

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA PROTEOMIQUE NON-CIBLÉE PAR LC-MS/MS APPLIQUÉE AUX ÉCHANTILLONS
BIOLOGIQUES COMPLEXES ET LA MISE AU POINT D'UNE MÉTHODE CIBLÉE POUR LA
QUANTIFICATION DE PROTÉINES S100 ET ANNEXINES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

À LA MAÎTRISE EN BIOCHIMIE

PAR

CARINA LIMA

AVRIL 2025

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens tout d'abord à remercier et à exprimer ma plus sincère gratitude à ma directrice de recherche Pre. Lekha Sleno pour m'avoir ouvert les portes de son laboratoire. Votre expertise, vos précieux conseils, votre dévouement et votre engagement m'ont permis de mener à bien ce magnifique projet. Votre amour pour la science a été pour moi une source d'inspiration et un exemple, me donnant la motivation à me surpasser et à persévérer.

Je souhaite également remercier tous mes collègues de laboratoire : Leanne Ohlund, Kathrina Mae Kumaresan, Nathan Ghafari, Said Matar, Oriana Zambito, Kahina Chabi et Dre. Maggy Lépine pour leur conseils et soutien tout au long de mon cheminement.

Je voudrais également remercier l'ensemble de nos collaborateurs dont les Pr. Marc Lussier, Pre. Élodie Boisselier, Pr. Laurent Cappadocia, Dre. Marie-Claude Robert et Pr. Christopher Rose, ainsi que les membres de leurs laboratoires pour leurs contributions essentielles et inestimables à ce mémoire.

Finalement, je tiens à exprimer ma reconnaissance la plus sincère à ma famille, mes parents et ma sœur, ainsi qu'à mes amis. Merci de m'avoir encouragée et épaulée tout au long de ce parcours. Sans votre soutien inestimable et votre présence je n'aurais pas pu surmonter les défis et les moments de doute.

TABLE DES MATIÈRES

REMERCIEMENTS.....	ii
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX.....	x
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xi
LISTE DES SYMBOLES ET DES UNITÉS.....	xv
RÉSUMÉ.....	xvi
ABSTRACT.....	xvii
INTRODUCTION.....	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	3
1.1 Les approches en protéomique.....	3
1.2 Études protéomiques par spectrométrie de masse.....	4
1.2.1 Extraction des protéines.....	5
1.2.2 Déplétion et enrichissement des protéines.....	6
1.2.3 Protéomique ascendante.....	8
1.3 Chromatographie liquide couplée à la spectrométrie de masse en tandem.....	9
1.3.1 Chromatographie liquide à haute performance.....	10
1.3.2 Spectrométrie de masse en tandem à haute résolution.....	11
1.3.3 Identification des protéines par séquençage.....	15
1.3.4 Les modes d'acquisition des données.....	16
1.4 Objectifs.....	19
CHAPITRE 2 ARTICLE SCIENTIFIQUE : ÉTUDE INTERACTOMIQUE BIOID COUPLÉE À L'ANALYSE PAR SPECTROMÉTRIE DE MASSE À HAUTE RÉOLUTION.....	21
2.1 Résumé.....	22
2.2 Introduction.....	22
2.3 Matériels et méthodes.....	25
2.3.1 Produits chimiques.....	25
2.3.2 Échantillons cellulaires.....	26
2.3.2.1 Culture des cellules Neuro2A.....	26
2.3.2.2 Production des culots cellulaires.....	26
2.3.3 Extraction des protéines.....	26
2.3.4 Protocole d'enrichissement des peptides biotinylées.....	27

2.3.5	Protocole d'enrichissement des protéines biotinylées	28
2.3.6	Analyses par LC-HRMS/MS.....	28
2.3.6.1	Acquisition Dépendante des Données (DDA).....	29
2.3.6.2	Acquisition Indépendante des Données (DIA-SWATH)	29
2.3.7	Traitement des données	29
2.4	Résultats et discussion	30
2.4.1	Élaboration d'une approche d'enrichissement pour l'étude des protéines biotinylées par LC-HRMS/MS.....	30
2.4.1.1	Identification des protéines par la méthode d'enrichissement des peptides.....	30
2.4.1.2	Identification des protéines par la méthode d'enrichissement des protéines	32
2.4.1.3	Comparaison des deux méthodes d'enrichissement	34
2.4.2	Étude interactomique de la protéine RNF13	38
2.4.2.1	Identification des interacteurs par une analyse en mode acquisition dépendante des données (DDA).....	38
2.4.2.2	Évaluation de l'enrichissement des protéines par une analyse en mode acquisition indépendante des données de type SWATH	39
2.4.2.3	Résultats de l'analyse combinée des flux de travail DDA et DIA-SWATH	40
2.4.2.4	Analyse STRING des potentiels interacteurs de RNF13	41
CHAPITRE 3 ARTICLE SCIENTIFIQUE : DÉVELOPPEMENT D'UNE MÉTHODE CIBLÉE PAR LC-MRM POUR L'ANALYSE QUANTITATIVE DES PROTÉINES S100 ET ANNEXINES ET SON APPLICATION DANS L'ANALYSE DE LARMES HUMAINES		44
3.1	Résumé	45
3.2	Introduction	45
3.3	Matériels et méthodes.....	48
3.3.1	Produits chimiques	48
3.3.2	Analyse des protéines S100 et ANX standards par LC-HRMS/MS	48
3.3.2.1	Expression et purification des protéines S100 et ANX.....	48
3.3.2.2	Digestions des protéines standards	49
3.3.2.3	Analyse LC-HRMS/MS et traitement des données.....	50
3.3.3	Analyse ciblée des protéines S100 et ANX dans les larmes par LC-MRM	51
3.3.3.1	Collecte des échantillons de larmes	51
3.3.3.2	Digestion des échantillons de larmes	51
3.3.3.3	Analyse LC-MRM des larmes et traitement des données	51
3.4	Résultats et discussion	52
3.4.1	Développement d'une méthode ciblée LC-MRM pour la quantification des protéines S100 et ANX.....	52
3.4.2	Application de la méthode LC-MRM pour l'analyse de larmes humaines	62
CHAPITRE 4 ARTICLE SCIENTIFIQUE : COMPARAISON DE MÉTHODES POUR L'ANALYSE NON-CIBLÉE DU PROTÉOME PLASMATIQUE PAR LC-HRMS/MS		67
4.1	Résumé	68

4.2 Introduction	68
4.3 Matériels et méthodes	70
4.3.1 Produits chimiques	70
4.3.2 Échantillons commerciaux de plasma	71
4.3.3 Préparation d'une courbe d'étalonnage adaptée à la matrice.....	71
4.3.4 Méthode 1 - Précipitation des protéines du plasma par l'acétone.....	71
4.3.5 Méthode 2 - Déplétion des IgG du plasma suivit d'une précipitation des protéines à partir d'acétone.....	72
4.3.6 Méthode 3 - Extraction des protéines à partir du kit de plaque filtrante ProTrap.....	72
4.3.7 Méthode 4 - Enrichissement des vésicules extracellulaires par l'approche MagNet	73
4.3.8 Méthode 5 - Déplétion des protéines abondantes à partir d'acide perchlorique.....	73
4.3.9 Analyse LC-HRMS/MS	74
4.3.9.1 Acquisition Dépendante des Données (DDA).....	75
4.3.9.2 Acquisition Indépendante des Données (DIA-SWATH)	75
4.3.10 Traitement des données	76
4.3.10.1 Identification des protéines à partir des données DDA et DIA-SWATH	76
4.3.10.2 Analyse des courbes d'étalonnages adaptée à la matrice	77
4.4 Résultats et discussion	77
4.4.1 Identification et caractérisation des protéines extraites	81
4.4.2 Évaluation de la précision quantitatives par l'analyse de courbes d'étalonnage adaptées à la matrice	86
4.4.3 Choix des méthodes pour des études <i>in vivo</i> futures	89
CONCLUSION ET PERSPECTIVES.....	91
ANNEXE A SPECTRES MS ET MS/MS DES EXPÉRIENCES BIOID.....	96
BIBLIOGRAPHIE.....	102

LISTE DES FIGURES

Figure 0.1: Les différentes sciences «omiques» et leur application	1
Figure 1.1: Les différentes étapes d'un flux de travail pour la protéomique ascendante	5
Figure 1.2: Complexe biotine-streptavidine.....	8
Figure 1.3: Les différentes composantes d'un spectromètre de masse	12
Figure 1.4: Principe de l'ionisation par électrospray (ESI).....	13
Figure 1.5: Les composantes d'un spectromètre de masse de type quadripôle-temps de vol (QqTOF) et son mode de fonctionnement.....	14
Figure 1.6: Les différents ions produit lors de la fragmentation du lien peptidique	15
Figure 1.7: Exemple de séquençage d'un peptide biotinylé par spectrométrie de masse en tandem à haute résolution	16
Figure 1.8: Comparaison des modes d'acquisition DDA et DIA pour les analyses non-ciblées..	17
Figure 1.9: Les composantes d'un spectromètre de masse de type triple quadripôle (QqQ) et son mode de fonctionnement	19
Figure 2.1: Principe du marquage des protéines par BioID et exemple d'un flux de travail pour l'analyse des protéines biotinyllées par LC-MS/MS	24
Figure 2.2: Procédure de l'extraction des protéines par une lyse cellulaire.....	27
Figure 2.3: Analyse des protéines biotinyllées par la méthode d'enrichissement des peptides... 31	
Figure 2.4: Résultats de l'identification des protéines biotinyllées par la méthode d'enrichissement des peptides	31
Figure 2.5: Analyse des protéines biotinyllées par la méthode d'enrichissement des protéines.. 32	
Figure 2.6: Résultats de l'identification des protéines biotinyllées par la méthode d'enrichissement des protéines.....	33
Figure 2.7: Comparaison des protéines identifiées dans la fraction PNB et PB pour la validation des protéines biotinyllées	34
Figure 2.8: Structure cristallographique comparant le complexe Tamadivine-Biotine (structure verte) et le complexe Streptavidine-Biotine (structure bleue)	35
Figure 2.9: Diagramme de Venn comparant les protéines identifiées en DDA.....	39

Figure 2.10: Analyse quantitative de l'enrichissement des protéines dans RNF13-UltraID +Biotine	40
Figure 2.11: Résultats combinées des différentes approches DDA et DIA pour l'analyse des interacteurs de RNF13.....	41
Figure 2.12: Analyse STRING combinant les 131 protéines identifiées en DDA et les 77 protéines quantifiées en DIA-SWATH	42
Figure 3.1: Flux de travail correspondant à l'optimisation de la méthode LC-MRM et exemple d'un chromatogramme correspondant à l'optimisation de l'énergie de collision pour la transition m/z 636,9 → 1088,6 (y9) associée au peptide ALNSIIDVYHK de la protéine S100A8.....	53
Figure 3.2: Exemple de transitions et de leur chromatogramme pour l'identification du peptide ALNSIIDVYHK appartenant à la protéine S100A8	54
Figure 3.3: Flux de travail correspondant à l'analyse ciblée LC-MRM des échantillons de larmes humaines.....	62
Figure 3.4: Chromatogrammes d'ion extrait pour certaines des protéines S100 et ANX.....	64
Figure 3.5: Carte thermique correspond aux variations de l'expression des protéines S100 et ANX observées chez les différents groupes d'échantillons	65
Figure 3.6: Principales protéines S100 impliquées dans les différents stades de développement de NALFD, NASH et HCC	66
Figure 4.1: Classement des protéines du plasma en fonction de leur concentration dans la matrice	70
Figure 4.2: Flux de travail associé au chargement des échantillons sur les Evotip Pure.....	75
Figure 4.3: Les différentes méthodes de préparation d'échantillon testées pour la protéomique plasmatique ascendante.....	80
Figure 4.4: Analyse comparative des différents flux de travail utilisés pour l'extraction des protéines du plasma	82
Figure 4.5: Caractérisation des protéines extraites par MagNet, PCA et ProTrap.....	83
Figure 4.6: Comparaison des logiciels de traitement des données DIA-NN et PEAKS pour l'identification des protéines à partir des données DIA-SWATH.....	86
Figure 4.7: Principe de l'analyse des courbes d'étalonnage adaptées à la matrice (Matrix-matched calibration curve)	87

Figure 4.8: Comparaison de la distribution des coefficients de variation (CV) pour la méthode MagNet faite manuellement et l'approche automatisée sur le robot Opentrons.....	89
Figure 5.1: Flux de travail correspondant à l'automatisation de la méthode MagNet	92
Figure 5.2: Carte thermique montrant la détectabilité des protéines S100 et ANX dans le plasma selon les différentes méthodes d'extractions testées	94
Figure A.1: Spectre MS et MS/MS du peptide ILGTFKSKK ²⁺ (FA83H) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS+Biotine pour la méthode d'enrichissement des peptides.....	96
Figure A.2: Spectre MS et MS/MS du peptide TILGAGPVLEAFGNAK ³⁺ (MYO9B) identifié dans la fraction peptide biotinylé (PB) de l'échantillon -FBS -Biotine pour la méthode d'enrichissement des peptides.....	96
Figure A.3: Spectre MS et MS/MS du peptide GQPLCVLSAMKMETVVTSPMEGTIR ³⁺ (PYC) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines	97
Figure A.4: Spectre MS et MS/MS du peptide FNVWDTAGQEKFGLR ³⁺ (RAN) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines	97
Figure A.5: Spectre MS et MS/MS du peptide MQASIEKGGSLPK ³⁺ (NPM) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines.....	98
Figure A.6: Spectre MS et MS/MS du peptide AGDSLVMMIAMKMEHTIK ⁴⁺ (MCCA) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines	98
Figure A.7: Spectre MS et MS/MS du peptide SKAESALLR ³⁺ (C2D2A) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines.....	99
Figure A.8: Spectre MS et MS/MS du peptide KQSASWALKCVK ²⁺ (TTC14) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines	99
Figure A.9 : Spectre MS et MS/MS du peptide GQPLCVLSAMKMETVVTSPMEGTIR ³⁺ (PYC) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS -Biotine pour la méthode d'enrichissement des protéines	100

Figure A.10: Spectre MS et MS/MS du peptide AGDSLVMVIAMKMEHTIK³⁺ (MCCA) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS -Biotine pour la méthode d'enrichissement des protéines 100

Figure A.11: Spectre MS et MS/MS du peptide STSATCGKK²⁺ (BCORL) identifié dans la fraction peptide biotinylé (PB) de l'échantillon -FBS -Biotine et pour la méthode d'enrichissement des protéines..... 101

LISTE DES TABLEAUX

Tableau 1.1: Les avantages et limitations liés aux approches de déplétion	7
Tableau 2.1 : Avantages et limitations liés aux deux méthodes d'enrichissement.....	37
Tableau 3.1: Les différents interactions formées entre les protéines S100 et ANX répertoriés dans la littérature.....	46
Tableau 3.2: Protéines S100 et ANX précédemment identifiées par LC-HRMS/MS dans les larmes de patients sains ou quantifiées par LC-MRM chez des patients atteints du syndrome de Sjögren ou de la pemphigoïde des membranes muqueuses.....	47
Tableau 3.3: Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines	55
Tableau 3.4: Identification des protéines S100 et ANX au travers des groupes d'échantillons ..	63
Tableau 4.1: Évaluation de la précision quantitative des différentes méthodes de préparations d'échantillons en plasma par l'analyse de courbes d'étalonnage adaptée à la matrice	88
Tableau 4.2: Analyse comparative des différents flux de travail de préparation d'échantillons de plasma testés	90
Tableau 5.1: Concentrations plasmatiques des protéines S100 et ANX	93

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ABC	Bicarbonate d'ammonium
ACN	Acétonitrile
ANX	Annexine
AP3B1	<i>AP-3 complex subunit beta-1</i>
AP3D1	<i>AP-3 complex subunit delta-1</i>
AP3M1	<i>AP-3 complex subunit mu-1</i>
AP3S1	<i>AP-3 complex subunit sigma-1</i>
ARF2	<i>ADP-ribosylation factor 2</i>
BCORL	<i>BCL-6 corepressor-like protein 1</i>
bioAMP	biotinyl-AMP
BioID	<i>Proximity-dependent biotin identification</i>
BTP	Bis-Tris Propane
CE	Énergie de collision (<i>Collision energy</i>)
CID	Dissociation induite par collision (<i>Collision induced dissociation</i>)
CUR	Gaz rideau
CV	Coefficient de variation
DBS	<i>Dynamic background subtraction</i>
DDA	Acquisition dépendante des données (<i>Data-dependent acquisition</i>)
DIA	Acquisition indépendante des données (<i>Data-independent acquisition</i>)
DP	Potentiel de dégroupage (<i>Declustering potential</i>)
DTT	Dithiothréitol
EB	Epidermolyse bulleuse
EMEM	<i>Eagle's minimal essential medium</i>
ESI	Électrospray (<i>Electrospray ionisation</i>)

F10A1	<i>Hsc70-interacting protein</i>
FA	Acide formique
FA83H	<i>Protein FAM83H</i>
FBS	<i>Fetal bovin serum</i>
FDR	<i>False discovery rate</i>
HCC	Carcinome hépatocellulaire
HPLC	Chromatographie liquide à haute performance (<i>High-performance liquid chromatography</i>)
HRMS	Spectrométrie de masse à haute résolution (<i>High resolution mass spectrometry</i>)
IAM	Iodoacétamide
IgG	Immunoglobuline G
IPTG	isopropyl- β -D-thiogalactopyranoside
LC-MS/MS	Chromatographie liquide couplée à la spectrométrie de masse en tandem
<i>m/z</i>	Rapport masse sur charge
MALDI	Désorption-ionisation laser assistée par matrice (<i>Matrix Assisted Laser Desorption Ionisation</i>)
MCCA	<i>Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial</i>
MeOH	Méthanol
MRM	Surveillance de réactions multiples (<i>Multiple reaction monitoring</i>)
MS	Spectrométrie de masse
MS/MS	Spectrométrie de masse en tandem
MYO9B	<i>Unconventional myosin-IXb</i>
NAFLD	Stéatose hépatique non alcoolique
NASH	Stéatohépatite non alcoolique
NPM	<i>Nucleophosmin</i>

ODR4	<i>Protein odr-4 homolog</i>
PB	Peptides biotinylés
PBS	Tampon phosphate salin (<i>Phosphate-buffered saline</i>)
PCA	Acide perchlorique
PLCX3	<i>PI-PLC X domain-containing protein 3</i>
PNB	Peptides non-biotinylés
PTM	Modification post-traductionnelle (<i>Post-translational modification</i>)
PYC	<i>Pyruvate carboxylase, mitochondrial</i>
Q	Quadripôle
QqQ	Triples-quadripôles
QqTOF	Quadripôle-temp de vol (<i>Quadrupole time of flight</i>)
RAN	<i>GTP-binding nuclear protein Ran</i>
RM16	<i>Large ribosomal subunit protein uL16m</i>
RNF13	<i>Ring Finger Protein 13</i>
RPLC	Chromatographie liquide en phase inverse (<i>Reverse phase liquid chromatography</i>)
RRAGA	<i>Ras-related GTP-binding protein A</i>
RS27A	<i>Ubiquitin-ribosomal protein eS31 fusion protein</i>
SAX	Échangeur anionique (<i>Strong Anion Exchange</i>)
SDS	Sodium dodecyl sulfate
sMRM	<i>Détection de réaction sélectionnée cédulée (scheduled multiple reaction monitoring)</i>
SPE	Extraction en phase solide (<i>solid phase extraction</i>)
SWATH	<i>Sequential Windows Acquisition of All Theoretical Spectra</i>
TCEP	Tris(2-carboxyéthyl)phosphine
TFA	Acide trifluoroacétique

TOF	Temp de vol (<i>Time of Flight</i>)
t_r	Temps de rétention
TRIPC	<i>E3 ubiquitin-protein ligase TRIP12</i>
UFSP2	<i>Ufm1-specific protease 2</i>
z	Charge

LISTE DES SYMBOLES ET DES UNITÉS

L	Litre
mL	Millilitre
μ L	Microlitre
g	Gramme
mg	Milligramme
μ g	Microgramme
M	Molaire
mM	Millimolaire
%	Pourcentage
h	Heure
min	Minute
s	Seconde
ms	Milliseconde
$^{\circ}$ C	Degré Celsius
pH	Potentiel hydrogène
rpm	Rotation par minute
<i>g</i>	Force gravitationnelle
mm	Millimètre
μ m	Micromètre
kDa	Kilodalton
Da	Dalton
V	Volt
psi	Livre-force par pouce carré
ppm	Parties par million
~	Environ

RÉSUMÉ

La protéomique basée sur la MS s'est avérée être un domaine essentiel à la biochimie moderne. L'étude des protéines à large échelle a permis d'améliorer notre compréhension sur la biologie de ces molécules, ouvrant la voie vers l'identification de nouveaux biomarqueurs et cibles thérapeutiques. Bien que la MS se soit révélée être un outil puissant afin d'étudier les protéines, leurs analyses reste extrêmement difficile sans l'utilisation de méthodes adaptées. Dans ce mémoire, nous avons exploré, à travers trois projets, différents concepts liés au développement de méthodes pour les études en protéomique ascendante par LC-MS/MS. Le chapitre 2, s'intéresse à l'élaboration d'une approche permettant l'analyse non ciblée des protéines biotinylées par LC-HRMS/MS dans un contexte d'étude interactomique BioID de la protéine RNF13. Deux approches d'enrichissement sur résine streptavidine ont été comparées : au niveau des protéines biotinylées et l'autre à partir des peptides. L'approche protéique, combinée aux analyses DDA et DIA-SWATH, s'est avérée être très efficace permettant d'identifier la modification et d'établir le niveau d'enrichissement des protéines. L'analyse de l'interactome de RNF13 a permis d'identifier plus d'une centaine d'interacteurs potentiels. Dans le chapitre 3, l'objectif était de développer une méthode LC-MRM ciblée pour la quantification de 17 protéines S100 et 10 annexines dans des larmes humaines. Une analyse initiale de protéines standards par une approche LC-HRMS/MS non ciblée a permis la sélection de peptides tryptiques uniques à optimiser par LC-MRM. La méthode développée a été appliquée afin d'analyser différents échantillons de larmes de patients, certains souffrant d'une maladie hépatique. Plusieurs de ces protéines se sont révélées fortement altérées chez ce groupe de patients. Pour le chapitre 4, l'objectif était de tester différentes méthodes de préparation d'échantillons pour la protéomique plasmaticque non ciblée. Nous avons comparé cinq approches: l'enrichissement des vésicules à partir de billes magnétiques SAX, une déplétion des protéines abondantes à l'acide perchlorique, une méthode de solubilisation des protéines plasmaticques suivie de leur précipitation à l'acétone, avec ou sans déplétion des IgG, et une extraction commerciale sur plaque filtrante. Les différentes méthodes ont fourni des résultats complémentaires, permettant d'augmenter le recouvrement de certaines protéines.

Mots clés : Protéomique ascendante, LC-MS/MS, analyses ciblées et non-ciblées, interactomique, BioID, RNF13, protéomique plasmaticque, larmes, annexines, S100

ABSTRACT

MS-based proteomics has proven to be an essential field in modern biochemistry. The study of proteins on a large scale has improved our understanding of the biology of these molecules, paving the way for the identification of new biomarkers and therapeutic targets. Although MS has proven to be a powerful tool for studying proteins, their analysis remains extremely difficult without the use of adapted methods. In this thesis, we explored, through three projects, different concepts related to the development of methods for bottom-up proteomic studies by LC-MS/MS. Chapter 2 focuses on the development of an approach allowing the untargeted analysis of biotinylated proteins by LC-HRMS/MS in the context of a BioID interactomic study of the RNF13 protein. Two enrichment approaches on streptavidin resin were compared: at the level of biotinylated proteins and the other from peptides. The protein approach, combined with DDA and DIA-SWATH analyses, proved to be very effective in identifying the modification and establishing the level of protein enrichment. Analysis of the RNF13 interactome identified more than a hundred potential interactors. In chapter 3, the objective was to develop a targeted LC-MRM method for the quantification of 17 S100 proteins and 10 annexins in human tears. An initial analysis of standard proteins by an untargeted LC-HRMS/MS approach allowed the selection of unique tryptic peptides to be optimized by LC-MRM. The developed method was applied to analyze different tear samples from patients, some suffering from liver disease. Several of these proteins were found to be highly altered in this group of patients. For chapter 4, the objective was to test different sample preparation methods for untargeted plasma proteomics. We compared five approaches: vesicle enrichment from SAX magnetic beads, depletion of abundant proteins with perchloric acid, a method of plasma protein solubilization followed by acetone precipitation, with or without IgG depletion, and a commercial filter plate extraction. The different methods provided complementary results, making it possible to increase the coverage of certain proteins.

Keywords: Bottom-up proteomics, LC-MS/MS, targeted and untargeted analyses, interactomics, BioID, RNF13, plasma proteomics, tears, annexins, S100

INTRODUCTION

De manière générale, les sciences dites «-omiques», telles que la génomique, la transcriptomique et la protéomique (Figure 0.1), ont pour but d'étudier dans son intégralité une classe de molécules composant un système biologique donné (Evans, 2000). Le terme protéomique fait plus particulièrement référence à la science qui cherche à caractériser l'ensemble des protéines exprimées soit par une cellule, un tissu ou un organisme à un temps donné (Chandramouli & Qian, 2009), et ce dans le but ultime d'établir leur rôle dans certains processus physiologiques et pathologiques (Cho, 2007). Il est aujourd'hui possible de diviser la protéomique en différents sphères d'application. En effet, celle-ci est non seulement appliquée afin d'identifier et quantifier les protéines, mais celle-ci est également utilisée afin d'étudier leurs modifications, leurs interactions, leur fonction ou encore leur structure (Graves & Haystead, 2002).

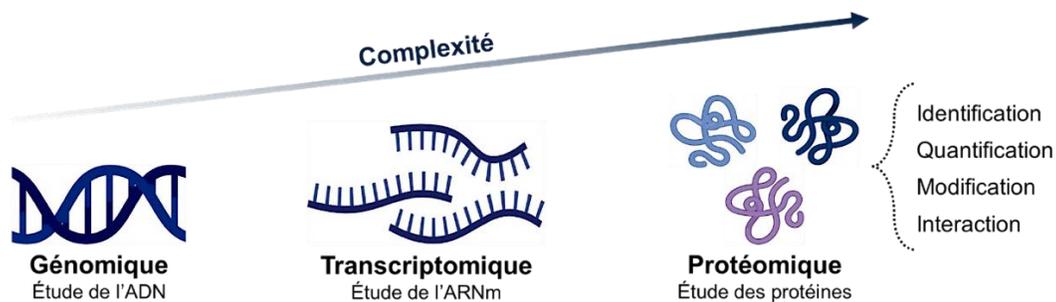


Figure 0.1: Les différentes sciences «omiques» et leur application

Les protéines étant l'unité fonctionnelle du génome (Yates, Ruse & Nakorchevsky, 2009), la protéomique permet alors de répondre à des questions que la génomique ou la transcriptomique ne peuvent à elles seules répondre (Wilhelm *et al.*, 2014; Graves & Haystead, 2002). En effet, l'étude unique des gènes ne permet pas de déterminer l'abondance, la structure ou l'état d'activité d'une protéine (Cho, 2007). Le génome étant une entité relativement stable, le protéome est quant à lui extrêmement dynamique, pouvant varier d'une cellule à l'autre, et ce en fonction du temps et de l'environnement (Conti & Alessio, 2015; Aslam *et al.*, 2016). De plus, les protéines étant continuellement synthétisées, dégradées et modifiées, ceci aura pour effet de moduler leur fonction et ainsi affecter l'état d'un organisme (Jensen, 2004). C'est pourquoi la protéomique est alors idéale afin de décrire les mécanismes biochimiques permettant de réguler un système

biologique, et ainsi d'identifier de potentiels biomarqueurs et cibles thérapeutiques associés à diverses maladies.

Les études en protéomique représentent encore à ce jour un immense défi. En effet, comparativement au génome, le protéome se caractérise également par sa complexité (Conti & Alessio, 2015). Par exemple, bien que le génome humain code pour 19 000 à 20 000 protéines, on estime qu'il en existerait en réalité plus de 1 000 000 (Aebersold *et al.*, 2018). Cette diversité résulte d'une part de l'épissage alternatif de l'ARN, mais également des différentes modifications post-traductionnelles (PTM) qu'une protéine peut subir (plus de 500 PTM différentes répertoriées) (Cho, 2007; Brandi *et al.*, 2022). De plus, les protéines n'étant pas exprimées à des niveaux égaux dans le protéome (Zhang *et al.*, 2010), celles-ci vont exister dans une plage en concentration excessivement vaste. Dans le plasma humain, par exemple, les concentrations de certaines protéines peuvent s'étendre des mg/mL (albumine et certaines immunoglobulines) jusqu'à des pg/mL pour d'autres (certaines cytokines) (Angel *et al.*, 2012). Par conséquent, la détection de protéines d'intérêt biologique présentes en faible abondance et composant certaines matrices biologiques complexes reste extrêmement difficile et ce, en raison des limitations liées aux techniques et aux instruments utilisés (Conti & Alessio, 2015).

Dans ce contexte, il nous est primordial de développer des méthodes d'analyse robustes, sensibles et efficaces, permettant d'identifier et de quantifier ces protéines dans les matrices biologiques complexes. Au cours de ce travail, nous explorerons différents aspects liés au développement de méthodes pour les études en protéomique par LC-MS/MS et ce, par l'entremise de trois projets de recherches distincts.

CHAPITRE 1

REVUE DE LITTÉRATURE

1.1 Les approches en protéomique

L'émergence et le développement d'outils biotechnologiques a grandement contribué à la croissance du domaine de la protéomique au fil des ans. Bien que les approches immuno-enzymatiques telles que ELISA et l'immunobuvardage de type Western (*Western-blot*) soient des approches couramment employées afin d'étudier les protéines, celles-ci se limitent à l'analyse d'un nombre restreint de protéines (Aslam *et al.*, 2016). L'introduction de la spectrométrie de masse (MS) a permis l'identification et l'étude de milliers de protéines en une seule analyse (Shi *et al.*, 2004), favorisant ainsi l'étude à large échelle de ces molécules. Toutefois, les analyses par spectrométrie de masse étant limitées par la complexité de la matrice biologique, une étape préalable de séparation est généralement nécessaire.

Traditionnellement, la séparation des protéines a été menée par électrophorèse sur gel bidimensionnelle (2-DE), soit une approche où les protéines sont préalablement séparées sur gel en fonction de leur point isoélectrique et de leur poids moléculaire (Zhang *et al.* 2010). L'identification subséquente des protéines est possible par leur coloration au bleu brillant de Coomassie ou au nitrate d'argent, suivie de leur excision et analyse par spectrométrie de masse (Zhang *et al.* 2010; Deracinois *et al.*, 2013; Wittmann-Liebold, Graack & Pohl, 2006). Bien que cette méthode permette l'étude des isoformes et de certaines PTM (Zhang *et al.* 2010), celle-ci présente également de nombreuses limitations. Il est par exemple difficile d'analyser les protéines hautement hydrophobes, acides ou basiques, tout comme les protéines présentes en faible abondance (Shi *et al.*, 2004). Le faible pouvoir de résolution de la méthode peut également affecter l'identification et la quantification des protéines. De plus, étant une approche fastidieuse et complexe, celle-ci est peu reproductible et limite grandement les analyses à haut débit (Zhang *et al.* 2010), un aspect important en protéomique où l'analyse simultanée de plusieurs échantillons biologiques est généralement nécessaire.

La chromatographie liquide couplée à la spectrométrie de masse (LC-MS) s'est alors positionnée comme une puissante alternative, et ce en raison de sa haute reproductibilité, de son pouvoir de résolution et de sa sensibilité d'analyse (Shi *et al.*, 2004; Zhang *et al.* 2010). La polyvalence de la méthode permet entre autres différents modes de séparation chromatographique, dont différentes phases stationnaires et mobiles (Shi *et al.*, 2004). De plus, comparativement aux approches sur

gel, la chromatographie liquide nécessite généralement moins de manipulations et d'étapes de préparation (Shi *et al.*, 2004) favorisant ainsi les analyses à haut débit (Zhang *et al.* 2010). Le principe de la LC-MS/MS sera détaillé dans la section 1.3 de ce chapitre.

1.2 Études protéomiques par spectrométrie de masse

Bien que les instruments en LC-MS/MS se soient considérablement améliorés au cours des dernières décennies, la qualité des résultats générés par cette méthode repose encore à ce jour sur une étape efficace de préparation d'échantillons (Rogers & Bomgarden, 2016). Dans un monde idéal, un protocole permettrait un recouvrement complet des protéines. Toutefois, les protéines étant des biomolécules complexes et aux propriétés physico-chimiques diversifiées, l'extraction et l'isolation de l'ensemble de ces molécules restent extrêmement difficiles (Varnavides *et al.*, 2022). De plus, il est important de mentionner qu'il n'existe pas de protocole universel pour les analyses en protéomique basée sur la spectrométrie de masse, et que le choix de certaines méthodologies reposera sur une multitude de facteurs : objectifs de l'étude, compatibilité des approches, coûts des méthodes, rapidité et simplicité d'exécution, etc (Macklin, Khan & Kislinger, 2020, Hughes *et al.*, 2019). Par conséquent, dans la plupart des cas, un flux de travail typique en protéomique basée sur la LC-MS/MS va comprendre une série d'étapes. La figure 1.1 présente un exemple de flux de travail pour la protéomique de type ascendante (*bottom-up*). Dans cet exemple, les protéines sont préalablement extraites et isolées par des approches mécaniques et/ou chimiques puis réduites, alkylées et digérées en peptides. Les peptides sont extraits sur SPE (*solid phase extraction*) avant d'être analysés par LC-MS/MS. L'identification des protéines est possible en comparant les données à des bases de données en séquence et ce, à l'aide de différents outils bio-informatiques tels que PEAKS (Bioinformatics Solutions) ou ProteinPilot (Sciex), par exemple.

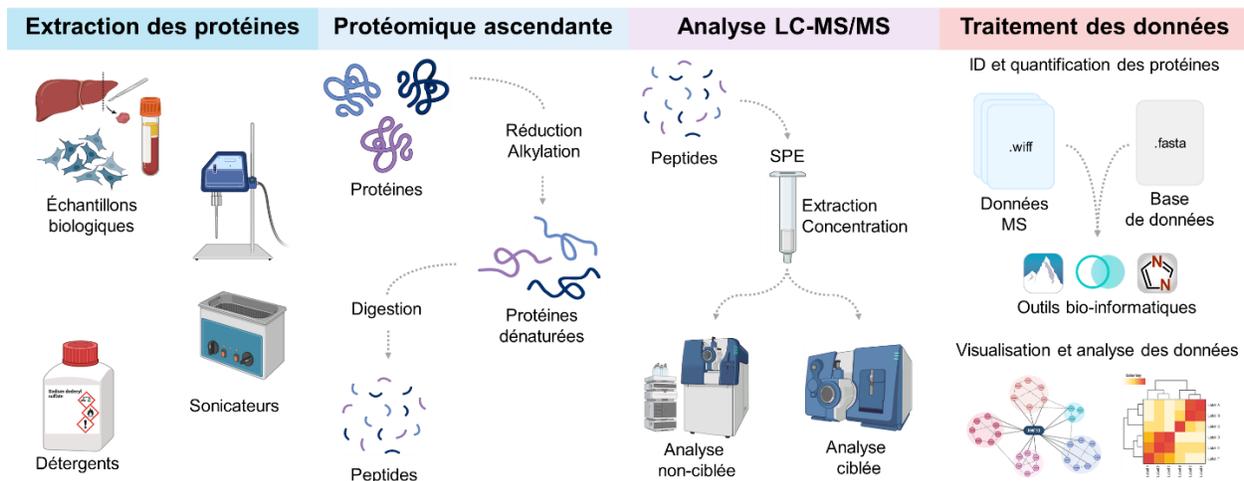


Figure 1.1: Les différentes étapes d'un flux de travail pour la protéomique ascendante

1.2.1 Extraction des protéines

La première étape d'un flux de travail en protéomique est l'extraction des protéines. C'est une étape extrêmement importante pouvant directement affecter le recouvrement de ces molécules. Lors de cette étape, il est important de prendre en considération le type d'échantillon biologique disponible (cellules, tissus, fluides biologiques) et ce, afin d'adapter la méthodologie. Par exemple, les échantillons cellulaires et tissulaires vont nécessiter une lyse physique (homogénéisation, sonication) afin de libérer leur contenu cellulaire (Rogers & Bomgarden, 2016). Dans certains cas, l'ajout de détergents ou d'agents chaotropes peut également contribuer à l'extraction des protéines en favorisant la solubilisation des protéines membranaires et des complexes protéiques (Jiang *et al.*, 2024; Rogers & Bomgarden, 2016). Cependant, il est également important de mentionner que plusieurs de ces composés peuvent être incompatibles avec les étapes ultérieures du flux de travail (Hughes *et al.* 2019). Par exemple, une concentration trop élevée en urée (>2M) peut affecter l'activité enzymatique de la trypsine lors de la digestion des protéines, nécessitant alors de diluer l'échantillon avant cette étape (Jiang *et al.*, 2024; Hughes *et al.* 2019). Le sodium dodécylsulfate (SDS) est entre autres incompatible pour les étapes de digestion enzymatique, mais encore plus néfaste pour la chromatographie liquide en phase inverse et la détection par spectrométrie de masse (Jiang *et al.*, 2024; Rogers & Bomgarden, 2016). C'est pourquoi il est alors important d'éliminer ce composé en procédant par exemple, à une étape de précipitation des protéines (Jiang *et al.*, 2024; Rogers & Bomgarden, 2016).

1.2.2 Déplétion et enrichissement des protéines

L'un des principaux obstacles en protéomique basée sur la LC-MS/MS réside dans la complexité de la matrice biologique, et plus particulièrement dans la plage dynamique de concentration en protéines. En effet, celle-ci peut s'étendre de 6 à 9 ordres de grandeur pour les cellules et les tissus, et jusqu'à plus de 12 ordres de magnitude pour le plasma (Rogers & Bomgarden, 2016). Par conséquent, la présence de protéines fortement abondantes aura pour effet de rendre la détection de protéines peu abondantes extrêmement difficile et ainsi limiter l'identification de potentiels biomarqueurs protéiques. De plus, ces plages se situant généralement hors des domaines de détection des instruments LC-MS/MS (Angel *et al.*, 2012), il est nécessaire, dans plusieurs cas, d'introduire une étape adéquate de préparation d'échantillon afin de réduire cette complexité et améliorer la sensibilité de détection des protéines moins abondantes. Diverses méthodes basées sur la déplétion des protéines très abondantes ou l'enrichissement des protéines de faibles concentrations peuvent ainsi être employées.

De manière générale, la déplétion consiste à éliminer les protéines abondantes qui pourraient causer la suppression du signal des protéines peu abondantes (Lee *et al.*, 2019). Il existe actuellement une multitude d'approches de ce type dont l'immunodéplétion et la déplétion basée sur la précipitation des protéines. L'immunodéplétion est une technique largement utilisée en protéomique du plasma. Cette méthode repose sur l'interaction spécifique entre un antigène et son anticorps afin de dépléter de manière générale jusqu'à 20 protéines les plus abondantes du plasma telles que l'albumine, plusieurs immunoglobulines, la transferrine, le fibrinogène et les apolipoprotéines (Jiang *et al.*, 2024; Geyer *et al.*, 2017; Rogers & Bomgarden, 2016). Étant une approche hautement spécifique et reproductible, celle-ci est toutefois excessivement dispendieuse pouvant ainsi limiter les études à large échelle (Zhao *et al.*, 2023). De plus, les interactions croisées entre certaines protéines abondantes et d'autres protéines d'intérêts biologiques peuvent également mener à leur déplétion (Zhao *et al.*, 2023). La déplétion basée sur la précipitation des protéines consiste quant à elle à induire l'agrégation et la précipitation des protéines en altérant leur solubilité à partir de différents agents : sels inorganiques, solvants inorganiques, agents réducteurs ou acides (Zhao *et al.*, 2023; Jiang *et al.*, 2024). Ainsi, par une étape de centrifugation, il est alors possible de récupérer le surnageant déplété et de mener une analyse protéomique de celui-ci (Zhao *et al.*, 2023). Contrairement à l'immunodéplétion, cette approche est très peu coûteuse et simple. Toutefois, celle-ci est très peu spécifique et présente des risques élevés de perte (Jiang *et al.*, 2024). Le tableau 1.1 résume les avantages et les limitations liées aux deux méthodes de déplétion présentées.

Tableau 1.1: Les avantages et limitations liés aux approches de déplétion

Approches	Avantages	Limitations
Immunodéplétion	<ul style="list-style-type: none"> • Hautement spécifique • Excellente reproductibilité 	<ul style="list-style-type: none"> • Dispendieuse • Risque de dépléter des protéines d'intérêts par leur interaction croisée avec les protéines cibles (abondantes)
Déplétion basée sur la précipitation des protéines	<ul style="list-style-type: none"> • Simplicité d'exécution • Méthode très peu coûteuse 	<ul style="list-style-type: none"> • Peu spécifique • Risque de perte en protéines

Les méthodes d'enrichissement ont pour but d'isoler une classe de protéines spécifiques, afin d'augmenter leur sensibilité de détection (Lee *et al.*, 2019). Par exemple, bien que la spectrométrie de masse soit l'approche de choix afin d'étudier les PTM, celles-ci restent extrêmement difficile à analyser sans une étape efficace d'enrichissement. En effet, les peptides modifiés étant présents à des niveaux sous-stœchiométriques par rapport à leurs homologues non-modifiés, autrement dit étant beaucoup moins abondants, ceux-ci ne sont généralement pas détectés et ce, en raison des effets de suppression ioniques (Zhang *et al.*, 2015). C'est pourquoi certaines méthodes d'enrichissement ont ainsi été spécifiquement développées afin d'enrichir différents types de protéomes: phosphoprotéome, glycoprotéome (Ignjatovic *et al.*, 2019), etc. D'autres approches ont également été développées afin d'isoler des protéines issues d'un compartiment cellulaire ou organite spécifique telles que les vésicules membranaires (Ignjatovic *et al.*, 2019). Dans ce mémoire, nous nous intéresserons entre autres à la biotinylation des protéines. L'une des méthodes d'enrichissement les plus fréquemment utilisées se base sur la forte affinité biotine-streptavidine ($K_d \approx 10^{-15}M$) (Cirri *et al.*, 2023). La streptavidine est une protéine homotétramérique pouvant lier jusqu'à quatre molécules de biotine (Dundas, Demonte & Park, 2013). Cette forte interaction est le résultat de la formation d'un nombre important de ponts hydrogènes soit huit au total. Par exemple, les résidus Asn23, Ser27 et Tyr43 forment chacun un pont hydrogène avec l'atome d'oxygène de la fonction uréido ($-NHCONH-$) de la biotine (Liu, Zhang & Mei, 2016). La figure 1.2 présente la structure cristallographique du complexe formé entre la streptavidine et la biotine (Figure 1.2A), les différentes interactions formées entre les deux molécules (Figure 1.2B) et la structure de la biotine (Figure 1.2C).

