a permis le travail sur la C9ORF72 et l'alpha-synucléine. Ainsi qu'aux Dr. Pierre-Olivier Fernagut et Benjamin Dehay de l'Université de Bordeaux.

Je tiens à remercier tout particulièrement les membres du jury pour leur lecture critique et l'évaluation de ce manuscrit.

Enfin, j'ai passé trois années excellentes et très enrichissantes au sein du SPI grâce à toute l'équipe, merci !

La team Hélène, Christelle et Lolo pour les bons moments passés au Labo et en dehors : le ski, Bruxelles, le Carnaval de Dunkerque, .. Merci à Hélène pour le très bon séjour aux Etats-Unis, Christelle pour les verres du vendredi soir et Lolo pour les bringues, on a bien rigolé avec Chantal. Mes amis et collègues de bureaux de départ : Michou, Pierre, Nono junior, Nono Senior, Samouel et la relève Kathleen, Thaïs, Anne-Sophie (oui je connais ton prénom !) et mon rayon de soleil Mimine !

Un grand merci à toute l'équipe du LEMM et de MedDay.

Merci à Coco et Coco mes amis de Lyon montés sur Paris faire une thèse aussi ! J'ai beaucoup apprécié toutes les soirées que j'ai passé en votre compagnie ! Merci à Coco pour m'avoir accueilli plus d'une fois pour diner, regarder les matches de Lyon et jouer à la Play !

Merci à mes amis Caro, Justine, Berki, Jojo et Adja d'être exceptionnels tout simplement ! Merci à ma famille pour leur soutien et d'être toujours repartis les bras chargés de bonne charcuterie Ardéchoise !

# Table des matières

Table des matières .....	3
Liste des figures.....	7
Liste des tableaux.....	12
INTRODUCTION GENERALE .....	14
I. Les biomarqueurs.....	15
1. Qu'est-ce qu'un biomarqueur ? .....	15
2. Les différents types de biomarqueurs moléculaires .....	15
a) Gènes et ARN.....	15
b) Protéines.....	17
c) Métabolites .....	19
3. Découverte et validation de nouveaux biomarqueurs.....	19
4. Protéines et protéoformes.....	21
a) La notion de protéoformes.....	21
b) Les types de protéoformes et leurs pertinences cliniques.....	22
(i) ADN : polymorphisme nucléotidique et mutations .....	24
(ii) ARN : épissage alternatif .....	26
(iii) Modifications post-traductionnelles, protéolyse endogène.....	27
(a) Phosphorylation .....	27
(b) Glycosylation .....	29
(c) Acetylation.....	30
(d) Acylation.....	33
(e) Méthylation .....	34
(f) Ubiquitination.....	36
(g) Sumoylation.....	38
(h) Protéolyse endogène.....	39
c) Complexité des protéoformes.....	43
II. L'analyse des biomarqueurs protéiques et de leurs protéoformes par spectrométrie de masse	46
1. La spectrométrie de masse .....	46
a) Principe.....	46
b) Apport de la spectrométrie de masse pour l'analyse des protéines et de leurs	
protéoformes .....	48
(i) Gamme de masse adaptée à la diversité des protéines.....	49
(ii) ELISA et spécificité apportée par la spectrométrie de masse .....	51

(iii) Linéarité et les effets matrices .....	54
2. Préparation des échantillons.....	55
a) Précipitation des protéines et déplétion des protéines majoritaires .....	56
b) Extraction en phase solide et fractionnement chromatographique .....	56
c) Immunoprécipitation (protéine/peptides).....	58
3. Les approches de protéomique et leur application à l'analyse des protéoformes.....	61
a) bottom-up .....	62
(i) Trypsine et « shotgun » proteomique.....	62
(ii) extended bottom up.....	68
b) middle-down .....	70
c) Top-down .....	72
(i) Instrumentation :.....	73
(ii) Fragmentation :.....	74
(iii) Préparation d'échantillon :.....	75
(iv) Séparation chromatographique : .....	76
(v) Analyse : .....	76
4. Les approches de quantification par spectrométrie de masse .....	77
a) Quantification relative par spectrométrie de masse .....	78
(i) Label free.....	78
(a) Résultats d'identification.....	78
(b) Les méthodes basées sur l'extraction de l'intensité du signal pour une gamme de m/z : XIC (Extracted ion chromatogram).....	79
(ii) Marquage isotopique SILAC (ou marquage métabolique) .....	80
(iii) Marquage chimique .....	83
(a) Isotope-coded affinity tag reagents (ICATs) : .....	83
(b) Tag isobarique : .....	84
(c) 18O labelling:.....	87
(d) Dimethyl labelling:.....	88
b) Quantification absolue des protéines .....	89
(i) Principe et standards internes .....	91
(a) Principes généraux : .....	91
(b) Comparaison des performances :.....	96
(c) Problèmes spécifiques à la quantifications absolue :.....	96
(ii) Méthodes ciblées : SRM et PRM .....	98
(iii) Quantification des protéines par Top-down/intact .....	101
(a) Instrumentation :.....	101

(b)	Traitement du signal.....	103
(c)	Etalon interne en top-down/intacte .....	104
(iv)	Quantification des protéoformes.....	107
(a)	Modifications post-traductionnelles : phosphorylation et glycosylation.....	107
(b)	Quantification des isoformes : .....	111
(c)	Troncations : .....	112
(d)	Approche intacte et comparaison bottom-up.....	113
	Conclusion de l'étude bibliographique et objectifs de la thèse .....	118
	Chapitre 1 : Quantification de protéines impliquées dans la physiopathologie des DFT. Apport du bottom-up pour la quantification d'isoformes et le développement de méthodes multiplexées. ....	122
I.	Introduction : les démences lobaires frontotemporales.....	123
1.	Mutation du gène <i>C9ORF72</i> .....	123
2.	Mutation du gène <i>GRN/PGRN</i> .....	125
3.	Mutation du gène <i>MAPT</i> .....	126
II.	Matériels et méthodes .....	127
1.	Réactifs .....	127
2.	Méthodes d'extraction.....	127
a)	immunoprécipitation.....	127
b)	Lyse ultrason.....	128
c)	Délipidation .....	128
3.	Méthodes LC-MS/MS multiplexe .....	128
a)	UHPLC.....	128
b)	Chromatographie capillaire.....	128
c)	DIA .....	129
III.	Partie Expérimentale .....	130
1.	Développement d'une méthode d'immunoprécipitation pour la <i>C9ORF72</i> .....	130
2.	Article : quantification de la <i>C9ORF72</i> dans le cortex frontal .....	136
3.	Extension de la méthode pour la quantification de la <i>C9ORF72</i> dans les PBMCs .....	158
a)	Extraction.....	159
b)	Evaluation de la méthode.....	159
c)	Applications à des échantillons humains .....	161
4.	Développement d'une méthode de quantification multiplexe par PRM.....	165
a)	Liste des protéines.....	165
b)	Optimisation de la préparation d'échantillon : .....	169
c)	Evaluation de la méthode et comparaison UHPLC/ chromatographie capillaire.....	170
d)	Résultat de quantification relative des échantillons de cortex frontal de la cohorte.....	176

e) Développement d'une méthode DIA .....	180
IV. Discussion et perspectives .....	183
Chapitre 2 : Développement méthodologique, bottom-up et top-down, pour la quantification des protéoformes de l'alpha-synucléine .....	187
I. Introduction.....	187
II. Matériels et méthodes .....	194
1. Réactifs .....	194
2. Méthodes LC-MS/MS .....	194
3. Conditions d'incubation de l'alpha-synucléine par les protéases.....	194
III. Article : quantification multiplexe de tau et de l'alpha-synucléine dans le LCR .....	196
IV. Développement d'une méthode intact/top-down pour la caractérisation de l'alpha-synucléine.....	231
1. Optimisation des paramètres MS.....	231
2. Développement d'une méthode pour la quantification en LCR .....	232
3. Développement d'une méthode de quantifications des tronctions de l'alpha-synucléine dans le cerveau.....	235
a) Génération de standards marqués tronqués .....	235
(i) Sélection des protéases.....	235
(ii) Optimisation des conditions d'incubation .....	236
(iii) Identification des fragments .....	237
b) Evaluation du protocole pour la quantification des formes tronquées en cerveau (souris) et comparaison des modes d'acquisition ciblée pour la quantification (SIM/PRM).....	240
c) Application à une cohorte .....	244
d) Quantification des protéases dans les échantillons de cerveaux.....	248
V. Discussion et perspectives .....	252
Conclusion générale .....	259
Références.....	265
Annexe 1.....	288
Annexe 2.....	289



# Liste des figures

Figure 1 Gènes causant ou étant un facteur de risque de la maladie d'Alzheimer. D'après Yamazaki et al. 2016 .....	17
Figure 2 Etapes de découverte et validation des biomarqueurs selon une approche triangulaire (gauche) ou rectangulaire (droite). D'après Bekker-jensen 2017 .....	20
Figure 3 Définition du terme protéoforme avec les sources de modifications créant leur diversité. D'après Lloyd M Smith, Neil L Kelleher & The Consortium for Top Down Proteomics 2013 .....	22
Figure 4 illustration des modifications post-traductionnelles les plus fréquentes. Adapté de Pagel et al. 2015 .....	23
Figure 5 Production de deux isoformes d'une protéine par épissage alternative de l'Exon 2. Adapté de Romero et al. 2006 .....	26
Figure 6 Schéma de principe de phosphorylation d'une protéine. La première étape consiste en la fixation de l'ATP dans le site actif de la Kinase puis le substrat (la protéine) se fixe à son tour. Un groupement phosphate est alors transféré vers une sérine, thréonine ou tyrosine. Une fois se transfert réalisé le substrat puis l'ADP sont libérées. Adapté de Ubersax et al. 2007.....	28
Figure 7 Principales glycosylation. D'après Lodish et al. 200.....	29
Figure 8 Acétylation des protéines en N-terminale et des lysines. Le groupement acétyle provient de l'acétyle-COA, ce transfert est catalysé par des enzymes différentes pour l'acétylation en N-terminale et l'acétylation des lysines. L'acétylation des lysines est réversible alors que l'acétylation en N-terminale est supposé pour l'instant irréversible. D'après Ree et al. 2018 .....	30
Figure 9 illustration de l'effet de l'acétylation sur la structure de la chromatine et son effet sur la transcription. Adapté de Verdin & Ott 2015 .....	32
Figure 10 Structure des acylations qui sont présentes sur les lysines ainsi que leurs déacetylases. D'après Choudhary et al. 2015 .....	34
Figure 11 Influence du degré de méthylation d'une lysine sur l'hydrophobicité et leur capacité à former des liaisons hydrogènes. Adapté de Murn et al. 2017 .....	35
Figure 12 Schéma de principe de l'attachement d'une ubiquitine à une protéine. Premièrement l'enzyme E1 utilise l'ATP pour se lier à une ubiquitine puis l'enzyme E1 transfère cette ubiquitine vers l'une des 40 enzymes E2. Alors E2 par interactions avec l'E3 (RING, HECT ou RNR) transfère l'ubiquitine vers la protéine cible. L'enzyme DUB permet d'enlever ce marquage. D'après Rape et al. 2018.....	36
Figure 13 Les différents motifs d'ubiquitinylation avec leurs fonctions. D'après Rape et al. 2018 .....	37
Figure 14 Schéma de principe de la SUMOylation des lysines. La protéine SUMO est dans un premiers temps activé par la protéase SENP qui expose les acides aminés GG. Puis cette protéine activée se lie à l'enzyme E1 qui est un complexe protéique SAE1 et SAE2 par l'intermédiaire de l'ATP. Puis la protéine SUMO est transféré sur l'enzyme E2 qui par l'intermédiaire de E3 la lie à la protéine cible. D'après Hendricks et al. 2016 .....	38
Figure 15 Activation de la caspase par clivage de la pro-caspase en C-terminale de spécifique acide aspartique. D'après Lavrik et al. 2005 .....	41
Figure 16 Formation des peptides amyloïdes beta. D'après thèse Bros P. ....	42
Figure 17 Mécanisme de dégradation des protéines par le protéasome. Les protéines poly-ubiquitinylés sont ciblées par le protéasome. La dégradation de la protéine est se fait en premier à partir d'une de ses régions non-structurés. D'après Inobe et al. 2014.....	43
Figure 18 Interactions entre les différentes modifications post-traductionnelles des histones.. Ces interactions peuvent être positives (flèche) ou négatives flèche à tête plate. D'après Bannister et al. 2011.....	45

Figure 19 Représentation schématique du principe de la spectrométrie de masse .....	46
Figure 20 Approches d'identification et de caractérisation des protéines par spectrométrie de masse. Adapté de Chait et al. 2006 .....	49
Figure 21 Spectre de masse de protéines et de complexes protéiques analysées en intactes (Rose et al. 2012) ainsi que leur structure cristallographique. Spectre de masse d'un IgG (149 kDa) (a), capsidie pentamérique et hexamérique du phage HK97 (210 kDa et 253 kDa) (b), complexe protéique du protéasome 20S d'une levure (730 kDa) (c) et la protéine GroEL (800kDa) (d).....	50
Figure 22 Représentation schématique du principe de la technique ELISA (Mistry et al. 2014). (i) l'anticorps de capture est greffé à la surface d'une plaque, (ii) l'échantillon contenant l'antigène est ensuite appliqué dans les puits de la plaque puis (iii) l'anticorps de détection est ajouté. (iv) Enfin, un anticorps secondaire lié à une enzyme est ajouté et va se lier à l'anticorps de détection. Les enzymes couramment utilisées sont la peroxydase de raifort et la phosphatase alcaline. (v) La détection se fait par la conversion d'un substrat avec l'enzyme en signal fluorescent.....	51
Figure 23 ELISA sans interférences pour la thyroglobuline (gris) qui est capté par l'anticorps blanc et est détectée par l'anticorps à pois (A). En cas d'autoanticorps (noir) la thyroglobuline n'est plus captée par les anticorps de capture (blanc), ce qui résulte en une sous-estimation de la quantité de thyroglobuline (B). Adapté de Hoofnagle et al. 2009.....	53
Figure 24 Gamme dynamique des concentrations des protéines plasmatiques. En jaune, les biomarqueurs utilisés en clinique et en rouge les protéines identifiées par le projet HUPO plasma proteome project (PPP) d'après Surinova et al. 2011 .....	54
Figure 25 Stratégies de préparation pour la détection des peptides pour améliorer la sensibilité et la spécificité des méthodes par dilution isotopiques. L'ajout de peptides marqués permet d'augmenter la précision la mesure. La déplétion seule des 12 à 70 protéines majoritaires du plasma permet d'atteindre des limites de quantifications de l'ordre de 100ng/mL. L'ajout d'une étape de fractionnement permet d'atteindre des limites de quantifications proches du ng/mL voir même 50 pg/mL pour la PSA. Un autre moyen d'atteindre des limites de quantifications aux alentours du ng/mL est l'immunoprécipitation au niveau de la protéine ou des peptides (SISCAPA). D'après Gillette et al. 2013 .....	55
Figure 26 Fréquence normalisé des sites de coupure de la trypsine pour les 45 paires acide aminés les plus courantes. D'après Rodriguez et al. 2008 .....	62
Figure 27 Représentation des 10 instruments les plus utilisés dont les données sont déposées sur ProteomeXchange .....	63
Figure 28 (A) Caractéristique des spectromètres de la famille Q Exactive, le temps d'analyse de l'Orbitrap du Q Exactive et du Q Exactive plus est de 512 ms à une résolution de 140 000 et de 512 ms à 240 000 de résolution pour le Q Exactive HF. Le Q Exactive HF à résolution équivalente est donc 2 fois plus rapide. Ce qui explique un nombre d'évènement plus important. (B) représentation des évènements MS et MS/MS en analyse DDA pour les différents Q-exactive. D'après Kelstrup et al. 2014.....	64
Figure 29 Comparaison du DDA et DIA. En DDA la fenêtre d'isolement des ions est restreinte 2Da contrairement au DIA 20 Da. Les spectres de masse des ions fragments peuvent être extrait à un point donnée en DDA alors qu'en DIA ils peuvent être suivit temporellement . Adapté de Egertson et al. 2015. ....	67
Figure 30 Séquence de l'albumine bovine (BSA), les zones colorées correspondent aux peptides détectés avec différentes enzymes. Dans l'ordre par (avec la couverture de séquence obtenue) : ArgC: 18.3%, AspN: 38.1%, chymotrypsin: 57.8%, GluC: 61.9%, LysC: 70.8%, LysN: 60.8% and trypsin: 78.4%. D'après Giansanti et al. 2015.....	68
Figure 31 Analyse protéomique bottom-up d'un lysat cellulaire HeLa par digestion à la neprosine, trypsine et Lys-C. (A) Diagramme de Venn des protéines identifiées (B) diagramme de Venn de la	

couverture de séquence pour les 1251 protéines identifiées par les trois enzymes. D'après Schröder et al. 2017 .....	70
Figure 32 Présentation des approches protéomiques, BUP (bottom-up), eBUP (extended Bottom-up), MDP (middle-down) et TDP (top-down) avec les méthodes de séparation chromatographique, les méthodes de fragmentation, l'instrumentation ainsi que les logiciels de traitement adaptés. D'après Laskay 2013 .....	71
Figure 33 Le Top-down permet de caractériser sans ambiguïté les protéoformes d'une protéine et ils peuvent être corrélés avec les phénotypes d'une maladie. Par exemple, on peut observer une augmentation d'un protéoforme particulier comme une méthylation et une phosphorylation simultanée. En revanche, le bottom-up ne permet pas de mettre en évidence cette méthylation et phosphorylation simultanée. En effet, l'information est perdue au cours de la digestion. D'après Savaryn et al. 2013 .....	73
Figure 34 La couverture de séquence des protéines est essentielle pour éviter les mauvaises attributions sur la position d'une modification post-traductionnelle et donc des protéoformes. (i) dans le cas d'un seul protéoforme et avec une couverture de séquence de 100% il n'y a pas d'ambiguïté dans l'attribution du protéoforme H4K5acK8ac (ii) dans le cas de deux protéoformes H4K5acK8ac et H4K5acK12ac si la couverture de séquence n'est pas suffisante il y a une ambiguïté dans la caractérisation des protéoformes. (iii) Dans le cas de spectres provenant de plusieurs protéoformes malgré une couverture de séquence totale il y a une ambiguïté sur le nombre de protéoformes et l'acétylation de la lysine en K12. D'après Dang et al. 2014 .....	74
Figure 35 Les principales méthodes de quantifications utilisées en spectrométrie de masse .....	78
Figure 36 Principe des méthodes SILAC (a) et super SILAC (b). Les cellules marquées sont utilisées dans ce cas comme standard pour comparer des échantillons. (c,d) chaque colonne représente un nombre de protéines. En combinant 5 types cellulaires (super SILAC) la tumeur est mieux représentée en terme de diversités de protéines. De plus (e,f) le super SILAC permet l'identification de plus de protéines avec des CV inférieurs au SILAC. D'après Geiger et al. 2010 .....	81
Figure 37 isotopologues de la Lysine commercialement disponibles (gauche) et exemple de spectre de masse pour la quantification multiplexe utilisant des lysines isotopologues .....	82
Figure 38 Représentation du réactif ICAT. D'après Gygi et al. 1999 .....	83
Figure 39 Les réactifs TMT utilisés pour les analyses multiplexes. Les atomes portant un marquage isotopique sont indiqués par une astérisque. ....	85
Figure 40 illustration de la compaction du ratio mesuré .....	86
Figure 41 Représentation du réactif EASI-tag, les atomes portant un marquage isotopique sont indiqués par une astérisque(gauche). Comparaison du taux de fragmentation des tag EASI et TMT par HCD (droite). D'après Virreira Winter et al. 2018 .....	87
Figure 42 Schéma réactionnel d'insertion de H <sub>2</sub> <sup>18</sup> O lors de la digestion à la trypsine d'une protéine. D'après Yao et al. 2011 .....	87
Figure 43 Marquage DML 5 fois multiplexé par l'ajout de Carbone 13C et de deutérium. Représentation schématique de la comparaison de 5 échantillons. D'après Wu et al. 2014 .....	88
Figure 44 Représentation de la justesse et de la fidélité sous forme de cibles .....	90
Figure 45 Stratégie de quantification absolue des protéines par bottom-up. Les étapes d'ajouts des étalons internes sont représentées. D'après Brun et al. 2009 .....	91
Figure 46 Comparaison des différentes méthodes de quantification, pour la thyroglobuline, sans calibration externe, tSIL (peptide AQUA), CSIL (peptide AQUA comprenant 6 acides aminés en N et C-terminale) et SIL-rTg (protéine recombinante SIL). La droite rouge représente la valeur cible et les lignes en pointillés la valeur cible +/- 20%. Sans calibration externe la protéine recombinante est celle permettant une quantification la plus juste suivit de CSIL puis de tSIL. Les conditions de digestion	

(DOC, Urée et TFE) ont aussi une grande influence sur la quantification à partir des peptides. D'après Shufford et al. 2017 .....	95
Figure 47 Effet du tampon d'équilibration sur la quantification absolue d'une protéine. D'après Pritchard et al. 2014 .....	97
Figure 48 Comparaison des méthodes de quantification avec calibration externe. En utilisant une protéine recombinante en DOC tous les peptides sont quantifiés avec une précision <20%. D'après Shufford et al. 2017 .....	98
Figure 49 Représentation schématique des modes SRM (A) et PRM (B) D'après Peterson et al. 2012	99
Figure 50 Profil d'états de charges pour le lysozyme humain et un anticorps (gauche), l'augmentation de la résolution de 2 000 et 58 000 permet d'avoir une séparation isotopique(droite) .....	103
Figure 51 Quantification des formes non-phosphorylées et phosphorylées des protéines AHNAK, CAMK2D et EGFR en fonction des conditions de culture des cellules U87. La condition est : 10%FCS (sérum de vœu fœtal), no FCS (les cellules sont en culture sans FCS pendant 3h), 30 min 10% FCS ( No FCS puis incubation avec du sérum de vœu fœtal pendant 30min) et 15 min 10% FCS ( No FCS puis incubation avec du sérum de vœu fœtal pendant 15min). Le taux de phosphorylation est représenté par un point rouge. D'après Dekker et al. 2018 .....	108
Figure 52 Processus pour la détermination du taux de phosphorylation des peptides. Wu et al. (gauche) et Tsai et al. (droite). .....	110
Figure 53 Quantification absolue multiplexe de la protéine tau dans le LCR. D'après Barthélemy et al. 2016.....	112
Figure 54 Peptides détectés pour la TTR en fonction de l'enzyme utilisée. D'après Vila rico et al. 2015 .....	113
Figure 55 Comparaison des résultats de quantification par bottom-up HR-XIC et en intacte. (A) quantification des modifications en Cys-10 et en excluant la forme majoritaire S-cys (à droite). (B) pourcentage des formes wt et V30M. D'après Vila rico et al. 2015.....	114
Figure 56 (a) quantification par SDS-PAGE à 5 ratio différents 1:5, 1:3, 1:1 ;, 3:1 et 5:1 pour l'ENH2(droite) et la $\beta$ -caséine (gauche). (b) quantification obtenus par spectrométrie de masse et (c) corrélation entre les résultats obtenus par SDS-PAGE et spectrométrie de masse. D'après Wu et al. 2018.....	115
Figure 57 quantification de la troponine cardiaque 1 porcine. D'après Zhang et al. 2011.....	116
Figure 58 Principe de l'immunoprécipitation pour la concentration d'une protéine cible.....	131
Figure 59 Isoformes de la C9ORF72 C9-L (haut) et C9-S (bas). Les peptides protéotypique sélectionnés sont soulignés. L'acide aminé différenciant C9-L et C9-S est encadré en rouge.....	131
Figure 60 Comparaison de la linéarité et de la limite de détection en plasma 1/20 -ème entre précipitation au méthanol et immunoprécipitation (gamme de 1 ng/mL à 250 ng/mL).....	134
Figure 61 Optimisation de la lyse des PBMCs. PBS 10s (n=4), PBS 30s (n=1) et PBS + triton 0.1% 10s (n=4).....	159
Figure 62 Evaluation de la stabilité de la C9ORF72 après lyse des cellules sur glace (n=2) .....	160
Figure 63 Gamme de calibration obtenu en plasma dilué au 5 <sup>ème</sup> . La gamme est de 5 ng/mL à 5 $\mu$ g/mL/ La LOQ pour TEIALSKG et DSTGSFVLPFR est de 50 ng/mL. ....	161
Figure 64 Ratio de concentration de la C9-total/C9-L entre les différents groupes .....	162
Figure 66 XICs peptide unique C9-S ainsi que de ses correspondants phosphorylés .....	163
Figure 67 Spectre MS des ions fragments de la forme sans phosphorylation et des peptides phosphorylés. Exemple pour l'ion $\gamma$ 17 2+ .....	163
Figure 68 Spectre MS des ions fragments de la forme sans phosphorylation et des peptides phosphorylés. Exemple pour l'ion $\gamma$ 14 2+ .....	164
Figure 68 Exemples de gamme de calibration pour deux peptides LPFPIIDDR et VTLDGMNPR .....	172
Figure 69 Corrélation C9ORF72 avec la paraplegin.....	179

Figure 70 Coefficient de variations des précurseurs identifiés dans les QC .....	181
Figure 72 Mécanisme de formation des corps de Lewy et propagation de cellule à cellule. D'après Lashuel al. 2013.....	189
Figure 73 Principales modification post-traductionnelles de l'alpha-synucléine qui ont été observées dans les corps de Lewy ou in vitro. Les épitopes des anticorps anti-alpha-synucléine sont aussi présentés. D'après Schmid et al. 2013 .....	191
Figure 74 recoupement des principales maladies neurodégénératives où la protéine Tau, Abeta ou l'alpha-synucléine sont impliquées. D'après Moussaud et al. 2014 .....	192
Figure 75 Optimisation des conditions de source: icid, température des gaz aux. et température du capillaire de transfert .....	231
Figure 76 Impact de l'icid et du nombre de $\mu$ scans sur la qualité du spectre.....	232
Figure 77 Droites de calibration obtenues pour les trois modes d'acquisition :FullMS, target-SIM et PRM .....	234
Figure 78 Spectres de masse déconvolué représentant les troncations générées dans pour les conditions d'incubation optimale pour chaque protéase .....	237
Figure 79 Exemple d'attribution basé sur le temps de rétention, le spectre MS et le sptres MS/MS pour la forme 76-140. Les spectres de masses sont attribués avec une précision de masse <5ppm. Les principaux fragments ont été attribués. ....	238
Figure 80 illustration du décalage en temps de rétention pour la troncation 1-57.....	239
Figure 81 Spectres MS/MS de la forme 1-57 montrant une dominance du spectre par les ions b pouvant rendre l'interprétation erronée. En effet, l'extrémité C-terminale n'est alors pas couverte. ....	240
Figure 82 XIC des peptides appartenant aux protéases.....	249
Figure 83 comparaison des profils pour Tau (haut) et de l'alpha-synucléine (bas) .....	254
Figure 84 Corrélation des médianes obtenues par PRM/SIM.....	257

# Liste des tableaux

Tableau 1 liste non exhaustive de protéines qui sont utilisées comme biomarqueurs de cancer .....	18
Tableau 2 Spécificité des Nat. D'après Ree et al. 2018 .....	33
Tableau 3 Les caractéristiques des principaux analyseurs ( Adapté de Junot et al. 2014) .....	48
Tableau 4 Orthogonalité et capacité de pics pratique pour différents couple chromatographique. Adapté de Gilar et al. 2005.....	57
Tableau 5 Comparaison de différentes méthodes de préparations d'échantillons en termes de sensibilité et de fidélité pour la quantification de la PSA plasmatique. Adapté de Makawita et al. 2010.....	60
Tableau 6 Les 6 protéases les plus communes en bottom-up avec leurs spécificités ainsi que leurs avantages et inconvénients. D'après Giansanti et al 2015 .....	69
Tableau 7 Impact du nombre d'état de charge utilisé pour la quantification sur la justesse et la fidélité (Qc n=3). Adapté de Lanshoeft et al. 2017.....	104
Tableau 8 intensité obtenue pour les trois peptides en cellules HEK et en plasma. L'immunoprécipitation a été réalisé avec 0.4µg d'anticorps par échantillon et ont été greffés à des billes magnétiques M-280 tosylactivated .....	132
Tableau 9 Aires des peptides de la C9ORF72 détecté dans deux homogénats de cerveau humain. Les peptides communs sont indiqués en rouges .....	133
Tableau 10 Comparaison des aires obtenues par immunoprécipitation avec la précipitation au méthanol en cerveau de souris et humain.....	134
Tableau 11 Evaluation de la répétabilité de la méthode pour 5 analyses le même et sur 3 jours consécutifs de PBMCs .....	160
Tableau 12 Liste des échantillons de PBMCs analysés.....	161
Tableau 13 liste des peptides sélectionnés pour la quantification multiplexe (1).....	167
Tableau 14 liste des peptides sélectionnés pour la quantification multiplexe (2).....	168
Tableau 15 Comparaison des protocoles de préparation d'échantillon .....	170
Tableau 16 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (1) .....	173
Tableau 17 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (2) .....	174
Tableau 18 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (3) .....	175
Tableau 19 Liste des protéines impactées avec la médiane des ratios des aires des peptides pour chaque groupe. Le test statistique utilisé est le test de Mann-Withney, les médianes sont considérées comme statistiquement différentes si $p < 0.05$ . .....	177
Tableau 20 protéines détectées en DIA .....	182
Tableau 21 Performances analytiques des modes d'acquisition pour la quantification de l'alpha-synucléine en plasma dilué au 200ème .....	234
Tableau 22 Rendement de SPE et de précipitation pour les formes tronquées de l'alpha-synucléine .....	241
Tableau 23 Comparaison des performances de quantification entre les modes SIM et PRM. LOQ est déterminé avec un biais de <20% .....	243
Tableau 24 Quantification des troncations par PRM .....	246

Tableau 25 Quantification des tronctions par SIM.....	247
Tableau 26 Paramètres MS pour la détection par PRM des peptides des protéases.....	250
Tableau 27 Quantification des protéases dans les cerveaux .....	251
Tableau 28 Liste des peptides générés digestion des formes tronquées observées, par la Glu-C ou par la trypsine. L'acide aminé du site de coupure est en rouge. L'acide aminé suivant (troncation C-ter) et précédant (N-ter) est aussi indiqué. Comme illustré, tous les peptides n'auront pas une taille adapté à l'analyse LC-MS, et des coupures endogènes sont communes à la trypsine ou à la Glu-C.....	258

# INTRODUCTION GENERALE



Cette partie introductive ne se limitera pas aux seules maladies neurodégénératives mais cherchera à présenter de manière plus large les méthodologies proposées pour la quantification de protéines et de leurs protéoformes. Par exemple, du fait du peu de biomarqueurs protéiques validés pour les maladies neurodégénératives les exemples ne seront pas limités à ce domaine mais s'étendront aux cancers où les biomarqueurs sont couramment utilisés pour orienter une thérapie par exemple. Enfin, pour les protéoformes l'exemple le plus parlant est le code des histones car ce sont les protéoformes les mieux décrits.

## I. Les biomarqueurs

La thèse a porté sur la quantification de protéines et de leurs protéoformes par spectrométrie de masse. Ces composés ont un intérêt particulier dans le domaine des biomarqueurs, que ce soit pour le diagnostic, le pronostic ou le suivi thérapeutique. Afin de préciser le contexte de ce travail, nous commencerons par expliciter le terme biomarqueur.

### 1. Qu'est-ce qu'un biomarqueur ?

Le National Institute of Health (NIH, USA) définit un biomarqueur comme étant une caractéristique pouvant être objectivement mesuré comme un indicateur d'un processus biologique normal, pathologique ou comme une réponse pharmacologique induit par un traitement ou intervention thérapeutique<sup>1</sup>. Un biomarqueur peut être donc simple comme la température corporelle, le pouls et la pression sanguine ou des biomarqueurs plus complexes comme des métabolites ou protéines.

Ici, nous nous intéresserons aux biomarqueurs dit moléculaires qui sont l'ADN (Acide DésoxyriboNucléique), l'ARN (Acide RiboNucléique), les protéines et les métabolites. Les exemples choisis ne se limiteront pas aux maladies neurodégénératives avec notamment des exemples basés sur les cancers dont les biomarqueurs sont validés et couramment utilisés en clinique.

### 2. Les différents types de biomarqueurs moléculaires

#### a) Gènes et ARN

- Les mutations des gènes sont impliquées dans un certain nombre de cancers. La recherche de mutations dans un gène peut orienter la stratégie thérapeutique comme c'est le cas pour le cancer du poumon non à petites cellules<sup>2</sup>. En effet, aux Etats-Unis 15% des patients

atteints de ce cancer ont une mutation dans le gène de l'EGFR sur les exons 19 ou 21. Ces mutations ont pour conséquence une activation de l'EGFR. Une stratégie ciblée peut donc être employée, en comparaison avec la chimiothérapie, consistant en des inhibiteurs de tyrosine kinase ou avec un anticorps monoclonal comme le Cetuximab qui empêche la dimérisation du récepteur EGFR. Ce traitement par le Cetuximab est aussi utilisé dans le cas de cancer colorectal<sup>3</sup>. En revanche, les patients présentant une mutation dans le gène K-ras ne vont pas répondre au traitement par anti-EGFR avec le Cetuximab. En effet, les mutations du gène K-ras ont pour conséquence de rendre cette protéine active sans activation du récepteur EGFR. Dans le cas d'une mutation K-ras d'autres stratégies thérapeutiques doivent être sélectionnées.

- Les modifications épigénétiques : La méthylation des bases de l'ADN joue un rôle important dans le contrôle des gènes. La méthylation a souvent lieu aux sites de contrôles des gènes. L'hyperméthylation des sites CpG est souvent associé avec une désactivation de l'expression du gène alors qu'une perte de méthylation est associée avec sa réactivation. Ces modifications peuvent conduire au développement d'une tumeur comme c'est le cas par exemple pour l'hyperméthylation du gène suppresseur de tumeurs p21 WAF1/CIP ou p16INK4 $\alpha$ . Il existe 14 kits commerciaux permettant la mesure de la méthylation de l'ADN sur 13 gènes différents. Parmi ces 14 tests disponibles seulement deux sont approuvés par la FDA pour la détection d'un cancer colorectale avec la mesure de la méthylation sur les gènes NDGRG4 et BMP3 pour le premier et SEPT9 pour le second<sup>4</sup>.
- Les ARN peuvent aussi être utilisés comme biomarqueurs, notamment pour les cancers<sup>5</sup>. Par exemple, un panel de 31 ARN messagers permet de prédire l'évolution du cancer de la prostate<sup>6</sup>.
- Concernant les maladies neurodégénératives nous pouvons prendre l'exemple de maladie d'Alzheimer. Des mutations dans les gènes *PSEN1*, *PSEN2* et *APP* sont responsables des formes précoces familiales de la maladie (voir figure 1). D'autres mutations comme par exemple du gène *APOE* induisent un risque plus élevé de développer une forme tardive de la maladie d'Alzheimer. En effet, un allèle *APOE $\epsilon$ 4* augmente le risque de développer la maladie d'Alzheimer d'un facteur 3 et deux allèles d'un facteur 12. D'autres mutations des gènes sont aussi des facteurs de risques de la maladie d'Alzheimer<sup>7</sup> (voir figure 1).

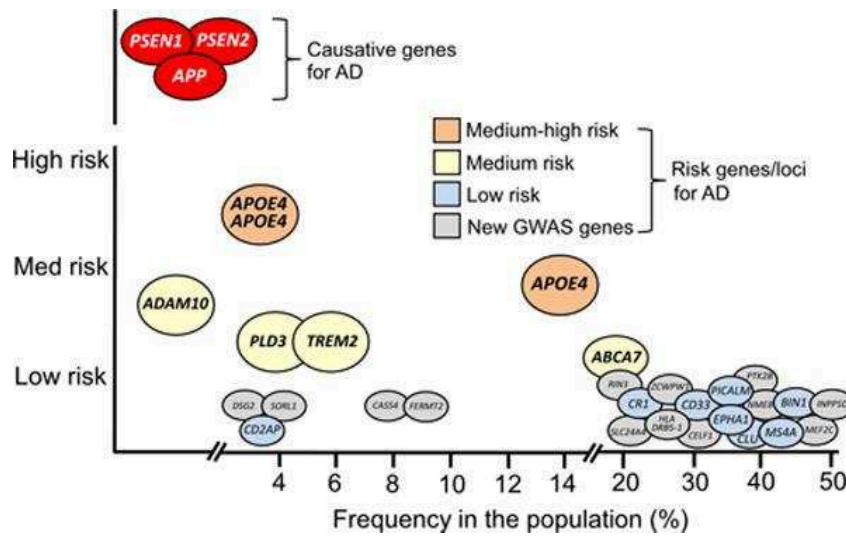


Figure 1 Gènes causant ou étant un facteur de risque de la maladie d'Alzheimer. D'après Yamazaki et al. 2016

## b) Protéines

Les protéines sont impliquées dans tous les processus cellulaires majeurs. Le protéome qui comprend l'ensemble des protéines d'un milieu biologique donné est composé de plus de 20 000 protéines. Ces protéines sont issues d'environ 20 000 gènes et la diversité des protéines provient par exemple de l'épissage alternatif ou des modifications post-traductionnelles qui seront décrites dans la partie I.3. L'étude de ces biomarqueurs se fait principalement selon trois approches : électrophorèse sur gel de polyacrylamide en deux dimensions, les tests immuno-enzymatique ELISA (enzyme-linked immunoassay sorbent) et les approches par spectrométrie de masse. Ces deux derniers seront décrits dans la partie II. Les protéines sont associées à un certain nombre de cancers<sup>8</sup> (voir tableau 1).

Tableau 1 liste non exhaustive de protéines qui sont utilisées comme biomarqueurs de cancer

Biomarqueur	Matrice biologique	Type de cancer	Utilisation clinique
$\alpha$ -Fetoprotein	Sérum	Tumeur non séminomateuse des testicules	Evaluation du stade du cancer
Human chorionic gonadotropin- $\beta$	Sérum	Testicules	Evaluation du stade du cancer
CA19-9	Sérum	Pancréas	Suivi
GA125	Sérum	Ovaires	Suivi
CEA	Sérum	Colon	Suivi
EGFR	Colon	Colon	Sélection de la thérapie
KIT	Tumeur gastrointestinal	GIST	Diagnostique et sélection de la thérapie
Thyroglobuline	Sérum	Thyroïde	Suivi
CA15-3	Sérum	Sein	Suivi
CA27-29	Sérum	Sein	Suivi
Cytokeratins	Tumeur du sein	Sein	Pronostic
Oestrogen receptor and progesterone receptor	Tumeur du sein	Sein	Sélection de la thérapie
HER2/NEU	Tumeur du sein	Sein	Pronostic et sélection de la thérapie
HER2/NEU	Serum	Sein	Suivi
NMP22	Urine	Vessie	Dépistage et suivi
Fibrin/FDP	Urine	Vessie	Suivi
BTA	Urine	Vessie	Suivi
High molecular weight CEA and mucin	Urine	Vessie	Suivi

Des biomarqueurs protéiques sont aussi couramment utilisés en clinique pour le diagnostic/confirmation de la maladie d'Alzheimer<sup>9</sup>. Avec la quantification dans le LCR de A $\beta$ -42, tau et p-tau dont les taux sont impactés.

### c) Métabolites

Le métabolome correspond à l'ensemble des métabolites contenues dans un milieu biologique donné. Les métabolites sont des composés organiques de faibles masses moléculaires issus des différents processus métaboliques. La métabolomique est l'analyse semi quantitative ou quantitative du métabolome. Les outils d'analyses de choix pour les analyses métabolomique sont la résonance magnétique nucléaire (RMN) et la spectrométrie de masse. Par exemple, Shaham et al.<sup>10</sup> ont mis en évidence des biomarqueurs métabolique dans le plasma pour une maladie liée à un dysfonctionnement de la chaîne respiratoire<sup>10</sup>. En effet, le lactate, l'alanine et la créatine sont significativement augmentés dans le plasma des patients en comparaison des contrôles. Ces biomarqueurs ont été mis en évidence par spectrométrie de masse. Les analyses métabolomiques ont aussi permis de mettre en évidence dans biomarqueurs pour les cancers<sup>11</sup> et pour les maladies neurodégénératives<sup>12</sup>.

### 3. Découverte et validation de nouveaux biomarqueurs

La découverte et la validation de nouveaux biomarqueurs se fait en trois étapes principales (voir figure 2).

- Phase de découverte : La phase découverte est la première étape, elle permet d'identifier les protéines candidats biomarqueurs sans a priori. Les approches non ciblées par spectrométrie de masse et l'électrophorèse bidimensionnelle sont des outils de choix pour obtenir l'empreinte protéique d'un échantillon. Cette phase dans une approche dite « triangulaire » comprend quelque dizaine d'échantillons et permet l'identification de 100 à plus de 10 000 protéines<sup>13</sup>. La seconde approche qui s'est développée récemment du fait des avancées technologiques (spectromètres de masse plus sensibles et plus rapides, automatisation de la préparation d'échantillon) est l'approche dite « rectangulaire » dont les cohortes de découverte et de validation ont la même taille d'une centaine à plusieurs milliers d'échantillons. Cette approche permet d'augmenter la puissance statistique des données de découvertes ainsi que de mettre en évidence des profils protéiques.
- Phase de vérification : Une méthode d'analyse ciblée par spectrométrie de masse de type MRM est mise en place. Seulement un faible nombre de candidates est retenu, généralement inférieur à 10 protéines. En revanche, le nombre d'échantillons est plus

important : plusieurs dizaines à plusieurs centaines d'échantillon sont analysés. Ces échantillons sont de préférence indépendants des échantillons analysés pendant la phase de découverte.

- La validation des biomarqueurs se fait sur une cohorte encore plus importante de plusieurs centaines à plusieurs milliers d'échantillons. La validation se fait soit par des méthodes immuno-enzymatique ou bien par spectrométrie de masse du fait de certaines limitation des méthodes immuno-enzymatique qui seront abordées dans la partie II. Les approches rectangulaires utilisent uniquement la spectrométrie de masse. Les candidats sélectionnés doivent répondre à des critères de spécificité de sensibilité. La spécificité correspond à la probabilité d'avoir un test négatif pour un contrôle et la sensibilité à avoir un test positif pour un patient atteint de la maladie recherchée. L'utilisation d'un biomarqueur unique commence à être remplacée par un profil protéique comprenant plusieurs protéines afin d'augmenter ces deux critères et de mieux décrire des pathologies complexes.

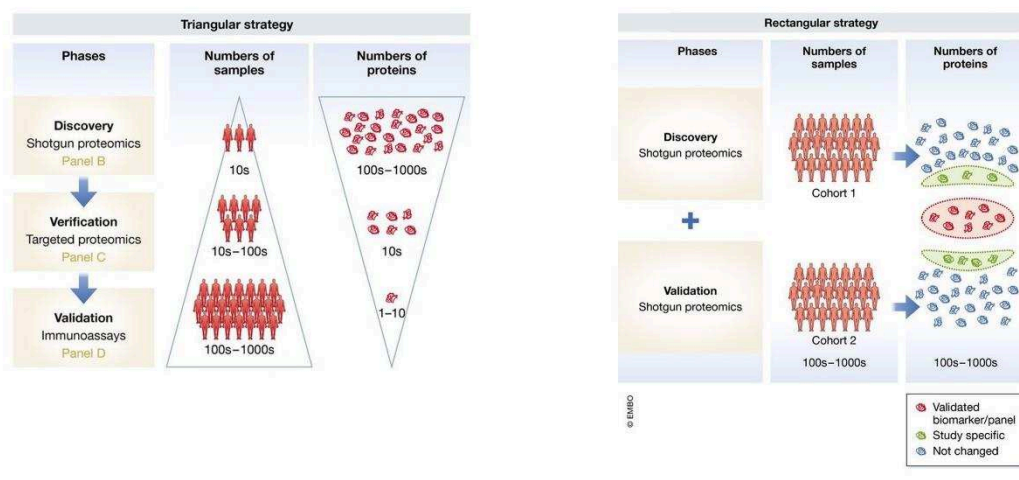


Figure 2 Etapes de découverte et validation des biomarqueurs selon une approche triangulaire (gauche) ou rectangulaire (droite). D'après Bekker-jensen 2017

Les biomarqueurs moléculaires sont de nature variée ADN, protéines et métabolites. La quantification des biomarqueurs se fait dans des matrices biologiques variées et le choix de la matrice est une étape primordiale pour les étapes de découverte et de validation des biomarqueurs. Le nombre de nouveaux tests approuvé par la FDA est de moins de 2 par an<sup>14</sup>. Ce qui montre la difficulté d'amener à candidat biomarqueur à un biomarqueur validé.

Des méthodes d'analyses robustes, précises et sensibles doivent donc être développés afin de caractériser et quantifier ces biomarqueurs (cf II.4).

Pour la suite nous nous intéresserons uniquement aux biomarqueurs protéiques et plus particulièrement aux protéoformes qui permettent de décrire plus précisément des phénotypes cliniques complexes.

#### **4. Protéines et protéoformes**

Chez l'humain le nombre de gènes codant pour des protéines est compris entre 19 587 et 20 245 soit sensiblement le même nombre que pour la drosophile et le *Caenorhabditis elegans*<sup>15</sup> et loin des 100 000 gènes initialement estimés. Ce qui implique que contrairement à ce qui a été postulé, la majorité de la complexité du protéome, dont la taille estimée est comprise entre 20 000 à plusieurs millions, ne provient pas d'un grand nombre de gènes mais plutôt de modifications au niveau des protéines. Cette complexité peut être créée au niveau de l'ADN, ARN ou directement au niveau des protéines et est représentée par un terme, introduit par Smith et al. en 2013<sup>16</sup>, englobant toutes les modifications possible d'une protéine : une protéoforme.

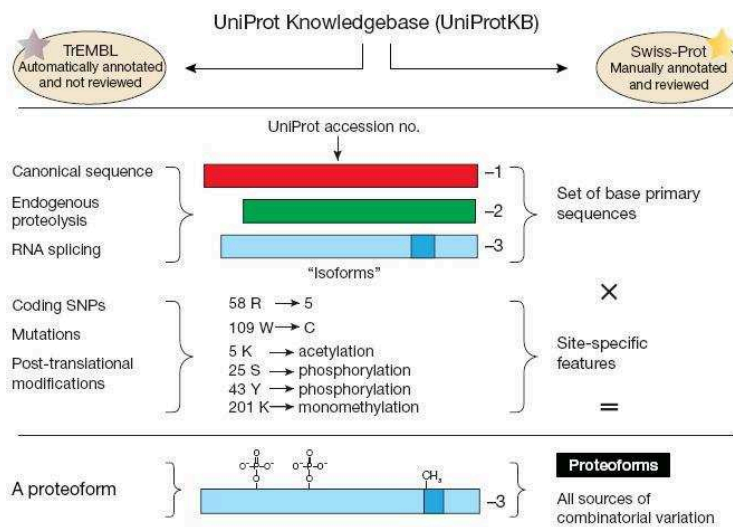
##### **a) La notion de protéoformes**

Le terme protéoforme est centré sur les gènes c'est-à-dire que les protéoformes d'une protéine proviennent tous d'un seul gène, en accord avec la base de données de séquences de protéines UniProt Knowledgebase.

Les modifications peuvent être faites à trois niveaux :

- ADN : variations alléliques : polymorphisme nucléotide, mutations de la séquence nucléotidique
- ARN : épissage alternatif pour 93% des gènes humains.
- protéines : modifications post-traductionnelles, protéolyse endogène

La combinaison de ces modifications représente les protéoformes d'une protéine comme illustré figure 3 :



**Figure 3 Définition du terme protéoforme avec les sources de modifications créant leur diversité.**  
 D'après Lloyd M Smith, Neil L Kelleher & The Consortium for Top Down Proteomics 2013

Plus de 33 000 isoformes issus d'épissage alternatif, 78 000 mutations d'un acide aminé et environ 53 000 modifications post-traductionnelles sont répertoriés dans la base de données humaine Swissprot. Ce qui représente un grand nombre de protéoforme théoriquement observable.

Les différents types de modifications post-traductionnelles majeures seront abordés dans le paragraphe suivant (voir figure 4).

### **b) Les types de protéoformes et leurs pertinences cliniques**

Les modifications post-traductionnelles permettent de répondre à a des besoins spécifiques des cellules et d'agir de façon réversible ou non réversible sur un processus. Bien que toutes les fonctions des mutations, épissages alternatif et modifications post-traductionnelles ne soit pas connues.



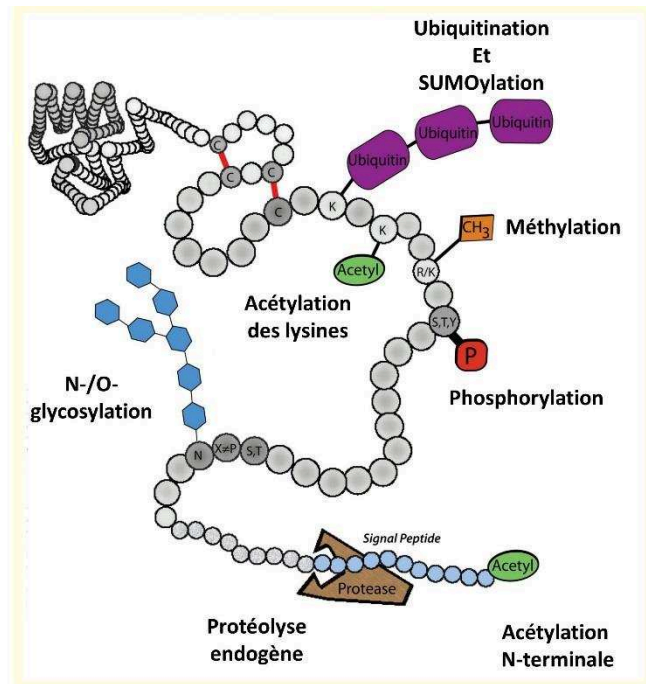


Figure 4 illustration des modifications post-traductionnelles les plus fréquentes. Adapté de Pagel et al. 2015

Les modifications post-traductionnelles peuvent être classés en deux groupes en fonction de leurs complexités structurales<sup>17</sup> :

- Simples : Phosphorylation, acétylation, méthylation, O-GlcNAc
- Complexes : Poylubiuitination, polySUMOylation, glycosylation

Les protéoformes ajoutent un degré supplémentaire d'analyse en combinant la complexité des modifications post-traductionnelles. Par exemple, pour l'histone H4 en ne prenant seulement en compte les 13 modifications post-traductionnelles les plus fréquentes le nombre théorique de protéoformes est de 98 304 or seulement 75 protéoformes ont été observés<sup>18</sup>. Le nombre théorique de protéoformes théoriques peut être calculé par :

$$nb \text{ protéoformes théo.} = \prod_{i=1}^n (\text{modification post traductionnelle possible à l'AA}_i + 1)$$

Ce qui donne :  $2^6$  (K5/8/12/16/31/91 acétylation)\* $3^1$  (R3 simple et double méthylation)\* $4^1$  (K20 simple, double et triple méthylation)\* $2^5$  (S1/S46/Y51/T79/Y87 phosphorylation)\* $2$  (N-terminal acétylation)\* $2$  (E63Q mutation)= 98 304 protéoformes<sup>19</sup>.

Ce qui pose la question de la diversité et du nombre réel des protéoformes ainsi que de la complexité de l'analyse de toutes ces protéoformes. Ponomarenko et al.<sup>20</sup>, l'ont évalué à 6,13 millions mais il est probable que ce chiffre soit réévalué dans les années à venir.

Dans cette section nous allons ici décrire les principales causes de diversité des protéines qui sont les mutations au niveau de l'ADN, l'épissage alternatif ainsi que les modifications post-traductionnelles les plus fréquentes.

### **(i) ADN : polymorphisme nucléotidique et mutations**

Le nombre de polymorphismes nucléotidiques et de mutations non synonymes présent dans les régions codantes de l'ADN, c'est-à-dire produisant une protéine dont la séquence en acides aminés diffère d'un acide aminé, est actuellement d'environ 135 000 validée sur SwissProt chez l'humain. Une simple substitution d'un acide aminé dans une protéine peut engendrer une perte d'association dans un complexe protéique ou favoriser son agrégation. Ces mutations sont directement liées à un nombre important de maladie comme par exemple la drépanocytose<sup>21</sup> (6 E->V), l'Amylose, la maladie de Parkinson et Alzheimer.

- Pour la drépanocytose, les deux allèles doivent être mutés, c'est la substitution en position 6 d'un acide aminé polaire chargé négativement (acide glutamique) par un acide aminé apolaire (valine) dans l'hémoglobine- $\beta$  qui induit une déformation des globules rouges et donc une incapacité à fixer le dioxygène ainsi qu'une durée de vie plus courte des cellules<sup>21</sup>.
- L'amylose est causée par la mutation du gène de la transthyrétine. Plus de 120 simples ou doubles mutations ont été reportés sur le gène de la transthyrétine dont la plus fréquente conduit à la substitution d'une valine par une méthionine en position 30 de la protéine. La transthyrétine est une peptide protéine de 127 acides aminés qui est présente sous forme tétramérique. La mutation en position 30 induit la dissociation du tétramère et la formation de fibrilles amyloïdes de transthyrétine dans divers organes et tissus<sup>22</sup>.
- Concernant les maladies neurodégénératives, notamment les maladies de Parkinson et d'Alzheimer, les mutations jouent un rôle prépondérant dans les cas familiaux. Dans les cas familiaux de la maladie d'Alzheimer, une mutation, par exemple, dans le gène PSEN1<sup>23</sup> conduit à une substitution d'un acide glutamique en valine en position 280 (acide aminé polaire chargé négativement -> acide aminé apolaire) et à l'apparition d'une forme précoce de cette maladie. La presenilin-1 fait partie d'un complexe protéique appelé

gamma sécrétase responsable entre autres du clivage de l'APP en amyloïde beta. Cette mutation semble augmentée la concentration d'une des formes toxiques de l'amyloïde beta : 1-42. Concernant la maladie de Parkinson, des mutations dans le gène SNCA codant pour la protéine alpha-synucléine ont été identifiées. Notamment, A30P, E46K, H50Q, G51D et A53T<sup>24</sup>. L'alpha-synucléine est une protéine de 140 acides aminés qui pathologiquement tend à s'agréger dans le cerveau pour former des agrégats insolubles appelés corps de Lewy. Flagmeier et al.<sup>24</sup>, ont ainsi mis en évidence que ces mutations accélèrent certaines étapes clés de l'agrégation de l'alpha-synucléine notamment l'initiation de la production de fibrilles induit par des vésicules lipidiques ainsi que leurs proliférations. Toutes ces mutations ont lieu dans la partie N-terminale amphipathique (hydrophile/hydrophobe) de l'alpha-synucléine qui est liée à la membrane lipidique.

Une autre catégorie de mutation est la présence d'une expansion variable d'une séquence nucléotidique répétée dans un gène<sup>25</sup>. Cette répétition est le plus couramment un trinuécléotide mais peut aussi être jusqu'à un dodécanucléotide. Ces mutations sont associées à plusieurs maladies dont la maladie de Huntington, les démences frontotemporales, la sclérose latérale amyotrophique, dystrophie myotonique de Steinert de type 1 et 2. La taille des expansions a généralement une influence sur le phénotype de la maladie. De plus, ces répétitions sont normalement présentes parmi la population mais une taille limite a pour conséquence d'induire une maladie. Par exemple, pour l'ataxie spinocérébelleuse de type 3 un nombre de répétition du trinuécléotide CAG de 12 à 44 est considéré comme normal alors qu'un nombre supérieur à 50 est associé à la maladie. De plus, le nombre d'expansion est associé à des sous-types de la maladie : athénie, ataxie, dystonie. Ces expansions peuvent induire une modification structurale de la protéine en ajoutant une queue poly-Q à l'Huntingtine par exemple. Cette chaîne est causée par une répétition (CAG), elle-même associée à 9 maladies. Ces expansions peuvent aussi ne pas induire de modifications structurales à la protéine comme pour les DFT avec une expansion (G<sub>4</sub>C<sub>2</sub>) du gène *C9ORF72*. Celles-ci seront dans le chapitre 1. En revanche dans le cas de cette expansion ainsi que pour le syndrome de l'X fragile, l'expansion est traduite, ce qui produit des protéines ayant tendance à agréger ainsi que des foci d'ARN<sup>25</sup>.

## (ii) ARN : épissage alternatif

Le pré-ARNm des eucaryotes est constitué de régions codantes, les exons et de régions non codantes les introns. L'épissage correspond au processus au cours duquel les introns sont enlevés pour former un ARNm mature constitué d'exons qui va servir à la production de protéines. L'épissage peut aussi enlever des exons ou inclure des introns (rétention d'intron), c'est ce qui s'appelle l'épissage alternatif<sup>26</sup> (voir figure 5).

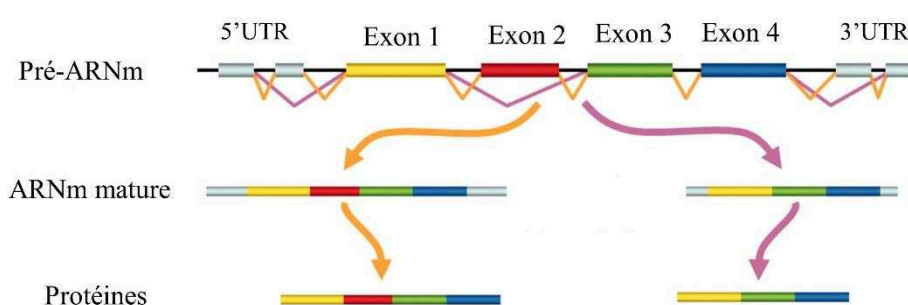


Figure 5 Production de deux isoformes d'une protéine par épissage alternative de l'Exon 2. Adapté de Romero et al. 2006

L'épissage alternatif permet d'ajouter un niveau de contrôle supplémentaire dans l'expression des gènes et d'augmenter la diversité des protéines. En effet, l'épissage alternatif concerne 92 à 94% des gènes, mais seulement 86% des gènes produisent une isoforme à une fréquence supérieure à 15% c'est-à-dire à un niveau d'expression non négligeable<sup>27</sup>. L'ARN messager et les protéines produits par épissage alternative peuvent avoir une structure, des fonctions et une localisation différente. Par exemple, chez la drosophile l'épissage alternatif permet de générer plus de 19 000 protéines de surface ayant des domaines extracellulaires différents et possédant donc des propriétés de liaisons différentes. L'épissage alternatif joue, dans ce cas, un grand rôle dans la reconnaissance cellulaire. L'épissage alternatif en changeant la structure de la protéine, peut en changer sa fonction comme c'est le cas pour la protéine Bcl-x qui est impliquée dans l'apoptose. Deux protéoformes sont produits : une forme de la protéine dite longue Bcl-xL comprenant les 4 domaines BH (Bcl-2 homology) et une forme de la protéine plus courte Bcl-xS ne comprenant que 2 domaines BH<sup>28</sup>. Alors que Bcl-xL est impliquée dans l'inhibition de l'apoptose, Bcl-xS au contraire inhibe la fonction de Bcl-xL. Ceci a pour conséquence une expression dépendante du tissu ou type cellulaire de l'ARN messager de la courte et de la longue. La courte est présente dans les cellules qui sont renouvelées rapidement et n'est que très peu présente dans les tissus alors que la longue est présente dans les cellules ayant une durée de vie plus longue comme celles du cerveau par exemple<sup>29</sup>. Une seconde conséquence d'un changement de structure des protéines

est un changement de localisation de la protéine comme c'est le cas pour la protéine GLI1 qui en perdant sa partie N-terminale par épissage alternatif passe du noyau au cytoplasme<sup>30</sup>. De plus, Yang et al.<sup>31</sup>, ont montré que les produits d'épissages alternatifs n'interagissent majoritairement pas avec les mêmes protéines et donc agissent comme s'ils étaient issus de gènes distincts. L'épissage alternatif est donc un processus bien régulé et les modifications de régulation de l'épissage alternatif conduisent à l'apparition de maladies<sup>32</sup>. Comme illustré chapitre I.DLFT, 53 mutations du gène MAPT ont été identifiées comme étant responsable du développement de DFT. La protéine TAU issue de ce gène est constituée de 6 isoformes et une dérégulation de la proportion de ces isoformes un serait responsable du processus pathologique.

### ***(iii) Modifications post-traductionnelles, protéolyse endogène***

Nous nous limiterons ici à la description des modifications post-traductionnelles les plus fréquentes, illustré par quelques exemples de pathologies liées à une modification de leur régulation.

#### **(a) Phosphorylation**

La première protéine phosphorylée enzymatiquement a été mise en évidence dès 1954 par Burnett et Kennedy. La phosphorylation est une modification post-traductionnelle réversible qui se caractérise par l'ajout d'un groupement phosphate ( $\text{PO}_3^{2-}$ ) principalement sur les sérines, thréonines et dans une moindre mesure les tyrosines. Le groupement phosphate possède à pH physiologique deux charges négatives. La phosphorylation est la modification post-traductionnelle la plus répandue, en effet il est estimé qu'au moins 30% des protéines seraient phosphorylées<sup>33</sup>.

La phosphorylation est un processus dynamique. En effet, les protéines sont phosphorylées par l'intermédiaire des kinases et sont déphosphorylées par l'intermédiaire des phosphatases. Les kinases peuvent être spécifiques comme la MEK (MAPK (Mitogen-Activated Protein Kinase)/ERK(Extracellular signal-Regulated Kinases) Kinase) qui assure la phosphorylation de deux sites sur ERK1 et ERK2 alors que les kinases CaM et CDKs permettent la phosphorylation de plus de 100 sites. Le génome code pour environ 500 kinases (400 serine et thréonine, 90 tyrosine et 40 serine/thréonine et tyrosine) et 200 phosphatases (40 Serine et thréonine, 100 tyrosine et 50 serine/thréonine et tyrosine). La phosphorylation se fait par le transfert d'un groupement phosphate de l'ATP qui est alors convertit en ADP<sup>33</sup> (voir figure 6).

Les kinases sont structurellement identiques : 250 acides aminés permettent la liaison avec l'ATP, l'adénosine est dans une poche hydrophobe formée par un lobe N-terminal en feuillet beta et un lobe C-terminal en hélice alpha. Ces deux lobes sont reliés par une boucle peptidique qui forme le site catalytique dont l'activation est induite par sa phosphorylation.

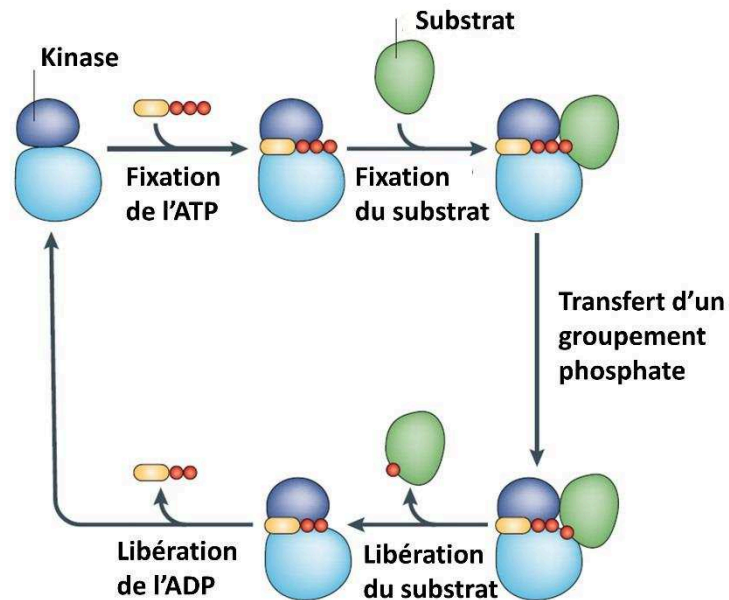


Figure 6 Schéma de principe de phosphorylation d'une protéine. La première étape consiste en la fixation de l'ATP dans le site actif de la Kinase puis le substrat (la protéine) se fixe à son tour. Un groupement phosphate est alors transféré vers une sérine, thréonine ou tyrosine. Une fois se transfert réalisé le substrat puis l'ADP sont libérés. Adapté de Ubersax et al. 2007

Le mode d'action pour les sérines et thréonines est identique. En revanche il est différent pour les tyrosines. En effet, la tyrosine a besoin d'un puit catalytique plus profond du fait de sa taille. Bien qu'il existe un recouvrement notamment dans le sens où les kinases des Ser/Thr peuvent phosphoryler les Tyr<sup>34</sup>.

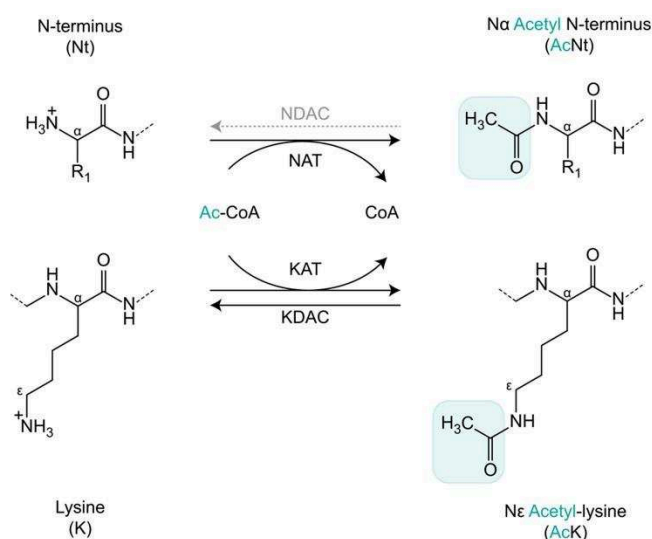
La phosphorylation est un évènement dynamique comme montré par kleiman et al. pour le récepteur de l'EGF<sup>35</sup>. En effet la demi-vie de la phosphorylation de ce récepteur est de 30 min mais quand la kinase responsable de sa phosphorylation est inhibée, le récepteur est rapidement déphosphorylé avec une demi-vie comprise entre 10 et 30s. Ce qui suppose que cet évènement de phosphorylation/déphosphorylation a lieu plus d'une centaine de fois au cours de son cycle. La phosphorylation des protéines est impliquées dans de nombreux processus cellulaire comme par exemple, la différenciation des cellules souches<sup>36</sup>, la régulation du cycle cellulaire<sup>37</sup> et l'apoptose<sup>38</sup>.



La chaîne glucidique est impliquée dans plusieurs mécanismes : interactions protéine-protéines, assure le repliement des protéines et sert de contrôle qualité<sup>42</sup>. De plus, la O-glycosylation semble aussi jouer un rôle dans la prévention de l'agrégation de l'alpha-synucléine<sup>43</sup>. Plusieurs biomarqueurs des cancers sont des glycoprotéines<sup>44</sup> comme par exemple la PSA, CA19-9 et la CEA (voir tableau 1).

### (c) Acétylation

L'acétylation consiste en l'ajout d'un groupement CO-CH<sub>3</sub> sur l'amine N-terminale d'une protéine ou bien sur l'amine de la chaîne latérale des lysines (voir figure 8). Le groupement acétyle neutralise alors la charge présente sur ce groupement. L'acétylation des protéines en N-terminale et des lysines est un processus distinct n'impliquant pas les mêmes enzymes<sup>45</sup>.



**Figure 8 Acétylation des protéines en N-terminale et des lysines. Le groupement acétyle provient de l'acétyle-COA, ce transfert est catalysé par des enzymes différentes pour l'acétylation en N-terminale et l'acétylation des lysines. L'acétylation des lysines est réversible alors que l'acétylation en N-terminale est supposée pour l'instant irréversible. D'après Ree et al. 2018**



Concernant l'acétylation des lysines, trois classes de molécules sont impliquées : les acétyltransférases qui catalysent le transfert d'un groupement acétyle vers une protéine, les désacétylases qui catalysent l'enlèvement du groupement acétyle et enfin des protéines se liant aux protéines acétylées qui permettent la transduction et la traduction du signal. L'acétylation des protéines a été étudié pour la première fois par Allfrey et al.<sup>46</sup>, en 1964. Allfrey et al., ont observés que les histones étaient acétylées et par ajout d'acétate radiomarqué (acétate-2-<sup>14</sup>C) ils ont trouvé que l'incorporation des groupements acétyles dans les histones se faisaient post-traditionnellement. Ils ont alors émis l'hypothèse que l'activation et la répression de la synthèse de l'ARN étaient gouvernés par l'acétylation des histones. Ce n'est qu'en 1995 que la première histone acétyltransférase a été découverte<sup>47</sup>. Les désacétylases sont divisées en deux familles : (HDAC-11) les désacétylases dépendantes du Zn<sup>2+</sup> présente uniquement dans le noyau et cytoplasme et les désacétylases SIRT1-7 dépendantes du métabolite NAD<sup>+</sup> localisées dans le noyau, cytoplasme et la mitochondrie. Les protéines reconnaissant les lysines acétylées contiennent un domaine spécifique appelé bromodomaine comprenant 110 acides aminés. Ces protéines reconnaissent donc les acétylations sur les lysines notamment en se liant à la queue des histones et ont un rôle dans l'activation de la transcription<sup>48,49</sup>. Le génome humain prévoit qu'il y a 46 protéines comprenant un bromodomaine.

Le niveau d'acétylation est fonction de la concentration en acétyl-CoA. En effet une baisse des nutriments entraîne une hausse de NAD<sup>+</sup> et une augmentation de l'activité catalytique des sirtuines, qui sont des désacétylases, et donc de la désacétylation. Au contraire une augmentation des nutriments induit une hausse d'acétyl-CoA et une augmentation de l'activité des acétyltransférases et donc une augmentation de l'acétylation<sup>50</sup>. L'acétylation des protéines est donc liée au métabolisme de la mitochondrie. L'acétylation est la modification post-traductionnelle majoritaire dans la mitochondrie. L'acétylation des lysines a lieu sur les parties structurées de la protéine : hélices  $\alpha$  et feuillets  $\beta$  ainsi que sur les protéines très conservées.

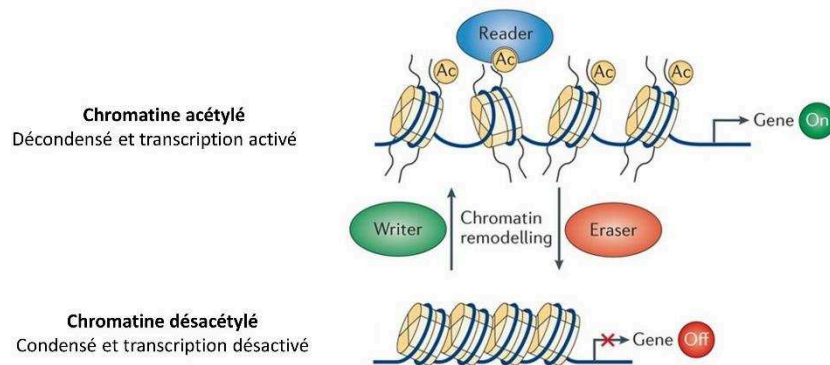


Figure 9 illustration de l'effet de l'acétylation sur la structure de la chromatine et son effet sur la transcription. Adapté de Verdin & Ott 2015

L'acétylation des lysines a de nombreuses fonctions notamment dans la compaction et l'activation de la chromatine<sup>51</sup> (voir figure 9), le changement de localisation des protéines ainsi que dans les interactions protéines-protéines. Dans une étude réalisée par le groupe de Mann, plus de 1750 protéines acétylées correspondant à 3 600 lysines acétylées ont été identifiées par spectrométrie de masse<sup>52</sup>.

Contrairement à l'acétylation des lysines, l'acétylation en N-terminale semble être un processus non-réversible. En effet, aucune enzyme permettant la désacétylation de l'amine N-terminale n'a été encore découverte. Il est estimé qu'environ 80% des protéines seraient acétylées en N-terminale<sup>45</sup>. Six enzymes permettent l'ajout de ce groupement : NatA, NatB, NatC, NatD, NatE et NatF. Les trois plus courantes étant NatA, NatB et NatC. La sélectivité de ces enzymes est déterminée par les deux premiers acides aminés d'une protéine.

La NatA : acétyle les acides aminés (A, S, T, C, V et G) des protéines dont la méthionine a été enlevée (cf. protéase). Contrairement aux 5 autres enzymes. La spécificité des enzymes est illustrée dans le tableau 2.

Tableau 2 Spécificité des Nat. D'après Ree et al. 2018

	NatA	NatB	NatC	NatD	NatE	NatF (membrane de l'appareil de Golgi)
Spécificité	A	MD	ML	Histones H2A et H4	MK	MK
	S	MN	MI		MV	MS
	T	ME	MF		MA	MV
	C	MQ	MW		MY	ML
	V		MV		MF	MQ
	G		MM		ML	MI
			MH		MS	MY
			MK		MT	MT

L'acétylation des protéines en N-terminale joue plusieurs rôles opposés en effet, elle peut protéger une protéine de la polyubiquitination et donc prolonger sa demi vie ou au contraire favoriser le recrutement de l'ubiquitine. L'acétylation peut influencer les interactions protéine-protéine, permettre le bon repliement d'une protéine et conduire les protéines à la membrane lipidique (directement ou indirectement). L'acétylation N-terminale est impliquée dans plusieurs maladies comme les cancers<sup>53</sup> et des troubles du développement<sup>45</sup> comme par exemple le syndrome d'Ogden<sup>54</sup>.

#### (d) Acylation

L'acétylation est un cas particulier d'acylation qui correspond à l'ajout d'un groupement acyle provenant d'un Acyl-CoA sur une lysine<sup>50,55</sup>. Cette partie traitera des autres acylations plus récemment identifiés comme : Formylation, succinylation, malonylation, butyrylation, propionylation, glutarylation et crotonylation (voir figure 10). L'acylation est une modification post-traductionnelle qui est supposée réversible et entraînant la neutralisation de la charge initialement présente sur la lysine. Ces modifications ont été pour le moment principalement trouvées sur les histones et les enzymes métaboliques de la mitochondrie. Les enzymes spécifiques de ces acylations et des déacylations non pas encore été toute identifiées. Des sirtuines comme par exemple la SIRT5 permet la déacylation des groupements acyles chargés négativement et la SIRT6 semble être spécifique de la myristoylation.

Les groupements acyle ont les propriétés suivantes :

- Propionylation et butyrylation: groupements hydrophobes, probablement comparable à l'acétylation en terme de fonctions.
- Crotonylation : groupement hydrophobe possédant une double liaison carbone-carbone, ce qui lui confère une organisation spatiale atypique planaire.
- Malonylation, succinylation et glutarylation : groupements acides possédant une charge négative (-1).

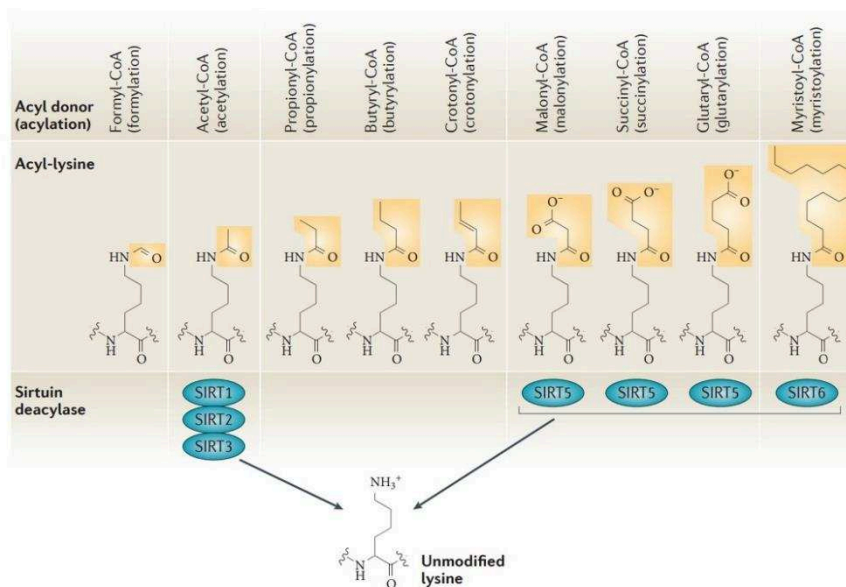


Figure 10 Structure des acylations qui sont présentes sur les lysines ainsi que leurs déacétylases. D'après Choudhary et al. 2015

### (e) Méthylation

La méthylation des protéines consiste en l'ajout d'un ou plusieurs groupement méthyle (CH<sub>3</sub>) sur les acides aminés suivants : lysine, arginine, acide aspartique, acide glutamique, méthionine, cystéine, histidine, glutamine et asparagine. Bien que la méthylation puisse se produire sur ces 9 acides aminés, la lysine et l'arginine sont les principaux acides aminés concernés et seront ceux abordés. Comme illustré par la figure 11, la lysine peut être mono-, di- ou tri-méthylés et l'arginine mono- ou-di-méthylés soit symétriquement ou asymétriquement. Contrairement à l'acétylation la charge de l'acide aminé est conservée. L'ajout des groupements méthyls augmentent leurs hydrophobicités tout en diminuant leurs capacités à former des liaisons hydrogènes<sup>56</sup> (voir figure

11). Des enzymes spécifiques sont responsables de ces méthylations (mono, di et tri). Par exemple plus de 10 enzymes transférant spécifiquement, bien qu'il y ait des recouvrements, un groupement méthyle ont été identifiées pour l'arginine chez les mammifères<sup>57</sup>. Cette modification post-traductionnelle est dynamique bien que plus lente que la phosphorylation et que l'acétylation. En effet, la moyenne des demi-vies des méthylations sur l'histone H3, par exemple, est de 1,3 jours<sup>58</sup> alors que la phosphorylation et l'acétylation sont de l'ordre de plusieurs minutes à quelques heures. La méthylation des protéines permet de moduler de nombreuses fonctions biologiques différentes<sup>59,60</sup>.

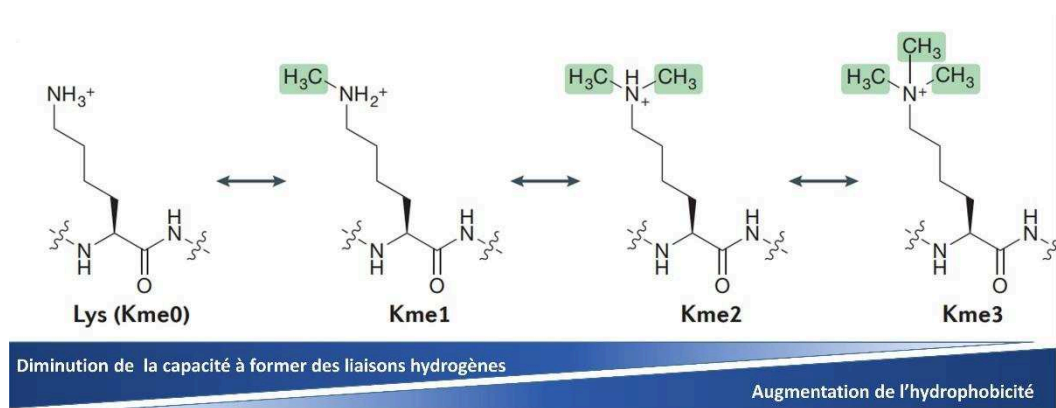


Figure 11 Influence du degré de méthylation d'une lysine sur l'hydrophobicité et leur capacité à former des liaisons hydrogènes. Adapté de Murn et al. 2017

La première enzyme : découverte permettant d'ajouter un groupement méthyle aux lysine, arginine, acide aspartique et acide glutamique est l'enzyme S-adenosylmethionine (SAM). D'autres enzymes sont responsables de la déméthylation des protéines comme la LSD1 qui est spécifique des lysines et de certaines histones : H3K4me1 ; H3K4me2 ; H3K9me1 et H3K9me2. Ces enzymes jouent un rôle clé dans la vie d'une cellule. En effet, la méthylation de H3K9 permet la répression du gène cycline E pendant la phase G1 mais ce gène ne doit plus être réprimé pendant la phase S et donc les groupements méthyle en H3K9 doivent être retirés pour être remis pour la suite du cycle cellulaire. La méthylation permet donc de contrôler en partie le cycle cellulaire. La méthylation est notamment importante pour les protéines se liant à l'ARN pour réguler leurs localisations, l'épissage de l'ARN pré-messager et la stabilité de l'ARN messager. Par exemple, la méthylation favorise l'export des protéines hnRNPs hors du noyau<sup>61</sup>. La méthylation d'une protéine peut permettre de bloquer son ubiquitination et donc de retarder sa dégradation ce qui implique une augmentation de sa demi-vie. Kim et al., ont montré que dans des cellules de

glioblastome la kinase AKT phosphoryle EZH2 qui à son tour catalyse l'ajout d'un groupement méthyle sur STAT3 et augmente sa phosphorylation, ce qui augmente son activité<sup>62</sup>. Cet exemple permet d'illustrer le lien entre méthylation et phosphorylation.

Plus de 5 500 protéines humaines correspondant à plus 16 000 lysines et arginines méthylées ont été observées par spectrométrie de masse montrant l'importance cette modification post-traductionnelle<sup>63</sup>.

#### (f) Ubiquitination

L'ubiquitine est une protéine composée de 76 acides aminés (8,5 kDa) localisée dans toutes les cellules humaines. Elle est présente sous forme de polyubiquitine (gènes UBB et UBC) ou attachée à une protéine ribosomale (RPS27 et UBA52). L'ubiquitine est conjuguée aux protéines afin de modifier leur stabilité, leurs interactions, localisations ou activités. L'ubiquitine est conjuguée aux protéines par l'intermédiaire d'une liaison polypeptidique entre sa glycine terminale et une lysine de la protéine à marquer. Trois enzymes sont impliquées dans ce processus : E1, E2 et E3. E1 active l'ubiquitine en la transférant sur l'enzyme de conjugaison E2. Enfin, l'enzyme E3 (RING, HECT ou RBR) permet le transfert de l'ubiquitine sur la protéine à marquer soit directement depuis l'enzyme E2 ou indirectement par conjugaison de l'ubiquitine sur une cystéine de l'E3 puis transfert sur la protéine cible (voir figure 12).

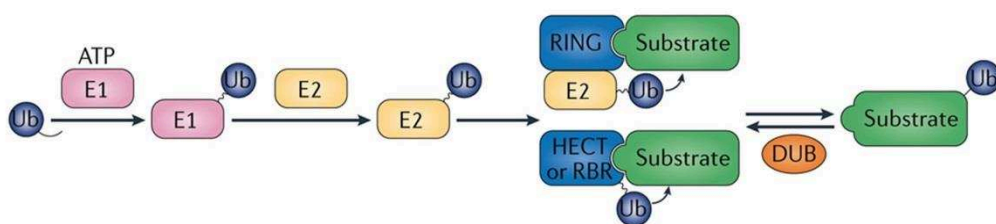


Figure 12 Schéma de principe de l'attachement d'une ubiquitine à une protéine. Premièrement l'enzyme E1 utilise l'ATP pour se lier à une ubiquitine puis l'enzyme E1 transfère cette ubiquitine vers l'une des 40 enzymes E2. Alors E2 par interactions avec l'E3 (RING, HECT ou RNR) transfère l'ubiquitine vers la protéine cible. L'enzyme DUB permet d'enlever ce marquage. D'après Rape et al. 2018

Les protéines peuvent être mono ou polyubiquitinylées. Dans le cas d'une seule protéine d'ubiquitine, cette modification est utilisée par le contrôle cellulaire pour modifier les interactions de la protéine cible avec d'autres protéines. Une ubiquitination plus complexe peut avoir lieu en conjuguant les ubiquitines entre elles en liant une des sept lysines ou la méthionine N-terminale avec la glycine d'une autre ubiquitine. Cette architecture complexe permet d'induire une réponse spécifique de la cellule. Par exemple, la dégradation de la protéine par le protéasome est induite par une chaîne polyubiquitine liée en K11 ou K48 ou K11/K48 alors qu'un enchaînement M1 ou K63 conduit à la régulation de l'endocytose. Ces enchaînements sont aussi impliqués dans la division cellulaire et peuvent être aussi modifiés par phosphorylation ou acétylation (voir figure 13). L'ubiquitination est donc un processus complexe permettant de contrôler de manière spécifique plusieurs fonctions cellulaires<sup>64,65</sup>.

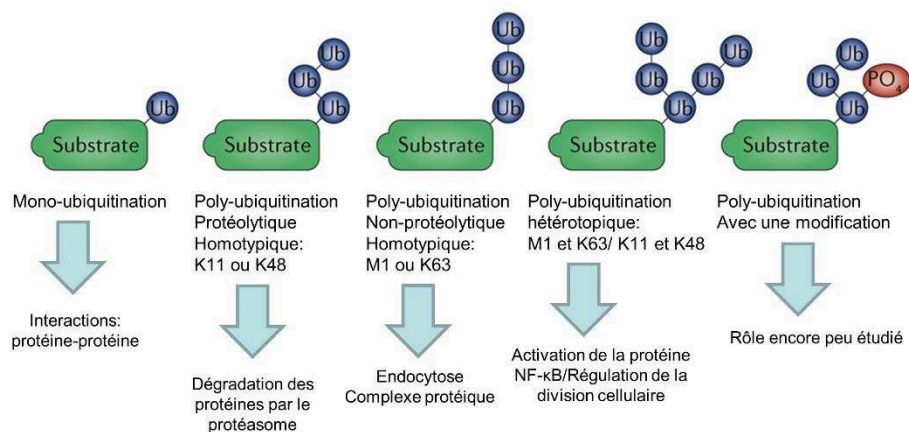


Figure 13 Les différents motifs d'ubiquitinylation avec leurs fonctions. D'après Rape et al. 2018

Dans certaines maladies neurodégénératives où les protéines agrègent comme la maladie d'Alzheimer ou de Parkinson l'ubiquitination semble jouer un rôle. En effet, pour la maladie d'Alzheimer, la formation de plaque amyloïde et de « tangles » de tau ubiquitinylé serait liée à un dysfonctionnement du protéasome. Concernant la maladie de Parkinson une mutation dans le gène Parkin qui régule l'ubiquitination de l'alpha-synucléine jouerait un rôle dans le développement de la maladie<sup>66,67</sup>.

(g) Sumoylation

Très proche de l'ubiquitination, la SUMOylation (small ubiquitin-like modifiers ≈12 kDa) est un processus servant principalement à contrôler les interactions de la protéine cible avec d'autres protéines<sup>68</sup>. Elles sont au nombre de 4 SUMO1, SUMO2, SUMO3 et SUMO4. Les SUMO forment comme l'ubiquitine des liaisons covalentes avec les lysines d'autres protéines et peuvent aussi se lier entre elles pour former des chaînes. La SUMOylation se passe principalement dans le noyau des cellules, en opposition avec l'ubiquitination qui a lieu dans tous les compartiments de la cellule. La SUMOylation est un processus dynamique<sup>69</sup> (voir figure 14). Le motif préférentiel de SUMOylation est [IV]-K-X-E, bien qu'en conditions de stress la SUMOylation devienne moins spécifique. Les protéines SUMO sont présentes sous une forme inactive et ont besoin d'une protéase comme la SENP, par exemple, pour l'activer en clivant les 2 à 11 acides aminés (ce nombre dépend de la protéine SUMO) situés avant les acides aminés GG en C-terminal. Une fois présente sous forme mature, la SUMO peut se lier par l'intermédiaire de l'ATP à une cystéine de l'enzyme E1 (ici SAE1-SAE2). La SUMO se conjugue ensuite avec l'enzyme E2 (UBC9) dont l'interaction est favorisée par le motif K-X-E. Les enzymes E3 permettent ensuite de former la liaison protéine-SUMO. Des protéases spécifiques, les SENP (sentrin-specific protease), permettent la déconjugaison des SUMO. D'une manière similaire à l'ubiquitination une chaîne polymérique de SUMO peut être formée.

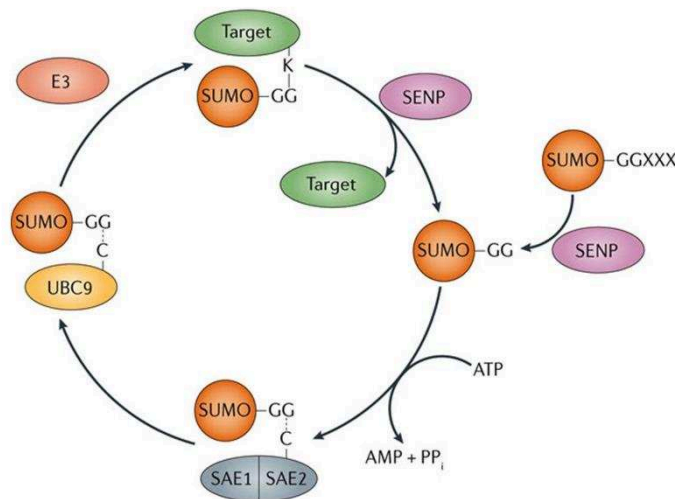


Figure 14 Schéma de principe de la SUMOylation des lysines. La protéine SUMO est dans un premiers temps activé par la protéase SENP qui expose les acides aminés GG. Puis cette protéine activée se lie à l'enzyme E1 qui est un complexe protéique SAE1 et SAE2 par l'intermédiaire de l'ATP. Puis la protéine SUMO est transféré sur l'enzyme E2 qui par l'intermédiaire de E3 la lie à la protéine cible. D'après Hendricks et al. 2016



Certaines protéines contiennent un motif d'interaction avec les SUMOs (SIMs : SUMO-interaction motifs). Ce motif est généralement hydrophobe avec une région acide (E ou D). Ces motifs ont plusieurs fonctions : créer une interaction SUMO-SIM, attirer l'E3 de l'ubiquitine pour que la protéine se fasse ubiquitinyler ou encore favoriser la SUMOylation d'une protéine. Comme l'ubiquitine les SUMOs peuvent être ubiquitinylés, phosphorylés, méthylés et acétylés. La SUMOylation semble être impliquée dans plusieurs maladies neurodégénératives comme la maladie d'Huntington<sup>70</sup> et les maladies d'Alzheimer<sup>71</sup>.

#### (h) Protéolyse endogène

Un certain nombre de protéases clivent les protéines en hydrolysant les liaisons peptidiques, ce qui génèrent une protéine tronquée ainsi qu'un peptide issu du clivage. Cette modification post traductionnelle est irréversible contrairement aux autres modifications post traductionnelles exposées précédemment. Le nombre de gènes codant pour les protéases humaines s'élève à 588, soit plus de 2% des gènes, divisé en 5 classes en fonction de leur activité catalytique : Aspartique (21), cystéine (164), métallo (192), sérine (184) et thréonine (27)<sup>72</sup>. Les protéases servent à dégrader des enzymes comme par exemple pour le renouvellement des enzymes marqué par l'ubiquitine mais elles génèrent aussi des protéines tronquées stables et peuvent permettre la libération d'un peptide signal.

Les protéases sont des protéines de tailles variées allant de 20 kDa à 6 MDa pour la métalloprotéase meprin. Elles peuvent aussi fonctionner en complexe protéique comme les protéasomes.

Les différents types de protéases sont les suivants :

- Aspartique : le site actif de ces protéases est composé d'un acide aminé chargé négativement acide aspartique ou acide glutamique et fonctionne donc mieux à pH acide. Cette réaction a besoin d'une molécule d'eau pour réaliser une réaction acide base avec la liaison peptidique à rompre. Les plus connus sont les pepsines et cathepsines.
- Cystéine : Leurs domaine catalytique est constitué de trois acides aminés et fonctionne par l'attaque nucléophile de la cystéine. Ces protéases fonctionnent en milieu légèrement réducteur pour ne pas oxyder la cystéine et pH (4.0-6.5). Par exemple : les calpaines et papaines.

- Métallo : deux grandes sous famille : les métalloprotéinases matricielles et les ADAMs. Elles sont caractérisées par la coordination, au niveau du site catalytique, d'un ion  $Zn^{2+}$  avec trois histidines et d'une molécule d'eau.
- Sérine : le site actif de ces protéases est composé de trois acides aminés une histidine, un acide aspartique et une sérine. La liaison peptidique se fait hydrolyser en commençant par une attaque nucléophile de la sérine. La trypsine est une protéase à sérine.
- Thréonine : le mécanisme est similaire aux protéases à sérine. Dans ce cas la triade catalytique est composée d'une thréonine, acide aspartique et d'une lysine. La première étape se fait par attaque nucléophile de la thréonine N-terminale.

L'activité des protéases est régulée par les inhibiteurs de protéases en bloquant directement le site catalytique des protéases, en se liant à une région adjacente du site catalytique ou en bloquant l'activité de la protéase (empêche sa dimérisation par exemple) ainsi que par leurs synthèses.

Les protéases répondent à de nombreux besoins cellulaires. Les protéines peuvent contenir un peptide signal qui détermine leur localisation dans la cellule. Ce peptide après la translocation est enlevé par les protéases<sup>73,74</sup>. Les protéases permettent aussi la production d'une protéine dite mature comme par exemple pour les SUMOs, l'insuline et l'angiotensine. Le ARN messager commence toujours par le codon AUG ce qui conduit à une méthionine en première position (N-terminale) de toutes les protéines. Cette méthionine est enlevée dans 55 à 70% des cas et le premier acide aminé de la protéine est un paramètre qui détermine la demi vie de la protéine<sup>75</sup>. Les protéases servent aussi à dégrader les protéines qui sont mal repliées notamment dans le réticulum endoplasmique.

L'apoptose recrute les caspases qui servent de régulateurs majeurs. Les caspases sont présentes dans les cellules sous une forme inactive qui est activé par un clivage au niveau de 3 acides aspartiques spécifiques <sup>76</sup>(voir figure 15). Les domaines p10 et p20 constituent alors un tétramères actif qui permet la dégradation des protéines

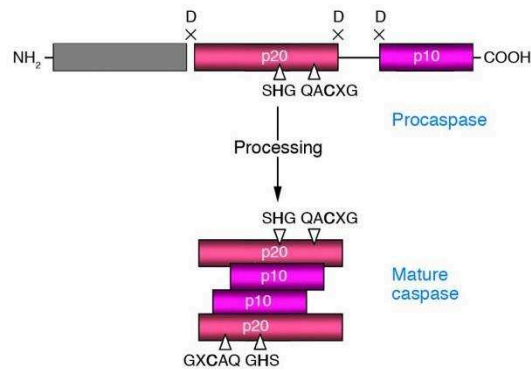


Figure 15 Activation de la caspase par clivage de la pro-caspase en C-terminale de spécifique acide aspartique. D'après Lavrik et al. 2005

Une dérégulation du clivage de l'APP, protéine transmembranaire de 87kDa, entraîne la formation de plaque amyloïde dans le cerveau, une des caractéristiques de la maladie d'Alzheimer. L'APP est clivé selon deux voies : la voie non amyloïdogénique, l' $\alpha$ -sécrétase clive le domaine extracellulaire de l'APP pour former sAPP  $\alpha$  puis son domaine intracellulaire est clivé par la  $\gamma$ -sécrétase. La voie amyloïdogénique où le domaine extracellulaire est premièrement clivé par  $\beta$ -sécrétase puis par la  $\gamma$ -sécrétase et l' $\alpha$ -sécrétase (voir figure 16). Ces clivages ne sont que peu spécifiques puisque plus de 30 peptides amyloïdes bêta sont produits. La voie amyloïdogénique a notamment pour conséquence une augmentation de la production de l'amyloïde bêta 1-42 qui a tendance à oligomériser. Le ratio 1-42/1-40 est utilisé en clinique comme biomarqueurs de la maladie d'Alzheimer car ce ratio est augmenté dans le LCR chez les patients Alzheimer.

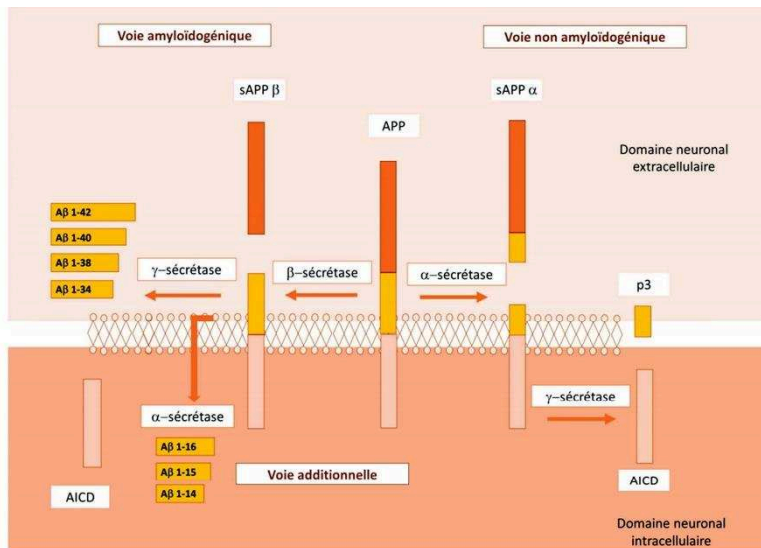


Figure 16 Formation des peptides amyloïdes beta. D'après thèse Bros P.

Une majorité des DLFT est caractérisée par des inclusions TDP-43 dans le cerveau. La TDP-43 est clivée par une protéase qui la clive notamment en N291. Or, des mutations dans des cas de DLFT (G290->A et S292->N) affectent des acides aminés situés autour du site de coupure, ce qui pourrait engendrer une mauvaise dégradation de TDP-43 et son agrégation.

Un cas particulier des protéases est le protéasome car il peut dégrader toutes les protéines. En effet, le protéasome dégrade toutes les protéines qui sont marquées par l'ubiquitine<sup>77</sup>(voir figure 17). Le protéasome permet de réguler la concentration d'une protéine et donc de répondre à un besoin cellulaire, ainsi que de dégrader les protéines ayant un mauvais repliement. Le protéasome est un complexe protéique de 2.5 MDa constitué de 33 sous-unités comprenant 3 sites catalytiques ayant des spécificités complémentaires. Le protéasome est une cible pour les maladies neurodégénératives dans lesquelles des protéines agrègent. En effet, dans ces maladies le protéasome ne remplirait plus ou mal sa fonction<sup>78</sup>.

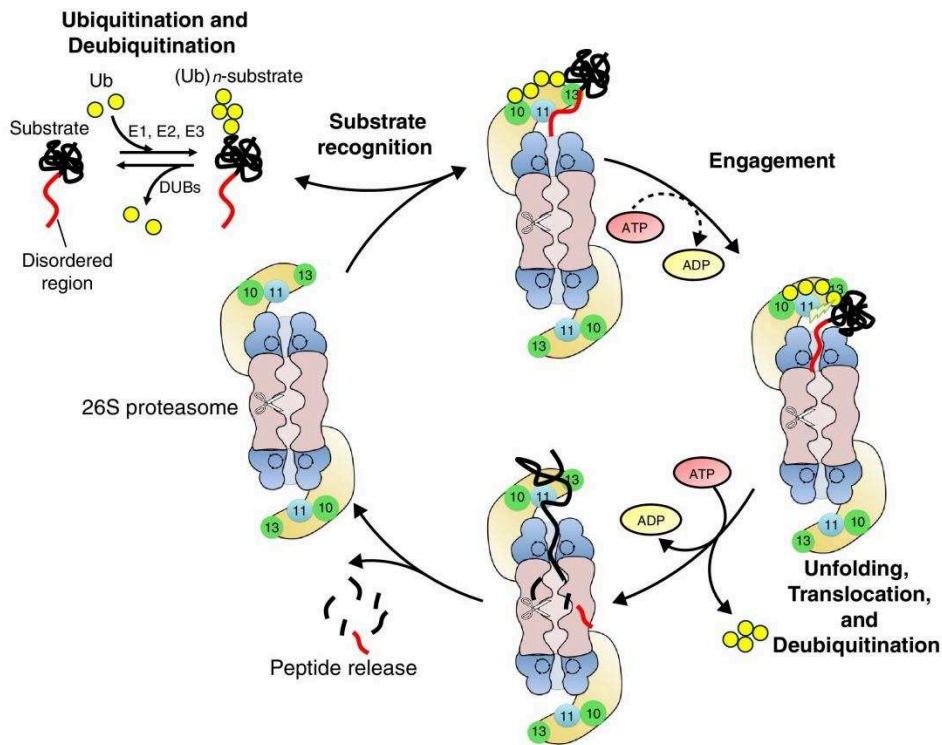


Figure 17 Mécanisme de dégradation des protéines par le protéasome. Les protéines poly-ubiquitinylés sont ciblées par le protéasome. La dégradation de la protéine est se fait en premier à partir d'une de ses régions non-structurés. D'après Inobe et al. 2014

### c) Complexité des protéoformes

Les protéoformes correspondent à des combinaisons de ces différentes modifications post-traductionnelles. De nombreuses études commencent à mettre en évidence une connectivité entre ces dernières notamment grâce au développement de méthode par spectrométrie de masse. Ces connectivités vont du cas simple de compétition entre deux modifications post-traductionnelle présente sur le même acide aminé, aux modifications promouvant la polyubiquitination, au cas le plus célèbre le code des histones. Les protéoformes permettent de répondre à des processus biologiques complexes et dans ce sens sont des biomarqueurs potentiels.

Un cas simple est la compétition entre deux modifications ayant lieu sur un même acide aminé. Comme c'est le cas entre la phosphorylation et la O-GlcNAc. Par exemple pour la protéine Tau, la O-glycosylation stabiliserait les microtubules alors que la phosphorylation déstabilise les microtubules<sup>79-81</sup>.

## Le code des histones

Les histones peuvent être modifiées par un certain nombre de modifications post-traductionnelles appelé « marques » comme par exemple les acétylations, les méthylations, les ubiquitinations et les phosphorylations. Ces modifications influencent la structure de la chromatine, donc la transcription, directement ou en permettant le recrutement d'autres molécules (facteurs) qui vont alors se lier à la chromatine. Notamment, l'acétylation neutralise la charge de la lysine ou la phosphorylation ajoute une charge négative sur une lysine et donc influence directement la structure compacte de la chromatine en défavorisant l'interaction des histones avec l'ADN. Concernant le recrutement de facteurs, ils reconnaissent certaines « marques » qui vont avoir un impact sur la structure de la chromatine. Par exemple, HP1 reconnaît la marque H3K9me3 et est associé avec la structure de la chromatine ainsi que l'extinction du gène<sup>82</sup>.

De plus ces modifications peuvent agir de façon antagoniste ou synergique c'est-à-dire qu'une modification peut en influencer une autre. L'antagonisme est principalement au niveau des lysines car plusieurs modifications post-traductionnelles peuvent y avoir lieu (acétylation, méthylation et ubiquitination) et ces modifications post-traductionnelles ont des propriétés différentes. De plus, deux modifications proches peuvent influencer de manière négative la liaison de l'histone avec un facteur. Comme par exemple, la phosphorylation de H3S10 qui empêche la liaison de HP1 avec la chromatine. Mais cet événement est régulé car il n'a lieu que pendant la mitose. Une modification peut être une condition pour la présence de la seconde. Par exemple, l'ubiquitination de H2B induit la di- et tri-méthylation de H3K4. Cette marque peut alors être reconnue par des facteurs ayant un domaine PHD<sup>83</sup>. Enfin, ces modifications peuvent coopérer pour lier plus efficacement des facteurs. Par exemple, Vermeulen et al. ont montré que le facteur PHF8 se liait plus fortement avec H3K4me3 quand les lysines K9 et K14 sont acétylées<sup>84</sup>. Ces marques et les connections entre elles forment le code des histones qui est impliquée dans la régulation de la transcription (voir figure 18).

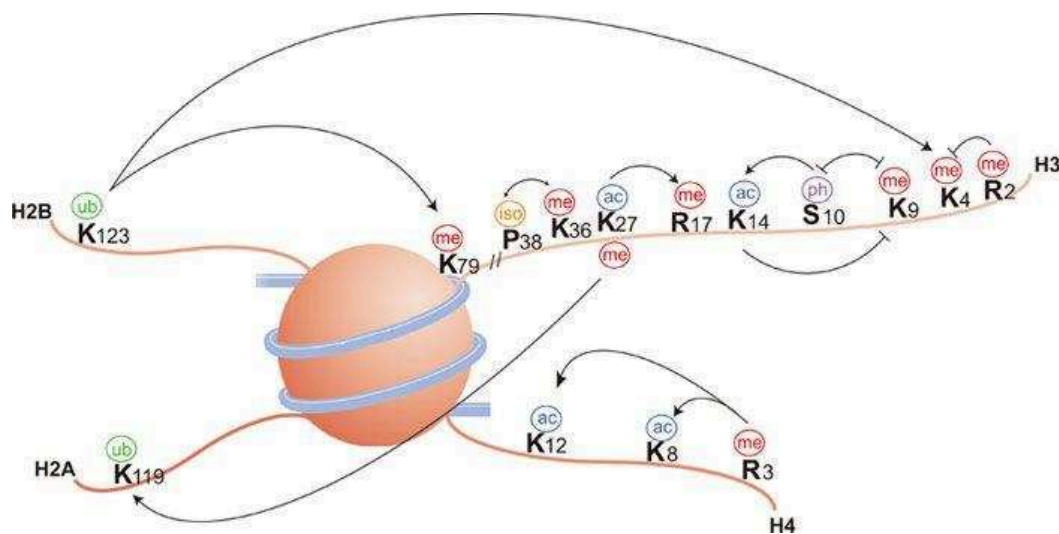


Figure 18 Interactions entre les différentes modifications post-traductionnelles des histones.. Ces interactions peuvent être positives (flèche) ou négatives (flèche à tête plate). D'après Bannister et al. 2011

## Conclusion :

La découverte de nouveaux biomarqueurs est d'un intérêt tout particulier dans le domaine clinique que ce soit comme outils de diagnostic ou pour suivre l'effet d'un traitement. Dans ce contexte les protéoformes sont une source de biomarqueurs potentiellement intéressante car ils pourraient traduire des phénotypes cliniques complexes<sup>19</sup>. De l'étape de découverte des biomarqueurs protéiques à leurs validations des techniques analytiques spécifiques doivent être employées pour les caractériser et les quantifier, dans le but d'aboutir à un test clinique pertinent. Les méthodes analytiques les plus couramment utilisées sont l'électrophorèse bidimensionnelle, les dosages immunologiques et la spectrométrie de masse. Les avantages et les inconvénients de deux dernières méthodes seront exposés dans la partie suivante.

Des avancées technologiques significatives ont été faites notamment dans le domaine de la spectrométrie de masse ce qui permet une analyse plus exhaustive des modifications post-traductionnelles et notamment des protéoformes. La spectrométrie de masse est un outil de choix pour analyser la complexité des protéoformes. Les méthodes de caractérisation des biomarqueurs protéiques et la quantification des protéoformes seront abordées dans la suite de ce manuscrit.

## II. L'analyse des biomarqueurs protéiques et de leurs protéoformes par spectrométrie de masse

L'analyse des biomarqueurs protéiques par spectrométrie de masse se fait en trois étapes principales. La première étape consiste à réduire la complexité de la matrice biologique soit par fractionnement soit en concentrant sélectivement les biomarqueurs ciblés, puis l'échantillon est analysé par spectrométrie de masse et finalement le signal obtenu est traité informatiquement afin d'identifier et quantifier la ou les protéines d'intérêt.

Dans un premier temps le principe de la spectrométrie de masse et ses caractéristiques seront exposés pour l'analyse des biomarqueurs protéiques. Puis les méthodes de préparation d'échantillon ainsi que les différentes stratégies d'analyse des protéines (avec digestion enzymatique : bottom-up et middle-down ; sans digestion enzymatique top-down) seront abordées. Finalement, les méthodes de quantification des protéines et leurs applications aux protéoformes seront exposées. Cette dernière partie traitera de la normalisation/standardisation nécessaire à la quantification ainsi que des méthodes d'acquisition utilisées, notamment le mode SRM et le mode PRM.

### 1. La spectrométrie de masse

#### a) Principe

La spectrométrie de masse introduite au début du 20<sup>ème</sup> siècle par J.J Thomson<sup>85</sup> permet la mesure du rapport masse sur charge ( $m/z$ ) d'ions en phase gazeuse. Un spectromètre de masse est constitué des éléments suivants (voir figure 19) :

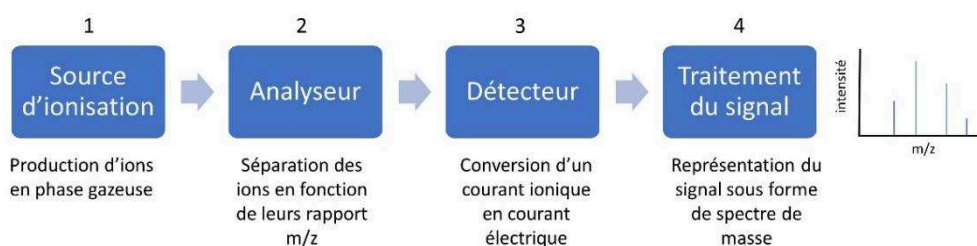


Figure 19 Représentation schématique du principe de la spectrométrie de masse



1) La source d'ionisation permet le transfert en phase gazeuse et l'ionisation d'une molécule. Il existe différents types de source selon la nature du composé à analyser et la matrice dans laquelle il se trouve. Pour les composés volatiles et présents sous forme gazeuse les deux principales sources d'ionisation sont :

- L'ionisation par électron : l'analyte est ionisé en traversant un faisceau d'électrons formé par le chauffage d'un filament.
- l'ionisation chimique : l'analyte réagit en phase gazeuse (acide/base) avec des ions précédemment formés par ionisation par électron.

Pour les composés non volatiles et en phase solide, les méthodes d'ionisation utilisées sont des méthodes de désorption par laser comme par exemple le MALDI (Matrix Assisted Laser Desorption Ionisation) ou par bombardement avec des molécules ionisées SIMS (Secondary Ion Mass Spectrometry).

Enfin pour les composés non volatiles comme les protéines les sources d'ionisation principales sont le MALDI et l'ionisation par électrospray (ESI). Pour l'ESI, les analytes, présents en phase liquide, sont transférés en phase gazeuse et ionisés par électronebulisation. Brièvement, une tension est appliquée à la pointe d'un capillaire ce qui crée une dispersion du liquide sous forme de microgouttelettes chargées. Le solvant contenu dans ces gouttelettes s'évapore laissant de plus en plus de charge à leur surface. Les microgouttelettes se divisent au moment où les forces de cohésions des microgouttelettes deviennent inférieures à la répulsion des charges (limite de Rayleigh). Ce processus est répété jusqu'à obtenir un ion nu.

2) L'analyseur permet la séparation des ions en fonction de leur rapport  $m/z$ . Il en existe plusieurs types basés sur des technologies différentes comme par exemple<sup>86</sup> : quadripôle, piège à ions, temps de vol (TOF), et orbitrap (voir tableau 3).

Tableau 3 Les caractéristiques des principaux analyseurs ( Adapté de Junot et al. 2014)

Type d'analyseur	Résolution (largeur à mi-hauteur)	Précision de masse
Quadripôle	7500 à m/z 508	/
Piège à ions	10 000	/
TOF	60 000 à m/z 1222	< 1-2 ppm avec une calibration interne
Orbitrap (q-Exactive)	140 000 à m/z 200	< 5 ppm avec une calibration externe

3) Le détecteur convertit un courant ionique en courant électrique. Pour les appareils à transformée de Fourier de type orbitrap, c'est l'image du courant formé par la fréquence d'oscillation des ions qui est mesurée et convertie en rapport masse sur charge.

4) Le signal est ensuite converti et représenté par son intensité en fonction du rapport m/z.

#### **b) Apport de la spectrométrie de masse pour l'analyse des protéines et de leurs protéoformes**

La spectrométrie de masse est un outil analytique de choix pour l'analyse des protéines. Cette technique apporte une **spécificité** unique en comparaison aux ELISA notamment pour l'analyse des protéoformes. Le second avantage est une **gamme dynamique** plus importante qui permet la caractérisation de protéines de faible abondance sous réserve d'une préparation d'échantillon adaptée (voir II.2). Finalement, la spectrométrie de masse permet l'analyse sans a priori d'un échantillon biologique (cf. Trypsine et « shotgun » protéomique).

L'analyse des protéines peut se faire de deux manières principales, soit en analysant les peptides produits par digestion enzymatique des protéines (Bottom-up), soit en les analysant intacts (Top-down)<sup>87</sup> (figure 20). Dans les deux cas, les ions (peptides ou protéines) sont fragmentés soit dans le but d'identifier leurs séquences en acides aminés ou dans un but de quantification. Cette étape de fragmentation confère à la spectrométrie de masse une dimension de spécificité supplémentaire.

Les spectromètres de masse doivent donc être capables d'analyser une grande variété de masse et le choix du spectromètre de masse va dépendre de l'approche analytique suivie.

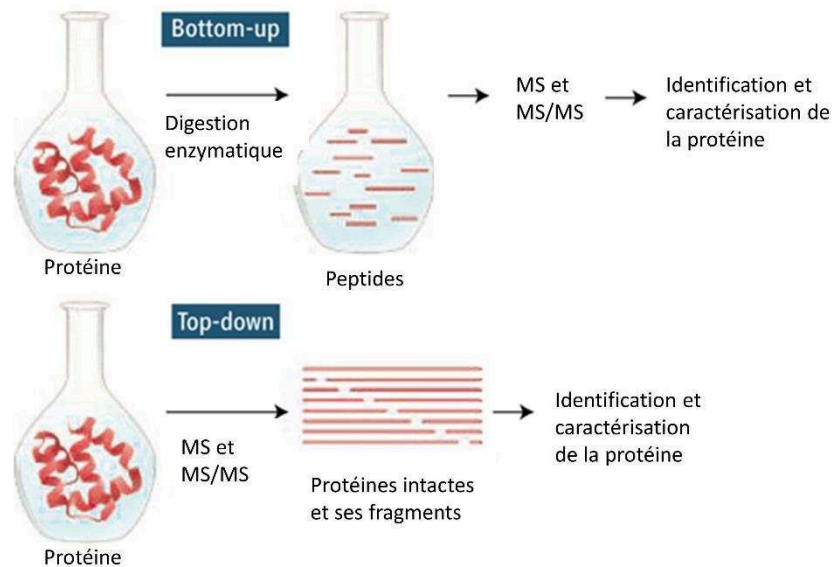
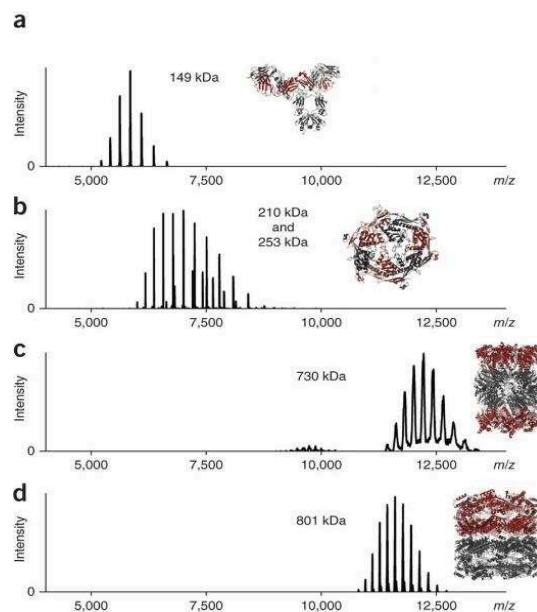


Figure 20 Approches d'identification et de caractérisation des protéines par spectrométrie de masse. Adapté de Chait et al. 2006

**(i) Gamme de masse adaptée à la diversité des protéines**

Concernant l'approche bottom-up : l'enzyme la plus couramment utilisée est la trypsine car elle génère des peptides ayant des propriétés physico-chimiques intéressantes pour l'analyse par spectrométrie de masse. En effet, la trypsine produit des peptides comprenant environ 8 à 25 acides aminés soit une masse d'environ 1,4 kDa<sup>88,89</sup>. Les peptides produits possèdent une lysine ou une arginine en C-terminale, du fait de la spécificité du clivage de la trypsine, ce qui conduit majoritairement à la formation de peptides di-chargés en mode d'ionisation positive. La majorité des peptides possèdent alors un rapport  $m/z$  compris entre 300 et 1400. Ces  $m/z$  sont facilement analysable par les spectromètres de masse conventionnels<sup>90</sup> tel que les quadripôles (10 -4000  $m/z$ ) ou les orbitraps (50-2000  $m/z$ ).

