



**HAL**  
open science

# Méthodes d'analyses protéomiques quantitatives pour la découverte et validation de biomarqueurs circulants : application aux maladies hépatiques

David Perez Compte

## ► To cite this version:

David Perez Compte. Méthodes d'analyses protéomiques quantitatives pour la découverte et validation de biomarqueurs circulants : application aux maladies hépatiques. Biotechnologie. Université Grenoble Alpes [2020-..], 2023. Français. NNT : 2023GRALV056 . tel-04434513

**HAL Id: tel-04434513**

**<https://theses.hal.science/tel-04434513>**

Submitted on 2 Feb 2024

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

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

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Biotechnologie

Unité de recherche : BGE - Laboratoire Biosciences et bioingénierie pour la Santé

**Méthodes d'analyses protéomiques quantitatives pour la découverte et validation de biomarqueurs circulants : application aux maladies hépatiques**

**Quantitative proteomic analysis methods for the discovery and validation of circulating biomarkers: application to liver diseases**

Présentée par :

**David PEREZ COMPTE**

Direction de thèse :

**Virginie BRUN**

DIRECTEUR DE RECHERCHE, Université Grenoble Alpes

Directrice de thèse

**Michel JAQUINOD**

Ingénieur-chercheur CEA, Université Grenoble Alpes

Co-encadrant de thèse

Rapporteurs :

**Jean ARMENGAUD**

DIRECTEUR DE RECHERCHE, CEA centre de Marcoule

**Marie-Pierre BOUSQUET**

PROFESSEURE DES UNIVERSITES, Université Toulouse 3 - Paul Sabatier

Thèse soutenue publiquement le **19 octobre 2023**, devant le jury composé de :

**Virginie BRUN**

DIRECTRICE DE RECHERCHE, CEA centre de Grenoble

Directrice de thèse

**Thomas DECAENS**

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,  
CHU Grenoble Alpes

Examineur

**Jean ARMENGAUD**

DIRECTEUR DE RECHERCHE, CEA centre de Marcoule

Rapporteur

**Marie-Pierre BOUSQUET**

PROFESSEURE DES UNIVERSITES, Université Toulouse 3 - Paul Sabatier

Présidente

Invités :

**Jérôme Boursier**

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER, CHU Angers





A la patience infinie et au soutien inestimable de ma compagne,  
de ma famille, de mes amis et de mes encadrants

## Remerciements :

Le fait que je considère cette thèse comme une aventure a été rendu possible par toutes les personnes extrêmement talentueuses et bienveillantes qui m'ont entouré pendant ces trois années. Je suis très reconnaissant d'avoir pu assouvir ma curiosité scientifique et d'avoir fait mes premiers pas dans la recherche dans un tel environnement.

Mich, y en a qui diront que t'es un ours, mais c'est parce qu'ils ne te connaissent pas beaucoup, tu es l'Ours avec un grand O qui n'accepte que les simples négations et qui connais toujours les meilleurs spots pour grimper, skier, cueillir des myrtilles ou des champignons. Un immense merci pour l'encadrement aux petits oignons tout au long de la thèse pendant laquelle tu as su être un mentor, c'est un mot lourd de sens pour moi car il décrit quelqu'un de très passionné et expérimenté qui non seulement transmet ses connaissances mais qui guide et fait grandir son apprenti, c'est une qualité terriblement rare mais dont je eu la chance de profiter. Je garderai des souvenirs plein la tête que j'espère on pourra ressortir lors de soirées pizza à Quaix pendant longtemps.

Virginie, un énorme merci pour m'avoir ouvert les portes de ton équipe il y a 5 ans parce que j'avais mis sur mon CV que je faisais du ski alpi et que j'allais forcément bien m'entendre avec Mich. Merci pour l'encadrement et la confiance que tu m'as fait pour les deux projets sur lesquels j'ai travaillé pendant la thèse. Tu as été un pilier fondamental lors des nombreuses réunions qu'on a faites, pas uniquement par tes qualités scientifiques mais surtout pour ton humanité. Je retiendrai pendant longtemps l'incroyable capacité pour motiver les gens qui travaillent autour de toi et pour mettre en valeur leurs qualités. Travailler avec toi a été une expérience profondément enrichissante.

Merci à mes premières collègues de cubicule Olga et Hélène pour m'avoir aussi bien accueilli (grand merci à toi Hélène pour la place à côté de la fenêtre avec vue sur le Néron). Puis merci aussi aux collègues de cubicule Victoria et Salomé que j'ai pu voir débarquer avec leur bonne humeur et leur joie de vivre puis qui sont reparties pour laisser la place au gang des thésards. Merci Lucas pour cette collaboration extrêmement enrichissante pendant laquelle j'ai pu goûter aux statistiques de haut niveau mais aussi pour ces weekends incroyables de RAID et de psicobloc. Merci Vaitson pour les débats sans fin, les rigolades, les repas à H3 et les verres en ville, pour ton énergie débordante à cause des cafés en trop et pour les pauses dès ton arrivée au labo. Hassan merci pour ton calme, tes conseils que je n'ai pas suivis (à tort) pour que je fasse mes graphes sur R, les pauses et repas mais surtout pour ta compagnie lors des congrès au Portugal et à Aix, ton seul défaut étant le manque d'équilibre sur un paddle-surf, je te nomme irrévocablement the EDyPest. Courage à vous pour la fin de votre thèse.

Merci Anne-Marie pour ta patience infinie avec le projet PSAQ<sup>+1</sup>, qui nous aura bien donné mal à la tête par moments mais dont nous pouvons être fiers de ce que nous avons accompli. Merci aussi à toi Christophe pour avoir su trouver des solutions aux impasses qui paraissaient impossibles à surmonter sur ce projet et pour la sérénité dont tu as fait malgré les centaines de fois où je suis venu te déranger. Encore merci à vous deux mais aussi à Yohann et Thomas pour ces réunions NASH et PSAQ<sup>+1</sup> qui ont représenté tout un défi pour ma part, avec un peu de recul, ces réunions sont pour moi des expériences inoubliables, qui m'ont permis d'apprendre à quoi ressemble la recherche scientifique de qualité avec des personnes qui foisonnent d'idées et de connaissances, vous aurez réussi à puiser toute mon énergie sans me faire faire du sport, j'en suis encore surpris mais toujours reconnaissant.

Sylvie, merci pour l'organisation des barbecues/jacuzzi et des sorties rando, pour les tartes aux myrtilles et à la rhubarbe et pour les confitures, je pense qu'il faut que je te remercie aussi, même si je n'étais pas là, d'avoir calmé l'Ours lorsqu'il s'arrachait les cheveux avec mes manuscrits ou mes erreurs sur le Qtrap.

Merci Marie pour ton aide au labo mais surtout pour m'avoir tenu compagnie pendant que je faisais mes tests ELISA, la musique sur Virage radio et Nostalgie était tout ce dont j'avais besoin à ces moments ;)

Un énorme merci à toi Benoit, avec qui j'ai pu partager quelques courses à la Bastille, plusieurs discussions en accéléré et surtout merci pour les fois où tu m'as sorti du pétrin avec tes macros Excel et ton appel pour préparer l'entretien alors que tu étais en vacances.

Merci à tous les autres intégrants de l'équipe EDyP pour toutes les discussions autour de la table lors des pauses café et pour avoir rendu le quotidien de ma thèse toujours aussi agréable. Les brioches, pains au chocolat et croissants seront un souvenir associé à EDyP pendant longtemps.

Merci Ben pour les dizaines de sorties escalade et ski, pour ta motivation inébranlable et pour repousser mes limites à chaque voie. Merci à toutes les personnes qui ont fait du ski et de l'escalade avec moi pendant la thèse, et tout particulièrement à Thibault, Thomas et Sam avec qui j'ai sans doute fait les meilleures sorties.

Gràcies als meus amics del Geriàtric : Fortinhus, Fiñi, Uri, Palu, Dami, Albert i Marc pels viatges, sopars i caps d'anys que em permetien agafar el bocinet d'aire necessari per tornar a submergir el cap durant uns mesos. Ens veiem poc, pero cada cop que ho fem és com si ens haguessim vist ahir, senyors ho hem de fer més sovint.

Gràcies als companys de fatigues, sobretot a tu Sergi per ser un coequipier de primera categoria pero també a l'Arnau, el Bartu i els altres companys de l'equip amb els que he pogut riure, entrenar i gaudir (i patir) fent esquí de muntanya per casa.

Je tiens aussi à remercier les entraîneurs de la Fédération Andorrane de ski et ceux du TIM, Thierry et Antoine tout particulièrement, car même si avec la thèse j'ai dû abandonner petit à petit les compétitions, je me suis accroché à cette passion autant que possible grâce à vous. Sans aucun doute, j'ai grandi avec le ski alpi et ça représentera toujours une partie de moi.

Merci à ma famille, à ma belle-famille, à mes amis et à chaque personne qui m'a soutenu. Plus particulièrement je veux remercier du fond du cœur mes parents, pour leur soutien sans faille malgré la distance. Je vous suis pleinement reconnaissant de m'avoir laissé choisir la voie que je voulais en venant à Grenoble pour faire mes études en sachant que ça impliquait de partir loin de la maison pendant des années.

Et enfin mes remerciements les plus tendres sont pour celle avec qui je partage ma vie, Julia. Merci pour ton soutien tout au long de la thèse, tu connais tous les échecs et réussites, tu as vécu avec moi les moments où tout se passait pour le mieux et ceux plus compliqués. C'est peut-être mon nom qui est écrit sur ce manuscrit mais sans toi cette thèse aurait été complètement différente. Tout cela m'est très précieux et j'ai hâte de continuer à partager la suite avec toi.

## Résumé :

L'objectif de mes travaux de thèse était de développer et mettre en application des méthodes de quantification en protéomique. Dans un premier volet, chez des patients atteints de MASLD une maladie hépatique métabolique touchant 25% de la population mondiale, des biomarqueurs plasmatiques indicateur de l'état de fibrose hépatique ont été découverts et validés. L'ALS et la LG3BP, deux de ces marqueurs, ont été validés sur deux cohortes indépendantes puis ont été introduites dans un nouveau panel dérivé du FibroTest. Le deuxième volet de cette thèse consistait à optimiser les quantifications absolues en LC-MS basées sur l'utilisation de protéines standards isotopiquement alourdis (PSAQ). L'originalité du projet PSAQ<sup>+1</sup> consistait à introduire la protéine standard assimilée à la protéine endogène dont toutes les arginines et lysines sont marquées uniquement avec un <sup>13</sup>C au début du pipeline de protéomique quantitative. Les résultats ont montré la fiabilité et la robustesse d'une telle approche apportant plus de spécificité, sensibilité et précision lors des quantifications absolues par LC-MS/MS utilisant des standards isotopiques.

Mots clés: Protéomique ; biomarqueurs ; MASLD ; quantification par dilution isotopique

## Summary :

The aim of my thesis was to develop and apply proteomic quantification methods. Firstly, plasma biomarkers were discovered and validated in patients suffering from MASLD, a metabolic liver disease that affects 25% of the world's population. ALS and LG3BP, two of these markers, were validated in two independent cohorts and then incorporated into a new panel derived from FibroTest. The second part of this work involved the optimisation of absolute LC-MS quantifications based on the use of isotopically enriched standard proteins (PSAQ). The originality of the PSAQ+1 project was to introduce the standard protein assimilated to the endogenous protein with all arginines and lysines labelled only with <sup>13</sup>C at the beginning of the quantitative proteomics pipeline. The results demonstrated the reliability and robustness of such approach, providing greater specificity, sensitivity and precision for absolute quantification by LC-MS/MS using PSAQ<sup>+1</sup>.

Keywords : Proteomics; biomarkers; MASLD; quantification by isotope dilution





## **Table des matières :**

Remerciements :	3
Résumé :	5
Summary :	5
Abréviations :	9
Liste des Figures :	11
Liste des Tableaux :	13
Liste des Equations :	13
Liste des Annexes :	14
Introduction :	15
1. Les différentes maladies hépatiques :	17
2. La stéatopathie métabolique ou MASLD :	18
3. Intérêt de la découverte de nouveaux biomarqueurs :	20
4. La protéomique :	24
Partie I :	27
Découverte, vérification et validation de biomarqueurs dans le cadre de la MASLD :	27
A. Analyse des cohortes :	27
1. Composition des cohortes :	27
2. Comparaison des deux cohortes :	29
3. Corrélation de variables :	30
4. Stratégie d'analyse des échantillons :	32
B. Etude découverte :	33
1. Stratégie de préparation des échantillons :	33
2. Détails sur les analyses MS et le process des données :	34
3. Résultats d'identification des protéines plasmatiques :	34
3. Résultats de l'étude semi-quantitative par DDA :	35
C. Vérification ou évaluation de l'utilité clinique des cibles protéiques :	36
1. Stratégie d'analyse de la cohorte en protéomique ciblée :	36
2. Choix des protéines, peptides et transitions :	38
3. Optimisation de la séparation des peptides en chromatographie liquide :	40
4. Utilisation de peptides marqués isotopiquement :	41
5. Modus operandi en SRM :	44
6. Traitement du signal et analyse des résultats SRM :	44
7. Vérification par ELISA :	46
D. Etude de validation :	47
1. Tendances des 3 études :	47

2.	Le cas du C7 : .....	48
3.	Traitement statistique : .....	50
E.	Création d'un panel : .....	52
1.	Introduction : .....	52
2.	Re-fittage FibroTest : .....	53
F.	Conclusion projet MASLD : .....	58
	Partie II : .....	61
	Projet PSAQ <sup>+1</sup> : optimisation d'une technique de quantification absolue par utilisation de standards protéiques : .....	61
A.	Introduction projet PSAQ <sup>+1</sup> : .....	61
1.	Généralités : .....	61
2.	Concepts et définitions : .....	62
3.	Méthodes de quantification par PSAQ : .....	64
4.	Quantification par utilisation de standards protéiques : .....	66
B.	Expérimentations PSAQ <sup>+1</sup> : .....	69
1.	Production des protéines recombinantes ALDOB non-marquée / PSAQ / PSAQ <sup>+1</sup> : .....	69
2.	Injections des différentes formes de façon indépendante : .....	69
3.	Corrections modèle théorique : .....	76
4.	Corrections modèle expérimental : .....	82
5.	Calculs de ratios forme non-marquée / PSAQ <sup>+1</sup> : .....	86
C.	Conclusion projet PSAQ <sup>+1</sup> : .....	90
	Conclusion générale : .....	92
	Annexes : .....	94
	Références : .....	119
	Article en cours de soumission : .....	147
	Abstract : .....	168
	Résumé: .....	168
	Summary: .....	168

## Abréviations :

AASLD : Association américaine pour l'étude des maladies du foie  
AAT : Alpha-1-antitrypsine  
AFEF : Association Française pour l'Étude du Foie  
ALD : Atteinte hépatique d'origine alcoolique (Alcoholic liver disease)  
ALDOB : Aldolase-B  
Alpha2m ou A2m : Alpha-2-macroglobuline  
ALS : Insulin-like growth factor-binding protein complex acid labile  
ALT : Alanine aminotransférase  
APOB : Apolipoprotéine B (APOM : Apolipoprotéine M ; APOL1 : ... ; APOD : ... ; APOA1;...)  
AQUA : Absolute quantification peptide  
ASH : Stéatohépatite d'origine alcoolique (Alcoholic steatohepatitis)  
AST : Aspartate aminotransférase  
AUROC : Aire sous la courbe (AUC) "receiver operating characteristic" (ROC)  
BIL : Bilirubine  
C7 : Complément 7  
CHC : Carcinome hépatocellulaire  
CHONS : Carbone, Hydrogène, Oxygène, Azote, Soufre  
CI : Intervalle de confiance  
CRP : Protéine C-réactive  
C<sub>ter</sub> : C-terminal  
CV : Coefficient de variation  
DDA : Data dependent acquisition  
DIA : Data independent acquisition  
EASL : European Association for the Study of the Liver  
ELF : Enhanced liver fibrosis panel  
FC : Ratio d'abondance (Fold change)  
FDR : Taux de fausse découverte (False discovery rate)  
GGT: Gamma-glutamyltransférase  
HA : Acide hyaluronique  
HCD : Dissociation collisionnelle à haute énergie (Higher energy collisional dissociation)  
HP : Haptoglobuline  
HPLC : Chromatographie liquide à haute performance (LC : Chromatographie liquide ; RPLC : Chromatographie liquide en phase inverse)  
IMC : Indice de masse corporel  
INR : International normalized ratio  
IRM : Imagerie par résonance magnétique  
LDL et HDL : Lipoprotéine de basse densité et haute densité  
LG3BP : Galectin-3-binding protein  
LT : Transplantation hépatique (Liver transplantation)  
MASH : Stéatohépatite métabolique (Metabolic dysfunction-associated steatohepatitis)  
MASLD : Stéatopathie métabolique (Metabolic dysfunction-associated steatotic liver disease)  
MEC : Matrice extracellulaire  
ML : Machine learning  
MS : Spectrométrie de masse (MS/MS : spectrométrie de masse en tandem)  
NAFLD : Maladie du foie gras d'origine non-alcoolique (Non-alcoholic fatty liver disease)  
NASH : Hépatite stéatosique d'origine non-alcoolique (Non-alcoholic steatohepatitis)  
NFS : NAFLD fibrosis score  
P0 : isotopologue 0 (P1 : isotopologue 1 ; P2 : ...)  
PRM : Parallel reaction monitoring  
PSAQ : Protein standard absolute quantification

QconCAT : Quantification concatemers  
RF: Re-fitté (Re-fitted)  
SAF: Stéatose, activité, fibrose  
SRM : Selection reaction monitoring (MRM : Multiple reaction monitoring)  
TP : Taux de prothrombine  
VHA: Virus de l'hépatite A (VHB : Virus de l'hépatite B ; VHC : ...)  
WPM : Well plate maker  
XIC : Extracted-ion chromatogram

## Liste des Figures :

Figure 1 : Schéma de l'architecture du foie.....	15
Figure 2 : Stades d'avancement du foie d'un patient MASLD.....	18
Figure 3 : Modèle de développement de la MASLD.....	19
Figure 4 : Coupe histologique issue d'une biopsie hépatique présentant une cirrhose.....	20
Figure 5 : Détails du calcul du score SAF en fonction des observations histologiques.....	23
Figure 6 : Schéma des marqueurs de MASLD présents dans la littérature scientifique.....	24
Figure 7 : Stratégie triangulaire pour la découverte, vérification et validation de biomarqueurs.....	25
Figure 8 : Résumé graphique du projet MASLD.....	27
Figure 9 : Statistiques descriptives des cohortes de Grenoble et Angers.....	30
Figure 10 : Corrélations entre les variables communes aux cohortes de Grenoble et Angers.....	31
Figure 11 : Conception de l'étude et utilisation des cohortes.....	33
Figure 12 : Schéma de la répartition des échantillons à partir de WPM.....	33
Figure 13 : Résultats de protéomique issus de l'étude de découverte sur la cohorte de Grenoble.....	35
Figure 14 : Volcano plot réalisé à partir des données obtenues par LC-MS/MS en mode DDA.....	36
Figure 15 : Stratégie « rectangulaire ».....	37
Figure 16 : Choix des protéines pour les différentes étapes de l'étude de vérification.....	38
Figure 17 : Fragmentation MS/MS d'un peptide selon la nomenclature Biemann et Roepstorff.....	40
Figure 18 : Schéma du principe de fonctionnement d'une LC.....	41
Figure 19 : Schéma général décrivant le principe de la quantification par peptides AQUA.....	42
Figure 20 : Spectre de masse de tous les peptides ciblés pour l'étude SRM.....	43
Figure 21 : Schéma du traitement des données sur Skyline.....	45
Figure 22 : Tendances obtenues pour les protéines ALS, C7 et LG3BP.....	47
Figure 23 : Tendances des protéines ALS et LG3BP en fonction des stades de fibrose.....	48
Figure 24 : Tendance du C7 en fonction des stades de fibrose.....	48
Figure 25 : Schéma du complexe d'attaque membranaire.....	49
Figure 26 : Comparaison des abondances moyennes de l'ALS et de la LG3BP.....	51
Figure 27 : Courbes ROC et AUC de l'ALS, la LG3BP et du FibroTest.....	52
Figure 28 : Courbes ROC et AUC du FibroTest et de la combinaison {ALS, LG3BP}.....	53
Figure 29 : Courbes ROC et AUROCs du FibroTest RF et {Alpha2m, GGT} avec ALS et LG3BP.....	55
Figure 30 : Techniques de quantification relative label-free utilisées en protéomique LC-MS.....	61
Figure 31 : Schématisation atomistique des isotopes du carbone ( $^{12}\text{C}$ et $^{13}\text{C}$ ).....	62
Figure 32 : Schématisation de la distribution isotopique des carbones $^{11}\text{C}$ , $^{12}\text{C}$ , $^{13}\text{C}$ et $^{14}\text{C}$ .....	63
Figure 33 : Distribution isotopique de l'état de charge $z=18$ de la myoglobine.....	64
Figure 34 : Pipeline de protéomique quantitative avec dilution isotopique.....	65
Figure 35 : Comparaison des fenêtres de sélection.....	66
Figure 36 : Représentation de l'arginine et la lysine dans un échantillon endogène (ou non marqué), un échantillon PSAQ et un échantillon PSAQ $^{+1}$ .....	67
Figure 37 : Schématisation des distributions isotopiques obtenues lors de l'analyse d'un peptide non-marqué et d'un peptide PSAQ $^{+1}$ .....	68
Figure 38 : Détails du code sur R studio permettant de calculer la proportion théorique de peptide non-marqué.....	68
Figure 39 : Comparaison des principes d'analyse PRM et SRM.....	70
Figure 40 : Schématisation de deux mélanges endogène / PSAQ (A) et endogène / PSAQ $^{+1}$ (B).....	71
Figure 41 : Schématisation des pics inclus dans l'analyse en fonction de différentes configurations de fenêtres de sélection.....	72

Figure 42 : Détail du calcul pour obtenir le pourcentage de différence entre le modèle théorique et expérimental .....	73
Figure 43 : Intensité de chaque fragment des 6 peptides suivis pour l'ALDOB .....	75
Figure 44 : Schéma du pipeline de quantification par PSAQ+1 lors de l'utilisation d'une fenêtre de sélection m/z de 1,6 .....	77
Figure 45 : Schéma du pipeline de quantification par PSAQ+1 lors de l'utilisation d'une fenêtre de sélection m/z de 1,6 tenant en compte un massif isotopique supplémentaire par rapport à la Figure 44 .....	78
Figure 46 : Détails de la formule pour calculer la quantité de fragments légers venant de précurseurs non-marqués avec différentes quantités de <sup>13</sup> C. ....	79
Figure 47 : Schéma du pipeline de quantification par PSAQ+1 lors de l'utilisation d'une fenêtre de sélection m/z de 0,4 .....	80
Figure 48 : Ratios théoriques en fonction des ratios expérimentaux de tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection m/z 0,4.....	81
Figure 49 : Comparaison des intensités obtenues en MS/MS lors d'une analyse de peptides de l'ALDOB non-marquée avec une fenêtre de sélection m/z 0,4 ciblée sur le P1 du massif isotopique. ....	83
Figure 50 : Représentation des pics obtenus lors d'une analyse d'un peptide chargé 2+ sans énergie de collision en MS <sub>1</sub> et en MS <sub>2</sub> .....	84
Figure 51 : Forme d'onde de la tension d'impulsion de la foudre d'après Ghosh, P. et al .....	85
Figure 52 : Courbe de concentration de la moyenne des peptides de l'ALDOB .....	87
Figure 53 : Ratios entre fragment non-marqué et marqué pour chacun des fragments de l'ALDOB... ..	88

## Liste des Tableaux :

Tableau 1 : Avantages et inconvénients des différentes méthodes de diagnostic de la MASLD.....	22
Tableau 2 : Caractéristiques des patients de la cohorte de Grenoble (n=160) et Angers (n=200) .....	28
Tableau 3 : Données générales sur les cohortes de Grenoble et Angers ainsi que sur les populations française, européenne et mondiale MASLD et non-MASLD .....	30
Tableau 4 : Tableau regroupant les variables corrélées entre les cohortes de Grenoble et d'Angers.	32
Tableau 5 : Exemple de séquences peptidiques et transitions choisies et des paramètres associés permettant de les cibler et les analyser en SRM.....	39
Tableau 6 : Protéines différentiellement abondantes en fonction du stade de fibrose .....	45
Tableau 7 : Tableau regroupant les protéines majoritaires observées dans un échantillon de standard de C7 après injection en LC-MS/MS en mode DDA.....	50
Tableau 8 : Tableau des résultats des likelihood ratio tests entre les modèles de FibroTest RF.....	54
Tableau 9 : P-valeurs des likelihood ratio tests pour le FibroTest RF et {Alpha2m, GGT} lors de l'ajout de l'ALS et la LG3BP ou de la combinaison {ALS ;LG3BP} sur les données d'Angers.....	55
Tableau 10 : Seuils de sensibilité et spécificité du modèle {Alpha2m, GGT, ALS, LG3BP} .....	56
Tableau 11 : Tableau de comparaison de l'AUROC obtenue sur plusieurs tests de 1 <sup>ère</sup> ligne, de panels issus de tests de 1 <sup>ère</sup> ligne, sur des tests de 2 <sup>ème</sup> ligne et sur les biomarqueurs et panels développés dans le projet MASLD. ....	57
Tableau 12 : Isotopes naturels des éléments présents dans les protéines.....	64
Tableau 13 : Pourcentage de pureté de l'échantillon PSAQ <sup>+1</sup> .....	70
Tableau 14 : Abondance de forme non-marquée lourde dans échantillon de forme non-marquée calculée à partir de données théoriques.....	73
Tableau 15 : Pourcentage de différence obtenu entre abondance théorique et expérimentale (intensité et aire) sur 3 fenêtres de m/z (0,4-0,8-1,6).....	74
Tableau 16 : Comparaison de deux conditions lors du calcul du pourcentage de différence entre abondance théorique et expérimentale .....	76
Tableau 17 : Comparaison du pourcentage de différence entre théorique et expérimental pour des intensités et des aires obtenues en MS <sub>1</sub> sur des fenêtres de sélection m/z de 1,6 et 0,4. ....	85
Tableau 18 : Dilutions et répliques réalisés pour le mélange d'ALDOB légère et d'ALDOB PSAQ <sup>+1</sup> .....	86
Tableau 19 : Nombre de peptides avec un coefficient de variation des intensités inférieur à 20% pour différentes conditions de tri.....	89

## Liste des Equations

Equation 1 : Equation du FibroTest. ....	52
Equation 2 : Equation du modèle {Alpha2m, GGT, ALS, LG3BP}. ....	56
Equation 3 : Equation permettant de calculer la proportion de molécules ne possédant qu'un seul isotope alourdi en fonction du nombre d'atomes ciblés de la molécule et de l'abondance de l'isotope alourdi. ....	63

## Liste des Annexes :

Annexe A : Résumé des différents panels décrits dans l'étude bibliographique et de leur AUROC.....	94
Annexe B : Résumé des différents biomarqueurs décrits la bibliographie et de leur AUROC.....	96
Annexe C : Schéma résumant les principes des modes d'acquisition DDA et DIA.....	101
Annexe D : Détails de l'injection SRM pour la quantification de peptides du projet NAFLD.....	101
Annexe E : Tableau avec la masse et le temps de rétention moyen de chaque protéine, peptide et fragment choisis pour l'étude SRM.....	102
Annexe F : Distribution des coefficients de variation de l'intensité pour les données SRM.....	104
Annexe G : Courbes par stade de fibrose des protéines C5, C6, C7, C8A, C8B, C8G et C9.....	104
Annexe H : Tableau de comparaison de l'AUROC obtenue sur plusieurs tests de 1ère ligne, de panels issus de tests de 1ère ligne, sur des tests de 2ème ligne et sur les biomarqueurs et panels développés dans le projet MASLD.....	105
Annexe I : Protocole de réduction, dénaturation, alkylation, digestion et dessalage pour le mix ALDOB sans marquage et PSAQ <sup>+1</sup> .....	106
Annexe J : Détails de l'injection LC-MS du mix ALDOB sans marquage et PSAQ <sup>+1</sup> .....	107
Annexe K : Comparaison du pourcentage de différence pour chaque fragment des peptides.....	107
Annexe L : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 0,4m/z.....	110
Annexe M : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 0,8 m/z.....	111
Annexe N : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 1,6 m/z.....	112
Annexe O : Pourcentage de différence entre ratio expérimental et théorique sur les données d'intensité obtenues en MS <sub>1</sub> pour chacun des peptides de l'ALDOB.....	113
Annexe P : Comparaison des intensités obtenues en MS <sub>2</sub> lors d'une analyse de peptides de l'ALDOB en PRM et sans énergie de collision.....	114
Annexe Q : Comparaison des intensités obtenues en MS <sub>2</sub> lors d'une analyse de peptides de l'ALDOB en PRM et sans énergie de collision.....	115
Annexe R : Représentation des pics obtenus lors d'une analyse d'un peptide chargé 2+ sans énergie de collision en MS <sub>1</sub> et en MS <sub>2</sub> .....	116
Annexe S : m/z des fragments inclus dans différentes conditions de sélection.....	117
Annexe T : Coefficients de variation moyens des conditions C et E pour les différentes concentrations et répliques de la courbe de calibration.....	118



## Introduction :

Les maladies hépatiques représentent un des grands enjeux médicaux au niveau mondial. Toutes les zones géographiques sont touchées différemment par les atteintes hépatiques<sup>1</sup>. Ainsi, les pays les moins développés sont davantage touchés par les hépatites virales ou les atteintes hépatiques dues aux aflatoxines à cause d'un manque de moyens sanitaires et économiques pour gérer ces maladies. D'un autre côté, les pays émergents et les pays développés subissent plutôt des atteintes hépatiques métaboliques et cancéreuses. Ces atteintes sont principalement liées à une surconsommation nutritionnelle, alcoolique et/ou médicamenteuse et au vieillissement de la population. Ainsi, les maladies du foie sont directement liées au mode de vie des personnes<sup>1</sup>. Le foie est un organe vital pour le fonctionnement normal de l'organisme. Chez l'Homme, cet organe densément vascularisé est composé de deux lobes qui sont alimentés par deux vaisseaux majeurs, l'artère hépatique et la veine portale. Parallèlement aux vaisseaux sanguins, le foie est parcouru par un grand nombre de voies biliaires permettant l'évacuation de la bile<sup>2</sup> (Figure 1). Les lobes sont composés de lobules hexagonaux permettant un ravitaillement efficace par la triade portale (canaux biliaires, branches de la veine porte et l'artère hépatique). La structure des lobules est maintenue par les cellules du foie, aussi appelés hépatocytes. En fonction de leur localisation, les hépatocytes seront approvisionnés par différents types de canaux et auront donc des fonctions distinctes<sup>3</sup> (Figure 1).

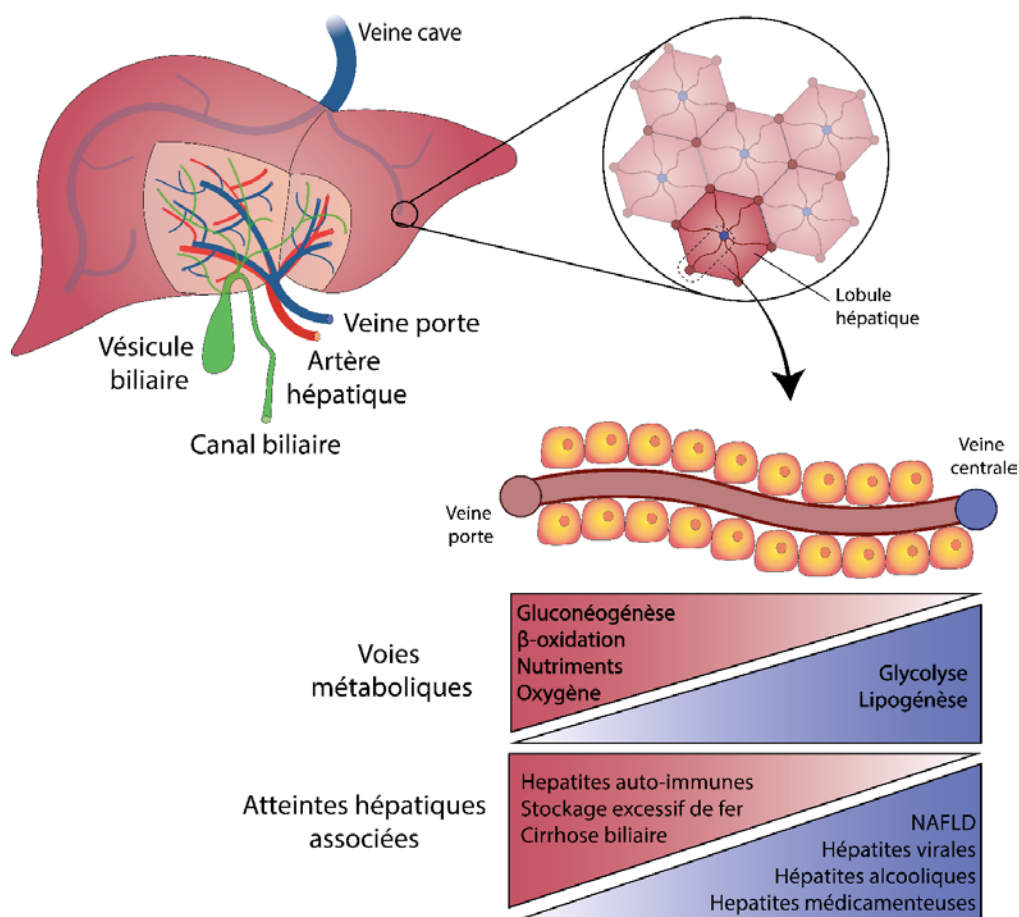


Figure 1 : Schéma de l'architecture du foie, des lobules hépatiques et des voies métaboliques associées aux différentes zones du lobule hépatique et les maladies associées à ces zones. Les hépatocytes autour de la veine centrale participeront à la glycolyse et à la production de lipides alors que ceux à proximité de la veine porte seront plus enclins à synthétiser du glucose, dégrader des acides gras et recevoir les apports en oxygène, hormones et nutriments. Les hépatocytes sont donc des cellules capables à la fois de synthétiser des molécules (lipides, glucides, protéines, acides biliaires...) et d'en cataboliser d'autres (hormones, lipides, bilirubine, ammoniac, alcool ou des xénobiotiques comme les médicaments, les drogues ou les molécules toxiques)<sup>4,5</sup>.