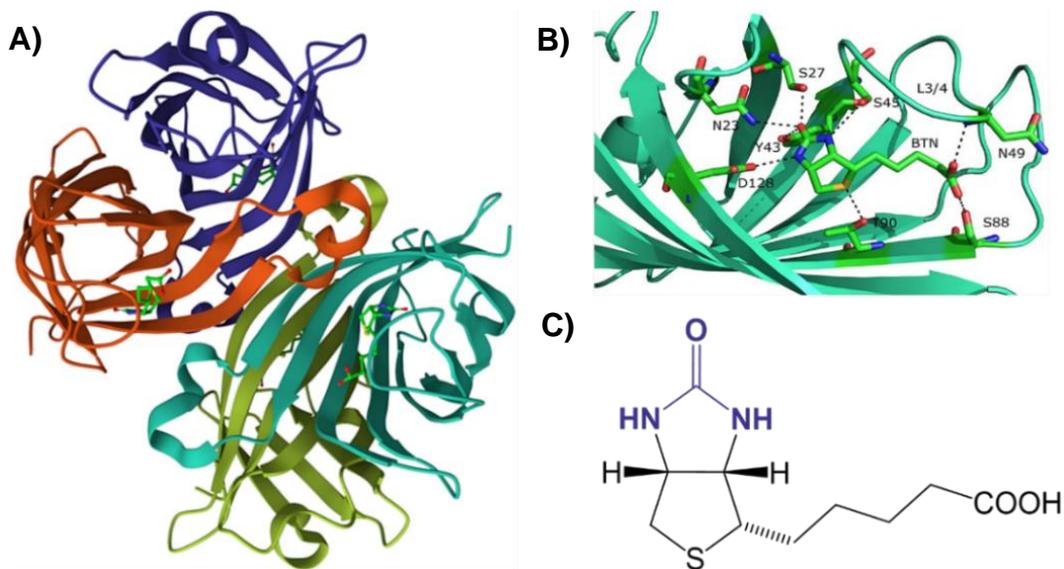


Figure 1.2: Complexe biotine-streptavidine. A) Structure cristallographique du complexe biotine-streptavidine (Wilhelm *et al.*, 1989). B) Les différentes interactions formées entre la biotine et la streptavidine (Liu, Zhang & Mei, 2016). C) Structure moléculaire de la biotine. Le groupement en bleu correspond à la fonction uréido ($-\text{NHCONH}-$) de la biotine.

1.2.3 Protéomique ascendante

Les analyses protéomiques basées sur la spectrométrie de masse reposent principalement sur deux approches: la protéomique de type descendante (*top-down*) et la protéomique de type ascendante (*bottom-up*) (Duong & Lee, 2023). L'approche *top-down* consiste à analyser directement les protéines sous leur forme intacte. En ce qui concerne l'approche *bottom-up*, les protéines sont identifiées à partir des peptides générés lors d'une étape de digestion enzymatique (Duong & Lee, 2023). Alors que l'approche *top-down* favorise l'identification d'isoformes et de protéoformes par une couverture de séquence généralement plus complète des protéines, cette approche se limite à l'analyse de protéines de faible poids moléculaire (<50 kDa) (Rogers & Bomgarden, 2016). De plus, les protéines étant des molécules aux propriétés physico-chimiques plus diversifiées que les peptides, elles sont alors plus difficiles à séparer et à identifier par LC-MS/MS (Gregorich, Chang & Ge, 2014; Rogers & Bomgarden, 2016).

L'approche *bottom-up* est la stratégie la plus couramment employée pour l'identification et la quantification en protéomique (Angel *et al.*, 2012). De manière générale, cette approche implique une série d'étapes préalables à la digestion enzymatique des protéines incluant la solubilisation et la dénaturation, ainsi que la réduction et l'alkylation. En effet, puisque la structure tridimensionnelle des protéines peut restreindre l'accès des enzymes de digestion à leurs sites de

clivage, il est alors important d'induire leur dénaturation afin de favoriser la formation de peptides (Jiang *et al.*, 2024). Les protéines peuvent ainsi être dénaturées à partir de différents d'agents chaotropiques tels que l'urée et la thiourée (Duong & Lee, 2023). L'ajout d'agents réducteurs tels que le dithiothreitol (DTT) ou le tris(2-carboxyethyl)phosphine (TCEP) permet quant à eux la réduction des ponts disulfures. En vue de prévenir leurs réoxydations, les thiols libres sont alors alkylés en utilisant par exemple l'iodoacétamide (IAM) (Duong & Lee, 2023; Jiang *et al.*, 2024).

L'étape de digestion enzymatique des protéines est une étape cruciale en protéomique *bottom-up*, responsable de générer les peptides nécessaires à l'identification des protéines et ainsi maximiser leur couverture de séquence. Pour être facilement analysés par spectrométrie de masse, les peptides ont idéalement une longueur comprise entre 7-35 acides aminés (Jiang *et al.*, 2024). En effet, les peptides trop longs sont difficiles à analyser par LC-MS/MS et sont généralement perdus lors de l'étape d'extraction sur phase solide (SPE), et ce en raison de leur interaction trop forte avec la phase stationnaire (Jiang *et al.*, 2024). À l'inverse, les peptides de petite taille sont peu retenus en chromatographie liquide en phase inverse (RPLC) et leur séquence peut difficilement être attribuée aux protéines en raison des homologies de séquence (Rogers & Bomgarden, 2016; Jiang *et al.*, 2024). La trypsine est l'enzyme protéolytique la plus couramment employée, en raison de sa spécificité, son efficacité et son faible coût (Rogers & Bomgarden, 2016; Jiang *et al.*, 2024). Cette enzyme clive plus particulièrement les protéines à l'extrémité C-terminale des lysines et des arginines uniquement si celles-ci ne sont pas précédées par une proline (K/R-X où X≠P) (Jiang *et al.*, 2024). Les peptides tryptiques générés sont idéaux afin d'être identifiés par spectrométrie de masse. En effet, la nature basique de ces deux acides aminés favorise l'ionisation et la fragmentation des peptides par MS/MS (Rogers & Bomgarden, 2016; Jiang *et al.*, 2024). De plus, les peptides ont généralement la taille et l'hydrophobicité nécessaire pour être séparés par chromatographie et analysés par spectrométrie de masse (Jiang *et al.*, 2024).

1.3 Chromatographie liquide couplée à la spectrométrie de masse en tandem

Présenté comme une alternative aux approches d'électrophorèse sur gel et aux approches d'immunoessais, la chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS) s'est taillé une place de choix en protéomique, et ce en raison de son incroyable potentiel d'analyse à haut débit ainsi que de sa haute sensibilité et spécificité.

1.3.1 Chromatographie liquide à haute performance

Malgré les nombreux avantages de l'approche *bottom-up*, celle-ci a pour effet d'augmenter de manière considérable la complexité de la matrice soit en générant des centaines de milliers de peptides, ce qui peut affecter leur identification (Zhang *et al.*, 2010; Gillet, Leitner & Aebersold, 2016). En effet, les spectromètres de masse étant sensibles aux effets de suppression ionique causés entre autres par la co-élution de peptides, une séparation physique de ces molécules favorise alors leur identification (Jiang *et al.*, 2024; Gillet, Leitner & Aebersold, 2016). La chromatographie liquide est une approche de séparation basée sur la vitesse d'élution des analytes à travers une phase stationnaire, et aidé par un gradient d'élution de phase mobile. Cette séparation est plus particulièrement possible grâce aux propriétés physico-chimiques différentielles des molécules, affectant leur affinité pour les différentes phases. Ainsi, les composés possédant une faible affinité pour la phase stationnaire auront un temps de rétention moindre que ceux de plus forte affinité (Allwood & Goodacre, 2010).

Plusieurs facteurs peuvent affecter la séparation des peptides tels que le type de colonne (composition chimique, longueur, taille des particules) ainsi que le gradient (solvants utilisés, débit) (Deracinois *et al.*, 2013). De manière générale, la séparation chromatographique peut être améliorée en utilisant des particules de petites tailles (<5 µm), ce qui permet d'augmenter la surface de contact entre les analytes et la phase stationnaire et ainsi améliorer leur rétention. Toutefois, ceci aura pour effet d'augmenter la résistance d'écoulement de la phase mobile. C'est pourquoi il est alors nécessaire d'appliquer une haute pression afin d'entraîner le solvant au travers de la phase stationnaire. La chromatographie liquide à haute performance (HPLC), à savoir l'approche employée ici, combine l'utilisation de pompes à haute pression et de colonnes chromatographiques renforcées afin de permettre ce type d'application (Allwood & Goodacre, 2010).

Selon le type de phase stationnaire et de phase mobile utilisée, différents modes de séparation chromatographiques sont possibles tels que la chromatographie par échangeurs ioniques, la chromatographie par exclusion stérique ou encore la chromatographie basée sur l'hydrophobicité (phase inverse) (Deracinois *et al.*, 2013). La méthode plus couramment employée en protéomique est la chromatographie liquide en phase inverse (RPLC), et ce en raison de la compatibilité des phases mobiles pour l'analyse en spectrométrie de masse (Shi *et al.*, 2004). Le principe général de la RPLC repose sur la séparation des protéines ou des peptides en fonction de leur hydrophobicité. Cette approche se base plus particulièrement sur la répartition des composés

entre une phase stationnaire hydrophobe, composée généralement de particules de silice C₁₈, et une phase mobile organique aqueuse (Deracinois *et al.*, 2013). Dans ce cas-ci, les composés apolaires auront une meilleure rétention que les composés polaires, nécessitant alors un gradient de polarité décroissante afin d'induire leur élution (Deracinois *et al.*, 2013). En ce qui concerne la phase mobile, celle-ci se compose essentiellement d'eau, d'un solvant organique miscible tel que l'acétonitrile et d'un acide volatil tel l'acide formique (Shi *et al.*, 2004). L'ajout de ce dernier permet d'une part de dénaturer les protéines ou les peptides, d'introduire une charge positive dans la molécule et de minimiser la formation d'interactions ioniques non spécifiques avec la phase stationnaire (Shi *et al.*, 2004).

1.3.2 Spectrométrie de masse en tandem à haute résolution

De manière générale, la spectrométrie de masse (MS) est une méthode analytique qui se démarque par sa puissance et sa polyvalence. En protéomique, cette approche est non seulement idéale pour l'identification et la quantification des protéines, mais elle est également bien adaptée pour la caractérisation de leurs modifications et de leurs interactions protéine-protéine (Sinha & Mann, 2020). De plus, la haute spécificité et sensibilité des instruments font de la spectrométrie de masse un outil de choix pour l'analyse de matrices biologiques complexes telles que le plasma (Zhao *et al.*, 2023).

Initialement introduite en 1910 par J.J. Thomson (McLafferty, 2011), et étant en constante évolution, la spectrométrie de masse permet l'analyse de molécules en fonction de leur rapport masse sur charge (m/z). Bien qu'il existe une multitude d'instrument, un spectromètre de masse se compose essentiellement de trois éléments : (I) d'une source d'ionisation qui permet de produire des ions en phase gazeuse, (II) d'un ou plusieurs analyseurs de masse qui permet de séparer les ions produits en fonction de leur rapport m/z , et (III) d'un détecteur qui permet de mesurer les ions produits (Haag, 2016). La figure 1.3 présente de manière schématisée les principales composantes d'un spectromètre de masse.

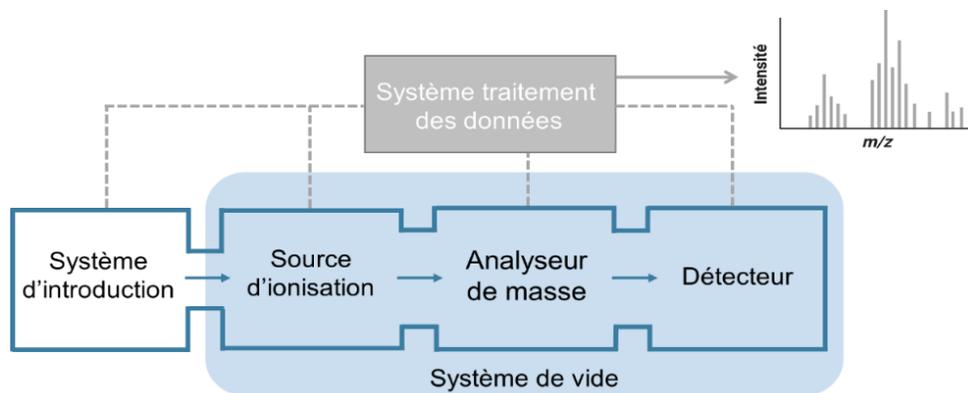


Figure 1.3: Les différentes composantes d'un spectromètre de masse. Figure adaptée de Jürgen H Gross (Jürgen H Gross, 2017)

Il existe plusieurs types de sources d'ionisation incluant l'ionisation par impact électronique, l'ionisation chimique, l'ionisation par désorption/ionisation laser assistée par matrice (MALDI) ou encore l'ionisation par électrospray (ESI). En protéomique, MALDI et ESI sont les sources les plus fréquemment employées (Karpievitch *et al.*, 2010). En effet, les deux sont des sources d'ionisation dite «douces», c'est-à-dire qu'elles vont limiter la fragmentation des molécules lors de l'étape d'ionisation et ainsi préserver la structure des protéines/peptides (Gregorich, Chang & Ge, 2014), soit un aspect important pour la caractérisation et l'identification subséquentes des molécules. De plus, contrairement au MALDI, l'ESI peut facilement être couplé à la chromatographie liquide (Coufalíková *et al.*, 2017) soit un aspect intéressant en protéomique ascendante où comme mentionné, une séparation préalable des peptides permet de favoriser leur identification.

L'ionisation est une étape importante en spectrométrie de masse, principalement lorsque celle-ci est couplée à la chromatographie liquide. En effet, les analyseurs opérants sous vide et à partir d'un champ électromagnétique, il est alors essentiel de convertir les peptides issus de la phase liquide en ions gazeux, soit en procédant à leur désolvatation (Walther & Mann, 2010, Sinha & Mann, 2020; Graves & Haystead, 2002). En ce qui concerne l'ionisation par ESI, soit le type de source utilisée dans cette étude, l'éluant provenant de la chromatographie liquide est directement injecté au travers soit d'une fine aiguille ou d'un microcapillaire maintenu à un haut potentiel. Le champ électrique produit permet alors d'induire la dissociation et la dispersion du solvant, résultant en la formation d'un aérosol composé de fines gouttelettes chargées. L'évaporation étant également assistée par la présence d'un gaz de séchage (N_2), les gouttelettes diminuent alors graduellement en taille causant ainsi une augmentation de la densité de charge à leur surface. Ce processus résulte en la désintégration des gouttelettes et en la libération d'ions gazeux soit

par un phénomène appelé «explosion coulombienne». Pour finir, les ions seront dirigés vers l'analyseur de masse par l'entremise du gaz N_2 , et ce afin d'être séparés et analyser en fonction de leur m/z (Figure 1.4) (Banerjee & Mazumdar, 2012; Ho *et al.*, 2003).

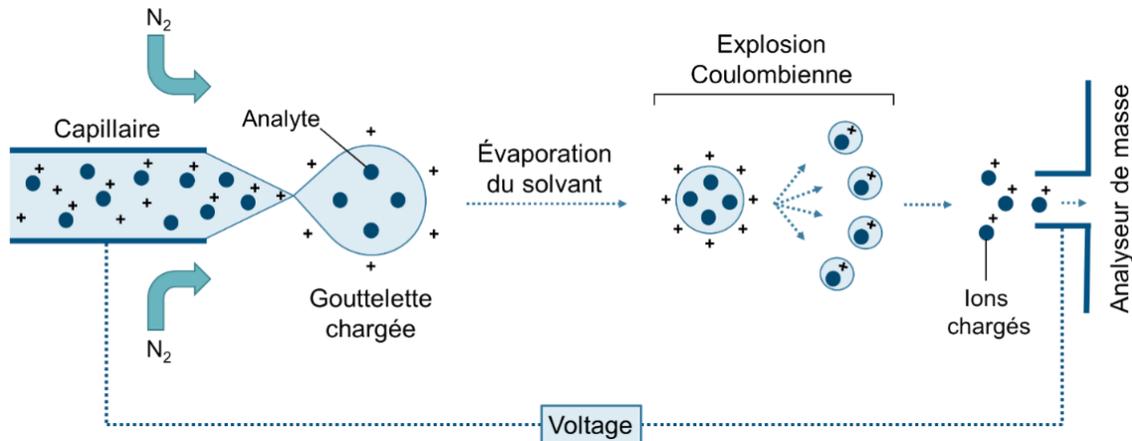


Figure 1.4: Principe de l'ionisation par électrospray (ESI). Figure adaptée de Banerjee & Mazumdar (Banerjee & Mazumdar, 2012)

L'analyseur est une composante essentielle du spectromètre de masse, soit en étant responsable de filtrer et de déterminer le rapport m/z des ions produits. Il existe une multitude d'analyseurs de masse tels que les quadripôles (Q), les trappes ioniques, les temps de vol (TOF) et les OrbiTrap, ainsi que plusieurs variantes hybrides telles que les triples-quadripôles (QqQ) et les quadripôle-temps de vol (QqTOF) (Haag, 2016; Karpievitch *et al.*, 2010). L'ensemble de ces analyseurs vont entre-autres se distinguer par leur sensibilité, leur résolution et leur précision. La sensibilité se définit comme étant la capacité d'un l'instrument à détecter des composés présents en faible abondance, alors que la résolution correspond à sa capacité à différencier des ions dont les masses exactes sont très proches les unes des autres. La précision se réfère quant à elle au potentiel de l'instrument à obtenir des mesures de masse proches de celles de la valeur théorique (Karpievitch *et al.*, 2010).

Les instruments hybrides ont l'avantage de combiner les propriétés et les caractéristiques de deux analyseurs (Dupree *et al.*, 2020). En protéomique, les instruments de type QqTOF et OrbiTrap sont communément employés, et ce, en raison de leur sensibilité, sélectivité, rapidité et stabilité (Sinha & Mann, 2020, Zhao *et al.*, 2023). Dans le contexte de ce projet, un spectromètre de masse de type QqTOF a été utilisé. Ce type d'instrument combine à la fois le haut potentiel de sélection

de m/z d'un quadripôle, avant l'étape de fragmentation, ainsi que la haute sensibilité, résolution et rapidité d'un TOF (Allen & McWhinney, 2019). Le premier quadripôle (Q1), dont la fonction est de filtrer les ions précurseurs selon leurs masses, permet de sélectionner de manière spécifique les ions à envoyer à la cellule de collision pour la fragmentation. La sélection est basée sur la trajectoire des ions au travers un champs électrique. Les ions qui auront une trajectoire stable seront envoyés à la cellule de collision alors que les ions qui auront une trajectoire instable seront eux neutralisés par l'analyseur (Haag, 2016). La cellule de collision, composée d'un quadripôle (q2), permet d'induire la fragmentation des ions précurseurs sélectionnés en Q1. Lors de cette étape, les ions rentrent en collision avec des molécules de gaz neutre présentes dans la cellule, et ce afin d'induire leur fragmentation par un processus appelé dissociation par collision (CID). Le dernier analyseur, se composant d'un TOF, permet quant à lui l'analyse des ions fragments en fonction de leur vitesse de déplacement au travers d'un tube de vol. Pour se faire, les ions sont préalablement accumulés puis accélérés avant d'être envoyés dans le tube. Les ions ayant alors acquis la même énergie cinétique, il est possible de distinguer leur masse en fonction de leur temps d'arrivé au détecteur. En effet, les ions de faible masse se déplaceront beaucoup plus rapidement, et auront donc un temps de vol plus court que les ions de masse plus élevée (Haag, 2016; Zhao *et al.*, 2023). L'ensemble de ce processus est présenté à la figure 1.5.

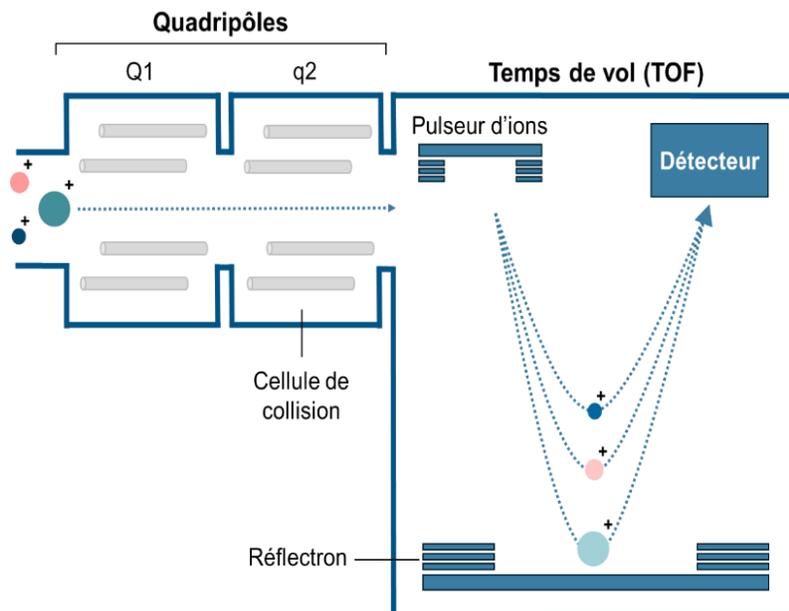


Figure 1.5: Les composantes d'un spectromètre de masse de type quadripôle-temps de vol (QqTOF) et son mode de fonctionnement. Figure adaptée de Allen & McWhinney (Allen & McWhinney, 2019)

1.3.3 Identification des protéines par séquençage

L'identification des protéines par MS/MS est alors possible en passant par le séquençage des peptides. En effet, les peptides ayant des patrons de fragmentation prévisibles, les différents ions produits contiennent alors l'information nécessaire afin de reconstituer la séquence. De plus, en fonction du mode de dissociation utilisé, différents mécanismes de fragmentation sont possibles. En ce qui concerne le mode CID, celui-ci produit majoritairement des ions b et y, soit des ions issus de la fragmentation du lien peptidique. Les ions b vont inclure l'extrémité N-terminale intacte du fragment tandis que, les ions y vont conserver l'extrémité C-terminale. La différence de masse entre deux ions b ou y adjacents peut ainsi être corrélée à la masse d'un acide aminé perdu lors de la fragmentation (Jiang *et al.*, 2024; Graves & Haystead, 2002) (Figure 1.6).

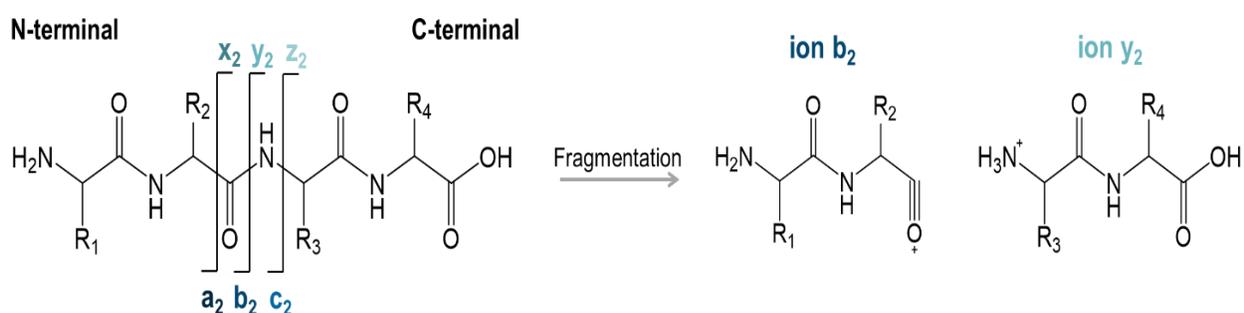


Figure 1.6: Les différents ions produit lors de la fragmentation du lien peptidique

La spectrométrie de masse en tandem est également idéale pour la caractérisation des modifications de protéines. En effet, la présence d'une modification dans la séquence va induire un déplacement de masse par rapport au fragment non modifié (Aebersold & Mann, 2016). De plus, la fragmentation peut également induire la formation de certains ions spécifiques à la modification (Renuse *et al.*, 2020). La figure 1.7B présente un exemple de séquençage pour un peptide biotinylé. Cette modification se produisant sur les résidus lysines, celle-ci induit une différence de masse de 354 Da comparativement à 128 Da pour une lysine non-modifiée. De plus, la biotine peut également se fragmenter afin de produire différents ions associés à la modification (figure 1.7A), dont l'ion dehydrobiotine avec un m/z de 227,08 Da (Renuse *et al.*, 2020).

Les données générées en protéomique par spectrométrie de masse sont très volumineuses et excessivement complexes, nécessitant alors l'utilisation d'outils bio-informatiques automatisés

afin d'en faire l'analyse et l'interprétation. Bien que ces outils aient démontré leur efficacité, de fausses identifications peuvent être attribuées, nécessitant alors une validation statistique afin de s'assurer de la qualité des résultats obtenus. Le taux de fausse découverte (*false discovery rate*, FDR) est une métrique permettant d'établir la confiance globale dans un ensemble de données (Aggarwal & Yadav, 2016).

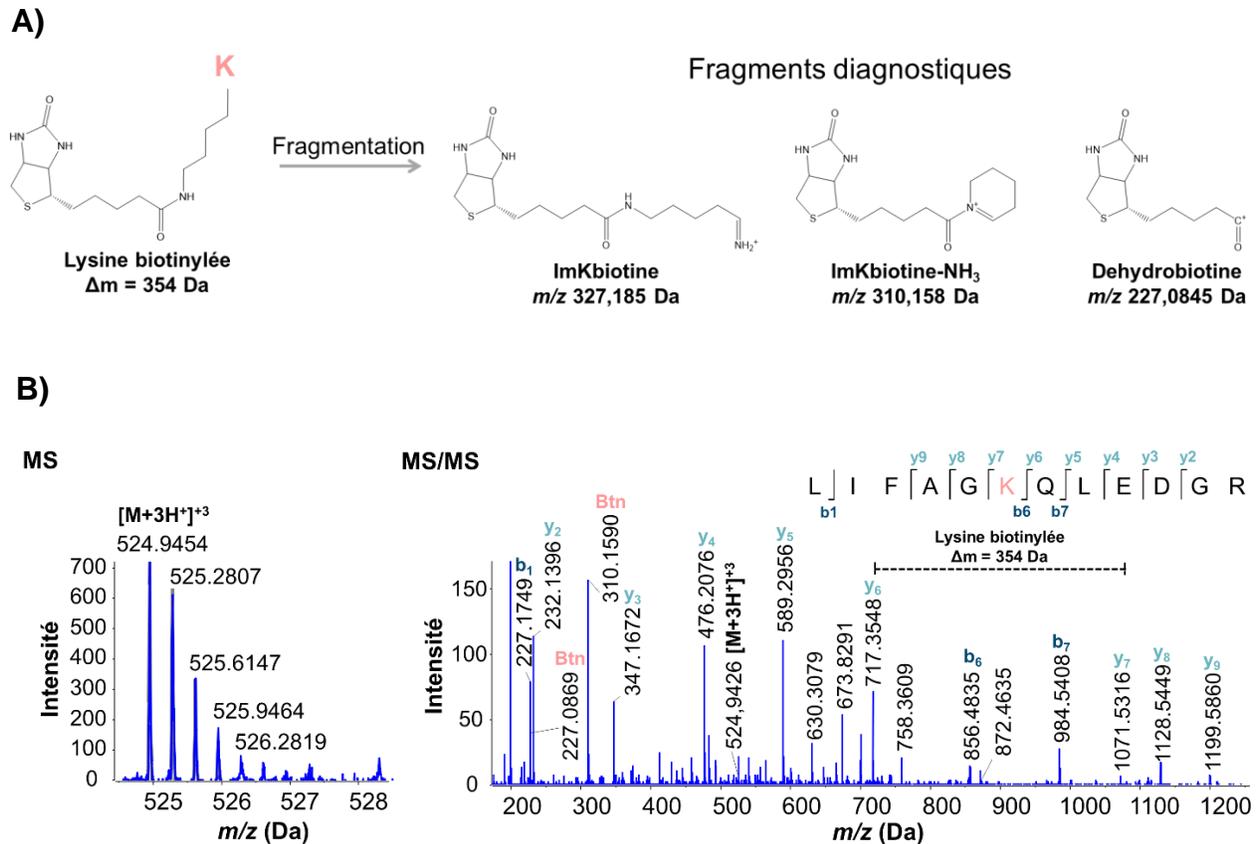


Figure 1.7: Exemple de séquençage d'un peptide biotinylé par spectrométrie de masse en tandem à haute résolution. (A) Les différents ions diagnostiques produits à la suite de la fragmentation d'une lysine biotinylée (Renuse et al., 2020). (B) Exemple de séquençage pour le peptide biotinylé LIFAGKQLEDGR⁺³. Données issues des expériences BiOLD (Chapitre 2).

1.3.4 Les modes d'acquisition des données

Contrairement à d'autres approches analytiques (ELISA, immunoessais), la spectrométrie de masse peut aussi bien être utilisée afin d'effectuer des analyses protéomiques ciblées que des analyses protéomiques non ciblées (Wenk *et al.*, 2024). Les analyses non ciblées ont pour but

d'identifier et/ou quantifier de façon relative un maximum de protéines dans un échantillon biologique donné (Sobsey *et al.*, 2020; Wenk *et al.*, 2024), et ce sans aucune connaissance préalable de sa composition moléculaire. C'est pourquoi, ce type d'analyse fait de la spectrométrie de masse un excellent outil afin d'identifier de potentiels nouveaux biomarqueurs (Wenk *et al.*, 2024). Les approches ciblées, qui nécessitent une connaissance initiale de la composition en protéines, sont idéales pour la validation des potentiels biomarqueurs puisqu'elles sont plus sélectives et sensibles (Wenk *et al.*, 2024).

En protéomique non ciblée, il est possible de distinguer deux modes d'acquisition des données : le mode d'acquisition dépendante des données (DDA) et le mode d'acquisition indépendante des données (DIA). La figure 1.8 montre une comparaison entre les deux modes d'acquisition. En ce qui concerne le mode DDA, seul un nombre prédéfini d'ions précurseurs sera sélectionné lors de l'analyse TOF-MS pour être fragmenté et produire un spectre MS/MS, soit généralement les ions les plus intenses et ceux qui changent dans le temps (confirmant des pics chromatographiques réelles) (Rozanova *et al.*, 2021; Krasny & Huang, 2021). Ce mode d'acquisition induit toutefois un biais lors de l'identification des protéines, soit en favorisant la détection de peptides abondants au détriment de peptides moins abondants (Rozanova *et al.*, 2021). De plus, la sélection étant semi-stochastique, l'identification non constante des ions au travers de répliquas aura pour effet d'affecter la reproductibilité des résultats (Li *et al.*, 2020).

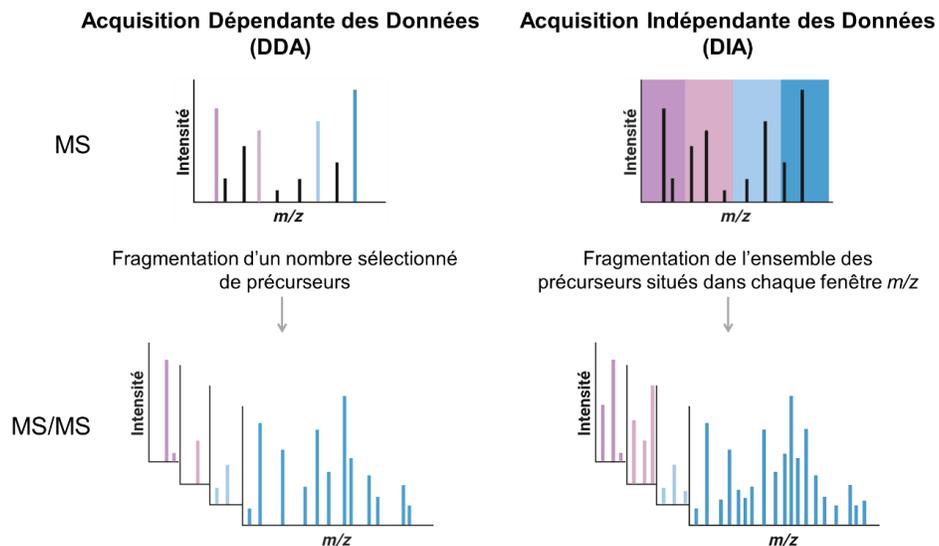


Figure 1.8: Comparaison des modes d'acquisition DDA et DIA pour les analyses non-ciblées. Figure adaptée de Krasny & Huang (Krasny & Huang, 2021).

Une alternative à ce mode est l'acquisition en mode DIA où l'ensemble des ions précurseurs situés dans une fenêtre de m/z est sélectionné et fragmenté afin de produire un spectre MS/MS composé de plusieurs précurseurs (Krasny & Huang, 2021). Bien que cette approche permette une meilleure reproductibilité, quantification et couverture protéomique comparativement au mode DDA, les données générées sont excessivement complexes, rendant leur analyse à partir de bases de données traditionnelles extrêmement difficiles (Rozanova *et al.*, 2021; Li *et al.*, 2020). En effet, la fragmentation simultanée d'un ensemble d'ions précurseurs fait en sorte que l'on perd la relation directe entre les ions précurseurs et leurs ions fragments, nécessitant alors des logiciels d'analyse plus sophistiqués (Rozanova *et al.*, 2021) et une librairie spectrale afin d'effectuer les recherches. Les données DIA peuvent être analysées de deux façons: (I) à partir d'une librairie spectrale générée à partir des données DDA ou bien (II) à partir d'une librairie générée *in silico* (Wen *et al.*, 2023). Les bibliothèques spectrales DDA ont l'avantage de contenir plusieurs informations essentielles à l'identification des protéines en DIA, soit le temps de rétention observé pour chaque peptide ainsi que les fragments identifiés (Li *et al.*, 2020). Toutefois, l'une des principales limitations de cette approche est que l'identification des protéines en DIA sera limitée au nombre de protéines préalablement identifiées en DDA (Rozanova *et al.*, 2021). L'approche *in silico* utilise une base de données de séquences en protéines (fichier fasta) et un logiciel spécialisé tel que DIA-NN afin de générer une librairie spectrale contenant les spectres MS/MS théoriques des peptides ainsi que leur temps de rétention théoriques dans certains cas (Wen *et al.*, 2023; Demichev *et al.*, 2020). L'identification n'étant plus limitée au nombre de protéines préalablement identifiées, celle-ci permet alors une meilleure couverture protéomique. Toutefois, les temps de rétention théoriques pouvant diverger des temps «réels», soit mesurés par les instruments, cette approche peut alors potentiellement mener à de fausses identifications.

Il existe également plusieurs modes d'acquisition des données pour la protéomique ciblée. Le mode surveillance de réactions multiples (MRM) est un mode fréquemment utilisé pour la quantification relative ou absolue des protéines (Kulyyassov *et al.*, 2021). Ce type d'analyse est généralement menée à l'aide d'un spectromètre de masse de type triple quadripôle (QqQ) et repose sur la quantification de peptides issus d'une protéine d'intérêt (Meng & Veenstra, 2011). Ainsi, contrairement aux modes DDA et DIA où un ensemble de peptides, issus d'une multitude de protéines, est analysé, le mode MRM consiste à sélectionner spécifiquement des transitions associées à un peptide, c'est-à-dire le m/z d'un ion précurseur et le m/z d'un ion fragment associées (Borràs & Sabidó, 2017). Pour ce faire, le premier quadripôle (Q1) sélectionne spécifiquement l'ion précurseur à envoyer à la cellule de collision (q2) pour être fragmenté. Le

deuxième analyseur (Q3) permet quant à lui de sélectionner l'ion fragment d'intérêt à analyser (Figure 1.9) (Ryu & Thomas, 2021). La quantification du peptide d'intérêt est alors possible en mesurant l'aire sous la courbe de la transition associée (Macklin, Khan & Kislinger, 2020). Il est important de mentionner ici que ce type d'analyse implique de connaître le rapport m/z des ions précurseurs et de leurs fragments (Macklin, Khan & Kislinger, 2020), nécessitant généralement d'effectuer au préalable des analyses non-ciblées de type DDA ou d'utiliser des bases de données telles SRMAtlas afin d'obtenir cette information. De plus, afin d'augmenter la sélectivité et la spécificité de la méthode plusieurs transitions (3 à 5 transitions par peptide) sont généralement utilisées (Meng & Veenstra, 2011).

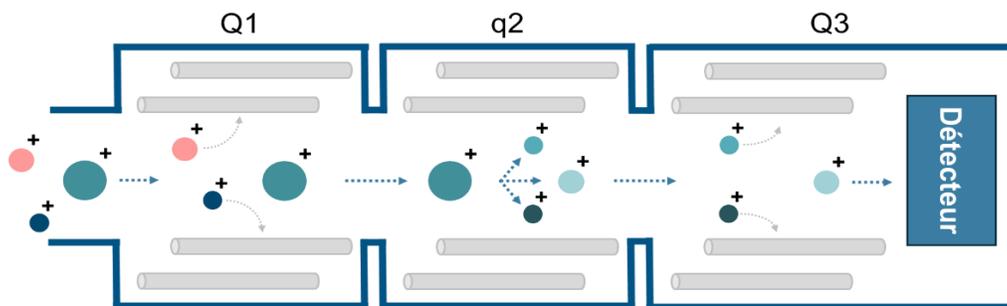


Figure 1.9: Les composantes d'un spectromètre de masse de type triple quadripôle (QqQ) et son mode de fonctionnement. Figure adaptée de Hristova & Svinarov (Hristova & Svinarov, 2022)

1.4 Objectifs

En protéomique, la qualité des résultats générés par LC-MS/MS repose principalement sur une étape de préparation des échantillons efficace et bien adaptée aux objectifs de l'étude. Cette étape pouvant directement affecter l'identification et la quantification des protéines, elle en fait alors une étape prédéterminante, notamment lorsque l'on travaille avec des protéines de faible abondance. Dans ce mémoire, nous nous sommes concentrés sur le développement de méthodes bioanalytiques via trois projets de recherche différents pour l'analyse protéomique par LC-MS/MS.

Dans le chapitre 2 de ce mémoire, nous présenterons l'élaboration d'une stratégie d'étude non ciblée des protéines biotinylées par chromatographie liquide couplée à la spectrométrie de masse en tandem à haute résolution (LC-HRMS/MS). Nous avons initialement comparé deux méthodes

d'enrichissement sur résine streptavidine, l'une ciblant directement les protéines biotinylées et l'autre leurs peptides modifiés. Le principal objectif était d'établir une approche d'enrichissement efficace et spécifique permettant l'analyse de ces protéines dans une matrice biologique complexe (cellules) et ce afin de l'appliquer ultérieurement à des études interactomiques de type BiOLD pour la protéine RNF13, un projet mené en collaboration avec le laboratoire du Professeur Marc Lussier à l'UQAM.

Dans le chapitre 3, nous présenterons le développement d'une méthode ciblée LC-MRM pour la quantification de 17 protéines S100 et 10 annexines (ANX). Pour ce faire, nous avons initialement analysé des protéines exprimées et purifiées avec une approche LC-HRMS/MS non ciblée en mode DDA, à partir d'une plateforme quadripôle-temps de vol. Cette analyse a permis la sélection des peptides et de leurs transitions à optimiser par la suite pour leur détection ciblée par chromatographie liquide couplée à la spectrométrie de masse en mode surveillance de réactions multiples (LC-MRM), sur une plateforme de type triple quadripôle. Le principal objectif était d'établir une méthode de quantification sensible, sélective et robuste pour analyser l'ensemble de ces protéines et de l'appliquer ensuite à l'analyse de plusieurs échantillons de larmes provenant de patients souffrant de diverses pathologies.

Pour le dernier chapitre de ce mémoire (Chapitre 4), nous présenterons la comparaison de différentes méthodes de préparation d'échantillon pour la protéomique plasmatique ascendante non-ciblée par LC-HRMS/MS. L'objectif était d'évaluer l'efficacité de cinq méthodes d'extraction en termes de recouvrement des protéines, de complémentarité et de précision quantitative et ce afin d'établir laquelle ou lesquelles seraient la mieux adaptées afin d'étudier les protéines du plasma dans des échantillons d'intérêts biologiques. Nous avons ainsi comparé une méthode d'enrichissement des vésicules membranaires utilisant des billes magnétiques SAX (MagNet), une méthode de déplétion des protéines abondantes par leur précipitation sélective avec de l'acide perchlorique (PCA), une méthode de solubilisation des protéines plasmatiques au SDS suivie de leur précipitation à l'acétone, avec ou sans déplétion des IgG (PPT+/-IgG), et enfin, une méthode commerciale d'extraction des protéines à l'aide d'une plaque filtrante 96 puits ProTrap.

CHAPITRE 2
**ARTICLE SCIENTIFIQUE : ÉTUDE INTERACTOMIQUE BIOID COUPLÉE À L'ANALYSE
PAR SPECTROMÉTRIE DE MASSE À HAUTE RÉOLUTION**

Carina Lima¹, Audrey Sénécal¹, Marc Lussier¹ et Lekha Sleno¹

¹Université du Québec à Montréal, Département de chimie, Montréal, Canada

Contribution des auteurs : Carina Lima a effectué le développement de la méthode bioanalytique, l'analyse des résultats, la réalisation des tableaux et des figures en plus de rédiger l'article. Audrey Sénécal a participé à la préparation des échantillons cellulaires et la conception de la méthode impliquant la culture des cellules utilisées dans le cadre de ce projet. Marc Lussier a participé à la conception du projet et l'encadrement de la préparation des échantillons cellulaires et la mise au point de la méthode BioID dans les cellules. Lekha Sleno a supervisé le développement de la méthode bioanalytique élaboré dans le cadre de ce projet et a révisé l'article.

2.1 Résumé

Les interactions protéine-protéine étant d'importants régulateurs de la fonction des protéines, la caractérisation de ces interactions est cruciale pour mieux comprendre certains processus biologiques. La méthode BioID est une approche puissante pour étudier les interactions protéine-protéine dans la cellule. Basée sur le marquage des protéines par l'entremise d'une biotine ligase fusionnée à une protéine d'intérêt, les protéines marquées à la biotine peuvent être analysées par spectrométrie de masse (MS) afin d'identifier de potentiels interacteurs de la protéine d'intérêt. Toutefois, en raison de la faible abondance des protéines biotinylées, leur détection par MS dans des matrices complexes reste pratiquement impossible sans une étape d'enrichissement efficace. L'objectif de cette étude était, d'une part, d'optimiser une méthode d'enrichissement permettant l'analyse non-ciblée des protéines biotinylées par chromatographie liquide couplée à la spectrométrie de masse en tandem à haute résolution (LC-HRMS/MS), et, d'autre part, d'appliquer la méthode optimisée pour étudier l'interactome de la protéine RNF13, à savoir une protéine connue pour jouer un rôle dans le développement de l'encéphalopathie développementale et épileptique. Dans cette étude, deux méthodes d'enrichissement sur résine streptavidine ont été testées et comparées : une méthode d'enrichissement des peptides biotinylés et une autre des protéines biotinylées. Dans la première méthode, les protéines sont initialement digérées avant d'effectuer l'enrichissement des peptides biotinylés sur la résine streptavidine. Dans la seconde approche, les protéines sont d'abord enrichies avant d'être directement digérées sur la résine. Dans ce cas, les peptides issus de la digestion n'intégreront pas la modification dans leur séquence; toutefois, par une étape d'élution, il est possible de libérer les peptides modifiés et ainsi obtenir cette information. Le flux de travail basé sur l'enrichissement des protéines, combiné aux modes d'acquisition DDA et DIA-SWATH, s'est avéré être très efficace, permettant d'identifier la modification, d'augmenter la couverture de séquence des protéines identifiées et d'établir leur niveau d'enrichissement. Finalement, à partir de ce flux de travail, il a été possible d'identifier plus de 130 interacteurs potentiels de la protéine RNF13, dont plusieurs sont en processus de validation.