A l'autre extrémité, l'analyse des protéines intactes peut se faire en conditions natives, c'est-à-dire en conditions non dénaturantes<sup>91</sup> (solvants organiques, acides, voltage source electrospray) ce qui permet de garder la structure native de la protéine et des complexes protéiques intactes ou en conditions dénaturantes. L'analyse des protéines en conditions natives est possible en utilisant principalement les appareils de type temps de vol (TOF) car n'ayant pas de limite théorique haute en  $m/z$ . Même si en pratique, il semble y avoir une limite pour avoir une résolution au niveau des états de charges, due probablement à une mauvaise désolvatation, à 20 MDa comme montré lors de l'analyse des capsides des bactériophages HK97<sup>92</sup>. Les spectromètres de masse de types orbitrap se sont aussi montrés efficaces pour l'analyse de protéines de haut poids moléculaire et  $m/z$ <sup>93</sup> avec une précision de l'ordre de 50 ppm, comme par exemple pour la protéine GroEL 800kDa ( $m/z$  of 10,000 à 12,000 pour les états de charges 68+ à 77+) (voir figure 21). Des spectromètres de masse moins conventionnels comme ceux à détection de charge permettent l'analyse de molécules au-delà du MDa en mesurant simultanément la charge et le rapport  $m/z$  comme par exemple le phage P22 de 50 MDa<sup>94</sup>.



**Figure 21** Spectre de masse de protéines et de complexes protéiques analysées en intactes (Rose et al. 2012) ainsi que leur structure cristallographique. Spectre de masse d'un IgG (149 kDa) (a), capside pentamérique et hexamérique du phage HK97 (210 kDa et 253 kDa) (b), complexe protéique du protéasome 20S d'une levure (730 kDa) (c) et la protéine GroEL (800kDa) (d).

La source electrospray crée des ions multichargés, il peut donc être intéressant de travailler en conditions dénaturantes c'est-à-dire en présence d'un solvant organique et d'un acide afin d'augmenter le nombre de charges présent sur la protéine et d'analyser des états de charges plus hauts donc de plus faible  $m/z$ , en comparaison avec les conditions natives<sup>91</sup>.

## (ii) ELISA et spécificité apportée par la spectrométrie de masse

La méthode sandwich immuno-enzymatique ELISA (enzyme-linked immunosorbent assay) est la technique analytique couramment utilisée en clinique pour la quantification des biomarqueurs protéiques. Cette technique permet la quantification d'une protéine avec une sensibilité pouvant atteindre le  $\text{pg/mL}$ <sup>95</sup> soit au moins une dizaine de fois plus sensible que les méthodes par spectrométrie de masse les plus sensibles<sup>96</sup>. Cette technique se base sur l'association de deux anticorps, un de capture et un de détection, dirigés contre un antigène pour sa détection (voir figure 22).

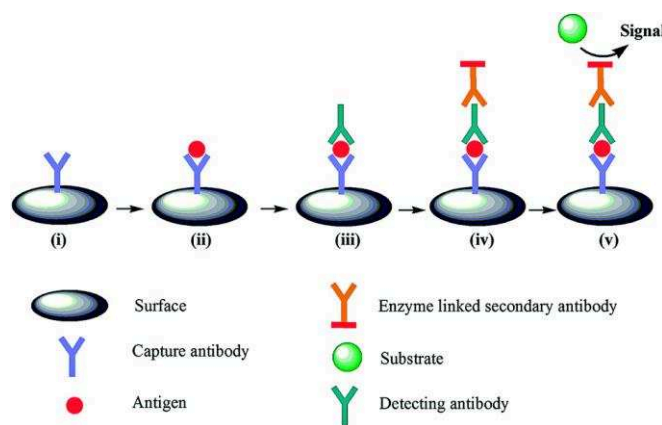


Figure 22 Représentation schématique du principe de la technique ELISA (Mistry et al. 2014). (i) l'anticorps de capture est greffé à la surface d'une plaque, (ii) l'échantillon contenant l'antigène est ensuite appliqué dans les puits de la plaque puis (iii) l'anticorps de détection est ajouté. (iv) Enfin, un anticorps secondaire lié à une enzyme est ajouté et va se lier à l'anticorps de détection. Les enzymes couramment utilisées sont la peroxydase de raifort et la phosphatase alcaline. (v) La détection se fait par la conversion d'un substrat avec l'enzyme en signal fluorescent.

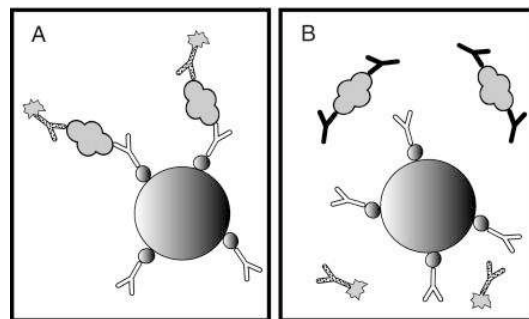
Lin et al.<sup>97</sup>, ont comparé ELISA et Immunoprécipitation-MRM pour des biomarqueurs du cancer dans du plasma artificiel (60mg/mL BSA). Les deux méthodes se sont montrées équivalentes en termes de fidélité des mesures avec des CV allant de 2% à 19%. En revanche la limite de quantification en ELISA est supérieure à la limite de détection par MRM avec un gain de sensibilité d'un facteur 6 à 36. Les ELISA ont l'avantage, en plus d'être généralement plus sensibles que les

méthodes par spectrométrie de masse, de ne pas avoir de besoin d'une préparation d'échantillon aussi fastidieuse, ce qui contribue à rendre les ELISA plus simples d'utilisation et l'analyse plus rapide. En effet, une simple dilution de l'échantillon biologique peut être suffisante. Des tests ELISA sont couramment utilisés en routine à l'hôpital pour mesurer par exemple l'hyperphosphorylation de tau sur les thréonines en position 181 et 231 dans le LCR comme biomarqueur de la maladie d'Alzheimer<sup>98,99</sup>. Les limites de quantification pour tau et p-tau 181 sont de 1,6 à 7,8 pg/mL dépendant du couple d'anticorps utilisé, avec une gamme dynamique de  $2 \log^{100}$ , les limites de quantification atteintes en ELISA pour la protéine tau sont supérieures à celles obtenues par spectrométrie de masse qui sont à 150 pg/mL pour le meilleur peptide<sup>101</sup>.

L'avantage de la spectrométrie de masse est qu'elle permet dans ce cas d'avoir une vue plus globale des sites de phosphorylation de tau dans le LCR, contrairement aux ELISA limité aux anticorps disponibles. En effet, Russel et al.<sup>40</sup>, ont développé une méthode de quantification de plus de 30 sites de phosphorylation de Tau et ont montré que 11 sites de phosphorylation sont augmentés d'au moins 40% chez les patients atteints maladie d'Alzheimer<sup>40</sup>. Pour les peptides amyloïdes bêtas, les méthodes ELISA ne peuvent pas quantifier la diversité des protéoformes et donc évaluer leur potentiel clinique. En effet, seulement des kits ELISA spécifiques pour 1-38,1-40 et 1-42 existent. Alors que, la spectrométrie de masse à haute résolution a permis l'identification de 26 protéoformes de abeta dans le cerveau humain<sup>102</sup> et 29 dans le LCR<sup>103</sup>. Les ELISA ne permettent que dans des cas précis la quantification de protéoformes comme par exemple pour l'hyper-phosphorylation de Tau sur la thréonine 181. En effet, la production d'un nouvel anticorps spécifique pour un antigène (ici un protéoforme) prend plus de 6 mois<sup>104</sup> et il paraît laborieux de produire un anticorps par protéoforme.

Les tests ELISA peuvent aussi souffrir de réactivité croisée. Par exemple dans le cas du peptide natriurétique de type B (BNP) qui est utilisé comme biomarqueur clinique pour l'insuffisance cardiaque. Ce peptide est produit par conversion de sa proHormone (proBNP) de 108 acides aminés en une forme active 1-32 et une forme peu ou pas active 1-76. Le peptide BNP 1-32 est ensuite protéolysé dans le plasma majoritairement en 3-32, 4-32 et 5-32. Les tests ELISA peuvent donc surestimer la concentration de BNP 1-32 par réaction croisée des anticorps avec les formes protéolysées de BNP 1-32 qui sont majoritaires, ainsi que la proBNP comme démontré par spectrométrie de masse. Dans une récente, Suzuki et al. ont prouvé la pertinence de suivre spécifiquement les formes 3-32, 4-32 et 5-32 comme outils diagnostique en utilisant la spectrométrie de masse<sup>105</sup>.

Les ELISA peuvent aussi conduire à des valeurs faussement basses comme dans le cas d'une production d'autoanticorps contre la protéine cible ou faussement élevées causées par la présence d'interférences hétérophiles. Les interférences hétérophiles sont dues à la réaction d'anticorps endogènes humains avec les anticorps animaux de capture et de détection<sup>106</sup>. Une seconde source de variabilité est la présence d'autoanticorps, qui sont des anticorps endogènes humains ayant pour cible une protéine endogène humaine comme c'est le cas pour la thyroglobuline<sup>107</sup>. Les autoanticorps anti-thyroglobuline sont présents chez 1 personne sur 10 et un quart des patients atteints de cancer de la thyroïde. La présence de ces anticorps anti-thyroglobuline induit une diminution de la thyroglobuline mesurée (voir figure 23).



**Figure 23** ELISA sans interférences pour la thyroglobuline (gris) qui est capté par l'anticorps blanc et est détectée par l'anticorps à pois (A). En cas d'autoanticorps (noir) la thyroglobuline n'est plus captée par les anticorps de capture (blanc), ce qui résulte en une sous-estimation de la quantité de thyroglobuline (B). Adapté de Hoofnagle et al. 2009

La spectrométrie de masse bien que moins sensible que les tests ELISA a l'avantage de pouvoir détecter et quantifier de manière spécifique les protéines et leurs protéoformes. La spécificité est apportée par :

- La résolution qui correspond à la capacité à séparer deux ions dont le rapport  $m/z$  est proche. Elle est définie par  $m/\Delta m$  où  $m$  :  $m/z$  de l'ion d'intérêt et  $\Delta m$  est l'écart entre l'ion d'intérêt et l'ion le plus proche dont il peut être séparé. Par exemple, les peptides AAAAAAQPK et AAAAAAKPK ont un  $\Delta m$  d'uniquement 0,03638Da, la résolution minimale nécessaire est donc de 11 000. Seuls les appareils de moyenne et haute résolution pourront distinguer ces deux peptides. La résolution est particulièrement importante dans le cas des protéoformes car elle permet de caractériser une substitution d'un acide aminé (Q->K), les isoformes et les modifications post-traductionnelles notamment une déamination par exemple qui induit une différence de masse de 0,984 Da.

- La précision sur la mesure du rapport m/z exprimée en partie par million (ppm) correspond à la capacité de l'analyseur à mesurer le rapport m/z aussi proche que possible du m/z théorique. Les analyseurs de type orbitrap ont une précision de masse de l'ordre de 5 ppm, de 1 à 2 ppm pour les TOF et 200 ppm pour les pièges à ions et quadripôle.
- La fragmentation permet d'augmenter la spécificité de l'analyse par l'analyse des ions fragments (MS/MS) ou des ions n fois fragmentés MS<sup>n</sup>. Ces méthodes permettent d'obtenir l'enchaînement des acides aminés composant les peptides et donc apportent un degré d'information supplémentaire.

### (iii) Linéarité et les effets matrices

Les matrices biologiques comme le plasma sont complexes avec une gamme dynamique<sup>108,109</sup> pour les protéines de l'ordre de 10<sup>12</sup> cf. Figure 24.

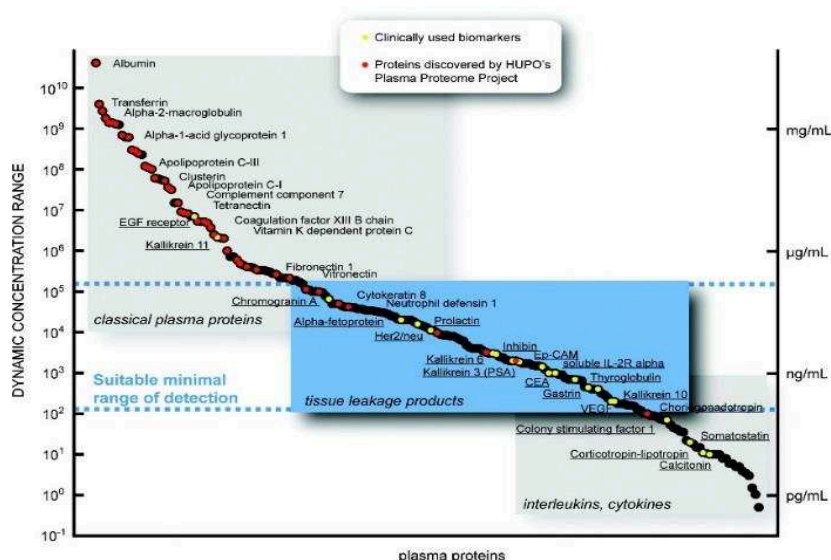


Figure 24 Gamme dynamique des concentrations des protéines plasmatiques. En jaune, les biomarqueurs utilisés en clinique et en rouge les protéines identifiées par le projet HUPO plasma proteome project (PPP) d'après Surinova et al. 2011

Seulement douze protéines représentent plus de 95% de la masse totale des protéines plasmatiques<sup>110</sup> dont plus de la moitié par la seule albumine qui a une concentration de l'ordre de 33-53mg/mL. Ces protéines vont avoir tendance à limiter la détection des protéines de faibles abondances car la source electrospray est un processus compétitif<sup>111</sup>. De plus dans les spectromètres de masse de type Orbitrap, la C-trap qui permet d'accumuler les ions avant injection dans l'orbitrap, va être remplie principalement d'ions abondants interférents, le temps d'accumulation de l'ion d'intérêt sera moins important. Ce temps d'accumulation plus court a pour

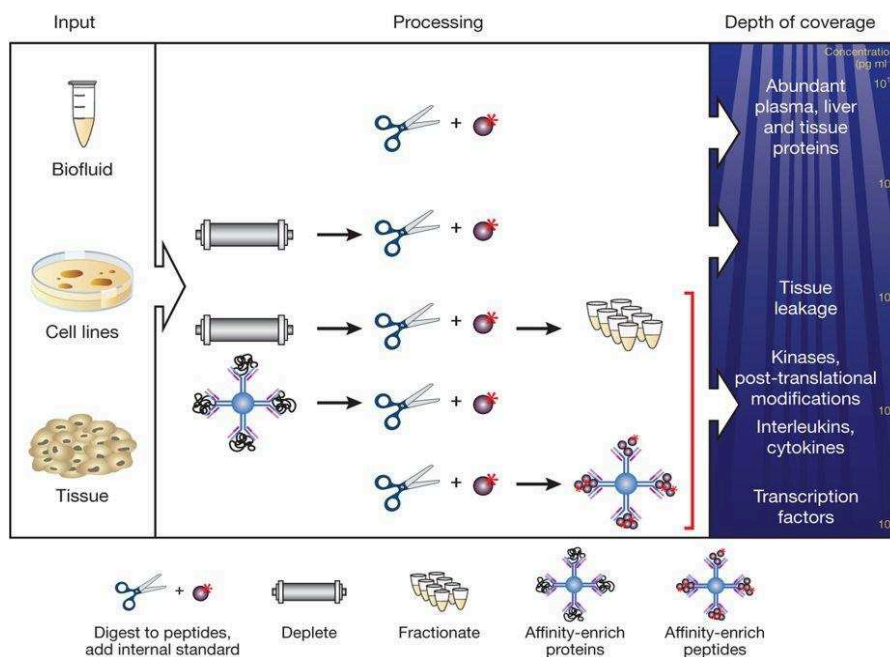


conséquence une diminution du rapport signal sur bruit. Ces deux processus constituent l'effet matrice.

## 2. Préparation des échantillons

Pour pallier aux défauts de la spectrométrie de masse, gamme dynamique inférieure à la gamme dynamique des protéines plasmatiques et effets matrices, des stratégies de préparations d'échantillons doivent être mises en place. La méthode de préparation d'échantillon est sélectionnée en fonction de la complexité de la matrice et de la concentration de la protéine d'intérêt. Plus la protéine est minoritaire plus la préparation d'échantillon doit être spécifique (voir figure 25).

Sans préparation d'échantillon autre que la digestion enzymatique des protéines, uniquement les protéines majoritaires sont détectables et quantifiables<sup>112</sup>. Comme par exemple, pour les hormones de croissances IGF-1 et IGFBP-3<sup>113</sup>.



**Figure 25** Stratégies de préparation pour la détection des peptides pour améliorer la sensibilité et la spécificité des méthodes par dilution isotopiques. L'ajout de peptides marqués permet d'augmenter la précision la mesure. La déplétion seule des 12 à 70 protéines majoritaires du plasma permet d'atteindre des limites de quantifications de l'ordre de 100ng/mL. L'ajout d'une étape de fractionnement permet d'atteindre des limites de quantifications proches du ng/mL voir même 50 pg/mL pour la PSA. Un autre moyen d'atteindre des limites de quantifications aux alentours du ng/mL est l'immunoprécipitation au niveau de la protéine ou des peptides (SISCAPA). D'après Gillette et al. 2013

### **a) Précipitation des protéines et déplétion des protéines majoritaires**

Pour quantifier les protéines majoritaires du plasma, une simple étape de précipitation peut suffire. En effet la précipitation permet une diminution de l'effet matrice en enlevant des métabolites, les petites protéines et certains lipides. De plus, une précipitation utilisant un dénaturant comme le méthanol améliore le rendement de digestion des protéines. Ces améliorations ont été rapportées par Ouyang et al.<sup>114</sup> qui ont obtenu un gain de sensibilité d'un facteur 15 par simple précipitation des protéines du plasma par le méthanol, suivi par une digestion du pellet de protéines, en comparaison à une digestion directe du plasma.

L'autre stratégie pour la quantification de protéines abondantes dans le plasma consiste à dépléter les protéines majoritaires. L'albumine représente la moitié des protéines du plasma, l'enlever réduit donc la complexité de la matrice. Pour cela il existe des kits commerciaux comme le Kit vendu par Thermo Fisher : Pierce Albumin Depletion Kit. Une méthode alternative d'extraction de l'albumine est basée sur sa solubilité dans l'isopropanol avec 1% d'acide trichloracétique<sup>115</sup>. Liu et al.<sup>115</sup>, ont démontré que dans leurs conditions cette méthode est plus efficace que les kits commerciaux qu'ils ont testés. En effet elle permet d'enlever 98% d'albumine contre 50% pour les kits commerciaux testés. De plus, le gain de sensibilité apporté par Liu et al, dans les conditions testées, par rapport à une simple précipitation au méthanol est d'un facteur 2.

Pour quantifier des biomarqueurs moins abondants ces méthodes doivent être combinées à des méthodes de fractionnement comme les extractions en phase solide ou fractionnements par chromatographie liquide.

### **b) Extraction en phase solide et fractionnement chromatographique**

Les biomarqueurs protéiques, par exemples la PSA et les troponines cardiaques, dans le plasma sont aux alentours du ng/mL. L'analyse du plasma par simple précipitation ou par déplétion des protéines majoritaires ne permet pas leur détection. Une étape supplémentaire de préparation d'échantillon, qui se fait généralement en complément d'une étape de précipitation ou d'extraction des protéines majoritaires, doit être ajoutée comme par exemple l'extraction en phase solide<sup>116</sup>. L'extraction en phase solide peut être faite soit au niveau de la protéine<sup>101</sup> soit, plus couramment, au niveau des peptides<sup>117</sup>. La SPE permet de reconcentrer

l'échantillon et de réduire les effets matrice. Une autre méthode est le fractionnement par chromatographie liquide. Cela correspond à une chromatographie liquide dite bidimensionnelle (2D). Les caractéristiques pour un fractionnement idéal<sup>118</sup> sont : une capacité de pics importante et l'orthogonalité du fractionnement avec la séparation analytique des peptides. L'orthogonalité est définie comme la répartition des peptides dans un espace géométrique en 2D. Une évaluation de ces paramètres a été effectuée en comparant les phases stationnaires suivantes C18 (phase inverse), HILIC (Chromatographie d'interaction hydrophile), SEC (Chromatographie d'exclusion stérique) et SCX (Chromatographie d'échange de cations) pour un mélange de protéines standards. Les résultats de cette comparaison sont à voir tableau 4.

**Tableau 4 Orthogonalité et capacité de pics pratique pour différents couple chromatographique. Adapté de Gilar et al. 2005**

	Orthogonalité	Capacité de pics pratique
C18 pH=10/ C18 pH=2.6	53%	5425
HILIC/ C18 pH=2.6	69%	4616
SEC/ C18 pH=2.6	54%	708
SCX/ C18 pH=2.6	54%	2430

Le couple C18 à haut et bas pH et l'HILIC/C18 à bas pH ont les meilleures performances en termes d'orthogonalité et de capacité de pics pratique. Les autres couples testés sont moins performants. En effet, la largeur des pics chromatographiques en SEC est supérieure aux autres colonnes (1min pour la SEC contre 0.35min pour la C18, 0.4 min pour l'HILIC et 0.6 min pour la SCX) ce qui induit une capacité de pics inférieure. De plus, à cause d'adsorption non spécifique il y a une perte de peptides conséquente, la taille des pores n'étant pas adapté à l'analyse des peptides. La SCX a une capacité de pics moins élevée que les colonnes C18 et HILIC car les pics sont plus larges. De plus la séparation est guidée par la charge des peptides, or les peptides tryptiques sont majoritairement des ions chargés 2+ et 3+, ils ont donc des temps de rétentions proches et donc une fenêtre d'élution restreinte. Le second inconvénient de la SCX est la présence de sels qui ont tendances à avoir un effet négatif sur l'ionisation par electrospray, il faut donc ajouter une étape de dessalage de l'échantillon, ce qui peut entraîner une perte des peptides par adsorption non spécifique. Ces

résultats ont été confirmés par Wang et al.<sup>119</sup> concernant la C18 et la SCX. En montrant que malgré une orthogonalité égale entre C18 à haut pH/C18 à bas pH et SCX/C18 bas pH, le nombre de peptides et de protéines identifiées est respectivement 1,8 et 1,6 fois supérieur. Cette augmentation du nombre de peptides et de protéines identifiées est due à une capacité de pics est plus élevée avec une colonne C18 en comparaison avec une colonne SCX.

Keshishian et al.<sup>120,121</sup>, ont néanmoins montré l'intérêt du fractionnement par SCX. Par déplétion des 12 protéines majoritaires du plasma au lieu des 7 majoritaires, la LOQ est augmenté d'un facteur 2. Par contre, un fractionnement par SCX, bien que ne comprenant que 6 fractions, permet d'augmenter la LOD et la LOQ jusqu'à un facteur 1000 en atteignant le ng/mL et permet la quantification de biomarqueurs cardiaque. En dépliant les 14 protéines majoritaires et en réalisant un fractionnement sur colonne C18 haut pH/ bas pH plus poussé avec 96 fractions qui sont ensuite combinées en 12 fractions, Shi et al.<sup>96</sup> ont atteint des limites de quantification en plasma pour la PSA de 50-100 pg/mL. Cette limite de quantification est la plus basse atteinte pour la PSA par spectrométrie de masse (voir tableau 5) mais résulte d'un fractionnement conséquent.

### **c) Immunoprécipitation (protéine/peptides)**

Les méthodes les plus sensibles, en générale, sont les méthodes d'immunoprécipitation des protéines. En effet, l'immunoprécipitation est la méthode de préparation la plus stricte qui permet d'avoir des limites de quantification inférieures à 1 ng/mL. La capture peut se faire avec un anticorps dirigé contre la protéine ou dirigé contre un peptide protéotypique de la protéine (SISCAPA : Stable Isotope Standards with Capture by Anti-Peptide Antibodies).

L'immunoprécipitation au niveau de la protéine peut être multiplexée et la capture peut être normalisée en utilisant une version marquée <sup>13</sup>C et/ou <sup>15</sup>N de la protéine<sup>122</sup>. Cette méthode peut être appliquée pour la quantification de protéoformes comme par exemple pour la quantification d'isoformes d'une toxine, l'abrine, en utilisant une combinaison de quatre anticorps monoclonaux comme montré par Hansbauer et al.<sup>123</sup>, ou dans le domaine des biomarqueurs en quantifiant deux isoformes de l'antigène prostatique spécifique<sup>124</sup>. Gao et al.<sup>125</sup> ont réalisé une immunoprécipitation multiplexe de 5 protéines, impliquée dans l'inflammation et l'insuffisance rénale, et ont quantifié leurs protéoformes grâce à la spécificité apportée par la spectrométrie de masse.

L'immunoprécipitation peut aussi se faire au niveau du peptide ce qui permet de normaliser l'immuncapture par l'ajout d'un peptide marqué. La méthode SISCAPA a été introduite par Anderson et al.<sup>126</sup> en 2004 pour la quantification de quatre protéines du plasma. Cette méthode a récemment été appliquée pour l'enrichissement multiplexe de 24 peptides<sup>127</sup> ou même plus de 150 peptides et la quantification de 41 protéines<sup>128</sup> voir 400 peptides<sup>129</sup>. Ces résultats montrent une capacité de multiplexage plus importante qu'avec une capture faite au niveau de la protéine. Contrairement aux ELISA, les anticorps auto induits n'interfèrent pas avec le SISCAPA et ont permis la quantification de la thyroglobuline<sup>130</sup> en plasma avec une limite de quantification de 0.5 ng/mL.

L'inconvénient majeure du SISCAPA par rapport à l'immunoprécipitation réalisé avec une protéine marquée aux isotopes stables (<sup>13</sup>C et/ou <sup>15</sup>N) est la non normalisation de la digestion, une des étapes les plus variables<sup>131</sup>, renforcé par le fait que la digestion se fait un milieu complexe pour le SISCAPA et donc une trypsine de moins bonne qualité doit être utilisée. La qualité de la trypsine influe sur la spécificité de son clivage. En effet, la normalisation par une protéine marquée<sup>132</sup> a permis d'augmenter la justesse d'un facteur 2 dans le cas de la quantification de la protéine S100B. Pour pallier à la variabilité causée par la digestion la méthode SISCAPA peut être automatisé<sup>131</sup>. En général seulement un à deux peptides par protéine est quantifié en SISCAPA, ce qui limite son utilisation pour l'étude des protéoformes. Bien que l'étude de modifications post-traductionnelles soit possible dans des cas précis, comme par exemple pour l'acétylation de la tubuline<sup>133</sup>.

La production d'anticorps est longue, couteuse et nécessite l'utilisation d'animaux ; des alternatives ont donc été développées comme les aptamères<sup>134</sup>, oligonucléotides facilement synthétisables et les affimers, petites protéines pouvant être produites par des bactéries ou cellules. Les aptamères sont sélectionnés d'après la méthode SELEX (Systematic Evolution of Ligands by EXponential enrichment) qui permet la sélection d'un aptamère d'après une banque d'ARN ou ADN, dont les séquences sont aléatoires, en 3 à 7 semaines<sup>135</sup>. Les affimers ont montré des performances similaires aux anticorps pour la quantification de la protéine sRAGE<sup>104</sup> avec une LOQ à 0.2 ng/mL.

En résumé, l'étape de préparation d'échantillon est une étape cruciale dans le développement d'une méthode de quantification des protéines par spectrométrie de masse. Moins la protéine est abondante, plus la préparation d'échantillon doit être spécifique. En règle générale les méthodes par immunoprécipitation sont les plus sensibles avec des LOQ  $\approx$  1 ng/mL mais comme illustré dans le tableau 5 pour la quantification de la PSA<sup>136</sup>, une méthode par déplétion des 14 protéines majoritaires et par fractionnement permet d'atteindre une limite de quantification environ 10 fois

supérieure à celle obtenue par immunoprécipitation. L'immunoprécipitation est, en effet, dépendante de la qualité de l'anticorps.

**Tableau 5 Comparaison de différentes méthodes de préparations d'échantillons en termes de sensibilité et de fidélité pour la quantification de la PSA plasmatique. Adapté de Makawita et al. 2010.**

Préparation d'échantillon	Sensibilité (LOQ, ng/mL)	Fidélité (CV)	Référence
Digestion directe	4500 ng/mL (LOD)	13%	Barnifge D.R. et al., 2004
Déplétion des 7 protéines majoritaires, digestion	124 ng/mL	19.7%	Keshishian H. et al., 2007
Déplétion des 12 protéines majoritaires, digestion	~46 ng/mL	8.5%	Keshishian H. et al., 2007
Déplétion des 12 protéines majoritaires, digestion, fractionnement par SCX	~2 ng/mL	2.8%	Keshishian H. et al., 2007
Déplétion de l'albumine, digestion, extraction en phase solide (MCX)	4,5 ng/mL	<10%	Fortin T. et al., 2009
Déplétion des 14 protéines majoritaires, digestion, fractionnement par high-pH RPLC	<0,05 ng/mL	~10%	Shi T. et al., 2012
Immunoprécipitation de la protéine, digestion	0,5 ng/mL	15.7%	Chen Y-T. et al., 2015

### **3. Les approches de protéomique et leur application à l'analyse des protéoformes.**

La spectrométrie de masse s'est imposée comme l'outil central pour l'identification, la caractérisation et la quantification des protéines provenant de milieux biologiques pour plusieurs raisons. Premièrement, les analyses peuvent être faites sans a priori ce qui est particulièrement important pour la découverte de nouveaux biomarqueurs ainsi que pour l'étude des protéoformes. Deuxièmement, la spectrométrie de masse apporte une spécificité unique en comparaison des ELISA pour l'identification et la quantification des protéines.

L'analyse des protéines peut se faire selon trois approches complémentaires : le bottom-up, le middle-down et le top-down. Le bottom-up consiste à digérer enzymatiquement une protéine ce qui génère des peptides (2 kDa) qui sont ensuite analysés pour identifier la protéine. En top-down les protéines sont analysées intactes et leurs séquences sont déterminées par identification des fragments produits dans le spectromètre de masse. Le middle-down est un compromis entre ces deux approches, les enzymes ou réactifs chimiques utilisés génèrent des plus gros peptides (3-15 kDa) que ceux obtenus par bottom-up.

a) **bottom-up**

(i) **Trypsine et « shotgun » proteomique**

La trypsine est l'enzyme la plus couramment utilisée pour l'approche bottom-up du fait de sa grande spécificité, sa disponibilité et sa robustesse d'utilisation. En effet, la trypsine est toujours active en présence de solvants organique à 80%<sup>137</sup>, urée 1M et jusqu'à 60°C. Elle coupe avec une grande spécificité en C terminale des lysines et arginines sauf si une proline est l'acide aminé en c terminale : KP ou KR. Bien que la trypsine possède une grande spécificité, la fréquence de coupure est dépendante de l'acide aminé qui suit, la fréquence de digestion est plus faible en présence d'un acide glutamique ou aspartique<sup>138</sup> ce qui a pour conséquence un nombre de non-clivages plus important<sup>139</sup>(voir figure 26).

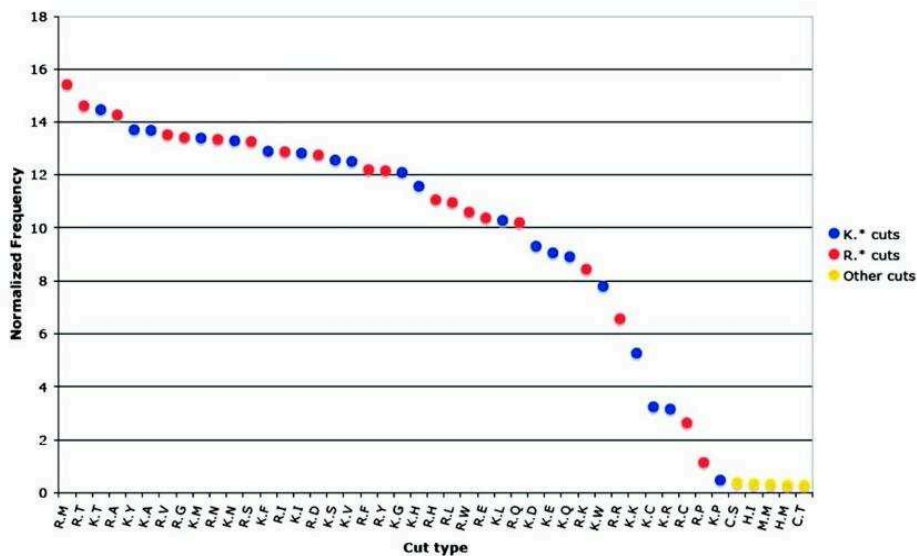


Figure 26 Fréquence normalisé des sites de coupure de la trypsine pour les 45 paires acide aminés les plus courantes. D'après Rodriguez et al. 2008



En protéomique « shotgun » les spectromètres de masse haute résolution de la famille des Q-Exactive sont actuellement les plus populaires comme l'illustre la figure 27 qui représente le nombre de jeux de données déposés dans la plateforme ProteomeXchange<sup>140</sup>. La méthode d'acquisition couramment utilisée est le data dependant acquisition (DDA).

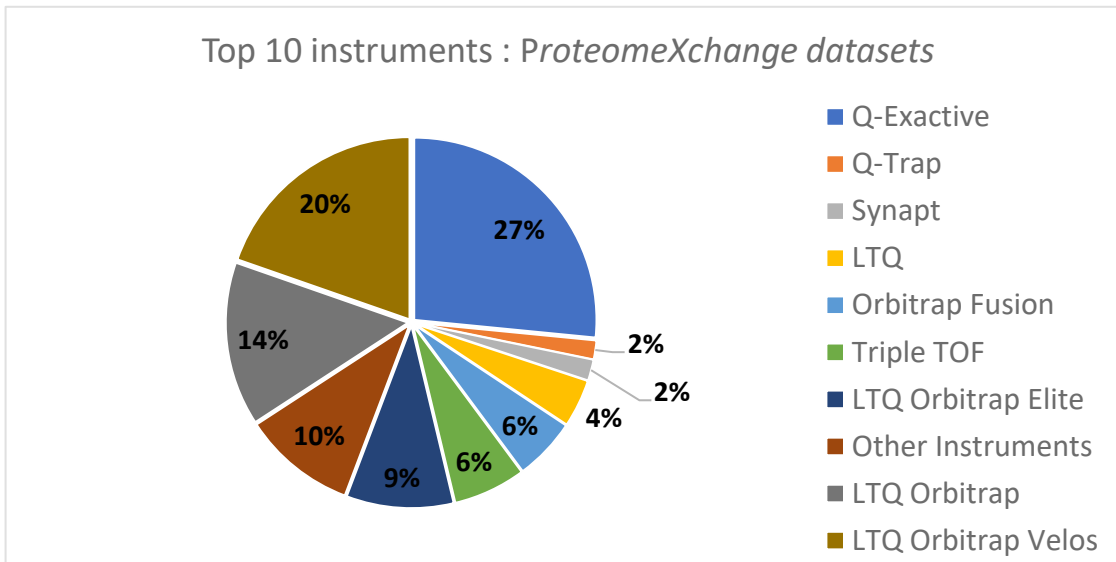


Figure 27 Représentation des 10 instruments les plus utilisés dont les données sont déposées sur ProteomeXchange

Cette méthode cyclique, comporte deux étapes par cycle :

- Un premier scan des ions peptides (MS)
- Les N ions les plus intense ions sont fragmentés par HCD ou CID (N est autour de 10) et les ions fragments sont analysés (MS/MS).

Le nombre d'ions pouvant être fragmenté entre deux scans MS des ions peptides est dépendant de la vitesse d'acquisition des instruments<sup>141</sup> (voir figure 28). Les spectres MS et MS/MS sont ensuite confrontés à des bases de données théoriques et/ou expérimentales, ce qui permet l'identification de la (ou les) protéine(s) de la (ou les) quelles il appartient. Si ce peptide est unique à une protéine alors ce peptide est appelé protéotypique. La vitesse d'acquisition est cruciale pour augmenter le nombre de MS/MS réalisées et augmenter le nombre de peptides identifiés. Kelstrup et al.<sup>141</sup> ont montré une augmentation du nombre de spectres d'ions fragmentés de 70% pour le Q-Exactive HF en comparaison du q-Exactive. Cette augmentation du nombre d'ions fragmentés résulte en un nombre de peptides identifiés plus important de 38%.

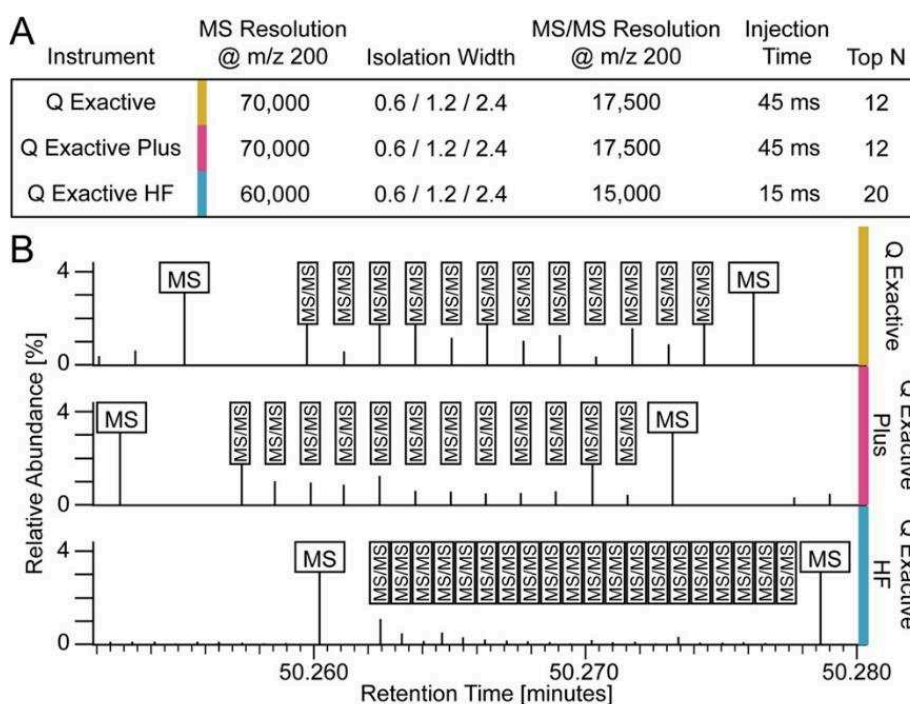


Figure 28 (A) Caractéristique des spectromètres de la famille Q Exactive, le temps d'analyse de l'Orbitrap du Q Exactive et du Q Exactive plus est de 512 ms à une résolution de 140 000 et de 512 ms à 240 000 de résolution pour le Q Exactive HF. Le Q Exactive HF à résolution équivalente est donc 2 fois plus rapide. Ce qui explique un nombre d'évènement plus important. (B) représentation des évènements MS et MS/MS en analyse DDA pour les différents Q-exactive. D'après Kelstrup et al. 2014

Cette méthode est particulièrement intéressante dans le domaine des biomarqueurs car elle permet en outre l'identification et la quantification relative des protéines (II.4) qui sont présentes dans des matrices biologiques variées (tissus ou biofluides) afin de trouver des biomarqueurs potentiels. Kim et al.<sup>142</sup>, dans l'étude la plus exhaustive, ont couvert plus de 84% des protéines encodées par 17 294 gènes à partir de tissus humains et de cellules souches hématopoïétiques. Cette étude a permis la détection de peptides spécifiques d'isoformes pour 2 861 protéines montrant la capacité de la spectrométrie de masse à détecter les isoformes.

Néanmoins, le DDA a plusieurs limitations. En effet, le DDA utilise un premier scan Full-MS qui sert à sélectionner les 10-20 ions les plus abondants ce qui biaise l'analyse envers les peptides abondants au détriment des peptides les moins abondants. De plus, cette sélection ne se fait que rarement à l'apex du pic. Tout cela a pour conséquence un sous-échantillonnage, des CV importants et donc une capacité de quantification limitée. C'est pourquoi une nouvelle méthode d'acquisition a été développée : le data independent analysis. Cette méthode d'acquisition sans a priori tend à répondre aux limitations intrinsèques du DDA<sup>143</sup> et à tirer avantages des méthodes ciblées (voir figure 29).

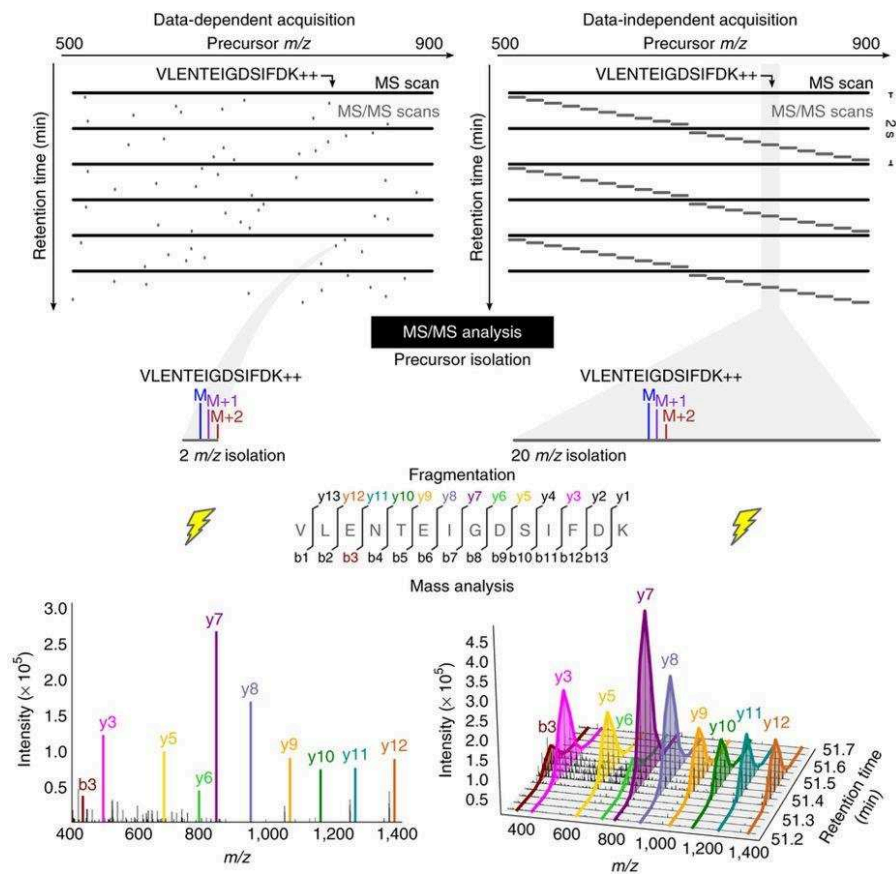
Le DIA, introduit en 2003<sup>144</sup>, enregistre cycliquement une suite de fenêtre d'isolement de taille défini pour couvrir une gamme de masse donnée, sans aucune autre sélection, où tous les ions inclus dans ces fenêtres sont fragmentés. La taille des fenêtres d'isolement impacte directement la performance de la méthode. En effet, plus les fenêtres sont larges et plus le temps de cycle va être court, ce qui peut permettre d'augmenter le temps d'accumulation des ions. Mais comme les fenêtres sont larges, plus d'ions différents sont fragmentés, ce qui a pour conséquence de complexifier les spectres de fragmentation et donc d'augmenter le nombre d'ions interférents. Venable et al.<sup>145</sup> en 2004 ont proposé une méthode DIA avec des fenêtres de 10 m/z, couvrant 1000 m/z, en utilisant un LTQ avec un gain en signal sur bruit de 3 à 5 entre une extraction MS/MS et MS. La gamme dynamique étant souvent une limite, Panchaud et al.<sup>146</sup>, ont développé PACIFIC qui consiste en une injection qui couvre seulement 15m/z avec 10 fenêtres de 2.5 m/z afin d'avoir une extraction spécifique, ce qui permet d'obtenir une gamme dynamique de 8 log. En revanche, cette méthode ne permet pas une analyse haut débit.

Les méthodes DIA actuelles comme le SWATH-MS Q-TOF consistent en 32 fenêtres de 25m/z<sup>147</sup> pour couvrir une gamme de masse de 800 m/z. L'identification des peptides se fait à partir d'une librairie de spectres qui permet d'avoir les fragments et le temps de rétention. Cette méthode a été montrée linéaire sur une gamme dynamique de 4 log et est plus sensible qu'une quantification en label free MS1 d'un facteur 2 à 10. En revanche, en comparaison avec une méthode ciblée comme le SRM le SWATH-MS est pour l'instant environ 10 fois moins sensible. Ce besoin de librairie de spectres pour le SWATH-MS a motivé le travail de Kang et al.<sup>148</sup> pour développer le SWATH-ID qui est capable d'identifier et quantifier les peptides sans avoir recours à une librairie. SWATH-ID correspond à un cycle de 4s comprenant 200 fenêtres de 3 m/z. Chaque précurseur est accumulé de 0 à 4 ms pour pouvoir remonter à l'identité du peptide. Deux autres méthodes permettent aussi de se passer de l'utilisation d'une librairie de spectres DIA-Umpire<sup>149</sup> et Spectronaut. Les librairies spectrales sont acquises en DDA, elles sont généralement spécifiques au projet et laborieuse à construire. L'analyse sans librairies en DIA permet de supprimer cette étape de générations de librairies et d'analyser l'échantillon sans a priori pour identifier des peptides pas nécessairement détectés en DDA. A noter, Zolg et al<sup>150</sup> ont réalisé la synthèse d'environ 1.4 million de peptides tryptiques permettant de couvrir la totalité du protéome humain. Ces peptides pourront à terme servir pour enrichir les librairies et améliorer les outils d'analyse.

Du fait des fenêtres larges utilisées en DIA, les spectres de fragmentation sont complexes et peuvent conduire à une mauvaise interprétation. Pour remédier à ce problème, Egertson et al.<sup>151</sup> ont exploité la capacité de multiplexage du q-Exactive pour sélectionner de manière aléatoire 5 fenêtres parmi 100 de 4 m/z chacune, ce qui a pour conséquence de diminuer le niveau des interférences. Lambert et al.<sup>152</sup> DIA a permis l'identification de 5089 peptides contre 2741 en DDA avec des CV<20% pour 82,1% des peptides en DIA contre 74,5% en DDA montrant l'intérêt du DIA par rapport à l'approche plus classique de DDA et des perspectives intéressantes pour la découverte de biomarqueurs. Une étude plus récente a montré des résultats similaires avec 2 fois plus de protéines identifiées en DIA qu'en DDA avec des CV 2 fois plus faibles<sup>153</sup>.

Bien que moins sensible que les méthodes ciblées comme le PRM et le SRM, le DIA a pour avantage de simplifier l'acquisition des données notamment dans le cas d'analyse multiplexe. De plus, les données des spectres MS1 et MS2 peuvent être analysées rétrospectivement pour extraire des données qui n'ont pas été ciblées auparavant<sup>154</sup>. Ce qui peut être particulièrement intéressant concernant l'analyse des modifications post-traductionnelles. Le DIA a notamment été appliqué à l'analyse des modifications des histones<sup>155,156</sup> et ainsi que pour l'analyse du phosphoprotéome<sup>157-159</sup>.

Dans la même idée que DIA mais pour améliorer la qualité des spectres MS1 est la méthode BoxCar qui partage un scan MS1 en plusieurs fenêtres d'acquisition restreintes ce qui permet d'augmenter le temps d'injection maximale global dans la C-trap et donc d'augmenter le signal sur bruit des spectres MS1<sup>160</sup>. La méthode BoxCar a pour but d'optimiser le temps d'accumulation des ions parents dans la C-trap avec la durée d'analyse de l'orbitrap. En général, pour un spectre MS1 en DDA le temps de stockage est autour de 1ms dans la C-trap alors que l'orbitrap met au moins 120 ms, ce qui veut dire que 119 ms sont perdues et que le signal correspondant est inutilisé d'où l'idée du BoxCar.



**Figure 29 Comparaison du DDA et DIA. En DDA la fenêtre d'isolement des ions est restreinte 2Da contrairement au DIA 20 Da. Les spectres de masse des ions fragments peuvent être extrait à un point donnée en DDA alors qu'en DIA ils peuvent être suivit temporellement . Adapté de Egertson et al. 2015.**

Ces deux méthodes DDA et DIA sont aussi grandement utilisé pour suivre des variations de niveau des modifications post traductionnelles comme les phosphorylations par exemple<sup>161-163</sup>. La quantification de ces modifications sera abordée dans II-4. De plus, afin de faciliter et d'améliorer la robustesse de l'analyse des modifications post-traductionnelles, plus de 5000 peptides synthétiques comprenant les 21 modifications post-traductionnelles les plus fréquentes, ont été synthétisés<sup>150</sup>. Ce projet prend part dans le ProteomeTools project qui a pour but de synthétiser les peptides compris dans le protéome humain. La fragmentation de ces peptides par différents modes de fragmentation ainsi que leur comportement chromatographique sont étudiées.

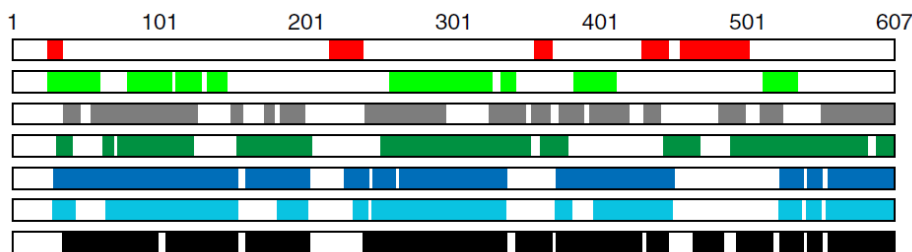
Bien que la trypsine soit toujours l'enzyme la plus utilisée, des études récentes ont pointé du doigt les limites des études protéomiques actuelles n'utilisant que la trypsine. Notamment en termes de diversité du protéome et pour l'étude des modifications post traductionnelles. En effet, comme montré par Wihelm et al.<sup>164</sup>, la couverture du protéome pour les chromosomes 21 et Y (81% et 57%) est bien en dessous des autres chromosomes (>90%). Ceci est dû au fait que le chromosome 21 contient beaucoup de protéines ne comprenant pas de peptides compatibles avec une

digestion à la trypsine et leur analyse par spectrométrie de masse. De plus, bien qu'il y ait une confirmation protéique pour 92% des gènes humains annotés dans Swiss-Prot, seulement 22% des isoformes de protéines listées dans UniProt sont détectées. L'utilisation d'autres enzymes permet d'augmenter la couverture de séquence d'une protéine et donc d'avoir des informations complémentaires à la trypsine.

**(ii) extended bottom up**

Du fait des limitations du bottom-up, en termes de couverture de séquence, de diversité du protéome accessible et des modifications post-traductionnelles détectables<sup>165,166</sup>, l'utilisation d'enzymes alternatives produisant des peptides légèrement plus gros<sup>167</sup> a été exploré ainsi que leur combinaisons<sup>168</sup>.

En effet, par exemple dans le cas de l'albumine de sérum de bovin l'utilisation de la trypsine seule permet d'avoir une couverture de séquence de seulement 78.4%. En combinant, cette dernière avec 6 autres enzymes (l'ArgC, l'AspN, chymotrypsine, GluC, LysC et la LysN), la couverture de séquence cumulée est de 94% (voir figure 30), montrant l'intérêt potentiel de combiner des enzymes pour les analyses protéomiques<sup>169-171</sup>. La spécificité des enzymes est indiquée tableau 6. La combinaison des enzymes est essentielle car individuellement la GluC, l'AspN, la LysC ou l'ArgC conduisent à moins de peptides identifiés que la trypsine<sup>167</sup>.



**Figure 30** Séquence de l'albumine bovine (BSA), les zones colorées correspondent aux peptides détectés avec différentes enzymes. Dans l'ordre par (avec la couverture de séquence obtenue) : ArgC: 18.3%, AspN: 38.1%, chymotrypsin: 57.8%, GluC: 61.9%, LysC: 70.8%, LysN: 60.8% and trypsin: 78.4%. D'après Giansanti et al. 2015

Nagaraj et al.<sup>172</sup> ont démontré l'intérêt de combiner la trypsine, Glu-C et Lys-C pour augmenter le nombre de protéines identifiées et la couverture du protéome de cellules HeLa. 85% des protéines identifiées par la trypsine et la Lys-C sont communes, la Glu-C apporte seulement 5% de protéines identifiées en plus.

Tableau 6 Les 6 protéases les plus communes en bottom-up avec leurs spécificités ainsi que leurs avantages et inconvénients. D'après Giansanti et al 2015

Protéase	Site de coupure	Avantages	Inconvénients
ArgC	C-terminale des R	Combinaisons avec d'autres enzymes pour l'étude des PTMs et augmentation de la couverture du protéome	Coupure avec une moins grande efficacité aussi en C-terminale des K
AspN	N-terminale des D	Fonctionne à pH acide et basique 4-9	Sensible à la présence de détergents, qui a pour conséquence des coupures non spécifiques en N-terminale des E Peu spécifique envers les E
Chymotrypsine	C-terminale des F,Y,L,W et M	Le protéome couvert est orthogonal à celui couvert par les peptides tryptiques	Les sites de coupures sont peu spécifiques
GluC	C-terminale des D et E	Combinaisons avec d'autres enzymes pour l'étude des PTMs et augmentation de la couverture du protéome	La spécificité de l'enzyme est dépendante du pH Peu spécifique. Sites de coupures
LysN	N-terminale des K	Robuste (dénaturants et température (70°C))	supplémentaires sont A, S et R
LysC	C-terminale des K	Très efficace et spécifique Particulièrement robuste aux dénaturants comme l'urée (8M)	Peu de différence avec la trypsine en termes de couverture du protéome

Une enzyme moins classique telle que la prolyl-endopeptidase (neprosine) a été proposée par Schröder et al<sup>173</sup>. La neprosine coupe préférentiellement les protéines après les prolines et alanines. La neprosine apporte ainsi 50% de couverture de séquence supplémentaire par rapport à la trypsine, ce qui a un intérêt pour chercher de nouvelles modifications post-traductionnelles (voir figure 31).

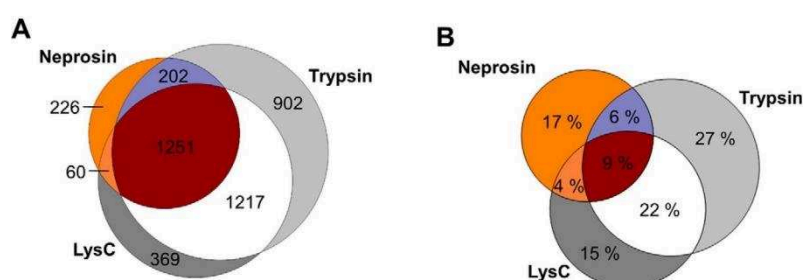


Figure 31 Analyse protéomique bottom-up d'un lysat cellulaire HeLa par digestion à la neprosine, trypsine et Lys-C. (A) Diagramme de Venn des protéines identifiées (B) diagramme de Venn de la couverture de séquence pour les 1251 protéines identifiées par les trois enzymes. D'après Schröder et al. 2017

## b) middle-down

La taille des peptides tryptiques est généralement < 2 kDa ce qui peut conduire à des difficultés à attribuer des peptides à une protéine unique car ils ne sont pas protéotypiques<sup>174</sup>. Le top-down permet de résoudre ce problème mais est moins efficace pour les protéines de haute masse. Le middle-down est un compromis entre les deux approches avec une gamme de masse comprise entre les deux approches 3-20 kDa.

Néanmoins, les limites concernant la taille des peptides définissant le middle-down n'est pas claire. De plus les protéases produisent une diversité de tailles de peptides dépendante du type de protéine (cf. histone), donc la classification d'une enzyme dans une catégorie (bottom-up, extended bottom-up ou middle-down) n'est pas évidente. Laslkay et al<sup>175</sup>. ont proposé de différencier ces catégories en prenant en compte le type de colonnes chromatographiques utilisées, le spectromètre de masse, le type de fragmentation et les moyens informatiques mis en place pour traiter les données. Ces considérations sont résumées dans la figure 32 ci-dessous :



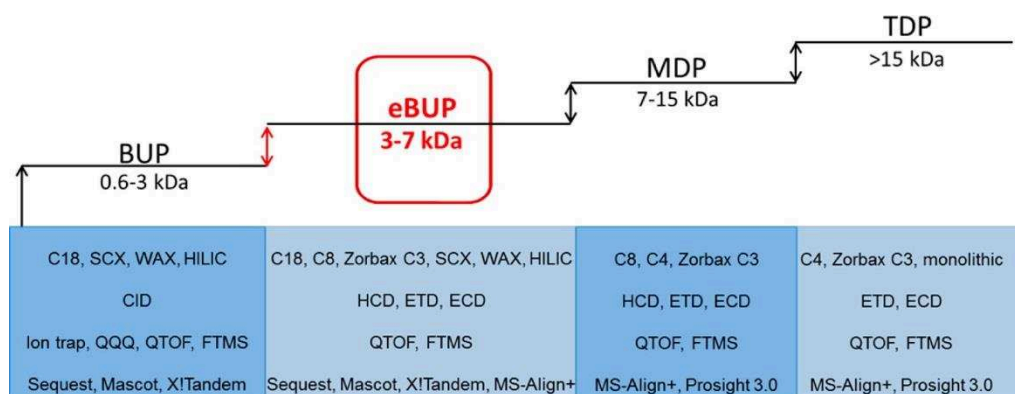


Figure 32 Présentation des approches protéomiques, BUP (bottom-up), eBUP (extended Bottom-up), MDP (middle-down) et TDP (top-down) avec les méthodes de séparation chromatographique, les méthodes de fragmentation, l'instrumentation ainsi que les logiciels de traitement adaptés. D'après Laskay 2013

Des enzymes avec de nouveaux sites de coupures non conventionnels (autre que K et R) ont dû être identifiées. Comme par exemple, les protéases dibasiques comme kex2<sup>176</sup> ou Outer membrane protease (OmpT) qui produisent des peptides de 6.3kDa en moyenne, coupent entre deux acides aminés dibasiques<sup>89</sup>. Une autre protéase dibasique est la sap9 qui génère des peptides de 3,2 kDa en moyenne<sup>177</sup> avec un nombre significatif de peptides dans la gamme de masse de 3 à 7 kDa. Cette enzyme a aussi été utilisée avec succès pour l'analyse d'anticorps monoclonaux<sup>178</sup>. Les outils de bio-informatique<sup>179</sup> ont mis en évidence l'intérêt du middle-down pour des coupures moins classiques comme en C-terminale des tryptophanes, cystéines ou des méthionines avec une couverture du protéome respectivement de 93,57%, 97,25% et 99,85%. Des alternatives chimiques ont été trouvées pour réaliser ces coupures comme le BNPS-Skatole pour le tryptophane avec une taille moyenne des peptides de 6,1 kDa, le NTCB pour les cystéines (4,3kDa) et le CNBr pour les méthionines (5,7 kDa)<sup>180,181</sup>. Ces agents sont des alternatives crédibles aux protéases usuellement utilisées mais le CNBr convertit la méthionine en un mélange d'homosérine et d'homosérine lactone.

Un cas particulier du middle-down est l'analyse des histones. Les histones constituent les éléments de base de la chromatine, elles sont au nombre de 8 (H3, H4, H2A et H2B) avec une histone supplémentaire H1 dite de liaison qui entoure l'ADN. La partie N-terminale des histones se situe à l'extérieur du nucléosome et contient de nombreuses modifications post-traductionnelles<sup>182</sup> comme par exemple les lysines peuvent être acétylées et mono-di ou tri méthylées. Ces modifications post-traductionnelles sont impliquées dans la régulation de l'état transcriptionnel de la chromatine et la combinaison de ces modifications est appelée le Code des histones. Le

« crosstalk » entre les modifications, c'est-à-dire qu'une modification peut entraîner un changement d'abondance d'une autre modification est important pour la régulation de l'état transcriptionnel de la chromatine. Le « crosstalk » peut modifier l'état transcriptionnel comme par exemple en supprimant l'interaction d'une enzyme avec la queue de l'histone<sup>183,184</sup> ou au contraire en catalysant leur interaction<sup>185</sup>. Les histones sont des protéines très basiques comprenant de nombreuses arginines et lysines, l'utilisation de la trypsine conduit donc à la production de petits peptides mal retenus sur les colonnes chromatographiques en phase inverse C18. Une solution consiste à dériver les lysines avec de l'anhydride propanoïque ce qui conduit à la production de peptides de 4 à 20 acides aminés rendant la quantification des protéoformes des histones possibles à condition de marquer aux isotopes stables l'anhydride propanoïque<sup>186</sup>. Une autre alternative, l'utilisation de la glu-C<sup>187,188</sup> ou AspN permet d'avoir les queues des histones intactes (50-60 aa), ce qui correspond au domaine de masse du middle-down. Les approches middle-down ont permis de suivre l'évolution des modifications post-traductionnelles N-terminales au cours de l'évolution de *C. elegans*<sup>189</sup>, identifier le rôle de nouvelles modifications post-traductionnelles comme par exemple H3K23me<sup>2190</sup> ainsi que d'identifier 233 protéoformes de l'histone H4 au cours du cycle cellulaire de cellules cancéreuses<sup>191</sup>.

### **c) Top-down**

Le top-down se limite, pour l'instant, aux protéines majoritaires car l'ionisation par électrospray des protéines et leurs détections sont moins efficaces que pour les peptides<sup>192</sup>. Des particularités concernant la préparation d'échantillons, au niveau instrumentation ou au traitement des données sont aussi à prendre en considération.

Malgré ces limitations, le top-down est la seule méthode permettant de caractériser les protéoformes d'une protéine sans ambiguïté en permettant une vue d'ensemble directe des

protéofomes<sup>193</sup> comme illustré dans la figure 33. En effet, avec les approches classiques par digestion enzymatique le lien entre les peptides et leurs protéofomes d'origine est perdu.

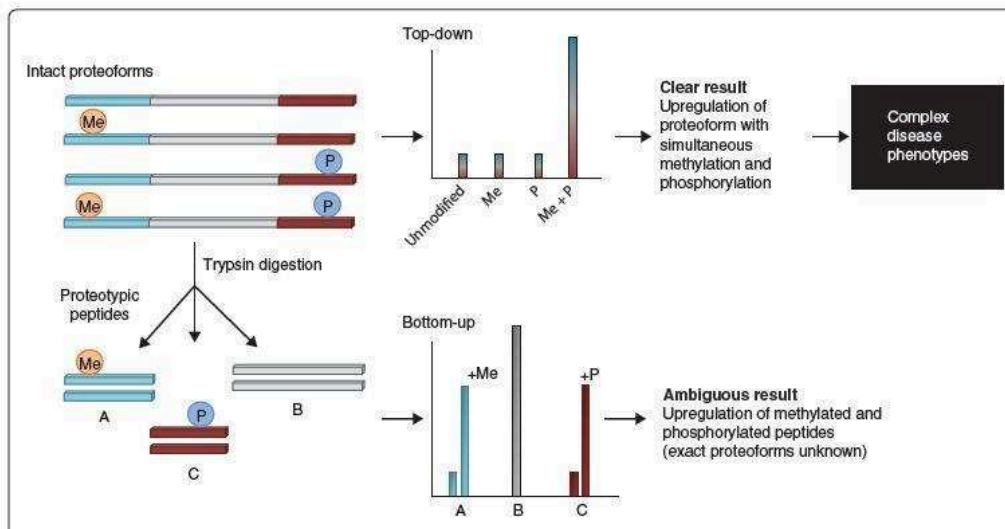


Figure 33 Le Top-down permet de caractériser sans ambiguïté les protéofomes d'une protéine et ils peuvent être corrélés avec les phénotypes d'une maladie. Par exemple, on peut observer une augmentation d'un protéofomes particulier comme une méthylation et une phosphorylation simultanée. En revanche, le bottom-up ne permet pas de mettre en évidence cette méthylation et phosphorylation simultanée. En effet, l'information est perdue au cours de la digestion. D'après Savaryn et al . 2013

**(i) Instrumentation :**

L'identification se fait sur la base de la MS1 et MS2. Bien qu'une résolution isotopique ne soit pas forcément nécessaire pour déconvoluer le signal à partir du spectre MS1, une résolution d'au moins 50000 est généralement requise pour pouvoir déconvoluer et attribuer les ions fragments MS2. Les ions fragments permettent de valider l'identité de la protéine et leurs spectres de masse peut être complexe. C'est pourquoi les spectromètres de masse de choix sont les TOF ou les instruments à transformée de Fourier. L'identification de la protéine se fait sur la base de sa masse et de ses fragments. Une bonne couverture de séquence et une bonne répartition des fragments sont donc primordiales notamment pour la caractérisation des protéofomes<sup>194</sup> comme illustré par la figure 34.

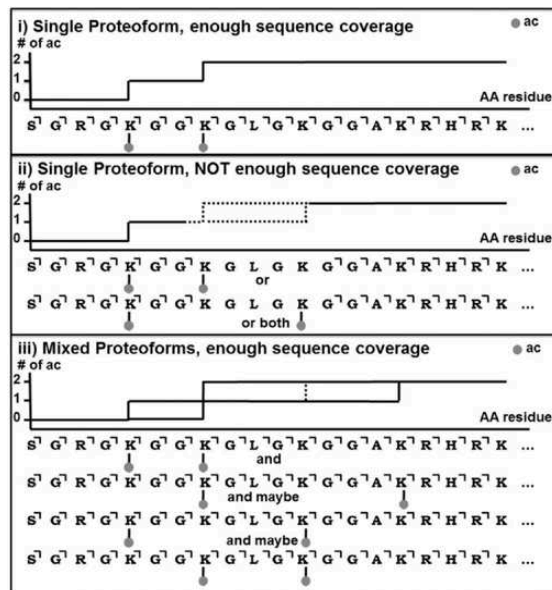


Figure 34 La couverture de séquence des protéines est essentielle pour éviter les mauvaises attributions sur la position d'une modification post-traductionnelle et donc des protéoformes. (i) dans le cas d'un seul proteoforme et avec une couverture de séquence de 100% il n'y a pas d'ambiguïté dans l'attribution du protéoforme H4K5acK8ac (ii) dans le cas de deux protéoformes H4K5acK8ac et H4K5acK12ac si la couverture de séquence n'est pas suffisante il y a une ambiguïté dans la caractérisation des protéoformes. (iii) Dans le cas de spectres provenant de plusieurs protéoformes malgré une couverture de séquence totale il y a une ambiguïté sur le nombre de protéoformes et l'acétylation de la lysine en K12. D'après Dang et al. 2014

## (ii) Fragmentation :

La qualité de la fragmentation des protéines est un élément déterminant pour leur caractérisation. Les modes de fragmentations classiquement utilisé pour les peptides sont le CID et le HCD. La fragmentation CID dans les pièges à ions et la fragmentation de type HCD (activation par collision) ont pour sites de fragmentation préférentiels en N-terminale des prolines et en C-terminale des acides aminés acides<sup>195</sup>, mais du fait d'une plus haute énergie de dissociation utilisé en HCD ces coupures sont moins spécifiques. C'est pourquoi, la fragmentation de type HCD s'est montrée plus informative que la fragmentation CID dans les pièges à ions pour l'alpha-synucléine<sup>196</sup>. De même que le CID dans les quadrupoles pour l'ubiquitine<sup>197</sup>.

Afin d'accroître la couverture de séquence des protéines d'autre mode de fragmentations ont été développés. Comme par exemple les modes de fragmentations basés sur le transfert d'électrons de basses énergies (< 1eV) comme l'ECD et l'ETD ou basée sur l'absorption d'un photon par la protéine : UVPD (ultraviolet photodissociation) qui utilise un laser ultraviolet à 193 nm.

- L'ECD correspond à la capture d'un capture électrons de basse énergie par un ion multichargé ce qui conduit à la fragmentation. En ETD, un électron est transféré vers un ion multichargé par interaction entre cet ion et un anion radicalaire. Ces deux méthodes produisent des fragmentations du squelette peptidique de type N-C $\alpha$ . L'inconvénient de ces méthodes est que le rendement de fragmentation est plus bas que celui obtenu par HCD. L'efficacité de fragmentation des protéines de l'ETD peut être augmenté par activation des ions avec un laser infrarouge<sup>198</sup>. De plus, l'ETD sur les spectromètres de masse hybride peut être couplé à l'HCD par exemple, les ions fragments produits par l'ETD sont alors fragmentés en HCD, ce qui permet d'augmenter la diversité des fragments produits.
- L'UVPD permet d'avoir un nombre plus important de fragments plus uniformément réparti sur la protéine en comparaison avec le CID, HCD et ETD comme montré par Cannon et al.<sup>199</sup> En effet, dans le cas de la protéine fluorescente verte la couverture de séquence est de 23,7% CID ; 27,1% HCD ; 38,1% ETD et 61,4% pour l'UVPD.

Un des inconvénients du CID et du HCD est qu'ils peuvent résulter en la perte de modifications post-traductionnelles labiles comme les phosphorylations<sup>200</sup>, le spectre de masse est alors dominé par les pertes de neutre HPO<sub>3</sub> ou H<sub>3</sub>PO<sub>4</sub> qui entre en compétition avec la fragmentation du squelette peptidique. Ce qui conduit à des erreurs d'interprétation sur la position du groupement phosphate ou sur la séquence du peptide<sup>201</sup>. En comparaison l'ETD, l'ECD et l'UVPD<sup>200</sup> ne fragmentent pas le groupement phosphate et donc l'information sur sa position est conservée. L'ETD, l'ECD et l'UVPD sont donc particulièrement intéressants pour l'étude des protéoformes. En revanche, le rendement de fragmentation de l'ECD et ETD est généralement moins bon et dépendant de l'état de charge fragmenté<sup>202</sup>.

Les modes de fragmentations exposés dans cette partie sont complémentaires du fait des mécanismes de fragmentation différents. Cette complémentarité a été montrée dans une étude globale sur des cellules HeLa, l'UVPD a montré une augmentation de la couverture de séquence jusqu'à 74% et du niveau de caractérisation des protéoformes de 9% en plus en comparaison avec l'HCD<sup>203</sup>.

### **(iii) Préparation d'échantillon :**

L'étape de préparation d'échantillon est cruciale pour couvrir au mieux le protéome et doit être adaptée à l'analyse des protéines intactes car les méthodes classiquement utilisés en bottom-up ne s'appliquent que dans des cas précis en top-down.

Pour les analyses non-ciblées : Tran et al.<sup>204</sup>, ont développé une méthode de séparation des protéines intactes basée sur leur poids moléculaire, car orthogonal à l'hydrophobicité et à la charge, appelé, Gel-Eluted liquid Fraction Entrapment Electrophoresis (GeLFrEE). Afin de pouvoir détecter simultanément des protéines de bas et haut poids moléculaires, deux dimensions supplémentaires ont été ajoutée par le groupe de Kelleher<sup>205</sup>avec une séparation basée sur le point isoélectrique des protéines (Focalisation isoélectrique en solution) avant le GeLFrEE et une séparation par chromatographie liquide en phase inverse après collection des fractions. Cette stratégie a été développée et appliquée par Tran et al. en 2011, identifiant plus de 3000 protéoformes appartenant à 1043 protéines provenant de cellules HELA. Les études récentes permettent l'identification de plus de 5000 protéoformes pour environ 1000 protéines<sup>206,207</sup>. D'autres méthodes de fractionnement ont aussi été appliquées avec succès comme par exemple la chromatographie d'exclusion stérique<sup>208</sup>. Concernant les méthodes ciblées l'immunoprécipitation des protéines restent la méthode la plus sensible et de choix pour l'analyse des protéines par top-down car c'est la méthode la plus sélective<sup>209,210</sup>.

#### **(iv) Séparation chromatographique :**

Des techniques séparatives particulières doivent aussi être mises en place pour l'analyse des protéines intactes. En effet les protéines ont tendances à accrocher dans les colonnes chromatographiques classiquement utilisées pour le bottom-up. Des méthodes orthogonales à la RPLC ont été évaluées pour l'analyse en top-down comme la chromatographie d'interaction hydrophile (HILIC)<sup>211</sup>, la chromatographie d'exclusion stérique<sup>212</sup>, chromatographie à échange d'ions<sup>213</sup>, la chromatographie d'interaction hydrophobe<sup>214</sup> et l'électrophorèse capillaire de zone<sup>207</sup>.

#### **(v) Analyse :**

Les analyses protéomique par top-down comme exposé ont permis la détection de plus de 5000 protéoformes. En revanche, une des limitations principales de ces approches est la limite en taille des protéines détectées généralement inférieures à 30 kDa pour plusieurs raisons : vitesse d'acquisition basse des spectromètres de masse, rapport signal sur bruit peu élevé, fragmentation des protéines moins efficace et augmentation de la complexité des spectres de masse. Pour pallier à ses inconvénients Anderson et al.<sup>215</sup>, ont utilisé un 21T FT ICR combiné à une séparation des protéines par GeLFrEE et une séparation par chromatographie en phase inverse pour l'identification de plus de 3284 protéoformes pour 684 protéines dont 372 protéoformes ayant des masses supérieures à 30kDa dans des cellules cancéreuses humaines.

Le top-down peut aussi être appliqué à l'étude d'une protéine (ciblée) et la caractérisation de ses protéoformes. En effet, le top-down apporte un niveau d'information supplémentaire au bottom-up comme c'est le cas pour la troponine cardiaque I humaine<sup>216</sup>. C'est une protéine de 24 kDa dont l'hyperphosphorylation est associée à une dysfonction cardiaque. Cette protéine comporte 23 sites de phosphorylations potentiels. Zabrouskov et al., après immunoprécipitation à partir de tissus cardiaques, ont démontré l'existence de 36 protéoformes. Ces protéoformes sont des combinaisons de tronctions, mono ou di-phosphorylation et oxydation. Plus récemment la troponine cardiaque T porcine a été caractérisé par Lin et al. du fait de l'utilisation croissante des porcs nains comme modèle animale pour les dysfonctionnements cardiaques. Montrant ainsi l'existence de 7 isoformes provenant d'épissage alternatif qui peuvent être monophosphorylés.

Les approches bottom-up, middle-down et top-down permettent d'identifier et de caractériser les protéines et les protéoformes provenant d'un échantillon complexe dans le but d'établir une liste de protéines pouvant être des biomarqueurs. Afin, d'évaluer le potentiel d'une protéine comme biomarqueur que ce soit pour sa découverte ou pour sa validation, des méthodes de quantification doivent être mise en place.

#### **4. Les approches de quantification par spectrométrie de masse**

Les approches protéomique exposées précédemment sont complémentaires et ont pour but de permettre la détection et la quantification des peptides/protéines. Les approches de quantification par spectrométrie de masse peuvent être classées en deux catégories : la quantification relative et la quantification absolue. Les méthodes de quantification relative sont principalement utilisées durant la phase de découverte des biomarqueurs alors que les méthodes de quantification absolue plus précise sont utilisées dans les phases ultérieures, c'est-à-dire pour la validation des biomarqueurs et en clinique. Ces deux approches utilisent soit le marquage isotopique (dilution isotopique) pour quantifier les protéines soit sans l'utilisation d'isotope (label free). Les différentes approches de quantification qui vont être abordées sont résumées dans la figure 35 ci-dessous :

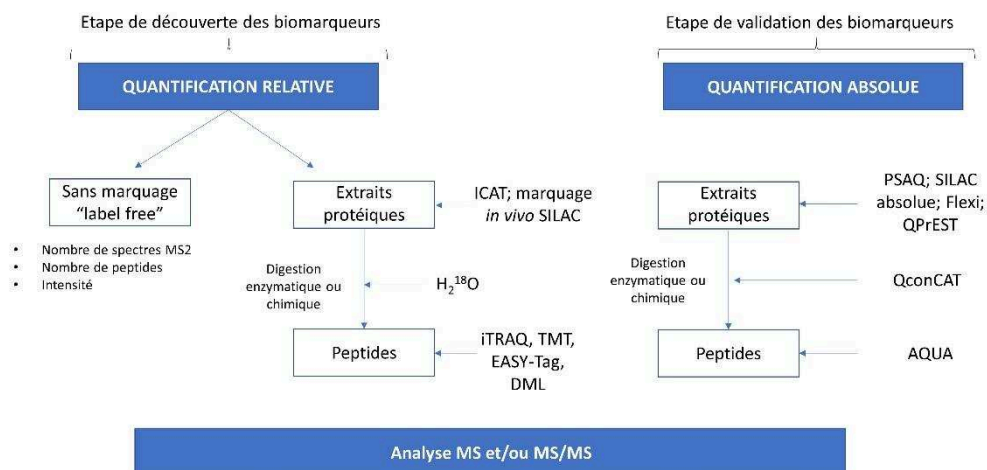


Figure 35 Les principales méthodes de quantifications utilisées en spectrométrie de masse

### a) Quantification relative par spectrométrie de masse

Les méthodes de quantification relative par spectrométrie de masse sont utilisées pour la découverte de biomarqueurs et se font majoritairement par bottom-up. Bien que le top-down soit une approche émergente (II-3). Ces méthodes utilisent essentiellement le data-indépendant acquisition qui permet l'identification et la quantification des peptides. La quantification se fait soit sans marquage « label free » ou par l'ajout d'une modification aux peptides ajoutant un incrément de masse.

#### (i) Label free

Les méthodes « label free » n'utilisent pas de peptides marqués aux isotopes stables ni d'ajout de modifications pour la quantification. La quantification relative se fait par la comparaison d'abondances des protéines de deux manières soit en se basant sur les données d'identification soit sur l'intensité des ions.

#### (a) Résultats d'identification

Nombre de peptides : Rappsilber et al.<sup>217</sup> ont proposé une méthode de comparaison d'abondance des protéines basée sur le ratio entre le nombre de peptides observés divisé par le nombre théorique de peptides tryptiques observables dans une gamme de masse donnée. Le Protein Abundance Index (PAI), cette méthode permet de prendre en compte la taille de la protéine et donc de rendre la comparaison entre différentes protéines possible. Ishihama et al.<sup>218</sup> ont proposé une version exponentielle du PAI égale à  $10^{\text{PAI}}$  moins un, basé sur



l'observation d'une corrélation entre le PAI et le logarithme de la concentration d'une protéine. Plus récemment, Sun et al.<sup>219</sup>, dans le but de prendre en compte la préparation d'échantillon, la séparation et l'analyse par spectrométrie de masse : le mSCI (modified spectral counting index) qui est égal au nombre de peptides observé divisé par RIPpro (protein relative identification possibility). Le RIPpro est calculé en considérant le poids moléculaire, le point isoélectrique et l'hydrophobicité de la protéine.

Nombre de spectres : Nombre de spectres MS2 associés à une protéine associant les informations redondantes d'identification d'un peptide comme par exemple les différents états de charge ou une nouvelle détection du peptide du fait de la fin de l'exclusion dynamique. La corrélation entre le nombre de spectres MS2 et la concentration de protéine, a été premièrement démontré par Liu et al.<sup>220</sup> permettant de voir des changements d'abondances de manière linéaire sur 2 log. Deux outils APEX<sup>221</sup> (absolute protein expression) et NSAF<sup>222</sup> (normalized spectral abundance factor) permettent la quantification à grande échelle de protéines en prenant en compte leur taille. Ces méthodes sont limitées par des effets de saturation à haute concentration. C'est pourquoi des variantes ont été proposées comme le MS2-TIC<sup>223</sup> qui combine le nombre de spectres MS2 avec la moyenne du TIC pour chaque spectre ou le normalized Spectral Index (SIn)<sup>224</sup> qui combine le nombre de peptides identifiés, le nombre de spectres et l'intensité des ions fragments.

(b) Les méthodes basées sur l'extraction de l'intensité du signal pour une gamme de m/z : XIC (Extracted ion chromatogram)

Deux étapes de pre-processing sont nécessaires pour les méthodes XIC. En effet, pour minimiser les décalages, les temps de rétentions doivent être alignés entre les analyses afin de faciliter l'identification des peptides et l'intensité du signal doit être aussi normalisée pour diminuer les biais. Il existe deux manières de déterminer l'abondance d'une protéine soit en moyennant soit en sommant les XIC des peptides. Pour la moyenne deux possibilités : moyenne de tous les peptides<sup>225</sup> ou la moyenne des trois peptides les plus intenses de chaque protéines<sup>226</sup>, ces méthodes sont répétables avec des CV < 10%. Pour la somme, généralement la somme des trois peptides les plus intenses est réalisée<sup>227</sup> et une étape de normalisation est ajoutée en normalisant soit par la taille des protéines soit par la somme des intensités de tous les peptides. Une alternative décrite par Schwanhausser et al.<sup>228</sup>, est la méthode iBaq (intensity based absolute quantification) où la somme de tous les peptides est divisée par le nombre de peptides théoriques observables. Cette méthode a été validée avec un mélange de 48 protéines standards montrant un R<sup>2</sup> de 0,88 pour la protéine UPS2.

L'avantage des méthodes label-free est leur relative facilité à mettre en œuvre par rapport aux méthodes de dérivation des peptides avec des tags. Le second avantage est que le nombre d'échantillons pouvant être comparé n'est pas limité comme pour les méthodes nécessitant un marquage qui vont être exposées par la suite. L'inconvénient de la quantification par label-free est qu'aucune étape, de la préparation d'échantillon à l'analyse par spectrométrie de masse, n'est normalisée et est donc la méthode la moins précise.

Les méthodes de quantification relative par marquage isotopique permettent d'améliorer la fidélité. Elles sont basées sur l'ajout d'un tag ou l'incorporation d'isotope stable comme le  $^2\text{H}$ ,  $^{18}\text{O}$ ,  $^{15}\text{N}$  ou  $^{13}\text{C}$ . L'incorporation d'isotope stable permet de conserver les propriétés physico-chimiques du peptide entre la version légère et la version lourde. Selon l'étape d'incorporation, l'extraction, la digestion et les étapes de préparations d'échantillons peuvent être normalisées.

#### **(ii) Marquage isotopique SILAC (ou marquage métabolique)**

Dans le SILAC les cellules se développent dans un milieu de culture manquant des acides aminés essentiels. Une version marquée aux isotopes stables ( $^{13}\text{C}$ ;  $^{15}\text{N}$ ;  $^2\text{H}$ ) de ces derniers est ajoutée dans le milieu de culture. Introduit en 2002 par Ong et al.<sup>229</sup>, le milieu de culture était dépourvu de leucine. De la leucine deutérée a donc été incorporée dans le milieu de culture et a permis la quantification relative de protéines impliquées dans la différenciation de cellules musculaires. Les limites de cette approche sont que tous les peptides tryptiques ne contiennent pas une leucine et ne sont donc pas marqués. De plus le deutérium induit un décalage en temps de rétention<sup>230</sup>. Pour pallier au problème de marquage des leucines, les acides aminés les plus couramment utilisés, maintenant, sont la lysine et l'arginine. En effet, dans la majorité des cas la digestion enzymatique est réalisée avec la trypsine qui laisse l'un de ces deux acides aminés en C-terminal du peptide. Le marquage est alors présent sur la majorité des peptides (voir figure 36). D'autres acides aminés

ont aussi employés comme les sérines, glycines, méthionines, valines, leucine, histidines et tyrosines<sup>231</sup>.

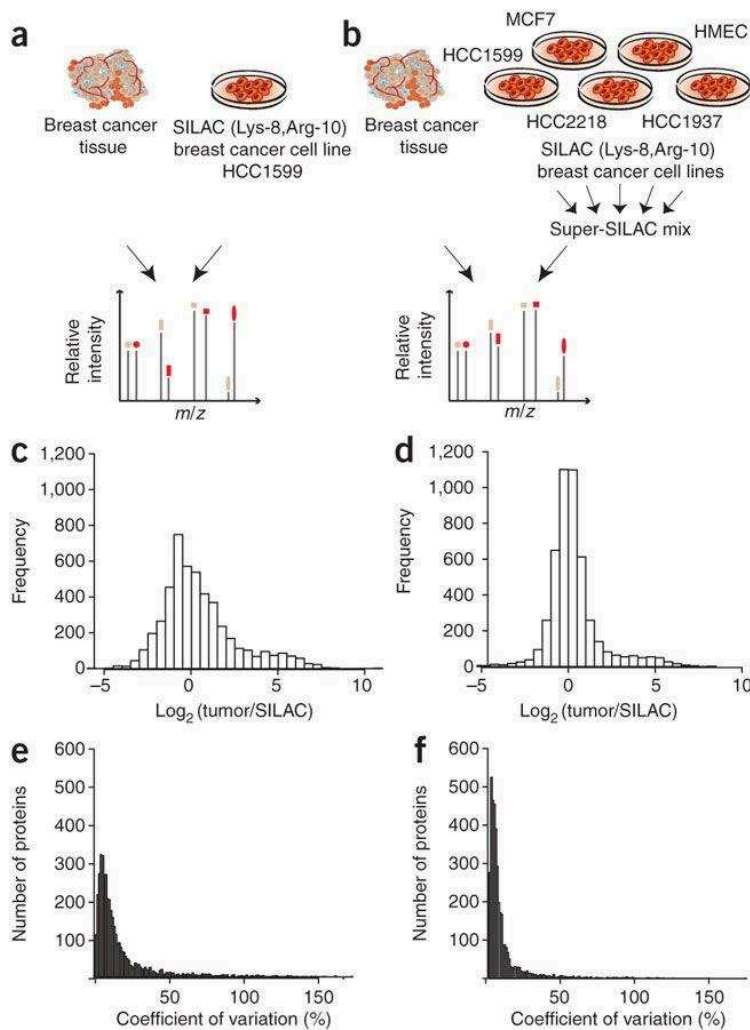


Figure 36 Principe des méthodes SILAC (a) et super SILAC (b). Les cellules marquées sont utilisées dans ce cas comme standard pour comparer des échantillons. (c,d) chaque colonne représente un nombre de protéines. En combinant 5 types cellulaires (super SILAC) la tumeur est mieux représentée en terme de diversité de protéines. De plus (e,f) le super SILAC permet l'identification de plus de protéines avec des CV inférieurs au SILAC. D'après Geiger et al. 2010

Le SILAC peut être multiplexé 3-plex<sup>232</sup> ce qui permet de comparer 2 conditions avec une condition standard ou jusqu'à 5-plex<sup>233,234</sup>.

Pour la quantification en tissus, des modèles animaux peuvent aussi avoir un régime spécial contenant des acides aminés marqués (souris, drosophile, nématode et poisson-zèbre)<sup>235</sup>. Mais la production d'animaux marqués demeure chère et laborieuse, ce qui limite les applications.

Le SILAC est limité aux échantillons pouvant être marqué métaboliquement mais peut être étendu aux tissus ou fluides biologiques en utilisant l'échantillon marqué comme un standard pour la quantification (Spike in)<sup>236</sup>. Pour pouvoir quantifier des tissus humains et notamment des tumeurs, cette méthode a été étendue et appelé Super-SILAC, en combinant 5 lignées cellulaires différentes afin de représenter au mieux la diversité des tissus<sup>237</sup>(voir figure 36).

Plus récemment, exploitant l'avantage de la haute résolution apporté par les spectromètres de masse Hebert et al.<sup>238</sup>, on introduit le neutron-encoded Neucode SILAC. Neucode est un marquage métabolique des cellules comme le SILAC mais plutôt que d'utiliser des isotopes séparés d'au moins 4 Da, NeuCode exploite la différence de masse existante entre <sup>13</sup>C (+1,0034), <sup>2</sup>H (+1,0062) et <sup>15</sup>N (+0,997) pour créer des isotopologues de la lysine (figure 37) séparés au minimum de 6mDa. Ces isotopologues sont indifférenciables à un pouvoir de résolution <100 000. Le Neucode a l'avantage de pouvoir atteindre un multiplexage de 30-plexes. Ce marquage a été appliqué avec succès pour la quantification en top-down de plus de 500 protéines de levure (*Saccharomyces cerevisiae*) avec deux isotopologues séparés de 36mDa<sup>239</sup>.

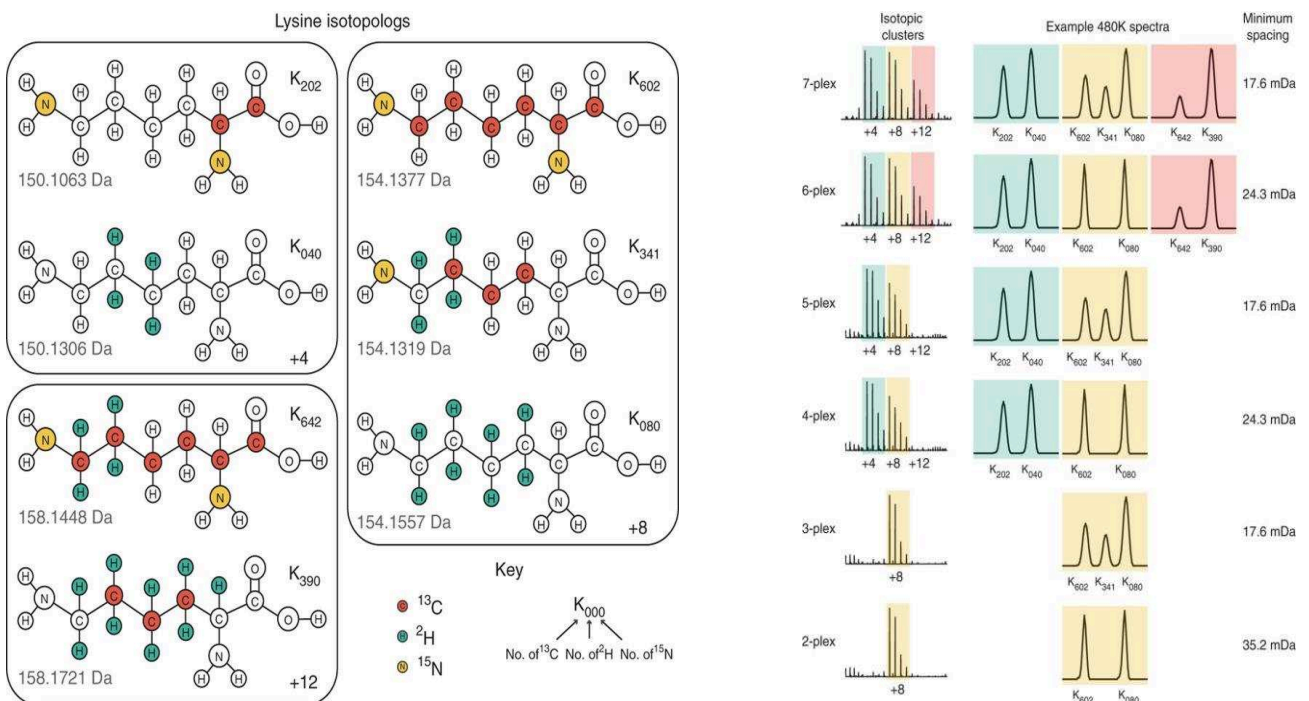


Figure 37 isotopologues de la Lysine commercialement disponibles (gauche) et exemple de spectre de masse pour la quantification multiplexe utilisant des lysines isotopologues

### (iii) Marquage chimique

#### (a) Isotope-coded affinity tag reagents (ICATs) :

Les réactifs ICATs ont été introduits en 1999 par Gygi et al.<sup>240</sup> (voir figure 38) et désignés pour se lier aux groupements thiol des cystéines au niveau des peptides. Ils sont composés de trois parties : un tag (ou étiquette) d'affinité qui est la biotine, un élément de liaison non marqué ou marqué aux isotopes stables (<sup>2</sup>H ou <sup>13</sup>C) et un groupement réactif qui réagit avec les thiols libres des cystéines. Après digestion, les peptides sont concentrés sur des colonnes de streptavidine, ce qui permet aussi de réduire par un facteur 10 le nombre de peptides<sup>241</sup>.

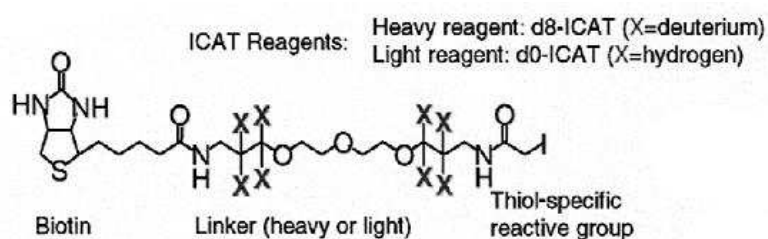


Figure 38 Représentation du réactif ICAT. D'après Gygi et al. 1999

La quantification relative de cellules saines et malades se fait par comparaison de l'intensité des peptides entre peptides marqués et non marqués.

Les désavantages concernant l'ICATs sont que la biotine interfère pour l'identification des peptides par les banques de données et rend les peptides plus hydrophobes, ce qui a pour conséquence qu'ils éluent avec un temps de rétention proche. Ce problème a été résolu par Li et al., en introduisant un élément de liaison clivable par l'acide permettant d'enlever la biotine. De plus l'éluion des peptides biotinilés n'est pas quantitative. Finalement, seuls les peptides ayant une cystéine peuvent être quantifiés, ce à quoi Pan et al., ont proposé une alternative en produisant un réactif ICAT qui catalysé par la trypsine se greffe spécifiquement en N terminale des peptides.

Sur le même principe Schimdt et al.<sup>242</sup>, ont développé un tag (ou étiquette) qui se lie aux fonctions amines libres présentes sur les protéines et n'est donc pas spécifique d'un acide aminé. Comme le marquage se fait au niveau de la protéine intacte, les conditions testées peuvent être rassemblées à une étape précoce de la préparation d'échantillon. Ce tag est appelé Isotope-coded protein labels (ICPL). Ce marquage a été appliqué à la quantification relative ciblée de 6 protéines

impliquées dans la protection cellulaire dans une lignée cellulaire humaine, afin de mettre en évidence des composés alimentaire augmentant l'abondance de ces protéines<sup>243</sup>.

(b) Tag isobarique :

Deux tags fonctionnant sur le même principe ont été développés le TMTs (tandem mass tags) par Thompson et al.<sup>244</sup> en 2003 (commercialisé par Thermo Scientific) et le iTRAQ ( Isobaric tag for relative and absolute quantitation) par Ross et al.<sup>245</sup> en 2004 (commercialisé par AB Sciex).

Les tags isobariques ne sont pas différenciables dans le spectre MS1, ce qui permet de ne pas complexifier le spectre de masse en combinant différents échantillons. Les tags sont uniquement différenciables au niveau de leur spectre MS2 et la fragmentation des tags permet de remonter à l'abondance relative des peptides. B

Le tag est constitué de trois parties (voir figure 39) :

- Un groupement ester activé avec le N-hydroxysuccinimide (NHS) qui réagit avec les amines libres (N-terminal des peptides et lysines) pour former une liaison covalente. Un marquage non spécifique peut aussi avoir lieu principalement sur les tyrosines et les sérines<sup>246</sup>.
- Un groupement central servant à contrebalancer la masse du groupement rapporteur afin d'avoir des tags isobariques.
- Un groupement rapporteur observable en MS/MS et marqué <sup>13</sup>C et/ou <sup>15</sup>N pour obtenir un delta de masse d'au moins 6,32 mDa<sup>247,248</sup>.

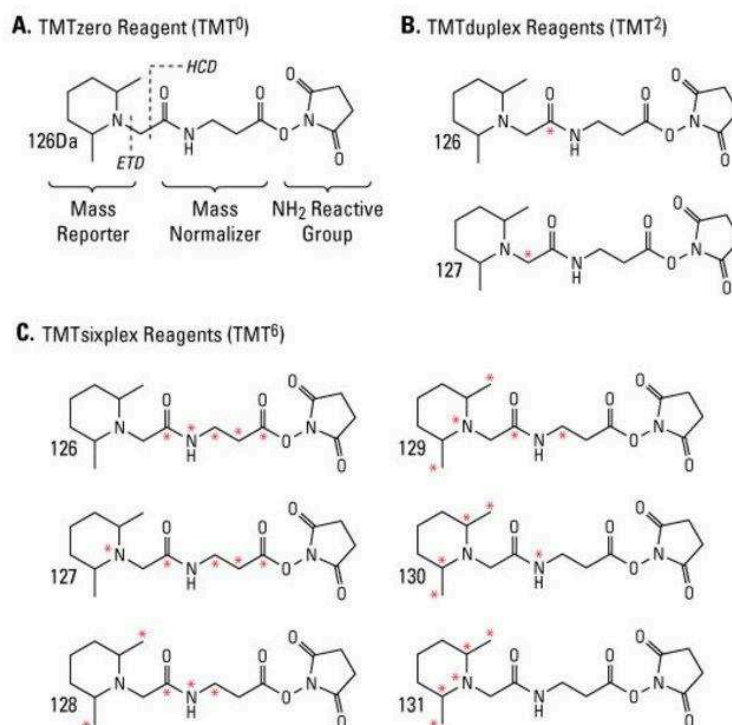


Figure 39 Les réactifs TMT utilisé pour les analyses multiplexes. Les atomes portant un marquage isotopique sont indiqués par une astérisque.

iTRAQ et TMT peuvent être multiplexé jusqu'à 8 et 10 respectivement. Pour le TMT 10-plex une résolution d'au moins 50 000 à 130 m/z est nécessaire pour pouvoir séparer et quantifier les ions rapporteurs isotopologues séparés uniquement par 6,32mDa<sup>247,248</sup>. Une des limites de cette approche est le possible co-isolément et fragmentation d'un ion ayant un rapport m/z compris dans la fenêtre d'isolement ce qui a pour conséquence d'interférer avec la mesure des ratios comme illustré figure 40. Ting et al.<sup>249</sup>, ont montré que d'ajouter une étape de sélection et de fragmentation MS3 permettaient d'éliminer presque toutes les interférences, l'identification se fait à partir des spectres MS et MS2 et la quantification avec le spectre MS3 (voir figure 40).

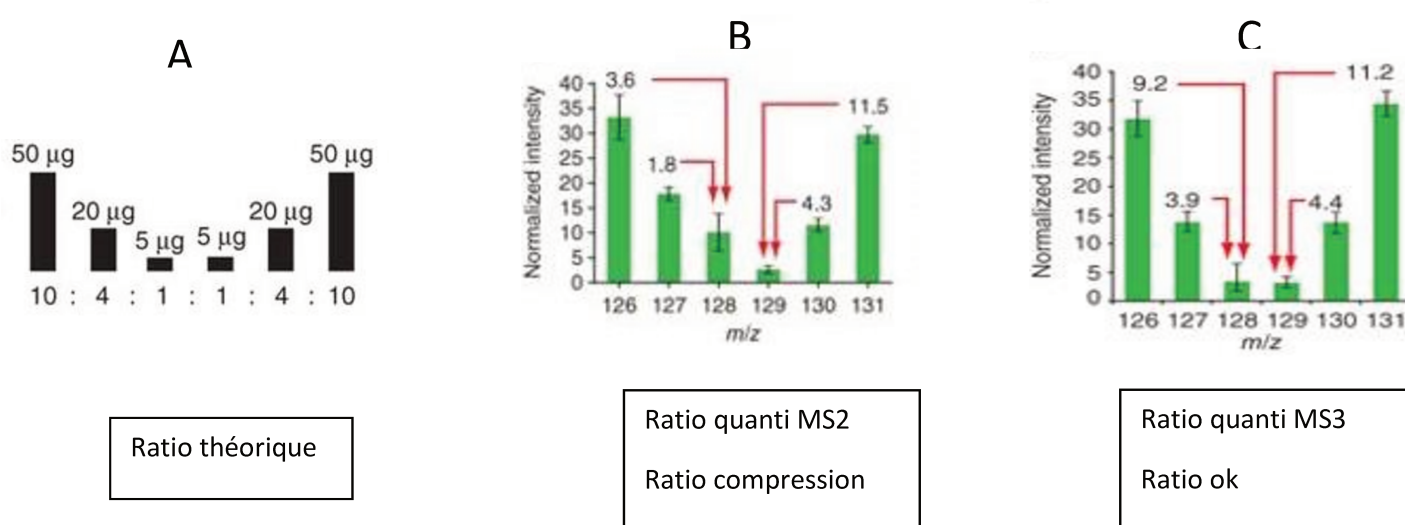


Figure 40 illustration de la compaction du ratio mesuré

En revanche pour obtenir une bonne sensibilité, une préparation d'échantillon laborieuse comprenant une étape de fractionnement (20 fractions) a dû être réalisée. Malgré cette préparation d'échantillon 12% de protéines en moins ont été quantifiées à cause d'un temps d'acquisition plus long entre MS3 et MS2<sup>249</sup>. La vitesse d'acquisition peut être augmentée sur les spectromètres de masse tribrid de type Orbitrap fusion en réalisant le scan MS3 dans le piège à ion linéaire et non dans l'orbitrap, augmentant ainsi le nombre d'identification de 66%<sup>250</sup>.

Dans le cas des protéoformes, ces tags ont principalement été utilisés pour étudier les peptides phosphorylés. Dans ce but Jiang et al.<sup>251</sup> ont développé une méthode, SPS-MS3 (synchronous precursor selection) brièvement un ion parent est sélectionné en MS1 pour être fragmenté en CID dans le piège à ion et jusqu'à 20 ions fragments MS2 peuvent être simultanément isolés pour être refragmentés en HCD et analysés par l'Orbitrap. Jiang et al., ont amélioré ces méthodes MS3<sup>249,252</sup>, en déclenchement le scan MS3 quand il y a une perte de neutre (sélection et fragmentation uniquement de la perte de neutre car elle domine généralement le spectre MS2 des peptides phosphorylés obtenus par CID dans un piège à ion. Berberich et al.<sup>253</sup> ont optimisé cette méthode de quantification relative (MS3-IDQ) pour les études de phosphoprotéomique par TMT en utilisant le scan MS2 pour l'identification (MS2-ID) et le scan MS3 pour la quantification des peptides phosphorylés (MS3-Q). Le spectre MS3 permet aussi d'obtenir des données d'identification supplémentaires. En effet, la MS3 permet d'éliminer les problèmes d'interférences et apporte des



fragments uniques permettant qu'augmenter la qualité des données d'identification. La valeur de NCE ainsi que la gamme de masse ont aussi été optimisées permettant d'augmenter de 50 à 70% le nombre de peptides phosphorylés quantifié.

Une alternative récente aux TMT et iTRAQ permettant de s'affranchir des problèmes d'interférences est le développement du EASI-tag par Virreira Winter et al<sup>254</sup> qui permet une quantification 6-plex. L'originalité de ce TAG est une fragmentation à plus basse énergie en HCD que le TMT et le iTRAQ résultant en une perte de neutre et la conservation intact de la structure peptidique (voir figure 41).

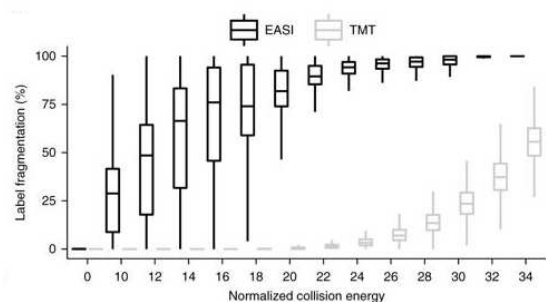
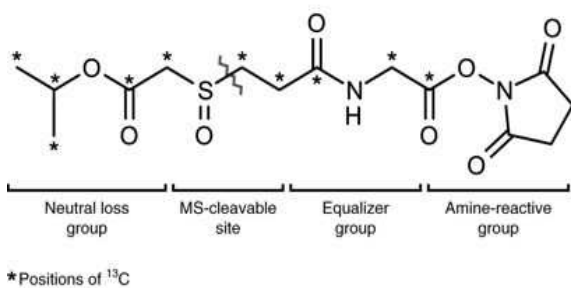


Figure 41 Représentation du réactif EASI-tag, les atomes portant un marquage isotopique sont indiqués par une astérisque(gauche). Comparaison du taux de fragmentation des tag EASI et TMT par HCD (droite). D'après Virreira Winter et al. 2018

(c) <sup>18</sup>O labelling:

Cette méthode présentée en 2001<sup>255</sup> permet l'introduction de deux isotopes d'oxygène 18 en C-terminale des peptides (figure 42). La réaction est catalysée par les protéases à sérine comme la trypsine, glu-C ou lys-C en présence d'eau H<sub>2</sub><sup>18</sup>O. Cette méthode universelle applicable à tout type d'échantillon permet de marquer tous les peptides de manière uniforme et économique. En revanche, seulement deux états peuvent être comparé à la fois et le marquage se fait au niveau des peptides, ce qui ne permet pas de normaliser la digestion.

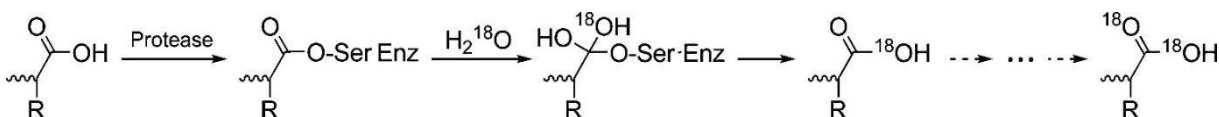


Figure 42 Schéma réactionnel d'insertion de H<sub>2</sub><sup>18</sup>O lors de la digestion à la trypsine d'une protéine. D'après Yao et al. 2011

Plus récemment Wang et al., ont proposé une variante en alkylant les cystéines avec de l'iodoacétamide marqué à l'oxygène 18. L'avantage est que le marquage se fait au niveau de la protéine mais uniquement les peptides contenant une cystéine peuvent être quantifiés.

(d) Dimethyl labelling:

Introduit en 2003 par Hsu et al.<sup>256</sup> ce marquage est économique et applicable à tout type d'échantillons<sup>257</sup>. Un groupement diméthyle s'attache en N-terminale des peptides et sur l'amine libre des lysines. Le marquage peut se faire en solution<sup>256</sup> ou sur colonne et en ligne<sup>258</sup>.

Cette méthode peut être multiplexée jusqu'à 5 fois<sup>259</sup>(voir figure 43).

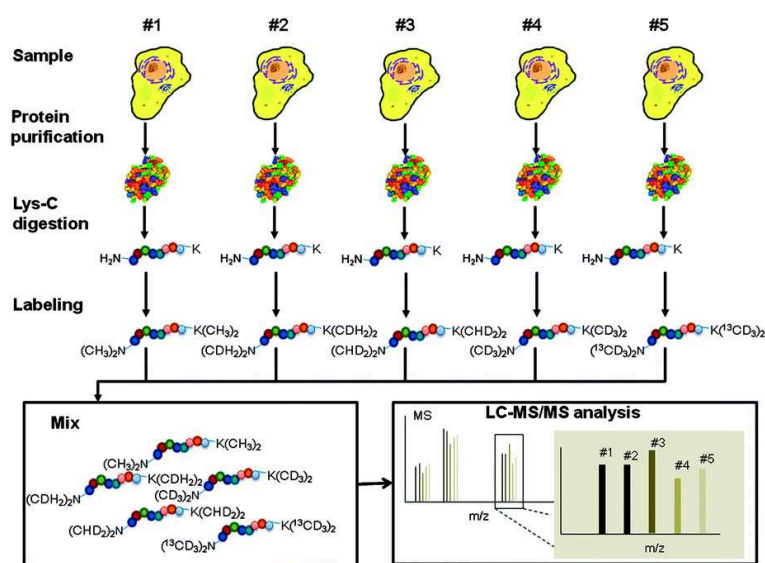


Figure 43 Marquage DML 5 fois multiplexé par l'ajout de Carbone 13C et de deutérium. Représentation schématique de la comparaison de 5 échantillons. D'après Wu et al. 2014

Lau et al.<sup>260</sup>, ont comparé les performances du SILAC et du diméthyl labelling (DML) pour la quantification de protéines provenant de cellules HeLa. Trois conditions de marquage (DML) ont aussi été testées afin d'évaluer l'efficacité des différentes méthodes de marquage existantes. Les trois conditions sont : en solution et marquage et sur colonne acidifiée ou non. Le marquage sur colonne non acidifiée est le plus efficace et a donc été comparé au SILAC. En termes de performance, le DML conduit à 23% de peptides en moins identifiés en comparaison avec le SILAC. L'étape de marquage se fait sur des stageTips C18 ce qui a pour conséquence une perte des peptides hydrophiles. Un autre avantage du SILAC est que les cellules sont combinées à un stade

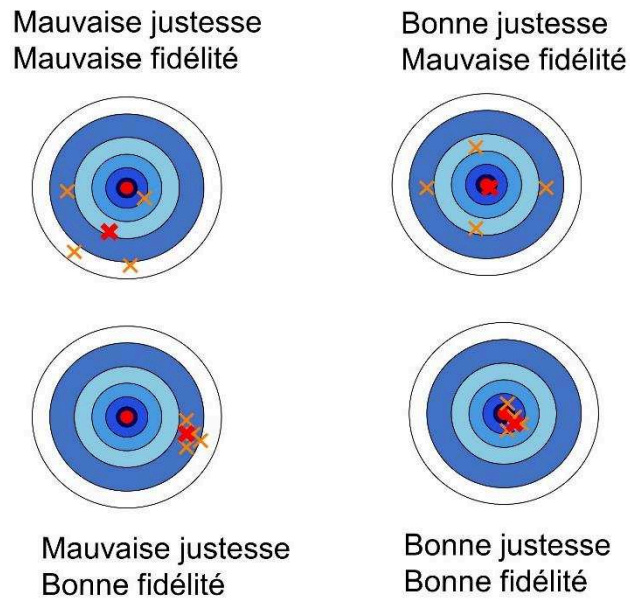
précoce, les étapes de digestion et de déssalage sur colonne sont normalisées contrairement au DML. Ce qui a pour conséquence une meilleure fidélité avec des standards de déviation en log2 de 0,175 contre 0,255 et une répétabilité du SILAC 4 fois supérieure au DML avec des standards de déviation de 0,0591 contre 0,204 pour le DML.

Conclusion : Les approches de quantification relative présentées dans cette partie sont majoritairement utilisées en combinaison des approches bottom-up. En effet, les marquages chimiques comme l'iTRAQ, le TMT, EASY-Tag et le DML ont été optimisés pour être utilisés au niveau des peptides. Bien que comme montré par Hung et al., le TMT semble pouvoir être utilisé pour marquer des protéines. Ces approches permettent de quantifier des modifications post-traductionnelles. Mais comme la quantification se fait au niveau du peptide (bottom-up) ces approches sont limitées pour l'étude des protéoformes. Le SILAC a été couramment employé en combinaison des méthodes bottom-up apportant une meilleure fidélité que les marquages chimiques car l'étape de digestion est normalisée. Le SILAC a aussi été évalué pour les approches top-down<sup>261,262</sup> bien que des limitations existent principalement concernant le taux d'incorporation des acides aminés lourds qui est dépendant de la taille des protéines. Le SILAC pourraient avoir un intérêt pour l'étude des protéoformes.

#### **b) Quantification absolue des protéines**

Les méthodes de quantification relative des protéines permettent de quantifier une protéine de manière fidèle. En revanche, la valeur absolue de la protéine ne peut pas être déterminée car il manque un étalon interne permettant de rendre la méthode juste. Cet étalonnage avec un étalon caractérisé permet de développer des méthodes de quantification absolue. Pour qu'une méthode de quantification absolue soit validée, elle doit être exacte c'est-à-dire juste et fidèle comme illustré figure 44. Alors que les méthodes de quantification relative doivent avoir uniquement une bonne fidélité. La justesse correspond à l'écart mesuré par rapport à une valeur de référence et la fidélité à la dispersion des valeurs mesurées obtenues par des mesures répétées.

Figure 44 Représentation de la justesse et de la fidélité sous forme de cibles



Les méthodes de quantification absolue utilisent un étalon interne lourd qui sert à la quantification et doit être au plus proche de la protéine endogène à quantifier. Les différents étalons lourds, ainsi que leur étape d'insertion dans la préparation d'échantillon, sont illustrés dans la figure 45 ci-dessous<sup>263</sup> :

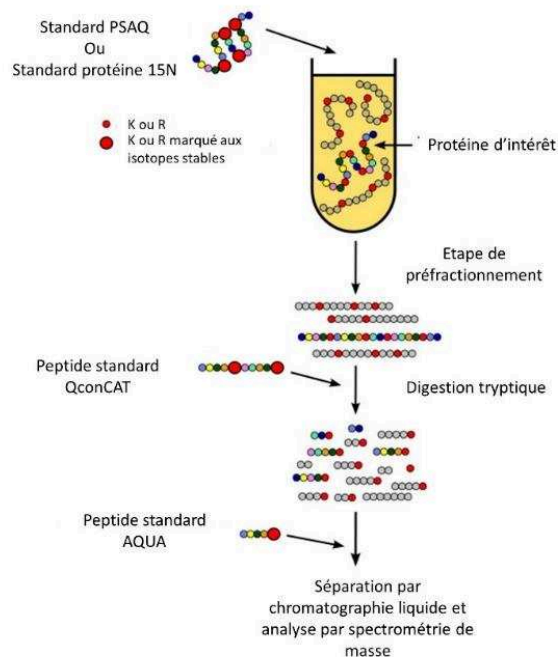


Figure 45 Stratégie de quantification absolue des protéines par bottom-up. Les étapes d'ajouts des étalons internes sont représentées. D'après Brun et al. 2009

### (i) Principe et standards internes

Les approches exposées précédemment s'intéressent à comparer un (ou plusieurs) échantillon(s) avec une référence sans donner une concentration exacte. La quantification absolue des protéines permet d'obtenir la valeur de concentration totale de la protéine et doit être fidèle et juste. Les performances de la méthode sont dépendantes du choix de peptide, du type de marquage et de la calibration.

#### (a) Principes généraux :

##### Choix du peptide

En premier lieu, concernant les approches par bottom-up un ou plusieurs peptides par protéine doit être choisis. Les peptides sélectionnés pour la quantification doivent être quantotypiques c'est-à-dire qu'ils représentent fidèlement la quantité de la protéine. Pour cela ils doivent répondre à plusieurs critères. Les peptides doivent être protéotypiques c'est à dire appartenant à une unique protéine et ne doivent pas contenir de méthionine, cystéine et tryptophane (oxydation), glutamine en N-terminale (formation d'un pyro-glu : cyclisation) et ne pas contenir

de modifications post-traductionnelles<sup>264,265</sup>. Les peptides sont sélectionnés en fonction d'un critère de taille, ils ne doivent être ni trop hydrophiles ni trop hydrophobes, ce qui correspond généralement à des peptides d'environ 10 acides aminés<sup>266</sup>.

### **Types d'étalons lourds internes**

La quantification est généralement obtenue par dilution isotopique. Il faut donc sélectionner un étalon interne lourd. Il peut être de différente nature et être ajouté à l'échantillon à différentes étapes comme illustré figure 45.

Etalons peptidiques :

- AQUA : Les peptides AQUA ( Absolute Quantification) sont généralement des peptides tryptiques marqués <sup>15</sup>N et <sup>13</sup>C au niveau de la lysine ou arginine en C-terminale. La pureté des peptides est évaluée par analyse des acides aminés<sup>267</sup>. Ce type de méthode ne prend pas en compte la digestion et est juste uniquement si la digestion est complète<sup>268</sup>. Le neucode peut aussi être appliqué dans ce cas pour une quantification multiplex absolue en ajoutant un étalon lourd AQUA avec une lysine isotopologue des lysines utilisées pour le marquage (SILAC) les cellules mais cette méthode n'est applicable qu'aux cellules marquées métaboliquement<sup>269</sup>.
- Le QconCAT qui est une concaténation de peptides tryptiques provenant de plusieurs protéines dans une même protéine<sup>270</sup>.

Etalons protéiques : Les protéines recombinantes peuvent être utilisées comme standards pour la quantification absolue des protéines endogènes. Il existe plusieurs types de protéines standards :

- Les protéines recombinantes qui sont entièrement marquées <sup>15</sup>N<sup>101</sup>,
- PSAQ<sup>271</sup> et Flexiquant<sup>272</sup> où seulement les arginines et lysines sont marquées <sup>15</sup>N et <sup>13</sup>C
- QPrEST<sup>273</sup> où 50 à 150 acides aminés sont identiques avec la protéine à quantifier. Ces 50 à 150 acides aminés comprennent au minimum 2 peptides tryptiques marqués <sup>15</sup>N et <sup>13</sup>C sur lysine ou arginine. A cela sont ajoutés : une étiquette 6-histidine pour la purification et une étiquette de 120 acides aminés assurant la solubilité de la protéine. Le QPrEST permet notamment la quantification de protéines non solubles et d'ajouter une dimension quantitative absolue au SILAC<sup>274</sup>.