Sans rentrer dans les détails, les fonctions du foie sont très diverses et d'une importance vitale pour la plupart des organes humains. En effet, les molécules synthétisées, métabolisées ou stockées par le foie sont nécessaires à l'organisme pour un fonctionnement normal. Celles-ci sont :

#### Le stockage :

Le foie agit comme organe de stockage de nombreux produits. Par exemple, le glucose est transformé en glycogène puis stocké jusqu'au moment où l'organisme aura un besoin énergétique. Le foie est également capable de stocker des métaux comme le fer provenant de l'hémoglobine ou encore le cuivre. Les vitamines liposolubles, telles que les vitamines A, D, E, K et B12, arrivent au foie après absorption intestinale puis sont métabolisées et emmagasinées avant d'être utilisées<sup>6</sup>. Par exemple, la vitamine K se retrouve dans l'enzyme hépatique gamma-glutamyl carboxylase et permet l'activation des facteurs de coagulation (II, VII, IX, X, XIX ou la protéine S)<sup>7</sup>.

#### La détoxification :

Certaines substances comme les hormones (œstrogènes, aldostérones, hormones antidiurétiques...), l'ammoniac, l'alcool ou autres xénobiotiques (médicaments, drogues, toxiques...)<sup>4,5</sup> sont dégradés par le foie. Pour cela, le foie transforme ces molécules lipophiles en molécules hydrophiles à travers plusieurs réactions<sup>6</sup>. Des variables telles que l'âge, le sexe ou la présence de maladies hépatiques peuvent faire varier l'efficacité de détoxification du foie<sup>8</sup>.

#### La production biliaire :

Le foie est également le lieu de synthèse des acides biliaires et d'autres composants de la bile intervenant dans la digestion des graisses. La bile a pour fonction de transporter jusqu'à l'intestin les produits liposolubles à éliminer après leur passage dans le foie<sup>6</sup>. La bile est sécrétée par les hépatocytes, puis stockée dans la vésicule biliaire. Elle sera ensuite relarguée dans l'intestin lors de la digestion, elle permet notamment de former des gouttelettes avec les graisses afin qu'elles soient dégradées<sup>9</sup>.

#### La production de protéines :

Les protéines sont constamment synthétisées et dégradées en fonction des besoins de l'organisme. Le foie est un organe fortement lié à ces processus. Par exemple, il obtient plus de la moitié de l'énergie nécessaire à son fonctionnement à partir de l'oxydation d'acides aminés<sup>10</sup>. Les protéines plasmatiques les plus abondantes sont l'albumine (~55%), les globulines (~38% dont plus de la moitié sont des immunoglobulines), le fibrinogène (~7%) puis les protéines ayant des rôles dans la régulation de l'expression des gènes (<1%) et la coagulation (<1%)<sup>11</sup>. D'autres types de protéines existent mais leurs abondances sont plusieurs ordres de grandeur plus faibles. La plupart des protéines sont produites dans le foie, à l'exception de la fraction  $\gamma$  de la globuline (produite dans la moelle rouge des os)<sup>10</sup>. En cas de dysfonctionnement hépatique, on observe généralement une dérégulation de protéines dans le sang, donnant lieu par exemple à des troubles de la coagulation pouvant aboutir à des hémorragies<sup>12</sup>.

#### Transformation de la bilirubine libre :

La bilirubine, produit de dégradation des hèmes, est une molécule relarguée par les globules rouges (hémoglobine) lorsqu'ils sont dégradés<sup>6</sup>. Plus précisément, l'hémoglobine sera décomposée en globine et en hème. L'hème subira plusieurs processus jusqu'à se transformer en bilirubine libre ou non-conjuguée véhiculée essentiellement par l'albumine puis sera conjuguée à l'acide gluconique dans

le foie. Une partie de la bilirubine conjuguée sera filtrée par le rein et une deuxième partie sera captée par les hépatocytes et introduite dans la bile. Une accumulation trop importante de bilirubine résultant d'une atteinte hépatique peut causer un jaunissement des yeux ou de la peau<sup>6</sup>.

### 1. Les différentes maladies hépatiques :

Le foie étant un organe avec une multitude de fonctions, il est exposé à des agressions multiples et variées. Les symptômes des atteintes hépatiques présentent une variabilité importante. En plus de la jaunisse déjà évoquée dans le cadre de l'accumulation de bilirubine, certains patients peuvent présenter une fatigue importante, des nausées ou des démangeaisons de la peau dans les stades précoces. Les atteintes hépatiques plus avancées peuvent à leur tour se manifester par un gonflement abdominal (ascite), des encéphalopathies ou des hypertensions au niveau de la veine porte<sup>13</sup>. Prises en charge précocement ces atteintes hépatiques, de par l'exceptionnelle capacité de régénération du foie peuvent être inversées<sup>14</sup>.

Les maladies du foie de plus en plus fréquentes, notamment du fait d'une mauvaise alimentation et d'un manque d'activité physique, sont bien souvent d'origine métabolique. Parmi les autres atteintes, la plus commune est d'origine virale. L'atteinte virale va se propager dans les hépatocytes et causer leur dysfonctionnement. Il existe plusieurs types d'hépatites virales : hépatite A, B, C, D, E et G (VHA, VHB, VHC, VHD, VHE et VHG respectivement), la forme VHB étant tout particulièrement répandue avec 3,5% de la population mondiale infectée en 2018<sup>15</sup>. Les hépatites virales peuvent se manifester sous forme aiguë ou chronique<sup>16</sup>.

D'autres types d'atteintes hépatiques sont celles d'origine génétique. Par exemple, la maladie de Wilson se caractérise par une insuffisance en céruloplasmine, une enzyme produite dans le foie, chargée de 90% du transport du cuivre dans le sang<sup>17</sup>. D'autres atteintes hépatiques d'origine génétique existent comme l'hémochromatose<sup>18</sup> ou la déficience en alpha-1-antitrypsine (AAT)<sup>19</sup>.

Les atteintes hépatiques toxiques sont également un problème de santé mondial et peuvent avoir différentes causes. Tout d'abord, les aflatoxines sont des mycotoxines produites par un champignon (*Aspergillus flavus* et *Aspergillus parasiticus*) provoquant des problèmes hépatiques, des malformations fœtales ou des problèmes immunitaires<sup>20</sup>. La surconsommation d'alcool peut également conduire à une atteinte hépatique appelée ALD (Atteinte hépatique d'origine alcoolique ou alcoholic liver disease) pouvant se dégrader en stéatohépatite alcoolique (ASH) ou cirrhose<sup>21</sup>. Finalement, des cas d'hépatites aiguës médicamenteuses, liées le plus souvent à une absorption excessive de médicaments comme le paracétamol, sont également une cause de dommage hépatique<sup>22</sup>.

Le cancer du foie est une atteinte maligne se développant à partir des tissus hépatiques endommagés par une maladie chronique. Le cancer hépatique peut être causé par des hépatites virales<sup>23</sup>, par l'excès de certains toxiques (alcool, fer...) <sup>24</sup> ou même par des maladies métaboliques<sup>25</sup>. La forme la plus commune est le carcinome hépatocellulaire (CHC) représentant 80% des cancers hépatiques et la cause de décès la plus fréquente chez les personnes atteintes de cirrhose<sup>23</sup>. Le CHC est le 6<sup>ème</sup> cancer avec l'incidence la plus élevée au niveau global et le 4<sup>ème</sup> cancer le plus mortel<sup>26</sup>. Le CHC peut se déclarer dans un foie totalement sain mais cette situation reste exceptionnelle.

Cependant, ce sont les atteintes d'origine métabolique qui affectent le plus de personnes dans le monde. Les premiers stades de celles-ci sont appelés MASLD, stéatopathie métabolique ou maladie du "foie-gras" non alcoolique (MASLD en anglais pour metabolic dysfunction-associated steatotic liver disease). Les atteintes les plus graves sont appelées MASH ou stéatohépatite métabolique (MASH pour metabolic dysfunction-associated steatohepatitis) (Figure 2). Les atteintes hépatiques métaboliques

ont récemment subi un changement de nomenclature, elles étaient jusque-là connues sous le nom de NAFLD (Non-alcoholic fatty liver disease) et NASH (Non-alcoholic steatohepatitis)<sup>27</sup>.

## 2. La stéatopathie métabolique ou MASLD :

La maladie dite du « foie gras » non-alcoolique ou encore MASLD, avec l'augmentation croissante de l'obésité est devenue l'une des causes les plus courantes de maladie hépatique métabolique dans le monde<sup>28</sup>. C'est un problème majeur de santé publique. Cette pathologie caractérisée par une accumulation importante de graisses dans le foie, survient en dehors de toute consommation excessive d'alcool. Elle englobe un spectre clinico-pathologique large de maladies suivant l'évolution suivante : (1) stéatose simple et progressive, (2) MASH, (3) cirrhose, (4) CHC<sup>29</sup> (Figure 2).

### Progression des atteintes hépatiques métaboliques

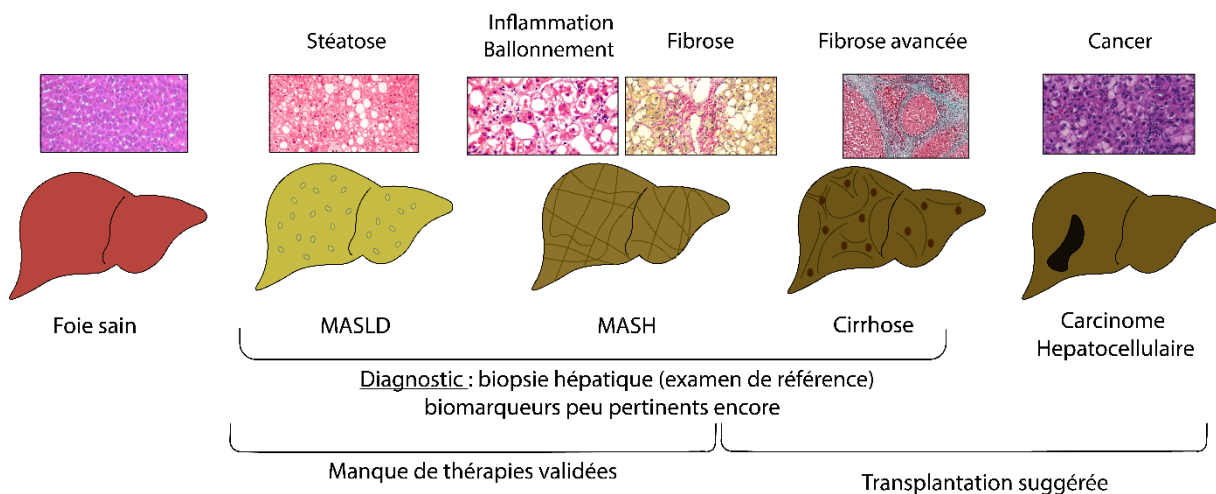


Figure 2 : Stades d'avancement du foie d'un patient MASLD. Images histologiques adaptées de Zakaria et al<sup>30</sup>, Hubscher et al<sup>31</sup>, Wikipedia<sup>32</sup> et El Jabbour et al<sup>33</sup>.

En 2023, les atteintes de type MASLD touchent 30% de la population mondiale dont 59% étaient à un stade plus avancé appelé MASH pour lequel l'indice de mortalité se situe à environ 15%<sup>34</sup>. Cette maladie est en constante augmentation, sa prévalence globale est passée de 15% en 2005 à 30% en 2023 ce qui peut se comprendre par l'augmentation de cas liés au surpoids et à l'obésité, ceux-ci étant également des facteurs de risque pour la MASLD<sup>28</sup>. En effet, l'obésité peut entraîner une hyperglycémie (taux de glucose en sang constamment supérieur à 7 mmol/L) ou une hypertriglycéridémie (taux de triglycérides en sang supérieur à 1,7 mmol/L) pouvant donc aboutir à une accumulation de gras dans le foie<sup>35,36</sup>. L'obésité peut aussi être une cause indirecte pour la MASLD car elle augmente la probabilité de souffrir d'un diabète de type 2<sup>37</sup>. En effet, les premiers stades de cette maladie se caractérisent par une hyper-insulinémie perturbant le métabolisme des acides gras intra-hépatiques entraînant in fine une stéatose<sup>38</sup>.

En termes de prévalence, la MASLD montre une augmentation constante dans la plupart des populations au cours des deux dernières décennies avec un effet plus marqué dans les pays d'Asie et de la région Asie-Pacifique : zones reflétant des changements rapides de mode de vie<sup>39</sup>. Les patients atteints de MASLD peuvent montrer une certaine prédisposition génétique. Par exemple le variant *Ile<sub>148</sub>Met* du gène *PNPLA3* codant pour une « patatin like phospholipase domain-containing protein 3 » présente diverses activités liées à l'accumulation de gras telles que des activités lipase et d'acylglycérol transacylase. A ce jour, il s'agit du variant génétique présentant la plus forte prédisposition pour la MASLD<sup>40</sup>. D'autres variants liés à la MASLD peuvent être la mutation des gènes *PHKA2* ou *PHKB*

impliqués dans le dérèglement du stockage du glucose ou encore une mutation du gène *ApoB* codant pour l'Apolipoprotéine B (APOB) ou du gène *MTP* codant pour la « microsomal triglyceride transfer protein » provoquant des bêta-lipoprotéinémies<sup>41-43</sup>. Des études récentes montrent que d'autres gènes tels que *MBOAT7*, *GCKR*, *HSD17B13* et *TM6SF2*<sup>40</sup> contribuent également à l'apparition de cette pathologie mais de façon plus modérée<sup>44</sup>.

Bien qu'il soit établi que la MASLD soit fortement liée à l'obésité et à l'insulino-résistance, sa physiopathologie reste mal connue et les options thérapeutiques sont très limitées<sup>45</sup>. Les atteintes hépatiques métaboliques se caractérisent par des dégradations hépatiques (stéatose, fibrose, ballonnement, inflammation, cytolysse, cancérogenèse) plus ou moins importantes qui sont des indicateurs de l'état d'avancement de la maladie (Figure 2). Des troubles systémiques peuvent également avoir lieu tels que des déséquilibres lipidiques ou encore des perturbations du système périphérique (Figure 3). Au stade le plus précoce, la MASLD se caractérise par une accumulation accrue de lipides au niveau du foie engendrant un processus de lipotoxicité<sup>46</sup>. Cette lipotoxicité se met en place lorsque le foie n'est plus capable de métaboliser ou d'assimiler les molécules du fait de leur concentration excessive.

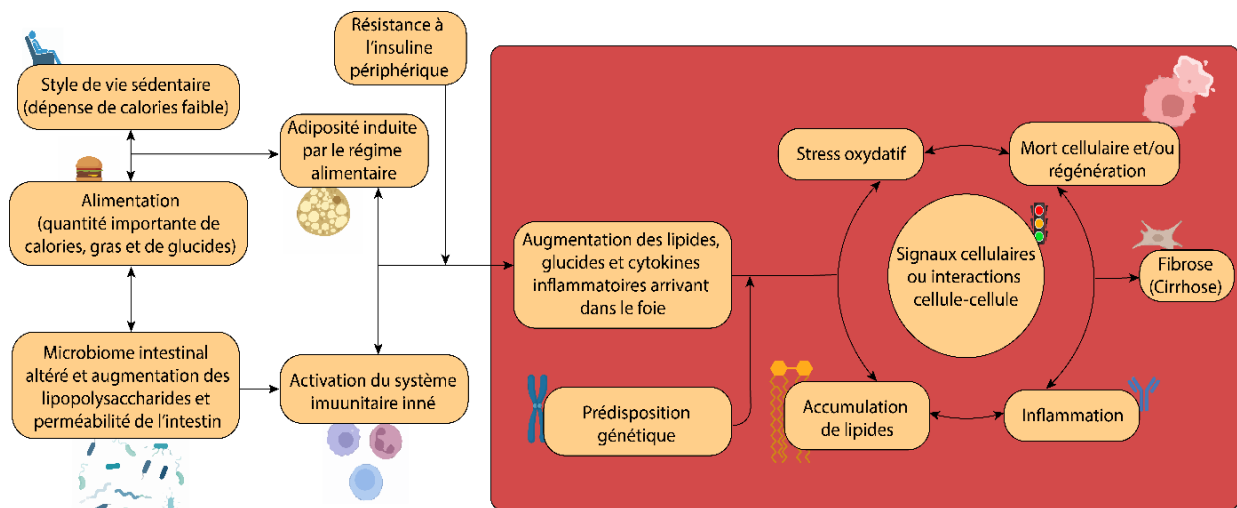


Figure 3 : Modèle de développement de la MASLD (adapté de Sanyal 2019<sup>47</sup>).

Toutefois, lors d'atteintes hépatiques chroniques telle que la MASLD, un dommage tissulaire persistant peut se produire entraînant une cicatrisation importante de tissus par le moyen d'une activation soutenue du système immunitaire et d'une réponse fibrillaire<sup>48</sup>. Cette réponse fibrillaire est aussi appelée fibrose et elle sera à l'origine d'une accumulation excessive de matrice extracellulaire (MEC) au niveau hépatique ainsi que de cellules appelées myofibroblastes. L'objectif de cette accumulation est de réparer les tissus hépatiques endommagés, la quantité et l'extension de la fibrose peut être un indicateur du degré d'atteinte hépatique. Lorsque la fibrose se produit pendant une longue période (en général 15 à 20 ans), celle-ci peut devenir irréversible entraînant le patient dans un stade dit de cirrhose. Au niveau histologique, des travées fibreuses peuvent être observées (Figure 4), celles-ci engendrent une obstruction des vaisseaux, le tissu hépatique n'est donc plus alimenté et par conséquent le foie ne peut plus fonctionner normalement. En effet, le manque d'irrigation sanguine ne permet pas une régénération correcte des lobules et des hépatocytes conduisant par ailleurs à une stagnation de la bile et une hypertension portale. Ceci va aboutir à une dégradation importante du foie et va être à l'origine de la mort cellulaire des hépatocytes ainsi qu'un arrêt progressif de toutes les fonctions hépatiques<sup>48</sup>.

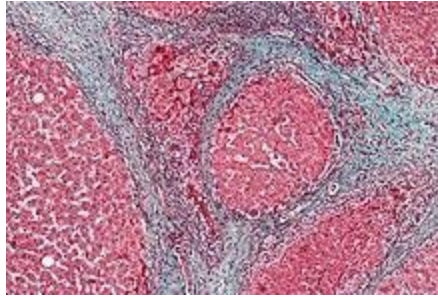


Figure 4 : Coupe histologique issue d'une biopsie hépatique présentant une cirrhose<sup>32</sup>.

Actuellement, les seuls traitements pour limiter les effets de la fibrose et la cirrhose hépatiques sont symptomatiques et essaient de prévenir de nouvelles complications (ex : dépistage CHC). Cela nécessite, entre autres, de contrôler la prise de médicaments pour les atteintes hépatiques avancées car les fonctions de filtrage du foie sont altérées par la maladie. Cependant, certaines molécules à visée thérapeutique, testées sur des essais cliniques avancés, suscitent de nombreux espoirs, mais à ce jour aucune autorisation de mise sur le marché n'a été délivrée<sup>49,50</sup>. Les bénéfices les plus importants pour le patient peuvent être tirés en identifiant puis limitant la cause de la fibrose (alimentation excessive, alcool, xénobiotiques, ...) et par la réalisation d'activité physique<sup>51</sup>. Le seul traitement curatif est la transplantation hépatique, mais cela reste une opération lourde et uniquement réalisable lorsqu'il n'y a pas de contre-indications et qu'il y a une disponibilité en organe<sup>52,53</sup>.

### 3. Intérêt de la découverte de nouveaux biomarqueurs :

Les biomarqueurs sont des indicateurs d'un état biologique<sup>54</sup>. Ils peuvent être de nature moléculaire, radiographique, histologique, physiologique ou numérique<sup>55</sup>. Les biomarqueurs peuvent être utilisés pour dépister ou diagnostiquer une maladie (ex : test génétique chez le nouveau-né pour la phénylcétonurie) mais également dans des contextes de surveillance, de pronostic, de prédiction ou encore pour évaluer la réponse à un traitement médical ou la toxicité d'une molécule<sup>56</sup>.

Selon les dernières directives établies par l'association américaine pour l'étude des maladies du foie (AASLD), la MASLD est diagnostiquée lorsque les 4 critères suivants sont remplis : (1) la présence de stéatose hépatique observée par imagerie ou histologiquement; (2) l'absence de consommation régulière et marquée d'alcool (apport d'éthanol <210 g / semaine pour les hommes et <140 g / semaine pour les femmes); (3) pas de présence d'autres facteurs induisant une stéatose hépatique; et (4) absence d'autres causes de maladie hépatique chronique<sup>57</sup>.

Actuellement, pour poser le diagnostic il existe plusieurs examens pour établir une atteinte MASLD que sont : l'échographie, la quantification de biomarqueurs sanguins, la mesure de la rigidité du foie ou encore la biopsie hépatique<sup>58</sup>. La biopsie hépatique est considérée comme l'examen de référence pour évaluer l'état du foie d'un patient<sup>59</sup>. Son principe consiste à prélever un morceau de foie à l'aide d'une aiguille afin d'être analysé par histologie (examen microscopique de coupes tissulaires) pour déterminer la pathologie du patient et son stade. La biopsie hépatique est un examen coûteux nécessitant une anesthésie locale, la mobilisation de plusieurs experts pour l'opération et l'occupation d'un lit hospitalier par le patient jusqu'à son rétablissement. Il s'agit également d'un examen invasif puisqu'un morceau de foie est prélevé et malgré la capacité de régénération hépatique, des risques pour la santé du patient peuvent survenir (hémorragie par exemple)<sup>60</sup>. Enfin, la biopsie est également associée à des erreurs d'échantillonnage. Pour les éviter, il serait essentiel de procéder au

prélèvement d'au moins deux échantillons d'une longueur et d'une quantité de tissu suffisants<sup>61</sup>. En effet, la biopsie est couramment réalisée à 15mm de profondeur ce qui comporte souvent des erreurs de jugement sur l'avancement de la fibrose hépatique du patient, en conséquence de quoi une biopsie de 25 mm semblerait préférable pour réaliser un examen<sup>62</sup>. De plus, une certaine variabilité a été observée en fonction du niveau d'expertise de la personne qui réalise et analyse la biopsie<sup>63</sup>. Par ailleurs, avec une MASLD touchant 30% de la population mondiale<sup>34</sup>, la biopsie ne conviendrait pas pour être l'examen utilisé pour un dépistage massif des atteintes hépatiques. La découverte de nouveaux marqueurs non-invasifs et pertinents est donc nécessaire.

Il existe des tests visant à évaluer l'état hépatique d'un patient sans nécessité d'une intervention invasive telle que la biopsie hépatique<sup>64</sup>. Pour estimer la performance de ces tests, il convient d'utiliser l'aire sous la courbe 'Receiver Operator Characteristic' (AUC-ROC ou AUROC) qui utilise les données sur la sensibilité et la spécificité du test à différents seuils pour donner un score de performance. La performance maximale (AUROC=1) est obtenue lorsque le test non-invasif arrive à déceler une atteinte hépatique aussi bien que la biopsie. A l'heure actuelle et malgré des performances en constante amélioration, aucun test non-invasif n'a atteint une AUROC de 1. Ils peuvent cependant être utilisés comme examens préliminaires pour avoir un premier avis sur l'état du foie du patient et dans certains contextes, ils peuvent donner des informations complémentaires à la biopsie. Ainsi, des examens tels que l'imagerie par résonance magnétique (IRM), l'échographie, l'élastographie ou la tomodensitométrie sont utilisés pour diagnostiquer la MASLD<sup>58</sup> (Tableau 1). Ces méthodes exploitent le fait que les propriétés élastiques du tissu impliqué ou ses proportions en eau se voient modifiées lors du processus pathologique. Ces examens présentent l'avantage d'être non-invasifs et peuvent être répétés sur une période de temps plus courte que la biopsie hépatique. Cependant, l'échographie permet de détecter de façon précise et fiable uniquement les concentrations hépatiques de gras modérées ou sévères ( $\geq 20\%$  d'infiltration graisseuse)<sup>65</sup>. Il est aussi possible de tenter de diagnostiquer des atteintes hépatiques par l'évaluation de l'abondance de biomarqueurs sériques, mais les performances sont toujours inférieures à une AUROC de 1. En conséquence, les atteintes MASLD sont difficiles à discriminer par des approches non-invasives et la biopsie reste pour l'heure incontournable pour les valider. À noter que le diagnostic et la stratification de la MASLD restent d'autant plus ardues que certains patients peuvent être asymptomatiques et présenter des concentrations en enzymes hépatiques normales<sup>66</sup>. Il existe également des cas de MASLD « maigre » où les patients ont une atteinte hépatique métabolique sans être en surpoids. Cette forme d'atteinte est estimée à 10-20% des cas de MASLD<sup>67</sup>.

Tableau 1 : Avantages et inconvénients des différentes méthodes de diagnostic de la MASLD <sup>58,61,65,68-72</sup>.

Test	Avantages	Inconvénients
Biopsie hépatique	Précision dans le diagnostic : stratification fiable de la stéatose, l'activité et la fibrose	Erreurs d'échantillonnage (1/50000 <sup>e</sup> du foie prélevé) Variabilité inter/intra-observateur Risques / complications (douleur, hémorragies, mortalité) Coûts importants (hospitalisation, instruments, mobilisation de plusieurs experts) Dépistage massif difficile Suivi limité dans le temps
Echographie	Non-invasif Simple Prix du dispositif abordable	Faible sensibilité sur des stades précoces Performance limitée sur patients présentant une obésité très importante Stratification difficile Dépistage massif difficile
Tomographie	Peu invasif Capable de mesurer le gras viscéral Bonne performance pour détection de quantités importantes de stéatose	Coût du dispositif élevé Radiation Faible sensibilité Stratification difficile Dépistage massif difficile
IRM	Non- invasif Précision très élevée pour la stéatose Bonne différenciation entre MASH et stéatose simple	Coût du dispositif élevé Stratification difficile Dépistage massif difficile
Elastographie	Non-invasif Dépistage massif possible Performance acceptable pour stades de fibrose très précoces ou avancés (F0 ou F4)	Difficile à réaliser sur patients obèses Stratification difficile
Biomarqueurs sériques	Peu invasifs Peu coûteux Tests réalisés en routine (rapides) Dépistage massif possible	Peu sensibles et spécifiques (surtout sur des stades précoces) Stratification difficile

Afin de juger de la sévérité de l'atteinte, plusieurs scores composites reposant sur l'établissement des différents stades de stéatose, d'activité et/ou de fibrose ont été établis<sup>73</sup>. Un des scores les plus répandus, largement recommandé par l'AFE (Association Française pour l'Étude du Foie) et l'EASL (European Association for the Study of the Liver), est le score SAF (pour Stéatose, Activité, Fibrose), permettant de positionner sur une échelle de 0 à 3 ou 4, en fonction du paramètre évalué, l'état du foie du patient à partir d'observations réalisées sur la coupe histologique issue de la biopsie hépatique <sup>74</sup> (Figure 5).



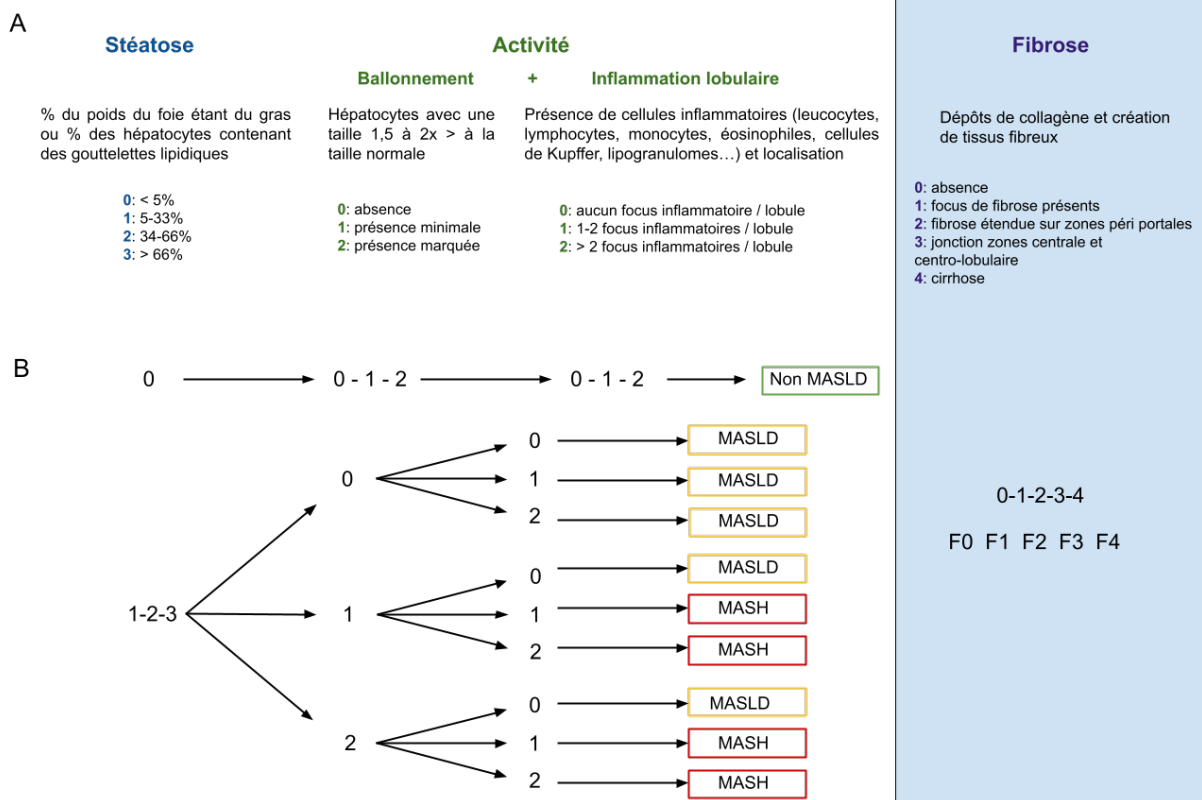


Figure 5 : Détails du calcul du score SAF en fonction des observations histologiques<sup>75-77</sup>. L'état MASLD ou MASH n'est diagnostiqué qu'en tenant compte des critères de stéatose et activité (inflammation lobulaire + ballonnement des hépatocytes), la fibrose est un paramètre indépendant mais qui est la plupart du temps relié aux atteintes hépatiques de tout type.

La fibrose est le seul paramètre qui permet de déterminer la morbidité liée à une atteinte hépatique de type MASLD<sup>78,79</sup>. De plus, en fonction du stade de fibrose, la prise en charge des patients sera adaptée, en effet pour les stades avancés (F3-F4) des traitements antifibrotiques<sup>80</sup> ou une transplantation hépatique peuvent être envisagés<sup>52,53</sup>. Comme cela a été vu dans le Tableau 1, la biopsie hépatique présente de nombreux risques et désavantages, elle n'est donc réalisée que lorsqu'il y a une forte suspicion de fibrose avancée.

L'évaluation de l'état du foie du patient s'effectue sur la base de l'exploration clinique, histologique et biologique, à la lumière des mécanismes pathologiques sous-jacents. La fibrose étant l'un des paramètres les plus importants à déterminer lors d'une atteinte MASLD, elle est devenue un point central dans la recherche de marqueurs non-invasifs. A ce jour, plusieurs marqueurs composites ou panels ont été développés dans l'objectif de prédire la fibrose. Le FibroTest largement utilisé, établit un score à partir de la concentration sérique de plusieurs variables que sont : l'alpha-2-macroglobuline (A2M), l'haptoglobine (Hp), l'apolipoprotéine A1 (ApoA1), la bilirubine et la gamma glutamyltransférase (GGT) en tenant compte de l'âge et du sexe du patient<sup>81</sup>. D'autres panels de ce type existent comme le NAFLD Fibrosis score<sup>82</sup>, l'Hepamet<sup>83</sup>, l'ADAPT<sup>84</sup> ou le Fib-4<sup>85</sup>. Le Tableau en Annexe A répertorie les principaux panels classés en fonction de leur performance. Cependant, tous ces tests n'atteignent pas la performance requise pour remplacer la biopsie hépatique<sup>86</sup>. Il existe donc un réel besoin pour les cliniciens de disposer d'un test non-invasif qui caractériserait correctement la fibrose hépatique.