2.2 Introduction

Les interactions protéines-protéines jouent un rôle central dans de nombreux processus physiologiques et pathologiques permettant de réguler par exemple la prolifération cellulaire, la différenciation et l'apoptose (Lu *et al.*, 2020). L'étude de ces interactions est alors essentielle afin de mieux comprendre la fonction des protéines au sein de la cellule (Rao *et al.*, 2014). L'approche BioID (*Proximity-dependent biotin identification*) se base sur le marquage proximal des protéines

par l'entremise d'une biotine ligase fusionnée à une protéine d'intérêt (Cheah & Yamada, 2017). Les protéines qui interagissent avec la protéine d'intérêt ou qui se situent à proximité seront par conséquent biotinylées par la ligase qui y est rattachée. Dans l'approche traditionnelle du BioID, une variante mutée R118G d'une biotine ligase provenant d'*E. coli* (Bir A) est utilisée afin de procéder au marquage des protéines (Cheah & Yamada, 2017; Kubitz *et al.*, 2022). Pour ce faire l'enzyme utilise l'ATP et la biotine afin de produire le biotinyl-AMP (bioAMP) (Cheah & Yamada, 2017; Kubitz *et al.*, 2022). Toutefois, comparativement à la ligase de type sauvage qui elle retient le bioAMP dans son site actif, BioID libère la molécule dans le milieu menant à la biotinylation non enzymatique des lysines composant les protéines situées dans un rayon ~10 nm (Cheah & Yamada, 2017; Kubitz *et al.*, 2022). Bien que cette approche soit un outil puissant afin de caractériser l'interactome d'une protéine d'intérêt, celle-ci présente quelques limitations. Tout d'abord, la taille volumineuse de la ligase (~36 kDa) peut affecter, dans certains cas, la localisation subcellulaire de la protéine de fusion exprimée par la cellule (Kubitz *et al.*, 2022; Kim *et al.*, 2016). De plus, cette ligase possédant une lente cinétique de marquage, celle-ci nécessite alors de longues heures d'exposition à la biotine (Kubitz *et al.*, 2022). Dans ce contexte, différentes variantes de l'approches ont été développées au fil des ans dont l'approche UltraID. L'approche UltraID utilise une variante tronquée et mutée d'une biotine ligase issue de *A. aeolicus* (19,7 kDa) (Kubitz *et al.*, 2022). Cette enzyme présente une cinétique de marquage de l'ordre de quelques minutes et est plus spécifique que les autres approches (Kubitz *et al.*, 2022).

Dans tous les cas, les protéines marquées à la biotine peuvent alors être identifiées par spectrométrie de masse et ce, dans le but d'établir une liste de potentiels interacteurs. Toutefois, l'identification des protéines biotinylées par MS/MS représente un défi important, en raison de leurs faibles abondances dans les matrices biologiques complexes. C'est pourquoi une étape efficace d'enrichissement est alors nécessaire avant leur analyse par LC-MS/MS. En ce qui concerne les protéines biotinylées, différentes stratégies d'enrichissement basées sur l'affinité biotine-streptavidine peuvent être utilisées (Cheah & Yamada, 2017). La figure 2.1 montre de manière schématisée le principe du marquage des protéines par la méthode BioID ainsi qu'un exemple de flux de travail pour l'identification des interacteurs par spectrométrie de masse.

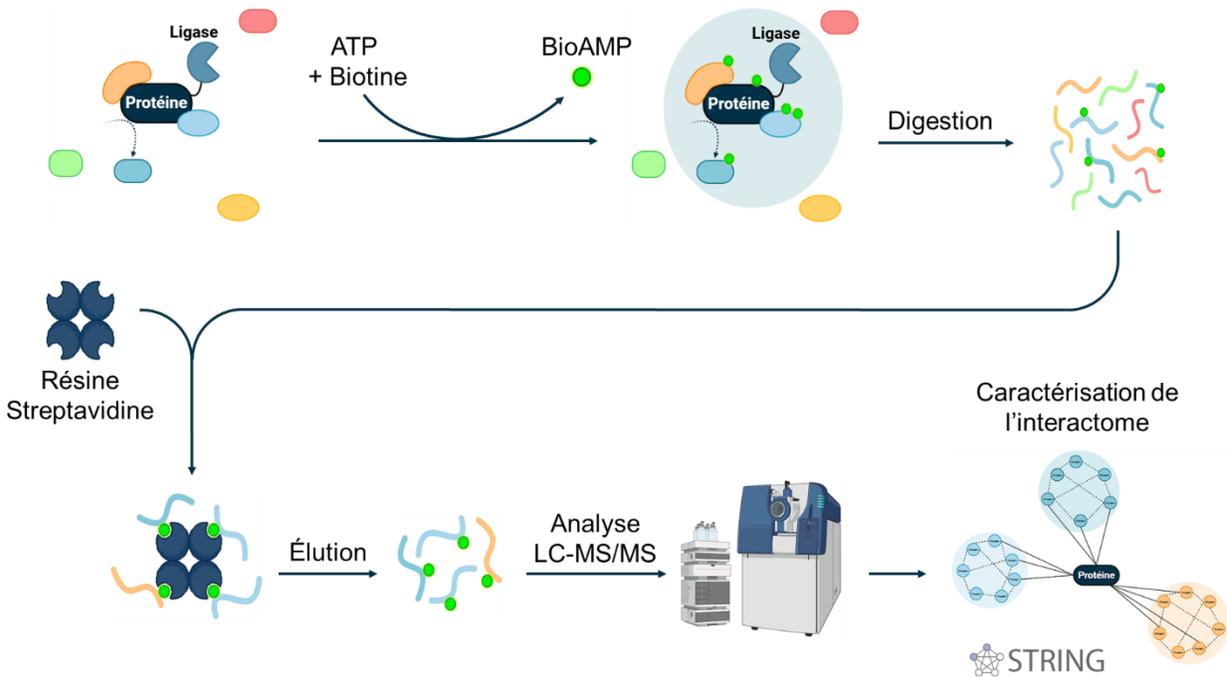


Figure 2.1: Principe du marquage des protéines par BioID et exemple d'un flux de travail pour l'analyse des protéines biotinylées par LC-MS/MS. Une protéine d'intérêt exprimant une biotine ligase biotinyle à partir de BioAMP les protéines situées à proximité. Après la digestion des protéines, les peptides modifiés peuvent être enrichis sur résine streptavidine avant d'être analysés par LC-MS/MS afin d'identifier de potentiels interacteurs de la protéine d'intérêt. Figure adaptée de BioRender (2020).

L'un des principaux objectifs de cette étude a été d'optimiser une méthode d'enrichissement permettant l'analyse des protéines biotinylées par chromatographie liquide couplée à la spectrométrie de masse en tandem à haute résolution (LC-HRMS/MS). Dans ce projet, deux méthodes d'enrichissement sur résine solide avec streptavidine immobilisée ont été testées : la première au niveau des peptides biotinylés et la seconde au niveau des protéines modifiées. Dans la première approche, les protéines sont préalablement digérées, puis les peptides formés sont enrichis sur la résine streptavidine et récupérés pour être analysés par LC-MS/MS. Dans la seconde approche, les protéines biotinylées sont enrichies avant d'être digérées directement sur la résine. Dans ce cas, les peptides issus de la digestion ne vont pas inclure la modification dans leur séquence, car la biotine reste liée à la résine. Toutefois, par une seconde étape d'éluion, il est possible de rompre l'interaction biotine-streptavidine et permettre la libération des peptides biotinylés couplés à la résine, fournissant ainsi cette information. Dans cette étude, nous discuterons des avantages et des limitations liés aux différentes approches testées, ainsi que du processus ayant mené à l'obtention de notre méthode optimisée.

Dans un second temps, une fois l'approche optimale établie, celle-ci a été appliquée afin d'étudier l'interactome associé à la protéine RNF13, dans le cadre d'un projet mené en collaboration avec le laboratoire du Pr. Lussier au département de chimie à l'UQAM. La protéine RNF13 (*Ring Finger Protein 13*) est une ubiquitine E3 ligase transmembranaire retrouvée entre autres au niveau du réticulum endoplasmique, des endosomes et des lysosomes (Cabana *et al.*, 2024). Ayant été associée à la prolifération cellulaire, la migration cellulaire, la tumorigenèse, la myogenèse et à l'autophagie (Cabana *et al.*, 2024), RNF13 serait également impliqué dans le développement de plusieurs maladies dont à l'encéphalopathie développementale et épileptique (Cabana *et al.*, 2021), la sclérose latérale amyotrophique (Khani *et al.*, 2023) et l'athérosclérose (Zhang *et al.*, 2020). Par exemple, une étude a permis de démontrer que les variants L311S et L312P de RNF13, soit deux variants impliqués dans l'encéphalopathie développementale et épileptique, affectent la localisation cellulaire des endosomes (Cabana *et al.*, 2021). Ces variants ont également démontré affecter l'interaction de RNF13 avec les sous-unités AP3D1 et AP3S1 du complexe AP3 (Cabana *et al.*, 2021). En effet, la principale fonction du complexe AP3 (*Adaptor protein complex 3*) est d'assurer le trafic cellulaire des protéines du trans-Golgi et des endosomes vers les endosomes tardifs et les lysosomes (Cabana *et al.*, 2024; Ma *et al.*, 2021). Les auteurs de l'étude ont conclu que le complexe AP3 serait alors important afin de réguler le trafic de RNF13 dans la cellule (Cabana *et al.*, 2021). Par conséquent, étudier l'interactome de RNF13 nous permettrait éventuellement de mieux comprendre sa fonction et les voies de signalisation cellulaire impliquées dans certaines de ces pathologies.

2.3 Matériels et méthodes

2.3.1 Produits chimiques

Le tampon phosphate salin 1X (PBS) provenait de Wisent Inc. (Saint-Jean-Baptiste, QC, Canada), la résine haute capacité Streptavidine Agarose a été obtenue auprès de Thermo Scientific (# produit 20361; capacité 10 mg BSA biotinylée/mL de résine), le bicarbonate d'ammonium (ABC), dithiothréitol (DTT), iodoacétamide (IAM) et la trypsine de grade séquençage du pancréas bovin (# produit T1426), ainsi que le méthanol (MeOH) et l'acétonitrile (ACN) de grade HPLC, l'acide formique (FA) et l'acide trifluoroacétique (TFA) ont tous été obtenus auprès de Sigma-Aldrich (Oakville, ON, Canada). Les cartouche SPE OASIS HLB de 30 mg proviennent de Waters (Milford, MA). L'eau ultra-pure provenait d'un système UV Millipore Synergy (Billerica, MA, USA).

2.3.2 Échantillons cellulaires

2.3.2.1 Culture des cellules Neuro2A

Les échantillons cellulaires nous ont été fournis par le laboratoire du Pr. Lussier du département de chimie à l'UQAM. Des cellules Neuro2A exprimant de manière stable les protéines d'intérêt (RNF13-UltraID ou UltraID seul), ont été cultivées dans un milieu de culture EMEM supplémenté avec 10% FBS. Les cellules ont été conservées à une confluence ~80% dans un incubateur maintenu à 37°C et dans une atmosphère contenant 5% CO₂.

2.3.2.2 Production des culots cellulaires

Environ 2,5x10⁶ cellules par 15 cm ont étéensemencées (6 pétris de 15 cm/lignée stable) et incubées pendant 24h. Les cellules ont été dépourvues de leur milieu de culture, rincées une fois avec PBS et affamées (*starved*) en changeant le milieu pour un milieu EMEM sans FBS. Les cellules ont été incubées pendant 24h. À la moitié des pétris (3x15 cm) ont été ajoutés 50 µM de biotine pour une durée de 24h. Les cellules ont été placées sur glace puis rincées deux fois avec PBS. Un volume de 5 mL de HBSS+Mg/+Cl a été ajouté à chaque pétri. Les cellules ont été grattées avec un grattoir stérile. Les cellules des 3 pétris (3x15 cm avec biotine; 3x15 cm sans biotine) ont été combinées. Les cellules ont été culotées en centrifugeant à 3000 g, pendant 10 min à 4°C. Le surnageant a été éliminé et les échantillons de cellules ont été entreposés à -80°C.

2.3.3 Extraction des protéines

Les cellules culotées (~35 x 10⁶ cellules) ont été resuspendues dans 2 mL de tampon PBS et séparées en aliquots de 1 mL (~17,5 x 10⁶ cellules/aliquot). Les cellules ont été lysées à partir d'un sonicateur Qsonica XL-200 (2 x 10 sec à une puissance de 3) puis centrifugées (5000 rpm, 4°C, 5 min). Finalement, pour les cellules contrôles, des aliquots du surnageant ont été récupérés afin de procéder soit à l'enrichissement des peptides biotinylés (8 x 200 µL) ou à l'enrichissement des protéines biotinylées sur résine streptavidine (2 x 200 µL). Pour les échantillons BioID d'intérêts, 2 aliquots de 980 µL ont été récupérés afin de procéder à l'enrichissement des protéines. L'ensemble de la procédure est décrit à la figure 2.2.

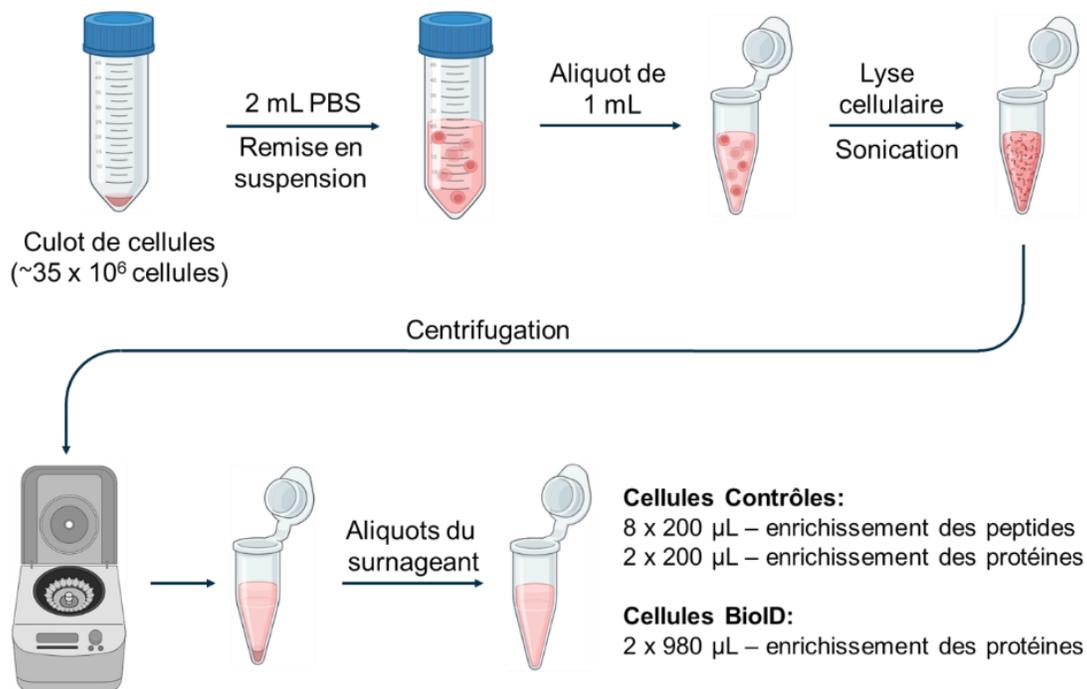


Figure 2.2: Procédure de l'extraction des protéines par une lyse cellulaire. Figure faite à partir de Biorender

2.3.4 Protocole d'enrichissement des peptides biotinylés

À 200 µL d'échantillon ont été ajoutés 400 µL de tampon 100 mM ABC, pH 8,5. Les ponts disulfures ont été réduits en ajoutant 10 µL de DTT 100 mM (37°C, 650 rpm, 20 min) et les thiols libres alkylés en ajoutant 15 µL d'IAM 100 mM (37°C, 650 rpm, 30 min dans le noir), avant l'ajout de la trypsine (10 µg) pour une digestion à 37°C pour 16-18h. Les échantillons de peptides ont été extraits par SPE (*solid phase extraction*) en utilisant des cartouches OASIS HLB de 30 mg (Waters, Milford, MA). Les cartouches ont été conditionnées avec 1 mL MeOH et 1 mL d'eau, à l'aide d'une pompe à vide. Après avoir appliqué les échantillons, les cartouches ont été lavées avec 1 mL d'eau et éluées avec deux aliquots de 500 µL MeOH, suivant une étape d'évaporation à sec au SpeedVac (Fisher Scientific, Mississauga, ON, Canada).

Suivant la digestion et l'extraction sur phase solide (SPE), les peptides ont été reconstitués avec 500 µL de tampon PBS et ajoutés à 50 µL de résine streptavidine (Sigma-Aldrich) préalablement équilibrés deux fois avec 300 µL de tampon PBS. Une incubation a été faite à température pièce, sous agitation, pendant 30 min. Les peptides non-biotinylés ont été écartés par centrifugation (10 000 rpm, 3 min) et les peptides biotinylés couplés à la résine ont été lavés deux fois avec 500 µL de tampon PBS, avec une étape de centrifugation entre chaque lavage. Les peptides

biotinylés couplés à la résine ont été élués deux fois avec 300 μL de tampon d'éluion (80 % ACN; 0,2 % TFA; 0,1 % FA) à 95°C pendant 5 min. Les éluats ont été combinés avant d'être évaporés au Speed Vac.

2.3.5 Protocole d'enrichissement des protéines biotinylées

Une quantité de 100 μL de résine streptavidine (capacité de 10 mg de BSA biotinylé/mL de résine) a d'abord été équilibrée deux fois avec 300 μL de tampon PBS, puis combinée aux échantillons de protéines (200 μL cellules contrôles ou 980 μL BioID). Les réactions ont été incubées à température pièce, sous agitation, pendant 30 min. Les protéines non-biotinylées ont été écartées par centrifugation (10 000 rpm, 3 min) et les protéines biotinylées couplées à la résine ont été lavées deux fois avec 500 μL de tampon PBS, avec une étape de centrifugation entre chaque lavage.

Les protéines biotinylées couplées à la résine ont été resuspendues dans 200 μL de tampon 100 mM ABC, pH 8,5. Les protéines ont été réduites en ajoutant 10 μL de DTT 100 mM (37°C, 650 rpm, 20 min) et alkylées en ajoutant 15 μL d'IAM 100 mM (37°C, 650 rpm, 30 min dans le noir). Pour finir, celles-ci ont été digérées à 37°C, 16h, avec 10 μg de trypsine. Les peptides non-biotinylés ont été récupérés par centrifugation (10 000 rpm, 3 min). La résine a été lavée avec 200 μL de tampon 100 mM ABC, puis centrifugée (10 000 rpm, 3 min). Les éluats ont été combinés avant d'être extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA). Les cartouches ont été conditionnées avec 1 mL MeOH et 1 mL d'eau. Après avoir appliqué les échantillons, les peptides ont été lavés avec 1 mL d'eau et élués deux fois avec 500 μL MeOH. Les échantillons ont finalement été évaporés au Speed Vac.

Les peptides biotinylés couplés à la résine ont été élués deux fois avec 300 μL de tampon d'éluion (80 % ACN; 0,2 % TFA; 0,1 % FA) incubé à 95°C pendant 5 min. Les éluats ont été combinés avant d'être directement évaporés au Speed Vac.

2.3.6 Analyses par LC-HRMS/MS

Les échantillons évaporés ont été reconstitués dans 200 μL d'une solution de 10% ACN; 0,2% FA. Les échantillons ont été injectés (30 μL) sur un système UHPLC Nexera (Shimadzu, Columbia, MD) muni d'une colonne analytique Aeris™ PEPTIDE XB-C18 (1,7 μm , 100 x 2,1 mm) et d'une colonne de garde SecurityGard ULTRA C18. Des phases mobiles composées d'eau (A) et d'ACN (B), contenant chacune 0,1% FA, ont été utilisées. La séparation a été faite à 40°C, à un débit de

0,3 mL/min, en suivant un gradient croissant en ACN. Le gradient a été le suivant : maintenir 5% B pendant 2,5 min, augmenter linéairement jusqu'à 30% B en 37,5 min, jusqu'à 50% B en 2 min puis à 90% B en 2 min, et maintenir à 90% B pendant 3 min. L'acquisition des spectres TOF-MS et MS/MS à haute résolution des peptides a été faite à partir d'un TripleTOF 5600 (quadripôle-temps de vol) (Sciex, Concord, ON) avec ionisation électrospray (ESI) en mode positif. Les paramètres pour la source d'ionisation ont été de 30 psi pour le gaz de rideau (CUR) et de 50 psi pour le gaz nébuliseur (azote) et le gaz de séchage (air sec), 500°C pour la température de la source, 5000 V pour le voltage de la source ESI et 80 V pour le potentiel de dégroupage (DP).

2.3.6.1 Acquisition Dépendante des Données (DDA)

L'identification des protéines a été faite en mode DDA. Pour chaque cycle d'acquisition, une première acquisition TOF-MS a été faite entre m/z 120-1500 avec un temps d'accumulation de 250 ms. Les 15 ions les plus intenses ont ensuite été sélectionnés, selon des critères d'intensité de signal, avec le *dynamic background subtraction* (DBS), pour l'acquisition des spectres MS/MS entre m/z 80-1250. Le temps d'accumulation pour chaque MS/MS a été de 50 ms et une énergie de collision de $30V \pm 10V$ a été appliquée. Le temps de cycle total était de 1,05 s.

2.3.6.2 Acquisition Indépendante des Données (DIA-SWATH)

La quantification relative des protéines biotinyllées a été effectuée par une analyse DIA-SWATH pour les fractions de peptides non-biotinyllés provenant de la méthode d'enrichissement des protéines. Une acquisition TOF-MS (m/z 140-1250) suivie de 100 expériences TOF-MS/MS avec des fenêtres d'acquisition Q1 variables (m/z 80-1500, de 25 ms chacune) a été réalisée. Une énergie de collision de $30V \pm 10V$ a été appliquée avec un temps de cycle total de 2,7 s.

2.3.7 Traitement des données

Les données DDA des fractions PNB (peptides non-biotinyllés) ont été générées afin de créer une librairie spectrale pour les analyses DIA-SWATH. Les données DDA ont tout d'abord été soumises à des recherches avec le logiciel ProteinPilot dans la suite OneOmics (version 3.1, Sciex) à partir de la base de données d'UniProtKB/Swiss-Prot (Mars 2019, 17 165 protéines) des séquences de protéines de souris. La recherche a été effectuée à l'aide de l'algorithme Paragon™ selon les critères suivants : alkylation des cystéines par l'iodoacétamide et digestion enzymatique par la trypsine. Le seuil d'identification des protéines et des peptides a été fixé à 1 % du FDR (*false discovery rate*) global.

À partir de la librairie spectrale DDA des fractions PNB (peptides non-biotinylés), une analyse quantitative DIA-SWATH a été menée. Les recherches ont été effectuées à partir du logiciel OneOmics Suite (Sciex) selon les critères de recherches suivants : protéines et peptides à 1% du FDR global, jusqu'à 4 peptides par protéine et 3 transitions par peptide, avec une fenêtre d'extraction (XIC) de 50 ppm et de 2 min.

Les données DDA des fractions PB (peptides biotinylés) ont été traitées à partir du logiciel ProteinPilot (version 5.0.2) avec une base de données d'UniProtKB/Swiss-Prot (Janvier 2022, 17 263 protéines) de souris. Les critères de recherche ont été les suivants : alkylation des cystéines par l'iodoacétamide et digestion enzymatique par la trypsine. L'algorithme de recherche des protéines a été modifié pour prendre en compte une probabilité plus haute (50%) pour la biotinylation des résidues de lysines.

2.4 Résultats et discussion

2.4.1 Élaboration d'une approche d'enrichissement pour l'étude des protéines biotinylées par LC-HRMS/MS

Pour cette partie de l'étude, les analyses ont été menées sur des cellules Neuro-2a exprimant UltraID et ayant été cultivées en absence ou en présence de FBS (*fetal bovine serum*) et de biotine. Sa composition pouvant affecter l'état de biotinylation des protéines, cette expérience a donc été menée à titre de contrôle. En effet, le FBS est un milieu de culture cellulaire contenant différents nutriments et vitamines, dont la biotine. Ainsi, les différents échantillons ont été soumis aux deux protocoles d'enrichissement sur résine streptavidine décrits précédemment, et ce dans le but d'établir leur efficacité. Dans ce cas précis, nous nous attendons à un nombre de protéines biotinylées plus important pour les échantillons cultivés en présence de FBS et de biotine (+FBS +Biotine), suivis de ceux cultivés en présence de FBS et en absence de biotine (+FBS - Biotine) et pour finir, ceux cultivés en absence de FBS et de biotine (-FBS -Biotine).

2.4.1.1 Identification des protéines par la méthode d'enrichissement des peptides

La méthode d'enrichissement des peptides biotinylés consiste en une digestion des protéines par la trypsine, suivie d'un enrichissement des peptides biotinylés sur résine streptavidine. Une étape de lavage avec du PBS aide à éliminer les interactions non spécifiques à la résine. Les peptides biotinylés couplés à la résine sont par la suite élués avec une solution 80 % ACN, 0,2 % TFA, 0,1 % FA le tout en chauffant à 95°C pendant 5 min. La figure 2.3 présente de manière schématisée l'entièreté de cette approche.

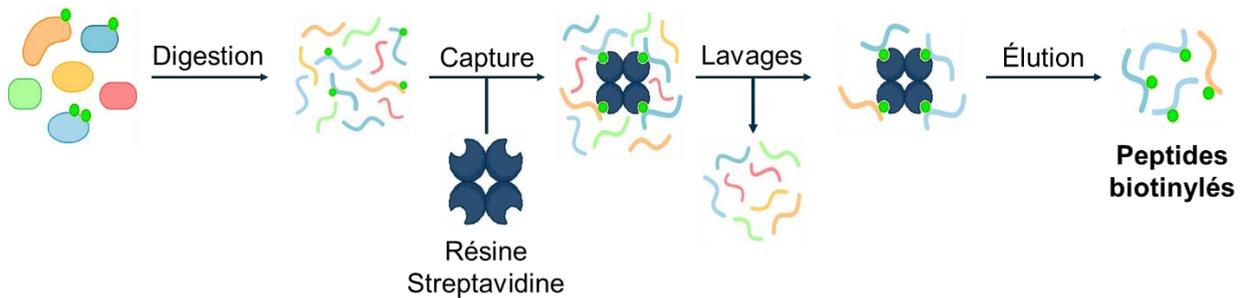


Figure 2.3: Analyse des protéines biotinylées par la méthode d'enrichissement des peptides. Les protéines sont préalablement digérées avant d'enrichir les peptides biotinylés sur résine streptavidine. Les peptides biotinylés couplés à la résine sont élués avant d'être analysés

Les résultats présentés à la figure 2.4 montrent et comparent l'ensemble des protéines identifiées dans les différentes conditions expérimentales pour la méthode d'enrichissement des peptides. Dans cette approche, l'identification des protéines biotinylées reposant uniquement sur l'identification de leurs peptides modifiés, celle-ci semble être l'étape limitante de la méthode. En effet, les résultats montrent qu'une seule protéine biotinylée a été identifiée dans l'échantillon +FBS +Biotine (FA83H) et celui de -FBS -Biotine (MYO98). En ce qui concerne +FBS -Biotine aucune protéine modifiée n'a été identifiée. À partir du spectre MS/MS du peptide biotinylé identifié pour la protéine FA83H, il a été possible d'identifier le fragment du peptide contenant la lysine biotinylée et donc de valider la modification (Figure A1). Pour la protéine MYO98 (Figure A2) aucun fragment contenant la lysine modifiée ou appartenant à la biotine n'a été identifiée, cette identification pouvant ainsi être un faux positif.

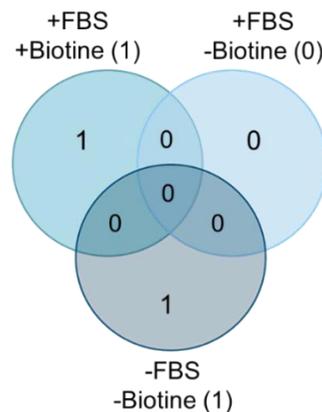


Figure 2.4: Résultats de l'identification des protéines biotinylées par la méthode d'enrichissement des peptides

2.4.1.2 Identification des protéines par la méthode d'enrichissement des protéines

En ce qui concerne l'approche basée sur l'enrichissement des protéines, les protéines biotinylées préalablement enrichies sur résine streptavidine, sont lavées avec PBS puis digérées directement sur la résine. Les peptides sont récupérés par centrifugation afin d'obtenir la fraction peptides non-biotinylés (PNB), soit une fraction composée de peptides non-modifiés mais étant issus de protéines biotinylées. Finalement, les peptides modifiés couplés à la résine sont élués à partir d'une solution 80 % ACN; 0,2 % TFA; 0,1 % FA, et ce afin d'obtenir la fraction peptides biotinylés (PB). Dans cette approche, l'identification des protéines biotinylées repose d'une part sur leur identification dans la fraction PNB et si possible sur les peptides modifiés identifiés dans la fraction PB, et ce afin de valider leur modification. La figure 2.5 présente le flux de travail pour l'enrichissement des protéines.

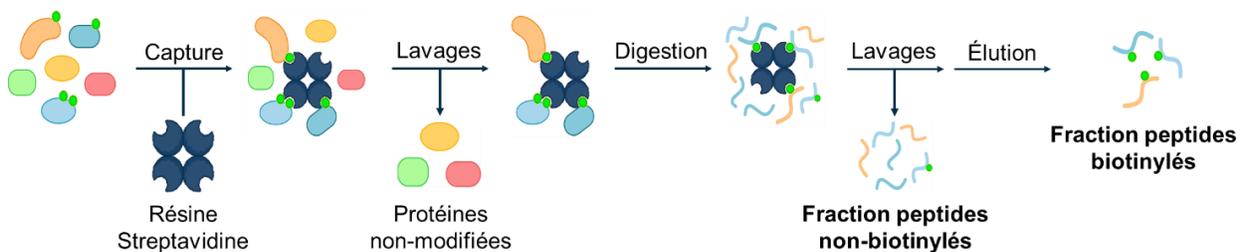


Figure 2.5: Analyse des protéines biotinylées par la méthode d'enrichissement des protéines. Les protéines sont préalablement enrichies, lavées puis directement digérées sur la résine. Les peptides non-biotinylés issus de protéines modifiées sont récupérés afin de former la fraction PNB. Les peptides biotinylés couplés à la résine sont élués afin de former la fraction PB. L'identification des protéines repose sur leur identification dans les deux fractions.

Les diagrammes de Venn présentés à la figure 2.6 montrent et comparent le nombre de protéines identifiées dans les fractions PNB et PB pour la méthode d'enrichissement des protéines. Pour la fraction PNB (Figure 2.6A), il est possible d'observer un nombre important de protéines identifiées pour chacune des conditions expérimentales. Celui-ci étant plus élevé pour l'échantillon +FBS +Biotine avec 1500 protéines identifiées et 311 protéines lui étant uniques, suivi de +FBS -Biotine avec 1391 protéines et 191 protéines uniques, et de -FBS -Biotine avec 1098 protéines et 140 protéines uniques, ces résultats permettent de soutenir l'hypothèse de départ, et donc de démontrer le potentiel de la méthode. Toutefois, il est également important de souligner qu'un nombre très important de protéines ont été identifiées dans l'échantillon sans biotine (-FBS -Biotine) et que la majorité de ces protéines se voient être partagées avec les deux autres

conditions (66, 814 et 77 protéines). L'identification de ces protéines pourrait être le résultat de leurs interactions non spécifiques avec la résine, montrant la nécessité d'augmenter l'efficacité de nos étapes de lavage. En ce qui concerne les résultats pour la fraction PB (Figure 2.6B), un nombre plus important de protéines ont encore une fois été identifiées dans l'échantillon +FBS +Biotine (6 protéines) suivi de +FBS -Biotine (2 protéines) et de -FBS-Biotine (1 protéine). Toutefois, l'identification des protéines biotinylées à partir de leurs peptides modifiés semble encore une fois être l'étape limitante.

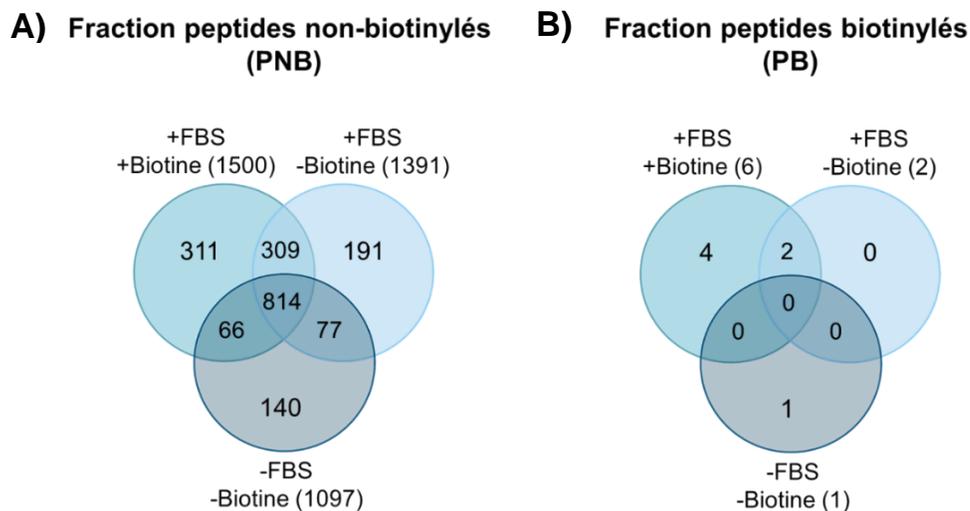


Figure 2.6: Résultats de l'identification des protéines biotinylées par la méthode d'enrichissement des protéines. A) protéines identifiées dans la fraction peptides non-biotinylés et B) dans la fraction peptides biotinylés.

La validation des protéines identifiées dans PNB devant toutefois passer par leur identification dans la fraction PB, il a été possible de valider l'identification de 4 protéines pour +FBS +Biotine (PYC, MCCA, NPM, RAN), 2 protéines pour +FBS -Biotine (PYC, MCCA) et 1 protéine pour -FBS -Biotine (BCORL) (Figure 2.7A-C). En ce qui concerne l'échantillon +FBS +Biotine, il a été possible de confirmer la modification de ces quatre protéines en regardant les spectres MS/MS de leurs peptides modifiés et ce, en identifiant différents fragments associés à la biotine ou contenant la lysine modifiée (Figure A3-A6). Pour les deux protéines uniques à la fraction PB de l'échantillon +FBS +Biotine (Figure 2.7A), leurs modifications n'ont pas pu être validées à partir de leurs spectres MS/MS (Figure A7-A8), celles-ci pouvant être des faux positifs. Pour les deux protéines validées dans l'échantillon +FBS -Biotine, il a également été possible de confirmer leur modification à partir des spectres MS/MS de leur peptide modifié (Figure A9-A10). Pour la protéine

validée dans l'échantillon -FBS -Biotine (BCORL), sa modification n'a pas pu être confirmée à partir du spectre MS/MS de son peptide modifié (Figure A11).

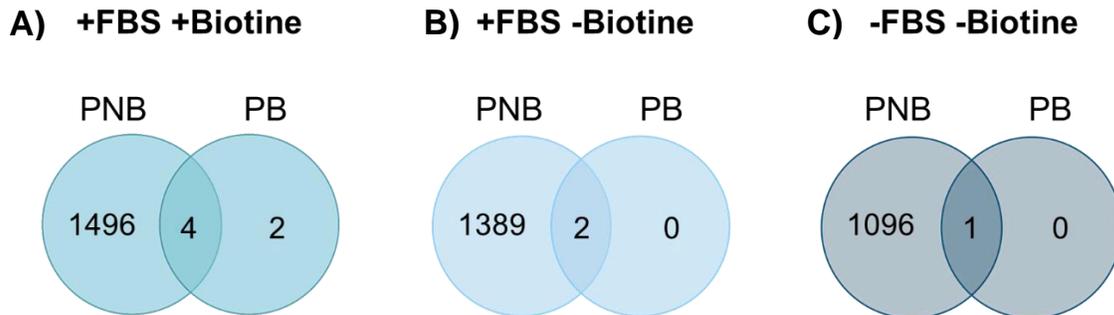


Figure 2.7: Comparaison des protéines identifiées dans la fraction PNB et PB pour la validation des protéines biotinylées. A) +FBS +Biotine, B) +FBS -Biotine et C) -FBS -Biotine. PNB : peptides non-biotinylés. PB : peptides biotinylés

2.4.1.3 Comparaison des deux méthodes d'enrichissement

À partir de cette étude, il a été possible de comparer et d'établir l'efficacité des deux approches d'enrichissement. Le tableau 2.1 résume les avantages et les limitations liées aux différentes méthodes testées. L'une des principales limitations, soit commune aux deux méthodes, réside dans l'identification des protéines biotinylées à partir de leurs peptides modifiés. Ces résultats peuvent entre autres être expliqués par le fait que les peptides sont difficilement élués de la résine, et ce en raison de la forte affinité entre la biotine et la streptavidine ($K_d \approx 10^{-15}M$) (Cirri *et al.*, 2023). Plusieurs approches sont possibles afin de faciliter l'éluion des peptides biotinylés et ainsi prévenir leur perte. De manière générale, de fortes conditions dénaturantes sont nécessaires afin de rompre l'interaction et donc permettre la libération des peptides biotinylés (Cheah & Yamada, 2017). Une alternative à la méthode présentée serait l'éluion des peptides à partir de biotine libre mise en excès, celle-ci permettant de compétitionner avec les peptides modifiés (Cheah & Yamada, 2017, Udeshi *et al.*, 2017).

De plus, le choix du type de résine peut également affecter la libération des peptides modifiés. Ainsi, favoriser l'utilisation de résine d'affinité moindre telles que différentes formes modifiées de l'avidine (avidine monomérique) (Udeshi *et al.*, 2017), de la Streptavidine (Takakura, Sofuku & Tsunashima, 2013) ou de la Tamavidine 2 (Motani & Kosako, 2020), permettrait potentiellement une libération plus efficace des peptides. Par exemple, la Tamavidine est une protéine de liaison

à la biotine d'affinité comparable à celle de la Streptavidine. Toutefois, comparativement à la Streptavidine qui elle forme jusqu'à huit ponts hydrogènes avec la biotine (Liu, Zhang & Mei, 2016), la Tamavidine en forme sept (Takakura, Sofuku & Tsunashima, 2013). Six de ces interactions se font par l'entremise du cycle bicyclique de la biotine, et sont également conservées chez la Streptavidine (Takakura *et al.*, 2009). De plus, contrairement à la Streptavidine qui forme jusqu'à deux ponts hydrogènes avec la fonction carboxylate de la biotine, la Tamavidine en forme seulement un (Takakura *et al.*, 2009). La Tamavidine 2-REV est une protéine d'ingénierie issue de la mutation S36A de la Tamadivine 2. Cette mutation permet de briser l'une des liaisons hydrogènes formées entre la protéine et la fonction uréido de la biotine, ceci permettant une interaction réversible entre les deux molécules (Takakura, Sofuku & Tsunashima, 2013; Motani & Kosako, 2020). La figure 2.8 compare les différentes interactions entre la biotine et les protéines Tamavidine et Streptavidine.

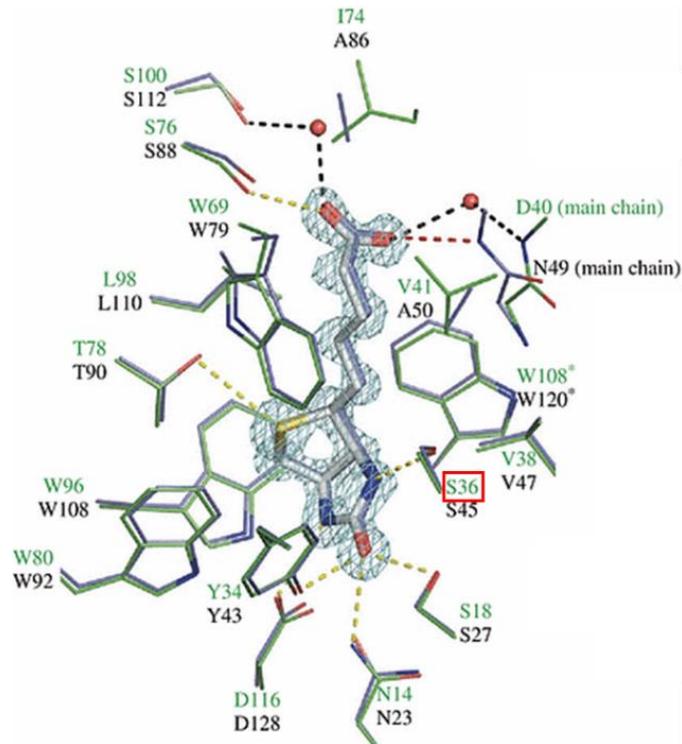


Figure 2.8: Structure cristallographique comparant le complexe Tamadivine-Biotine (structure verte) et le complexe Streptavidine-Biotine (structure bleue). Les interactions en jaunes sont les interactions conservées entre les deux protéines et celle en rouge unique à la Streptavidine. L'encadré rouge montre l'acide aminé muté dans la Tamavidine 2-Rev. Figure adaptée de Takakura *et al.*(Takakura *et al.*, 2009).

D'autres hypothèses peuvent également expliquer le faible recouvrement des peptides biotinylés. Dans le cas de la méthode d'enrichissement des peptides, une augmentation de la complexité du mélange, induite par l'étape de digestion enzymatique, pourrait affecter la liaison des peptides modifiés à la résine. Relativement à la méthode d'enrichissement des protéines, une dégradation de la streptavidine lors de l'étape de digestion sur résine (Motani & Kosako, 2020) pourrait expliquer la perte de certains peptides. Une liaison incomplète des sites biotinylés lors de la capture pourrait également causer la perte de certains peptides modifiés, et donc limiter leur identification. En effet, l'encombrement stérique causé par la conformation tridimensionnelle des protéines ou par la liaison de complexes protéiques à la résine pourrait limiter la liaison de certains sites modifiés à la résine.

De manière générale, la couverture de séquence peut également affecter l'identification des protéines par spectrométrie de masse. L'enrichissement au niveau des peptides nous limite à avoir seulement les peptides ayant les sites modifiés, alors l'identification des protéines est plus difficile. De plus, il est également important de mentionner que certains des peptides biotinylés identifiés peuvent potentiellement appartenir à plus d'une protéine, en raison des homologies de séquence en acides aminés. Il est par conséquent difficile d'attribuer l'identification à une protéine en particulier. Toutefois, en recourant à l'approche basée sur l'enrichissement des protéines, il est possible de palier à ce problème. En effet, la fraction PNB permet une couverture de séquence plus importante des protéines. Il est alors possible de valider l'identification de ces protéines à partir de la fraction PB.

De plus, il est important de mentionner que l'ensemble des résultats présentés dans cette section proviennent uniquement de l'identification des protéines en DDA. En complément à cette étude, il serait également intéressant de mener des analyses quantitatives de type DIA-SWATH, et plus particulièrement à partir de la fraction PNB de la méthode d'enrichissement des protéines. En effet, bien que cette fraction ne nous permette pas d'obtenir d'information sur les sites modifiés d'une protéine, il serait toutefois possible d'établir et comparer le niveau d'enrichissement des protéines extraites entre les différentes conditions, par exemple entre +FBS +Biotine vs -FBS -Biotine. Dans ce cas-ci, une protéine fortement enrichie dans +FBS +Biotine comparativement à -FBS -Biotine pourrait potentiellement avoir des niveaux plus importants de la modification dans sa séquence. Par conséquent, il serait possible de corréliser à partir de ces données le niveau d'enrichissement des protéines à leur état sans devoir nécessairement recourir à leurs peptides modifiés.

Par conséquent, en considérant les différents avantages et limitations liés aux différentes méthodes testées dans cette section, il est possible de conclure à ce stade que l'approche basée sur l'enrichissement des protéines biotinylées semble être la mieux adaptée afin de répondre aux objectifs ultérieurs de l'étude.

Tableau 2.1 : Avantages et limitations liés aux deux méthodes d'enrichissement

Approches	Avantages	Limitations
Enrichissement de peptides biotinylés	<ul style="list-style-type: none"> • Protéines biotinylées directement identifiées à partir des peptides modifiés. 	<ul style="list-style-type: none"> • Les peptides biotinylés sont difficiles à éluer. • Très faible couverture de séquence des protéines identifiées • Les peptides biotinylés identifiés peuvent appartenir à plus d'une protéine (homologies). • La haute complexité du mélange peut affecter la liaison des peptides modifiés à la résine.
Enrichissement de protéines biotinylées	<ul style="list-style-type: none"> • La fraction PNB permet une meilleure couverture de séquence des protéines. • Validation des protéines identifiées en comparant les fractions PNB et PB. • Analyse quantitative statistique du niveau d'enrichissement des protéines à partir de la fraction PNB. 	<ul style="list-style-type: none"> • Interactions non spécifiques avec la résine streptavidine. • La conformation 3D et les complexes de protéines peuvent nuire à l'interaction des protéines avec la résine. • Possible perte des peptides biotinylés (sites non liés à la résine des protéines intactes). • Potentielle dégradation de la streptavidine lors de la digestion sur billes.