### **Choix calibration externe ou non**

Enfin la quantification peut être réalisée avec ou sans calibration externe. Ce paramètre dépend de la disponibilité d'un calibrant externe ainsi qu'une matrice vierge adapté. Pour la calibration externe une gamme de calibration doit être faite en ajoutant une quantité variable d'étalon léger sur 2 à 3 log<sub>10</sub>. Une quantité fixe d'étalon lourd est ajouté et doit être, dans l'idéal, le plus proche possible de la quantité endogène. La courbe de calibration est obtenue en déterminant le ratio d'aires étalon léger/aire étalon lourd en fonction de la concentration de l'étalon léger. La gamme de calibration est réalisée dans une matrice mime ne contenant pas de façon endogène la protéine cible et reproduisant les effets matrices de la matrice d'intérêt. Finalement la quantification des échantillons se fait par ajout de la même quantité d'étalon lourd que celle ajouté pour la gamme de calibration. La concentration est obtenue par extrapolation de l'aire de l'endogène/l'aire de l'étalon lourd contre la gamme de calibration (sans calibration externe, la quantité absolue de la protéine est déterminée directement par le ratio d'aire endogène/étalon lourd).

Les étalons internes et externes sont quantifiés en acide aminé après hydrolyse acide, en utilisant de l'acide chlorhydrique, des peptides ou protéines<sup>267</sup>. Uniquement la proline, valine, isoleucine, phénylalanine, arginine, alanine et leucine sont utilisés et des versions marquées de la proline, valine, isoleucine et phénylalanine sont ajoutés pour la quantification. La concentration des acides aminés permet d'obtenir la concentration de l'étalon.

Les problèmes rencontrés pour parvenir à cette quantification absolue résultent du fait que l'analyte est souvent différent de l'étalon lourd ou du calibrant léger en début d'expérience (peptides vs. Protéines ou protéines ayant des conformations différentes). Ce qui a pour conséquence de biais lors de la préparation d'échantillon (par exemple : stabilité de la protéine, rendement SPE, rendement immunoprécipitation et effets matrices) qui ne sont pas pris en compte<sup>268</sup>.

### **Sans calibration externe :**

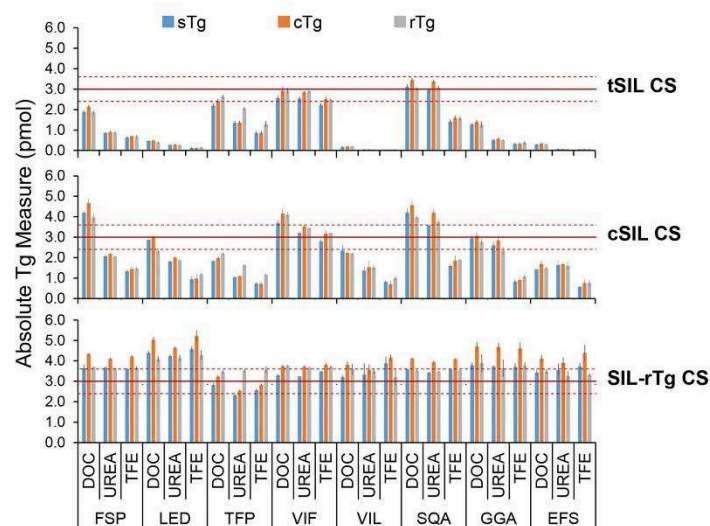
Pour la quantification de biomarqueurs en plasma, notamment dans le cas de la quantification multiplexe, les standards de protéines servant à faire une gamme de calibration externe ne sont pas toujours disponibles ou trop coûteux. De plus une matrice vierge n'est pas forcément disponible et l'utilisation d'une matrice mime en BSA ne peut parfaitement mimer la complexité du plasma.

Les peptides AQUA sont couramment utilisés pour la quantification de protéines sans courbe de calibration externe. Afin de contourner les deux problèmes de calibration et de matrices, LeBlanc et al.<sup>275</sup> ont développé une méthode double SIS multiplexe, pour la quantification de 31 peptides, contenant 2 peptides lourds isotopologues, le premier est utilisé comme étalon interne lourd et le second comme standard de calibration ajouté à concentration variable pour réaliser une gamme de calibration. La performance analytique pour les 31 peptides est un CV <12,1% et une justesse <10%. Une méthode classiquement utilisée est de réaliser des gammes de calibration inverse c'est à dire que l'étalon lourd (peptide AQUA) est ajouté en quantité variable et que le peptide endogène sert d'étalon interne. L'inconvénient majeure de cette méthode est que l'étalon utilisé comme standard interne est différent entre celui utilisé pour faire la gamme inverse (peptide endogène provenant d) et les échantillons (peptide AQUA) contrairement à la méthode double SIS. Ce qui résulte en 1 peptide ayant une erreur > 20%, 4 peptides >10% et 2 peptides dont la courbe ne peut être obtenue car les peptides endogènes sont inférieurs à la LLOQ alors que pour la méthode double SIS les 31 peptides ont une erreur <10%.

Afin d'augmenter le multiplexage et donc le nombre de protéines pouvant être quantifiées, Beynon et al.<sup>270</sup>, ont introduit le (QconCAT) qui est une concaténation de peptides tryptiques provenant de plusieurs protéines dans une même protéine. Le QconCAT souffre du même problème de digestion que les peptides AQUA, en effet les acides aminés autour des sites de coupures sont différents entre le QconCAT et la protéine endogène or la trypsine est sensible aux acides aminés entourant les sites de clivages<sup>138</sup>. Scott et al.<sup>276</sup> ont donc évalué le nombre nécessaire d'acides aminés à incorporer aux extrémités des peptides constituant la protéine QconCAT afin de reproduire l'environnement local du peptide endogène. Cet ajout a pour but de normaliser la digestion. Pour cela 0, 2, 4 ou 6 acides aminés ont été ajoutés et comparés avec l'utilisation d'une protéine marquée <sup>15</sup>N. La digestion a été optimisée avec l'ajout d'un surfactant le Rapigest à 0,1% pour dénaturer la protéine. Sans ajout d'acides aminés (+0) les résultats de quantification sont dépendants des peptides avec des pentes de calibration au-dessus ou en dessous de l'unité montrant une efficacité de digestion différente en comparaison avec la protéine recombinante. L'ajout de 6 acides aminés (+6) permet de rendre quantitatif les peptides qui ne l'étaient pas avec +0 et permet donc de quantifier avec une meilleure justesse plus de peptides. La même comparaison a été faite pour les peptides AQUA montrant des meilleures performances quantitatives avec l'ajout de six aminés avec une pente à +0 égale à  $0,04 \pm 0,00$ , à +4 égale à  $0,015 \pm 0,02$  et à +6 égale à  $0,77 \pm 0,06$  pour le peptide IPVALGLK. L'utilisation d'un peptide AQUA résultera donc dans une surestimation de la quantité de la protéine et ne peut donc pas être utilisé



pour de la quantification. L'ajout de 6 acides aminés supplémentaires permet de s'approcher ici de l'unité et permet d'utiliser ce peptide qui a une cinétique de digestion lente pour de la quantification. En revanche, le peptide DDKPTLQLESVDPK a une pente proche de l'unité peu importe l'étalon lourd utilisé et dans ce cas un peptide AQUA est suffisant pour la quantification. En effet, celui-ci a une cinétique de digestion rapide et sa quantité après avoir atteint un plateau reste stable au cours du temps. L'avantage des peptides avec des acides aminés ajoutés aux extrémités des peptides marqués est une amélioration de la justesse de la méthode<sup>130</sup> principalement à partir de 6 acides aminés. Bien que cette méthode permette la réduction de la variabilité liée à la digestion, elle ne permet pas de normaliser les étapes de préparation d'échantillon faites au niveau de la protéine.



**Figure 46** Comparaison des différentes méthodes de quantification, pour la thyroglobuline, sans calibration externe, tSIL (peptide AQUA), CSIL (peptide AQUA comprenant 6 acides aminés en N et C-terminale) et SIL-rTg (protéine recombinante SIL). La droite rouge représente la valeur cible et les lignes en pointillés la valeur cible +/- 20%. Sans calibration externe la protéine recombinante est celle permettant une quantification la plus juste suivit de CSIL puis de tSIL. Les conditions de digestion (DOC, Urée et TFE) ont aussi une grande influence sur la quantification à partir des peptides. D'après Shufford et al. 2017

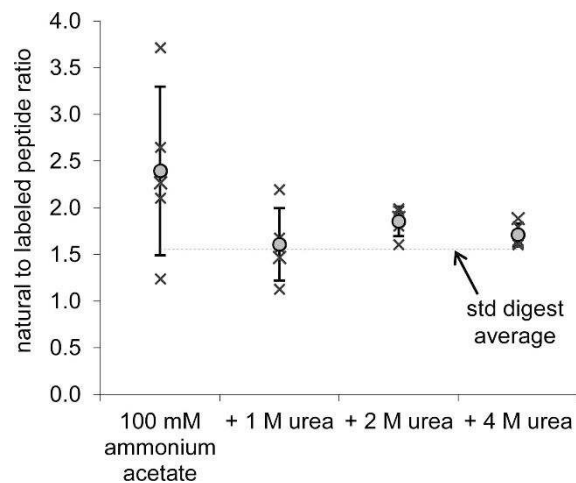
(b) Comparaison des performances :

Oeckl et al.<sup>277</sup> ont comparé les performances des différentes méthodes PSAQ, QPrEST, AQUA et AQUA avec deux acides aminés en N-terminale et C-terminale pour la quantification de l'alpha-synucléine. Ces quatre approches sont équivalentes en termes de CV avec des CV<15%. En revanche, concernant la justesse seul le QPrEST s'est montré comparable au PSAQ. Donc dans le cas d'une petite protéine (14 kDa) naturellement non replié le QPrEST mime la protéine endogène et peut être une alternative au PSAQ. Les deux approches peptidiques sont dépendantes du peptide quantifié et peuvent avoir des déviations jusqu'à 54%.

De même, Brun et al.<sup>271</sup>, ont comparé AQUA, QconCAT et PSAQ pour la quantification d'une toxine. Les méthodes basées sur les peptides AQUA et QconCAT ne compensent pas la variabilité de la digestion ni de la préparation d'échantillon notamment de l'étape préfractionnement par SDS-PAGE. La quantité de toxine est donc sous-estimée. Contrairement au PSAQ qui permet de normaliser ces étapes et donc de donner une valeur absolue de la quantité de protéines.

(c) Problèmes spécifiques à la quantifications absolue :

Bien que la quantification absolue en utilisant des standards protéines marquées soit considérée comme la méthode la plus exacte, des précautions sont nécessaires pour s'en assurer. En effet, la structure tertiaire de la protéine standard doit être caractérisée pour plusieurs raisons. Premièrement, un changement de structure tertiaire peut influencer sur l'adsorption non spécifique de la protéine. Deuxièmement, la protéine peut être susceptible de se dégrader plus rapidement. Finalement, un biais peut être introduit pendant l'étape de préparation d'échantillon notamment pour l'immunoprécipitation. Il est donc nécessaire de casser la structure tertiaire de la protéine endogène et du standard le plus tôt possible lors de la préparation d'échantillon. Une équilibration du standard dans un tampon dénaturant peut permettre de réduire les biais de quantifications<sup>278</sup> (voir figure 47.)



**Figure 47 Effet du tampon d'équilibration sur la quantification absolue d'une protéine. D'après Pritchard et al. 2014**

Cette observation est renforcée par l'étude de Shuford et al. pour la quantification de la thyroglobuline humaine<sup>268</sup>. Trois méthodes de normalisation évaluées : protéines, peptides tryptiques et peptides tryptiques avec des acides aminés en plus en N et C terminale et 3 tampons de dénaturation DOC, urée et trifluoroéthanol. En absence d'effets matrices et avec une calibration externe avec une protéine thyroglobuline recombinante les trois méthodes ont des performances équivalentes en termes de précision. En revanche, un des huit peptides a un biais négatif en urée et trifluoroéthanol suggérant une digestion moins efficace de la protéine endogène par rapport à la protéine recombinante. Ce problème est résolu par l'ajout de DOC qui permet d'avoir un rendement de digestion identique, le choix du tampon de dénaturation est donc primordial pour avoir des données quantitatives (voir figure 46 et 48).

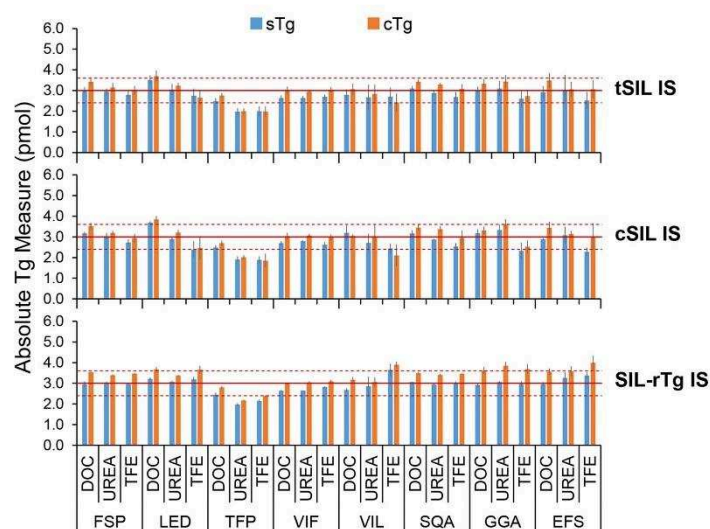


Figure 48 Comparaison des méthodes de quantification avec calibration externe. En utilisant une protéine recombinante en DOC tous les peptides sont quantifiés avec une précision <20%. D'après Shufford et al. 2017

Ces approches de quantification absolue nécessitent une détection MS reproductible et sensible. Les approches de types DDA ne sont donc pas adaptées et des méthodes dites ciblées ont été développées.

### (ii) Méthodes ciblées : SRM et PRM

Les approches de types shotgun manquent de sensibilité et de reproductibilité pour déterminer de manière précise la concentration d'une protéine et les méthodes d'acquisition de type DDA sont biaisées envers la détection des protéines de forte abondance, ne fragmentant pas les peptides les moins intenses. C'est pourquoi des méthodes ciblées, c'est-à-dire qu'uniquement une liste d'ions préalablement sélectionné ne sera acquise, ont été développées pour la quantification des protéines.

Pour la quantification des protéines la méthode de choix est le SRM (selected reaction monitoring) ou MRM (multiple reaction monitoring) avec des appareils de type triple quadripôle. Le premier quadripôle (Q1) sert à sélectionner un ion précurseur, le second (Q2) à le fragmenter en CID et le troisième (Q3) permet de sélectionner consécutivement un ou plusieurs fragments appelés transitions<sup>279,280</sup>. Généralement 3 à 5 transitions<sup>281</sup> donnant le signal le plus intense et ayant le

moins d'interférences sont sélectionnées, après optimisation de l'énergie de collision. Du fait de la basse résolution des quadripôles ces méthodes souffrent de problèmes d'interférences notamment en matrices complexes<sup>282,283</sup>. Depuis 2012, les spectromètres de masse haute résolution de type Q-exactive<sup>284</sup> se sont imposés comme une alternative crédible à ceux de type triple quadripôles (voir figure49). Brièvement, le Q-exactive est constitué de 4 parties : un premier quadripôle sert à sélectionner un ion précurseur, puis les ions sont accumulés dans la C-trap pour ensuite être soit analysés intacts dans l'orbitrap ou envoyés dans la cellule HCD (higher energy collision dissociation) pour être fragmentés et finalement analysés par l'orbitrap. L'orbitrap est à analyseur de masse à haute résolution et permet d'atteindre une précision de masse <5 ppm. Le Q-exactive en mode ciblée peut être utilisé selon deux modes single ion monitoring (SIM) et parallel reaction monitoring (PRM). Pour le mode SIM, les ions précurseurs, après sélection dans le quadripôle, sont stockés dans la C-trap puis ils sont envoyés intact dans l'orbitrap. Pour le mode PRM, le plus utilisé en matrice complexe, une étape de fragmentation dans une cellule HCD est ajoutée après stockage des ions dans la C-trap et tous les ions fragments sont analysés en parallèle ; le choix des transitions se fait donc post-acquisition.

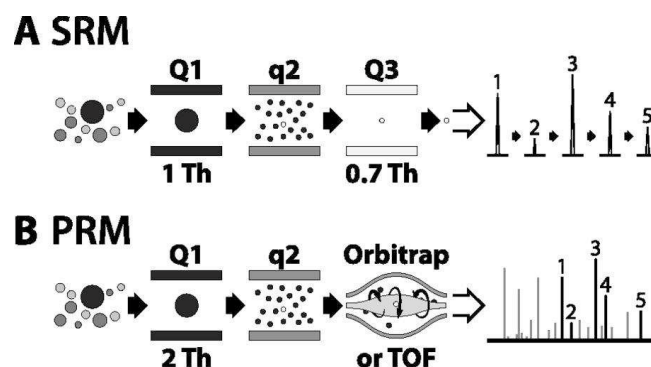


Figure 49 Représentation schématique des modes SRM (A) et PRM (B) D'après Peterson et al. 2012

Plusieurs études de comparaison des performances du SRM et du PRM pour la quantification des biomarqueurs ont été publiées. Ces études bien que parfois contradictoires tendent vers le fait que plus la matrice est complexe plus l'apport de la haute résolution est important en termes de gamme dynamique de quantification car la haute résolution permet d'éliminer des interférences. Alors qu'en tampon la gamme dynamique de quantification du SRM et PRM est équivalente<sup>284</sup>. En matrices complexes (digestat tryptique de levure et urine) Peterson et al.<sup>284</sup>, et Gallien et al.<sup>285</sup>,

ont montré l'apport de la haute résolution, le PRM exhibant une gamme dynamique plus importante que le SRM. Le SRM présentait des interférences pour des peptides alors que ces interférences étaient absentes en PRM, démontrant le gain apporté par la haute-résolution. La haute résolution ne permet pas de résoudre tous les problèmes d'interférences notamment en mode SIM (single ion monitoring, où les ions ne sont pas fragmentés dans la cellule HCD). Le mode SIM à 70 000 de résolution à 200 m/z ne s'affranchit pas des interférences, ce qui réduit sa gamme dynamique de quantification démontrant la nécessité de quantifier les peptides par MS2 et non MS1<sup>285,286</sup>. Dans une matrice moins complexe Ronsein et al. ont montré que pour la quantification des protéines dans les lipoprotéines de haute densité les deux modes de quantification ciblés PRM et SRM ont des performances équivalentes en terme de linéarité, gamme dynamique, précision et répétabilité<sup>286</sup>. Dans le cas de la quantification en sérum de HSP90α<sup>287</sup> le mode PRM et les ELISA atteignent les mêmes limites de quantification : 1,5 ng/mL pour le peptide 1 et 1,6 ng/mL pour le peptide 2 en mode PRM et pour les ELISA 1,2 ng/mL. Les CV sont équivalents pour les deux méthodes (<10%). Les auteurs ont aussi, dans cet exemple, évalué le mode SRM qui a une limite de quantification 5 fois plus élevée que les deux précédents avec des CV<11,8%. Les ELISA ont l'avantage de ne pas avoir besoin de préparation d'échantillon alors que pour les analyses par spectrométrie de masse une extraction par SPE a été nécessaire. Néanmoins, dans ce cas, le test ELISA utilisé est loin d'atteindre les limites de quantifications généralement atteintes (≈ 10 pg/mL)

Les méthodes SRM et PRM peuvent être hautement multiplexées ce qui permet la quantification d'un grand nombre de peptides en matrices biologiques. Les deux modes ont des performances similaires pour le multiplexe :

Concernant le SRM, Burgess et al.<sup>129</sup>, ont développé une méthode permettant la quantification de 400 peptides endogènes en suivant aussi une version marquée aux isotopes stables de ces peptides, ce qui représentent en tout 800 peptides ciblés et 2400 transitions suivies avec un gradient de 180 min et un dwell time compris entre 8 et 107 ms. En raison de la variabilité des temps de rétention d'une injection à l'autre, Rougemont et al.<sup>288</sup>, ont développé une méthode appelé SCOUT-SRM. 19 peptides répartis le long du chromatogramme enclenchent l'acquisition d'un groupe de transitions (150 max), ce qui a permis la quantification de 782 peptides en suivant 2346 transitions avec un gradient de 120 min. Pour le multiplexage en PRM Gallien et al.<sup>289</sup>, ont proposé une méthode pour la quantification de 600 peptides en 66 min avec l'idée aussi de résoudre le problème des temps de rétention variables. Ici, le peptide marqué déclenche l'acquisition du peptide endogène pour cela deux modes sont utilisés : un mode surveillance avec

un temps de remplissage de la C-trap de 20 ms et un mode de quantification avec un temps de remplissage de 110 ms permettant d'augmenter le signal sur bruit pour les peptides de faible abondance. Le PRM montre des performances sensiblement identiques au SRM en un temps 2 fois plus faible car la haute résolution assure une plus grande sélectivité et donc la séparation chromatographique peut être moins bonne.

Ces méthodes ciblées ont été développées principalement pour la quantification des peptides après digestion. Pour la quantification en top-down ou en intacte d'autres méthodes ont été explorées.

### **(iii) Quantification des protéines par Top-down/intact**

L'analyse en top-down permet la caractérisation des protéines sans passer par une étape de digestion, qui est introduit une variabilité. Les protéines intactes sont analysées par ESI ou MALDI-MS pour obtenir leurs masses moyenne ou monoisotopique. Puis des informations sur leurs séquences sont obtenues par fragmentation. L'intérêt majeur du top-down est pour l'analyse des protéoformes (Introduction). La quantification en top-down se fait généralement sur des appareils haute résolution comme les Q-TOF, orbitrap ou FT-ICR. Bien qu'il soit possible de le faire sur des appareils moins résolutifs de type triple quadripôle<sup>290</sup>.

La quantification des protéines n'est pas aussi simple que la quantification des peptides du fait de la dispersion du signal sur plusieurs états de charges de 6 pour le lysozyme humain à plus de 36 pour un anticorps. De plus, pour chaque état de charge, si la résolution est suffisante, un massif isotopique apparaît (voir figure 50). La question est donc quel état de charge faut-il choisir pour que la méthode soit quantitative et quels isotopes ?

#### **(a) Instrumentation :**

La quantification au niveau de la protéine intacte ou en top-down se fait principalement sur des spectromètres de masse à haute résolution. En effet, cette méthode analyse des protéines entières avec des masses de plusieurs kDa. Pour les différencier il est donc nécessaire d'employer des spectromètres de masse à haute résolution ayant une grande précision de masse. De même pour les spectres MS de fragmentations des protéines qui peuvent être complexe. L'analyse peut se faire en mode non ciblée<sup>215</sup> ou en mode ciblée comme le PRM<sup>192</sup> ou SRM.

La quantification en top-down des protéines par SRM est nécessite plus d'optimisation que pour les approches par bottom-up<sup>290</sup>. En effet, pour les protéines le CID produit un grand nombre d'ion fragment de basse masse et la fragmentation n'est pas reproductible. Wang et al., ont montré

qu'en diminuant l'énergie de collision et la pression des gaz dans la cellule CID, la fragmentation des protéines devenait reproductible pour les 6 protéines testées (8-17 kDa). Concernant la quantification, l'état de charge le plus intense est filtré par le premier quadripôle et deux fragments sont sélectionnés. Pour les 6 protéines standards : la gamme dynamique, la limite de quantification, la justesse et la fidélité ont été évaluées. Les 6 standards n'ont pas la même réponse en masse impliquant une limite de quantification variable de 3 à 50 µg/mL ainsi qu'une gamme dynamique variant de 1 à 3 log. En revanche, la méthode de quantification développée s'est montrée performante en termes de justesse (erreur <5%) et de fidélité <15%. D'après cet exemple, le mode SRM peut donc être utilisé pour quantifier les protéines < 20 kDa.

Des alternatives à la source electrospray permettent aussi la quantification de protéines intactes comme le MALDI ou l'ICP-MS. L'ICP-MS permet l'analyse des hétéroatomes présent dans les protéines comme le soufre (cystéine et méthionine). La quantification absolue des protéines est obtenue par dilution isotopique en corrigeant les erreurs d'injection par un étalon interne BOC-Met et les effets matrices par infusion de <sup>34</sup>S, avec une LOD 1,2-300 pmol de soufre<sup>264</sup>. La séparation chromatographique est primordiale en effet si deux protéines co-eluent la quantification va être fautive. Cette méthode a été appliquée avec succès à l'analyse d'un venin de serpent. Comme l'ICP-MS ne permet pas l'identification des protéines, leur identification est réalisée en parallèle par un Q-TOF.



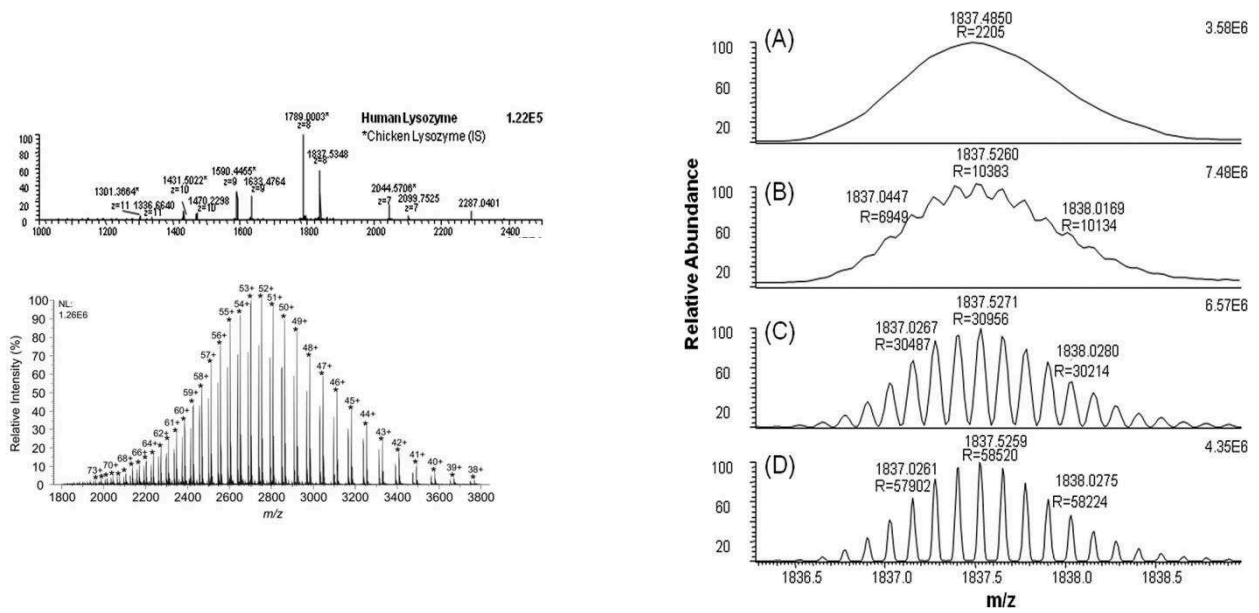


Figure 50 Profil d'états de charges pour le lysozyme humain et un anticorps (gauche), l'augmentation de la résolution de 2 000 et 58 000 permet d'avoir une séparation isotopique(droite)

(b) Traitement du signal

Ruan et al.<sup>291</sup>, ont sélectionnée, un état de charge, le plus intense en faisant un XIC des 10 isotopes les plus intenses pour le lysozyme humain. Alors que, Gucinski et al.<sup>292</sup>, ont trouvé des résultats similaires où la gamme dynamique de quantification est légèrement plus importante en utilisant uniquement un état de charge (25-2500 mM) plutôt qu'en sommant les quatre états charges les plus intenses (50-2000 mM) en réalisant un XIC sur l'isotope le plus intense. L'idée de sommer plusieurs états de charges vient de l'observation que la distribution des états de charges est dépendante de la concentration de la protéine. En effet, plus la concentration augmente plus le massif d'états de charges se déplacent vers un massif comprenant plus de bas états de charges, ce qui peut biaiser la quantification relative comme montré par Van den Broeck<sup>293</sup> et al. et Ramagiri et al.<sup>294</sup>. Le fait que la gamme dynamique ne diminue que légèrement peut être intéressant en matrices complexes, en effet en cas d'interférences un autre état de charge peut être choisit sans impact trop négatif sur les performances de quantification. Pour la quantification d'anticorps (avec un standard d'anticorps <sup>13</sup>C), Lanshoeft et al.<sup>295</sup>, ont montré que le choix des états de charges n'impactait ni la fidélité ni la justesse de leur méthode (voir tableau 7). De plus, l'ajout de plus d'isotopes n'augmente pas la gamme dynamique de quantification de l'hormone de croissance somatotropine (20 kDa). De façon assez intéressante l'ajout de 5, 7 et 10 états de charges diminue la gamme dynamique, ce qui peut être expliqué par une augmentation du bruit.

Tableau 7 Impact du nombre d'état de charge utilisé pour la quantification sur la justesse et la fidélité (Qc n=3). Adapté de Lanshoeft et al. 2017

		Concentration des QCs en sérum de rat ( $\mu\text{g/mL}$ )				
		8.00	5.00	0.250	0.100	
Nombre d'états de charges		Justesse et fidélité intrajour (n=3)				
	<b>3 (51+ à 53+)</b>	Concentration moyenne ( $\mu\text{g/mL}$ )	<b>8.7</b>	<b>5.44</b>	<b>0.273</b>	<b>0.109</b>
	$r^2= 0.9891$	Justesse intrajour (biais %)	8.7	8.7	9.4	9.0
		Fidélité intrajour (CV%)	3.2	11.3	8.4	1.9
<b>6 (51+ à 56+)</b>		Concentration moyenne ( $\mu\text{g/mL}$ )	8.09	5.45	0.265	0.112
	$r^2= 0.9876$	Justesse intrajour (biais %)	1.1	9.0	5.9	12.3
		Fidélité intrajour (CV%)	3.6	3.5	4.1	6.7
<b>9 (48+ à 56+)</b>		Concentration moyenne ( $\mu\text{g/mL}$ )	8.51	5.55	0.289	0.115
	$r^2= 0.9876$	Justesse intrajour (biais %)	6.4	11.0	15.6	15.4
		Fidélité intrajour (CV%)	6.1	3.9	3.4	7.7
<b>18 (42+ à 59+)</b>		Concentration moyenne ( $\mu\text{g/mL}$ )	8.32	5.53	0.291	0.119
	$r^2= 0.9928$	Justesse intrajour (biais %)	4.0	10.6	16.6	19.2
		Fidélité intrajour (CV%)	6.2	1.0	1.6	3.3

(c) Etalon interne en top-down/intacte

Les méthodes de quantification « label-free » qui sont basées sur l'extraction du signal (intensité et aire) ont été comparées avec une méthode de comptage du nombre de spectre couramment utilisé en bottom-up <sup>296</sup>. Plusieurs critères sont essentiels pour son application notamment un nombre élevé de spectres est nécessaire pour l'identification d'un protéoforme, ce qui conduit à diminuer le temps d'exclusion d'un ion, ainsi qu'une analyse à haute résolution afin de déterminer

la masse monoisotopique de la protéine. Trois protéines TNF- $\alpha$  (17.3 kDa), RNase (13.7 kDa) et CAH (29 kDa) ont été ajoutés en quantité croissante de 1 à 50 ng pour la TNF- $\alpha$  et la RNase et de 10 à 100 ng pour la CAH, la linéarité des trois méthodes a été comparée ainsi qu'à mesurer des ratios d'abondances supérieurs ou égaux à 2. Les trois méthodes ont une tendance à la linéarité avec des  $R^2$  variant de 0.66 à 0.90. En utilisant le spectral counting dans 18%, 27% et 63% des cas un changement de ratio  $\geq 2$  d'une protéine n'est pas détecté pour le TNF- $\alpha$ , RNase et CAH; pour l'intensité des pics 50%, 33%, 36% et pour l'aire des pics 13%, 31% et 72%. Ces résultats mettent en valeur une performance légèrement supérieure pour le spectral counting par rapport aux deux autres méthodes. Le spectral counting peut donc être utilisé pour la quantification en top-down des protéoformes notamment dans les études protéomiques de recherche de biomarqueurs. Acosta-Martin et al.<sup>297</sup>, ont proposé une méthode en SRM (piège à ion) pour la quantification de l'hémoglobine HbA<sub>2</sub> qui est composée de 2 chaînes  $\alpha$  et 2 chaînes  $\delta$  et dont la proportion par rapport au taux d'hémoglobine globale (HbA, HbA<sub>2</sub> et HbF) augmente passant à des valeurs  $< 3.5\%$  à des valeurs comprises entre 4-6%. Les auteurs de cette étude ont montré l'intérêt de la quantification à partir des ions fragments en effet, le signal sur bruit du spectre MS de fragmentation permet une amélioration du signal sur bruit d'un facteur 68 à 90. La quantification est alors obtenue à partir des ions fragments spécifiques des chaînes  $\alpha$  et  $\delta$ , en traçant le ratio  $\delta/\alpha$  en fonction du % HbA<sub>2</sub>,  $\delta$  est unique de la forme HbA<sub>2</sub> et  $\alpha$  représente l'ensemble des formes. L'imprécision de la méthode est de 3.5% à 5.3% et permet donc de différencier les personnes atteintes d'un trouble de l'hémoglobine ( $> 3.7\%$ ) des personnes saines ( $< 3.5\%$ ). La chaîne  $\alpha$  de l'hémoglobine est utilisée comme standard interne. Le SILAC neucode permet aussi la quantification en top-down. Le marquage se fait au niveau des lysines  $^{13}\text{C}_6$  $^{15}\text{N}_2$ -lysine ou  $^2\text{H}_8$ -lysine des protéines. Différents ratios ont été testés avec succès mais il est nécessaire d'avoir une haute résolution ce qui n'est pas adapté aux études protéomiques par top-down et cette méthode est pour l'instant limitée à l'annotation manuelle des protéines<sup>239</sup>. Une seconde approche de marquage métabolique a été développée par Quijada et al.<sup>298</sup> en cultivant les levures avec du glucose  $^{13}\text{C}$  ou de l'eau  $^2\text{H}$ . Cette méthode est linéaire sur 2 log et a des performances comparables au SILAC. En revanche, il y a un risque de superpositions des massifs isotopiques des protéines « légères » et « lourdes », une attention particulière doit être portée à l'annotation. Contrairement à l'approche bottom-up l'utilisation de protéines marquées aux isotopes stables n'est pas encore répandue pour la quantification intacte/top-down. Lanshoeft et al., ont développé une méthode pour la quantification d'anticorps thérapeutique dans le sérum de rat par

immunoprécipitation. Cette méthode à l'originalité d'utiliser un anticorps  $^{13}\text{C}$ -labeled pour la quantification.

En conclusion, le Top-down est une méthode moins sensible que le bottom-up et seulement dans des cas précis le top-down peut s'approcher des limites de sensibilités du bottom-up<sup>192</sup>. Les méthodes top-down ne nécessitent pas une étape de digestion enzymatique et demande donc un temps d'analyse plus court. En termes de performance analytique la gamme dynamique est de seulement  $2\log^{292}$  et une répétabilité  $< 10\%^{295}$ , la gamme dynamique est donc plus limitée que pour les approches en bottom-up. La préparation d'échantillons doit être adaptée aux protéines intactes ainsi que la séparation chromatographique. Contrairement au bottom-up le choix d'étalons internes est limité aux protéines marquées aux isotopes stables comme le PSAQ. En effet, les marquages chimiques avec des tags ne semblent pour l'instant pas adaptés à l'analyse par top-down en matrices biologiques. De même, pour le marquage métabolique dont les massifs isotopiques entre deux contributions peuvent se superposer. C'est pourquoi les analyses par top-down sont majoritairement réalisées en « label-free ».

Plus récemment, le mode PRM (haute résolution) a montré des performances similaires au bottom-up pour la quantification de la Somatropine (22 kDa) avec une LOD de 0,57 fmol pour le top-down contre 0,41 pour le bottom-up fmol. La quantification par les méthodes ciblées utilisées avec les approches bottom-up est donc possible, mais la sensibilité des méthodes top-down dépend fortement de la protéine analysée. Les méthodes de quantification, basées sur la dérivation de l'échantillon, exposées précédemment comme l'iTRAQ<sup>299</sup>, le TMT ou le DML peuvent aussi être appliquées aux protéines intactes pour leurs quantifications. Dans ce cas le marquage se fait donc au niveau de la protéine et non plus au niveau du peptide. Un TMT 6-plex a été appliqué avec succès pour des protéines  $< 35\text{kDa}$  en tampon avec une gamme dynamique de 1 log10 par Hung et al.<sup>300</sup> L'application du DML a été faite avec la myoglobine (17 kDa) montrant une gamme dynamique linéaire 5 fois plus importante que celle du TMT et cette méthode a été appliquée à la quantification de 33 protéines provenant d'une lignée cellulaire hépatique. Bien que ces méthodes semblent prometteuses pour la quantification en top-down/intactes, leur application en matrices biologiques complexes reste encore à démontrer.

Comme il n'y a pas d'étape de digestion l'analyse des protéines intactes permet donc de gagner du temps et est particulièrement intéressante pour l'analyse des protéoformes car il n'y a pas de pertes d'informations.

#### **(iv) Quantification des protéoformes**

La quantification des protéoformes ne peut se faire en théorie qu'en développant une approche top-down. En effet, l'étape de digestion des approches bottom-up et middle-down induit un risque de perte d'un peptide contenant une modification et surtout la perte de l'information permettant de lier les modifications entre elles. Il existe des exceptions comme, par exemple, pour la quantification des histones par middle-down, comme exposé précédemment, où la majorité des modifications est présente sur la queue des histones. Un point important pour la quantification des protéoformes est que le traitement de l'échantillon biologique ne doit pas dégrader des modifications post-traductionnelles (phosphorylation : utilisation d'anti-phosphatase) ou induire de nouvelles modifications (déamidation ou oxydation).

Ici, nous nous intéresserons à la particularité de la quantification des protéoformes par rapport aux méthodes de quantification exposées précédemment.

Dans un premier temps les stratégies développées pour la quantification des modifications post-traductionnelles par bottom-up vont être exposées et ensuite les méthodes récentes en top-down pour la quantification des protéoformes.

#### **Quantification modifications post-traductionnelles et des isoformes par bottom-up**

Bien que la caractérisation complète des protéoformes par bottom-up ne soit pas possible<sup>301</sup>, le bottom-up a des avantages pour la quantification des modifications post-traductionnelles notamment en termes de sensibilité et sur le choix des méthodes disponibles pour la quantification.

##### **(a) Modifications post-traductionnelles : phosphorylation et glycosylation**

De nombreuses études se sont penchées sur la phosphorylation car environ 30% des protéines sont phosphorylées à un moment donné, ce qui fait de la phosphorylation une des modifications post-traductionnelles intracellulaire les plus fréquentes. Bien que ce soit une modification fréquente, les peptides phosphorylés restent minoritaires. Une étape de préconcentration est généralement employée avec soit des colonnes IMAC (immobilized metal affinity chromatography) ou des anticorps anti-phosphorylation S, T ou Y. La sensibilité peut être impactée car les peptides phosphorylés ne sont pas nécessairement des peptides répondant aux critères des peptides protéotypique, ayant les meilleures propriétés physico-chimiques pour une

analyse en LC-MS. Une autre difficulté provient du fait que la trypsine ne coupe pas en présence d'un groupement phosphate voisin.

Les peptides phosphorylés peuvent être quantifiés par le biais de peptides AQUA. Dekker et al.<sup>302</sup> ont étudié le changement de niveau de phosphorylation dans des cellules U87 qui sont privées de sérum de veau fœtal en utilisant des peptides marqués des formes phosphorylées et non-phosphorylées pour les quantifier (AQUA), ce qui permet de calculer le taux de phosphorylation. La quantification se fait avec le mode PRM (voir figure 51).

$$\text{Taux phosphorylation (\%)} = \frac{\text{Quantité protéine phosphorylée}}{\text{quantité protéine phosphorylée} + \text{quantité protéine nonphosphorylée}} \times 100$$

L'intérêt de l'approche est la mesure absolue des sites de phosphorylation et du taux de phosphorylation qui permet de quantifier de manière plus précise des changements de phosphorylation. En effet, le taux de phosphorylation des sites de phosphorylation est d'un intérêt particulier car un changement de quantité d'un peptide phosphorylé ne reflète pas nécessairement un changement de phosphorylation mais un changement de quantité de la protéine<sup>303</sup>.

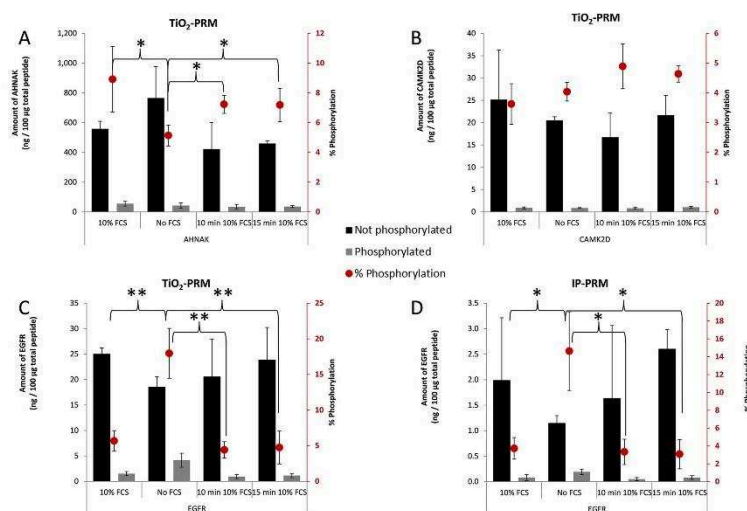


Figure 51 Quantification des formes non-phosphorylées et phosphorylées des protéines AHNAK, CAMK2D et EGFR en fonction des conditions de culture des cellules U87. La condition est : 10%FCS (sérum de veau fœtal), no FCS (les cellules sont en culture sans FCS pendant 3h), 30 min 10% FCS (No FCS puis incubation avec du sérum de veau fœtal pendant 30min) et 15 min 10% FCS (No FCS puis incubation avec du sérum de veau fœtal pendant 15min). Le taux de phosphorylation est représenté par un point rouge. D'après Dekker et al. 2018

L'inconvénient des méthodes ciblées utilisant des peptides AQUA est la limite dans le nombre de peptides pouvant être quantifié (coût important des peptides AQUA). Des méthodes non ciblées, de quantification relative, ont été développées pour la mesure de la stœchiométrie des sites de phosphorylation. En effet, Wu et al.<sup>304</sup>, Tsai et al.<sup>305</sup> ont développé des approches similaires pour déterminer la stœchiométrie des sites de phosphorylation (par marquage chimique) (voir figure 52). Brièvement, l'échantillon est séparé en deux, une moitié est traitée avec une phosphatase alors que la seconde moitié est traitée avec une anti-phosphatase pour garder les groupements phosphates. Dans une seconde étape les peptides tryptiques phosphorylés et non phosphorylés sont marqués soit par un groupement propionyl soit par un groupement diméthyle. Dans le cas de Wu et al., les peptides non phosphorylés légers et lourds sont directement comparés pour obtenir le niveau de phosphorylation d'un peptide. Tsai et al. ont cherché à quantifier et à déterminer des sites de phosphorylation peu abondants, pour cela une première étape consiste à enrichir les peptides non phosphorylés puis à phosphoryler certains motifs peptidiques avec des kinases pour ensuite les enrichir avec une colonne IMAC. La stœchiométrie est déterminée avec l'abondance des peptides lourds et légers. Le désavantage de ces deux méthodes est que si plusieurs sites de phosphorylation sont présents sur un même peptide, uniquement la stœchiométrie globale peut être déterminée et non la stœchiométrie de chaque site de phosphorylation du peptide. De plus ces méthodes sont complexes et des biais peuvent être introduits notamment aux étapes de phosphatases et de réaction avec les kinases.

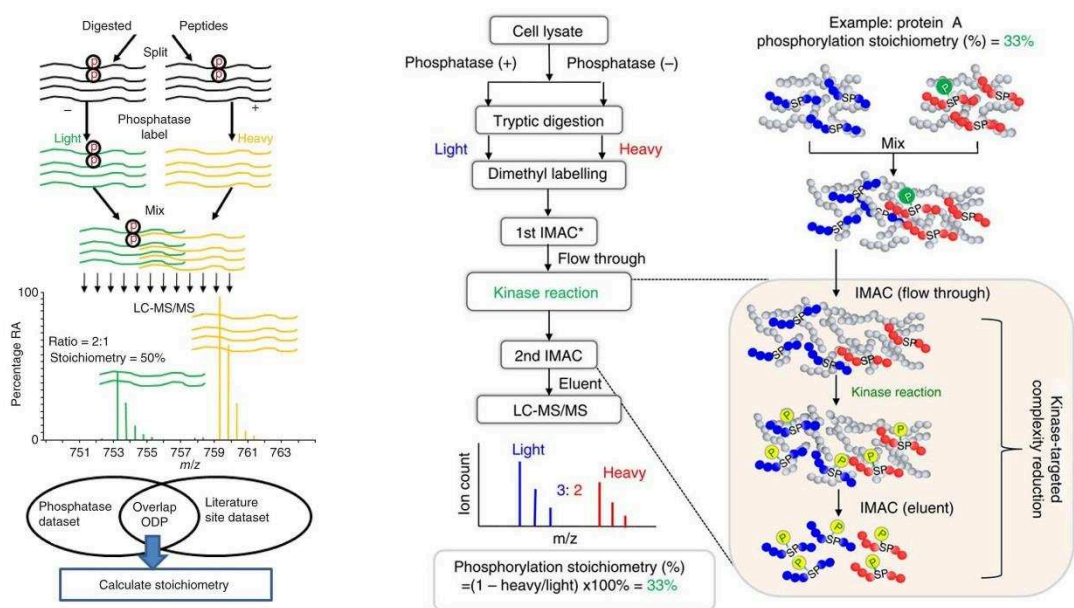


Figure 52 Processus pour la détermination du taux de phosphorylation des peptides. Wu et al. (gauche) et Tsai et al. (droite).

Quatre méthodes de quantification relatives exposés précédemment ont été comparé par Hogrebe et al.<sup>306</sup> pour la quantification des taux de phosphorylation. Les quatre méthodes comparées sont : label-free, SILAC, TMT-MS2 et TMT-MS3. Le label-free comme attendu est la méthode la moins adéquate. Le SILAC a le meilleur compromis entre précision et fidélité mais à de trop nombreux désavantages par rapport au TMT, dans ce cas, comme par exemple un moins bon multiplexage, un nombre de phosphopeptides identifiés moins importants et les changements subtils de phosphorylation ne sont pas détectables. Pour ces raisons le TMT est la méthode de choix et plus particulièrement le TMT-MS2 qui bien que plus interférée que la quantification MS3, permet la quantification de deux fois plus de sites de phosphorylation ce qui est un avantage non négligeable pour les études protéomiques.

Des méthodes du même type peuvent être appliquée à d'autre modifications post-traductionnelles comme les glycosylations par exemple. Hülsmeier et al.<sup>307</sup>, ont développé une méthode de quantification du taux de glycosylation en quantifiant le peptide déglycosylé en présence d' $H_2^{18}O$  par rapport à un peptide synthétique correspondant. Mais comme la PNGaseF, une enzyme permettant la Déglycosylation, peut induire une déamination de l'asparagine (JASMS 2017). Zhu et al.<sup>308</sup> ont développé une méthode sans étape de déglycosylation. La quantification se fait alors en choisissant un peptide représentant la valeur de la protéine totale et ne comportant pas de modifications post-traductionnelles ainsi que la version non glycosylée du peptide glycosylé



à quantifier. Le taux d'occupation est donc 1-(quantité peptide non glycosylé/ quantité protéine totale).

Les méthodes de quantification relative et absolue classiquement utilisées en bottom-up ont donc été appliquées avec succès pour la quantification de modifications post-traductionnelles comme par exemple pour les phosphorylations, glycosylation et également l'acétylation<sup>133</sup>.

(b) Quantification des isoformes :

Les isoformes peuvent aussi être quantifiées par bottom-up. De même que pour les phosphorylations les peptides uniques à une isoforme ne sont pas nécessairement quantotypique ni prototypique, le choix est donc limité. Dans ce cas, l'abondance des différents isoformes peuvent être quantifié par différence, comme par exemple pour les 3 isoformes de la proGRP (proGastrin releasing peptide) impliquées dans les cancers bronchiques à petites cellules.

Isoforme 1

MRGREPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSVSEKSLKQQL  
REYIRWEEAAR

**NLLGLIEAKENRNHQPPQPKALGNQQPSWDESSNFKDVGSKGKVGRLSAPGSQREGRNPQLNQ**

Isoforme 2

MRGREPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSVSEKSLKQQL  
REYIRWEEAAR

**NLLGLIEAKENRNHQPPQPKALGNQQPSWDESSNFKDVGSKGKGSQREGRNPQLNQ**

Isoforme 3

MRGREPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSVSEKSLKQQL  
REYIRWEEAAR

**NLLGLIEAKENRNHQPPQPKALGNQQPSWDESSNFKDLVDSLLQVLNVKEGTPS**

Les isoformes de la proGRP ont des séquences identiques à plus de 83%, ce qui ne laisse que peu de choix pour avoir un peptide unique à chaque isoforme. Les isoformes 1 et 3 sont quantifiés par un peptide unique. Comme l'isoforme 2 n'a pas de peptide prototypique, sa quantification se fait par différence par quantification d'un peptide commun aux 3 isoformes (NLLGLIEAK) : (proGRP isoforme 2) = (proGRP totale) - (proGRP isoforme 1) - (proGRP isoforme 3). Cette méthode a été appliquée avec succès pour la quantification des isoformes de la proGRP en sérum<sup>309</sup>.

(c) Troncations :

Les troncations des protéines sont issues des protéases (cf Introduction). Ces protéases produisent des formes tronquées des protéines qui peuvent être abondantes. En revanche, contrairement aux isoformes où la séquence est connue, les sites de troncations d'une protéine par les protéases endogènes ne sont pas prédictibles. C'est pourquoi pour leur identification des méthodes sans *a priori* doivent être développées.

La quantification directe des modifications post-traductionnelles pour les protéines faiblement n'est pas évidente comme c'est le cas pour la protéine tau dans le LCR (< ng/mL) uniquement les approches ciblées par bottom-up peuvent permettre, avec les technologies actuelles, sa quantification. L'idée original de Barthélemy et al.<sup>101</sup>, a été de couvrir au mieux la séquence de la protéine en quantifiant 17 peptides par rapport à une protéine marquée 15N, intacte et non modifiée. Comme illustré par la figure 53, les peptides centraux sont les plus concentrés (156 à 224) en comparaison avec les extrémités N- et C-terminale suggérant l'existence d'abondantes modifications post-traductionnelles comme des troncations dans ces régions. Sur le même modèle et idée, Mair et al.<sup>161</sup> ont montré pour la protéine tau que les différences de concentration entre les peptides peuvent être expliqués aussi par des phosphorylations, ce qui permet une détermination indirecte du taux de modifications des peptides. De plus, les peptides spécifiques des isoformes 1N et 2N ; 3R et 4R ont aussi été quantifiés.

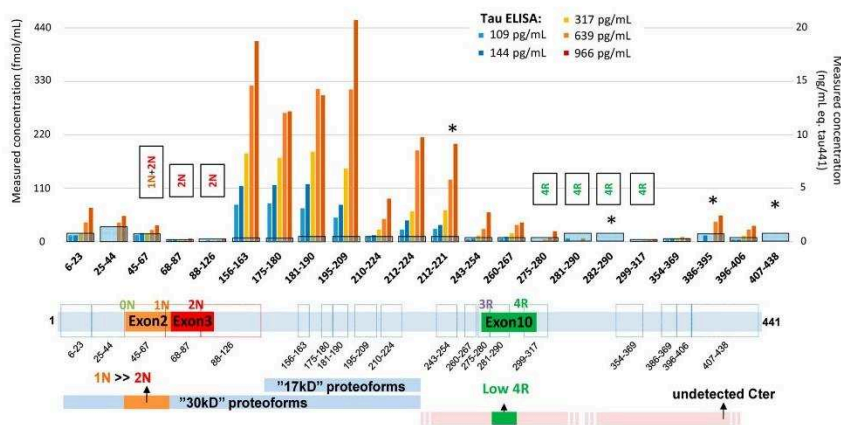


Figure 53 Quantification absolue multiplexe de la protéine tau dans le LCR. D'après Barthélemy et al. 2016

(d) Approche intacte et comparaison bottom-up

La transthyréline (TTR) est une protéine plasmatique extrêmement modifiée entre 85-90%. Du fait de sa taille 16 kDa et sa concentration 0.2-0.4 µg/mL la transthyréline est un bon modèle pour comparer la quantification des protéoformes par bottom-up et par top-down<sup>310</sup>. Une immunoprécipitation de la TTR du sérum est réalisée comme préparation d'échantillon.

Pour le bottom-up, la TTR est aussi un bon exemple où combiner les enzymes est nécessaire pour pouvoir quantifier certaines modifications post-traductionnelles. En effet, la cystéine (qui est un site de modification) en position 10 est située entre deux lysines en 9 et 15 ce qui par digestion à la trypsine produit un peptide trop petit, non détecté (voir figure 54). L'enzyme sélectionnée est l'ArgC.

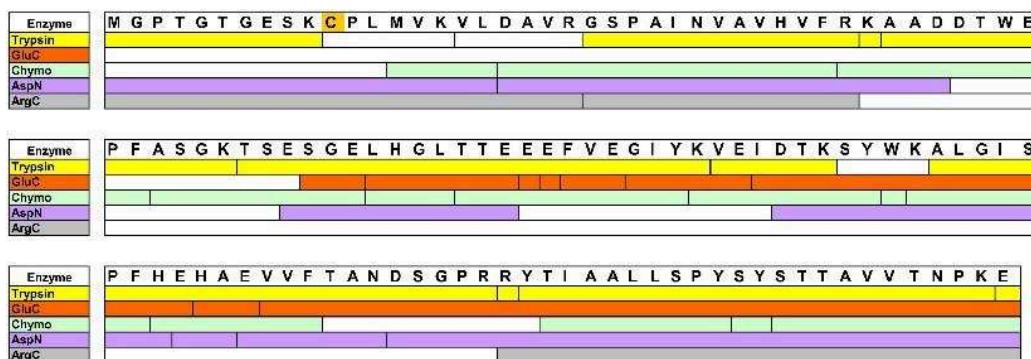
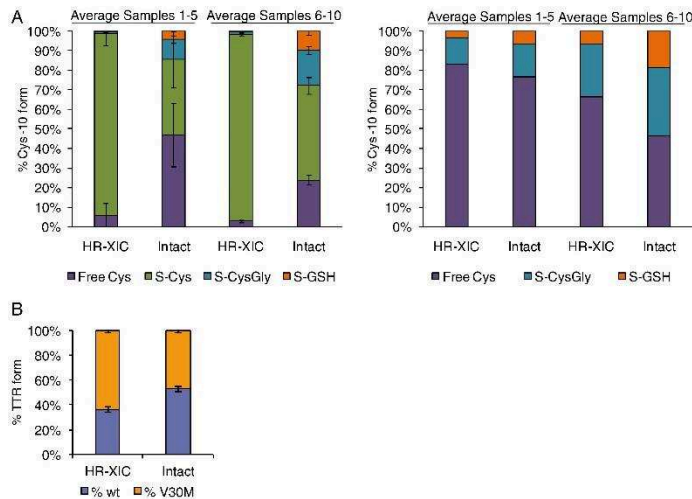


Figure 54 Peptides détectés pour la TTR en fonction de l'enzyme utilisée. D'après Vila rico et al. 2015

La quantification par bottom-up se fait, après séparation chromatographique, sur la base du XIC en MS1 en utilisant un standard AQUA pour chaque modification post-traductionnelle. Pour la quantification en intact des protéoformes se fait par infusion et l'analyse par un 7T-LTQ-FT ultra sur la base d'un état de charge ( $z +14$ ). La comparaison des approches (voir figure 55) montre des résultats différents qui peuvent être expliqués par l'absence de standard pour l'intact et donc que l'hypothèse faite sur une égale réponse des protéoformes est fautive. Il semble nécessaire pour avoir une quantification absolue d'ajouter des standards pour une quantification en intacte. De plus, en intact les m/z de la forme S-sulfo de la TTR V30M et S-cys w.t. se recouvre, dans ce cas une approche top-down semblerait plus pertinente.



**Figure 55** Comparaison des résultats de quantification par bottom-up HR-XIC et en intacte. (A) quantification des modifications en Cys-10 et en excluant la forme majoritaire S-cys (à droite). (B) pourcentage des formes wt et V30M. D'après Vila rico et al. 2015

Une question importante pour la quantification des protéoformes en intactes est l'influence des modifications sur la réponse en spectrométrie de masse. Par exemple, la phosphorylation impacte l'ionisation par électrospray ce qui rend la quantification directe d'un peptide avec son correspondant phosphorylé impossible<sup>311</sup>. Au niveau de la protéine une monophosphorylation n'affecte pas l'électrospray (pour une protéine de 26 kDa) et donc une quantification relative de la phosphorylation d'une protéine est possible. En revanche, pour une protéine multi phosphorylée comme la caséine  $\beta$  qui comprend 5 groupement phosphates, l'ionisation par électrospray est impactée et donc la proportion de caséine  $\beta$  ne peut pas être déterminée à partir de la forme non phosphorylée<sup>304</sup> (voir figure 56). Avec la même idée, le groupe de Kelleher a introduit le Protein ion relative ratio (PIRR)<sup>312</sup> qui somme les 5 isotopes les plus abondants pour comparer l'abondance des protéoformes qui co-eluent. Le PIRR a permis la quantification d'histone et mis en évidence qu'une histone acétylée aura une réponse équivalente à sa version non-acétylée au niveau de la protéine alors qu'au niveau peptidique la variation est plus grande. Ces résultats montrent les précautions à prendre pour la quantification par top-down des protéoformes.

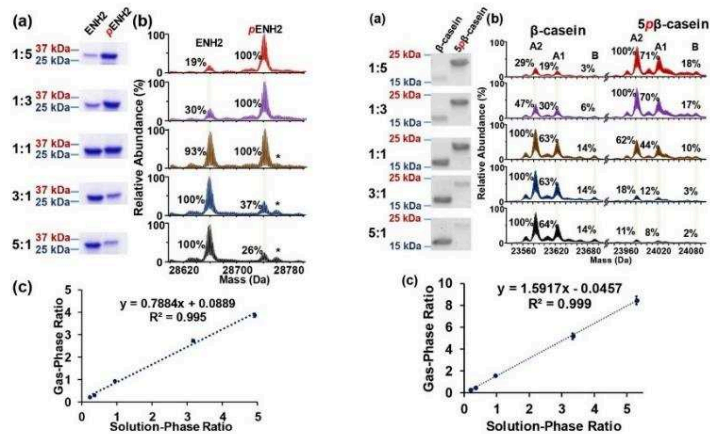


Figure 56 (a) quantification par SDS-PAGE à 5 ratio différents 1:5, 1:3, 1:1 ; 3:1 et 5:1 pour l'ENH2(droite) et la  $\beta$ -caséine (gauche). (b) quantification obtenus par spectrométrie de masse et (c) corrélation entre les résultats obtenus par SDS-PAGE et spectrométrie de masse. D'après Wu et al. 2018

Le bottom-up ne permet pas toujours l'obtention de peptides spécifique d'une isoforme ou d'une forme tronquée d'une protéine. C'est pourquoi Kellie et al.<sup>210</sup>, ont quantifié la proportion des différentes formes tronquées de l'alpha-synucléine, qui est une protéine majoritaire, avec une protéine marquée, intacte et non modifiée 15N, dans le cerveau humain. Cette matrice biologique étant complexe une immunoprécipitation de l'alpha-synucléine est réalisée avec un anticorps monoclonal dont l'épitope est 91-99.

La quantification se fait en utilisant le modèle suivant :

$\text{Log}_{10}(\text{AUC}_{ij}) = \mu + Z_i + S_j + \varepsilon_{ij}$  avec :  $\text{AUC}_{ij}$  est l'aire du pic de l'état de charge  $i$  de l'échantillon  $j$  ;  $Z_i$  est un terme représentant l'état de charge  $i$  ;  $S_j$  est un terme représentant la variation de concentration de la protéine de l'échantillon  $j$  ;  $\mu$  est la moyenne ;  $\varepsilon_{ij}$  correspond à l'erreur résiduelle.

Le modèle assume une réponse égale des différentes protéoformes de l'alpha-synucléine. La limite de cette méthode est son application en milieux biologiques complexes sans immunoprécipitation.

Cette méthode a permis la détection et la quantification de 11 protéoformes (troncations) de l'alpha-synucléine. Aucune troncation ne s'est avérée discriminante de la maladie de Parkinson, dans le cortex frontal, ce qui peut être expliqué par le fait que c'est un évènement normal dans le cerveau<sup>313</sup>. De plus l'immunoprécipitation sélectionnée ne semble pas optimale car l'épitope de

l'anticorps monoclonal utilisé est 91-99 or l'alpha-synucléine est connue pour avoir de nombreuses tronctions autour de l'acide aminé 91<sup>314</sup>, ces formes ne seront donc pas captées par l'anticorps.

Des approches ciblées peuvent aussi être mise en place. Dans ce cas l'immunoprécipitation classiquement utilisé en bottom-up est une méthode de choix et a été appliqué avec succès pour la quantification de protéoformes comme par exemple la quantification multiplexe des protéoformes de la protéine C réactive, sérum amyloïde A, la cystatine C et les protéines S100A8/9<sup>125</sup>, la quantification des protéoformes de KRAS4b qui est impliquée dans divers cancers<sup>209</sup> et pour la quantification de la troponine cardiaque I porcine<sup>315</sup>. Dans ce dernier exemple Zhang et al<sup>315</sup> ; ont mis en évidence une corrélation en le taux de phosphorylation de la troponine I la sévérité de la maladie dans des tissus post-mortem et des tissus sains. En effet, le taux de phosphorylation diminue avec la sévérité de la maladie.

Un autre exemple concerne la capacité élevée ou basse d'efflux du cholestérol pour la lipoprotéine de haute densité ApoA-1. Onze modifications post traductionnelles différentes réparties en 18 protéoformes ont été détectées, parmi ces protéoformes 6 comportent plus d'une modification. De plus, 6 protéoformes sont significativement élevées ( $p < 0.005$ ) pour l'ApoA-1 dans le cas d'une capacité élevée d'efflux du cholestérol dont un détectable et quantifiable uniquement par top-down comprenant une tronction et une acylation par l'acide palmitique<sup>316</sup>.

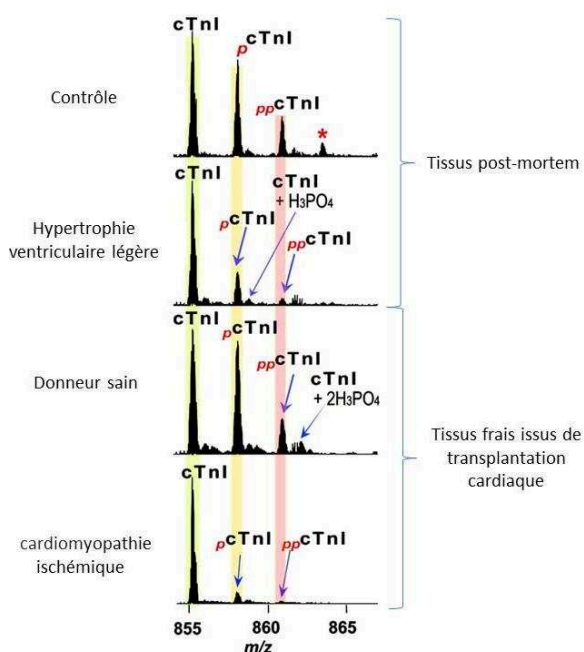


Figure 57 quantification de la troponine cardiaque 1 porcine. D'après Zhang et al. 2011

Les approches bottom-up du fait de leurs sensibilités permettent de quantifier des modifications post-traductionnelles minoritaires mais les récentes avancées concernant la préparation des échantillons et des spectromètres de masses ont favorisé le développement des approches top-down et intactes. La quantification des protéoformes est, pour l'instant, essentiellement basé sur des méthodes de « discovery » sans standards internes du fait du coût de synthétiser une protéine marquée par protéoforme ainsi que de la complexification du spectre pour les approches SILAC. Récemment<sup>316</sup>, une approche ciblée concernant la préparation des échantillons (immunoprécipitation) a permis de mettre en évidence un « cross-talk » entre les différentes modifications post-traductionnelles de KRAS4b ce qui met en évidence l'intérêt des approches par top-down pour comprendre des processus biologiques complexes. Bien que l'immunoprécipitation soit l'approche permettant généralement d'atteindre les meilleures sensibilités son application dans le cadre des protéoformes doit être caractérisée car les anticorps peuvent discriminer des formes comme c'est le cas pour l'alpha-synucléine<sup>210</sup>. Dans la perspective de quantification des biomarqueurs un travail doit être effectué pour la recherche de standards internes pour le top-down car les approches actuelles sont limitées<sup>300</sup>.

# Conclusion de l'étude bibliographique et objectifs de la thèse

Nous avons mis en évidence lors de l'étude bibliographique la diversité des protéines et de leurs protéoformes, et quelques exemples du lien entre cette diversité et la physiopathologie. Le développement de méthodes analytiques sensibles, robustes et spécifiques pour caractériser et quantifier ces protéoformes dans les échantillons biologiques est capital. La quantification des biomarqueurs protéiques et de surcroît leurs protéoformes est actuellement limitée par plusieurs obstacles. La gamme dynamique importante des protéines en matrices biologiques ( i.e. plasma : 10log) implique la nécessité d'une préparation d'échantillon spécifique et/ou de méthodes d'analyses performantes, pour la quantification de protéines minoritaires. De ce fait, la spécificité et la capacité de multiplexage de la spectrométrie de masse, et notamment à haute-résolution, en font un outil de choix. La stratégie bottom-up, qui consiste à découper la protéine en peptides, est l'approche la plus couramment utilisée. En effet, les approches de type discovery ou quantification relative ont bénéficié d'avancées technologiques récentes en termes de sensibilité (nouvelles méthodes d'acquisition comme le DIA, par exemple) et de protocole de quantification par marquage isotopique. De plus, les méthodes ciblées quantitatives par bottom-up permettent maintenant d'atteindre des protéines minoritaires ( $\approx$  ng/mL) de manière robuste. En revanche, en découpant la protéine en peptides le lien entre les modifications post-traductionnelles est généralement perdu, ce qui en conséquence limite la caractérisation des protéoformes. Une méthode émergente pour l'étude des protéoformes est l'approche top-down où l'étape de digestion des protéines est supprimée. Cette méthode encore récente est plus complexe, notamment pour la préparation des échantillons et l'analyse des données. De plus, elle souffre pour le moment de la comparaison en termes de sensibilité avec les approches bottom-up. Probablement pour ces raisons, cette approche est principalement utilisée en discovery. Des méthodes de quantification absolue par top-down se développent en matrices biologiques, mais reposent souvent sur une immunoprécipitation préalable de l'échantillon, ce qui pose la question de la disponibilité d'un anticorps de bonne affinité et de la possible discrimination des protéoformes à cette étape. Il faut également noter le besoin de standards internes adaptés à la quantification par top-down.



Ce travail de thèse a porté sur le développement de méthodes de quantification par spectrométrie de masse pour des protéines et des protéoformes impliquées dans des maladies neurodégénératives. La découverte de nouveaux biomarqueurs et la meilleure compréhension de la physiopathologie des maladies neurodégénératives est un enjeu crucial dans le domaine biomédical. En effet, les mécanismes impliqués dans les maladies neurodégénératives sont encore peu compris. Mieux comprendre ces mécanismes aiderait d'une part au développement de thérapies, en identifiant des cibles potentielles, et d'autre part à la découverte de biomarqueurs. Le besoin de biomarqueurs est important tant pour détecter le plus précocement possible une maladie que pour discriminer des maladies cliniquement proches à un stade précoce ou pour suivre l'effet d'un traitement thérapeutique. De plus en plus, le domaine des biomarqueurs se tourne vers la combinaison de plusieurs biomarqueurs plutôt que l'analyse d'un simple biomarqueur afin d'augmenter la spécificité et la sensibilité des tests. L'idée de cette thèse a été d'exploiter les progrès technologiques des approches bottom-up et top-down en tentant de combler certaines des limites actuelles, notamment du top-down, pour apporter de nouvelles données sur l'expression de protéines et de leurs proteoformes, en lien avec deux maladies neurodégénératives : les démences frontotemporales et les synucléinopathies. Il sera ainsi présenté le développement d'une méthode de quantification des isoformes de la protéine C9ORF72 par une approche bottom-up sans anticorps, dans le cortex frontal de sujet sains ou atteints de démences frontotemporales, avec ou sans une expansion hexanucléotitique  $(G_4C_2)_n$  présent dans le gène *C9ORF72*. Cette méthode a ensuite été étendue à la quantification multiplexe de 49 protéines dans la même cohorte d'échantillons avec pour objectif de mettre en évidence des protéines impactées dans cette maladie afin de mieux comprendre la physiopathologie de la maladie et potentiellement, d'identifier des cibles thérapeutiques. Une approche non ciblée de type DIA a ici été évaluée en parallèle du mode PRM. La seconde partie de la thèse a été consacrée à l'étude de l'alpha-synucléine, protéine hétérogène par de nombreuses modifications post-traductionnelles. De nouvelles approches de quantification top-down et bottom-up ont été développées en s'affranchissant de l'étape d'immuno-affinité, optimisées puis évaluées dans des échantillons de cerveau et liquide céphalo rachidien (LCR). Nous nous sommes notamment intéressés à l'identification puis à la quantification des tronçons de la partie C-terminale de la protéine.

Une mutation liée à une expansion d'un hexanucléotide  $(G_4C_2)_n$  dans le gène *C9ORF72* est la première cause familiale de démences frontotemporales et de scléroses latérales amyotrophiques. De plus, cette protéine fait le lien entre les deux maladies puisque 15% des

personnes atteintes de SLA développeront une DFT<sup>317</sup>. Ce gène produit par épissage alternatif deux isoformes, une forme courte (222 AA) et une forme longue (481 AA) de la protéine C9ORF72 dont la fonction n'est pas clairement définie. De plus, l'existence dans le cerveau de l'isoforme court est toujours sujet à débat. Les données de quantification au niveau de la protéine et de ses isoformes ont été jusqu'à présent uniquement réalisées par Westernblot avec des anticorps non commercialement disponibles ce qui limite l'applicabilité des méthodes proposés. De plus, la spécificité des anticorps utilisés est questionnable notamment pour l'isoforme court dont les données divergent. Nous nous sommes donc intéressés au développement d'une méthode de quantification de ces isoformes par LC-MS/MS en s'affranchissant d'anticorps et pouvant être facilement transposable à d'autres laboratoires. La méthode a été validée et appliquée à une cohorte d'échantillon de cerveaux humains. Ce travail a donné lieu à une publication dans le journal *Frontiers in Neuroscience*. Le second objectif de ce travail a été d'élargir la liste des protéines ciblées pour aider à une meilleure compréhension des mécanismes impliqués les DFT. Pour cela une méthode de quantification multiplexe a été développée comprenant 56 protéines, sélectionnées d'après des données transcriptomiques et une étude bibliographique. Une méthode de protéomique non-ciblée de type DIA a aussi été évaluée dans cette application. Ce travail constitue le premier chapitre de la partie résultat du manuscrit.

L'alpha-synucléine est une protéine intrinsèquement désordonnée et impliquée dans un groupe de maladie neurodégénératives appelé synucléinopathies. Les synucléinopathies regroupent notamment la maladie de Parkinson, les démences à corps de Lewy et l'atrophie multisystématisée. Elles sont caractérisées par la présence de corps de Lewy dans le cerveau où des agrégats d'alpha-synucléine sont retrouvés. L'alpha-synucléine est une protéine hétérogène comprenant de nombreuses modifications post-traductionnelles dont certaines comme les troncations, causées par la protéolyse endogène, ou la phosphorylation semblent impacter sa capacité à former des agrégats. Bien que l' $\alpha$ -synucléine semble jouer un rôle central dans le développement des synucléinopathies, son utilité en tant que biomarqueur n'est pas encore prouvé pour plusieurs raisons. Premièrement les résultats des test ELISA divergent ce qui est probablement lié à un manque de spécificité des anticorps utilisés. De plus, l'alpha-synucléine est hétérogène et il n'est pas à exclure qu'un ou des protéoformes soient discriminants et pas d'autres ce qui pose problème pour les ELISA. Les méthodes de quantification par spectrométrie de masse publiées jusqu'à présent ne permettent pas la quantification de la partie C-terminale de la protéine où de nombreuses troncations ont pourtant été rapportées. D'autre part, les maladies neurodégénératives peuvent être cliniquement proches notamment à un stade précoce de la

maladie. La combinaison de biomarqueurs permettrait d'établir un diagnostic plus précis des cas suspects. C'est pourquoi nous avons développé une méthode de quantification multiplexe par LC-MS/MS de Tau et de l'alpha-synucléine, en incluant la partie C-terminale. En parallèle une méthode top-down a été développée pour l'identification et la quantification des troncations de l'alpha-synucléine dans le cerveau. Ces résultats seront discutés dans le chapitre 2 de ce manuscrit.

Chapitre 1 : Quantification de protéines impliquées dans la physiopathologie des DFT. Apport du bottom-up pour la quantification d'isoformes et le développement de méthodes multiplexées.

# I. Introduction : les démences lobaires frontotemporales

Les maladies neurodégénératives sont un problème de santé publique. En 2050, il est estimé que plus de 100 millions de personnes seront atteintes d'une démence, résultant d'une augmentation de la population et de son vieillissement<sup>318</sup>. Il y a un besoin de marqueurs précoce ainsi que d'une meilleure compréhension des mécanismes impliqués dans les maladies neurodégénératives. Dans ce manuscrit les démences lobaires frontotemporales (DLFT) et les synucléinopathies seront abordées.

Les démences lobaires frontotemporales présentent une grande diversité clinique : une forme comportementale, les variants non fluents/agrammatiques et sémantiques d'aphasie primaire progressive. Les démences lobaires frontotemporales sont liées à l'apparition d'inclusion de protéines dans les neurones. Ces agrégats insolubles de protéines sont de trois types : TDP-43 (TAR DNA-binding protein) dans 50 à 60 % des cas<sup>157</sup>, TAU dans 30-40% et enfin la protéine FUS dans 10% des cas<sup>320</sup>. La protéine TDP-43 est une protéine nucléaire qui dans ces démences se délocalise vers le cytoplasme où elle agrège. Le passage de la TDP-43 du noyau au cytoplasme est aussi présent dans d'autres maladies neurovégétatives où des protéines agrègent comme la maladie d'Alzheimer, de Parkinson et d'Huntington<sup>321</sup>. Les protéines TDP-43 et FUS sont des protéines se liant à l'ADN et à l'ARN. Alors que la protéine TAU est associée à la stabilisation des microtubules. Les DLFT sont à 70% sporadique et 30% familiale. A ce jour, 19 gènes ont été trouvés comme impliqués dans les DLFT. Ici, seulement les 3 mutations les plus fréquentes, qui représentent 80% des cas familiaux et 10-15% des formes sporadiques, seront abordées : *C9ORF72*, *GRN/PGRN* et *MAPT*.

Les DLFT sont aussi liées à la sclérose latérale amyotrophique (SLA). 15-18% des patients ayant une SLA ont aussi une DLFT<sup>322</sup>. D'un point de vue génétique les deux maladies se ressemblent<sup>323</sup>.

## 1. Mutation du gène *C9ORF72*

Cette mutation bien que présente dans 25% des formes familiales de DFT<sup>324,325</sup> (n'a été que découverte récemment en 2011. La mutation correspond à une expansion d'un hexanucléotide (G<sub>4</sub>C<sub>2</sub>)<sub>n</sub> dans le premier intron du gène de la *C9ORF72* pouvant aller jusqu'à plus milliers. Le nombre d'expansion est dépendant des tissus<sup>326</sup>. Il n'y a pas de corrélation claire entre début de la maladie et la taille des expansions bien qu'il semble qu'un nombre important de répétition soit

associé à un âge de début de maladie plus tardif<sup>327</sup>. Un nombre élevé d'expansions est aussi corrélé avec l'âge de décès, il n'est donc pas à exclure que le nombre de répétitions augmentent avec l'âge. Trois mécanismes de toxicité sont supposés comme pouvant conduire au développement de la maladie : une perte de fonction de la C9ORF72, les foci d'ARN et la traduction sens et antisens de l'hexanucléotide induisant la production de protéines dipeptides.

- La protéine C9ORF72 est encodée par 3 transcrits issus d'épissage alternatif du gène *C9ORF72* produisant 2 isoformes une forme courte de 222 acides aminés (C9-S) et une forme longue de 481 acides aminés (C9-L). Un transcrit V1 encode pour la C9-S et 2 transcrits V2, V3 encodent pour la C9-L. Il a été montré que le transcrit V2 est plus abondant que V1 et V3 dans le système nerveux central. La C9ORF72 n'a pas une fonction clairement définie, il est supposé qu'elle joue un rôle dans l'autophagie en interagissant avec des facteurs d'échange de Rab<sup>328</sup>. Il a été montré que l'ARN messager de ces transcrits est diminué chez les patients ainsi qu'au niveau de la protéine d'environ 50% dans le cortex frontal par Westernblot . La présence ou non de l'isoforme C9-S dans le cerveau est encore matière à débat du fait du manque d'anticorps commerciaux de bonne qualité, bien que cet isoforme semble minoritaire dans le cerveau . Shi et al.<sup>331</sup>, ont récemment montré qu'une haploinsuffisance de la C9ORF72 conduisait au développement d'une neurodégénérescence dans des cellules souche pluripotentes induites en neurone moteurs. Par deux mécanismes : accumulation de récepteurs au glutamate (excitotoxicité) et une mauvaise dégradation des protéines dipeptides produits par la traduction non AUG dépendante de (G<sub>4</sub>C<sub>2</sub>)<sub>n</sub>. Le rôle de la C9ORF72 est encore peu connu. La C9ORF72 interagit avec les endosomes, permet le trafic normal des vésicules et des lysosomes . Les modèles de souris K.O. C9ORF72 présente aussi un défaut d'autophagie<sup>332</sup>. L'hypothèse d'une haploinsuffisance peut être soutenue par une étude récente montrant que les neurones ont pu être sauvés en restaurant le rôle de la C9ORF72 par l'activation de Rab5<sup>331</sup>.
- Les foci d'ARN séquestrent les protéines se liant à l'ARN<sup>333</sup> et empêchent la maturation de nombreux ARN.
- Les protéines dipeptides sont produites par traduction non AUG dépendante de l'hexanucléotide. Elles sont au nombre de 6 (P-R ;P-A ;G-P ;G-A ;G-R et G-P)<sup>334</sup>. Ces protéines s'accumulent et agrègent dans les neurones. Ces inclusions sont positives à la sequestosome 1 et négatives à la protéine TDP-43. Le rôle toxique de ces agrégats n'est pas encore clair en effet, ces agrégats peuvent être toxiques ou au contraire peuvent

avoir un rôle de neuroprotection et alors plutôt les formes solubles seraient toxiques. Les poly-GP<sup>335-337</sup> sont significativement augmentés dans les cas de démence avec expansion C9 dans le LCR et pourraient être utilisés comme biomarqueurs. Les poly-GP ont été choisis car ils sont abondants, stables, produits par sens et anti-sens et ils sont plus solubles que les plus abondants poly-GA. En revanche, ils ne permettent pas la discrimination entre patients symptomatiques et asymptomatiques ce qui suggère qu'ils sont produits par les cellules saines. Les neurofilaments légers et les neurofilaments lourds phosphorylés sont des marqueurs de la dégradation axonale et donc permettent de différencier les symptomatiques des asymptomatiques. Ils pourraient être utilisés en combinaison des poly-GP.

## **2. Mutation du gène *GRN/PGRN***

Ces mutations ont été identifiées en 2006<sup>338,339</sup> et sont présentes dans 15-20% des cas familiaux de DLFT et dans 5% des cas sporadiques en France. Ce gène produit une glycoprotéine de 63.5 kDa la progranuline<sup>340</sup>, la progranuline peut être ensuite découpée par des protéases (métalloprotéinases matricielles) en 7 (1/2) granulines de 6 kDa contenant 12 cystéines et 6 ponts disulfure. La progranuline serait impliquée dans l'inflammation, la réparation des tissus et aurait un rôle neurotrophique et neuroprotecteur. La pro granuline et les granules semblent avoir un rôle différent notamment dans l'inflammation où les granules seraient pro-inflammatoire contrairement à la progranuline. 150 mutations sont connues sur ce gène et ces mutations conduisent à la production d'un ARN messager muté ne possédant pas de codon STOP ce qui entraîne sa dégradation et ce qui conduit à une perte de fonction partielle de la progranuline qui n'est alors traduite qu'à hauteur de 50%. Une haploinsuffisance de la progranuline est suspectée comme étant la cause de neurodégénérescence.

Le dosage de la progranuline plasmatique permet de confirmer une suspicion de DLFT associée à une mutation du gène PGRN.

### **3. Mutation du gène *MAPT***

Les mutations de ce gène ont été identifiées en 1998 et sont présentes, en France, dans 10% des cas familiaux de DLFT et dans 3% des cas sporadiques.

La protéine TAU est constituée 6 isoformes qui sont produits par épissage alternatif de l'ARN messenger. Notamment l'épissage alternatif de l'exon 10 produit 3 isoformes comprenant trois répétitions « microtubules binding repeats » (3R) et 3 isoformes en comprenant 4 (4R). Les exons 2 et 3 peuvent aussi être insérés dans la protéine : 0N aucuns des exons, 1N : exon2, 2N : exons 2 et 3. La concentration des isoformes 3R et 4R dans le cerveau est équivalente. 53 mutations ont été identifiées et elles sont principalement localisées dans le domaine codant pour les « microtubules binding domains » qui permettent de stabiliser les microtubules<sup>341</sup>. Des agrégats de la protéine TAU peut être observés dans les cerveaux des patients ayant une mutation MAPT.



## II. Matériels et méthodes

Le matériel correspondant à la quantification des isoformes de la protéine C9ORF72 est décrit dans l'article (I.). Sont indiqués ici, les conditions d'immunoprécipitation, et les paramètres des analyses multiplexes. Les points particuliers qui ont nécessités des optimisations sont décrits dans la partie expérimentale

### 1. Réactifs

L'anticorps polyclonale provient de chez GeneTex (reference GTX119776) et l'anticorps monoclonal de chez ProteinTech (cat#66140-1-iG). Les billes fonctionnalisées M-280 tosylactivated (Dynabeads M-280 tosylactivated) et protéine G (dynabeads protein G) ont été obtenues chez Invitrogen (Carlsbad, CA, USA).

### 2. Méthodes d'extraction

#### a) immunoprécipitation

M-280 tosylactivated : Pour 9  $\mu\text{L}$  de la solution d'anticorps, 31  $\mu\text{L}$  de billes sont prélevées puis lavées deux fois avec 500  $\mu\text{L}$  de tampon phosphate. Ensuite, les bille sont reprises dans 41  $\mu\text{L}$  de tampon phosphate et 9  $\mu\text{L}$  d'anticorps sont ajoutés. La solution est incubée à 37°C sur la nuit pour fixer les anticorps sur les billes. Les billes sont alors fixées sur le tube eppendorf à l'aide d'un aimant et la solution est retirée. Les billes sont alors incubées avec du PBS +0,5% BSA pendant 1 heure à 37°C. Après cette étape, les billes sont lavées deux fois avec 500  $\mu\text{L}$  de PBS+ 0,1% BSA et sont reprises dans 44  $\mu\text{L}$  de ce tampon. La solution est alors dispensée dans les échantillons et les échantillons sont mis à agiter sur table rotative pendant 1 heure. Les billes sont alors lavées par 2\*500  $\mu\text{L}$  de PBS puis une fois 200  $\mu\text{L}$  d'eau. Enfin les billes sont mises à sec. Pour la digestion enzymatique, les biles sont reprises dans du rapigest 0,05% (bicarbonate d'ammonium 50 mM) puis 10  $\mu\text{L}$  de dithiothréitol à 20 mM est ajouté. La solution est alors agitée 15 min à 95°C. Après 10  $\mu\text{L}$  à 45 mM de iodoacétamide est ajouté et la réaction est faite dans le noir pendant 45min. Puis 5  $\mu\text{L}$  de trypsine (Promega) à 100  $\mu\text{g}/\text{mL}$  est ajoutée et la digestion est faite à 37°C pendant la nuit. Enfin, 5  $\mu\text{L}$  de chlorure d'hydrogène à 1M est ajoutée et la solution est agitée 45 min à 37°C. Après une étape de centrifugation, les billes sont retirées et la solution mis en vial.

Protéine G : protocole identique au protocole M-280 sauf que le greffage de l'anticorps se fait par agitation sur une table rotative pendant 30 min dans un tampon PBS.

### **b) Lyse ultrason**

Les PBMCs sont repris dans 200  $\mu\text{L}$  de tampon de lyse puis homogénéisé à une amplitude de 30% pendant 10s. L'instrument utilisé est l'Ultra Vibra cell 75185 avec une sonde de 3 mm.

### **c) Délipidation**

Ajout de 150  $\mu\text{L}$  d'un mélange Méthanol : chloroforme (2 :1) à l'homogénat de cerveau (50  $\mu\text{L}$ ). Formation d'un « pellet » de protéine entre les deux phases. Après une courte centrifugation de 5s, les phases liquides sont enlevées et le « pellet » est re suspendus dans 20  $\mu\text{L}$  de rapigest 0,05% (bicarbonate d'ammonium 50 mM)

## **3. Méthodes LC-MS/MS multiplexe**

### **a) UHPLC**

Méthode chromatographique similaire à la méthode publiée pour la quantification de la C9ORF72 dans le cortex frontal. Les paramètres d'acquisition qui ont été changés sont le maximum injection time (60 ms) et la résolution (17 500 à 200 m/z).

### **b) Chromatographie capillaire**

La séparation par chromatographie liquide a été réalisé avec une colonne en phase inverse Acclaim PepMap 300 C18 (150 mm  $\times$  0.3 mm; 5  $\mu\text{m}$ , 300 Å; Thermo Fisher Scientific, Bremen, Germany) en 90 min à un débit de 6 $\mu\text{L}/\text{min}$ . Les phases mobiles sont pour la phase A : eau +0.1 % acide formique et pour la phase B 80% acétonitrile, 20% eau et 0.1% acide formique. Après un palier isocratique de 3 min à 4% B, le gradient est augmenté jusqu'à 35% de B pendant 60 min. Puis le pourcentage de B est augmenté jusqu'à 95% pendant 6 min. Enfin la colonne est rééquilibrée à 4% de B pendant 21 min.

Les paramètres de sources sont : débit du sheath gas à 10 a.u., tension du capillaire à 2.8 kV, température du capillaire à 320 °C. Les analyses PRM ont été réalisé en mode d'ionisation positif. La résolution de l'orbitrap est 70 000 à 200 m/z (full width at half-maximum) avec un maximum injection time (MIT) de 250 ms de 5 à 24 min et de 56 à 80 min. De 24 à 56 min la résolution a été abaissée à 35 000 avec un maximum injection time de 200 ms. L' automatic gain control (AGC) est constant à 1e6 et la fenêtre d'isolation des ions de 2 m/z.

### c) DIA

Le DIA a les mêmes paramètres chromatographiques que la méthode PRM multiplexe-capillaire. Les paramètres du Q-exactive sont les suivants : 1 scan full-MS : résolution à 70 000 ( 200m/z), AGC de 3e6 et MIT de 20 ms. Vingt et une fenêtres DIA : résolution à 17 500 (200m/z), AGC 3e6, MIT auto et step NCE : 15,25. Les fenêtres DIA ont des fenêtres d'isolation variables avec un recouvrement de 0.5 Da et sont les suivantes : 320-382.5, 381.5-416.5, 415.5-440.5, 439.5-462.5,461.5-484.5, 483.5-504.5, 503.5-524.5, 523.5-544.5, 543.5-562.5, 561.5-580.5, 579.5-598.5, 597.5-616.5, 615.5-634.5, 633.5-654.5, 653.5-678.5, 677.5-702.5, 701.5-730.5, 729.5-764.5, 763.5-804.5, 803.5-844.5, 843.5-900.

Nous nous intéresserons dans ce premier chapitre à la quantification de protéines et de protéoformes impliquées dans les DFLT, avec l'objectif d'aider à mieux comprendre les mécanismes impliqués dans cette maladie neurodégénérative. Un premier travail a consisté à développer une méthode de quantification des isoformes de la protéine C9ORF72 par une approche bottom-up, ensuite appliquée à une cohorte d'échantillon de cortex frontal. Par la suite, nous avons souhaité diversifier les protéines ciblées et étendre l'analyse à 49 protéines cérébrales en exploitant la capacité de multiplexage de la spectrométrie de masse.

### III. Partie Expérimentale

#### 1. Développement d'une méthode d'immunoprécipitation pour la C9ORF72

Afin de mieux explorer le mécanisme de toxicité lié à une perte de fonction de la protéine C9ORF72 (cf chapitre 1 I.), il est nécessaire de disposer d'une méthode de quantification précise de cette protéine. Au démarrage de la thèse, seule la technique Western Blot était proposée<sup>329</sup> qui repose sur des anticorps dont la spécificité est questionnable. Aucune méthode basée sur la spectrométrie de masse n'avait été publiée, malgré l'intérêt de cette approche (Introduction générale).

Dans un premier temps, nous avons souhaité développer une méthode de détection de la C9ORF72 par immunoprécipitation (cf figure 58) en matrices biologiques (plasma, homogénat de cerveau) suivie d'une détection par spectrométrie de masse ciblée, le mode PRM sur qExactive. Le choix de l'immunoprécipitation s'explique par les faibles concentrations de la protéine C9ORF72, comme par exemple dans le plasma où sa concentration est proche du ng/mL<sup>342</sup>. Les méthodes par immunoprécipitation sont généralement les plus sensibles mais sont dépendantes de la qualité de l'anticorps (cf Introduction II.3). Néanmoins, le couplage immunocapture-

spectrométrie de masse permet de tolérer une moindre spécificité des anticorps, contrebalancé par la détection finale par spectrométrie de masse.

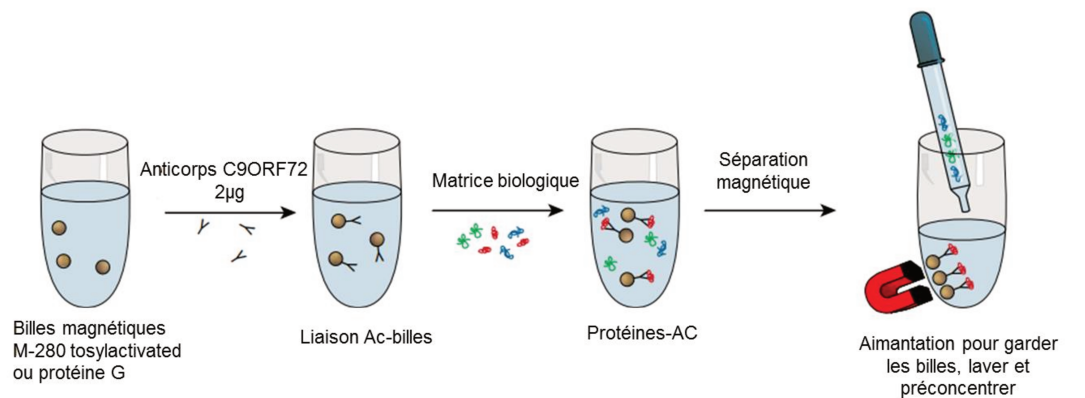


Figure 58 Principe de l'immunoprécipitation pour la concentration d'une protéine cible.

Après mise au point des conditions LC et PRM, ciblant 4 peptides de la protéine communs, ou spécifiques des isoformes court ou long (cf figure 59 et article New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers), nous avons développé les conditions de l'immunocapture



Figure 59 Isoformes de la C9ORF72 C9-L (haut) et C9-S (bas). Les peptides protéotypique sélectionnés sont soulignés. L'acide aminé différentiel C9-L et C9-S est encadré en rouge.

Nous avons pour commencer utiliser un anticorps polyclonal de GeneTex (reference GTX119776) pour développer notre méthode de détection de la C9ORF72. Il est dirigé contre une séquence commune aux deux isoformes, il permet donc de capturer les isoformes courtes et longues qui sont ensuite différenciées par spectrométrie de masse. Nous sommes partis d'un protocole couramment utilisé au laboratoire pour développer cette méthode<sup>122,123</sup>. Nous avons alors réalisé un premier test de détection sur des cellules HEK transfectées par C9-S ou C9-L, en plasma humain et cerveaux de souris avec ou sans surcharge de protéine recombinante de la forme courte (voir tableau 8).

Tableau 8 intensité obtenue pour les trois peptides en cellules HEK et en plasma. L'immunoprécipitation a été réalisée avec 0.4µg d'anticorps par échantillon et ont été greffés à des billes magnétiques M-280 tosylactivated

	HEK 293 courte	HEK 293 longue	Plasma	Plasma + recombinante C9-S 100 ng/mL	Homogénat cerveau de souris	Homogénat cerveau de souris + C9-S 100 ng/mL
TEIALSGK	4.7E5	1E5	nf	1.1E6	5.6E3	1e4
IILEGTER	7.3E5	1.5E5	6E2	1E6	5.1E3	3.3E3
DSTGSFVLPFR	nf	7E4	5E2	Nf	1E3	1E4

Ce test n'a pas été concluant. En effet, bien que nous observions une capture de la C9ORF72 dans les cellules HEK (qui nous servent ici de témoins) et une capture de la protéine recombinante en plasma, le niveau endogène en plasma est trop faible pour être détecté. Nous avons donc augmenté la quantité d'anticorps utilisé par échantillon d'un facteur 5 soit 2µg, en nous focalisant sur l'homogénat de cerveau de souris ou un faible signal était tout de même détectable (Tableau 8).

En augmentant la quantité d'anticorps, nous nous sommes aperçus de la détection de peptides de la C9-S humaine dans l'échantillon de cerveau de souris. Nous avons alors digéré uniquement la solution d'anticorps et nous avons observé les peptides de la C9-S humaine. La solution d'anticorps contient donc les peptides de la forme courte C9-S, probablement utilisée pour immunopurifier l'anticorps polyclonal. Néanmoins, nous avons tout de même pu détecter par PRM 3 peptides unique à la souris IVLEGTER, TEIALSGESPLLAATFAYWDNILGPR et DATGSSFVLPFR. L'anticorps permet donc apparemment de capter la C9ORF72 de souris en milieu complexe.

Nous avons ensuite testé un anticorps monoclonal provenant de ProteinTech (cat#66140-1-iG) qui lui n'est pas contaminé par la C9-S. Afin de comparer les deux anticorps, deux homogénats de cerveaux humains avec et sans surcharge ont été analysés.

Tableau 9 Aires des peptides de la C9ORF72 détecté dans deux homogénats de cerveau humain. Les peptides communs sont indiqués en rouges

	Polyclonal				Monoclonal				
	Ech-1	Ech-1	Ech-2	Ech-2	+C9-	Ech-	Ech-1	Ech-	Ech-2
		+C9-S 100		S	100	1	+C9-S 100	2	+C9-S 100
		ng/mL		ng/mL			ng/mL		ng/mL
TEIALSGK	45720	605352	46895	512047		nf	160353	nf	9784
IILEGTER	69329	764736	74008	691134		nf	196643	nf	8847
DSTGSFVLPFR	2212	3863	5629	6969		nf	nf	nf	nf

Malgré la pollution de l'anticorps polyclonal, la capture s'est montrée plus répétable qu'avec l'anticorps monoclonal sur les échantillons surchargés en C9-S (tableau 9). De plus, un signal faible est détecté pour le peptide unique de la C9-L (DSTGSFVLPFR) de la C9ORF72 endogène avec l'anticorps polyclonal. Ce résultat montre une mauvaise affinité de l'anticorps monoclonal pour la C9ORF72. Nous avons donc continué les optimisations avec l'anticorps polyclonal, en tentant de réduire le signal lié à la pollution.

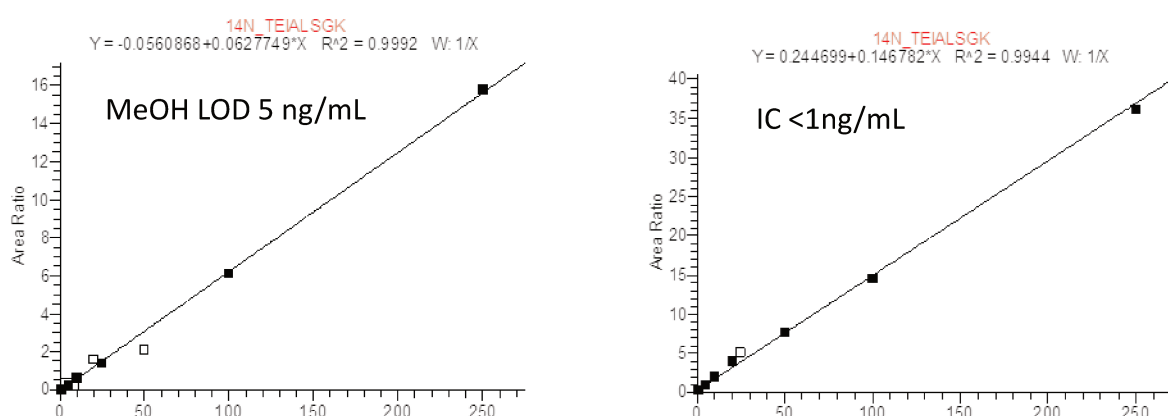
Nous avons notamment remplacé les billes M280 « tosylactivated » (greffage covalent de l'anticorps) par les billes magnétiques protéine G où seuls les anticorps anti-C9ORF72 doivent se lier (et non la protéine C9-S). Ces billes ont permis une diminution de la contamination d'un facteur 4. La présence de tween 20 à 0.5% a permis aussi de diminuer le signal d'un facteur 2. Des protocoles de dénaturation à l'acide des anticorps ont aussi été testés, sans succès car nous n'arrivions pas à renaturer l'anticorps pour une capture efficace de la C9ORF72. Après optimisation de ces paramètres et augmentation de la quantité d'anticorps d'un facteur 5 nous avons pu améliorer le signal du peptide DSTGSFVLPFR d'un facteur  $\approx 10$  et réduire le signal de la pollution de pour les peptides TEIALSGK et IILEGTER d'un facteur 4.

Pour évaluer notre protocole par immunocapture (Ac polyclonal), nous l'avons comparé à un protocole par précipitation des protéines au méthanol (qui sera décrit chapitre 1 II. 2.), donc sensé être moins efficace en termes de purification. Ce test a été réalisé d'abord en plasma diluée (1/20) et surchargé en C9-S recombinante, puis sur un échantillon plus complexe, des homogénats de cerveaux (souris et humain). En matrice peu complexe (plasma diluée), la méthode par

immunoprécipitation présente une LOD inférieure au protocole par précipitation des protéines au méthanol : 5 ng/mL par précipitation et <1 ng/mL par immunocapture (Figure 60). Par contre, le protocole simple par précipitation au méthanol s'est révélé plus sensible pour la détection des peptides de la C9ORF2 par un facteur 4-5 en homogénat de cerveaux (cerveau souris et cerveau humain, Tableau 10). Cette intensité des signaux a été confirmée lors des optimisations ultérieures du protocole par précipitation. Ces résultats contradictoires entre le plasma et l'homogénat de cerveau peuvent être attribués à plusieurs facteurs : une meilleure capture par l'anticorps de la protéine recombinante par rapport à la protéine endogène (à relier à une conformation particulière ou à la présence de modifications post-traductionnelles de la protéine endogène qui impactent la reconnaissance par l'anticorps) ou un effet matrice lors de l'étape d'immunocapture en homogénat de cerveaux, expliqué par exemple par une protéine cérébrale captée par l'anticorps par réaction croisée.

**Tableau 10 Comparaison des aires obtenues par immunoprécipitation avec la précipitation au méthanol en cerveau de souris et humain**

	Immunoprécipitation	Méthanol	Gain méthanol
Souris (SELTAFWR)	60771	320945	5
Humain (DSTGSFVLPFR)	49129	213433	4



**Figure 60 Comparaison de la linéarité et de la limite de détection en plasma 1/20 -ème entre précipitation au méthanol et immunoprécipitation (gamme de 1 ng/mL à 250 ng/mL)**



## Conclusion :

Nous n'avons pas réussi à développer une méthode suffisamment sensible pour la quantification de la C9ORF72 cérébrale par immunoprécipitation, malgré le test de deux anticorps commerciaux un monoclonal et un polyclonal, et une optimisation des conditions de capture. L'anticorps monoclonal s'est montré moins performant que l'anticorps polyclonal pour la détection de la C9ORF72. En revanche, nous avons observé une contamination de cet anticorps par les peptides communs de C9-S et C9-L. Malgré cette pollution nous avons pu suivre la détection du peptide unique de la forme longue DSTGSFVLPFR dans des homogénats de cerveau humains. Nous avons alors comparé cette méthode à un protocole de précipitation des protéines par le méthanol et observé un signal diminué d'un facteur 4. Du fait de la pollution de l'anticorps, et de la sensibilité inférieure en homogénat de cerveau pour la C9ORF72 endogène nous avons opté pour le développement d'une méthode de quantification sans anticorps. Le protocole d'immunocapture n'a pas été caractérisé plus en détails (rendement de capture, répétabilité). Nous avons orienté le travail vers le développement d'un protocole par précipitation. Son application à la quantification de la C9ORF72 et la validation analytique sont présentées dans l'article suivant.

**2. Article : quantification de la C9ORF72 dans le cortex frontal**



# New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers

Arthur Viodé<sup>1</sup>, Clémence Fournier<sup>2,3</sup>, Agnès Camuzat<sup>2,4</sup>, François Fenaille<sup>1</sup>, NeuroCEB Brain Bank, Morwena Latouche<sup>2,4</sup>, Fanny Elahi<sup>5</sup>, Isabelle Le Ber<sup>2,3,6</sup>, Christophe Junot<sup>1</sup>, Foudil Lamari<sup>7,8</sup>, Vincent Anquetil<sup>2,3</sup> and François Becher<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Leo P. Sugrue,  
University of California,  
San Francisco, United States

### Reviewed by:

Chiara F. Valori,  
Helmholtz-Gemeinschaft Deutscher  
Forschungszentren (HZ), Germany  
Dapeng Chen,  
University of Maryland, College Park,  
United States

### \*Correspondence:

François Becher  
francois.becher@cea.fr

### Specialty section:

This article was submitted to  
Neurodegeneration,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 15 March 2018

**Accepted:** 06 August 2018

**Published:** 28 August 2018

### Citation:

Viodé A, Fournier C, Camuzat A, Fenaille F, NeuroCEB Brain Bank, Latouche M, Elahi F, Le Ber I, Junot C, Lamari F, Anquetil V and Becher F (2018) New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers. *Front. Neurosci.* 12:589. doi: 10.3389/fnins.2018.00589

<sup>1</sup> Service de Pharmacologie et Immunoanalyse, Laboratoire d'Etude du Métabolisme des Médicaments, Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Institut National de la Recherche Agronomique, Université Paris Saclay, Gif-sur-Yvette, France, <sup>2</sup> Inserm U 1127, CNRS UMR 7225, Institut du Cerveau et de la Moelle Épineuse, ICM, Sorbonne Université, Paris, France, <sup>3</sup> Assistance Publique – Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Paris, France, <sup>4</sup> Ecole Pratique des Hautes Etudes, PSL Research University, Paris, France, <sup>5</sup> Department of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, CA, United States, <sup>6</sup> National Reference Center for Rare or Early Dementias, Institute of Memory and Alzheimer's Disease (IM2A), Department of Neurology, AP-HP – Hôpital Pitié-Salpêtrière, Paris, France, <sup>7</sup> Assistance Publique – Hôpitaux de Paris, Service de Biochimie Métabolique, Hôpitaux Universitaires Pitié Salpêtrière – Charles Foix, Paris, France, <sup>8</sup> GRC 13 Neurométabolisme – UPMC, Sorbonne Université, Paris, France

Frontotemporal dementia (FTD) is a fatal neurodegenerative disease characterized by behavioral and language disorders. The main genetic cause of FTD is an intronic hexanucleotide repeat expansion ( $(G_4C_2)_n$ ) in the *C9ORF72* gene. A loss of function of the *C9ORF72* protein associated with the allele-specific reduction of *C9ORF72* expression is postulated to contribute to the disease pathogenesis. To better understand the contribution of the loss of function to the disease mechanism, we need to determine precisely the level of reduction in *C9ORF72* long and short isoforms in brain tissue from patients with *C9ORF72* mutations. In this study, we developed a sensitive and robust mass spectrometry (MS) method for quantifying *C9ORF72* isoform levels in human brain tissue without requiring antibody or affinity reagent. An optimized workflow based on surfactant-aided protein extraction and pellet digestion was established for optimal recovery of the two isoforms in brain samples. Signature peptides, common or specific to the isoforms, were targeted in brain extracts by multiplex MS through the parallel reaction monitoring mode on a Quadrupole–Orbitrap high resolution mass spectrometer. The assay was successfully validated and subsequently applied to frontal cortex brain samples from a cohort of FTD patients with *C9ORF72* mutations and neurologically normal controls without mutations. We showed that the *C9ORF72* short isoform in the frontal cortices is below detection threshold in all tested individuals and the *C9ORF72* long isoform is significantly decreased in *C9ORF72* mutation carriers.

**Keywords:** frontotemporal dementia (FTD), frontotemporal lobar degeneration (FTLD), amyotrophic lateral sclerosis (ALS), C9ORF72, TDP-43, TDP43, mass spectrometry (MS), GRN

## INTRODUCTION

Frontotemporal dementia (FTD) is the second most prevalent neurodegenerative disease before the age of 65, after Alzheimer disease. FTD are caused by frontal and temporal lobar degeneration leading to behavioral, socioemotional, language disorders, and progressive loss of autonomy and death approximately 10 years after disease onset (Rascovsky et al., 2011). Amyotrophic lateral sclerosis (ALS) caused by motor neuron degeneration is associated with FTD in 15% of patients or families. Familial forms of FTD, accounting for about 20–50% of cases (Rosso, 2003) are mainly caused by mutations in three major genes: granulin (*GRN*), microtubule-associated protein tau (*MAPT*) and C9 open reading frame 72 (*C9ORF72*).

*C9ORF72*, the most frequent genetic etiology, represents 25% of familial FTD cases (DeJesus-Hernandez et al., 2011; Renton et al., 2011), 80% of FTD-ALS familial cases and also explains 40% of pure familial ALS (without FTD) (Majounie et al., 2012). An expansion of hundreds to thousands hexanucleotide ( $G_4C_2$ ) repeats is present in the first intron of the gene in patients, while healthy individuals carry less than 24 ( $G_4C_2$ ) repeats. While the exact mechanisms of disease remain unknown, two main modes of toxicity are proposed. The expansion may be deleterious through formation of nuclear RNA foci by RNA containing the hexanucleotide expansion ( $G_4C_2$ ) with sequestration of RNA-binding proteins (Lagier-Tourenne et al., 2013), and/or translation of polydipeptides proteins that aggregate in the brain (Mori et al., 2013). A loss of function of the *C9ORF72* protein associated with a decrease in *C9ORF72* allele-specific expression was also hypothesized to contribute to the pathogenesis based on reduction of *C9ORF72* mRNA transcript levels in patients (Ciura et al., 2013). The three alternatively spliced *C9ORF72* transcripts encode two *C9ORF72* protein isoforms, a 222 amino acids (AA) protein isoform called C9-short (C9-S) and a 481AA protein isoform called C9-long (C9-L). The two protein isoforms have been shown by immunofluorescence to have distinct cellular localization with the C9-S localizing to the nuclear membrane and C9-L to the cytoplasm (Xiao et al., 2016). This observation suggests that the two isoforms have a different function, while the precise function of the protein has not been clearly determined. A potential role in endosomal transport and autophagy was reported through interaction with Rab-GTPases (Zhang et al., 2012; Levine et al., 2013).

A better understanding of the contributions of the loss of function to the disease mechanism requires a precise quantification of reduction in levels of *C9ORF72* isoforms. *C9ORF72* protein has been so far only quantified in human tissues by Western blot (Waite et al., 2014). Several studies pointed out the poor affinity and selectivity of commercially available antibodies (Waite et al., 2014; Davidson et al., 2018), requiring laborious in-house generation of antibodies. Targeted mass spectrometry (MS) is a powerful alternative to Western blot and enzyme-linked immunosorbent assay (ELISA) for quantification of proteins. It provides accurate quantification, high level of specificity, avoiding interference due to cross-reactivity of antibodies, and the ability to discriminate between isoforms (Chen D. et al., 2015; Chen Y.-T. et al., 2015;

Jedrychowski et al., 2015; Lesur et al., 2015). Nevertheless, MS-based detection of low-abundant proteins in complex fluids or tissues remains challenging without efficient sample preparation protocols. The gold standard relies on the combination with immunoprecipitation to selectively enrich the analyte of interest prior to MS (Chen Y.-T. et al., 2015), but is applicable only when antibodies with sufficient specificity and affinity for the target protein are available.

We have developed a sensitive and robust antibody-free MS assay for quantification of *C9ORF72* isoforms in brain samples. The protocol consists in an optimized tissue lysis protocol followed by pellet digestion of extracted brain proteins and specific monitoring of common and isoform specific peptides by targeted high-resolution MS in the parallel reaction monitoring mode (PRM). Reproducibility and linearity were demonstrated, as well as equivalent isoform recovery from brain tissue samples and stability during sample preparation. This new assay allowed for the first time the quantification of the *C9ORF72* long isoform in *post mortem* frontal cortex brain samples from a cohort of FTD patients harboring a *C9ORF72* mutation and highlighted a significant decrease in concentrations in mutation carriers. The short isoform was found to be below the sensitivity threshold of the method.

## MATERIALS AND METHODS

### Patients Information, Tissue Collection, and Consent

Frozen tissue from frontal cortex (Brodmann area 9/10) of 21 FTD (with or without secondarily developed ALS) patients carrying *C9ORF72* expansion, of 10 patients with non-genetic FTD (with or without secondarily developed ALS) pathologically characterized by TDP-43-positive neuronal inclusions, and of 12 neurologically healthy controls were studied. The brain samples were collected through a brain donation program dedicated to neurodegenerative dementias coordinated by the NeuroCEB Brain Bank Network. The informed consent for post-mortem examination and research studies was signed by the legal representative of each patient in patient's name, as allowed by the French law and approved by the local ethics committee and the brain bank has been officially authorized to provide samples to scientists (agreement AC-2013-1887). All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committees and with the 1964 Helsinki declaration. The brain banks fulfill criteria from the French Law on biological resources including informed consent, ethics review committee and data protection (article L1243-4 du Code de la Santé publique, August 2007). The Neuro-CEB brain bank (BioResource Research Impact Factor number BB-0033-00011) has been declared to the Ministry of Research and Higher Education, as required by French law.

### Chemical and Materials

*C9ORF72* short isoform was purchased from Proteintech (cat# ag21080) (Proteintech Group, Chicago, IL, United States)

and C9ORF72 long isoform from Abnova (cat#00203228-P01) (Abnova Le Perray En Yvelines, France). Trypsin from bovine pancreas TPCK Treated (reference T1426), ammonium hydroxide, ammonium bicarbonate were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). RapiGest SF Surfactant and SPE Oasis Max ICC/30 mg were purchased from Waters Corporation (Milford, MA United States). Labeled peptides for quantification were synthesized in Absolute QUAntitation (AQUA) ultimate quality by Thermo Fisher Scientific (Paisley, United Kingdom). Water (ChromaSolve LC-MS), acetonitrile (HPLC-grade), and formic acid were obtained from Honeywell/Riedel-de Haen (Seelze, Germany) and VWR chemicals (Fontenay sous Bois, France), respectively. All other chemicals were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France) or VWR Chemicals (Fontenay sous Bois, France). Pierce BCA protein Assay kit was purchased from Pierce (Rockford, IL, United States). For all reactions, LoBind Eppendorf tubes (Dutscher, Brumath, France) were used.

## Sample Preparation

### Brain Protein Extraction Protocol

Lysis buffer containing, trizma-base 20 mM; NaCl 150 mM; cComplete Protease Inhibitor Cocktail 1X and 1% triton, was added to single pieces of whole brain tissue (~100 mg) at a ratio of 5  $\mu$ L per 1 mg of tissue. Brain samples were homogenized by beads beating using a precellys soft tissue CK14 2 mL (3\*30S at 6,500 rpm). The lysate was then centrifuged at 4,000 rpm for 15 min at 4°C. Fifty microliters from the supernatant was used for analysis. A 5  $\mu$ L aliquot was used for total protein concentration, determined by the Pierce BCA Protein Assay kit with a sample to working reagent ratio 1:20. Two percent SDS were added to eliminate interference from lipids.

### Pellet Digestion

The lysate was precipitated by adding 150  $\mu$ L of methanol (ratio 3:1), followed by vortex-mixing and briefly centrifuged 5 s. The supernatant was discarded. Twenty microliters of rapigest 0.05% in ammonium bicarbonate 50 mM were added to the pellet. Aqua peptides were added at this step, at 10 ng/mL final concentration. After mixing for approximately 15 min, reduction was performed with 10  $\mu$ L DTT (20 mM) and incubation at 60°C for 30 min. Alkylation was performed with 10  $\mu$ L iodoacetamide (45 mM) and 45 min incubation at room temperature. Proteins were digested overnight at 37°C with 40  $\mu$ g of trypsin.

### Solid Phase Extraction (SPE)

Tryptic digests were diluted by addition of 300  $\mu$ L of 5% ammonium hydroxide before SPE extraction on oasis MAX 1 cc/30 mg, previously conditioned with 1 mL of Methanol and equilibrated with 1 mL of water. Samples were loaded and washed with 500  $\mu$ L of 5% NH<sub>4</sub>OH and 2\*250  $\mu$ L of methanol. Peptides were eluted with 3\*250  $\mu$ L of methanol containing 10% formic acid. Extracts were then evaporated to dryness with a Turbopap instrument (Biotage, Uppsala, Sweden) (5–15 psi, 40°C for 1 h). The dry residue was re-dissolved in 95% water 5% acetonitrile 0.1% formic acid and centrifuged at 4°C for 10 min at 15,000  $\times$  g, prior injection into the LC system.

## LC-MS/MS Analysis

LC-MS/MS was performed on a Dionex Ultimate 3,000 chromatography system coupled to a Q-exactive Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Ten microliters of sample was loaded onto the column. Chromatographic separation was performed on an Aeris peptide XBC18 reverse phase column (150 mm  $\times$  2.1 mm; 1.7  $\mu$ m; 100 Å; phenomenex, Le Pecq, France) and achieved in a 30 min gradient at a flow rate of 500  $\mu$ L/min. A gradient of mobile phase consisting of LC-MS-grade water with 0.1% formic acid (phase A) and acetonitrile with 0.1% formic acid (phase B) was delivered. After an isocratic step of 0.5 min at 5% B, the gradient was ramped to 25% over the next 19.5 min then to 50% over the next 4min. Then acetonitrile was increased to 95% for the next 2 min. Column re-equilibration at 5% B was realized for 4 min.

Instrument parameters of the electrospray ionization source were set as follows: sheath gas flow rate at 70 a.u., spray voltage at 4 kV, capillary temperature at 320°C. The Q-exactive instrument was operated in positive ion mode under time-scheduled sequential PRM acquisition. Endogenous peptides precursor ions and AQUA peptides were selected in the quadrupole with an isolation mass window of 1.5 m/z. Precursors were fragmented in the HCD cell using nitrogen as collision gas and the optimized normalized collision energy (**Supplementary Table S2**). All fragment ions were transferred to the Orbitrap. Resolution was set to 70,000 at m/z 200 (full width at half-maximum), automatic gain control to 1e6, and maximum injection time to 240 ms. Xcalibur 2.2 software (Thermo Fisher Scientific, Bremen, Germany) was used for instrument control and processing of the data files.