Les atteintes hépatiques métaboliques perturbent le foie à différents niveaux, avec un potentiel impact fonctionnel : dégradation de la MEC, inflammation, apoptose, résistance à l'insuline, stress

oxydatif et troubles de la coagulation. Chacun de ces processus pathogéniques met en jeu des acteurs moléculaires et en particulier des protéines par le biais de modifications de répertoires protéiques, de variations d'abondances protéiques ou de modifications post traductionnelles des protéines (chimiques, enzymatiques...). Si ces changements protéiques se retrouvent dans le sang (fraction plasmatique) suite à un relargage ou une sécrétion hépatique, alors ils peuvent constituer de potentiels biomarqueurs ou indicateurs de processus pathogéniques variés et à différents stades. Ainsi, les biomarqueurs peuvent être prédictifs s'ils sont retrouvés lors d'atteintes précoces ou stratifiants si l'abondance du biomarqueur augmente en fonction du degré d'atteinte du patient. D'un côté, les biomarqueurs prédictifs ont pour objectif de prévoir une atteinte plus avancée pour initier une prise en charge du patient précoce, alors que les biomarqueurs stratifiants permettent de savoir exactement à quel stade de la maladie se situe le patient.

Le Tableau en Annexe B, recense les biomarqueurs protéiques potentiels ou avérés de MASLD et/ou de MASH identifiés à ce jour. Une vue générale des fonctions y étant associées est donnée dans la Figure 6. Les différents types de biomarqueurs protéiques ont été analysés en fonction de leur AUROC puis de leur capacité à être prédictifs, stratifiants, sensibles et spécifiques par rapport à la pathologie.

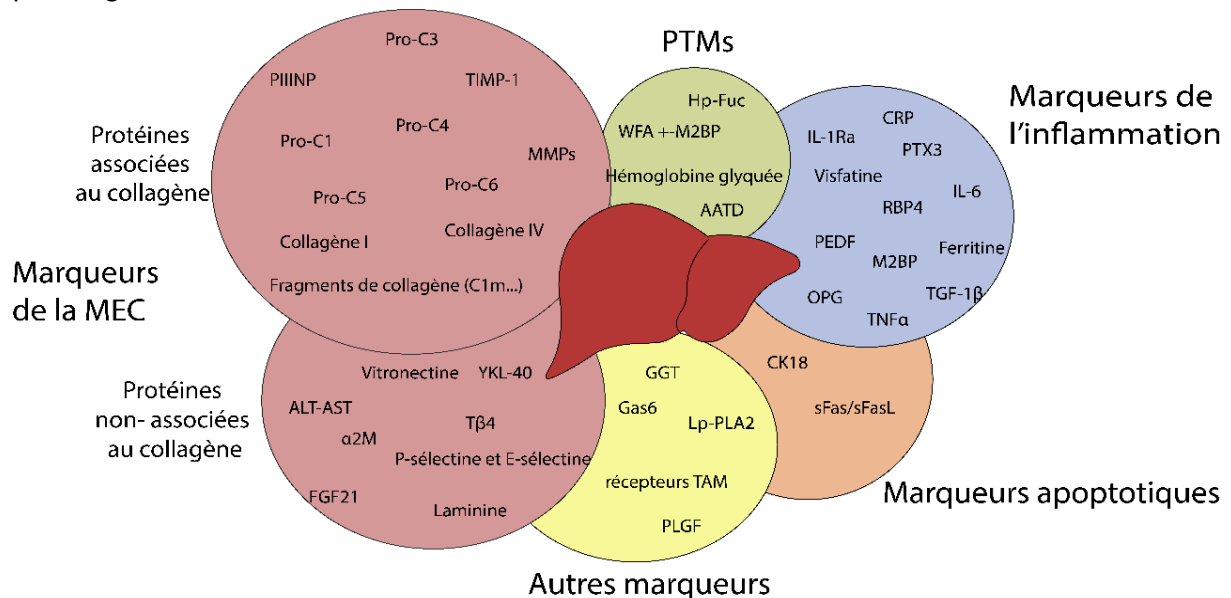


Figure 6 : Schéma des marqueurs de MASLD présents dans la littérature scientifique en fonction de leurs caractéristiques biologiques (MEC : Matrice extracellulaire ; PTMs : Modifications Post-Traductionnelles).

Beaucoup de protéines sont liées aux différents processus physiopathologiques impliqués dans la progression de la MASLD. Une manière d'appréhender les changements d'abondance liés à la progression de la pathologie requiert de mettre en œuvre des méthodologies de protéomique quantitative par exemple.

#### 4. La protéomique :

La protéomique est la discipline chargée de l'exploration de l'ensemble des protéines, leurs isoformes, leurs modifications et leurs interactions<sup>87</sup>. L'analyse protéomique est basée sur l'utilisation de techniques extrêmement sensibles et pouvant fonctionner à très haut débit, en particulier la spectrométrie de masse (MS)<sup>87</sup>. Il est possible d'appliquer la protéomique dans le domaine clinique, notamment dans la recherche de biomarqueurs<sup>88</sup>. Les échantillons analysés en protéomique clinique peuvent être très divers, allant de fluides biologiques (sang, urines, salive...)<sup>89</sup> à des tissus solides (tumeurs, organes...)<sup>90</sup>. Les différents couplages et paramétrages liés à la spectrométrie de

masse permettront de réaliser différents types d'analyses. L'utilisation d'une chromatographie liquide couplée à la MS (HPLC-MS) est l'une des associations les plus classiques permettant une identification et quantification performantes de protéines dans un échantillon liquide. En général, la découverte de biomarqueurs en protéomique suit une stratégie dite de « l'entonnoir »<sup>88</sup>. Parmi l'ensemble des protéines quantifiées dans de petites cohortes, de potentiels biomarqueurs sont décelés puis vérifiés pour enfin être validés sur des cohortes de taille plus importantes par des analyses ciblées (Figure 7).

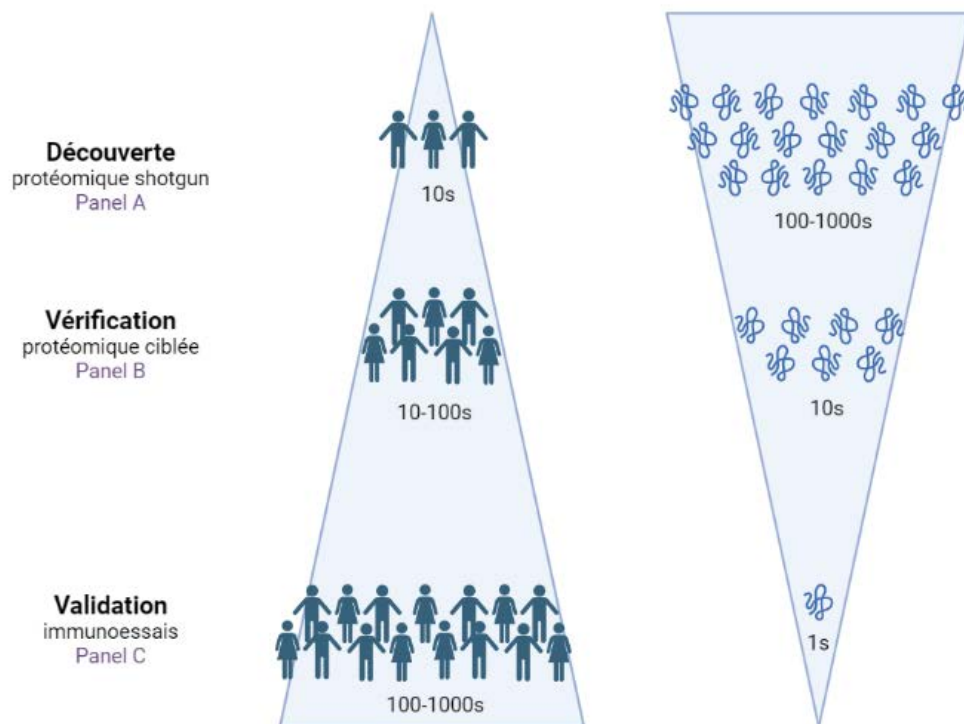


Figure 7 : Stratégie triangulaire ou de l'entonnoir pour la découverte, vérification et validation de biomarqueurs<sup>88</sup>.

Le foie étant un organe fortement vascularisé et produisant la plupart des protéines plasmatiques, la protéomique peut être un outil pertinent pour identifier les protéines différenciellement abondantes en fonction des processus physiopathologiques. La fibrose étant associée à la MASLD et impliquant de nombreuses protéines<sup>91</sup>, des études ont tenté de déceler des biomarqueurs protéiques plasmatiques spécifiques de ce processus<sup>92</sup>.

En fonction de la phase de l'étude, différents types d'études protéomique peuvent être réalisées. Comme montré dans la Figure 7, les études de découverte sont souvent réalisées en mode shotgun. Il s'agit d'une approche analytique utilisée pour étudier l'ensemble des protéines présentes dans un échantillon biologique de façon non-ciblée. Pour utiliser cette approche avec une méthode LC-MS, il faut « décomposer » l'échantillon biologique en peptides (fragments de protéines) à l'aide d'une enzyme, ces peptides seront ensuite analysés par spectrométrie de masse. De par son fonctionnement, le spectromètre de masse ne peut identifier et quantifier correctement que les protéines majoritaires. Cela peut être problématique dans des échantillons complexes comme le plasma qui est composé majoritairement par de l'albumine (55%), des globulines (38%) et du fibrinogène (7%) mais dont le nombre total de protéines serait supérieur à 10500 d'après la base de données « Plasma Proteome Database » compilant les résultats de plus de 500 études<sup>93</sup>. La « profondeur » d'analyse du spectromètre de masse en mode shotgun, ne permet pas d'analyser les protéines les moins

abondantes de façon fiable. Cependant, des méthodes existent pour réaliser des analyses plus exactes sur des concentrations faibles de protéines. En effet, des approches utilisant une déplétion ou « pêche » de protéines peuvent permettre de réaliser des analyses plus profondes ou sur des types de protéines souhaités. D'un autre côté, il est également possible de réaliser des analyses dites ciblées qui ne vont mesurer que les protéines sélectionnées préalablement. L'avantage de ces méthodes c'est qu'elles vont offrir de meilleurs résultats pour les protéines ciblées (sensibilité, précision et reproductibilité accrues, réduction du bruit de fond, ...). Ce type d'analyses sont surtout utilisées lors d'études plus poussées de vérification ou validation de biomarqueurs car elles ne permettent d'analyser que des protéines choisies. En général, les approches shotgun permettent d'identifier de potentiels biomarqueurs parmi les protéines majoritaires puis les approches ciblées seront utiles pour vérifier et valider ces biomarqueurs. Cette approche est utilisée pour limiter certains biais des analyses d'échantillons cliniques. En effet, les échantillons cliniques tels que le sang, l'urine ou les tissus peuvent présenter une variabilité biologique importante à cause de facteurs génétiques, environnementaux ou pathologiques, il est donc important d'avoir une vue d'ensemble avec une étude de découverte shotgun pour pouvoir réaliser une étude ciblée de qualité.

Dans le cadre de ma thèse, j'ai identifié de potentiels biomarqueurs plasmatiques de la fibrose grâce une étude de protéomique shotgun semi-quantitative réalisée sur une cohorte de patients MASLD (n=158). Parmi les protéines différentiellement abondantes identifiées à travers les stades de fibrose, une sélection a été réalisée basée sur des critères statistiques, techniques et bibliographiques. La variation d'abondance plasmatique d'une quinzaine de protéines candidates a été confirmée sur cette même cohorte grâce à deux techniques de quantification ciblées (utilisation de peptides marqués isotopiquement et test ELISA). Deux de ces protéines ont été validées sur une deuxième cohorte indépendante (n=200) par ELISA et ont donné lieu à l'élaboration d'un panel basé sur le FibroTest. Cette étude fera l'objet de la Partie I de ce manuscrit. En parallèle, j'ai pu travailler sur un projet visant à optimiser une technique de quantification absolue en HPLC-MS basée sur l'utilisation de protéines standard marquées isotopiquement (PSAQ). Ces standards pourraient améliorer la précision, la sensibilité et la spécificité des quantifications absolues par HPLC-MS. Les études de validation de biomarqueurs nécessitent une quantification absolue de qualité, cette approche a pour objectif de repousser les limites de ce domaine et sera détaillée dans la Partie II.

## Partie I :

### Découverte, vérification et validation de biomarqueurs dans le cadre de la MASLD :

Ce projet a pour objectif d'identifier et valider de nouveaux biomarqueurs plasmatiques de la fibrose chez des patients atteints de MASLD (Figure 8). Comme sommairement décrit en introduction, le plus important est de différencier les stades de fibrose avancée des stades précoces. La stratégie suivie pour mettre en évidence les protéines dont la concentration reflèterait le stade de fibrose du foie comporte plusieurs étapes. Tout d'abord, une étude protéomique de découverte reposant sur l'exploitation des données de protéomique semi-quantitative établie à partir d'une cohorte de 158 patients a permis d'identifier 72 protéines différemment exprimées. Parmi celles-ci, 15 protéines ont été choisies pour confirmer les données préalablement acquises par l'étude de découverte. Enfin deux protéines avec un fort potentiel pour être des biomarqueurs ont été testées en ELISA sur cette première cohorte puis validées par la même méthode sur une cohorte indépendante de 200 patients. Enfin, ces deux protéines ont été introduites dans un panel inspiré par le FibroTest afin d'obtenir une performance diagnostique plus élevée.

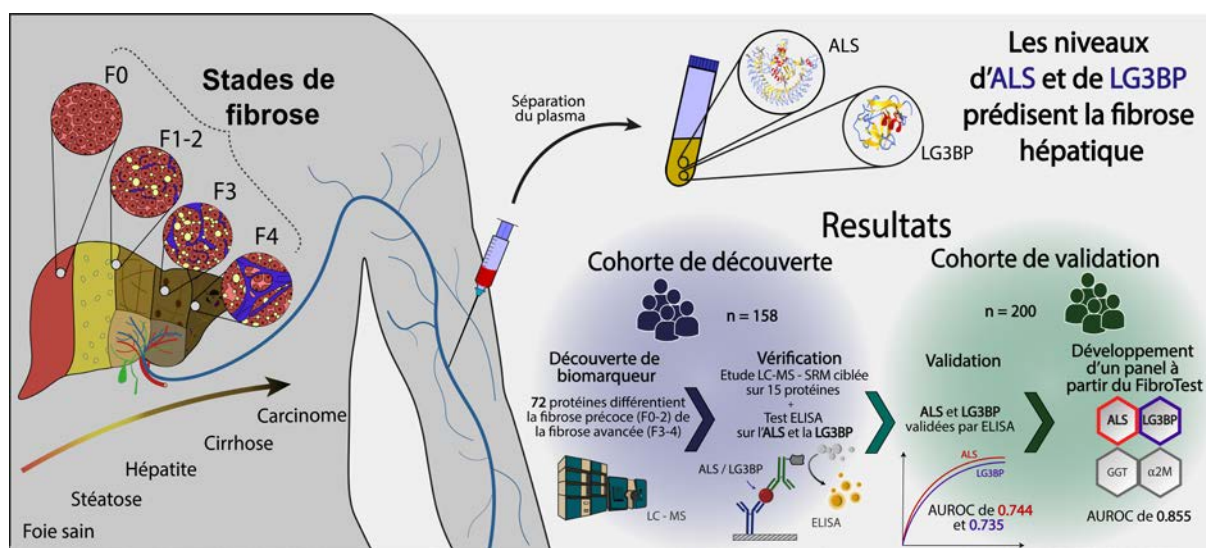


Figure 8 : Résumé graphique du projet MASLD.

#### A. Analyse des cohortes :

##### 1. Composition des cohortes :

Le projet MASLD inclut deux cohortes d'échantillons plasmatiques provenant de patients pour lesquels une atteinte hépatique MASLD a été établie par biopsie hépatique. Le but de ce projet est de trouver et valider des biomarqueurs protéiques qui permettraient de stratifier la fibrose chez ces patients.

La première cohorte est composée de 158 échantillons de plasma provenant du CHU de Grenoble, la deuxième cohorte est quant à elle composée de 200 échantillons de nature similaire provenant du CHU d'Angers. Les échantillons obtenus après centrifugation de l'échantillon de sang total ont été collectés entre 2014 et 2019 pour la cohorte grenobloise et entre 2013 et 2021 pour la cohorte angevine. Chaque échantillon est issu d'un prélèvement effectué dans les minutes/heures

suivant la biopsie dans des tubes traités avec de l'anticoagulant (EDTA). Celui-ci a ensuite été centrifugé entre 1000 et 2000 g pendant 10-15 minutes pour séparer les cellules sanguines et les plaquettes du plasma. Tous les patients des deux cohortes ont fait l'objet d'une biopsie et d'une étude approfondie des variables biologiques, métaboliques et histologiques de routine du service hépatologie (127 variables pour la cohorte de Grenoble et 55 variables pour celle d'Angers). En préambule, il faut noter que seules certaines variables ont été analysées dans le cadre de l'étude MASLD et résumées dans le Tableau 2. De surcroît, seuls les patients avec suspicion d'atteinte hépatique métabolique sont inclus dans la cohorte. Ainsi, pour éviter des biais, tous les patients ont été testés pour des atteintes hépatiques d'origine non-métabolique (virales, alcooliques, médicamenteuses...). Les critères d'inclusion varient légèrement entre les deux cohortes, notamment concernant la quantité d'alcool consommée, pour la cohorte de Grenoble, la limite est de 350 g d'alcool par semaine, alors que pour la cohorte d'Angers la limite est de 210 g par semaine. Les données biologiques des deux cohortes sont résumées dans le tableau ci-dessous (Tableau 2). Comme nous pouvons l'observer ces deux cohortes sont biologiquement très proches car la plupart des variables ne sont pas statistiquement différentielles entre les deux cohortes. Cependant certaines variables comme le tour de taille, l'indice de masse corporelle, le taux d'albumine ou le taux de cholestérol présentent des p-valeurs inférieures à 1%. Ces différences s'expliquent par le fait que les deux cohortes ont été établies de façon indépendante et malgré le fait que des critères stricts ont été choisis pour inclure les patients dans les cohortes, il existe une diversité importante.

*Tableau 2 : Caractéristiques cliniques, biologiques et histologiques des patients de la cohorte de Grenoble (n=160) et Angers (n=200). IMC calculé à partir de la formule poids/taille<sup>2</sup> (kg/m<sup>2</sup>). ALT et AST : alanine et aspartate aminotransférases ; GGT : gamma-glutamyl-transférase ; CRP : protéine C-réactive ; LDL et HDL : lipoprotéine de basse densité et de haute densité. L'activité est la somme du score histologique attribué au ballonnement hépatocellulaire et de celui de l'inflammation lobulaire. La variable MASH correspond aux patients avec au moins un score de 1 sur la stéatose et l'activité (ballonnement + inflammation lobulaire). Les p-valeurs ont été estimées par le test de Mann-Whitney.*

<b>Variab</b> les	<b>Grenoble</b> <b>158</b>	<b>Angers</b> <b>200</b>	<b>p-valeurs</b>
Age (ans)	55 ± 12	57 ± 12	0,029
Poids (kg)	89 ± 18	94 ± 19	0,023
Taille (cm)	169 ± 10	167 ± 10	0,231
Tour de taille (cm)	108 ± 15	113 ± 14	0,0054
Indice de masse corporelle (IMC)(Kg/m <sup>2</sup> )	30,9 ± 5,5	33,4 ± 6,1	0,00014
<b>Caractéristiques biologiques et cliniques</b>			
Haptoglobine (g/L)	1,25 ± 0,61	1,24 ± 0,55	0,859
Alpha2m (g/L)	2,25 ± 0,87	2,06 ± 0,82	0,027
ApoA1 (g/L)	1,46 ± 0,25	1,39 ± 0,21	0,029
Albumine (g/L)	44,8 ± 4,0	42,4 ± 3,7	9,00E-10
CRP (mg/L)	4,3 ± 5,5	4,3 ± 4,9	0,677
ALT UI/L	68,4 ± 34,3	63,7 ± 41,9	0,030
AST UI/L	41,2 ± 21,6	45,5 ± 26,3	0,099
GGT UI/L	153 ± 187	117 ± 170	0,186
Cholestérol total (g/L)	1,96 ± 0,54	1,86 ± 0,45	0,061
Cholestérol HDL (g/L)	0,53 ± 0,47	0,43 ± 0,11	0,0076
Cholestérol LDL (g/L)	1,16 ± 0,43	1,20 ± 0,42	0,328
Triglycérides (g/L)	1,77 ± 1,22	1,70 ± 1,12	0,976

Urée (g/L)	5,4 ± 1,5	5,6 ± 2,0	0,603
Bilirubine (µmol/L)	11,2 ± 6,1	12,8 ± 7,2	0,028
Caractéristiques histologiques			
Grade de stéatose, n (%)			
0	10 (6.3%)	15 (7.5%)	
1	31 (19.6%)	91 (45.5%)	
2	54 (34.2%)	63 (31.5%)	
3	63 (39.9%)	31 (15.5%)	
Activité			
Ballonnement hépatocellulaire, n (%)			
0	29 (18.3%)	37 (18.5%)	
1	72 (45.6%)	83 (41.5%)	
2	57 (36.1%)	80 (40.0%)	
Inflammation lobulaire, n (%)			
0	49 (31.0%)	37 (18.5%)	
1	84 (53.2%)	136 (67.5%)	
2	25 (15.8%)	27 (13.5%)	
Fibrose, n (%)			
0	24 (15.2 %)	22 (11.0%)	
1	30 (19.0%)	49 (24.5%)	
2	46 (29.1%)	54 (27.0%)	
3	33 (20.9%)	61 (30.5%)	
4	25 (15.8%)	14 (7.0%)	
MASH, n (%)	102 (64.6%)	148 (74.0%)	

## 2. Comparaison des deux cohortes :

Le Tableau 3 montre que les cohortes de Grenoble et Angers ont un nombre de femmes et hommes atteints de MASLD similaire à celui des populations française, européenne et mondiale. De même, l'âge moyen de ces cohortes est assez similaire, même s'il est légèrement supérieur à l'âge moyen de la population MASLD générale. Les données sur l'IMC et le tour de taille chez les patients MASLD étant difficilement accessibles sur les populations française, européenne et mondiale, seulement les chiffres des cohortes avec ceux des populations non-MASLD ont été comparés. L'IMC et le tour de taille des patients des cohortes sont supérieurs à ceux de la population générale. Cette observation était attendue car la MASLD touche surtout des populations en surpoids<sup>94</sup>. Une comparaison en fonction du groupe ethnique aurait été intéressante car des études montrent que la prévalence de MASLD est haute chez les Hispaniques, modérée chez les Blancs et faible chez les Noirs<sup>95</sup> mais ces données n'étaient pas disponibles pour les cohortes analysées. Enfin, même si les données des deux cohortes semblent représentatives des populations française, européenne et mondiale, il n'est pas possible de conclure sur ce point car nous n'étions pas en mesure de faire des tests statistiques sur ces trois populations en raison de l'indisponibilité des données à partir desquelles ont été établies les moyennes.

Tableau 3 : Données générales (nombre, âge, sexe, IMC, tour de taille) sur les cohortes de Grenoble et Angers ainsi que sur les populations française, européenne et mondiale MASLD et non-MASLD <sup>96;97;98;99;100;101;102;103;104;105;106;107</sup>(articles lus le 03/2023).

		Grenoble (moyenne)	Angers (moyenne)	France (moyenne)		Union européenne (moyenne)		Monde (moyenne)	
				Population générale <sup>96</sup>	Population NAFLD (% de la population totale française - % parmi la population NAFLD)	Population générale	Population NAFLD	Population générale	Population NAFLD
Nombre	femmes	64 (40,0%)	65 (32,5%)	48,3	~4 millions <sup>95</sup> (6% - 33%)	~228 millions (51%)	Prévalence plus importante chez les hommes <sup>102</sup>	49,6% <sup>104</sup>	Prévalence plus importante chez les hommes
	hommes	96 (60,0%)	135 (67,5%)	51,7	~8 millions <sup>95</sup> (12% - 66%)	~219 millions (49%)		50,4% <sup>104</sup>	
	total	160	200	~ 67 millions	~12 millions (18,2% <sup>95</sup> )	~447 millions <sup>97</sup>	20-30% <sup>101</sup>	~8 milliards <sup>103</sup>	~2,5 milliards (30%) <sup>34</sup>
Age	femmes	56	55,6	43,6	Pas de données - prévalence augmente avec l'âge sauf à partir de 80ans	45,5 <sup>98</sup>	Pas de données	31,1 <sup>105</sup>	Pas de données
	hommes	53,4	59,7	40,8		42,6 <sup>98</sup>		29,6 <sup>105</sup>	
	total	54,4	57,0	42,2		44,0 <sup>99</sup>		[44-53] en fonction de l'étude <sup>102</sup>	
IMC	femmes	30,5	33,1	23,6	Pas de données	Pas de données - 52,7% de surpoids <sup>100</sup>	Pas de données	25 <sup>106</sup>	Pas de données
	hommes	31,2	33,9	24,9				24,5 <sup>106</sup>	
	total	30,9	33,3	24,3				24,8 <sup>106</sup>	
Tour de taille	femmes	104,8	113,0	88,7	Pas de données	Pas de données	Pas de données	Pas de données	Pas de données
	hommes	110,7	112,6	95,6					
	total	108,0	112,9	92,3					

La Figure 9 montre que la répartition des groupes de fibrose, sexe et âge entre les deux cohortes semble similaire. Cependant, nous constatons dans les deux groupes une proportion plus élevée de patients présentant des niveaux intermédiaires de fibrose (F1-F2-F3) par rapport à la population générale. En effet, à mesure que le stade de fibrose progresse, le nombre de patients atteints diminue<sup>108</sup>. Cette sous-représentation des stades précoces pourrait s'expliquer par le fait que seuls les patients suspectés de fibrose avancée ont subi une biopsie, et que les stades précoces sont souvent asymptomatiques donc peu diagnostiqués. Cependant, des patients sans fibrose (F0) ou avec des stades précoces sont quand même inclus dans l'étude. Dans ces situations, la suspicion de fibrose avancée pourrait découler d'autres types de lésions hépatiques détectées par des méthodes diagnostiques distinctes de la biopsie.

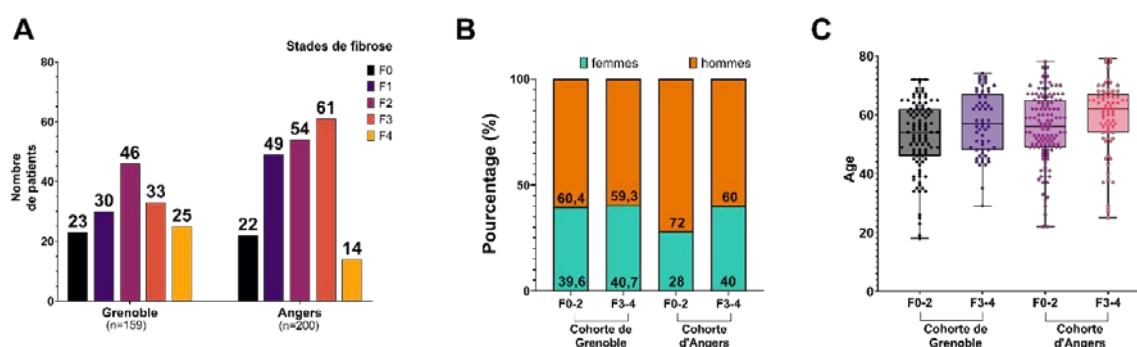


Figure 9 : Statistiques descriptives des cohortes de Grenoble et Angers. A. Nombre de patients par stade de fibrose dans les cohortes de Grenoble et d'Angers. B. Répartition par sexes entre fibrose précoce et avancée pour les cohortes de Grenoble et Angers. C. Distribution de l'âge entre fibrose précoce et avancée pour les cohortes de Grenoble et Angers (p-valeurs Mann-Whitney : (1) F0-2 vs F3-4 Grenoble : 0.0535 ; (2) F0-2 vs F3-4 Angers : 0.0222 ; (3) Grenoble vs Angers : 0.0293).

### 3. Corrélation de variables :

L'objectif de cette partie est de vérifier si parmi les données fournies par les hôpitaux, il existe des variables qui sont corrélées entre elles. Cela permet d'approfondir les connaissances sur les données tout en vérifiant si les marqueurs utilisés en routine clinique sont pertinents pour diagnostiquer les pathologies visées.





Tableau 4 : Tableau regroupant les variables les plus fortement corrélées négativement et positivement entre les cohortes de Grenoble et d'Angers.

Corrélations négatives	Variables corrélées	Corrélations positives	Variables corrélées
-1,0	∅	1,0	∅
< - 0,9	INR / TP	> 0,9	Activité / ballonnement hépatocytes NAS / activité LDL / cholestérol
< - 0,8	∅	> 0,8	MASH calculé / activité NAS / ballonnement hépatocytes NAS / stéatose Activité / inflammation lobulaire FibroTest calculé / alpha2m HOMA / insuline Hba1c / glucose
< - 0,7	∅	> 0,7	MASH calculé / NAS MASH calculé / inflammation lobulaire NAS / inflammation lobulaire
< - 0,6	∅	> 0,6	MASH calculé / ballonnement hépatocytes
< - 0,5	TP / FibroTest calculé	> 0,5	Fibrose / ballonnement hépatocytes Fibrose / activité Fibrose / inflammation portale Fibrose / NAS FibroTest calculé / INR

#### 4. Stratégie d'analyse des échantillons :

Les échantillons des deux cohortes n'étant pas arrivés en même temps au laboratoire, ont été traités selon la séquence décrite dans l'explication suivante.

Comme mentionné dans l'introduction (partie E), la stratégie que nous avons décidé de mettre en œuvre est de type triangulaire (Figure 7). Dans un premier temps, une étude de découverte a été réalisée sur la cohorte de Grenoble sur la totalité des patients afin d'identifier les protéines majoritaires du plasma. Dans un second temps, une analyse différentielle de l'abondance protéique en fonction des stades de fibrose a été menée. Puis une étude de vérification de quelques candidats a été diligentée sur la base d'une méthode SRM (Selection reaction monitoring) puis par une méthode ELISA sur des protéines triées en fonction de plusieurs critères. Finalement, une étude de validation de deux biomarqueurs particulièrement prometteurs a été réalisée sur la cohorte d'Angers par technique ELISA. Pour conclure les résultats de ces biomarqueurs ont été introduits dans un nouveau panel (multidonnées) basé sur le FibroTest (Figure 11).

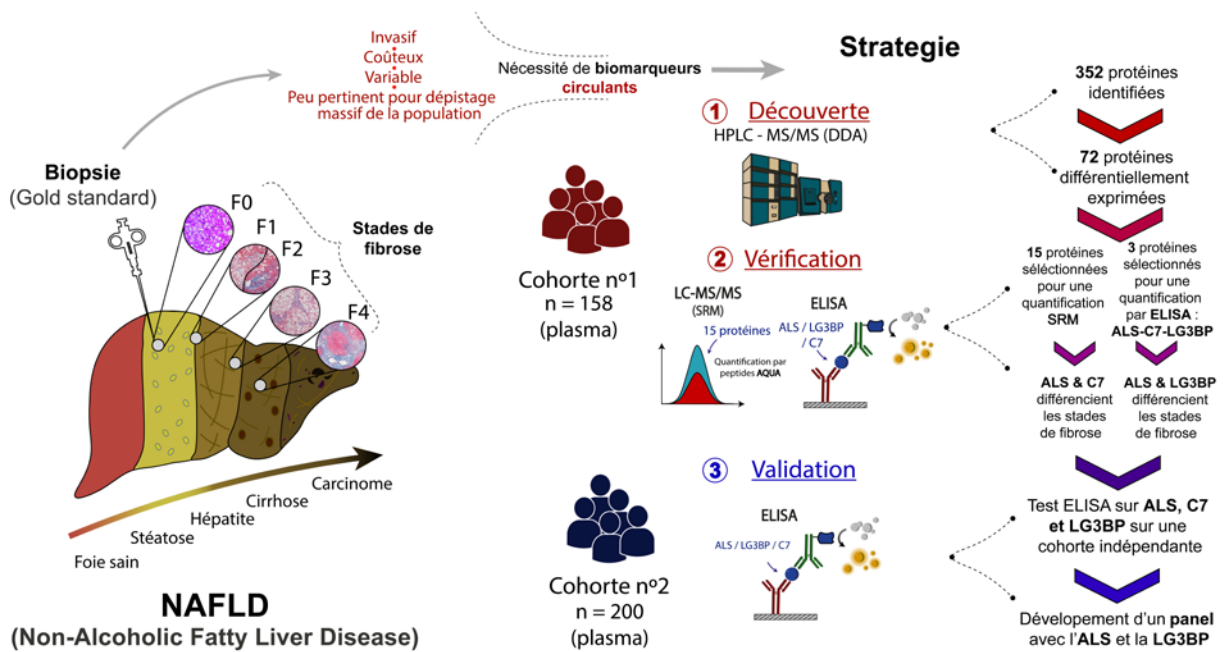


Figure 11 : Conception de l'étude et utilisation des cohortes : vue d'ensemble de la conception de l'étude et des différentes étapes de filtrage dans la sélection des protéines à utiliser comme marqueurs potentiels de la pathologie.