PNB : fraction peptides non-biotinylés. PB : fraction peptides biotinylés

2.4.2 Étude interactomique de la protéine RNF13

Une fois la méthode d'analyse des protéines biotinylées testée, celle-ci a été appliquée afin d'étudier l'interactome associé à la protéine RNF13. Dans ce contexte, les analyses ont été menées sur des cellules Neuro-2a exprimant soit la protéine recombinante RNF13-UltraID ou UltraID seul. Les cellules ayant également été cultivées soit en présence ou en absence de biotine, celles-ci ont ainsi été sujettes au protocole d'enrichissement des protéines biotinylées décrit précédemment. De plus, il est important de mentionner que deux expériences et analyses BioID indépendantes ont été menées afin de générer les résultats présentés dans cette section.

Pour cette partie de l'étude, une série de contrôles ont été mis en place afin de consolider l'identification des potentiels interacteurs de RNF13 lors des analyses par LC-MS/MS. L'un des principaux contrôles est une condition dans laquelle les cellules expriment uniquement l'UltraID, et ce afin d'exclure les protéines qui interagissent de manière non spécifique avec la ligase dans la construction RNF13-UltraID (May *et al.*, 2020). De plus, la biotinylation étant une modification observée naturellement chez les protéines, l'absence de biotine dans le milieu de culture permet de distinguer les niveaux basaux de ceux induits par la ligase (Cho *et al.*, 2020).

2.4.2.1 Identification des interacteurs par une analyse en mode acquisition dépendante des données (DDA)

L'identification des interacteurs a tout d'abord été faite par une analyse en mode DDA des fractions PNB et PB. Ce mode d'acquisition des données, permettant de caractériser la composition générale en protéines, celui-ci a ainsi été utilisé afin d'identifier les protéines dans chacun des échantillons suivants : (1) RNF13-UltraID +Biotine; (2) RNF13-UltraID -Biotine (3) UltraID +Biotine et (4) UltraID -Biotine. Le diagramme de Venn présenté à la figure 2.9 compare l'ensemble des protéines identifiées par LC-MS/MS suivant la digestion des protéines retenues sur la résine des différents échantillons. Parmi les 1297 protéines identifiées dans RNF13-UltraID +Biotine, 131 ont été uniquement identifiées dans cet échantillon. En comparant ces protéines à celles identifiées dans la fraction PB, il a ainsi été possible de valider l'identification de 3 protéines biotinylées (ARF2, RS27A et une autre protéine X qui reste confidentielle en raison d'une publication éventuelle à venir) et donc d'identifier 3 potentiels interacteurs à la protéine RNF13.

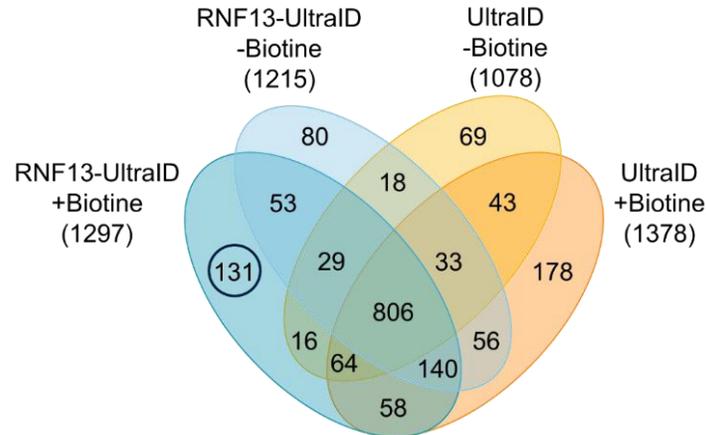


Figure 2.9: Diagramme de Venn comparant les protéines identifiées en DDA

2.4.2.2 Évaluation de l'enrichissement des protéines par une analyse en mode acquisition indépendante des données de type SWATH

Bien que l'identification des protéines par DDA soit une approche somme toute efficace afin d'identifier de potentiels interacteurs, celle-ci présente également quelques limitations. En effet, comme mentionné dans l'introduction, ce type d'acquisition est aléatoire et seul un nombre restreint de protéines sera analysé par spectrométrie de masse, ceci pouvant ainsi mener à des valeurs manquantes lors de l'identification (Li *et al.*, 2020). Par conséquent, une protéine ayant été uniquement identifiée dans RNF13-UltraID +Biotine peut également se retrouver dans l'une des autres conditions. De plus, ce mode d'acquisition se basant uniquement sur l'identification, une protéine biotinyllée qui a été identifiée dans RNF13-UltraID +Biotine, mais qui a également été identifiée dans l'une des autres conditions, peut tout de même être un potentiel interacteur de RNF13. En effet, la protéine peut être biotinyllée à l'état basale, mais présenter des niveaux plus élevés de la modification causée par son interaction à la protéine RNF13.

Dans ce contexte, il est possible d'effectuer une analyse complémentaire de type DIA-SWATH à partir de la fraction PNB pour une analyse quantitative afin de comparer les différents échantillons. Il est alors possible de mener des analyses statistiques afin d'établir le niveau d'enrichissement des protéines. Dans cette étude, deux analyses DIA-SWATH ont été réalisées afin de comparer le niveau d'enrichissement des protéines entre la condition RNF13-UltraID +Biotine et le contrôle UltraID +Biotine. Ainsi, les protéines dont le niveau était statistiquement plus haut dans l'échantillon RNF13-UltraID +Biotine ($p\text{-value} < 0,05$ et un signal relatif $\geq 1,5$ comparer au contrôle) ont été considérées comme étant significativement enrichies dans RNF13-UltraID +Biotine. Au total, 51 protéines ont été filtrées de cette manière lors de la première analyse DIA-SWATH et 32

lors de la seconde, pour un total de 77 protéines d'intérêts (Figure 2.10A). De ces protéines, 6 ont été quantifiées lors des deux analyses DIA-SWATH à savoir les protéines X (confidentielle), TRIPC, F10A1, UFSP2, RM16 et ODR4. Il faut savoir que la protéine X a été identifiée comme faisant partie des 131 protéines uniquement identifiées dans RNF13-UltraID+Biotine lors des analyses DDA et qu'il a également été de confirmer sa modification à partir de la fraction PB. Les résultats quantitatifs présentés dans cette section permettent ainsi de consolider l'identification.

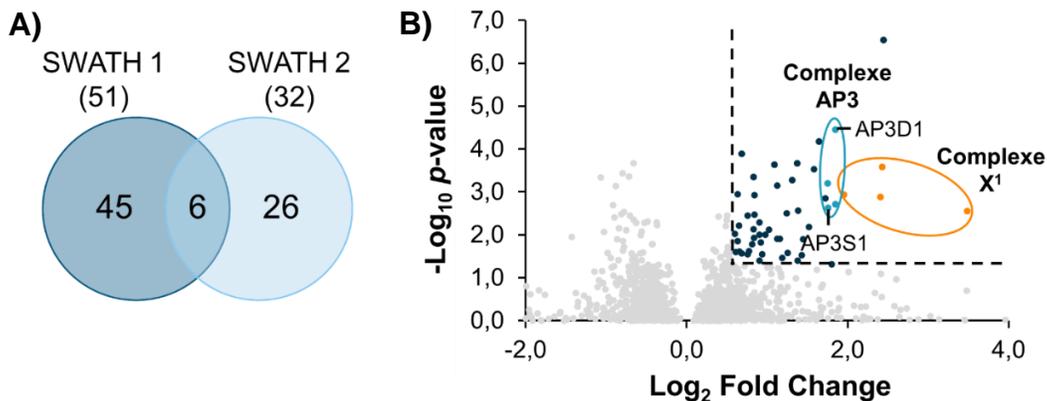


Figure 2.10: Analyse quantitative de l'enrichissement des protéines dans RNF13-UltraID +Biotine. A) Diagramme de Venn comparant les protéines identifiées comme étant statiquement enrichies entre les deux analyses DIA-SWATH. B) Exemple d'un graphique «volcan» montrant la distribution des protéines enrichies lors de l'analyse SWATH 1. Les protéines avec un $p\text{-value} < 0,05$ et un signal relatif $\geq 1,5$ ont été considérées comme étant enrichies dans RNF13-UltraID+Biotine comparativement à UltraID+Biotine. ¹Le nom du complexe X ne peut être divulgué pour des raisons de confidentialité.

Il est également intéressant de mentionner que quatre protéines des 77 quantifiées appartiennent au complexe AP3 (AP3D1, AP3M1, AP3B1 et AP3S1) (Figure 2.10B). Il est pertinent de souligner que les sous-unités AP3D1 et AP3S1 ont été validées dans une étude précédente comme étant des interacteurs de RNF13 (Cabana *et al.*, 2021) démontrant l'efficacité de la méthode choisie. Qui plus est, les protéines AP3B1 et AP3S1 ont également été identifiées comme faisant partie des 131 protéines uniquement identifiées dans RNF13-UltraID+Biotine lors des analyses DDA, toutefois il n'a pas été possible de valider leur modification à partir de la fraction PB.

2.4.2.3 Résultats de l'analyse combinée des flux de travail DDA et DIA-SWATH

En comparant l'ensemble des résultats obtenus en DDA et DIA-SWATH (Figure 2.11), il est alors possible de consolider l'identification des potentiels interacteurs de RNF13. En considérant

l'ensemble des protéines identifiées en DDA pour RNF13-UltraID+Biotine (1297), 62 de ces protéines présentaient un sur-enrichissement en DIA-SWATH. De ces protéines, 4 d'entre elles présentaient également la modification (PYC, F10A1, RRAGA, X1). En considérant uniquement les protéines uniques à RNF13-UltraID+Biotine (131), cinq protéines étaient sur-enrichies par rapport à UltraID, +Biotine (X1, PLCX3, X2, AP3S1, AP3B1). Toutefois, uniquement l'une de ces protéines présentait la modification (X1). Des 126 protéines uniques à RNF13-UltraID+Biotine, mais qui n'ont pas été quantifiées en DIA-SWATH, deux protéines ont également été identifiées dans la fraction PB (RS27A et ARF2).

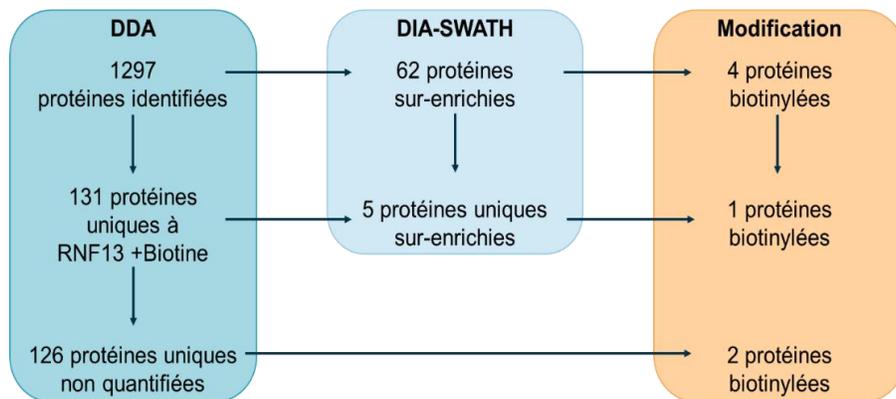


Figure 2.11: Résultats combinés des différentes approches DDA et DIA pour l'analyse des interacteurs de RNF13

2.4.2.4 Analyse STRING des potentiels interacteurs de RNF13

Afin d'identifier de potentielles interactions protéines-protéines entre RNF13 et les protéines identifiées dans cette étude, nous avons effectué une analyse STRING. STRING (<https://string-db.org/>) est une base de données en ligne répertoriant à la fois les interactions protéines-protéines physiques ainsi que les associations fonctionnelles (Szklarczyk *et al.*, 2023). La figure 2.12 présente une analyse interactomique STRING pour les protéines uniques à RNF13-UltraID +Biotine (131) et les protéines sur-enrichies dans RNF13-UltraID +Biotine par DIA-SWATH (77). Cette analyse montre un réseau complexe d'interactions entre les différentes protéines. Il est important de mentionner que BioID ne permet pas d'identifier une interaction protéine-protéine réelle. En effet, en considérant le mécanisme d'action de l'approche, différents types d'interactions peuvent être caractérisés: interactions directes (transitoires ou stables), interactions indirectes et protéines situées à proximité (pas d'interaction directe ou indirecte) (Sears, May & Roux, 2019;

Roux *et al.*, 2012). Dans ce cas précis, la protéine AP3S1, précédemment validée par Cabana *et al.* (Cabana *et al.*, 2021), et la protéine RS27A ont été identifiées comme étant des interacteurs directs de RNF13 dans l'analyse STRING. Bien qu'il n'y ait pas d'évidence directe de l'interaction de RNF13 et RS27A dans la littérature, l'analyse STRING révèle que différents homologues putatifs de la protéine RS27A sont des interacteurs de RNF13 chez différentes espèces. Par exemple UBE2D2, un homologue de RS27A, formerait un complexe avec RNF13 chez l'humain (Sarkar *et al.*, 2019). Sachant ceci, il est fort probable que RS27A puisse interagir dans ce cas-ci avec RNF13, mais ceci reste toutefois à être validé expérimentalement, par exemple avec des expériences de co-immunoprécipitation ou de BRET. En ce qui concerne les protéines RRAGA, PCX, ARF2 ou encore F10A1 (St13), celles-ci pourraient être des interacteurs indirects de RNF13. Par exemple, RRAGA pourrait interagir avec RNF13 par l'entremise de RS27A tout comme ARF2 qui pourrait interagir entre autres via le complexe AP3. Le complexe X pourrait quant à lui interagir avec RNF13 via le complexe AP3.

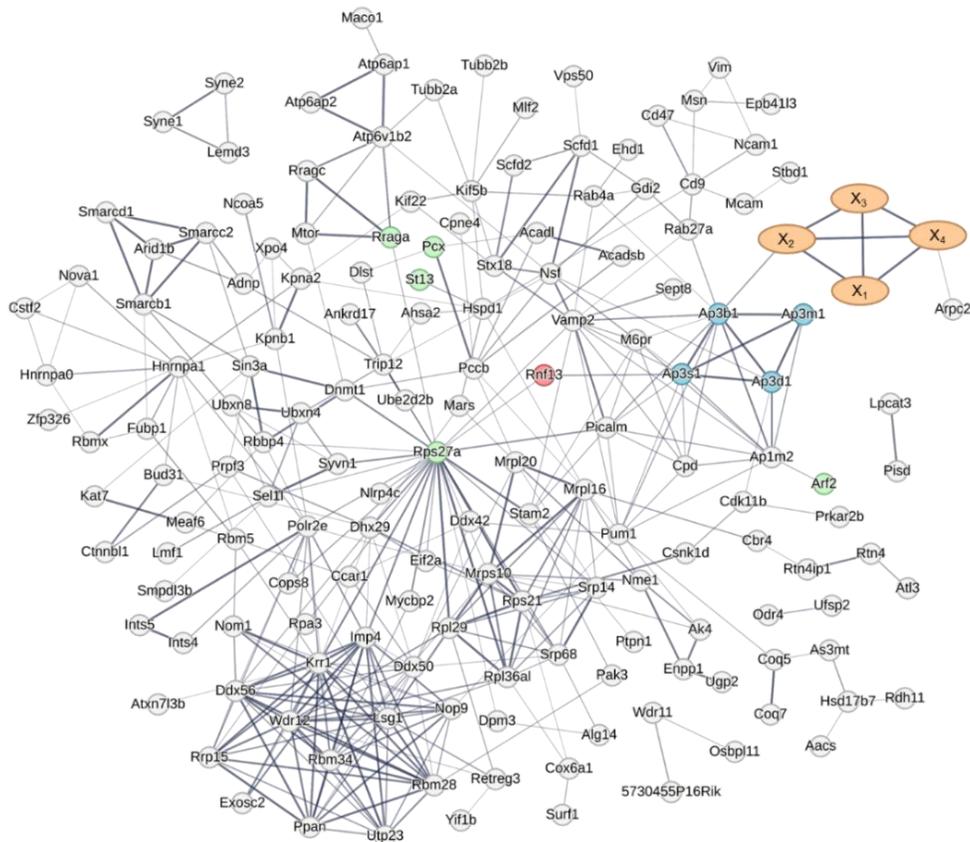


Figure 2.12: Analyse STRING combinant les 131 protéines identifiées en DDA et les 77 protéines quantifiées en DIA-SWATH. Les protéines sans interacteurs ont été éliminées de l'analyse.

En conclusion, différentes méthodes d'analyse des protéines biotinyllées ont été testées et comparées dans cette étude. Le flux de travail basé sur l'enrichissement des protéines biotinyllées, suivi de leur digestion sur résine streptavidine, présente de nombreux avantages, notamment l'augmentation de la couverture de séquence des protéines et la possibilité de détecter le site de modification. De plus, les analyses DDA et DIA-SWATH offrent une grande complémentarité afin de valider et consolider l'identification des potentiels interacteurs de RNF13. Dans cette étude, l'analyse de l'interactome de la protéine RNF13 a permis de révéler différents interacteurs connus de RNF13 (AP3S1 et AP3D1) mais aussi de potentiels nouveaux candidats qui font actuellement l'objet d'études plus approfondies.

De plus, l'une des principales limitations de notre approche réside dans l'identification des protéines moins solubles dont les protéines transmembranaires. En effet, notre approche d'extraction favorise l'extraction des protéines plus solubles au détriment des protéines moins solubles. Ainsi, afin d'identifier ces protéines, il serait intéressant d'ajouter des détergents (par exemple le SDS) afin de favoriser leur solubilisation, mais cela implique nécessairement d'éliminer ces détergents avant l'analyse par LC-MS/MS car ils nuisent à la détection et à la chromatographie des peptides (Jiang *et al.*, 2024; Rogers & Bomgarden, 2016). De plus, sachant que RNF13 est une protéine transmembranaire composant entre autres le réticulum endoplasmique, les endosomes et les lysosomes (Cabana *et al.*, 2021), il serait probablement possible d'identifier d'autres interacteurs situés à la membrane de ces compartiments. Ainsi, en ajoutant une étape d'ajout de détergent durant l'étape d'extraction des protéines et avant leur digestion, il serait possible d'obtenir cette information.

CHAPITRE 3

ARTICLE SCIENTIFIQUE : DÉVELOPPEMENT D'UNE MÉTHODE CIBLÉE PAR LC-MRM POUR L'ANALYSE QUANTITATIVE DES PROTÉINES S100 ET ANNEXINES ET SON APPLICATION DANS L'ANALYSE DE LARMES HUMAINES

Carina Lima^{1,5}, Maggy Lépine^{1,5}, Souleimen Jmii¹, Mélanie Tremblay², Marie-Claude Robert³,
Laurent Cappadoccia^{1,5}, Christopher Rose², Élodie Boisselier^{4,5} et Lekha Sleno^{1,5}

¹Université du Québec à Montréal, Département de chimie, Montréal, Canada

²CR-CHUM, Université de Montréal, département médecine, laboratoire hépato-neuro, Montréal, Canada

³CR-CHUM, Université de Montréal, département Ophtalmologie, Montréal, Canada

⁴Université Laval-CHU de Québec, département d'ophtalmologie et d'oto-rhino-laryngologie, Québec, Canada

⁵Regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, PROTEO, Montréal, Canada

Contributions des auteurs : Carina Lima a effectué la recherche bibliographique, le développement de méthode, l'analyse des résultats, la réalisation des tableaux et des figures en plus de rédiger l'article. Maggy Lépine a participé au développement de méthode et à l'analyse des résultats. Souleimen Jmii a participé à l'expression et la purification de protéines. Mélanie Tremblay a participé à la collecte de larmes. Marie-Claude Robert a participé à la collecte de larmes. Laurent Cappadoccia a encadré les aspects du projet pour l'expression et la purification de protéines. Christopher Rose a participé à la collecte de larmes. Élodie Boisselier a participé à la conception du projet et l'encadrement de l'expression de certaines protéines standards. Lekha Sleno a encadré et supervisé le projet en plus d'avoir révisé l'article.

3.1 Résumé

Les protéines S100 et annexines forment deux importantes classes de protéines permettant de lier le calcium. Intervenant dans différents processus physiologiques, elles sont connues pour jouer un rôle central dans le développement de plusieurs pathologies humaines dont le cancer, les maladies neurodégénératives, les troubles inflammatoires et certaines maladies oculaires. Les larmes offrant une alternative non invasive pour le diagnostic de plusieurs maladies oculaires et systémiques, la quantification précise de ces protéines pourrait ainsi fournir des informations essentielles sur diverses maladies. L'objectif de cette étude était de développer une méthode ciblée LC-MRM pour quantifier de manière sensible et robuste 17 protéines S100 et 10 annexines dans les larmes de patients. L'analyse préalable de protéines standard issues d'expression bactérienne a été réalisée à l'aide d'une approche LC-MS/MS non-ciblée à haute résolution sur une plateforme quadripôle-temps de vol (QqTOF). Cette analyse a permis de dresser une liste initiale en peptides tryptiques uniques à optimiser par la suite en LC-MRM sur une plateforme de type triple quadripôle (QqQ). En fonction des résultats d'optimisation de 2 à 5 peptides par protéine et 2 transitions par peptide ont été sélectionnés pour la méthode finale. La méthode LC-MRM optimisée a finalement été appliquée afin d'analyser les peptides issus d'une digestion à la trypsine des protéines lacrymales provenant de patients atteints soit d'une maladie oculaire, d'épidermolyse bulleuse ou d'une atteinte hépatique. Il a été possible de détecter efficacement la majorité de ces protéines dans les larmes de patients. Qui plus est, plusieurs de ces protéines se voient être fortement augmentées chez le groupe de patient souffrant d'une maladie hépatique.

3.2 Introduction

Les protéines S100 et annexines (ANX) forment deux groupes distincts de protéines liant le calcium (Weisz & Uversky, 2020). La famille des S100, regroupant au total 21 protéines, est impliquée dans la régulation d'une multitude de processus intra- et extracellulaires tels l'apoptose, la prolifération, la différenciation, la migration, le métabolisme énergétique et l'inflammation (Weisz & Uversky, 2020; Gonzalez, Garrie & Turner, 2020). De plus, ces protéines étant connues pour former des complexes dimériques entre elles (homo- et hétérodimérique), cette particularité leur permet de diversifier leurs propriétés fonctionnelles (Weisz & Uversky, 2020). En ce qui concerne la famille des ANX, qui se compose de 12 protéines, celle-ci permet essentiellement d'assurer l'intégrité du réseau membranaire des cellules, dont le trafic membranaire, l'endo- et l'exocytose, les interactions membrane-cytosquelette et la réparation des membranes (Weisz & Uversky, 2020; Lizarbe *et al.*, 2013). Les ANX sont également impliquées dans la régulation d'une

multitude de fonctions extracellulaires telle que l'inflammation, la coagulation et la fibrinolyse (Weisz & Uversky, 2020).

Jouant un rôle primordial au sein de la cellule, plusieurs de ces protéines ont été associées au développement d'une multitude de maladies : cancers, maladies neurodégénératives, troubles inflammatoires, diabète, maladies oculaires, etc (Weisz & Uversky, 2020; Tong *et al.*, 2014; Soria *et al.*, 2013). Par exemple, les protéines S100A8 et A9 sont des marqueurs associés à plusieurs maladies pro-inflammatoires chroniques et aiguës (Bode *et al.*, 2008). L'expression de ces protéines a également été associée au cancer du sein et au cancer de la prostate (Bode *et al.*, 2008). De plus, il est possible de corréliser la multifonctionnalité et la polypathogénicité de ces protéines à leurs capacités à interagir avec différents partenaires (Weisz & Uversky, 2020). Par exemple, plusieurs membres de ces deux familles de protéines sont connus pour interagir ensemble et former différents complexes dimériques (S100A6/ANXA2, S100A6/ANXA6, ANXA6/S100A1) (Weisz & Uversky, 2020). Le tableau 3.1 résume quelques-uns des complexes répertoriés dans la littérature. L'étude de ces protéines pourrait alors améliorer notre compréhension sur une multitude de voies biologiques impliquées dans diverses maladies.

Tableau 3.1: Les différents interactions formées entre les protéines S100 et ANX répertoriés dans la littérature

Protéine	Protéines cibles
S100A1	ANXA5 ¹ ; ANXA6 ^{1,2,3} ; S100P (forme un hétérodimère) ¹⁸
S100A4	ANXA2 ⁴ ; S100A4 (forme un tétramère) ⁵
S100A6	ANXA2 ^{6,7} ; ANXA11 ⁸
S100A8	S100A9 ^{9,10} ; ANXA6 ¹⁰
S100A11	ANXA1 ^{11,12} ; ANXA2 ¹³ ; ANXA6 ¹⁴
S100A13	ANXA2 ¹⁵
S100B	ANXA6 ^{1,2}
S100A10	ANXA2 ^{16,17}

¹Garbuglia, Verzini & Donato, 1998; ²Arcuri *et al.*, 2002; ³Garbuglia *et al.*, 2000; ⁴Semov *et al.*, 2005; ⁵Burkitt *et al.*, 2003; ⁶Zeng, Gerke & Gabius, 1993; ⁷Filipek & Wojda, 1996; ⁸Sudo, & Hidaka, 1998; ⁹Eue *et al.*, 2002; ¹⁰Bode *et al.*, 2008; ¹¹Poeter *et al.*, 2013; ¹²Réty *et al.*, 2000; ¹³Rintala-Dempsey *et al.*, 2006; ¹⁴Chang *et al.*, 2007; ¹⁵Daily & Kumar, 2010; ¹⁶Bharadwaj, Kempster & Waisman, 2021; ¹⁷Rezvanpour, Santamaria-Kisiel, & Shaw, 2011; ¹⁸Wang *et al.*, 2004

La santé oculaire constitue l'un des principaux sujets d'intérêt de notre laboratoire de recherche. Certains travaux menés par notre groupe ont permis de détecter dans les larmes des centaines de protéines, certaines présentant des changements significatifs chez des patients atteints du syndrome de Sjögren ou de la pemphigoïde des membranes muqueuses (Lépine, Zambito & Sleno, 2023; Lépine, Robert & Sleno, 2024; Lépine, 2024). Le tableau 3.2 répertorie l'ensemble des protéines S100 et ANX qui ont été identifiées antérieurement par notre groupe de recherche dans l'analyse de différentes cohortes de patients et de volontaires sains. D'autres groupes ont également associé l'altération de l'expression de certaines de ces protéines au développement de plusieurs autres maladies oculaires, dont au dysfonctionnement des glandes Meibomius, à la sécheresse oculaire ou encore au ptérygion (Tong *et al.*, 2014; Soria *et al.*, 2013).

Tableau 3.2: Protéines S100 et ANX précédemment identifiées par LC-HRMS/MS dans les larmes de patients sains ou quantifiées par LC-MRM chez des patients atteints du syndrome de Sjögren ou de la pemphigoïde des membranes muqueuses

Protéines S100	Patients sains¹	MMP²	SS³	Protéines Annexines	Patients sains¹	MMP²	SS³
S100A2	X	↑		ANXA1	X		
S100A4	X	↑		ANXA2	X		
S100A6	X	↑		ANXA3	X	↑	
S100A8	X		↑	ANXA4	X	↑	
S100A9	X		↑	ANXA5	X	↑	↑
S100A11	X		↑	ANXA7	X		
S100A13	X			ANXA11	X		
S100P	X						

¹Lépine, M., Zambito, O., & Sleno, L., 2023; ²Lépine, M., Robert, M. C., & Sleno, L., 2025; ³Lépine, M., Robert, M. C., & Sleno, L., 2024. X : Identifiée par LC-HRMS/MS; ↑ : Statistiquement augmentée dans la maladie; MMP : Pemphigoïde des membranes muqueuses; SS : Syndrome de Sjögren

Par conséquent, compte tenu de l'importance que tiennent les protéines S100 et ANX à la surface oculaire, la quantification précise de ces protéines dans les larmes pourrait nous fournir de l'information essentielle sur différentes maladies oculaires, mais également sur plusieurs maladies systémiques. Dans ce chapitre, nous présenterons le développement d'une méthode d'analyse ciblée LC-MRM pour la quantification de 17 protéines S100 et 10 protéines ANX. La méthode développée a été appliquée afin d'analyser les larmes de patients atteints d'une variété de pathologies, dont différentes maladies oculaires, d'épidermolyse bulleuse ou encore de différentes maladies hépatiques. L'application de la méthode développée nous a servi de données préliminaires offrant une base pour des études futures en permettant d'examiner chez un groupe varié d'individus la détectabilité de ces protéines dans les larmes humaines.

3.3 Matériels et méthodes

3.3.1 Produits chimiques

La trypsine de qualité séquençage a été obtenue auprès de Promega (produit V5111, Madison, WI, USA) ou de Sigma-Aldrich (produit T1426, Oakville, ON, Canada). Le bicarbonate d'ammonium (ABC), dithiothréitol (DTT), iodoacétamide (IAM), ainsi que le méthanol (MeOH) et l'acétonitrile (ACN) de grade HPLC, l'acide formique (FA) et l'acide trifluoroacétique (TFA) ont tous été obtenus auprès de Sigma-Aldrich (Oakville, ON, Canada). Les cartouche SPE OASIS HLB de 30 mg proviennent de Waters (Milford, MA). L'eau ultra-pure provenait d'un système UV Millipore Synergy (Billerica, MA, USA).

3.3.2 Analyse des protéines S100 et ANX standards par LC-HRMS/MS

3.3.2.1 Expression et purification des protéines S100 et ANX

La protéine S100A10 a été produite dans le laboratoire de Pre. Boisselier à l'Université Laval. L'expression et la purification de S100A10 ont été faite selon la méthode décrite par Yan *et al.* (Yan *et al.*, 2019; Yan *et al.*, 2021). Brièvement, la protéine GST-S100A10 a été surexprimée dans la souche *E. coli* BL21-Codon Plus (DE3)-RIL. Les cellules ont été cultivées dans 1 L de milieu LB contenant de l'ampicilline et incubées à 37°C jusqu'à ce que la densité optique à 600 nm (OD₆₀₀ nm) atteigne 0,8. La surexpression a été initiée en ajoutant 10 mL de 100 mM d'isopropyl-β-D-thiogalactopyranoside (IPTG). Les cultures ont été maintenues à 21°C pendant 16h. Les cellules ont été récoltées par centrifugation (3270 g, 30 min, 4°C). La lyse cellulaire a été réalisée à partir de lysozyme dans du PBS 1X, suivi par 3 cycles de gel-dégel et de sonication. Les cellules ont été centrifugées à 15 000 g, 4°C et pendant 30 min. Le surnageant et le culot ont été mis en

suspension dans le même volume de PBS. Après vérification sur gel 12% SDS-PAGE, le surnageant a été chargé sur deux colonnes GSTrap FF (5 mL) connectées et maintenues à 4 °C pour purifier GST-S100A10. Les colonnes ont été lavées et les fractions éluées ont été collectées. Les éluats ont été analysés sur gel 12% SDS-PAGE. Les fractions contenant GST-S100A10 ont été mélangées ensemble. Pour éliminer l'excès de glutathion, le tampon a été échangé par un tampon basique (Tris 50 mM et NaCl 100 mM, pH 9,5) et une filtration par centrifugation a été effectuée avant de cliver la GST à 4°C. Après 2 heures de clivage, l'échantillon a été chargé sur une colonne GSTrap FF (5 ml) connectée à une colonne GSTrap HP (5 mL) pour purifier S100A10 (11,203 kDa). Plusieurs fractions ont été collectées par le lavage des colonnes. Les échantillons ont été déposés sur gel SDS-PAGE 17%. L'image du gel a été numérisée et analysée par ImageJ pour déterminer la pureté de la protéine S100A10. Les fractions contenant S100A10 avec une pureté ≥ 95 % ont été entreposées à -20°C.

Les protéines ANX (A1, A3, A5, A6, A7, A8, A9, A10 et A11) et les protéines S100 (A2, A3, A4, A5, A6, A7, A11, A13 et A14) ont été produites par le laboratoire du Pr. Cappadocia à l'UQAM. Une réaction en chaîne par polymérase (PCR) a amplifié les ANX (ID Uniprot : P04083, P12429, P08758, P08133, P20073, P13928, O76027, Q9UJ72 et P50995) et les protéines S100 (ID Uniprot : P29034, P33764, P26447, P33763, P06703, P31151, P31949, Q99584, Q9HCY8) à partir d'une bibliothèque d'ADNc générée à partir de cellules HeLa. Les séquences ont été clonées par assemblage Gibson dans le site BamHI d'un vecteur pRSF-Trx. Les constructions ont été validées par séquençage puis transformées dans la souche *E.coli* BL21 DE3 CodonPlus RIL. Les cellules ont été cultivées dans un milieu Super Broth à 37°C jusqu'à atteindre un A_{600} de 1,0. L'expression des protéines a été réalisée à 18°C pendant 24h en utilisant d'IPTG 0,3 mM. Les cellules ont été collectées par centrifugation, mises en suspension dans un tampon de lyse contenant 20 mM de Tris-HCl, pH 8,0, 500 mM de NaCl, 20 mM d'imidazole, pH 8,0, 5 mM de β -mercaptoéthanol et 100 μ g/mL de lysozyme. Les cellules en suspension ont été lysées par congélation et sonication. Après centrifugation à 40 000 g pendant 30 min, le surnageant a été appliqué sur une résine Ni-QIAGEN Ni-NTA. Les protéines ont été éluées avec un tampon d'éluion contenant Tris-HCl 20 mM, pH 8,0, NaCl 500 mM, imidazole 500 mM pH 8,0 et du β -mercaptoéthanol 5 mM.

3.3.2.2 Digestions des protéines standards

À 50 μ g de protéines standards ont été ajoutés jusqu'à 400 μ L de tampon 100 mM ABC, pH 8,5. Les protéines ont été réduites en ajoutant 5 μ L de DTT 100 mM (37°C, 600 rpm, 15 min) et

alkylées en ajoutant 10 μL d'IAM 100 mM (37°C, 600 rpm, 30 min dans le noir). Pour finir, celles-ci ont été digérées à 37°C, 16h, avec 1 μg de trypsine. Les peptides ont été extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA). Les cartouches ont été conditionnées avec 1 mL MeOH et 1 mL d'eau. Après avoir appliqué les échantillons, les cartouches ont été lavées avec 1 mL d'eau et éluées deux fois avec 500 μL MeOH. Les échantillons ont finalement été évaporés au Speed Vac (Fisher Scientific, Mississauga, ON, Canada). Les échantillons ont été reconstitués dans 110 μL d'une solution 10% ACN avec 0,2% d'acide formique, avant l'analyse LC-MS.

3.3.2.3 Analyse LC-HRMS/MS et traitement des données

Les échantillons ont été injectés (25 μL) sur un système UHPLC Nexera (Shimadzu, Columbia, MD) muni d'une colonne Aeris™ PEPTIDE XB-C18 (1,7 μm , 100 x 2,1 mm) et d'une colonne de garde SecurityGuard ULTRA C-18. Une phase mobile composée d'eau (A) et d'ACN (B), toutes deux contenant 0,1% FA, a été utilisée. L'élution a été faite à 40°C, à un débit de 0,3 mL/min et en suivant un gradient croissant en ACN. Le gradient a été le suivant : une étape initiale à 5% B pendant 2,5 min, suivie par une augmentation linéaire jusqu'à 30% B en 37,5 min, jusqu'à 50% B en 2 min puis à 90% B en 2 min, qui a été maintenu pendant 3 min pour bien laver la colonne avant l'étape de rééquilibration de la colonne aux conditions initiales. L'analyse à haute résolution des peptides a été faite à partir d'un TripleTOF 5600 (quadripôle-temps de vol) (Sciex, Concord, ON) en utilisant l'ionisation ESI en mode positif. Les paramètres pour la source d'ionisation ont été de 30 psi pour le gaz rideau (*curtain gaz*), de 50 psi pour le gaz nébuliseur (azote) et le gaz de séchage, 500°C pour la température de la source, 5000 V pour le voltage ESI et 80 V pour le potentiel de dégroupage (DP).

Pour les expériences DDA, une première analyse des ions en TOF-MS a été faite entre m/z 140-1250 avec un temps d'accumulation de 250 ms. Les 15 ions les plus intenses ont ensuite été sélectionnés pour acquérir des spectres MS/MS à haute résolution entre m/z 80-1300, chacun avec un temps d'accumulation de 50 ms et une énergie de collision de $30 \pm 10\text{V}$. Le cycle d'acquisition total a été de 1,05 sec.

Les données ont été acquises à l'aide du logiciel Analyst TF 1.7 et les données brutes ont été visualisées avec PeakView 2.2 et Masterview 1.1 (Sciex). Les données DDA ont été recherchées dans ProteinPilot OneOmics Suite (Sciex, version 3.1) à partir de la base de données d'UniProtKB/Swiss-Prot (Mars 2019, 20 572 protéines) de l'humain. La recherche a été effectuée selon les critères suivants : alkylation des cystéines par l'iodoacétamide, digestion enzymatique

par la trypsine et l'humain comme espèce. Le seuil d'identification des protéines et des peptides a été fixé à 1 % du FDR global. Les peptides identifiés par LC-HRMS/MS ont été utilisés comme base afin d'élaborer la méthode ciblée LC-MRM. Cette liste a également été complétée à partir des données répertoriées dans SRMATlas (Kusebauch *et al.*, 2016). La plateforme en ligne neXtProt (Schaeffer *et al.*, 2017) a été utilisée afin d'établir l'unicité des peptides pour les protéines d'intérêt.

3.3.3 Analyse ciblée des protéines S100 et ANX dans les larmes par LC-MRM

L'ensemble de ce projet a été réalisé en conformité avec le comité d'éthique de la recherche du Centre de recherches cliniques de l'Université de Montréal (CHUM) (CE 19.212), et suivant les principes de la Déclaration d'Helsinki, et avec la biobanque de matériel biologique à des fins de recherche sur les maladies hépatiques (22.276) et l'empreinte protéomique lacrymale chez une cohorte de patients atteints d'une maladie hépatique à des fins d'étude exploratoire (23.082).

3.3.3.1 Collecte des échantillons de larmes

Les larmes des patients ont été collectées sur des bandelettes lacrymales Schirmer (World Vision). Chaque bandelette a été remplie jusqu'à 23 mm et collectées pendant une période minimale de 5 min. La partie de la bandelette en contact direct avec la surface oculaire a été coupée et jetée. La bandelette restante a ensuite été coupée en petits morceaux dans un tube en polypropylène de 1,5 mL (Eppendorf, Mississauga, ON, Canada) et conservée à -80°C.

3.3.3.2 Digestion des échantillons de larmes

Aux tubes ont été ajoutés 200 µL de tampon 100 mM ABC, (pH 8,5) avant de les soniquer 15 min dans un bain à ultrasons. Les protéines ont été réduites en ajoutant 5 µL de DTT 100 mM (37°C, 600 rpm, 20 min) et alkylées en ajoutant 10 µL d'IAM 100 mM (37°C, 600 rpm, 30 min dans le noir). Les réactions ont été arrêtées en ajoutant 5 µL de DTT 100 mM. Les protéines ont été digérées à 37°C, durant la nuit (16-18h), avec l'ajout de 1 µg de trypsine. Les peptides ont été extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA) selon la méthode décrite précédemment.

3.3.3.3 Analyse LC-MRM des larmes et traitement des données

Les échantillons ont été injectés en utilisant le même modèle d'UHPLC et la même colonne que pour les analyses ci-haut. Le gradient chromatographique utilisé a été le suivant : une étape

initiale à 5% B pendant 1 min, suivie par une augmentation linéaire jusqu'à 30% B en 24 min, jusqu'à 50% B en 2 min puis à 90% B en 0,1 min, qui a été maintenu pendant 1,5 min pour bien laver la colonne avant l'étape de rééquilibration de la colonne aux conditions initiales. Un spectromètre de masse de type triple quadripôle-trappe linéaire Sciex QTRAP 5500, équipé d'une source d'ionisation Turbo IonSpray en mode positif, a été utilisé pour effectuer l'analyse par détection de réaction sélectionnée cédulée (sMRM, *scheduled multiple reaction monitoring*), avec les mêmes paramètres de source décrits ci-dessus. Une méthode finale compilant 27 protéines (17 S100 et 10 ANX) a été conçue pour la détection de 2 à 5 peptides par protéine et de deux transitions MRM par peptide. Pour chaque transition, le temps de maintien (*dwell time*) minimum a été fixé à 5 ms, avec une fenêtre de temps de rétention de 150 ms et un temps de cycle cible total de 1,25 s. L'acquisition des données a été faite en utilisant le logiciel avec Analyst 1.7 (Sciex).

Les pics LC-MRM ont manuellement été intégrés à partir du logiciel SciexOS (version 2.0.1) et visualisés à partir de PeakView 2.2 (Sciex). Les données quantitatives ont été normalisées en calculant le rapport entre la surface du pic et la surface la plus élevée trouvée dans l'ensemble des échantillons. Les données normalisées, correspondant aux ratios des aires d'intégration, ont été utilisées afin de construire la carte thermique. Celle-ci a été générée à partir de NG-CHM Builder (MD Anderson Cancer Center) (Ryan *et al.*, 2019). Le clustering hiérarchique (*Hierarchical clustering*) a été appliqué pour les lignes en spécifiant la métrique de distance Euclidienne et l'agglomération Ward. Aucune covariable n'a été attribuée.

3.4 Résultats et discussion

3.4.1 Développement d'une méthode ciblée LC-MRM pour la quantification des protéines S100 et ANX

Dans cette section, nous présentons tout d'abord le développement de la méthode ciblée LC-MRM pour l'analyse des protéines S100 et ANX. Pour se faire, une analyse préalable des protéines d'intérêt au moyen d'une approche non-ciblée par LC-HRMS/MS a permis la sélection des peptides et de leurs transitions et ce pour leur optimisation subséquente par LC-MRM. Grâce à deux collaborations menées respectivement avec les laboratoires de la Pre. Boisselier au département d'ophtalmologie de l'Université Laval et celui du Pr. Cappadocia au département de chimie à l'UQAM, nous avons eu accès à des protéines standards issues d'expressions bactériennes (17 S100 et 10 ANX). Ces protéines ont tout d'abord été soumises à un protocole de digestion enzymatique à la trypsine avant d'être analysées en mode DDA par LC-HRMS/MS. À partir de ces données, une première liste de peptides identifiés a été compilée. Ici, seuls les

peptides tryptiques uniques à chaque protéine ont été sélectionnés au moyen de neXtProt (Schaeffer *et al.*, 2017). Cette liste a également été complétée à partir de données issues de SRMATlas (Kusebauch *et al.*, 2016), soit une base de données répertoriant les peptides et transitions dérivant d'études antérieures. La Figure 3.1 représente le flux de travail de l'expression et purification des protéines standards jusqu'à leur optimisation par LC-MRM.