## PRM and Quantification

A time-scheduled sequential PRM method was established targeting the following C9ORF72 peptides TEIALSGK, ILL-EGTER, DSTGSFVLPFR, and SHSVPEEIADIADTVLNDDDI GDSCHEGFLK (**Supplementary Table S2**). To increase the signal to noise ratio and assay sensitivity, the signal of up to 6 major and non-interfered fragment ions identified with high resolution (5 ppm) from a common peptide precursor were summed up to provide one extracted ion chromatogram (XIC) for each peptide. Isotope-labeled synthetic peptides with labeled amino acids <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>2</sub>-labeled lysine and <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>4</sub>-labeled arginine were used for signal normalization and quantification of C9ORF72 peptides. Raw MS data were exported to Skyline 3.7 (MacLean et al., 2010) for verification of the transitions ratio of unlabeled and labeled peptides. Xcalibur 2.2 software (Thermo Fisher Scientific, Bremen, Germany) was used for quantitative data analysis.

External calibration curve was made by digesting the C9ORF72 long isoform recombinant protein in a surrogate matrix (mice brain) and SIL peptides were spiked before reduction and alkylation of the pellet. Linear regression with 1/x weighting was applied to generate a standard curve.

## HEK293 Cells Transfection

Plasmids expressing EGFPN-tagged C9ORF72 long or short isoforms were assembled as follow. cDNAs coding for NP\_060795 (long) and NP\_659442 (short) were ordered from DNA2.0/ATUM in the pCS2 vector. Fifty nanograms of each plasmid was PCR amplified with AccuPrime Pfx Supermix (Invitrogen) according to the manufacturer's protocol. Forward primer was CACCTCGACTCTTTGCCACC and reverse primers were, respectively, for the long and short isoforms CTAAAAAGTCATTAGAACATCTCGTTCTTGCACAC and C TACTTGAGAAGAAAGCCTTCATGACAGC. Purified and sequenced PCR products were cloned into pENTR/D Gateway TOPO according to manufacturer's protocol (Thermo Fisher Scientific). Purified and sequenced entry long and short C9ORF72 clones were recombined with pgLAP1 destination vector (Addgene Plasmid #19702) with LR Clonase Enzyme Mix according to the protocol given by the manufacturer (Thermo Fisher Scientific). After sequencing, the two plasmids, EGFP long C9ORF72 and EGFP short C9ORF72, were transfected with lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's protocol. Briefly, T-75 flasks of HEK293 cells at 70% confluency were transfected with 15  $\mu$ g EGFP long C9ORF72 and 15  $\mu$ g EGFP plasmids, or 15  $\mu$ g EGFP short C9ORF72 and 15  $\mu$ g EGFP plasmids or 15  $\mu$ g EGFP long C9ORF72 and 15  $\mu$ g EGFP short C9ORF72 plasmids, or 30  $\mu$ g EGFP control plasmid. Cells were collected 48 h after transfection. PBS washed pellets were stored at minus 80°C until protein extraction.

## Statistical Analyses

All statistical analysis was performed using Graphpad Prism software (version 5.01). Data were compared with a Mann-Whitney test and medians were considered significantly different if  $p < 0.05$ . Data were represented with median and interquartile range.

## RESULTS

### Assay Design

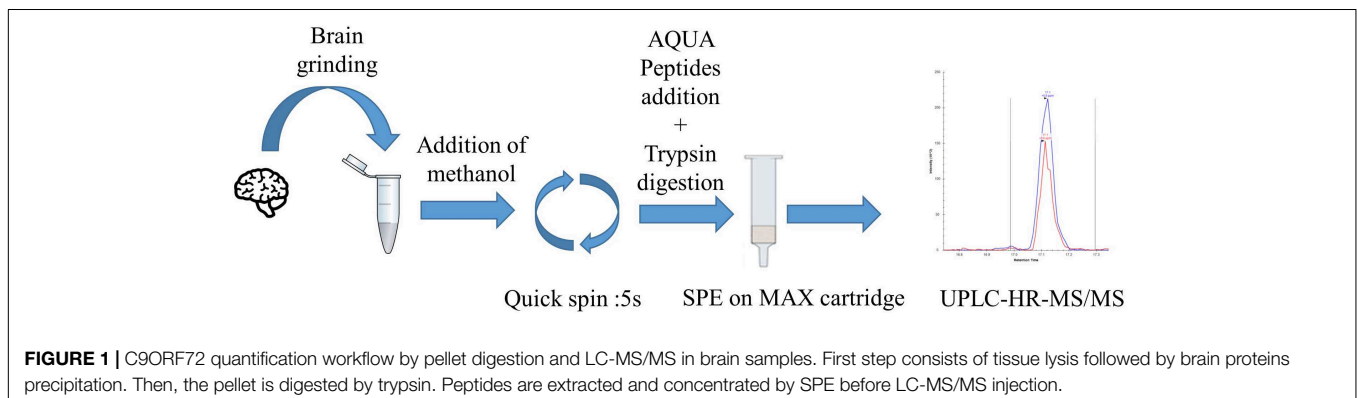
First step of the assay (**Figure 1**) consists in an optimized brain sample homogenization and extraction of C9ORF72 isoforms

using bead-beating tubes with the presence of a nonionic surfactant Triton X-100 to further disrupt lipidic cell membranes. Then, a pellet digestion protocol was adapted to protein digestion in brain extracts. C9ORF72 isoforms are denatured and equivalently precipitated with methanol. After removal of the supernatant containing the surfactant, resuspension and addition of AQUA peptides, i.e., stable isotope peptide, the pellet was digested with trypsin. Signature peptides were then extracted using mixed-mode anion exchange cartridges. Finally, after evaporation and resuspension in 5% acetonitrile with 0.1% formic acid, samples were injected into the LC-MS/MS where three peptides, common to both isoforms or unique to the long isoform, were quantified with specificity in the PRM mode. A specific peptide from the short isoform was also monitored for detection purposes only.

## Analytical Procedure

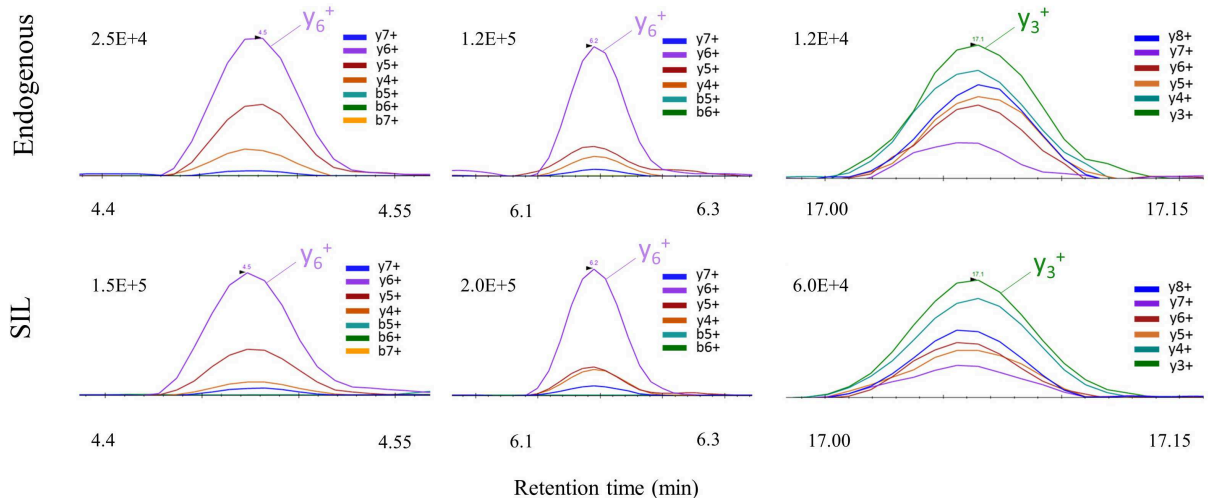
### Proteotypic Peptides Selection and LC-MS/MS Detection

Two isoforms of the C9ORF72 protein are reported (DeJesus-Hernandez et al., 2011) (Uniprot sequence entries: Q96LT7-1 and Q96LT7-2). Amino-acid sequences are illustrated in **Figure 2**. Short isoform amino acid sequence is shared with the long one with the exception of its last residue, i.e., residue 222 (N→K). For a comprehensive assay of C9ORF72, common and isoform-specific peptides have to be identified. To this end, data-dependent analysis (DDA) experiments on human frontal cortices were performed (described in Supporting Information). No signal was detected for the C9ORF72 isoforms, similarly, to another recent study (Umoh et al., 2018), illustrating their low abundance in brain and the need to develop a more sensitive targeted method. A tryptic digest of recombinant C9ORF72 long isoform was then used for selection of best responding peptides. Three peptides were selected based on signal intensity after digestion, size (i.e., between 6 and 20 amino acids), and lack of cysteine, methionine and glutamine residues. Specificity of the selected peptides was assessed using Basic Local Alignment Search Tool (BLAST) against the UniprotKB/Swissprot human database. Among the three peptides selected for quantification (**Figure 2**), two are common to both the short and long isoforms and one is unique to the long isoform. The assay based on peptide 1 (TEIALSGK) and 2 (IILEGTER) quantify total





**FIGURE 2** | C9ORF72 long (top) and short (bottom) isoform (respectively, Q96LT7-1 and Q96LT7-2). Proteotypic peptides selected are underlined. A red box displays the single amino acid differentiating the common sequence.



**FIGURE 3** | PRM signal of C9ORF72 proteotypic peptides with their corresponding co-eluting SIL peptide in human brain samples.

C9ORF72 whereas the assay based on peptide 3 (DSTGSFVLPFR) differentiates the long isoform. Internal standards AQUA peptides of the three selected peptides for quantification were synthesized by incorporating stable isotopes at the C-terminal amino acid residues ( $^{13}\text{C}_6$ ,  $^{15}\text{N}_2$ -labeled lysine and  $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$ -labeled arginine). AQUA peptides were prepared with high purity (>95%) and well-defined concentrations. The unique 30 amino acids long peptide specific for the short isoform (peptide 4), containing residue 222 (Figure 2), did not meet the stringent selection criteria for inclusion in the quantitative method, and was selected for detection only. A stable isotope-labeled (SIL) analog of peptide 4 was nonetheless synthesized for unambiguous identification of C9-S in the brain samples. In the absence of a suitable quantitative peptide for C9-S, quantification is obtained by difference as previously reported for progastrin-releasing peptide (ProGRP) isoforms (Torsetnes et al., 2013).

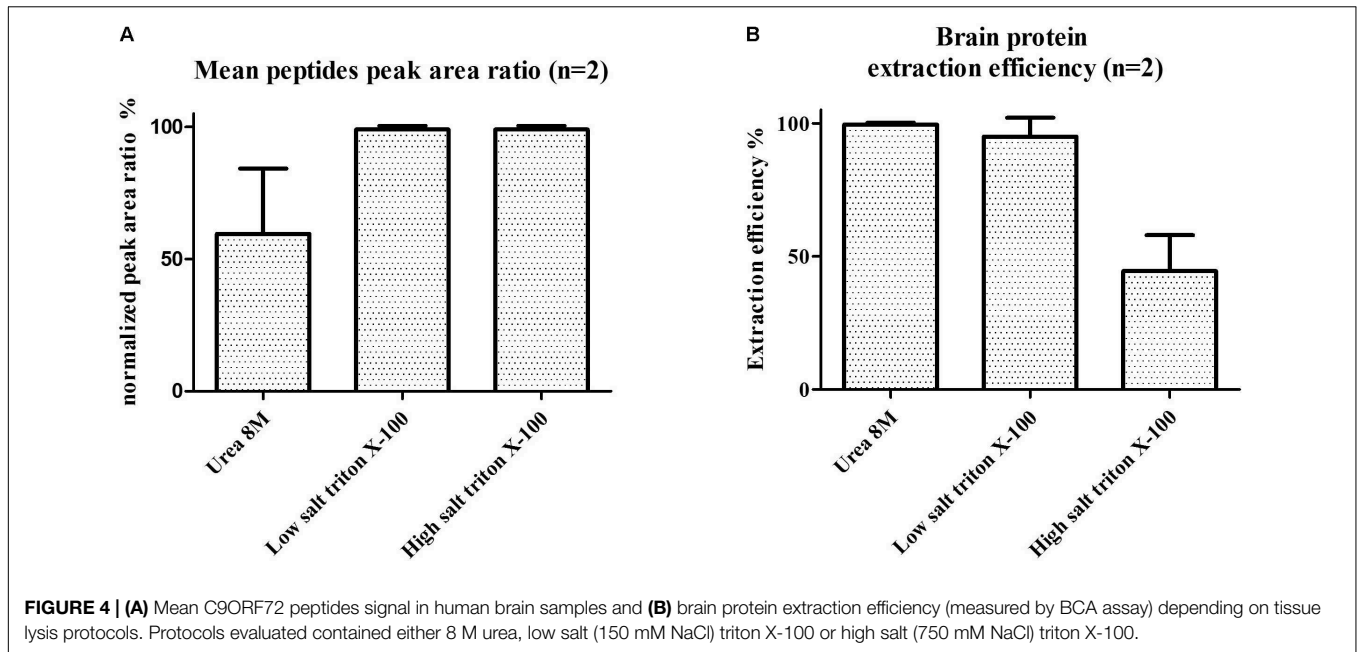
Simplification of sample preparation oriented our choice toward Parallel Reaction Monitoring (PRM) over Selected Reaction Monitoring (SRM) for the increased specificity regarding fragment ion detection in complex matrixes provided by the high-resolution Orbitrap mass analyzer. A PRM method was established targeting the three selected peptides for quantification, and the unique C9-S peptide for detection only. Ultra-high performance liquid chromatography was performed on a C18 column for peptide separation with a total runtime of 30 min. Shorter gradients resulted in interferences at the retention time of peptide DSTGSFVLPFR (SIL version) and a decrease of signal intensity up to a factor of 2 in brain samples (Supplementary Figure S1 and Supplementary Table S1). Data

treatment increased sensitivity by summing the signals of up to six major and non-interfered fragment ions to provide one XIC for each targeted peptide (Dupré et al., 2015; Figure 3 and Supplementary Table S2). Each endogenous peptide and their corresponding isotope-labeled form (AQUA peptides) must strictly co-elute with similar transition ratio across the different samples in comparison with a standard. If not, the transition was excluded.

Ability of the targeted method to detect the isoforms was demonstrated in HEK293 cells transfected with plasmids expressing either C9-S or C9-L. All common (peptides 1 and 2) and the two unique isoform-specific peptides of C9-L and C9-S (peptides 3 and 4) were detected in corresponding HEK293 cells (Supplementary Figure S3). Mock-transfected HEK293 cells were also analyzed and displayed a lower level of C9ORF72.

## Sample Preparation of Human Brain Samples Prior to LC-PRM

The main steps of sample preparation consist of tissue homogenization for optimal protein extraction followed by a pellet digestion and peptide enrichment with SPE for lowering ion suppression/matrix effects and increasing sensitivity. Protein extraction protocols with tissue homogenization by mechanical shear recently published for C9ORF72 (Xiao et al., 2015), containing either low/high-salt content (i.e., 150 mM NaCl or 750 mM NaCl), 8 M urea or Triton X-100, were individually evaluated on tissues samples from frontal cortex (Figure 4). These protocols were adapted to single use



bead-beating Precellys tubes to parallelize protein extraction and avoid contamination between samples. Protocols were evaluated based on total protein extraction and signal from C9ORF72 peptides. The low salt protocol without detergent resulted in low protein extraction yield and no signal for C9ORF72 peptides, although Western blot signal was previously reported for the short isoform (Xiao et al., 2015). Signals were observed for the three quantitative peptides selected for quantification (peptides 1–3) when applying the Triton containing mixtures, the latter showing higher intensity than that containing urea (Figure 4A). We finally selected the low-salt Triton protocol because it allowed the extraction of more total proteins than the high-salt Triton protocol (Figure 4B). However, no signal was detected for the C9-S unique peptide (peptide 4) with any of the evaluated protocols.

A recently published convenient pellet digestion protocol for monoclonal antibody quantification in human plasma (Becher et al., 2017) was evaluated for the detection and quantification of C9ORF72 isoforms in brain proteins lysates. Recombinant protein of C9-S like and C9-L spiked into mice brain lysate, as a substitute to human brain, before and after methanolic precipitation demonstrated high and equivalent recovery for both recombinant protein isoforms with equivalent precipitation yield above 70% and coefficients of variation below 10% ( $n = 2$ ) (Table 1). Brain lysate is a highly complex matrix resulting in deleterious matrix effects. To further reduce its complexity, SPE clean-up was evaluated in human brain samples based on AQUA peptides signal after SPE (Supplementary Figure S2). Among Oasis Hydrophilic-Lipophilic Balance (HLB) operated at high pH, Mixed-mode Cation-eXchange (MCX), and Mixed-mode Anion-eXchange (MAX) cartridges, Oasis MAX has proven to give the higher peptide signal intensity (data

**TABLE 1 |** Precipitation yield and variation of C9-S and C9-L isoforms in mice brain extracts.

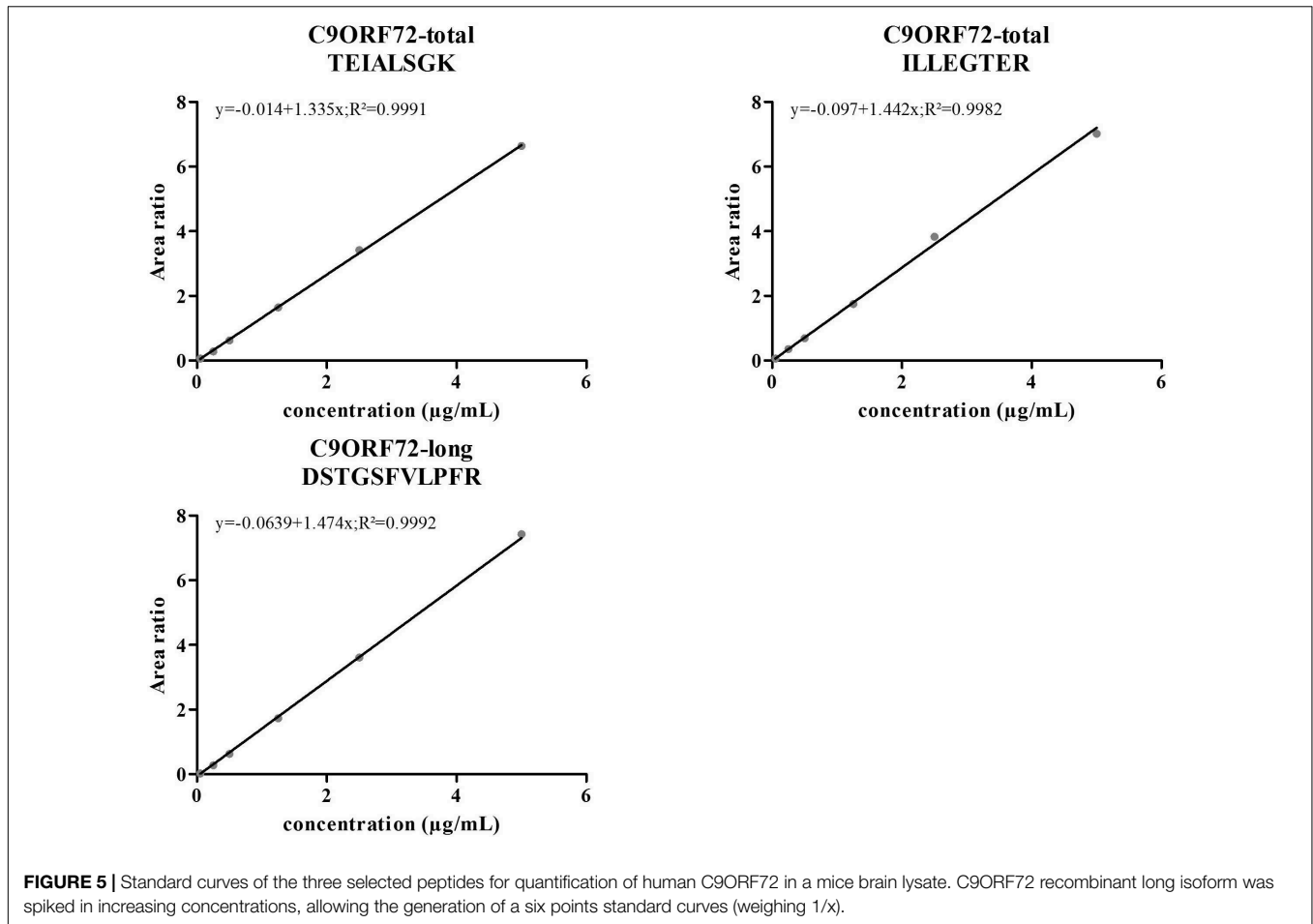
	Short isoform		Long isoform	
	Mean	CV	Mean	CV
Precipitation yield ( $n = 2$ )	88%	8%	74%	9%

not shown), in line with the acidic isoelectric point of the peptides ranging from 3.9 to 5.8. Signal enhancement resulting from SPE was about two folds. SPE yield was then determined in the final conditions in mice brain lysate. Yield measured by aqua peptides spiked before and after the SPE extraction were between 45 and 70% (Supplementary Table S3), which is in line with previous works (Gong et al., 2015). For method robustness, aqua peptides were spiked early in the protocol, before trypsin digestion of the pellet, and therefore extracted by SPE, similarly, to the C9ORF72 peptides.

## Method Validation

The assay was then evaluated for C9ORF72 quantification in human brain tissues, including the main items of linearity, sensitivity, inter-, and intra-day assay precision, stability during sample processing, i.e., over 90 min at room temperature and matrix effect. Linearity and sensitivity were determined with standard curves prepared in mice brain as a surrogate to human brain samples. Indeed, the three proteotypic peptides selected for quantification are not present in mice due to single point mutations, despite 85% sequence identity with human (Supplementary Figure S4). Other items were tested with quality control samples prepared in human brain extract, spiked with recombinant C9-L. Recombinant C9-L was selected for





validation experiments because it contains all three quantitative peptides.

Linearity and sensitivity of the method were evaluated with a 6-point calibration curve of the recombinant C9-L. Mice brain proteins were extracted with the same protocol as for human brain. The method was shown to be linear from 50 to 5,000 ng/mL with a Lower Limit Of Quantification (LLOQ) observed at 50 ng/mL (Figure 5). LLOQ was defined based on accuracy between 80 and 120%. To determine a potential matrix effect between mice and human brains, recombinant C9ORF72 was spiked into human brain extract in triplicates at 500 ng/mL and confronted against the calibration curve in mice (Table 2). The three peptides displayed good accuracy in the range of 85–115%, which demonstrate the suitability of mice brain extract as a surrogate matrix for C9ORF72 determination in human brain. Intraday repeatability of the analytical method was evaluated in human brain extract. Briefly, proteins from a human brain sample were extracted and divided in five aliquots for protein precipitation, digestion and LC-MS/MS analysis. Precision was observed below 10% for each peptide ( $n = 5$ ) (Table 3). Interday precision was also evaluated by analyzing three different brain extracts on three different days. Variability was found to be acceptable, with CV% between 10 and 26% (Table 4). Stability of the C9ORF72 protein in brain extract

during the sample preparation is an important parameter for quantification. Brain extracts were either directly processed or left on ice for 90 min, corresponding to the time to process around 100 samples from protein extraction to protein precipitation. Both conditions showed similar area ratio for the three peptides demonstrating stability of C9ORF72 in our conditions (Figure 6).

Taking all results together, the protocol was found efficient for determination of C9ORF72 in human brain samples. The new method demonstrated robustness with variability and accuracy below 20%.

## C9ORF72 Isoforms Determination in Human Brain Tissue

Levels of C9ORF72 protein were investigated in frontal cortices of FTD patients (with or without ALS) carrying C9ORF72 expansions ( $n = 21$ ), patients with non-genetic FTD, pathologically characterized by TDP-43-positive neuronal inclusions ( $n = 10$ ), and neurologically normal controls ( $n = 12$ ) determined by pathologists within the network of the NeuroCeb brain bank.

The three peptides selected for quantification were detected above LLOQ in all samples (Figure 7). However, no signal was detected for the unique C9-S peptide in any of the

**TABLE 2** | Matrix effect between mice and human brain, evaluated for each quantitative peptide.

QC in human brain (n = 3)	TEIALSGK				ILLEGTER				DSTGSFVLPFR			
	Mean measured (μg/mL)	CV %	Theoretical value (μg/mL)	Bias %	Mean Measured (μg/mL)	CV%	Theoretical value (μg/mL)	Bias %	Mean measured (μg/mL)	CV %	Theoretical value (μg/mL)	Bias %
Blank (endogenous)	0.285	6%			0.292	2%			0.348	5%		
Spike (+0.5 μg/mL)	0.767	4%	0.785	2%	0.815	1%	0.792	3%	0.785	3%	0.848	7%

Human brain lysate was spiked with 0.5 μg/mL of C9ORF72 recombinant long isoform and back calculated against a calibration curve prepared in mice brain lysate. Endogenous C9ORF72 concentration was added to the spiked amount for calculations of bias. Bias and CV% are reported.

**TABLE 3** | Intraday variability of the measured C9ORF72 concentrations by replicate analysis of a control human brain sample (n = 5).

Intraday validation (n = 5)	TEIALSGK		ILLEGTER		DSTGSFVLPFR	
	Mean (μg/mL)	CV	Mean (μg/mL)	CV	Mean (μg/mL)	CV
Control-1	0.37	4%	0.43	7%	0.414	6%

**TABLE 4** | Interday variation of the measured C9ORF72 concentrations in 2 C9ORF72 human brains and one human control brain samples; (n = 3).

Interday variation (n = 3)	TEIALSGK		ILLEGTER		DSTGSFVLPFR	
	Mean (μg/mL)	CV	Mean (μg/mL)	CV	Mean (μg/mL)	CV
Control-2	0.514	18%	0.545	10%	0.563	14%
C9-1	0.344	26%	0.382	12%	0.390	22%
C9-2	0.153	11%	0.182	16%	0.197	10%

brain samples. Furthermore, similar amounts of total C9ORF72 and C9-L were measured, considering assay accuracy and variability of 20%, revealing by difference, the low abundance of C9-S (**Supplementary Table S4**). Differences in measured concentrations between groups were assessed by a Mann-Whitney nonparametric test. Quantification based on any of the three peptides demonstrated a significant decrease in total C9ORF72 in FTD patient with C9ORF72 expansions ( $p < 0.0001$ ) compared to controls and non-genetic FTD ( $p < 0.05$ ) (**Figure 7**). The decrease in concentration of about 42%, which was observed for the three peptides (i.e., the two peptides representing total C9ORF72 and the one specific to the long isoform), corroborates prior Western blot findings.

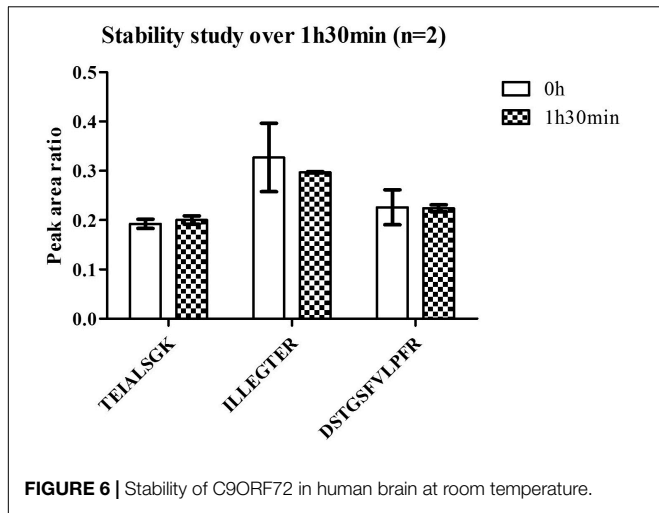
## DISCUSSION

C9ORF72 isoforms have solely been investigated by Western blot which relies on the availability of antibodies whose specificity has to be characterized and validated (Liu et al., 2006; Davidson et al., 2018). In this respect, we developed a MS assay avoiding antibodies and the gold standard immunoprecipitation for protein quantification in complex matrixes. An efficient protocol, with optimized sample preparation steps, i.e., protein

extractions from tissue, pellet digestion and SPE, was successfully implemented for the first time to quantitate C9ORF72 isoforms level in a cohort of human brain samples from C9ORF72 or non-genetic FTD patients and control individuals.

Preparation of tissue lysates and protein extraction is a key issue for proteome coverage, especially the more challenging membrane or nuclear proteins (Cox and Emili, 2006; Wiśniewski et al., 2009). Taking into account the reported distinct subcellular localizations of C9ORF72 isoforms with localization of C9-S to the nuclear membrane and C9-L in the cytoplasm (Xiao et al., 2015), lysis protocols were investigated. We found that Triton X-100 facilitated the extraction of the C9ORF72 proteins, in agreement with higher extraction yield reported with detergent based protocols in comparison to organic solvents or chaotropic reagents such as urea, in fatty tissue such as the brain (Shevchenko et al., 2012). In addition, the Triton X-100 protein extraction protocol probably enhanced the subsequent trypsin digestion as previously reported (An et al., 2015). A simple pellet digestion protocol (Becher et al., 2017) was used here for removal of potentially interfering matrix components such as small molecules, phospholipids, peptides (Ouyang et al., 2012) and the added Triton X-100 surfactant which could otherwise have a dramatic impact on MS sensitivity (Cox and Emili, 2006). Considering the molecular weight difference between C9-S and C9-L, i.e., 25 and 54 kDa, and the potential solubility of smaller proteins in organic precipitation solvents (Lopes et al., 2014), we ascertained the equivalent recovery for both isoforms, ensuring accurate quantification. The signal observed for both C9-S and C9-L unique peptides in transfected HEK293 cells further confirmed the correct extraction and precipitation of C9ORF72 isoforms.

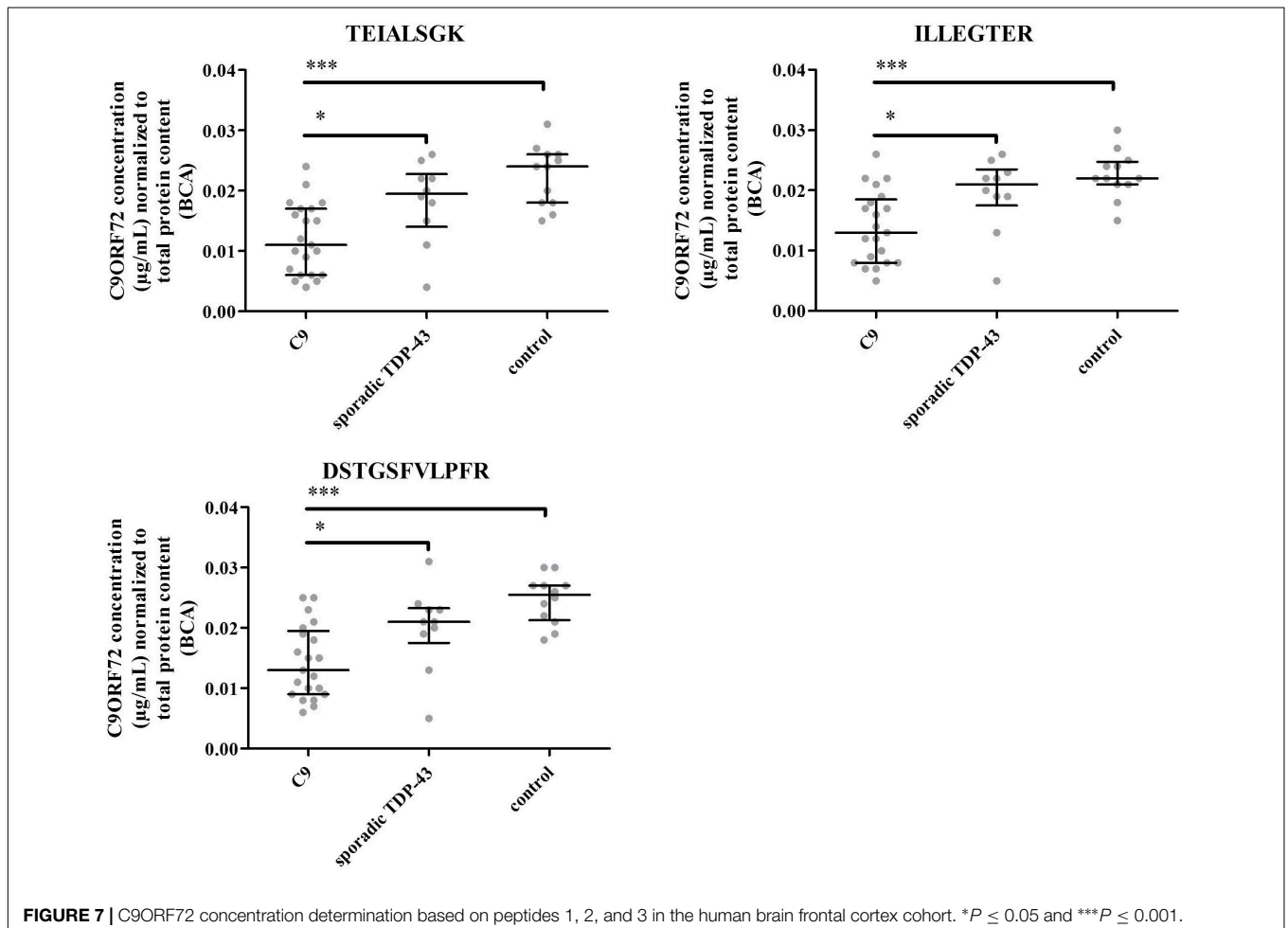
The validation experiments demonstrated that the precision provided by the method was satisfactory, with coefficients of variation below 20% and its ability to determine C9ORF72 relative concentration in human brain samples with a LLOQ at 50 ng/mL. Prior work with pellet digestion reported LLOQ around 1 μg/mL for quantification of therapeutic antibodies in plasma (Ouyang et al., 2012; Becher et al., 2017). In the present work, the use of PRM mode for peptide detection offered additional selectivity linked to the high resolution and mass precision measurements in the Orbitrap and the opportunity to accumulate fragment ions in the C-trap, eventually summed-up for higher signal intensity. Regarding signature peptide selection, chimeric or humanized therapeutic antibodies are more and more alike and so only a few peptides are unique



fivefold lower protein content in brain extract found by BCA may have contributed to reduced deleterious matrix effects compared to plasma. Most probably, a combination of these factors explains the gain in sensibility obtained for C9ORF72.

So far, C9ORF72 quantification was only accomplished by Western blot using in-house generated antibodies. The new MS assay allowed for the first time monitoring of C9ORF72 isoforms in human brain samples, without potential interference due to cross-reactivity of antibodies. First, we observed equivalent brain levels for C9-total and C9-L, indicating a low abundance of C9-S within the assay variability, i.e., a concentration below 20% of C9-L. Here, a similar peptide release between the recombinant and endogenous C9-S/C9-L was assumed since the protocol denature proteins during tissue lysis and pellet digestion which enhances digestion by protein unfolding (Pritchard et al., 2014). The low abundance of C9-S is further strengthened by the undetected unique C9-S peptide regardless of the evaluated brain protein extraction protocol, including those previously published for C9ORF72 (Xiao et al., 2015), whereas the peptide was well detectable in transfected HEK293 cells. Considering the lower abundance of the short isoform, deeper fractionation of the sample, for instance through enrichment of gray matter

to one therapeutic antibody. These unique peptides do not necessarily have the best physicochemical properties to be efficiently ionized by electrospray ionization whereas C9ORF72 peptides were selected based on signal intensity. Also, the



or subcellular fractions, could be considered to increase assay sensitivity. Also, post-translational modifications could impact peptide detection by our targeted method, even though none was reported for any of the selected peptides by now. In previous Western blot reports, Waite et al. (Waite et al., 2014) found that C9-S was in lower abundance than C9-L but questioned the specificity of their antibody, two bands being present at 27 and 29 kDa. Although, Xiao et al. (2015) detected the C9-S in frontal cortex using isoform specific antibodies, the relative abundance between C9-L and C9-S was not determined, probably in relation to the inherent limitation of protein quantification by Western blotting which depends on the affinity and specificity of the reagents (Aebersold et al., 2013). Next, we were able to confirm in a collection of 43 frontal cortices the diminution of the C9ORF72 long protein concentration in C9ORF72 FTD patients corroborating previous observations by the two Westernblot studies (Waite et al., 2014; Xiao et al., 2015) and at the mRNA level. The age of onset of the disease is highly variable as well as the number of expansion, the precise determination of C9ORF72 levels afforded by the new assay can be used to investigate correlations between the length of expansion and the levels of C9ORF72.

In summary, an efficient protocol was developed for quantification of C9ORF72 isoforms in brain samples by MS. Combination of optimal sample preparation and targeted high-resolution MS demonstrated robust and efficient quantification ability. This new assay has the advantage of being based on MS, avoiding the potential cross-reactivity of antibodies and simplifying implementation in various laboratories (Addona et al., 2009). C9ORF72 long isoform was significantly decreased in carriers of C9ORF72 expansion in comparison with controls and non-genetic FTD patients with or without ALS, corroborating prior observations made by Western blot and at the mRNA level. Whereas some studies reported a short isoform, here it represents less than 20% of the long one, suggesting possible non-specificity or cross-reactivity of antibodies. To our knowledge, this is the first report of a MS-based quantification assay for C9ORF72 proteins. This method needs to be further applied to other biological matrixes of a more relevant diagnostic nature and potentially to follow treatment efficacy in the future. This method could be easily implemented to mice models of C9ORF72 FTD or other animal model owing to a highly conserved sequence, in order to advance understanding of the contribution of C9ORF72 to disease mechanisms.

## REFERENCES

- Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., et al. (2009). Multi-Site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* 27, 633–641. doi: 10.1038/nbt.1546
- Aebersold, R., Burlingame, A. L., and Bradshaw, R. A. (2013). Western blots versus selected reaction monitoring assays: time to turn the tables? *ASBMB* 12, 2381–2382.
- An, B., Zhang, M., Johnson, R. W., and Qu, J. (2015). Surfactant-aided precipitation/on-pellet-digestion (SOD) procedure provides robust and rapid sample preparation for reproducible, accurate and sensitive LC/MS quantification of therapeutic protein in plasma and

## AUTHOR CONTRIBUTIONS

AV, CF, ILB, FL, VA, and FB contributed conception and design of the study. AV, CF, AC, ML, and VA conducted the experiments. NeuroCEB brain Bank provided samples. AV and FB wrote the first draft of the manuscript. ILB, CF, and VA wrote sections of the manuscript. CF, FF, FE, ILB, CJ, and FL critically reviewed the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

The research leading to these results has received funding from the program “Investissements d’avenir” ANR-10-IAIHU-06.

## ACKNOWLEDGMENTS

The authors thank the “NeuroCEB” Brain Bank (GIE NeuroCEB BB-0033-00011); the brain donation program is funded by a consortium of patient associations: ARSEP (association for research on multiple sclerosis), CSC (association for research on cerebellar ataxias), France Parkinson, France Alzheimer and LECMA-Vaincre Alzheimer. The NeuroCEB Brain bank network includes: Prof. Franck Letournel (CHU Angers), Prof. Anne Vital (CHU Bordeaux), Prof. Françoise Chapon (CHU Caen), Dr. Catherine Godfraind (CHU Clermont-Ferrand), Prof. Claude-Alain Maurice (CHU Lille), Dr. Vincent Deramecourt (CHU Lille), Dr. David Meyronnet (CHU Lyon), Dr. Nathalie Streichenberger (CHU Lyon), Dr. André Maues de Paula (CHU Marseille), Dr. Valérie Rigau (CHU Montpellier), Dr. Fanny Vandebos-Burel (Nice), Prof. Charles Duyckaerts (CHU PS Paris), Prof. Danielle Seilhean (CHU PS, Paris), Dr. Véronique Sazdovitch (CHU PS Paris), Dr. Serge Milin (CHU Poitiers), Dan Christian Chiforeanu (CHU Rennes), Prof. Annie Laquerrière (CHU Rouen), Dr. Béatrice Lannes (CHU Strasbourg).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00589/full#supplementary-material>

- tissues. *Anal. Chem.* 87, 4023–4029. doi: 10.1021/acs.analchem.5b00350
- Becher, F., Ciccolini, J., Imbs, D. C., Marin, C., Fournel, C., Dupuis, C., et al. (2017). A simple and rapid LC-MS/MS method for therapeutic drug monitoring of cetuximab: a GPCO-UNICANCER proof of concept study in head-and-neck cancer patients. *Sci. Rep.* 7:2714. doi: 10.1038/s41598-017-02821-x
- Chen, D., Wang, Y., and Chin, E. R. (2015). Activation of the endoplasmic reticulum stress response in skeletal muscle of G93A<sup>SOD1</sup> amyotrophic lateral sclerosis mice. *Front. Cell. Neurosci.* 9:170. doi: 10.3389/fncel.2015.00170
- Chen, Y. T., Tuan, L. P., Chen, H. W., Wei, I. A., Chou, M. Y., Chen, H. M., et al. (2015). Quantitative analysis of prostate specific antigen isoforms using immunoprecipitation and stable isotope labeling mass spectrometry. *Anal. Chem.* 87, 545–553. doi: 10.1021/ac503306g

- Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., et al. (2013). Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Ann. Neurol.* 74, 180–187. doi: 10.1002/ana.23946
- Cox, B., and Emili, A. (2006). Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nat. Protoc.* 1, 1872–1878. doi: 10.1038/nprot.2006.273
- Davidson, Y. S., Robinson, A. C., Rollinson, S., Pickering-Brown, S., Xiao, S., Robertson, J., et al. (2018). Immunohistochemical detection of C9orf72 protein in frontotemporal lobar degeneration and motor neurone disease: patterns of immunostaining and an evaluation of commercial antibodies. *Amyotrophic Lateral Sclerosis Frontotemporal Degener.* 19, 102–111. doi: 10.1080/21678421.2017.1359304
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-Linked FTD and ALS. *Neuron* 72, 245–256. doi: 10.1016/j.neuron.2011.09.011
- Dupré, M., Gilquin, B., Fenaille, F., Feraudet-Tarisse, C., Dano, J., Ferro, M., et al. (2015). Multiplex quantification of protein toxins in human biofluids and food matrices using immunoprecipitation and high-resolution targeted mass spectrometry. *Anal. Chem.* 87, 8473–8480. doi: 10.1021/acs.analchem.5b01900
- Gong, C., Zheng, N., Zeng, J., Aubry, A. F., and Arnold, M. E. (2015). Post-pellet-digestion precipitation and solid phase extraction: a practical and efficient workflow to extract surrogate peptides for ultra-high performance liquid chromatography – tandem mass spectrometry bioanalysis of a therapeutic antibody in the low ng/mL range. *J. Chromatogr. A* 1424, 27–36. doi: 10.1016/j.chroma.2015.10.049
- Jedrychowski, M. P., Wrann, C. D., Paulo, J. A., Gerber, K. K., Szpyt, J., Robinson, M. M., et al. (2015). Detection and quantitation of circulating human irisin by tandem mass spectrometry. *Cell Metab.* 22, 734–740. doi: 10.1016/j.cmet.2015.08.001
- Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H. R., et al. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4530–E4539. doi: 10.1073/pnas.1318835110
- Lesur, A., Ancheva, L., Kim, Y. J., Berchem, G., van Oostrum, J., and Domon, B. (2015). Screening protein isoforms predictive for cancer using immunoaffinity capture and fast LC-MS in PRM mode. *PROTEOMICS – Clin. Appl.* 9, 695–705. doi: 10.1002/prca.201400158
- Levine, T. P., Daniels, R. D., Gatta, A. T., Wong, L. H., and Hayes, M. J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics (Oxford, England)* 29, 499–503. doi: 10.1093/bioinformatics/bts725
- Liu, Y. T., Kardosh, A., Cooc, J., and Schönthal, A. H. (2006). Potential misidentification of cyclooxygenase-2 by western blot analysis and prevention through the inclusion of appropriate controls. *Mol. Biotechnol.* 34, 329–335.
- Lopes, F., Cowan, D. A., Thevis, M., Thomas, A., and Parkin, M. C. (2014). Quantification of intact human insulin-like growth factor-i in serum by nano-ultrahigh-performance liquid chromatography/tandem mass spectrometry: quantification of insulin-like growth factor-I by nanoUHPLC-MS/MS. *Rapid Commun. Mass Spectrom.* 28, 1426–1432. doi: 10.1002/rcm.6908
- McLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. doi: 10.1093/bioinformatics/btq054
- Majounie, E., Renton, A. E., Mok, K., Dopper, E. G. P., Waite, A., Rollinson, S., et al. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* 11, 323–330. doi: 10.1016/S1474-4422(12)70043-1
- Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., et al. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science* 339, 1335–1338. doi: 10.1126/science.1232927
- Ouyang, Z., Furlong, M. T., Wu, S., Slecza, B., Tamura, J., Wang, H., et al. (2012). Pellet digestion: a simple and efficient sample preparation technique for LC-MS/MS quantification of large therapeutic proteins in plasma. *Bioanalysis* 4, 17–28. doi: 10.4155/bio.11.286
- Pritchard, C., Groves, K. J., Biesenbruch, S., O'Connor, G., Ashcroft, A. E., Arsene, C., et al. (2014). Quantification of human growth hormone in serum with a labeled protein as an internal standard: essential considerations. *Anal. Chem.* 86, 6525–6532. doi: 10.1021/ac501032q
- Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., et al. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 134, 2456–2477. doi: 10.1093/brain/awr179
- Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Raphael Gibbs, S. R. J., Schymick, J. C., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268. doi: 10.1016/j.neuron.2011.09.010
- Rosso, S. M. (2003). Frontotemporal dementia in the netherlands: patient characteristics and prevalence estimates from a population-based study. *Brain* 126, 2016–2022. doi: 10.1093/brain/awg204
- Shevchenko, G., Musunuri, S., Wetterhall, M., and Bergquist, J. (2012). Comparison of extraction methods for the comprehensive analysis of mouse brain proteome using shotgun-based mass spectrometry. *J. Proteome Res.* 11, 2441–2451. doi: 10.1021/pr201169q
- Torsetnes, S. B., Nordlund, M. S., Paus, E., Halvorsen, T. G., and Reubsæet, L. (2013). Digging deeper into the field of the small cell lung cancer tumor marker ProGRP: a method for differentiation of its isoforms. *J. Proteome Res.* 12, 412–420. doi: 10.1021/pr300751j
- Umoh, M. E., Dammer, E. B., Dai, J., Duong, D. M., Lah, J. J., Levey, A. I., et al. (2018). A proteomic network approach across the ALS-FTD disease spectrum resolves clinical phenotypes and genetic vulnerability in human brain. *EMBO Mol. Med.* 10, 48–62. doi: 10.15252/emmm.201708202
- Waite, A. J., Bäumer, D., East, S., Neal, J., Morris, H. R., Ansorge, O., et al. (2014). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol. Aging* 35, 1779.e5–1779.e13. doi: 10.1016/j.neurobiolaging.2014.01.016
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322
- Xiao, S., MacNair, L., McGoldrick, P., McKeever, P. M., McLean, J. R., Zhang, M., et al. (2015). Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis: C9orf72 isoforms in ALS. *Ann. Neurol.* 78, 568–583. doi: 10.1002/ana.24469
- Xiao, S., MacNair, L., McLean, J., McGoldrick, P., McKeever, P., Soleimani, S., et al. (2016). C9orf72 isoforms in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Brain Res.* 1647, 43–49. doi: 10.1016/j.brainres.2016.04.062
- Zhang, D., Iyer, L. M., He, F., and Aravind, L. (2012). Discovery of novel DENN proteins: implications for the evolution of eukaryotic intracellular membrane structures and human disease. *Front. Genet.* 3:283. doi: 10.3389/fgene.2012.00283

**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.

The handling Editor declared a shared affiliation, though no other collaboration, with one of the authors FE.

Copyright © 2018 Viodé, Fournier, Camuzat, Fenaille, NeuroCEB Brain Bank, Latouche, Elahi, Le Ber, Junot, Lamari, Anquetil and Becher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



## SUPPORTING INFORMATION

### **New antibody-free mass spectrometry-based quantification reveals that C9ORF72 long protein isoform is reduced in the frontal cortex of hexanucleotide-repeat expansion carriers.**

Arthur Viodé<sup>1</sup>, Clemence Fournier<sup>2,6</sup>, Agnès Camuzat<sup>2,3</sup>, François Fenaille<sup>1</sup>, NeuroCEB brain bank, Morwena Latouche<sup>3</sup>, Fanny Elahi<sup>4</sup>, Isabelle Le Ber<sup>2,5,6</sup>, Christophe Junot<sup>1</sup>, Foudil Lamari<sup>7,8</sup>, & Vincent Anquetil<sup>2,6</sup>, François Becher<sup>1\*</sup>

1. Service de Pharmacologie et Immunoanalyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments (LEMM), CEA, INRA, Université Paris Saclay, F-91191 Gif-sur-Yvette cedex, France.
2. Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, F-75013, Paris, France.
3. Ecole Pratique des Hautes Etudes - EPHE, PSL Research University, 75014, Paris, France.
4. Memory and Aging Center, Department of Neurology, University of California, San Francisco, 675 Nelson Rising Lane, Suite 190, San Francisco, California 94158, USA
5. Institute of Memory and Alzheimer's Disease (IM2A), Centre of excellence of neurodegenerative disease (CoEN), ICM, APHP Department of Neurology, Hopital Pitié-Salpêtrière, University Paris 6, Paris, France.
6. Assistance Publique – Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Centre de référence Démences Rares, F-75013, Paris, France.
7. Assistance Publique – Hôpitaux de Paris Hôpitaux universitaires Pitié-Salpêtrière-Charles-Foix Service de Biochimie Métabolique, 75013, Paris, France.
8. GRC 13 Neurométabolisme – UPMC, Sorbonne Université, F-75013, Paris, France

#### **Table of content:**

Supplementary Method .....	150
Table S1.....	<b>Error! Bookmark not defined.</b>
Table S2.....	152
Table S3.....	<b>Error! Bookmark not defined.</b>
Table S4.....	154
Figure S1 .....	<b>Error! Bookmark not defined.</b>
Figure S2. ....	<b>Error! Bookmark not defined.</b>
Figure S3. ....	<b>Error! Bookmark not defined.</b>
Figure S4. ....	<b>Error! Bookmark not defined.</b>

### **Supplementary Method: C9ORF72 peptides identification by Data-Dependent Acquisition**

The identification of C9ORF72 peptides was done manually. An in-silico-digest of C9orf72 isoforms was performed. The m/z ratio of the resulting peptides was searched within the full MS scan following data-dependent acquisition (top 5 ions) on the Q-Exactive. Peptides were assumed as identified, if the m/z ratio of the parent measured in the Orbitrap was within 5 ppm to the theoretical m/z ratio, and if at least 3 fragment ions from the corresponding MS/MS spectrum could be assigned. The same method was applied to a brain sample using database search with Proteome Discoverer 1.3 (Thermo Scientific) and SEQUEST.



## Supplementary Tables and Figures

**Table S1.** Effect of gradient length on the peak area of targeted peptides of endogenous C9ORF72 in a human brain sample.

	<b>TEIALSGK</b>	<b>ILLEGTER</b>	<b>DSTGSFVLPFR</b>
	Peak area (a.u.)	Peak area (a.u.)	Peak area (a.u.)
Short-gradient (15 min total run time)	123403	68044	142012
Long-gradient (30 min total run time)	159265	124394	195284
Gain (long vs. short)	30%	80%	40%

**Table S2.** MS parameters for PRM detection of C9ORF72 peptides as well as corresponding SIL internal standards

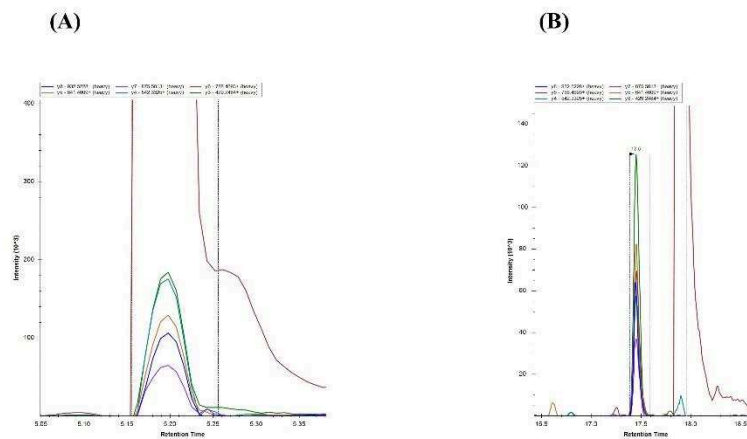
PEPTIDE SEQUENCE	AVERAGE MASS (Da)	PRECURSOR ION m/z	PRODUCT ION m/z*	NORMALIZED COLLISION ENERGY %	RETENTION TIME (MIN)
TEIALSGK	817.94	409.9 (z=2)	404.2504 (y4)	17	4.5
			475.2875 (y5)		
			588.3715 (y6)		
TEIALSGK <sup>[13C6; 15N2]</sup>	825.88	413.74 (z=2)	412.2646 (y4)	17	4.5
			483.3017 (y5)		
			596.3857 (y6)		
IILEGTER	930.08	465.97 (z=2)	462.2307 (y4)	17	6.2
			591.2733 (y5)		
			704.3573 (y6)		
IILEGTER <sup>[13C6; 15N4]</sup>	940.00	470.77 (z=2)	472.2390 (y4)	17	6.2
			601.2815 (y5)		
			714.3656 (y6)		
DSTGSFVLPFR	1225.37	613.31 (z=2)	419.2401 (y3)	20	17.1
			532.3242 (y4)		
			631.3926 (y5)		
			778.4610 (y6)		
			865.4930 (y7)		
DSTGSFVLPFR <sup>[13C6; 15N4]</sup>	1235.3	618.32 (z=2)	429.2484 (y3)	20	17.1
			542.3325 (y4)		
			641.4009 (y5)		
			455.4693 (y6)		
			875.5013 (y7)		
SHSVPEEIDIADTVLNDDDIGDSCHGFLK	3441.7	861.3 (z=4)	917.8992 (y16 +2)	20	20.7
			974.4413 (y17 +2)		
			1132.0128 (y20 +2)		
			1167.5314 (y21 +2)		
			921.9063 (y16 +2)		
SHSVPEEIDIADTVLNDDDIGDSCHGFLK <sup>[13C6; 15N2]</sup>	3449.7	863.15 (z=4)	921.9063 (y16 +2)	20	20.7
			978.4484 (y17 +2)		
			1136.0199 (y20 +2)		
			1171.5385 (y21 +2)		

**Table S3.** SPE extraction recovery and CV% measured for each quantitative peptide.

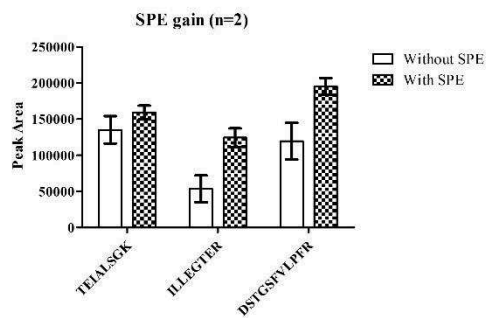
	<b>TEIALSGK</b>		<b>ILLEGTER</b>		<b>DSTGSFVLPFR</b>	
	Mean	CV	Mean	CV	Mean	CV
Recovery	68%	7%	45%	3%	70%	16%

**Table S4.** Mean difference between C9ORF72 protein concentrations obtained by peptides 1 and 2, peptides 1 and 3, peptides 2 and 3, normalized to the mean total concentration (mean of concentrations determined by peptides 1 and 2).

<b>Patient classification</b>	<b>Pep 1 – Pep 2</b>	<b>Pep 1 – Pep 3</b>	<b>Pep 2- Pep 3</b>
C9	14%	16%	2%
TDP-43	6%	10%	4%
Control	1%	11%	10%

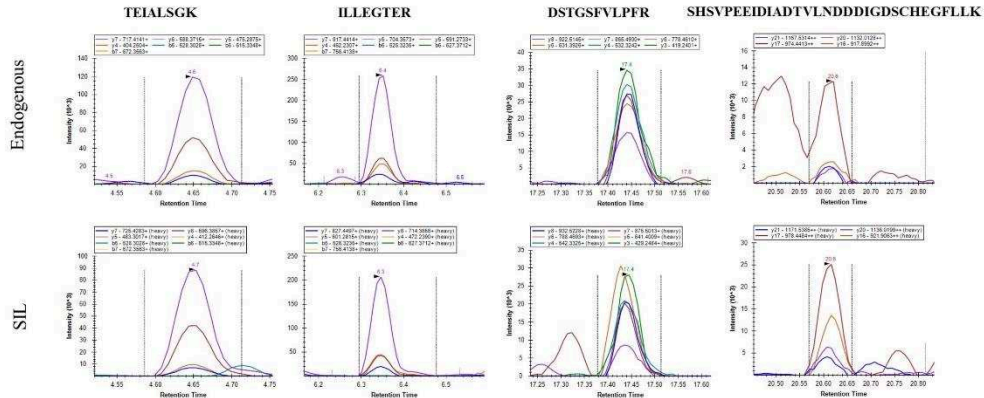


**Figure S1.** Effect of gradient length on the separation of peptide DSTGGSFVLPFR (SIL version) from interference in human brain sample, **A)** total run time of 15min and **B)** total run time of 30 min.

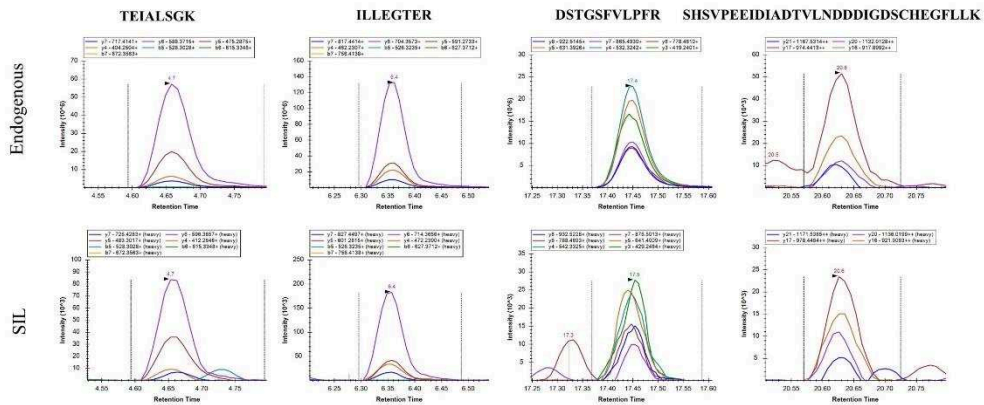


**Figure S2.** SPE gain in a human brain sample, measured for each quantitative peptide.

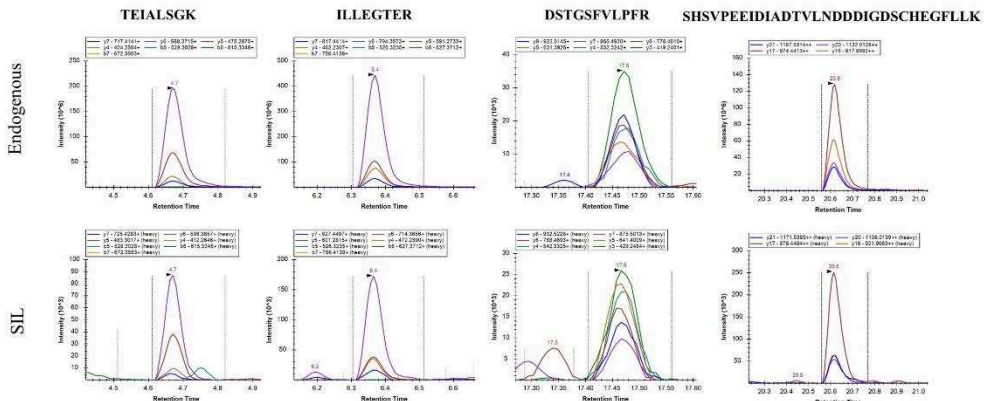
HEK CONTROL



HEK C9-L



HEK C9-S



**Figure S3.** C9ORF72 peptides detection in HEK293 cells mock-transfected (HEK control) and transfected with plasmids coding for the long (HEK C9-L) or the short (HEK C9-S) isoforms.

Q6DFW0	CI072_MOUSE	1	NSTICPPSPAVAKTETALSGLPLLAATFAYNDNILGPRVRHINAPKTEQVLLSDGEIT	60
Q96LT7	CI072_HUMAN	1	NSTLCPPSPAVAKTETALSGLPLLAATFAYNDNILGPRVRHINAPKTEQVLLSDGEIT	60
Q6DFW0	CI072_MOUSE	61	FLAHHTLNGEILRNAESGAIKVKFFVLSEKGVIIIVSLIFDGNWNGDRSTYGLSIIILPQTE	120
Q96LT7	CI072_HUMAN	61	FLAHHTLNGEILRNAESGAIKVKFFVLSEKGVIIIVSLIFDGNWNGDRSTYGLSIIILPQTE	120
Q6DFW0	CI072_MOUSE	121	LSFYLP.LHRVCVDR.LTHIIRKGRD.WHKERQENVKD.LLESTERMEDQGSIIIPMLTGEV	180
Q96LT7	CI072_HUMAN	121	LSFYLP.LHRVCVDR.LTHIIRKGRD.WHKERQENVKD.LLESTERMEDQGSIIIPMLTGEV	180
Q6DFW0	CI072_MOUSE	181	IPWHELLASHKSHSVPEIDIDTIVLNDODIGDSCHGEGFLNIAISSHLQTCGCSVWVGSS	240
Q96LT7	CI072_HUMAN	181	IPWHELLSSNKSHSVPEEIDIDTIVLNDODIGDSCHGEGFLNIAISSHLQTCGCSVWVGSS	240
Q6DFW0	CI072_MOUSE	241	AEKWKIVRTLCLFLTPAERKCSR.LCEAESSFKYESGLFVQGLLKQISFVLPFRQWY	300
Q96LT7	CI072_HUMAN	241	AEKWKIVRTLCLFLTPAERKCSR.LCEAESSFKYESGLFVQGLLKQISFVLPFRQWY	300
Q6DFW0	CI072_MOUSE	361	APVPTTHIDVDVNTVQPPCHEIYNQRVHRS.ELTAFWRATSEEDMAQDTIIYDESF	360
Q96LT7	CI072_HUMAN	361	APVPTTHIDVDVNTVQPPCHEIYNQRVHRS.ELTAFWRATSEEDMAQDTIIYDESF	360
Q6DFW0	CI072_MOUSE	361	TPDLNIFQDVLHRD.LVKAFLDQVFHLKPGLSLRSTFLAQFL.LLHRRKALTLIKYIEDDT	420
Q96LT7	CI072_HUMAN	361	TPDLNIFQDVLHRD.LVKAFLDQVFQKPGLSLRSTFLAQFL.LLHRRKALTLIKYIEDDT	420
Q6DFW0	CI072_MOUSE	421	-----	420
Q96LT7	CI072_HUMAN	421	QKGGKPKSLRNLIKIDLDLTAEGDLNIIIMALAEIKRPLGHSFIFGRPFYTSVQERDVLMT	480
Q6DFW0	CI072_MOUSE	421	-	420
Q96LT7	CI072_HUMAN	481	F	481

**Figure S4.** Sequence alignment of human C9-L vs. mouse C9orf72.

### **Conclusion de l'article :**

Nous avons développé une méthode de quantification des isoformes de la C9ORF72 par spectrométrie de masse. La particularité de cette méthode est qu'elle ne nécessite pas l'utilisation d'anticorps qui sont connus pour être peu spécifique. Après optimisation du protocole d'homogénéisation du cerveau pour s'assurer de l'extraction optimale de la C9ORF72 et des autres protéines. Nous avons opté pour un protocole de précipitation des protéines au méthanol<sup>114,117</sup> suivi d'une extraction par SPE. Cette méthode, après optimisation, a été appliquée à l'analyse d'une cohorte de cortex frontal comprenant des témoins, des DFT avec expansion (G<sub>4</sub>C<sub>2</sub>)<sub>n</sub> et des DFT sporadique avec agrégats de TDP-43. L'analyse de cette cohorte nous a permis de mettre en évidence une diminution de la C9-L (environ 40%) pour les cas ayant une expansion (G<sub>4</sub>C<sub>2</sub>)<sub>n</sub> dans le gène de la C9ORF72. En revanche, nous n'avons pas pu conclure sur la présence ou non de l'isoforme C9-S. En effet, nous n'avons pas détecté le peptide unique de la forme courte dans les échantillons et nous pouvons uniquement conclure sur le fait qu'il représente moins de 20% (variabilité de la méthode de quantification par différence) de C9-L. Néanmoins, cette méthode facilement transposable dans d'autres laboratoires répond à un besoin de quantification de cette protéine notamment pour le suivi thérapeutique dans des modèles animaux. En effet, un des mécanismes de toxicité de cette expansion serait liée à une haploinsuffisance et donc rétablir son niveau permettrait de mitiger la neurodégénérescence<sup>331</sup>. Avec cette idée en tête nous avons exploré des matrices plus accessibles comme les PBMCs où l'expansion est aussi retrouvée<sup>343</sup>. Nous avons aussi exploré l'applicabilité de notre méthode pour la quantification de la C9ORF72 en cellules.

### **3. Extension de la méthode pour la quantification de la C9ORF72 dans les PBMCs**

Nous avons développé une méthode de quantification de la C9ORF72 dans les PBMCs. La taille de l'expansion est dépendante du type de tissus et donc la taille de l'expansion en PBMCs ne reflète pas nécessairement ce qui se passe dans le cerveau . Nous nous sommes donc plutôt intéressés à un changement de ratio entre les deux isoformes. Une stratégie similaire à celle publiée à été développée : précipitation des protéines au méthanol et SPE. Dans un premier temps nous avons optimisé la lyse des cellules. Les développements ont été fait à partir de 20\*10<sup>6</sup> cellules afin d'avoir un signal suffisant pour sa quantification.



### a) Extraction

Afin d'extraire au mieux la C9ORF72 des cellules nous avons testé différents tampons de lyse et temps de sonication (voir figure 61). Le PBS+0.1% de triton X-100 permet une meilleure extraction de la C9ORF72.

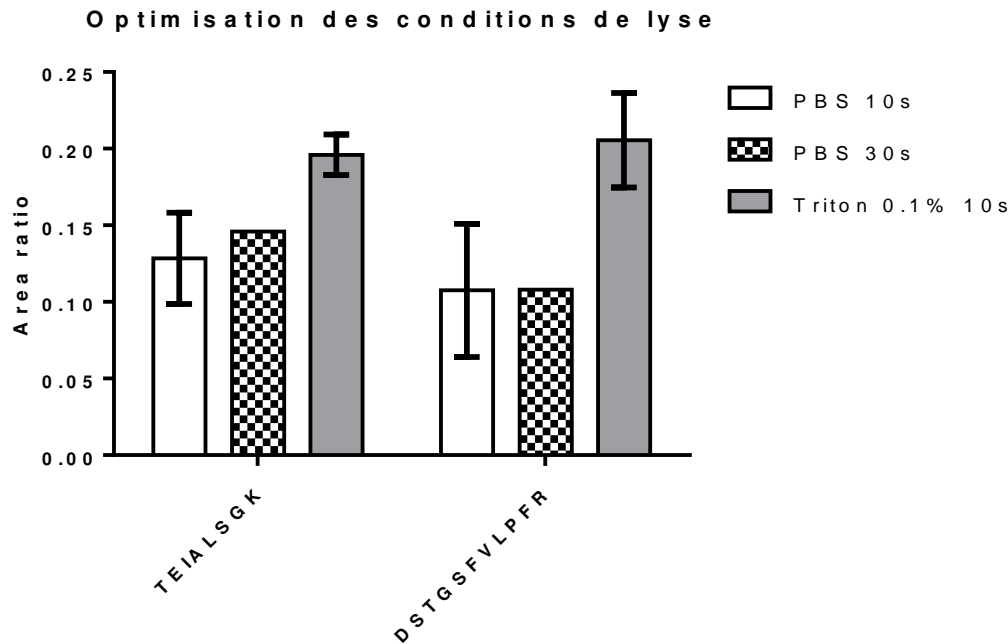


Figure 61 Optimisation de la lyse des PBMCs. PBS 10s (n=4), PBS 30s (n=1) et PBS + triton 0.1% 10s (n=4)

### b) Evaluation de la méthode

La stabilité de la C9ORF72 est un élément important pour sa quantification. Nous avons dans un premier temps évalué sa stabilité après lyse aux ultrasons avec une quantité d'anti-protéase 1X (identique au protocole publié pour la quantification de la C9ORF72 dans le cortex frontal). Les protéines ont soit été directement précipitées ou après 60 min sur glace (voir figure 62). Les résultats ont montré une nette diminution de la quantité de C9ORF72 après 60 min probablement causée par des protéases. Afin de remédier à ce problème la quantité d'anti-protéase a été augmentée d'un facteur 5. Dans ces conditions la C9ORF72 est stable même après 60 min. Contrairement à ce que nous avons pu observer pour le cerveau, la protéine C9ORF72 se dégrade après lyse des cellules à 4°C. Nous avons donc augmenté la quantité d'anti-protéase d'un facteur 5.

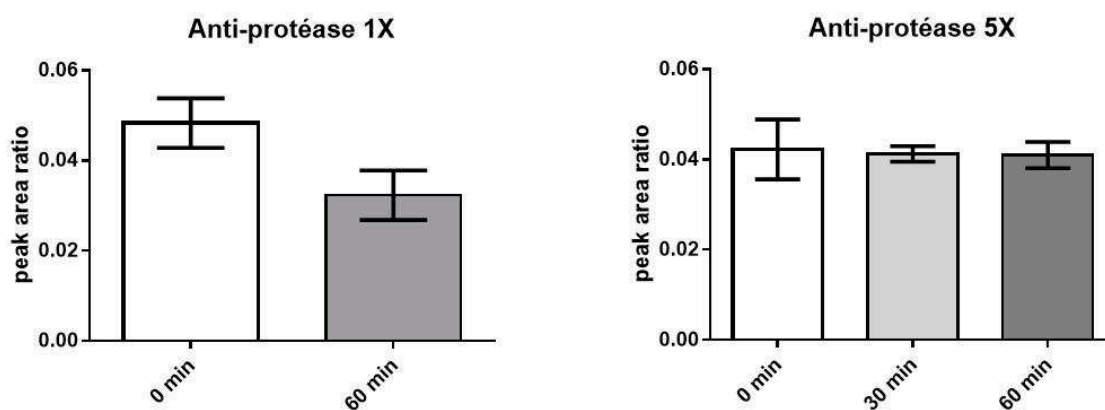


Figure 62 Evaluation de la stabilité de la C9ORF72 après lyse des cellules sur glace (n=2)

Dans un second temps nous avons évalué la répétabilité de la méthode, pour cela 5 échantillons de PBMCs ont été analysés sur 3 jours consécutifs, préparés selon le protocole développé au laboratoire<sup>345</sup>. La répétabilité sur un même jour est acceptable et comprise entre 5 à 21%, en recherche la variabilité sur les 3 jours est moins bonne et est comprise entre 21% et 25%. Cette variabilité peut être expliquée par un manque de sensibilité.

Tableau 11 Evaluation de la répétabilité de la méthode pour 5 analyses le même et sur 3 jours consécutifs de PBMCs

Interday validation (n=5)	TEIALSGK		ILLEGTER		DSTGSFVLPFR	
	Mean	CV	Mean	CV	Mean	CV
Jour 1	0.043	18%	0.025	18%	0.082	18%
Jour 2	0.069	13%	0.037	10%	0.108	14%
Jour 3	0.048	21%	0.027	5%	0.072	15%
<b>Interday</b>	<b>0.053</b>	<b>25%</b>	<b>0.03</b>	<b>23%</b>	<b>0.087</b>	<b>21%</b>

Enfin la méthode a été évaluée en terme de linéarité. Une gamme de calibration de 10 points (de 5 ng/mL à 5 µg/mL) a été réalisée avec la protéine recombinante C9-L. La méthode est linéaire de 50 ng/mL à 5 µg/mL ( cf figure 63).

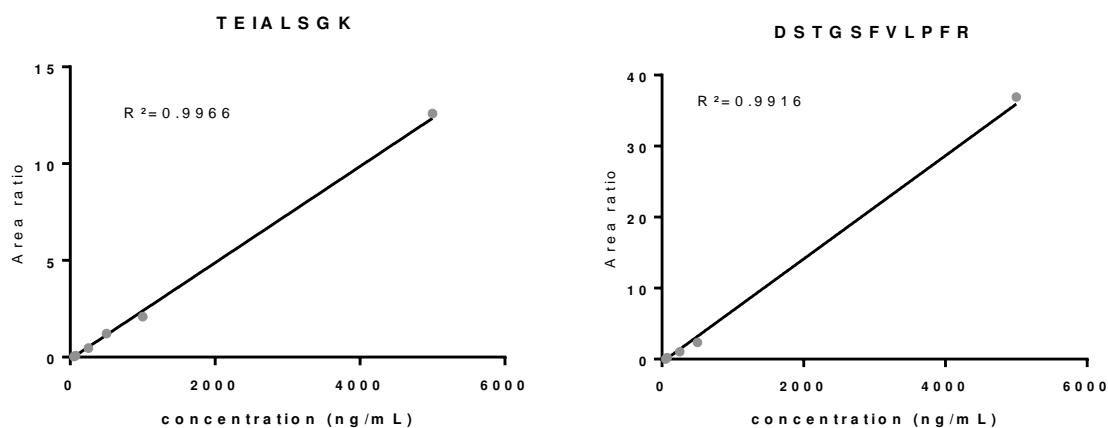


Figure 63 Gamme de calibration obtenu en plasma dilué au 5<sup>ème</sup>. La gamme est de 5 ng/mL à 5 µg/mL/ La LOQ pour TEIALSGK et DSTGSFVLPFR est de 50 ng/mL.

### c) Applications à des échantillons humains

Après évaluation de la méthode, nous l'avons appliqué à des échantillons sanguins provenant de patients ayant des démences frontotemporales (voir tableau 12).

Tableau 12 Liste des échantillons de PBMCs analysés

Type d'échantillons	Nombre
Contrôle	3
Maladie d'Alzheimer	8
Mutation PGRN	5
DTF sporadique	5
C9ORF72	9
C9ORF72 préemptomatique et suspicion	4

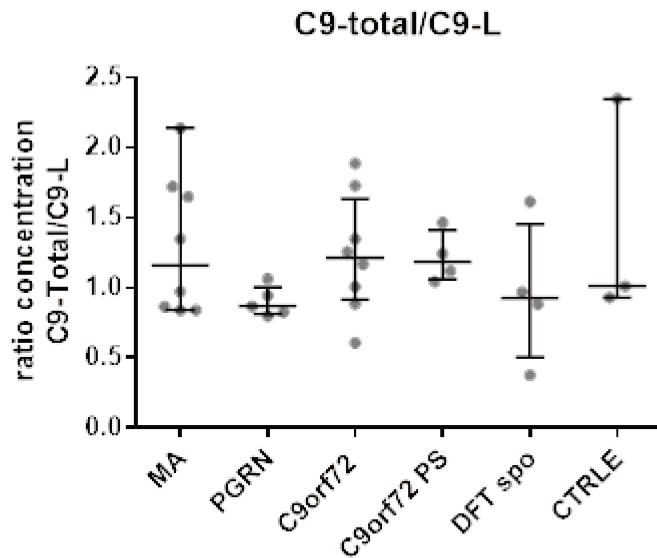


Figure 64 Ratio de concentration de la C9-total/C9-L entre les différents groupes

La méthode développée pour la quantification de la C9ORF72 dans les PBMCs et appliquée à des échantillons de patients s'est avérée pas suffisamment répétable et sensible. En effet, dans certains cas nous trouvons un ratio inférieur à 1 (voir figure64), ce qui n'est pas cohérent. La méthode pour être appliquée pour la quantification de la C9ORF72 en PBMCs devra être améliorée, notamment par le développement d'une méthode par immunoprécipitation.

## Perspectives

### Recherche de modifications post-traductionnelles dans les cellules HEK :

Dans les cellules HEK 293 transfectées exprimant les formes courtes et longues nous avons cherché par une analyse non ciblée (DDA) à identifier des modifications post-traductionnelles sur les isoformes de la C9ORF72 (figure 66). Pour cela nous avons utilisé le logiciel protein discoverer v1.4. Nous avons ensuite confirmé les identifications manuellement (cf figure 67 et 68) afin d'identifier 2 sites de phosphorylations sur le peptide unique de la C9-S (SHSVPEEIDIADTVLNDDDIGDSCHEGFLK) en serine 3 et 23.

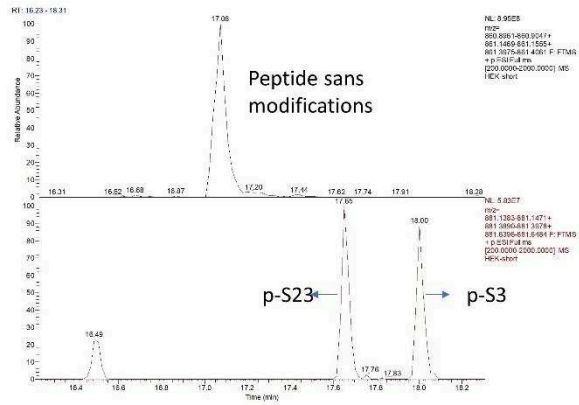


Figure 65 XICs peptide unique C9-S ainsi que de ses correspondants phosphorylés

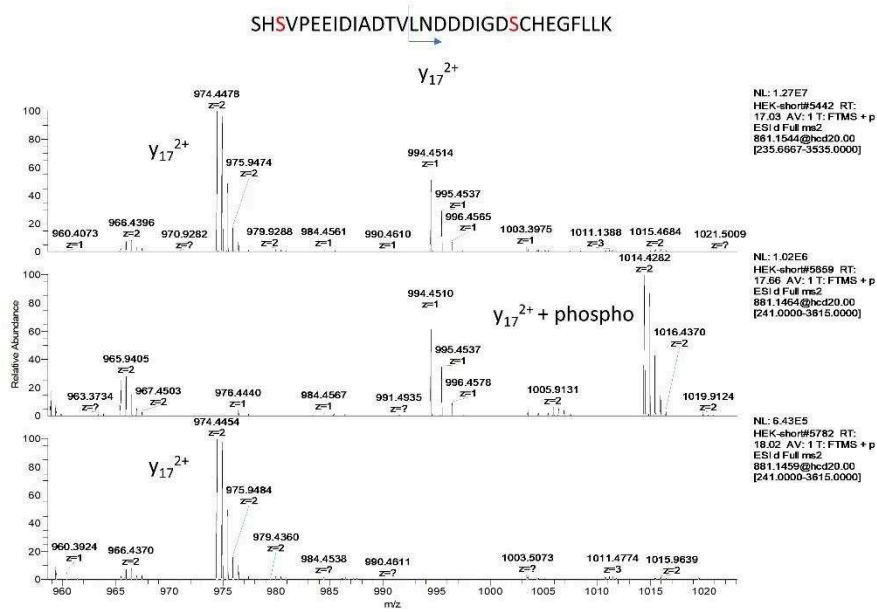


Figure 66 Spectre MS des ions fragments de la forme sans phosphorylation et des peptides phosphorylés. Exemple pour l'ion y<sub>17</sub> 2+

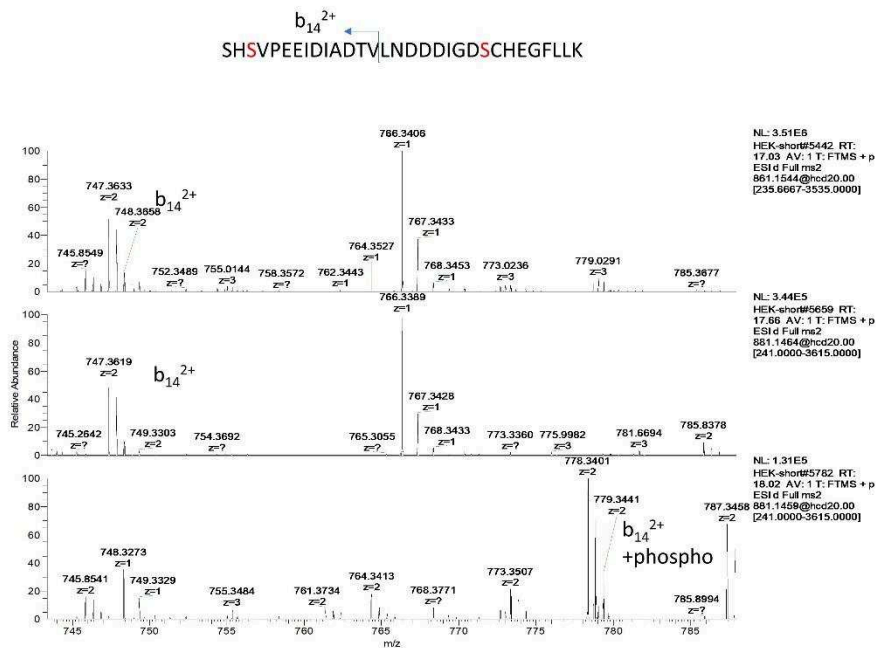


Figure 67 Spectre MS des ions fragments de la forme sans phosphorylation et des peptides phosphorylés. Exemple pour l'ion  $y_{14}^{2+}$

Le peptide comprenant les deux sites de phosphorylation simultanément n'a pas été observé. Ces deux sites de phosphorylation ne sont pas reportés dans la littérature. En ce qui concerne la forme longue, le peptide tryptique formé est composé de 54 acides aminés et n'a donc pas les bonnes propriétés physico-chimiques pour être analysé dans nos conditions. Nous n'avons pas pu identifier ces phosphorylations sur C9-L. Il faudrait pouvoir confirmer ce résultat sur C9-S à un niveau endogène. Pour cela nous pourrions enrichir les peptides phosphorylés avec une colonne IMAC ou en utilisant des anticorps p-S en combinaison avec une méthode de masse ciblée par PRM. Il faudrait également améliorer la sensibilité globale de la méthode pour détecter le peptide unique de C9-S. Néanmoins, nous ne pouvons pas exclure que ces modifications soient un artefact causé par la transfection des cellules HEK 293

Afin d'avoir une meilleure compréhension des mécanismes impliqués dans les DFTs ainsi que de trouver des biomarqueurs potentiels nous avons développé une méthode de quantification multiplex par PRM pour 59 protéines. Ces protéines sont impactées au niveau de leur ARN messenger ou sont des marqueurs de cellules ou elles sont connues comme étant impliquées dans les DFT. Nous avons appliqué cette méthode à la quantification de la même cohorte que pour la quantification de la C9ORF72 dans le cortex frontal. Cette méthode a été optimisée en termes de préparation d'échantillon puis évalué en termes de linéarité et fidélité.

#### **4. Développement d'une méthode de quantification multiplexe par PRM**

Les modes d'acquisition SRM<sup>288</sup> et PRM<sup>129</sup> permettent la quantification multiplexe de plusieurs centaines de peptides au cours d'une analyse. Du fait d'une meilleure sensibilité en comparaison des modes non ciblées, DDA et DIA, et du nombre de protéines ciblées, nous avons opté pour le développement d'une méthode multiplexe par PRM.

En effet, nous avons développé une méthode de quantification relative ciblant 59 protéines dans des homogénats de cortex frontal humain par chromatographie liquide couplée à la spectrométrie de masse à haute résolution. Pour chaque protéine au moins 1 peptide est ciblée ainsi que son correspond marqué aux isotopes stables (<sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>2</sub>-labeled lysine et <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>4</sub>-labeled arginine).

Une optimisation du protocole de lyse des tissus cérébraux et de l'extraction des protéines à partir de l'homogénat a été effectuée dans un premier temps. Puis, deux types de débits chromatographique ont été comparée (UHPLC : 500 µL/min et chromatographie capillaire : 6 µL/min). Enfin, une méthode non-ciblée DIA a été mise au point et appliquée aux mêmes échantillons.

##### **a) Liste des protéines**

Nous avons établi une liste de protéine à quantifier basée sur les données de l'ICM (ARN), de la littérature<sup>346</sup> ainsi que des protéines marqueurs des types cellulaire spécifiques. Par exemple, la TDP-43 a été choisi car elle forme des inclusions cérébrales dans les DFT. Le nombre de protéines constituant la liste initiale était de 65 protéines. Plusieurs critères ont permis d'établir une première liste de peptides. Premièrement, les peptides sélectionnés doivent être protéotypique. Deuxièmement, les peptides contenant une méthionine ou une cystéine n'ont pas été sélectionnée prioritairement, de même pour les peptides ayant une glutamine en N-terminale. Enfin, les peptides contenant une séquence K.K, K.R., R.R et R.K ont été exclus. A partir de ces critères une première sélection des peptides s'est faite soit à partir

des données d'une analyse non ciblée en DDA soit à partir de la base de données peptide atlas en se basant sur les scores de détectabilités des peptides. Ces scores sont calculés à partir de données empiriques ou ils sont prédits par de algorithmes.

Pour sélectionner les peptides, nous avons vérifié leur détectabilité en cerveau avec le protocole développé pour la quantification de la C9ORF72. Pour cela, une première analyse DDA par UHPLC-MS a été réalisée sur un échantillon de cerveau, ce qui a permis l'identification de 18 protéines et la sélection de 35 peptides. Enfin, des analyses par PRM ont été effectuées ciblant 10 peptides, issus de peptide Atlas, par analyse afin d'avoir un temps d'accumulation dans la C-trap suffisant (60ms). Si uniquement un peptide pour une protéine est détecté alors un nouveau peptide est sélectionné. Si, ce nouveau peptide n'est pas détecté alors un autre peptide est sélectionnée en se basant sur son score de détectabilité. Si uniquement un peptide est identifié avec un doute sur son identification alors un seul peptide est sélectionnée pour la protéine. Enfin si aucun peptide n'est détecté alors la protéine est retirée de la liste. Pour ces analyses PRM le peptide est considéré comme détecté si au moins deux fragments (<5ppm) co-éluent. Cette méthode a permis d'éliminer 9 protéines et de commander 122 peptides marqués (PEPotec thermo Fisher) (cf. tableaux 13 et 14). Nous avons opté pour des peptides PEPotec (concentration non déterminée par AAA et non purifié) car ils sont suffisants pour le développement d'une méthode de quantification relative. De plus, les peptides marqués permettent de confirmer l'identification d'un peptide : temps de rétention et fragmentation.



Tableau 13 liste des peptides sélectionnés pour la quantification multiplexe (1)

Uniprot	Protein Name	Provenance	Peptide Sequence	Identification
P06733	Alpha-enolase	ICM	EGLELLK IGAEVYHNLK	ddMS2 ddMS2
P30086	Phosphatidylethanolamine-binding protein 1	ICM	GNDIISSTVLSDYVGSPPK NRPTSISWDGLDSGK	ddMS2 ddMS2
P04406	Glyceraldehyde-3-phosphate dehydrogenase	ICM	LISWYDNEFGYSNR VGVNGFGR	ddMS2 ddMS2
P30041	Peroxiredoxin 6	ICM	LSILYPATTGR LPFPIIDDR	ddMS2 ddMS2
P52565	Rho GDP-dissociation inhibitor 1	ICM	AEYEFLTPVEEAPK EALLGR	ddMS2 ddMS2
P62805	Histone H4	ICM	ISGLIYEETR VFLENVIR	peptide atlas peptide atlas
O95817	BAG family molecular chaperone regulator 3	ICM	ELLALDSVDPEGR VEAILEK	peptide atlas peptide atlas
P25685	DnaJ homolog subfamily B member 1	ICM	DGSDVIYPAR TIPVVFK	peptide atlas peptide atlas
P50454	Serpin H1	ICM	LYGPSSVSFADDFVR LFYADHPFFLVR	peptide atlas peptide atlas
Q09666	Neuroblast differentiation-associated protein AHNAK	ICM	AEGPEVDVNLPK VDIETPNLEGLTGPR	peptide atlas peptide atlas
Q9UMX0	Ubiquilin-1	ICM	SHTDQLVIFAGK DQDTLSQHQHIDGLTVHLVIK	peptide atlas peptide atlas
O95232	Luc7-like protein 3	ICM	NEVNGTSEDIK STTSTIESFAAQEK	peptide atlas peptide atlas
O75175	CCR4-NOT transcription complex subunit 3	ICM	AAISSGIEDPVPTLHLTER EGFTFEYR	peptide atlas peptide atlas
Q9UQ90	Paraplegin	ICM	EFVDYLK AAEDELNIEAK	peptide atlas peptide atlas
P35268	60S ribosomal protein L22	ICM	ITVTSEVPFSK AGNLGGGVVTIER	peptide atlas peptide atlas
P62244	40S ribosomal protein S15a	ICM	ESYELR HGYIGEFEIIDDHR	peptide atlas peptide atlas
P25398	40S ribosomal protein S12	ICM	IVVNLGTGR TALIHDLGLAR	peptide atlas peptide atlas
P18124	60S ribosomal protein L7	ICM	DVIEEYFK EANNFLWPFK	peptide atlas peptide atlas
Q9UJY1	Heat shock protein beta-8	ICM	IALTDNALIAR FGVPAEGR	peptide atlas peptide atlas
P36222	Chitinase-3	ICM	LLDDGFGMDPPFDLTASWPDWALPR SFTLASSSETGVGAPISGPGIPGR	peptide atlas peptide atlas
P47972	Neuronal pentraxin-2	ICM	EAGTLAYYEICDFLR ANVSNAGLPGDFR	peptide atlas peptide atlas
Q92598	Heat shock protein 105 kDa	ICM	FDATQAFVGLSQFNIWDR DPGHVVEQLSR	peptide atlas peptide atlas
Q13501	Sequestosome-1	ICM	VLGTAFDPFLLGGK IEVPLYSLLEQTHLK	peptide atlas peptide atlas
Q96AE4	Far upstream element-binding protein 1	ICM	AGGIETIANEFSDR LTPVSPSSSTEEK	peptide atlas peptide atlas
Q15427	Splicing factor 3B subunit 4	ICM	NYDIGAALDTIQYSK IGGNEGIDVPIPR	peptide atlas peptide atlas
Q92989	Polyribonucleotide 5'-hydroxyl-kinase Clp1	ICM	IQIAPDSGGLPER IQFKPDDGTTPER	peptide atlas peptide atlas
P62266	40S ribosomal protein S23	ICM	NQDATVYVGGLEDEK VTGQHQQYGVFEFLSEEDADYAIK	peptide atlas peptide atlas
P24298	Alanine aminotransferase 1	ICM	TEVAYVSK DLPHFVR	peptide atlas peptide atlas
			VANVSLALYK AHLGTALK	peptide atlas peptide atlas
			ALELEQELR VLTLDGMNPR	peptide atlas peptide atlas

Tableau 14 liste des peptides sélectionnés pour la quantification multiplexe (2)

Uniprot	Protein Name	Provenance	Peptide Sequence	Identification
P10809	60 kDa heat shock protein, mitochondrial	ICM	NAGVEGSLIVEK VGEVIVTK	ddMS2 ddMS2
P25705	ATP synthase subunit alpha, mitochondrial	ICM	HALIYDLSK TGAIVDPVGEELLGR	ddMS2 ddMS2
P18669	Phosphoglycerate mutase 1	ICM	NLKPQPMQFLGDEETVR HLEGLSEEAIMEELNPTGPIVYELDK	ddMS2 ddMS2
Q86YJ6	Threonine synthase-like 2	ICM	FPEAVLAAGLTPETPAEIVALEHK	peptide atlas
Q8NFB5	Nucleoporin NUP53	ICM	SPLLAGGSPQPVPVPAHK	peptide atlas
B2RNN3	Complement C1q and tumor necrosis factor-related protein 9B	ICM	FPSDVPK	peptide atlas
Q9H5Z1	Probable ATP-dependent RNA helicase DHX35	ICM	FWRPGTEGPGVSISEER	peptide atlas
Q09161	Nuclear cap-binding protein subunit 1	ICM	ATNDEIFSLK	peptide atlas
P56524	Histone deacetylase 4	ICM	DQPVLLNPAR	peptide atlas
Q9ULM6	CCR4-NOT transcription complex subunit 6	ICM	SHWAELEISGR	peptide atlas
Q8N1G2	Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1	ICM	DYLFVAVNIK GDGPPVQGIINFEQK	peptide atlas ddMS2
P00441	Superoxide dismutase [Cu-Zn]	littérature	DGVADVSIEDSVISLGDHCIIGR	peptide atlas
Q13148	TDP-43	littérature	GIVHISNAEPK TSDLIVLGLPWK	peptide atlas peptide atlas
P35637	RNA-binding protein FUS	littérature	EFSGNPIK TGQPMINLYTDR	peptide atlas peptide atlas
P02649	Apolipoprotein E	littérature	SELEEQLTPEVAETR WVQTLSEQVQELLSSQVTQELR	peptide atlas ddMS2
P25311	Zn- $\beta$ -2-glycoprotein	littérature	HVEDVPAFQALGSLNDLQFFR AGEVQPELR	ddMS2 peptide atlas
P07196	Neurofilament light	littérature	FTVLTESAAK VLEAELLVLR SAYSSYAPVSSLSVR DEPPSEGEAEIEEK SNHEEEVADLLAQIQASHITVER	peptide atlas peptide atlas peptide atlas peptide atlas ddMS2
P07197	Neurofilament medium	littérature	HNHDLSSYQDTIQLENELR SPVPKSPVEEK SPVPKSPVEEAK HQADIASYQEAIQQLDAELR	peptide atlas peptide atlas ddMS2 ddMS2
P12036	Neurofilament heavy	littérature	LEQEHLLEDIAHVR EEAKSPAEEK EEAKSPAEEK	ddMS2 ddMS2 ddMS2
P51991	Heterogeneous nuclear ribonucleoprotein A3	littérature	GFAFVTFDDHDTVDK LFIGGLSFETDDSLR	peptide atlas peptide atlas
P62857	40S ribosomal protein S28	littérature	EGDVLTLLESER GDVLTLLSESR	ddMS2 peptide atlas
P30050	60S ribosomal protein L12	littérature	HSGNITFDEIVNIAR EILGTAQSVGCNVDR	peptide atlas peptide atlas
P32969	60S ribosomal protein L9	littérature	TILSNQTVDIPENVDTLK GVTLGFR	peptide atlas peptide atlas
P46777	60S ribosomal protein L5	littérature	VLEQLTGQTPVFSK NNFSDTGNFGGQIEHIDLGK	peptide atlas peptide atlas
P08247	Synaptophysin	marqueur des neurones	ETGWAAPFLR EPLGFVK	peptide atlas peptide atlas
P78352	Disks large homolog 4	marqueur des neurones	IIPGGAAAQDGR EVTHSAAVEALK LDQLTANSAR	peptide atlas peptide atlas ddMS2
P14136	Glial fibrillary acidic protein	ICM +marqueur des astrocytes	LADVYQAELR LEAENLAAYR FASYIEK	ddMS2 peptide atlas peptide atlas
P43003	Excitatory amino acid transporter 1,EAAT1	marqueur des astrocytes	VTAADAFDLIR TTTNVLGDSLGAIVEHLR	ddMS2 ddMS2
P02686	Myelin basic protein,MBP	marqueur des oligodendrocyte	HGFLPR DTGILDSIGR	ddMS2 ddMS2
O75508	OSP	marqueur des oligodendrocyte	LDELGSK FYTAGSSSPTTHAK	ddMS2 ddMS2
P11215	CD11B	marqueur de la microglie	LFTALFPFEK YVIGVGDAFR	peptide atlas peptide atlas

## **b) Optimisation de la préparation d'échantillon :**

Les données obtenues par spectrométrie de masse ont été analysées avec le logiciel librement accessible Skyline<sup>347</sup>. Les optimisations ont été faites à partir de « pools » de tous les échantillons d'homogénat de cerveau pour chaque condition d'homogénéisation.

Dans un premier temps nous avons testé l'impact de la méthode d'homogénéisation sur le nombre de protéine identifiées ainsi que sur l'aire des peptides afin de pouvoir quantifier le plus grand nombre de protéines possibles. Les méthodes de lyses que nous avons testées sont une lyse mécanique où le tissu est homogénéisé par des billes en céramiques (identique au protocole publié pour la quantification de la C9ORF72) et la lyse du tissu par une sonde à ultrason qui permet la lyse des membranes cellulaires (identique à celle présentée dans la partie sur la quantification de la C9ORF72). Le tampon de lyse est le même pour ces deux conditions. Ces deux protocoles ont été testés notamment pour s'assurer de l'extraction des protéines nucléaires. Les résultats obtenus sont présentés tableau 15. Les peptides identifiés sont identiques et le ratio d'aire entre les deux conditions est égal à 1. Les deux protocoles d'homogénéisation du tissu cérébrale sont donc équivalents dans nos conditions.

Le cerveau étant un tissu gras de 36% à 81% du poids sec selon le type de tissus (substance blanche, grise et myéline), nous avons voulu évaluer l'intérêt d'une étape de délipidation à la place de la simple précipitation au méthanol<sup>348</sup> afin de réduire les effets matrices. Pour cela nous avons adapté un protocole utilisé au sein du laboratoire. Les résultats sont présentés tableau 15. Le protocole de délipidation avec homogénéisation par ultrason n'apporte pas de gain en termes de peptides identifiés et de sensibilité en comparaison avec le protocole de précipitation des protéines avec le méthanol. Au regard de ces résultats nous avons opté pour le protocole déjà établi et validé pour la quantification de la C9ORF72 dans le cortex frontal qui consiste en une homogénéisation mécanique (Precellys) puis d'une précipitation des protéines de l'homogénat par le méthanol.

Tableau 15 Comparaison des protocoles de préparation d'échantillon

Ratio aires des peptides	Precellys/Ultrason MeOH	Ultrason Delipidation/ ultrason MeOH
Moyenne	1.05	0.98
CV	29%	24%

L'étape de SPE utilisée dans le protocole de quantification de la C9ORF72 a aussi été évaluée. Les peptides marqués ont été ajoutés avant et après la SPE et l'aire du peptide endogène a servi de référence. Le rendement de SPE a donc pu être évalué pour les peptides sélectionnés. Ce rendement varie de 9% à 90% avec une moyenne à 61%. Ce rendement moyen est satisfaisant car égal au rendement moyen des trois peptides quantifiés de la C9ORF72 obtenu après optimisation des différentes conditions SPE. Nous n'avons donc pas testé d'autres types de SPE.

Ces étapes ont permis de déterminer le niveau endogène des peptides et d'ajuster la quantité de peptides marqués pour être au plus proche possible d'un ratio de 1. Les peptides marqués ont été divisés en trois niveaux : haut, moyen (dilution 10X : solution mère) et bas (dilution 100X : solution mère). Les peptides marqués étant des peptides PEPotec nous ne pouvons pas donner leur concentration.

### **c) Evaluation de la méthode et comparaison UHPLC/ chromatographie capillaire**

Afin d'évaluer la méthode de quantification multiplexe pour l'UHPLC et la capillaire une gamme de calibration inverse a été réalisée sur 2 log<sub>10</sub> (voir figure 69) dans un homogénat de cortex frontal représentant l'ensemble des échantillons. La fidélité de la méthode a été évalué par un niveau de QC en quintuplé. Enfin, des QC dilués 5 fois dans un homogénat de cerveau de souris ont été analysés dans le but d'évaluer la performance de la méthode envers des variations de concentration d'une protéine. Ce facteur de dilution est un cas extrême de ce qui peut être attendu en termes de variation d'une protéine entre les différents groupes. De plus, ce QC est intéressant car les protéines sont diluées et non les peptides. Les résultats sont présentés tableaux 16, 17 et 18. La matrice choisit pour diluer les QC n'est pas optimale en effet uniquement 25 peptides détectés ont une séquence différente entre l'humain et la souris, la détermination de la

fidélité des QC dilués n'est donc pas possible. Nous aurions pu diluer l'homogénat par une solution de BSA à la concentration du pool d'homogénat mais la BSA ne mimera pas parfaitement la complexité de la matrice ou comme montré par Chen et al.<sup>349</sup> l'utilisation de plasma dilué comme mime des tissus.

La fidélité et les biais pour les QCs sont équivalents pour les deux conditions testées avec 9% et 13% pour l'UHPLC et 9% et 14% pour la chromatographie capillaire (moyennes obtenues à partir des résultats des tableaux 16, 17 et 18). Nous observons aussi un léger gain de sensibilité par la chromatographie capillaire en comparaison avec l'UHPLC avec les trois points de gamme les plus bas représentant 77% contre 65% des LOQ. Ce gain est décevant, en effet le gain théorique est de  $(du/dc)^2$  soit 49, nous pouvions nous attendre à mieux. La complexité de la matrice peut être responsable, une fenêtre isolation du quadrupole plus stricte des ions 0,7 Da plutôt que 2 Da aurait permis de gagner en sensibilité en accumulant plus spécifiquement les ions dans la C-trap. Une méthode de trapping des peptides en amont de la colonne chromatographique aurait aussi pu être installée afin de réduire les effets matrices. La différence entre les deux méthodes est le temps d'accumulation maximale des ions dans la C-trap, 60 ms pour l'UHPLC et de 200 à 250 ms pour la chromatographie capillaire. Si la matrice est trop complexe alors la C-trap est rapidement rempli <60ms et il n'y a pas de gain apporté par le temps maximal d'accumulation des ions.

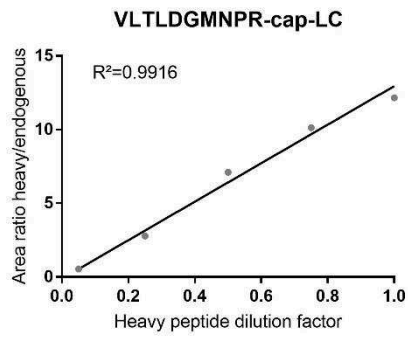
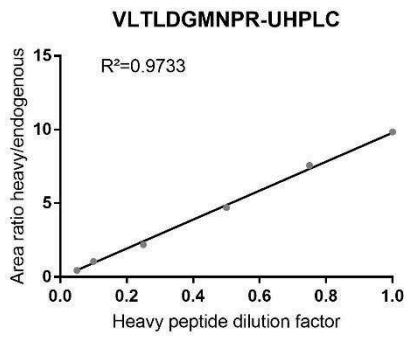
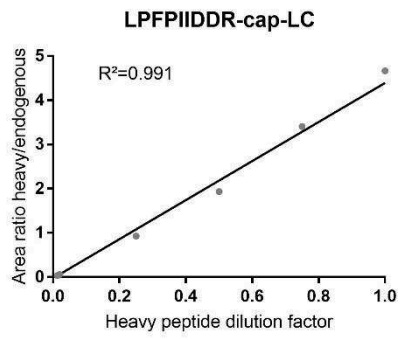
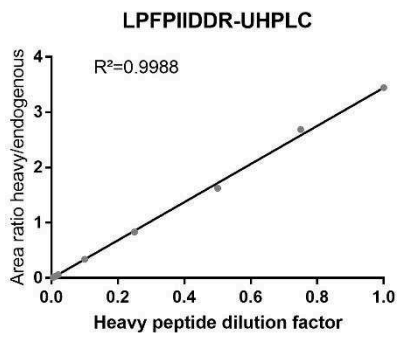


Figure 68 Exemples de gamme de calibration pour deux peptides LPFPIIDDR et VLTLGDMNPR

Tableau 16 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (1)

Protéines	Peptide Sequence	Coefficient de corrélation R <sup>2</sup>	UHPLC				Capillaire						
			LOQ	Fidélité	biais	QC-dil	LOQ	Fidélité	biais	QC-dil			
Alpha-enolase	EGLELLK	0,9877	G9	13%	5%	8%	14%			6%		6%	31%
	IGAENVYHNLK	0,9929	G9		2%			0,9889	G10	12%	6%		
Phosphatidylethanolamine-binding protein 1	GNDISSGTVLSDYVVGSGPPK	0,9858	G7	10%	2%			0,9893	G9	19%	9%	19%	30%
	NRPTSISWDGLDSGK	0,9901	G10	8%	9%	13%	16%	0,9893	G9	8%	1%	8%	30%
Glyceraldehyde-3-phosphate dehydrogenase	LISWYDNEFGYSNR	0,9965	G10	2%	13%	9%	17%	0,9869	G9	4%	6%	4%	20%
	VGVNGFGR	0,9926	G5	8%	29%			0,9861	G6	11%	6%		
Peroxiredoxin 6	LSILYPATTGR							0,9934	G9	5%	2%	19%	24%
	LPFPIIDDR	0,9988	G10	7%	14%	21%	5%	0,991	G9	19%	34%		
Rho GDP-dissociation inhibitor 1	AEEYFLTPVEEAPK	0,9969	G10	3%	11%	7%	9%	0,9912	G9	6%	5%	6%	21%
	EALLGR	0,9988	G10	5%	8%			0,9952	G10	7%	0%		
Histone H4	ISGLIYEETR	0,9973	G10	5%	3%			0,9926	G9	4%	16%		
	VFLENVIR	0,9974	G10	5%	11%			0,9956	G9	5%	0%		
BAG family molecular chaperone regulator 3	ELLALDSVDPEGR	0,9959	G10	5%	8%			0,9968	G9	8%	10%		
	VEAILEK	0,9952	G7	3%	19%			0,9738	G8	7%	15%		
DnaJ homolog subfamily B member 1	DGSDVIYPAR	0,9983	G9	5%	6%			0,9948	G7	12%	7%		
	TIPVVFK	0,9977	G10	2%	6%			0,9979	G10	3%	19%		
Serpine H1	LYGPSSVSFADDFVR	0,9809	G5	5%	27%			0,9967	G6	20%	12%		
	LFYADHPFIFLVR							0,9893	G4	28%	10%		
Neuroblast differentiation-associated protein AHNAK	AEGPEVDVNLPK	0,9856	G9	12%	18%			0,9842	G9	10%	8%		
	VDIETPNLEGLTGPR	0,9954	G7	26%	13%			0,9781	G6	16%	20%		
Ubiquilin-1	SHTDQLVLIIFAGK						0,9957	G7	11%	5%			
Luc7-like protein 3	STTSTIESFAAQEK						0,9876	G9	8%	45%			
CCR4-NOT transcription complex subunit 3	AAISSGIEDPVPTLHLTER	0,9602	G5	13%	6%			0,9951	G7	24%	13%		
	EFVDYLK	0,9842	G5	14%	8%			0,9772	G6	21%	5%		
Paraplegin	AAEDELNIEAK	0,9932	G5	13%	11%			0,9546					
	ITVTSEVPFSK	0,9904	G9	5%	10%			0,9941	G10	13%	14%		
60S ribosomal protein L22	AGNLGGGVVTIER	0,9984	G9	4%	8%			0,9932	G10	4%	9%		
	ESYELR			3%				0,9962	G9	6%	1%		

Tableau 17 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (2)

Protéines	Peptide Sequence	Coefficient de corrélation R <sup>2</sup>	UHPLC				Capillaire						
			LOQ	QC		QC-dil		LOQ	QC		QC-dil		
				Fidélité	biais	Fidélité	biais		Fidélité	biais	Fidélité	biais	
40S ribosomal protein S15a	HGYIGEFEIIDDHR	0,9947	G7	4%	2%			0,9925	G10	4%	7%		
	IVVNLTR							0,9841	G8	5%	22%		
40S ribosomal protein S12	TALIHDLGLAR	0,9965	G10	5%	11%			0,9935	G10	10%	2%		
	DVIEEYFK	0,9927	G9	8%	3%			0,9916	G9	9%	6%		
60S ribosomal protein L7	IALTDNALIAR	0,9935	G9	5%	9%	10%	17%	0,9987	G9	3%	8%	12%	19%
Heat shock protein beta-8 Chitinase-3	FGVPAEGR	0,998	G9	6%	7%			0,9931	G10		8%		
	SFTLASSSETGVGAPISGPGIPGR							0,9937	G8	7%	3%		
Heat shock protein 105 kDa	VLGTAFDPFLGGK	0,9973	G9	5%	11%			0,9966	G9	5%	10%		
	IEVPLYSLEQTHLK	0,916	/	12%	49%	34%	22%	0,9717	x	11%	46%	11%	25%
	AGGIETIANEFSDR							0,994	G9	3%	4%		
Sequestosome-1	LTPVSPSSSTEK	0,9614	G5	32%	25%	23%	47%	0,9894	G6	12%	1%	12%	38%
	NYDIGAALDTIQYSK	0,9827	G5	15%	37%			0,9923	G7	10%	3%		
Far upstream element-binding protein 1 Splicing factor 3B subunit 4	IGGNEGIDVPIPR	0,9997	G6	15%	43%			0,9928	G8	4%	11%		
	IQIAPDSGGLPER	0,9889	G8	13%	35%			0,9482	G8	5%	64%		
40S ribosomal protein S23 Alanine aminotransferase 1	NQDATVYVGGGLDEK							0,9919	G9	13%	9%		
	AHLGTALK	0,9921	G7	6%	11%			0,9944	G8	5%	1%		
60 kDa heat shock protein, mitochondrial	VLTLDGMNPR	0,9972	G6	12%	20%			0,9916	G6	10%	11%		
	NAGVEGSLIVEK	0,9967	G9	4%	11%			0,985	G9	5%	8%		
	VGEVIVTK	0,9802	G9	8%	39%			0,9864	G9	7%	18%		
ATP synthase subunit alpha, mitochondrial	HALIYDDLK	0,997	G9	3%	10%			0,99338	G9	5%	10%		
	TGAIVDVVPVGEELLGR	0,9978	G9	3%	8%			0,9975	G9	5%	2%		
Phosphoglycerate mutase 1 Nucleoporin NUP53	NLKPQKPMQFLGDEETVR							0,9858	G9	7%	13%		
	SPLLAGGSPQPVPVPAHK							0,9814	G8	16%	28%		
Nuclear cap-binding protein subunit 1	ATNDEIFSILK	0,9749	G5	30%	20%			0,9899	G6	13%	8%		
	GDPVQGIINFEQK	0,9833	G7	13%	2%	56%	12%	0,9865	G5	11%	29%	11%	13%
Superoxide dismutase [Cu-Zn]	DGVADVSIEDSVISLSDHCCIIGR							0,9973	G8	20%	55%		
	GISVHISNAEPK	0,9985	G6	4%	15%			0,9981	G6	7%	13%		
TDP-43	TSDLIVLGLPWK	0,9933	G5	10%	18%								
	EFGNPIK	0,9732	G5	8%	9%			0,9995	G8	6%	1%		
RNA-binding protein FUS	TGQPMINLYTDR	0,9727	G5	17%	5%								



Tableau 18 Evaluation de l'UHPLC et du capillaire en termes de linéarité et de LOQ. La fidélité et le biais des QC (n=5) ont aussi été évalués. La fidélité correspond au coefficient de variation des QCs et le biais correspond à l'écart par rapport à la valeur théorique attendue (3)

Protéines	Peptide Sequence	Coefficient de corrélation R <sup>2</sup>	UHPLC					Capillaire					
			LOQ	QC		QC-dil		Coefficient de corrélation R <sup>2</sup>	LOQ	QC		QC-dil	
				Fidélité	biais	Fidélité	biais			Fidélité	biais	Fidélité	biais
Apolipoprotein E	SELEEQLTPVAEETR	0,9716	G6	15%	10%	36%	24%	0,9804	G6	17%	12%	17%	6%
	HVEDVPAFQALGSLNDLQFFR	0,9754	G5	23%	7%	22%	69%						
Zn-α-2-glycoprotein	AGEVQPELRL	0,9566	G5	16%	10%	29%	37%	99%	G5	18%	0%		
	FTVLTESAAK	0,9954	G9	4%	6%			0,9891	G9	2%	8%		
Neurofilament light	VLEAELLVLR	0,993	G9	3%	5%			0,9937	G9	9%	13%		
	SNHEEEVADLLAQIQASHITVER							0,9765	G9	7%	20%		
	HNHDLSSYQDTIQLENELR							0,9925	G6	6%	8%		
	SPVPKSPVEEK	0,9982	G9	6%	26%	16%	2%	0,9525	G9	10%	39%	10%	7%
Neurofilament medium	SPVPKSPVEEAK	0,9942	G10	11%	12%	13%	9%	0,9973	G10	8%	10%	8%	3%
	HQADIASYQEAIQQLDAELR	0,9836	G6	65%	33%	38%	31%	0,9902					
Neurofilament heavy	LEQEHLLEDIAHVR	0,9945	G10	4%	11%			0,9902	G9	8%	8%		
	GFAFVTFDDHDTVSK	0,9621	G8	9%	38%			0,9735	G8	16%	47%		
Heterogeneous nuclear ribonucleoprotein A3	LFIGGLSFETDDSLR							0,9659	G8	8%	33%	34%	
40S ribosomal protein S28	EGDVLTLLESER	0,9948	G9	3%	0%			0,9984	G9	10%	1%		
60S ribosomal protein L12	EILGTAQSVGCNVDGR	0,979	G5	19%	3%			0,9906	G6	14%	2%		
60S ribosomal protein L9	TILSNQTVDIPENVDTLK	0,9977	G8	4%	1%	13%	25%	0,9947	G8	5%	4%	5%	7%
60S ribosomal protein L5	VLEQLTGQTPVFSK	0,9939	G9	5%	8%			0,9958	G6	13%	9%		
	ETGWAAPFLR	0,9913	G8	8%	1%	11%	2%	99%	G10	5%	1%	5%	4%
Synaptophysin	EPLGFVK	0,9851	G10	9%	13%			0,9774	G9	10%	19%		
	IIPGGAAAQDGR	0,9982	G10	3%	7%			0,9959	G10	5%	7%		
Disks large homolog 4	EVTHSAAVEALK	0,9806	G9	4%	16%			0,974	G8	5%	13%		
	LDQLTANSAR	0,9997	G10	5%	21%	15%	20%	0,9962	G9	11%	29%	11%	22%
	LADVYQAEELR							0,9921	G9	3%	9%	3%	24%
Glial fibrillary acidic protein	LEAENLAAYR	0,969	G8	11%	42%	27%	2%	0,9988	G9	13%	34%	13%	9%
	FASYIEK	0,9916	G10	4%	7%	8%	3%	0,9873	G9	12%	0%	12%	23%
	VTAADAFDLIR	0,9977	G10	4%	8%	10%	17%	0,9939	G9	6%	8%	6%	24%
Excitatory amino acid transporter 1, EAAT1	TTNVLGDSLGAIVEHLSR	0,9975	G10	5%	13%			0,9933	G9	14%	9%		
	HGFLPR	0,9967	G10	3%	9%			0,9974	G10	5%	1%		
Myelin basic protein, MBP	DTGILDSIGR	0,9953	G9	7%	14%			0,988	G9	6%	30%		
	LDELGSK	0,9912	G9	4%	2%	15%	6%	0,9877	G9	5%	4%	5%	22%
OSP	FYYTAGSSSPTHAK	0,9902	G9	6%	2%	13%	23%	0,9849	G9	1%	4%	1%	30%
CD11B	YVIGVGDAFR							0,9702	G7	22%	78%		

#### **d) Résultat de quantification relative des échantillons de cortex frontal de la cohorte**

L'UHPLC a permis de quantifier 72 peptides appartenant à 45 protéines. La chromatographie capillaire a permis de quantifier 82 peptides appartenant à 49 protéines. La chromatographie capillaire n'a permis un gain de seulement 4 protéines, sachant que le temps d'analyse est triplé. De plus, les temps de rétention sont plus variables avec un débit capillaire (voir annexe) ce qui peut amener à devoir réinjecter des échantillons. Notre stratégie gagnerait à être adaptée avec les méthodes Scout-SRM<sup>288</sup> ou IS-PRM<sup>289</sup> afin d'adapter au mieux les fenêtres d'isolement et de ne plus perdre de peptides raison de la variabilité du temps de rétention. Notre stratégie de sélection de peptides nous a permis de détecter 67% des peptides ciblées ce qui est un meilleur rendement que la méthode proposée par Kim et al<sup>350</sup> où le rendement n'était seulement que de 22% et donc permet de diminuer le coût de la méthode par peptide.