## B. Etude découverte :

### 1. Stratégie de préparation des échantillons :

Les échantillons de plasma ont été répartis sur 3 plaques de 96 puits conformément à l'ordre fourni par l'outil informatique Well plate maker (WPM) permettant de limiter les effets de plaque<sup>112</sup>. Pour faire cela, les patients ont été distribués en fonction de leur score de fibrose (F0-F4) de façon à ce que deux patients appartenant au même groupe ne se retrouvent pas côte à côte verticalement ou horizontalement (contrainte N-S-E-O)(Figure 12).

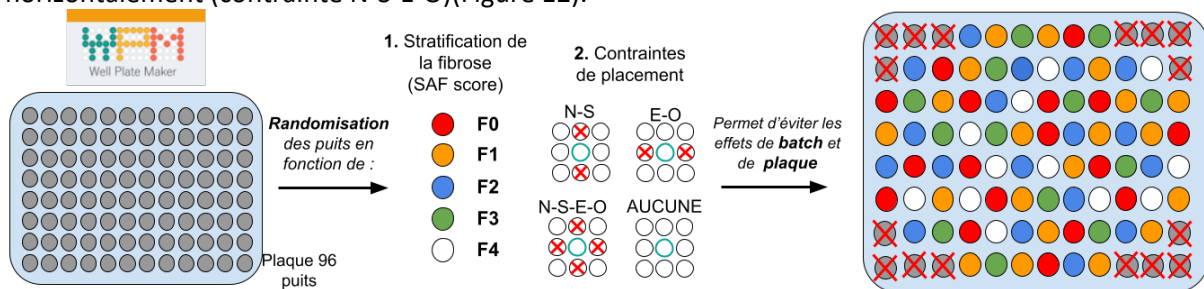


Figure 12 : Schéma de la répartition des échantillons à partir de WPM<sup>97</sup>. N,S,E,O représentent les contraintes de placement nord, sud, est, ouest, donc deux échantillons avec les mêmes stades de fibrose ne peuvent se retrouver côte à côte verticalement ou horizontalement sous ces contraintes.

Dans chaque puits, 3  $\mu$ l de plasma ont été déposés afin d'y être dénaturés, réduits et alkylés en utilisant de la guanidine hydrochloride (8M), du TCEP (tris(2-carboxyethyl)phosphine) (100mM) et du chloroacétamide (500mM) dilués dans du bicarbonate d'ammonium (500mM). Une incubation à 95° C pendant 5 minutes a par la suite été réalisée. Les échantillons ont ensuite été digérés en deux étapes grâce à un mix de 3  $\mu$ g de trypsine-LysC pendant une heure à 37° C puis par un autre ajout de 3  $\mu$ g de trypsine-LysC pour une digestion overnight. Pour finir les digestats ont été dessalés sur une colonne StrataX Microelution 96-Well SPE Plate, aliquotés et séchés pour une conservation optimale à -80° C avant analyse.

Dans notre approche, des analyse LC-MS/MS en mode Data dependent acquisition (DDA) et Data independent acquisition (DIA) ont été réalisées par la plateforme du laboratoire EDyP (description en

Annexe C). Les premières analyses ont été réalisées essentiellement sur les données DDA en amont de mon arrivée au laboratoire dans le cadre d'une thèse en bio-informatique<sup>113</sup>. Les données DDA ont été les seules à être traitées car l'examen des données DIA nécessiterait un ajustement de la méthodologie, tant dans les phases de prétraitement que dans l'analyse statistique. Ceci aurait nécessité une perspective et des compétences différentes pour évaluer adéquatement la pertinence de cette démarche. Ainsi, dans la continuité du travail réalisé sur le projet de thèse précédent, nous avons analysé les données DDA pour la cohorte de Grenoble.

## 2. Détails sur les analyses MS et le process des données :

Un aliquot de chaque échantillon a été remis en suspension dans 170µl de solution à 4% d'acétonitrile et 0,1% d'acide trifluoroacétique contenant des standards HRM-IRT. De ce mélange, 1,5µl (environ 500ng de matériel) ont été analysés par nanoLC-MS/MS : Ultimate 3000 RSLCnano et Q-Exactive HF respectivement, tous deux de Thermo Fisher Scientific. Les peptides ont été injectés sur une précolonne PepMap C18 de 300 µm x 5 mm (Thermo Fisher Scientific) et séparés sur une colonne C18 de 75 µm 250 mm (Reprosil-Pur 120 C18-AQ, 1,9 µm, Dr. Maisch HPLC GmbH). La méthode nano-LC consistait en un gradient multilinéaire de 60 minutes allant de 5 à 42 % d'acétonitrile dans 0,1 % d'acide formique à un débit de 300 nL/min. Le voltage du spray était fixé à 2 kV et la température du capillaire était réglée à 270 °C. Des spectres de surveillance en full-scan MS ( $m/z = 400-1600$ ) ont été acquis à une résolution de 60 000 ( $m/z 200$ ) après accumulation de  $3 \times 10^6$  ions (temps d'injection maximal de 55 ms), avec masse bloquée activée. Les 15 ions les plus intenses ont été fragmentés par dissociation collisionnelle à haute énergie (HCD) après accumulation de  $10^5$  ions (temps d'injection maximal : 120 ms). La résolution MS/MS a été fixée à 30 000 ( $m/z 200$ ). Les données MS et MS/MS ont été acquises en mode DDA à l'aide du logiciel Q Exactive HF Tune, v. 2.8.1 (Thermo Scientific).

Les données ont été traitées de façon automatique en utilisant le logiciel Mascot Distiller (version 2.7.1.0, Matrix Science). Les peptides et protéines ont été identifiées en utilisant le logiciel Mascot (version 2.6) à travers de recherches concomitantes à partir de Swiss-Prot (taxonomie Homo sapiens, téléchargée en Novembre 2019), une base de données de contaminants classiques (faite maison) et leur base de données correspondantes inversées. La trypsine était l'enzyme choisie dans le logiciel et un maximum de deux miss cleavages étaient autorisés. Les tolérances d'erreur de masse du précurseur et du fragment ont été fixées à 10 ppm et 25 mmu, respectivement. Les modifications peptidiques autorisées au cours de la recherche étaient les suivantes : (1) carbamidométhylation de la cystéine (fixe) ; (2) acétylation de l'extrémité N-terminale de la protéine (variable), et (3) oxydation de la méthionine (variable). Le logiciel Proline (version 2.1) a été utilisé pour combiner les données de tous les patients. Après combinaison, les résultats ont été filtrés en conservant les correspondances peptides-spectres de rang 1 avec une longueur minimale de 7 acides aminés et un score Mascot minimal de 25. Avec ces paramètres, le taux de fausse découverte (FDR) pour les identifications de correspondances peptide-spectre était inférieur à 1% selon l'approche target-decoy. Un minimum de deux peptides était nécessaire pour identifier le « protein group ». Proline a ensuite été utilisé pour effectuer une quantification label-free par MS<sub>1</sub> des groupes identifiés sur la base de l'abondance de leurs peptides spécifiques.

## 3. Résultats d'identification des protéines plasmatiques :

L'analyse de l'ensemble des résultats de DDA a donné lieu à l'identification de 352 protéines et à leur quantification sur la base des 5095 peptides identifiés par Mascot (Figure 13.B et C). Le nombre de protéines et de peptides identifiés ne varie pas de manière significative entre les stades de fibrose (p-valeurs 0,3017 et 0,1380 respectivement). En moyenne, on compte 14 peptides par protéine identifiée (Figure 13.A). Il y a 1258 peptides et 156 protéines identifiées de façon simultanée dans tous les échantillons, ce qui représente environ 25% et 44% du total de peptides et de protéines identifiées (Figure 13.D et E). Ces chiffres peuvent sembler relativement bas, cependant plusieurs raisons

pourraient les expliquer. Tout d'abord et de façon générale, la LC-MS donnera des signaux plus fiables sur les protéines les plus abondantes. Ainsi, l'identification de protéines dont la concentration plasmatique est très faible peut s'avérer difficile, donnant lieu à des résultats variables. De plus, la cohorte analysée présente des patients avec des degrés d'atteinte hépatique très disparates impliquant que différentes protéines se retrouvent dans les limites de détection et donc peu de protéines/peptides sont identifiées dans l'ensemble des échantillons.

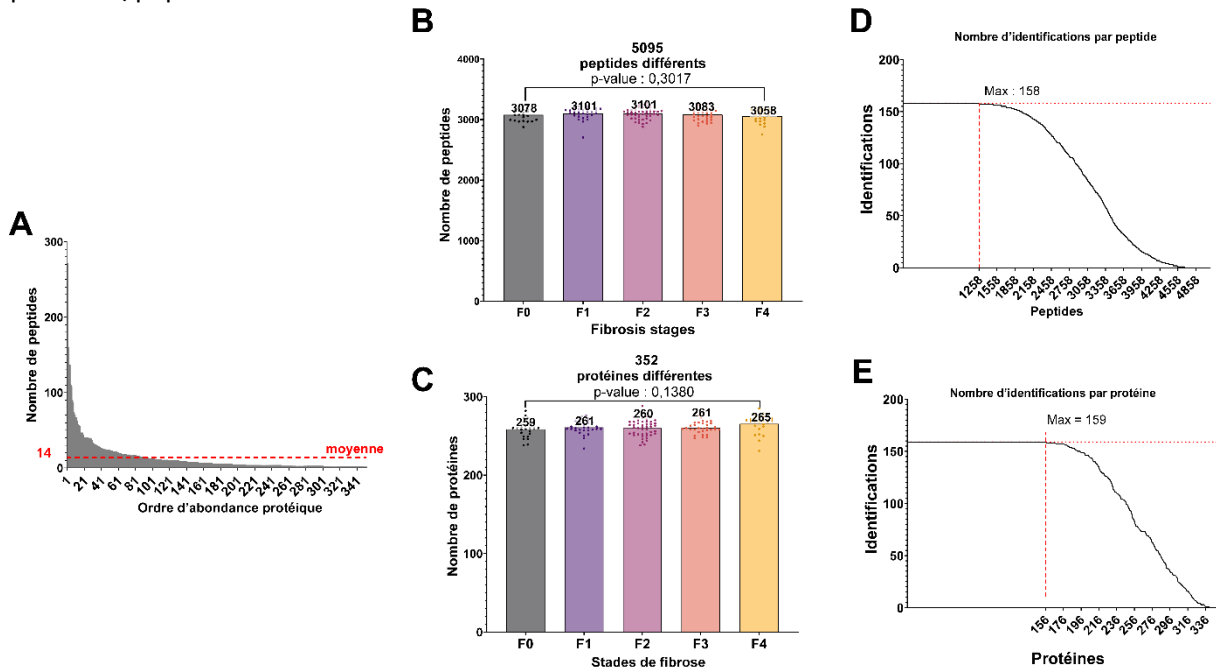


Figure 13 : Résultats de protéomique issus de l'étude de découverte sur la cohorte de Grenoble. A. Nombre de patients identifiés par protéine en mode DDA sur la cohorte de découverte. B. Nombre de peptides moyens identifiés par stade de fibrose après analyse LC-MS/MS en mode DDA sur la cohorte de découverte (les points représentent le nombre de peptides pour chaque patient). Test ANOVA utilisé pour calculer la p-valeur. C. Nombre de protéines moyen identifiées par stade de fibrose après analyse LC-MS/MS en mode DDA sur la cohorte de Grenoble (les points représentent le nombre de protéines pour chaque patient). Test ANOVA utilisé pour calculer la p-valeur. D. Nombre d'identifications par peptide. E. Nombre d'identifications par protéine.

### 3. Résultats de l'étude semi-quantitative par DDA :

Le traitement statistique de la cohorte de Grenoble a été réalisé en utilisant la suite logicielle Prostar<sup>114</sup>. Seules les protéines identifiées sur au moins 70% des patients ont été conservées pour des raisons de qualité de la quantification et de généralisation des identifications au travers des patients. Les abondances relatives de ces protéines ont été transformées par un  $\text{Log}_2$  afin de calculer des ratio d'abondance ou « fold changes » (FC) plus facilement comparables entre valeurs croissantes et décroissantes. Ensuite les abondances ont été normalisées par la méthode « normalisation par stabilisation de la variance » ou vsn (variance stabilization normalization)<sup>115</sup>. Cette méthode peut être utile soit pour simplifier les considérations dans l'analyse graphique des données exploratoires, soit pour permettre l'application de techniques basées sur la régression ou l'analyse de la variance. Ici, cette méthode consiste à transformer les intensités mesurées afin de limiter la dépendance entre variance et moyenne des intensités<sup>116</sup>. La méthode vsn a été choisie car il s'agit d'une approche robuste même lorsqu'il y a une proportion non négligeable de valeurs manquantes (comme c'est le cas en protéomique label-free) tout en conservant la variabilité biologique entre les groupes d'échantillons<sup>117</sup>. Les valeurs manquantes ont été imputées par la méthode des « structured least squares algorithm » car elle est adaptée pour analyser ce type de données<sup>118,119</sup>. Seulement deux modalités de fibrose ont été évaluées : fibrose précoce (F0-2) contre fibrose avancée (F3-4). Les tests statistiques choisis pour évaluer les deux modalités de fibrose sont celui de Limma<sup>120,121</sup> et le Fold Change sur base  $\text{Log}_2$  ( $\text{Log}_2$  FC). Le test de Limma est une adaptation d'un test de Student (t-test) estimant la variance sur l'ensemble des protéines. Il a été démontré que ce test apporte une sensibilité accrue lors de l'analyse

de données de protéomique en label-free par rapport à des méthodes plus classiques comme le t-test ou l'ANOVA, c'est pour cette raison que nous l'avons utilisé dans ce cas<sup>122</sup>.

Le Volcano plot, en Figure 14, résume les analyses réalisées avec le test de Limma. Celui-ci permet de visualiser les protéines différentiellement abondantes entre une fibrose précoce (F0-2) et une fibrose avancée (F3-4). Le seuil de significativité a été défini pour une p-valeur permettant d'atteindre un FDR inférieur à 5% suivant la procédure de Benjamini-Hochberg entre les deux états de fibrose<sup>123</sup>.

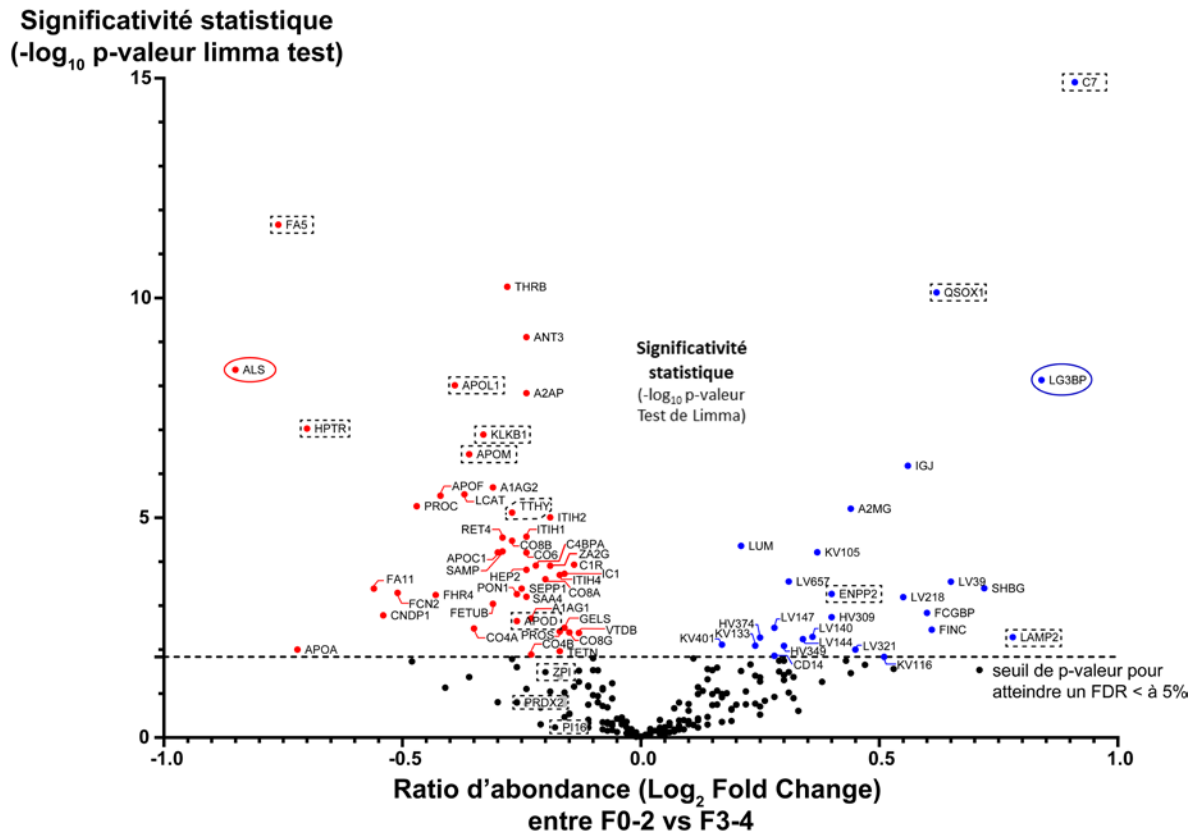


Figure 14 : Volcano plot réalisé à partir des données obtenues par LC-MS/MS en mode DDA sur la cohorte de découverte (cohorte de Grenoble). L'axe Y représente la performance statistique de chaque protéine pour différencier un stade de fibrose précoce d'un stade avancé calculé suivant la formule :  $-\log(p\text{-valeur})$ , la p-valeur a été obtenue avec le test de limma (t-test). Le seuil de significativité statistique a été établi pour une p-valeur de 0,05. L'axe des abscisses représente le  $\log_2$  FC entre les stades de fibrose précoce et avancée. Chaque point représente la valeur obtenue pour une protéine, les protéines présentant une diminution de leur concentration en fonction du stade de fibrose sont en rouge, les protéines dont la concentration augmente sont données en bleu. Les rectangles quadrillés indiquent les protéines utilisées dans l'étude de vérification par SRM et les protéines encerclées correspondent à celles validées par une méthode ELISA.

### C. Vérification ou évaluation de l'utilité clinique des cibles protéiques :

#### 1. Stratégie d'analyse de la cohorte en protéomique ciblée :

Une stratégie robuste de développement de biomarqueurs doit contenir les phases de découverte, vérification et validation<sup>124</sup>. L'étape de vérification, a pour objectif de vérifier que les potentiels biomarqueurs découverts par des méthodes à haut débit restent pertinents lorsqu'on utilise des méthodes plus ciblées. La vérification permet de montrer l'utilité clinique potentielle d'un biomarqueur plasmatique candidat pour justifier d'un investissement supplémentaire du ou des candidat(s) en vue d'études de validation clinique ceci du fait du coût élevé des études de validation. La vérification ou « accréditation du candidat » est souvent réalisée sur une cohorte indépendante contenant un nombre de patients plus élevé que pour la cohorte de découverte. La cohorte de Grenoble ayant déjà un nombre de patients important (n=158) et face à l'impossibilité d'obtenir une

cohorte indépendante supplémentaire, nous avons jugé qu'elle pouvait être utilisée pour l'étude de vérification. Contrairement à la stratégie de « l'entonnoir » présentée en introduction, nous nous situons plutôt sur une approche « rectangulaire » grâce à la cohorte indépendante de validation contenant 200 patients (Figure 15). La stratégie dite de l'entonnoir a une puissance statistique faible du fait du petit nombre d'échantillons de la phase de découverte la rendant plus sujette aux biais du groupe d'échantillons, tels que les variations pré-analytiques et biologiques, qui pourraient entraîner le passage de candidats erronés à la validation. L'avantage de ce type de méthode dite « rectangulaire » c'est de « mieux séparer les vrais biomarqueurs en minimisant les effets cohorte-spécifiques »<sup>88</sup>. Ainsi, les probabilités pour que les potentiels biomarqueurs de l'étude de découverte arrivent à l'étape de validation sont plus élevées<sup>124</sup>.

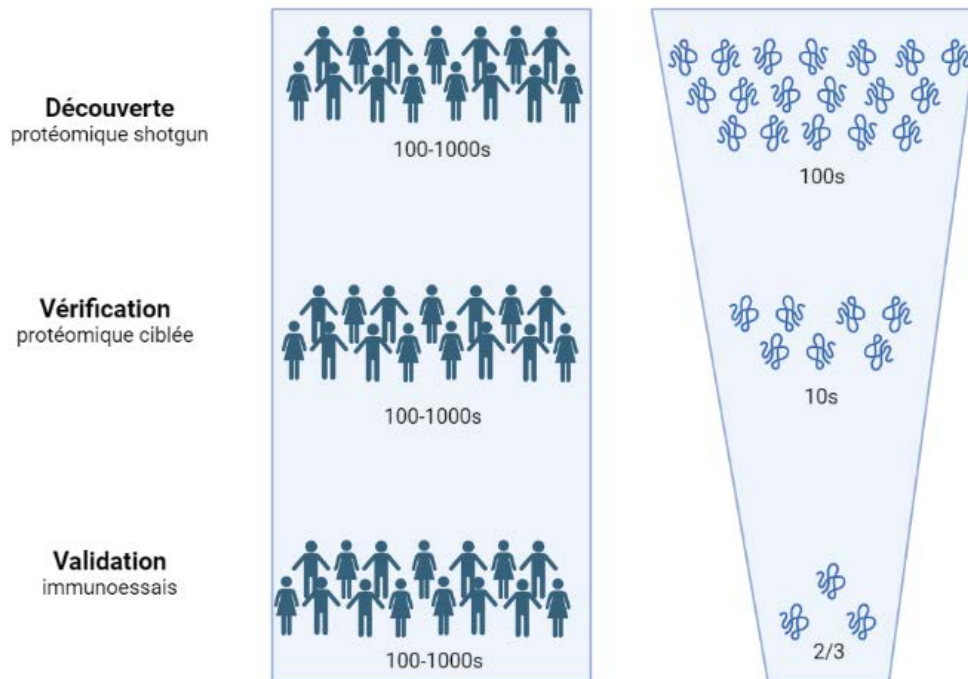


Figure 15 : Stratégie « rectangulaire » pour la découverte, vérification et validation de biomarqueurs.

Pour l'étude de vérification, il a été initialement prévu de la mener en utilisant une méthode LC-MS/MS ciblée appelée SRM (Selected Reaction Monitoring). Il s'agit d'une technique permettant d'obtenir une reproductibilité<sup>125</sup> et répétabilité élevées lors de l'analyse d'espèces peu abondantes dans des milieux très complexes<sup>126</sup>. Elle est également capable d'acquérir des données protéomiques avec une sensibilité remarquable pour des concentrations protéiques infimes<sup>127</sup>. La SRM nécessite l'utilisation d'un spectromètre de masse à triple quadrupole pour sélectionner un ion et son fragment afin cibler uniquement les molécules d'intérêt. Il est donc important de choisir avec soin les protéines, peptides et transitions qui seront analysés.

Des critères précis ont été définis pour choisir les protéines qui allaient être ciblées dans l'étude de vérification par SRM (voir partie Choix des protéines). Cependant, d'autres protéines ne remplissant pas ces critères mais dont les performances statistiques étaient intéressantes ont été choisies a posteriori pour une étude de vérification par ELISA sur la cohorte de Grenoble. Finalement, 3 protéines ont été analysées par ELISA sur la cohorte de Grenoble (Figure 16).

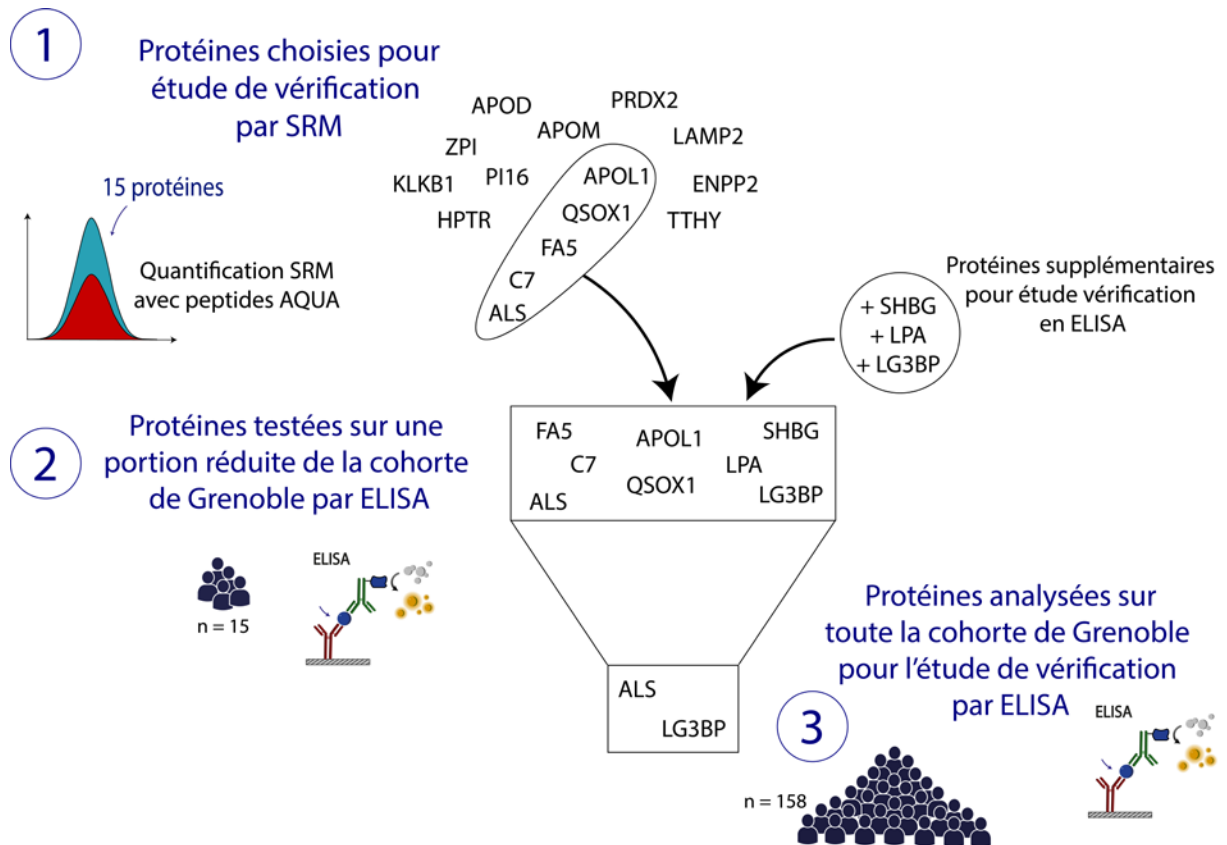


Figure 16 : Choix des protéines pour les différentes étapes de l'étude de vérification.

## 2. Choix des protéines, peptides et transitions :

### Choix des protéines :

Les critères de sélection des protéines incluses dans l'étude de vérification étaient les suivants : (1) p-valeur inférieure à 2% sur le test de limma et un  $\text{Log}_2$  (FC) inférieur à -0,5 ou supérieur à 0,5 entre fibrose précoce et avancée ; (2) expression de la protéine dans le foie (vérifiée sur Protein Atlas <sup>128</sup> en octobre et novembre 2020) ; (3) protéine non-utilisée en routine clinique pour diagnostiquer des événements physiopathologiques dans la MASLD (ex : GGT ou Alpha2m déjà utilisées dans le FibroTest). Les protéines choisies sont listées dans la partie 1 de la Figure 16.

### Choix des peptides :

Une seule protéine pouvant aboutir à la génération de dizaines de peptides après digestion enzymatique, il est important que les peptides sélectionnés soient protéotypiques, c'est-à-dire qu'ils appartiennent uniquement à la protéine d'intérêt, ces peptides sont aussi nommés quantotypiques (ou Qpeptide). Par exemple, les peptides contenant des PTMs ou comportant des résidus hautement modifiables ne sont pas ciblés car ils sont généralement présents sous plusieurs formes. C'est tout particulièrement le cas des peptides contenant des méthionines (M) en effet ce résidu est facilement oxydable<sup>129</sup>. D'un autre côté, les peptides contenant des glutamines (Q) ou des asparagines (N) sont sujets à la déamidation et donnent des glutamates (E) et des aspartates (D) ou iso-aspartate respectivement. De même, les peptides présentant une glutamine (Q) en N-terminal sont écartés car ils peuvent se transformer en pyroglutamate par cyclisation. Ainsi, si l'abondance de 2 peptides appartenant à une même protéine diverge cela peut être l'indicateur de la présence d'une PTM <sup>126</sup> ou de la présence d'une hydrolyse enzymatique partielle (miss cleavage). Des sites web pour prédire le greffage de PTMs existent ([www.uniprot.org/](http://www.uniprot.org/), [www.phosphositeplus.org/](http://www.phosphositeplus.org/), <http://phosphopep.org/>).



Il peut arriver que les peptides aient des “erreurs” de digestion aussi appelées miss cleavages. Pour une quantification robuste, il est conseillé d’éviter les peptides susceptibles d’avoir des miss cleavages car leur proportion peut varier en fonction des échantillons et donc faire varier les résultats de quantification absolue<sup>126</sup>. Les peptides contenant deux acides aminés basiques à la suite (KK, RR, KR, RK, KxD ou KxE) ont plus de chances de subir des erreurs lors de la digestion trypsique<sup>126</sup>. De même, les résidus acides (D ou E) sont susceptibles de s’exposer à des miss cleavages. Une donnée intéressante de Kuster et al. c’est que tous les peptides n’ont pas les mêmes chances d’être observés après digestion enzymatique, des référentiels en ligne donnant accès aux peptides observés le plus fréquemment pour une protéine sélectionnée peuvent être consultés pour trier ses peptides<sup>130</sup> (PeptideAtlas, HumanProteinpedia, GPM Proteomics Database, PRIDE). Ces sites utilisent des données issues d’un grand nombre d’expériences pour créer leur base de données.

Si l’analyse se fait sur un organisme “non séquencé” ou absent des bases de données, il est possible d’utiliser des algorithmes de prédiction pour savoir si les peptides sont protéotypiques<sup>131</sup> ou si leur signal sera suffisamment important pour être quantifié<sup>132</sup>. Ce type d’algorithmes utilise les données sur la charge du peptide, leur hydrophobicité mais aussi sur la séquence secondaire de ce dernier pour prédire celui qui aura les caractéristiques optimales. Des méthodes innovantes comme le deep-learning sont utilisées dans ce sens<sup>132</sup>. Cependant ces algorithmes manquent encore de précision, il est donc préférable d’utiliser les sites rassemblant des données empiriques pour construire une analyse en SRM<sup>133</sup>.

#### Choix des transitions :

Avant de commencer l’analyse par LC-SRM, les transitions à suivre doivent être renseignées (Tableau 5). En SRM, les paramètres m/z précurseur; m/z fragment et ions suivis doivent être précautionneusement choisis. Pour choisir les transitions les plus adaptées, l’analyse du spectre MS/MS du peptide d’intérêt est en général utilisée pour repérer les transitions les plus intenses et exemptes d’interférences. La transition en SRM représente la combinaison des valeurs m/z du 1er et 3e quadrupole pour un peptide et son fragment. Dans un souci d’optimisation du temps d’analyse, l’idéal est de prendre entre 2 et 4 transitions par peptide<sup>126</sup>.

*Tableau 5 : Exemple de séquences peptidiques et transitions choisies et des paramètres associés permettant de les cibler et les analyser en SRM. Tableau entier contenant toutes les transitions (précurseur/fragment) en Annexe E. Le dwell time est un terme désignant la quantité de temps que le spectromètre de masse met à analyser chaque ion, permet de déduire combien de points par pic seront disponibles.*

Séquence	m/z précurseur	m/z fragment	ion fragment	Energie de collision [eV]	dwell time
AFLTPR	409.250163	599.3875	+2y5.light	80	5ms
AFLTPR	409.250163	486.303458	+2y4.light	80	5ms
AFLTPR	414.254297	609.3958	+2y5.heavy	80	5ms
AFLTPR	414.254297	496.3117	+2y4.heavy	80	5ms
NILTSNNIDVK	615.838	1003.5419	+3y9.light	80	5ms
NILTSNNIDVK	615.838	890.4578	+3y8.light	80	5ms
NILTSNNIDVK	619.8452	1011.5560	+3y9.heavy	80	5ms
NILTSNNIDVK	619.8452	898.4720	+3y8.heavy	80	5ms

Les transitions choisies ont le plus souvent un  $m/z$  supérieur à celui du précurseur pour éviter les interférences, il est possible de prendre des transitions dont le  $m/z$  est inférieur à celui du précurseur mais le bruit de fond est souvent plus important <sup>126</sup>. En effet cela est possible, car lors de la fragmentation, une espèce dont la charge est +2 ou +3 se trouve "fractionnée" et donc le  $m/z$  peut être supérieur à celui du précurseur (Figure 17). Pour obtenir des résultats avec des données le plus sensibles possibles, il faut cibler les fragments avec les intensités les plus élevées possible (cette transition est aussi connue sous le nom de quantifier). Dans le cas où la quantifier serait contaminée, il est possible de cibler des transitions avec des intensités plus faibles mais qui ne contiennent pas de contaminations. Il faut aussi que les résultats soient le plus spécifiques possible, pour cela il faut sélectionner une transition signature appartenant uniquement au peptide d'intérêt. De plus les fragments préférentiels sont les  $y$  (Figure 17) du fait du mode de fragmentation choisi (HCD : higher energy collisional dissociation).