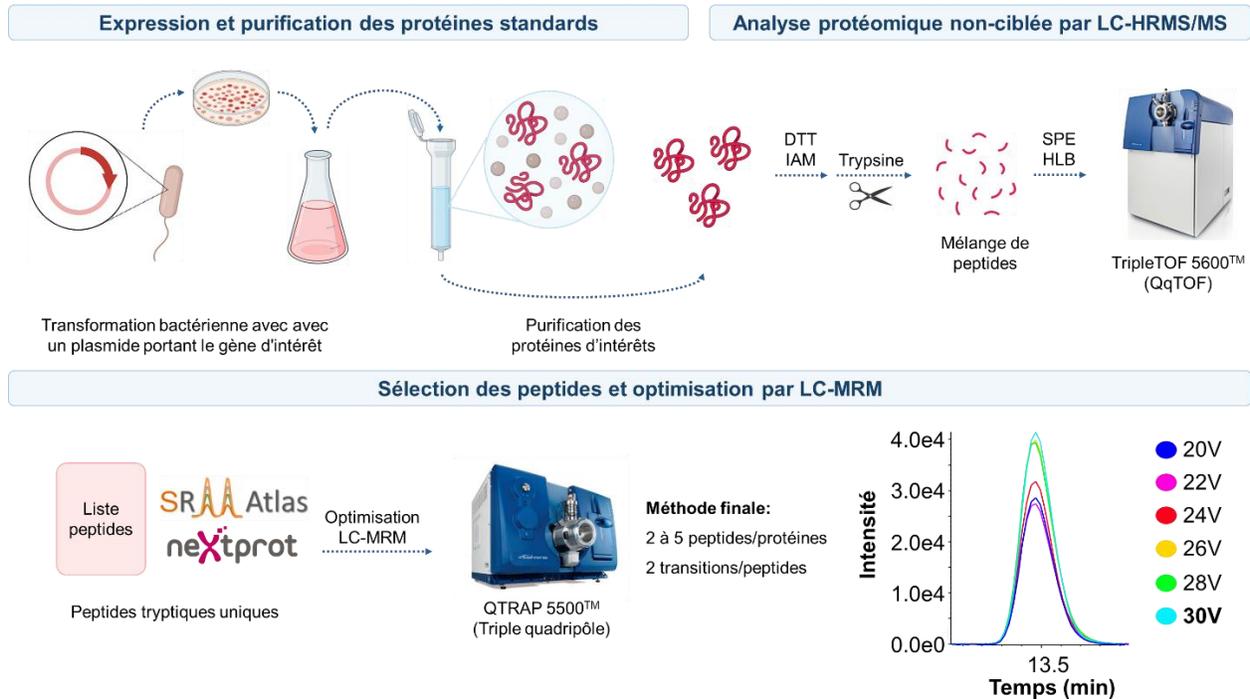


Figure 3.1: Flux de travail correspondant à l'optimisation de la méthode LC-MRM et exemple d'un chromatogramme correspondant à l'optimisation de l'énergie de collision pour la transition m/z 636,9 \rightarrow 1088,6 (y9) associée au peptide ALNSIIDVYHK de la protéine S100A8

Par la suite, sur la base de cette liste, une méthode ciblée LC-MRM a été développée. L'énergie de collision pour chaque transition (ion précurseur et ion fragment associé) a été optimisée et ce afin de permettre une sensibilité de détection optimale des transitions (Figure 3.1B). Finalement, de deux à cinq peptides par protéine et deux transitions par peptides ont été sélectionnés lors du développement de la méthode. La figure 3.2 présente un exemple de transitions pour le peptide ALNSIIDVYHK appartenant à la protéine S100A8. La première transition est utilisée pour la quantification du peptide alors que la seconde est utilisée pour l'analyse qualitative de ce dernier. Il est important de mentionner que leurs co-élution est nécessaire afin de confirmer l'identification

lors des analyses. Le tableau 3.3 répertorie l'ensemble des peptides et transitions choisies et optimisées constituant la méthode finale.

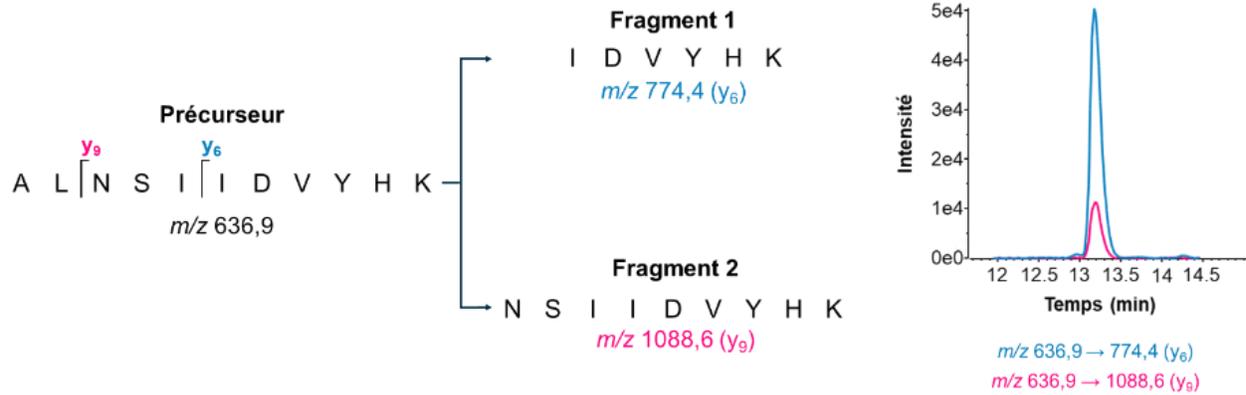


Figure 3.2: Exemple de transitions et de leur chromatogramme pour l'identification du peptide ALNSIIDVYHK appartenant à la protéine S100A8

Tableau 3.3: Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)		
P04083	Annexin A1	ALTGHLEEVVLALLK	26,1	3	536	711,4	y13 ²⁺	20		
			26,1	3	536	656,5	y6	22		
		GTDVNVFNTILTTR	19,9	2	775,9	818,5	y7	36		
			19,9	2	775,9	965,5	y8	36		
		GVDEATIIDILTK	22,9	3	463,3	589,4	y5	18		
			22,9	3	463,3	702,4	y6	16		
		ILVALCGGN	14,6	2	458,8	247,1	y3	16		
			14,6	2	458,8	407,1	y4	14		
		TPAQFDADELK	11,9	2	631,8	270,2	b3	36		
			11,9	2	631,8	865,4	y7	34		
		Q9UJ72	Annexin A10	ALLAICAGDAEDY	18,1	2	691,3	482,3	b5	16
					18,1	2	691,3	900,3	y8	18
DTLMNLVQGTR	16,3			2	624,3	560,3	y5	30		
	16,3			2	624,3	787,4	y7	30		
GVGTDENCLIEILASR	21,3			2	873,9	446,3	y4	44		
	21,3			2	873,9	688,4	y6	44		
LYSAIHDFGFHNK	11,9			2	774,9	636,8	y11 ²⁺	36		
	11,9			2	774,9	261,2	y2	40		
TMLQMILCNK	18,8			2	626,3	233,1	b2	30		
	18,8			2	626,3	778,4	y6	28		
P50995	Annexin A11			AHLVAVFNEYQR	12,4	3	482,9	492,3	b5	20
					12,4	3	482,9	591,4	b6	16
		DAQELYAAGENR	8,2	2	668,8	475,2	y4	20		
			8,2	2	668,8	780,4	y7	26		
		GTITDAPGFDPLR	15,4	2	680,4	385,3	y3	44		
			15,4	2	680,4	872,5	y8	34		
		SETDLLDIR	14	2	531,3	217,1	b2	24		
			14	2	531,3	288,2	y2	34		
		TPVLFDIYEIK	21,3	2	669,4	298,2	b3	34		
			21,3	2	669,4	618,9	y10 ²⁺	30		

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P07355	Annexin A2	AEDGSVIDYELIDQDAR	16,5	3	637	604,3	y5	24
			16,5	3	637	717,4	y6	24
		GDLENAFLNLVQCIQNKPLYFADR	27	3	946,8	990	y16 ²⁺	36
			27	3	946,8	1120,1	y18 ²⁺	34
		GVDEVTIVNILTNR	21,7	2	771,9	730,4	y6	34
			21,7	2	771,9	829,5	y7	28
		QDIAFAYQR	10,9	2	556,3	537,3	y4	26
			10,9	2	556,3	755,4	y6	26
P09525	Annexin A4	AASGFNAMEDAQTLR	13,4	2	791,4	389,3	y3	38
			13,4	2	791,4	832,4	y7	42
		AEIDMLDIR	15,8	2	538,3	288,2	y2	32
			15,8	2	538,3	762,4	y6	16
		GLGTDEDAIISVLAYR	22,6	2	846,9	708,4	y6	40
			22,6	2	846,9	821,5	y7	36
		SDTSFMFQR	12,3	2	559,8	728,4	y5	22
			12,3	2	559,8	815,4	y6	24
P08758	Annexin A5	SETSGSFEDALLAIVK	24	3	556,3	430,3	y4	16
			24	3	556,3	656,5	y6	20
		GLGTDEESILTLLTSR	22,2	2	853	476,3	y4	34
			22,2	2	853	589,4	y5	34
		VLTEIIASR	11,8	2	501,3	446,3	y4	32
			11,8	2	501,3	559,4	y5	26
		DENQSINHQMAQEDAQR	6,1	3	672	764,4	y13 ²⁺	30
			6,1	3	672	746,3	y6	30
P20073	Annexin A7	GAGTDDSTLVR	5,3	2	546,3	690,4	y6	28
			5,3	2	546,3	805,4	y7	26
		GFGTDEQAIVDVVANR	18,5	3	564,3	360,2	y3	18
			18,5	3	564,3	459,3	y4	20
		SEIDLVIQIK	13,8	2	522,8	217,1	b2	22
			13,8	2	522,8	715,4	y6	22
		VLIEILCTR	17,2	2	558,8	213,2	b2	28
			17,2	2	558,8	904,5	y7	24

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P13928	Annexin A8	GIGTNEQAIDVLTk	20,4	2	786,4	688,4	y6	38
			20,4	2	786,4	872,6	y8	38
Q5VT79	Annexin A8-like protein 1	AWIEQEGVTVK	12,1	2	630,3	347,2	y3	24
			12,1	2	630,3	889,5	y8	28
P13928/ Q5VT79	Annexin A8-like protein 1/ Annexin A8	NALLSLVGSDP	19,6	2	543,3	412,3	b4	16
			19,6	2	543,3	231,1	y2	16
O76027	Annexin A9	SSSHFNPDPAETLYK	10,6	3	603,3	468,7	y8 ²⁺	22
			10,6	3	603,3	936,5	y8	22
		ASDSAVDVAIEILATR	24,6	3	544,3	347,2	y3	14
			24,6	3	544,3	460,3	y4	20
		LHQALQETEPNYQVLIR	14,4	3	684,7	628,4	y5	36
			14,4	3	684,7	501,8	y8 ²⁺	24
		MAPSLTQEILSHLGLASK	24,2	3	632,7	847,5	y16 ²⁺	24
			24,2	3	632,7	565,3	y16 ³⁺	24
		SLQAALSGNLER	12,8	2	629,8	588,3	y5	24
			12,8	2	629,8	788,4	y7	28
STGQELEEAVQNR	10,5	2	730,9	417,2	y3	34		
	10,5	2	730,9	845,4	y7	38		
P23297	Protein S100-A1	ELLQTELSGFLDAQK	22	2	846,5	778,4	y7	34
			22	2	846,5	978,5	y9	36
		MGSELETAMETLINVFHAHSGK	22	3	601,3	878,5	y5	26
P60903	Protein S100-A10	DPLAVDK	5,6	2	379,2	262,1	y2	26
			5,6	2	379,2	545,3	y5	18
		EFPGFLENQK	14,8	2	604,8	466,7 ²⁺	y8	24
			14,8	2	604,8	932,5	y8	24

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P31949	Protein S100-A11	CIESLIAVFQK	20,8	2	654,4	422,2	y3	30
			20,8	2	654,4	1034,6	y9	26
		DGYNYTLISK	8,3	2	530,8	234,2	y2	24
			8,3	2	530,8	611,3	y5	20
		DPGVLDR	5,7	2	386,2	559,3	y5	20
			5,7	2	386,2	328,7	y6 ²⁺	20
		ISSPTETER	3,5	2	510,3	732,4	y6	24
			3,5	2	510,3	906,4	y8	24
		TEFLSFMNTELAFTK	26,6	3	617,3	231,1	b2	26
			26,6	3	617,3	650,4	y6	20
P80511	Protein S100-A12	ELANTIK	4,7	2	394,7	214,6	b4 ²⁺	20
			4,7	2	394,7	475,3	y4	18
		GHFDTLSK	4	3	302,2	234,2	y2	12
			4	3	302,2	347,2	y3	12
Q99584	Protein S100-A13	DSLVSNEFK	11,2	2	519,8	203,1	b2	20
			11,2	2	519,8	836,5	y7	16
		DVGSLDEK	4,3	2	431,7	215,1	b2	16
			4,3	2	431,7	648,3	y6	16
		ELVTQQLPHLLK	15,6	2	709,9	720,5	y6	34
			15,6	2	709,9	848,5	y7	34
		SLDVNQDSELK	8,5	2	624,3	201,1	b2	26
			8,5	2	624,3	833,4	y7	28
Q9HCY8	Protein S100-A14	DLVTQQLPHLMPSNCGLEEK	17,8	3	770,4	229,1	b2	30
			17,8	3	770,4	991,5	y17 ²⁺	28
		IANLGSCNDSK	4,3	2	589,8	767,3	y7	26
			4,3	2	589,8	994,4	y9	24
		NFHQYSVEGGK	5,2	2	633,3	503,2	b8 ²⁺	30
			5,2	2	633,3	576,3	y6	34
		SANAEDAQEFSDVER	9,4	2	834,4	752,4	y6	38
			9,4	2	834,4	881,4	y7	38
SFWELIGEAAK	21,2	2	625,8	475,3	y5	24		
	21,2	2	625,8	701,4	y7	26		

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P25815	Protein S100-P	DLDANGDAQVDFSEFIVFVAaitsachk	27,5	3	1014,2	229,1	b2	44
			27,5	3	1014,2	1021,5	y18 ²⁺	40
		ELPGFLQSGK	13,7	2	538,3	736,4	y7	28
			13,7	2	538,3	417,2	y8 ²⁺	20
		MTELETAMGMIIDVFSR	27,7	3	648,7	409,2	y3	26
			27,7	3	648,7	736,4	y6	20
		YSGSEGSTQTLTK	4,6	2	679,8	835,5	y8	34
			4,6	2	679,8	964,5	y9	32
Q96FQ6	Protein S100-A16	AVIVLVENFYK	20,5	2	647,9	284,2	b3	28
			20,5	2	647,9	799,4	y6	24
		ELNHMLSDTGNR	7,1	3	462,9	346,2	y3	22
			7,1	3	462,9	562,3	y5	18
		ISFDEYWTLLIGGITGPIAK	27	3	694,4	485,3	y5	18
			27	3	694,4	813,5	y9	26
		LIHEQEQQSSS	13,9	2	643,3	503,8	b8 ²⁺	24
			13,9	2	643,3	547,3	b9 ²⁺	22
		LIQNLDANHDGR	6,3	3	455,9	570,3	y10 ²⁺	16
			6,3	3	455,9	506,2	y9 ²⁺	20
P29034	Protein S100-A2	ELPSFVGEK	12,1	2	503,3	382,2	y7 ²⁺	20
			12,1	2	503,3	763,4	y7	20
		VDEEGLK	3,1	2	395,2	215,1	b2	16
			3,1	2	395,2	575,3	y5	16
P33764	Protein S100-A3	DCEVDFVEYVR	16,9	2	715,8	927,5	y7	28
			16,9	3	477,6	566,3	y4	18
		ELATWTPTEFR	16,1	2	675,8	750,4	y6	30
			16,1	2	675,8	936,5	y7	30
		FMSVLDTNK	12	2	527,8	590,3	y5	24
			12	2	527,8	776,4	y7	18
MARPLEQAVAAIVCTFQEYAGR	27,3	3	827,8	569,8	b11 ²⁺	34		
	27,3	3	827,8	303,2	y3	30		
SLACLCLYCHEYFK	17	3	622	796,8	y11 ²⁺	24		
	17	3	622	832,4	y12 ²⁺	22		

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P26447	Protein S100-A4	ALDVMVSTFHK	12,9	3	416,6	532,3	y4	14
			12,9	3	416,6	718,4	y6	20
		ELPSFLGK	14,4	2	445,8	204,1	y2	16
			14,4	2	445,8	324,7	y6 ²⁺	14
		LMSNLDSNR	6,3	2	525,3	376,2	y3	36
			6,3	2	525,3	805,4	y7	24
		MACPLEK	5,6	2	424,7	486,3	y4	24
			5,6	2	424,7	646,3	y5	16
		TDEAAFQK	3,4	2	455,2	217,1	b2	20
			3,4	2	455,2	693,4	y6	16
P33763	Protein S100-A5	ALTTMVTTFHK	11,7	2	625,3	633,3	y5	26
			11,7	2	625,3	732,4	y6	26
		ELCLGEMK	10,8	2	490,2	385,2	b3-18	20
			10,8	2	490,2	464,2	y4	20
		ESSIDLMK	11,7	2	519,3	506,3	y4	20
			11,7	2	519,3	821,4	y7	20
		NSDQEIDFK	7,7	2	548,3	409,2	y3	24
7,7	2		548,3	894,4	y7	18		
P06703	Protein S100-A6	ELTIGSK	5,3	2	374,2	404,3	y4	18
			5,3	2	374,2	505,3	y5	18
		LMEDLDR	8,1	2	446,2	518,3	y4	20
			8,1	2	446,2	647,3	y5	18
		LQDAEIAR	5,4	2	458,3	242,2	b2	20
			5,4	2	458,3	225,1	b2-17	22
		MACPLDQAIGLLVAIFHK	27,7	3	666,4	363,1	b3	26
27,7	3		666,4	818	y15 ²⁺	26		
P31151	Protein S100-A7	GTNYLADVFEK	17	2	628,8	821,4	y7	26
			17	2	628,8	984,5	y8	26
		IDFSEFLSLLGDIATDYHK	27,5	3	728,7	666,8	y12 ²⁺	28
			27,5	3	728,7	796,9	y14 ²⁺	26

Tableau 3.3 : Transitions MRM optimisées pour l'analyse ciblée des protéines S100 et annexines (suite)

Numéro Accession	Protéine	Peptide	tr (min)	z	Q1 m/z	Q3 m/z	Fragment	CE (V)
P31151/ Q86SG5	Protein S100-A7/ Protein S100-A7A	ENFPNFLSACDK	16,4	2	721,3	244,1	b2	36
			16,4	2	721,3	1051,5	y9	32
		QSHGAAPCSGGSQ	18,6	2	622,3	216,1	b2	40
			18,6	2	622,3	353,2	b3	26
		SIIGMIDMFHK	20,2	3	431,2	790,4	y6	18
			20,2	3	431,2	489,7	y8 ²⁺	16
P05109	Protein S100-A8	ALNSIIDVYHK	13,2	2	636,9	774,4	y6	30
			13,2	2	636,9	1088,6	y9	30
		GADVWFK	12	2	411,7	244,1	b3	18
			12	2	411,7	694,4	y5	14
		GNFHAVYR	4,7	3	321,8	396,7	y6 ²⁺	10
			4,7	3	321,8	445,2	y7-17 ²⁺	14
		LLETECPQYIR	11,9	2	711,4	676,4	y5	34
			11,9	2	711,4	1066,5	y8	32
		MLTELEK	8,5	2	432,2	245,1	b2	18
			8,5	2	432,2	619,3	y5	16
P06702	Protein S100-A9	DLQNFLK	8,7	2	439,2	521,3	y4	18
			8,7	2	439,2	649,4	y5	18
		LGHPDTLNQGEFK	4,6	3	485,9	621,3	b6	20
			4,6	3	485,9	353,2	y6-17 ²⁺	20
		LTWASHEK	19,9	2	486,3	500,3	y4	24
			19,9	2	486,3	757,4	y6	22
		NIETIINTFHQYSVK	14,3	3	603	725,9	y12 ²⁺	22
			14,3	3	603	790,4	y13 ²⁺	20
P04271	Protein S100-B	VIEHIMEDLDTNADK	14,3	2	871,9	479,3	b4	30
			14,3	2	871,9	548,3	y5	26
		AMVALIDVFHQYSGR	20,8	3	569,6	667,9	y11 ²⁺	20
			20,8	3	569,6	703,4	y12 ²⁺	18
		ELINNELSHFLEEIK	23,8	3	610	679,9	y11 ²⁺	24
			23,8	3	610	736,9	y12 ²⁺	24

3.4.2 Application de la méthode LC-MRM pour l'analyse de larmes humaines

Grâce à une collaboration menée avec la Dre Marie-Claude Robert, ophtalmologue au Centre hospitalier de l'Université de Montréal (CHUM) ainsi qu'avec le Dr Christopher Rose du laboratoire hépato-neuro au centre de recherche du CHUM (crCHUM), nous avons eu accès à différents échantillons de larmes de patients. La méthode ciblée développée a ainsi été appliquée afin d'analyser 69 échantillons de larmes provenant de 49 patients atteints d'une variété de pathologies, incluant certaines maladies oculaires, d'épidermolyse bulleuse (EB) ou encore de différentes maladies hépatiques. L'épidermolyse bulleuse correspond à un groupe hétérogène de dermatose rare et génétiquement héréditaire, caractérisée par des lésions au niveau de la peau et des muqueuses, et par la formation de cloques (Bardhan et *al.*, 2020). Les différents échantillons ont été sujets à un protocole de digestion à la trypsine avant d'être analysés par LC-MRM à partir d'un triple quadripôle QTRAP 5500 (figure 3.3). Cette analyse nous a permis de valider la méthode développée, et ce afin d'établir quelles sont les protéines qui peuvent être détectées de façon adéquate dans les larmes.

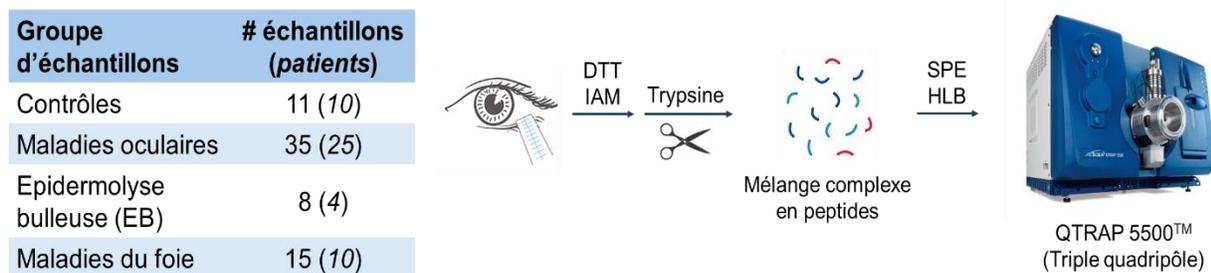


Figure 3.3: Flux de travail correspondant à l'analyse ciblée LC-MRM des échantillons de larmes humaines

Parmi les 27 protéines ciblées, seules deux d'entre elles, à savoir S100A10 et S100B, n'ont été identifiées dans aucun des échantillons. Les protéines S100A1, A3, A5 et A9 ont quant à elles été identifiées dans moins de 40% des échantillons (tableau 3.4). Différentes raisons peuvent expliquer ceci. Les effets de matrice causés par la complexité des échantillons peuvent affecter l'ionisation des peptides rendant leur identification par spectrométrie de masse extrêmement difficile. De plus, la faible abondance de ces protéines dans les larmes pourrait également expliquer ces résultats. En effet, il peut être difficile de distinguer le signal induit par nos peptides d'intérêt du bruit de fond lors de l'intégration. La figure 3.4 présente des chromatogrammes

représentatifs pour certaines des transitions associées aux protéines S100 et ANX identifiées soit dans l'un des échantillons contrôles (bleu) ou dans un échantillon issu du groupe maladie hépatique (rose). Pour la majorité des données, les transitions choisies présentent un très bon rapport signal sur bruit de fond permettant une très bonne identification des protéines dans les larmes. Toutefois, pour certaines de ces protéines (S100A3, S100B, ANXA8L1, ANXA9), il est possible d'observer beaucoup d'interférences rendant l'identification et la quantification de ces protéines impossible.

Tableau 3.4: Identification des protéines S100 et ANX au travers des groupes d'échantillons

Protéine	# échantillons				Total	% des échantillons
	Contrôle	Maladies oculaires	EB	Maladies hépatiques		
S100A7	11	35	8	15	69	100,0
S100A8	11	35	8	15	69	100,0
S100A11	11	35	8	15	69	100,0
ANXA11	11	35	8	15	69	100,0
ANXA10	11	35	8	15	69	100,0
ANXA2	11	35	8	15	69	100,0
S100A6	11	35	8	15	69	100,0
ANXA7	11	35	8	15	69	100,0
S100A4	11	35	8	15	69	100,0
S100P	11	35	8	15	69	100,0
S100A16	11	35	8	15	69	100,0
ANXA11	11	34	8	15	68	98,6
ANXA5	11	34	8	15	68	98,6
S100A2	10	35	8	15	68	98,6
S100A14	11	35	8	14	68	98,6
ANXA8	11	33	8	14	66	95,7
ANXA4	11	32	8	14	65	94,2
S100A13	10	33	8	14	65	94,2
ANXA9	9	29	6	15	59	85,5
S100A12	10	25	6	14	55	79,7
ANXA8L1	8	29	4	11	52	75,4
S100A9	3	13	5	1	22	31,9
S100A3	3	8	3	5	19	27,5
S100A1	1	2	0	11	14	20,3
S100A5	0	0	0	1	1	1,5
S100B	0	0	0	0	0	0,0
S100A10	0	0	0	0	0	0,0

Nombre d'échantillons où les protéines ont été identifiées dans les différents groupes.

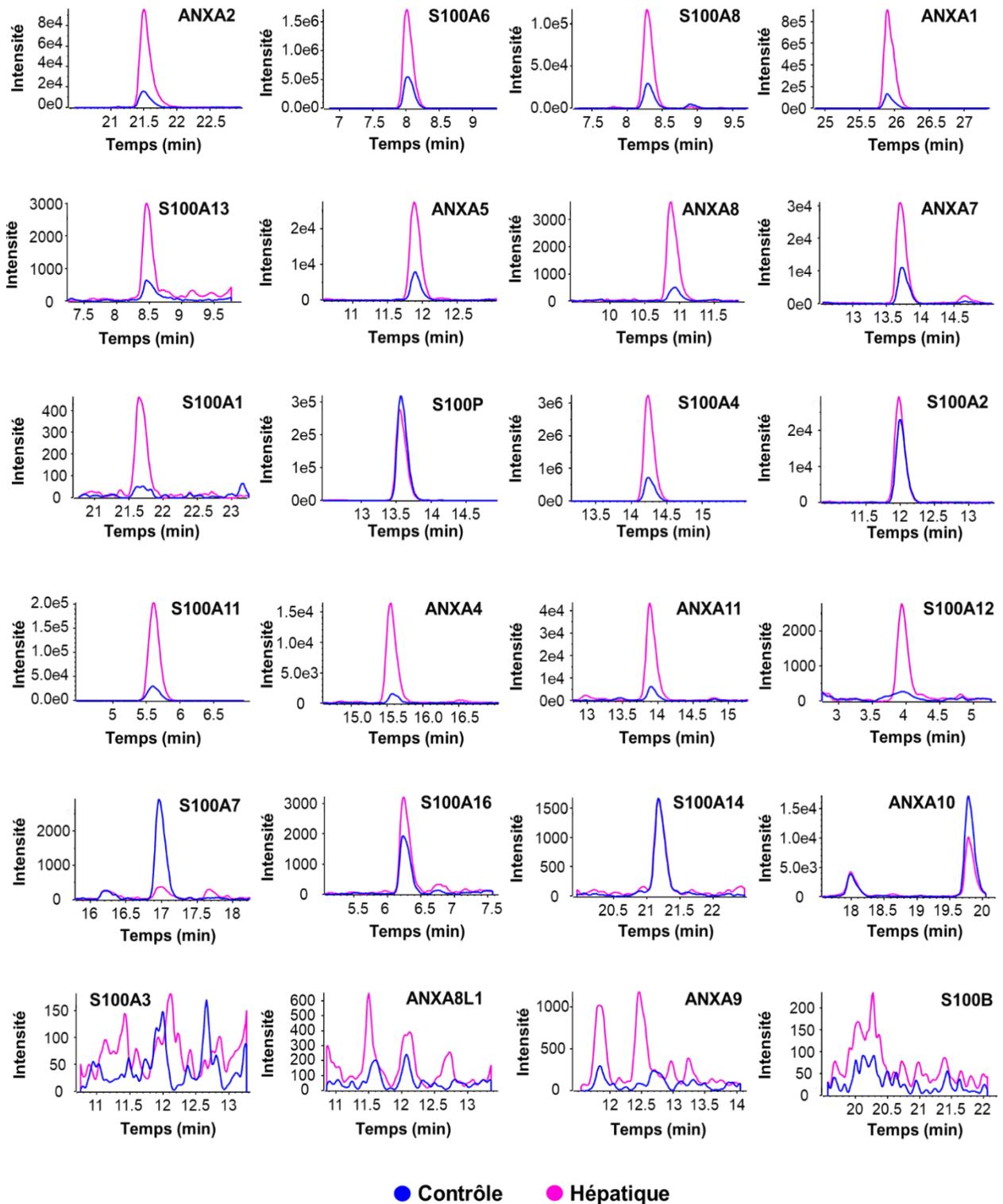


Figure 3.4: Chromatogrammes d'ion extrait pour certaines des protéines S100 et ANX. Exemple pour un échantillon contrôle (bleu) et un échantillon issu du groupe maladie hépatique (rose)

Un aspect intéressant à mentionner, et comme l'en démontre la carte thermique présentée à la figure 3.5, plusieurs de ces protéines semblent fortement impactées chez le groupe de patients souffrant d'une maladie hépatique. Tout d'abord, il faut savoir que le foie joue un rôle clé non seulement dans le métabolisme, mais également dans la réponse immunitaire et inflammatoire, et qu'il communique avec d'autres organes par l'entremise de ces mécanismes afin de réguler certains processus physiopathologiques (Wu *et al.*, 2023; Yuan *et al.*, 2021). De plus, un dysfonctionnement hépatique peut souvent être associé à un phénotype pathologique systémique notamment aux niveaux des yeux (Wu *et al.*, 2023).

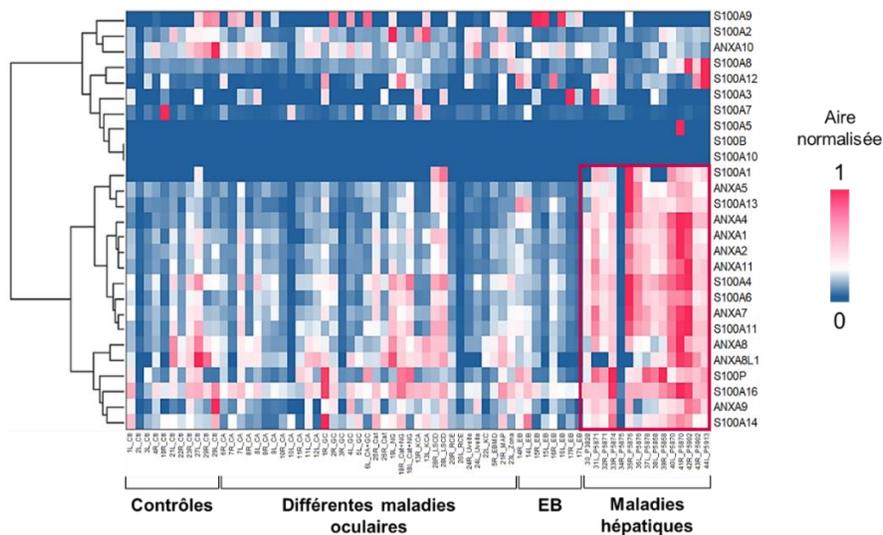


Figure 3.5: Carte thermique correspond aux variations de l'expression des protéines S100 et ANX observées chez les différents groupes d'échantillons. Les protéines en rouge correspondent aux protéines augmentées dans les larmes des patients.

Il est également important de souligner le potentiel rôle que peuvent tenir certaines de ces protéines dans le développement et la progression d'une multitude de maladies hépatiques notamment dans la stéatose hépatique non alcoolique (NAFLD), la cirrhose ou encore le carcinome hépatocellulaire (HCC). Par exemple, les niveaux sériques de S100A4 se voient être extrêmement élevés chez les patients atteints de cirrhose (Chen *et al.*, 2015), tout comme les taux de S100A9 chez des modèles murins atteints de NAFLD (Liu *et al.*, 2015). Les protéines S100A11 et S100A8 se voient également être augmentées dans le foie de patients atteints de NAFLD ou dans différents modèles de souris atteints de la maladie (Teng *et al.*, 2021; Sobolewski *et al.*, 2020; Mukai *et al.*, 2016). Une étude menée par Zhuang *et al.*, et dont le but était d'analyser

les données publiées de l'expression des ANXA dans le cancer du foie, a révélé que les ANXA1, A2, A3, A4 et A5 se voyaient être surexprimées dans le HCC alors que l'ANXA10 se voit être sous-exprimée (Zhuang *et al.*, 2019). L'altération de l'expression de l'ARNm de plusieurs protéines S100 (S100P, S100A2, S1005, S100A6, S100A7, S100A8, S1009, S10010, S100A11, S100A12, S100A17, S100A14, S100G et S100Z) a également été associées au HCC (Zhang *et al.*, 2021). La figure 3.6 présente les principales protéines S100 connues pour être impliquées dans la pathogénèse de NAFLD, NASH (stéatohépatite non alcoolique) et HCC.

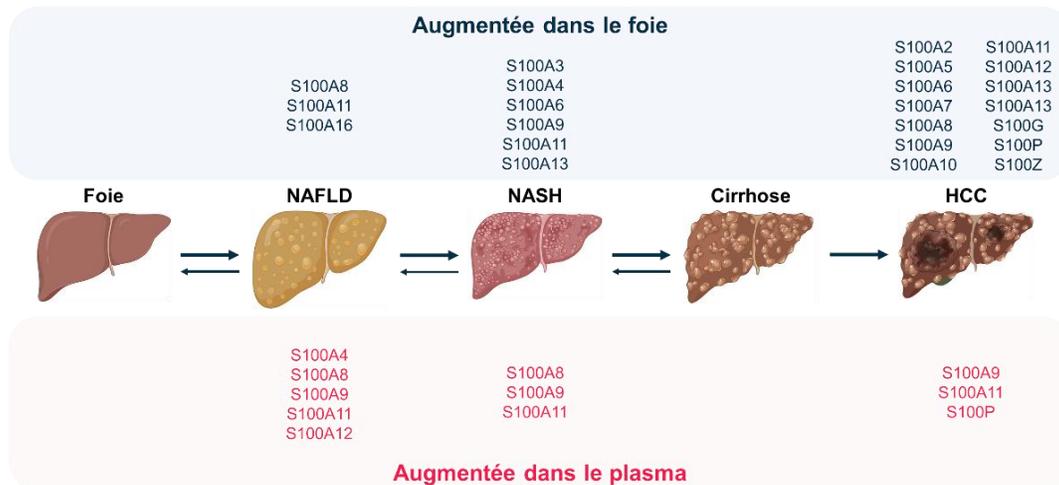


Figure 3.6: Principales protéines S100 impliquées dans les différents stades de développement de NALFD, NASH et HCC. En bleu se sont les protéines qui sont augmentées dans le foie et en rouges celles augmentées dans le plasma. Figure adaptée de Delangre *et al.* (Delangre *et al.*, 2022)

La pertinence clinique de ces protéines en tant que biomarqueurs potentiels pour le diagnostic et le pronostic des maladies hépatiques ne peut ainsi être sous-estimée. Nous avons montré dans cette étude que plusieurs des protéines S100 et ANX étudiées se sont avérées être fortement augmentées dans les larmes des patients souffrant de maladies hépatiques, soulignant le rôle de ces protéines dans le développement des maladies du foie. Ces résultats préliminaires sont très prometteurs dans la mesure où les larmes pourraient offrir une alternative intéressante afin d'étudier ces différentes maladies hépatiques, ceci constituant à notre connaissance une première dans ce contexte. Ces données devront toutefois être validées dans le futur à partir de cohortes de patients plus grandes. De plus, la méthode développée peut être appliquée afin d'analyser différents types d'échantillons incluant le plasma, l'urine et les tissus, et ce dans le contexte de différentes maladies ou de traitements.

CHAPITRE 4
**ARTICLE SCIENTIFIQUE : COMPARAISON DE MÉTHODES POUR L'ANALYSE NON-
CIBLÉE DU PROTÉOME PLASMATIQUE PAR LC-HRMS/MS**

Carina Lima¹ et Lekha Sleno¹

¹Université du Québec à Montréal, Département de chimie, Montréal, Canada

Contributions des auteurs : Carina Lima a effectué la recherche bibliographique, la réalisation de toute les manipulations, l'analyse des résultats, la réalisation des tableaux et des figures en plus de rédiger l'article. Lekha Sleno a encadré et supervisé le projet en plus d'avoir révisé l'article.

4.1 Résumé

Le plasma constitue une source incroyable de biomarqueurs protéiques, faisant de lui un outil de choix pour étudier les perturbations biologiques. Toutefois, les analyses protéomiques du plasma par spectrométrie de masse vont souffrir de l'extrême complexité de la matrice, rendant l'analyse de certaines protéines peu abondantes extrêmement difficile. Dans ce contexte, de nombreux efforts ont été déployés pour développer des flux de travail efficaces pour couvrir le protéome plasmatique. Dans ce travail, nous comparons cinq approches de préparation d'échantillon pour des analyses protéomiques ascendantes non ciblées par chromatographie liquide couplée à la spectrométrie de masse à haute résolution (LC-HRMS/MS). Nous avons comparé ici : (1) une méthode d'enrichissement des vésicules membranaires à partir de billes magnétiques fonctionnalisées de groupements ammonium quaternaire (MagNet), (2) une méthode de déplétion des protéines abondantes par leur précipitation à l'acide perchlorique (PCA), (3) une méthode de solubilisation des protéines au SDS suivie d'une précipitation à l'acétone assistée par sel (PPT), (4) une méthode combinant une déplétion des IgG à partir de billes magnétiques avec une protéine G recombinante immobilisée suivie d'une précipitation des protéines déplétées à l'acétone (PPT-IgG) et (5) un ensemble commercial d'extraction des protéines à partir d'une plaque filtrante 96 puits (ProTrap). Chaque méthode a été évaluée pour son efficacité et sa performance : recouvrement protéique, reproductibilité, complémentarité, précision quantitative, coûts, automatisation, etc. La méthode ProTrap offre la meilleure couverture protéomique (~325) et la meilleure reproductibilité, tandis que MagNet et PCA offrent une grande complémentarité avec un nombre de protéines uniques plus élevé. L'analyse de courbes d'étalonnage adaptées à la matrice a révélé que les méthodes PPT et ProTrap permettent une meilleure quantification des protéines identifiées comparativement aux méthodes basées sur la déplétion et l'enrichissement.

4.2 Introduction

Le sang et ses différents dérivés (plasma et sérum) offrent une voie prometteuse afin d'étudier les processus biologiques et pathologiques chez un groupe d'individus, étant ainsi souvent utilisés comme matrice biologique dans le but d'identifier de nouveaux biomarqueurs associés à diverses maladies. Le sang étant en contact avec les différents organes et tissus de l'organisme par l'entremise du système vasculaire, sa composition moléculaire peut ainsi offrir un aperçu global de l'état physiologique d'un individu (Deutsch *et al.*, 2021; Ward *et al.*, 2024; Kaur *et al.*, 2021). En effet, une multitude de maladies peuvent induire des effets systémiques sur l'organisme, qui peuvent altérer la composition moléculaire du sang et ainsi fournir des informations essentielles sur ces pathologies (Ward *et al.*, 2024).

La collecte de sang est simple, peu invasive et faite sur une base régulière en clinique (Kaur *et al.*, 2021). Le plasma, soit le sujet de cette étude, est un merveilleux outil de diagnostic, faisant de lui l'un des biofluides les plus couramment prélevés et utilisés (~40% des analyses cliniques) (Viode *et al.*, 2023). Ce dérivé du sang est obtenu après l'ajout d'agents anti-coagulants tels l'EDTA ou le citrate de sodium, suivi de sa centrifugation (Ignjatovic *et al.*, 2019). Par conséquent, le plasma correspond à la composante liquide du sang (91-92% H₂O), dépourvue de tout matériel cellulaire (érythrocytes, leucocytes et thrombocytes) (Mathew, Sankar & Varacallo, 2018; Ignjatovic *et al.*, 2019). D'un point de vue moléculaire, le plasma sanguin est une matrice biologique excessivement complexe et hétérogène composé d'une grande variété de molécules : lipides, métabolites, hormones, peptides et protéines (Ward *et al.*, 2024; Ignjatovic *et al.*, 2019). En ce qui concerne le protéome plasmatique, celui-ci serait une source incroyable en biomarqueurs (Rice & Belani, 2024). En effet, on estime qu'il contiendrait plus de 10 000 protéines distinctes, ce nombre excluant les différents protéoformes issues des modifications post-traductionnelles (PTM) (Blume *et al.*, 2020; Keshishian *et al.*, 2017).

La LC-MS/MS s'est positionnée comme une méthode de choix afin d'étudier le protéome plasmatique, et ce en raison de sa spécificité et de son impartialité (*unbiased*). Toutefois, le potentiel du plasma comme matrice biologique sera limité par son immense complexité, faisant de lui un fluide biologique excessivement difficile à analyser de façon complète (Metatla *et al.*, 2024). En effet, l'abondance extrêmement élevée d'un nombre restreint de protéines plasmatiques (22 protéines), et qui composent à elles seules plus de 99% de la masse totale en protéines, limite grandement l'identification de protéines d'intérêt présentes en plus faible quantité (Ward *et al.*, 2024; Viode *et al.*, 2023). De plus, la gamme dynamique des concentrations en protéines, qui peuvent s'étendre de 10 à 12 ordres de grandeur (Figure 4.1), vont largement dépasser les limites de détection des instruments en LC-MS/MS, limitant également l'identification et la quantification de ces protéines (Keshishian *et al.*, 2017; Ward *et al.*, 2024). C'est pourquoi différentes stratégies ont été développées au fil des ans et ce afin de réduire la complexité de la matrice et ainsi favoriser l'identification des protéines du plasma. Parmi celles-ci on retrouve une multitude d'approches basées sur le préfractionnement des échantillons, la déplétion des protéines abondantes ou l'enrichissement des protéines peu abondantes (Lee *et al.*, 2019).

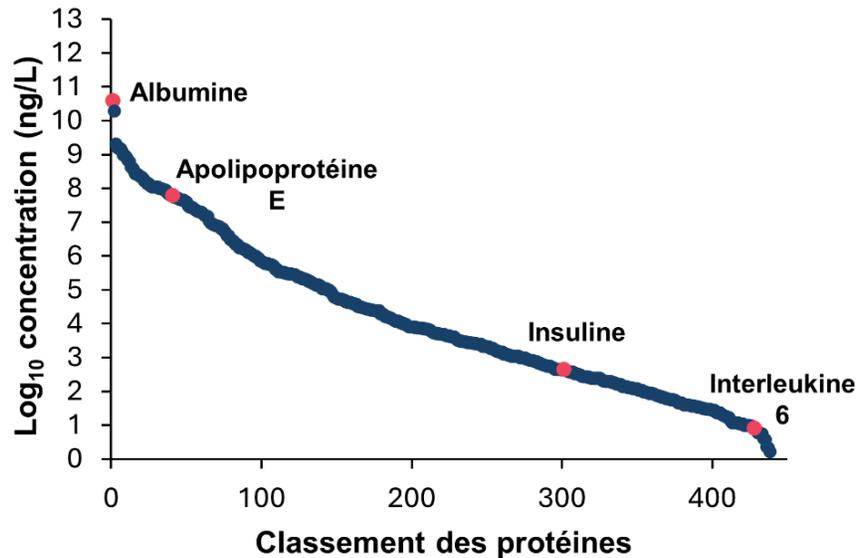


Figure 4.1: Classement des protéines du plasma en fonction de leur concentration dans la matrice. Les concentrations sont exprimées sur l'axe des y en valeur absolues (Log10). Quantification des protéines du plasma par immunoessais. Figure adaptée de *The Human Protein Atlas* (Uhlén et al., 2015).

Ainsi, dans l'optique d'établir une approche efficace afin d'étudier les protéines du plasma dans des échantillons biologiques, différentes méthodologies d'extractions et d'analyses ont été testées et comparées dans ce travail. Nous présenterons et comparerons cinq méthodes différentes, soit une méthode d'enrichissement des vésicules membranaires à partir de billes magnétiques à échangeurs anioniques, une méthode de déplétion des protéines abondantes à partir d'acide perchlorique, une méthode de précipitation des protéines du plasma à l'aide d'acétone avec et sans déplétion des immunoglobulines G (IgG), et pour finir une méthode commerciale d'extraction des protéines à partir d'une plaque filtrante 96 puits.