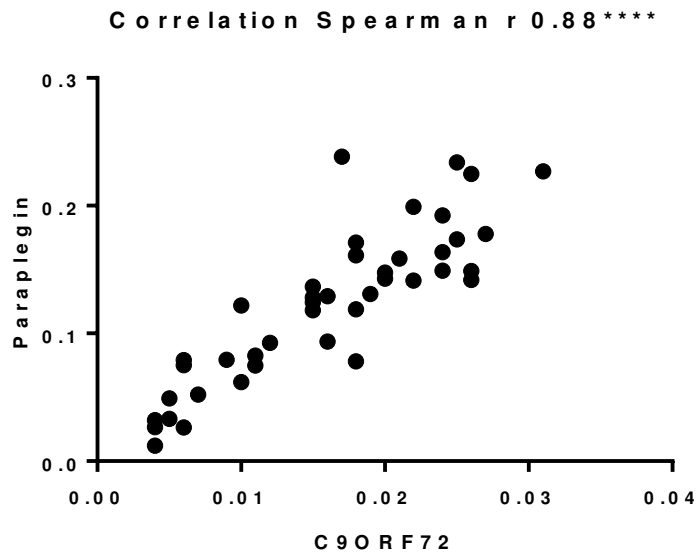
Au final, en comparant les données de quantification des trois groupes C9ORF72, DFT sporadique et témoins, 19 protéines sont significativement impactées dans l'un des groupes testés (voir tableau 19). Pour les analyses statistiques les médianes ont été comparées par le test Mann-Whitney. Les différentes protéines sont impliquées dans plusieurs processus biologiques comme l'inflammation ou la protection contre les espèces oxydantes par exemple. Mis à part quelque exception comme la sequestosome 1 et la Disks large homolog 4 nous observons une tendance identique entre les différents peptides d'une même protéine, ce qui renforce les résultats obtenus.

Tableau 19 Liste des protéines impactées avec la médiane des ratios des aires des peptides pour chaque groupe. Le test statistique utilisé est le test de Mann-Withney, les médianes sont considérées comme statistiquement différentes si  $p < 0.05$ .

Protéines	Peptide Sequence	Mediane			p value			ratio		
		C9	sporadic-TDP43	témoins	sporadic vs. C9	témoins vs. C9	témoins vs. Sporadique	sporadic vs. C9	témoins vs. C9	témoins vs. Sporadique
Peroxiredoxin 6	LSILYPATTGR	1,17E-01	6,51E-02	7,22E-02	0,0006	0,002	0,5387	0,55	0,61	1,11
	LPFFIIDDR	1,67E-01	1,02E-01	1,08E-01	0,013	0,0006	0,7844	0,61	0,65	1,05
Histone H4	ISGLIYEETR	1,79E-01	7,96E-02	2,18E-01	0,1552	0,8417	0,0575	0,45	1,22	2,74
	VFLENVIR	9,74E-02	3,70E-02	8,11E-02	0,1426	0,73	0,0666	0,38	0,83	2,19
BAG family molecular chaperone regulator 3	ELLALDSVDPEGR	4,23E-02	2,71E-02	2,93E-02	0,0009	0,0114	0,2839	0,64	0,69	1,08
Luc7-like protein 3	STTSTIESFAAQEK	8,89E-04	1,55E-03	2,20E-03	0,1426	0,0015	0,0503	1,74	2,48	1,42
CCR4-NOT transcription complex subunit 3	AAISSGIEDPVPTLHLTER	6,67E-03	1,06E-02	1,94E-02	0,0768	0,0001	0,0093	1,59	2,91	1,84
Paraplegin	EFVDYLK	7,31E-02	1,30E-01	1,24E-01	0,0822	0,0482	0,7713	1,78	1,70	0,95
	AAEDELNIEAK	8,11E-02	1,45E-01	1,49E-01	0,0244	0,0114	0,9274	1,78	1,84	1,03
Sequestosome-1	LTPVSPSSSTEEK	5,86E-02	6,13E-02	8,06E-02	0,8121	0,0269	0,0883	1,05	1,37	1,31
	NYDIGAALDTIQYSK	8,86E-03	5,55E-03	7,99E-03	0,0096	0,3658	0,0503	0,63	0,90	1,44
Far upstream element-binding protein 1	IGGNEGIDVPIPR	2,79E-02	3,29E-02	4,76E-02	0,2307	0,0002	0,1229	1,18	1,71	1,45
40S ribosomal protein S23	AHLGTALK	3,85E-02	4,14E-02	4,75E-02	0,5884	0,0081	0,0503	1,08	1,23	1,15
Alanine aminotransferase 1	VLTLDGMNPR	4,79E-02	2,72E-02	1,98E-02	0,0216	0,0002	0,101	0,57	0,41	0,73
Nucleoporin NUP53	SPLLAGSGPPQVVPAAHK	8,08E-04	7,38E-04	1,58E-03	>0,9999	0,0001	0,0009	0,91	1,96	2,14
Superoxide dismutase [Cu-Zn]	GDGPVQGIINFEQK	1,81E-01	1,34E-01	1,70E-01	0,0001	0,8989	0,0121	0,74	0,94	1,27
	DGVADVSIEDSVISLGDHCIIGR	2,97E-02	1,98E-02	3,19E-02	0,389	0,2182	0,099	0,67	1,07	1,61
TDP-43	GISVHISNAEPK	2,56E-02	1,89E-02	1,99E-02	0,0244	0,0071	0,7844	0,74	0,78	1,05
	TSDLIVLGLPWK	7,03E-02	6,47E-02	7,09E-02	0,7787	0,4437	0,4458	0,92	1,01	1,10
RNA-binding protein FUS	EFSGNPIK	2,16E-02	2,39E-02	2,76E-02	0,3265	0,0116	0,2316	1,11	1,28	1,15
	TGQPMINLYTDR	9,95E-02	1,31E-01	1,33E-01	0,267	0,053	0,4833	1,32	1,33	1,01
Neurofilament light	FTVLTESAAK	7,05E-02	5,78E-02	9,18E-02	0,5884	0,0327	0,0776	0,82	1,30	1,59
	VLEAELLVLR	1,82E-01	1,43E-01	2,17E-01	0,6496	0,08	0,1151	0,78	1,19	1,52
Neurofilament medium	SNHEEEVADLLAQIQASHITVER	6,65E-02	5,38E-02	9,06E-02	0,8193	0,0386	0,0358	0,81	1,36	1,68
	HNHDLSSYQDTIQLENELR	6,29E-01	4,77E-01	7,03E-01	0,8518	0,2741	0,1593	0,76	1,12	1,47
Neurofilament heavy	LEQEHLLEDIAHVR	9,25E-02	8,49E-02	1,36E-01	0,9828	0,0435	0,1828	0,92	1,47	1,61
Heterogeneous nuclear ribonucleoprotein A3	GFAFVTFDDHDTVDK	6,12E-02	7,56E-02	1,02E-01	0,3735	0,0022	0,0303	1,24	1,67	1,35
	LFIGGLSFETDDSLR	3,84E-02	4,90E-02	6,68E-02	0,1045	0,0006	0,01151	1,27	1,74	1,37
Synaptophysin	ETGWAAPFLR	2,50E-02	2,93E-02	3,32E-02	0,4223	0,1581	0,6049	1,17	1,33	1,13
	EPLGFVK	1,39E-02	1,58E-02	2,26E-02	0,6811	0,1275	0,2569	1,14	1,62	1,43
Disks large homolog 4	IIPGGAAAQDGR	4,02E-02	4,16E-02	4,37E-02	>0,9999	0,7445	0,8718	1,03	1,09	1,05
	EVTHSAAVEALK	1,58E-02	2,14E-02	2,52E-02	0,1307	0,0063	0,4562	1,36	1,60	1,18
Glial fibrillary acidic protein	LDQLTANSAR	5,93E-01	2,68E-01	3,44E-01	0,0045	0,008	0,208	0,45	0,58	1,28
	LADVYQAE LR	1,57E+00	7,56E-01	7,39E-01	0,0141	0,0023	0,9229	0,48	0,47	0,98
	LEAENNLAAAYR	4,29E-01	1,95E-01	2,11E-01	0,0023	0,0071	0,5224	0,45	0,49	1,08
	FASYIEK	7,12E-01	3,65E-01	2,94E-01	0,0045	0,0016	0,9274	0,51	0,41	0,80
Myelin basic protein,MBP	HGFLPR	3,22E-01	1,50E-01	4,29E-01	0,1307	0,1474	0,0358	0,47	1,33	2,85
	DTGILDSIGR	1,75E+00	8,19E-01	2,02E+00	0,1196	0,1694	0,0178	0,47	1,15	2,47
OSP	LDELGSK	9,91E-02	5,12E-02	9,73E-02	0,1829	0,8702	0,1306	0,52	0,98	1,90
	FYYTAGSSSPTHAK	7,65E-02	3,56E-02	8,72E-02	0,0906	0,8989	0,0575	0,47	1,14	2,45

- La GFAP est significativement augmentés chez les patients ayant une expansion C9 en comparaison avec les deux autres groupes. La GFAP est le composant principal des neurofilaments intermédiaires des astrocytes<sup>351</sup>. Une augmentation de cette protéine est liée à une astrogliose réactive qui est un changement de morphologie et une augmentation du nombre d'astrocytes causée par un stress ou une maladie neurodégénérative<sup>352</sup>. Cette protéine est aussi un marqueur de l'inflammation. L'astrogliose est au départ un mécanisme de protection. L'augmentation de la GFAP peut aussi être causée par un dysfonctionnement de l'autophagie et protéasome<sup>353</sup> et l'astrogliose réactive a été montrée comme capable de promouvoir la neurodégénérescence de type SLA<sup>354</sup>. Concernant les DFT avec mutation C9 l'inflammation peut être induite par l'agrégation des protéines dipeptides qui sont le produit de la traduction de l'expansion comme par exemple pour le poly-GA<sup>355</sup>. La C9ORF72 semble joué un rôle dans l'autophagie et une haploinsuffisance pourrait entrainer son dysfonctionnement et expliquer l'augmentation de la GFAP<sup>356</sup>. La GFAP possède trois isoformes principaux dont les fonctions sont encore peu connues mais il semble que les isoformes jouent des rôles différents dans la structuration des filament intermédiaires<sup>357</sup>.
- La peroxiredoxin 6 est une protéine impliquée dans la réponse à un stress oxydatif où elle y joue un rôle protecteur<sup>358</sup>. La présence de protéines poly-GR augmente le stress oxydatif dans des cellules souche pluripotente induite de neurones moteurs<sup>359</sup> et donc l'augmentation de la peroxiredoxin observé serait une réponse à ce stress. La peroxiredoxin 6 a aussi une activé phospholipase A2 qui est régulée par sa phosphorylation<sup>360</sup>. De même, nous observons une augmentation de la protéine chaperonne BAG family molecular chaperone regulator 3. Cette protéine chaperonne les protéines HSP70 et HSC70 qui protègent les cellules contre les stress et empêche l'agrégation des protéines<sup>361</sup>. La protéine HSP70 interagit aussi avec les microtubules.
- Pour l'histone H4 nous observons une tendance vers sa diminution dans les cas sporadique comprenant des inclusions TDP-43. La protéine TDP-43 en conditions non pathologique interagit avec l'ARN et est nucléaire. Or dans les DFT, cette protéine est déplacée vers le cytoplasme du fait de modifications post-traductionnelles<sup>362</sup>. Amlie-Wolf et al.<sup>362</sup>, ont observé que cette mauvaise localisation de la protéine TDP-43 avait un impact sur les transcrits régulant la chromatine. Il serait alors intéressant de chercher des modifications post-traductionnelles des histones qui régulent leurs états transcriptionnels.

- La paraplegin est une protéine qui est impliquée dans la perméabilité des pores de la membrane de la mitochondrie<sup>363</sup>. Un dysfonctionnement de la mitochondrie est observé dans les DFT et SLA<sup>364</sup>. De plus, les DPR interagissent avec la mitochondrie et induisent sont dysfonctionnement<sup>359</sup>. Il est intéressant de noter une corrélation de la paraplegin avec le niveau de la C9 (voir figure 69).



**Figure 69** Corrélation C9ORF72 avec la paraplegin

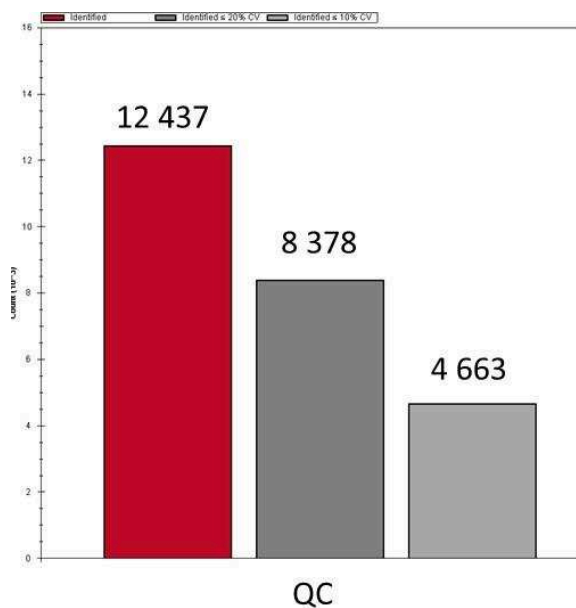
- Alanine amino transferase 1 augmentation chez les C9. Cette protéine catalyse, de façon réversible, la réaction suivante : L-alanine + 2-oxoglutarate = pyruvate + L-glutamate. Une augmentation du glutamate dans les maladies neurodégénératives est courante et associée à une excitotoxicité<sup>365</sup>. Ici, nous observons une augmentation de l’alanine amino transferase 1 uniquement chez les porteurs de l’expansion C9. Cela peut être causé par une diminution de la C9ORF72 qui a été montré comme induisant une augmentation des récepteurs aux glutamates<sup>331</sup>.
- La nucleoprine NUP 53 est significativement diminuée dans les DFTs en comparaison avec les témoins. Les nucléoporines sont une famille de protéine constituant la structure des pores nucléaires. Chou et al.<sup>366</sup>, ont montré que la protéine TDP-43, dans les DFT et ALS, induisait un changement de localisation des nucléoporines et leurs agrégations. Le transport nucléaire des protéines est alors diminué. De même, l’expansion C9 provoque un disfonctionnement de ce transport<sup>367</sup>.
- La far upstream binding protein 1 et la luc-7 like protein sont impliquées dans l’épissage alternatif de l’ARN. Ces deux protéines sont impactées dans les DFTs.

- La myelin basic protein et l'OSP sont des protéines marqueurs des oligodendrocytes. Ces deux protéines sont impactées dans le cas sporadique uniquement. Deux études ont montré un dysfonctionnement des oligodendrocytes et une diminution de la densité de la MBP pour la SLA<sup>368,369</sup>. Dans le cas des C9, ce sont les marqueurs synaptiques (disk large homolog protein 4 et Synaptophysine) qui sont diminués. Cette diminution pourrait être une conséquence de l'accumulation de glutamate dans les synapses<sup>370</sup>.
- Heterogeneous nuclear ribonucleoprotein A3 est significativement diminué chez les C9. Cette protéine est impliquée dans l'export de l'ARN. Cette protéine se lie aux expansions G<sub>4</sub>C<sub>2</sub><sup>371</sup> ainsi qu'avec les DPR<sup>372</sup> où elle agrège.
- CCR4-NOT transcription complex subunit 4, constitue un complexe CCR4-NOT qui a un rôle dans le métabolisme de l'ARN messenger<sup>373</sup>.

Les résultats obtenus sont en accord avec une analyse non ciblée par DDA et sans standards internes<sup>374</sup> où une augmentation de la GFAP et de la peroxiredoxin-6 ainsi qu'une diminution des protéines synaptiques ont été observées. De plus, nos résultats vont dans le même sens que des travaux récents, comme par exemple le rôle du glutamate dans les DFT<sup>331,370</sup> ou l'implication de l'heterogeneous nuclear ribonucleoprotein A3.

#### e) Développement d'une méthode DIA

Pour élargir encore cette approche et s'affranchir de la sélection des peptides une méthode sans a priori a été développée. En effet, la perspective principale de cette méthode multiplexe est le développement d'une méthode DIA. Nous avons opté pour le DIA pour ses meilleures performances analytiques en comparaison du DDA (voir Introduction). L'inconvénient du DIA est qu'il faut une librairie de spectres or ces bibliothèques sont peu disponibles pour les protéines contenues dans le cerveau humain notamment car c'est un tissu précieux<sup>375</sup>. Pour pallier à ce problème Bruderer et al.<sup>375</sup>, ont développé une librairie basée sur l'analyse de cerveaux de souris en ne sélectionnant que les peptides communs entre l'humain et la souris. Ces analyses sans *a priori* peuvent être réanalysées pour chercher des peptides d'isoformes non présent dans la librairie ou des modifications post-traductionnelles particulières. Nous avons choisi de faire une méthode comprenant un scan Full-MS suivi de 21 événements MS/MS. Ces 21 événements ont des fenêtres d'isolements variables afin d'avoir le même nombre de précurseur dans chaque fenêtre (en se basant sur une analyse DDA). Nous avons évalué la fidélité de la méthode par injection de 5 QCs (voir figure 70). Plus de 67% des précurseurs identifiés dans les QC ont un CV inférieur à 20%.



**Figure 70** Coefficient de variations des précurseurs identifiés dans les QC

Nous avons analysé les mêmes échantillons que pour l'analyse multiplexe. Cette stratégie nous a permis de détecter 26 soit 53% des protéines quantifiées par la méthode multiplexe et seulement 6 protéines impactées (voir tableau 20). Le DIA est dans nos conditions bien moins sensible que la méthode PRM développée, ce qui peut être expliqué par la complexité de la matrice. Néanmoins, le nombre de peptides pour les 27 protéines est supérieur à 2, ce qui est intéressant dans le contexte des protéoformes. De plus, l'analyse DIA permet la détection sans *a priori* des peptides ce qui conduit à l'identification de plus de 800 protéines et pourrait permettre d'avoir une vue globale des mécanismes impliqués dans les DFT sans être limité par une liste de protéines à cibler.

Ces résultats sont préliminaires et nécessitent une analyse de données plus approfondie.

Tableau 20 protéines détectées en DIA

Uniprot	protéines
P06733	Alpha-enolase
P30086	Phosphatidylethanolamine-binding protein 1
P04406	Glyceraldehyde-3-phosphate dehydrogenase
P30041	Peroxiredoxin 6
P52565	Rho GDP-dissociation inhibitor 1
P62805	Histone H4
Q09666	Neuroblast differentiation-associated protein AHNAK
P18124	60S ribosomal protein L7
Q92598	Heat shock protein 105 kDa
P10809	60 kDa heat shock protein, mitochondrial
P25705	ATP synthase subunit alpha, mitochondrial
P18669	Phosphoglycerate mutase 1
P00441	Superoxide dismutase [Cu-Zn]
P02649	Apolipoprotein E
P07196	Neurofilament light
P07197	Neurofilament medium
P12036	Neurofilament heavy
P51991	Heterogeneous nuclear ribonucleoprotein A3
P30050	60S ribosomal protein L12
P32969	60S ribosomal protein L9
P08247	Synaptophysin
P78352	Disks large homolog 4
P14136	Glial fibrillary acidic protein,GFAP
P43003	Excitatory amino acid transporter 1,EAAT1
P02686	Myelin basic protein,MBP
O75508	Claudin-11,OSP

### Conclusion :

Nous avons développé une méthode de quantification multiplexe de 56 protéines dans le cortex frontal. Les peptides ont été sélectionnés à partir de données acquises sans *a priori* en DDA et à partir de la base de données peptide atlas couplée à une détection sans standard par PRM. Cette stratégie de sélection nous a permis de détecter 82 peptides sur les 122 sélectionnés correspondant à 49 protéines. Cette méthode a ensuite été évaluée en termes de linéarité et de fidélité avant d'être appliquée à l'analyse d'une cohorte de patients atteints de DFT avec expansion C9ORF72 ou sporadique ayant des inclusions TDP-43 ainsi qu'un groupe témoin. L'analyse de cette cohorte a permis de mettre en évidence 19 protéines impactées. Enfin, une méthode DIA a été mise au point. Nous retrouvons 53% des protéines quantifiés par la méthode multiplexe-PRM. Les résultats obtenus sont préliminaires et nécessitent une analyse plus poussée.



## IV. Discussion et perspectives

La première cause des DFT familiale est une expansion d'un hexanucléotide ( $(G_4C_2)_n$ ) dans le gène *C9ORF72*. Trois hypothèses, non concurrentes, ont été proposées pour les mécanismes de toxicité engendrée par cette expansion : une haploinsuffisance de la *C9ORF72*, la formation de foci d'ARN et la production de protéines dipeptides. Dans ce manuscrit nous nous sommes intéressés à développer des outils analytiques pour la quantification des isoformes de la *C9ORF72*. En effet, jusqu'à présent cette quantification était réalisée uniquement par Western Blot à partir d'anticorps non commercialement disponibles, ce qui limite la reproduction des analyses dans différents laboratoires. De plus, la spécificité de ces anticorps est discutable en effet, Xiao et al., ont détecté l'isoforme C9-S dans le cortex frontal contrairement à Frick et al. Nous avons eu donc pour but de développer une méthode par spectrométrie de masse permettant la quantification de ces deux isoformes. Cette méthode a été appliquée à une cohorte comprenant des témoins, des cas sporadiques avec inclusions TDP-43 et des DFT avec expansion C9, dans le cortex frontal. Les résultats ont montré une diminution de la C9-L en comparaison des témoins et des cas sporadiques avec inclusions TDP-43. En revanche nous n'avons pas pu détecter et quantifier l'isoforme C9-S qui est sous la limite de détection de notre méthode. Nous avons alors étendu l'application de notre méthode à différents types cellulaires ainsi qu'aux PBMCs. La méthode doit être améliorée pour être appliquée au suivi du niveau de la *C9ORF72* en cellules<sup>331</sup>. En revanche elle peut être appliquée à la quantification de la *C9ORF72* dans des cerveaux d'animaux afin de suivre l'effet d'un traitement contre l'haploinsuffisance de la *C9ORF72* soit restaurant son niveau ou en augmentant l'efficacité de sa fonction<sup>331</sup>.

Nous avons pu répondre, sans ambiguïté, à la question de la diminution de la C9-L dans le cortex frontal. Néanmoins, nous n'avons pas pu conclure de manière satisfaisante pour la C9-S. Les points d'amélioration de la méthode vont être discutés.

- Dans notre méthode de quantification des isoformes par différence, les biais des peptides s'additionnent ce qui ne nous permet pas de quantifier un isoforme minoritaire<sup>309</sup> (<20%). Ce qui montre la limite de cette approche de quantification indirecte. Il faudrait alors quantifier le peptide unique de la courte. L'analyse de ce peptide, en plus de permettre la quantification directe de l'isoforme courte pourrait se révéler intéressante du fait de phosphorylations observées en cellules HEK. Ce peptide est plus long que les peptides tryptiques classiquement analysés. En effet, il est composé de 31 acides aminés (3.3 kDa). Pour son analyse une colonne de type C8, C4 ou Zorbax C3 pourrait être plus appropriée<sup>175</sup>.

Enfin, le rendement SPE pour ce peptide était le plus faible des quatre peptides, une optimisation des conditions pour ce peptide serait pertinente. Une approche par SISCAPA pourrait aussi être développée en remplacement de l'étape de SPE. En effet, cette méthode présentée en introduction (II-2) permet d'atteindre des limites de quantifications proche du ng/mL. De plus, une plus grande quantité de tissus cérébrale pourrait être analysée sans augmenter les effets matrices.

- La quantification par top-down est-elle envisageable ? Premièrement, la taille des isoformes de la C9ORF72 est de 25 kDa et de 54 kDa. La C9-S rentre dans la gamme de masse des protéines facilement analysable par top-down (Introduction II-3) alors que la C9-L est au-dessus (>30 kDa). Les conditions d'ionisations et d'analyse par spectrométrie de masse devront donc être particulièrement optimisées pour l'analyse de la C9-L. De même que pour les conditions chromatographiques avec l'utilisation de colonnes monolithiques, C4 ou Zorbax C3. Une fois ces conditions optimisées une attention particulière devra être portée à l'extraction de la C9ORF72 de la matrice biologique. En effet, ce n'est pas une protéine majoritaire du cerveau comme montré par nos analyses avec des concentrations de l'ordre de 100 ng/mL après lyse, l'extraction doit donc être sélective pour les analyses par top-down. La méthode GelFrEe développée par le groupe de Kelleher<sup>205</sup> pourrait être envisagée mais cette méthode semble plutôt être adaptée pour des analyses protéomiques. Le développement d'une immunoprécipitation, ce que nous n'avons pas réussi, permettrait l'extraction sélective de la C9ORF72. Pour cela, l'utilisation d'un nouvel anticorps est requise comme par exemple celui récemment publié par Frick et al<sup>330</sup>. L'immunoprécipitation est moins gênée par les effets matrices donc la quantité de tissus pourrait être augmentée afin de gagner en sensibilité. De plus, une méthode top-down suffisamment sensible permettrait la quantification directe des deux isoformes ainsi que la recherche d'autres modifications post-traductionnelles encore non décrites.

La méthode que nous avons développée va être appliquée à la quantification de la C9ORF72 à d'autres régions du cerveau qui ne sont pas impactées par la DFT comme montré, par les résultats préliminaires obtenus, avec le lobe occipital. Cela permettra de faire une cartographie du niveau des isoformes de la C9ORF72 et de chercher la présence l'isoforme C9-S dans différentes régions cérébrales. Il serait aussi intéressant d'inclure les protéines quantifiées avec la méthode multiplexe afin de voir si les modifications du niveau de ces protéines sont dépendantes de la région cérébrale.

La C9ORF72 n'a pas pour but d'être un biomarqueur diagnostique car le diagnostic est déjà réalisé par quantification du nombre d'expansion. Afin de mieux comprendre la physiopathologie des DFT, de trouver des biomarqueurs potentiels d'évolution ou des cibles thérapeutiques nous avons développé une méthode de quantification multiplexe par bottom-up de 49 protéines. Nous avons trouvé 19 protéines impactées dans les DFT en comparaison avec des témoins. Une méthode ciblée de ces 19 protéines pourrait être développée pour leur quantification dans le LCR ou le plasma. La sensibilité de la méthode pourrait être améliorée en utilisant une chaîne nano-LC avec une préparation d'échantillon plus sélective. La digestion des protéines pourrait aussi être optimisée afin de gagner en sensibilité et fidélité.

Le gène *C9ORF72* est particulièrement intéressant puisqu'il fait le lien entre les DFT et les SLA (DFT-> ALS et DFT+ALS)<sup>332</sup>. Dans le cas des DFT le cortex frontal est touché alors que pour les SLA ce sont les neurones moteurs<sup>332</sup>. Il serait alors intéressant d'identifier des protéoformes de la C9ORF72 qui pourraient être différents entre les deux maladies et représentatif de ces maladies.

Les protéines dipeptides produites par la traduction de l'expansion sont des biomarqueurs potentiels<sup>337,376</sup>. Leurs analyses par spectrométrie de masse semblent être un challenge pour plusieurs raisons. Premièrement, l'hétérogénéité de la taille de l'expansion qui est de plusieurs dizaine à plusieurs milliers de répétition. Cette hétérogénéité va être traduite au niveau de la protéine et rendre son analyse par des méthodes ciblées de spectrométrie de masse difficile. De plus, au moins 6 protéines sont produites avec des propriétés physico-chimique différents en effet certaines formes sont très hydrophobes (GA)<sub>n</sub> alors que les formes (PR)<sub>n</sub> et (GR)<sub>n</sub> vont être plutôt hydrophiles. Les (PR)<sub>n</sub> et (GR)<sub>n</sub> ont une séquence propice à l'ionisation par electrospray. Ces deux protéines vont être extrêmement chargées positivement ce qui pourrait être un avantage pour la préparation d'échantillon, par exemple SPE échange de cations. Bien que l'analyse par spectrométrie de masse ne semble pas optimale il y a un réel intérêt à développer des méthodes de caractérisation de ces protéines.

#### **Perspectives plus globales :**

La TDP-43 est une protéine qui forment des inclusions neuronales dans les maladies neurodégénératives, les DFT en particulier, au même titre que la protéine tau et l'alpha-synucléine. La TDP-43 n'est pas une protéine majoritaire du cerveau comme l'alpha-synucléine et une méthode par top-down par simple précipitation des protéines ne semble pas être adaptée. Plusieurs modifications, comme des troncations ou phosphorylations, sont connues comme augmentant la capacité de la TDP-43 à former des inclusions, notamment en induisant un

changement de sa localisation<sup>377</sup>. De plus, la TDP-43 a des propriétés de prion<sup>378,379</sup> comme l'alpha-synucléine. Du fait de la similarité entre ces deux protéines le développement d'une méthode similaire à tau ou l'alpha-synucléine (cf. chapitre 2) serait pertinente<sup>380</sup>.

# Chapitre 2 : Développement méthodologique, bottom-up et top-down, pour la quantification des protéoformes de l'alpha-synucléine

## I. Introduction

Les synucléinopathies constituent un groupe de maladies neurodégénératives comprenant la maladie de Parkinson, la démence à corps de Lewy et l'atrophie multisystématisée. Ces maladies sont caractérisées par l'accumulation anormale d'alpha-synucléine. Dans le cas de la maladie de Parkinson et des démences à corps de Lewy, l'alpha-synucléine forme des inclusions neuronales appelées corps de Lewy. Alors que dans l'atrophie multisystématisée, les inclusions sont présentes dans le cytoplasme des oligodendrocytes. Des mutations dans le gène SNCA sont responsables des synucléinopathies familiales et une même mutation peut induire la maladie de Parkinson ou une atrophie multisystématisée<sup>381</sup>.

La famille des synucléines est constituée de trois protéines : l'alpha, la beta et la gamma-synucléine. L'alpha-synucléine est homologue à 59% à la beta-synucléine et à 50% avec la gamma-synucléine. L'alpha-synucléine joue un rôle central dans le développement des synucléinopathies.

L'alpha-synucléine est une petite protéine de 14.5 kDa qui est présente sous plusieurs formes en condition physiologique. En effet, elle existe sous forme d'un monomère intrinsèquement désordonnée<sup>382</sup>, sous forme dimérique ou tétramérique et sa partie N-terminale est sous forme d'une hélice alpha liée aux membranes lipidiques<sup>383,384</sup>. Elle est principalement exprimée dans les tissus neuronaux et est présente aux terminaux présynaptiques. La fonction exacte de l'alpha-synucléine n'est pas connue mais il semble qu'elle soit impliquée dans le trafic vésiculaire, la différenciation neuronale, la plasticité synaptique et le métabolisme de la dopamine<sup>385</sup>. Les modèles de souris K.O. soit simple K.O. alpha-synucléine<sup>386</sup>, double alpha-beta-synucléine<sup>387</sup>, ou triple<sup>388</sup> montrent une diminution du niveau de dopamine ainsi que des changements dans la composition des protéines synaptique et de la structure des axones. Une diminution de la taille des vésicules synaptiques de l'ordre de 30% a aussi été observée. Les souris sans l'une ou toutes les synucléines sont viables ce qui suggère une redondance des fonctions des synucléines et qu'elles sont plutôt responsables de la viabilité à long terme du système nerveux. En effet, les souris ne développent pas les phénotypes d'une maladie neurodégénérative contrairement aux modèles sur-exprimant l'alpha-synucléine.

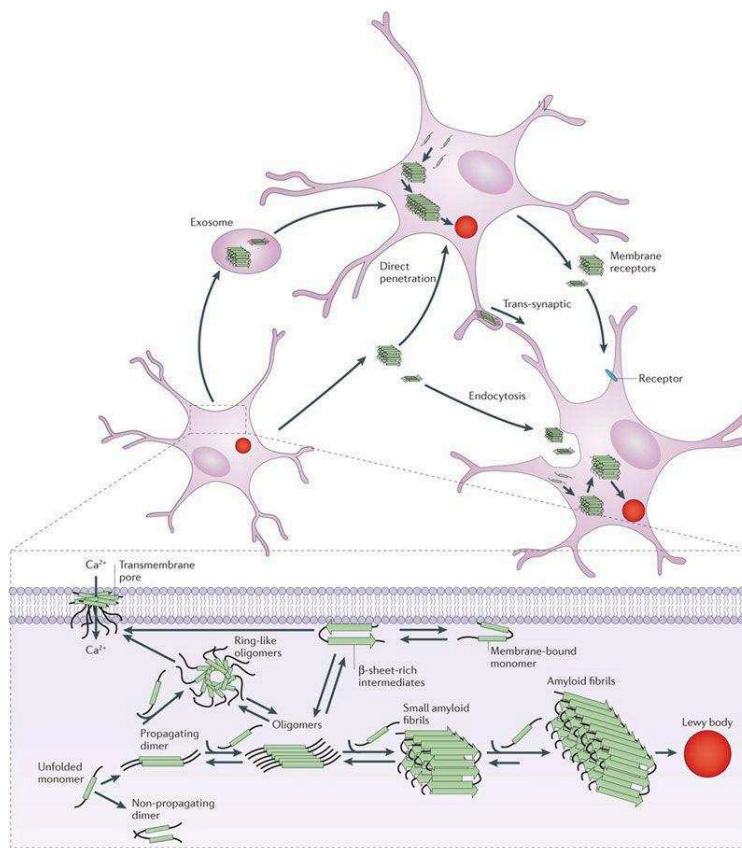
L' $\alpha$ -synucléine est constituée de trois parties :

- 1-60 : Une partie N-terminale acétylée comportant un motif répétitif de 6 acides aminés KTKEGV présent dans  $\alpha\beta\gamma$ -synucléine, pouvant se lier aux membranes lipidiques. Ce motif serait particulièrement important pour la formation des tétramères<sup>389</sup>.
- 61-95 : une région centrale qui est hydrophobe, appelé le non-amyloïde component (NAC)
- 96 -140 : une partie C-terminale comprenant de nombreux acides aminés comportant une charge négative

L' $\alpha$ -synucléine possède 4 isoformes principaux issus d'épissage alternatif de l'exon 3 et 5 :  $\alpha$ 140 alpha-synucléine intact,  $\alpha$ 112 (exon 5 : il manque les AA 103-130),  $\alpha$ 126 (exon3 : il manque les AA 41-54) et  $\alpha$ 98 (exon 5+ exon3 : il manque les AA 41-54 et 103-130). Ces formes conservent la partie hydrophobe centrale et ont été principalement étudiées au niveau de l'ARN messager du faite d'une faible concentration protéique<sup>314</sup>. Les isoformes issus de l'épissage alternatif sont à des niveaux plus faibles que la forme 1-140 dans le cerveau<sup>390</sup>. L'épissage alternatif de l'alpha-synucléine semble être impliqué dans le développement des synucléinopathies. En effet, de nombreuses études ont montré une diminution ou une augmentation des ARN messagers issus de l'épissage alternatifs dans le cas des synucléinopathies et de la maladie d'Alzheimer. Ces différences sont dépendantes de l'aire cérébrale étudié<sup>391-396</sup>. Récemment, ces isoformes ont été étudiées pour l'atrophie multisystématisée<sup>396</sup>. Les ARN messagers des isoformes  $\alpha$ 140 et  $\alpha$ 112 sont surexprimés dans les ganglions de la base et le cervelet des patients MSA en comparaison avec les contrôles et les parkinsoniens. En revanche, l'isoforme  $\alpha$ 126 est diminuée chez les MSA et parkinson en comparaison avec les contrôles. Au niveau protéique seul les isoformes  $\alpha$ 140,  $\alpha$ 112 sont observées par Western blot. L' $\alpha$ 140 est plus élevée dans toutes les régions du cerveau étudié en comparaison des contrôles et est plus élevée que chez les parkinsoniens dans le cortex préfrontal et le cortex cérébral. Concernant l'isoforme  $\alpha$ 112, elle semble aussi être augmentée en comparaison avec les parkinsoniens et les contrôles dans plusieurs aires cérébrales. Mais il n'y a pas d'anticorps spécifique contre les isoformes  $\alpha$ 112,  $\alpha$ 126 et  $\alpha$ 98 juste des exons, donc on ne peut pas affirmer que ce qui est observé ne soit pas une troncation. Des preuves à l'échelle de la protéine manque toujours pour l'étude des isoformes de l'alpha-synucléine, notamment par spectrométrie de masse qui permettrait de discriminer sans ambiguïté un isoforme d'une forme tronquée. Un nouvel isoforme de 41AA issu de l'épissage alternatif de l'exon 3 et 4 avec un codon stop prématuré a été observée<sup>397</sup>. La bêta-synucléine a aussi des isoformes mais ils ne sont que peu étudiés. Les transcrits ARN pourraient être utilisés comme biomarqueur notamment celui

produisant l'isoforme  $\alpha 12$  qui est significativement diminué dans le sang des Parkinsoniens en comparaison avec des contrôles<sup>398</sup>.

Les corps de Lewy seraient formés par un changement d'équilibre entre une forme soluble de l' $\alpha$ -synucléine et une forme oligomérique insoluble<sup>399</sup> (voir figure 71). Les oligomères forme ensuite des protofibrilles qui poursuivent leurs agrégations pour former des fibrilles insolubles et finalement des corps de Lewy. Il semblerait que les oligomères soient toxiques et non les corps de Lewy<sup>400,401</sup>. Les causes de la neurotoxicité causé par l'agrégation sont encore débattues entre : un mécanisme de gain de fonction et une perte de fonction de  $\alpha$ -synucléine ou une somme des deux mécanismes<sup>402</sup>.



**Figure 71 Mécanisme de formation des corps de Lewy et propagation de cellule à cellule. D'après Lashuel al. 2013**

Le mécanisme de gain de fonction est soutenu par une inhibition de l'autophagie par les lysosomes ne permettant plus la dégradation des protéines toxiques<sup>402</sup>. D'autre part Les modèles animaux surexprimant l' $\alpha$ -synucléine développent des troubles neurologiques contrairement aux animaux K.O.  $\alpha$ -synucléine. Quant à la perte de fonction, elle serait causée par une perte de fonctionnalité

de la synucléine soit liée à une conformation non fonctionnelle ou une à mauvaise localisation : migration des terminaux présynaptique vers le péricaryon<sup>403</sup>. Mais sa fonction est globalement mal connue<sup>385</sup>.

Plusieurs modifications de l' $\alpha$ -synucléine semblent favoriser la formation d'oligomères et son agrégation et donc son passage sous une forme pathologique. Les principales modifications post-traductionnelles de l'alpha-synucléine sont résumées figure 72.

- Les mutations : comme présenté dans l'introduction, des formes familiales de la maladie Parkinson sont causées par des mutations d'un acide aminé dans la partie N-terminale de l'alpha-synucléine.
- Phosphorylations : La phosphorylation en S129 est à un niveau faible chez une personne saine (<4%) alors que dans les corps de Lewy le taux de phosphorylation de l'alpha-synucléine est supérieure à >90%<sup>404,405</sup>. Bien qu'elle soit majoritairement trouvée dans les corps de Lewy, il n'est pas clair si la phosphorylation augmente ou diminue la capacité d'agrégation de l' $\alpha$ -synucléine. En effet, Zhou et al<sup>406</sup> ont montré que la phosphorylation serait un événement tardif et apparaîtrait après la formation de fibrilles. Néanmoins, Walker et al., ont montré une corrélation entre le niveau de phosphorylation en S129 et la sévérité des symptômes. De plus, Le niveau de phosphorylation semble permettre de distinguer les types de synucléinopathies entre elles. En effet, dans le LCR, Foulds et al.<sup>407</sup>, ont montré une augmentation de cette phosphorylation MSA en comparaison de la maladie de Parkinson et des démences à corps de Lewy. Cette étude a été étendue au plasma où une augmentation de la phosphorylation a été observée chez les personnes atteintes de la maladie de Parkinson en comparaison avec un groupe contrôle<sup>408</sup>. A noter que d'autres sites de phosphorylation de l'alpha-synucléine sont présents S87 et Y125<sup>409</sup>.
- Troncations : Des formes tronquées de l'alpha-synucléine sont présentes dans les corps de Lewy<sup>410</sup>. Le clivage de l'alpha-synucléine dans le cerveau est un événement normal et ces troncations ont lieu dans les parties N-terminale et C-terminale de l'alpha-synucléine<sup>313</sup>. Plusieurs enzymes sont responsables du clivage de l'alpha-synucléine comme par exemple le protéasome, la cathepsin D et L, la calpaïne 1, la caspase 1 et la MMP-3<sup>314</sup>. Les formes tronquées en C-terminale semblent être les formes les plus toxiques. D'une part, cette partie C-terminale contient 14 acides aminés acides lui conférant une charge négative qui est en partie responsable de la non structuration de l'alpha-synucléine. Ainsi, les formes tronquées en C-terminale forment plus rapidement des fibrilles en comparaison des formes



N-terminales<sup>411</sup>. Par exemple, il a été montré que les formes tronquées de l'alpha-synucléine produites par la caspase-1 favorise l'agrégation de l'alpha-synucléine. D'autre part, ces troncations C-ter ont été montrées comme favorisant l'agrégation de l'alpha-synucléine *in vitro*<sup>412</sup> et étant cytotoxiques pour les cellules SH-SY5Y<sup>412,413</sup>. Bassil et al.<sup>414</sup>, ont montré que réduire les troncations en C-terminale en inhibant la caspase 1 dans un modèle murin de MSA permettait de réduire la neurodégénérescence. Des résultats similaires ont été obtenus par immunothérapies<sup>415</sup>.

- Glycosylation : L'alpha-synucléine à 9 sites de glycosylation O-GlcNAc<sup>416</sup>. Deux de ces sites T72 et S87 sont respectivement un site clivage de l'alpha-synucléine par la calpaine et un site de phosphorylation. Cette glycosylation a été montrée comme importante pour diminuer la capacité de l'alpha-synucléine a agréger en empêchant son clivage par la calpaine<sup>416</sup>.
- Ubiquitination : L'alpha-synucléine contient plusieurs sites d'ubiquitination<sup>405,417</sup>. L'alpha-synucléine monoubiquitynilé est dégradée par le protéasome<sup>418,419</sup> alors que l'alpha-synucléine polyubiquitynilé est dégradée par les lysosomes<sup>417</sup>. Elle peut être aussi SUMOylée et une dérégulation de sa SUMOylation conduirait à l'accumulation de l'alpha-synucléine en bloquant sa voie de dégradation via l'ubiquitine<sup>418</sup>.
- Nitration : cette modification a été montrée comme favorisant<sup>420,421</sup> la formation d'oligomères et l'agrégation de l'alpha-synucléine *in vitro*<sup>422,423</sup>.

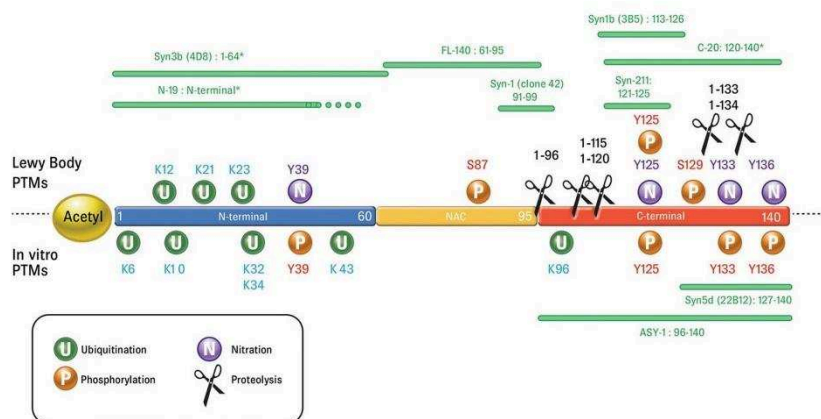


Figure 72 Principales modification post-traductionnelles de l'alpha-synucléine qui ont été observées dans les corps de Lewy ou *in vitro*. Les épitopes des anticorps anti-alpha-synucléine sont aussi présentés. D'après Schmid et al. 2013

Enfin, l'alpha-synucléine semble être aussi impliquée aussi dans la maladie d'Alzheimer. En effet, dans de nombreux cas des corps de Lewy sont trouvés dans les cerveaux de personnes atteintes de la maladie d'Alzheimer et des agrégats de tau chez les personnes atteintes d'une synucléinopathie, ainsi que des plaques amyloïdes<sup>424</sup>. Ces deux protéines interagissent et la forme 1-120 de l'alpha-synucléine semble promouvoir l'agrégation tau *in vitro*<sup>425</sup>.

La maladie d'Alzheimer et les démences à corps de Lewy peuvent être difficile à discerner<sup>426</sup> d'un point de vue clinique<sup>427</sup>, notamment aux stades précoces (voir figure 73). Il y a donc un besoin de biomarqueurs discriminant. Dans la maladie d'Alzheimer le déclin cognitif a été corrélé avec le niveau d'alpha-synucléine dans le cerveau de modèle murin<sup>428</sup> et l'alpha-synucléine serait augmentée dans le LCR d'après quelques études<sup>429,430</sup>. En comparaison, l'alpha-synucléine semble être diminué dans le LCR des synucléinopathies bien que les résultats soient variables<sup>431</sup>. Cette variabilité peut être causée par les anticorps utilisé pour les tests ELISA qui n'ont pas tous les mêmes épitopes ou dont l'épitope n'est pas clairement défini. De plus, l'alpha-synucléine est présente, en partie, sous forme tronquée dans le cerveau et les formes tronquées en C-ter favoriseraient l'agrégation et donc ces protéoformes de l'alpha-synucléine dans le LCR seraient potentiellement de meilleurs marqueurs que la forme intacte. Enfin, la combinaison de ces deux biomarqueurs pourrait permettre une meilleure différenciation de ces deux pathologies.

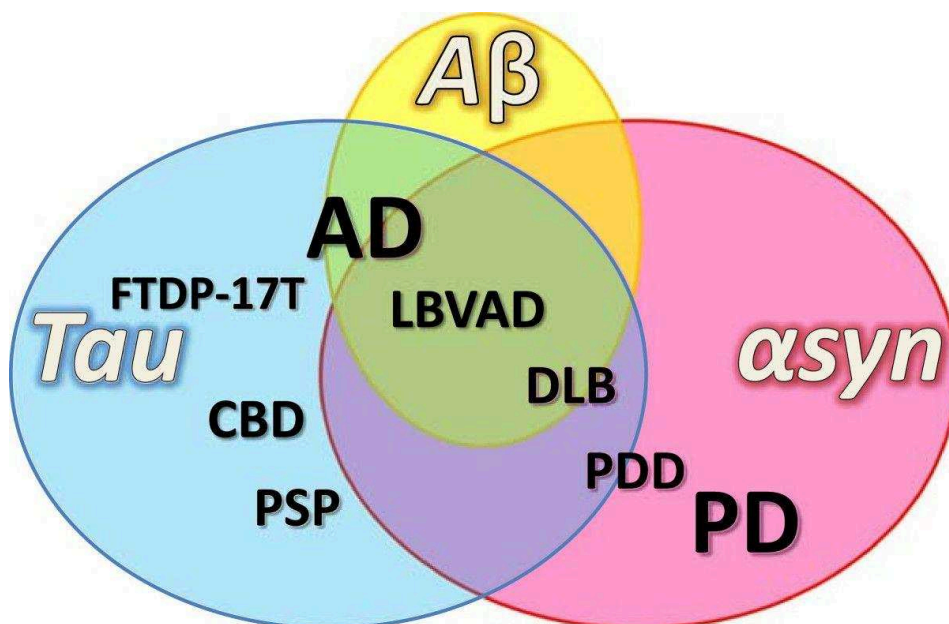


Figure 73 recoupement des principales maladies neurodégénératives où la protéine Tau, Abeta ou l'alpha-synucléine sont impliquées. D'après Moussaud et al. 2014

Dans ce contexte, ce chapitre va être composé de deux parties. Une première partie portera sur le développement d'une méthode de quantification multiplexe de l'alpha-synucléine et de Tau dans le LCR. Cette méthode a été appliquée à l'analyse de 23 échantillons de LCR contenant des démences à corps de Lewy, des contrôles et des cas d'Alzheimer. Un objectif dans le développement de cette première méthode est aussi de couvrir au mieux la séquence de l'alpha-synucléine par une approche de type Bottom-up, technique considérée comme la plus sensible et donc potentiellement apte à détecter les taux endogènes du LCR. Cette couverture de séquence, notamment la partie C-terminale, nous est apparue essentielle étant donné l'hétérogénéité de la protéine. Nous avons mis en place une approche similaire à celle développée par Barthelemy et al<sup>101</sup> où l'abondance relative des peptides permet d'estimer de manière l'indirecte la présence de modifications post-traductionnelles. Puis, la seconde partie traitera du développement d'une méthode d'analyse de l'alpha-synucléine sous forme intacte (top-down), permettant l'identification puis la quantification directe des formes tronquées de l'alpha-synucléine dans le putamen et le cortex frontal humains. Cette seconde partie a pour but de mettre en avant des formes tronquées impactées dans la MSA et de pouvoir à termes les rechercher dans le LCR.

## II. Matériels et méthodes

Le matériel correspondant aux analyses bottom—up est décrit dans l'article (III.). Concernant les analyses top-down, les conditions sont décrites ci-dessous. Les points particuliers qui ont nécessité des optimisations sont décrits dans la partie expérimentale.

### 1. Réactifs

Protéases : cathepsin-L : 25 µg from human liver de chez Sigma ( 6854-25ug), MMP-3 recombinant de chez Sigma (SRP7788-10ug), cathepsinD from human liver de chez Sigma C8696-25ug, calpain-1 de chez sigma C6108-100 ug et caspase-1 recombinants 25 UN C5482-25UN ( Sigma).

### 2. Méthodes LC-MS/MS

La séparation par chromatographie liquide a été réalisé avec une colonne en phase inverse C4 Acquity UPLC protein BEH C4 column, 300Å, 1.7 µm, 2.1 mm\*15 mm en 60 min à un débit de 500 µL/min. Les phases mobiles sont pour la phase A : eau +0.1 % acide formique et pour la phase B acétonitrile, 0.1% acide formique. Après un palier isocratique de 0.5 min à 5% B, le gradient est augmenté jusqu'à 20% de B en 10 min. Puis le pourcentage de B est augmenté jusqu'à 30% pendant 30 min. Le pourcentage de B est ensuite augmenté jusqu'à 50% en 20 min. Le pourcentage de B% est ensuite augmenté à 95% pendant 4 min. Enfin la colonne est rééquilibrée à 5% de B pendant 6 min. Paramètres MS : multiplexe SIM : Résolution 70 000 (250 m/z) , MIT = 250 ms, fenêtre d'isolement 3 m/z ; Multiplexe PRM : Résolution 35 000, MIT 120 ms et fenêtre d'isolement 3.0 m/z.

### 3. Conditions d'incubation de l'alpha-synucléine par les protéases

5 µL d'alpha-synucléine 15N à 1 mg/mL + 5µL d'acétonitrile+ 5µL de protéase + 80 µL de tampon spécifique à la protéase. Ces solutions sont ensuite mises à incuber. La digestion est arrêtée par 5 µL d'HCL à 1M.

Tampon cathepsines D et L : acétate de sodium à 50 mM pH=4.0 ; 40°C ; 45 min et 1h ;250µg/mL

Tampon caspase 1 : 50 mM HEPES (pH=7.4), 100 mM NaCl, 10 mM DTT, 1mM EDTA ; 42°C, 24 h, 250 µg/mL

Tampon MMP-3 : 20 mM tris-HCl (pH=7.5), 150 mM NaCl, 5mM CaCl<sub>2</sub>, 0.5mM ZnCl<sub>2</sub> ; 37°C, 30 min,100 µg/mL

Tampon calpain-1 : 40 mM HEPES ( pH=7.2), 5mM DTT et 1mM CaCl<sub>2</sub> ; 37°C, 15min , 100µg/mL

Ces 5 solutions sont ensuite rassemblées et 12.5  $\mu\text{L}$  est ajoutée par échantillons pour l'analyse de la cohorte.

### III. Article : quantification multiplexe de tau et de l'alpha-synucléine dans le LCR

## Antibody-free quantification of CSF Tau and $\alpha$ -synuclein by multiplex high resolution targeted mass spectrometry.

Viodé A<sup>1</sup>, Benyounes I<sup>2</sup>, Epelbaum S<sup>4</sup>, Verny M<sup>5</sup>, Fenaille F<sup>1</sup>, Junot C<sup>1</sup>, Foudil L<sup>2,3</sup>, Becher F<sup>1</sup>

9. Service de Pharmacologie et Immunoanalyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments (LEMM), CEA, INRA, Université Paris Saclay, F-91191 Gif-sur-Yvette cedex, France.
10. Service de Biochimie métabolique, Hôpitaux universitaires Pitié-Salpêtrière, AP-HP, 75013, Paris, France.
11. GRC-2011 Neurométabolisme, Université Pierre et Marie-Curie –Sorbonne, 75006, Paris, France.
12. Institut de la mémoire et maladie d'Alzheimer (IM2A), Département de neurologie, Hôpitaux universitaires Pitié-Salpêtrière, AP-HP, 75013, Paris, France
13. Service de Gériatrie, Hôpitaux universitaires Pitié-Salpêtrière, AP-HP, 75013, Paris, France

### Abstract

Alzheimer disease (AD) and Dementia with Lewy bodies (DLB) are two clinically overlapping neurodegenerative diseases characterized by the abnormal aggregation of proteins, respectively tau and alpha-synuclein. Both proteins are promising biomarkers candidate for diagnosis or therapeutics response. However, the level of total-alpha-synuclein reported among studies are inconsistent and recent mass spectrometry (MS) method failed to monitored peptides from the C-terminal part of the protein where numerous post-translational modifications are found. We have developed a sensitive and robust MS based assay for the simultaneous quantification of tau and alpha-synuclein in cerebrospinal fluid (CSF). The workflow was optimized for optimal sequence coverage of alpha-synuclein with an emphasis on its C-termini by the combination of two enzymes, Glu-C and trypsin. The extraction protocol is based on protein precipitation by perchloric acid and thus avoid the use of affinity reagents. Regarding alpha-synuclein up to 7 peptides were quantified, including the extreme C-termini peptide (132-140), resulting in a sequence coverage of 54% in CSF. The lower limits of quantification (LLOQ) ranged from 0.1 ng/mL to 1 ng/mL depending on the peptide. Concerning tau, 4 peptides, common to all isoforms, were monitored with LLOQ ranging from 0.5 ng/mL to 0.75 ng/mL. The method was successfully applied to CSF samples from AD and DLB patients. Tau and alpha-

synuclein concentration showed opposite trends in AD and DLB patients suggesting the benefit of combining the two biomarkers to increase the diagnosis accuracy.



## INTRODUCTION

Microtubule associated protein Tau and  $\alpha$ -synuclein are implicated in the pathogenesis of major neurodegenerative diseases characterized by intraneuronal proteins aggregates. Neurofibrillary tangles containing tau protein are observed in Alzheimer Disease (AD)<sup>432</sup>. Abnormal accumulation of misfolded  $\alpha$ -synuclein leads to the presence of Lewy Bodies in Parkinson Disease (PD), Dementia with Lewy Bodies (DLB) or glial cytoplasmic inclusions in Multiple System Atrophy (MSA)<sup>433</sup>. An overlap between tau and  $\alpha$ -synuclein is reported, based on mutual interaction, cross-seeding and concomitant accumulation in intracellular inclusions<sup>434</sup>. Notably, co-existence of  $\alpha$ -synuclein and tau abnormal deposition was observed in the brains of patients with AD, PD and DLB<sup>435,436</sup>.

In addition to aggregation, tau and  $\alpha$ -synuclein share common important structural features. Tau and synuclein belong to the family of natively unfolded proteins<sup>437</sup>, exhibit highly modified sequences by allele-specific isoforms and post-translational modifications. Some of these modifications, especially phosphorylation and truncations<sup>210,438–440</sup> are thought to promote aggregation.

The continuous exchange of cerebrospinal fluid (CSF) with brain interstitial fluid in direct contact with neurons supports the investigation of changes in CSF protein concentrations for diagnostic research. Increased CSF tau can be synonymous with AD or Creutzfeldt-Jakob disease and assumed to be linked to neuronal injury<sup>441,442</sup>. Interestingly, higher CSF  $\alpha$ -synuclein concentrations in patients with AD were reported in some studies<sup>429,443</sup>. Consequently, CSF  $\alpha$ -synuclein was proposed as an additional molecular candidate biomarker of early AD pathophysiology<sup>430</sup>. Reduced CSF  $\alpha$ -synuclein is suggested in synucleinopathies, although inconsistent results and discrepancies in absolute concentrations were reported across studies<sup>444,445</sup>. CSF concentrations were determined by immunoassays in the majority of the published clinical studies<sup>444</sup>. Disparate affinity of the different antibody pairs for proteoforms, matrix effects and lack of standardization may account for the differences observed between measurements<sup>444,446</sup>.

Targeted mass spectrometry has proven to be a powerful alternative for precise and robust quantification of proteins in biological matrices, that serve to improve assay standardization between laboratories or to get better insight into proteoform abundance<sup>278,447,448</sup>. Nonetheless, application to low-abundant proteins still requires efficient sample clean-up protocols. Recently, a few MS methods have met the challenging detection of CSF tau or  $\alpha$ -synuclein thanks to different purification approaches, based either on immunocapture, protein precipitation or peptide fractionation, but none was able to measure simultaneously both proteins<sup>101,161,346,449,450</sup>.

The aim of this work was first to develop a common sample preparation protocol to quantify CSF tau and  $\alpha$ -synuclein in the same mass spectrometry analysis, considering the overlap between the proteins in neurodegenerative diseases. In addition, we sought to provide high sequence coverage of  $\alpha$ -synuclein, especially at the C-terminal part where disease related PTMs are reported<sup>314</sup>. In the present work, the previously reported assay of CSF tau<sup>101</sup> was improved for simultaneous quantification of  $\alpha$ -synuclein down to 100 pg/mL. A two-step digestion strategy was implemented for quantitative monitoring in the parallel reaction monitoring (PRM) mode of seven  $\alpha$ -synuclein peptides distributed along the sequence, resulting in 54% sequence coverage. CSF Tau and  $\alpha$ -synuclein were successfully determined in patients with AD or DLB patients, two neurodegenerative dementia with overlapping clinical symptoms.

## **MATERIALS AND METHODS**

### **CSF Collection**

Participants gave written consent for participation in the study, which was approved by the ethics committee. CSF was collected using standardized collection, centrifugation, and storage protocols. CSF samples were transferred to the laboratories in <4 h, centrifuged (1000g, 10 min, at 4–8 °C, without breaks), and aliquoted in polypropylene tubes before storage at –80 °C. CSF  $\alpha$ -synuclein concentration was measured using ELISA U-PLEX Kit from MesoScale Discovery (MSD) according to the manufacturer's instructions

### **Chemical and materials:**

<sup>14</sup>N recombinant tau (isoform 441) and  $\alpha$ -synuclein were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France) (>90% purity by SDS PAGE). Lyophilized samples were suspended in water to a final concentration of 100  $\mu$ g/mL. The solution of <sup>15</sup>N-tau-441 recombinant protein, prepared as detailed elsewhere, was a gift from Guy Lippens (CNRS - Université de Lille 1, UMR8576, Villeneuve-d'Ascq, France)<sup>451</sup>. <sup>15</sup>N-  $\alpha$ -synuclein was purchased from AlexoTech AB, Umea, Sweden.

Sequencing Grade Modified Trypsin was from Promega, Madison, USA, Endoproteinase Glu-C, ammonium hydroxide, ammonium bicarbonate were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). RapiGest SF Surfactant and  $\mu$ SPE Oasis HLB sorbent plates were purchased from Waters Corporation (Milford, MA). Labeled peptides were synthesized in Pepotec quality by ThermoFisher Scientific (Paisley, UK). Water (ChromaSolve LC-MS), acetonitrile (HPLC-grade), and formic acid were obtained from Honeywell/Riedel-de Haen (Seelze, Germany) and VWR chemicals (Fontenay sous Bois, France), respectively. All other chemicals were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France) or VWR Chemicals (Fontenay sous Bois, France). Pierce BCA protein Assay kit was purchased from Pierce (Rockford, IL, USA). For all reactions, LoBind Eppendorf tubes (Dutscher, Brumath, France) were used.

**Sample preparation:**

CSF samples, calibration standards, and quality control samples in 0.5% serum were spiked with 15N-tau-441 and 15N- $\alpha$ -synuclein immediately after thawing (20  $\mu$ L addition to 200  $\mu$ L of sample volume). Ten microliters of 70% perchloric acid was added to precipitate proteins and samples were kept on ice for 15 min before centrifugation (15 min, 4 °C, and 16 000g). Supernatants were collected and mixed with 20  $\mu$ L of 1% trifluoroacetic acid (TFA) before  $\mu$ SPE extraction on Oasis HLB sorbent, previously conditioned with 300  $\mu$ L of methanol, and equilibrated with 500  $\mu$ L of 0.1% trifluoroacetic acid (TFA). Samples were loaded and washed with 500  $\mu$ L of 0.1% TFA containing 10% acetonitrile. Tau and  $\alpha$ -synuclein were eluted with 100  $\mu$ L of 50% acetonitrile in 0.1% TFA. Extracts were evaporated to dryness with a Turbovap instrument (Biotage, Uppsala, Sweden) (10 psi of nitrogen, 50 °C, 2 h). The dry residue was re-dissolved in Rapigest 0.05%, digested overnight at 25 °C with 10  $\mu$ L Glu-C at 500 ng/ $\mu$ L and during 4 hr at 37°C with 5  $\mu$ L trypsin at 100 ng/ $\mu$ L. Finally, the digest was acidified with 5  $\mu$ L of HCl 1M, incubated 45 min à 37°C before centrifugation at 14000 rpm for 10 min and stored at -20 °C prior to LC-MS/HRMS analysis.

**LC-MS/MS analysis:**

LC-MS/MS was performed on a Dionex Ultimate 3000 capillary chromatography system coupled to a Q-exactive Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Ten microliters of sample was loaded onto the column. Chromatographic separation was performed on a Acclaim PepMap 300 C18 reverse phase column (150 mm  $\times$  0.3 mm; 5  $\mu$ m, 300 Å; Thermo Fisher Scientific, Bremen, Germany) and achieved in a 60 min gradient at a flow rate of 6  $\mu$ L/min. A gradient of mobile phase consisting of LC-MS-grade water with 0.1% formic acid (phase A) and acetonitrile with 0.1% formic acid and 20% water (phase B) was delivered. After an isocratic step of 3 min at 4% B, the gradient was ramped to 35% over the next 60 min. Then acetonitrile was increased to 95% for the next 6 min. Column re-equilibration at 4% B was realized for 21 min.

Instrument parameters of the electrospray ionization source were set as follows: sheath gas flow rate at 10 a.u., spray voltage at 2.8 kV, capillary temperature at 320 °C. The Q-exactive instrument was operated in positive ion mode under time-scheduled sequential PRM acquisition. Endogenous peptides precursor ions and labeled counterparts were selected in the quadrupole with an isolation mass window of 2 m/z. Precursors were fragmented in the HCD cell using nitrogen as collision gas and the optimized normalized collision energy (Table 3). All fragment ions were transferred to the Orbitrap. Resolution was set to 70 000 at m/z 200 (full width at half-maximum), automatic gain control to 1e6, and maximum injection time between 500 and 1000 ms depending on the number of eluting peptides. Xcalibur 2.2 software (Thermo Fisher Scientific, Bremen, Germany) was used for instrument control and processing of the data files.

#### **PRM analysis and quantification:**

A time-scheduled sequential PRM method was established targeting the following tau and  $\alpha$ -synuclein peptides (Table 3). To increase the signal to noise ratio and assay sensitivity, the signal of up to 6 major and non-interfered fragment ions identified with high resolution (5 ppm) from a common peptide precursor were summed up to provide one extracted ion chromatogram (XIC) for each peptide. Isotope-labeled peptides counterparts with  $^{15}\text{N}$  labeled amino acids were used for signal normalization and quantification. For specific quantification in individual CSF extract, we applied the following inclusion criteria (i) similar peak shape and retention time to  $^{15}\text{N}$ -tau/ $^{15}\text{N}$   $\alpha$ -synuclein peptides from internal standard ( $\pm 0.05$  min), (ii) detection of the most intense PRM transitions, and (iv) similar area ratio among transitions between endogenous and spiked  $^{14}\text{N}$ - tau/ $^{14}\text{N}$   $\alpha$ -synuclein for signals with enough intensity. Raw mass spectrometry data were exported to Skyline 4.1<sup>347</sup> for verification of the transitions ratio of unlabeled and labeled peptides. Xcalibur 2.2 software (Thermo Fisher Scientific, Bremen, Germany) was used for quantitative data analysis. Unlabeled/labeled peak area ratios were calculated for the different peptides and used for further calculation. External calibration curve was made by spiking tau and  $\alpha$ -synuclein in 0.5% human serum as artificial CSF from 0.25 to 50 ng/mL and

0.1 to 20 ng/ml, respectively. Linear regression with 1/x weighting was applied to generate a standard curve.

### **Statistical analysis**

All statistical analysis was performed using Graphpad Prism software (version 5.01). Data were compared with a Mann–Whitney test and medians were considered significantly different if  $p < 0.05$ .

## RESULTS AND DISCUSSION

### METHOD DEVELOPMENT FOR $\alpha$ -SYNUCLEIN

Efficient sample clean-up is a prerequisite for detection of endogeneous  $\alpha$ -synuclein concentration in CSF<sup>450,452</sup>. The antibody-free extraction protocol developed by Barthelemy et al<sup>101</sup> for tau protein was selected considering the similar natively unfolded structure of  $\alpha$ -synuclein and unchanged conformation by pH and chemical denaturant<sup>437</sup>. Protein precipitation by perchloric acid was combined with microsolid phase extraction ( $\mu$ SPE). Elution solvent strength was increased to 50% acetonitrile for higher recovery of  $\alpha$ -synuclein. In this condition, recovery of  $\alpha$ -synuclein was determined at  $63\pm 7\%$  by spiking experiments with recombinant  $\alpha$ -synuclein in artificial CSF, before and after processing (Table 1). It was concluded that perchloric acid protocol provides efficient extraction of  $\alpha$ -synuclein.

$\alpha$ -synuclein is highly complex in terms of structural modifications with numerous proteoforms reported, including truncations and post-translational modifications<sup>314</sup>. Notably, PTMs within the C-terminal region are thought to favor aggregation and toxicity<sup>210,444</sup>. In this context, MS-based assay should detect as many peptides as possible, covering the main site of modifications. The large C-terminal peptide released by trypsin, covering residues 104 to 141, was generally not detected by latest MS-based assays, due to its size and acidic nature<sup>277,314,452</sup>. Glu-C digestion was suggested as an alternative to trypsin for improved sequence coverage because of several glutamic acid residues distributed along the sequence<sup>314</sup>. However,  $\alpha$ -synuclein quantification in CSF using Glu-C digestion was never reported up to now. Peptides identified from recombinant  $\alpha$ -synuclein following Glu-C digestion are indicated in Table 2. Interestingly, a digest of extracted human whole blood lysate showed similar peptide profile, including the acetylated N-termini peptide 1-13<sup>453</sup>(Table 2). Sequence coverage was found at 86.4% (Table 2) impacted by two undetected basic peptides from the N-terminal region of  $\alpha$ -synuclein i.e. KTKQGVAE (21-28) and AAGKTKE (29-35). Considering the C-terminal region only, sequence coverage reached 100%. The results highlight the effectiveness of Glu-C for both recombinant and endogeneous  $\alpha$ -synuclein. Four of the released peptides by Glu-C were identified with missed-cleavage sites i.e. Glu or Asp residues<sup>454</sup>, at positions 98/104 (peptide

GAGSIAATGFVKKDQLGKNEE), 110 (peptide GAPQEGILE), 126/130 (peptide AYEMPSEE), 139 (peptide GYQDYEPEA) (Table 2). Signal was below 10 % of the parent peptides for the cleaved product of peptides GAGSIAA..., AYEMPSEE and GYQDYEPEA, demonstrating only minor cleavage at positions 104, 126 and 130<sup>455</sup>. Significant cleavage was observed at position 110 illustrated by signal of the cleaved peptide GILE at about 62 % of GAPQEGILE. GAPQEGILE peptide was therefore selected in the final method, but not considered as quantotypic peptide<sup>456,457</sup>. GILE peptide was excluded because of questionable specificity of small peptide sequence.

### **CONDITIONS OPTIMISATION FOR SIMULTANEOUS DETECTION OF TAU AND $\alpha$ -SYNUCLEIN**

Regarding the multiplex detection of  $\alpha$ -synuclein with tau, Glu-C digestion was not appropriated due to a deficit of cleavage site in the central core of tau protein where peptides fulfilling the criteria for accurate quantification are located<sup>101</sup>. A sequential Glu-C/trypsin digestion was therefore optimized for maximum and reproducible peptides release. Following investigation of Glu-C quantity and temperature during incubation, optimal digestion was achieved by overnight incubation with Glu-C at 25°C followed by 4 hours incubation with trypsin at 42°C (Figure 1).  $\alpha$ -Synuclein peptides signal following sequential Glu-C/trypsin digestion was confronted to Glu-C only protocol. Similar responses were observed in both conditions, showing no detrimental impact of the sequential protocol on the release and/or detection of  $\alpha$ -synuclein peptides (Figure 2). The final list of peptides selected for tau and  $\alpha$ -synuclein are provided in Table 3. Nine  $\alpha$ -synuclein peptides provided a sequence coverage of 66% from position 14 to 140. Recombinant <sup>15</sup>N-labeled tau and  $\alpha$ -synuclein were digested in the same conditions, <sup>15</sup>N isotopologue version of the 13 selected peptides were identified (Table 3). In addition, peptide VLSPADK from hemoglobin was best detected in human whole blood and included in the selection (Table 3) as a marker of hemolysis.

A scheduled PRM acquisition method was established targeting the 28 selected peptides in total for quantification of tau,  $\alpha$ -synuclein and hemoglobin including isotope-labeled versions (Table 3).



Acquisition parameters were thoroughly adjusted to yield the best detection sensitivity and multiplexing capacity. Briefly, high fill times were applied (from 300 to 1200 ms) to maximize the accumulation of targeted peptide, considering also cycle time to allow collecting around ten data points to define the chromatographic peaks (Figure 3). In this objective, capillary chromatography conditions for peptides separation were carefully adjusted to minimize possible coelutions. Data processing was optimized by summing the signals of up to six major and non-interfered fragment ions to provide one XIC for each targeted peptide<sup>122</sup> (supplementary Table S2). Each endogenous peptide and their corresponding isotope-labeled form must strictly co-elute with similar transition ratio across the different samples in comparison with a standard. If not, the transition was excluded. Finally, the ratios of PRM signal from unlabeled and labeled peptides were used for establishment of the corresponding calibration curves and further quantification in CSF samples. The final protocol for simultaneous quantification of tau and  $\alpha$ -synuclein is illustrated in Figure 4.

## METHOD PERFORMANCE ASSESSMENT

The assay was assessed for Tau/  $\alpha$ -synuclein quantification in human CSF by evaluating each of the selected peptide. Linearity, sensitivity, intra-day assay precision, and matrix effect were assessed. Standard curves and QCs were prepared in artificial CSF, as a surrogate to human CSF, for linearity, sensitivity and intra-day variation evaluation. Quality control samples prepared in human CSF spiked or not with recombinant Tau/  $\alpha$ -synuclein were used for matrix effect determination. Linearity and sensitivity of the method were evaluated with a 9-point calibration curve. The method was shown to be linear from 0.1 to 20 ng/mL for  $\alpha$ -synuclein and 0.5 to 50 ng/mL for Tau, based on best peptides of each protein (Figure 4, Table 4). LLOQ was determined based on accuracy between 80 and 120 %, and ranged from 0.1 to 1 ng/mL for the 7 best monitored  $\alpha$ -synuclein peptides, and 0.5 to 0.75 ng/mL for tau peptides (Table 4). Unfortunately, LLOQ was above 20 ng/mL for peptides DMPVDPDNE and AYEMPSEE of synuclein, above the reported CSF  $\alpha$ -synuclein concentrations<sup>450,452,458</sup>, most probably in relation to bad ionization properties or methionine oxidation. Regarding tau quantification, sensitivity of the multiplex assay was confronted to the original protocol using trypsin only<sup>101</sup>. We observed similar range of linearity and LLOQ, showing no impact of the simultaneous quantification (Figure 6, Table 4). Intraday repeatability was evaluated by the replicate analysis of spiked artificial CSF at low and high concentrations. Accuracy and precision were determined in the range 80-120% and below 20% for each peptide, respectively, in line with recommendation<sup>459</sup> (Table 4). High precision was observed for peptide GAPGEGILE i.e. CV below 10%, despite the missed-cleavage site. To determine a potential matrix effect between artificial and human CSF, and demonstrate accurate quantification in human CSF, recombinant Tau/  $\alpha$ -synuclein were spiked into human CSF in triplicates at low and high concentration, i.e. 1 – 2.5 ng/mL and 2.5 – 12.5 ng/mL respectively and confronted against the calibration curve in artificial CSF (Table 5). Peptides displayed accuracy of back-calculated concentrations in the range of 85-115%, showing the suitability of artificial CSF as a surrogate matrix for Tau/  $\alpha$ -synuclein determination in human CSF. Taking all results together, the protocol was found

efficient for simultaneous determination of Tau/  $\alpha$ -synuclein in human CSF. The new method demonstrated robustness with variability and accuracy below 20%, with 7 synuclein peptides reaching a LLOQ at or below 1 ng/mL and covering 54% of the sequence.

### CSF SAMPLE ANALYSIS

The multiplex method was applied to 23 CSF samples of patients with Alzheimer disease (n=6), Dementia with Lewy Bodies (DLB) (n=10) or controls with non-neurodegenerative disorders (n=7). Tau and /  $\alpha$ -synuclein were determined in parallel by ELISA.

Concentrations of  $\alpha$ -synuclein were in the range 0.52 - 3.40 ng/mL (peptide GAGSIAATGFVK), similar or slightly higher than previously reported<sup>444,446,452</sup>. Two samples with  $\alpha$ -synuclein concentrations above 5 ng/mL, were excluded for suspected interference by hemolysis/blood contamination. Contamination was further confirmed by the intense and significantly increased signal of the hemoglobin peptide (Figure 7).  $\alpha$ -synuclein levels were reliably and consistently quantified with the 7 best peptides in nearly 100% of patients (Table 6), resulting in a total sequence coverage 54%, from position 14 to the last amino acid in position 140. This result confirms the good MS sensitivity of those 7 peptides in CSF samples. Peptides AYEMPSEE and DMPVSPDNE were not detected, as anticipated from the spiking experiments with recombinant  $\alpha$ -synuclein. Mean molar  $\alpha$ -synuclein concentrations determined with each of the 7 quantitative peptides are illustrated on Figure 8. N-terminal peptides GVVAAGE and GVLVVGSK are common to  $\alpha$ -synuclein and  $\beta$ -synuclein, resulting in higher relative abundance to central or C-terminal peptides. Lower abundance was observed for peptides 62-80 and 84-96 belonging to the central region compared to the C-termini (mean at 1.09 and 2.92 ng/mL respectively, vs 2.92 ng/mL,  $p < 0.0001$ ), which could be related to GlcNAcylation, phosphorylation at Ser 87, or truncations at position 65, 66, 68, 78, 93<sup>210,438</sup>. Peptide GYQDYEEA 132-140 was well detected, at level above LLOQ in 20 samples, suggesting intact C-termini of CSF  $\alpha$ -synuclein. The higher abundance of peptide GAPGEGILE 106-114 could be related either to a different behavior of recombinant and endogenous  $\alpha$ -synuclein<sup>278</sup>, e.g. different digestion at the missed-cleavage site, or

occurrence in CSF of a short truncated form of  $\alpha$ -synuclein containing peptide GAPGEGILE. We will develop intact mass spectrometry analysis for further investigation in future works. Regarding tau, all peptides from the central region were above LLOQ in 100% of patients. Concentrations were within the same range as Barthelemy et al <sup>101</sup>, with similar higher abundance compared to the C-terminal peptide SPVVSAGDTSPR (Table 6).

Quantitative results obtained for  $\alpha$ -synuclein and tau appeared to be highly correlated with the ELISA measurements (Figure 9a), despite higher absolute MS levels than those from ELISA. Deviations in absolute values was already reported for tau <sup>101</sup>, and attributed to differences in the standard used for calibration or the selection of distinct proteoforms depending on the analytical strategy. The lower difference observed for  $\alpha$ -synuclein, i.e. MS values from 1.58 to 4.38 (peptide 62-80 and 132-140) times higher than those from ELISA, could be associated in part with a lower structural heterogeneity, as suggested in this study by peptide abundances along the sequence.

$\alpha$ -Synuclein and tau concentrations appeared correlated (Figure 9b) and significantly increased in AD patients compared to DLB and controls (Figure 10), in line with some studies.  $\alpha$ -Synuclein is highly enriched in synapses<sup>460</sup>, and increased levels of  $\alpha$ -synuclein in AD could be related to the synaptic degeneration<sup>429,430</sup>. Consequently, CSF  $\alpha$ -syn was suggested as an additional molecular candidate biomarker of early AD pathophysiology<sup>430</sup>. Interestingly,  $\alpha$ -synuclein was also found significantly or significantly close ( $p=0.06$ ) lower in DLB compared to controls patients, depending on the selected peptide (Figure 10). In summary, disparate alterations of CSF concentration in DLB and AD were observed by MS, which might constitute promising discriminating markers. Further investigations in large cohorts of patients should confirm the clinical utility of the new multiplex assay for differentiation between AD, DLB and controls.

## CONCLUSIONS

Simultaneous CSF tau and  $\alpha$ -synuclein quantification by mass spectrometry was established for the first time. Signature peptides were efficiently released through a combined Glu-C/trypsin digestion protocol and assay reliability was demonstrated. Up to 7 CSF  $\alpha$ -synuclein peptides were quantified along the sequence including the peptide in position 132-140, resulting in a sequence coverage of 54%. Comparison of peptide abundance suggests a lower structural heterogeneity in comparison to CSF tau, especially at the C-termini.

CSF levels of tau and  $\alpha$ -synuclein showed opposite trends in AD and DLB patients, which indicate potential differentiation of two clinically overlapping neurodegenerative dementia by combination of both markers. Simultaneous quantification of the two markers in a single MS assay provides reduced CSF sampling, lower processing time and higher robustness. The new antibody-free method will also facilitate harmonization of measurements between laboratories. Further validation of the assay in larger cohorts is needed to confirm clinical utility. In parallel, future effort may be directed toward more sensitive mass spectrometry analysis, which would enable quantification of peptide 124-131, released by Glu-C but not detected in CSF, and potentially the abundance of phosphorylation on Ser 129.

Table 1: Total recovery of  $\alpha$ -synuclein extraction protocol

$\alpha$ -Synuclein peptides*	Recovery**
EGVVHGVATVAEK	62 $\pm$ 11%
EGVVAAAEEK	66 $\pm$ 9%
EGVLYVGSK	63 $\pm$ 9%
EQVTNVGGAVVTGVTAVAQK	60 $\pm$ 1%
TVEGAGSIAAATGFVK	65 $\pm$ 3%
Mean	63 $\pm$ 4%

\*Trypsin was used for digestion of  $\alpha$ -synuclein in this experiment

\*\* after extraction in artificial CSF; n=2

Table 2: Peptides identified following Glu-C digestion of recombinant and whole blood  $\alpha$ -synuclein

Peptides	Position	m/z (observed)	m/z (theoretical)	z	
Acetyl-MDVFMKGLSKAKE*	1-13	509.2637	509.2633	3	
GVVAAAE	14-20	616.3304	616.3301	1	
KTKQGVAE	21-28	ND			
AAGKTKE	29-35	ND			
GVLYVGSKTKE	36-46	394.2254	394.2239	3	
GVVHGVATVAE	47-57	519.7831	519.7825	2	
QVTNVGGAVVTGVTAVAQKTVE	62-83	710.0600	710.0585	2	
GAGSIAAATGFVKKDQLGKNEE	84-105	548.5391	548.5393	4	
GAPQEGILE	106-114	457.2530	457.2349	2	
DMPVDPDNE	115-123	516.2036	516.2029	2	
AYEMPSEE	124-131	478.1898	478.1893	2	
GYQDYEPEA	132-140	536.2176	536.2169	2	
<i>Sequence coverage</i>	<i>86.4%</i>				
					<i>Signal intensity (relative)</i>
<i>Cleaved peptides</i>					
GAGSIAAATGFVKKD	84-98	464.9206	464.9209	3	<10%
GAGSIAAATGFVKKDQLGKNE	84-104	516.2796	516.2786	4	<10%
GAPQE	106-110	ND			
GILE	111-114	431.2500	431.2500	1	60%
AYEMPSE	124-130	ND			
MPSEE	127-131	592.2283	592.2283	1	<10%
GYQDYEPE	132-139	500.6986	500.6984	2	<10%

\*: identified in human whole blood

ND: not detected

Table 3: Targeted  $\alpha$ -synuclein and tau peptides and MS conditions

$\alpha$ -Synuclein	Precursor ion m/z	NCE (%)	Product ion type
14N GVVAAAE	616.33 (z=1)	20	b3 ; b4 ; b5 ; b6 ; y4
15N GVVAAAE	6223.31 (z=1)	20	b3 ; b4 ; b5 ; b6 ; y4
14N GVLYVGSK	412.00 (z=2)	15	b3 ; y6 <sup>+2</sup> ; y6 <sup>+2</sup> (P+1) ; y5 ; y5 (P+1) ; y6 ; y6 (P+1)
15N GVLYVGSK	416.23 (z=2)	15	b3 ; y6 <sup>+2</sup> ; y6 <sup>+2</sup> (P+1) ; y5 ; y5 (P+1) ; y6 ; y6 (P+1)
14N GVVHGVATVAE	519.78 (z=2)	20	b9-H2O +2; b9 +2 ; b10+2 ; b7 ; b8- H2O ; b9-H2O ; b9-H2O P+1 ; b9 ; b9 p+11 ; b10 ; b10 p+1
15N GVVHGVATVAE	526.26 (z=2)	20	b9-H2O +2; b9 +2 ; b10+2 ; b7 ; b8- H2O ; b9-H2O ; b9-H2O P+1 ; b9 ; b9 p+11 ; b10 ; b10 p+1
14N QVTNVGGAVVTGVTAVAQK	600.34 (z=3)	17	y3 ; y4 ; y5 ; y6 ; y8 ; y9 ; y10 ; b7 ; b8 ; b9
15N QVTNVGGAVVTGVTAVAQK	607.98 (z=3)	17	y3 ; y4 ; y5 ; y6 ; y8 ; y9 ; y10 ; b7 ; b8 ; b9
14N GAGSIAAATGFVK	575.32 (z=2)	22	b4 ; b5 ; y4 ; y5 ; y6 ; y7 ; y8 ; y9 ; y10 ; y11
15N GAGSIAAATGFVK	582.29 (z=2)	22	b4 ; b5 ; y4 ; y5 ; y6 ; y7 ; y8 ; y9 ; y10 ; y11
14N GAPQEGILE	457.23 (z=2)	17	y2 ; b4 ; y4 ; b5 ; y5 ; b7 ; b8 ; b8 P+1 ; b7
15N GAPGEGILE	462.22 (z=2)	17	y2 ; b4 ; y4 ; b5 ; y5 ; b7 ; b8 ; b8 P+1 ; b7
14N DMPVDPDNE	516.2 (z=2)	17	y4 ; y5 ; y7 ; b8
15N DMPVDPDNE	521.19 (z=2)	17	y4 ; y5 ; y7 ; b8
14N AYEMPSEE	478.19 (z=2)	17	y3 ; y4 ; b4 ; b7
15N AYEMPSEE	482.18 (z=2)	17	y3 ; y4 ; b4 ; b7
14N GYQDYEPEA	536.22 (z=2)	17	y3 ; b5 ; b6 ; b7 ; b8
15N GYQDYEPEA	541.22 (z=2)	17	y3 ; b5 ; b6 ; b7 ; b8
<b>Tau</b>	Precursor ion m/z	NCE (%)	Product ion type
14N TPPAPK	305.68 (z=2)	10	y5 +2 ; y4 ; y5 ; y5 P+1
15N TPPAPK	309.17 (z=2)	10	y5 +2 ; y4 ; y5 ; y5 P+1
14N SGYSSPGSPGTPGSR	697.73 (z=2)	20	b3 ; y4 ; b5 ; y7 ; y7 P+1 ; b8 ; y10 ; y10 P+1 ; y11 P+1 ; y12 ; y12 P+1
15N SGYSSPGSPGTPGSR	706.29 (z=2)	20	b3 ; y4 ; b5 ; y7 ; y7 P+1 ; b8 ; y10 ; y10 P+1 ; y11 P+1 ; y12 ; y12 P+1

14N TPSLTPPTR	533.80 (z=2)	25	y4 ;y6 ;y6 P+1 ; y7 ; y8 ;y8 P+1
15N TPSLTPPTR	540.28 (z=2)	25	y4 ;y6 ;y6 P+1 ; y7 ; y8 ;y8 P+1
14N SPVVSGDTSPR	551.28 (z=2)	25	y4 ; y6 ; y6 P+1 ;y7 ;y8 ; y8 P+1 ;y9 ;y9 P+1
15N SPVVSGDTSPR	558.26	25	y4 ; y6 ; y6 P+1 ;y7 ;y8 ; y8 P+1 ;y9 ;y9 P+1
<b>Hemoglobin</b>	Precursor ion m/z	NCE (%)	Product ion type
VLSPADK	365.2107	20	y6 ;y5 ;y4



Table 4: Assay performances of each  $\alpha$ -synuclein and tau peptides in artificial CSF

Peptide	QC Low			QC High	
	LLOQ (ng/mL)	Accuracy (%)	CV%	Accuracy (%)	CV%
<b><math>\alpha</math>-Synuclein</b>					
GVVAAAE	0.2	100.3	15.4	102.8	8.6
GVLVVGSK	0.2	104.8	13.6	96.3	3.0
GVVHGVATVAE	0.4	115.3	8.4	99.7	0.8
QVTNVGGAVVTGVTAVAQK	0.4	120.8	14.7	98.1	1.9
GAGSIAAATGTFVK	0.1	83.3	11.6	98.4	3.4
GAPQEGILE	1.0	NA	NA	98	9.0
DMPVVPVNE	>20	NA	NA	NA	NA
AYEMPSEE	>20	NA	NA	NA	NA
GYQDYEPEA	1.0	NA	NA	102.1	6.8
<b>Tau</b>					
TPPAPK	0.75	114.9	21.8	95.2	8.7
SGYSSPGSPGTPGSR	0.75	94.9	5.5	92.6	2.8
TPSLPTPPTR	0.50	103.6	17.7	93.3	8.0
SPVVSGDTSR	0.75	98.7	9.5	91.6	3.9

Accuracy and CV% were determined with n=3 replicates

LLOQ was determined as the lowest concentration with bias<20%

QC Low: 0.4 and 1 ng/mL for  $\alpha$ -Synuclein and tau respectively

QC High: 4 and 10 ng/mL for  $\alpha$ -Synuclein and tau respectively

NA: not applicable

Table 5: Evaluation of matrix effect in human CSF-for synuclein a) and tau b)

a)

QC human CSF (n=3)	GVVAAAE			GVLYVGSK			GVVHGVATVAE			GAGSIAAATGFVK			GYQDYEPEA		
	Mean ng/mL	Bias	CV	Mean ng/mL	Bias	CV	Mean ng/mL	Bias	CV	Mean ng/mL	Bias	CV	Mean ng/mL	Bias	CV
Blank (endogenous)	1.79	/	21%	1.58		15%	0.38	/	13%	0.33	/	7.9%	LLOQ	/	/
Spike (+1 ng/mL)	2.86	7.9%	2.9%	2.71	13.0%	9.5%	1.35	3.1%	1.9%	1.46	13.0%	4.9%	LLOQ	/	/
Spike (+5 ng/mL)	6.7	1.6%	3.8%	6.35	4.7%	5.6%	4.94	8.7%	12%	5.0	6.6%	7.5%	5.24	1.7%	4.4%

b)

QC human CSF (n=3)	TPPAPK			SGYSSPGTPGSR			TPSLTPPTR			SPVSGDTSR		
	Mean (ng/mL)	Bias	CV	Mean (ng/mL)	Bias	CV	Mean (µg/mL)	Bias	CV	Mean (µg/mL)	Bias	CV
Blank (endogenous)	11.4	/	8%	2.59		20%	4.37	/	13%	LLOQ	/	/
Spike (+2.5 ng/mL)	/	/	/	5.32	9.1%	10.0%	7.13	10.0%	10.0%	1.93	23%	17%
Spike (+12.5ng/mL)	24	-0.3%	1.5%	14.8	2.2%	2.1%	17.3	-3.7%	9.2%	12.0	8.8%	3.3%

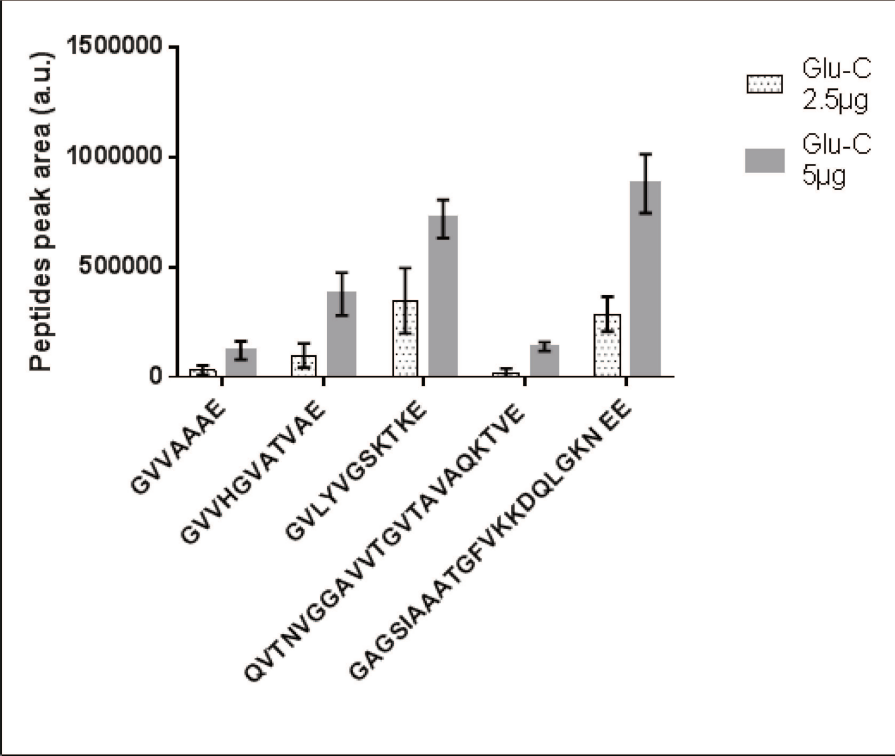
Human CSF was spiked with a mix of  $\alpha$ -synuclein and tau (low/high concentrations) and back calculated against a calibration curve prepared in artificial CSF. Endogenous  $\alpha$ -synuclein and tau concentrations were added to the spiked amount for calculations of bias. Bias and CV% are reported

Table 6: Proportion (%) of CSF sample above LLOQ for each peptide of  $\alpha$ -Synuclein (21 CSF samples) and tau (23 CSF samples), range of measured concentrations.

Peptide			
$\alpha$ -Synuclein	Detected	>LLOQ	Range (ng/mL)
GVVAAAE	100	100	2.7-18.5
GVLVVGSK	100	100	1.53-13.7
GVVHGVATVAE	100	100	1.05-4.12
QVTNVGGAVVTGVTAVAQK	100	90	0.40-2.83
GAGSIAAATGFVK	100	100	0.52-3.40
GAPQEGILE	100	100	2.65-22.75
GYQDYEPEA	95	95	1.17-7.43
Tau			
TPPAPK	100	100	5.12-46.9
SGYSSPGSPGTPGSR	100	100	3.20-48.7
TPSLTPPTR	100	100	7.21-45.8
SPVVSGDTSPR	100	70	0.75-4.90

Figure 1: Optimization of sequential Glu-C/trypsin digestion: Impact of Glu-C quantity a) and temperature of incubation b) on representative peptide signal (n=2 in each conditions)

a)



b)

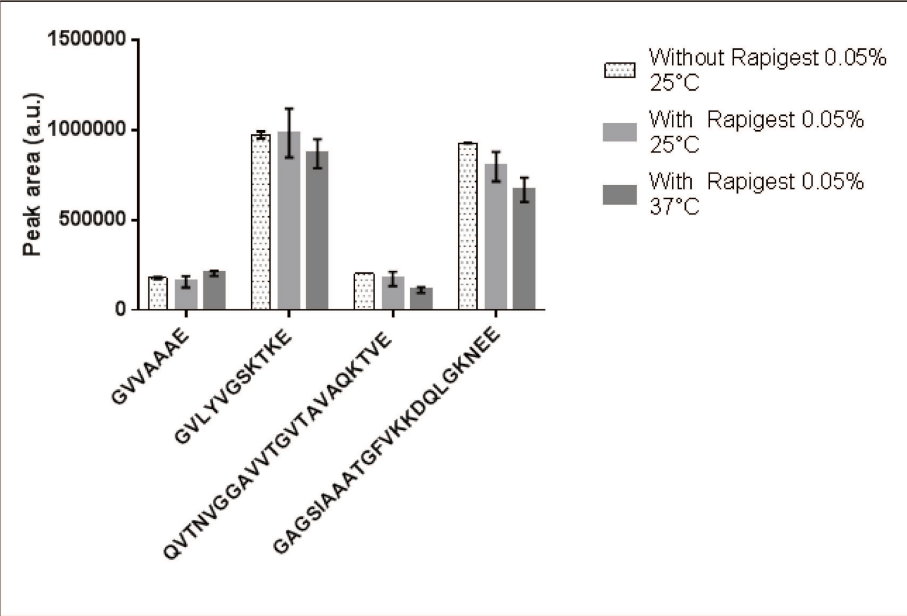


Figure 2: Impact of the sequential Glu-C/trypsin protocol on the release and/or detection of  $\alpha$ -synuclein peptides. Mean peak area of  $\alpha$ -synuclein peptides (n = 2) is illustrated.

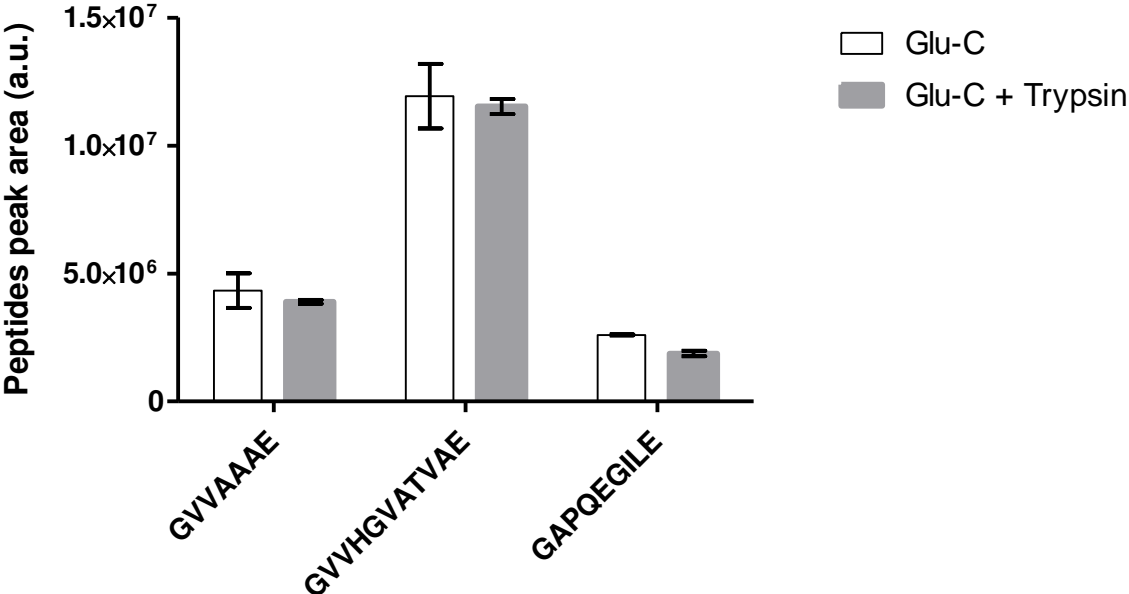


Figure 3: Extracted ion chromatogram (XIC) of the 14 peptides from tau,  $\alpha$ -synuclein and hemoglobin in CSF

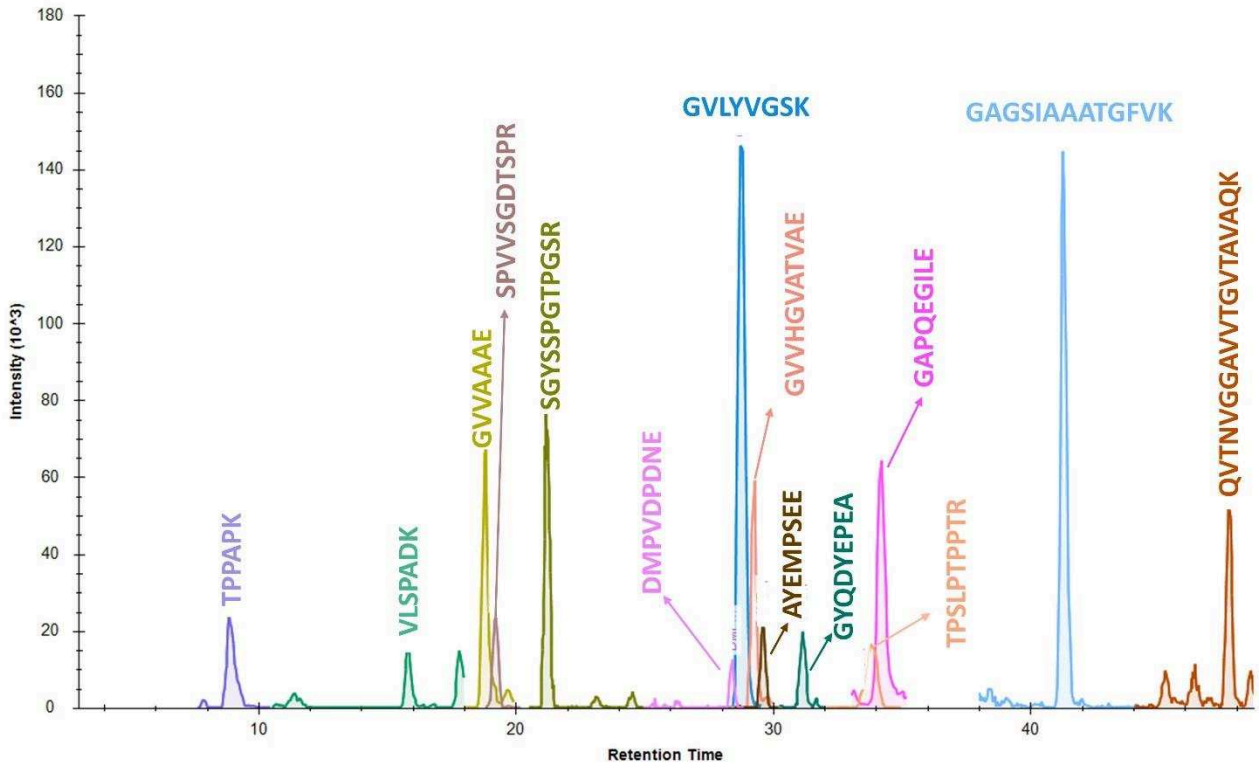


Figure 4: Workflow for the simultaneous quantification of tau and  $\alpha$ -synuclein in CSF

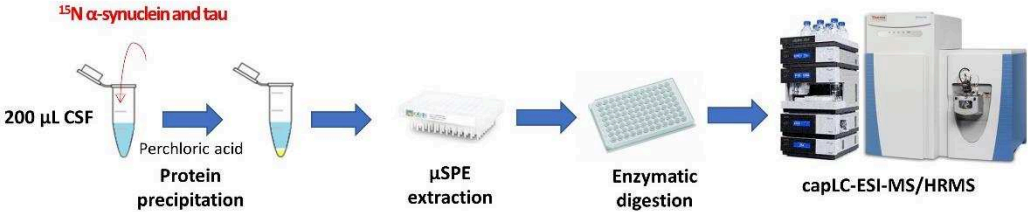
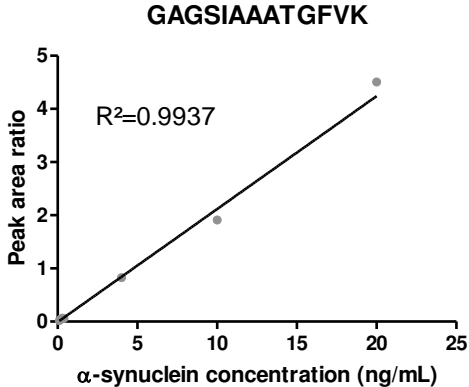
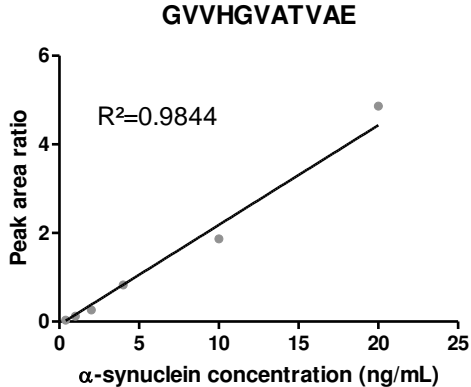


Figure 5: Linearity of  $\alpha$ -Synuclein and tau measurements. Calibration curve were established in artificial CSF. Representative peptides of  $\alpha$ -Synuclein a) and tau b) are illustrated

a)



b)

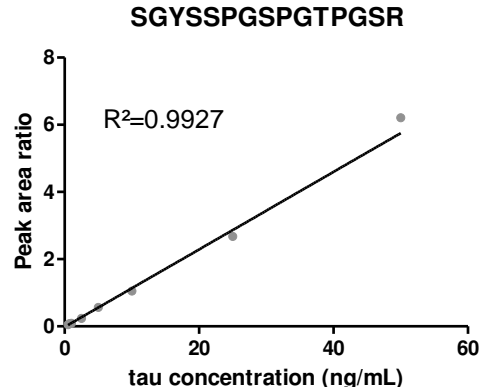
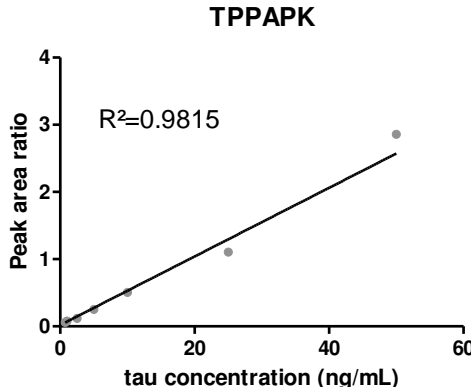




Figure 6: Linearity of tau (peptide TPSLPTPPTTR) in artificial CSF by overnight digestion and trypsin only digestion (sample preparation as in Barthelemy N, J.Prot Res. 2016). LLOQ was determined at 0.5 ng/mL

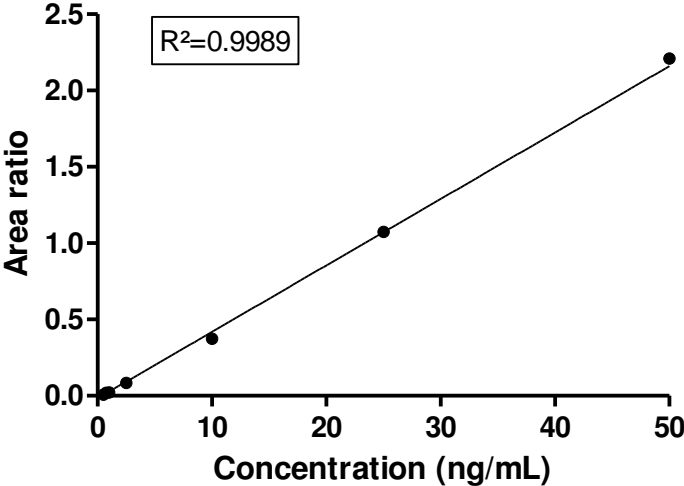


Figure 7: Evaluation of blood contamination based on signal intensity of peptide VLSPADK of hemoglobin.

CSF sample was considered as contaminated when signal of peptide VLSPADK exceed the highest intensity observed in the calibration curve prepared in artificial CSF. Samples N°9 and 19 were excluded.

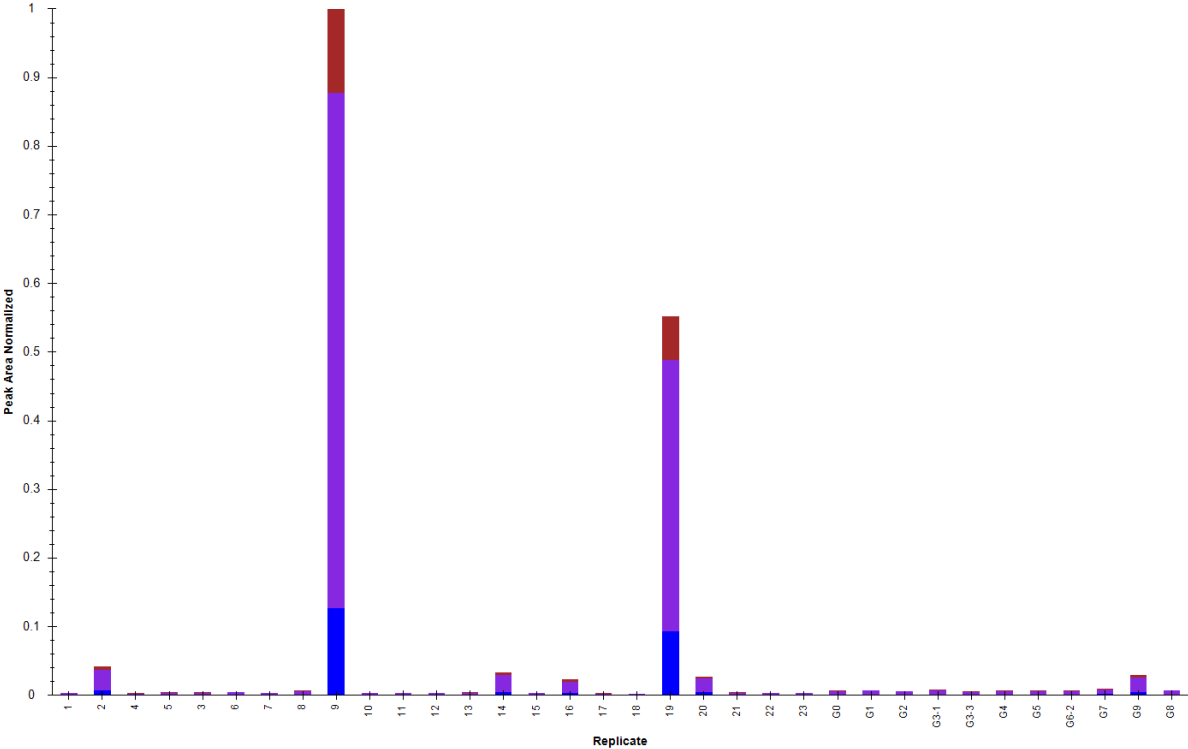


Figure 8:  $\alpha$ -Synuclein peptides abundance along the protein sequence in human CSF samples

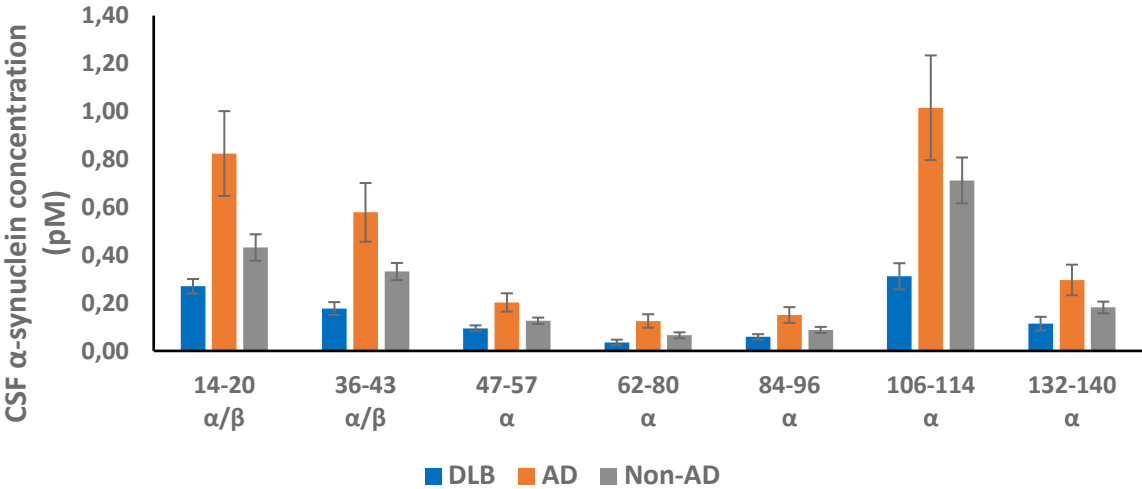
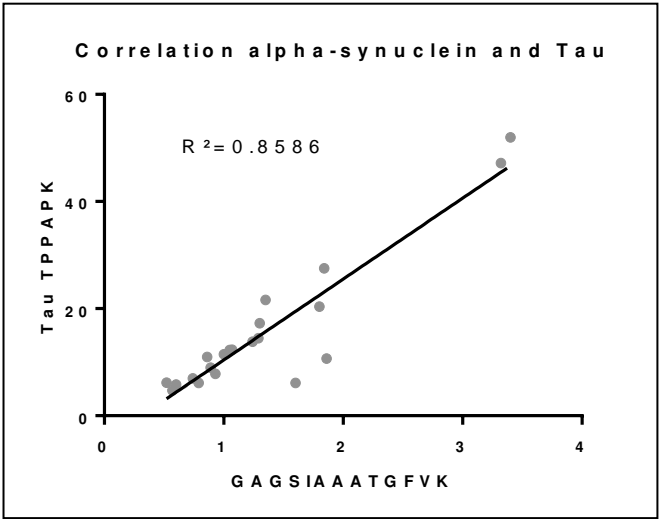
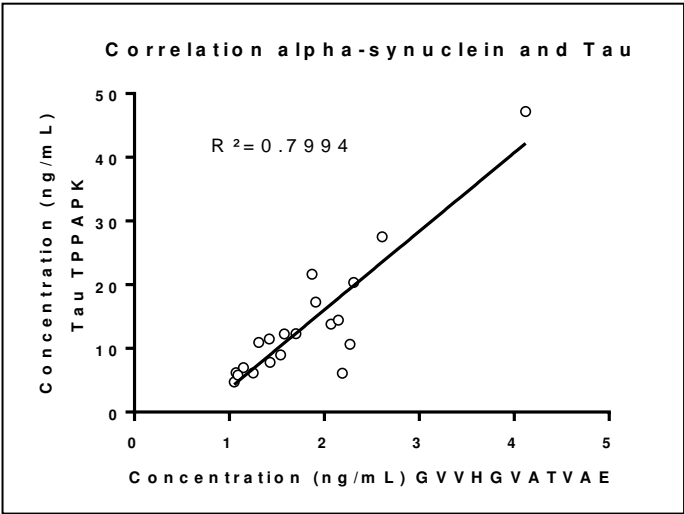
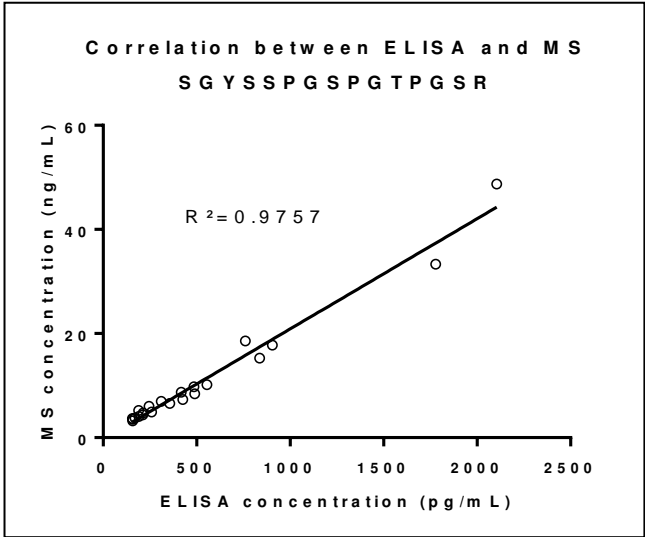
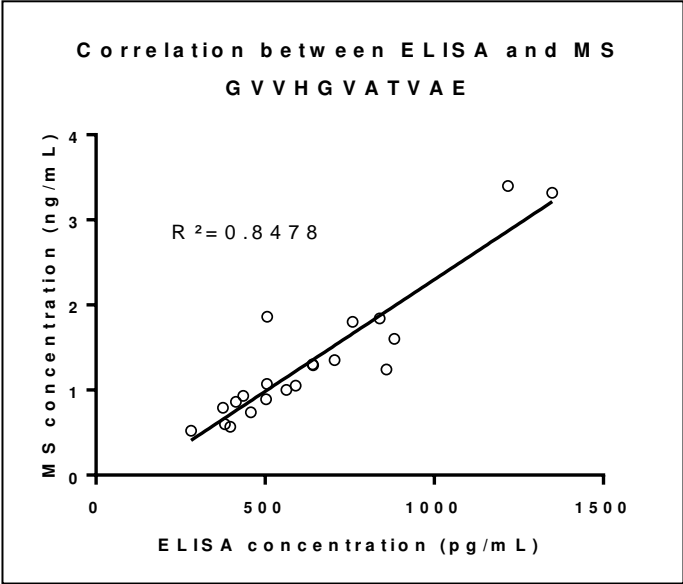


Figure 9: Correlation between Elisa and mass spectrometry measurements for  $\alpha$ -Synuclein a), tau b) and between  $\alpha$ -Synuclein and tau mass spectrometry determination c)



c)

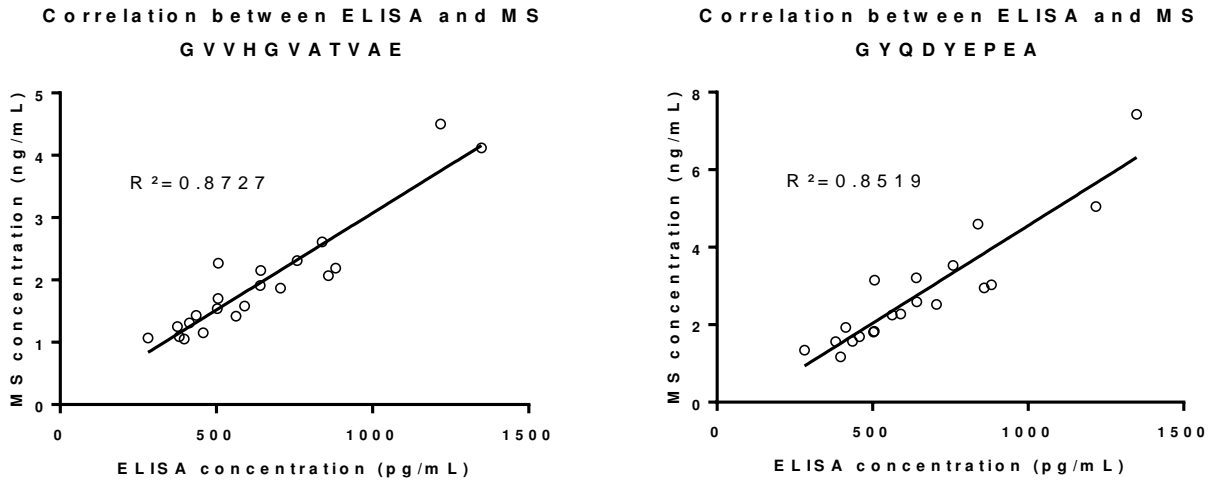
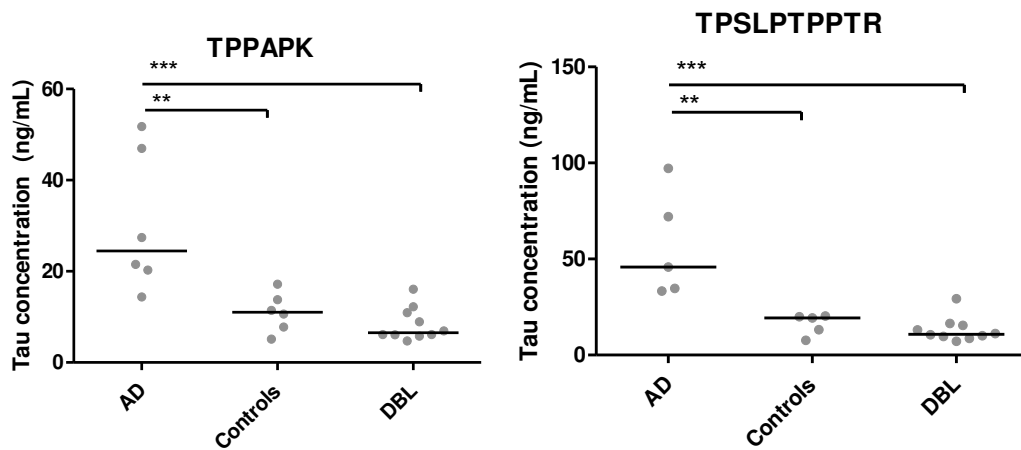
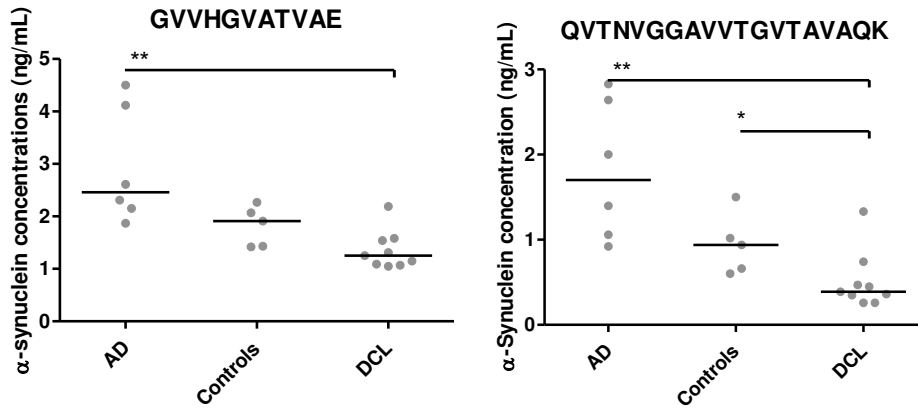


Figure 10: Boxplot of representative tau a) and  $\alpha$ -synuclein b) peptide levels in the AD, DLB and controls subgroups. \* $P \leq 0.05$  and \*\* $P \leq 0.01$

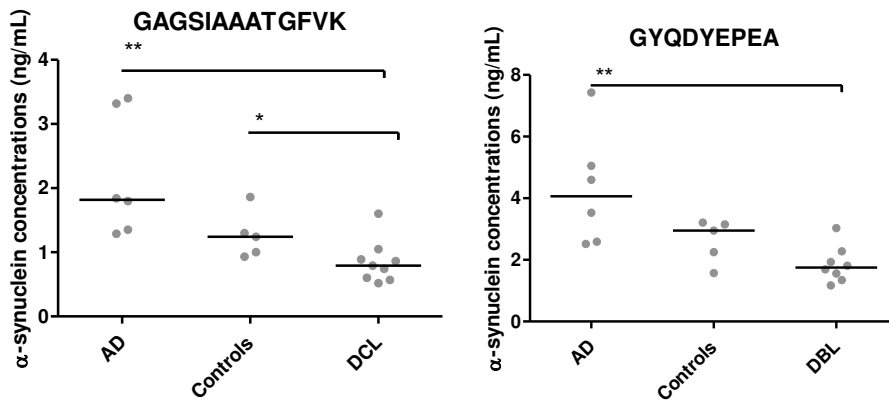
a)



b)



c)



Conclusion article :

L'alpha-synucléine et la protéine tau sont deux protéines impliquées dans la maladie d'Alzheimer ainsi que dans les synucléinopathies. Ces deux protéines interagissent<sup>461</sup> et peuvent notamment favoriser l'agrégation de l'autre<sup>420,425,462</sup>. Nous avons développé une méthode de quantification multiplexe de l'alpha-synucléine et de tau dans le LCR. La méthode proposée est une extension de la publication de Barthélemy et al. En effet, l'extraction de la synucléine se fait par précipitation des protéines par l'acide perchlorique, l'alpha-synucléine et tau restent solubles car ce sont deux protéines intrinsèquement désordonnées. Avec l'idée de couvrir au mieux la séquence de l'alpha-synucléine, le protocole de digestion à la trypsine a été adapté pour la quantification de l'alpha-synucléine, en combinant la glu-C puis la trypsine. En effet pour l'alpha-synucléine, une digestion à la trypsine génère un peptide C-terminale (102-140) long de 38 acides aminés (4.3 kDa) ne contenant pas de lysine. Ce peptide n'a pas de bonnes propriétés physico-chimiques pour son analyse par LC-MS ayant notamment de nombreux acides aminés chargés négativement. Il n'a pas été détecté dans la méthode proposée par Oeckl et al.<sup>452</sup> malgré des limites de quantifications inférieure à 100 pg/mL pour les meilleurs peptides. Afin de détecter les peptides présents dans la partie C-terminale de l'alpha-synucléine Yang et al.<sup>450</sup>, ont développé une méthode combinant une digestion à la trypsine et à l'Asp-N. Cette combinaison d'enzyme a permis une détection des peptides jusqu'à l'acide aminé 114 et donc ne couvrant toujours pas de manière satisfaisante la partie C-terminale de l'alpha-synucléine. Au regard de ces résultats nous avons sélectionnée la Glu-C<sup>314</sup> comme enzyme pour digérer l'alpha-synucléine. En effet, cette enzyme génère quatre peptides au lieu d'un seul avec la trypsine. Afin de profiter de la méthode publiée pour tau nous avons combiné la digestion à la Glu-C avec une digestion à la trypsine et sélectionné 4 des meilleurs peptides tryptiques qui ne sont pas impactés par la Glu-C pour la protéine tau. Après optimisation des conditions de digestion nous avons obtenu une couverture de la séquence de la synucléine à la LLOQ de 54%. Cependant, l'inconvénient de la combinaison des deux enzymes est que deux peptides deviennent alors communs avec la beta-synucléine, ce qui diminue la couverture de séquence pour l'alpha-synucléine. La méthode a été validée en termes de répétabilité et d'effets matrices. Enfin, la méthode a été appliquée à la quantification de 23 LCR. Les résultats obtenus sont encourageant pour la discrimination des 3 groupes (MA, DCL et témoins). De plus, nous avons observé une corrélation entre tau et l'alpha-synucléine, comme précédemment rapporté<sup>430</sup>. La méthode permet la quantification de l'extrémité C-terminale de l'alpha-synucléine (132-140) qui ne semble pas être diminuée par rapport aux peptides centraux comme ce qui a pu être observé précédemment pour la protéine tau. En revanche, deux peptides de la partie C-terminale ne sont

pas détectés (DMPVDPDNE et AYEMPSEE) du fait d'un manque de sensibilité de notre méthode envers ces peptides. Ce qui est décevant car le peptide AYEMPSEE contient le site de phosphorylation S129. A noter d'une part une plus faible abondance des peptides centraux QVT... et GAG..., potentiellement liée à des modifications post-traductionnelles sur ces peptides (troncations, phosphorylation, etc) et une augmentation du peptide présent sur la partie C-terminale GAPQEGILE. Ce dernier peut être expliqué une digestion variable produisant GAPQEGILE et GILE. Nous pouvons tout de même souligner que Yang et al. ont aussi observé une augmentation pour ce peptide NEEGAPQEGILE. Néanmoins ce peptide était le moins sensible et celui possédant le CV le plus élevé.