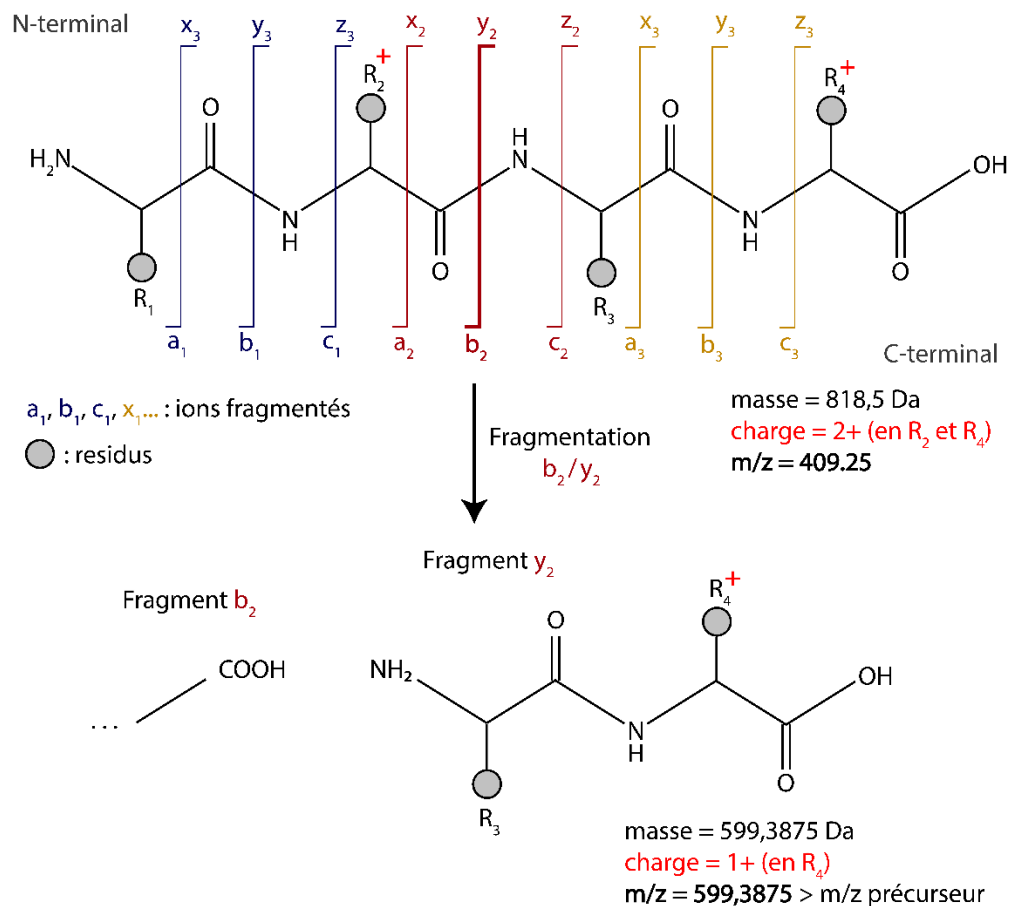


Figure 17 : Fragmentation MS/MS d'un peptide selon la nomenclature Biemann et Roepstorff et obtention d'un fragment avec une  $m/z$  supérieur à celui du précurseur.

Dans notre étude, nous avons utilisé une modalité de la SRM appelée scheduled SRM. Ce mode de fonctionnement permet d'analyser un plus grand nombre de transitions que la SRM classique, car l'analyse se focalise sur un temps d'éluion du peptide d'intérêt, augmentant la sensibilité de détection et donnant lieu à une quantification de meilleure qualité <sup>126,134</sup>.

### 3. Optimisation de la séparation des peptides en chromatographie liquide :

La chromatographie liquide (LC) est le premier élément par lequel vont être séparés les analytes, elle est couplée au spectromètre de masse. Dans le cadre de ma thèse, j'ai été amené à utiliser une phase inverse pour la séparation des peptides (Reverse phase liquid chromatography)

(Figure 18). Elle permet d'éluer les espèces dans un ordre d'hydrophobicité croissante en augmentant progressivement la part de solvant organique (tampon B) au détriment de la part de solvant inorganique (tampon A). Elle est composée d'une phase stationnaire à base de chaînes d'alkyles (-CH<sub>2</sub>)<sub>18</sub> attachées à des particules de silice recouvertes d'une fonction méthyle pour éviter la création de fonctions silanol (Si-OH) engendrant des interactions hydrophiles rendant les résultats peu reproductibles. Les peptides contenus dans l'échantillon vont être introduits dans la LC et vont être retenus par les chaînes alkylées. A mesure que la proportion de solvant organique augmente, celui-ci entre en compétition avec les liaisons hydrophobes créées entre les peptides et les chaînes alkyles, rompant les liaisons. Les peptides les plus hydrophobes vont être davantage retenus et vont être élués plus tardivement.

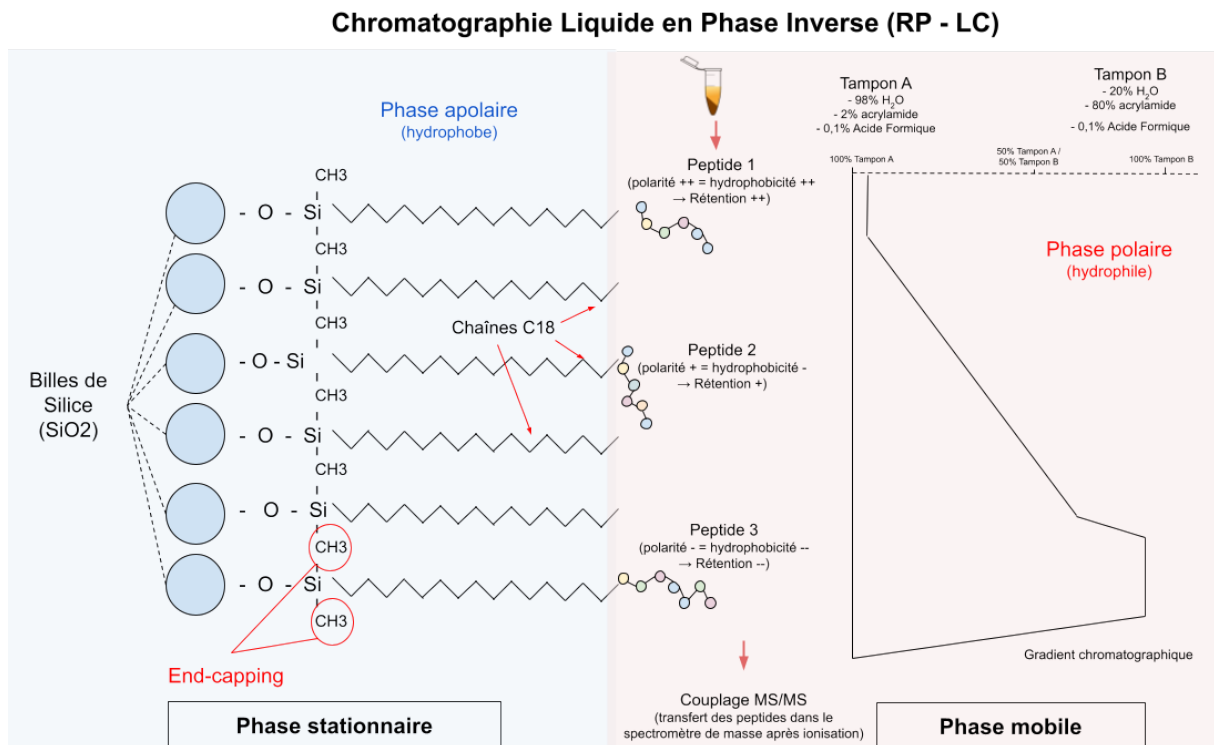


Figure 18 : Schéma du principe de fonctionnement d'une LC avec une colonne fonctionnalisée avec des chaînes alkyles C<sub>18</sub>.

La LC est couplée au spectromètre de masse, ainsi à mesure que les peptides sont élués, ils vont être introduits dans le spectromètre pour y être analysés (Annexe D pour les détails de l'injection scheduled SRM).

#### 4. Utilisation de peptides marqués isotopiquement :

En complément d'une espèce marquée isotopiquement, la SRM peut être utilisée pour obtenir une quantification très précise<sup>135</sup>. Le plus souvent, les espèces marquées utilisées en protéomique quantitative sont des peptides isotopiquement alourdis appelés AQUA (Absolute quantification peptides) (Figure 19). Par exemple l'utilisation d'une arginine marquée en <sup>13</sup>C-<sup>15</sup>N occasionne une incrémentation de masse de 10 Da. En général, seulement les arginines (R) et les lysines (K) possèdent des <sup>13</sup>C et des <sup>15</sup>N car ce sont les acides aminés pour lesquels un clivage de la liaison peptidique en C-ter s'opère lors de la digestion enzymatique à la trypsine. L'avantage des peptides dits alourdis est qu'ils possèdent les mêmes caractéristiques physico-chimiques que leurs homologues non-alourdis. Ceci permet que l'élution des deux peptides après séparation en LC se réalise en même temps (une exception existe pour les isotopes alourdis au deutérium, qui vont conférer une acidité un peu plus importante à la molécule et donc retenir légèrement plus la molécule lors du passage en LC en C<sub>18</sub>). S'ensuit l'analyse MS/MS en SRM qui va cibler les deux peptides et leurs transitions correspondantes.

Le courant d'ions des fragments suivis permettra d'établir un ratio entre l'intensité ou l'aire (plus fiable en général) de la forme alourdie et de la forme endogène. La forme alourdie ayant été introduite en quantités connues dans l'échantillon, il sera possible de calculer la concentration de peptide endogène.

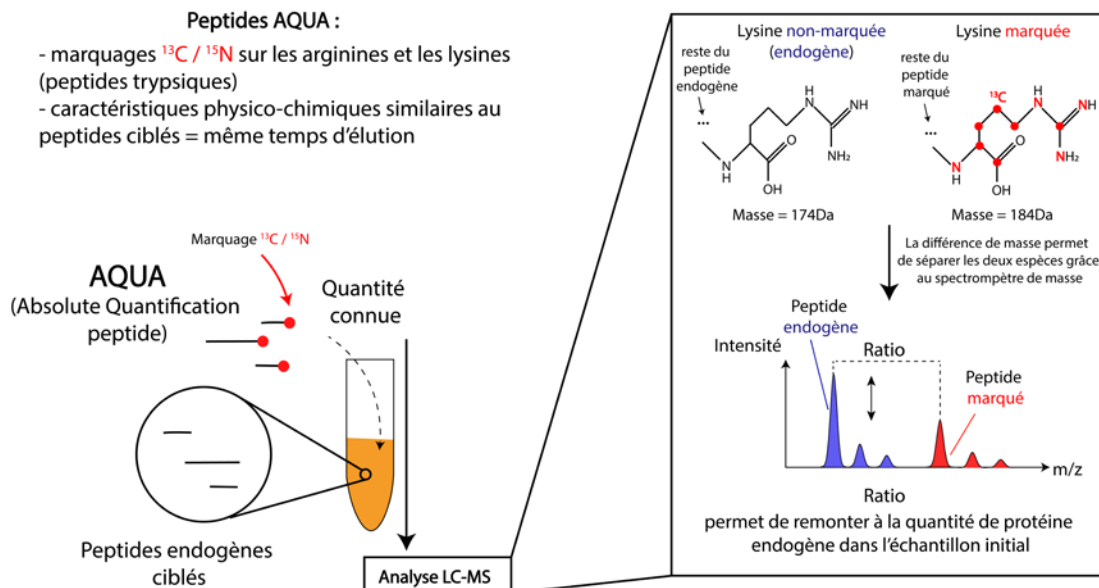


Figure 19 : Schéma général décrivant le principe de la quantification par utilisation de peptides AQUA.

Pour revenir à notre étude ciblée par SRM, 33 peptides dits "uniques" appartenant à 15 protéines présentant des différences de concentrations entre les différents stades de fibrose ont été choisis selon les critères énoncés précédemment (Choix des peptides) (Figure 20 et Annexe E pour tous les détails sur la masse et le temps de rétention de chaque peptide/transition).

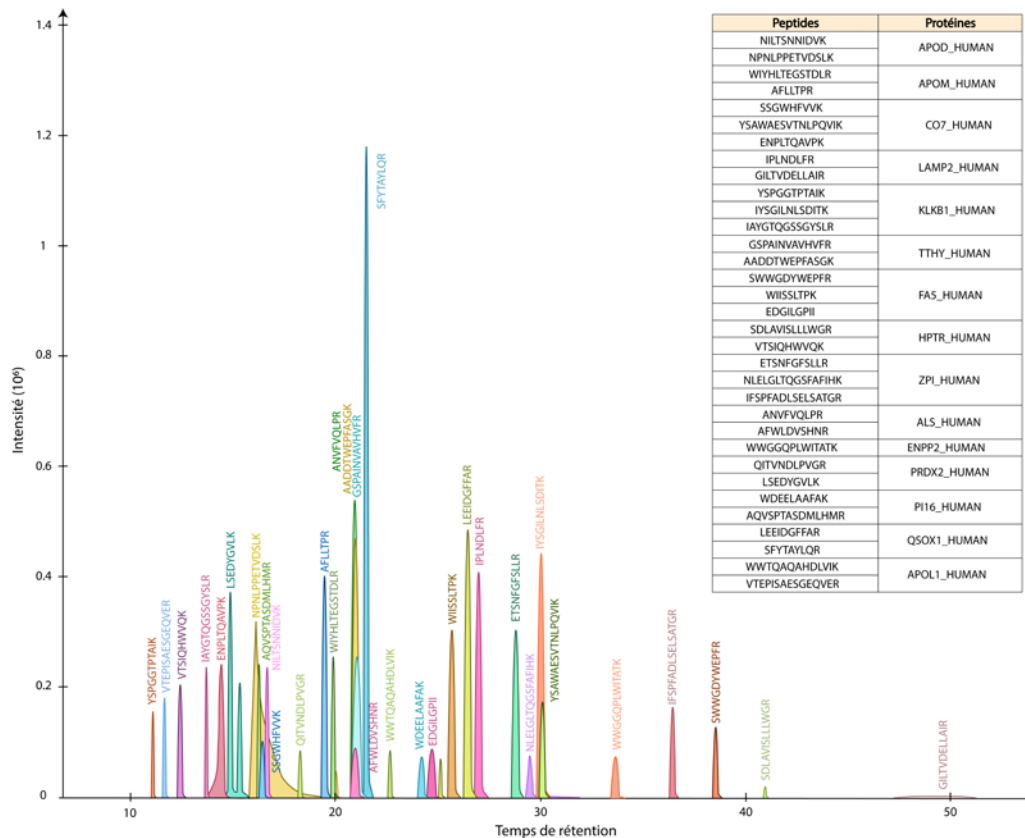


Figure 20 : Spectre de masse de tous les peptides ciblés pour l'étude SRM et tableau des protéines correspondant à chaque peptide.

Parmi les protéines d'origine hépatique choisies, trois sont des apolipoprotéines (APOL1, APOM et APOD), il s'agit de protéines dont le rôle est de se complexer avec les lipides et de les transporter<sup>136</sup>. PI16 est quant à elle une protéine dont la fonction est peu connue mais certaines études relèvent une expression de cette protéine par des cellules immunitaires<sup>137</sup> ou des cellules cardiaques<sup>138</sup>, d'autres études ont montré qu'elle pourrait servir en tant que marqueur pour le cancer de la prostate<sup>139</sup> ou de carcinome hépatocellulaire<sup>140</sup>. La QSOX1 catalyse l'oxydation de groupes sulfhydriles dans les thiols de peptides et protéines, par ce biais elle pourrait contrôler la composition et certaines fonctions de la matrice extracellulaire<sup>141</sup>. Certains articles montrent des concentrations différemment abondantes de QSOX1 pour des atteintes hépatiques alcooliques<sup>142</sup>. La PRDX2 pourrait être impliquée dans le stress oxydatif<sup>143</sup>, des homologues comme la PRDX4 et la PRDX6 semblent avoir un effet protecteur contre la MASLD<sup>144,145</sup>. De son côté, l'ENPP2 se lie avec des molécules lipidiques<sup>146</sup>, des études montrent qu'elle pourrait être liée à la fibrose chez les patients MASLD<sup>147-150</sup>. L'ALS est impliquée dans des processus de régulation de l'insuline<sup>151-154</sup>, des niveaux bas en ALS ont été associés avec la progression de la fibrose chez des patients avec des hépatites C<sup>155</sup> ou des hépatites alcooliques<sup>156</sup>. La protéine ZPI fait partie des inhibiteurs de protéases<sup>157</sup>, il s'agit d'une protéine peu connue dans le domaine des biomarqueurs hépatiques. La protéine HPTR s'associe aux apolipoprotéines et à l'hémoglobine<sup>158</sup>, aucune évidence bibliographique n'associe l'HPTR à des atteintes métaboliques, seulement une forme glycosylée de l'haptoglobine a pu être mise en relation avec la MASLD<sup>159</sup>. Le FA5 est un facteur de coagulation<sup>160</sup> peu associé à la MASLD contrairement à ses homologues FA6, FA9, FA11 et FA12<sup>161</sup>. La TTHY transporterait des protéines du sang au cerveau<sup>162</sup>, très peu d'informations sur un possible lien avec MASLD à part le fait qu'elle formerait un complexe avec RBP4 qui est largement étudiée dans ce domaine<sup>163-166</sup>. Le gène *KLKB1* code pour une glycoprotéine qui participe à l'activation de la coagulation sanguine, de la fibrinolyse, de la génération

de kinine et de l'inflammation. Dans le tissu cardiaque, KLKB1 fait partie d'un complexe ayant un rôle dans un mécanisme de déposition de gouttelettes lipidiques ce qui en ferait une potentielle cible pour un médicament contre la MASLD<sup>167,168</sup>. La protéine LAMP2 est une protéine membranaire avec de nombreuses interactions protéine-protéine<sup>169,170</sup>, seulement des articles montrant une diminution de cette protéine lors d'exercice physique pour des patients MASLD sont disponibles<sup>171,172</sup>. Finalement le C7 est une protéine du complexe d'attaque membranaire (MAC) qui est essentielle pour le système immunitaire inné et adaptatif<sup>173</sup>, plusieurs études trouvent un lien entre la MASLD et le C7<sup>174-176</sup>.

Pour la plupart des protéines choisies dans le cadre de l'étude ciblée, des publications ont déjà fait état de niveaux différentiels dans le cadre de la MASLD. En effet, un nombre important d'études de protéomique sur plasma de patients MASLD ont déjà été réalisées<sup>175,177,178</sup>. Sans préparation spécifique de l'échantillon, la protéomique permet d'identifier uniquement les protéines plasmatiques les plus abondantes, or une grande partie de ces protéines sont déjà connues dans la littérature. Atteindre des protéines plus inédites et moins abondantes, nécessiterait un enrichissement de certains groupes protéiques ou une déplétion des protéines les plus abondantes. Cependant, ces méthodes sont difficiles à mettre en œuvre sur des cohortes importantes et très peu de tests ELISA commerciaux fiables pour la validation de celles-ci sont actuellement disponibles. L'objectif final de ce projet étant de valider des protéines à des fins d'utilisation en milieu clinique, il est nécessaire de pouvoir avoir à disposition des méthodes de validation aisément déployables en laboratoire d'analyse hospitalier. Il est vrai que de nouvelles techniques telle que Somascan<sup>179</sup> ou Olink<sup>180</sup>, permettraient d'analyser de façon quantitative des milliers de protéines plasmatiques grâce à l'utilisation d'aptamères. Pour l'heure, leurs résultats ne sont pas toujours corrélés à ceux obtenus en spectrométrie de masse ( $r=0,58$  pour le protéome de cellules souches mésenchymateuses) et l'analyse de cohortes avec des centaines de patients est environ 100 fois plus chère en comparaison avec une technique ELISA qui ciblerait une seule protéine<sup>181</sup>.

#### 5. Modus operandi en SRM :

Avant l'introduction des peptides AQUA dans les échantillons plasmatiques, des tests d'équilibrage ont été réalisés de façon à ce que la transition la plus intense soit comprise dans un rapport de 1 à 10 entre le peptide endogène et la forme isotopiquement alourdie. Trois échantillons de plasma de patients à divers stades de fibrose (F0, F3 et F4) ont été choisis pour équilibrer le signal du peptide lourd dans la fourchette des concentrations des patients. Les 33 peptides AQUA ont été mélangés et conservés à  $-80^{\circ}\text{C}$  à une concentration 100x plus élevée que la concentration visée pour l'analyse afin de limiter les problèmes d'adsorption. Cette étape nous a de même permis d'éliminer les transitions présentant des interférences. Notre mélange de peptides AQUA finalisé, celui-ci a été ajouté aux 158 échantillons afin d'être injectés de façon randomisée sur une période de 4 à 5 mois en LC-MS (scheduled SRM). Afin de s'assurer du bon fonctionnement du couplage et de la stabilité de la chromatographie liquide, 3 contrôles au cytochrome C sont analysés avant les échantillons spikés de peptides AQUA alternés avec des blancs pour éviter les risques de carry-over. La quantité de peptide AQUA introduite dans les échantillons a été adaptée pour obtenir un chromatogramme équilibré, les quantités vont de  $1,2^{\text{E}}-4$  ng à 300 ng en fonction du peptide AQUA considéré.

#### 6. Traitement du signal et analyse des résultats SRM :

Le traitement des données SRM a été fait dans Skyline, un programme développé par le laboratoire de MacCoss de l'université de Washington<sup>182</sup>. Les XICs (extracted-ion chromatogram) afférents à chaque fragment de peptide sont utilisés pour définir la transition qui sera exploitée pour la quantification des peptides (quantiflier). Pour obtenir une quantification aussi précise que possible, il est nécessaire de « borner » les limites de cette transition afin de définir le signal des ions fragments mais également de repérer les possibles contaminations (Figure 21).

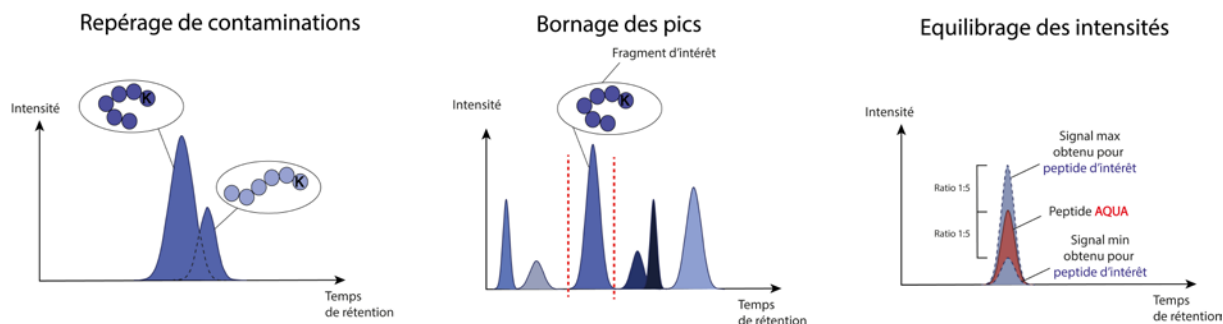


Figure 21 : Schéma du traitement des données sur Skyline avec le repérage de contaminations, le bornage des pics et l'équilibrage des intensités avec un ratio 1 à 5. Ces étapes permettent d'accéder à des signaux dont la qualité est suffisante pour réaliser une quantification satisfaisante.

Les signaux (lourd-léger) sont comparés deux à deux. Connaissant la concentration du peptide lourd, il est possible de déterminer la concentration du peptide endogène grâce un simple calcul de ratio. Par la suite, ces abondances ont été analysées avec Prostar®<sup>114</sup>, un logiciel développé par EDyP. Les protéines différenciellement exprimées (avec une p-valeur < 0,05, test de Student) entre les stades de fibroses sont regroupées dans le Tableau 6. Les résultats obtenus en SRM sur les 15 protéines sélectionnées confortent l'étude de découverte semi-quantitative. Étant donné l'ampleur du travail réalisé et le temps consacré à cette première phase de l'étude de vérification nous nous sommes focalisés sur les cinq protéines (ALS, C7, FA5, APOL1 et QSOX1) les plus différenciellement exprimées sur le plus de stades de fibrose.

Tableau 6 : Protéines différenciellement abondantes (p-valeur < 0,05) en fonction du stade de fibrose selon un test de Student après l'étude LC-MS/MS en mode scheduled SRM sur la cohorte de Grenoble.

Stade de Fibrose	Protéines différentielles (p-valeur < 0,05 t-test)
F0 vs F1	∅
F0 vs F2	C7
F0 vs F3	ALS / C7 / HPTR
F0 vs F4	ALS / C7 / HPTR / KLKB1 / APOD / APOL1 / TTHY / APOM / FA5 / QSOX1
F1 vs F2	∅
F1 vs F3	ALS / C7 / HPTR / APOM
F1 vs F4	ALS / C7 / HPTR / KLKB1 / APOD / TTHY / APOM / FA5 / QSOX1 / ZPI / PI16
F2 vs F3	∅
F2 vs F4	ALS / C7 / HPTR / KLKB1 / APOD / TTHY / APOM / FA5 / QSOX1 / ZPI / PI16
F3 vs F4	ALS

A ce stade, L'ALS et le C7 sont les protéines qui permettraient de différencier le plus de stades de fibrose, elles ont donc été conservées pour réaliser une validation par ELISA sur une deuxième cohorte.

Cependant, en prévision du changement de méthode analytique et pour conforter ces premiers résultats, nous avons décidé de ré-analyser la cohorte de Grenoble par ELISA. Nous avons d'abord

décidé d'analyser plusieurs protéines supplémentaires (LG3BP, SHBG et LPA) en plus des protéines montrant de bonnes performances en SRM (ALS, C7, FA5, APOL1 et QSOX1) sur une portion réduite de la cohorte de Grenoble. Les critères pour inclure ces nouvelles protéines dans l'étude étaient les suivants : (1) la protéine doit être quantifiée sur 70% des échantillons ; (2) les protéines doivent être différentiellement abondantes en ayant un seuil de p-valeur permettant d'atteindre un FDR inférieur à 5% suivant la procédure de Benjamini-Hochberg ; (3) critères bibliographiques.

Finalement, seuls les tests ELISA ciblant l'ALS, le C7 et la LG3BP montraient des résultats avec un fort potentiel, elles ont donc été choisies pour être analysées sur la totalité de la cohorte de Grenoble (Figure 16).

Pour ce qui est de la LG3BP, il s'agit d'une protéine se trouvant différentiellement abondante lors de l'étude de découverte. Dans un premier temps, celle-ci fut écartée du fait de l'important nombre de publications montrant qu'une forme glycosylée de la LG3BP appelée WFA+*-*LG3BP pourrait être un biomarqueur efficace pour la fibrose chez des patients MASLD<sup>183-186</sup>. Cependant, la forme non-glycosylée de la LG3BP pourrait également être un biomarqueur pertinent pour la MASLD mais pour l'instant une seule étude a pu vérifier cette information<sup>187</sup>.

#### 7. Vérification par ELISA :

Pour les mêmes raisons que pour l'étude SRM, les échantillons plasmatiques ont été randomisés pour réaliser des tests ELISA grâce à Well Plate Maker<sup>112</sup>. Les résultats obtenus pour l'ALS et la LG3BP sont très similaires à ceux obtenus lors de l'étude de découverte et SRM (Figure 22). Cela nous conforte dans l'idée de choisir cette méthode pour essayer de valider les résultats sur une deuxième cohorte. Par contre, les résultats pour le C7 sont très différents entre les deux techniques. Nous avons cependant décidé d'analyser les 3 protéines sur la cohorte angevine.



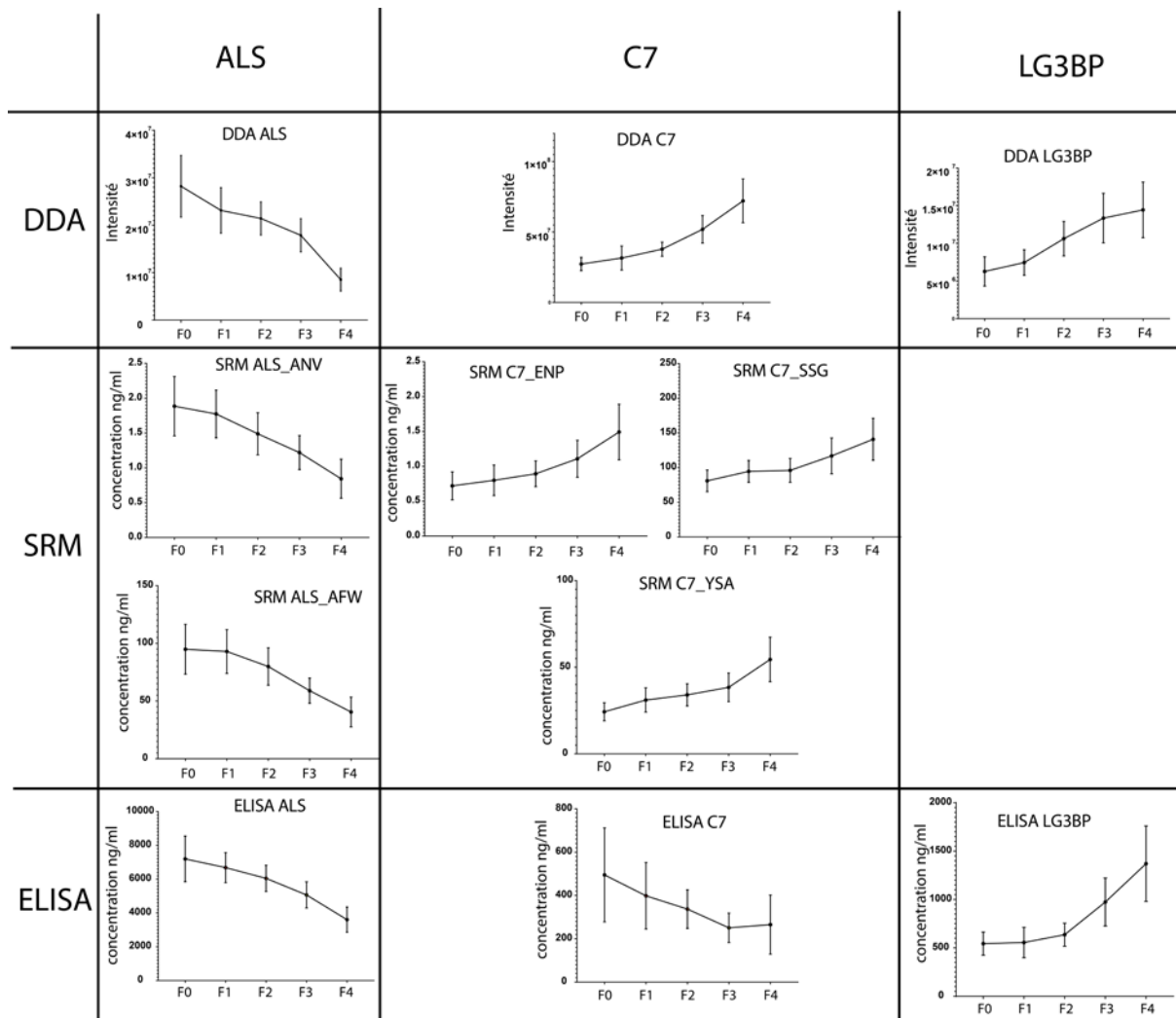


Figure 22 : Tendances obtenues pour les protéines ALS, C7 et LG3BP en fonction des stades de fibrose lors des analyses en DDA, SRM et ELISA sur la cohorte de Grenoble

#### D. Etude de validation :

##### 1. Tendances des 3 études :

La cohorte d'Angers a également été randomisée grâce à Well Plate Maker. Les résultats obtenus sont à leur tour très similaires à ceux obtenus avec la cohorte de Grenoble en méthode ELISA. L'ALS et la LG3BP se comportent de la même façon quelle que soit la méthode ou la cohorte utilisée lorsqu'on les compare aux stades de fibrose (Figure 23). Les résultats issus de la LC-MS correspondent à la moyenne de l'intensité de tous les peptides identifiés pour chaque protéine.

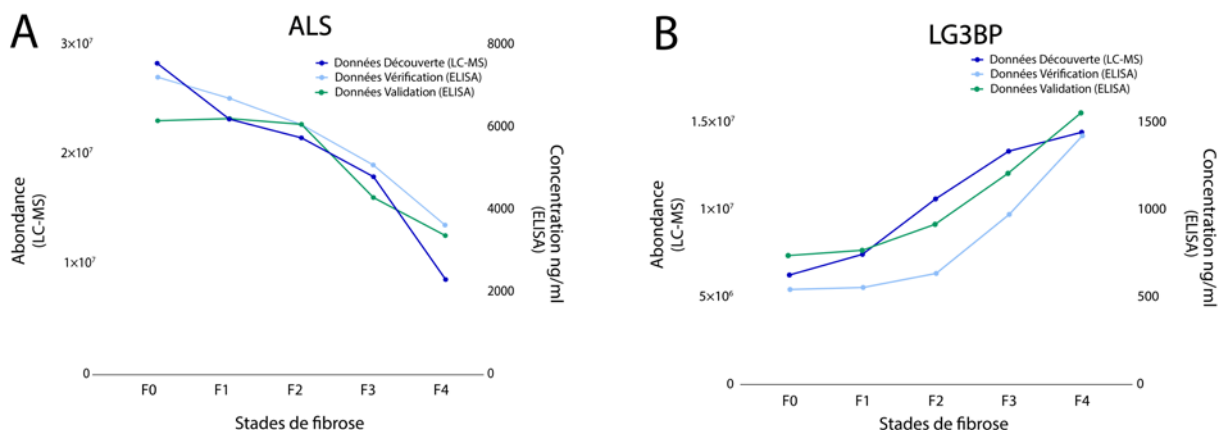


Figure 23 : Tendances des protéines ALS et LG3BP en fonction des stades de fibrose pour les études de découverte, de vérification et de validation sur les cohortes de Grenoble et Angers.