4.3 Matériels et méthodes

4.3.1 Produits chimiques

Le tampon phosphate salin 1X (PBS) provenait de Wisent Inc. (Saint-Jean-Baptiste, QC, Canada). Les billes magnétiques *Dynabeads Protein G* ont été obtenues auprès de ThermoFisher Scientific (# produit 10003D) et les billes magnétiques à échangeurs anionique (SAX) MagReSyn® auprès de Resyn Bioscience (# produit MR-SAX005, Pretoria, Afrique du Sud). L'ensemble de plaque 96 puits ProTrap HT avec tampon SDS et tampon de digestion ont été obtenu auprès d'Allumiqs (# cat. PWP-R001, Halifax, NS, Canada). Le plasma lyophilisé humain (# produit P9523), le plasma

lyophilisé de poulet (# produit P3266), le Bis-Tris Propane (BTP), Tris(2-carboxyéthyl)phosphine (TCEP), sodium dodecyl sulfate (SDS), bicarbonate d'ammonium (ABC), dithiothréitol (DTT), iodoacétamide (IAM) ainsi que le méthanol (MeOH), l'acétonitrile (ACN) et l'acétone de grade HPLC ainsi que l'acide formique (FA) et l'acide trifluoroacétique (TFA) ont tous été acquis auprès de Sigma-Aldrich (Oakville, ON, Canada). La trypsine de qualité séquençage a été obtenue auprès de Promega (produit V5111, Madison, WI, USA). Les cartouche SPE OASIS HLB de 30 mg proviennent de Waters (Milford, MA). L'eau ultra-pure provenait d'un système UV Millipore Synergy (Billerica, MA, USA).

4.3.2 Échantillons commerciaux de plasma

Une quantité de 5 g de plasma lyophilisé humain (plasma contenant du citrate de sodium à 4% comme anticoagulant) ou de poulet (plasma contenant du citrate de sodium à 3,8% comme anticoagulant) a été reconstituée dans 5 mL d'eau chacun, puis aliquotée dans des tubes en polypropylène de marque Eppendorf (1 mL) avant d'être entreposée à -20°C jusqu'à leur utilisation.

4.3.3 Préparation d'une courbe d'étalonnage adaptée à la matrice

Le plasma humain a été dilué dans du plasma de poulet par mélange volumétrique. La courbe de calibration comprend 6 points de ratio plasma humain : plasma poulet (0:100, 10:90, 25:75, 50:50, 75:25 et 100:0). Chaque niveau d'étalonnage a été préparé en triplicata (n=3) pour un total de 18 dilutions.

4.3.4 Méthode 1 - Précipitation des protéines du plasma par l'acétone

Les protéines du plasma ont tout d'abord été solubilisées. À 5 µL de plasma ont été ajoutée 20 µL 0,5% SDS. Les échantillons ont été chauffés à 95°C pendant 5 min puis mis à température pièce. Finalement, 75 µL de NaCl 75 mM a été ajoutés aux échantillons. Les protéines ont été précipitées en ajoutant 400 µL d'acétone. Les mélanges ont été incubés à -20°C pendant 30 min puis centrifugés (14 000 rpm, 4°C, 8 min) afin d'écarter le surnageant. Les culots en protéines ont été lavés avec 400 µL d'acétone et centrifugés (14 000 rpm, 4°C, 2 min). Un second lavage avec 80% d'acétone a été fait selon la même approche. Les culots ont finalement été séchés environ 10 min au Speed Vac. Les protéines ont été réduites et solubilisées dans 50 µL de DTT 20 mM (préparé dans une solution 7M urée, 2M thiourée) et soniquées 10 min dans un bain à ultrasons. Les échantillons ont été dilués en ajoutant 100 µL de tampon 100 mM ABC, pH 8,5. Les protéines ont été alkylées en ajoutant 15 µL 100 mM IAM (37°C, 650 rpm, 15 min dans le noir) puis dilués à

nouveau en ajoutant 250 µL de tampon 100 mM ABC, pH 8,5. Finalement, les protéines ont été digérées à 37°C, 16h avec 2 µg de trypsine. Les peptides ont été extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA). Les cartouches ont été conditionnées avec 1 mL MeOH et 1 mL d'eau. Après avoir appliqué les échantillons, les peptides ont été lavés avec 1 mL d'eau et élués deux fois avec 500 µL MeOH. Les échantillons ont finalement été évaporés au Speed Vac.

4.3.5 Méthode 2 - Déplétion des IgG du plasma suivit d'une précipitation des protéines à partir d'acétone

25 µL de plasma ont tout d'abord été dilués dans 75 µL de tampon 1X PBS. Par la suite, 100 µL de billes magnétiques *Dynabeads Protein G* (ThermoFischer Scientific) ont été équilibrés trois fois avec 500 µL de tampon 1X PBS puis combinés aux échantillons de plasma dilué (100 µL). Les réactions ont été incubées à température pièce, sous agitation, pendant 30 min. Le plasma déplété en IgG (100 µL) a été récupéré sur un séparateur magnétique et les billes ont été lavées avec 50 µL de tampon 1X PBS. Le lavage a été récupéré et combiné aux protéines déplétées. À 25 µL de plasma déplété ont été ajoutés 25 µL de SDS 1% afin de solubiliser les protéines. Les échantillons ont été chauffés à 95°C pendant 5 min puis mis à température pièce. 50 µL de NaCl 100 mM ont été ajoutés avant de procéder à la précipitation des protéines et à leur digestion selon la méthode décrite précédemment.

4.3.6 Méthode 3 - Extraction des protéines à partir du kit de plaque filtrante ProTrap

Pour cette expérience, les protéines du plasma ont été extraites à partir du kit de plaque filtrante 96 puits ProTrap de la compagnie Allumiqs. Pour ce faire, 5 µL de plasma ont été dilués dans 95 µL de tampon SDS de la compagnie. Les échantillons ont été mélangés avant d'être transférés (90 µL) dans les puits. Les protéines ont d'abord été réduites en ajoutant 5 µL 200 mM DTT préparé dans un tampon 100 mM ABC, pH 8,5 (37°C, faible agitation, 30 min) et alkylées en ajoutant 5 µL 300 mM IAM également préparé dans un tampon 100 mM ABC, pH 8,5 (37°C, faible agitation, 30 min dans le noir). Les protéines ont été précipitées au fond des puits en ajoutant 400 µL d'une solution 87,5% acétone; 12,5% MeOH. Les mélanges ont été mélangés manuellement (10 sec) avant d'être incubés à température pièce pendant 30 min. Le solvant a été éliminé en appliquant un vide (*manifold*). Les protéines ont été lavées deux fois avec 200 µL MeOH. Les protéines ont été digérées à 37°C, 16h à partir de 2 µg de trypsine (préparée dans le tampon de digestion de la compagnie). Les peptides ont finalement été récupérés en appliquant un vide (*manifold*) et extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA) selon la procédure décrite précédemment.

4.3.7 Méthode 4 - Enrichissement des vésicules extracellulaires par l'approche MagNet

L'enrichissement des vésicules membranaires du plasma a été fait selon l'approche MagNet (Wu *et al.*, 2023). En résumé, 50 µL de plasma ont été dilués dans 50 µL de tampon 100 mM BTP, 150 mM NaCl, pH 6,3. Par la suite, 25 µL de billes magnétiques SAX (ReSyn Biosciences), ont été équilibrées deux fois avec 200 µL de tampon 50 mM BTP, 150 mM NaCl, pH 6,3 puis combinées aux échantillons de plasma dilué (100 µL). Les réactions ont été incubées à température pièce, sous agitation, pendant 45 min. Les protéines non liées (~100 µL) ont été récupérées sur un séparateur magnétique pour une analyse protéomique subséquente (méthode PCA). Les vésicules membranaires liées à la résine ont été lavées trois fois avec 200 µL 50 mM BTP, 150 mM NaCl, pH 6,3. Le premier lavage (~200 µL) a été récupéré et combiné aux protéines non liées pour la méthode PCA. Les vésicules membranaires liées aux billes ont été réduites et solubilisées en ajoutant 100 µL de tampon 50 mM Tris-HCl pH 8,1; 1% SDS, 10 mM TCEP (37°C, 600 rpm, 10 min) et alkylées en ajoutant 15 µL d'IAM 100 mM (37°C, 600 rpm, 30 min dans le noir). La réaction a été finalement quenchée en ajoutant 10 µL de DTT 100 mM (37°C, 600 rpm, 15 min). Les protéines ont été précipitées sur les billes en ajoutant 325 µL ACN 100%. Les échantillons ont été légèrement mélangés et incubés à température pièce pendant 10 min. Les billes ont été lavées trois fois avec 200 µL d'ACN 95%. Pour finir, les billes ont été resuspendues dans 400 µL de tampon 100 mM ABC, pH 8,5 et les protéines ont été digérées à 37°C, 16h, avec 2 µg de trypsine. Les peptides ont été extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA) selon la procédure décrite précédemment.

4.3.8 Méthode 5 - Déplétion des protéines abondantes à partir d'acide perchlorique

La déplétion des protéines abondantes a été faite selon une version modifiée de l'approche décrite par Viode *et al.* (Viode *et al.*, 2023). Aux 300 µL de protéines non liées (MagNet) ont été ajoutés 200 µL H₂O et 25 µL de PCA. Les mélanges ont été vortexés avant d'être incubés à -20°C pendant 15 min. La déplétion des protéines abondantes a été menée en centrifugeant à faible vitesse les mélanges (3200 g, 4°C, 60 min). Au surnageant déplété a été ajouté 40 µL 1% TFA. Les protéines ont ensuite été extraites sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA). Dans ce protocole, les cartouches ont été conditionnées avec 1 mL MeOH et 1 mL 0,1% TFA. Après avoir appliqué les échantillons, les protéines ont été lavées avec 1 mL de TFA 0,1% et éluées deux fois avec 500 µL d'une solution 90% ACN; 0.1% TFA. Les échantillons ont finalement été évaporés au Speed Vac. Les protéines ont été reconstituées dans 200 µL de tampon 100 mM ABC, pH 8,5 et soniquées 10 min dans un bain à ultrasons. Les protéines ont été réduites en ajoutant 10 µL de DTT 100 mM (37°C, 650 rpm, 20 min) et alkylées en ajoutant 15 µL d'IAM 100 mM (37°C, 650 rpm,

30 min dans le noir). Pour finir, celles-ci ont été digérées à 37°C, 16h, avec 2 µg de trypsine. Les peptides ont été extraits sur cartouche OASIS HLB de 30 mg (Waters, Milford, MA) selon la procédure décrite précédemment.

4.3.9 Analyse LC-HRMS/MS

Les échantillons ont été reconstitués dans 100 µL d'une solution 10% ACN; 0,2% FA. Les échantillons ont par la suite été dilués soit 1/20 dans de l'eau contenant 0,1% FA (PPT, PPT-IgG, ProTrap et MagNet) ou 1/10 (PCA). 20 µL de chaque échantillon dilué ont été chargés sur des embouts EvoTip Pure™ (EVOSEP) et analysés en utilisant un système microLC Evosep One (EVOSEP) couplé à un ZenoTOF 7600 (quadripôle-temps de vol) (Sciex, Concord, ON). Les échantillons ont été analysés à partir de la méthode Evosep One 30 SPD (*sample per day*) avec un gradient de 44 min. Des phases mobiles composées d'eau (A) et d'ACN (B), contenant chacune 0,1% FA, ont été utilisées. La séparation a été faite à 40 °C, à un débit de 0,5 µL/min et en suivant un gradient croissant en ACN. Les embouts EvoTip Pure intègrent à la fois une étape de dessalage, de chargement et d'élution des peptides dans le système Evosep One. Une fois les échantillons chargés sur les embouts, l'Evosep One élue une fraction de l'échantillon dans son système, retenant certaines molécules hydrophobes et de grosse taille dans les embouts, permettant de diminuer les effets de *carry-over* entre les analyses. L'intégralité de la procédure de préparation de la pointe EvoTip Pure est répertoriée dans la figure 4.2.

L'acquisition des spectres TOF-MS et MS/MS à haute résolution des peptides a été faite à partir d'un ZenoTOF 7600 (quadripôle-temps de vol) (Sciex, Concord, ON) à ionisation électrospray (ESI) en mode positif. Les paramètres pour la source d'ionisation ont été de 35 psi pour le gaz de rideau (CUR), 10 psi pour le gaz nébuliseur (azote) et 25 psi pour le gaz de séchage (air sec), 150°C pour la température de la source, 4500 V pour le voltage de la source ESI et 80 V pour le potentiel de dégroupage (DP).

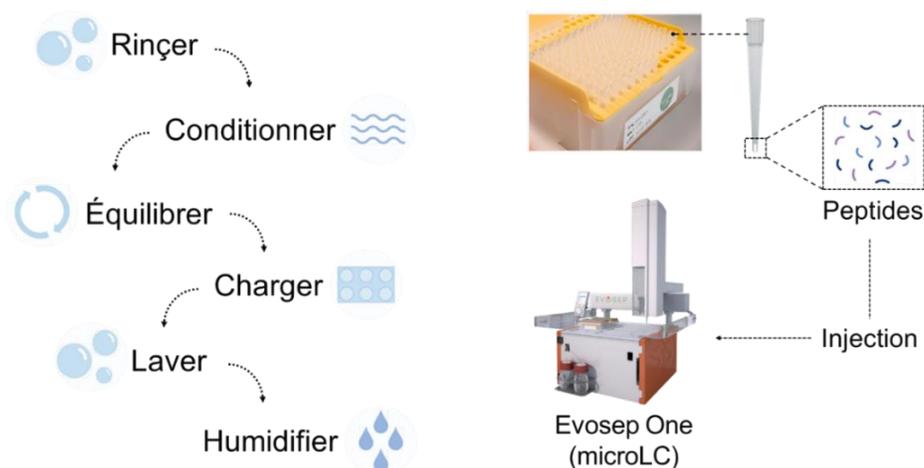


Figure 4.2: Flux de travail associé au chargement des échantillons sur les Evtip Pure. Rincer les embouts avec 20 μl de solvant B (10% ACN, 0,1% FA) et centrifuger (800 g, 1min). Conditionner les pointes des embouts en les trempant dans l'isopropanol jusqu'à ce que toutes les pointes soient d'un blanc pâle. Équilibrer les embouts avec 20 μl de solvant A (H_2O , 0,1% FA) et centrifuger (800 g, 1min). Charger 20 μl d'échantillon et centrifuger (800 g, 1min). Laver les embouts avec 20 μl de solvant B et centrifuger (800 g, 1min). Transférer 200 μl de solvant A et centrifuger pendant 10 s à 800 g pour garder les pointes humides. Les échantillons maintenus dans les embouts sont alors directement injectés dans le système microLC Evosep One à partir de la méthode 30 SPD (gradient de 44 min).

4.3.9.1 Acquisition Dépendante des Données (DDA)

L'identification des protéines a été faite en mode DDA. Pour chaque cycle d'acquisition, une première acquisition TOF-MS a été faite entre m/z 250-1300 avec un temps d'accumulation de 250 ms. Les 25 ions les plus intenses ont ensuite été sélectionnés pour l'acquisition des spectres MS/MS entre m/z 80-1500. Le temps d'accumulation pour chaque MS/MS a été de 25 ms et une énergie de collision de 10V a été appliquée. Le temps de cycle total était de 1,024 s.

4.3.9.2 Acquisition Indépendante des Données (DIA-SWATH)

La quantification relative des protéines a été effectuée par une analyse DIA-SWATH. Une acquisition TOF-MS (m/z 250-1250 avec un temps d'accumulation de 100 ms) suivie de 95 expériences TOF-MS/MS (m/z 100-1500 avec un temps d'accumulation de 8 ms) avec des fenêtres d'acquisition Q1 de 7Da (m/z 350-850) et de 15Da (m/z 850-1200) a été réalisée. Une énergie de collision de 10V a été appliquée avec un temps de cycle total de 1.333 s.

4.3.10 Traitement des données

4.3.10.1 Identification des protéines à partir des données DDA et DIA-SWATH

Les données DDA issues des dilutions 100% plasma humain ont été analysées à partir du logiciel commercial PEAKS® Studio 12 (Bioinformatics Solutions Inc., Waterloo, Canada). Une base de données compilant les protéines humaines (UniProt, 2024, 20 754 séquences) a été utilisée afin d'effectuer les recherches. Les recherches ont été effectuées selon les critères mentionnés précédemment. Le seuil d'identification des protéines et des peptides a été fixé à 1 % FDR.

Les données DIA issues des dilutions 100% plasma humain ont été analysées à partir du logiciel commercial PEAKS® Studio 12 (Bioinformatics Solutions Inc., Waterloo, Canada). Une base de données compilant les protéines humaines du plasma (UniProt, janvier 2024, 6270 entrées) ou une librairie spectrale produite à partir des données DDA ont été utilisées afin d'effectuer les recherches. Les recherches ont été effectuées selon les critères mentionnés précédemment. Le seuil d'identification des protéines et des peptides a été fixé à 1 % FDR.

Les données DIA issues des dilutions 100% plasma humain ont été analysées à partir du logiciel de bureau DIA-NN (version 1.8.1) en mode sans librairie (*library-free*). Une base de données compilant les protéines humaines du plasma (UniProt, janvier 2024, 6270 entrées) a été utilisée pour créer la librairie spectrale *in silico*. Les données DIA-SWATH ont par la suite été recherchées à partir de cette librairie selon les critères de recherches suivants : protéase (Trypsine), jusqu'à un clivage manqué (*missed cleavages*) et aucune PTM variable par peptide. La carbamidométhylation (C) et l'excision des méthionines en N-terminal ont été fixées comme modifications. Les répliques pour chaque condition individuelle ont été recherchées simultanément avec la correspondance entre les exécutions (*Match Between Runs*).

Le logiciel Panther (Mi *et al.*, 2013) a été utilisé pour l'analyse de la classification des protéines (*protein class*). PeptideAtlas (*Human Plasma 2023-04*) (Desiere *et al.*, 2006) a été utilisé pour l'analyse des concentrations en protéines. Les cartes thermiques (*heatmaps*) ont été générées à partir de NG-CHM Builder (MD Anderson Cancer Center) (Ryan *et al.*, 2019). Le clustering hiérarchique (*Hierarchical clustering*) a été appliqué pour les lignes en spécifiant la métrique de distance Euclidienne et l'agglomération Ward. Aucune covariable n'a été attribuée.

4.3.10.2 Analyse des courbes d'étalonnages adaptée à la matrice

Les données DDA issues des dilutions 100% plasma humain ont été analysées à partir du logiciel commercial PEAKS® Studio 12 (Bioinformatics Solutions Inc., Waterloo, Canada). Une première librairie spectrale a été créée à partir des données DDA des échantillons 100% en plasma humain. Les recherches ont été effectuées à partir d'une base de données compilant l'ensemble des protéines humaines et de poulet répertoriées dans UniProt (juillet 2024, 38 960 entrées). Une tolérance de masse MS de 15 ppm et MS/MS de 0,05 Da ont été autorisées. Les recherches ont été effectuées en spécifiant la trypsine comme enzyme et jusqu'à 2 sites de clivage manqués. La carbamidométhylation des cystéines a été définie comme modification fixe, et l'oxydation des méthionines et la désamidation des asparagines comme variables. Un maximum de 2 PTM a été autorisé par peptide. Le seuil d'identification des protéines et des peptides a été fixé à 1 % FDR. Une fois la librairie créée, les peptides partagés entre les deux espèces ont été manuellement retirés de celle-ci. La librairie spectrale finale, contenant les séquences uniques à l'humain, a été utilisée afin d'effectuer les recherches sur les données quantitatives DIA-SWATH pour l'ensemble des dilutions.

L'ensemble des données DIA-SWATH (dilution 0 à 100%) ont été analysées à partir du logiciel commercial PEAKS® Studio 12 (Bioinformatics Solutions Inc., Waterloo, Canada). Les recherches ont été effectuées à partir de la librairie spectrale contenant les séquences uniques à l'humain. Une tolérance de masse MS de 15 ppm et MS/MS de 0,05 Da ont été autorisées. Les recherches ont été effectuées selon les critères mentionnés précédemment. Les différentes conditions ont été recherchées simultanément avec la correspondance entre les exécutions (*Match Between Runs*). Le seuil d'identification des protéines et des peptides a été fixé à 1 % FDR.

4.4 Résultats et discussion

Au courant de la dernière décennie, le domaine de la protéomique plasmatique a subi un regain en popularité, suscitant l'intérêt de plusieurs groupes de recherches et ayant mené au développement d'une très grande variété d'approches (enrichissement, déplétion, fractionnement). Le choix étant tellement vaste, nous avons cherché à établir et comparer l'efficacité et la performance de plusieurs de ces méthodes, et ce dans le but d'établir quelle méthode serait la mieux adaptée afin de procéder à des analyses sur des échantillons d'intérêt biologique (chapitre 5). Dans ce projet, différentes méthodes commerciales ont initialement été testées dont plusieurs ensembles d'extraction et d'enrichissement de la compagnie PreOmics

(iST-BCT, ENRICH-iST et SP3-iST Add-on) ainsi qu'une approche d'immunodéplétion sur cartouche MARS (*Multiple Affinity Removal System*) de la compagnie Agilent. Bien que ces approches ne se sont pas particulièrement démarquées par rapport aux méthodes que nous allons présenter, celles-ci sont également excessivement dispendieuses et dans certains cas extrêmement fastidieuses (MARS) limitant largement les études à grande échelle.

La figure 4.3 présente de manière schématisée les cinq méthodes qui seront comparées dans ce chapitre. Des cinq méthodes présentées, nous allons distinguer trois approches basées sur l'extraction des protéines à partir de solvants organiques à savoir PPT, PPT-IgG et ProTrap. Dans la première approche (PPT), les protéines du plasma sont tout d'abord solubilisées à partir de SDS puis précipitées à l'aide d'acétone. Les protéines vont finalement être resolubilisées, réduites, alkylées et digérées par la trypsine. Dans la seconde approche (PPT-IgG), le plasma est incubé à température pièce pendant 30 min avec des billes *Dynabeads Protein G* et ce afin de dépléter les immunoglobulines de type G (IgG). Le plasma déplété est alors récupéré avant d'être soumis aux mêmes étapes de solubilisation, précipitation et digestion enzymatique que pour l'approche PPT. En ce qui concerne l'approche ProTrap, les protéines plasmatiques vont être extraites à partir de l'ensemble commercial de plaques filtrantes 96 puits ProTrap de la compagnie Allumiqs. Cette méthode consiste en une solubilisation des protéines avec du SDS, suivie de leur alkylation réductrice, de leur précipitation à partir d'une solution acétone : MeOH et pour finir de leur digestion par la trypsine, le tout sur une plaque filtrante 96 puits.

Une approche de déplétion à partir d'acide perchlorique (PCA) a également été testée. Cette méthode développée par Viode *et al.*, consiste en une précipitation sélective des protéines abondantes du plasma à l'aide d'acide perchlorique (Viode *et al.*, 2023). Par une étape de centrifugation à basse vitesse (3200 g, 60 min), il est alors possible de récupérer un surnageant déplété en protéines. Bien que son mécanisme d'action ne soit pas bien compris, certaines études suggèrent que l'acide perchlorique favoriserait la solubilisation des glycoprotéines (Lee *et al.*, 2016).

Le plasma étant composé de 1 à 2% en vésicules extracellulaires (exosomes, microvésicules, corps apoptotiques) (Wu *et al.*, 2023), celles-ci représentent également une potentielle source en biomarqueurs. L'enrichissement de ces compartiments a été effectué selon l'approche MagNet, initialement décrite par Wu *et al* (Wu *et al.*, 2023). Cette méthode utilise des billes fonctionnalisées de groupements ammonium quaternaire afin d'enrichir les vésicules. Dans notre méthode, le

plasma est incubé avec les billes SAX (*Strong Anion Exchange*) selon un ratio plasma : billes de 2 : 1, et à température pièce pendant 45 min. Les vésicules liées sont alors lysées à partir de SDS. Les protéines libérées sont finalement réduites, alkylées, précipitées sur billes et digérées à l'aide de trypsine.

Lors de ce test, nous avons décidé de combiner l'approche MagNet et PCA dans la mesure où les protéines abondantes non-liées par MagNet ont été récupérées afin de procéder à leur déplétion. L'idée de leur combinaison présente de nombreux avantages. Cette approche nécessitant moins de plasma comparativement à deux expériences menées indépendamment, ceci serait idéal lorsque le volume d'échantillon est limité. De plus, il serait potentiellement possible d'augmenter le recouvrement de certaines protéines qui n'auraient pas été extraites par l'une ou l'autre des méthodes.

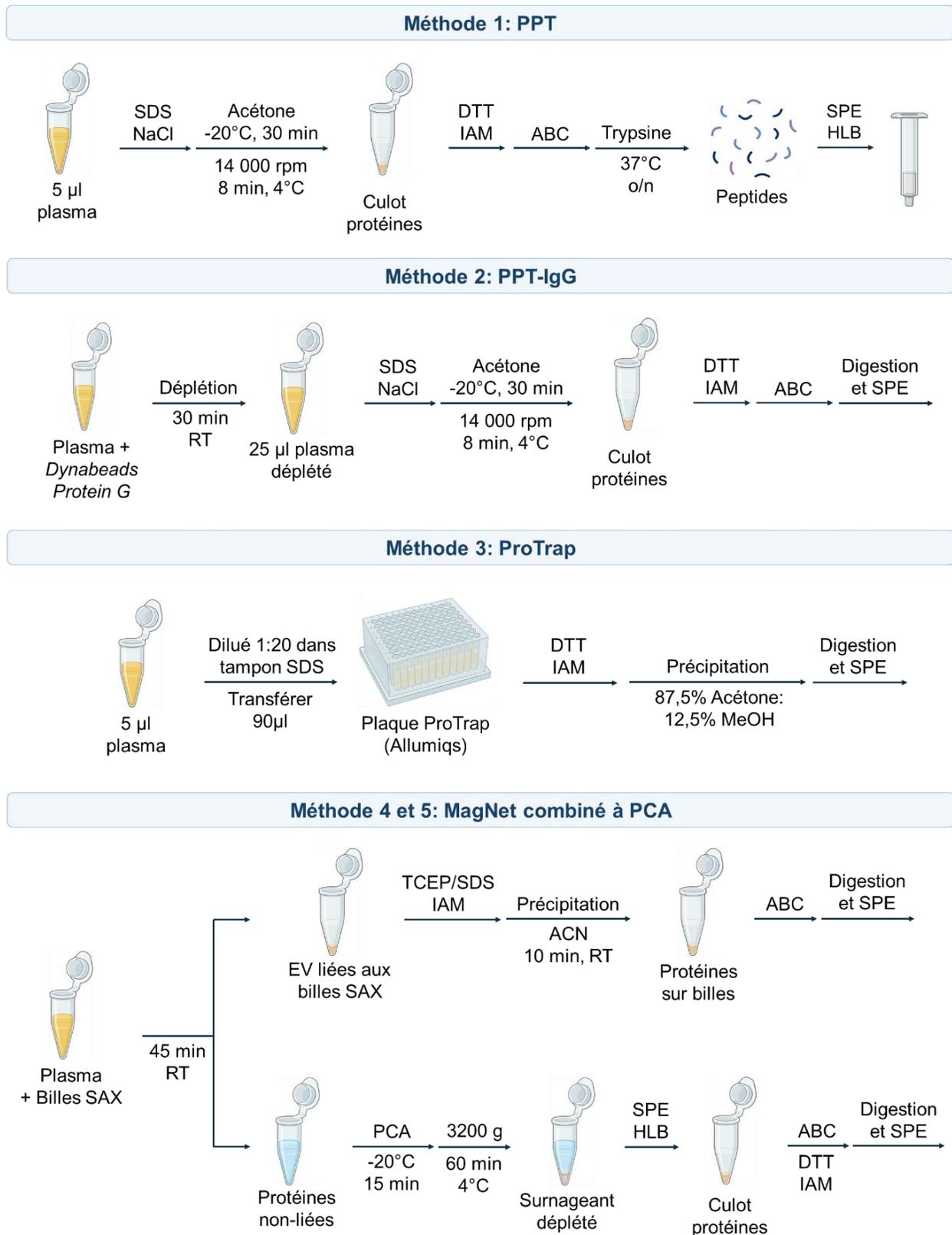


Figure 4.3: Les différentes méthodes de préparation d'échantillon testées pour la protéomique plasmatique ascendante

4.4.1 Identification et caractérisation des protéines extraites

Pour cette partie de l'étude, les échantillons composés à 100% en plasma humain ont été analysés en DDA et DIA-SWATH afin d'établir l'efficacité et la performance des différentes méthodes testées. Le diagramme à barres présenté à la figure 4.4A montre le nombre de protéines identifiées par chacune des méthodes ainsi que leur reproductibilité. Comme présenté, ProTrap permet un meilleur recouvrement des protéines avec 325 protéines identifiées en moyenne au travers des répliques comparativement à PPT (291), mais également par rapport à MagNet (256), PCA (253) et PPT-IgG (266), soit nos approches basées sur l'enrichissement ou la déplétion des protéines. Il est également possible d'observer une excellente reproductibilité dans le nombre de protéines identifiées dans les trois répliques pour la méthode ProTrap.

Par la suite, nous avons cherché à étudier plus en détail la relation entre le recouvrement protéomique et le type de protéines extraites. Le diagramme de Venn présenté à la figure 4.4B compare l'ensemble des protéines identifiées par les différentes méthodes. Bien que ProTrap permet un meilleur recouvrement des protéines, une grande proportion des protéines ont été identifiées comme étant unique aux approches PCA (19.8%) et MagNet (7,4%). De plus, il est également intéressant de mentionner que plusieurs des protéines extraites sont partagées entre les différentes approches basées sur la précipitation des protéines (ProTrap, PPT et PPT-IgG). Ces résultats suggèrent qu'en fonction des méthodes utilisées, celles-ci permettraient d'isoler différents groupes de protéines montrant ainsi une complémentarité entre les différentes approches testées, mais montrant également la complexité du protéome plasmatique. Ces observations peuvent également être faites à partir des résultats quantitatifs DIA-SWATH présentés à la figure 4.4C où il est possible d'observer des patrons d'extraction distinct entre les méthodes basées sur la précipitation des protéines (ProTrap, PPT et PPT-IgG) de ceux induit par MagNet et PCA. Par conséquent, compte tenu des résultats de couverture protéomique obtenus pour ProTrap et de la grande complémentarité offerte par les méthodes MagNet et PCA, seules ces trois méthodes seront prises en considération pour les analyses subséquentes.

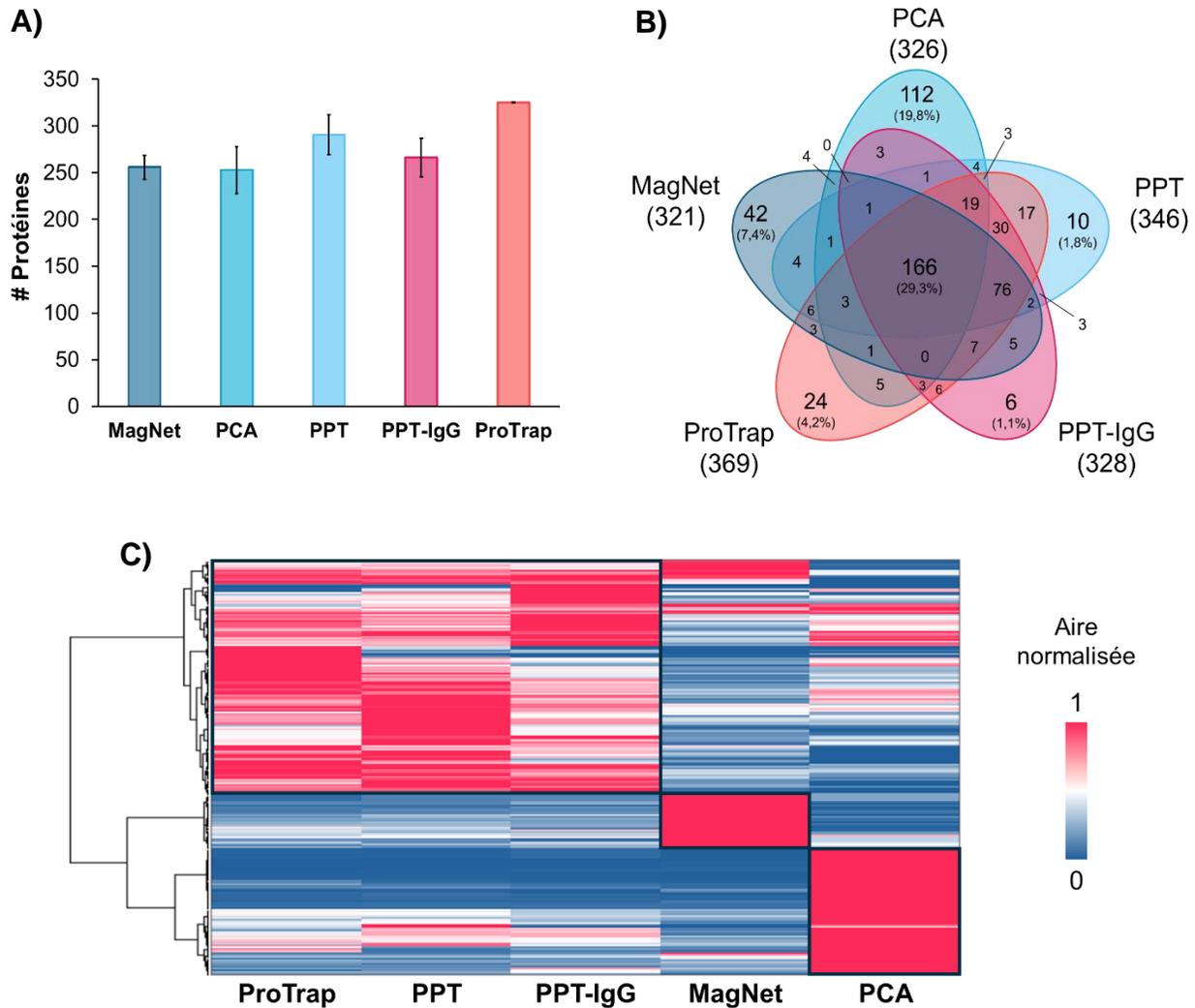


Figure 4.4: Analyse comparative des différents flux de travail utilisés pour l'extraction des protéines du plasma. A) Diagramme à bande montrant la reproductibilité des méthodes, B) Diagramme de Venn comparant les protéines identifiées en DDA pour chacune des méthodes et C) analyse quantitative DIA-SWATH comparant les protéines enrichies. Chaque méthode a été testées en triplicata (n=3).

Dans le but de caractériser les protéines uniques à MagNet, PCA et ProTrap identifiées en DDA, différentes analyses ont été réalisées. Le logiciel en ligne PANTHER a été utilisé pour effectuer une analyse de type classification des protéines (*protein class*). Les protéines ont également été recherchées dans PeptideAtlas (*Human Plasma 2023-04*) afin d'établir la plage de concentrations des protéines identifiées. Comme l'en démontre le diagramme en violon présenté à la figure 4.5B, les approches PCA et MagNet permettent, de manière générale, d'identifier des protéines présentes en plus faible concentration dans le plasma comparativement à ProTrap. En effet,

comme anticipé, ProTrap extrait majoritairement du plasma des protéines abondantes connues (Figure 4.5B), celles-ci étant généralement associées à l'immunité (39) (Figure 4.5A) et étant majoritairement des immunoglobulines (37). À l'inverse, la combinaison de MagNet et de PCA permet quant à elle de révéler la diversité fonctionnelle du protéome plasmatique en identifiant une série de protéines associées par exemple à la liaison de la chromatine/régulation, au trafic membranaire, à l'adhésion cellulaire ou encore à régulation transcriptionnelle, pour ne citer que celles-ci (Figure 4.5A). Ceci est un aspect très intéressant et important d'un point de vue biologique. Il serait par exemple possible d'identifier à partir de ces deux méthodes certaines voies biochimiques inconnues et potentiellement altérées dans certaines maladies.

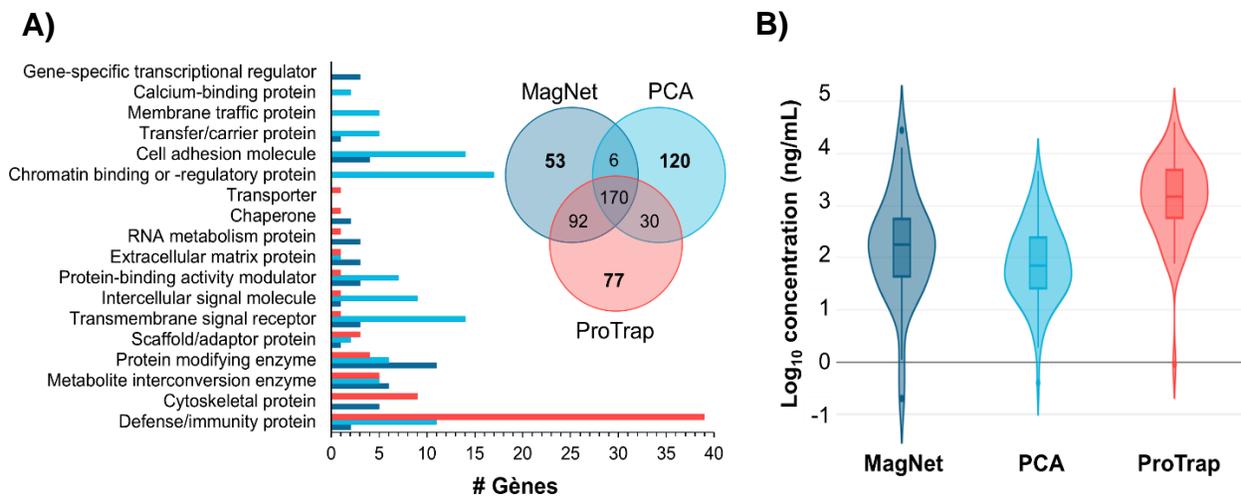


Figure 4.5: Caractérisation des protéines extraites par MagNet, PCA et ProTrap. A) Analyse Panther (protein class) pour les protéines uniques à chacune des méthodes et B) plage de concentrations couverte par ces protéines. Les concentrations de chaque protéine proviennent de PeptideAtlas (*Human Plasma 2023-04*)

Il est possible d'identifier de façon fiable, sans méthode de déplétion ou d'enrichissement, environ 300 protéines plasmatiques (Viode *et al.*, 2023). Les résultats d'identification pour l'approche PCA varient grandement d'une étude à l'autre. Par exemple, Viode *et al.* ont identifié plus de 1500 protéines à partir de la méthode de précipitation au PCA (Viode *et al.*, 2023), alors que Rice & Belani ont identifié environ 200 protéines en DDA et 600 en DIA avec cette même méthode (Rice & Belani, 2024). Dans la même perspective, selon une étude récente, l'approche MagNet permettrait de détecter plus de 4000 protéines plasmatiques (Wu *et al.*, 2023). Il est important de mentionner ici qu'il est difficile de comparer l'ensemble de ces résultats à ceux que nous avons

obtenus dans cette étude. En effet, plusieurs facteurs peuvent affecter l'identification des protéines et expliquer ces différences dont le type de système LC-MS/MS utilisé, le débit de la phase mobile HPLC, le mode d'acquisition des données (DDA vs DIA), le choix des logiciels de traitement de données, le type de librairie spectrale utilisé pour effectuer les recherches DIA, et bien plus. Par exemple, dans le cas de notre étude, nous avons utilisé un microLC comparativement à un nanoLC pour Rice & Belani, et Wu *et al.* (Rice & Belani, 2024; Wu *et al.*, 2023). Les systèmes nanoLC, opérant à des débits beaucoup plus faibles, ceux-ci sont connus pour améliorer la sensibilité de détection des peptides en favorisant entre autres leur ionisation par ESI (Long *et al.*, 2025; Bian *et al.*, 2020). Toutefois, les systèmes nanoLC sont peu stables dans le temps, pouvant affecter la reproductibilité et la robustesse de la méthode (Bian *et al.*, 2020).

Le traitement des données peut également influencer sur l'identification des protéines par spectrométrie de masse. Afin d'illustrer ceci, nous avons comparé ici trois méthodes de traitement des données pour l'identification des protéines à partir des données DIA-SWATH à savoir : (1) recherche des données à partir d'une librairie générée *in silico* dans DIA-NN, (2) recherche en mode *library-free* à partir du logiciel PEAKS et (3) recherche dans PEAKS à partir d'une librairie spectrale générée à partir des données DDA. La base de données compilant les protéines humaines du plasma (Uniprot, janvier 2024, 6270 entrées) a été utilisée afin de générer la librairie *in silico* dans DIA-NN et pour effectuer les recherches en mode *library-free* dans PEAKS (PEAKS DB). Pour les recherches effectuées à partir de la librairie spectrale DDA (PEAKS DDA), les données DDA issues des échantillons 100% plasma humain ont tout d'abord été recherchées à partir de la base de données compilant les protéines humaines (Uniprot, 2024, 20 754 entrées). La librairie générée a par la suite été utilisée afin d'effectuer les recherches des données DIA-SWATH dans PEAKS.

La figure 4.6A présente le nombre de protéines identifiées par ces différentes approches. De manière générale, il a été possible d'identifier quelques protéines de plus à partir de la librairie DDA (PEAKS DDA) comparativement aux recherches menées à partir de DIA-NN et PEAKS DB. Ceci peut être expliqué par le fait que plus de protéines composaient la base de données initialement utilisée pour créer la librairie DDA. Il aurait été possible d'effectuer les recherches DIA-NN et PEAKS DB à partir de cette même base de données et ce, afin d'augmenter le nombre d'identifications. Toutefois, il faut savoir qu'une meilleure couverture protéomique vient souvent au détriment de temps de recherches plus longues pouvant limiter les analyses à larges échelles.

L'approche utilisant une librairie DDA est également connue pour être une approche très longue. En effet, celle-ci implique d'injecter les échantillons à la fois en mode DDA et DIA-SWATH, de traiter initialement les données DDA afin de créer la librairie et pour finir d'utiliser cette librairie pour rechercher les données DIA-SWATH.

En ce qui concerne le nombre de protéines identifiées par DIA-NN et PEAKS DB (Figure 4.6A), les résultats sont somme toute très similaires, à l'exception de PPT où il a été possible d'identifier environ 300 protéines dans DIA-NN et 200 protéines dans PEAKS. Toutefois, en comparant l'ensemble de ces identifications (Figure 4.6B), il est possible d'observer des différences majeures entre DIA-NN et PEAKS DB et ce, malgré le fait que la même base de données ait été utilisée pour effectuer les recherches. Il faut savoir qu'il existe une multitude d'outils bio-informatiques permettant d'analyser les données DIA, tous étant basés sur différents principes de recherche et utilisant différents algorithmes. Par exemple DIA-NN, un logiciel *open-source* conçu par le laboratoire du Pr. Lilley à l'université de Cambridge au Royaume-Uni, utilise une approche de recherche dite centrée sur les peptides (*peptide-centric*), c'est-à-dire que le logiciel génère à partir d'une base de données de séquences en protéines une librairie prédite *in silico* contenant des spectres et des temps de rétention théoriques (Rice & Belani, 2022; Demichev *et al.*, 2020). À l'inverse, PEAKS, un outil commercial développé par la compagnie Bioinformatics Solutions Inc., utilise une approche dite sans librairie (*library-free*) où les données DIA sont directement analysées par séquençage *de novo* à partir d'une base de données en protéines (Zhang *et al.*, 2020).

De plus, plusieurs des protéines identifiées par DIA-NN et PEAKS DB leur sont uniques. En effet, comme mentionné précédemment, le biais induit par le mode DDA peut mener à des valeurs manquantes lors de l'identification des protéines en DIA lorsqu'on utilise une librairie spectrale DDA. Par exemple, pour la méthode PPT-IgG (Figure 4.6B), ~19% de toutes les protéines identifiées sont uniques à DIA-NN et PEAKS DB (32 + 15 + 20), et ne sont donc pas identifiées en DDA. Toutefois, bien que ces outils aient démontré leur immense potentiel dans l'identification impartiale (*unbiased*) des protéines, il faut être conscient des risques associés à l'identification de faux positifs ou négatifs (Zhang *et al.*, 2020). Les librairies spectrales DDA permettent quant à elles une identification robuste des protéines, et ce à un très haut niveau de confiance puisque étant basée sur l'identification antérieure des protéines selon nos conditions LC-MS/MS.