En résumé, nous avons réussi à développer une méthode de quantification multiplexe de tau et de l'alpha-synucléine dans le LCR. Néanmoins, notre approche par bottom-up est limitée pour la caractérisation des protéoformes de l'alpha-synucléine dans le LCR, alors que leur quantification pourrait être plus spécifique pour la discrimination des maladies neurodégénératives. En effet, nous obtenons une couverture de séquence uniquement de 54% et les modifications d'abondances des peptides ne permettent qu'une quantification indirecte, par différence, des formes modifiées. D'autre part, il serait préférable de détecter directement les modifications post-traductionnelles afin de les identifier. Afin de mieux caractériser ces protéoformes nous avons développé une approche top-down pour l'analyse de l'alpha-synucléine dans le cerveau, approche dont les avantages ont été présentés dans la partie bibliographique. Cette méthode top-down a été appliquée à une cohorte de cerveaux humains. Ces résultats pourront servir à cibler des formes tronquées de l'alpha-synucléine dans le LCR.



## IV. Développement d'une méthode intact/top-down pour la caractérisation de l'alpha-synucléine

### 1. Optimisation des paramètres MS

Pour les approches top-down une attention particulière doit être portée aux paramètres de sources et instrumentaux. En effet, l'identification de protéoformes peu abondants dépend de la qualité du spectre de masse (signal sur bruit)<sup>463</sup>. Les paramètres de sources que nous avons optimisés sont le « In-source collision-induced dissociation » (icid), la température du gaz auxiliaire et la température du capillaire de transfert (voir figure 74). Les paramètres ont été optimisés à un débit de 500  $\mu\text{L}/\text{min}$ . Le icid améliore la désolvatation et augmente le signal sur bruit mais si la tension appliquée est trop forte l'ion peut fragmenter comme c'est le cas pour l'alpha-synucléine ( $z=15$ ) à partir de 20V (voir figure 74). L'optimal a été trouvé à 10V. La température du gaz auxiliaire améliore l'efficacité d'évaporation des gouttelettes formés par la source electrospray<sup>464</sup>, l'optimal a été trouvé à 400°C. Enfin, une augmentation de la température du capillaire permet d'augmenter la désolvatation des gouttelettes<sup>465</sup>, nous avons sélectionné une température de 320°C.

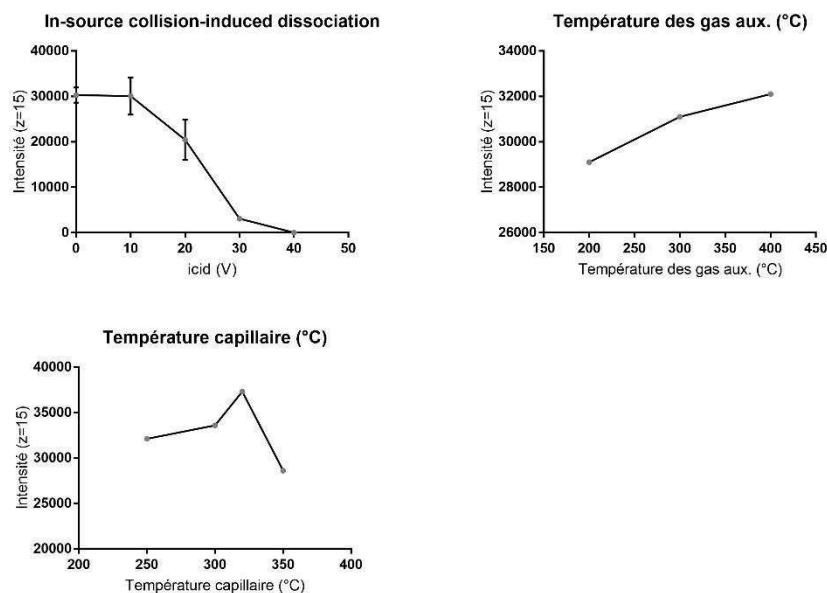


Figure 74 Optimisation des conditions de source: icid, température des gaz aux. et température du capillaire de transfert

Enfin, le dernier paramètres que nous avons optimisés est le nombre de microscans. Un microscan correspond à un scan par l'orbitrap et l'augmentation du nombre de microscans permet de moyenniser le spectre et d'améliorer le signal sur bruit. Bien qu'avantageux pour

améliorer la qualité du spectre de masse notamment pour les ions peu abondants, le temps d'analyse par l'orbitrap est multiplié par le nombre de microscans (voir figure 75).

Ces paramètres sont dépendants des protéines et donc bien qu'optimale pour l'alpha-synucléine 1-140, ils ne seront pas nécessairement optimaux pour ses protéoformes.

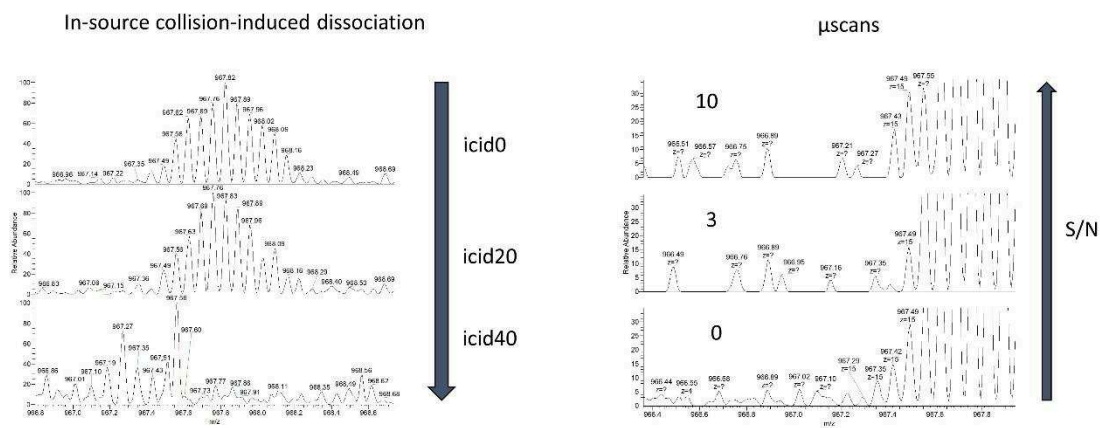


Figure 75 Impact de l'icid et du nombre de μscans sur la qualité du spectre

Une fois les paramètres MS optimisés, nous avons évalué la sensibilité des approches intactes/top-down pour la quantification de l'alpha-synucléine dans le LCR.

## 2. Développement d'une méthode pour la quantification en LCR

La méthode de quantification développée par bottom-up pour l'alpha-synucléine n'est pas satisfaisante en termes de couverture de séquence, notamment car 2 peptides sont communs avec la bêta-synucléine et deux peptides de la partie C-terminale ne sont pas détectés. Nous avons donc cherché à évaluer la sensibilité des approches intactes/top-down pour la quantification de l'alpha-synucléine dans le LCR. Ces approches permettront d'avoir une vision plus globale des protéoformes présents dans le LCR. Afin d'évaluer les approches par top-down pour la quantification de l'alpha-synucléine dans le LCR, nous avons appliqué la même préparation d'échantillon que pour la méthode bottom-up (précipitation à l'acide perchlorique et μSPE HLB) en ajoutant de la BSA à 1 mg/mL dans les échantillons pour prévenir l'adsorption non spécifique de l'alpha-synucléine sur les tubes eppendorf. Nous avons aussi opté pour une colonne C4 avec des pores de 300Å au regard de la taille de l'alpha-synucléine<sup>175</sup>. Une gamme d'alpha-synucléine

recombinante dans du LCR mime (plasma dilué au 200<sup>ème</sup> cf. alpha-synucléine bottom-up) a été réalisée avec un standard interne l'alpha-synucléine <sup>15</sup>N recombinante. Les concentrations testées dans un premier temps vont de 5 ng/mL à 10 µg/mL. Cette gamme exploratoire ne descend pas aux concentrations endogènes de l'alpha-synucléine. Dans le but d'évaluer la répétabilité de ces approches 2 niveau de QC (n= 3) ont été analysés (200 ng/mL et 2 000 ng/mL). Trois modes d'acquisition ont été comparés : deux ciblés (target-SIM et PRM) et un non ciblé (Full-MS) (voir figure 76). Pour le mode SIM et PRM l'état de charge le plus intense a été ciblée. Pour le mode SIM la quantification a été faite à partir des isotopes représentant au minimum 30% de l'isotope maximum. Pour le mode PRM, les fragments les plus intenses ont été sommés après optimisation de l'énergie de collision.

Comme attendu le mode target-SIM est le plus sensible des trois modes avec une LOQ à 50 ng/mL soit 4 fois inférieure au mode PRM et 10 fois inférieure au mode Full-MS (voir tableau 21). Le mode PRM et target-SIM ont des performances similaires en termes de fidélité et biais. Néanmoins, bien que moins sensible le mode PRM pourrait se révéler avantageux en matrices biologiques plus complexes.

La limite de quantification du mode d'acquisition le plus sensible est de 50 ng/mL dans nos conditions. Cette limite de quantification est environ 50 fois trop élevée pour la quantification de l'alpha-synucléine dans le LCR dont la concentration est proche du ng/mL. Nous n'avons donc pas poursuivi cette approche pour la quantification dans le LCR. Cette méthode pourrait être améliorée en utilisant une colonne chromatographique nano. En effet, Mohr et al. ont montré que dans leurs conditions une diminution du diamètre de la colonne d'un facteur 10 ( 1mm à 0.1mm) permettait d'avoir un gain de sensibilité de 2.5 à 25 qui est dépendant de la protéine analysée<sup>466</sup>. Le gain apporté par une diminution du diamètre de la colonne ne pourrait pas être suffisant. Nous pourrions alors envisager une immunoprécipitation couplant plusieurs anticorps ayant des épitopes différents afin de capturer la diversité des formes de l'alpha-synucléine comme proposé par Rogeberg et al.<sup>103</sup> pour les peptides amyloïde-béta. Par exemple nous pourrions sélectionner les anticorps suivants : un N-terminal (N-19), un central (FL 140) et un C-terminal (Syn5d). Une autre solution envisageable serait de réduire la distribution d'états de charge de la protéine en utilisant par exemple des réactif permettant d'augmenter le nombre moyen de charges <sup>467</sup>.

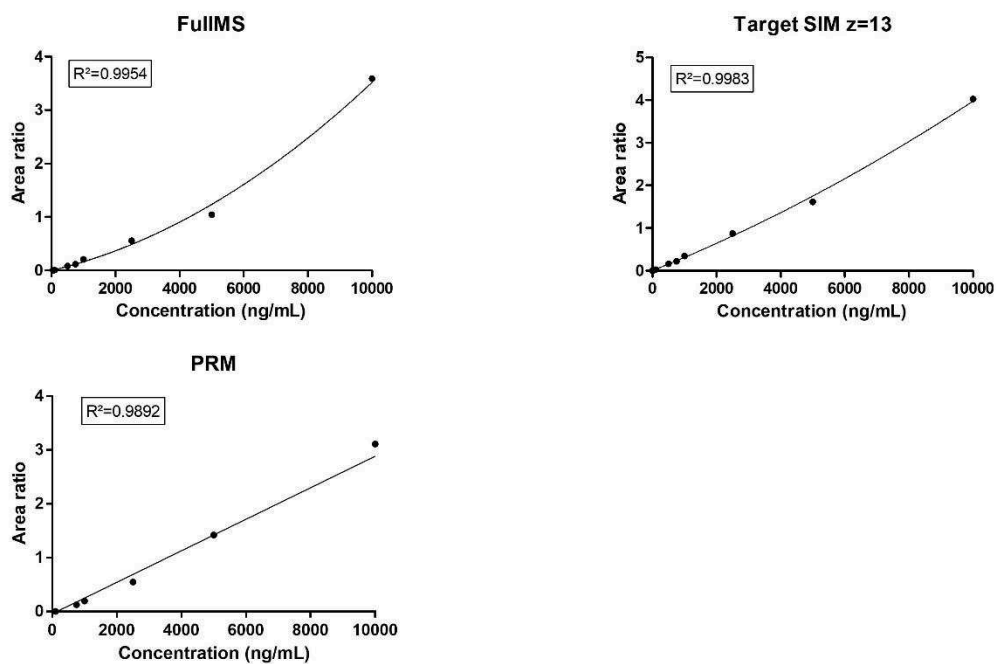


Figure 76 Droites de calibration obtenues pour les trois modes d'acquisition :FullIMS, target-SIM et PRM

Tableau 21 Performances analytiques des modes d'acquisition pour la quantification de l'alpha-synucléine en plasma dilué au 200ème

Acquisition	LOQ	QC bas 200ng/mL (n=3)		QC haut 2000 ng/mL (n=3)	
		Fidélité	Biais	Fidélité	Biais
Full-MS	500 ng/mL	9%	28%	6.3%	1%
SIM	50 ng/mL	18%	11%	0.4%	13%
PRM	200 ng/mL	3%	12%	9.5%	13%

### **3. Développement d'une méthode de quantifications des troncations de l'alpha-synucléine dans le cerveau**

Les approches par top-down/intactes n'étant pas directement applicables dans le LCR, nous avons développé ces approches pour la caractérisation et la quantification des troncations de l'alpha-synucléine dans le cerveau car l'alpha-synucléine y est abondante et le cerveau est plus proche de la physiopathologie que le LCR. Le but étant de mettre en évidence des formes étant impactées dans les synucléinopathies (ici nous nous concentrerons sur la MSA, suite à une collaboration avec l'équipe de Benjamin DEHAY et Pierre- Olivier FERNAGUT de Institut des Maladies Neurodégénératives, Bordeaux). De plus ces protéoformes pourront être recherchées dans le LCR (à condition d'avoir un peptide protéotypique). Nous nous sommes focalisés sur les formes tronquées dans cette première étude car les corps de Lewy sont constitués à environ 15% de formes tronquées et ces troncations peuvent promouvoir l'agrégation de l'alpha-synucléine.

Les approches top-down/intactes pour la quantification des protéoformes se font généralement sans standards internes, ce qui peut induire un biais causé par les effets matrices et les étapes de préparation d'échantillon. La solution à ce problème que nous proposons ici est la génération de formes tronquées à partir du standard (1-140) <sup>15</sup>N de l'alpha-synucléine (cf. III.) par des protéases connues pour cliver l'alpha-synucléine.

#### **a) Génération de standards marqués tronqués**

##### **(i) Sélection des protéases**

Il a été observé une dérégulation de la dégradation de l'alpha-synucléine dans les synucléinopathies<sup>468</sup>. En effet, Webb et al.<sup>469</sup> et Wong et al.<sup>470</sup>, ont montré un dysfonctionnement des lysosomes et du protéasome. Les lysosomes s'occupent de la dégradation des formes formant les fibrilles amyloïdes ainsi que de la régulation de la concentration de l'alpha-synucléine<sup>471</sup>. Le protéasome lui régule l'alpha-synucléine soluble. Nous avons sélectionné différentes classes de protéases (lysosomale, métalloprotéinase matricielle, dépendante du calcium et liée à la mort cellulaire) Nous avons sélectionné des protéases pour lesquels un lien a été établi avec l'alpha-synucléine : capacité à dégrader la synucléine, co-localisation dans les corps de Lewy et/ou génération de forme clivées en C-terminale potentiellement toxiques.

- Cathepsines D et L :

Les cathepsines sont responsables de la dégradation de l'alpha-synucléine dans les lysosomes<sup>471</sup>. Par exemple, la cathepsine D dégrade l'alpha-synucléine qui est liée à la

membrane alors que la cathepsine L dégrade la forme monomérique de l'alpha-synucléine ainsi que les fibrilles<sup>471,472</sup>. La cathepsin D aurait un rôle protecteur<sup>473,474</sup> puisqu'une perte de fonction de la cathepsin-D conduirait à un dysfonctionnement des lysosomes et à la propagation des agrégats de l'alpha-synucléine. Nous avons sélectionné ces deux cathepsines, la cathepsine D qui est une protéase aspartique et la cathepsine L qui est une protéase à cystéine<sup>475</sup>.

- MMP-3 :

La MMP-3 est activée dans les cellules dopaminergiques quand un stress leur est appliquée<sup>476</sup>. Cette enzyme a été montrée comme colocalisée avec l'alpha-synucléine dans les corps de Lewy<sup>438</sup>. La MMP-3 forme des fragments C-ter de l'alpha-synucléine qui seraient les fragments favorisant son agrégation<sup>477</sup>. La capacité de la MMP-3 à cliver l'alpha-synucléine dans les cellules dopaminergiques pourraient jouer un rôle crucial dans le développement des corps de Lewy<sup>438</sup>.

- Calpaïne-1 :

La calpaïne-1 est une protéase intracellulaire dépendant du calcium. Elle produit aussi des fragments favorisant l'agrégation<sup>478</sup>. Une inhibition de l'activité de la calpaïne-1 par un inhibiteur spécifique induit, dans un modèle murin, une réduction de l'agrégation de l'alpha-synucléine<sup>479</sup>. L'immunothérapie semble aussi prometteuse<sup>415</sup>. Il a été montré que l'activité de la calpaïne-1 est augmentée dans le putamen des personnes atteintes de MSA<sup>480</sup>.

- Caspase-1 :

La caspase-1 peut être activée par un signal inflammatoire<sup>481</sup>. Wang et al.<sup>383,414</sup> et Bassil et al., ont mis en évidence le lien entre les tronctions C-terminal de l'alpha-synucléine causée par la caspase-1 et la toxicité de ces formes *in vivo* et *in vitro*. Ils ont notamment montré qu'inhiber le clivage de l'alpha-synucléine par le VX-765 permettait de mitiger la neurodégénérescence.

Ces protéases vont servir à générer individuellement des standards <sup>15</sup>N de formes tronquées de l'alpha-synucléine.

### **(ii) Optimisation des conditions d'incubation**

Premièrement la pureté des enzymes a été évaluée par digestion à la trypsine et analyse par Proteome Discoverer 1.4, car les cathepsines D/L et la calpaïne-1 sont des protéines purifiées. Nous avons observé de la cathepsine-B avec la cathepsine-L et des unités du protéasome

présentes avec la calpaine-1, donc une possibilité de coupure non lié uniquement à l'enzyme sélectionnée. Ensuite, la quantité d'enzymes et le temps d'incubation de chaque enzyme ont été optimisés afin de générer des fragments variés et de tailles majoritairement supérieures à 5 kDa afin d'avoir des formes, en taille, proche de celles observées par Kellie et al. Les tronctions principales obtenues dans ces conditions sont présentées figure 77. Les tronctions ont été attribuées selon leur masse et leur spectre MS/MS. Des exemples d'attribution sont présentés figure 77. Pour la caspase-1, nous n'avons pas réussi à produire des formes tronquées majoritaires (<10%) même après 20h d'incubation.

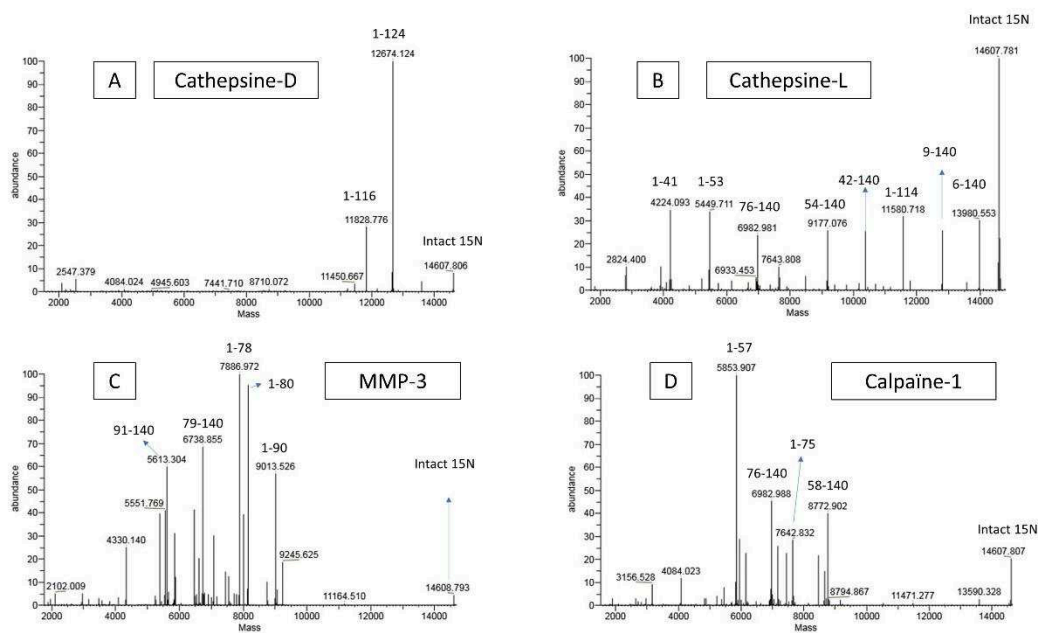


Figure 77 Spectres de masse déconvolué représentant les tronctions générées dans pour les conditions d'incubation optimale pour chaque protéase

### (iii) Identification des fragments

Nous avons dans un premier temps identifié/ caractérisé les tronctions de l'alpha-synucléine dans des cerveaux de souris modèle de la MSA sur-exprimant l'alpha-synucléine humaine. Dans une deuxième étape, nous avons ciblé et quantifié les formes identifiées dans des tissus cérébraux humains précieux. Pour ce faire la stratégie a été en deux étapes. Premièrement, les standards de formes tronquées nous ont permis d'obtenir un temps de rétention et l'état de charge majoritaire de l'ion, ce qui nous a aidé à identifier les formes en matrices biologiques. Néanmoins, l'alpha-synucléine endogène est acétylée en N-terminale contrairement à notre standard <sup>15</sup>N, ce qui induit

un décalage en temps de rétention et rend l'identification plus compliquée. Deuxièmement, nous avons comparé les spectres MS obtenus pour les souris w.t. avec les souris sur-exprimant l'alpha-synucléine humaine pour identifier les signaux provenant uniquement de l'alpha-synucléine humaine. Cette stratégie nous a permis d'identifier 20 formes tronquées de l'alpha-synucléine dans les modèles de souris dont des exemples sont présentés figures 78, 79 et 80. A partir de ces identifications et des formes tronquées des standards nous avons établi une méthode PRM et SIM pour la quantification. Les standards tronqués 15N nous ont donc aidé à identifier les formes tronquées dans les cerveaux des souris modèle, à quantifier les formes tronquées dans les cerveaux humains, et également à évaluer notre protocole comme indiqué ci-après.

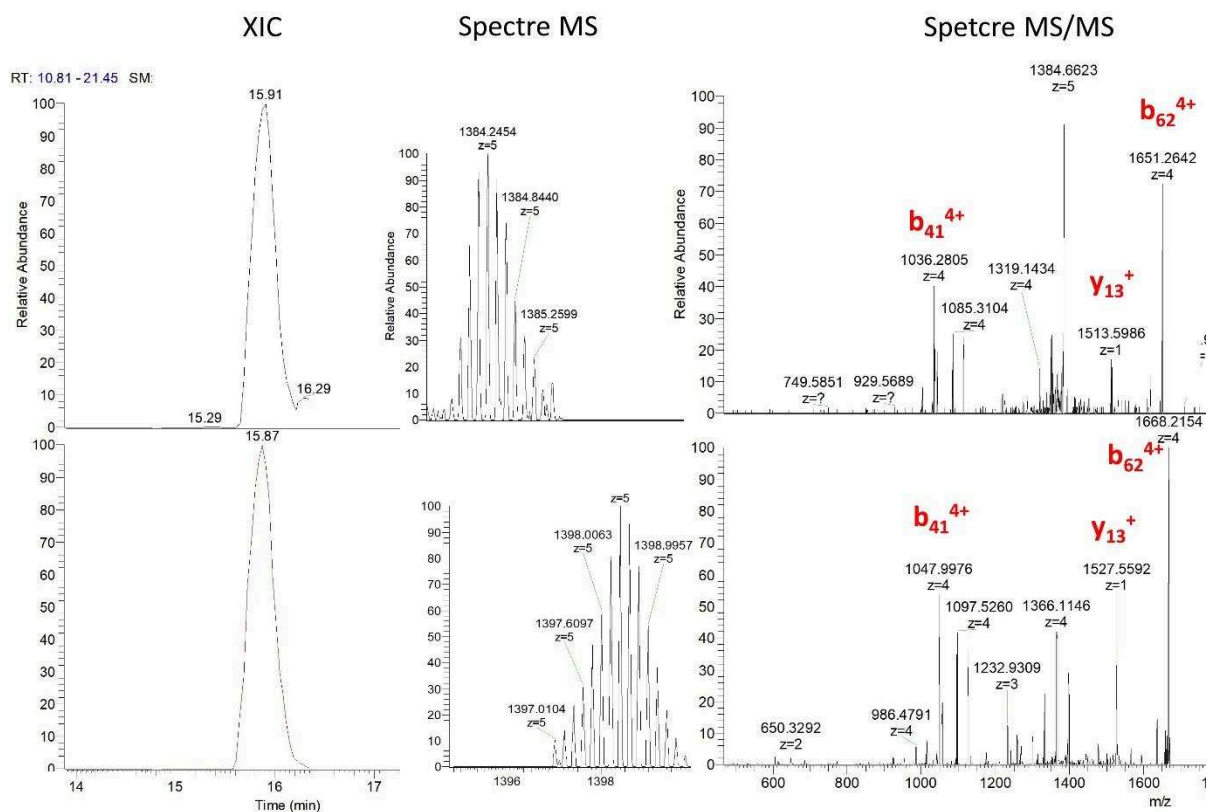


Figure 78 Exemple d'attribution basé sur le temps de rétention, le spectre MS et le spectres MS/MS pour la forme 76-140. Les spectres de masses sont attribués avec une précision de masse <5ppm. Les principaux fragments ont été attribués.



### XIC

### Spectre MS

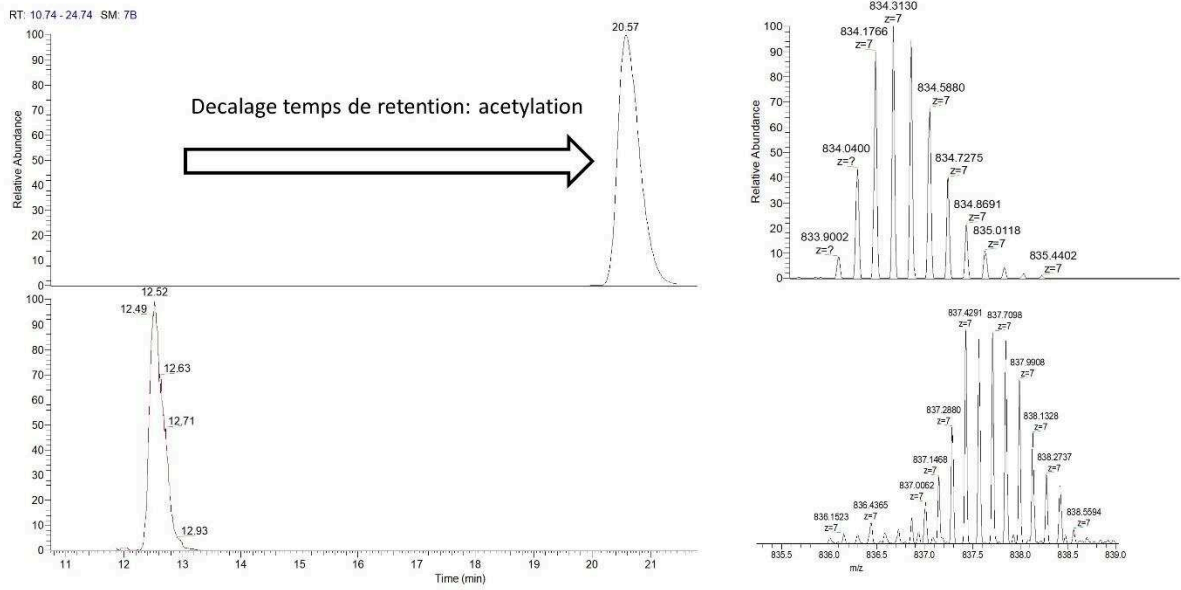


Figure 79 illustration du décalage en temps de rétention pour la troncation 1-57

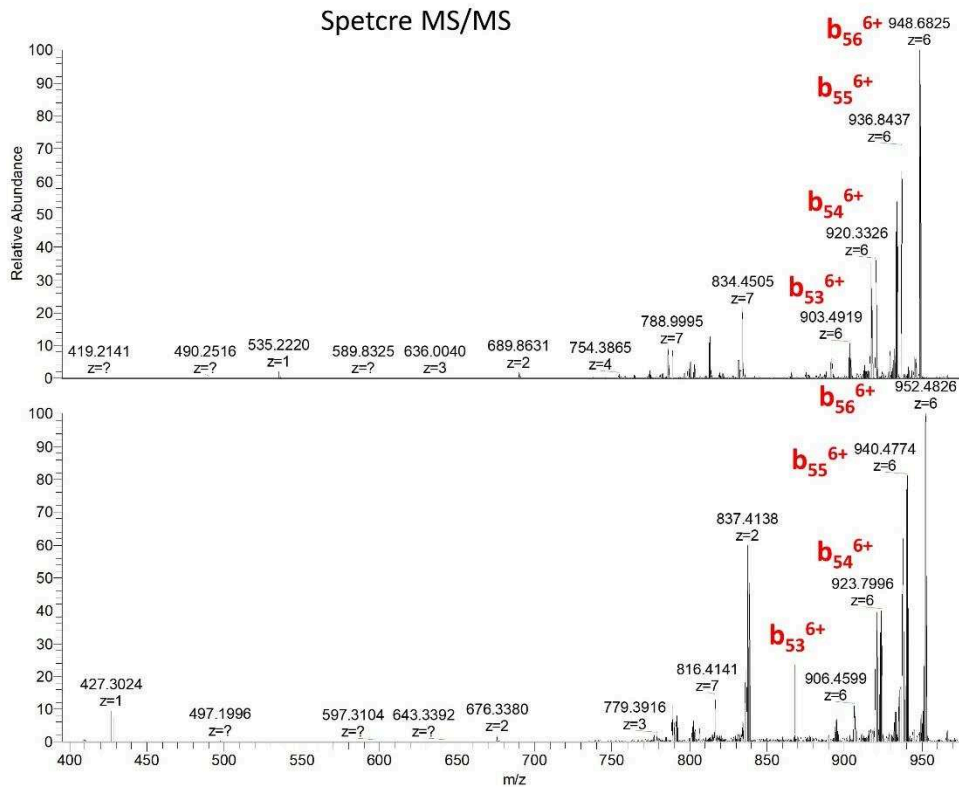


Figure 80 Spectres MS/MS de la forme 1-57 montrant une dominance du spectre par les ions b pouvant rendre l'interprétation erronée. En effet, l'extrémité C-terminale n'est alors pas couverte.

**b) Evaluation du protocole pour la quantification des formes tronquées en cerveau (souris) et comparaison des modes d'acquisition ciblée pour la quantification (SIM/PRM)**

Afin d'évaluer notre méthode nous avons mesurer le rendement d'extraction total (étape de précipitation et SPE) de notre méthode selon une stratégie similaire à celle développé pour le bottom-up. Le rendement moyen est de 57% avec un minimum à 22% pour la forme 1-124 et un maximum pour la forme 94-140. Il semble que notre protocole a un meilleur rendement envers les formes tronquées courtes montrant l'intérêt de normaliser le signal avec un standard interne.

Tableau 22 Rendement de SPE et de précipitation pour les formes tronquées de l'alpha-synucléine

Troncations C-terminale	Rendement total	Troncations N-terminale	Rendement total
1-27	60%	94-140	108%
1-31	76%	91-140	76%
1-35	97%	87-140	82%
1-38	59%	84-140	80%
1-40	72%	79-140	51%
1-41	70%	76-140	47%
1-50	48%	74-140	80%
1-53	59%	68-140	59%
1-57	85%	66-140	33%
1-60	75%	58-140	72%
1-64	45%	54-140	54%
1-65	56%	51-140	32%
1-67	63%	42-140	46%
1-69	54%	18-140	35%
1-73	60%	6-140	30%
1-78	53%		
1-79	52%		
1-80	50%		
1-83	56%		
1-86	67%		
1-90	38%		
1-103	27%		
1-124	22%		
1-116	33%		
1-119	28%		

Enfin, la linéarité, la sensibilité, la répétabilité des modes PRM et t-SIM a été évaluée par dilution des formes 15N tronquée de 1X à 100X (voir tableau 22) dans un homogénat de cerveau de souris, les formes endogènes sont utilisées comme standard internes. Malgré une matrice plus complexe que le LCR mime, le mode PRM à des limites de quantification proches de celles atteintes par le mode t-SIM pour les formes tronquées en N-terminale, 7 formes ayant une LOQ à 100X pour les deux modes. En revanche, due à une moins bonne fragmentation le mode PRM est moins sensible que le mode SIM pour la quantification des formes tronquées en C-terminale. En effet, le mode SIM permet d'atteindre une dilution d'un facteur 100 pour 10 formes alors que le mode PRM seulement pour 3 formes. D'autres méthodes de fragmentations pourraient être explorées (ETD, ECD ou UVPD) pour augmenter la sensibilité du mode PRM. Néanmoins, le CV et le biais moyen des QC est identique entre PRM et t-SIM : 14,6% et 9,8% contre 13,8% et 11,4% (voir tableau 23).

**Tableau 23 Comparaison des performances de quantification entre les modes SIM et PRM. LOQ est déterminé avec un biais de <20%**

Troncation	SIM				PRM			
	Coefficient de corrélation R <sup>2</sup>	LOQ (dilution factor)	QC		Coefficient de corrélation R <sup>2</sup>	LOQ (dilution factor)	QC	
			Fidélité	Biais			Fidélité	Biais
1-103	0,9834	10X	86,70%	45,80%				
1-116					0,9519	40X	8,30%	7%
1-119	0,9897	10X	16,30%	12,90%	0,9999	10X	55,90%	1%
1-124					0,9682	40X	3,60%	9%
1-140	0,9791	40X	16,70%	15,20%	0,9962	40X	8%	5%
1-31	0,9914	100X	6%	8%	0,9039	40X	40,40%	27%
1-38	0,9906	20X	11,62%	8,60%	0,9725	10X	15,40%	10%
1-39	0,9107	100X	25,80%	20,60%	0,9565	40X	17,30%	12%
1-41	0,9756	100X	12,70%	9,40%	0,9909	10X	12,10%	10%
1-50	0,994	20X	7,40%	6,10%	0,98	10X	6,20%	5%
1-57	0,9877	100X	5%	3,30%	0,9942	100X	14,90%	12%
1-67	0,9931	100X	8,90%	7,40%	0,993	10X	8,90%	6%
1-69	0,9975	100X	7,80%	6,50%				
1-73	0,9941	100X	4,60%	5,10%				
1-75	0,9862	100X	14,40%	30,20%				
18-140	0,9887	20X	16,20%	12,60%				
1-83	0,9969	100X	4%	5,90%	0,9898	20X	5,80%	6%
1-86	0,9962	20X	6,80%	6,20%				
1-90	0,9986	100X	1%	2,40%	0,9945	100X	5,10%	5%
42-140	0,9906	100X	16,50%	12,70%	0,9474	100X	24,20%	20%
51-140	0,9928	20X	4,30%	6,80%				
54-140	0,9418	100X	12,60%	9,10%	0,9727	100X	3,90%	8%
58-140	0,9911	100X	8,40%	7,20%	0,9702	100X	11,70%	10%
74-140	0,9781	100X	15,20%	12,30%	0,9904	100X	11,10%	9%
76-140	0,9733	100X	13,40%	10,70%	0,9783	100X	6,20%	5%
84-140	0,9871	100X	4%	3,20%	0,96	40X	13,90%	13%
91-140	0,9888	40X	13%	10%	0,9807	100X	12,60%	9%
94-140	0,9577	100X	18,40%	18,60%	0,9497	100X	20,50%	16%
<b>moyenne</b>			<b>13,8%</b>	<b>11,4%</b>			<b>14,6%</b>	<b>9,8%</b>

Après évaluation de notre méthode pour la quantification des formes tronquées de l'alpha-synucléine nous avons appliqué cette méthode à leur quantification en cerveau humain. Du fait, d'une plus faible quantité d'alpha-synucléine dans le cerveau humain en comparaison des cerveaux de souris, nous avons utilisé une quantité de cerveau 5 fois plus importante.

### c) Application à une cohorte

La cohorte est constituée de 7 échantillons de patients atteints de MSA et 6 témoins. Pour chaque échantillon, nous disposons de deux aires cérébrales le putamen et le cortex frontal (voir tableaux 24 et 25). Le putamen est une aire cérébrale fortement affecté alors que le cortex frontal n'est que peu affecté par la maladie<sup>414,482</sup>. Nous avons quantifié cette cohorte en utilisant le mode SIM et PRM. Nous avons pu quantifier 28 protéoformes avec le mode SIM et 21 avec le mode PRM. Ce résultat est sensiblement meilleur que celui obtenu par Kellie et al. (10 formes). Le mode PRM permet d'apporter une confirmation d'identification. Une première normalisation des formes tronquées a été faite à partir du standard <sup>15</sup>N tronquée correspondant ou celui le plus proche en temps de rétention quand la forme correspondante n'était pas détectée. En effet, le standard <sup>15</sup>N n'étant pas acétylé il y a un décalage des temps de rétention pour les formes tronquées en C-terminale. Enfin, le ratio d'aire a été rapporté en fonction de la quantité d'alpha-synucléine présente afin de pouvoir normaliser la quantité des formes tronquées. Du fait du faible nombre d'échantillons par groupe et de la variabilité analytique il n'est pas évident de mettre en évidence des changements significatifs entre les formes. Néanmoins, nous avons pu mettre en évidence des changements significatif par le mode SIM. En effet, 7 formes : 1-50, 1-60, 1-65, 54-140, 68-140, 91-140 et 94-140 sont significativement augmentées dans le putamen des MSA en comparaison avec les témoins. Cinq de ces formes sont aussi augmentées dans le cortex frontal. En revanche aucune de ces formes n'est impactées dans le cortex frontal suggérant une place centrale du putamen dans le développement de la MSA. En comparaison, avec le mode PRM uniquement une forme (51-140) est significativement impactée (augmentée) entre le groupe MSA et le groupe contrôle. Deux autres formes sont significativement augmentées (1-90 et 94-140) dans le putamen en comparaison du cortex frontal. Ce résultat est intéressant car l'alpha-synucléine est connue pour s'y agréger dans les MSA<sup>482</sup> et la forme 1-90 expose le NAC de l'alpha-synucléine favorisant la formation de fibrilles. La forme 1-90 augmentée par quantification n'est pas retrouvée comme significativement impactée par quantification par le mode SIM., une explication peut être le faible le nombre d'échantillon ou que l'augmentation observé par PRM est un artefact. Ces résultats bien que préliminaires semblent indiquer des différences de proportion des formes tronquées entre MSA et témoins ainsi qu'entre les aires cérébrales. Il

serait alors intéressant de confirmer ces résultats par bottom-up, plus sensible, en ciblant des formes tronquées spécifiques en cerveau pour valider ces résultats.

Tableau 24 Quantification des troncations par PRM

Troncation	Mediane				p value				ratio			
	Putamen- MSA	Putamen- ctrl	Cortex Frontal- MSA	Cortex Frontal- Ctrl	Putamen- Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen	Putamen- Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen
1-119	8,92E-02	1,57E-01	5,85E-02	2,43E-02	0,6905	0,2343	0,1061	0,0173	1,76	0,42	0,66	0,16
1-31	3,40E-01	3,54E-01	2,62E-01	1,02E-01	0,8333	0,5338	>0,9999	0,6623	1,04	0,39	0,77	0,29
1-38	7,40E-01	1,17E+00	7,90E-01	6,75E-01	0,1775	0,4691	>0,9999	0,1775	1,58	0,85	1,07	0,58
1-41	7,83E-02	8,26E-02	9,96E-02	7,20E-02	0,9307	0,2949	0,7308	0,5368	1,06	0,72	1,27	0,87
1-50	9,22E+00	3,82E+00	3,61E+00	2,48E+00	0,1255	0,1707	0,2343	0,8896	0,41	0,69	0,39	0,65
1-53	4,14E-01	5,17E-01	3,06E-01	2,76E-01	0,7922	0,7308	0,6282	0,1255	1,25	0,90	0,74	0,53
1-57	3,15E-01	4,11E-01	4,53E-01	5,68E-01	>0,9999	0,2949	0,6282	0,2468	1,31	1,26	1,44	1,38
1-67	1,75E-02	3,90E-02	4,80E-02	4,50E-02	0,329	0,655	0,1265	0,4525	2,23	0,94	2,74	1,15
1-86	1,43E-01	1,85E-01	1,10E-01	1,10E-01	0,9307	0,7308	0,9452	0,4286	1,29	1,00	0,77	0,60
1-90	5,00E+00	3,12E-01	1,05E+00	5,86E-01	0,1775	0,1807	0,035	0,7922	0,06	0,56	0,21	1,87
42-140	5,10E-03	1,05E-02	7,50E-03	7,75E-03	0,5476	0,8357	0,3434	0,4589	2,06	1,03	1,47	0,74
51-140	1,76E-01	6,90E-02	9,72E-02	6,67E-02	0,0303	0,5338	0,0734	0,9307	0,39	0,69	0,55	0,97
54-140	2,60E-01	2,37E-01	1,83E-01	1,45E-01	0,329	0,7308	0,5338	0,9307	0,91	0,79	0,70	0,61
58-140	8,80E-01	8,52E-01	4,00E-01	2,83E-01	0,7922	0,2949	0,8357	0,2468	0,97	0,71	0,45	0,33
68-140	1,75E-01	1,56E-01	2,57E-01	2,34E-01	0,9307	0,5338	0,366	0,4286	0,89	0,91	1,47	1,50
74-140	1,60E-02	2,19E-02	3,53E-02	3,69E-02	0,3095	0,7308	0,5303	0,9307	1,37	1,04	2,21	1,68
76-140	2,19E-02	1,99E-02	1,95E-02	2,64E-02	0,9307	0,6282	0,9452	0,6623	0,91	1,35	0,89	1,33
84-140	9,36E-02	5,73E-02	7,33E-02	8,73E-02	0,9307	0,9552	0,6282	0,1255	0,61	1,19	0,78	1,52
87-140	1,74E-02	3,04E-02	1,11E-01	1,13E-01	0,5476	0,9452	0,0732	0,2468	1,75	1,02	6,38	3,72
91-140	1,50E-01	3,93E-02	6,07E-02	5,57E-02	0,1775	0,366	0,5338	0,4286	0,26	0,92	0,41	1,42
94-140	1,07E-01	1,41E-02	1,48E-02	5,95E-03	0,0519	0,0734	0,0221	0,2662	0,13	0,40	0,14	0,42



Tableau 25 Quantification des troncations par SIM

Troncation	Mediane				p value				ratio			
	Putamen- MSA	Putamen- ctrl	Cortex Frontal- MSA	Cortex Frontal- Ctrl	Putamen- Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen	Putamen- Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen
1-103	8,60E-02	5,72E-02	2,66E-02	2,95E-02	0,0513	0,9452	0,0041	0,0026	0,66	1,11	0,31	0,52
1-119	2,20E-01	3,12E-01	1,55E-01	1,57E-01	0,2343	>0,9999	0,1649	0,0043	1,42	1,01	0,71	0,50
1-31	6,84E-01	1,06E+00	3,38E-01	2,47E-01	0,366	0,6556	0,0973	0,026	1,54	0,73	0,49	0,23
1-38	4,44E-01	7,93E-01	4,67E-01	6,92E-01	0,2949	0,1375	0,8048	0,3939	1,78	1,48	1,05	0,87
1-41	6,40E-01	6,89E-01	6,62E-01	7,05E-01	0,9452	0,7308	0,62	0,8182	1,08	1,07	1,03	1,02
1-50	2,85E+00	9,89E-01	1,23E+00	8,77E-01	0,035	0,1807	0,053	0,5887	0,35	0,72	0,43	0,89
1-53	2,95E-01	4,56E-01	3,12E-01	3,57E-01	0,366	0,9452	0,831	0,3939	1,55	1,14	1,06	0,78
1-57	1,85E+00	5,63E-01	6,94E-01	6,98E-01	0,1014	0,9452	0,2086	0,6991	0,30	1,01	0,37	1,24
1-60	3,30E-01	7,35E-02	1,05E-01	6,25E-02	0,035	0,1807	0,0379	0,9372	0,22	0,59	0,32	0,85
1-64	5,12E-02	3,92E-02	4,36E-02	3,01E-02	0,4452	0,1807	0,2593	0,4848	0,77	0,69	0,85	0,77
1-65	1,45E-01	4,87E-02	5,98E-02	5,46E-02	0,0221	0,6282	0,028	0,5887	0,34	0,91	0,41	1,12
1-67	3,76E-01	2,56E-01	2,80E-01	3,25E-01	0,4452	0,7308	0,62	0,5887	0,68	1,16	0,74	1,27
1-86	6,17E-01	1,13E+00	3,38E+00	9,76E-01	0,3152	0,6667	0,0023	0,8571	1,84	0,29	5,47	0,86
1-90	1,13E+00	2,13E-01	6,39E-01	2,45E-01	0,0513	0,1375	0,2593	0,9372	0,19	0,38	0,57	1,15
40-140	7,98E-02	6,62E-02	4,73E-02	5,95E-02	0,5338	0,8357	>0,9999	0,9372	0,83	1,26	0,59	0,90
42-140	3,00E-02	2,15E-02	2,30E-02	2,20E-02	0,4286	0,7016	0,8642	>0,9999	0,72	0,96	0,77	1,02
51-140	9,03E-01	8,32E-01	7,97E-01	4,81E-01	0,2468	0,8357	0,366	0,6623	0,92	0,60	0,88	0,58
5-140	6,01E-02	3,31E-02	2,02E-02	2,28E-02	0,0931	0,4286	0,0411	0,5368	0,55	1,13	0,34	0,69
54-140	4,17E-01	1,63E-01	1,46E-01	1,28E-01	0,0082	0,3124	0,0023	0,2251	0,39	0,88	0,35	0,79
58-140	1,21E+00	2,08E+00	5,42E-01	5,11E-01	0,366	0,7308	0,0973	0,026	1,72	0,94	0,45	0,25
68-140	4,42E+00	1,13E+00	3,95E+00	2,26E+00	0,0303	0,4452	0,8763	0,132	0,26	0,57	0,89	1,99
74-140	2,70E-02	3,50E-02	2,50E-02	2,75E-02	>0,9999	0,6282	0,8316	0,513	1,30	1,10	0,93	0,79
76-140	2,55E-02	2,35E-02	1,20E-02	1,55E-02	0,9372	0,655	0,4225	0,2879	0,92	1,29	0,47	0,66
79-140	2,45E-02	1,40E-02	1,50E-02	1,20E-02	0,119	0,4266	0,2191	0,7857	0,57	0,80	0,61	0,86
84-140	7,90E-02	4,25E-02	4,30E-02	3,85E-02	0,2949	0,6037	0,4557	0,4589	0,54	0,90	0,54	0,91
87-140	1,80E-02	2,30E-02	2,50E-02	2,35E-02	0,9697	0,3864	0,3869	0,8506	1,28	0,94	1,39	1,02
91-140	1,50E-01	5,80E-02	5,40E-02	4,20E-02	0,035	0,4219	0,0233	0,7338	0,39	0,78	0,36	0,72
94-140	8,47E-01	1,00E-01	1,61E-01	8,35E-02	0,0012	0,2949	0,007	0,587	0,12	0,52	0,19	0,84

#### **d) Quantification des protéases dans les échantillons de cerveaux**

En complément, afin de pouvoir lier une protéase à une forme tronquée de l'alpha-synucléine, nous avons cherché à explorer le lien entre ces tronctions et un changement de concentration de ces protéases entre les MSA et contrôle ou entre les aires cérébrales (putamen vs. Cortex frontal). Pour cela nous avons développé une méthode de quantification relative de ces 5 protéases. Ce protocole est basé sur une précipitation des protéines (pellet digestion, cf C9ORF72) par l'acide perchlorique à partir de 10mg de tissus. L'idée était de pouvoir faire l'analyse de l'alpha-synucléine intacte et des protéases (qui précipitent) à partir de la même préparation initiale. Après 15 min sur glace, les échantillons sont centrifugés à 16 000g. Afin de pouvoir re-suspendre le culot et neutraliser le pH, le culot est repris dans du bicarbonate d'ammonium 50mM et les échantillons sont passés au bain à ultrason pendant 5min. Le reste du protocole est identique au protocole publié pour la C9ORF72.

Les peptides ont été sélectionnés par digestion enzymatique à la trypsine des enzymes utilisées pour générer les formes tronquées <sup>15</sup>N et selon les critères exposés précédemment. La liste des peptides sélectionnés et leurs paramètres MS sont présentés tableau 26. Concernant la séparation chromatographique le gradient utilisé pour la quantification de la C9ORF72 a été conservé (voir figure 81). Contrairement aux méthodes exposées précédemment, les résultats présentés ici sont préliminaires et nous n'avons pas encore évalué la méthode en termes de linéarité, LOQ, fidélité et biais. Cependant, nous avons vérifié la détectabilité des 5 enzymes dans le cerveau humain. Seule la MMP-3 n'a pu être détectée en raison probablement d'un manque de sensibilité

Bien que ces résultats soit préliminaire nous observons une diminution des protéases présentent dans les lysosomes. Notamment les cathepsines-L et D sont diminuées, respectivement, dans le cortex frontal et le putamen des patients atteints de MSA. Globalement la cathepsine D est diminuée pour les cas MSA en comparaison des contrôles. Ces résultats peuvent indiquer un dysfonctionnement de la dégradation des protéines par le lysosome, ce qui pourrait induire l'accumulation de formes toxiques de l'alpha-synucléine. De même la calpaine-1 est diminuée dans le putamen des MSA. Kiely et al<sup>480</sup>, ont en revanche observé une augmentation de l'activité de la calpaine-1 dans le putamen des patients atteints de MSA. Ce qui pose la question du lien entre la concentration d'une protéase et son activité, sachant que l'activité de la calpaine-1 est dépendante de la présence de calcium. Nous

pouvons noter des différences dans les augmentations/diminution entre les peptides d'une même protéase, notamment pour les peptides de la calpaïne-1. La méthode n'étant pas validée, il est possible que certains peptides ne soient pas quantotypique.

Au regard des résultats de quantification des formes tronquées le lien entre une diminution d'une enzyme et l'augmentation d'une forme tronquée n'est pas évidente. Par exemple les formes 1-90, 4-140 et 91-140 sont majoritairement produites par la MMP-3 or nous n'avons pas pu quantifier cette enzyme. La forme 1-50 est majoritairement produite par la calpaïne-1, cette forme est augmentée alors que la quantité de calpaïne diminue. Ces résultats suggèrent une relation moins directe entre concentration d'une protéase et troncations *in vivo*.

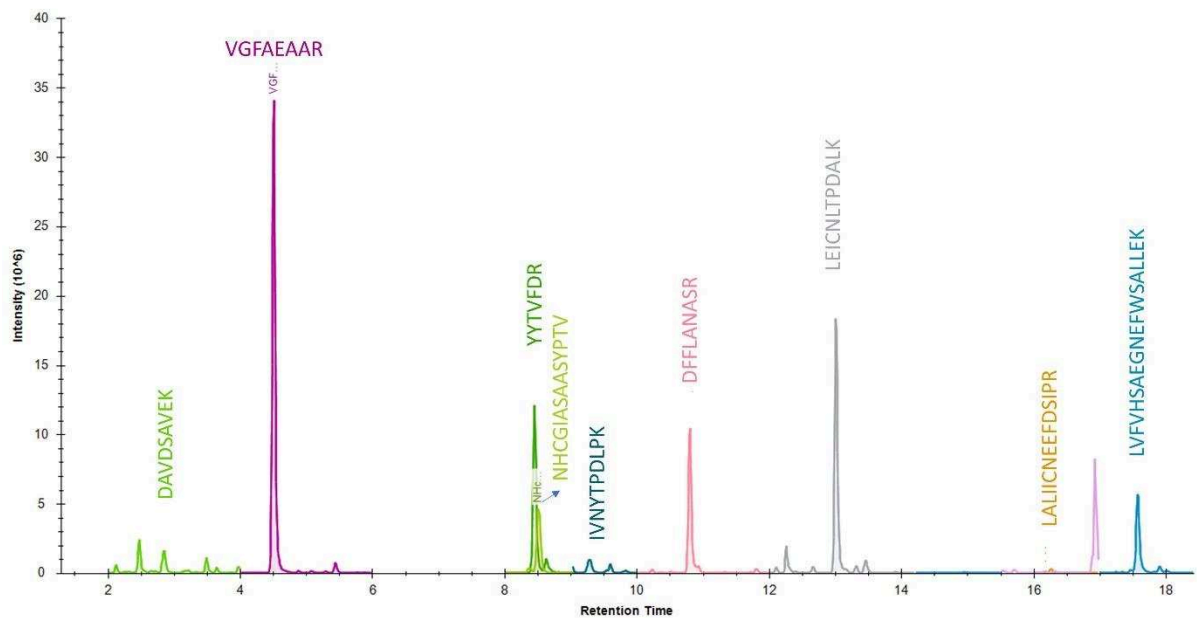


Figure 81 XIC des peptides appartenant aux protéases.

Tableau 26 Paramètres MS pour la détection par PRM des peptides des protéases

Protein	peptide sequence	Precursor ion m/z	product ion type	product ion m/z	NCE (%)	Retention time (min)
cathepsin L	NHCGIASAASYPTV	724,3354 (z=2)	y3	316,1867	15	8,5
			b8	811,3515		
			b9	882,3887		
			b10	969,4207		
			b11	1132,484		
	NHCGIASAASYPTV[ <sup>13</sup> C5, <sup>15</sup> N1]	727,3423 (z=2)	y3	322,2005	15	8,5
			b8	811,3515		
			b9	882,3887		
			b10	969,4207		
			b11	1132,484		
cathepsin D	VGFAEAAR	410,7192 (z=2)	y7	721,3628	20	4,5
			y5	517,2729		
	VGFAEAAR[ <sup>13</sup> C6, <sup>15</sup> N4]	415,7234 (z=2)	y4	446,2358	20	4,5
			y7	731,371		
	YYTVFDR	482,2322 (z=2)	y5	527,2812	15	8,5
			y4	456,244		
	YYTVFDR[ <sup>13</sup> C6, <sup>15</sup> N4]	487,2363 (z=2)	y6	800,3937	15	8,5
			y5	637,3304		
			y4	536,2827	15	8,5
			y6	810,402		
		y5	647,3387	15	8,5	
		y4	546,291			
MMP-3	DAVDSAVEK	467,2298 (z=2)	y7	747,3883	15	n.d.
			y6	648,3199		
			b2	187,0713		
			b3	286,1397	15	2,9
	DAVDSAVEK[ <sup>13</sup> C6, <sup>15</sup> N2]	471,2369 (z=2)	y7	755,4025		
			y6	656,3341		
			b2	187,0813	15	n.d.
	IVNYTPDLK	580,3215 (z=2)	b3	286,1397		
			y8	947,4833		
			y7	833,4403	15	n.d.
			y6	670,377		
			b3	327,2027		
			b5	591,3137	15	9,3
	IVNYTPDLK[ <sup>13</sup> C6, <sup>15</sup> N2]	584,3286 (z=2)	y8	955,4975		
			y7	841,4545		
		y6	678,3912	15	9,3	
		b3	327,2027			
		b5	591,3137			
caspase-1	DFIAFCSSTPDNVSWR	951,428 (z=2)	y12	1455,6321	20	15
			y11	1308,5637		
			y7	873,4213		
			y3	448,2203	20	15
	DFIAFCSSTPDNVSWR [13C6,15N4]	956,4321 (z=2)	b3	376,1867		
			y12	1465,6404		
			y11	1318,572	20	15
			y7	883,4296		
			y3	458,2386		
			b3	376,1867	20	16,3
	LALICNEEFDSIPR	895,4613 (z=2)	y11	1379,626		
			y10	1266,5419		
			b3	298,2125	20	16,3
	LALICNEEFDSIPR[13C6,15N4]	900,4654 (z=2)	y11	1389,6343		
			y10	1276,5502		
		b3	298,2125	15	10,8	
LVFVHSAEGNEFWSALLEK	726,0406 (z=3)	y8	993,5404			
		y7	846,472			
		y17 (+2)	982,4811	15	10,8	
		b11	1183,5742			
		b12	1333,6426			
LVFVHSAEGNEFWSALLEK[13C6,15N2]	728,7120 (z=3)	y8	1001,5546	15	10,8	
		y7	8544862			
		y17 (+2)	986,4882			
		b11	1183,5742	20	13	
		b12	1330,6426			
LEICNLTPDALK	693,8685 (z=2)	y10	1144,6031			
		y9	1031,519	20	13	
		y8	871,4884			
		y7	757,4454			
		y5	543,3137	20	13	
LEICNLTPDALK[13C6,15N2]	697,8756 (z=2)	y10	1152,6173			
		y9	1039,5332			
		y8	879,5026	20	13	
		y7	765,4596			
		y5	551,3279			
		y7	778,4206	10	10,8	
DFFLANASR	520,7616 (z=2)	y6	631,352			
		b4	523,2551			
		b5	594,2922	10	10,8	
		b7	779,3723			
DFFLANASR[13C6,15N4]	525,7658 (z=2)	y7	788,4289			
		y6	641,3605	10	10,8	
		b4	523,2551			
		b5	594,2922			
		b7	779,3723			

Tableau 27 Quantification des protéases dans les cerveaux

Protéines	Peptide Sequence	Mediane				p value				ratio			
		Putamen-MSA	Putamen-ctrl	Cortex Frontal-MSA	Cortex Frontal-Ctrl	Putamen-Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen	Putamen-Ctrl vs.MSA	CF-Ctrl vs. MSA	MSA CF vs. Putamen	Ctrl CF vs. Putamen
Cathepsin-L	NHCGIASAASYPTV	2,54E-01	3,99E-01	3,89E-01	4,74E-01	0,366	0,0221	0,8048	0,3095	1,57	1,22	1,53	1,19
Cathepsin-D	VGFAEAAR	2,09E+01	3,23E+01	2,45E+01	3,08E+01	0,0082	0,0513	0,9015	0,4848	1,54	1,26	1,17	0,95
	YYTVFDR	1,93E-01	1,90E+00	1,44E+00	1,98E+00	0,035	0,0513	0,1282	0,4848	9,83	1,38	7,43	1,04
Caspase-1	LALIICNEEFDSIPR	9,99E-02	1,14E-01	6,91E-02	9,19E-02	0,4286	0,2403	0,329	0,6991	1,14	1,33	0,69	0,81
	LVFVHSAEGNEFWSALLEK	4,41E-02	2,63E-01	2,60E-01	2,50E-01	0,0221	0,6282	0,0262	0,8182	5,95	0,96	5,88	0,95
Calpain-1	LEICNLTPDALK	1,60E-01	4,77E-01	3,94E-01	4,67E-01	0,014	0,2949	0,053	>0,9999	2,97	1,18	2,46	0,98
	DFFLANASR	5,07E-01	7,47E-01	5,58E-01	6,07E-01	0,0513	0,366	0,9015	0,0649	1,47	1,09	1,10	0,81

Conclusion top-down :

Les corps de Lewy contiennent environ 15% de formes tronquées d'alpha-synucléine. De plus, ces tronctions notamment de la partie C-terminale favorise la formation de fibrille amyloïde. L'alpha-synucléine est connue pour avoir un grand nombre de sites de tronctions<sup>314</sup> causé par des protéases comme la calpaïne-1. Kellie et al., ont récemment une dizaine de formes de l'alpha-synucléine dans le cortex frontal humain, en revanche la méthode de quantification proposé utilisait un anticorps afin de concentrer l'alpha-synucléine. Cet anticorps a un épitope 91-99 ne permettant pas la détection de forme plus courte de l'alpha-synucléine. Nous avons eu donc pour but de développer une méthode sans anticorps. Nous avons alors simplement utilisé le protocole d'extraction par l'acide perchlorique. Le tissu cérébral étant un tissu hétérogène nous avons généré des formes tronquées à partir d'un standard <sup>15</sup>N de l'alpha-synucléine pour normaliser les signaux. Pour cette génération nous avons sélectionnée 5 protéases connues pour couper l'alpha-synucléine. Les fragments générés en plus de permettre de normaliser le signal endogène à faciliter l'identification des formes de l'alpha-synucléine. Cette méthode a été appliqué à une cohorte de 7 MSA et 6 contrôles comprenant deux aires cérébrales : putamen et cortex frontal. L'analyse nous a permis d'identifier 28 protéoformes. Les résultats étant hétérogène et le nombre d'échantillon faible nous n'avons pas pu mettre en évidence des différences significatives claires entre les groupes communs aux deux modes de quantification. Néanmoins, une tendance se dégage en SIM, notamment, pour une augmentation de 7 formes dans le putamen des MSA en comparaison des contrôles. A noter que ces formes sont tronquées aussi bien Cter que Nter, alors que les formes C-terminale seraient les plus toxiques.

## V. Discussion et perspectives

Nous avons développé une méthode de quantification multiplexe des protéines tau et de l'alpha-synucléine dans le LCR. Ces deux protéines semblent être à la fois impliquées dans la maladie d'Alzheimer et les synucléinopathies. Un recouvrement clinique existe entre les démences à corps de Lewy et la maladie d'Alzheimer avec ou sans démences à corps de Lewy<sup>426</sup>. Il y a donc un besoin de biomarqueurs pour le diagnostic, et particulièrement aux stades précoces. En effet, le diagnostic clinique des démences à corps de Lewy à une bonne spécificité 96% mais a une sensibilité faible à 12%<sup>483</sup>. Notre méthode de quantification de tau et de l'alpha-synucléine

permettrait de discriminer les cas d'Alzheimer des cas de DCL. En effet, dans le cas de la maladie d'Alzheimer les deux biomarqueurs sont augmentés alors que dans les DCL l'alpha-synucléine est diminuée ne comparaison d'un groupe contrôle. Ce résultat reste à valider sur un plus grand nombre d'échantillons, notamment la diminution controversée des taux d'alpha-synucléine par rapport aux témoins (cf article III). Afin de permettre une meilleure classification de ces maladies l'ajout de biomarqueurs comme par exemple p181-tau, ou p-S129  $\alpha$ -synucléine pourrait permettre d'augmenter la spécificité et la sensibilité de notre méthode. En effet l'hyperphosphorylation de tau est un biomarqueur classiquement utilisé en clinique et une augmentation de la phosphorylation en S129 de l'alpha-synucléine est observée uniquement pour les synucléinopathies<sup>483</sup>. L'extension de notre méthode serait l'ajout de ces formes phosphorylées. Afin de quantifier ces phosphorylations la sensibilité de notre méthode doit être améliorée. En effet, la phosphorylation de p-tau 181 représente 10 à 30% de la forme totale<sup>484</sup> et la phosphorylation de l'alpha-synucléine en S129 représente entre 9 % et 23%<sup>483</sup>. Pour gagner en sensibilité, et détecter p-tau 181, nous pourrions simplement utiliser un volume de LCR plus important (500  $\mu$ L gain d'un facteur 2.5) En effet, cette phosphorylation est sur la partie centrale de tau et devrait donc être autour de 300 pg/mL. Pour les peptides de tau nous avons des LLOQ de 0.5 à 0.75 ng/mL, il faudrait donc gagner un facteur d'environ 3. En revanche, pour l' $\alpha$ -synucléine nous ne sommes pas parvenu à quantifier dans le LCR le peptide AYEMPSEE sans phosphorylation. Une étape d'oxydation des méthionines pourrait permettre d'augmenter la sensibilité en évitant une dispersion du signal entre une forme nono-oxydée et mono/di-oxydée, à combiner avec une analyse en nano-LC, pour améliorer les seuils de détection. Les peptides générés par la glu-C pour la partie C-terminale contiennent de nombreux acides aminés chargés négativement, nous pourrions aussi envisager de les analyser en mode d'ionisation négatif afin de tirer avantage de leur séquence peu adaptée à l'analyse en mode d'ionisation positif.

Bien que nous ne soyons pas parvenus à quantifier tous les peptides de la partie C-terminale de l'alpha-synucléine, nos résultats suggèrent une hétérogénéité des formes présentes dans le LCR bien que moins marquée que pour la protéine tau (voir figure 82). En effet, la protéine tau est une protéine très hétérogène, présente dans le LCR sous la forme de deux fragments majoritaires de 17 kDa et de 30 kDa, d'après les résultats de Barthelemy et al (Figure 82) et des analyses par Western Blot. Cette protéine serait donc tronquée par des protéases en N-terminale et C-terminale, laissant la partie centrale intacte. Contrairement à la protéine tau, les résultats obtenus pour la quantification de l'alpha-synucléine suggèrent que des modifications de l'alpha-synucléine auraient lieu dans sa partie centrale (NAC) 62-96. Cette observation est concordante avec les

résultats, de quantification de l'alpha-synucléine, obtenus par Oeckel et al<sup>452</sup>. La région NAC comprend, en effet, plusieurs sites de troncations 65-140, 66-140, 68-140, 71-140, 1-78,1-83,1-91 et 1-93<sup>210,485</sup> et un site de phosphorylation S87. Le peptide C-terminale 132-140 et le peptide N-terminale 47-57 ont des niveaux identiques dans le LCR suggérant des troncations centrales et non comme Tau aux extrémités. La partie C-terminale de l'alpha-synucléine semble moins modifié que ce qui a pu être envisagé<sup>452</sup>. Néanmoins, ce résultat est à pondérer puisque nous quantifions uniquement un peptide pour la partie N-terminale et C-terminale.

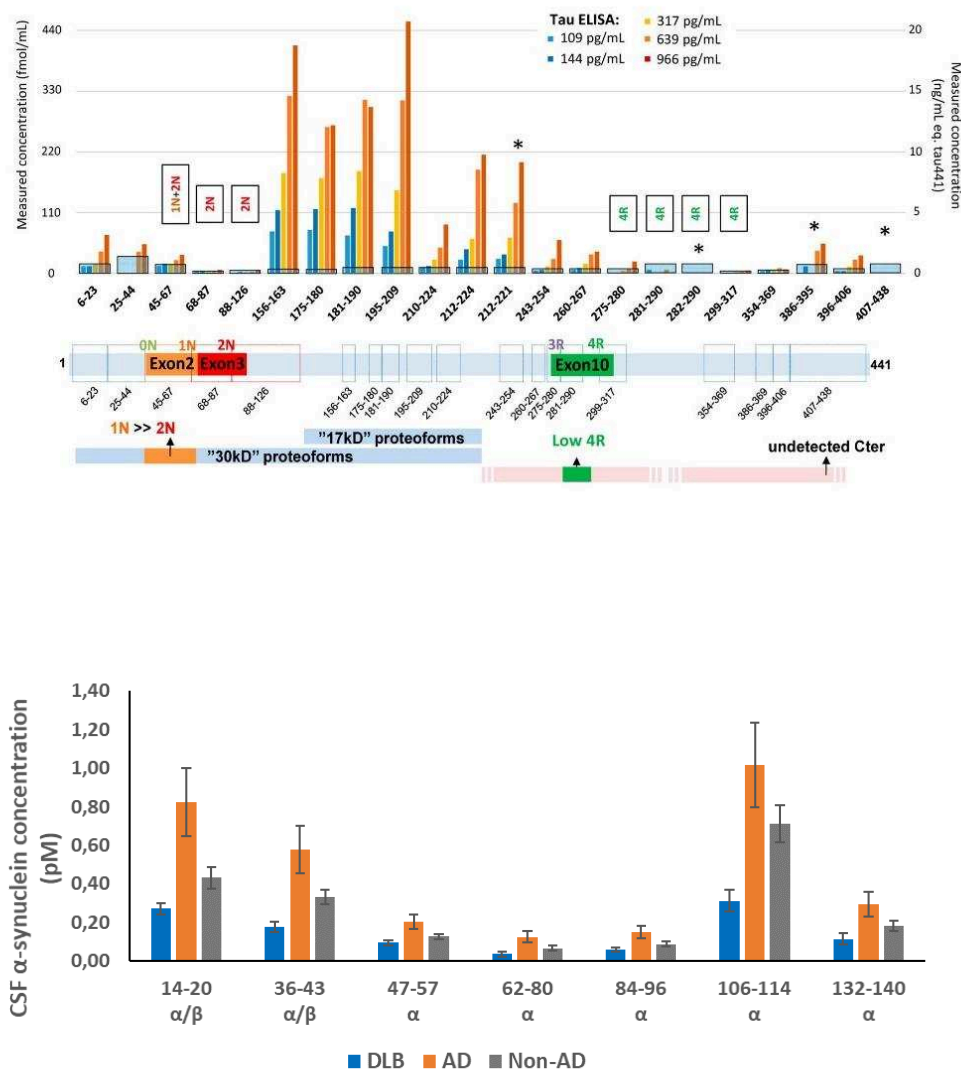


Figure 82 comparaison des profils pour Tau (haut) et de l'alpha-synucléine (bas)



Concernant les méthodes publiées pour la quantification de l'alpha-synucléine dans le LCR. La première<sup>452</sup> n'a pas observé de différence entre les synucléinopathies et la maladie d'Alzheimer alors que dans la seconde<sup>486</sup> il a été observé une différence entre le groupe PD et le groupe contrôle. Il est possible que cette différence provienne de la préparation d'échantillon. En effet, dans le cas de Oeckl et al. le LCR est digéré puis les peptides sont fractionnés alors que dans le cas de yang et al. les protéines sont précipitées avec de l'acide trichloracétique. Cette précipitation à l'acide, pourrait, comme nous l'avons observé avec notre protocole, favoriser des protéoformes (cf mesure des rendements de précipitation et SPE) et donc potentiellement la différence observée serait liée à la préparation d'échantillon.

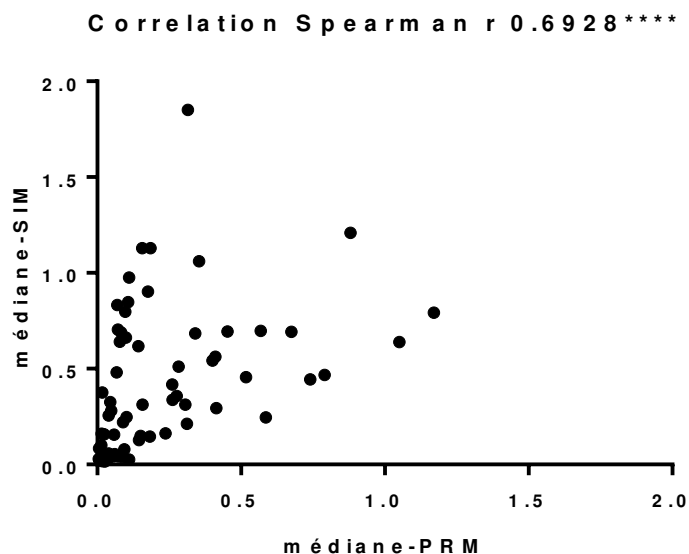
Au regard de ces résultats, une approche top-down permettrait de caractériser les protéoformes de l'alpha-synucléine dans le LCR. Nous avons évalué la sensibilité de cette approche, dans du LCR mime, à 50 ng/mL pour la synucléine intacte, avec la méthodologie top-down ciblée (mode tSIM) la plus sensible. Or nous avons trouvé des concentrations proches du ng/mL par bottom-up et des concentrations différentes entre les peptides indiquant des modifications post-traductionnelles. Le gain nécessaire pour une approche top-down serait au minimum de 50. Ce gain semble difficilement atteignable en gardant cette préparation d'échantillon. Une immunoprécipitation semblerait plus adaptée, par exemple en combinant plusieurs anticorps (N-terminale, NAC et C-terminale) afin de capter la diversité des formes présentes dans le LCR.

Du fait de ces limitations, nous avons développé une approche top-down pour la quantification de l'alpha-synucléine dans le cerveau où elle y est plus abondante. Cette méthode bien que préliminaire nous a permis de détecter et de quantifier plus de 28 formes tronquées chez l'humain en SIM et 21 en PRM. Cette quantification a été réalisée par l'ajout de standard marqué <sup>15</sup>N, qui ont été générés par digestion avec 5 protéases. Ce standard permet de normaliser les effets matrices et la préparation d'échantillon. Néanmoins, notre standard <sup>15</sup>N ne mime pas parfaitement l'alpha-synucléine endogène. En effet, cette dernière est acétylée en N-terminale ce qui conduit à un décalage en temps de rétention avec la protéine recombinante <sup>15</sup>N. Nous aurions pu envisager la méthode proposée très récemment par Rovere et al.<sup>487</sup> pour l'acétylation N-terminale de l'alpha-synucléine recombinante par la NatB<sup>45</sup>. De plus, cette acétylation nous aurait aussi permis de quantifier le peptide N-terminale par bottom-up.

La méthode de quantification a été appliquée à l'analyse de 26 homogénats de cerveaux humains. En comparaison avec la quantification de l'alpha-synucléine dans le cortex frontal de Parkinson,

réalisé par Kellie et al, nous quantifions 2 fois plus de protéoformes dont 3 sont communes 68-140, 1-119 et 5-140. Kellie et al., n'avaient pas mis en évidence de protéoformes impactées dans le cortex frontal des Parkinson en comparaison d'un groupe témoins ce qui peut être expliquée par le fait que les tronctions de l'alpha-synucléine par des protéases est un processus normal<sup>313</sup>. De même, nous n'avons pas trouvé de protéoformes significativement impactés dans le cortex frontal (MSA vs. Témoins). En revanche, plusieurs formes semblent être augmentées dans le putamen de personnes atteintes de MSA. Les formes mis en évidence sont tronquées en N-terminale et C-terminale. Ces deux types de tronctions peuvent induire une pathologie de type synucléinopathies chez la souris tout en ayant des cinétiques d'agrégation différentes<sup>411</sup>.

Les résultats obtenus entre PRM et SIM sont globalement corrélés Spearman  $r=0.6928$  et  $p<0.0001$  (comparaison des médianes, figure 83). Cependant, le niveau de corrélation aurait pu être plus marqué. Ce résultat pourrait être lié à un manque de sensibilité. La méthode a été développée en cerveaux de souris sur-exprimant la synucléine, et appliquée ensuite aux échantillons humains plus précieux. Des adaptations au niveau de la SPE ou du poids de cerveau prélevé seraient nécessaires. D'autre part, contrairement au bottom-up où nous vérifions le ratio des fragments peptidiques à l'aide du logiciel Skyline, nous n'avons pas pu vérifier aussi facilement le ratio des fragments issus de l'analyse PRM des protéines entières. De même pour le mode SIM, il faudrait contrôler l'abondance relative des isotopes pour s'assurer de la bonne extraction du signal. Une adaptation de ce logiciel au traitement des données top-down serait utile. C'est pourquoi ces résultats pourraient être confirmées par bottom-up pour certaines formes (voir tableau 28), et si possible, sur un plus grand nombre d'échantillons.



**Figure 83** Corrélation des médianes obtenues par PRM/SIM

Nos résultats suggèrent une augmentation des tronctions de l'alpha-synucléine dans le putamen, or nous n'avons observé que des diminutions de concentration des protéases dans cette aire cérébrale. Le lien entre la concentration en enzymes et les formes tronquées n'est donc pas évident.

Une autre limite de nos résultats est la probable contamination par les globules rouges, riches en alpha-synucléine. Nous avons retiré deux échantillons de LCR pour cette raison (signal intense du peptide de l'hémoglobine). Nous n'avons pas pu faire ce test sur les échantillons de cerveau. Cette contamination peut interférer dans la quantification de certaines formes tronquées et fausser leur quantification. Enfin, comme indiqué précédemment, les résultats sont à prendre avec précautions en raison du faible nombre d'échantillon analysé par groupe (7 MSA et 6 contrôles).

Après confirmation, les formes impactées par la pathologie, pourraient être recherchées dans le LCR dans une optique diagnostic (voir tableau 28). Pour la quantification de ces formes, la combinaison de deux analyses une en Glu-C et une en trypsine devra être réalisée pour en quantifier le plus possible (tableau 28). Ces deux enzymes seront complémentaires pour générer des peptides des formes tronquées détectables par LC-MS.

Tableau 28 Liste des peptides générés digestion des formes tronquées observées, par la Glu-C ou par la trypsine. L'acide aminé du site de coupure est en rouge. L'acide aminé suivant (troncation C-ter) et précédant (N-ter) est aussi indiqué. Comme illustré, tous les peptides n'auront pas une taille adapté à l'analyse LC-MS, et des coupures endogènes sont communes à la trypsine ou à la Glu-C

	GluC-C	blast	Trypsine	blast
1-103	GAGSIAAATGFVKKDQLGK <b>N</b> EE	unique	X	
1-119	DMPV <b>D</b> P		NEEGAPQEGILEDMPV <b>D</b> P	unique
1-31	AA <b>G</b> K		QGVAAEA <b>G</b> K	unique
1-38	GV <b>L</b> Y		EGV <b>L</b> Y	
1-41	GVLYV <b>G</b> S	beta-synucléine	EGVLYV <b>G</b> S	beta-synucléine
1-50	GVV <b>H</b> G		EGVV <b>H</b> G	
1-53	GVVHG <b>V</b> AT	unique	EGVVHG <b>V</b> AT	unique
1-57	GVVHG <b>V</b> ATV <b>A</b> EK	unique	EGVVHG <b>V</b> ATV <b>A</b> EK	unique
1-60	KT <b>K</b> E		T <b>K</b> E	
1-64	QV <b>T</b> N		EQV <b>T</b> N	
1-65	QVT <b>N</b> V		EQVT <b>N</b> V	
1-67	QVTNV <b>G</b> G		EQVTNV <b>G</b> G	unique
1-86	GA <b>G</b> S		TVEGA <b>G</b> S	
1-90	GAGSIA <b>A</b> A		TVEGAGSIA <b>A</b> A	unique
40-140	Y <b>V</b> GSKTKE	unique	Y <b>V</b> GSK	
42-140	G <b>S</b> KTKE		<b>G</b> SK	
51-140	H <b>G</b> VATVAE	unique	H <b>G</b> VATVAEK	unique
5-140	F <b>M</b> KGLSKAKE	unique	F <b>M</b> KGLSK	
54-140	A <b>T</b> VAE		A <b>T</b> VAEK	
58-140	E <b>K</b> TKE		<b>E</b> K	
68-140	G <b>G</b> AVVTGVTAVAQKTVE	unique	G <b>G</b> AVVTGVTAVAQK	unique
74-140	G <b>V</b> TAVAQKTVE	unique	G <b>V</b> TAVAQKT	unique
76-140	T <b>A</b> VAQKTVE	unique	T <b>A</b> VAQK	
79-140	A <b>Q</b> KTVE		A <b>Q</b> K	
84-140	E <b>G</b> AGSIAAATGFVKKDQLGKNEE	X	E <b>G</b> AGSIAAATGFVK	unique
87-140	G <b>S</b> IAAATGFVKKDQLGKNE	unique	G <b>S</b> IAAATGFVK	unique
91-140	A <b>A</b> TGFVKKDQLGKNE	unique	A <b>A</b> TGFVK	
94-140	G <b>F</b> VKKDQLGKNE	unique	G <b>F</b> VK	

# Conclusion générale

La découverte de nouveaux biomarqueurs et la meilleure compréhension de la physiopathologie des maladies neurodégénératives est un enjeu crucial dans le domaine biomédical. Les biomarqueurs existants comme par exemple les peptides amyloïdes bêta, la protéine tau et la phosphorylation de la thréonine 181 sont quantifiés par des tests immunoenzymatiques (ELISA). La sensibilité et la relative facilité d'utilisation des ELISA en font un outil de choix pour une application en clinique. Néanmoins, les ELISA ne permettent pas de capturer l'hétérogénéité d'une protéine, appelée protéoforme. Une protéoforme correspond à toutes les modifications possibles d'une protéine que ce soit pendant la traduction (mutation, isoforme) ou après : modifications post-traductionnelles (phosphorylation, troncations, etc.). Les protéoformes permettraient de mieux décrire des phénotypes cliniques complexes. Dans ce domaine la spectrométrie de masse s'impose comme un outil de choix de par la spécificité apportée en comparaison des ELISA. Plus largement, la découverte de nouveaux biomarqueurs par spectrométrie de masse se fait essentiellement selon des approches non ciblées. Ces approches non ciblées souffrent de limitation en termes de sensibilité et de justesse, c'est pourquoi pour la validation des approches de quantification ciblée sont mises en place. Ces approches ciblées sont normalisées par un étalon interne qui permet de rendre la méthode juste et fidèle. La quantification se fait généralement par bottom-up (digestion enzymatique des protéines) mais dans le cadre des protéoformes le lien entre les modifications est perdu. Les analyses en top-down (sans digestion enzymatique) sont donc dans ce cas particulièrement intéressantes puisqu'elles permettent d'analyser directement les protéoformes. Mais cette approche est généralement moins sensible que le bottom-up et il y a un manque de standards internes pour la quantification des protéoformes.

Au travers de ce travail de thèse, nous avons souhaité évaluer l'apport de la spectrométrie de masse bottom-up et top-down pour la quantification de protéoformes de biomarqueurs protéiques impliqués dans diverses maladies neurodégénératives. Nous nous sommes intéressés d'une part à mieux caractériser une protéine hétérogène régulièrement citée comme biomarqueur potentielle des synucléinopathies, l'alpha-synucléine et d'autre part, à la quantification des isoformes de la protéine C9ORF72, protéine encore peu étudiée et dont le lien avec une forme familiale de DFT a été récemment découvert. Nous avons également profité de la capacité de multiplexage des approches bottom-up pour cibler un plus grand nombre de protéines impliquées dans les DFT.

La première partie de ce travail de thèse a été consacré à la quantification des isoformes de la C9ORF72 dans le cortex frontal de sujet sains ou atteints de démences frontotemporales avec ou sans expansion  $(G_4C_2)_n$  présent dans le gène *C9ORF72*. Cette expansion a été découverte récemment en 2011 et représente la première cause génétique de DFT et de SLA. Trois mécanismes de toxicité, non exclusif, sont supposés comme conduisant à l'apparition de la maladie. Premièrement, une haploinsuffisance de la C9ORF72 qui ne remplirait alors plus sa fonction. Deuxièmement, la production de foci d'ARN qui séquestreraient les protéines liées à l'ARN. Enfin, la production de protéines dipeptides, s'agréant dans le cerveau, résultant de la traduction de l'expansion. Nous avons souhaité apporter un outil analytique permettant de quantifier la C9ORF72 dans diverses matrices biologiques et de confirmer sa diminution dans le cortex frontal de sujet ayant l'expansion  $(G_4C_2)_n$ . La C9ORF72 possède deux isoformes : une forme longue de 481 AA et une forme courte de 222. La forme courte possède seulement un AA différent en comparaison avec la forme longue. Du fait de sa faible abondance dans le cerveau nous avons opté pour une approche de quantification par bottom-up. Nous avons développé une méthode de quantification des deux isoformes par différence en quantifiant deux peptides communs à la C9-S et la C9-L et un peptide unique de la C9-S. Du fait de la similarité des deux isoformes nous n'avons pu cibler qu'un seul peptide unique de la C9-S, mais qui ne remplissait pas tous les critères des peptides protéotypiques. Après validation, nous avons appliqué cette méthode à une cohorte comportant 49 sujets, dont des DFT avec expansion (n=21), sans expansion (n=10) et des sujets sains (n=13). Nous avons pu mettre en valeur une diminution d'environ 40% de la quantité de la C9-L dans le cortex frontal des sujets ayant une expansion  $(G_4C_2)_n$ . Ce résultat confirme les observations précédentes faites par Western Blot, mais probablement avec une meilleure précision liée à la spectrométrie de masse. Il pourra par la suite servir à établir des corrélations entre l'expression de la protéine et le dossier clinique du patient, le nombre d'expansion présente ou l'efficacité de nouveau traitement. Cette méthode est facilement transposable à d'autres laboratoires et pourrait permettre de suivre l'impact d'un traitement visant l'haploinsuffisance de la C9ORF72 sur l'abondance de la C9ORF72 dans des modèles animaux. En revanche, notre méthode n'est pas assez sensible pour la quantification de la C9-S. En effet, cette isoforme semble minoritaire en comparaison de l'isoforme C9-L (<20%). Du fait de la faible abondance de la C9-S une méthode de quantification par différence n'est pas optimale comme nous l'avons observé. Il faudrait alors gagner en sensibilité, par exemple par une méthode d'immunoprécipitation avec de nouveaux anticorps pour tenter de détecter cette C9-S, et potentiellement tenter une approche

top-down (poids moléculaire de 25kDa compatible avec le top-down). Le cerveau étant une matrice biologique précieuse et non adaptée au suivi pronostique ou thérapeutique chez l'homme, nous avons par la suite tenter d'étendre notre méthode à la quantification de la C9ORF72 en PBMC. Afin d'avoir une meilleure compréhension des mécanismes mis en jeu dans les DFT, nous avons profité des capacités de multiplexage de la spectrométrie de masse pour réaliser le suivi de 56 protéines. Les protéines ont été sélectionnée à partir de données transcriptomiques et de la littérature. A partir de cette liste de protéines les peptides ont été sélectionnés par une analyse DDA ou par des analyses ciblées par PRM. Cette stratégie nous a permis d'éliminer 5 protéines et de sélectionner 122 peptides. Après optimisation du protocole d'extraction, la méthode a été évaluée en termes de répétabilité et de linéarité. Enfin, la méthode a été appliquée à l'analyse de la même cohorte d'échantillons que pour la quantification de la C9ORF72. Nous avons pu quantifier 49 protéines et trouvées 19 protéines impactées dans les DFT. Ces protéines sont impliquées par exemple dans le métabolisme du glutamate, dans la réponse à un stress ou des dans l'inflammation, et nous avons pu ainsi relier nos données protéomiques à des mécanismes physiopathologiques récemment proposés dans les DFT. Cette méthode sera appliquée à d'autres régions cérébrales. Pour finir, nous avons évalué le mode DIA sur les mêmes échantillons dans une perspectives de quantification exhaustives, mais pour le moment l'approche est apparemment moins sensible (préliminaires) et plus complexe au niveau traitement des données.

La seconde partie de ce travail de thèse a porté sur la caractérisation des protéoformes de l'alpha-synucléine. L'alpha-synucléine est impliquée dans un groupe de maladie neurodégénératives appelé synucléinopathies regroupant la maladie de Parkinson, l'atrophie multisystématisée et les démences à corps de Lewy. Ce groupe est caractérisé par l'accumulation anormal d'alpha-synucléine dans le cerveau qui forme des agrégats appelés corps de Lewy. L'alpha-synucléine dans le LCR est un biomarqueur contesté des synucléinopathies du fait de résultats contradictoires obtenues principalement par ELISA, bien que plusieurs études montrent une diminution de l'alpha-synucléine dans le LCR chez les sujets atteints de Parkinson ou de démences à corps de Lewy, en comparaison de témoins. Les protéoformes de l'alpha-synucléine présents dans le LCR sont suspectés d'être une des raisons de discordance entre les différentes études. D'autre part, les troncations de l'alpha-synucléine ont été montrées comme favorisant son agrégation et la formation de fibrilles. L'alpha-synucléine semble être aussi impliquée dans d'autres maladies neurodégénératives comme la maladie d'Alzheimer où des agrégats d'alpha-synucléine sont

retrouvés dans le cerveau pour environ 50% des cas. De plus, il existe un recouvrement clinique à un stade précoce de la maladie d'Alzheimer et des démences à corps de Lewy.

Du fait de ce recouvrement et de l'intérêt des formes tronquées nous avons eu pour but de coupler la quantification de l'alpha-synucléine et de la protéine tau dans le LCR en couvrant le plus possible la séquence de l'alpha-synucléine notamment pour la partie C-terminale. Pour ce faire nous avons adapté le protocole développé et publié par Barthélémy et al. en digérant successivement avec la Glu-C puis avec la trypsine. Après validation de la méthode nous l'avons appliqué à la quantification d'une cohorte comprenant 23 patients. Nous avons pu quantifier 7 peptides de l'alpha-synucléine correspondant à une couverture de séquence de 54%. Les résultats de quantification de chaque peptide semblent indiquer que l'alpha-synucléine est moins hétérogène que la protéine tau dans le LCR, avec une diminution d'abondance plutôt observée au niveau des peptides centraux. La concentration de l'alpha-synucléine dans le LCR a été retrouvée augmentée chez les sujets atteints de la maladie d'Alzheimer et diminuée chez les sujets atteints de démences à corps de Lewy en comparaison d'un groupe témoin. Ces résultats encourageants en terme de différenciation des groupes de patients seront confirmés sur un nombre plus important d'échantillons. En parallèle une approche top-down a été développée pour la quantification de l'alpha-synucléine afin de mieux caractériser ses protéoformes. La sensibilité obtenue en matrice LCR est seulement de 50 ng/mL et ne pourra donc pas être appliquée pour la quantification de l'alpha-synucléine du LCR (~ ng/mL). Cette approche a donc été développée dans le cerveau où l'alpha-synucléine est plus abondante. C'est aussi dans cet organe, au plus près de la pathologie, que nous avons le plus grand espoir de mettre en évidence des formes tronquées toxiques. Afin de rendre l'approche par top-down quantitative, nous avons généré des standards marqués <sup>15</sup>N de formes tronquées par digestion du standard <sup>15</sup>N de l'alpha-synucléine par des protéases suspectées d'être impliquées dans le clivage de l'alpha-synucléine. Nous avons pu caractériser plus d'une vingtaine de formes tronquées de l'alpha-synucléine dans le cerveau humain (cortex frontal et putamen). Sept de ces formes ont été trouvées comme significativement augmentés dans le putamen de sujets MSA en comparaison avec des témoins. En parallèle, nous avons quantifié dans les mêmes échantillons les protéases utilisées pour la génération des standards, mais nous n'avons pas pu faire le lien avec l'augmentation des formes tronquées chez les sujets MSA.

Au niveau analytique, ce travail met en évidence les limites des approches bottom-up, malgré des avantages indéniables comme par exemple la précision de quantification obtenue avec la méthode multiplexe. La plus importante limite du bottom-up que nous avons observé est



certainement la perte d'information. Cela se manifeste au travers du travail de quantification de la C9ORF72, avec la difficile différenciation de la forme courte soit de manière directe (difficulté à identifier un peptide protéotypique) ou indirecte (variabilité de la quantification par différence). La même limitation se retrouve avec la synucléine où nous n'avons pu quantifier qu'un seul peptide terminal, ni détecter le peptide comportant le site de phosphorylation S129, pourtant essentiel au niveau physiopathologique. De la même manière, nous avons observé une plus faible abondance des peptides centraux, mais cette observation est seulement constatée par différence avec les peptides terminaux, et sans information sur les modifications post-traductionnelles impliquées. A l'inverse, nous avons vu tout l'intérêt du top-down pour identifier de manière spécifique les formes tronquées de la synucléine dans le cerveau. Ce résultat aurait difficilement pu être obtenu par une méthode bottom-up, même s'il semble possible d'appliquer cette approche à l'avenir pour une quantification ciblée et sensible de certaines troncations dans le LCR.

Un point essentiel pour l'analyse des biomarqueurs et de leur protéoformes souvent minoritaires est la mise en place d'une stratégie de préparation d'échantillon adaptée pour atteindre la sensibilité nécessaire. Nous avons cherché au cours de cette thèse à nous affranchir des anticorps, d'une part pour éviter d'orienter l'analyse vers des protéoformes en particuliers selon l'affinité des anticorps, mais aussi pour des problèmes de disponibilité de « bon » anticorps comme observé au début du travail sur le C9ORF72. Les approches sans anticorps ont montré de bonnes performances en termes de robustesse et de linéarité, mais nous pourrions certainement améliorer nos sensibilités et gagner en exhaustivité pour les formes minoritaires par une extraction par immunocapture.

En résumé, la spectrométrie de masse a permis de caractériser et quantifier des protéoformes de protéines impliquées dans les DFT ou les synucléinopathies, notamment des isoformes et des troncations. Cette quantification s'est faite essentiellement par bottom-up étant donné sa sensibilité, la disponibilité des standards (peptides marqués aux isotopes stables) et des solutions de traitement de données. Le top-down permet de conserver le lien entre les modifications post-traductionnelles et apparaît ainsi beaucoup plus informatif, notamment ici pour des formes tronquées, mais il est pour l'instant majoritairement appliqué à la caractérisation de protéoformes majoritaires. Cette perte de sensibilité est la contrepartie d'une information plus détaillée, en lien avec la dispersion du signal lié aux multiples états de charge et aux modifications post-traductionnelles. D'autre part, le top-down est aussi limité au niveau standardisation, même si nous avons montré, de manière encore imparfaite, la possibilité de générer de multiples formes

tronquées marquées. Les développements technologiques et méthodologiques dans les années à venir permettront de réduire l'écart entre ces deux approches.

# Références

1. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* **69**, 89–95 (2001).
2. Keedy, V. L. *et al.* American Society of Clinical Oncology Provisional Clinical Opinion: Epidermal Growth Factor Receptor ( EGFR ) Mutation Testing for Patients With Advanced Non–Small-Cell Lung Cancer Considering First-Line EGFR Tyrosine Kinase Inhibitor Therapy. *Journal of Clinical Oncology* **29**, 2121–2127 (2011).
3. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
4. Koch, A. *et al.* Analysis of DNA methylation in cancer: location revisited. *Nature Reviews Clinical Oncology* **15**, 459 (2018).
5. Xi, X. *et al.* RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Non-Coding RNA* **3**, 9 (2017).
6. Cuzick, J. *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol.* **12**, 245–255 (2011).
7. Yamazaki, Y., Painter, M. M., Bu, G. & Kanekiyo, T. Apolipoprotein E as a Therapeutic Target in Alzheimer’s Disease: A Review of Basic Research and Clinical Evidence. *CNS Drugs* **30**, 773–789 (2016).
8. McDermott, J. E. *et al.* Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data. *Expert Opin Med Diagn* **7**, 37–51 (2013).
9. Lashley, T. *et al.* Molecular biomarkers of Alzheimer’s disease: progress and prospects. *Disease Models & Mechanisms* **11**, dmm031781 (2018).
10. Shaham, O. *et al.* A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. *Proceedings of the National Academy of Sciences* **107**, 1571–1575 (2010).
11. Griffin, J. L. & Shockcor, J. P. Metabolic profiles of cancer cells. *Nature Reviews Cancer* **4**, 551–561 (2004).
12. Kori, M., Aydin, B., Unal, S., Arga, K. Y. & Kazan, D. Metabolic Biomarkers and Neurodegeneration: A Pathway Enrichment Analysis of Alzheimer’s Disease, Parkinson’s Disease, and Amyotrophic Lateral Sclerosis. *OMICS: A Journal of Integrative Biology* **20**, 645–661 (2016).
13. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems* **4**, 587–599.e4 (2017).
14. Anderson, N. L., Ptolemy, A. S. & Rifai, N. The Riddle of Protein Diagnostics: Future Bleak or Bright? *Clinical Chemistry* **59**, 194–197 (2013).
15. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
16. Smith, L. M. *et al.* Proteoform: a single term describing protein complexity. *Nature Methods* **10**, 186–187 (2013).
17. Pagel, O., Loroch, S., Sickmann, A. & Zahedi, R. P. Current strategies and findings in clinically relevant post-translational modification-specific proteomics. *Expert Rev Proteomics* **12**, 235–253 (2015).
18. Dang, X. *et al.* The first pilot project of the consortium for top-down proteomics: A status report. *Proteomics* **14**, 1130–1140 (2014).
19. Aebersold, R. *et al.* How many human proteoforms are there? *Nature Chemical Biology* **14**, 206–214 (2018).
20. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* (2016). doi:10.1155/2016/7436849
21. Kato, G. J. *et al.* Sick cell disease. *Nat Rev Dis Primers* **4**, 18010 (2018).

22. Bataille, J. *et al.* On-a-chip tryptic digestion of transthyretin: a step toward an integrated microfluidic system for the follow-up of familial transthyretin amyloidosis. *Analyst* **143**, 1077–1086 (2018).
23. Acosta-Baena, N. *et al.* Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. *Lancet Neurol* **10**, 213–220 (2011).
24. Flagmeier, P. *et al.* Mutations associated with familial Parkinson's disease alter the initiation and amplification steps of  $\alpha$ -synuclein aggregation. *PNAS* **113**, 10328–10333 (2016).
25. Paulson, H. Repeat expansion diseases. in *Handbook of Clinical Neurology* **147**, 105–123 (Elsevier, 2018).
26. Romero, P. R. *et al.* Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8390–8395 (2006).
27. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
28. Akgul, C., Moulding, D. A. & Edwards, S. W. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell. Mol. Life Sci.* **61**, 2189–2199 (2004).
29. Schwerk, C. & Schulze-Osthoff, K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol. Cell* **19**, 1–13 (2005).
30. Shimokawa, T. *et al.* Novel Human Glioma-associated Oncogene 1 (GLI1) Splice Variants Reveal Distinct Mechanisms in the Terminal Transduction of the Hedgehog Signal. *J Biol Chem* **283**, 14345–14354 (2008).
31. Yang, X. *et al.* Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**, 805–817 (2016).
32. Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1–30 (2013).
33. Ubersax, J. A. & Ferrell, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8**, 530–541 (2007).
34. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genetics* **26**, 283–289 (2000).
35. Kleiman, L. B., Maiwald, T., Conzelmann, H., Lauffenburger, D. A. & Sorger, P. K. Rapid Phospho-Turnover by Receptor Tyrosine Kinases Impacts Downstream Signaling and Drug Binding. *Mol Cell* **43**, 723–737 (2011).
36. Phosphorylation dynamics during early differentiation of human embryonic stem cells. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/19664995>. (Accessed: 28th October 2018)
37. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10762–10767 (2008).
38. Kuwahara, H., Nishizaki, M. & Kanazawa, H. Nuclear localization signal and phosphorylation of Serine350 specify intracellular localization of DRAK2. *J. Biochem.* **143**, 349–358 (2008).
39. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279–3290 (2007).
40. Russell, C. L. *et al.* Comprehensive Quantitative Profiling of Tau and Phosphorylated Tau Peptides in Cerebrospinal Fluid by Mass Spectrometry Provides New Biomarker Candidates. *J. Alzheimers Dis.* **55**, 303–313 (2017).
41. Lodish, H. *et al.* Protein Glycosylation in the ER and Golgi Complex. *Molecular Cell Biology. 4th edition* (2000).
42. Glycosylation-directed quality control of protein folding. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/26465718>. (Accessed: 28th October 2018)
43. Marotta, N. P. *et al.* O-GlcNAc modification blocks the aggregation and toxicity of the protein  $\alpha$ -synuclein associated with Parkinson's disease. *Nature Chemistry* **7**, 913–920 (2015).
44. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).

45. Ree, R., Varland, S. & Arnesen, T. Spotlight on protein N-terminal acetylation. *Exp. Mol. Med.* **50**, 90 (2018).
46. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS\*. *Proc Natl Acad Sci U S A* **51**, 786–794 (1964).
47. Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase. Available at: <http://www.jbc.org/content/270/42/24674.long>. (Accessed: 28th October 2018)
48. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
49. Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J.* **19**, 6141–6149 (2000).
50. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Reviews Molecular Cell Biology* **15**, 536–550 (2014).
51. Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* **16**, 258–264 (2015).
52. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
53. Kalvik, T. V. & Arnesen, T. Protein N-terminal acetyltransferases in cancer. *Oncogene* **32**, 269–276 (2013).
54. Myklebust, L. M. *et al.* Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum. Mol. Genet.* **24**, 1956–1976 (2015).
55. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nature Reviews Molecular Cell Biology* **18**, 90–101 (2017).
56. Murn, J. & Shi, Y. The winding path of protein methylation research: milestones and new frontiers. *Nature Reviews Molecular Cell Biology* **18**, 517–527 (2017).
57. Lee, Y.-H. & Stallcup, M. R. Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. *Mol. Endocrinol.* **23**, 425–433 (2009).
58. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35**, 618–626 (2010).
59. Erce, M. A., Pang, C. N. I., Hart-Smith, G. & Wilkins, M. R. The methylproteome and the intracellular methylation network. *Proteomics* **12**, 564–586 (2012).
60. Paik, W. K., Paik, D. C. & Kim, S. Historical review: the field of protein methylation. *Trends in Biochemical Sciences* **32**, 146–152 (2007).
61. Shen, E. C. *et al.* Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev.* **12**, 679–691 (1998).
62. Kim, E. *et al.* Phosphorylation of EZH2 Activates STAT3 Signaling via STAT3 Methylation and Promotes Tumorigenicity of Glioblastoma Stem-like Cells. *Cancer Cell* **23**, 839–852 (2013).
63. Afjehi-Sadat, L. & Garcia, B. A. Comprehending dynamic protein methylation with mass spectrometry. *Current Opinion in Chemical Biology* **17**, 12–19 (2013).
64. Rape, M. Ubiquitylation at the crossroads of development and disease. *Nature Reviews Molecular Cell Biology* **19**, 59–70 (2018).
65. Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).
66. Petroski, M. D. The ubiquitin system, disease, and drug discovery. *BMC Biochem* **9**, S7 (2008).
67. Atkin, G. & Paulson, H. Ubiquitin pathways in neurodegenerative disease. *Front Mol Neurosci* **7**, (2014).
68. Hickey, C. M., Wilson, N. R. & Hochstrasser, M. Function and regulation of SUMO proteases. *Nature Reviews Molecular Cell Biology* **13**, 755–766 (2012).
69. Hendriks, I. A. & Vertegaal, A. C. O. A comprehensive compilation of SUMO proteomics. *Nature Reviews Molecular Cell Biology* **17**, 581–595 (2016).

70. Ochaba, J. *et al.* PIAS1 Regulates Mutant Huntingtin Accumulation and Huntington's Disease-Associated Phenotypes In Vivo. *Neuron* **90**, 507–520 (2016).
71. Yang, Y. *et al.* Protein SUMOylation modification and its associations with disease. *Open Biol* **7**, (2017).
72. Pérez-Silva, J. G., Español, Y., Velasco, G. & Quesada, V. The Degradome database: expanding roles of mammalian proteases in life and disease. *Nucleic Acids Res* **44**, D351–D355 (2016).
73. Chung, J.-J., Shikano, S., Hanyu, Y. & Li, M. Functional diversity of protein C-termini: more than zipcoding? *Trends in Cell Biology* **12**, 146–150 (2002).
74. López-Otín, C. & Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **283**, 30433–30437 (2008).
75. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *CMLS, Cell. Mol. Life Sci.* **61**, 1455–1474 (2004).
76. Lavrik, I. N., Golks, A. & Krammer, P. H. Caspases: pharmacological manipulation of cell death. *J Clin Invest* **115**, 2665–2672 (2005).
77. Inobe, T. & Matouschek, A. Paradigms of protein degradation by the proteasome. *Curr. Opin. Struct. Biol.* **24**, 156–164 (2014).
78. Gadhve, K. *et al.* The ubiquitin proteasomal system: a potential target for the management of Alzheimer's disease. *J. Cell. Mol. Med.* **20**, 1392–1407 (2016).
79. Arnold, C. S. *et al.* The Microtubule-associated Protein Tau Is Extensively Modified with O-linked N-acetylglucosamine. *J. Biol. Chem.* **271**, 28741–28744 (1996).
80. Yuzwa, S. A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **8**, 393–399 (2012).
81. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annual Review of Biochemistry* **80**, 825–858 (2011).
82. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
83. Kim, J. *et al.* RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* **137**, 459–471 (2009).
84. Vermeulen, M. *et al.* Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142**, 967–980 (2010).
85. Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **80**, 5678–5683 (2008).
86. Junot, C., Fenaille, F., Colsch, B. & Bécher, F. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrometry Reviews* **33**, 471–500 (2014).
87. Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **314**, 65–66 (2006).
88. Olsen, J. V. *et al.* A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics* **8**, 2759–2769 (2009).
89. Wu, C. *et al.* A protease for 'middle-down' proteomics. *Nature Methods* **9**, 822–824 (2012).
90. Han, X., Aslanian, A. & Yates, J. R. Mass Spectrometry for Proteomics. *Curr Opin Chem Biol* **12**, 483–490 (2008).
91. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods* **5**, 927–933 (2008).
92. Snijder, J., Rose, R. J., Veesler, D., Johnson, J. E. & Heck, A. J. R. Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angewandte Chemie International Edition* **52**, 4020–4023 (2013).
93. Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. R. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods* **9**, 1084–1086 (2012).
94. Keifer, D. Z., Motwani, T., Teschke, C. M. & Jarrold, M. F. Measurement of the accurate mass of a 50 MDa infectious virus. *Rapid Commun. Mass Spectrom.* **30**, 1957–1962 (2016).
95. Rusling, J. F., Kumar, C. V., Gutkind, J. S. & Patel, V. Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer. *Analyst* **135**, 2496–2511 (2010).

96. Shi, T. *et al.* Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *PNAS* **109**, 15395–15400 (2012).
97. Lin, D., Alborn, W. E., Slebos, R. J. C. & Liebler, D. C. Comparison of Protein Immunoprecipitation-Multiple Reaction Monitoring with ELISA for Assay of Biomarker Candidates in Plasma. *Journal of Proteome Research* **12**, 5996–6003 (2013).
98. Vanmechelen, E. *et al.* Quantification of tau phosphorylated at threonine 181 in human cerebrospinal fluid: a sandwich ELISA with a synthetic phosphopeptide for standardization. *Neurosci. Lett.* **285**, 49–52 (2000).
99. Blennow, K., Hampel, H., Weiner, M. & Zetterberg, H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology* **6**, 131–144 (2010).
100. Jr, J. E. M. *et al.* Characterization of Novel CSF Tau and ptau Biomarkers for Alzheimer’s Disease. *PLOS ONE* **8**, e76523 (2013).
101. Barthélemy, N. R. *et al.* Tau Protein Quantification in Human Cerebrospinal Fluid by Targeted Mass Spectrometry at High Sequence Coverage Provides Insights into Its Primary Structure Heterogeneity. *J. Proteome Res.* **15**, 667–676 (2016).
102. Wildburger, N. C. *et al.* Diversity of Amyloid-beta Proteoforms in the Alzheimer’s Disease Brain. *Scientific Reports* **7**, 9520 (2017).
103. Rogeberg, M., Almdahl, I. S., Wettergreen, M., Nilsson, L. N. G. & Fladby, T. Isobaric Quantification of Cerebrospinal Fluid Amyloid- $\beta$  Peptides in Alzheimer’s Disease: C-Terminal Truncation Relates to Early Measures of Neurodegeneration. *J. Proteome Res.* **14**, 4834–4843 (2015).
104. Klont, F., Hadderingh, M., Horvatovich, P., ten Hacken, N. H. T. & Bischoff, R. Affimers as an Alternative to Antibodies in an Affinity LC–MS Assay for Quantification of the Soluble Receptor of Advanced Glycation End-Products (sRAGE) in Human Serum. *J. Proteome Res.* **17**, 2892–2899 (2018).
105. Suzuki, T. *et al.* Prognostic Role of Molecular Forms of B-Type Natriuretic Peptide in Acute Heart Failure. *Clinical Chemistry* **63**, 880–886 (2017).
106. Hoofnagle, A. N. & Wener, M. H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *Journal of Immunological Methods* **347**, 3–11 (2009).
107. Spencer, C. A. *et al.* Serum Thyroglobulin Autoantibodies: Prevalence, Influence on Serum Thyroglobulin Measurement, and Prognostic Significance in Patients with Differentiated Thyroid Carcinoma. **83**, 7 (1998).
108. Beck, H. C., Overgaard, M. & Melholt Rasmussen, L. Plasma proteomics to identify biomarkers – application to cardiovascular diseases. *Translational Proteomics* **7**, 40–48 (2015).
109. Surinova, S. *et al.* On the Development of Plasma Protein Biomarkers. *Journal of Proteome Research* **10**, 5–16 (2011).
110. Hortin, G. L., Sviridov, D. & Anderson, N. L. High-Abundance Polypeptides of the Human Plasma Proteome Comprising the Top 4 Logs of Polypeptide Abundance. *Clinical Chemistry* **54**, 1608–1616 (2008).
111. Taylor, P. J. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clinical Biochemistry* **38**, 328–334 (2005).
112. Anderson, L. & Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Molecular & Cellular Proteomics* **5**, 573–588 (2006).
113. Kirsch, S., Widart, J., Louette, J., Focant, J.-F. & De Pauw, E. Development of an absolute quantification method targeting growth hormone biomarkers using liquid chromatography coupled to isotope dilution mass spectrometry. *Journal of Chromatography A* **1153**, 300–306 (2007).
114. Ouyang, Z. *et al.* Pellet digestion: a simple and efficient sample preparation technique for LC–MS/MS quantification of large therapeutic proteins in plasma. *Bioanalysis* **4**, 17–28 (2012).