## 2. Le cas du C7 :

Le C7 a également été analysé par ELISA sur la cohorte d'Angers mais contrairement à l'ALS et la LG3BP, il ne suit pas la même tendance que lors des analyses en DDA et en SRM (Figure 22). Les peptides ciblés en SRM suivent tous la même tendance. Par contre, comme cela peut être observé sur la Figure 24, le C7 se comporte de façon similaire entre les deux cohortes en ELISA.

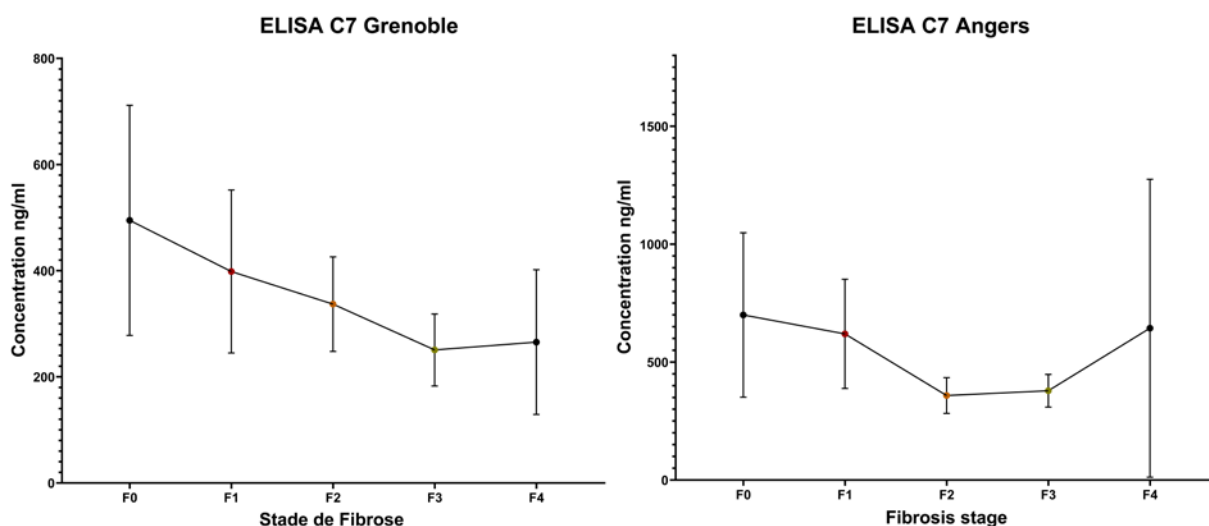


Figure 24 : Tendance du C7 en fonction des stades de fibrose pour l'étude de vérification et de validation par ELISA sur les cohortes de Grenoble et Angers respectivement.

Plusieurs raisons permettraient d'expliquer ce phénomène. Comme il a été dit précédemment, le C7 fait partie du MAC, en effet il va s'associer avec les protéines C5b – C6 – C8 – C9 pour former une structure multi-protéique impliquée dans le système immunitaire<sup>188</sup> (Figure 25). Une raison qui pourrait expliquer la disparité des résultats en F4, c'est qu'il s'agit d'un stade hépatique très avancé et par conséquent des dérèglements très importants commencent à survenir de façon généralisée<sup>189</sup>. Comme montré dans la Figure en Annexe F, la distribution des protéines est beaucoup plus variable dans les stades avancés.

Les patients atteints de MASLD présentent souvent une activation du système immunitaire, il a été démontré que les protéines du complément ont un rôle dans cette maladie notamment dans la reconnaissance de lésions tissulaires dans le cadre de l'immunité innée<sup>183</sup>. Nous nous attendons à ce que les protéines du complément se retrouvent en quantités croissantes au fur et à mesure que la

fibrose progresse. Dans le cadre de l'étude de découverte ceci est effectivement le cas pour le C7 mais aussi pour d'autres études de protéomique utilisant des systèmes LC-MS<sup>176</sup>. Par ailleurs, les résultats ont été confirmés par l'étude ciblée LC-SRM. Par contre, les tests ELISA montrent plutôt que le C7 a tendance à diminuer. De plus, des intervalles de confiance très importants sur les stades avancés sont également observables (Figure 24). La différence de résultats pourrait s'expliquer du fait que les deux techniques n'analysent pas exactement la même chose. En effet, la LC-MS va plutôt permettre d'analyser les peptides provenant de la digestion trypsique du C7 après dénaturation et donc par dissociation des possibles complexes. Ces mesures tendent à estimer la quantité totale de C7 dans le plasma, présent sous forme libre ou sous forme complexée. Bien que le MAC soit davantage connu sous forme de complexe membranaire, il peut également se retrouver sous forme soluble<sup>190</sup> ce qui expliquerait que l'on puisse le retrouver dans le plasma.

D'un autre côté, le test ELISA, de par son mécanisme anticorps/antigène, va cibler la partie libre du C7, or une partie non négligeable pourrait ne pas être disponible si le C7 est sous forme de complexe. D'après les résultats obtenus, une hypothèse peut être émise : la quantité de C7 dans le plasma augmente à mesure que la fibrose progresse comme le montrent les résultats LC-MS mais par contre la quantité de C7 sous forme libre tend à diminuer, sûrement parce que la quantité de C7 sous forme de complexe augmente pour faire face à une atteinte hépatique (Figure 25).

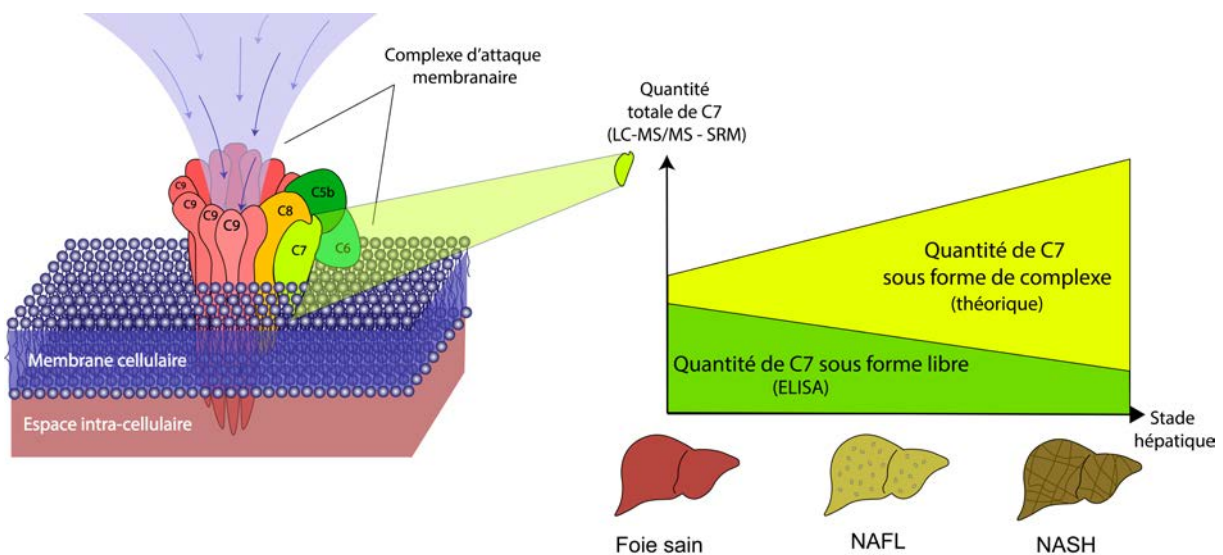


Figure 25 : Schéma du complexe d'attaque membranaire lors de la perforation d'une membrane cellulaire et représentation des quantités théoriques de C7 libre et sous forme de complexe lors des différents stades MASLD.

Une étude multi-omique sur la MASLD regroupant des tests génomiques, transcriptomiques et protéomiques a été réalisée sur deux cohortes de 35559 et 47151 patients pour identifier des biomarqueurs de la MASLD<sup>191</sup>. Les données disponibles de cet article ne montrent pas de différences significatives pour le C7 au niveau génomique et transcriptomique. Par contre, au niveau protéomique, les quantités de C7 sont significativement plus abondantes lors de cirrhoses hépatiques. Malgré l'utilisation d'une technique de protéomique différente (aptamères), les résultats vont dans le même sens que notre étude de découverte.

Nous avons donc émis l'hypothèse que le test ELISA utilisé sur les deux cohortes était défectueux. Pour essayer de vérifier cela, nous avons analysé en MS/MS en mode découverte, les standards du test ELISA. Malgré une injection de 10 ng de solution de standard dans le spectromètre de masse, le C7 n'a

pas été identifié. Les seules protéines qui ont été retrouvées correspondent à des protéines majoritaires du sérum bovin (Tableau 7).

*Tableau 7 : Tableau regroupant les protéines majoritaires observées dans un échantillon de standard de C7 après injection en LC-MS/MS en mode DDA. La colonne iBaq correspond à la somme des intensités des peptides observés divisé par le nombre de peptides théoriquement observables. Cette métrique peut donner une idée de l'abondance relative des protéines, elles ont été classées en fonction de la métrique iBaq dans la colonne % Total.*

Rang	Protéine	Espèce	Identifiant	iBaq (Abondance)	% Total
1	Précurseur de l'albumine sérique bovine	Bos taurus (bœuf)	P02769	9.82E+08	93%
2	Précurseur de l'alpha-1-glycoprotéine acide	Bos Taurus (bœuf)	Q3SZR3	2.21E+07	2%
3	Ubiquitine-protéine ligase	Homo sapiens (humain)	O75150	2.10E+07	2%
4	Trypsine	Sus crofa (cochon)	P00761	1.03E+07	1%
5	Précurseur de l'endopine-1	Bos taurus (bœuf)	Q9TTE1	2.47E+06	0%
...	...	...	...	...	...
59	Semblable au précurseur du complément C4-A	Bos taurus (bœuf)	P00000007350	1.48E+03	0%

Il est cependant important de souligner que ce n'est pas parce que le C7 n'a pas été observé, qu'il n'est pas présent dans le standard. En effet, malgré la capacité de l'instrumentation LC-MS actuelle de détecter des protéines dans une gamme dynamique importante, il est possible que des protéines dont l'abondance est très faible passent inaperçues. Suite à ces résultats, des interrogations sont apparues quant aux informations sur la concentration en C7 données par le fournisseur car les quantités injectées étaient largement suffisantes pour qu'il soit identifié. Des analyses supplémentaires par des méthodes LC-MS ciblées ou un test de la spécificité des anticorps pourraient permettre d'obtenir des réponses plus précises sur la nature du problème rencontré avec ce test ELISA.

Enfin, un résultat intéressant qui peut être retrouvé en Annexe G, c'est qu'en DDA, les protéines C5, C6, C8A, C8B, C8G et C9 ont une tendance opposée au C7. On pourrait donc penser qu'une erreur s'est produite lors de l'analyse en DDA, or les résultats vont dans le même sens lors de l'analyse en LC-SRM et les études bibliographiques confortent aussi la tendance observée. Il est donc possible que le C7 ait un comportement différent des autres protéines composant le MAC lors d'une atteinte hépatique fibrotique.

### 3. Traitement statistique :

Pour revenir aux résultats sur l'ALS et la LG3BP, lorsque les stades de fibrose précoce (F0-2) sont comparés aux stades avancés (F3-4) avec les 3 études nous obtenons des résultats statistiquement différentiels avec des p-valeurs de 9,4E-07, 4.4E-07, 3.0E-09, et 1.3E-06, 4.1E-07, 2.8E-09 pour les études de découverte, de vérification et validation pour les protéines ALS (Figure 26. A, B et C) et LG3BP respectivement (Figure 26. D, E et F). Malgré l'exactitude de nos résultats il n'est pas possible d'être discriminant, c'est-à-dire qu'il n'y a pas de concentration qui permette de définir le stade de fibrose du patient avec certitude. Ceci peut être observé sur la Figure 26 où l'on observe un recouvrement

entre les concentrations des stades précoces et des stades avancés. Pour essayer d'obtenir des résultats plus discriminants, souvent des panels regroupant plusieurs paramètres (dont la concentration de la protéine en question) sont créés.

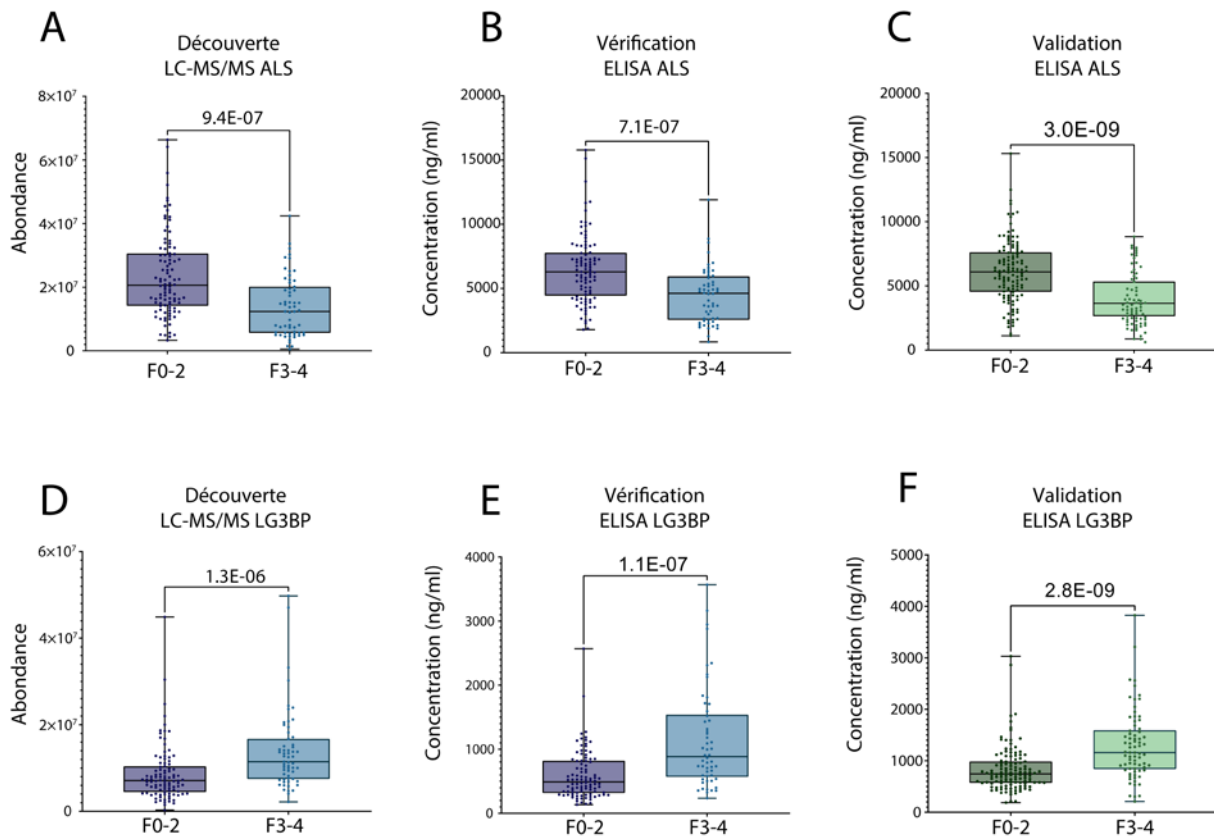


Figure 26 : Comparaison des abondances moyennes de l'ALS et de la LG3BP dans les échantillons de plasma en fonction des stades de fibrose et des phases de l'étude sur les biomarqueurs. Abondance MS (étude de découverte, unité arbitraire) et concentration en ng/ml (études de vérification et de validation, mesures ELISA) de l'ALS (A) et du LG3BP (B).

Le test statistique choisi pour comparer les moyennes des deux groupes de fibrose est le t-test de Mann-Whitney. Il s'agit d'un test non-paramétrique pour lequel l'hypothèse nulle  $H_0$  est qu'il n'y a pas de différence entre les deux moyennes et l'hypothèse alternative  $H_1$  dit que les distributions ne sont pas identiques. L'étude réalisée par Christin, C. et al. montre que le test de Mann-Whitney serait plus adapté que d'autres tests ayant pour objectif de discriminer des variables comme le principal component discriminant analysis ou le « nearest shrunken centroid » parmi d'autres<sup>192</sup>, surtout pour des études avec un nombre important d'échantillons.

La méthode que nous avons choisie pour analyser la performance des biomarqueurs est l'aire sous la courbe (AUC) d'une courbe ROC (receiver operating characteristic). La courbe ROC reflète la sensibilité et la spécificité d'un test à différents seuils. D'un côté, la sensibilité représente le nombre de vrais positifs, dans le cas du projet MASLD, cela se traduirait par le pourcentage de patients diagnostiqués avec une fibrose avancée par un biomarqueur réellement à ce stade de la maladie. D'un autre côté, la spécificité est le pourcentage de faux négatifs, c'est-à-dire les personnes sans fibrose avancée identifiées par le biomarqueur. Chaque point de la courbe ROC représente un seuil de décision. Lorsque tous les points de la courbe sont déterminés, une courbe peut être tracée et son AUC correspondante peut être calculée. Dans le cas de notre étude, on estime que la biopsie hépatique a une AUC de 1 et donc que tous les patients sont correctement diagnostiqués. Cependant, ceci n'est pas nécessairement vrai, mais faute de disposer d'une meilleure technique de diagnostic, ce postulat doit être posé tout en connaissant les limites qu'il suscite quant à la pertinence des biomarqueurs

étudiés par la suite. Malgré le fait que certaines études aient suggéré que la courbe ROC pourrait être insuffisante pour identifier de petites mais potentiellement importantes discriminations<sup>193-195</sup>, d'autres études démontrent le contraire et continuent à préconiser son utilisation<sup>196-198</sup>.

Nous avons dans un premier temps comparé la performance des deux biomarqueurs avec celle du FibroTest car il s'agit d'un des panels les plus répandus pour estimer la fibrose en cas de suspicion de MASLD et dont l'ensemble des paramètres nécessaires pour en calculer le score étaient disponibles pour les deux cohortes (Figure 27). Les performances sont calculées sur les données ELISA de la cohorte de validation (Angers).

*Figure 27 : Courbes ROC et AUC de l'ALS, la LG3BP et du FibroTest avec leurs intervalles de confiance (IC) à 95 % issus des quantifications ELISA ALS/LG3BP et du score FibroTest sur la cohorte d'Angers, pour la classification entre les stades de fibrose précoces (F0-2) et les stades avancés (F3-4).*

Cette Figure montre que les performances sur la cohorte d'Angers de l'ALS (0.735 [0.661 – 0.81]) et la LG3BP (0.744 [0.673 – 0.816]) prises de façon indépendante sont quasiment équivalentes à celles du FibroTest (0.758 [0.691 – 0.825]), qui est un panel issu de 6 variables.

#### E. Création d'un panel :

##### 1. Introduction :

En général, pour essayer d'améliorer les résultats d'un biomarqueur, un panel regroupant différentes variables est créé. Plusieurs méthodologies existent pour faire cela. La technique la plus classique est la régression logistique. Elle a pour objectif d'expliquer au mieux une variable binaire (fibrose précoce vs fibrose avancée) grâce à d'autres variables qui se voient appliquer des coefficients. De nombreux panels existent pour déterminer le stade de fibrose chez des patients MASLD de façon non-invasive : Fib-4<sup>199</sup>, NAFLD Fibrosis score (NFS)<sup>82</sup>, European liver fibrosis panel (ELF)<sup>200</sup>, Fibrometer<sup>201</sup> or Fibrotest<sup>202</sup> entre autres (voir Annexe A).

L'équation pour calculer le score du FibroTest est issue d'une régression logistique :

$$Z = 4.467 \times [A2m(g/L)] - 1.357 \times [Hp(g/L)] + 1.017 \times [GGT(IU/L)] + 0.0281 \times [Age] + 1.737 \times [Bil(\mu mol/L)] - 1.184 \times [APOA(g/L)] + 0.301 \times \beta - 5.54$$

*Equation 1 : Equation du FibroTest. Alpha2m : alpha-2-macroglobuline ; HP : haptoglobuline ; GGT : gamma-glutamyltransférase ; Bil : bilirubine ; APOA : Apolipoprotéine ;  $\beta$  : sexe.*

Des méthodes plus modernes utilisant du « machine learning » (ML) peuvent également être mises en place. Cependant, le problème avec le ML est qu'il nécessite trois cohortes pour être utilisé correctement. Ainsi une première cohorte servira d'entraînement, une deuxième cohorte est appelée de validation et une troisième servira en tant que test. Dans notre approche nous n'avons que deux cohortes, et même s'il serait envisageable d'en diviser une de façon équilibrée, cela pourrait avoir des répercussions sur la généralisation du test<sup>203</sup>. En effet, nos données ELISA étant biaisées car nous savons que ces protéines (ALS et LG3BP) permettent de distinguer correctement la fibrose avancée de la fibrose précoce à partir des données DDA et SRM, il existe un risque d'obtenir une bonne performance sur la cohorte d'entraînement mais que la performance sur la cohorte de validation soit médiocre. Un autre problème que nous pourrions rencontrer est que lorsqu'une cohorte est utilisée en tant que cohorte de validation avec une technique de ML, elle ne peut pas être réutilisée dans des études à postériori car elle créerait un biais chez l'expérimentateur qui connaît en avance les résultats que fourniraient ces échantillons. La cohorte d'Angers étant incluse dans des études parallèles, nous avons pris la décision de ne pas utiliser de méthode ML pour créer un panel mais plutôt une régression logistique. Le FibroTest étant l'un des panels les plus répandus et dont l'ensemble des paramètres nécessaires pour en calculer le score étaient disponibles pour les deux cohortes, nous avons décidé de l'utiliser pour réaliser notre panel.

Dans un deuxième temps et suite aux résultats très encourageants de la Figure 26, nous avons combiné l'ALS et la LG3BP afin de la comparer au FibroTest (Figure 28). La combinaison des deux variables a été réalisée par une méthode de régression logistique.

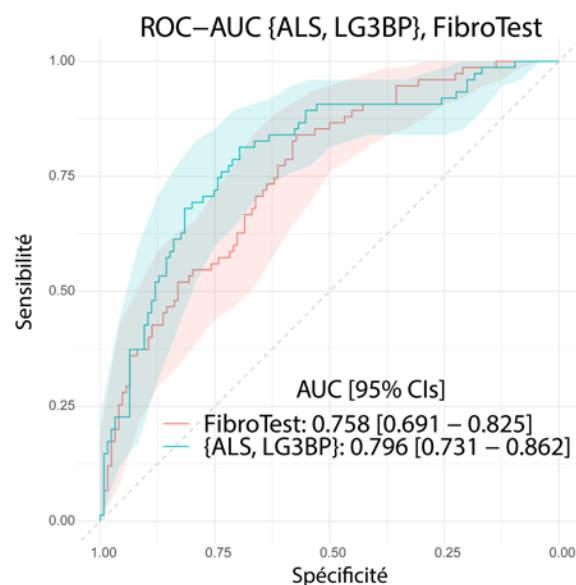


Figure 28 : Courbes ROC et AUC du FibroTest et de la combinaison {ALS, LG3BP} avec leurs intervalles de confiance (IC) à 95% issus de la combinaison des quantifications ELISA ALS/LG3BP et du score FibroTest sur la cohorte d'Angers, pour la classification entre les stades de fibrose précoces (F0-2) et les stades avancés (F3-4).

Nous pouvons observer que la combinaison des deux biomarqueurs validés présente une performance plus élevée (0.796 [0.731 – 0.862]) que celle du FibroTest (0.758 [0.691 – 0.825]) sur la cohorte de validation (Figure 28).

## 2. Re-fittage FibroTest :

Suite aux résultats obtenus, nous avons émis l'hypothèse que la combinaison de l'ALS et la LG3BP avec le FibroTest (ou du moins avec certaines variables du FibroTest) pourrait améliorer la performance de stratification de la fibrose. Cependant, la validation de cette hypothèse doit être faite avec prudence car la modification d'un panel préexistant par l'ajout (ou la suppression) de variables

nécessite une évaluation précise de l'implication de chaque variable dans les performances du panel. Une méthode appelée « nested hypothesis » a depuis longtemps été utilisée dans ce but <sup>204</sup>, cependant nous n'avons pas pu l'utiliser car le FibroTest n'a pas été entraîné sur la même cohorte que celle que l'on s'apprête à utiliser. Pour y remédier, nous avons créé un objet mathématique intermédiaire appelé FibroTest RF (pour Re-Fitted ou réadapté) qui consiste essentiellement à redéterminer le poids des variables du FibroTest en utilisant un modèle de régression logistique sur la cohorte de découverte. Le FibroTest RF a par la suite été testé sur la cohorte d'Angers et nous avons obtenu une AUC significativement plus élevée (0,843 [0,789 - 0,898]) que celle du FibroTest original (0,758 [0,691 - 0,825]) (Figure 29).

Ces résultats nous ont permis d'émettre l'hypothèse que toutes les variables du FibroTest RF n'étaient pas forcément pertinentes pour distinguer une fibrose précoce d'une fibrose avancée sur les cohortes de Grenoble et Angers. Nous avons donc retiré une par une les variables contribuant le moins au modèle FibroTest RF puis testé la similarité de chaque nouveau modèle en comparaison avec la version précédente en utilisant un test de likelihood ratio (Tableau 8). Pour illustrer ces propos, sur la première colonne il est possible de voir que la variable qui contribue le moins au FibroTest RF est l'haptoglobine. S'en suivent la bilirubine, le sexe, l'apolipoprotéine et l'âge par ordre croissant de contribution au FibroTest RF. Les p-valeurs obtenues pour chacune de ces variables étant lors de chaque comparaison au-dessus du seuil  $\alpha$  de 0,05 ; elles peuvent être retirées du modèle sans différences significatives. Il ne reste donc que deux variables contribuant de manière significative au FibroTest RF. Celles-ci sont l'Alpha2m et la GGT.

Tableau 8 : Tableau des résultats des likelihood ratio tests entre les différents modèles de FibroTest RF

		Modèle						FibroTest complet
		FibroTest RF complet	Alpha2m, GGT, âge, bil, sexe, APOA	Alpha2m, GGT, âge, sexe, APOA	Alpha2m, GGT, âge, APOA	Alpha2m, GGT, âge	Alpha2m, GGT	
Variables testées pour élimination	Alpha2m	3,47E-06	2,84E-06	2,44E-06	2,65E-06	1,14E-06	1,44E-07	0,8067
	GGT	0,001898	0,001862	0,00169	0,001663	0,00283	0,0029	
	Age	0,3319	0,3377	0,3047	0,2771	0,3669		
	APOA	0,2731	0,2597	0,2927	0,3006			
	Sexe	0,6453	0,656	0,6881				
	Bilirubine	0,7202	0,6581					
	Haptoglobuline	0,8155						

Nous obtenons donc une version réduite du FibroTest RF avec uniquement deux variables : {Alpha2m, GGT}. La performance AUROC de cette version réduite (0.840 [0.785 – 0.895]) est très similaire à celle du FibroTest RF (0.843 [0.789 – 0.898]) (Figure 29).



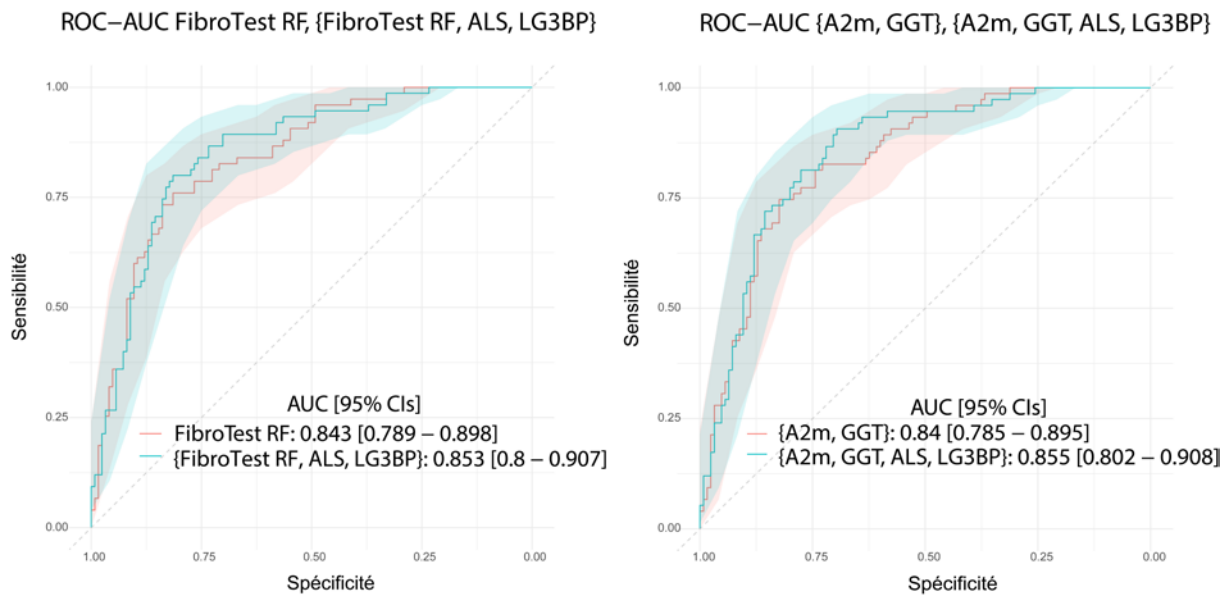


Figure 29 : Courbes ROC et AUROCs avec leurs 95% CI respectifs du FibroTest RF et {Alpha2m, GGT} combinés avec ALS et LG3BP. Les modèles ont été ajustés sur la cohorte de Grenoble et les ROC-AUROC ont été mesurés sur la cohorte d'Angers.

Par la suite, nous avons introduit l'ALS et la LG3BP dans les nouveaux modèles FibroTest RF et {Alpha2m, GGT} afin d'estimer si ces nouveaux biomarqueurs apportaient un gain significatif à la performance du modèle. Nous pouvons voir sur la Figure 29 que l'ALS et la LG3BP apportent respectivement un gain de 0,01 et 0,015 en AUROC aux deux modèles. Il est cependant difficile d'estimer avec cette figure si cela représente un gain ou pas car les intervalles de confiance se superposent de façon considérable. Pour évaluer la pertinence de l'addition de l'ALS et la LG3BP, nous avons à nouveau réalisé des likelihood ratio tests uniquement ajustés sur la cohorte d'Angers (Tableau 9). Ce qui a été testé dans ce tableau c'est si les modèles incluant l'ALS, la LG3BP ou les deux, s'ajustent significativement mieux aux modèles FibroTest RF et {Alpha2m, GGT} que lorsque ces variables ne sont pas contenues dans ces modèles. Pour l'ALS, nous obtenons des p-valeurs de 0,004 et 0,011 pour le FibroTest RF et le modèle {Alpha2m, GGT} respectivement, alors que pour la LG3BP les p-valeurs sont de 0,06 et 0,02. Lorsque l'on introduit l'ALS et la LG3BP ensemble dans les modèles, les p-valeurs obtenues sont de 0,007 et 0,009. En somme, l'ajout de l'ALS et la LG3BP dans le modèle {Alpha2m, GGT} ou de l'ALS dans le modèle FibroTest RF donne lieu à des modèles significativement différents, il est donc important de conserver ces variables dans chacun des modèles décrits.

Tableau 9 : P-valeurs des likelihood ratio tests pour le FibroTest RF et {Alpha2m, GGT} lors de l'ajout de l'ALS et la LG3BP ou de la combinaison {ALS ;LG3BP} sur les données d'Angers. L'hypothèse nulle pour une valeur p donnée à la i-ème ligne et à la j-ème colonne est que la ou les variables à la j-ème colonne n'apportent aucune contribution lorsqu'elles sont ajoutées au i-ème modèle. L'hypothèse alternative est que la (les) variable(s) j-ème apporte(nt) une amélioration significative au modèle i-ème.

		Variable(s) ajoutée(s)		
		ALS	LG3BP	LG3BP & ALS
Modèle par défaut	FibroTest RF	0.004152	0.06582	0.007259
	{Alpha2m, GGT}	0.01107	0.02777	0.009031

Le principe du rasoir d'Ockham (Ockham's razor) recommande d'utiliser des explications construites avec le plus petit nombre d'éléments possible pour résoudre un problème. C'est en suivant cette prémisse que nous avons proposé le modèle {Alpha2m, GGT, ALS, LG3BP} comme nouveau panel

pour différencier les stades de fibrose précoces des stades avancés pour des patients MASLD. La performance AUROC de ce modèle dans la cohorte de validation est de 0.855 [0.802 – 0.908] et les détails sur les poids de chaque variable ainsi que les seuils de sensibilité et spécificité sont donnés dans l'Equation 2 et le Tableau 10.

$$Z = -3.3570909 + 3.0592294 \times \text{Log}_{10}[\text{Alpha2m}(g/L)] - 0.0002089 \times [\text{ALS}(g/L)] \\ + 0.9868492 \times \text{Log}_{10}[\text{GGT}(IU/L)] + 0.0012399 \times [\text{LG3BP}(g/L)]$$

*Equation 2* : Equation du modèle {Alpha2m, GGT, ALS, LG3BP} ayant pour objectif de différencier les fibroses précoces des fibroses avancées pour des patients MASLD.