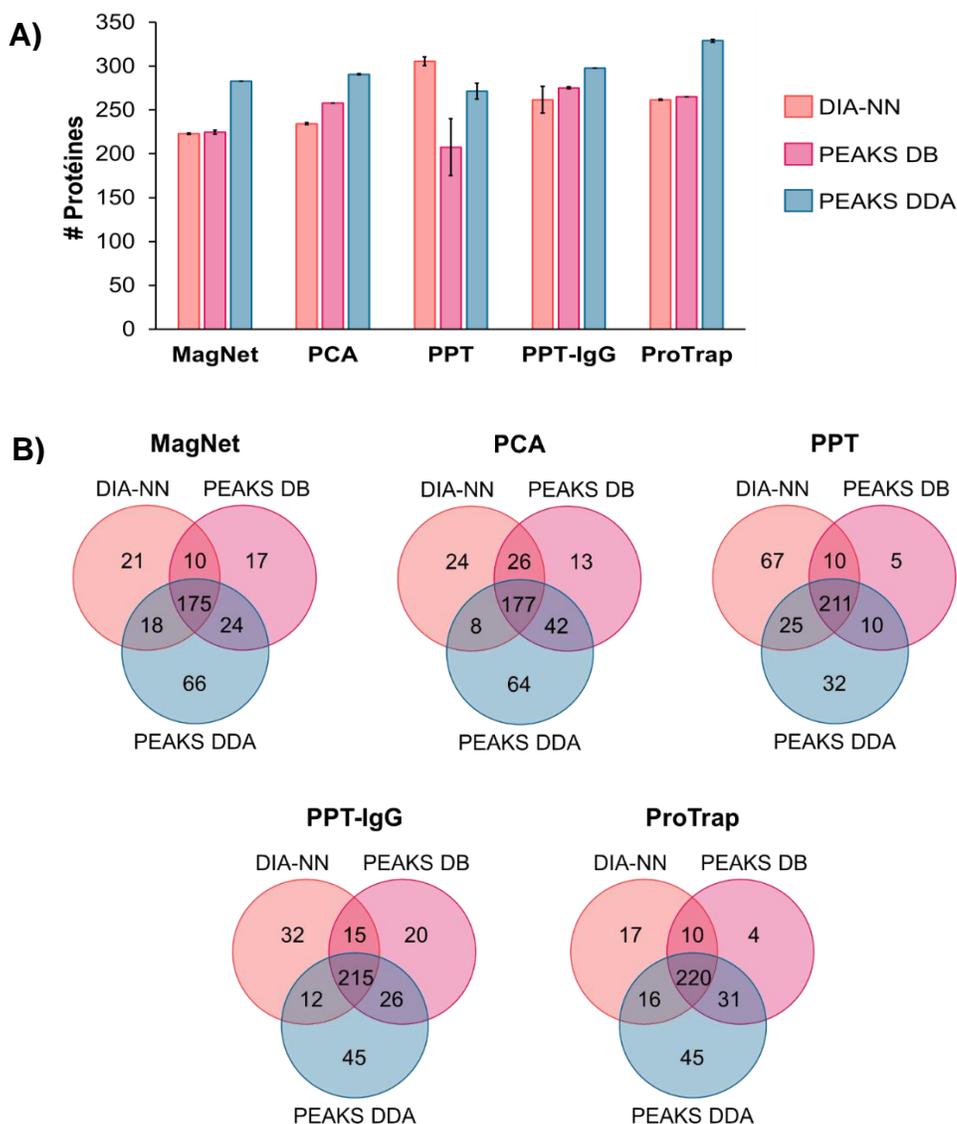


Figure 4.6: Comparaison des logiciels de traitement des données DIA-NN et PEAKS pour l'identification des protéines à partir des données DIA-SWATH. A) Diagramme à bande montrant le nombre de protéines identifiées et B) diagrammes de Venn comparant l'ensemble des identifications. Les données DIA-SWATH issues des échantillons composés à 100% en plasma humain ont été recherchées dans DIA-NN ou PEAKS DB à partir de la base de données des protéines plasmatiques humaines (Uniprot, janvier 2024, 6270 entrées).

4.4.2 Évaluation de la précision quantitatives par l'analyse de courbes d'étalonnage adaptées à la matrice

Bien que l'identification des protéines par LC-MS/MS soit une mesure importante afin d'établir l'efficacité d'une méthode, il est également important d'établir la précision quantitative de celle-ci. En effet, en protéomique, la quantification des protéines est un aspect tout aussi important que l'identification, dans la mesure où une protéine/peptide peut être identifié, mais ne pas être mesuré

de manière quantitative. Par conséquent, afin d'établir la précision quantitative des différentes méthodes, nous avons procédé à l'analyse de courbes d'étalonnage adaptées à la matrice (*Matrix-matched calibration curve*). Cette approche consiste à diluer la matrice d'intérêt, le plasma humain, dans une matrice de complexité comparable, dans ce cas-ci du plasma de poulet (Wu *et al.*, 2023; Pino *et al.*, 2020). Dans cette expérience, trois courbes d'étalonnage préparées indépendamment, et comprenant chacune 6 points de dilution humain : poulet, ont été préparées, pour un total de 18 dilutions. Ainsi, chaque dilution a été soumise aux 5 méthodes décrites précédemment (Figure 4.7A).

Pour cette partie de l'expérience, une librairie spectrale contenant les séquences uniques à l'humain a dû être créée. Pour se faire, les données DDA issues des échantillons 100% plasma humain ont tout d'abord été recherchées dans une base de données combinant l'ensemble des protéines humaines et de poulet répertoriées dans UniProt. Une fois la librairie créée, les peptides partagés entre les deux espèces ont été manuellement retirés de celle-ci. La librairie spectrale finale, contenant les séquences uniques à l'humain (16 577 entrées) a été utilisée afin d'effectuer les recherches sur les données quantitatives DIA-SWATH pour l'ensemble des dilutions de la courbe. À partir de ces données, nous avons calculé le coefficient de détermination R^2 de chaque peptide et protéine quantifié et ce afin d'établir la relation entre leur aire et le % de plasma humain attendu (Figure 4.7B).

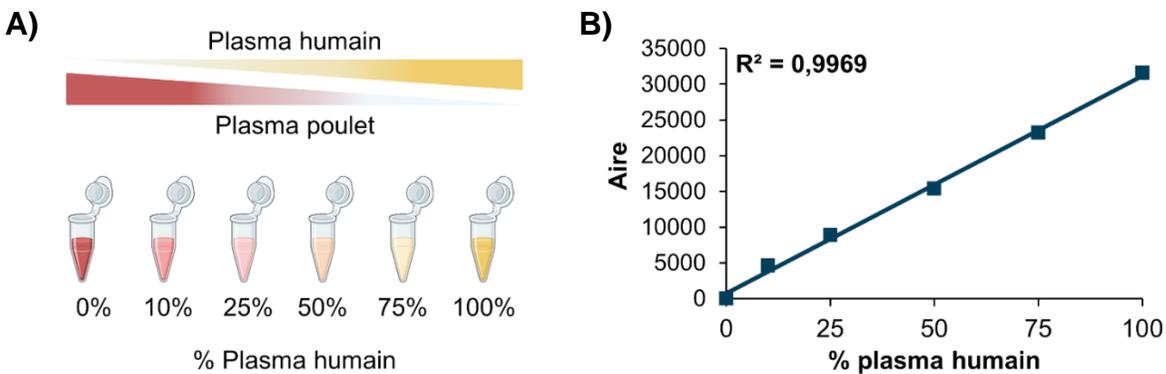


Figure 4.7: Principe de l'analyse des courbes d'étalonnage adaptées à la matrice (*Matrix-matched calibration curve*). A) Dilutions plasma humain : plasma de poulet utilisées pour construire la courbe d'étalonnage. Chaque courbe a été préparée en triplicata (n=3). B) Relation entre le % en plasma humain et l'aire du signal permettant de déterminer le coefficient de détermination R^2 . Exemple pour la protéine Apolipoprotéine M identifiée dans PCA. Les protéines/peptides dont le R^2 est supérieur à 0,9 ont été considérés comme ayant une bonne corrélation.

Selon les résultats présentés dans le tableau 4.1, les méthodes PPT et ProTrap montrent une meilleure linéarité dans la quantification des peptides et des protéines comparativement aux différentes méthodes basées sur la déplétion et l'enrichissement des protéines. En effet, 48% des peptides identifiés par PPT et 78,7% des protéines présentent un R^2 supérieur à 0,9. Pour ProTrap, ce sont 41% des peptides et 68% des protéines. Les résultats pour MagNet sont d'environ 33% pour les peptides et 56% pour les protéines. En ce qui concerne PCA, ce sont environ 39% des peptides et 67% des protéines qui possèdent un R^2 supérieur à 0,9. Pour terminer, environ 21% des peptides et 37% des protéines présentaient un R^2 supérieur à 0,9 pour PPT-IgG.

Par conséquent, il serait possible de conclure à ce stade que PPT et ProTrap permettraient une quantification plus juste des protéines et des peptides. Toutefois, il est important d'établir les limites liées à l'approche utilisée afin de déterminer la précision quantitative des approches testées. Bien que cette approche permette de construire une courbe d'étalonnage pour l'ensemble des peptides détectés dans la matrice de référence, celle-ci sera limitée par l'abondance naturelle du peptide dans cette même matrice. Pour des peptides présents en faible abondance, ceci aura pour effet de restreindre leur plage linéaire. De plus, ces peptides vont généralement produire un faible signal qui, lui, peut devenir nul lorsque dilué (Pino *et al.*, 2020). Comme les résultats l'ont démontré, MagNet et PCA favorisent justement l'identification de protéines peu abondantes comparativement à ProTrap, et potentiellement par rapport à PPT aussi si nous prenons en considération son mécanisme d'action. Ceci pourrait alors expliquer les résultats obtenus ici.

Tableau 4.1: Évaluation de la précision quantitative des différentes méthodes de préparations d'échantillons en plasma par l'analyse de courbes d'étalonnage adaptée à la matrice

Méthodes	Peptides			Protéines		
	# total ID	$R^2 > 0,9$	%	# total ID	$R^2 > 0,9$	%
MagNet	7105	2312	32,5	184	103	56
PCA	4437	1723	38,8	153	102	66,7
PPT-IgG	8257	1750	21,2	167	62	37,1
PPT	8380	4020	48	174	137	78,7
ProTrap	7828	3209	41	182	124	68,1

total ID : nombre total de peptides ou protéines identifiés. $R^2 > 0,9$ correspond au nombre de peptides ou de protéines dont le coefficient R^2 calculé est supérieur à 0,9. % : rapport en pourcentage entre le nombre d'identification ayant un $R^2 > 0,9$ et le nombre total d'identification.

4.4.3 Choix des méthodes pour des études *in vivo* futures

Comme l'ont démontré les résultats, la complémentarité entre les différentes approches permettrait d'augmenter le recouvrement des protéines plasmatiques. Dans ce cas-ci, le choix final des méthodes MagNet, PCA et PPT-IgG a été motivé par une série de facteurs. Le tableau 4.2 résume les résultats obtenus et fournit une comparaison globale des méthodes testées. De manière générale, le coût par échantillon est un élément important à prendre en considération lors du choix, celui-ci pouvant constituer l'une des principales limitations lors d'analyses protéomiques à large échelle. Les approches MagNet et PCA sont très peu dispendieuses comparativement à d'autres approches actuellement offertes et comme nous l'avons montré, celles-ci peuvent facilement être combinées entre elles. De plus, bien que l'approche MagNet semble extrêmement complexe et fastidieuse en termes de manipulations, nous avons été en mesure d'adapter l'ensemble du flux de travail afin de permettre l'automatisation de l'approche sur l'Opentrons Flex (Opentrons Labworks, Inc.), le tout en maintenant une très bonne reproductibilité dans la méthode comme l'en démontre les coefficients de variation (CV) présentés à la figure 4.8.

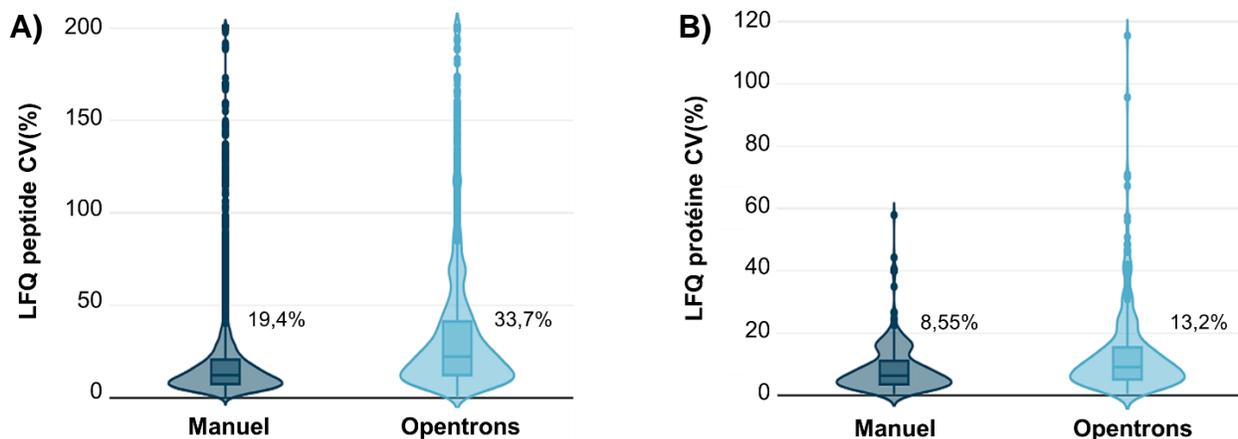


Figure 4.8: Comparaison de la distribution des coefficients de variation (CV) pour la méthode MagNet faite manuellement et l'approche automatisée sur le robot Opentrons. Coefficient de variation pour la quantification A) des peptides et B) des protéines. Dans cette expérience, du sérum de souris a été utilisé comme matrice et des réplicats techniques ont été préparés (n=4).

Dans cette étude, l'approche ProTrap a montré de bons résultats en termes d'identification, reproductibilité et quantification. En ce qui concerne PPT-IgG, bien que celle-ci ne se soit pas particulièrement démarquée dans cette partie de l'étude, il serait possible d'augmenter l'efficacité

de l'approche en utilisant moins de plasma. Ayant initialement commencé avec un volume de 25 μ l, donc environ 250 μ g d'IgG si nous prenons en considération sa concentration dans le plasma (10 mg/mL) (Bondt *et al.*, 2021), celle-ci dépasse largement la capacité de liaison de la résine qui est de 8 μ g d'IgG par mg de billes. En effet, nous avons utilisé dans cette expérience 100 μ L de billes *Dynabeads Protein G* à 30 mg/mL, donc 3 mg de billes. Alors, un maximum de 24 μ g d'IgG seront déplétés dans ce cas. En utilisant par exemple un volume de 2 μ l de plasma (\sim 20 μ g d'IgG) et 50 μ l de billes, il serait possible de dépléter \sim 12 μ g d'IgG et potentiellement augmenter l'efficacité de l'approche. De plus, cela permettrait également de diminuer de moitié les coûts associés à la méthode.

Tableau 4.2: Analyse comparative des différents flux de travail de préparation d'échantillons de plasma testés

Méthode	Identification	Reproductibilité	Coûts	Temps de préparation	Automatisation possible
PPT	291	*	\$	\sim 1h30	Non
PPT-IgG	266	*	\$\$\$\$	\sim 2h00	Non
ProTrap	325	***	\$\$	\sim 1h30	Non
MagNet	256	**	\$\$\$\$	\sim 2h30	Oui
PCA	253	*	\$	\sim 4h00	Non

Le nombre d'étoiles pour la quantification et la reproductibilité évalue la performance des méthodes : trois étoiles (***) correspondent à une bonne performance et une étoile (*) à une performance faible. Les temps de préparation ne prennent pas en considération l'étape de digestion à la trypsine.

CONCLUSION ET PERSPECTIVES

Les protéines jouant un rôle central dans la régulation d'une multitude de processus physiopathologiques, l'étude de ces molécules est alors indispensable afin de mieux comprendre leur rôle et leur fonction, et ce, en vue de découvrir de nouveaux biomarqueurs et cibles thérapeutiques. Bien que la spectrométrie de masse soit l'outil de choix pour étudier les protéines en raison de sa polyvalence, sa spécificité et sa haute sensibilité, l'étude de ces biomolécules reste difficile sans l'utilisation de méthodes de préparation d'échantillons efficaces et adaptées. Dans ce mémoire, nous avons abordé différents concepts liés au développement de méthodes pour les analyses protéomiques par LC-MS/MS, à travers trois projets de recherche différents. Nous nous sommes plus particulièrement concentrés sur le développement de méthodes analytiques afin: (1) d'étudier par LC-MS/MS les interactions protéines-protéines de la protéine RNF13, (2) de quantifier par une méthode ciblée les protéines S100 et ANX dans les larmes de patients, et (3) d'analyser les protéines du plasma par une approche LC-MS/MS non-ciblée.

Dans le premier projet de ce mémoire (Chapitre 2), nous avons tout d'abord établi une stratégie afin d'étudier les protéines biotinylées par LC-MS/MS et ce afin de l'appliquer ultérieurement à des études interactomiques de type BioID. Deux méthodes d'enrichissement sur résine streptavidine ont été comparées, une au niveau des peptides modifiés et l'autre directement à partir des protéines. Cette dernière approche s'est avérée plus efficace, permettant d'une part d'identifier les sites modifiés mais également d'augmenter la couverture de séquence des protéines identifiées et ainsi permettre la validation de leur identification. Pour finir, l'application de cette méthode afin d'étudier l'interactome associé à la protéine RNF13 a permis d'identifier plus de 130 potentiels interacteurs de cette protéine, dont plusieurs sont présentement en processus de validation.

Dans le second projet de ce mémoire (Chapitre 3), nous avons développé une méthode ciblée par LC-MRM pour la quantification de différentes protéines S100 et ANX dans les larmes de patients. Les travaux se sont initialement concentrés sur l'analyse de 17 S100 et 10 ANX standards issus d'expression bactérienne. Une analyse préalable de ces protéines par une approche non-ciblée et à partir d'un instrument de type quadripôle-temps de vol a permis la sélection de peptides tryptiques uniques à optimiser pour leur sensibilité et spécificité d'analyse par LC-MRM. Dans la méthode finale, entre 2 et 5 peptides par protéine et 2 transitions par peptide ont été sélectionnés. La méthode développée a été, par la suite, appliquée afin d'analyser les

larmes de différents groupes de patients, incluant certaines pathologies oculaires et hépatiques. Plusieurs de ces protéines se sont révélées être fortement impactées chez le groupe de patient atteints d'une maladie du foie, mettant en évidence le potentiel rôle que peuvent tenir certaines de ces protéines dans le développement de ces maladies.

Pour le dernier projet de ce mémoire (Chapitre 4), nous avons testé et comparé différentes approches de préparation d'échantillons pour l'analyse des protéines du plasma par LC-MS/MS. Dans ce chapitre, nous avons comparé cinq approches différentes à savoir ProTrap, PPT, PPT-IgG, MagNet et PCA. La sélection de ces deux dernières approches a été motivée par une série de facteurs. Mise à part leur faible coût, celles-ci peuvent facilement être combinées entre elles, permettant de réduire les quantités de plasma initialement nécessaires et offrant une belle complémentarité dans les résultats obtenus. De plus, nous avons été en mesure d'automatiser l'entièreté de l'approche MagNet, ceci permettant de simplifier grandement les analyses à haut débit.

En termes de perspective, nous envisageons appliquer à court terme plusieurs de ces méthodes afin d'analyser différents types d'échantillons biologiques. Par exemple, grâce à deux collaborations menées respectivement avec le laboratoire de la Pre. Catherine Mounier et celui du Pr. Nicolas Pilon au département des sciences biologie de l'UQAM, nous avons eu accès à différents types d'échantillons de plasma de souris provenant soit d'un modèle souffrant d'un dysfonctionnement métabolique et ayant été soumis à deux diètes différentes, ou d'un modèle murin de la maladie de Hirschsprung. L'application de nos différents flux de travail pour le plasma, à savoir MagNet et PCA (figure 5.1), nous permettrait d'étudier les perturbations liées à chacune de ces conditions, et ce afin d'établir les voies biologiques potentiellement altérées.

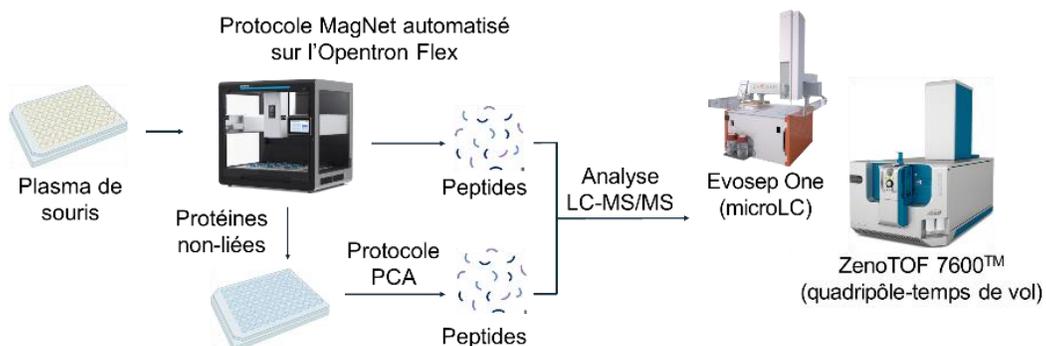


Figure 5.1: Flux de travail correspondant à l'automatisation de la méthode MagNet

En complément à cette étude, il serait également intéressant d'appliquer la méthode ciblée LC-MRM développée au chapitre 3 afin d'analyser les protéines S100 et ANX dans ces différents échantillons de plasma. En effet, comme l'en démontre les concentrations répertoriées au tableau 5.1, plusieurs de ces protéines sont présentes en très faible quantité dans le plasma et pourraient potentiellement ne pas être identifiées avec les méthodes non-ciblées présentées ci-haut. Par ailleurs, nous avons établi dans un premier temps la détectabilité de ces protéines en appliquant la méthode LC-MRM aux différents échantillons de plasma préparés au chapitre 4 (correspondant aux méthodes MagNet, PCA, PPT, PPT-IgG et Protrap). Comme le démontre la carte thermique présentée à la figure 5.2, plusieurs de ces protéines peuvent être bien identifiées dans le plasma grâce à nos méthodes PPT et MagNet, qui offrent elles une complémentarité dans l'identification de ces protéines.

Tableau 5.1: Concentrations plasmatiques des protéines S100 et ANX

Protéine	Concentration (ng/mL)	Protéine	Concentration (ng/mL)
S100A1	0.2	S100A16	0.2
S100A2	0.1	S100B	0.2
S100A3	0.1	S100P	1.8
S100A4	9.3	ANXA1	1200.0
S100A5	pas détectée	ANXA2	23.0
S100A6	4.8	ANXA4	1.3
S100A7	17.0	ANXA5	5.9
S100A8	270.0	ANXA7	0.7
S100A9	950.0	ANXA8	0.1
S100A10	0.2	ANXA8L1	-
S100A11	5.0	ANXA9	pas détectée
S100A12	5.8	ANXA10	0.0
S100A13	0.0	ANXA11	0.2
S100A14	0.3		

Concentrations provenant de PeptideAtlas (*Human plasma* 2023-04)

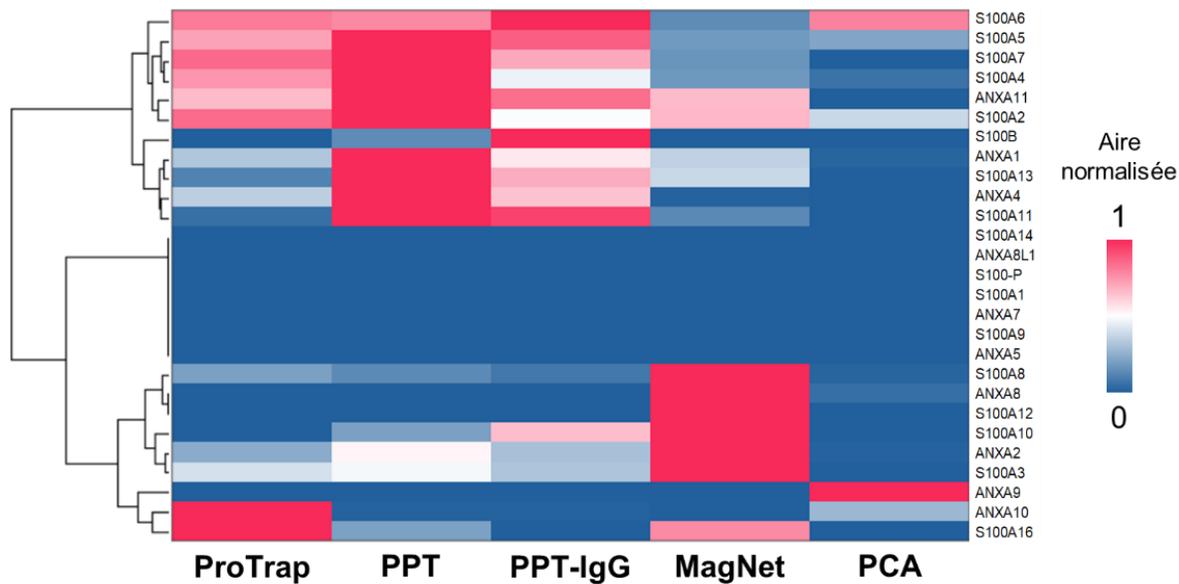


Figure 5.2: Carte thermique montrant la détectabilité des protéines S100 et ANX dans le plasma selon les différentes méthodes d'extractions testées

De plus, l'étude de ces protéines pourrait également s'avérer être très pertinente. Tout d'abord, il faut savoir qu'il existe un lien étroit entre le métabolisme et la réponse immunitaire (Hotamisligil, 2006), et qu'une multitude de conditions médicales associées à un dysfonctionnement métabolique (obésité, diabète de type 2, résistance à l'insuline) peuvent souvent être associée à une réponse pro-inflammatoire chronique (Hotamisligil, 2006). Les protéines S100 sont connues pour être d'importants médiateurs de la réponse immunitaire et inflammatoire. Par exemple, les membres A8 et A9 peuvent favoriser l'expression de plusieurs cytokines pro-inflammatoires par l'entremise de la voie NF- κ B (Shabani *et al.*, 2018). Les isoformes A12 et A13 sont également impliqués dans la production de cytokines et d'agents antimicrobiens (Singh & Ali, 2020). Qui plus est, certaines de ces protéines ont également été associées au développement de ces différentes conditions. Les membres S100B et S100A16 par exemple contribueraient au développement de l'obésité en favorisant respectivement la résistance à l'insuline et en promouvant l'adipogenèse (Gonzalez, Garrie & Turner, 2020). Par conséquent, il serait intéressant de comparer l'expression de ces différentes protéines entre les souris de types sauvages (WT) et celles mutées (dysfonctionnement métabolique) et d'établir l'influence de leur diète sur l'expression de ces protéines.

Les protéines S100 pourraient également être impliquées dans la réponse inflammatoire observée chez les patients atteints de la maladie de Hirschsprung. La maladie de Hirschsprung est une maladie congénitale rare caractérisée par un développement anormal du système nerveux entérique (Granström, Ludvigsson & Wester, 2021). Ce dernier, permettant de réguler la motilité intestinale, celui-ci reste contracté ce qui résulte en une obstruction intestinale. L'inflammation de l'intestin et une forte prévalence à développer une septicémie sont également quelques-uns des symptômes liés à la maladie (Soret *et al.*, 2020). Il existerait un lien étroit entre le système neuro-entérique et le système immunitaire, et une étude récente a permis de mettre en évidence le potentiel rôle de la voie NF- κ B dans le développement de la maladie (Elkrewi *et al.*, 2024). Comme mentionné précédemment, les protéines S100 sont des régulateurs de la voie NF- κ B, il serait alors intéressant d'étudier le niveau d'expression de ces protéines afin d'établir si celles-ci sont oui ou non impliquées dans la modulation de la réponse pro-inflammatoire.

Par conséquent, les méthodes présentées dans ce mémoire sont très polyvalentes et pourraient être appliquées dans différents contextes de recherches. Par exemple, celles-ci pourraient s'avérer être très utiles afin d'étudier différentes maladies, d'identifier des voies biologiques potentiellement altérées ou d'établir les effets d'un traitement sur une maladie. Ces méthodes peuvent également être appliquées à différentes matrices biologiques dont le plasma, les cellules et les tissus.

ANNEXE A

SPECTRES MS ET MS/MS DES EXPÉRIENCES BIOID

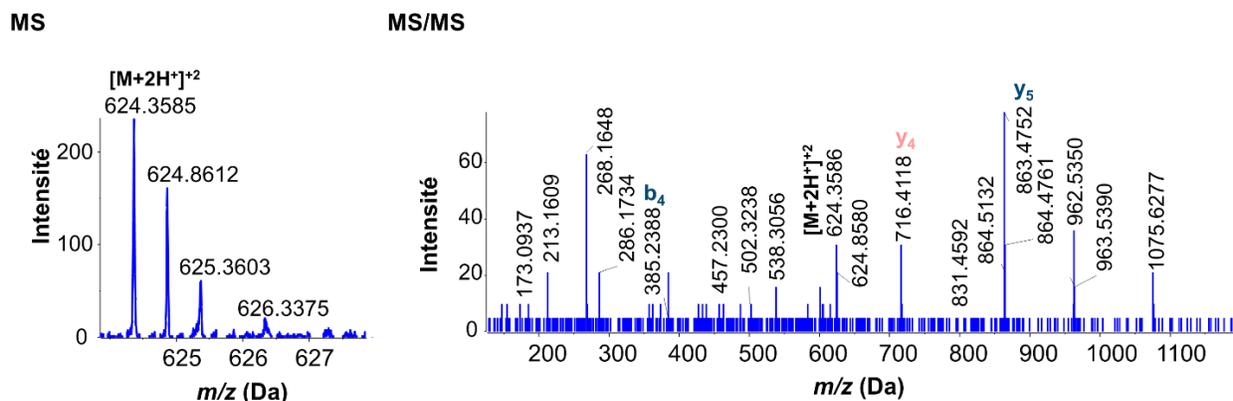


Figure A.1: Spectre MS et MS/MS du peptide ILGTFKSKK²⁺ (FA83H) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS+Biotine pour la méthode d'enrichissement des peptides. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés

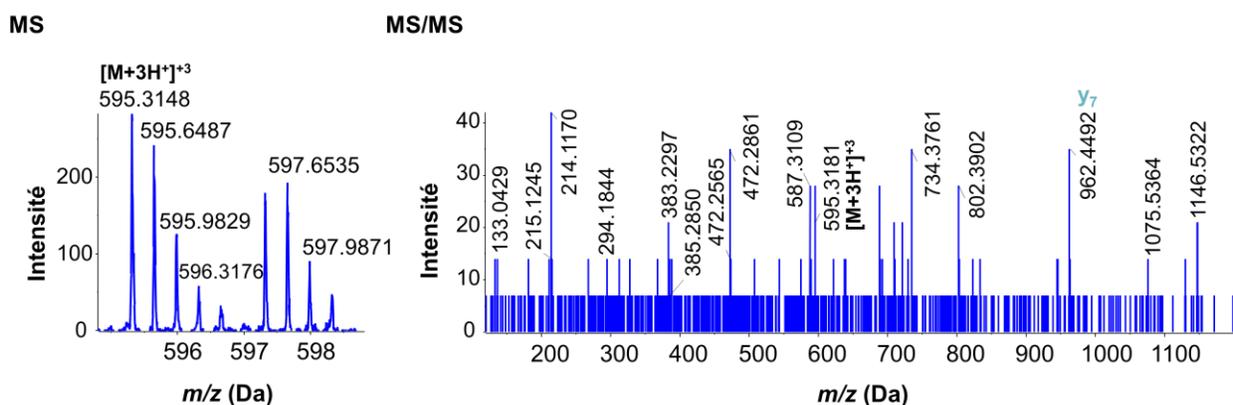


Figure A.2: Spectre MS et MS/MS du peptide TILGAGPVLEAFGNAK³⁺ (MYO9B) identifié dans la fraction peptide biotinylé (PB) de l'échantillon -FBS -Biotine pour la méthode d'enrichissement des peptides. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés

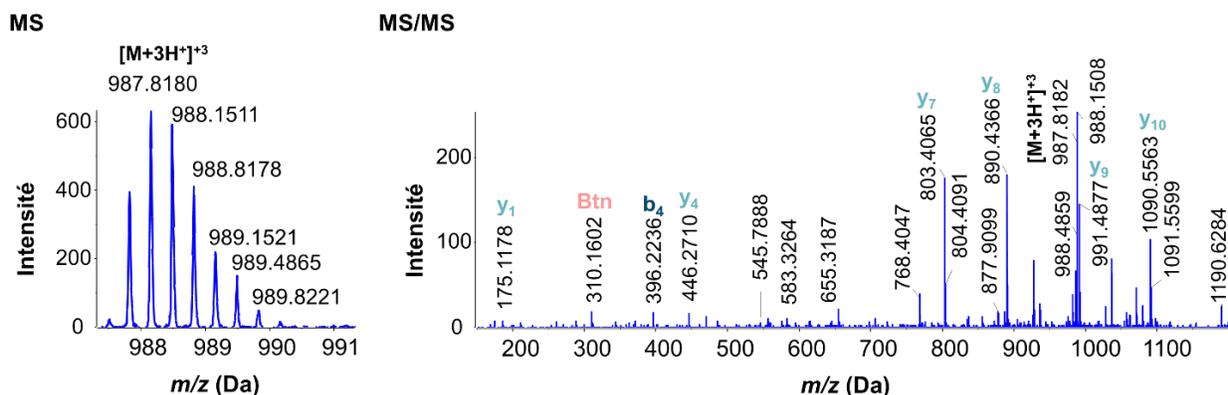


Figure A.3: Spectre MS et MS/MS du peptide GQPLCVLSAMKMETVVTSPMEGTIR³⁺ (PYC) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés.

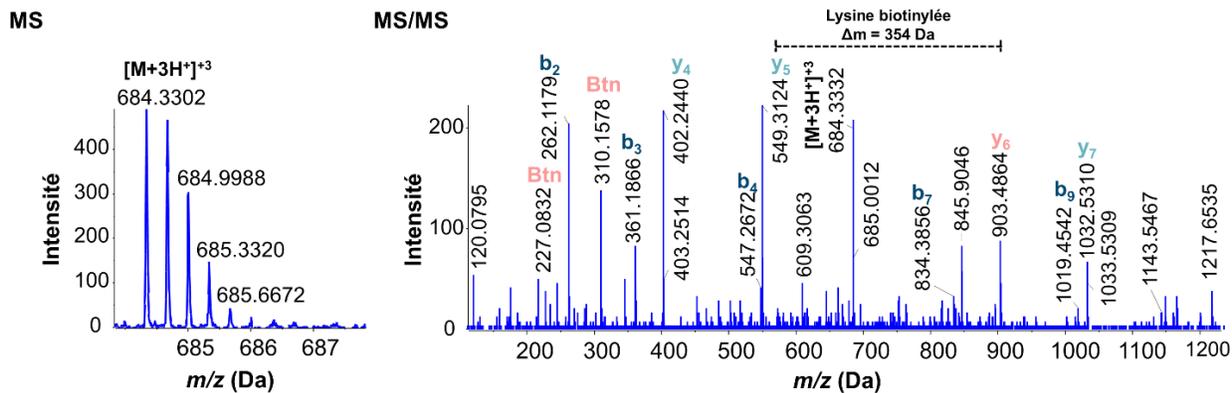


Figure A.4: Spectre MS et MS/MS du peptide FNVWDTAGQEKFGGLR³⁺ (RAN) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés.

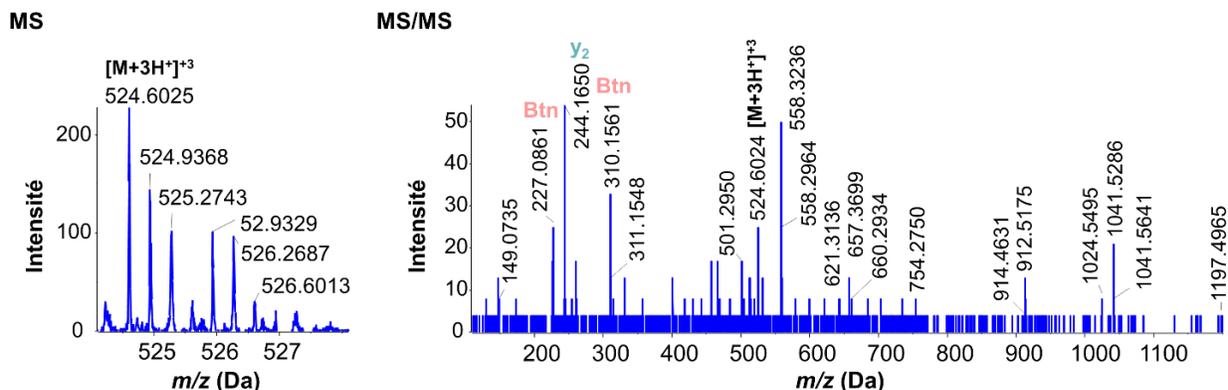


Figure A.5: Spectre MS et MS/MS du peptide MQASIEKGGSLPK³⁺ (NPM) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés.

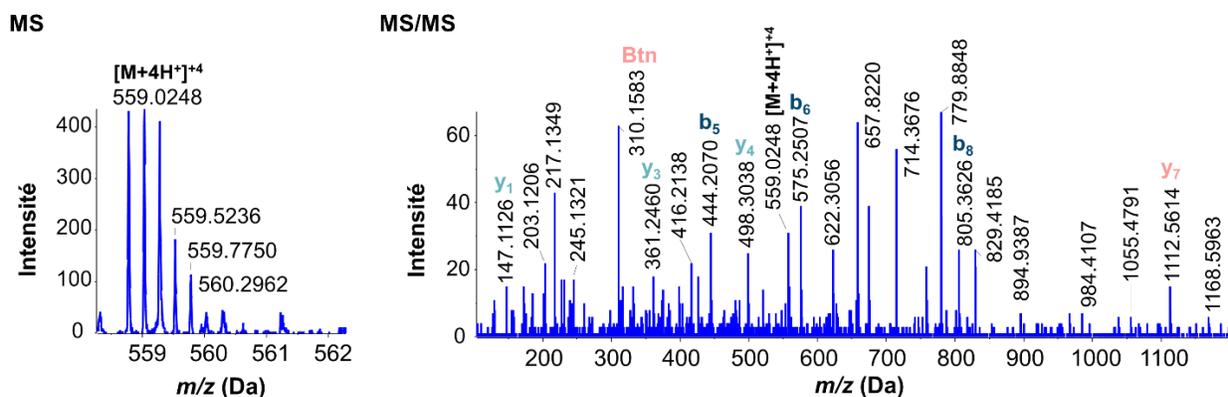


Figure A.6: Spectre MS et MS/MS du peptide AGDSLMIAMIAMKMEHTIK⁴⁺ (MCCA) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés

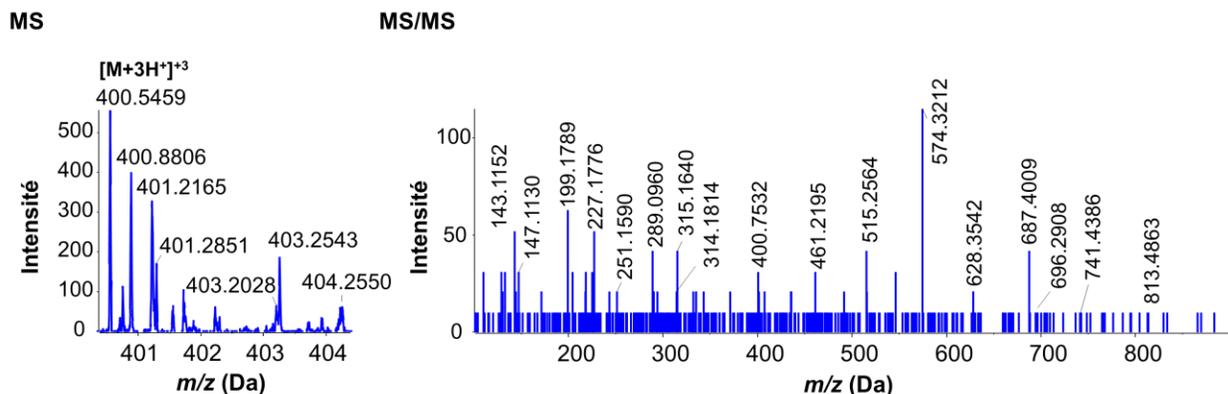


Figure A.7: Spectre MS et MS/MS du peptide SKAESALLR³⁺ (C2D2A) identifié dans la fraction peptide biotylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines.