115. Liu, G. *et al.* A Novel and Cost Effective Method of Removing Excess Albumin from Plasma/Serum Samples and Its Impacts on LC-MS/MS Bioanalysis of Therapeutic Proteins. *Analytical Chemistry* **86**, 8336–8343 (2014).
116. Fortin, T. *et al.* Multiple Reaction Monitoring Cubed for Protein Quantification at the Low Nanogram/Milliliter Level in Nondepleted Human Serum. *Analytical Chemistry* **81**, 9343–9352 (2009).
117. Becher, F. *et al.* A simple and rapid LC-MS/MS method for therapeutic drug monitoring of cetuximab: a GPCO-UNICANCER proof of concept study in head-and-neck cancer patients. *Scientific Reports* **7**, 2714 (2017).
118. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Analytical Chemistry* **77**, 6426–6434 (2005).
119. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *PROTEOMICS* **11**, 2019–2026 (2011).
120. Keshishian, H., Addona, T., Burgess, M., Kuhn, E. & Carr, S. A. Quantitative, Multiplexed Assays for Low Abundance Proteins in Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. *Molecular & Cellular Proteomics* **6**, 2212–2229 (2007).
121. Keshishian, H. *et al.* Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell Proteomics* **8**, 2339–2349 (2009).
122. Dupré, M. *et al.* Multiplex Quantification of Protein Toxins in Human Biofluids and Food Matrices Using Immunoextraction and High-Resolution Targeted Mass Spectrometry. *Analytical Chemistry* **87**, 8473–8480 (2015).
123. Hansbauer, E.-M. *et al.* Rapid Detection of Abrin Toxin and Its Isoforms in Complex Matrices by Immuno-Extraction and Quantitative High Resolution Targeted Mass Spectrometry. *Analytical Chemistry* **89**, 11719–11727 (2017).
124. Chen, Y.-T. *et al.* Quantitative Analysis of Prostate Specific Antigen Isoforms Using Immunoprecipitation and Stable Isotope Labeling Mass Spectrometry. *Analytical Chemistry* **87**, 545–553 (2015).
125. Gao, J., Meyer, K., Borucki, K. & Ueland, P. M. Multiplex Immuno-MALDI-TOF MS for Targeted Quantification of Protein Biomarkers and Their Proteoforms Related to Inflammation and Renal Dysfunction. *Analytical Chemistry* **90**, 3366–3373 (2018).
126. Anderson, N. L. *et al.* Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *Journal of Proteome Research* **3**, 235–244 (2004).
127. Hsiao, Y.-C. *et al.* Development of a Multiplexed Assay for Oral Cancer Candidate Biomarkers Using Peptide Immunoaffinity Enrichment and Targeted Mass Spectrometry. *Molecular & Cellular Proteomics* **16**, 1829–1849 (2017).
128. Ippoliti, P. J. *et al.* Automated Microchromatography Enables Multiplexing of Immunoaffinity Enrichment of Peptides to Greater than 150 for Targeted MS-Based Assays. *Analytical Chemistry* **88**, 7548–7555 (2016).
129. Burgess, M. W., Keshishian, H., Mani, D. R., Gillette, M. A. & Carr, S. A. Simplified and Efficient Quantification of Low-abundance Proteins at Very High Multiplex via Targeted Mass Spectrometry. *Molecular & Cellular Proteomics* **13**, 1137–1149 (2014).
130. Kushnir, M. M. *et al.* Measurement of Thyroglobulin by Liquid Chromatography–Tandem Mass Spectrometry in Serum and Plasma in the Presence of Antithyroglobulin Autoantibodies. *Clinical Chemistry* **59**, 982–990 (2013).
131. Razavi, M., Leigh Anderson, N., Pope, M. E., Yip, R. & Pearson, T. W. High precision quantification of human plasma proteins using the automated SISCAPA Immuno-MS workflow. *New Biotechnology* **33**, 494–502 (2016).
132. Kuhn, E. *et al.* Interlaboratory Evaluation of Automated, Multiplexed Peptide Immunoaffinity Enrichment Coupled to Multiple Reaction Monitoring Mass Spectrometry for Quantifying Proteins in Plasma. *Molecular & Cellular Proteomics* **11**, M111.013854 (2012).



133. Yang, X. *et al.* Mass Spectrometric Quantitation of Tubulin Acetylation from Pepsin-Digested Rat Brain Tissue Using a Novel Stable-Isotope Standard and Capture by Anti-Peptide Antibody (SISCAPA) Method. *Anal. Chem.* **90**, 2155–2163 (2018).
134. Baird, G. S. & Hoofnagle, A. N. A Novel Discovery Platform: Aptamers for the Quantification of Human Proteins. *Clinical Chemistry* **63**, 1061–1062 (2017).
135. Zhuo, Z. *et al.* Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *International Journal of Molecular Sciences* **18**, 2142 (2017).
136. Makawita, S. & Diamandis, E. P. The Bottleneck in the Cancer Biomarker Pipeline and Protein Quantification through Mass Spectrometry–Based Approaches: Current Strategies for Candidate Verification. *Clinical Chemistry* **56**, 212–222 (2010).
137. Chen, E. I., Cociorva, D., Norris, J. L. & Yates, J. R. Optimization of Mass Spectrometry-Compatible Surfactants for Shotgun Proteomics. *J. Proteome Res.* **6**, 2529–2538 (2007).
138. Rodriguez, J., Gupta, N., Smith, R. D. & Pevzner, P. A. Does Trypsin Cut Before Proline? *J. Proteome Res.* **7**, 300–305 (2008).
139. Vandermarliere, E., Mueller, M. & Martens, L. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrometry Reviews* **32**, 453–465 (2013).
140. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology* **32**, 223–226 (2014).
141. Kelstrup, C. D. *et al.* Rapid and Deep Proteomes by Faster Sequencing on a Benchtop Quadrupole Ultra-High-Field Orbitrap Mass Spectrometer. *Journal of Proteome Research* **13**, 6187–6195 (2014).
142. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
143. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nature Protocols* **10**, 887–903 (2015).
144. Purvine, S., Eppel\*, J.-T., Yi, E. C. & Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *PROTEOMICS* **3**, 847–850 (2003).
145. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods* **1**, 39–45 (2004).
146. Panchaud, A. *et al.* Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Analytical Chemistry* **81**, 6481–6488 (2009).
147. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis\* □S. **17**
148. Kang, Y., Burton, L., Lau, A. & Tate, S. SWATH-ID: An instrument method which combines identification and quantification in a single analysis. *PROTEOMICS* **17**, 1500522 (2017).
149. Tsou, C.-C., Tsai, C.-F., Teo, G. C., Chen, Y.-J. & Nesvizhskii, A. I. Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers. *PROTEOMICS* **16**, 2257–2271 (2016).
150. Zolg, D. P. *et al.* Building ProteomeTools based on a complete synthetic human proteome. *Nature Methods* **14**, 259–262 (2017).
151. Egertson, J. D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nature Methods* **10**, 744–746 (2013).
152. Lambert, J.-P. *et al.* Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods* **10**, 1239–1245 (2013).
153. Muntel, J. *et al.* Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *Journal of Proteome Research* **14**, 4752–4762 (2015).
154. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology* **14**, e8126 (2018).

155. Krautkramer, K. A., Reiter, L., Denu, J. M. & Dowell, J. A. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. *Journal of Proteome Research* **14**, 3252–3262 (2015).
156. Sidoli, S. *et al.* Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. *Molecular & Cellular Proteomics* **14**, 2420–2428 (2015).
157. Lawrence, R. T., Searle, B. C., Llovet, A. & Villén, J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nature Methods* **13**, 431–434 (2016).
158. Rosenberger, G. *et al.* Inference and quantification of peptidofoms in large sample cohorts by SWATH-MS. *Nature Biotechnology* **35**, 781–788 (2017).
159. Parker, B. L. *et al.* Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. *Sci. Signal.* **8**, rs6–rs6 (2015).
160. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J. & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nature Methods* **15**, 440–448 (2018).
161. Mair, W. *et al.* FLEXITau: Quantifying Post-translational Modifications of Tau Protein *in Vitro* and in Human Disease. *Analytical Chemistry* **88**, 3704–3714 (2016).
162. Parker, B. L., Larsen, M. R., Edvinsson, L. I. & Povlsen, G. K. Signal Transduction in Cerebral Arteries after Subarachnoid Hemorrhage—A Phosphoproteomic Approach. *Journal of Cerebral Blood Flow & Metabolism* **33**, 1259–1269 (2013).
163. Ruprecht, B. *et al.* Lapatinib Resistance in Breast Cancer Cells Is Accompanied by Phosphorylation-Mediated Reprogramming of Glycolysis. *Cancer Research* **77**, 1842–1853 (2017).
164. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
165. Giansanti, P. *et al.* An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Reports* **11**, 1834–1843 (2015).
166. Gauci, S. *et al.* Lys-N and Trypsin Cover Complementary Parts of the Phosphoproteome in a Refined SCX-Based Approach. *Anal. Chem.* **81**, 4493–4501 (2009).
167. Swaney, D. L., Wenger, C. D. & Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *Journal of Proteome Research* **9**, 1323–1329 (2010).
168. Guo, X., Trudgian, D. C., Lemoff, A., Yadavalli, S. & Mirzaei, H. Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics. *Molecular & Cellular Proteomics* **13**, 1573–1584 (2014).
169. Aye, T. T. *et al.* Proteome-wide protein concentrations in the human heart. *Molecular BioSystems* **6**, 1917 (2010).
170. Benevento, M. *et al.* Adenovirus Composition, Proteolysis, and Disassembly Studied by In-depth Qualitative and Quantitative Proteomics. *Journal of Biological Chemistry* **289**, 11421–11430 (2014).
171. Low, T. Y. *et al.* Quantitative and Qualitative Proteome Characteristics Extracted from In-Depth Integrated Genomics and Proteomics Analysis. *Cell Reports* **5**, 1469–1478 (2013).
172. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* **7**, 548–548 (2014).
173. Schröder, C. U. *et al.* Neprosin, a Selective Prolyl Endoprotease for Bottom-up Proteomics and Histone Mapping. *Molecular & Cellular Proteomics* **16**, 1162–1171 (2017).
174. Nesvizhskii, A. I. & Aebersold, R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Molecular & Cellular Proteomics* **4**, 1419–1440 (2005).
175. Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V. & Tsybin, Y. O. Proteome Digestion Specificity Analysis for Rational Design of Extended Bottom-up and Middle-down Proteomics Experiments. *Journal of Proteome Research* **12**, 5558–5569 (2013).

176. Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S. & Matsuo, H. Characterization of KEX2-encoded endopeptidase from yeast *Saccharomyces cerevisiae*. *Biochemical and Biophysical Research Communications* **159**, 305–311 (1989).
177. Laskay, Ü. A., Srzentić, K., Monod, M. & Tsybin, Y. O. Extended bottom-up proteomics with secreted aspartic protease Sap9. *Journal of Proteomics* **110**, 20–31 (2014).
178. Srzentić, K. *et al.* Advantages of Extended Bottom-Up Proteomics Using Sap9 for Analysis of Monoclonal Antibodies. *Analytical Chemistry* **86**, 9945–9953 (2014).
179. Srzentić, K. *et al.* Chemical-Mediated Digestion: An Alternative Realm for Middle-down Proteomics? *Journal of Proteome Research* **17**, 2005–2016 (2018).
180. Andreev, Y. A., Kozlov, S. A., Vassilevski, A. A. & Grishin, E. V. Cyanogen bromide cleavage of proteins in salt and buffer solutions. *Analytical Biochemistry* **407**, 144–146 (2010).
181. Morla, A. *et al.* Analysis of high molecular mass proteins larger than 150 kDa using cyanogen bromide cleavage and conventional 2-DE. *ELECTROPHORESIS* **29**, 4158–4168 (2008).
182. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
183. Fischle, W. *et al.* Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).
184. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline Isomerization of Histone H3 Regulates Lysine Methylation and Gene Expression. *Cell* **126**, 905–916 (2006).
185. Clements, A. *et al.* Structural Basis for Histone and Phosphohistone Binding by the GCN5 Histone Acetyltransferase. *Molecular Cell* **12**, 461–473 (2003).
186. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols* **2**, 933–938 (2007).
187. Garcia, B. A., Thomas, C. E., Kelleher, N. L. & Mizzen, C. A. Tissue-Specific Expression and Post-Translational Modification of Histone H3 Variants. *J. Proteome Res.* **7**, 4225–4236 (2008).
188. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols* **2**, 933–938 (2007).
189. Sidoli, S., Vandamme, J., Salcini, A. E. & Jensen, O. N. Dynamic changes of histone H3 marks during *Caenorhabditis elegans* lifecycle revealed by middle-down proteomics. *PROTEOMICS* **16**, 459–464 (2016).
190. Vandamme, J. *et al.* H3K23me2 is a new heterochromatic mark in *Caenorhabditis elegans*. *Nucleic Acids Research* gkv1063 (2015). doi:10.1093/nar/gkv1063
191. Jiang, T. *et al.* Middle-Down Characterization of the Cell Cycle Dependence of Histone H4 Posttranslational Modifications and Proteoforms. *PROTEOMICS* **18**, 1700442 (2018).
192. Chen, Y., Mao, P. & Wang, D. Quantitation of Intact Proteins in Human Plasma Using Top-Down Parallel Reaction Monitoring-MS. *Anal. Chem.* **90**, 10650–10653 (2018).
193. Savaryn, J. P., Catherman, A. D., Thomas, P. M., Abecassis, M. M. & Kelleher, N. L. The emergence of top-down proteomics in clinical research. *Genome Medicine* **5**, 53 (2013).
194. Dang, X. *et al.* The first pilot project of the consortium for top-down proteomics: A status report. *PROTEOMICS* **14**, 1130–1140 (2014).
195. Liu, Z. & Schey, K. L. Fragmentation of multiply-charged intact protein ions using MALDI TOF-TOF mass spectrometry. *Journal of the American Society for Mass Spectrometry* **19**, 231–238 (2008).
196. Chanthamontri, C., Liu, J. & McLuckey, S. A. Charge state dependent fragmentation of gaseous  $\alpha$ -synuclein cations via ion trap and beam-type collisional activation. *International Journal of Mass Spectrometry* **283**, 9–16 (2009).
197. Xia, Y., Liang, X. & McLuckey, S. A. Ion Trap versus Low-Energy Beam-Type Collision-Induced Dissociation of Protonated Ubiquitin Ions. *Analytical Chemistry* **78**, 1218–1227 (2006).
198. Riley, N. M., Westphall, M. S. & Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Analytical Chemistry* **87**, 7109–7116 (2015).
199. Cannon, J. R., Kluwe, C., Ellington, A. & Brodbelt, J. S. Characterization of green fluorescent proteins by 193 nm ultraviolet photodissociation mass spectrometry. *PROTEOMICS* **14**, 1165–1173 (2014).

200. Fort, K. L. *et al.* Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Analytical Chemistry* **88**, 2303–2310 (2016).
201. Palumbo, A. M. *et al.* Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrometry Reviews* **30**, 600–625 (2011).
202. Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry\*□S. 10
203. Cleland, T. P. *et al.* High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *Journal of Proteome Research* **16**, 2072–2079 (2017).
204. Tran, J. C. & Doucette, A. A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Analytical Chemistry* **80**, 1568–1573 (2008).
205. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
206. Catherman, A. D. *et al.* Large-scale Top-down Proteomics of the Human Proteome: Membrane Proteins, Mitochondria, and Senescence. *Molecular & Cellular Proteomics* **12**, 3465–3473 (2013).
207. McCool, E. N. *et al.* Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the *Escherichia coli* Proteome. *Analytical Chemistry* **90**, 5529–5533 (2018).
208. Capriotti, A. L., Cavaliere, C., Foglia, P., Samperi, R. & Laganà, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *Journal of Chromatography A* **1218**, 8760–8776 (2011).
209. Ntai, I. *et al.* Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk. *Proceedings of the National Academy of Sciences* **115**, 4140–4145 (2018).
210. Kellie, J. F. *et al.* Quantitative Measurement of Intact Alpha-Synuclein Proteoforms from Post-Mortem Control and Parkinson's Disease Brain Tissue by Intact Protein Mass Spectrometry. *Scientific Reports* **4**, 5797 (2014).
211. Gargano, A. F. G. *et al.* Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics. *Analytical Chemistry* **90**, 6601–6609 (2018).
212. Muneeruddin, K., Nazzaro, M. & Kaltashov, I. A. Characterization of Intact Protein Conjugates and Biopharmaceuticals Using Ion-Exchange Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry and Top-Down Tandem Mass Spectrometry. *Analytical Chemistry* **87**, 10138–10145 (2015).
213. Muneeruddin, K., Nazzaro, M. & Kaltashov, I. A. Characterization of Intact Protein Conjugates and Biopharmaceuticals Using Ion-Exchange Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry and Top-Down Tandem Mass Spectrometry. *Analytical Chemistry* **87**, 10138–10145 (2015).
214. Chen, B. *et al.* Online Hydrophobic Interaction Chromatography–Mass Spectrometry for Top-Down Proteomics. *Analytical Chemistry* **88**, 1885–1891 (2016).
215. Anderson, L. C. *et al.* Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *Journal of Proteome Research* **16**, 1087–1096 (2017).
216. Zabrouskov, V., Ge, Y., Schwartz, J. & Walker, J. W. Unraveling Molecular Complexity of Phosphorylated Human Cardiac Troponin I by Top Down Electron Capture Dissociation/Electron Transfer Dissociation Mass Spectrometry\*□S. 12
217. Rappsilber, J. Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Research* **12**, 1231–1245 (2002).
218. Ishihama, Y. *et al.* Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein\*□S. 8

219. Sun, A. *et al.* Modified Spectral Count Index (mSCI) for Estimation of Protein Abundance by Protein Relative Identification Possibility (RIPpro): A New Proteomic Technological Parameter. *Journal of Proteome Research* **8**, 4934–4942 (2009).
220. Liu, H., Sadygov, R. G. & Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry* **76**, 4193–4201 (2004).
221. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology* **25**, 117–124 (2007).
222. Zybailov, B. *et al.* Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research* **5**, 2339–2347 (2006).
223. Asara, J. M., Christofk, H. R., Freemark, L. M. & Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *PROTEOMICS* **8**, 994–999 (2008).
224. Griffin, N. M. *et al.* Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature Biotechnology* **28**, 83–89 (2010).
225. Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P. & Hale, J. E. Comprehensive Label-Free Method for the Relative Quantification of Proteins from Biological Samples. *Journal of Proteome Research* **4**, 1442–1450 (2005).
226. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute Quantification of Proteins by LCMSE : A Virtue of Parallel ms Acquisition. *Molecular & Cellular Proteomics* **5**, 144–156 (2006).
227. Ning, K., Fermin, D. & Nesvizhskii, A. I. Comparative Analysis of Different Label-Free Mass Spectrometry Based Protein Abundance Estimates and Their Correlation with RNA-Seq Gene Expression Data. *J. Proteome Res.* **11**, 2261–2271 (2012).
228. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
229. Ong, S.-E. *et al.* Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **1**, 376–386 (2002).
230. Zhang, R., Sioma, C. S., Wang, S. & Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Anal. Chem.* **73**, 5142–5149 (2001).
231. Beynon, R. J. & Pratt, J. M. Metabolic Labeling of Proteins for Proteomics. *Molecular & Cellular Proteomics* **4**, 857–872 (2005).
232. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127**, 635–648 (2006).
233. Dengjel, J. *et al.* Quantitative proteomic assessment of very early cellular signaling events. *Nature Biotechnology* **25**, 566–568 (2007).
234. Tzouros, M. *et al.* Development of a 5-plex SILAC Method Tuned for the Quantitation of Tyrosine Phosphorylation Dynamics. *Molecular & Cellular Proteomics* **12**, 3339–3349 (2013).
235. Shenoy, A. & Geiger, T. Super-SILAC: current trends and future perspectives. *Expert Review of Proteomics* **12**, 13–19 (2015).
236. Geiger, T. *et al.* Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature Protocols* **6**, 147–157 (2011).
237. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* **7**, 383–385 (2010).
238. Hebert, A. S. *et al.* Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods* **10**, 332–334 (2013).
239. Rhoads, T. W. *et al.* Neutron-Encoded Mass Signatures for Quantitative Top-Down Proteomics. *Analytical Chemistry* **86**, 2314–2319 (2014).
240. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994–999 (1999).

241. Gygi, S. P., Rist, B., Griffin, T. J., Eng, J. & Aebersold, R. Proteome Analysis of Low-Abundance Proteins Using Multidimensional Chromatography and Isotope-Coded Affinity Tags. *Journal of Proteome Research* **1**, 47–54 (2002).
242. Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *PROTEOMICS* **5**, 4–15 (2005).
243. Östreicher, C., Bartenbacher, S. & Pischetsrieder, M. Targeted proteome analysis with isotope-coded protein labels for monitoring the influence of dietary phytochemicals on the expression of cytoprotective proteins in primary human colon cells. *Journal of Proteomics* **166**, 27–38 (2017).
244. Thompson, A. *et al.* Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* **75**, 1895–1904 (2003).
245. Ross, P. L. *et al.* Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* **3**, 1154–1169 (2004).
246. Böhm, G. *et al.* Low-pH Solid-Phase Amino Labeling of Complex Peptide Digests with TMTs Improves Peptide Identification Rates for Multiplexed Global Phosphopeptide Analysis. *Journal of Proteome Research* **14**, 2500–2510 (2015).
247. McAlister, G. C. *et al.* Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* **84**, 7469–7478 (2012).
248. Werner, T. *et al.* High-Resolution Enabled TMT 8-plexing. *Analytical Chemistry* **84**, 7188–7194 (2012).
249. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8**, 937–940 (2011).
250. Liu, J. M., Sweredoski, M. J. & Hess, S. Improved 6-Plex Tandem Mass Tags Quantification Throughput Using a Linear Ion Trap–High-Energy Collision Induced Dissociation MS<sup>3</sup> Scan. *Analytical Chemistry* **88**, 7471–7475 (2016).
251. Jiang, X. *et al.* Sensitive and Accurate Quantitation of Phosphopeptides Using TMT Isobaric Labeling Technique. *Journal of Proteome Research* **16**, 4244–4252 (2017).
252. McAlister, G. C. *et al.* MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Analytical Chemistry* **86**, 7150–7158 (2014).
253. Berberich, M. J., Paulo, J. A. & Everley, R. A. MS3-IDQ: Utilizing MS3 Spectra beyond Quantification Yields Increased Coverage of the Phosphoproteome in Isobaric Tag Experiments. *Journal of Proteome Research* **17**, 1741–1747 (2018).
254. Virreira Winter, S. *et al.* EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nature Methods* **15**, 527–530 (2018).
255. Yao, X., Freas, A., Ramirez, J., Demirev, P. A. & Fenselau, C. Proteolytic <sup>18</sup>O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Analytical Chemistry* **73**, 2836–2842 (2001).
256. Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Analytical Chemistry* **75**, 6843–6852 (2003).
257. Kovanich, D. *et al.* Applications of stable isotope dimethyl labeling in quantitative proteomics. *Analytical and Bioanalytical Chemistry* **404**, 991–1009 (2012).
258. Boersema, P. J., Geiger, T., Wiśniewski, J. R. & Mann, M. Quantification of the N-glycosylated Secretome by Super-SILAC During Breast Cancer Progression and in Human Blood Samples. *Molecular & Cellular Proteomics* **12**, 158–171 (2013).
259. Wu, Y. *et al.* Five-plex isotope dimethyl labeling for quantitative proteomics. *Chemical Communications* **50**, 1708 (2014).
260. Lau, H.-T., Suh, H. W., Golkowski, M. & Ong, S.-E. Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics. *Journal of Proteome Research* **13**, 4164–4174 (2014).

261. Collier, T. S., Sarkar, P., Rao, B. & Muddiman, D. C. Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *Journal of the American Society for Mass Spectrometry* **21**, 879–889 (2010).
262. Waanders, L. F., Hanke, S. & Mann, M. Top-down quantitation and characterization of SILAC-labeled proteins. *Journal of the American Society for Mass Spectrometry* **18**, 2058–2064 (2007).
263. Brun, V., Masselon, C., Garin, J. & Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *Journal of Proteomics* **72**, 740–749 (2009).
264. Calderón-Celis, F., Encinar, J. R. & Sanz-Medel, A. Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrometry Reviews* **37**, 715–737 (2018).
265. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* **4**, (2008).
266. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555–566 (2012).
267. Rutherford, S. M. & Gilani, G. S. Amino Acid Analysis. in *Current Protocols in Protein Science* (eds. Coligan, J. E., Dunn, B. M., Speicher, D. W. & Wingfield, P. T.) 11.9.1-11.9.37 (John Wiley & Sons, Inc., 2009). doi:10.1002/0471140864.ps1109s58
268. Shuford, C. M. *et al.* Absolute Protein Quantification by Mass Spectrometry: Not as Simple as Advertised. *Analytical Chemistry* **89**, 7406–7415 (2017).
269. Potts, G. K. *et al.* Neucode Labels for Multiplexed, Absolute Protein Quantification. *Analytical Chemistry* **88**, 3295–3303 (2016).
270. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods* **2**, 587–589 (2005).
271. Brun, V. *et al.* Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics. *Molecular & Cellular Proteomics* **6**, 2139–2149 (2007).
272. Singh, S., Springer, M., Steen, J., Kirschner, M. W. & Steen, H. FLEXIQuant: A Novel Tool for the Absolute Quantification of Proteins, and the Simultaneous Identification and Quantification of Potentially Modified Peptides. *Journal of Proteome Research* **8**, 2201–2210 (2009).
273. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M. & Mann, M. A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines\*□S. 13
274. Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. *Journal of Proteome Research* **7**, 1118–1130 (2008).
275. LeBlanc, A. *et al.* Multiplexed MRM-Based Protein Quantitation Using Two Different Stable Isotope-Labeled Peptide Isotopologues for Calibration. *Journal of Proteome Research* **16**, 2527–2536 (2017).
276. Scott, K. B., Turko, I. V. & Phinney, K. W. Quantitative Performance of Internal Standard Platforms for Absolute Protein Quantification Using Multiple Reaction Monitoring-Mass Spectrometry. *Analytical Chemistry* **87**, 4429–4435 (2015).
277. Oeckl, P., Steinacker, P. & Otto, M. Comparison of Internal Standard Approaches for SRM Analysis of Alpha-Synuclein in Cerebrospinal Fluid. *Journal of Proteome Research* **17**, 516–523 (2018).
278. Pritchard, C. *et al.* Quantification of Human Growth Hormone in Serum with a Labeled Protein as an Internal Standard: Essential Considerations. *Analytical Chemistry* **86**, 6525–6532 (2014).
279. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555–566 (2012).
280. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature Methods* **10**, 28–34 (2013).
281. Vidova, V. & Spacil, Z. A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Analytica Chimica Acta* **964**, 7–23 (2017).

282. Abbatiello, S. E., Mani, D. R., Keshishian, H. & Carr, S. A. Automated Detection of Inaccurate and Imprecise Transitions in Peptide Quantification by Multiple Reaction Monitoring Mass Spectrometry. *Clinical Chemistry* **56**, 291–305 (2010).
283. Sherman, J., McKay, M. J., Ashman, K. & Molloy, M. P. How specific is my SRM?: The issue of precursor and product ion redundancy. *PROTEOMICS* **9**, 1120–1123 (2009).
284. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics\*□S. **14**
285. Gallien, S. *et al.* Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **11**, 1709–1723 (2012).
286. Ronsein, G. E. *et al.* Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of Proteomics* **113**, 388–399 (2015).
287. Güzel, C. *et al.* Comparison of Targeted Mass Spectrometry Techniques with an Immunoassay: A Case Study for HSP90 $\alpha$ . *PROTEOMICS - Clinical Applications* **12**, 1700107 (2018).
288. Rougemont, B. *et al.* Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by *Dickeya dadantii* Proteomic Analysis during Plant Infection. *Anal. Chem.* **89**, 1421–1426 (2017).
289. Gallien, S., Kim, S. Y. & Domon, B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Molecular & Cellular Proteomics* **14**, 1630–1644 (2015).
290. Wang, E. H., Combe, P. C. & Schug, K. A. Multiple Reaction Monitoring for Direct Quantitation of Intact Proteins Using a Triple Quadrupole Mass Spectrometer. *Journal of The American Society for Mass Spectrometry* **27**, 886–896 (2016).
291. Ruan, Q., Ji, Q. C., Arnold, M. E., Humphreys, W. G. & Zhu, M. Strategy and Its Implications of Protein Bioanalysis Utilizing High-Resolution Mass Spectrometric Detection of Intact Protein. *Analytical Chemistry* **83**, 8937–8944 (2011).
292. Gucinski, A. C. & Boyne, M. T. Evaluation of Intact Mass Spectrometry for the Quantitative Analysis of Protein Therapeutics. *Analytical Chemistry* **84**, 8045–8051 (2012).
293. van den Broek, I., Niessen, W. M. A. & van Dongen, W. D. Bioanalytical LC–MS/MS of protein-based biopharmaceuticals. *Journal of Chromatography B* **929**, 161–179 (2013).
294. Ramagiri, S. & Garofolo, F. Large molecule bioanalysis using Q-TOF without predigestion and its data processing challenges. *Bioanalysis* **4**, 529–540 (2012).
295. Lanshoeft, C., Cianférani, S. & Heudi, O. Generic Hybrid Ligand Binding Assay Liquid Chromatography High-Resolution Mass Spectrometry-Based Workflow for Multiplexed Human Immunoglobulin G1 Quantification at the Intact Protein Level: Application to Preclinical Pharmacokinetic Studies. *Analytical Chemistry* **89**, 2628–2635 (2017).
296. Geis-Asteggiante, L., Ostrand-Rosenberg, S., Fenselau, C. & Edwards, N. J. Evaluation of Spectral Counting for Relative Quantitation of Proteoforms in Top-Down Proteomics. *Analytical Chemistry* **88**, 10900–10907 (2016).
297. Acosta-Martin, A. E. *et al.* Quantitative Mass Spectrometry Analysis of Intact Hemoglobin A2 by Precursor Ion Isolation and Detection. *Anal. Chem.* **85**, 7971–7975 (2013).
298. Quijada, J. V., Schmitt, N. D., Salisbury, J. P., Auclair, J. R. & Agar, J. N. Heavy Sugar and Heavy Water Create Tunable Intact Protein Mass Increases for Quantitative Mass Spectrometry in Any Feed and Organism. *Analytical Chemistry* **88**, 11139–11146 (2016).
299. Wiese, S., Reidegeld, K. A., Meyer, H. E. & Warscheid, B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *PROTEOMICS* **7**, 340–350 (2007).
300. Hung, C.-W. & Tholey, A. Tandem Mass Tag Protein Labeling for Top-Down Identification and Quantification. *Analytical Chemistry* **84**, 161–170 (2012).
301. Compton, P. D., Kelleher, N. L. & Gunawardena, J. Estimating the Distribution of Protein Post-Translational Modification States by Mass Spectrometry. *Journal of Proteome Research* **17**, 2727–2734 (2018).



302. Dekker, L. J. M. *et al.* Determination of Site-Specific Phosphorylation Ratios in Proteins with Targeted Mass Spectrometry. *Journal of Proteome Research* **17**, 1654–1663 (2018).
303. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. *Journal of Proteome Research* **17**, 2226–2236 (2018).
304. Wu, Z. *et al.* Impact of Phosphorylation on the Mass Spectrometry Quantification of Intact Phosphoproteins. *Analytical Chemistry* **90**, 4935–4939 (2018).
305. Tsai, C.-F. *et al.* Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nature Communications* **6**, (2015).
306. Högberg, A. *et al.* Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nature Communications* **9**, (2018).
307. Hülsmeyer, A. J., Paesold-Burda, P. & Hennot, T. N-Glycosylation Site Occupancy in Serum Glycoproteins Using Multiple Reaction Monitoring Liquid Chromatography-Mass Spectrometry. *Molecular & Cellular Proteomics* **6**, 2132–2138 (2007).
308. Zhu, Z., Go, E. P. & Desaire, H. Absolute Quantitation of Glycosylation Site Occupancy Using Isotopically Labeled Standards and LC-MS. *J. Am. Soc. Mass Spectrom.* **25**, 1012–1017 (2014).
309. Torsetnes, S. B., Nordlund, M. S., Paus, E., Halvorsen, T. G. & Reubsæet, L. Digging Deeper into the Field of the Small Cell Lung Cancer Tumor Marker ProGRP: A Method for Differentiation of Its Isoforms. *Journal of Proteome Research* **12**, 412–420 (2013).
310. Vilà-Rico, M. *et al.* Quantitative analysis of post-translational modifications in human serum transthyretin associated with familial amyloidotic polyneuropathy by targeted LC-MS and intact protein MS. *Journal of Proteomics* **127**, 234–246 (2015).
311. Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N. & Kirschner, M. W. Phosphorylation Analysis by Mass Spectrometry: Myths, Facts, and the Consequences for Qualitative and Quantitative Measurements. *Molecular & Cellular Proteomics* **5**, 172–181 (2006).
312. Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Analytical Chemistry* **78**, 4271–4280 (2006).
313. Muntané, G., Ferrer, I. & Martínez-Vicente, M.  $\alpha$ -synuclein phosphorylation and truncation are normal events in the adult human brain. *Neuroscience* **200**, 106–119 (2012).
314. Schmid, A. W., Fauvet, B., Moniatte, M. & Lashuel, H. A. Alpha-synuclein Post-translational Modifications as Potential Biomarkers for Parkinson Disease and Other Synucleinopathies. *Molecular & Cellular Proteomics* **12**, 3543–3558 (2013).
315. Zhang, J. *et al.* Top-Down Quantitative Proteomics Identified Phosphorylation of Cardiac Troponin I as a Candidate Biomarker for Chronic Heart Failure. *Journal of Proteome Research* **10**, 4054–4065 (2011).
316. Seckler, H. dos S. *et al.* A Targeted, Differential Top-Down Proteomic Methodology for Comparison of ApoA-I Proteoforms in Individuals with High and Low HDL Efflux Capacity. *Journal of Proteome Research* **17**, 2156–2164 (2018).
317. Prado, L. de G. R. *et al.* C9ORF72 and the FTD-ALS spectrum: a systematic review of neuroimaging studies. *Dementia & Neuropsychologia* **9**, 413–421 (2015).
318. Prince, M. *et al.* The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **9**, 63-75.e2 (2013).
319. Neumann, M. *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science* **314**, 130–133 (2006).
320. Neumann, M. *et al.* Abundant FUS-immunoreactive pathology in neuronal intermediate filament inclusion disease. *Acta Neuropathol* **118**, 605 (2009).
321. Wang, W. *et al.* The inhibition of TDP-43 mitochondrial localization blocks its neuronal toxicity. *Nature Medicine* **22**, 869–878 (2016).
322. Ringholz, G. M. *et al.* Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* **65**, 586–590 (2005).

323. Ciura, S. *et al.* Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis: *Annals of Neurology* Vol 73 No 6 June 2013. *Annals of Neurology* n/a-n/a (2013). doi:10.1002/ana.23946
324. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
325. Renton, A. E. *et al.* A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
326. Nordin, A. *et al.* Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Human Molecular Genetics* **24**, 3133–3142 (2015).
327. van Blitterswijk, M. *et al.* Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *The Lancet Neurology* **12**, 978–988 (2013).
328. Levine, T. P., Daniels, R. D., Gatta, A. T., Wong, L. H. & Hayes, M. J. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics* **29**, 499–503 (2013).
329. Xiao, S. *et al.* Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis: C9orf72 Isoforms in ALS. *Annals of Neurology* **78**, 568–583 (2015).
330. Frick, P. *et al.* Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathologica Communications* **6**, 72 (2018).
331. Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nature Medicine* **24**, 313–325 (2018).
332. Balendra, R. & Isaacs, A. M. C9orf72 -mediated ALS and FTD: multiple pathways to disease. *Nature Reviews Neurology* **14**, 544 (2018).
333. Kumar, V., Hasan, G. M. & Hassan, M. I. Unraveling the Role of RNA Mediated Toxicity of C9orf72 Repeats in C9-FTD/ALS. *Frontiers in Neuroscience* **11**, (2017).
334. Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences* **110**, E4968–E4977 (2013).
335. Lehmer, C. *et al.* Poly-GP in cerebrospinal fluid links C9orf72 -associated dipeptide repeat expression to the asymptomatic phase of ALS/FTD. *EMBO Molecular Medicine* **9**, 859–868 (2017).
336. Gendron, T. F. *et al.* Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. *Acta Neuropathologica* **130**, 559–573 (2015).
337. Balendra, R., Moens, T. G. & Isaacs, A. M. Specific biomarkers for C9orf72 FTD/ALS could expedite the journey towards effective therapies. *EMBO Molecular Medicine* **9**, 853–855 (2017).
338. Baker, M. *et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442**, 916–919 (2006).
339. Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442**, 920–924 (2006).
340. Abella, V. *et al.* Progranulin as a biomarker and potential therapeutic agent. *Drug Discovery Today* **22**, 1557–1564 (2017).
341. Ghetti, B. *et al.* Invited review: Frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging. *Neuropathology and Applied Neurobiology* **41**, 24–46 (2015).
342. Fourier, A. *et al.* C9orf72 Protein Plasmatic Concentrations Are Similar between **C9ORF72** Expansion Carriers and Noncarriers in Frontotemporal Dementia. *Dementia and Geriatric Cognitive Disorders* 180–185 (2018). doi:10.1159/000492963

343. Van Mossevelde, S., van der Zee, J., Cruts, M. & Van Broeckhoven, C. Relationship between C9orf72 repeat size and clinical phenotype. *Current Opinion in Genetics & Development* **44**, 117–124 (2017).
344. Blitterswijk, M. van *et al.* Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *The Lancet Neurology* **12**, 978–988 (2013).
345. Becher, F. *et al.* Development of a Direct Assay for Measuring Intracellular AZT Triphosphate in Humans Peripheral Blood Mononuclear Cells. *Anal. Chem.* **74**, 4220–4227 (2002).
346. Oeckl, P., Steinacker, P., Feneberg, E. & Otto, M. Cerebrospinal fluid proteomics and protein biomarkers in frontotemporal lobar degeneration: Current status and future perspectives. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1854**, 757–768 (2015).
347. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
348. Shevchenko, G., Musunuri, S., Wetterhall, M. & Bergquist, J. Comparison of Extraction Methods for the Comprehensive Analysis of Mouse Brain Proteome using Shotgun-based Mass Spectrometry. *J. Proteome Res.* **11**, 2441–2451 (2012).
349. Chen, S., Wu, J.-T. & Huang, R. Evaluation of surrogate matrices for standard curve preparation in tissue bioanalysis. *Bioanalysis* **4**, 2579–2587 (2012).
350. Kim, Y. J. *et al.* Verification of the Biomarker Candidates for Non-small-cell Lung Cancer Using a Targeted Proteomics Approach. *J. Proteome Res.* **14**, 1412–1419 (2015).
351. Kamphuis, W. *et al.* GFAP Isoforms in Adult Mouse Brain with a Focus on Neurogenic Astrocytes and Reactive Astroglia in Mouse Models of Alzheimer Disease. *PLoS ONE* **7**, e42823 (2012).
352. Sofroniew, M. V. Molecular dissection of reactive astroglia and glial scar formation. *Trends in Neurosciences* **32**, 638–647 (2009).
353. Middeldorp, J. *et al.* Intermediate filament transcription in astrocytes is repressed by proteasome inhibition. *The FASEB Journal* **23**, 2710–2726 (2009).
354. Tripathi, P. *et al.* Reactive Astrocytes Promote ALS-like Degeneration and Intracellular Protein Aggregation in Human Motor Neurons by Disrupting Autophagy through TGF- $\beta$ 1. *Stem Cell Reports* **9**, 667–680 (2017).
355. Schludi, M. H. *et al.* Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. *Acta Neuropathologica* **134**, 241–254 (2017).
356. Webster, C. P., Smith, E. F., Grierson, A. J. & De Vos, K. J. C9orf72 plays a central role in Rab GTPase-dependent regulation of autophagy. *Small GTPases* **9**, 399–408 (2018).
357. Moeton, M. *et al.* GFAP isoforms control intermediate filament network dynamics, cell morphology, and focal adhesions. *Cellular and Molecular Life Sciences* **73**, 4101–4120 (2016).
358. Stephens, A. N. *et al.* Post-Translational Modifications and Protein-Specific Isoforms in Endometriosis Revealed by 2D DIGE. *Journal of Proteome Research* **9**, 2438–2449 (2010).
359. Lopez-Gonzalez, R. *et al.* Poly(GR) in C9ORF72 -Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron* **92**, 383–391 (2016).
360. Wu, Y. *et al.* Mitogen-activated protein kinase-mediated phosphorylation of peroxiredoxin 6 regulates its phospholipase A<sub>2</sub> activity. *Biochemical Journal* **419**, 669–679 (2009).
361. Fontaine, S. N. *et al.* The active Hsc70/tau complex can be exploited to enhance tau turnover without damaging microtubule dynamics. *Human Molecular Genetics* **24**, 3971–3981 (2015).
362. Amlie-Wolf, A. *et al.* Transcriptomic Changes Due to Cytoplasmic TDP-43 Expression Reveal Dysregulation of Histone Transcripts and Nuclear Chromatin. *PLoS ONE* **10**, e0141836 (2015).
363. Shanmughapriya, S. *et al.* SPG7 Is an Essential and Conserved Component of the Mitochondrial Permeability Transition Pore. *Molecular Cell* **60**, 47–62 (2015).
364. Smith, E. F., Shaw, P. J. & De Vos, K. J. The role of mitochondria in amyotrophic lateral sclerosis. *Neuroscience Letters* (2017). doi:10.1016/j.neulet.2017.06.052
365. Lewerenz, J. & Maher, P. Chronic Glutamate Toxicity in Neurodegenerative Diseases—What is the Evidence? *Frontiers in Neuroscience* **9**, (2015).

366. Chou, C.-C. *et al.* TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. *Nature Neuroscience* **21**, 228–239 (2018).
367. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61 (2015).
368. Rohan, Z., Matej, R., Rusina, R. & Kovacs, G. G. Oligodendroglial Response in the Spinal Cord in TDP-43 Proteinopathy with Motor Neuron Involvement. *Neurodegenerative Diseases* **14**, 117–124 (2014).
369. Philips, T. *et al.* Oligodendrocyte dysfunction in the pathogenesis of amyotrophic lateral sclerosis. *Brain* **136**, 471–482 (2013).
370. Starr, A. & Sattler, R. Synaptic dysfunction and altered excitability in C9ORF72 ALS/FTD. *Brain Research* **1693**, 98–108 (2018).
371. Mori, K. *et al.* hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathologica* **125**, 413–423 (2013).
372. Davidson, Y. S. *et al.* Heterogeneous ribonuclear protein A3 (hnRNP A3) is present in dipeptide repeat protein containing inclusions in Frontotemporal Lobar Degeneration and Motor Neurone disease associated with expansions in C9orf72 gene. *Acta Neuropathologica Communications* **5**, (2017).
373. Shirai, Y.-T., Suzuki, T., Morita, M., Takahashi, A. & Yamamoto, T. Multifunctional roles of the mammalian CCR4–NOT complex in physiological phenomena. *Frontiers in Genetics* **5**, (2014).
374. Umoh, M. E. *et al.* A proteomic network approach across the ALS-FTD disease spectrum resolves clinical phenotypes and genetic vulnerability in human brain. *EMBO Molecular Medicine* **10**, 48–62 (2018).
375. Bruderer, R. *et al.* New targeted approaches for the quantification of data-independent acquisition mass spectrometry. *PROTEOMICS* **17**, 1700021 (2017).
376. Meeter, L. H. H. *et al.* POLY-GP DIPEPTIDE REPEATS AND NEUROFILAMENT LIGHT CHAIN AS BIOMARKERS IN PRESYMPTOMATIC AND SYMPTOMATIC FRONTOTEMPORAL DEMENTIA CAUSED BY C9ORF72 REPEAT EXPANSIONS. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **13**, P1228–P1229 (2017).
377. Kametani, F. *et al.* Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains. *Scientific Reports* **6**, (2016).
378. Porta, S. *et al.* Patient-derived frontotemporal lobar degeneration brain extracts induce formation and spreading of TDP-43 pathology in vivo. *Nature Communications* **9**, (2018).
379. Nonaka, T. *et al.* Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. *Cell Reports* **4**, 124–134 (2013).
380. Majumder, V., Gregory, J. M., Barria, M. A., Green, A. & Pal, S. TDP-43 as a potential biomarker for amyotrophic lateral sclerosis: a systematic review and meta-analysis. *BMC Neurology* **18**, (2018).
381. Fanciulli, A. & Wenning, G. K. Multiple-System Atrophy. *New England Journal of Medicine* **372**, 249–263 (2015).
382. Fauvet, B. *et al.*  $\alpha$ -Synuclein in Central Nervous System and from Erythrocytes, Mammalian Cells, and *Escherichia coli* Exists Predominantly as Disordered Monomer. *Journal of Biological Chemistry* **287**, 15345–15364 (2012).
383. Wang, W. *et al.* Caspase-1 causes truncation and aggregation of the Parkinson's disease-associated protein  $\alpha$ -synuclein. *Proceedings of the National Academy of Sciences* **113**, 9587–9592 (2016).
384. Bartels, T., Choi, J. G. & Selkoe, D. J.  $\alpha$ -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107–110 (2011).
385. Emamzadeh, F. Alpha-synuclein structure, functions, and interactions. *Journal of Research in Medical Sciences* **21**, 29 (2016).
386. Al-Wandi, A. *et al.* Absence of  $\alpha$ -synuclein affects dopamine metabolism and synaptic markers in the striatum of aging mice. *Neurobiology of Aging* **31**, 796–804 (2010).

387. Chandra, S. *et al.* Double-knockout mice for  $\alpha$ - and  $\beta$ -synucleins: Effect on synaptic functions. *Proceedings of the National Academy of Sciences* **101**, 14966–14971 (2004).
388. Greten-Harrison, B. *et al.*  $\alpha\beta\gamma$ -Synuclein triple knockout mice reveal age-dependent neuronal dysfunction. *6*
389. Dettmer, U., Newman, A. J., von Saucken, V. E., Bartels, T. & Selkoe, D. KTKEGV repeat motifs are key mediators of normal  $\alpha$ -synuclein tetramerization: Their mutation causes excess monomers and neurotoxicity. *Proceedings of the National Academy of Sciences* **112**, 9596–9601 (2015).
390. McLean, J. R., Hallett, P. J., Cooper, O., Stanley, M. & Isacson, O. Transcript expression levels of full-length alpha-synuclein and its three alternatively spliced variants in Parkinson's disease brain regions and in a transgenic mouse model of alpha-synuclein overexpression. *Molecular and Cellular Neuroscience* **49**, 230–239 (2012).
391. Beyer, K. *et al.* Differential expression of alpha-synuclein isoforms in dementia with Lewy bodies. *Neuropathology and Applied Neurobiology* **30**, 601–607 (2004).
392. Beyer, K. *et al.* Identification and characterization of a new alpha-synuclein isoform and its role in Lewy body diseases. *Neurogenetics* **9**, 15–23 (2008).
393. Beyer, K. *et al.* Differential expression of alpha-synuclein, parkin, and synphilin-1 isoforms in Lewy body disease. *neurogenetics* **9**, 163–172 (2008).
394. Ana Gámez-Valero & Katrin Beyer. Alternative Splicing of Alpha- and Beta-Synuclein Genes Plays Differential Roles in Synucleinopathies. *Genes* **9**, 63 (2018).
395. Cardo, L. F. *et al.* Alpha-synuclein transcript isoforms in three different brain regions from Parkinson's disease and healthy subjects in relation to the SNCA rs356165/rs11931074 polymorphisms. *Neuroscience Letters* **562**, 45–49 (2014).
396. Brudek, T. *et al.* Altered  $\alpha$ -synuclein, parkin, and synphilin isoform levels in multiple system atrophy brains. *Journal of Neurochemistry* **136**, 172–185 (2016).
397. Vinnakota, R. L. *et al.* Identification of an Alternatively Spliced  $\alpha$ -Synuclein Isoform That Generates a 41-Amino Acid N-Terminal Truncated Peptide, 41-syn: Role in Dopamine Homeostasis. *ACS Chem. Neurosci.* (2018). doi:10.1021/acscchemneuro.8b00140
398. Locascio, J. J. *et al.* Association between  $\alpha$ -synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain* **138**, 2659–2671 (2015).
399. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of  $\alpha$ -synuclein: from structure and toxicity to therapeutic target. *Nature Reviews Neuroscience* **14**, 38–48 (2013).
400. Bengoa-Vergniory, N., Roberts, R. F., Wade-Martins, R. & Alegre-Abarrategui, J. Alpha-synuclein oligomers: a new hope. *Acta Neuropathologica* **134**, 819–838 (2017).
401. Ingelsson, M. Alpha-Synuclein Oligomers—Neurotoxic Molecules in Parkinson's Disease and Other Lewy Body Disorders. *Frontiers in Neuroscience* **10**, (2016).
402. Benskey, M. J., Perez, R. G. & Manfredsson, F. P. The contribution of alpha synuclein to neuronal survival and function - Implications for Parkinson's disease. *Journal of Neurochemistry* **137**, 331–359 (2016).
403. Osterberg, V. R. *et al.* Progressive Aggregation of Alpha-Synuclein and Selective Degeneration of Lewy Inclusion-Bearing Neurons in a Mouse Model of Parkinsonism. *Cell Reports* **10**, 1252–1260 (2015).
404. Fujiwara, H. *et al.*  $\alpha$ -Synuclein is phosphorylated in synucleinopathy lesions. *Nature Cell Biology* **4**, 160–164 (2002).
405. Anderson, J. P. *et al.* Phosphorylation of Ser-129 Is the Dominant Pathological Modification of  $\alpha$ -Synuclein in Familial and Sporadic Lewy Body Disease. *Journal of Biological Chemistry* **281**, 29739–29752 (2006).
406. Zhou, J. *et al.* Changes in the solubility and phosphorylation of  $\alpha$ -synuclein over the course of Parkinson's disease. *Acta Neuropathol* **121**, 695–704 (2011).
407. Foulds, P. G. *et al.* Post mortem cerebrospinal fluid  $\alpha$ -synuclein levels are raised in multiple system atrophy and distinguish this from the other  $\alpha$ -synucleinopathies, Parkinson's disease and Dementia with Lewy bodies. *Neurobiology of Disease* **45**, 188–195 (2012).

408. Foulds, P. G. *et al.* A longitudinal study on  $\alpha$ -synuclein in blood plasma as a biomarker for Parkinson's disease. *Scientific Reports* **3**, (2013).
409. Paleologou, K. E. *et al.* Phosphorylation at S87 Is Enhanced in Synucleinopathies, Inhibits - Synuclein Oligomerization, and Influences Synuclein-Membrane Interactions. *Journal of Neuroscience* **30**, 3184–3198 (2010).
410. Baba, M. *et al.* Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am J Pathol* **152**, 879–884 (1998).
411. Terada, M. *et al.* The effect of truncation on prion-like properties of  $\alpha$ -synuclein. *Journal of Biological Chemistry* **293**, 13910–13920 (2018).
412. Wateren, I. M. van der, Knowles, T. P. J., Buell, A. K., Dobson, C. M. & Galvagnion, C. C-terminal truncation of  $\alpha$ -synuclein promotes amyloid fibril amplification at physiological pH. *Chem. Sci.* **9**, 5506–5516 (2018).
413. Ma, L. *et al.* C-terminal truncation exacerbates the aggregation and cytotoxicity of  $\alpha$ -Synuclein: A vicious cycle in Parkinson's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1864**, 3714–3725 (2018).
414. Bassil, F. *et al.* Reducing C-terminal truncation mitigates synucleinopathy and neurodegeneration in a transgenic model of multiple system atrophy. *Proceedings of the National Academy of Sciences* **113**, 9593–9598 (2016).
415. Games, D. *et al.* Reducing C-Terminal-Truncated Alpha-Synuclein by Immunotherapy Attenuates Neurodegeneration and Propagation in Parkinson's Disease-Like Models. *Journal of Neuroscience* **34**, 9441–9454 (2014).
416. Levine, P. M. *et al.* O-GlcNAc modification inhibits the calpain-mediated cleavage of  $\alpha$ -synuclein. *Bioorganic & Medicinal Chemistry* **25**, 4977–4982 (2017).
417. Tofaris, G. K. *et al.* Ubiquitin ligase Nedd4 promotes  $\alpha$ -synuclein degradation by the endosomal–lysosomal pathway. *CELL BIOLOGY* **6**
418. Rott, R. *et al.*  $\alpha$ -Synuclein fate is determined by USP9X-regulated monoubiquitination. **6**
419. Abeywardana, T., Lin, Y. H., Rott, R., Engelender, S. & Pratt, M. R. Site-Specific Differences in Proteasome-Dependent Degradation of Monoubiquitinated  $\alpha$ -Synuclein. *Chemistry & Biology* **20**, 1207–1213 (2013).
420. Giasson, B. I. Initiation and Synergistic Fibrillization of Tau and Alpha-Synuclein. *Science* **300**, 636–640 (2003).
421. Chavarría, C. & Souza, J. M. Oxidation and nitration of  $\alpha$ -synuclein and their implications in neurodegenerative diseases. *Archives of Biochemistry and Biophysics* **533**, 25–32 (2013).
422. Burai, R., Ait-Bouziad, N., Chiki, A. & Lashuel, H. Elucidating the role of site-specific nitration of  $\alpha$ -synuclein in the pathogenesis of Parkinson's disease via protein semisynthesis and mutagenesis. *Journal of the American Chemical Society* **135**, 1207–1213 (2013).
423. Liu, Y., Qiang, M., Wei, Y. & He, R. A novel molecular mechanism for nitrated  $\alpha$ -synuclein-induced cell death. *Journal of Molecular Cell Biology* **3**, 239–249 (2011).
424. Goldman, J. G., Williams-Gray, C., Barker, R. A., Duda, J. E. & Galvin, J. E. The spectrum of cognitive impairment in Lewy body diseases: Cognitive Impairment in Lewy Body Diseases. *Movement Disorders* **29**, 608–621 (2014).
425. Guo, J. L. *et al.* Distinct  $\alpha$ -Synuclein Strains Differentially Promote Tau Inclusions in Neurons. *Cell* **154**, 103–117 (2013).
426. Thomas, A. J. *et al.* Improving the identification of dementia with Lewy bodies in the context of an Alzheimer's-type dementia. *Alzheimer's Research & Therapy* **10**, 27 (2018).
427. Moussaud, S. *et al.* Alpha-synuclein and tau: teammates in neurodegeneration? *Molecular Neurodegeneration* **9**, 43 (2014).
428. Larson, M. E. *et al.* Soluble  $\alpha$ -Synuclein Is a Novel Modulator of Alzheimer's Disease Pathophysiology. *Journal of Neuroscience* **32**, 10253–10266 (2012).
429. Slaets, S. *et al.* Increased CSF  $\alpha$ -synuclein levels in Alzheimer's disease: Correlation with tau levels. *Alzheimer's & Dementia* **10**, S290–S298 (2014).

430. Vergallo, A. *et al.* Association of cerebrospinal fluid  $\alpha$ -synuclein with total and phospho-tau181 protein concentrations and brain amyloid load in cognitively normal subjective memory complainers stratified by Alzheimer's disease biomarkers. *Alzheimer's & Dementia* (2018). doi:10.1016/j.jalz.2018.06.3053
431. Zhou, B., Wen, M., Yu, W.-F., Zhang, C.-L. & Jiao, L. The Diagnostic and Differential Diagnosis Utility of Cerebrospinal Fluid  $\alpha$ -Synuclein Levels in Parkinson's Disease: A Meta-Analysis. *Parkinson's Disease* (2015). doi:10.1155/2015/567386
432. Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuropathological Alterations in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine* **1**, a006189–a006189 (2011).
433. Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nature Reviews Neurology* **9**, 13–24 (2013).
434. Dolatshahi, M. *et al.* Longitudinal Alterations of Alpha-Synuclein, Amyloid Beta, Total, and Phosphorylated Tau in Cerebrospinal Fluid and Correlations Between Their Changes in Parkinson's Disease. *Frontiers in Neurology* **9**, (2018).
435. Vekrellis, K., Xilouri, M., Emmanouilidou, E., Rideout, H. J. & Stefanis, L. Pathological roles of  $\alpha$ -synuclein in neurological disorders. *The Lancet Neurology* **10**, 1015–1025 (2011).
436. Colom-Cadena, M. *et al.* Confluence of  $\alpha$ -synuclein, tau, and  $\beta$ -amyloid pathologies in dementia with Lewy bodies. *Journal of Neuropathology & Experimental Neurology* **72**, 1203–1212 (2013).
437. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T. NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded. *Biochemistry* **35**, 13709–13715 (1996).
438. Choi, D.-H. *et al.* Role of Matrix Metalloproteinase 3-mediated  $\alpha$ -Synuclein Cleavage in Dopaminergic Cell Death. *Journal of Biological Chemistry* **286**, 14168–14177 (2011).
439. Murray, I. V. J. *et al.* Role of  $\alpha$ -Synuclein Carboxy-Terminus on Fibril Formation in Vitro <sup>†</sup>. *Biochemistry* **42**, 8530–8540 (2003).
440. Buée, L., Bussièrè, T., Buée-Scherrer, V., Delacourte, A. & Hof, P. R. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders<sup>11</sup>These authors contributed equally to this work. *Brain Research Reviews* **33**, 95–130 (2000).
441. Blennow, K. *et al.* Tau protein in cerebrospinal fluid. *Molecular and Chemical Neuropathology* **26**, 231–245 (1995).
442. Otto, M. *et al.* Elevated levels of tau-protein in cerebrospinal fluid of patients with Creutzfeldt–Jakob disease. *Neuroscience Letters* **225**, 210–212 (1997).
443. Wang, Z.-Y. *et al.* Use of CSF  $\alpha$ -synuclein in the differential diagnosis between Alzheimer's disease and other neurodegenerative disorders. *International Psychogeriatrics* **27**, 1429–1438 (2015).
444. Simonsen, A. H. *et al.* The utility of  $\alpha$ -synuclein as biofluid marker in neurodegenerative diseases: a systematic review of the literature. *Biomarkers in Medicine* **10**, 19–34 (2016).
445. Mollenhauer, B. Quantification of  $\alpha$ -synuclein in cerebrospinal fluid: How ideal is this biomarker for Parkinson's disease? *Parkinsonism & related disorders* **20**, S76–S79 (2014).
446. Mollenhauer, B. *et al.* Antibody-based methods for the measurement of  $\alpha$ -synuclein concentration in human cerebrospinal fluid - method comparison and round robin study. *Journal of Neurochemistry* (2018). doi:10.1111/jnc.14569
447. Viodé, A. *et al.* New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers. *Front. Neurosci.* **12**, (2018).
448. Shi, T. *et al.* Sensitive Targeted Quantification of ERK Phosphorylation Dynamics and Stoichiometry in Human Cells without Affinity Enrichment. *Analytical Chemistry* **87**, 1103–1110 (2015).
449. McAvoy, T. *et al.* Quantification of Tau in Cerebrospinal Fluid by Immunoaffinity Enrichment and Tandem Mass Spectrometry. *Clinical Chemistry* **60**, 683–689 (2014).
450. Yang, L. *et al.* An alpha-synuclein MRM assay with diagnostic potential for Parkinson's disease and monitoring disease progression. *PROTEOMICS - Clinical Applications* **11**, 1700045 (2017).

451. Lippens, G. *et al.* Proline-Directed Random-Coil Chemical Shift Values as a Tool for the NMR Assignment of the Tau Phosphorylation Sites. *ChemBioChem* **5**, 73–78 (2004).
452. Oeckl, P. *et al.* Alpha-, Beta-, and Gamma-synuclein Quantification in Cerebrospinal Fluid by Multiple Reaction Monitoring Reveals Increased Concentrations in Alzheimer's and Creutzfeldt-Jakob Disease but No Alteration in Synucleinopathies. *Molecular & Cellular Proteomics* **15**, 3126–3138 (2016).
453. Iyer, A. *et al.* The Impact of N-terminal Acetylation of  $\alpha$ -Synuclein on Phospholipid Membrane Binding and Fibril Structure. *Journal of Biological Chemistry* **291**, 21110–21122 (2016).
454. Tsiatsiani, L. & Heck, A. J. R. Proteomics beyond trypsin. *FEBS Journal* **282**, 2612–2626 (2015).
455. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nature Protocols* **11**, 993–1006 (2016).
456. Worboys, J. D., Sinclair, J., Yuan, Y. & Jørgensen, C. Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nature Methods* **11**, 1041–1044 (2014).
457. Lawless, C. & Hubbard, S. J. Prediction of Missed Proteolytic Cleavages for the Selection of Surrogate Peptides for Quantitative Proteomics. *OMICS: A Journal of Integrative Biology* **16**, 449–456 (2012).
458. Majbour, N. K. *et al.* Oligomeric and phosphorylated alpha-synuclein as potential CSF biomarkers for Parkinson's disease. *Molecular Neurodegeneration* **11**, (2016).
459. Grant, R. P. & Hoofnagle, A. N. From Lost in Translation to Paradise Found: Enabling Protein Biomarker Method Transfer by Mass Spectrometry. *Clinical Chemistry* **60**, 941–944 (2014).
460. Burré, J. The Synaptic Function of  $\alpha$ -Synuclein. *Journal of Parkinson's Disease* **5**, 699–713 (2015).
461. Yan, X., Uronen, R.-L. & Huttunen, H. J. The interaction of  $\alpha$ -synuclein and Tau: A molecular conspiracy in neurodegeneration? *Seminars in Cell & Developmental Biology* (2018). doi:10.1016/j.semcdb.2018.05.005
462. Castillo-Carranza, D. L., Guerrero-Muñoz, M. J., Sengupta, U., Gerson, J. E. & Kaye, R.  $\alpha$ -Synuclein Oligomers Induce a Unique Toxic Tau Strain. *Biological Psychiatry* **84**, 499–508 (2018).
463. Kilpatrick, L. E. & Kilpatrick, E. L. Optimizing High-Resolution Mass Spectrometry for the Identification of Low-Abundance Post-Translational Modifications of Intact Proteins. *Journal of Proteome Research* **16**, 3255–3265 (2017).
464. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization—principles and practice. *Mass Spectrometry Reviews* **9**, 37–70 (1990).
465. Gabelica, V., Galic, N., Rosu, F., Houssier, C. & Pauw, E. D. Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry. *Journal of Mass Spectrometry* **38**, 491–501 (2003).
466. Mohr, J., Swart, R., Samonig, M., Böhm, G. & Huber, C. G. High-efficiency nano- and micro-HPLC – High-resolution Orbitrap-MS platform for top-down proteomics. *PROTEOMICS* **10**, 3598–3609 (2010).
467. Zenaidee, M. A. & Donald, W. A. Extremely supercharged proteins in mass spectrometry: profiling the pH of electrospray generated droplets, narrowing charge state distributions, and increasing ion fragmentation. *Analyst* **140**, 1894–1905 (2015).
468. Xilouri, M., Brekk, O. R. & Stefanis, L. Alpha-synuclein and Protein Degradation Systems: a Reciprocal Relationship. *Mol Neurobiol* **47**, 537–551 (2013).
469. Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N. & Rubinsztein, D. C.  $\alpha$ -Synuclein Is Degraded by Both Autophagy and the Proteasome. *Journal of Biological Chemistry* **278**, 25009–25013 (2003).
470. Wong, E. & Cuervo, A. M. Integration of Clearance Mechanisms: The Proteasome and Autophagy. *Cold Spring Harb Perspect Biol* **2**, a006734 (2010).
471. McGlinchey, R. P. & Lee, J. C. Cysteine cathepsins are essential in lysosomal degradation of  $\alpha$ -synuclein. *Proceedings of the National Academy of Sciences* **112**, 9322–9327 (2015).
472. McGlinchey, R. P., Dominah, G. A. & Lee, J. C. Taking a Bite Out of Amyloid: Mechanistic Insights into  $\alpha$ -Synuclein Degradation by Cathepsin L. *Biochemistry* **56**, 3881–3884 (2017).



473. Bae, E.-J. *et al.* Haploinsufficiency of cathepsin D leads to lysosomal dysfunction and promotes cell-to-cell transmission of  $\alpha$ -synuclein aggregates. *Cell Death & Disease* **6**, e1901–e1901 (2015).
474. Qiao, L. *et al.* Lysosomal enzyme cathepsin D protects against alpha-synuclein aggregation and toxicity. *Molecular Brain* **1**, 17 (2008).
475. Turk, V. *et al.* Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1824**, 68–88 (2012).
476. Sung, J. Y. *et al.* Proteolytic Cleavage of Extracellular Secreted  $\alpha$ -Synuclein via Matrix Metalloproteinases. *Journal of Biological Chemistry* **280**, 25216–25224 (2005).
477. Levin, J. *et al.* Increased  $\alpha$ -synuclein aggregation following limited cleavage by certain matrix metalloproteinases. *Experimental Neurology* **215**, 201–208 (2009).
478. Dufty, B. M. *et al.* Calpain-Cleavage of  $\alpha$ -Synuclein. *The American Journal of Pathology* **170**, 1725–1738 (2007).
479. Diepenbroek, M. *et al.* Overexpression of the calpain-specific inhibitor calpastatin reduces human alpha-Synuclein processing, aggregation and synaptic impairment in [A30P] $\alpha$ Syn transgenic mice. *Human Molecular Genetics* **23**, 3975–3989 (2014).
480. Kiely, A. P. *et al.* Exploring the putative role of kallikrein-6, calpain-1 and cathepsin-D in the proteolytic degradation of  $\alpha$ -synuclein in multiple system atrophy. *Neuropathology and Applied Neurobiology* (2018). doi:10.1111/nan.12512
481. Nuber, S. & Selkoe, D. J. Caspase-1 clipping causes complications for  $\alpha$ -synuclein. *PNAS* **113**, 9958–9960 (2016).
482. Tong, J. *et al.* Brain  $\alpha$ -synuclein accumulation in multiple system atrophy, Parkinson’s disease and progressive supranuclear palsy: a comparative investigation. *Brain* **133**, 172–188 (2010).
483. Nelson, P. T. *et al.* Low sensitivity in clinical diagnoses of dementia with Lewy bodies. *J Neurol* **257**, 359–366 (2010).
484. Kapaki, E. N. *et al.* Cerebrospinal fluid tau, phospho-tau<sub>181</sub> and  $\beta$ -amyloid<sub>1742</sub> in idiopathic normal pressure hydrocephalus: a discrimination from Alzheimer’s disease. *European Journal of Neurology* **14**, 168–173 (2007).
485. Beyer, K. & Ariza, A. Alpha-Synuclein Posttranslational Modification and Alternative Splicing as a Trigger for Neurodegeneration. *Molecular Neurobiology* **47**, 509–524 (2013).
486. Yang, L. *et al.* An alpha-synuclein MRM assay with diagnostic potential for Parkinson’s disease and monitoring disease progression. *PROTEOMICS – Clinical Applications* **11**, 1700045 (2017).
487. Rovere, M., Powers, A. E., Patel, D. S. & Bartels, T. pTSara-NatB, an improved N-terminal acetylation system for recombinant protein expression in *E. coli*. *PLOS ONE* **13**, e0198715 (2018).

# Annexe 1

## ORIGINAL ARTICLE

# $\beta$ APP Processing Drives Gradual Tau Pathology in an Age-Dependent Amyloid Rat Model of Alzheimer's Disease

Mickael Audrain<sup>1,2,3</sup>, Benoit Souchet<sup>1,3,4</sup>, Sandro Alves<sup>1,3</sup>, Romain Fol<sup>1,2,3</sup>, Arthur Viode<sup>5</sup>, Alexis Haddjeri<sup>6</sup>, Satoru Tada<sup>1,3</sup>, Nicola S. Orefice<sup>1,3</sup>, Charlène Joséphine<sup>3,7</sup>, Alexis-Pierre Bemelmans<sup>3,7</sup>, Thierry Delzescaux<sup>3,7</sup>, Nicole Déglon<sup>8,9</sup>, Philippe Hantraye<sup>3,7,10</sup>, Yvette Akwa<sup>11</sup>, François Becher<sup>5</sup>, Jean-Marie Billard<sup>6</sup>, Brigitte Potier<sup>6</sup>, Patrick Dutar<sup>6</sup>, Nathalie Cartier<sup>1,3</sup> and Jérôme Braudeau<sup>1,3</sup>

<sup>1</sup>INSERM UMR1169, Université Paris-Sud, Université Paris-Saclay, Orsay 94100, France, <sup>2</sup>Université Paris Descartes, Paris, France, <sup>3</sup>CEA, DRF, Institut François Jacob, MIRCen, Fontenay-aux-Roses 92265, France, <sup>4</sup>Université Paris Saclay, Paris, France, <sup>5</sup>CEA, Institut Frédéric Joliot, Service de Pharmacologie et d'Immunoanalyse, 91191, Gif-sur-Yvette, France, <sup>6</sup>INSERM UMR894, Centre de Psychiatrie et Neurosciences, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, <sup>7</sup>CNRS UMR9199, Fontenay-aux-Roses 92265, Université Paris-Sud, Université Paris-Saclay, Orsay 94100, France, <sup>8</sup>Department of Clinical Neurosciences, Laboratory of Cellular and Molecular Neurotherapies, Lausanne University Hospital, Lausanne, Switzerland, <sup>9</sup>Neuroscience Research Center, Laboratory of Cellular and Molecular Neurotherapies, Lausanne University Hospital, Lausanne, Switzerland, <sup>10</sup>INSERM UMS27, Fontenay-aux-Roses 92265, Université Paris-Sud, Université Paris-Saclay, Orsay 94100, France and <sup>11</sup>INSERM U1195 and Université Paris Sud and Université Paris-Saclay, 80 rue du général Leclerc, 94276 Le Kremlin-Bicêtre, France

Address correspondence to Jérôme Braudeau. Email: jerome.braudeau@agent-biotech.com.

Mickael Audrain and Benoit Souchet are joint first authors.

Jean-Marie Billard, Brigitte Potier and Patrick Dutar are contributed equally to this work. Nathalie Cartier and Jérôme Braudeau are contributed equally to this work.

## Abstract

The treatment of Alzheimer's disease (AD) remains challenging and requires a better in depth understanding of AD progression. Particularly, the link between amyloid protein precursor (APP) processing and Tau pathology development remains poorly understood. Growing evidences suggest that APP processing and amyloid- $\beta$  ( $A\beta$ ) release are upstream of Tau pathology but the lack of animal models mimicking the slow progression of human AD raised questions around this mechanism. Here, we described that an AD-like  $\beta$ APP processing in adults wild-type rats, yielding to human APP,  $\beta$ CTF and  $A\beta$  levels similar to those observed in AD patients, is sufficient to trigger gradual Tauopathy. The Tau hyperphosphorylation begins several months before the formation of both amyloid plaques and tangle-like aggregates in aged rats and without

associated inflammation. Based on a longitudinal characterization over 30 months, we showed that extrasynaptic and emotional impairments appear before long-term potentiation deficits and memory decline and so before A $\beta$  and Tau aggregations. These compelling data allowed us to (1) experimentally confirm the causal relationship between  $\beta$ APP processing and Tau pathology in vivo and without Tau transgene overexpression, (2) support the amyloidogenic cascade and (3) propose a 4-step hypothesis of prodromal AD progression.

**Key words:** Alzheimer's disease, amyloid pathology, hippocampus, pre-clinical AD, Tau pathology

## Introduction

Alzheimer's disease (AD), the most common cause of dementia, is characterized by a progressive accumulation of  $\beta$ -amyloid peptide (A $\beta$ ), Tau hyper-, and abnormal phosphorylation, and displays a decline in cognitive functions followed by senile plaques and tangles formation. A large body of evidence suggest that there is a long infra-clinical phase of AD, in which pathological changes begin decades before classical hallmarks (senile plaques and neurofibrillary tangles) (Rajan et al. 2015). Evidence suggests that soluble A $\beta$  peptides play a pivotal role during this infra-clinical phase (Ferreira et al. 2015). Toxicity of soluble A $\beta$  directly affects synaptic plasticity even before amyloid plaque deposition and the formation of neurofibrillary tangles (Selkoe 2002; Selkoe and Hardy 2016). This infra-clinical phase has been poorly studied, despite increasing interest, largely due to the absence of pertinent animal models. Early stage animal models with a wide therapeutic window before amyloid and Tau deposition are highly needed, and would be more relevant in terms of human AD pathology than current transgenic models. Such models must fulfil specific requirements. First, the period prior to reaching an AD-like pathology is as essential as the full-blown disease. AD is an age-related disorder and models that display amyloid plaques or tangles too early in the life of the animal may not be relevant in terms of disease progression in humans (Kitazawa et al. 2012). Furthermore, most currently used AD models, generated over the last 20 years, are transgenic mice (Lee and Han 2013) and are based on very high transgene expression that hugely exceeds human levels (Audrain et al. 2016). Such overproduction is especially responsible for the supra-physiological release of amyloid peptides resulting in the early onset of histological brain lesions such as senile plaques, one of the hallmarks of AD. Several transgenic AD rat models have more recently been reported (Abbott 2004). Not only are rats widely used in toxicity studies, but their longer lifespan and their cognitive abilities makes them ideal for investigating the mechanisms of age-dependent neurodegenerative disease. The larger size of their brains (relative to mice) also facilitates translational procedures such as surgical procedures, magnetic resonance imaging, or positron emission tomography. Biological fluid collection such as plasma and cerebrospinal (CSF) is also simplified, reflecting the growing interest in the use of such models (Do Carmo and Cuello 2013). Strikingly, the gradual emergence of AD rat models has followed the same pattern as that for the prior development of transgenic mice in that they mostly focus on amyloid plaque formation. The first AD rat models did not develop senile plaques and were not considered to reproduce an AD phenotype (Echeverria et al. 2004; Ruiz-Opazo et al. 2004; Folkesson et al. 2007). A few years later, 2 transgenic Sprague Dawley rat models were developed in which amyloid plaques were triggered at 17 and 7 months of age (Flood et al. 2009). More recently, McGill-R-Thy1-amyloid protein precursor (APP) rats have been shown to display senile plaques as early as 6 months of age

(Leon et al. 2010). A drawback of these more recently developed models is that they form senile plaques relatively early in life.

In this study, we developed and characterized an original rat model generated by AAV-mediated gene transfer (the AAV-AD rat) to induce the production of amyloid derivatives with levels as close as those found in human AD hippocampi and without a significant transgene overexpression. Soluble A $\beta$ 42 levels and the A $\beta$ 42/40 ratio gradually increased in rat hippocampi and decreased in CSF. Moreover, amyloid plaque deposition as cerebral amyloid angiopathy (CAA) progressively occurred until very late in the life of the animals, that is, 2.5 years after injection. More importantly, endogenous Tau was progressively both hyper- and abnormally phosphorylated over time, and associated with increased levels of GSK3 $\beta$  and DYRK1A kinases, resulting finally in phosphorylated Tau aggregate formation (AT8 and AT100 positive) in old rats. The biochemical, histological, electro-physiological, and behavioural evaluation of the AAV-AD rat highlighted the critical phases during infra-clinical disease progression implicated in AD and allowed us to propose a sequential AD progression hypothesis during infra-clinical stages, especially regarding  $\beta$ APP processing and Tau pathology links before inflammation processes and misfolded protein pathologies (senile plaques and tangles).

## Methods

### Plasmid Design and Vector Production

We used a human double-mutant APP751 cDNA containing the Swedish and London mutations (codon optimized and containing a Kozak sequence; GeneArt, Life Technologies, Saint Aubin, France), and a human PS1 cDNA containing the M146L mutation (pENTR4-PS1-S182M146L). The APP<sub>SL</sub> and the PS1<sub>M146L</sub> sequences were cloned into an AAV2 plasmid with a CAG promoter to generate AAV2-CAG-APP<sub>SL</sub> or -CAG-PS1<sub>M146L</sub>. AAV vectors were produced as previously described (Berger et al. 2015), except that the AAV packaging plasmid was replaced with a plasmid construct containing the rep gene of AAV2 and the cap gene of AAV9.

### Animals

We used 130 male Wistar rats (8-week-old; SARL JanvierLabs, Le Genest Saint Isle, France) in this study. All experiments were conducted in accordance with the ethical standards of French and European regulations (European Communities Council Directive 2010/63/EU, authorization number APAFIS#4449-2016031012491697).

### Bace 1 Inhibitor Study

Bace1 inhibitor LY2886721 was purchased from Apexbio Technology LLC, suspended in a DMSO/water mix (1:9) at 15 mg/mL, and orally administered by gavage in 1-month

postinjected rats (15 mg/kg). Animals were killed 4 h later and tissues were collected as previously described. DMSO/water mix without the LY2886721 was used as vehicle for the control group.

### Stereotactic Injections of AAVs

Rats were anesthetized by an intraperitoneal injection of ketamine/xylazine and placed in a stereotactic frame (Stoelting). Stereotactic intracerebral injections of AAVs into the hippocampi of both hemispheres were performed, using the following coordinates: antero-posterior:  $-3.6$  mm, lateral:  $\pm 2.5$  mm, ventral:  $-3.3$  mm relative to bregma. We injected  $4 \mu\text{L}$  of viral preparation into each site ( $2.5 \times 10^{10}$  vg/site and  $5 \times 10^{10}$  vg/site for PS1 and APP, respectively) at a rate of  $0.25 \mu\text{L}/\text{min}$ . Two groups of animals were initially created, according to the viral vectors injected or co-injected: AAV-CAG-PS1M146L (control rats), and AAV-CAG-APPSL + AAV-CAG-PS1M146L (AAV-AD rats).

### Human Brain Samples

Postmortem samples were obtained from brains collected as part of the Brain Donation Program of the GIE-Neuro-CEB Brain Bank, Pitié-Salpêtrière Hospital, Paris, France. Autopsies were carried out by accredited pathologists, after informed consent had been obtained from the relatives, in accordance with French Bioethics laws. Five hippocampal samples from 5 patients with sporadic forms of AD (male and female; Braak 6/Thal 5; aged between 69 and 89 years, with a postmortem interval (PMI) of 30–59 h) and 5 hippocampus samples from 5 age-matched control subjects (male and female, aged between 69 and 92 years, PMI of 6–63 h) were used in this study (Table 1).

### Tissue Collection and Sample Preparation

Rats were anesthetized with ketamine/xylazine and were first placed in a stereotactic frame to collect CSF by puncturing the cisterna magna with a low dead space syringe. One hemisphere was postfixed by incubation for 72 h in 4% PFA, cryoprotected in 30% sucrose in PBS, and cut into  $40 \mu\text{m}$  sections with a freezing microtome (Leica) for histological analyses. The contralateral hemisphere was dissected for isolation of the hippocampus and cortex. Samples were homogenized in a lysis buffer (150 mM NaCl and 1% Triton in Tris-buffered saline) containing phosphatase (Pierce) and protease (Roche) inhibitors and centrifuged for

20 min at  $15\,000 \times g$ . The same procedure was applied to human samples.

### ELISA

Extracted A $\beta$  was quantified using the MSD Human A $\beta$ 42 V-PLEX Kit and the triplex A $\beta$  Peptide Panel 1 (6E10) V-PLEX Kit (Meso Scale Diagnostics). Rat endogenous A $\beta$ 42 was assessed using the IBL Mouse/Rat Amyloid-beta (1-42) ELISA (IBL International GmbH).  $\beta$ CTF was quantified using the IBL APP  $\beta$ CTF Assay Kit (IBL International GmbH). Hyperphosphorylated Tau was quantified using the Innogenetics Phospho-Tau 181 P kit (Fujirebio Europe) and the MSD Phospho (Thr231)/Total Tau Kit (Meso Scale Diagnostics). The ELISA was performed according to the kit manufacturer's instructions in each case.

### Western Blotting

Equal amounts of protein ( $30 \mu\text{g}$ ) were separated by electrophoresis in NuPAGE Bis-Tris Gels (Life Technologies) and transferred to nitrocellulose membranes. The membranes were hybridized with various primary antibodies: APP 6E10 (1/500, Covance), PS1 (1/1000, Millipore), APP C-ter (1/500, Millipore), GAPDH (1/1000, Abcam), P-Tau Thr 212 (1/1000, Invitrogen), P-Tau Thr 422 (1/1000, Invitrogen), CDK5 (1/1000, Abcam), GSK3 $\beta$  (1/1000, Abcam), DYRK1A (1/250, Abnova), P-CamkII Thr 286 (1/1000, Abcam), GFAP (1/500, DAKO). Various secondary antibodies were also used: ECL horseradish peroxidase linked anti-rabbit, ECL horseradish peroxidase linked Anti-mouse, ECL horseradish peroxidase linked anti-rat (all 1/2000, GE Healthcare) (Table 2).

### Immunohistochemistry and Image Acquisition

Cryosections were washed with 0.25% Triton in PBS and saturated by incubation (0.25% Triton in PBS/5% goat serum). They were then incubated with one or more of the following primary antibodies: APP C-ter, (1/500, Millipore), 4G8-Biotin (1/1000, Covance), 6E10 (1/500, Covance), BAM10 (1/500, ThermoFisher), CD31 (1/200, Abcam), AT8 P-Tau pSer202/Thr205 (1/1000, ThermoFisher), and AT100 P-Tau pThr212/Ser214 (1/1000, ThermoFisher). For nonfluorescent immunostainings, endogenous peroxidase was quenched with PBS containing 3% H $_2$ O $_2$  for 5 min followed by amplification using the ABC system (VECTASTAIN Elite ABC HRP Kit, Vector Laboratories). Horseradish peroxidase conjugate and 3,3'-diaminobenzidine were then used according to the manufacturer's manual

**Table 1** Details of human brain samples used.

	Gender	Age (years)	PMI (hours)	Clinical AD status	Diagnosis
	M	73	54	Braak V, Thal IV, CAA	AD
	F	89	59	Braak VI, Thal V, CAA	AD
	M	70	30	Braak VI, Thal V, CAA	AD
	M	69	30	Braak VI, Thal V, CAA	AD
	M	72	44	Braak VI, Thal V	AD
Mean	M (80%)	74.6	43.4	Braak VI, Thal V, CAA	AD
	F	92	21	Braak I	Ctrl
	M	58	6	Braak I-II	Ctrl
	M	82	63	Braak II, Thal I	Ctrl
	F	89	12	/	Ctrl
	M	69	6	/	Ctrl
Mean	M (60%)	74.5	21.75	Braak I, Thal I	Ctrl

**Table 2.** Table of used antibodies.

Detected antigen	Antibodies	Dilution/technique	Suppliers
Human APP (6E10)	Purified anti- $\beta$ -Amyloid, 1-16 Antibody	1:500/WB	Biologend
Total APP (APP Cter)	Anti-APP A4 Antibody, a.a. 66–81 of APP [NT], clone 22C11	1:500/WB	Millipore
PS1	Anti-Presenilin-1 Antibody, loop, a.a. 263–378, CT, clone PS1-loop	1:1000/WB	Millipore
GAPDH	Anti-GAPDH antibody [6C5] (ab8245)	1:1000/WB	Abcam
p-CAMKII Thr286	Anti-CaMKII (phospho T286) ab32678	1:500/WB	Abcam
rodent APP (4 G8)	Purified anti- $\beta$ -Amyloid, 17–24 Antibody	1:250/IH	Covance
Human APP (6E10)	Purified anti- $\beta$ -Amyloid, 1–16 Antibody	1:250/IH	Covance
BAM10	beta Amyloid Monoclonal Antibody (BAM10)	1:250/IH	Thermo Scientific
CD31	Anti-CD31 antibody [EPR3094] (ab76533)	1:100/IH	Abcam
p-TAU Ser422	Phospho-Tau (Ser422) Polyclonal Antibody 44-764 G	1:1000/WB	Thermo Scientific
p-TAU Thr212	Phospho-Tau (Thr212) Antibody (44-740 G)	1:1000/WB	Thermo Scientific
AT8	Phospho-Tau (Ser202, Thr205) Monoclonal Antibody (AT8)	1:500/WB	Millipore
AT100	Phospho-Tau (Thr212, Ser214) Monoclonal Antibody (AT100)	1:500/WB	Millipore
GSK3 $\beta$	Anti-GSK3 beta antibody (ab73173)	1:500/WB	Abcam
DYRK1A	DYRK1A monoclonal antibody (M01), clone 7D10	1:250/WB	Abnova
CDK5	Anti-Cdk5 antibody [EP715Y] (ab40773)	1:1000/WB	Abcam
AT8	Phospho-Tau (Ser202, Thr205) Monoclonal Antibody (AT8)	1:250/IH	Millipore
AT100	Phospho-Tau (Thr212, Ser214) Monoclonal Antibody (AT100)	1:100/IH	Millipore

(Vector® DAB, Vector Laboratories). Images were captured using a Nikon Eclipse Ti Microscope or a Leica TCS SP8 confocal microscope and analysed using ImageJ software (NIH).

### 3D Reconstruction

In order to visualize 3D organization at cellular level, several image processing steps were performed using BrainVISA software (<http://brainvisa.info>). Series of images acquired with the confocal system were first stacked to produce 3D images and objects of interest were segmented. Based on the high contrast available in the images of fluorescence, an adaptive threshold followed by a median filtering operation were realized under expert supervision to produce proper binary segmentations. In the last step, surfaces were derived from these segmentations to generate 3D views of both cells and pathological aggregates enabling to assess accurately their spatial organization in 3D.

### LC-MS/MS Analyses

Sample preparation was adapted from Barthelemy et al. (2016). Briefly, 20  $\mu$ L of 70% perchloric acid were added to 20  $\mu$ L of hippocampal extracts for protein precipitation. Samples were kept on ice for 15 min before centrifugation (15 min, 4 °C, and 16 000 g). Supernatants were collected and mixed with 40  $\mu$ L of 1% trifluoroacetic acid (TFA) before  $\mu$ SPE extraction on Oasis HLB sorbent. Samples were loaded and washed with 500  $\mu$ L of 0.1% TFA in 10% acetonitrile on the sorbent, previously conditioned and equilibrated. Tau proteins were eluted with 100  $\mu$ L of 27.5% acetonitrile in 0.1% TFA. Extracts were evaporated to dryness with a Turbopap instrument (Biotage) (10 psi of nitrogen, 50 °C, 2 h). Samples were reconstituted with 35  $\mu$ L of 50 mM ammonium bicarbonate and digested overnight at 37 °C with 5  $\mu$ L of 100 ng/ $\mu$ L trypsin solution (sequencing grade modified trypsin, Promega, Madison, WI) in 50 mM ammonium bicarbonate. Finally, the digest was acidified with 5  $\mu$ L of 1 M chloridric acid and used for subsequent LC-MS/MS analyses following addition of 2.5 ng of the 3 labeled PEPotec peptides.

Capillary LC-HRMS was performed on an Ultimate 3000 chromatography system coupled to a Q-Exactive Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Ten microliters of

the sample extract were injected and loaded for 3 min at a flow rate of 10  $\mu$ L/min on a  $\mu$ PreColumn C18 Pepmap100 (0.3 mm  $\times$  5 mm, 5  $\mu$ m) (Thermo Scientific). Peptide separation was achieved in 60 min using an Acclaim Pepmap RSCL C18-100A column (0.3 mm  $\times$  150 mm, 2  $\mu$ m) (Thermo Scientific) at a flow rate of 6  $\mu$ L/min. The column oven temperature was set to 35 °C. The mobile phases were (A) 0.1% formic acid in water and (B) 0.1% formic acid in 80% MeOH. After an isocratic step of 3 min at 4% phase B, a linear gradient from 4% to 55% B was run for 30 min, then increased to 90% B. An equilibration step was performed at 4% B for 20 min. Data were acquired in the positive ion mode with an ion spray voltage of 4 kV, a capillary temperature set to 320 °C, and sheath gas at 5. S-lens RF were set at 60. The AGC target was set to 1e6 and Orbitrap resolution to 70 000 at 200 m/z for all experiments.

A time-scheduled sequential PRM method was established targeting the following p-tau peptides corresponding to Tau protein phosphorylation on residues Ser202, Thr217, Thr231/Ser235 (Table 3). The signal of up to 11 major and noninterfered fragment ions identified with high resolution from a common peptide precursor were summed up to provide one extracted ion chromatogram (XIC) for each p-tau peptide (Fig. 7C). Isotope-labeled synthetic peptides with labeled amino acids lysine (K, label 13C615N2) or arginine (R, label 13C615N4) were used for signal normalization and relative quantification of p-tau peptides. Raw mass spectrometry data were exported to Skyline 3.7 for verification of the transitions ratio of unlabeled and labeled peptides. Xcalibur 2.2 software (Thermo Fisher Scientific) was used for quantitative data analysis.

### Behavioural Assessment

**Open-field:** The apparatus consisted of an open-topped, clear plexiglas box measuring 90  $\times$  90  $\times$  70 cm placed in a room with controlled dim lighting (40 lux) and constant white noise at 40 dB. The rats were placed in the centre of the arena and a video recording was made over a period of 5 min. The behaviour of the animals was analysed from this video. The arena was divided into a central region and a peripheral region, and the time spent in the centre and periphery of the open-field was measured.

**Morris water maze:** Experiments were performed in a tank of 180 cm in diameter and 50 cm deep, filled with opacified water

**Table 3.** MS parameters for LC-MS/MS analysis of Ser 202; Thr 217 and Thr231/Ser235 as well as their corresponding labeled peptides.

Peptide sequence	Precursor ion m/z	Product ion m/z	Normalized collision energy %	Retention time (min)
SGYSSPGsPGTPGSR	737.30 (z = 2)	1166.4838 (y12) 1079.4517 (y11) 992.4197 (y10) 895.3669 (y9) 838.3455 (y8) 671.3471 (y7) 416.2252 (y4) 496.7135 (y10 + 2) 308.1241 (b3)	22	17.7
SGYSSPGSPGtPGSR[ <sup>6</sup> C <sub>13</sub> ; <sup>4</sup> N <sub>15</sub> ]	742.31 (z = 2)	1176.4920 (y12) 1089.4600 (y11) 1002.4280 (y10) 905.3752 (y9) 848.3537 (y8) 761.3217 (y7) 426.2335 (y4) 501.7176 (y10 + 2) 308.1241 (b3)	22	17.7
TPSLPtPPTREPK	500.92 (z = 3)	1399.6981 (y12) 1102.5293 (y9) 824.4625 (y7) 727.4097 (y6) 630.3570 (y5) 529.3093 (y4) 700.3527 (y12 + 2) 651.8263 (y11 + 2) 608.3103 (y10 + 2) 551.7683 (y9 + 2) 467.2376 (y12 + 3)	22	23.2
TPSLPtPPTREPK[ <sup>6</sup> C <sub>13</sub> ; <sup>2</sup> N <sub>15</sub> ]	504.26 (z = 3)	1409.7064 (y12) 1112.5375 (y9) 834.4707 (y7) 737.4180 (y6) 640.3652 (y5) 539.3175 (y4) 705.3568 (y12 + 2) 656.8304 (y11 + 2) 613.3144 (y10 + 2) 556.7724 (y9 + 2) 470.5736 (y12 + 3)	22	23.2
VAVVrtPPKsPSSAK	561.95 (z = 3)	978.4656 (y9) 792.8810 (y14 + 2) 743.8925 (y14-98 + 2) 757.3624 (y13 + 2) 708.3740 (y13-98 + 2) 489.7364 (y9 + 2) 440.7480 (y9-98 + 2) 706.3648 (b6)	22	16.5
VAVVrtPPKSPSsAK[ <sup>6</sup> C <sub>13</sub> ; <sup>2</sup> N <sub>15</sub> ]	564.6173 (z = 3)	986.4798 (y9) 796.8881 (y14 + 2) 747.8996 (y14-98 + 2) 761.3695 (y13 + 2) 712.3811 (y13-98 + 2) 493.7435 (y9 + 2) 444.7551 (y9-98 + 2) 706.3648 (b6)	22	17.3

kept at 21 °C, and equipped with a platform of 18 cm in diameter, submerged 1 cm below the surface of the water. Visual clues were positioned around the pool and the luminosity was maintained at 350 lux. The rats were initially exposed to a learning phase, which consisted of daily sessions (3 trials per

session) on 5 consecutive days. The starting position varied pseudo-randomly, between the 4 cardinal points. A mean interval of 20 min was left between trials. The trial was considered to have ended when the animal reached the platform. Long-term spatial memory was assessed 72 and 120 h after the last

training trial in a probe trial in which the platform was no longer available. Animals were monitored using EthoVision software.

### Ex Vivo Electrophysiology

Rats were anesthetized with halothane and decapitated. The brain was rapidly removed from the skull and placed in chilled (0–3 °C) artificial cerebrospinal fluid (ACSF) containing 124 mM NaCl, 3.5 mM KCl, 1.5 mM MgSO<sub>4</sub>, 2.5 mM CaCl<sub>2</sub>, 26.2 mM NaHCO<sub>3</sub>, 1.2 mM NaH<sub>2</sub>PO<sub>4</sub>, 11 mM glucose. Transverse slices (300–400 μm thick) were cut with a vibratome and placed in ACSF in a holding chamber, at 27 °C, for at least 1 h before recording. Each slice was individually transferred to a submersion-type recording chamber and submerged in ACSF continuously superfused and equilibrated with 95% O<sub>2</sub>, 5% CO<sub>2</sub>.

**Tonic current:** The biophysical properties of the tonic current generated by the activation of extrasynaptic NMDA receptors with ambient glutamate were evaluated. Whole-cell patch-clamp recordings of CA1 pyramidal cells were performed at room temperature, with borosilicate patch pipettes (5 MΩ) filled with a buffer containing 140 mM CsCH<sub>3</sub>CO<sub>3</sub>S, 6 mM CsCl, 2 mM MgCl<sub>2</sub>, 10 mM HEPES, 1.1 mM EGTA, 5 mM QX-314, and 4 mM ATP, at a pH of 7.3 at 290 mosm. Transmembrane currents were acquired and filtered through an amplifier (AxoPatch 1-D, Axon Instruments), stored on a computer, and digitized using WinLTP software for analysis. The tonic current was recorded at a holding potential of +40 mV, in the presence of TTX (1 μM), NBQX (10 μM), and bicuculline (10 μM), to isolate the NMDA component of the holding current (hc). After recording a stable control hc for 3–5 min, APV (50 μM) was added to the superfusion medium. The hc fell to a new stable value under the effect of APV, and the difference between the control hc and its value in the presence of APV determined the amplitude of the tonic current.

**Long-term potentiation (LTP) and depression (LTD).** For the LTP, theta-burst stimulation (TBS), mimicking the natural stimulation at the theta frequency from the medial septum to the hippocampus, consisted of 5 trains of four 100 Hz pulses each, separated by 200 ms and delivered at the test intensity. The sequence was repeated 3 times, with an interburst interval of 10 s. Testing with a single pulse was resumed for 60 min (TBS) to determine the level of LTP. The LTD was induced by the application of low frequencies stimulation (LFS): 1 Hz for 15 min or 2 Hz for 10 min. Responses were recorded for at least for 45 min after LFS.

### Statistical Analysis

Data are expressed as the mean ± SEM. Statistical analyses were performed using GraphPad Prism (GraphPad Software) or Statistica (StatSoft, Inc.) software and the statistical significance was set to a *P*-value < 0.05 for all tests. One-way ANOVA followed by Tukey's post hoc test or 2-way ANOVA followed by Sidak's post hoc test were used to determine the significance of differences between groups. Student's *t*-test was used when only 2 groups were analysed.

## Results

### AAV-AD Induction in Wild-Type Adult Rat Hippocampi

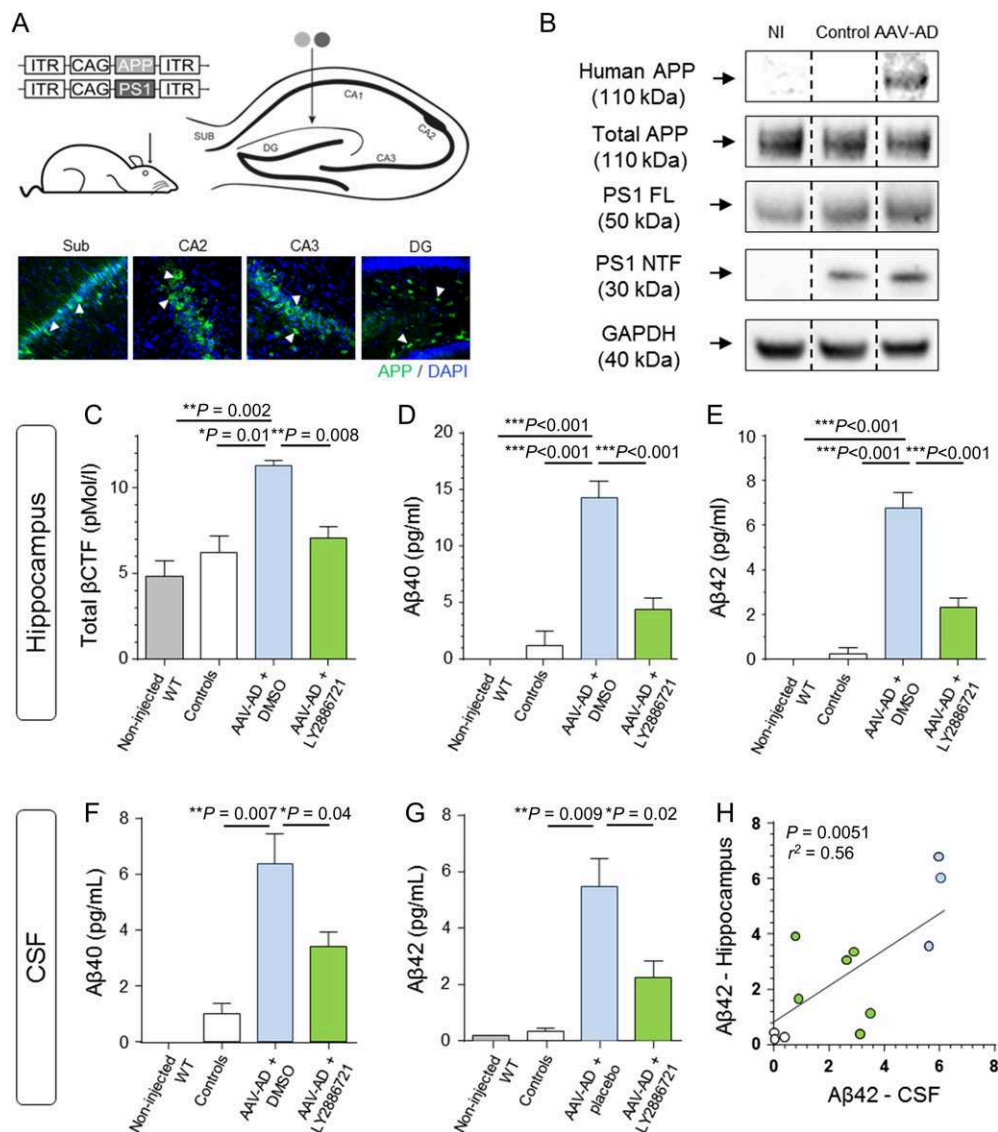
We generated AAV vectors encoding human mutant PS1<sup>M146L</sup> and human mutant APP<sup>SL</sup> (Swedish & London mutations) that were bilaterally co-injected into the stratum lacunosum of the hippocampus of 8-week-old male Wistar rats (Fig. 1A). We used the AAV9 capsid known to efficiently transduce the

hippocampus (Fol et al. 2015; Alves et al. 2016). We injected AAV-AD rats (*n* = 8) with both APP<sup>SL</sup> and PS1<sup>M146L</sup> AAV constructs, whereas the control group (*n* = 4) was only injected with the AAV-PS1 virus. We evaluated vector-mediated APP and PS1 expression in the hippocampus 1-month later. Western blot analysis of whole hippocampal extracts revealed comparable PS1 full-length (PS1 FL) levels in injected and non-injected rats. A 30-kDa PS1 N-terminal fragment (PS1 NTF) was mostly detectable in injected hippocampi, confirming the formation of the PS1 active site responsible for the formation of Aβ species, without PS1 FL accumulation. Similarly, human APP was expressed from 1-month after injection without a significant increase of total APP (human + rat forms; Fig. 1B). We assessed the engagement of the amyloidogenic pathway by first confirming the production of βCTF (C99 fragment; Control: 6.20 ± 0.97 pMol/L; AAV-AD: 11.28 ± 0.31 pMol/L; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.012), the first cleavage product of APP by BACE1 (Fig. 1C) (Hardy and Higgins 1992). We also confirmed that AAV-AD induction resulted in the substantial release of both soluble Aβ40 (Control: 1.25 ± 1.25 pg/mL; AAV-AD: 14.31 ± 1.46 pg/mL; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.0001) and Aβ42 peptides (Control: 0.25 ± 0.25 pg/mL; AAV-AD: 6.78 ± 0.68 pg/mL; one-way ANOVA followed by Tukey's post hoc test: *P* = 0.0001) relative to control rats (Fig. 1D,E). Aβ levels in the cerebrospinal fluid (CSF) of AD patients has a diagnostic potential as a biomarker (Dubois et al. 2014). We evaluated the dispersion of both Aβ peptides from the hippocampus to the CSF 1-month after injection. Despite the expected absence of βCTF in the CSF (data not shown), AAV-AD induction led to an increase of both Aβ40 (Control: 0.94 ± 0.34 pg/mL; AAV-AD: 6.00 ± 1.03 pg/mL; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.007) and Aβ42 (Control: 0.35 ± 0.09 pg/mL; AAV-AD: 5.46 ± 0.97 pg/mL; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.009) levels in CSF (Fig. 1F,G). We administered the β-secretase (also known as BACE) inhibitor, LY2886721, diluted in DMSO, by gavage in 1-month postinjected AAV-AD rats to confirm the involvement of BACE in AAV-AD induction. We observed a complete inhibition of βCTF production 4 h after gavage (Fig. 1C; LY2886721-treated AAV-AD: 7.06 ± 0.65 pMol/L; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.008) confirming BACE1-dependent βCTF production. In agreement with the amyloidogenic cascade hypothesis (Hardy and Higgins 1992), decreasing βCTF levels also led to decreases in both Aβ40 (LY2886721-treated AAV-AD: 4.40 ± 0.99 pg/mL; 1-way ANOVA followed by Tukey's post hoc test: *P* < 0.0001) and Aβ42 (LY2886721-treated AAV-AD: 2.30 ± 0.41 pg/mL; one-way ANOVA followed by Tukey's post hoc test: *P* < 0.0001) production (Fig. 1D,E). In a similar manner, both Aβ40 (LY2886721-treated AAV-AD: 3.23 ± 0.49 pg/mL; one-way ANOVA followed by Tukey's post hoc test: *P* = 0.04) and Aβ42 (LY2886721-treated AAV-AD: 2.25 ± 0.57 pg/mL; 1-way ANOVA followed by Tukey's post hoc test: *P* = 0.02) levels decreased in CSF 4 h after gavage with LY2886721 (Fig. 1F,G). Altogether, these data indicate BACE1-dependent engagement of the amyloidogenic cascade from 1-month postinjection. In addition, a positive linear correlation between hippocampal and CSF Aβ42 levels (Pearson correlation coefficient test: *P* = 0.0051, *r*<sup>2</sup> = 0.56; Fig. 1H) suggests that Aβ42 may passively spread from the hippocampus to the CSF.

### AAV-AD Induction Leads to Progressive Cognitive Impairment

We evaluated whether human-like APP processing in AAV-AD rats is associated with progressive cognitive and behavioural

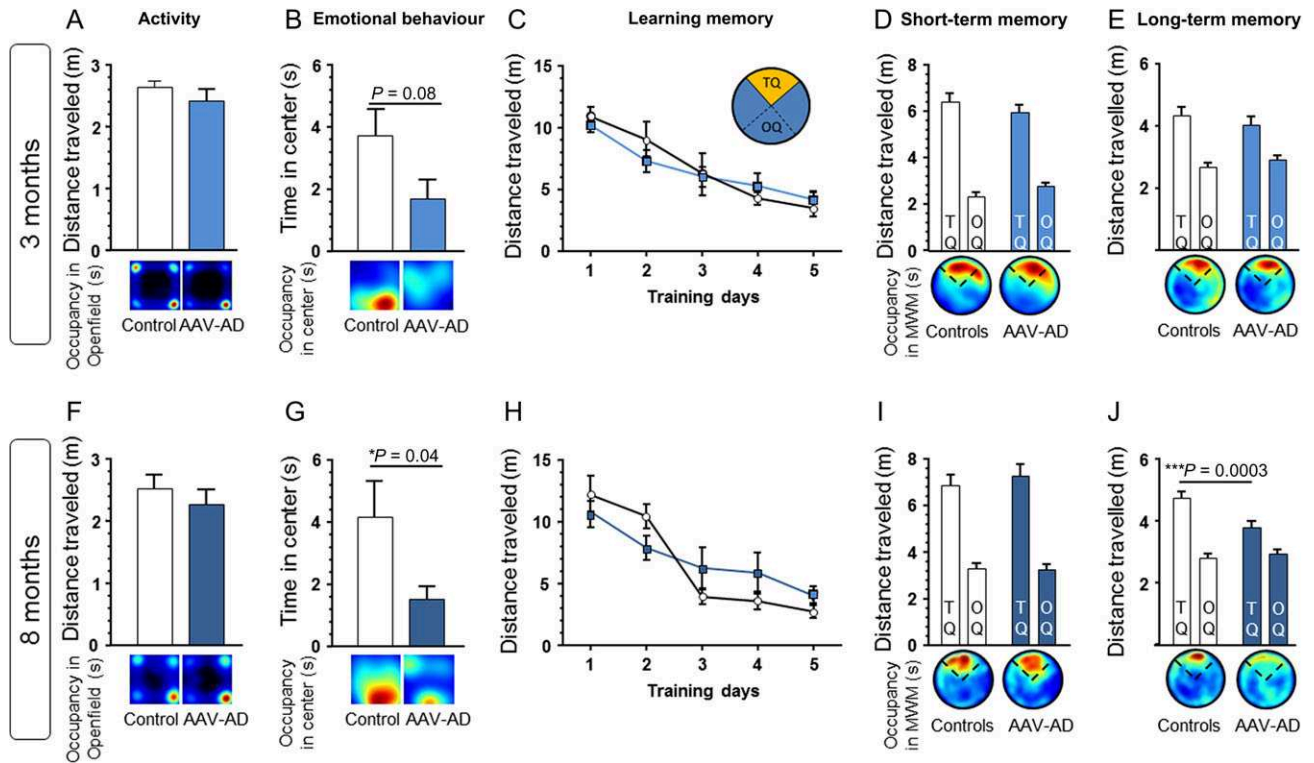




**Figure 1.** AAV-AD induction leads to BACE1-dependent APP processing in the hippocampus of adult rats. We generated AAV vectors encoding mutated human PS1M146L and mutated human APPSL, and injected them bilaterally into the stratum lacunosum of the rat hippocampus. Rats received either the AAV-CAG-PS1M146L vector (control,  $n = 4$ ) or both AAV-CAG-PS1M146L and AAV-CAG-APPSL vectors (AAV-AD,  $n = 8$ ). Rats were killed 1-month after vector injection for transgene expression analysis in the hippocampus. (A) Upper panel: schematic representation of the injection site in the rat hippocampus. Bottom panel: Confocal microscopy acquisition performed in sections from AAV-AD hippocampus: human APP (green) was expressed in sparse neurons in CA2/CA3 layers (blue) but also in the subiculum (Sub) and dentate gyrus (DG) subfields (APP C-ter antibody). (B) Representative western blots showing the expression of human APP (6E10 antibody), total APP (rat + human forms; APP C-ter antibody) and PS1 (full-length: PS1 FL and N-terminal fragment: PS1 NTF), confirming transgene expression 1-month after injection. All western blots were performed on whole hippocampal extracts (TBS-Tx soluble fraction). (C–E) Comparative analysis of TBS-Tx soluble  $\beta$ CTF (C),  $A\beta_{40}$  (D) and  $A\beta_{42}$  (E) levels by ELISA and MSD multiplex immunoassay in hippocampi from non-injected, controls and AAV-AD rats. A fourth group of AAV-AD rats treated with the Bace1 inhibitor LY2886721 has been also studied. (F–H) Comparative analysis of CSF  $A\beta_{40}$  (F) and CSF  $A\beta_{42}$  (G) levels by MSD multiplex in the same groups previously described.  $A\beta_{42}$  content in hippocampus strongly correlated with CSF  $A\beta_{42}$  (H). Bars represent the mean  $\pm$  SEM. Statistical analyses were performed using 1-way ANOVA followed by Tukey's post hoc test.

dysfunction over time, relative to controls at 3 ( $n = 20$ ) and 8 months ( $n = 20$ ) after AD induction. We first analysed anxiety-like emotional behaviour in a new environment using the open-field task, 3 months after AAV-AD induction. The control and AAV-AD rats travelled similar total distances, confirming the absence of motor abnormalities or hyperactivity (Fig. 2A). We observed no significant differences in time spent in the centre of the area, despite a downward trend (Control:  $3.73 \pm 0.93$  s; AAV-AD:  $1.70 \pm 0.62$  s; Student's  $t$ -test:  $P = 0.08$ ; Fig. 2B). We tested AAV-AD and control rats in the Y-maze and Novel Object Recognition (NOR) tasks to evaluate their working and

short-term memories. We observed no differences between the 2 groups (data not shown). Finally, we assessed spatial learning and short and long-term memories using the Morris Water Maze paradigm. There was no difference in the distance travelled to find the hidden platform or the time spent in the target quadrant (data not shown), that contains the platform, during the 5 consecutive training days (Fig. 2C). We then performed 3 probe trials to estimate short (4 h after the last trial) and long-term memories (mean of 72 h and 120 h probes; Fig. 2D,E). We detected no statistical differences between the groups.



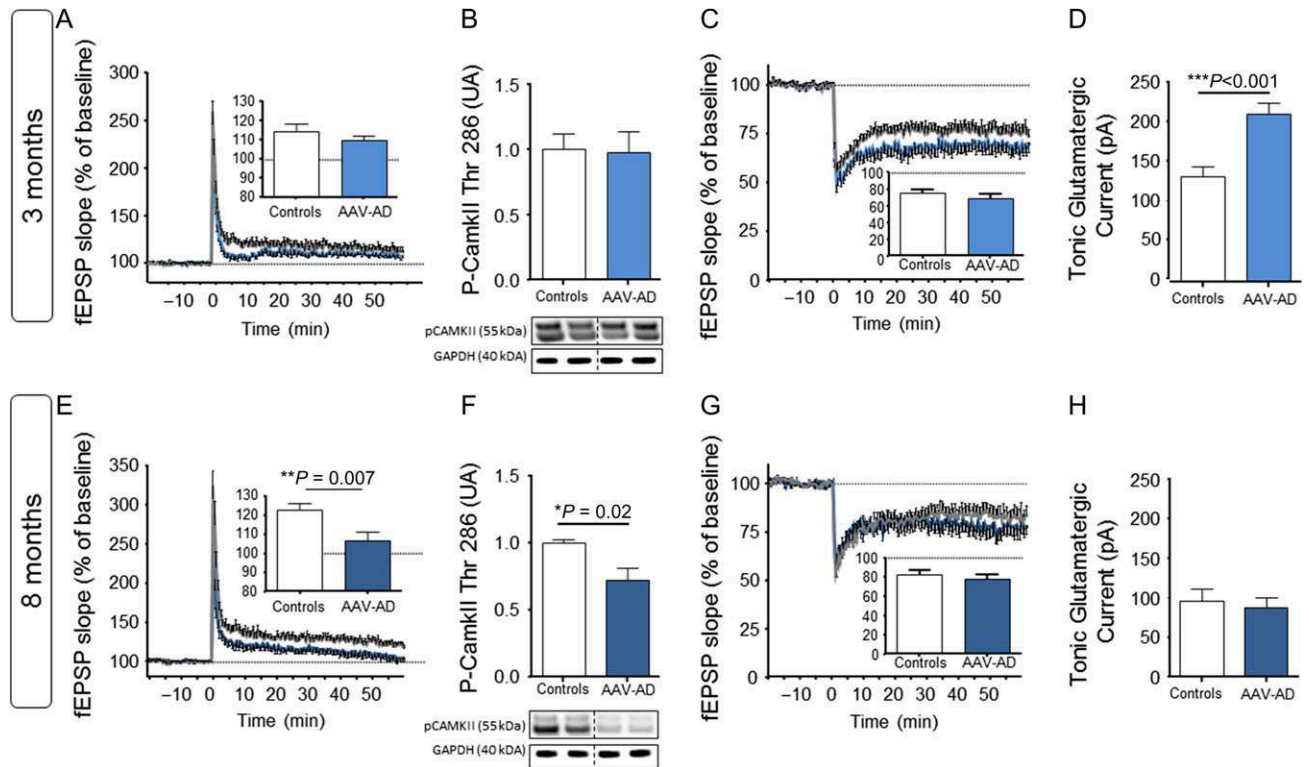
**Figure 2.** Early emotional disturbances and progressive long-term memory impairment. We evaluated behavioural impairment in AAV-AD rats relative to controls at 3 ( $n = 20$ ) and 8 months ( $n = 20$ ) after AD induction. (A, B) Open-field assay of 3-months AAV-AD rats. (A) Distance travelled during the open-field task. (B) Time spent in the centre of the open-field area. (C–E) Morris water maze assay of 3-months AAV-AD rats. (C) Learning curves representing the travelled distance during the 5 training days in the 4 quadrants (TQ = target quadrant that housed the platform during the training sessions, OQ = mean of other quadrants). (D) Probe trial performance at 4 h after the last training session. (E) Probe trial performance at 72 h and 120 h after the last training session. (F–G) Open-field assay of 8-months AAV-AD rats. (F) Distance travelled during the open-field task. (G) Time spent in the centre of the open-field area showing a change in emotional behaviour when faced to a new environment in AAV-AD rats. (H–J) Morris water maze assay of 8-month AAV-AD rats. (H) Travelled distance during the 5 training days in the 4 quadrants. (I) Probe trial performance at 4 h after the last training session. (J) Probe trial performance at 72 and 120 h after the last training session showing a long-term memory impairment in AAV-AD rats. For each figure, bottom group occupancy plots allow to visualize the areas in which the animals spent the most time during the test. Bars represent the means  $\pm$  SEM. Statistical analysis were performed using one-way ANOVA followed by Tukey's post hoc test or 2-way ANOVA followed by Sidak's post hoc test depending on the experimental design.

We then tested AAV-AD rats at 8 months after induction using the same approaches. Despite the absence of locomotor hypo-activity (Fig. 2F), AAV-AD rats spent significantly less time in the centre of the open-field apparatus (Control:  $4.15 \pm 1.17$  s; AAV-AD:  $1.50 \pm 0.43$  s; Student's *t*-test:  $P = 0.04$ ; Fig. 2G). We detected no behaviour differences in the Y-Maze or NOR tasks (data not shown). We detected no statistical differences during the learning session and the 4 h probe test (short-term memory; Fig. 2H,I) of the Morris Water Maze task, but did observe a significant decrease of the mean time spent (long-term memory: Control:  $22.40 \pm 0.92$  s; AAV-AD:  $18.54 \pm 0.99$  s; 2-way ANOVA followed by Sidak's post hoc test:  $P = 0.0064$ ; Supplementary Fig. 1) and mean distance travelled (long-term memory: Control:  $4.74 \pm 0.22$  m; AAV-AD:  $3.81 \pm 0.18$  m; 2-way ANOVA followed by Sidak's post hoc test:  $P = 0.0003$ ; Fig. 2J) in the target quadrant during the 72 h/120 h probes. Altogether, AAV-AD rats present emotional and memory impairment between 3 and 8 months postinduction, before both amyloid and phosphorylated Tau aggregations.