*Tableau 10* : Seuils de sensibilité et spécificité du modèle {Alpha2m, GGT, ALS, LG3BP} basés sur la courbe ROC de la cohorte de validation

<i>Spécificité</i>	<i>Sensibilité (95% CI)</i>	<i>Seuil</i>
50%	94.7% (89.3 – 98.7%)	0.217
75%	81.3% (70.7 – 93.3%)	0.360
90%	53.3% (26.7 – 76%)	0.590
95%	28% (12 – 54.7%)	0.755

A titre de comparaison, une AUROC a été calculée sur plusieurs variables et panels utilisés normalement en clinique dans le but de prédire la fibrose hépatique (Tableau 11). L'AUROC a été calculée sur la cohorte de validation (cohorte d'Angers). Parmi les tests comparés, il y a des tests de 1<sup>ère</sup> ligne, qui sont des tests sanguins réalisés lorsque le patient arrive au service d'hépatologie. Les panels issus des tests de première ligne peuvent également contenir des paramètres comme l'âge, le sexe ou l'IMC des patients et sont normalement utilisés pour obtenir des meilleurs scores que ceux obtenus pour les tests de 1<sup>ère</sup> ligne, or ce n'est pas forcément le cas lorsqu'ils sont testés avec la cohorte de validation de l'étude MASLD. Les tests de deuxième ligne ont pour but normalement de donner des informations complémentaires aux médecins lorsque les tests de 1<sup>ère</sup> ligne ne suffisent pas à émettre un diagnostic ou prendre une décision. Il s'agit souvent de tests non inclus dans la routine hospitalière et qui doivent être réalisés dans des laboratoires de biologie médicale.

Tableau 11 : Tableau de comparaison de l'AUROC obtenue sur plusieurs tests de 1<sup>ère</sup> ligne, de panels issus de tests de 1<sup>ère</sup> ligne, sur des tests de 2<sup>ème</sup> ligne et sur les biomarqueurs et panels développés dans le projet MASLD. L'AUROC a été calculée sur la cohorte de validation (cohorte d'Angers) et indique la performance de chaque test pour prédire une fibrose précoce ou avancée chez des patients diagnostiqués MASLD. HA : acide hyaluronique ; INR : international normalized ratio ; NFS (NAFLD fibrosis score) : âge + IMC + diabète + ASAT/ALAT + plaquettes + albumine ; APRI : AST to platelet ratio index ; AAR : AST to ALT ratio index ; FibroQ : INR + ASAT + ALAT + plaquettes + âge ; Forns index : plaquettes + âge + GGT + cholestérol, HUI score : IMC + bilirubine + albumine + plaquettes ; LOK index : plaquettes + AST + INR

	Test	AUROC F0-2 vs F3-4 (Cohorte de validation)
<b>Tests de 1<sup>ère</sup> ligne</b>	ASAT	0,723
	ALAT	0,628
	Plaquettes	0,492
	INR	0,726
<b>Panels issus de tests de 1<sup>ère</sup> ligne</b>	Fib-4	0,641
	NFS	0,676
	APRI	0,696
	AAR	0,581
	FibroQ	0,631
	Forns index	0,624
	HUI score	0,577
LOK index	0,677	
<b>Tests ou panels de 2<sup>ème</sup> ligne</b>	Fibroscan	0,856
	HA	0,738
	FibroTest	0,758
<b>Tests et panels développés sur l'étude MASLD</b>	ALS	0,744
	LG3BP	0,735
	{ALS; LG3BP}	0,798
	A2M	0,741
	GGT	0,663
	A2M ; GGT	0,840
	FibroTest RF	0,843
	{A2M; GGT; ALS; LG3BP}	0,855

Le Tableau 11 montre que la performance obtenue par le panel {A2M; GGT; ALS; LG3BP} (0,855) développé dans le cadre du projet MASLD est du même calibre que la performance du fibroscan (0,856). Ce modèle a été créé sur une cohorte indépendante de la cohorte sur lequel il a été testé, ce qui permet d'avoir une confiance relativement importante sur la capacité de généralisation de ce panel. Cependant, et comme cela peut être observé sur ce même tableau, les panels n'arrivent pas toujours à obtenir des performances supérieures que celles obtenues avec les variables qui les composent. Un Tableau plus complet, comparant la performance de chacune de ces variables sur leur capacité à prédire des stades avancés d'activité, inflammation lobulaire, ballonnement des hépatocytes, stéatose et NASH peut être retrouvé en Annexe H. Ce tableau complémentaire montre que les panels développés ne sont utilisables que pour prédire la fibrose car les performances sont très faibles pour déterminer les autres types d'atteintes hépatiques.

## F. Conclusion projet MASLD :

La MASLD est un problème de santé mondial du fait de sa prévalence, de ses conséquences, de l'absence de thérapies validées et du manque de méthodes de diagnostic précoce non invasives mais fiables. En ce qui concerne le diagnostic, la fibrose est le paramètre le plus important à déterminer, car elle permet aux cliniciens de déterminer la morbidité liée au foie ainsi que la stratégie la plus adaptée et le traitement du patient<sup>78</sup>. Il est donc urgent d'identifier des biomarqueurs capables de différencier les stades précoces (F0-2) des stades les plus avancés (F3-4) de la fibrose afin de mieux surveiller la maladie. À cette fin, nous avons conçu une étude protéomique de découverte visant à détecter les variations de l'abondance des protéines individuelles dans le plasma des patients atteints de MASLD. Nous avons identifié l'ALS et la LG3BP comme biomarqueurs potentiels parce qu'elles étaient différentiellement abondantes dans le plasma de patients présentant une fibrose précoce et avancée, mais aussi parce qu'elles sont principalement exprimées dans le foie et que des tests ELISA efficaces ciblant ces protéines étaient disponibles dans le commerce. Les deux protéines ont été validées dans une cohorte indépendante de 200 échantillons de plasma à l'aide de tests ELISA. Ces deux protéines sont impliquées dans des mécanismes biologiques liés aux processus physiopathologiques de la MASLD, tels que la régulation de l'insuline<sup>205</sup>, l'activation du système immunitaire ou les interactions avec la matrice extracellulaire<sup>206</sup>. La diminution des niveaux d'ALS a déjà été associée à la progression de la fibrose chez des patients atteints du virus de l'hépatite C<sup>155</sup> ou d'hépatite alcoolique<sup>156</sup>. Pour la LG3BP, des études antérieures ont montré des différences de glycosylation liées à la MASLD et à la fibrose<sup>187</sup>. Notre étude montre que les taux plasmatiques d'ALS et de LG3BP peuvent différencier les stades précoces des stades avancés de la fibrose.

A ce jour, de nombreux panels ont été développés pour mieux différencier la fibrose hépatique dans la MASLD<sup>82,200,202</sup>. Nous avons démontré que les performances de l'ALS (0,744 [0,673 - 0,816]) et du LG3BP (0,735 [0,661 - 0,81]) prises indépendamment étaient similaires à celles du FibroTest (0,758 [0,691 - 0,825]) dans l'identification de la fibrose avancée dans la cohorte de validation (Figure 26). Cependant, la combinaison de l'ALS et du LG3BP (0,796 [0,731 - 0,862]) a conduit à une amélioration significative de la performance par rapport au FibroTest. Nous avons, par la suite combiné les résultats afférents à l'ALS et la LG3BP avec le FibroTest pour améliorer les performances dans l'identification de la fibrose avancée. Pour cela, nous avons réentraîné les poids des variables du FibroTest sur la cohorte de Grenoble afin d'obtenir un FibroTest RF dont la performance était nettement améliorée par rapport aux autres modèles (0,843 [0,789 - 0,898]) (Figure 28). Nous avons ensuite démontré que 5 variables du FibroTest pouvaient être supprimées sans perte de performance significative. Les deux variables restantes (Alpha2m, GGT) combinées à ALS et LG3BP permettent d'obtenir un AUROC de 0,855 [0,802 - 0,908]. Cependant, nous mettons en garde contre les risques de surinterprétation : il n'est pas possible d'interpréter le FibroTest RF comme une amélioration par rapport à sa version originale, car nous n'avons pas comparé les performances entre les différentes cohortes (y compris celle sur laquelle le FibroTest a été validé). Par conséquent, le FibroTest RF devrait essentiellement être interprété comme un modèle mathématique abstrait, fournissant un terrain pour des comparaisons équitables : toute modification du panel produisant une AUROC plus importante que le FibroTest RF pourrait être considérée comme améliorant le FibroTest original, même dans le pire des scénarios où la différence de performances provenant du changement de cohorte ne serait pas prise en compte. Une fois encore, les résultats présentés démontrent que la combinaison de biomarqueurs usités en routine avec l'ALS et le LG3BP permet d'obtenir un panel plus performant pour diagnostiquer les stades précoces et avancés de la fibrose mais des expériences sur des cohortes complémentaires sont nécessaires.

Bien que notre étude ait été réalisée sur deux cohortes indépendantes et que les résultats soient confirmés par deux méthodes d'analyse différentes, nous notons quelques limitations. Tout d'abord, le stade de fibrose a été déterminé par une biopsie du foie, qui n'a été effectuée que sur des patients soupçonnés d'être atteints d'une fibrose avancée. Ceci peut affecter nos résultats étant donné la distribution du nombre de patients avec un stade précoce de fibrose dans nos deux cohortes. Par conséquent, même si certains patients ont peu ou pas de fibrose et ont été testés pour une hépatite virale, ils sont susceptibles d'avoir une maladie du foie qui pourrait affecter nos résultats. Une interrogation qui émerge suite à ces observations est : est-ce que les biomarqueurs et panels développés permettraient de différencier une fibrose chez des patients pour lesquels une fibrose avancée n'aurait pas été suspectée ? Pour répondre à ce questionnement, des biopsies sur des patients réunissant ces conditions devraient être réalisées, mais cela pose des problématiques éthiques car un test aussi invasif que la biopsie ne devrait être réalisé que lorsque l'état du patient l'exige.

Deuxièmement, des études antérieures ont montré que l'erreur d'échantillonnage dans les biopsies du foie peut conduire à des erreurs importantes de diagnostic et de stratification dans la MASLD parce que les lésions histologiques sont réparties de manière inégale dans le parenchyme hépatique. Les résultats des biopsies hépatiques peuvent également être plus ou moins précis en fonction de la longueur de l'échantillon<sup>207</sup>, de la distribution des lésions histologiques dans le foie<sup>72</sup> ou de la variabilité inter- et intra-observateur<sup>61</sup>, ce qui entraîne des erreurs importantes de diagnostic et de stratification de la MASLD. Ces avec ces raisonnements en tête qu'une étude a déterminé que l'AUC maximale pouvant être obtenue pour un biomarqueur de la fibrose hépatique est de 0,9<sup>208</sup>. Néanmoins, l'obtention de résultats équivalents en termes de performance entre les deux techniques différentes utilisées dans cette étude, telles que la MS et l'ELISA, garantit la validité des deux biomarqueurs tout en évitant les biais instrumentaux. L'avantage de la technique ELISA par rapport à la technique LC-MS dans l'analyse de grandes cohortes est qu'elle peut être réalisée en moins de temps et qu'elle est plus adaptée à une utilisation dans un laboratoire clinique hospitalier. En outre, l'analyse LC-MS nécessite un contrôle instrumental rigoureux en raison de sa complexité inhérente et de sa sensibilité aux variations techniques<sup>209</sup>.

D'autres limitations, notamment au niveau de la représentativité des cohortes, ont été décrites dans le manuscrit. En effet, même si l'âge et le sexe des cohortes analysées semblent représentatifs des populations française, européenne et mondiale, un test statistique démontrant cela aurait été pertinent. De même, comme cela est montré en Figure 9, il y a une surreprésentation de certains stades de fibrose (F1-2-3) dans les cohortes par rapport à la vraie distribution de la fibrose dans la population française. Cela peut limiter la généralisation des biomarqueurs et des panels développés dans le cadre de cette étude. De même, les deux cohortes, même si elles sont indépendantes, manquent d'un certain degré de diversité géographique. En effet, elles proviennent toutes les deux d'hôpitaux français, on pourrait donc se demander si les biomarqueurs et panels développés marcheraient aussi bien sur des patients provenant d'autres continents et dont les modes de vie et caractéristiques biologiques seraient différents.

Les résultats concernant l'étude de découverte, la vérification par ELISA de l'ALS et la LG3BP ainsi que leur validation et le développement du panel ont été rédigés sous forme d'article et soumis à un journal d'hépatologie. Les résultats de la technique SRM n'ont pas été décrits dans l'article parce que la LG3BP n'a pas été analysée par cette méthodologie. Cependant, la SRM a été très utile pour établir un premier filtre tout en réalisant une quantification plus fiable des protéines triées sur l'étude de découverte. Un avantage évident de la SRM par rapport aux tests de type ELISA est sa capacité de multiplexage. En effet, il est possible d'analyser un nombre important de protéines/peptides en même temps. Cependant, l'utilisation de peptides AQUA pour réaliser une quantification dans un échantillon

soulève plusieurs problèmes. Tout d'abord, les échantillons ne sont pas analysés en même temps contrairement aux tests ELISA, en effet, l'analyse de 158 échantillons en SRM a été réalisée sur une période de 5 mois alors que les tests ELISA pour une protéine ont pris 3 jours pour ce même nombre d'échantillons. Le temps peut nuire à la sensibilité de l'instrumentation MS mais peut également induire une dégradation des échantillons ce qui est problématique pour la robustesse des résultats. Cependant, la protéomique de haut débit étant en constant progrès, une analyse d'une cohorte de plusieurs centaines d'échantillons pourrait être faite de façon plus rapide et fiable en ce moment qu'au début de ma thèse. Ces progrès pourraient également faire l'objet prochainement d'une implémentation généralisée de l'instrumentation LC-MS dans le milieu hospitalier, qui actuellement utilise plutôt des tests ELISA pour l'analyse de biomarqueurs. Il faut noter que le fait d'introduire un nouveau test ELISA en milieu hospitalier est un processus de validation long qui peut prendre de nombreuses années. Enfin, une dernière limitation concernant l'instrumentation LC-MS lors de l'analyse de biomarqueurs c'est la difficulté d'obtenir des quantifications absolues de qualité. En effet, les peptides AQUA étant introduits à un stade tardif de la préparation d'échantillons, les quantifications protéiques réalisées par ce moyen ne peuvent pas donner une idée précise de la concentration réelle de protéine dans l'échantillon. Au contraire, les tests ELISA s'affichent comme des tests pouvant arriver à ce but même si nous avons montré au travers de l'analyse du C7 qu'ils peuvent également donner des résultats erronés. Le domaine de la LC-MS a un réel besoin de disposer d'une quantification absolue de meilleure qualité. La deuxième partie du manuscrit décrit les expériences que nous avons menées pour essayer de répondre à ce besoin.

## Partie II :

### Projet PSAQ<sup>+1</sup> : optimisation d'une technique de quantification absolue par utilisation de standards protéiques :

#### A. Introduction projet PSAQ<sup>+1</sup> :

##### 1. Généralités :

En protéomique, une quantification fiable des protéines est parfois nécessaire, particulièrement en recherche clinique où la concentration des biomarqueurs donne des informations cruciales dans le diagnostic, le pronostic et le suivi thérapeutique<sup>210</sup>. Les études protéomiques qui aboutissent à la validation d'un biomarqueur sont divisées en plusieurs étapes<sup>209,211</sup>. Dans la première des étapes, dite de découverte, un nombre important de protéines sont identifiées et quantifiées par différentes approches<sup>88</sup>. Lorsque la phase de découverte est réalisée avec une instrumentation LC-MS, on effectue en général une quantification en label-free. Ce type de quantification utilise l'intensité du signal spécifique aux différents peptides ou le comptage de spectres appartenant à un peptide (Figure 30)<sup>212-214</sup>. Il s'agit donc d'une quantification relative car l'abondance de chaque peptide ou protéine n'est connue que par rapport aux autres protéines présentes dans l'échantillon.

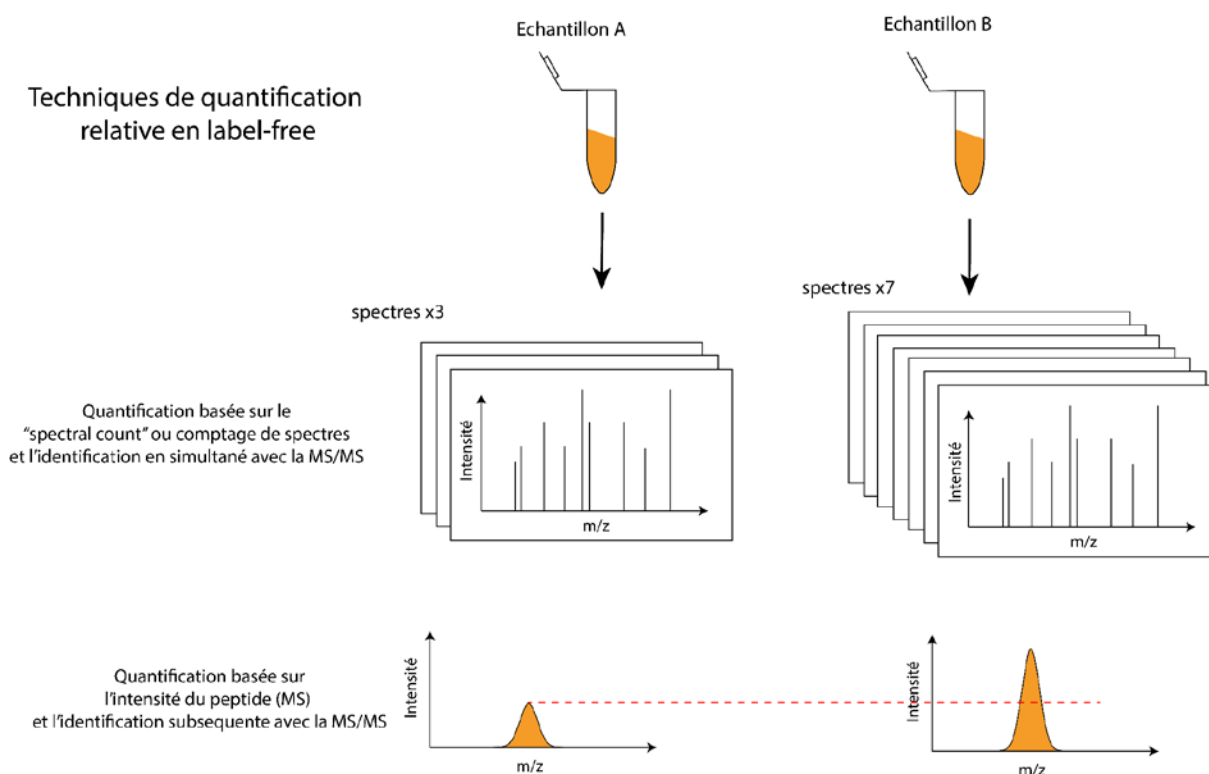


Figure 30 : Techniques de quantification relative label-free utilisées en protéomique LC-MS pour des études de découverte. Sont représentées la quantification par le moyen du « spectral count » ou comptage de spectres et par le moyen de l'intensité des peptides.

Des méthodes de quantification plus précises sont nécessaires dans les étapes plus avancées de recherche de biomarqueurs que sont les étapes de vérification et validation. Une technique très couramment utilisée lors d'analyses ciblées en protéomique clinique est l'utilisation de standards marqués isotopiquement<sup>215-218</sup>. Le principe est d'utiliser une molécule isotopiquement alourdie

(isotopologue) et de l'introduire dans l'échantillon en quantités connues, celle-ci servira de point de référence pour connaître la quantité de molécule(s) ciblée(s). La molécule marquée isotopiquement s'apparente à la molécule ciblée : la formule chimique est la même excepté les éléments sous forme d'isotopes. En protéomique, différents types de molécules marquées isotopiquement ont été développées : AQUA<sup>219</sup>, QconCAT<sup>220</sup>, PSAQ<sup>218</sup>. La méthode AQUA (pour Absolute quantification peptides) est amplement utilisée de par sa disponibilité commerciale et son prix abordable. Il s'agit de peptides s'apparentant à des peptides tryptiques dont les carbones et/ou les azotes des lysines et des arginines sont isotopiquement alourdis. Le choix d'alourdir ces acides aminés spécifiquement s'explique par le fait que les protéines d'intérêt sont clivées en C-terminal (C<sub>ter</sub>) des lysines et arginines par la trypsine et/ou la LysC lors de la digestion enzymatique. Ces acides aminés sont donc retrouvés dans tous les peptides en position C<sub>ter</sub>. Ce type de peptide a été utilisé pour l'étude SRM du projet NAFLD (Vérification ou évaluation de l'utilité clinique des cibles protéiques). Ainsi introduit dans l'échantillon, le peptide AQUA, de même séquence que le peptide à suivre (dit endogène ou non-marqué), se différencie par la présence des isotopes. La subtilité de cette technique réside dans le fait que le peptide AQUA et le peptide non-marqué auront les mêmes propriétés physico-chimiques mais des masses différentes. Ces caractéristiques sont intéressantes, car ces deux peptides sont élués en même temps par la chromatographie liquide, mais sont différenciés par leur masse lors de l'analyse par spectrométrie de masse. Si la quantité de peptide alourdi est connue, il suffit de comparer son intensité avec celle du peptide ciblé pour en déduire son abondance (Figure 19). Il s'agit d'une technique largement utilisée en protéomique quantitative<sup>218</sup> et plus précisément dans la recherche et validation de biomarqueurs en recherche clinique<sup>217,221-223</sup>.

Les techniques susmentionnées reposent sur la technique d'analyse par dilution isotopique, méthode quantitative de dosage d'éléments ou d'espèces dans un échantillon. La dilution isotopique est une méthode à étalon interne. Ici, le standard (la forme alourdie de l'espèce à analyser) est ajouté à l'échantillon. Contrairement aux méthodes semi-quantitatives reposant sur l'intensité du signal, la dilution isotopique s'appuie sur la mesure du rapport des signaux (endogène et alourdi). Dans cette partie de manuscrit, je suis amené à présenter les concepts et résultats obtenus dans le cadre de mon projet PSAQ<sup>+1</sup>.

## 2. Concepts et définitions :

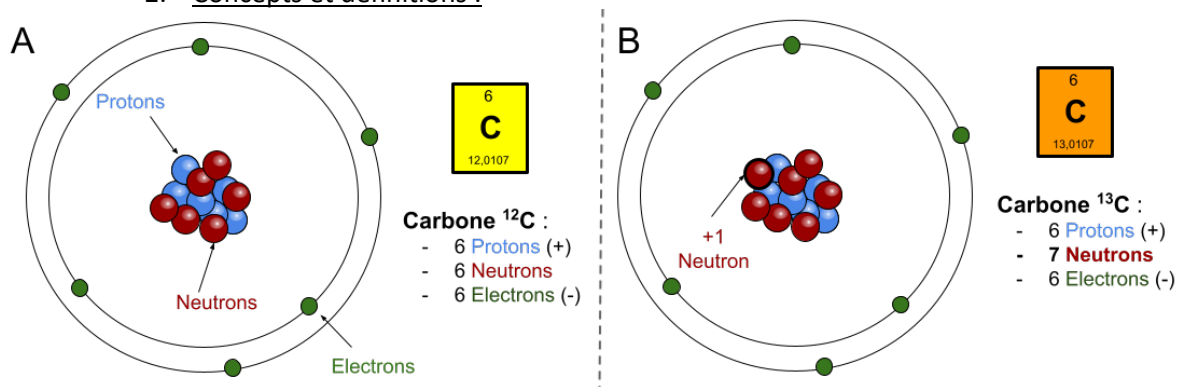


Figure 31 : Schématisation atomistique des isotopes du carbone (<sup>12</sup>C et <sup>13</sup>C). A. Représentation d'un atome de carbone <sup>12</sup>C avec 6 protons, 6 neutrons et 6 électrons. B. Représentation d'un atome de carbone <sup>13</sup>C avec 6 protons, 7 neutrons (ajout d'un neutron supplémentaire) et 6 électrons.

Comprendre, le projet PSAQ<sup>+1</sup> nécessite de revenir à des notions de chimie de base. La masse de tout élément du tableau périodique de Mendeleïev est définie par la somme du nombre de protons et de neutrons (Figure 31.A). Certains éléments peuvent avoir une quantité plus ou moins importante de neutrons tout en partageant le même nombre de protons, c'est ce que l'on appelle des isotopes. Par



exemple, le carbone possède une masse moyenne de 12,0107 g/mol mais possède différents isotopes avec des masses différentes, le carbone  $^{12}\text{C}$  ( $^{12}\text{C}$ ) a une masse de 12,0107 g/mol alors que le carbone  $^{13}\text{C}$  a une masse de 13,0107 g/mol grâce à un neutron supplémentaire (Figure 31.B).

De façon naturelle, des isotopes se forment pour certains éléments : le carbone 12 a une abondance naturelle de 98,93%, le carbone 13 est présent à 1,07% puis les carbones 14 ( $^{14}\text{C}$ ) et 11 ( $^{11}\text{C}$ ) peuvent également être retrouvés mais uniquement sous forme de traces (Figure 32).

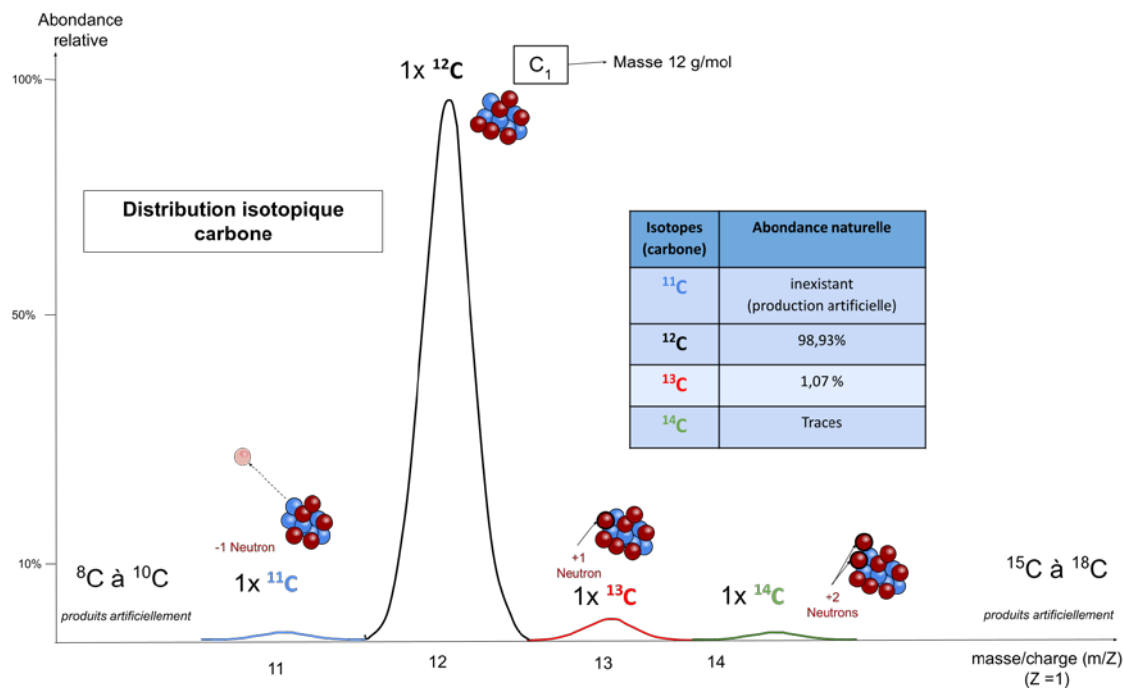


Figure 32 : Schématisation de la distribution isotopique des carbones  $^{11}\text{C}$ ,  $^{12}\text{C}$ ,  $^{13}\text{C}$  et  $^{14}\text{C}$  ainsi que de leur abondances naturelles.

Dans le cas d'une molécule dont on connaît le nombre de carbones et en s'appuyant sur la distribution isotopique théorique du carbone, il est possible d'établir la quantité de cette molécule possédant 1 atome  $^{13}\text{C}$ , 2 atomes  $^{13}\text{C}$  ou n atomes  $^{13}\text{C}$  (Figure 32). Une formule de statistiques combinatoires permet de déterminer la proportion de la molécule ne possédant qu'un seul carbone  $^{13}\text{C}$  dans une molécule (Equation 3).

Légende :

$$I = A^{X-1} \cdot B \cdot X$$

- I : probabilité de trouver une molécule avec un certain nombre d'isotopes
- A : abondance atome non isotopique (probabilité exprimée de 0 à 1)
- B : abondance isotope (probabilité exprimée de 0 à 1)
- X : nombre d'atomes ciblés dans la molécule

Exemple 1 :

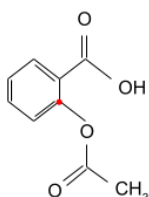
Proportion d'aspirine (9 atomes de carbone) avec un atome  $^{13}\text{C}$  :

$$I = A^{X-1} \cdot B \cdot X$$

$$I = 0,99^{9-1} \cdot 0,01 \cdot 9$$

$$I = 0,083$$

$$I = 8,3\%$$



Exemple 2 :

Proportion de glucagon (153 atomes de carbone) avec un atome  $^{13}\text{C}$

$$I = A^{X-1} \cdot B \cdot X$$

$$I = 0,99^{153-1} \cdot 0,01 \cdot 153$$

$$I = 0,332$$

$$I = 33,2\%$$

Equation 3 : Equation permettant de calculer la proportion de molécules ne possédant qu'un seul isotope alourdi en fonction du nombre d'atomes ciblés de la molécule et de l'abondance de l'isotope alourdi.

D'un point de vue statistique, plus la molécule est composée de carbones, plus la proportion de molécule sans isotope  $^{13}\text{C}$  va diminuer. Au contraire, la proportion de cette molécule avec 1 ou n isotopes  $^{13}\text{C}$  va augmenter. La Figure 33 illustre l'exemple de la myoglobine qui se compose de 759 atomes de carbone. Lors d'une analyse en spectrométrie de masse, si la résolution de l'appareil le permet, les différents isotopes d'une molécule peuvent être séparés. De plus, l'intensité associée à chaque isotope sera proportionnelle à son abondance (Figure 33).

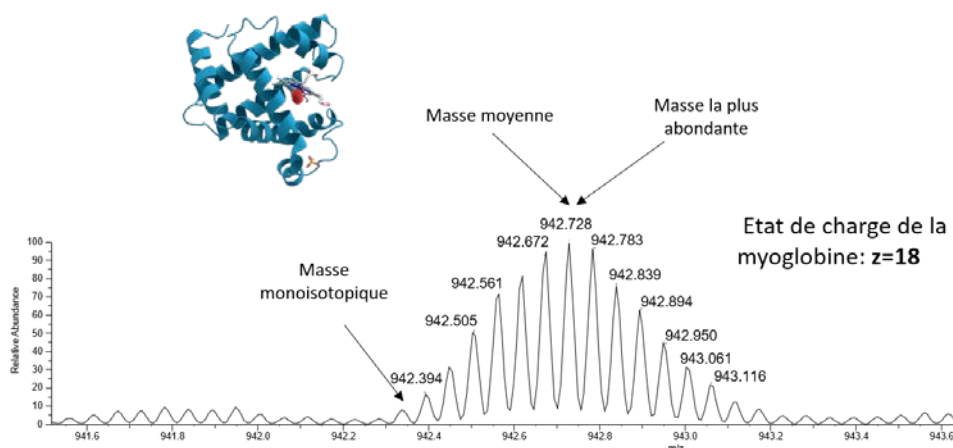


Figure 33 : Distribution isotopique de l'état de charge  $z=18$  de la myoglobine avec des indications sur la localisation de sa masse monoisotopique, sa masse moyenne et sa masse la plus abondante.

Dans la plupart des molécules, il existe des isotopes pour d'autres atomes que le carbone, qui sont également présents dans les acides aminés des protéines. Les principaux atomes des acides aminés sont l'hydrogène, le carbone, l'oxygène, l'azote et le soufre, dans l'ordre du plus abondant au moins abondant. Le Tableau 12 regroupe les atomes pouvant être retrouvés dans les acides aminés et les abondances de leurs isotopes naturels.

Tableau 12 : Isotopes naturels des éléments présents dans les protéines (les cases vert foncées encadrent l'isotope stable le plus abondant, les cases vert clair le deuxième le plus abondant, en rouge on retrouve les isotopes apportant un incrément de masse de 1Da par rapport à la forme la plus abondante).