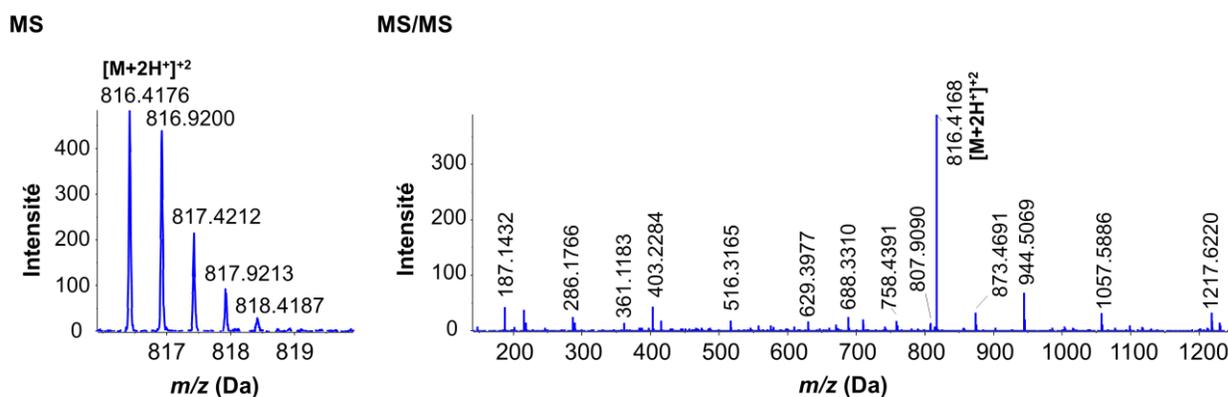


Figure A.8: Spectre MS et MS/MS du peptide KQSASWALKCVK²⁺ (TTC14) identifié dans la fraction peptide biotylé (PB) de l'échantillon +FBS +Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm \leq 10 ont été considérés

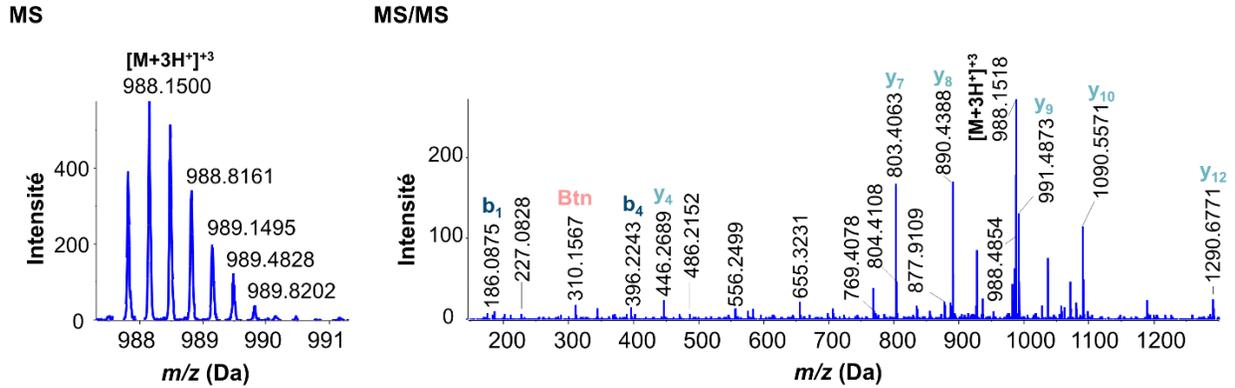


Figure A.9 : Spectre MS et MS/MS du peptide GQPLCVLSAMKMETVVTSPMEGTIR³⁺ (PYC) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS -Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés

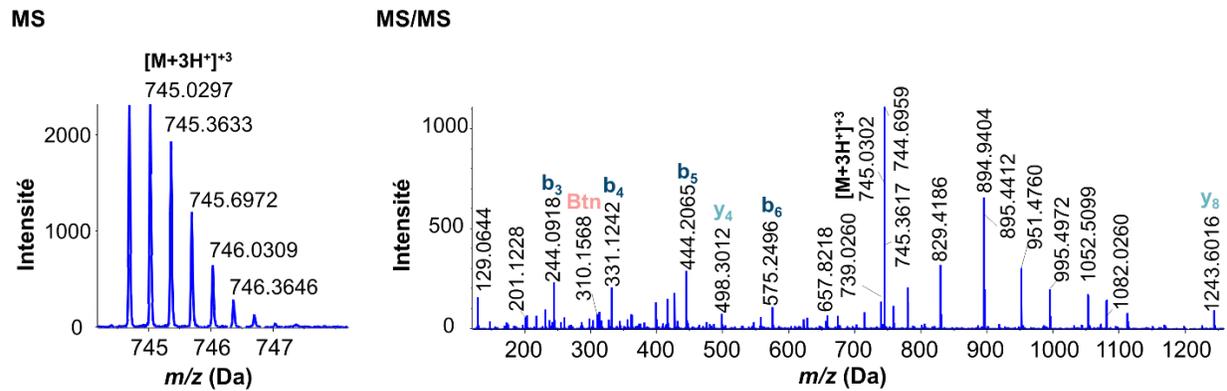
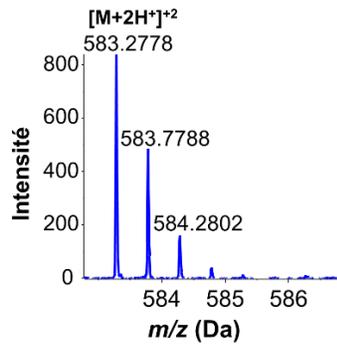


Figure A.10: Spectre MS et MS/MS du peptide AGDSLMIAMKMEHTIK³⁺ (MCCA) identifié dans la fraction peptide biotinylé (PB) de l'échantillon +FBS -Biotine pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés

MS



MS/MS

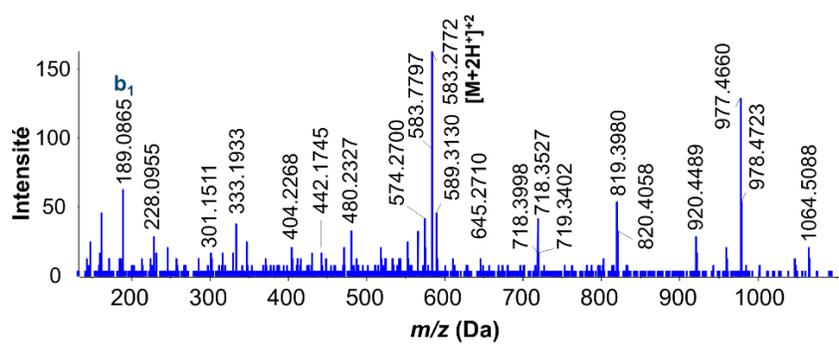


Figure A.11: Spectre MS et MS/MS du peptide STSATCGKK²⁺ (BCORL) identifié dans la fraction peptide biotinylé (PB) de l'échantillon -FBS -Biotine et pour la méthode d'enrichissement des protéines. Seuls les fragments identifiés avec un ppm ≤ 10 ont été considérés.

BIBLIOGRAPHIE

- Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347-355. <https://doi.org/10.1038/nature19949>
- Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., ... & Zhang, B. (2018). How many human proteoforms are there?. *Nature chemical biology*, 14(3), 206-214. <https://doi.org/10.1038/nchembio.2576>
- Aggarwal, S., & Yadav, A. K. (2016). False discovery rate estimation in proteomics. *Statistical Analysis in Proteomics*, 119-128. https://doi.org/10.1007/978-1-4939-3106-4_7
- Allen, D. R., & McWhinney, B. C. (2019). Quadrupole time-of-flight mass spectrometry: a paradigm shift in toxicology screening applications. *The Clinical Biochemist Reviews*, 40(3), 135. DOI: [10.33176/AACB-19-00023](https://doi.org/10.33176/AACB-19-00023)
- Allwood, J. W., & Goodacre, R. (2010). An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques*, 21(1), 33-47. <https://doi.org/10.1002/pca.1187>
- Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., & Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, 41(10), 3912-3928. <https://doi.org/10.1039/C2CS15331>
- Arcuri, C., Giambanco, I., Bianchi, R., & Donato, R. (2002). Annexin V, annexin VI, S100A1 and S100B in developing and adult avian skeletal muscles. *Neuroscience*, 109(2), 371-388. [https://doi.org/10.1016/S0306-4522\(01\)00330-X](https://doi.org/10.1016/S0306-4522(01)00330-X)
- Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2016). Proteomics: technologies and their applications. *Journal of chromatographic science*, 1-15. <https://doi.org/10.1093/chromsci/bmw167>
- Banerjee, S., & Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry*, 2012(1), 282574. <https://doi.org/10.1155/2012/282574>
- Bardhan, A., Bruckner-Tuderman, L., Chapple, I. L., Fine, J. D., Harper, N., Has, C., ... & Heagerty, A. H. (2020). Epidermolysis bullosa. *Nature Reviews Disease Primers*, 6(1), 78. <https://doi.org/10.1038/s41572-020-0210-0>
- Bharadwaj, A., Kempster, E., & Waisman, D. M. (2021). The annexin A2/S100A10 complex: the mutualistic symbiosis of two distinct proteins. *Biomolecules*, 11(12), 1849. <https://doi.org/10.3390/biom11121849>
- Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y. C., Meng, C., ... & Kuster, B. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. *Nature communications*, 11(1), 157. <https://doi.org/10.1038/s41467-019-13973-x>

- BioRender (2020). BioID Assay. <https://app.biorender.com/biorender-templates/figures/all/t-5f4d70fcd3e13800adf070a0-bioid-assay>
- Blume, J. E., Manning, W. C., Troiano, G., Hornburg, D., Figa, M., Hesterberg, L., ... & Farokhzad, O. C. (2020). Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nature communications*, 11(1), 3662. <https://doi.org/10.1038/s41467-020-17033-7>
- Bode, G., Lüken, A., Kerkhoff, C., Roth, J., Ludwig, S., & Nacken, W. (2008). Interaction between S100A8/A9 and annexin A6 is involved in the calcium-induced cell surface exposition of S100A8/A9. *Journal of Biological Chemistry*, 283(46), 31776-31784. DOI: [10.1074/jbc.M803908200](https://doi.org/10.1074/jbc.M803908200)
- Bondt, A., Hoek, M., Tamara, S., de Graaf, B., Peng, W., Schulte, D., ... & Heck, A. J. (2021). Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell systems*, 12(12), 1131-1143. DOI: [10.1016/j.cels.2021.08.008](https://doi.org/10.1016/j.cels.2021.08.008)
- Borràs, E., & Sabidó, E. (2017). What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics*, 17(17-18), 1700180. <https://doi.org/10.1002/pmic.201700180>
- Brandi, J., Noberini, R., Bonaldi, T., & Cecconi, D. (2022). Advances in enrichment methods for mass spectrometry-based proteomics analysis of post-translational modifications. *Journal of Chromatography A*, 1678, 463352. <https://doi.org/10.1016/j.chroma.2022.463352>
- Burkitt, W. I., Derrick, P. J., Lafitte, D., & Bronstein, I. (2003). Protein–ligand and protein–protein interactions studied by electrospray ionization and mass spectrometry. *Biochemical Society transactions*, 31(5), 985-989. <https://doi.org/10.1042/bst0310985>
- Cabana, V. C., Bouchard, A. Y., Sénécal, A. M., Ghilarducci, K., Kourrich, S., Cappadocia, L., & Lussier, M. P. (2021). RNF13 dileucine motif variants L311S and L312P interfere with endosomal localization and AP-3 complex association. *Cells*, 10(11), 3063. <https://doi.org/10.3390/cells10113063>
- Cabana, V. C., Sénécal, A. M., Bouchard, A. Y., Kourrich, S., Cappadocia, L., & Lussier, M. P. (2024). AP-1 contributes to endosomal targeting of the ubiquitin ligase RNF13 via a secondary and novel non-canonical binding motif. *Journal of Cell Science*, 137(18), jcs262035. <https://doi.org/10.1242/jcs.262035>
- Chandramouli, K., & Qian, P. Y. (2009). Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics: HGP, 2009*, 239204 <https://doi.org/10.4061/2009/239204>
- Chang, N., Sutherland, C., Hesse, E., Winkfein, R., Wiehler, W. B., Pho, M., ... & Walsh, M. P. (2007). Identification of a novel interaction between the Ca²⁺-binding protein S100A11 and the Ca²⁺-and phospholipid-binding protein annexin A6. *American Journal of Physiology-Cell Physiology*, 292(4), C1417-C1430. <https://doi.org/10.1152/ajpcell.00439.2006>

- Cheah, J. S., & Yamada, S. (2017). A simple elution strategy for biotinylated proteins bound to streptavidin conjugated beads using excess biotin and heat. *Biochemical and biophysical research communications*, 493(4), 1522-1527. <https://doi.org/10.1016/j.bbrc.2017.09.168>
- Chen, L., Li, J., Zhang, J., Dai, C., Liu, X., Wang, J., ... & Qin, Z. (2015). S100A4 promotes liver fibrosis via activation of hepatic stellate cells. *Journal of Hepatology*, 62(1), 156-164. <https://doi.org/10.1016/j.jhep.2014.07.035>
- Cho, K. F., Branon, T. C., Udeshi, N. D., Myers, S. A., Carr, S. A., & Ting, A. Y. (2020). Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nature Protocols*, 15(12), 3971-3999. <https://doi.org/10.1038/s41596-020-0399-0>
- Cho, W. C. (2007). Proteomics technologies and challenges. *Genomics, Proteomics and Bioinformatics*, 5(2), 77-85. [https://doi.org/10.1016/S1672-0229\(07\)60018-7](https://doi.org/10.1016/S1672-0229(07)60018-7)
- Cirri, E., Knautt, H., Di Fraia, D., Pömpner, N., Rahnis, N., Heinze, I., ... & Dau, T. (2024). Optimized Automated Workflow for BioID Improves Reproducibility and Identification of Protein–Protein Interactions. *Journal of Proteome Research*, 23(10), 4359-4368. <https://doi.org/10.1021/acs.jproteome.4c00308>
- Conti, A., & Alessio, M. (2015). Comparative proteomics for the evaluation of protein expression and modifications in neurodegenerative diseases. *International Review of Neurobiology*, 121, 117-152. <https://doi.org/10.1016/bs.irn.2015.05.004>
- Coufalíková, K., Benešová, I., Vaculovič, T., Kanický, V., & Preisler, J. (2017). LC coupled to ESI, MALDI and ICP MS—A multiple hyphenation for metalloproteomic studies. *Analytica Chimica Acta*, 968, 58-65. <https://doi.org/10.1016/j.aca.2017.03.016>
- Daily, A. E., & Kumar, T. K. S. (2010). Defining the Interaction Between S100A13 and Annexin II Peptide: Insight into Non-Classical Secretion. *Biophysical Journal*, 98(3), 57a. DOI: [10.1016/j.bpj.2009.12.327](https://doi.org/10.1016/j.bpj.2009.12.327)
- Delangre, E., Oppliger, E., Berkcan, S., Gjorgjieva, M., Correia de Sousa, M., & Foti, M. (2022). S100 proteins in fatty liver disease and hepatocellular carcinoma. *International journal of molecular sciences*, 23(19), 11030. <https://doi.org/10.3390/ijms231911030>
- Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods*, 17(1), 41-44. DOI: [10.1038/s41592-019-0638-x](https://doi.org/10.1038/s41592-019-0638-x)
- Deracinois, B., Flahaut, C., Duban-Deweer, S., & Karamanos, Y. (2013). Comparative and quantitative global proteomics approaches: an overview. *Proteomes*, 1(3), 180-218. <https://doi.org/10.3390/proteomes1030180>
- Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., ... & Aebersold, R. (2006). The peptidatlas project. *Nucleic acids research*, 34(suppl_1), D655-D658. <https://doi.org/10.1093/nar/gkj040>
- Deutsch, E. W., Omenn, G. S., Sun, Z., Maes, M., Pernemalm, M., Palaniappan, K. K., ... & Schwenk, J. M. (2021). Advances and utility of the human plasma proteome. *Journal of proteome research*, 20(12), 5241-5263. <https://doi.org/10.1021/acs.jproteome.1c00657>

- Dundas, C. M., Demonte, D., & Park, S. (2013). Streptavidin–biotin technology: improvements and innovations in chemical and biological applications. *Applied microbiology and biotechnology*, 97, 9343-9353. <https://doi.org/10.1007/s00253-013-5232-z>
- Duong, V. A., & Lee, H. (2023). Bottom-up proteomics: advancements in sample preparation. *International journal of molecular sciences*, 24(6), 5350. <https://doi.org/10.3390/ijms24065350>
- Dupree, E. J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B. A., & Darie, C. C. (2020). A critical review of bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes*, 8(3), 14. <https://doi.org/10.3390/proteomes8030014>
- Elkrewi, E. Z., Al Abdulqader, A. A., Khasanov, R., Maas-Omlor, S., Boettcher, M., Wessel, L. M., ... & Tapia-Laliena, M. Á. (2024). Role of Inflammation and the NF-κB Signaling Pathway in Hirschsprung's Disease. *Biomolecules*, 14(8), 992. <https://doi.org/10.3390/biom14080992>
- Eue, I., König, S., Pior, J., & Sorg, C. (2002). S100A8, S100A9 and the S100A8/A9 heterodimer complex specifically bind to human endothelial cells: identification and characterization of ligands for the myeloid-related proteins S100A9 and S100A8/A9 on human dermal microvascular endothelial cell line-1 cells. *International immunology*, 14(3), 287-297. <https://doi.org/10.1093/intimm/14.3.287>
- Evans, G. A. (2000). Designer science and the “omic” revolution. *Nature Biotechnology*, 18(2), 127-127. <https://doi.org/10.1038/72480>
- Filipek, A., & Wojda, U. (1996). p30, a novel protein target of mouse calcyclin (S100A6). *Biochemical Journal*, 320(2), 585-587. <https://doi.org/10.1042/bj3200585>
- Garbuglia, M., Verzini, M., & Donato, R. (1998). Annexin VI binds S100A1 and S100B and blocks the ability of S100A1 and S100B to inhibit desmin and GFAP assemblies into intermediate filaments. *Cell Calcium*, 24(3), 177-191. [https://doi.org/10.1016/S0143-4160\(98\)90127-0](https://doi.org/10.1016/S0143-4160(98)90127-0)
- Garbuglia, M., Verzini, M., Hofmann, A., Huber, R., & Donato, R. (2000). S100A1 and S100B interactions with annexins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1498(2-3), 192-206. [https://doi.org/10.1016/S0167-4889\(00\)00096-3](https://doi.org/10.1016/S0167-4889(00)00096-3)
- Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Molecular systems biology*, 13(9), 942. <https://doi.org/10.15252/msb.20156297>
- Gillet, L. C., Leitner, A., & Aebersold, R. (2016). Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. *Annual review of analytical chemistry*, 9(1), 449-472. <https://doi.org/10.1146/annurev-anchem-071015-041535>
- Gonzalez, L. L., Garrie, K., & Turner, M. D. (2020). Role of S100 proteins in health and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1867(6), 118677. <https://doi.org/10.1016/j.bbamcr.2020.118677>

- Granström, A. L., Ludvigsson, J. F., & Wester, T. (2021). Clinical characteristics and validation of diagnosis in individuals with Hirschsprung disease and inflammatory bowel disease. *Journal of Pediatric Surgery*, 56(10), 1799-1802. <https://doi.org/10.1016/j.jpedsurg.2020.11.015>
- Graves, P. R., & Haystead, T. A. (2002). Molecular biologist's guide to proteomics. *Microbiology and molecular biology reviews*, 66(1), 39-63. <https://doi.org/10.1128/membr.66.1.39-63.2002>
- Gregorich, Z. R., Chang, Y. H., & Ge, Y. (2014). Proteomics in heart failure: top-down or bottom-up?. *Pflügers Archiv-European Journal of Physiology*, 466, 1199-1209. <https://doi.org/10.1007/s00424-014-1471-9>
- Haag, A. M. (2016). Mass analyzers and mass spectrometers. *Modern proteomics—Sample preparation, analysis and practical applications*, 157-169. https://doi.org/10.1007/978-3-319-41448-5_7
- Ho, C. S., Lam, C. W. K., Chan, M. H., Cheung, R. C. K., Law, L. K., Lit, L. C., ... & Tai, H. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist Reviews*, 24(1), 3. PMID: [18568044](https://pubmed.ncbi.nlm.nih.gov/18568044/)
- Hotamisligil, G. S. (2006). Inflammation and metabolic disorders. *Nature*, 444(7121), 860-867. <https://doi.org/10.1038/nature05485>
- Hristova, J., & Svinarov, D. (2022). Enhancing precision medicine through clinical mass spectrometry platform. *Biotechnology & Biotechnological Equipment*, 36(1), 107-117. <https://doi.org/10.1080/13102818.2022.2053342>
- Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., & Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature protocols*, 14(1), 68-85. <https://doi.org/10.1038/s41596-018-0082-x>
- Ignjatovic, V., Geyer, P. E., Palaniappan, K. K., Chaaban, J. E., Omenn, G. S., Baker, M. S., ... & Schwenk, J. M. (2019). Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. *Journal of proteome research*, 18(12), 4085-4097. <https://doi.org/10.1021/acs.jproteome.9b00503>
- Jensen, O. N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current opinion in chemical biology*, 8(1), 33-41. <https://doi.org/10.1016/j.cbpa.2003.12.009>
- Jiang, Y., Rex, D. A. B., Schuster, D., Neely, B. A., Rosano, G. L., Volkmar, N., ... & Meyer, J. G. (2024). Comprehensive overview of bottom-up proteomics using mass spectrometry. *ACS Measurement Science Au*, 4(4), 338-417. <https://doi.org/10.1021/acsmeasuresciau.3c00068>
- Jürgen H Gross. (2017). *Mass Spectrometry: A Textbook* (3e éd.). Springer Cham. <https://doi.org/10.1007/978-3-319-54398-7>
- Karpievitch, Y. V., Polpitiya, A. D., Anderson, G. A., Smith, R. D., & Dabney, A. R. (2010). Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. *The annals of applied statistics*, 4(4), 1797. DOI: [10.1214/10-AOAS341](https://doi.org/10.1214/10-AOAS341)

- Kaur, G., Poljak, A., Ali, S. A., Zhong, L., Raftery, M. J., & Sachdev, P. (2021). Extending the depth of human plasma proteome coverage using simple fractionation techniques. *Journal of proteome research*, 20(2), 1261-1279. <https://doi.org/10.1021/acs.jproteome.0c00670>
- Keshishian, H., Burgess, M. W., Specht, H., Wallace, L., Clauser, K. R., Gillette, M. A., & Carr, S. A. (2017). Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nature protocols*, 12(8), 1683-1701. <https://doi.org/10.1038/nprot.2017.054>
- Khani, M., Nafissi, S., Shamshiri, H., Moazzeni, H., Taheri, H., & Elahi, E. (2023). Identification of RNF13 as cause of recessively inherited ALS in a multi-case pedigree. *Journal of Medical Genetics*, 60(3), 301-309. <https://doi.org/10.1136/img-2022-108645>
- Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., & Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Molecular biology of the cell*, 27(8), 1188-1196. <https://doi.org/10.1091/mbc.E15-12-0844>
- Krasny, L., & Huang, P. H. (2021). Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Molecular omics*, 17(1), 29-42. DOI: [10.1039/D0MO00072H](https://doi.org/10.1039/D0MO00072H)
- Kubitz, L., Bitsch, S., Zhao, X., Schmitt, K., Deweid, L., Roehrig, A., ... & Béthune, J. (2022). Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells. *Communications biology*, 5(1), 657. <https://doi.org/10.1038/s42003-022-03604-5>
- Kulyyassov, A., Fresnais, M., & Longuespée, R. (2021). Targeted liquid chromatography-tandem mass spectrometry analysis of proteins: Basic principles, applications, and perspectives. *Proteomics*, 21(23-24), 2100153. <https://doi.org/10.1002/pmic.202100153>
- Kusebauch, U., Campbell, D. S., Deutsch, E. W., Chu, C. S., Spicer, D. A., Brusniak, M. Y., ... & Moritz, R. L. (2016). Human SRMATlas: a resource of targeted assays to quantify the complete human proteome. *Cell*, 166(3), 766-778. DOI: [10.1016/j.cell.2016.06.041](https://doi.org/10.1016/j.cell.2016.06.041)
- Lee, C. S., Taib, N. A. M., Ashrafzadeh, A., Fadzli, F., Harun, F., Rahmat, K., ... & Hashim, O. H. (2016). Unmasking heavily O-glycosylated serum proteins using perchloric acid: identification of serum proteoglycan 4 and protease C1 inhibitor as molecular indicators for screening of breast cancer. *PLoS One*, 11(2), e0149551. <https://doi.org/10.1371/journal.pone.0149551>
- Lee, P. Y., Osman, J., Low, T. Y., & Jamal, R. (2019). Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery. *Bioanalysis*, 11(19), 1799-1812. <https://doi.org/10.4155/bio-2019-0145>
- Lépine, M., Robert, M. C., & Sleno, L. (2024). Discovery and Verification of Sjögren's Syndrome Protein Biomarkers in Tears by Targeted LC-MRM. *Journal of Proteome Research*, 23(6), 2219-2229. <https://doi.org/10.1021/acs.jproteome.4c00163>

- Lépine, M., Robert, M. C., & Sleno, L. (2025). Tear Protein Biomarkers for Ocular Mucous Membrane Pemphigoid Uncovered Using Targeted LC-MS/MS. *Journal of Proteome Research*, 24(3), 1275-1284 <https://doi.org/10.1021/acs.jproteome.4c00924>
- Lépine, M., Zambito, O., & Sleno, L. (2023). Targeted Workflow Investigating Variations in the Tear Proteome by Liquid Chromatography Tandem Mass Spectrometry. *ACS omega*, 8(34), 31168-31177. <https://doi.org/10.1021/acsomega.3c03186>
- Li, K. W., Gonzalez-Lozano, M. A., Koopmans, F., & Smit, A. B. (2020). Recent developments in data independent acquisition (DIA) mass spectrometry: application of quantitative analysis of the brain proteome. *Frontiers in molecular neuroscience*, 13, 564446. <https://doi.org/10.3389/fnmol.2020.564446>
- Liu, F., Zhang, J. Z., & Mei, Y. (2016). The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific reports*, 6(1), 27190. <https://doi.org/10.1038/srep27190>
- Liu, X., Wang, Y., Ming, Y., Song, Y., Zhang, J., Chen, X., ... & Mao, Y. (2015). S100A9: a potential biomarker for the progression of non-alcoholic fatty liver disease and the diagnosis of non-alcoholic steatohepatitis. *PLoS one*, 10(5), e0127352. <https://doi.org/10.1371/journal.pone.0127352>
- Lizarbe, M. A., Barrasa, J. I., Olmo, N., Gavilanes, F., & Turnay, J. (2013). Annexin-phospholipid interactions. Functional implications. *International journal of molecular sciences*, 14(2), 2652-2683. <https://doi.org/10.3390/ijms14022652>
- Long, Z., Zhao, Z., Fan, X., & Luo, X. (2025). Comparison of analytical-flow, micro-flow and nano-flow LC-MS/MS for sub-proteome analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 252, 116484. <https://doi.org/10.1016/j.jpba.2024.116484>
- Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., & Shi, J. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal transduction and targeted therapy*, 5(1), 213. <https://doi.org/10.1038/s41392-020-00315-3>
- Ma, Z., Islam, M. N., Xu, T., & Song, E. (2021). AP-3 adaptor complex-mediated vesicle trafficking. *Biophysics Reports*, 7(2), 91. DOI: [10.52601/bpr.2021.200051](https://doi.org/10.52601/bpr.2021.200051)
- Macklin, A., Khan, S., & Kislinger, T. (2020). Recent advances in mass spectrometry based clinical proteomics: applications to cancer research. *Clinical proteomics*, 17(1), 17. <https://doi.org/10.1186/s12014-020-09283-w>
- Mathew J, Sankar P, Varacallo M. Physiology, Blood Plasma. In: StatPearls. StatPearls Publishing, Treasure Island (FL); 2023. PMID: [30285399](https://pubmed.ncbi.nlm.nih.gov/30285399/).
- May, D. G., Scott, K. L., Campos, A. R., & Roux, K. J. (2020). Comparative application of BioID and TurboID for protein-proximity biotinylation. *Cells*, 9(5), 1070. <https://doi.org/10.3390/cells9051070>
- McLafferty, F. W. (2011). A century of progress in molecular mass spectrometry. *Annual review of analytical chemistry*, 4(1), 1-22. <https://doi.org/10.1146/annurev-anchem-061010-114018>

- Meng, Z., & Veenstra, T. D. (2011). Targeted mass spectrometry approaches for protein biomarker verification. *Journal of proteomics*, 74(12), 2650-2659. <https://doi.org/10.1016/j.jprot.2011.04.011>
- Metatla, I., Roger, K., Chhuon, C., Ceccacci, S., Chapelle, M., Schmit, P. O., ... & Guerrero, I. C. (2024). Neat plasma proteomics: getting the best out of the worst. *Clinical Proteomics*, 21(1), 22. <https://doi.org/10.1186/s12014-024-09477-6>
- Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*, 8(8), 1551-1566. <https://doi.org/10.1038/nprot.2013.092>
- Motani, K., & Kosako, H. (2020). BioID screening of biotinylation sites using the avidin-like protein Tamavidin 2-REV identifies global interactors of stimulator of interferon genes (STING). *Journal of Biological Chemistry*, 295(32), 11174-11183. DOI: [10.1074/jbc.RA120.014323](https://doi.org/10.1074/jbc.RA120.014323)
- Mukai, K., Miyagi, T., Nishio, K., Yokoyama, Y., Yoshioka, T., Saito, Y., ... & Takehara, T. (2016). S100A8 production in CXCR2-expressing CD11b+ Gr-1high cells aggravates hepatitis in mice fed a high-fat and high-cholesterol diet. *The Journal of Immunology*, 196(1), 395-406. <https://doi.org/10.4049/jimmunol.1402709>
- Pino, L. K., Searle, B. C., Yang, H. Y., Hoofnagle, A. N., Noble, W. S., & MacCoss, M. J. (2020). Matrix-matched calibration curves for assessing analytical figures of merit in quantitative proteomics. *Journal of proteome research*, 19(3), 1147-1153. <https://doi.org/10.1021/acs.jproteome.9b00666>
- Poeter, M., Radke, S., Koese, M., Hessner, F., Hegemann, A., Musiol, A., ... & Rescher, U. (2013). Disruption of the annexin A1/S100A11 complex increases the migration and clonogenic growth by dysregulating epithelial growth factor (EGF) signaling. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(7), 1700-1711. <https://doi.org/10.1016/j.bbamcr.2012.12.006>
- Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. S. (2014). Protein-protein interaction detection: methods and analysis. *International journal of proteomics*, 2014(1), 147648. <https://doi.org/10.1155/2014/147648>
- Renuse, S., Madugundu, A. K., Jung, J. H., Byeon, S. K., Goldschmidt, H. L., Tahir, R., ... & Pandey, A. (2020). Signature fragment ions of biotinylated peptides. *Journal of the American Society for Mass Spectrometry*, 31(2), 394-404. <https://doi.org/10.1021/jasms.9b00024>
- Réty, S., Osterloh, D., Arié, J. P., Tabaries, S., Seeman, J., Russo-Marie, F., ... & Lewit-Bentley, A. (2000). Structural basis of the Ca²⁺-dependent association between S100C (S100A11) and its target, the N-terminal part of annexin I. *Structure*, 8(2), 175-184. DOI: [10.1016/S0969-2126\(00\)00093-9](https://doi.org/10.1016/S0969-2126(00)00093-9)
- Rezvanpour, A., Santamaria-Kisiel, L., & Shaw, G. S. (2011). The S100A10-annexin A2 complex provides a novel asymmetric platform for membrane repair. *Journal of Biological Chemistry*, 286(46), 40174-40183. DOI: [10.1074/jbc.M111.244038](https://doi.org/10.1074/jbc.M111.244038)

- Rice, S. J., & Belani, C. P. (2022). Optimizing data-independent acquisition (DIA) spectral library workflows for plasma proteomics studies. *Proteomics*, 22(17), 2200125. <https://doi.org/10.1002/pmic.202200125>
- Rice, S. J., & Belani, C. P. (2024). Characterization of effective, simple, and low-cost precipitation methods for depleting abundant plasma proteins to enhance the depth and breadth of plasma proteomics. *Proteomics*, 24(15), 2400071. <https://doi.org/10.1002/pmic.202400071>
- Rintala-Dempsey, A. C., Santamaria-Kisiel, L., Liao, Y., Lajoie, G., & Shaw, G. S. (2006). Insights into S100 target specificity examined by a new interaction between S100A11 and annexin A2. *Biochemistry*, 45(49), 14695-14705. <https://doi.org/10.1021/bi061754e>
- Rogers, J. C., & Bomgardner, R. D. (2016). Sample preparation for mass spectrometry-based proteomics; from proteomes to peptides. *Modern Proteomics—Sample Preparation, Analysis and Practical Applications*, 43-62. https://doi.org/10.1007/978-3-319-41448-5_3
- Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of cell biology*, 196(6), 801-810. <https://doi.org/10.1083/jcb.201112098>
- Rozanova, S., Barkovits, K., Nikolov, M., Schmidt, C., Urlaub, H., & Marcus, K. (2021). Quantitative mass spectrometry-based proteomics: an overview. *Quantitative methods in proteomics*, 85-116. https://doi.org/10.1007/978-1-0716-1024-4_8
- Ryan, M. C., Stucky, M., Wakefield, C., Melott, J. M., Akbani, R., Weinstein, J. N., & Broom, B. M. (2020). Interactive Clustered Heat Map Builder: An easy web-based tool for creating sophisticated clustered heat maps. *F1000Research*, 8, ISCB-Comm. DOI: [10.12688/f1000research.20590.2](https://doi.org/10.12688/f1000research.20590.2)
- Ryu, J., & Thomas, S. N. (2021). Quantitative mass spectrometry-based proteomics for biomarker development in ovarian cancer. *Molecules*, 26(9), 2674. <https://doi.org/10.3390/molecules26092674>
- Sarkar, S., Behera, A. P., Borar, P., Banka, P. A., & Datta, A. B. (2019). Designing active RNF4 monomers by introducing a tryptophan: avidity towards E2~ Ub conjugates dictates the activity of ubiquitin RING E3 ligases. *Biochemical Journal*, 476(10), 1465-1482. <https://doi.org/10.1042/BCJ20180883>
- Schaeffer, M., Gateau, A., Teixeira, D., Michel, P. A., Zahn-Zabal, M., & Lane, L. (2017). The neXtProt peptide uniqueness checker: a tool for the proteomics community. *Bioinformatics*, 33(21), 3471-3472. <https://doi.org/10.1093/bioinformatics/btx318>
- Sears, R. M., May, D. G., & Roux, K. J. (2019). BioID as a tool for protein-proximity labeling in living cells. *Enzyme-Mediated Ligation Methods*, 299-313. https://doi.org/10.1007/978-1-4939-9546-2_15

- Semov, A., Moreno, M. J., Onichtchenko, A., Abulrob, A., Ball, M., Ekiel, I., ... & Alakhov, V. (2005). Metastasis-associated protein S100A4 induces angiogenesis through interaction with Annexin II and accelerated plasmin formation. *Journal of Biological Chemistry*, 280(21), 20833-20841. DOI: [10.1074/jbc.M412653200](https://doi.org/10.1074/jbc.M412653200)
- Shabani, F., Farasat, A., Mahdavi, M., & Gheibi, N. (2018). Calprotectin (S100A8/S100A9): a key protein between inflammation and cancer. *Inflammation Research*, 67, 801-812. <https://doi.org/10.1007/s00011-018-1173-4>
- Shi, Y., Xiang, R., Horváth, C., & Wilkins, J. A. (2004). The role of liquid chromatography in proteomics. *Journal of Chromatography A*, 1053(1-2), 27-36. <https://doi.org/10.1016/j.chroma.2004.07.044>
- Singh, P., & Ali, S. A. (2022). Multifunctional role of S100 protein family in the immune system: an update. *Cells*, 11(15), 2274. <https://doi.org/10.3390/cells11152274>
- Sinha, A., & Mann, M. (2020). A beginner's guide to mass spectrometry-based proteomics. *The Biochemist*, 42(5), 64-69. <https://doi.org/10.1042/BIO20200057>
- Sobolewski, C., Abegg, D., Berthou, F., Dolicka, D., Calo, N., Sempoux, C., ... & Foti, M. (2020). S100A11/ANXA2 belongs to a tumour suppressor/oncogene network deregulated early with steatosis and involved in inflammation and hepatocellular carcinoma development. *Gut*, 69(10), 1841-1854. <https://doi.org/10.1136/gutjnl-2019-319019>
- Sobsey, C. A., Ibrahim, S., Richard, V. R., Gaspar, V., Mitsa, G., Lacasse, V., ... & Borchers, C. H. (2020). Targeted and untargeted proteomics approaches in biomarker development. *Proteomics*, 20(9), 1900029. <https://doi.org/10.1002/pmic.201900029>
- Soret, R., Schneider, S., Bernas, G., Christophers, B., Souchkova, O., Charrier, B., ... & Pilon, N. (2020). Glial cell-derived neurotrophic factor induces enteric neurogenesis and improves colon structure and function in mouse models of Hirschsprung disease. *Gastroenterology*, 159(5), 1824-1838. <https://doi.org/10.1053/j.gastro.2020.07.018>
- Soria, J., Durán, J. A., Etxebarria, J., Merayo, J., González, N., Reigada, R., ... & Suárez, T. (2013). Tear proteome and protein network analyses reveal a novel pentamarker panel for tear film characterization in dry eye and meibomian gland dysfunction. *Journal of proteomics*, 78, 94-112. <https://doi.org/10.1016/j.jprot.2012.11.017>
- Sudo, T., & Hidaka, H. (1998). Regulation of calyculin (S100A6) binding by alternative splicing in the N-terminal regulatory domain of annexin XI isoforms. *Journal of Biological Chemistry*, 273(11), 6351-6357. DOI: [10.1074/jbc.273.11.6351](https://doi.org/10.1074/jbc.273.11.6351)
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., ... & Von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, 51(D1), D638-D646. <https://doi.org/10.1093/nar/gkac1000>
- Takakura, Y., Sofuku, K., & Tsunashima, M. (2013). Tamavidin 2-REV: an engineered tamavidin with reversible biotin-binding capability. *Journal of biotechnology*, 164(1), 19-25. <https://doi.org/10.1016/j.jbiotec.2013.01.006>

- Takakura, Y., Tsunashima, M., Suzuki, J., Usami, S., Kakuta, Y., Okino, N., ... & Yamamoto, T. (2009). Tamavidins—novel avidin-like biotin-binding proteins from the Tamogitake mushroom. *The FEBS journal*, 276(5), 1383-1397. <https://doi.org/10.1111/j.1742-4658.2009.06879.x>
- Teng, F., Jiang, J., Zhang, J., Yuan, Y., Li, K., Zhou, B., ... & Zhang, H. (2021). The S100 calcium-binding protein A11 promotes hepatic steatosis through RAGE-mediated AKT-mTOR signaling. *Metabolism*, 117, 154725. <https://doi.org/10.1016/j.metabol.2021.154725>
- Tong, L., Lan, W., Lim, R. R., & Chaurasia, S. S. (2014). S100A proteins as molecular targets in the ocular surface inflammatory diseases. *The ocular surface*, 12(1), 23-31. <https://doi.org/10.1016/j.jtos.2013.10.001>
- Udeshi, N. D., Pedram, K., Svinkina, T., Fereshetian, S., Myers, S. A., Aygun, O., ... & Carr, S. A. (2017). Antibodies to biotin enable large-scale detection of biotinylation sites on proteins. *Nature methods*, 14(12), 1167-1170. <https://doi.org/10.1038/nmeth.4465>
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419. DOI: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419)
- Varnavides, G., Madern, M., Anrather, D., Hartl, N., Reiter, W., & Hartl, M. (2022). In search of a universal method: a comparative survey of bottom-up proteomics sample preparation methods. *Journal of proteome research*, 21(10), 2397-2411. <https://doi.org/10.1021/acs.jproteome.2c00265>
- Viode, A., van Zalm, P., Smolen, K. K., Fatou, B., Stevenson, D., Jha, M., ... & IMPACC Network. (2023). A simple, time-and cost-effective, high-throughput depletion strategy for deep plasma proteomics. *Science advances*, 9(13), eadf9717. DOI: [10.1126/sciadv.adf9717](https://doi.org/10.1126/sciadv.adf9717)
- Walther, T. C., & Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *Journal of Cell Biology*, 190(4), 491-500. <https://doi.org/10.1083/jcb.201004052>
- Wang, G., Zhang, S., Fernig, D. G., Spiller, D., Martin-Fernandez, M., Zhang, H., ... & Barraclough, R. (2004). Heterodimeric interaction and interfaces of S100A1 and S100P. *Biochemical Journal*, 382(1), 375-383. <https://doi.org/10.1042/BJ20040142>
- Ward, B., Pyrdit Ruys, S., Balligand, J. L., Belkhir, L., Cani, P. D., Collet, J. F., ... & Elens, L. (2024). Deep plasma proteomics with data-independent acquisition: A fastlane towards biomarkers identification. *bioRxiv*, 2024-02. <https://doi.org/10.1101/2024.02.23.581160>
- Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., & Salemme, F. R. (1989). Structural origins of high-affinity biotin binding to streptavidin. *Science*, 243(4887), 85-88. <https://doi.org/10.1126/science.2911722>
- Weisz, J., & Uversky, V. N. (2020). Zooming into the dark side of human annexin-S100 complexes: dynamic alliance of flexible partners. *International journal of molecular sciences*, 21(16), 5879. <https://doi.org/10.3390/ijms21165879>

- Wen, C., Wu, X., Lin, G., Yan, W., Gan, G., Xu, X., ... & Zhong, C. Q. (2023). Evaluation of DDA Library-Free Strategies for Phosphoproteomics and Ubiquitinomics Data-Independent Acquisition Data. *Journal of Proteome Research*, 22(7), 2232-2245. <https://doi.org/10.1021/acs.jproteome.2c00735>
- Wenk, D., Zuo, C., Kislinger, T., & Sepiashvili, L. (2024). Recent developments in mass-spectrometry-based targeted proteomics of clinical cancer biomarkers. *Clinical Proteomics*, 21(1), 6. <https://doi.org/10.1186/s12014-024-09452-1>
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., ... & Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502), 582-587. <https://doi.org/10.1038/nature13319>
- Wittmann-Liebold, B., Graack, H. R., & Pohl, T. (2006). Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. *Proteomics*, 6(17), 4688-4703. <https://doi.org/10.1002/pmic.200500874>
- Wu, C. C., Tsantilas, K. A., Park, J., Plubell, D., Sanders, J. A., Naicker, P., ... & MacCoss, M. J. (2024). Mag-Net: Rapid enrichment of membrane-bound particles enables high coverage quantitative analysis of the plasma proteome. *bioRxiv*, 2023-06. <https://doi.org/10.1101/2023.06.10.544439>
- Wu, J., Duan, C., Yang, Y., Wang, Z., Tan, C., Han, C., & Hou, X. (2023). Insights into the liver-eyes connections, from epidemiological, mechanical studies to clinical translation. *Journal of Translational Medicine*, 21(1), 712. <https://doi.org/10.1186/s12967-023-04543-3>
- Yan, X., Kumar, K., Milette Lamarche, R., Youssef, H., Shaw, G. S., Marcotte, I., ... & Boisselier, E. (2021). Interactions between the cell membrane repair protein S100A10 and phospholipid monolayers and bilayers. *Langmuir*, 37(32), 9652-9663. <https://doi.org/10.1021/acs.langmuir.1c00342>
- Yan, X., Lebel-Beaucage, M. F., Tremblay, S., Cantin, L., Shaw, G. S., & Boisselier, E. (2019). Optimized transformation, overexpression and purification of S100A10. *BioTechniques*, 67(5), 246-248. <https://doi.org/10.2144/btn-2019-0081>
- Yates, J. R., Ruse, C. I., & Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annual review of biomedical engineering*, 11(1), 49-79. <https://doi.org/10.1146/annurev-bioeng-061008-124934>
- Yuan, T. H., Yue, Z. S., Zhang, G. H., Wang, L., & Dou, G. R. (2021). Beyond the liver: Liver-eye communication in clinical and experimental aspects. *Frontiers in Molecular Biosciences*, 8, 823277. <https://doi.org/10.3389/fmolb.2021.823277>
- Zeng, F. Y., Gerke, V. and Gabius, H. J. (1993) Identification of annexin II, annexin VI and glyceraldehyde-3-phosphate dehydrogenase as calyculin-binding proteins in bovine heart. *Int. J. Biochem.* 25, 1019–1027. [https://doi.org/10.1016/0020-711X\(93\)90116-V](https://doi.org/10.1016/0020-711X(93)90116-V)
- Zhang, C., Yao, R., Chen, J., Zou, Q., & Zeng, L. (2021). S100 family members: Potential therapeutic target in patients with hepatocellular carcinoma: A STROBE study. *Medicine*, 100(3), e24135. DOI: [10.1097/MD.00000000000024135](https://doi.org/10.1097/MD.00000000000024135)

- Zhang, F., Ge, W., Ruan, G., Cai, X., & Guo, T. (2020). Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics*, 20(17-18), 1900276. <https://doi.org/10.1002/pmic.201900276>
- Zhang, K., Qin, X., Zhou, X., Zhou, J., Wen, P., Chen, S., ... & Zhuang, J. (2020). Analysis of genes and underlying mechanisms involved in foam cells formation and atherosclerosis development. *PeerJ*, 8, e10336. <https://doi.org/10.7717/peerj.10336>
- Zhang, X., Fang, A., Riley, C. P., Wang, M., Regnier, F. E., & Buck, C. (2010). Multi-dimensional liquid chromatography in proteomics—a review. *Analytica chimica acta*, 664(2), 101-113. <https://doi.org/10.1016/j.aca.2010.02.001>
- Zhang, Y., Zhang, C., Jiang, H., Yang, P., & Lu, H. (2015). Fishing the PTM proteome with chemical approaches using functional solid phases. *Chemical Society Reviews*, 44(22), 8260-8287. <https://doi.org/10.1039/C4CS00529E>
- Zhao, Y., Xue, Q., Wang, M., Meng, B., Jiang, Y., Zhai, R., ... & Fang, X. (2023). Evolution of mass spectrometry instruments and techniques for blood proteomics. *Journal of Proteome Research*, 22(4), 1009-1023. <https://doi.org/10.1021/acs.jproteome.3c00102>
- Zhuang, C., Wang, P., Sun, T., Zheng, L., & Ming, L. (2019). Expression levels and prognostic values of annexins in liver cancer. *Oncology Letters*, 18(6), 6657-6669. <https://doi.org/10.3892/ol.2019.11025>