### Transient Extrasynaptic Deficits Precede LTP Impairment

Here, we evaluated synaptic plasticity and extrasynaptic receptor activation essential for newly formed memories. We first induced LTP at the Schaffer collateral to CA1 pathway by TBS in

control ( $n = 13$ ) and AAV-AD ( $n = 16$ ) rats, 3 months after induction (Fig. 3A). We observed no significant difference in LTP between the 2 groups (Student *t*-test:  $P > 0.05$ ), confirmed by the absence of the decrease in Thr 286 phosphorylation of calcium/calmodulin-dependent kinase II (CamkII), implicated in LTP mechanisms (Fukunaga et al. 1995; Lledo et al. 1995) (Fig. 3B). We then performed LTD recordings and again observed no significant differences between the AAV-AD and control groups (Student's *t*-test:  $P > 0.05$ ; Fig. 3C). In parallel, we recorded the tonic current generated in CA1 pyramidal cells by the ambient levels of glutamate acting on extrasynaptic NMDA receptors (NMDAR). The amplitude of the tonic current was significantly greater in the AAV-AD rats than in the control group (Control:  $133.8 \pm 11.84$  pA,  $n = 11$ ; AAV-AD:  $213.0 \pm 13.60$  pA,  $n = 16$ ; Student's *t*-test:  $P = 0.0004$ ; Fig. 3D). We recorded the same parameters in 8-month postinjected rats. AAV-AD rats exhibited significantly lower induction and maintenance of LTPs compared with the control rats (Control:  $122.8 \pm 3.35\%$ ,  $n = 13$ ; AAV-AD:  $106.6 \pm 4.41\%$ ,  $n = 11$ ; Student's *t*-test:  $P = 0.007$ ; Fig. 3E). This LTP impairment was associated with a decrease of CamkII phosphorylation (Control:  $1.0 \pm 0.02$ ; AAV-AD:  $0.72 \pm 0.08$ ; Student's *t*-test:  $P = 0.007$ ; Fig. 3F). In contrast, LTD, as the tonic glutamatergic current, was not impaired at this age (Fig. 3G,H). Altogether, these data suggest that transient extrasynaptic activation occurs before LTP impairment in the early phases.



**Figure 3.** Transient extrasynaptic activation preceded LTP impairment. (A–D) Electro-physiological properties of Glutamate receptors, 3 months after AAV-AD induction. The LTP (A), the P-CamKII levels (B), the long-term depression (LTD, C), and the tonic current representing the activation of extrasynaptic NMDA receptors (NMDAR, D) were evaluated in 3-month AAV-AD rats ( $n = 10$ ) relative to control rats ( $n = 10$ ). (E–H) Electro-physiological properties of Glutamate receptors, 8 months after AAV-AD induction: the LTP (E), the P-CamKII levels (F), the LTD (G), and the activation of extrasynaptic NMDAR (H) were quantified in 8-month AAV-AD rats ( $n = 13$ ) compared with control rats ( $n = 13$ ). Bars represent the mean  $\pm$  SEM. Statistical analyses were performed using the Student's *t*-test.

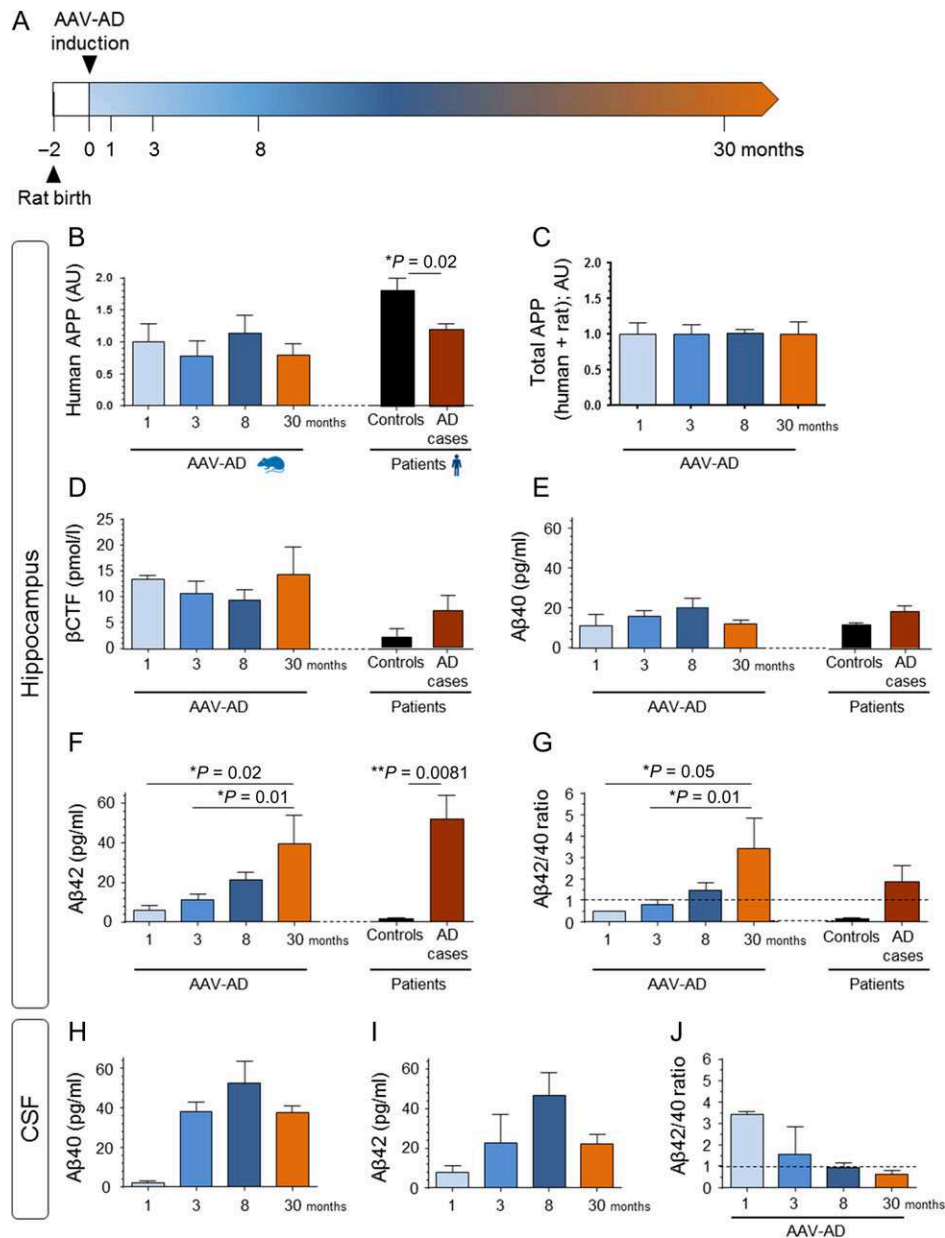
### AAV-AD Rats Develop Progressive AD-Like APP Processing Over Time in the Hippocampus

We established that human APP of AAV-AD rats is processed following the amyloidogenic pathway. We next measured the levels of these metabolites in the hippocampi of AAV-AD rats 1 ( $n = 4$ ), 3 ( $n = 15$ ), 8 ( $n = 15$ ), and 30 months ( $n = 5$ ) after AD induction to evaluate their evolution over time (Fig. 4A). These were then compared to those measured in human AD hippocampi (5 age-matched controls + 5 AD Braak 6/Thal 5 patients), which served as a reference. Human APP as total APP levels, did not increase with time in AAV-AD rats (Fig. 4B,C). Human APP content was lower in AD cases than in age-matched controls (Student's *t*-test:  $P = 0.02$ ) and remarkably close to the APP content in the hippocampi of AAV-AD rats (Fig. 4B; AAV-AD: 1 month:  $1.0 \pm 0.28$ , 3 months:  $0.78 \pm 0.24$ , 8 months:  $1.13 \pm 0.29$ , 30 months:  $0.79 \pm 0.18$ ; human controls:  $1.77 \pm 0.19$ ; AD patients:  $1.17 \pm 0.10$ ). Similarly,  $\beta$ CTF levels did not increase over time (Fig. 4D). We did not observe a significant increase in A $\beta$ 40 production, whereas A $\beta$ 42 content progressively increased (Fig. 4E,F; 1 month:  $5.93 \pm 2.58$  pmol/L; 3 months:  $11.41 \pm 2.75$  pmol/L; 8 months:  $20.94 \pm 4.26$  pmol/L; 30 months:  $39.26 \pm 14.40$  pmol/L; 1-way ANOVA: time effect:  $P = 0.01$ ), leading to a steady increase in the A $\beta$ 42/40 ratio in the hippocampi of AAV-AD rats from 1 to 30 months after injection (Fig. 4G; 1 month:  $0.47 \pm 0.02$ ; 3 months:  $0.81 \pm 0.22$ ; 8 months:  $1.45 \pm 0.37$ ; 30 months:  $3.412 \pm 1.41$ ; 1-way ANOVA: time effect:  $P = 0.02$ ). The A $\beta$ 40 and A $\beta$ 42 content, as well as the A $\beta$ 42/40 ratio at 30 months Post-injection were comparable to those observed in the hippocampus of late stage AD patients (Fig. 4E,G).

A similar analysis of the CSF confirmed an increase of both human amyloid peptides in AAV-AD rats at 1, 3, and 8 months postinjection (A $\beta$ 40: 1 month:  $2.14 \pm 1.02$  pg/mL; 3 months:  $38.07 \pm 4.79$  pg/mL; 8 months:  $52.23 \pm 11.45$  pg/mL; A $\beta$ 42: 1 month:  $7.43 \pm 3.75$  pg/mL; 3 months:  $22.44 \pm 14.54$  pg/mL; 8 months:  $46.37 \pm 11.54$  pg/mL). In contrast, both A $\beta$ 40 ( $37.47 \pm 3.65$  pg/mL) and A $\beta$ 42 ( $21.89 \pm 495$  pg/mL) tended to decrease in CSF samples at 30 months postinjection (Fig. 4H,I). The CSF A $\beta$ 42/40 ratio progressed in the opposite direction relative to that of the hippocampus, decreasing over time (Fig. 4J; 1 month:  $3.40 \pm 0.13$ ; 3 months:  $1.54 \pm 1.32$ ; 8 months:  $0.91 \pm 0.25$ ; 30 months:  $0.60 \pm 0.19$ ). In summary, we showed that our model faithfully mimics human amyloid processing (A $\beta$  levels and A $\beta$ 42/40 ratio) without APP accumulation in either the hippocampus or CSF.

### Human APP Processing Induces a Delayed Metabolism of the Endogenous APP in Rat Hippocampus

We then evaluated the consequences on the endogenous rodent APP processing. Despite absence of differences during the 8 first months, endogenous rat A $\beta$ 42 peptides belatedly increased in 30-month-old AAV-AD rat hippocampus relative to age-matched controls (2-way ANOVA followed by Sidak's post hoc test: Interaction effect:  $P = 0.006$ ; 1 month: ns; 3 months: ns; 8 months: ns; 30 months:  $P = 0.0005$ ; Supplementary Fig. 2B). By contrast, similar concentrations of rodent A $\beta$ 42 peptides was measured in CSF between controls and AAV-AD rats (2-way ANOVA followed by Sidak's post hoc test: Interaction effect: ns; Group effect: ns; Supplementary Fig. 2C). These results suggest a secondary engagement of the rodent APP



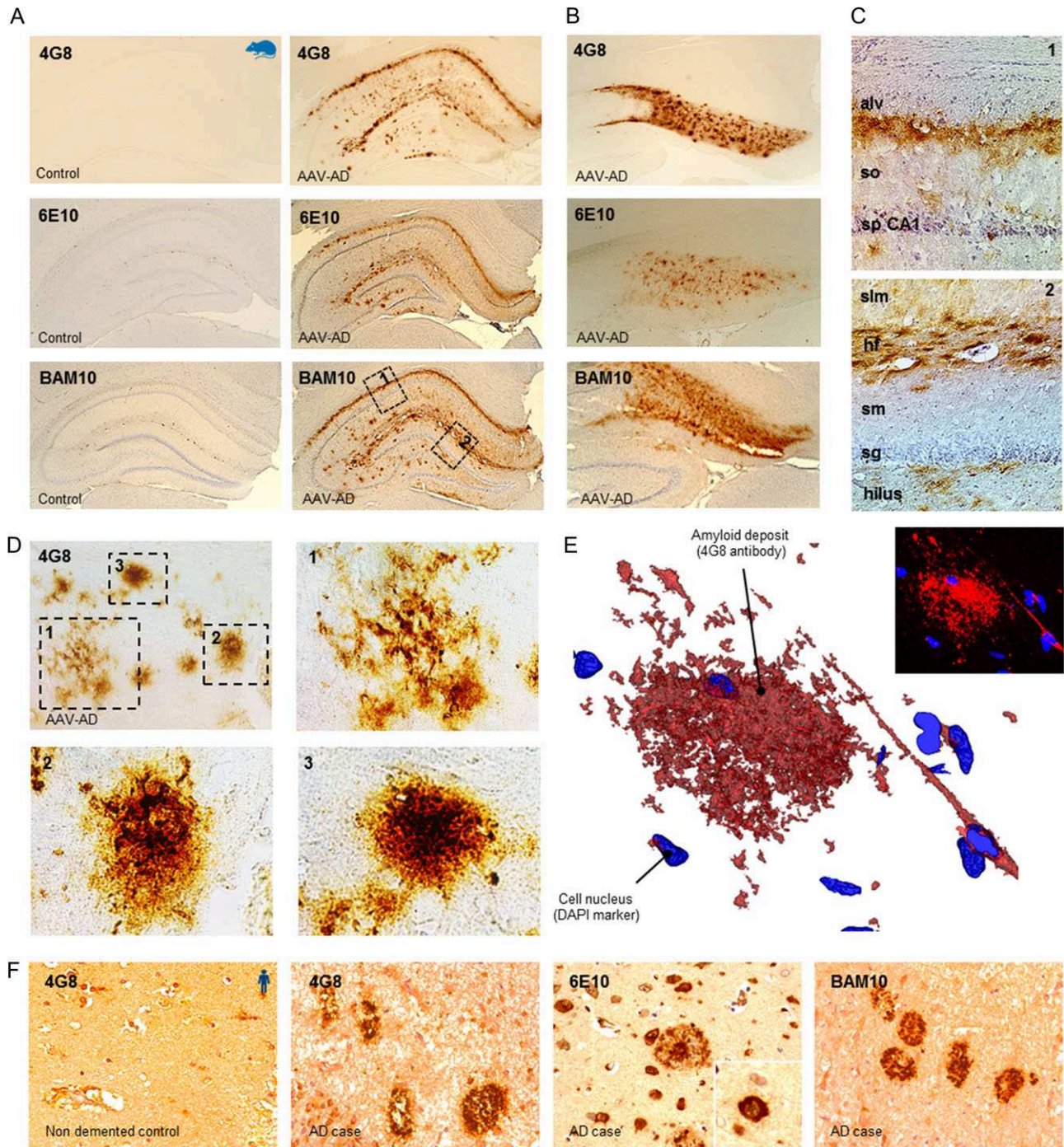
**Figure 4.** Aβ42/40 ratio increases in hippocampus and decreases in CSF from 1 to 30 months. Metabolites resulting from the amyloidogenic pathway were released into the hippocampus of AAV-AD rats. Rats were killed at 1 ( $n = 4$ ), 3 ( $n = 15$ ), 8 ( $n = 15$ ), and 30 ( $n = 5$ ) months after AAV-AD induction. Human hippocampal samples were obtained from late-onset AD cases (Braak 6, Thal 5,  $n = 5$ ) and age-matched controls ( $n = 5$ ). (A) Schematic representation of the experimental design for the progressive AAV-AD rat model. (B–C) Densitometric analyses of western blots of human APP (6E10 antibody, B) and total APP (rat + human forms; 22C11 antibody, C) of the hippocampus of AAV-AD rats for each time-point and for human samples ( $n = 5$  per group). (D–F) Comparative analysis of TBS-Tx soluble βCTF (D), Aβ40 (E), and Aβ42 (F) levels by Elisa and MSD multiplex respectively. (G) Representation of the Aβ42/40 ratio determined by multiplex MSD immunoassay. (H–J) Comparative analysis of soluble Aβ40 (H) and Aβ42 (I) CSF levels by MSD multiplex. (J) Representation of the CSF Aβ42/40 ratio over time. Bars represent the means  $\pm$  SEM. Statistical analyses were performed using 1-way ANOVA followed by Tukey's post hoc test.

processing in the AAV-AD rat hippocampus led by the human APP metabolism.

### Hippocampal Senile Plaques and CAA Appear at Late Stages in the AAV-AD Rat Model

Most AD animal models develop early amyloid plaques due to massive overexpression of human transgenes, resulting in Aβ accumulation over time that largely exceeds human pathophysiological levels. Here, we evaluated amyloid deposition in AAV-AD

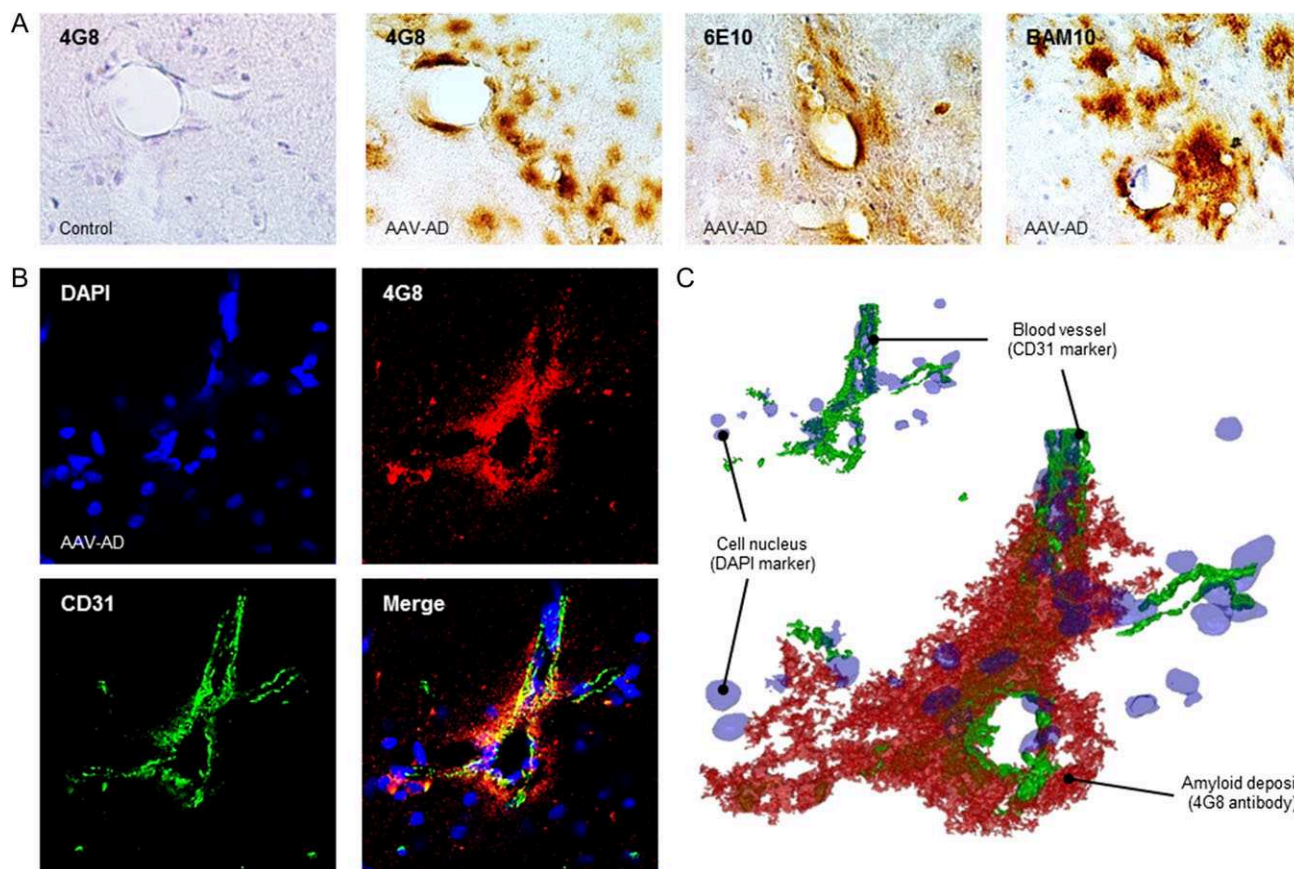
rats with several antibodies that detect Aβ. Immunostaining with 4G8, 6E10, and BAM10 antibodies revealed hippocampal amyloid plaques at 30 months postinjection (Fig. 5A). The plaques were mostly located in the subiculum of the hippocampus (Fig. 5B), and were also present at the junction of different tissue layers (Fig. 5C). Nonetheless, we detected plaques at different stages of their formation (Fig. 5D) from diffuse to dense core plaques. Confocal imaging and 3D reconstruction of 4G8 immunostaining lesions confirmed the presence of amyloid lesions in the characteristic form of senile plaques (Fig. 5E). We used 4G8, 6E10, and BAM10



**Figure 5.** AD-like cerebral amyloid lesions in 30-month-old AAV-AD rats. (A) Immunohistochemistry with 4G8, 6E10, and BAM10 antibodies showing amyloid deposition in coronal sections of AAV-AD rat hippocampi. (B) Plaque pathology in the subiculum of AAV-AD rats using the antibodies described above. (C) Higher magnification of amyloid deposition at the junction of tissue layers. alv, alveus; so, stratum oriens; sp CA1, stratum pyramidale of the CA1 region; slm, stratum lacunosum moleculare; hf, hippocampal fissure; sm, stratum moleculare; sg, stratum granulosum. (D) Examples of amyloid plaques at different maturation stages from a diffuse state (box 1) to a dense morphology (box 3). (E) 3D reconstruction illustrating the typical form of a 4G8 positive senile plaque in hippocampus of AAV-AD rat. (F) Immunohistochemistry with the antibodies described above revealing amyloid plaque lesions in AD patients vs non demented controls.

antibodies to confirm the presence of amyloid deposits in human AD tissue (Fig. 5F). We also carried out immunostaining and western blotting for GFAP to analyse potential astrocytic responses induced by amyloid APP processing in the AAV-AD rat model. We did not observe astrogliosis at any time-point after AAV-AD induction (Supplementary Fig. 3).

In order to characterize the cerebral angiopathy in the AAV-AD rat hippocampus, we performed immunostainings with the same previously described antibodies. In 30-month-old AAV-AD rats, we observed characteristic CAA around the blood vessels layers (Fig. 6A). In order to better characterize the vessel-surrounding aggregation, we proceed to a double



**Figure 6.** APP processing triggers CAA in 30-month-old AAV-AD rats. (A) Immunohistochemistry with 4G8, 6E10, and BAM10 antibodies showing respectively endogenous and exogenous CAA in coronal sections of AAV-AD rat hippocampi. (B) Laser confocal microscopy showing aggregation of endogenous amyloid (4G8) surrounding cerebral vessel identified with endothelial cells marker (CD31) in coronal sections of AAV-AD rat. (C) 3D reconstruction illustrating the CAA in sections of AAV-AD rat.

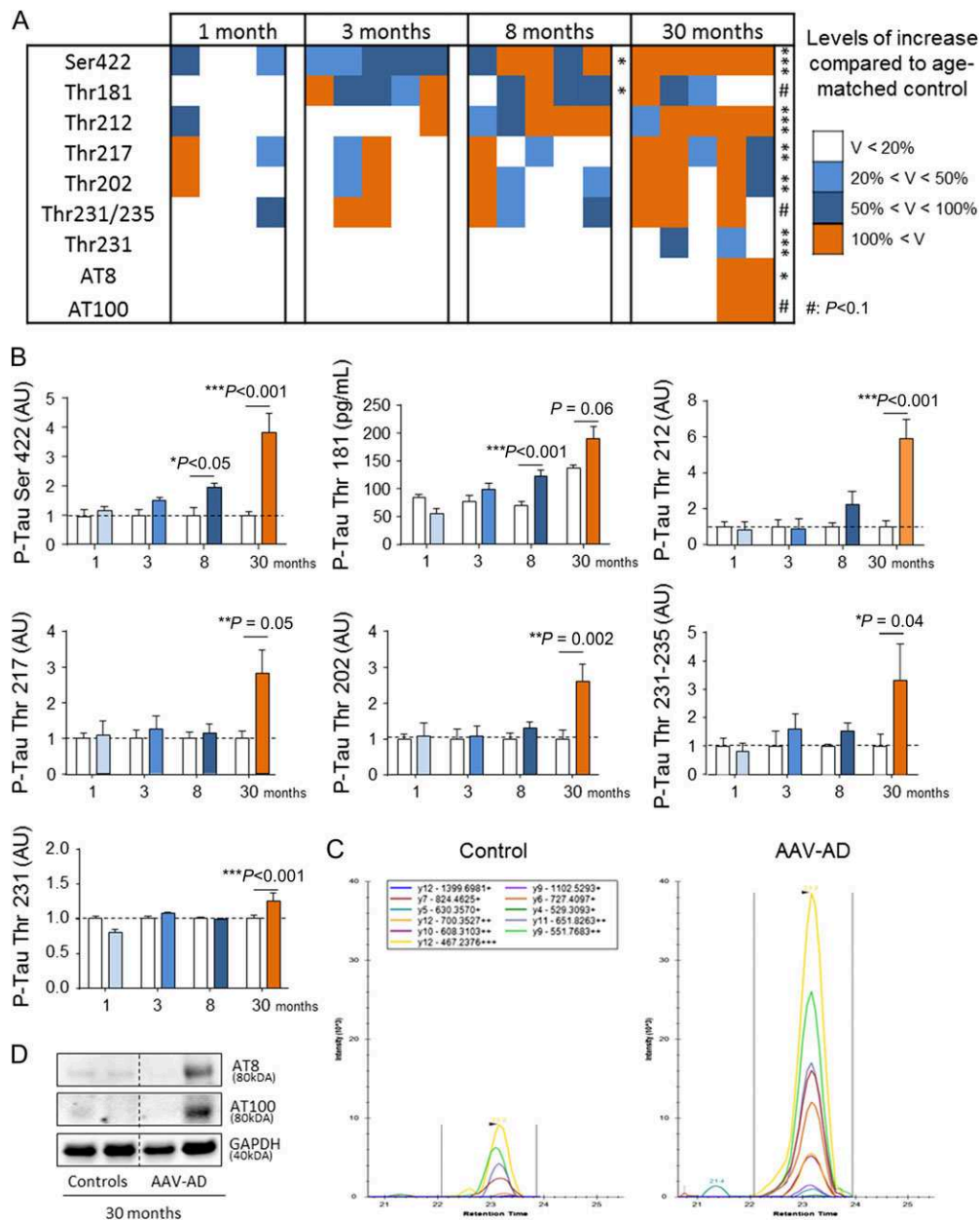
staining with both 4G8 and CD31 (a vessel marker) antibodies (Fig. 6B). Thus, we observed a blood vessel-surrounding amyloid sheath, observation confirmed by the 3D reconstruction (Fig. 6C). Altogether, these data indicate that a model exhibiting a slow progression to the AD phenotype can provide information about the initiation of amyloid deposition in the brain parenchyme as plaques and around vessels as CAA.

### Progressive Hyper- and Abnormal Phosphorylations of the Endogenous Tau Protein

We analysed the consequences of the amyloidogenic pathway on the phosphorylation of Tau. We used 4 different approaches (ELISA, western blot, HPLC and MSD multiplex immunoassay) to evaluate the phosphorylation of the endogenous Tau protein (P-Tau) on both hyperphosphorylated (thr181, thr212, thr217, thr202, thr231/235, and thr231) or abnormally phosphorylated sites (ser 422) (Augustinack et al. 2002) at 1 ( $n = 5$ ), 3 ( $n = 5$ ), 8 ( $n = 5$ ), and 30 months ( $n = 5$ ) postinduction. The global trend of Tau phosphorylation was equivalent for each studied phosphorylated site, and gradually increased over time relative to age-matched controls (Fig. 7A) (Time effect: thr181:  $P < 0.001$ , thr212:  $P < 0.002$ , thr217:  $P = 0.07$ , thr202:  $P < 0.04$ , thr231/235:  $P = 0.21$ , ser 422:  $P < 0.0001$  and thr231:  $P < 0.001$ ; Group effect: thr181:  $P < 0.05$ , thr212:  $P < 0.001$ , thr217:  $P = 0.04$ , thr202:  $P < 0.03$ , thr231/235:  $P = 0.08$ , ser 422:  $P < 0.0001$  and thr231: ns; Interaction effect: thr181:  $P = 0.05$ , thr212:  $P < 0.001$ , thr217:  $P = 0.05$ , thr202:  $P < 0.04$ , thr231/235:  $P = 0.08$ , ser 422:  $P < 0.0001$ ,

and thr231:  $P < 0.001$ ). The thr181 and ser422 epitopes early appeared to be phosphorylated by 8 months (thr181: Control:  $69.8 \pm 7.48$  pg/mL; AAV-AD:  $124.30 \pm 12.01$  pg/mL; 2-way ANOVA followed by Sidak's post hoc test  $P < 0.001$ ; ser422: Control:  $1.00 \pm 0.28$ ; AAV-AD:  $1.96 \pm 0.14$ ; 2-way ANOVA followed by Sidak's post hoc test  $P = 0.017$ ). Moreover, the phosphorylation of all studied epitopes was higher than in age-matched controls by 30 months after AAV-AD induction (2-way ANOVA followed by Sidak's post hoc test: thr181:  $P = 0.06$ , thr217:  $P = 0.005$ , thr202:  $P = 0.002$ , thr231/235:  $P = 0.04$ , ser422:  $P < 0.0001$ , thr212:  $P < 0.001$ , and thr231:  $P < 0.001$ ) (Fig. 7B,C).

Biochemical assessment of hippocampus samples using AT8 and AT100 antibodies confirmed the hyper- and abnormal phosphorylation of the endogenous Tau protein but only in half of 30-month-old AAV-AD rats (Fig. 7D). Thus, western blotting revealed characteristic bands close to 80 kDa using both antibodies in these animals. Interestingly these antibodies are currently used in clinics to deliver a definitive diagnosis in postmortem brain samples because of their capacity to mark tangles structures. Finally, we evaluated 3 kinases involved in the AD phosphorylation of Tau by western blotting: cyclin-dependent kinase 5 (CDK5), glycogen synthase kinase 3 $\beta$  (GSK3 $\beta$ ), and Dual specificity tyrosine-phosphorylation-regulated kinase 1A (DYRK1A). We observed a continuous increase of GSK3 $\beta$  and DYRK1A levels from 1 to 30 months postinduction (2-way ANOVA: Time effect: GSK3 $\beta$ :  $P < 0.0036$  and DYRK1A:  $P < 0.001$ ), whereas CDK5 levels remained stable (2-way ANOVA: Time effect:  $P = 0.14$ ) (Fig. 7E,F; Supplementary Fig. 4).



**Figure 7.** APP processing triggers sequential endogenous hyper- and abnormal Tau phosphorylation from 8 months postinjection. (A and B) Comparative analysis of TBS-Tx soluble rat phosphorylated Tau protein at residue Ser422, Thr212 (by western blot), Thr181 (by ELISA), Thr217, Thr, 202, 231-235 (by HPLC), and Thr231 (by MSD immunoassay). (C) LC-MS/MS signal of phosphorylated Thr 217 peptide ion fragments used for quantification in hippocampal samples from 30-month-old AAV-AD rat and age-matched control. (D) Representative western blot showing both AT8 and AT100 bands in AAV-AD rats hippocampi. (E, F) Densitometric analyses of western blots showing the expression of CDK5, GSK3 $\beta$ , and DYRK1A in each group and at different time points ( $n = 5$  rats per group). Bars represent the means  $\pm$  SEM. Statistical analyses were performed using 2-way ANOVA followed by Sidak's post hoc test.

### Hippocampal Tangle-Like Aggregates Appeared at Late Stages in the AAV-AD Rat Model

We performed immunohistochemistry with the AT8 (hyperphosphorylated epitopes: ser202/thr205) and AT100 (abnormal epitopes: thr212/ser214) antibodies to assess the potential aggregation of phosphorylated Tau measured in 30 months AAV-AD rats. We observed AT8 and AT100 positive cells in the *hilus* and *subiculum* (Fig. 8A). Strikingly, we observed a very similar pattern in hippocampal slices of AD patients, displaying AT8 and AT100 positive core dense cells (Fig. 8B). Confocal imaging and 3D

reconstruction of AT8 immunostaining confirmed the presence of hyperphosphorylated endogenous Tau surrounding nucleus in the characteristic form of misfolded protein aggregates (Fig. 8C,D). All these compelling data demonstrate that continuous human-like  $\beta$ APP processing during years induces accretion of phosphorylated Tau protein into tangle-like aggregates.

### Discussion

Understanding the infra-clinical phases and finding new early biomarkers will be decisive in the development of effective

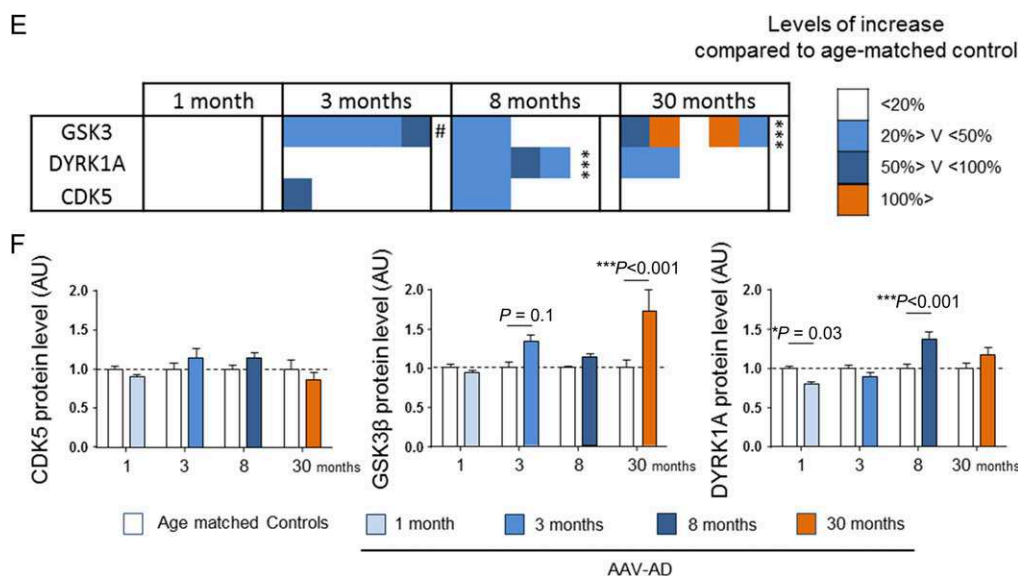


Figure 7. Continued.

preventive or therapeutic molecules to counteract AD (Langbaum et al. 2013). It is thus essential to dispose models that closely mimic human AD whether on the evolution over time or on the amyloid peptides levels produced (Sala Frigerio and De Strooper 2016). To address this challenge, we developed an original AD rat model by co-transfer of the APP<sub>SL</sub> and PS1<sub>M146L</sub> genes using hippocampal AAV vector delivery. This model shows a specific age-dependent progression pattern. Longitudinal analysis of this model has allowed us to experimentally confirm the causal relationship between human-like  $\beta$ APP processing and Tau pathology including Tau tangle-like aggregation in order to propose a 4-step hypothesis of prodromal AD progression (Fig. 9).

We generated AAV vectors encoding mutated human PS1<sub>M146L</sub> or APP<sub>SL</sub>, which were bilaterally injected into young adult rat hippocampus to generate the AAV-AD rat model. AAV-PS1<sub>M146L</sub> injected rats were used as controls. AAV-AD rats produced APP from 1-month after vector injection. Human APP immunoreactivity was found in sparse neurons of the CA2/CA3 layers, subiculum, and dentate gyrus. This pattern of expression may mimic the genomic mosaicism recently described in sporadic forms of human AD, in which an increase in copy number was observed for the APP gene in a limited subset of neurons (Bushman et al. 2015). The absence of APP accumulation in the hippocampus of AAV-AD rats, which was at levels similar to those in both control rats and AD patients, suggests that human APP was quickly catabolized by the  $\beta$ -secretase complex. This was corroborated by the rapid effect of the BACE inhibitor observed in less than 4 h. Moreover,  $\beta$ CTF, A $\beta$ 40, and A $\beta$ 42 peptides were produced in the hippocampus from 1-month after AAV-AD induction, whereas the levels of APP remained unchanged (Fig. 9, Step 1). Reducing A $\beta$ 42 peptide levels or normalizing the A $\beta$ 42/40 ratio by inhibiting amyloid APP processing appears to be the best therapeutic strategy at this first step of disease progression.

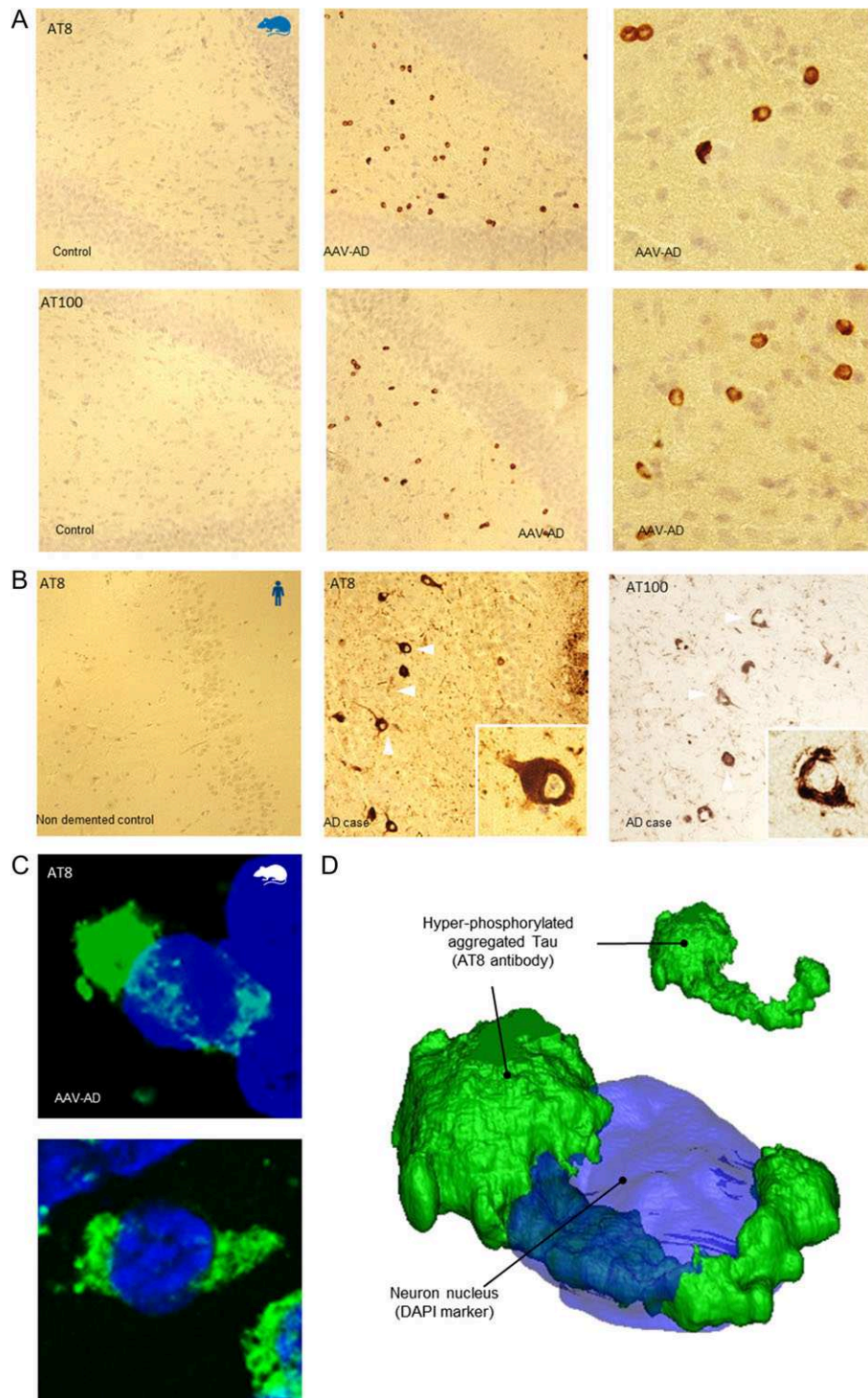
The concentration of soluble A $\beta$ 42 and the A $\beta$ 42/40 ratio gradually increased over time, whereas  $\beta$ CTF and A $\beta$ 40 levels remained stable. By 30 months after induction, the relative amount of A $\beta$  in the brains of the rats was similar to that found in AD patients. Analysis of the CSF revealed A $\beta$  diffusion from the hippocampus to the CSF. In contrast to the increased

hippocampal A $\beta$ 42/40 ratio, the CSF A $\beta$ 42/40 ratio progressively decreased over time (Fig. 9, Step 1). These data suggest that the ability of soluble A $\beta$ 42 to spread from the hippocampus to the CSF decreased over time, leading to an increase in A $\beta$ 42 levels in the hippocampus and a reduction in the CSF. This may be due to progressive CAA leading to endothelial membrane modifications as we observed in 30-month AAV-AD rats: as the level of vascular angiopathy increases, so does the level of A $\beta$ 42 in the hippocampus, whereas it decreases in the CSF (Ikeda et al. 2012). However, some therapeutic strategies (i.e., A $\beta$  immunotherapy) developed to remove cerebral A $\beta$ 42 aggregates also increase the severity of CAA (Patton et al. 2006; Weller et al. 2009). Strategies that prevent vascular amyloidosis and rebalance hippocampal and CSF A $\beta$  ratios may be beneficial (Agyare et al. 2014).

Emotional behaviour was altered without memory defects within 3 months after induction (Fig. 9, step 2). This observation suggests that there may be specific emotional neuropsychological disturbances that could be used as a potential readout in clinic practice as a very early cognitive symptom in AD patients (McCade et al. 2011).

The amplitude of the tonic current was greater in AAV-AD rats 3 months after induction as a direct consequence of AAV-AD induction. Progression of the disease resulted in secondary normalization of extrasynaptic NMDA receptor (eNMDAR) activation. Accordingly, transient extrasynaptic alterations appeared shortly after the A $\beta$ 42-induced neuronal stress. Such observations have already been described after the application of A $\beta$ 42 to mixed neuronal/glial rat cerebrocortical cultures (Talantova et al. 2013). Several events, such as a decrease of scaffolding proteins, a reduction of the glutamate transporter GLT-1, or an increase in glutamate liberation by astrocytes can lead to increased stimulation of the eNMDARs (Masliah et al. 1996; Li et al. 2011). Extrasynaptic NMDA receptor activation is implicated in the pathogenesis of several CNS disorders and A $\beta$  disturbs the balance between synaptic and eNMDAR activation (Wang et al. 2013). We show here, for the first time in vivo in our knowledge, that these transient events appear at the earliest stages of the disease course, before the onset of classical hallmarks such as amyloid plaques, neurofibrillary tangles or memory and LTP impairment (Fig. 9, step 2).

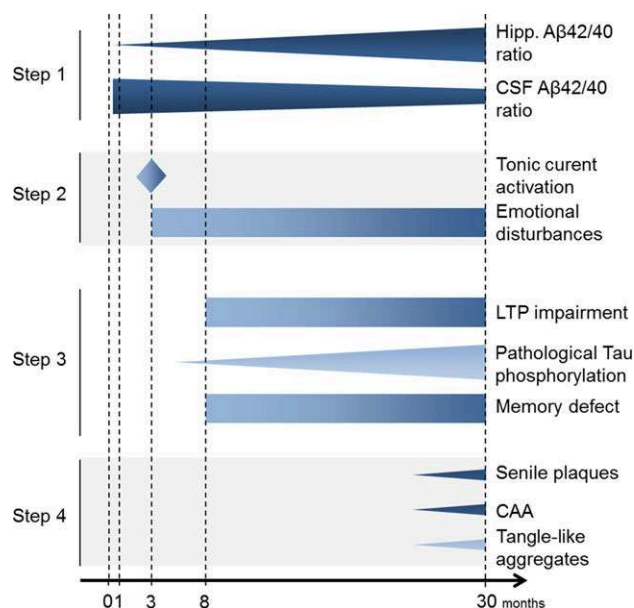




**Figure 8.** APP processing triggers terminal endogenous phosphorylated Tau aggregation in 30-month-old AAV-AD rats. (A) Immunohistochemistry with AT8 and AT100 antibodies showing respectively pretangle and tangle neurons located in the hilus of the DG in AAV-AD rat coronal sections. (B) Immunohistochemistry with AT8 and AT100 antibodies showing respectively pretangle and tangle lesions in AD patients. (C) Laser confocal microscopy showing AT8 positive cells in coronal sections of AAV-AD rat. (D) 3D reconstruction illustrating the typical helicoidal form of an AT8 positive pretangle neuron in hippocampus of AAV-AD rat.

The third step of AD progression may be related to the first memory problems in patients before clinical diagnosis (Rajan et al. 2015). We evaluated the LTP of glutamate synaptic transmission in AAV-AD rats at 3 and 8 months postinduction. We observed no defects in LTP in 3-month AAV-AD rats. In contrast, a large decrease in LTP may be attributable to AD progression in

8-month AAV-AD rats (Fig. 9, step 3). Long-term memory was intact 3 months after induction, as opposed to after 8 months, in which the AAV-AD rats displayed memory impairment (Fig. 9, step 3). Indeed, we observed long-term memory impairment in association with LTP defects from 8 months postinduction and later, suggesting progressive memory disturbances.



**Figure 9.** Age-dependent progression of pre-clinical AD stages: the 4-steps hypothesis.

We then investigated the potential impact of the amyloidogenic pathway on endogenous Tau phosphorylation. ELISA analysis demonstrated higher levels of Tau protein phosphorylation on residues thr181 and ser422 in the AAV-AD rat hippocampus than in control rats, from 8 months postinduction. These results confirm that the continuous increase of A $\beta$  levels, resulting from APP processing, was sufficient to trigger Tau pathway engagement as observed in AD patients (Fig. 9, step 3) (Selkoe and Hardy 2016). In addition, this engagement of the Tau pathway was associated with an increased level of specific kinases, such as GSK3 $\beta$  and DYRK1A. Strikingly, almost all amyloid transgenic mice do not develop progressive tau pathology in the brain (Kitazawa et al. 2012). Based on these experimental results, rejection of the amyloid cascade hypothesis was suggested arguing that pathological APP processing is not sufficient to drive tau pathology in rodent brain. However, our results clearly link APP and Tau pathologies and strengthens the amyloidogenic cascade hypothesis.

The next phase of AD disease progression consists of the appearance of AD-related cerebral lesions. Senile plaques and tangle-like aggregates appeared in AAV-AD hippocampus at least 30 months after AAV-AD induction (Fig. 9, step 4). These findings confirm the role of gradually increasing levels of the soluble forms of A $\beta$ , and potentially the A $\beta$ 42/40 ratio, starting from adulthood in both senile plaque and tangle formation, several years after induction. This may explain, in part, why subtle cognition/biological defects are present at least 18 years before clinical diagnosis (Rajan et al. 2015). Our results are once again in agreement with the evolution of the disease in humans. Thus, in experimental models with APP pathology close to AD patient, senile plaque were associated with Tau aggregation process as observed in AD brains. Unfortunately, it was not possible to observe the following steps of AD progression using AAV-AD rats because of their limited lifespan. It is possible that step 4 correspond to the time of AD diagnosis in patients.

How to explain that AAV-AD rats develop both APP and Tau pathologies while usual amyloid models failed? This main difference compared to classic AD models could be explained by

the technology used. Because induction is realized on adult animals, AAV-AD rats do not suffer from developmental compensation that could result in neuroprotection mechanisms breaking down Tau pathology. Another argument could be that induced APP pathology appear similar to human in terms of amyloid peptides amounts and A $\beta$ 42/40 ratio resulting in pathophysiology mechanisms including Tau pathology. The last difference lies in slow progression in setting the APP pathology, allowing time for Tau pathology development without early inflammation and plaque formation interfering.

In summary, this study reports the first experimental demonstration of amyloid cascade in rodents thanks to an inducible progressive AD rat model named AAV-AD rat. We propose a 4-step hypothesis for infra-clinical AD progression based on the study of the AAV-AD model over time. Step 1 consists of the enhancement of APP processing resulting in a steady increase of A $\beta$  levels and the A $\beta$ 42/40 ratio in the hippocampus starting from adulthood. Step 2 is characterized by the transient activation of extrasynaptic glutamate receptors and the onset of altered emotional behaviour. During Step 3, classical biological alterations gradually appear such as a defect in LTP, hyperphosphorylation of endogenous Tau, and long-term memory deficits. Finally, old AAV-AD rats develop specific histological lesions consisting of amyloid plaques and tangle-like aggregates (Fig. 9, step 4). This 4-step hypothesis highlights the need for pleiotropic therapies acting at least on both amyloid and Tau pathologies. Our AAV-AD rat model may be a flexible *in vivo* experimental setting to increase the understanding of AD onset, identify early AD biomarkers, while evaluating the efficacy of therapeutic strategies to counteract AD progression.

## Authors' Contributions

M.A., B.S., and J.B. carried out the design of the study and wrote the manuscript. M.A., B.S., and J.B. performed the biochemical, behavioural, histological, and statistical analyses. R.F., S.A., N.S.O., S.T. participated in the biochemical analyses and helped to draft the manuscript. Y.A. helped for the Bace1 inhibitor experiment as well as the behavior. S.A. performed the immunostaining on human slices. P.D., B.P., J.-M.B. and A.H. performed the electro-physiological recordings and helped to design the study and draft the manuscript. A.V. and F.B. realized the HPLC experiments. T.D. performed the 3D reconstruction. A.-P.B. and C.J. helped with the viral production. N.C., P.H. and N.D. read the manuscript. All authors read and approved the final manuscript.

## Supplementary Material

Supplementary data are available at *Cerebral Cortex* online.

## Funding

This work was supported by DIM Biotherapies, the Région Ile-de-France, IDEX Paris Saclay, INSERM Transfert, CEA "Technologie Pour la Santé" and by a grant from "Investissement d'Avenir—ANR-11-INBS-0011"—NeurATRIS: A Translational Research Infrastructure for Biotherapies in Neurosciences. Human post-mortem samples were obtained from the GIE-Neuro-CEB brain bank, which is run by a consortium of patient associations.

## Notes

We thank Noëlle Dufour, and Gwenaëlle Auregan for their help in producing the viral vectors, Gaëlle Dufayet-Chaffaud for her

help in the biochemical analyses, and Karine Cambon for advice on animal behaviour. *Conflict of interest:* The authors have no competing financial interests to declare.

## References

- Abbott A. 2004. Laboratory animals: the Renaissance rat. *Nature*. 428:464–466.
- Agyare EK, Jaruszewski KM, Curran GL, Rosenberg JT, Grant SC, Lowe VJ, Ramakrishnan S, Paravastu AK, Poduslo JF, Kandimalla KK. 2014. Engineering theranostic nanovehicles capable of targeting cerebrovascular amyloid deposits. *J Control Release*. 185:121–129.
- Alves S, Bode J, Bemelmans A-P, Kalle Von C, Cartier N, Tews B. 2016. Ultramicroscopy as a novel tool to unravel the tropism of AAV gene therapy vectors in the brain. *Sci Rep*. 6:28272.
- Audrain M, Fol R, Dutar P, Potier B, Billard J-M, Flament J, Alves S, Burlot M-A, Dufayet-Chaffaud G, Bemelmans A-P, et al. 2016. Alzheimer's disease-like APP processing in wild-type mice identifies synaptic defects as initial steps of disease progression. *Mol Neurodegener*. 11:5.
- Augustinack JC, Schneider A, Mandelkow E-M, Hyman BT. 2002. Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathol*. 103:26–35.
- Barthélemy NR, Fenaille F, Hirtz C, Sergeant N, Schraen-Maschke S, Vialaret J, Buée L, Gabelle A, Junot C, Lehmann S et al. 2016. Tau Protein Quantification in Human Cerebrospinal Fluid by Targeted Mass Spectrometry at High Sequence Coverage Provides Insights into Its Primary Structure Heterogeneity. *J Proteome Res*. 15(2):667–676.
- Berger A, Lorain S, Josephine C, Desrosiers M, Peccate C, Voit T, Garcia L, Sahel JA, Bemelmans AP. 2015. Repair of Rhodopsin mRNA by Spliceosome-Mediated RNA Trans-Splicing: A New Approach for Autosomal Dominant Retinitis Pigmentosa. *Mol Ther*. 23:918–30.
- Bushman DM, Kaeser GE, Siddoway B, Westra JW, Rivera RR, Rehen SK, Yung YC, Chun J. 2015. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife*. 4:e05116.
- Do Carmo S, Cuello AC. 2013. Modeling Alzheimer's disease in transgenic rats. *Mol Neurodegener*. 8:37.
- Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, et al. 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol*. 13:614–629.
- Echeverria V, Ducatenzeiler A, Alhonen L, Janne J, Grant SM, Wandosell F, Muro A, Baralle F, Li H, Duff K, et al. 2004. Rat transgenic models with a phenotype of intracellular Abeta accumulation in hippocampus and cortex. *J Alzheimers Dis*. 6:209–219.
- Ferreira ST, Lourenco MV, Oliveira MM, De Felice FG. 2015. Soluble amyloid- $\beta$  oligomers as synaptotoxins leading to cognitive impairment in Alzheimer's disease. *Front Cell Neurosci*. 9:191.
- Flood DG, Lin Y-G, Lang DM, Trusko SP, Hirsch JD, Savage MJ, Scott RW, Howland DS. 2009. A transgenic rat model of Alzheimer's disease with extracellular Abeta deposition. *Neurobiol Aging*. 30:1078–1090.
- Fol R, Braudeau J, Ludewig S, Abel T, Weyer SW, Roederer J-P, Brod F, Audrain M, Bemelmans A-P, Buchholz CJ, et al. 2015. Viral gene transfer of APPs $\alpha$  rescues synaptic failure in an Alzheimer's disease mouse model. *Acta Neuropathol*. 131:247–266.
- Folkesson R, Malkiewicz K, Kloskowska E, Nilsson T, Popova E, Bogdanovic N, Ganten U, Ganten D, Bader M, Winblad B, et al. 2007. A transgenic rat expressing human APP with the Swedish Alzheimer's disease mutation. *Biochem Biophys Res Commun*. 358:777–782.
- Fukunaga K, Muller D, Miyamoto E. 1995. Increased phosphorylation of Ca<sup>2+</sup>/calmodulin-dependent protein kinase II and its endogenous substrates in the induction of long-term potentiation. *J Biol Chem*. 270:6119–6124.
- Hardy JA, Higgins GA. 1992. Alzheimer's disease: the amyloid cascade hypothesis. *Science*. 256:184–185.
- Ikeda M, Hirayanagi K, Arai M, Kakuda S, Makioka K, Furuta N, Takai E, Kasahara H, Tsukagoshi S, Fujita Y, et al. 2012. Encephalopathy with amyloid angiopathy and numerous amyloid plaques with low levels of CSF A $\beta$ 1-40/A $\beta$ 1-42. *Amyloid*. 19:186–190.
- Kitazawa M, Medeiros R, LaFerla FM. 2012. Transgenic mouse models of Alzheimer disease: developing a better model as a tool for therapeutic interventions. *Curr Pharm Des*. 18:1131–1147.
- Langbaum JB, Fleisher AS, Chen K, Ayutyanont N, Lopera F, Quiroz YT, Caselli RJ, Tariot PN, Reiman EM. 2013. Ushering in the study and treatment of preclinical Alzheimer disease. *Nat Rev Neurol*. 9:371–381.
- Lee J-E, Han P-L. 2013. An update of animal models of Alzheimer disease with a reevaluation of plaque depositions. *Exp Neurobiol*. 22:84–95.
- Leon WC, Canneva F, Partridge V, Allard S, Ferretti MT, DeWilde A, Vercauteren F, Atifeh R, Ducatenzeiler A, Klein W, et al. 2010. A novel transgenic rat model with a full Alzheimer's-like amyloid pathology displays pre-plaque intracellular amyloid-beta-associated cognitive impairment. *J Alzheimers Dis*. 20:113–126.
- Li S, Jin M, Koeglsperger T, Shepardson NE, Shankar GM, Selkoe DJ. 2011. Soluble A $\beta$  oligomers inhibit long-term potentiation through a mechanism involving excessive activation of extrasynaptic NR2B-containing NMDA receptors. *J Neurosci*. 31:6627–6638.
- Lledo PM, Hjelmstad GO, Mukherji S, Soderling TR, Malenka RC, Nicoll RA. 1995. Calcium/calmodulin-dependent kinase II and long-term potentiation enhance synaptic transmission by the same mechanism. *Proc Natl Acad Sci USA*. 92:11175–11179.
- Masliah E, Alford M, DeTeresa R, Mallory M, Hansen L. 1996. Deficient glutamate transport is associated with neurodegeneration in Alzheimer's disease. *Ann Neurol*. 40:759–766.
- McCade D, Savage G, Naismith SL. 2011. Review of emotion recognition in mild cognitive impairment. *Dement Geriatr Cogn Disord*. 32:257–266.
- Patton RL, Kalback WM, Esh CL, Kokjohn TA, Van Vickle GD, Luehrs DC, Kuo Y-M, Lopez J, Brune D, Ferrer I, et al. 2006. Amyloid-beta peptide remnants in AN-1792-immunized Alzheimer's disease patients: a biochemical analysis. *Am J Pathol*. 169:1048–1063.
- Rajan KB, Wilson RS, Weuve J, Barnes LL, Evans DA. 2015. Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology*. 85:898–904.
- Ruiz-Opazo N, Kosik KS, Lopez LV, Bagamasbad P, Ponce LR, Herrera VL. 2004. Attenuated hippocampus-dependent learning and memory decline in transgenic TgAPPswe Fischer-344 rats. *Mol Med*. 10:36–44.

- Sala Frigerio C, De Strooper B. 2016. Alzheimer's Disease Mechanisms and Emerging Roads to Novel Therapeutics. *Annu Rev Neurosci.* 39:57–79.
- Selkoe DJ. 2002. Alzheimer's disease is a synaptic failure. *Science.* 298:789–791.
- Selkoe DJ, Hardy J. 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med.* 8:595–608.
- Talantova M, Sanz-Blasco S, Zhang X, Xia P, Akhtar MW, Okamoto S-I, Dziewczapolski G, Nakamura T, Cao G, Pratt AE, et al. 2013. A $\beta$  induces astrocytic glutamate release, extrasynaptic NMDA receptor activation, and synaptic loss. *Proc Natl Acad Sci USA.* 110:E2518–E2527.
- Wang ZC, Zhao J, Li S. 2013. Dysregulation of synaptic and extrasynaptic N-methyl-D-aspartate receptors induced by amyloid-beta. *Neurosci Bull.* 29:752–760.
- Weller RO, Boche D, Nicoll JAR. 2009. Microvasculature changes and cerebral amyloid angiopathy in Alzheimer's disease and their potential impact on therapy. *Acta Neuropathol.* 118:87–102.

## Annexe 2

Cite this: *Analyst*, 2018, **143**, 1077

## On-a-chip tryptic digestion of transthyretin: a step toward an integrated microfluidic system for the follow-up of familial transthyretin amyloidosis

Jeanne Bataille,<sup>a,b</sup> Arthur Viodé,<sup>c</sup> Iago Pereiro,<sup>id</sup><sup>d</sup> Josiane P. Lafleur,<sup>id</sup><sup>e</sup> Fanny Varenne,<sup>a,b</sup> Stéphanie Descroix,<sup>id</sup><sup>d</sup> François Becher,<sup>c</sup> Jörg P. Kutter,<sup>e</sup> Camille Roesch,<sup>f</sup> Christian Poüs,<sup>g</sup> Myriam Taverna,<sup>a,b</sup> Antoine Pallandre,<sup>a,h,i</sup> Claire Smadja,<sup>id</sup><sup>\*†a,b</sup> and Isabelle Le Potier<sup>†a,j</sup>

A microfluidic microreactor for trypsin mediated transthyretin (TTR) digestion has been developed as a step towards the elaboration of a fully integrated microdevice for the detection of a rare and disabling disease, the familial transthyretin amyloidosis (ATTR) which is related to specific TTR mutations. Therefore, an enzymatic microreactor coupled to an analytical step able to monitor the mutation of TTR on specific peptide fragments would allow an accurate monitoring of the treatment efficiency of ATTR. In this study, two types of immobilized trypsin microreactors have been investigated: a new miniaturized, microfluidic fluidized bed packed with trypsin functionalized magnetic particles (MPs), and a thiol-ene (TE) monolith-based chip. Their performances were first demonstrated with *N*-benzoyl-DL-arginine-4-nitroanilide hydrochloride BApNA, a low molecular weight substrate. High reaction yields (75.2%) have been reached within 0.6 min for the TE-based trypsin microreactor, while a lower yield (12.4%) was obtained for the micro-fluidized bed within a similar residence time. Transposition of the optimized conditions, developed with BApNA, to TTR digestion in the TE-based trypsin microreactor was successfully performed. We demonstrated that the TE-chip can achieve an efficient and reproducible digestion of TTR. This has been assessed by MS detection. In addition, TTR hydrolysis led to the production of a fragment of interest allowing the therapeutic follow-up of more than twenty possible ATTR mutations. High sequence coverage (90%), similar to those obtained with free trypsin, was achieved in a short time (2.4 min). Repeated experiments showed good reproducibility (RSD = 6.8%). These promising results open up the route for an innovative treatment follow-up dedicated to ATTR.

Received 21st October 2017,  
Accepted 26th December 2017

DOI: 10.1039/c7an01737e

rsc.li/analyst

<sup>a</sup>Faculté de Pharmacie, Université Paris-Sud, Université Paris-Saclay, Châtenay-Malabry, France. E-mail: claire.smadja@u-psud.fr; Tel: +33 146835462

<sup>b</sup>CNRS, UMR 8612, Institut Galien Paris-Sud, Université Paris-Saclay, Châtenay-Malabry, France

<sup>c</sup>Service de Pharmacologie et d'Immunoanalyse (SPI), Laboratoire d'Etude du Métabolisme des Médicaments, CEA, INRA, Université Paris Saclay, F-91191 Gif-sur-Yvette cedex, France

<sup>d</sup>CNRS, UMR 168, Institut Curie, Paris, France

<sup>e</sup>Department of Pharmacy, University of Copenhagen, Copenhagen, Denmark  
<sup>f</sup>IZON Sciences Ltd, Lyon, France

<sup>g</sup>Hôpital Antoine Bécère, AP-HP, Service de Biochimie, Clamart, France

<sup>h</sup>Faculté des Sciences, Université Paris-Sud, Université Paris-Saclay, Orsay, France

<sup>i</sup>CNRS, UMR 8000, Laboratoire de Chimie Physique, Université Paris-Sud, Université Paris-Saclay, Orsay, France

<sup>j</sup>CNRS, UMR 9001, Centre de Nanosciences et de Nanotechnologies, Université Paris-Sud, Université Paris-Saclay, Marcoussis, France

† Both authors contributed equally to this work.

## Introduction

Therapeutic follow-up of patients often requires the development of efficient sample pretreatments of biological samples before the analysis of the targeted biomolecule or biomarker. Miniaturized analysis technologies have the ability to automate the preparation of complex samples by including on-line analyte enrichment, concentration, chemical or enzymatic reactions, mixing and/or tagging steps.

In this work, we studied the sample preparation and particularly the enzymatic digestion of a protein: transthyretin (TTR). This protein is of particular interest since several of its peptide sequences are known to be biomarkers of a rare and disabling disease named familial transthyretin amyloidosis (ATTR).<sup>1–3</sup> TTR circulates at high concentration (0.2–0.4 g L<sup>-1</sup>) in the plasma where it transports some substances such as thyroxine or retinol-binding protein–vitamin A complex. TTR is produced mainly by the liver, the retinal pigment of

epithelial cells and the choroid plexus.<sup>2</sup> It is a tetrameric protein, which has four identical subunits composed by 127 amino acids (55 kDa). Transthyretin amyloidosis can be hereditary (familial) or secondary to senile systemic amyloidosis.<sup>4</sup> ATTR occurs when the mutation of transthyretin leads to the dissociation of the TTR tetramer into monomers, the first step towards amyloidosis. There are more than hundred genetic mutations,<sup>5–8</sup> which lead to pathologic variants. Different diagnostics<sup>9</sup> are proposed to detect the disease, such as neurological consultation, genetic testing that allows us to confirm the presence of mutations in the organism, and also biopsies to observe any abnormal deposits of fibrils in the organs. Nowadays, different treatments limiting this physio-pathological process are proposed.<sup>2,10</sup> One of them is the transplantation of an amyloidogenic TTR-producing liver with one that produces only wild type TTR with the aim of decreasing the circulating concentration of unstable and pathological TTR and prevent amyloid formation. Therefore, to confirm the treatment efficiency, there is a need for rapid and sensitive analysis allowing disease and treatment follow-up. Miniaturized and quantitative molecular analysis of this biomarker could provide an innovative tool to monitor treatment efficiency. In this line, enzymatic microreactors could represent a promising tool to distinguish the mutated from the non-mutated TTR. The common dimensions used in microfluidic devices lead to high surface to volume ratios while the small volumes of the reaction chambers lead to high efficiency enzymatic reactions and improved mass transfer and thereby digestion times. Numerous microfluidic enzymatic reactions have been proposed to digest proteins in miniaturized analytical devices.<sup>11–13</sup> Several of them have also demonstrated that microfluidic enzymatic digestion allows successful discrimination between native and pathological proteins.<sup>11</sup> These microsystems consist of enzyme immobilization on the surface of particles, beads, channels walls or within solid supports such as monoliths, disks, and membranes.<sup>14</sup> Several microfluidic devices rely on magnetic particles (MPs).<sup>11,15–17</sup> MPs offer a large surface area for enzyme grafting. In addition, with the help of a magnetic field they can be easily manipulated and removed from the microfluidic chip when the magnetic field is suppressed. Very recently, a new miniaturized, microfluidic fluidized bed was proposed, allowing a continuous particle recirculation. This is due to the combination of two forces, *i.e.* magnetic attraction and pressure-driven hydrodynamic forces.<sup>18</sup> This was successfully applied to the specific immunocapture and fluorescent derivatization of amyloid peptides.<sup>19</sup> It has been hypothesized that a continuous recirculation of particles bio-functionalized with enzyme would provide a very efficient interaction between enzyme and targeted proteins, leading thereby to an accurate therapeutic follow-up for TTR neuropathy.

Monolithic supports are also widely employed for enzymatic microreactor development.<sup>13,20–23</sup> Indeed, monolith-based microreactors are particularly interesting due to their large surface, enabling a high concentration of immobilized enzyme. Most monolith supports are based on methacrylate

monomers.<sup>14</sup> However, poor adhesion of the monolith in the microchannel wall is a current issue and leads often to detachment of monolith from the micro channel wall. Therefore, to overcome these drawbacks, a new support based on thiol-ene (TE) has been recently developed.<sup>24</sup> This monolith combines several advantages, *e.g.*, easy processing and adhesive properties preventing monolith detachment. Several authors have developed successful enzymatic microreactors with galactose oxidase, peptide-*N*-glycosidase F, or pepsin immobilized.<sup>25,26</sup>

In this work, two approaches have been selected to compare and optimize the on-a-chip enzymatic digestion. Both strategies (MPs and monolith) offer attractive alternatives to a traditional solution-based proteolysis reaction. Two immobilized trypsin microreactors were developed as a step towards the elaboration of a fully integrated microfluidic system enabling the protein extraction, protein digestion and peptide separation for treatment monitoring: the first one is a microfluidized bed packed with trypsin functionalized MPs,<sup>18</sup> and the second one is a TE trypsin grafted monolith-based chip. For the TE microreactor, trypsin immobilization was performed for the first time by adapting the protocol reported by Lafleur *et al.*<sup>25</sup>

In order to compare the digestion efficiency of both in solution-free trypsin and immobilized trypsin (batch-wise, microfluidized bed and TE-chip), *N*-benzoyl-*DL*-arginine-4-nitroanilide hydrochloride (BAPNA) was selected as a model substrate to perform kinetic studies. The influence of various parameters such as substrate flow rate was investigated for the on-a-chip digestion approach. Finally, the optimized procedure was applied to tryptic digestion of TTR. The performances of the selected microreactor based on the sequence coverage and the rate of hydrolysis reaction were then studied using mass spectrometry (MS). In particular, the yield of the production of a 22-aa fragment (called WP), which could be a relevant fragment for the treatment follow-up of patients suffering from more than twenty different ATTR mutations, has been evaluated by ultra-high performance liquid chromatography coupled to MS (UHPLC-MS).

## Experimental

### Chemicals and materials

Phosphate-buffered saline 10× (PBS 10×), *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC), *N*-hydroxysulfosuccinimide sodium (S-NHS), trypsin from bovine pancreas (TPCK-trypsin), *N*-benzoyl-*DL*-arginine-4-nitroanilide hydrochloride (BAPNA), *p*-nitroaniline (*p*-NA), dimethyl sulfoxide (DMSO), prealbumin from human plasma (TTR), epoxy-poly-(dimethyl acrylamid) (EpDMA), bovine serum albumin (BSA), triallyl-1,3,5-triazine-2,4,6-(1*H*,3*H*,5*H*-trione) (triallyl), pentaerythritoltetrakis (3-mercaptopropionate) (tetrathiol), 2-*tert* (butoxycarbonylamino)ethanethiol (*t*-BOC protected amino(ethanethiol)), methanol (MeOH), benzamidine, *t*-octyl-phenoxy-polyethoxy-ethanol (Triton X-100), pluronic F127 (PF 127), and sodium deoxycholate (DOC) were

obtained from Sigma-Aldrich (Munich, Germany). Lucirin TPO-L was purchased from BASF (Ludwigshafen, Germany). Hydrochloric acid 37% (HCl) was from VWR (Fontenay-sous-Bois, France). The synthetic peptide WP (TSESGELHGLTTEEFVEGIYK, WP, GPS1344) was synthesized by Genepep (Prades-le-Lez, France). Ascorbic acid (ASA), formic acid, acetonitrile (ACN) and ammonium bicarbonate were purchased from Fluka (Buchs, Switzerland). A micro bicinchoninic acid protein assay kit (BCA) was obtained from Thermo Fisher (Waltham, MA, USA). A Sylgard 184-poly(dimethylsiloxane) (PDMS) elastomer kit was obtained from Dow Corning (Midland, MI, USA). Magnetic beads surface-bound with carboxylic acid groups (Dynabeads® M-270 carboxylic acid, diameter 2.8  $\mu\text{m}$ , 30  $\text{mg mL}^{-1}$ ) were purchased from Life Technologies (Carlsbad, CA, USA).

A UV spectrophotometer (Biotek EL 800) and the Gen5 software were used for measuring the absorbance of samples and for data processing.

A qNano instrument was used to investigate the surface charge of the trypsin-functionalized particles. Measurements were performed using Tunable Resistive Pulse Sensing technology (TRPS, qNano, Izon Science, Christchurch, New Zealand) and acquired using the IZON Control Suite 3.2.2.268 software. The same software was then used to process the recorded data.

For the digestions performed using trypsin-immobilized magnetic particles in solution (batch-wise experiments), the functionalized beads were separated from the supernatant by a magnetic separator (Adem-Mag MSV, Ademtech, France). For the digestions performed on trypsin-immobilized magnetic particles within a micro-fluidized bed, fluid flow was pressure-regulated using a pressure controller (MFCS™, Fluigent) and for digestions realized on trypsin-immobilized monoliths, flow rates were controlled using a syringe pump (Model NE-1000 Multi-Phaser, New Era Pump Systems Inc., USA).

UHPLC-MS was performed on a Dionex Ultimate 3000 ultra-high performance liquid chromatography system coupled to a qExactive mass spectrometer equipped with an electrospray ionization (ESI) source operated in the positive ion mode (Thermo Fisher Scientific, Bremen, Germany). An Aeris peptide XB-C18 column (1.7  $\mu\text{m}$  particles size, 150  $\times$  2.1 mm, 100 Å porosity) from Phenomenex (Le Pecq, France) was used in peptide separation. X-calibur software (version 1.4SR1) was used for the data processing, identification and quantification of the peptides.

## Methods

### Trypsin immobilization on carboxylic-functionalized MPs

For optimal conditions (ratio of trypsin/MP equal to 15/1), the grafting was performed as follows: 10  $\mu\text{L}$  of MPs stock solution (30  $\text{mg mL}^{-1}$ , 300  $\mu\text{g}$ ) were washed 3 times with 1 $\times$  PBS solution. The method for covalent coupling of trypsin to COOH-functionalized carriers by EDC and S-NHS was adapted from Staros *et al.*<sup>27</sup> and Voyksner *et al.*<sup>28</sup> 450  $\mu\text{L}$  of a trypsin solution (10  $\text{mg mL}^{-1}$  in 1 $\times$  PBS), 200  $\mu\text{L}$  of an EDC solution (10

$\text{mg mL}^{-1}$  in 1 $\times$  PBS) and 200  $\mu\text{L}$  of an S-NHS solution (10  $\text{mg mL}^{-1}$  in 1 $\times$  PBS, 200  $\mu\text{L}$ ) were added to the particle suspension. Trypsin immobilization was achieved by gentle stirring of the suspension for 3 hours at room temperature. The grafting reaction was then stopped and the unbound enzymes and activator reagents were eliminated by washing the trypsin-immobilized MPs 3 times with 1 $\times$  PBS solution. These particles were finally stored at 4 °C in a 1 $\times$  PBS solution.

### Characterization of trypsin-functionalized MPs

After each grafting experiment, the determination of the total protein content was performed using a Micro BCA Protein Assay kit. In order to monitor the quality of the trypsin-functionalized MPs, the zeta potential ( $\zeta$ -potential) was measured using the qNano instrument.<sup>29</sup> Both calibration particles and unknown MPs were diluted to appropriate levels in 1  $\times$  PBS, used as an electrolyte. Measurements were performed in triplicate before and after each step of surface modification. Current pulse signals were acquired over a few minutes.

### Assay of trypsin activity – determination of the kinetic parameters

The hydrolysis efficiency of immobilized trypsin was evaluated and compared to the free trypsin hydrolysis protocol. For these assays performed at 25 °C, BApNA was selected as a model substrate to mimic the enzymatic reaction. In principle, BApNA was cleaved by trypsin to produce *p*-nitroaniline (*p*-NA), a yellow compound that can be detected with a microplate reader UV spectrophotometer at 405 nm. The absorbance of solutions and/or supernatants was measured at 0, 30 and 60 minutes, and the *p*-NA concentration was then calculated from a calibration curve.

The trypsin/BApNA ratio (w/w) has been optimized by varying the enzyme concentration from 4  $\text{mg mL}^{-1}$  to 0.016  $\text{mg mL}^{-1}$  with a fixed BApNA concentration (4  $\text{mg mL}^{-1}$ ) (ratio ranging from 1/1 to 1/250). The ratio 1/17 was selected since it was in the linearity range. For free trypsin experiments 6.25  $\mu\text{L}$  of BApNA stock solution (32  $\text{mg mL}^{-1}$  in DMSO) were added to 2  $\mu\text{L}$  (11.7  $\mu\text{g}$ ) of trypsin stock solution (5.85  $\text{mg mL}^{-1}$  in 100 mM pH 7.9 ammonium bicarbonate buffer (AB)) and filled in AB buffer to obtain the final volume of 50  $\mu\text{L}$ .

For trypsin-functionalized MPs in batch-wise mode, 6.25  $\mu\text{L}$  of a BApNA stock solution (32  $\text{mg mL}^{-1}$  in DMSO) were added with 300  $\mu\text{g}$  of trypsin-grafted MPs and filled into the digestion buffer to obtain a final volume of 50  $\mu\text{L}$ . Different digestion buffers were evaluated: AB, AB/ACN 50/50 v/v, AB with added Triton X-100 (0.001%), AB with added PF 127 (0.001%), AB/ACN 50/50 v/v with either PF 127 (0.01%) or DOC (0.1%). For AB buffer, the value of the Michaelis–Menten constant  $K_M$  was determined from the Lineweaver–Burk plot obtained using different substrate concentrations (from 10 to 2000  $\mu\text{g mL}^{-1}$ ).

For all the experiments performed on a micro-fluidized bed, the chip was first filled with the digestion buffer (AB/ACN 50/50 v/v buffer) prior to the introduction, by means of a pipette and in the presence of a magnet placed below the chip, of 300  $\mu\text{g}$  of trypsin-immobilized MPs re-suspended in the



digestion buffer. Once the MPs were loaded into the chamber, the whole system was placed in an ultrasound bath (30 s, 45 kHz, 80 W) to disaggregate particles. After that, a permanent magnet (ChenYang Technologies N50) was aligned with the chamber to form the fluidized bed (see Fig. 1). The BApNA solution (4 mg mL<sup>-1</sup> in AB/ACN 50/50 v/v) was passed through the fluidized MPs with a pressure-controller at a flow rate of 1 μL min<sup>-1</sup>.

For trypsin-monolith-based microchip experiments, the BApNA solution (4 mg mL<sup>-1</sup> in AB buffer) was passed through the microreactor with a syringe pump at 0.7 μL min<sup>-1</sup>.

For these digestion modes, the digestion yield was calculated using eqn (1):

$$\text{digestion yield (\%)} = \frac{QpNA_x}{QBAPNA_0} \times 100 \quad (1)$$

where  $QpNA_x$  corresponds to the *p*-NA mole number estimated by measuring the absorbance of supernatants or solutions, and  $x$  varies from 0 to 60 min and  $QBAPNA_0$  represents the initial mole number of BApNA. The initial velocity of the reaction was deduced from the curve representing the *p*-NA concentration as a function of the digestion time.

### Fabrication of the enzymatic microreactors

**Micro-fluidized bed.** The micro-fluidized bed employed in this work was adapted from a design reported recently.<sup>18</sup> The chip contains a diamond-shaped pattern (13 mm in length, 2.9 mm in maximum width and 50 μm high) with an inlet and an outlet at opposite ends. The chips were manufactured from PDMS (Sylgard® 184) mixed with its linker (ratio 10 : 1 w : w).

**TE monolithic trypsin microreactor.** The process used to fabricate the TE microfluidic chips has recently been described in detail.<sup>25</sup> For the monolith preparation, an emulsion of an off-

stoichiometric (OSTE) mixture (40% ene-excess; OSTE-allyl) in methanol (80% w/w) was magnetically stirred in a sealed container for 5 min. 7.5 μL of photoinitiator (10% v/v Lucirin TPO-L in ethanol) was added to the emulsion and stirred for 1 min. 50 μL of the emulsion were directly injected into the TE microfluidic chip by the inlet with the use of a pipette. The TE chip was immediately cured under the collimated UV-lamp (60 s, 20.5 mW cm<sup>-2</sup> at 365 nm, LS 100 3C2, 350 W near-UV, Bachur & Associates, Santa Clara, CA, USA). The excess of monomers and methanol was removed by rinsing immediately in ultrapure water using a syringe pump (5 min, 30 μL min<sup>-1</sup>). Fluidic access to the chip was achieved through a 3D-printed interface. The interface was modelled following a method recently described.<sup>26</sup> The inlet and the outlet of the chip were sealed with adhesive tape for its storage at 4 °C to prevent drying of the monolith.

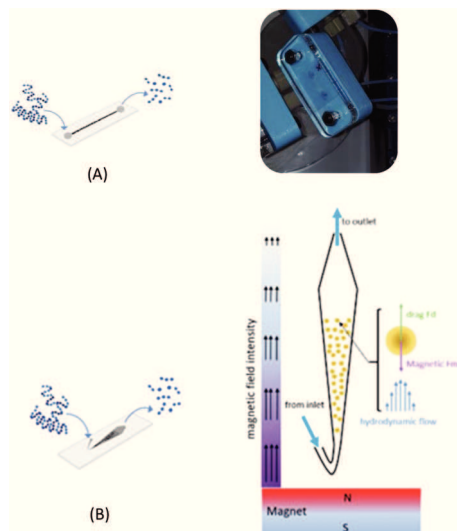
Trypsin immobilization on OSTE-allyl monoliths was achieved using the TE click-chemistry, as previously described by Lafleur *et al.*<sup>25</sup> The same method was adapted here for trypsin immobilization. To introduce an amino functionality at the surface of the OSTE-allyl monolith, the chip was filled with a *t*-BOC protected amino(ethanethiol) solution containing 5% v/v of lucirin TPO-L using a syringe pump at a flow rate of 10 μL min<sup>-1</sup>. The chip was then exposed to collimated UV light (30 s, 20.5 mW cm<sup>-2</sup> at 365 nm). Excess compounds were washed out with ultrapure water (5 min, 10 μL min<sup>-1</sup>). Deprotection of amine groups was performed with 4 M HCl overnight (12 hours at 100 μL h<sup>-1</sup>). After washing with water (5 min, 30 μL min<sup>-1</sup>), the monolith was conditioned with L-ascorbic acid (ASA) (~7% ASA in 66% methanol) for 5 min at a flow rate of 10 μL min<sup>-1</sup>. After incubation for 30 min at room temperature, allowing the linking of ASA as a cross-linker to the NH<sub>2</sub>-functionalized monolith, unreacted ASA was removed by rinsing immediately with ultrapure water (5 min, 30 μL min<sup>-1</sup>). In order to immobilize the trypsin *via* covalent bonds to the NH<sub>2</sub> groups, the channel was then filled with a trypsin solution (10 mg mL<sup>-1</sup> in ultrapure water, 20 μL min<sup>-1</sup>), sealed with adhesive tape, and incubated for 24 hours at 4 °C. After 24 hours incubation time, unreacted trypsin was removed by rinsing with ultrapure water (2 min, 20 μL min<sup>-1</sup>).

### Residence time determination

The influence of residence time was evaluated. For batch experiments, it is equal to the incubation time. For on-a-chip experiments, the average residence time,  $T_{\text{res}}$ , was computed using eqn (2).

$$T_{\text{res}} = \frac{V_c - V_p}{Q} \quad (2)$$

where  $Q$  represents the flow rate imposed either by a syringe pump (TE microfluidic chip experiments) or by a pressure-controller (micro-fluidized bed experiments),  $V_c$  corresponds to the chamber volume and  $V_p$  is the volume of particles introduced [ $V_p = V_1 \text{ particle} \times \text{particles number}$ ]. For the micro-fluidized bed, the volume of the chamber and the volume of particles were equal to 0.85 μL and 0.15 μL, respectively. For the



**Fig. 1** Schematic representation of enzymatic digestion performed on (A) a TE microfluidic chip and an interface with two-point clamping, (B) a micro-fluidized bed and principle (recirculation particles thanks to two forces). For microreactor dimensions, see the text.

TE microfluidic chip, the volume of the chamber was equal to 0.40  $\mu\text{L}$  and the volume inside the TE monoliths was calculated according to the procedure reported by Lafleur *et al.*<sup>25</sup> and was equal to  $8.1 \times 10^{-8}$   $\mu\text{L}$ .

### Enzymatic digestion of transthyretin

**TTR digestion in solution.** 40  $\mu\text{L}$  of a TTR solution ( $0.8 \text{ g L}^{-1}$  in 100 mM ammonium bicarbonate buffer pH 7.9 (AB), heated at 95 °C for 15 min) and 2  $\mu\text{L}$  of the trypsin solution ( $7.8 \text{ mg mL}^{-1}$  in AB buffer) were mixed to obtain the final volume of 42  $\mu\text{L}$ . Hydrolysis was performed at 25 °C for 4 hours. The TTR digest was 100-fold diluted in an aqueous–organic solution containing ACN (5.6%) and formic acid (0.1%) prior to UHPLC-MS analyses.

**TTR digestion using trypsin-immobilized magnetic particles.** For trypsin-functionalized MPs in batch-wise experiments, 400  $\mu\text{g}$  of trypsin-immobilized particles (13.4  $\mu\text{L}$  of MPs stock solution at  $30 \text{ mg mL}^{-1}$ ) were prepared. 40  $\mu\text{L}$  of a TTR solution ( $0.8 \text{ g L}^{-1}$  in either AB or AB/ACN 50/50 v/v buffer, heated at 95 °C for 15 min) were added to 400  $\mu\text{g}$  of a trypsin-functionalized MPs suspension. A final volume of 42  $\mu\text{L}$  was obtained by adding the digestion buffer (AB or AB/ACN 50/50 v/v buffer). TTR digestion was performed at 25 °C for 4 hours without stirring. After digestion, the trypsin-immobilized magnetic particles were easily separated from the TTR digest by magnetic rack separation. According to the used digestion buffer, the supernatant was 10-fold diluted in either an aqueous–organic solution containing ACN (5.6%) and formic acid (0.1%) or an aqueous–organic solution containing formic acid (0.1%) prior to UHPLC-MS analyses.

**TTR digestion in a TE monolithic trypsin microreactor.** The TTR solution ( $0.8 \text{ g L}^{-1}$  in 100 mM ammonium bicarbonate buffer (AB), pH 7.9, heated at 95 °C for 15 min) was injected into the enzymatic microreactor with a syringe pump at several flow rates: 10, 50 and 100  $\mu\text{L h}^{-1}$  corresponding to a residence time of 2.4 minutes, 28 and 14 seconds, respectively. 20  $\mu\text{L}$  of the outlet collected volume were 10-fold diluted in an aqueous–organic solution containing ACN (5.6%) and formic acid (0.1%) for UHPLC-MS analyses.

### UHPLC-ESI-qExactive-MS method for the identification and quantification of tryptic fragments

Peptides separation was achieved in 10 min on the Aeris peptide XB-C18 capillary column at a flow rate of 500  $\mu\text{L min}^{-1}$ , using the LC gradient shown in Table 1 and a column temperature of 50 °C. The sampler temperature was set to 5 °C. Regarding MS detection, data were acquired in the positive ion mode with an ion spray voltage of 4.00 kV (HSEI-II, Probe, Thermo Fisher Scientific). The source temperature was set at 40 °C, the capillary temperature was set at 320 °C and sheath and auxiliary gases (nitrogen) were set at 70 and 10 au, respectively. S-lens RF was set at 55. The AGC target was set at  $1 \times 10^6$  and the Orbitrap resolution was set at 35 000 at 200  $m/z$  for all experiments. The full scan mode was used for the identification of the TTR peptides. For the quantification of WP, the qExactive mass spectrometer was operated in the

**Table 1** The LC gradient profile for UHPLC-ESI-qExactive-MS. Solvent A consisted of water with 0.1% (v/v) formic acid and solvent B consisted of acetonitrile with 0.1% (v/v) formic acid

Step	Total time (min)	Flow rate ( $\mu\text{L min}^{-1}$ )	% A	% B
0	0.0	500	95	5
1	1.0	500	95	5
2	8.0	500	45	55
3	8.1	500	5	95
4	10.0	500	5	95
5	10.1	500	95	5
6	14.0	500	95	5

parallel reaction monitoring (PRM) mode with the following parameters: 2  $m/z$  isolation width, 120 ms maximum injection time, and 17% normalized collision energy.

In order to compare the different digestion protocols, the digestion rate ( $Y$ ) was calculated by monitoring the amount of WP produced by TTR hydrolysis according to eqn (3):

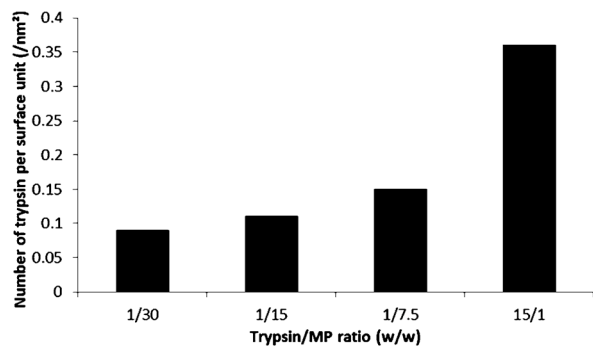
$$Y = \frac{Q_{\text{WP}}}{Q_{\text{TTR}}} \times 100\% \quad (3)$$

where  $Q_{\text{WP}}$  is the mole number of WP in the sample digest and  $Q_{\text{TTR}}$  is the initial mole number of TTR monomer subjected to enzymatic hydrolysis. To determine  $Q_{\text{WP}}$ , a concentration series of WP, from 25 to 5000  $\text{ng mL}^{-1}$ , was prepared in 100 mM ammonium bicarbonate with 5% v/v of acetonitrile and analyzed under the same conditions as the digest samples. The equation of the calibration curve was calculated by plotting the WP peak area *versus* WP peptide concentration ( $\text{ng mL}^{-1}$ ) and a satisfactory correlation coefficient has been obtained ( $R^2 = 0.9979$ ).

## Results and discussion

### Evaluation of trypsin-functionalized magnetic particle performance

Immobilization of trypsin on the MPs was performed using EDC and S-NHS as activation reagents for amide coupling between carboxylic groups onto the surface of MPs and amino groups of trypsin.<sup>27,28</sup> With the aim of grafting the optimal amount of trypsin, and thus maximizing the digestion efficiency, the reaction was carried out with different trypsin/bead ratios: 1/30, 1/15, 1/7.5, and 15/1 (w/w). The estimation of the enzyme content onto the MPs surface after the covalent coupling was performed by a protein assay using a Micro BCA kit. As shown in Fig. 2, an increase of the amount of grafted enzyme with the increase of the trypsin/MP ratio was observed. It was concluded that performing the coupling reaction with an excess of enzyme enhances their grafting and thus the number of potential active sites available for substrate biocatalysis. As a consequence, a higher BAPNA digestion efficiency has been observed for a trypsin/MP ratio of 15/1 compared to those obtained at ratios 4/1 and 1/30 (6- and 10-fold decrease, respectively).



**Fig. 2** Characterization of trypsin-functionalized MPs. The amount of trypsin per surface unit was estimated by using a Micro BCA Protein Assay kit (see Experimental section).

The trypsin/MP ratio of 15/1 was therefore selected for the following experiments. In order to characterize the trypsin-functionalized MPs, the zeta potential of particles was measured by tunable resistive pulse sensing (TRPS) technology (before grafting =  $-163 \pm 11$  mV; after grafting =  $-61 \pm 23$  mV). An increase of  $\zeta$ -potential was observed after the coupling reaction. Trypsin ( $pI \sim 10.1$ ), which carries a net positive charge in the employed phosphate buffer (pH 7.4), compensates for the negative charge of carboxylate groups leading to an increase of the surface charge of the functionalized particles.

The enzyme activity of trypsin-functionalized MPs in batch-wise experiments was evaluated at 25 °C in 100 mM ammonium bicarbonate pH 7.9 (AB) by investigating the hydrolysis of BAPNA as the substrate.<sup>31–34</sup> In principle, BAPNA is cleaved in the presence of trypsin to produce *p*-nitroaniline (*p*-NA), which exhibits a strong absorbance at 405 nm. Thus, the activity of the immobilized enzyme has been measured by a photometric assay. The estimated Michaelis constant  $K_M$  was equal to 0.4 mM. In order to optimize the digestion efficiency, different buffer additives (agent against trypsin autolysis, organic solvent and surfactants, listed in Table 2) were introduced either in the coupling buffer (PBS 1×) or in the AB digestion buffer. The hydrolysis reaction yields at 30 and 60 min, as well as the initial velocity of the reaction, have been calculated under these various conditions. The results are summarized in Table 2.

The addition of benzamidine, a competitive inhibitor of trypsin frequently used to prevent autocatalytic activity of trypsin during the immobilisation process,<sup>20,21,33,35–37</sup> was tested. No significant improvement in the reaction yield was observed irrespective of the blocking reagent concentration used (varying from 0.5 to 15-fold trypsin concentration). The yield and the initial velocity of the reaction were greatly enhanced (7-fold for both parameters) when ACN was added to the incubation buffer. These results are in good agreement with previous works<sup>30,36,38,39</sup> reporting the usefulness of aqueous–organic solvents for efficient protein digestion.

With the aim of minimizing the self-aggregation of MPs, which is detrimental to the enzyme availability and also to the homogeneity of the substrate for further microchip appli-

**Table 2** Summary of BAPNA hydrolysis parameters (buffer, incubation time, reaction yield, initial velocity) with immobilized trypsin in batch-wise experiments

Coupling buffer/ incubation medium	Incubation time (min)	Yield (%)	Initial velocity ( $\mu\text{g mL}^{-1} \text{min}^{-1}$ )
Ammonium bicarbonate (AB)	30	2.1	4.9
AB + benzamidine(*)	60	4.5	
	30	1.7	4.1
	60	3.9	
AB + ACN 50/50 v/v	30	14.8	34.3
	60	31.4	
AB + Triton X-100	30	3.6	9.0
	60	8.4	
AB + PF127	30	5.0	10.4
	60	9.2	
AB + ACN + PF127	30	20.2	94.1
	60	100	
AB + ACN + DOC	30	12.0	23.0
	60	20.5	

(\*) Benzamidine was added into the coupling buffer at 15-fold trypsin concentration. For this reaction condition, the BAPNA digestion was performed in 100 mM ammonium bicarbonate buffer pH 7.9.

cations, three surfactants, Triton X-100, PF-127 and DOC (at concentrations ranging from 0.001% to 1% v/v), were also investigated. A 2-fold increase of both the reaction yield and initial velocity values was observed when either Triton X-100 or PF-127 was added to the incubation buffer. The digestion efficiency was slightly higher with PF-127 at a concentration of 0.01% than those obtained for other concentrations. As shown in Table 2, by combining the additions of PF-127 (0.01%) and acetonitrile 50/50 v/v, we achieved a digestion efficiency superior to that obtained with each individual condition (100% of BAPNA digestion after 60 min). Like many surfactants, PF-127 is known to suppress peptide ionization<sup>40,41</sup> in MS and could not be used in our case. The results obtained with DOC (0.1%, the best conditions) were not as good as expected since both the reaction yield and the initial velocity of the reaction were less efficient than those obtained with ammonium bicarbonate buffer/acetonitrile 50/50 v/v. The overall results confirm that by limiting the aggregation of trypsin-functionalized MPs and modifying the type of interaction involved in the enzyme–substrate-binding event, through the addition of both the surfactant and the organic solvent, the digestion efficiency can be successfully tuned to reach 100% after only 1 h of digestion. These results are in agreement with previous studies.<sup>42–45</sup>

### Performance of microreactors

**Application of the microreactors to BAPNA hydrolysis.** The performances of the microreactors were then evaluated by comparing the digestion efficiency obtained for the hydrolysis of BAPNA. For the strategy based on trypsin-functionalized MPs, the batch-wise mode was downscaled into a micro-fluidized bed. MPs are expected to continuously re-circulate inside the bed due to the opposing effects of two driving forces, *i.e.* magnetic attraction and pressure driven hydrodynamic forces

**Table 3** Summary of BAPNA hydrolysis parameters (residence time and reaction yield) in a micro-fluidized bed, a TE-monolithic trypsin micro-reactor and free trypsin in solution

Digestion mode	Residence time (min)	Yield (%)
Micro-fluidized bed (*)	0.7	12.4
TE-chip (**)	0.6	75.2
Free trypsin (**)	30	92.6
	60	100

BAPNA digestion was performed with ammonium bicarbonate buffer/ACN 50/50 v/v (\*) and ammonium bicarbonate buffer (\*\*). The digestion conditions were described in the Experimental section.

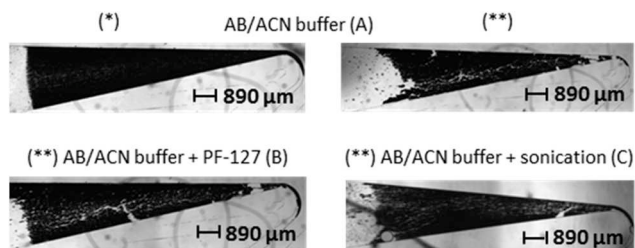
(see Fig. 1). The contact between trypsin-functionalized MPs and the substrate is therefore expected to be improved with this configuration compared to a classical design where particles are maintained stationary inside a microchannel.<sup>18,46–48</sup> The BAPNA sample was prepared in ammonium bicarbonate buffer pH 7.9/ACN 50/50 v/v and then percolated through the bed, initially pre-conditioned with the same buffer, at a flow rate of 1.0  $\mu\text{L min}^{-1}$ , which afforded a residence time of 0.7 min. As shown in Table 3, the hydrolysis rate was similar to that obtained using batch-wise mode (see Table 2). Indeed, we observed a yield of 12.4% (residence time 0.7 min) for the microreactor and 14.8% (incubation time 30 min) for the batch-wise experiments. This indicates that, as expected, the microreactor configuration drastically reduced the substrate residence time (by 40-fold). However, we hypothesized that detrimental aggregation of trypsin-immobilized MPs occurred, limiting the access of the substrate to the active sites of the enzyme and thereby the hydrolysis efficiency. This hypothesis was confirmed by examining, by video-microscopy, the bed which was packed either with trypsin-functionalized MPs or with non-functionalized MPs and percolated at flow rates ranging from 100 to 3000  $\text{nL min}^{-1}$  with AB/ACN 50/50 v/v buffer supplemented or not with PF-127 or DOC. As illustrated in Fig. 3, a homogeneous distribution of the non-functionalized MPs was observed for all conditions. In contrast, trypsin-functionalized MPs were not homogeneously distributed and fractures in the magnetic plug occurred in the aqueous-organic medium independent of the flow rate. No improvement was obtained by adding any surfactant, while the particle

homogeneity was slightly enhanced by sonicating the chamber for 30 s (Fig. 3) and no increase of the BAPNA hydrolysis yield was obtained.

The performances of the TE-based trypsin microreactor were evaluated in a similar way by investigating the BAPNA hydrolysis. The BAPNA sample was percolated through the TE-microchip, initially pre-conditioned with ammonium bicarbonate buffer pH 7.9, at 40  $\mu\text{L h}^{-1}$  corresponding to a residence time of 0.6 min. To avoid potential degradation of the TE-monolith, the organic solvent was omitted in our case. By comparing the results obtained with the micro-fluidized bed with those obtained with the TE-chip, the reaction yield was seen to be greatly enhanced (6-fold) in the latter case (Table 3). The TE-chip reaction yield was almost as good as that of a free-trypsin solution despite the fact that for free trypsin the digestion time was 30 min whereas in the TE-chip the residence time was only 0.6 min. The result obtained with the TE-chip experiment is promising for use in performing high throughput digestions. For this reason, the TE-chip has been selected for TTR digestion studies. Therefore, in this work, we focused on the performance evaluation of this enzymatic microreactor.

**TTR digestion on TE-chip.** The optimized protocol was then applied to TTR digestion. The tryptic digestion of TTR produces 10 theoretical fragments whose amino acid sequences are presented in Table 4. Among them, the fragment 49–70 (WP) is of particular interest for clinical studies. Thus, as previously reported<sup>49</sup> for the T49A ATTR, which is linked to the production of the T49A pathological variant (threonine at position 49 is replaced by alanine), the simultaneous quantitative assay of WP and of its mutated form T49A could be of great value for monitoring the evolution of T49A mutant concentration during therapy protocols.

To demonstrate the performance of immobilized trypsin on a monolithic support, a TTR solution (0.8  $\text{g L}^{-1}$  in 100 mM pH 7.9 AB buffer, heated for 15 min at 95 °C) was percolated through the microreactor at different flow rates, resulting in residence times ranging from 0.2 to 2.4 min. The effluent of the micro-reactor was collected and analyzed by LC-MS. The results obtained showed that whatever the residence time, all expected fragments (Table 4) could be identified, with the exception of fragment 1–9. This results in a sequence coverage of 90%. Repeated experiments ( $n = 3$ ) demonstrated very good reproducibility since the sequence coverage remained constant. This result of sequence coverage is comparable to stan-



**Fig. 3** Images of the micro-fluidized bed packed either with non-functionalized MPs (\*) or with trypsin-functionalized MPs (\*\*) and percolated at 1  $\mu\text{L min}^{-1}$  with an ammonium bicarbonate buffer/ACN 50/50 v/v mixture (A) supplemented with PF-127 (0.1%) (B) and after sonication (C).

**Table 4** TTR expected peptides after trypsin digestion

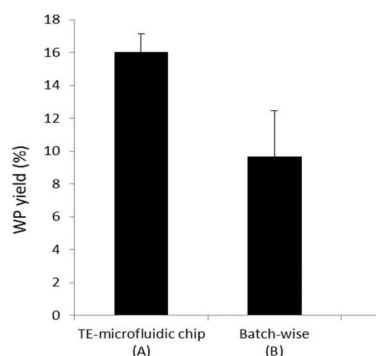
1–9	10–15
GPTGTGESK	CPLMVK
16–21	22–34
VLDVVR	GSPAINVAHVFR
36–48	49–70 (WP)
AADDTWEPFASGK	TSESGELHGLTTEEFVEGIYK
71–76	77–80
VEIDTK	SYWK
81–123	105–126
ALGISPFHEHAEEVFTANDSGPR	YTIAALLSPYSYSTTAVVTNPK

standard digestion with free trypsin or with trypsin-functionalized MPs in batch-wise but the latter two require 4 hours of digestion time to obtain a similar sequence coverage.

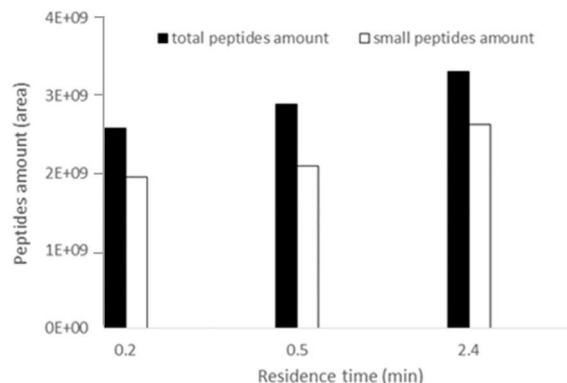
The digestion rate was calculated by monitoring by LC-MS the amount of WP produced by TTR hydrolysis. The best digestion efficiency was achieved with the TE-microfluidic chip for a residence time of 2.4 min. Under these conditions, the WP yield was around 15% and the reproducibility was good (RSD = 6.8%,  $n = 3$ ). In contrast, the experiments performed with a lower residence time were not reproducible (RSD > 15%), the WP yield being equal to 9% and 12% for a residence time of 0.5 and 0.2 min, respectively. Although the hydrolysis rate was lower for the trypsin-monolith based chip than those obtained for free trypsin, it is important to keep in mind that the residence time was 100-fold lower in microchip experiments. This could further be explained by an altered trypsin protein conformation or steric hindrance of the catalytic site when the enzyme is immobilized on the TE-microfluidic chip. This was assessed by the poor digestion efficiency observed in batch-wise experiments (*i.e.* MPs strategy, see Fig. 4).

We have also evaluated the missed cleavages corresponding to the WP fragment of interest. Only the 28-aa fragment 49–76 (MC) was detected by MS under all the experimental conditions tested. The MC/WP peak area ratio was then calculated for all digestion protocols. The results obtained showed that this ratio was below 0.15 when the digestion is performed by free trypsin and in TE-chip with a residence time of 2.4 min. The ratio slightly increased with lower residence times in TE-chip, while it was 10-fold higher after TTR with MPs strategy. This result confirms the superiority of TE-chip in monitoring the WP level.

Finally, we evaluated the impact of the flow rate on TTR digestion in TE-chips by comparing the amount of peptides identified, as well as the amount of the smallest peptides (*i.e.*, fragments 10–15, 16–21, 71–76, and 77–80) *vs.* residence time. The identified peptides correspond to the peptides presented in Table 4. In the measured range (10 to 100  $\mu\text{L h}^{-1}$ ), we can conclude that the largest amount of peptides was obtained at a



**Fig. 4** WP yield (%) obtained for TTR digestion performed with a TE monolithic trypsin microreactor (A) and trypsin-functionalized MPs in batch-wise mode (B). The residence time was equal to 2.4 min (A) and 240 min (B), respectively. The WP yield obtained in solution for free trypsin (residence time 240 min) was equal to 46.4% (data not shown).



**Fig. 5** The number of peptides *vs.* residence time in trypsin bio-functionalized TE-chip.

flow-rate of 10  $\mu\text{L h}^{-1}$  corresponding to a residence time of 2.4 min (Fig. 5). These results showed that TE-chip bio-functionalized by trypsin represents a promising tool to monitor WP and therefore opens the door to a rapid and accurate treatment follow-up of TTR neuropathy.

## Conclusions

In this work, different approaches have been proposed to perform the on-a-chip enzymatic digestion of transthyretin. Two immobilized trypsin microreactors, *i.e.* a micro-fluidized bed packed with trypsin functionalized magnetic particles and a thiol-ene monolith-based chip, were developed and compared using a model substrate, BAPNA. BAPNA has been successfully digested by using a micro-fluidized bed. However, BAPNA digestion performed following MPs strategy was less satisfactory compared to TE-chip. Therefore, transposition of the optimized conditions, developed with BAPNA, for TTR digestion in TE-based trypsin microreactors has been successfully performed. We have demonstrated that TE-chips can achieve an efficient and reproducible digestion of TTR. A high sequence coverage (90%), similar to those obtained with free trypsin, was achieved in a short time (2.4 min). Repeated experiments showed high reproducibility and led, with a satisfactory yield, to the production of a fragment of interest for the therapeutic follow-up of more than twenty different ATTR mutations, including T49A ATTR which is one of the most frequent in France. We believe that this work is a significant step towards the elaboration of a fully integrated microfluidic system enabling us to monitor the evolution of T49A mutant concentration during therapy protocols. Further work will focus on integrating consecutive units, allowing the upstream TTR extraction of the plasma and the downstream electrokinetic separation of tryptic digests.

## Conflicts of interest

The authors have declared no conflict of interest.

## Author contributions

JB: all the experiments related to enzymatic digestions in miniaturized devices. AV and FB: mass-spectrometry experiments; IP and SD: experiments on a micro-fluidized bed; FV and CR: magnetic particle characterization; CP brings his experience in TTR; JL and JK: TE-chip microfabrication; MT: TTR peptides mapping; AP: enzymatic digestions on micro-devices; CS and ILP: TTR digestion by trypsin.

## Acknowledgements

The author J. B. gratefully acknowledges funding from a fellowship programme of DIM Analytics and this work was supported in part by DIGIDIAG projects supported by the French National Research Agency (ANR) as part of the Investissements d'Avenir programme (reference number: ANR 10-NANO 0207).

## Notes and references

- F. L. Ruberg and J. L. Berk, *Circulation*, 2012, **126**(10), 1286–1300.
- Y. Ando, T. Coelho, J. L. Berk, *et al.*, *Orphanet J. Rare Dis.*, 2013, **8**, 31.
- M. A. Gertz, M. D. Benson, P. J. Dyck, *et al.*, *J. Am. Coll. Cardiol.*, 2015, **66**(21), 2451–2466.
- P. Westermark, K. Sletten, B. Johansson and G. G. Cornwell, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**(7), 2843–2845.
- V. Planté-Bordeneuve and G. Said, *Lancet Neurol.*, 2011, **10**(12), 1086–1097.
- C. Rapezzi, C. C. Quarta and L. Obici, *Eur. Heart J.*, 2013, **34**(7), 520–528.
- M. D. Benson and J. C. Kincaid, *Muscle Nerve*, 2007, **36**(4), 411–423.
- L. H. Connors, A. Lim, T. Prokaeva, *et al.*, *Amyloid*, 2003, **10**, 160–184.
- D. Mereles, M. Aurich, S. Greiner, *et al.*, *Expert Opin. Orphan Drugs*, 2016, **4**(6), 639–648.
- A. Baranczak and J. W. Kelly, *Curr. Opin. Chem. Biol.*, 2016, **32**, 10–21.
- A. Le Nel, N. Minc, C. Smadja, *et al.*, *Lab Chip*, 2008, **8**(2), 294–301.
- V. N. Luk, L. K. Fiddes, V. M. Luk, *et al.*, *Proteomics*, 2012, **12**(9), 1310–1318.
- S. Wu, L. Sun, J. Ma, *et al.*, *Talanta*, 2011, **83**(5), 1748–1753.
- M. Safdar, J. Spross and J. Jänis, *J. Chromatogr., A*, 2014, **1324**, 1–10.
- M. Slovakova, N. Minc, Z. Bilkova, *et al.*, *Lab Chip*, 2005, **5**(9), 935–942.
- W. P. Jiang, Z. Wang, L. X. Xu, *et al.*, *Biosci. Rep.*, 2015, **35**(3), 1–9.
- J. Shi, W. Zhao, Y. Chen, *et al.*, *Electrophoresis*, 2014, **33**(14), 2145–2151.
- I. Pereiro, S. Tabnaoui, M. Fermigier, *et al.*, *Lab Chip*, 2017, **17**(9), 1603–1615.
- T. D. Mai, I. Pereiro, M. Hiraoui, *et al.*, *Analyst*, 2015, **140**(17), 5891–5900.
- J. Krenkova and F. Svec, *J. Sep. Sci.*, 2009, **32**(5–6), 706–718.
- J. Krenkova, N. A. Lacher and F. Svec, *Anal. Chem.*, 2009, **81**(5), 2004–2012.
- B. Wouters, I. Dapic, T. S. E. Valkenburg, *et al.*, *J. Chromatogr., A*, 2017, **1491**, 36–42.
- M. Naldi, U. Černigoj, A. Štrancar, *et al.*, *Talanta*, 2017, **167**, 143–157.
- J. P. Lafleur, R. Kwapiszewski, T. G. Jensen, *et al.*, *Analyst*, 2013, **138**(3), 845–849.
- J. P. Lafleur, S. Senkbeil, J. Novotny, *et al.*, *Lab Chip*, 2015, **15**(10), 2162–2172.
- A. Jönsson, R. R. Svejdal, N. Bøgelund, *et al.*, *Anal. Chem.*, 2017, **89**(8), 4573–4580.
- J. V. Staros, *Biochemistry*, 1982, **21**(17), 3950–3955.
- R. D. Voyksner, D. C. Chen and H. E. Swaisgood, *Anal. Biochem.*, 1990, **188**(1), 72–81.
- I. Kozak, S. F. Oster, M. A. Cortes, *et al.*, *Ophthalmology*, 2011, **118**(6), 1119–1124.
- M. Chiari, M. Cretich, F. Damin and L. Ceriotti, *Electrophoresis*, 2000, **21**(5), 909–916.
- G. Bayramoglu, O. Celikbicak, M. Yakup Arica and B. Salih, *Ind. Eng. Chem. Res.*, 2014, **53**(12), 4554–4564.
- G. Bayramoğlu, M. Yilmaz, A. Ü. Şenel and M. Yakup Arica, *Biochem. Eng. J.*, 2008, **40**(2), 262–274.
- D. Horák, J. Kučerová, L. Korecká, *et al.*, *Macromol. Biosci.*, 2012, **12**(5), 647–655.
- J. Wu, M. Luan and J. Zhao, *Int. J. Biol. Macromol.*, 2006, **39**(4–5), 185–191.
- D. S. Peterson, T. Rohr, F. Svec and J. M. J. Fréchet, *Anal. Chem.*, 2002, **74**(16), 4081–4088.
- J. R. Freije, P. P. Mulder, W. Werkman, *et al.*, *J. Proteome Res.*, 2005, **4**(5), 1805–1813.
- L. Sun, Y. Li, P. Yang, *et al.*, *J. Chromatogr., A*, 2012, **1220**, 68–74.
- W. K. Russell, Z. Y. Park and D. H. Russell, *Anal. Chem.*, 2001, **73**(11), 2682–2685.
- L. M. Simon, K. László, A. Vértesi, *et al.*, *J. Mol. Catal. B: Enzym.*, 1998, **4**(1–2), 41–45.
- A. Dhabaria, P. Cifani, C. Reed, *et al.*, *J. Proteome Res.*, 2015, **14**(8), 3403–3408.
- U. Jakop, B. Fuchs, R. Süß, *et al.*, *Lipids Health Dis.*, 2009, **8**, 49.
- T. Masuda, M. Tomita and Y. Ishihama, *J. Proteome Res.*, 2008, **7**(2), 731–740.
- I. R. León, V. Schwämmle, O. N. Jensen and R. R. Sprenger, *Mol. Cell. Proteomics*, 2013, **12**(10), 2992–3005.
- Y. Zhang, D. Bottinelli, F. Lisacek, *et al.*, *Anal. Biochem.*, 2015, **484**, 40–50.
- J. L. Proc, M. A. Kuzyk, D. B. Hardie, *et al.*, *J. Proteome Res.*, 2010, **9**(10), 5422–5437.

- 46 I. Pereiro, A. Bendali, S. Tabnaoui, *et al.*, *Chem. Sci.*, 2017, **8**, 1329.
- 47 D. Ferraro, J. Champ, B. Teste, *et al.*, *Sci. Rep.*, 2016, **6**, 25540.
- 48 J. Srbova, P. Krulisova, L. Holubova, *et al.*, *Electrophoresis*, 2017, DOI: 10.1002/elps.201700257, in press.
- 49 S. Korchane, A. Pallandre, C. Przybylski, *et al.*, *Electrophoresis*, 2014, **35**(7), 1050–1059.

**Titre :** Quantification multiplexe de biomarqueurs d'intérêt clinique et de leurs protéoformes par spectrométrie de masse. Application à l'analyse de cohortes médicales

Les protéoformes désignent toutes les formes sous lesquelles une protéine peut être présente. Cela inclut les formes portant des modifications post-traductionnelles, les isoformes et les formes résultant d'épissages alternatifs. Ces modifications peuvent influencer sur la fonction d'une protéine, d'où l'intérêt de développer des méthodes sensibles, spécifiques et robustes de quantification de protéoformes pour une meilleure compréhension de mécanismes pathologiques ou la recherche de biomarqueurs. L'objectif de ce travail a été d'exploiter les avantages de la spectrométrie de masse à haute résolution pour la caractérisation et la quantification de protéoformes de protéines associées à des maladies neurodégénératives. Nous nous sommes principalement intéressés à deux protéines, la C9ORF72 et l'alpha-synucléine. Dans un premier temps, une méthode de quantification des deux isoformes de la C9ORF72 a été développée et appliquée à une cohorte de 43 cerveaux humains comprenant des cas de démences fronto-temporale (DFT) avec ou sans mutation C9ORF72. Les résultats obtenus montrent pour la première fois par spectrométrie de masse une diminution d'environ 50% de l'isoforme longue de la C9ORF72 en présence de la mutation. Dans une seconde étape, nous avons élargi l'analyse à la quantification multiplexe de 49 protéines cérébrales potentiellement impliquées dans les DFT. En parallèle, nous nous sommes intéressés aux formes tronquées de l'alpha-synucléine. Leur quantification a été réalisée par une approche top-down dans des tissus cérébraux et une approche par bottom-up dans le liquide céphalorachidien (LCR). Enfin, l'analyse a été étendue à la quantification multiplexe de l'alpha-synucléine et de la protéine tau du LCR.

**Mots clés :** protéoformes, spectrométrie de masse, quantification, maladie neurodégénératives, biomarqueurs

**Title :** Multiplexed mass spectrometry based quantification of clinical biomarkers proteoforms. Application to clinical cohorts

Proteoforms describe the complexity of protein forms. This includes forms with post-translational modifications, isoforms and forms resulting from alternative splicing. These modifications can influence the function of a protein, hence the interest in developing sensitive, specific and robust methods for the quantification of proteoforms for a better understanding of pathological mechanisms or biomarkers discovery. The objective of this work was to take advantage of high-resolution mass spectrometry for the characterization and quantification of proteoforms associated with neurodegenerative diseases. We mainly focused on two proteins, C9ORF72 and alpha-synuclein. First, a method for quantifying the two C9ORF72 isoforms was developed and applied to a cohort of 43 human brains including cases of frontotemporal dementia (FTD) with or without C9ORF72 mutation. The results obtained show for the first time by mass spectrometry a decrease of about 50% of the long isoform of C9ORF72 in the presence of the mutation. Then, we extended the analysis to the multiplex quantification of 49 brain proteins potentially involved in FTD. In parallel, we focused on the truncated forms of alpha-synuclein. Their quantification was performed by a top-down approach in brain tissue and a bottom-up approach in cerebrospinal fluid (CSF). Finally, the analysis was extended to the multiplex quantification of alpha-synuclein and tau protein in CSF.

**Keywords :** proteoform, mass spectrometry, quantification, neurodegenerative diseases, biomarkers