Carbone		Azote		Oxygène		Hydrogène		Soufre	
Isotope	Abondance	Isotope	Abondance	Isotope	Abondance	Isotope	Abondance	Isotope	Abondance
$^{11}\text{C}$	traces	$^{14}\text{N}$	99,6%	$^{16}\text{O}$	99,8%	$^1\text{H}$	99,98%	$^{32}\text{S}$	94,93%
$^{12}\text{C}$	98,93%	$^{15}\text{N}$	0,4%	$^{17}\text{O}$	~0,04%	$^2\text{H (D)}$	0,015%	$^{33}\text{S}$	0,76%
$^{13}\text{C}$	1,07%			$^{18}\text{O}$	~0,2%	$^3\text{H}$	traces	$^{34}\text{S}$	4,29%
$^{14}\text{C}$	traces							$^{35}\text{S}$	Traces
								$^{36}\text{S}$	0,02%

### 3. Méthodes de quantification par PSAQ :

En quantification absolue par HPLC-MS, la méthode la plus répandue reste l'utilisation de peptides AQUA<sup>224</sup>. Cependant, du fait de l'introduction des peptides AQUA dans l'échantillon juste avant l'injection en LC-MS cette technique est souvent inexacte. Cette imprécision dans l'approche AQUA est due au fait que les étapes de préparation des échantillons sont évitées pour le peptide alourdi contrairement au peptide endogène ciblé qui provient d'une protéine dénaturée, réduite, alkylée et digérée. Cela aboutit souvent à des inexactitudes dans la détermination des abondances du fait que les étapes de préparation d'échantillon n'ont pas un rendement parfait<sup>225</sup>. Par exemple, des

miss clivages enzymatiques peuvent survenir lors de la digestion<sup>226,227</sup>, ainsi que des dénaturations, réductions et alkylations partielles ou encore une perte de peptides lors du dessalage<sup>228,229</sup>. Des adsorptions peuvent aussi se produire sur les parois des tubes destinés à la préparation des échantillons<sup>230</sup>. Ces effets ont pour conséquence que la quantité de protéine mesurée est différente de la quantité réelle et initiale de la protéine dans l'échantillon. Cette erreur lors d'une quantification par peptide AQUA est difficilement évaluable. De plus, il est difficile d'obtenir une couverture complète de la séquence protéique ce qui peut limiter la fiabilité statistique de la quantification<sup>225</sup>. Malgré l'ajout d'une quantité identique de peptides issus d'une même protéine, les signaux observés lors de l'analyse en LC-MS peuvent varier de plusieurs ordres de grandeur (Vérification ou évaluation de l'utilité clinique des cibles protéiques). Même si les termes peuvent porter à confusion, ces conditions font que l'on situe sur une quantification dite relative plutôt que sur une quantification absolue.

Pour obtenir une quantification plus robuste, les solutions mises en œuvre tentent de diminuer ces effets. Pour cela, l'introduction de molécules marquées isotopiquement à des étapes plus précoces de la préparation de l'échantillon peuvent s'avérer utiles. Par exemple, le QconCAT (Quantification concatemers) est une concaténation de peptides tryptiques alourdis isotopiquement sur les acides aminés lysine et arginine. Les QconCAT sont issus de l'expression d'un gène artificiel dans un vecteur d'*Escherichia coli*.<sup>220</sup> S'apparentant à une séquence protéique, le QconCat est introduit en amont de la digestion de l'échantillon, il ne va donc pas subir les étapes de préparation de l'échantillon mise à part la digestion. Cependant, la molécule introduite le plus en amont reste le PSAQ (Protein standard absolute quantification), elle a été développée dans à EDyP pour remédier à un manque de précision dans les quantifications par marquage isotopique<sup>218</sup>. Il s'agit d'une protéine recombinante également alourdie au niveau des lysines et des arginines par des isotopes <sup>13</sup>C-<sup>15</sup>N. Cette protéine est introduite dans l'échantillon avant toute étape de préparation, ainsi elle va subir en même temps que la protéine ciblée les différents processus préalables à l'analyse. Les deux protéines ayant les mêmes caractéristiques physico-chimiques, elles pourront être comparées de façon plus exacte (Figure 34).

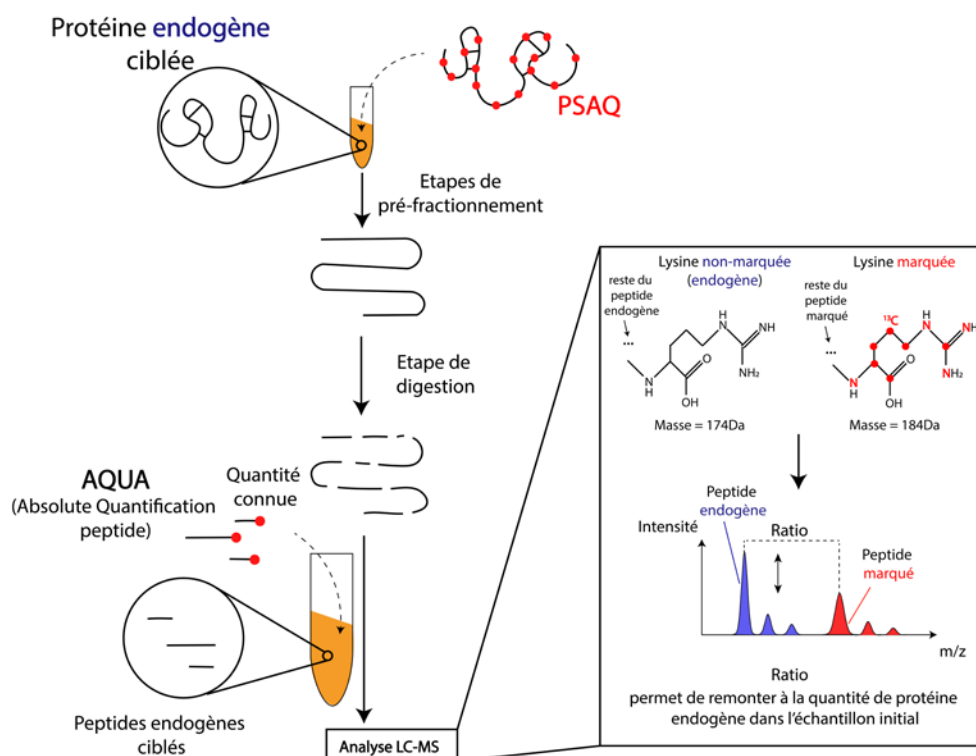


Figure 34 : Pipeline de protéomique quantitative avec dilution isotopique, ici sont référencés les différents types de molécules isotopiquement alourdis destinées à la quantification de peptides ou de protéines<sup>218</sup>

Cependant, lors d'une analyse HPLC-MS/MS, même avec l'utilisation d'une protéine PSAQ, il y a une étape où la forme marquée et ciblée seront séparées : il s'agit de l'étape de MS<sub>1</sub>. En effet, à ce niveau, les peptides (PSAQ et endogène) seront analysés de façon indépendante. Les différences de masse entre les deux molécules seront de +8 ou +10Da en fonction de la nature de l'acide aminé marqué en N-terminal. Ceci aura pour conséquence que les deux peptides seront analysés avec des fenêtres de masse distinctes (Figure 35). Afin de s'affranchir de cette limite, nous avons élaboré une stratégie appelée PSAQ<sup>+1</sup>, permettant de ne réaliser qu'une sélection unique au niveau du précurseur.

#### 4. Quantification par utilisation de standards protéiques :

##### Concept de la méthode PSAQ<sup>+1</sup>

Le PSAQ<sup>+1</sup> consiste à ajouter en quantité définie dans un échantillon biologique une protéine marquée avec un seul <sup>13</sup>C sur les lysines et les arginines. Cette protéine, analogue de la protéine endogène à doser est utilisée comme standard de quantification. Les massifs isotopiques de la protéine endogène et du PSAQ<sup>+1</sup> se chevauchent. L'analyse consiste à cibler uniquement, dans une fenêtre de masse restreinte, le signal composite issu du premier pic isotopique du PSAQ<sup>+1</sup> et du deuxième pic isotopique de la protéine endogène à cibler. Afin d'effectuer la quantification de la protéine endogène, il est nécessaire de déterminer grâce à l'analyse MS/MS du signal composite la contribution de chaque forme (PSAQ<sup>+1</sup> et protéine endogène). Pour cela, nous avons étudié et optimisé les étapes de l'analyse simultanée du PSAQ<sup>+1</sup> et de la protéine endogène.

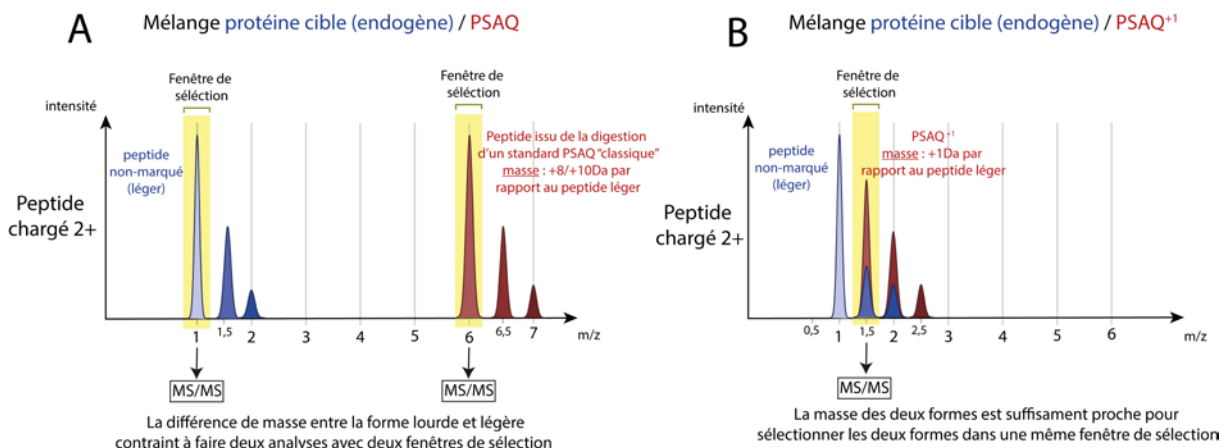


Figure 35 : Comparaison des fenêtres de sélection utilisées entre l'utilisation d'un standard protéique classique de type PSAQ et le PSAQ<sup>+1</sup>.

La protéine PSAQ<sup>+1</sup> est produite de façon à ce que les peptides tryptiques obtenus par digestion ne possèdent qu'un seul carbone ou azote isotopiquement alourdi (Figure 36). De cette façon, les propriétés du PSAQ<sup>+1</sup> restent identiques à celles du PSAQ classique, à la différence près que lors de l'analyse en MS, une seule fenêtre de sélection sera nécessaire pour sélectionner les deux espèces. Ceci permettra de réduire le temps de cycle, ce qui pourrait permettre d'augmenter le temps dédié à l'analyse des peptides et potentiellement améliorer le nombre de points par pic et/ou le signal provenant du peptide analysé. Cependant, la problématique principale qui émerge à ce moment, c'est que le pic sélectionné pour le mélange protéine cible / PSAQ<sup>+1</sup> est composite et il faut pouvoir différencier les deux parties. Pour cela, il faudra aller du côté de la MS/MS pour essayer de distinguer la partie provenant de la protéine ciblée non-marquée de la protéine PSAQ<sup>+1</sup>. Pour interpréter les spectres MS/MS nous avons été amenés à développer tout un pipeline permettant d'obtenir une quantification fiable et robuste.

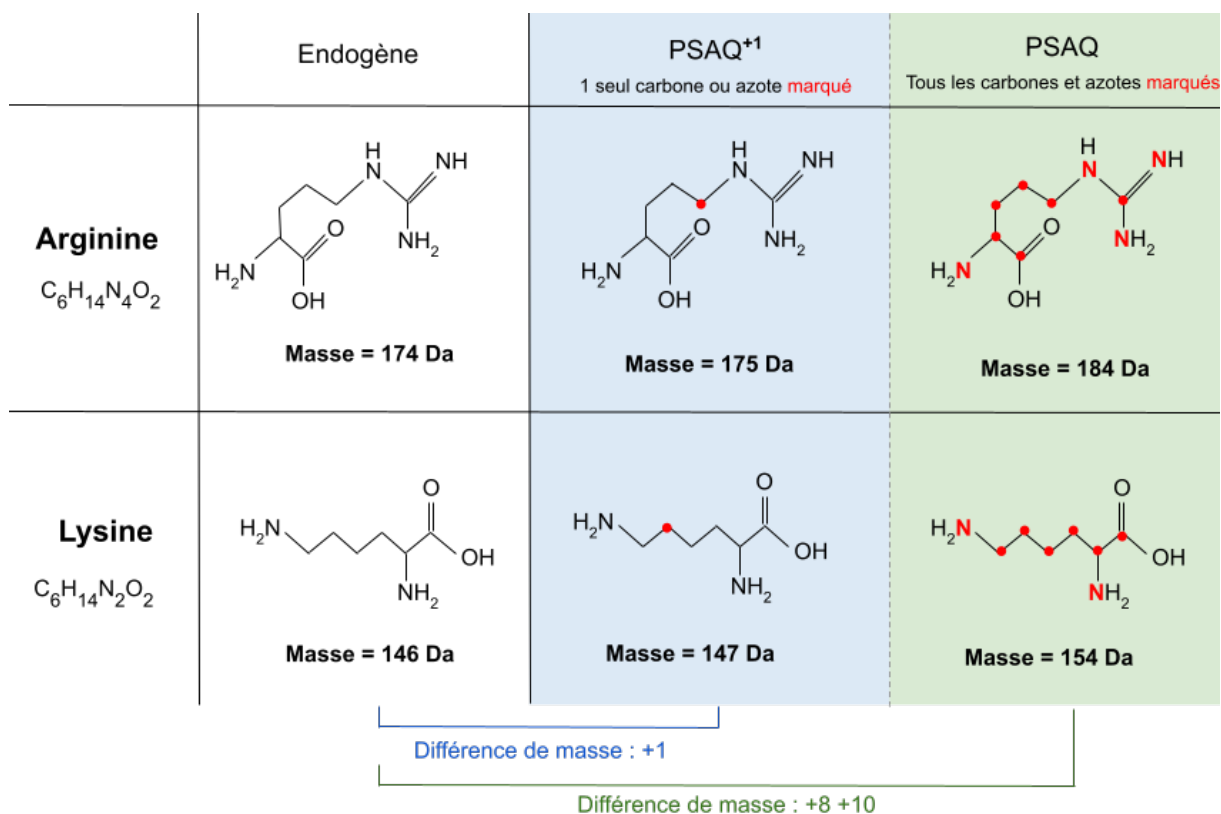


Figure 36 : Représentation de l'arginine et la lysine dans un échantillon endogène (ou non marqué), un échantillon PSAQ et un échantillon PSAQ<sup>+1</sup>. Les points en rouge illustrent la présence d'un isotope <sup>13</sup>C, les azotes en rouge schématisent quant à eux un <sup>15</sup>N.

Tout comme le PSAQ classique, le principe du PSAQ<sup>+1</sup> consiste à introduire en concentration connue la protéine marquée isotopiquement dans l'échantillon contenant la forme non-marquée à quantifier. La co-sélection comporte cependant quelques limites au niveau théorique car la présence des isotopes naturels (<sup>13</sup>C ; <sup>2</sup>H ; <sup>17</sup>O ; <sup>15</sup>N ; <sup>33</sup>S) dans la nature est à prendre en compte (Tableau 12). Effectivement, une proportion conséquente des peptides résultant de la digestion enzymatique de la protéine non-marquée ciblée seront sous forme d'isotopologues et pourront se confondre avec les peptides marqués résultant de la digestion du PSAQ<sup>+1</sup>. De même, évaluer la pureté du PSAQ<sup>+1</sup> est une étape nécessaire du fait de la possible présence d'espèce non alourdie pouvant conduire à un biais lors de la quantification de la protéine non-marquée. La quantification reposant sur cette co-sélection d'espèces nécessite d'évaluer aussi finement que possible la proportion provenant du fragment non-marqué et la proportion originarie du fragment marqué dans le spectre MS/MS.

L'objectif lors de l'introduction d'une forme marquée isotopiquement est de déterminer le ratio entre la forme marquée et non-marquée. Dans le cas du PSAQ<sup>+1</sup> et du fait du chevauchement entre les massifs isotopiques des formes marquée et non-marquée (Figure 37), il est nécessaire de pouvoir estimer la proportion d'aire provenant d'au moins une des deux formes (l'autre pouvant être déduite en soustrayant l'aire de la première à l'aire totale du pic). Dans un premier temps, la seule forme pouvant être estimée dans le pic composite est celle de la forme non-marquée lourde. Celle-ci peut être estimée à partir de l'aire sous le pic ou de l'intensité du pic provenant de la forme non-marquée légère. La proportion entre forme légère et forme lourde équivaut à la probabilité d'obtenir une forme légère ou lourde dans un échantillon analysé. Cette proportion est calculée à partir de la probabilité de présence d'un isotope donné dans la molécule. Comme montré dans le Tableau 12, chaque élément à une probabilité différente d'être sous forme alourdie. Cependant, tous les isotopes ne sont pas utiles dans ce calcul de probabilités. Les seuls pertinents sont les isotopes apportant un incrément de masse

d'un seul Dalton (en rouge Tableau 12) par rapport à la forme légère car ce sont ceux qui vont se chevaucher avec la forme marquée isotopiquement.

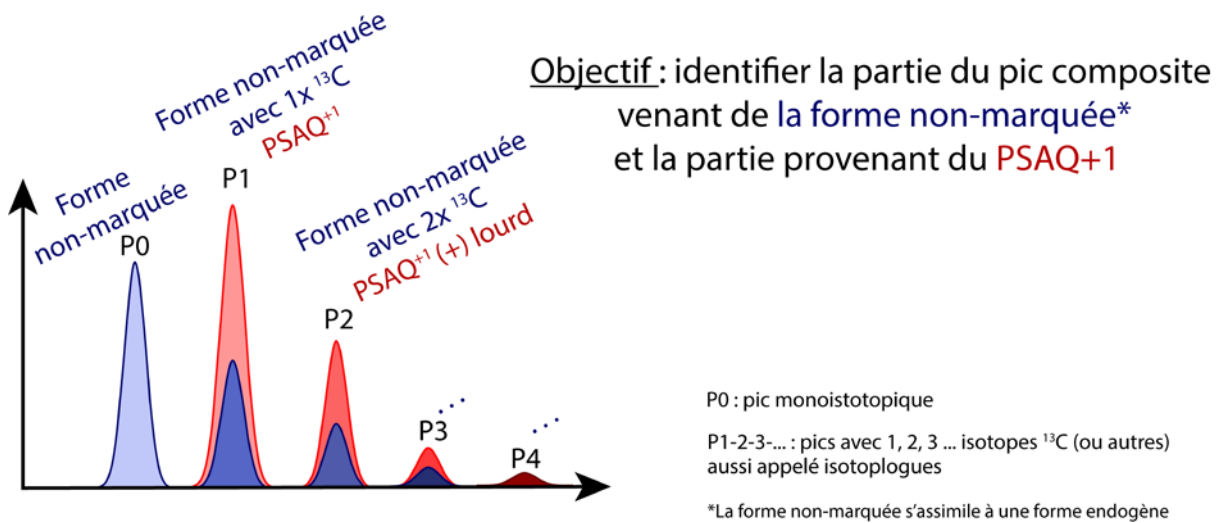


Figure 37 : Schématisation des distributions isotopiques obtenues lors de l'analyse d'un peptide non-marqué et d'un peptide PSAQ<sup>+1</sup>

Certains isotopes ont une contribution d'intensité relative plus importante tels que l' $^{18}\text{O}$  (0,19%) ou le  $^{34}\text{S}$  (4,29%). Leur abondance comparativement à la forme n'apportant qu'un incrément de 1Da ( $^{17}\text{O}$  – 0,038% et  $^{33}\text{S}$  – 0,79%) est plus élevée. Cependant ces isotopes ne peuvent pas être confondus avec le PSAQ<sup>+1</sup> si on prend en compte le pic P1 car leurs masses sont trop élevées.

Détail des calculs :

Calculer la probabilité d'obtention d'un isotope lourd, nécessite d'établir le nombre d'atomes de carbone, d'hydrogène, d'oxygène, d'azote et de soufre (CHONS) composant la molécule. Dans notre approche, le package R psinib a été utilisé pour calculer la probabilité d'avoir une forme lourde et/ou légère en fonction du nombre d'atomes de CHONS<sup>231</sup> (Figure 38). La fonction dsinib(*n\_iso*, size = *n\_espece*, prob = *proba\_isotop*) permet de réaliser ce calcul. *n\_iso* correspond au nombre d'isotopes recherchés, dans notre cas c'est 1. *Size* correspond au nombre d'atomes de CHONS et *prob* à la probabilité pour chacun des CHONS d'obtenir son isotope lourd apportant un incrément de 1Da à l'isotope stable le plus abondant.

```
# Pour installer le package
install.packages("sinib")

# A lancer une fois
library(sinib)

proba_isotop = c(0.0107, 0.000184, 0.00364, 0.000380781, 0.007948959) # Probabilité d'apparition
des isotopes apportant un incrément de 1Da à l'isotope stable le plus abondant.
n_espece = c(38, 56, 10, 11, 0) # Nombre d'atomes de chaque espèce dans le peptide LFSYGR
n_iso = 0 # Nombre d'isotopes dans la molécule
dsinib(n_iso, size = n_espece, prob = proba_isotop)
```

Figure 38 : Détails du code sur R studio permettant de calculer la proportion théorique de peptide non-marqué contenant un seul isotope à partir du nombre de chaque élément le composant et de la probabilité d'apparition de leur isotope.

En spectrométrie de masse, le détecteur est capable de fournir un signal correctement défini (5/10x supérieur au bruit de fond, forme centroïde...) sur 4 à 5 ordres de grandeur. Cela signifie que, pour une molécule donnée le nombre d'ions détectés est suffisamment important pour que les isotopes du massif isotopique soient représentatifs de la quantité d'isotopes présents dans la nature et équivalents aux probabilités théoriques d'obtention d'un isotope.

## B. Expérimentations PSAQ<sup>+1</sup>:

La protéine choisie pour réaliser les expérimentations est l'Aldolase-B humaine (Fructose-bisphosphate aldolase B, P05062) une protéine cytosolique se trouvant dans les hépatocytes (cellules du foie)<sup>232</sup>. Elle se trouve en quantités différentiellement abondantes lors d'atteintes hépatiques très avancées telles que les cirrhoses, les CHC ou les hépatites médicamenteuses<sup>178,233-236</sup>. S'agissant d'une protéine cytosolique, les taux observés sont intimement liés aux phénomènes de nécrose ou mort cellulaire qui arrivent en quantités massives lors d'atteintes hépatiques aiguës. Le laboratoire EDyP possède des échantillons de plasma de patients ayant subi des hépatites aiguës médicamenteuses, le PSAQ<sup>+1</sup> d'ALDOB pourrait donc être utilisé pour réaliser une preuve de concept en tant que biomarqueur de cette maladie (projet futur)<sup>215</sup>.

### 1. Production des protéines recombinantes ALDOB non-marquée / PSAQ / PSAQ<sup>+1</sup>:

L'ALDOB sous forme PSAQ<sup>+1</sup> a été synthétisée dans un système d'expression « cell-free » RTS 500 Proteomaster *E.Coli* HY kit utilisant des milieux contenant de la lysine et de l'arginine avec un seul <sup>13</sup>C. Pour valider notre approche d'autres isoformes ont été produites, l'une avec un marquage de tous les carbones des arginines et des lysines (marquage <sup>13</sup>C) et l'autre sans marquage afin de mimer une protéine endogène.

### 2. Injections des différentes formes de façon indépendante :

Principe de la PRM (Parallel reaction monitoring) :

Dans un premier temps, nous avons analysé l'échantillon contenant uniquement la protéine recombinante non-marquée par LC-MS/MS. L'objectif était d'évaluer le modèle permettant de calculer la probabilité d'avoir un peptide avec un isotope unique et donc présentant un incrément de masse de +1Da par rapport à la forme monoisotopique. Pour cela, de l'ALDOB dite légère a été préparée selon le pipeline classique de protéomique (dénaturation, réduction, alkylation, digestion (LysC/trypsine) et dessalage) puis a été analysée par LC-MS/MS. L'instrument utilisé est un Orbitrap Exploris (davantage de détails peuvent être retrouvés dans les Annexe I et J). L'avantage de ce type d'instrumentation est de pouvoir fonctionner en mode PRM<sup>237</sup>. Le principe de cette technique s'apparente à celui de la SRM/MRM, mais elle est plus pratique pour le développement d'essais en vue d'une quantification absolue de peptides et permet notamment d'obtenir une résolution plus élevée par rapport à un Qtrap ou un triple quadrupole. Cette méthode permet d'analyser tous les fragments provenant du peptide précurseur lors de la MS<sub>2</sub> contrairement à la SRM réalisée lors du projet NASH qui analyse uniquement les fragments préalablement sélectionnés (Figure 39). En résumé la SRM, seules trois à cinq transitions sont normalement sélectionnées, alors qu'en PRM, un spectre MS/MS complet contenant tous les ions fragments est acquis et confirmant par la même, l'identité du peptide ciblé.

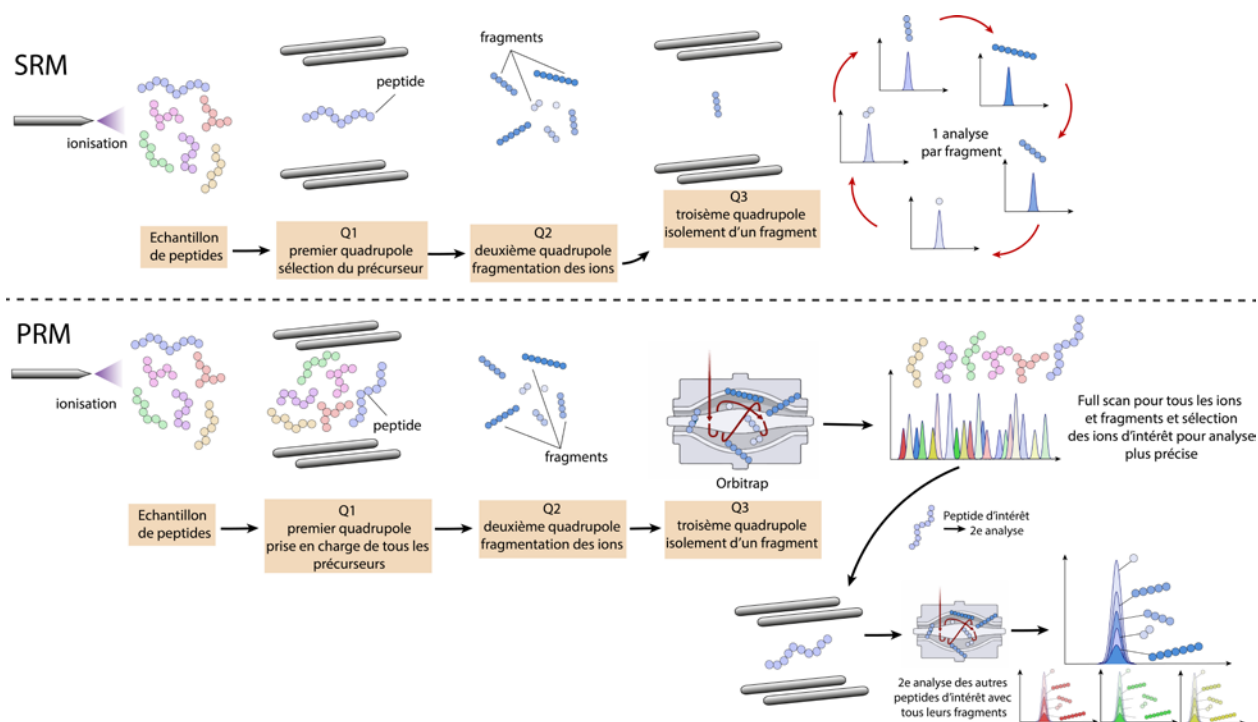


Figure 39 : Comparaison des principes d'analyse PRM et SRM

#### Pureté du PSAQ<sup>+1</sup> :

Pour évaluer le degré de pureté du PSAQ<sup>+1</sup> (proportion possédant concrètement un seul <sup>13</sup>C dans chaque peptide), une analyse LC-MS de peptides résultant de la digestion du PSAQ<sup>+1</sup> a été réalisée. Dans le Tableau 13 sont données les intensités et les aires provenant d'une forme légère de 4 peptides de l'ALDOB ainsi que d'une forme lourde correspondant au PSAQ<sup>+1</sup>. Le pourcentage de pureté est calculé en réalisant un rapport entre le signal léger (P0) et lourd (P1) (Tableau 13). Ce pourcentage est de l'ordre de 1%. Ce résultat devra être pris en compte lors des quantifications utilisant du PSAQ<sup>+1</sup> dans le calcul des ratios entre forme légère et lourde.

Tableau 13 : Pourcentage de pureté de l'échantillon PSAQ<sup>+1</sup> pour les peptides GILAADESVGTMGNR, ALQASALAAWGGK, YTPEQVAMATVTALHR, LFSYGR calculé à partir de l'intensité et l'aire sous la courbe des pics légers et lourds provenant d'un échantillon contenant uniquement du PSAQ<sup>+1</sup>.

Séquence du peptide	Forme	Intensité (coups)	Pourcentage de pureté	Aire (coups)	Pourcentage de pureté
GILAADESVGTMGNR	P0	6.0E+07	1.13%	8.6E+06	1.13%
	Lourde (1x <sup>13</sup> C)	5.3E+09		7.6E+08	
ALQASALAAWGGK	P0	1.8E+07	0.35%	3.1E+06	0.31%
	Lourde (1x <sup>13</sup> C)	5.0E+09		1.1E+09	
YTPEQVAMATVTALHR	P0	4.0E+07	1.27%	7.8E+06	1.23%
	Lourde (1x <sup>13</sup> C)	3.0E+09		6.4E+08	
LFSYGR	P0	4.0E+07	1.15%	6.9E+06	1.09%
	Lourde (1x <sup>13</sup> C)	3.0E+09		6.3E+08	

#### Fenêtre de sélection des précurseurs :

Lors de l'analyse de l'échantillon non-marqué, plusieurs paramètres ont été évalués afin de tester les limites de la méthode de quantification. L'un des paramètres était d'optimiser la largeur de la



fenêtre de sélection des précurseurs pour sélectionner les deux isotopologues P1 (PSAQ<sup>+1</sup>) et P0 dans une même fenêtre masse avec la meilleure efficacité possible. Le gain en temps d'analyse est atteint en utilisant une seule fenêtre de sélection au lieu des deux nécessaires normalement pour analyser deux espèces de masses différentes. Ce gain en temps d'analyse permet d'augmenter le temps dédié à l'acquisition de signal pour chaque précurseur et par conséquent d'augmenter le nombre de points par pic. Finalement, il est possible de faire varier la « largeur » de la fenêtre de sélection, en principe plus elle est étroite plus elle sera sélective car la quantité de molécules avec un temps d'éluion et une masse similaire à ceux de la molécule ciblée seront moins importants (Figure 40).

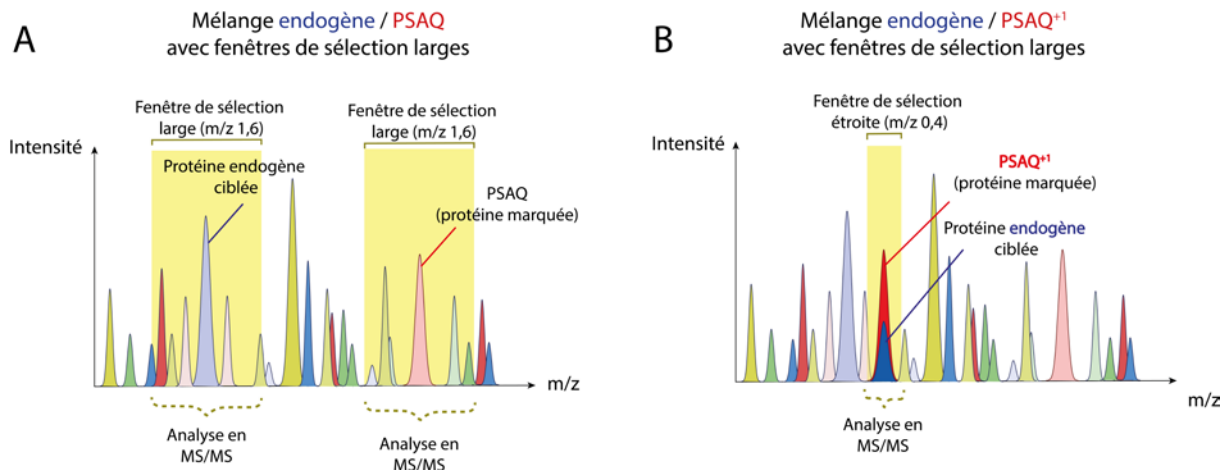


Figure 40 : Schématisation de deux mélanges endogène / PSAQ (A) et endogène / PSAQ<sup>+1</sup> (B) ainsi que des fenêtres de sélection plus ou moins importantes (de 0,4 m/z (B) à 1,6 m/z (A)) pour représenter la sélectivité et les contaminations pouvant être présentes dans chacun des modèles.

Nous avons testé trois largeurs de fenêtres de sélection de nos précurseurs chargés 2+ : 0,4 m/z ; 0,8 m/z et 1,6m/z. L'utilisation de fenêtres plus ou moins larges génère des différences notables sur la sélection des précurseurs et par conséquent sur la quantification. La fenêtre à 0,4 m/z est la fenêtre la plus étroite qui peut être atteinte avec l'instrumentation MS utilisée. En principe, il n'est pas possible d'analyser la forme non-marquée (P0) et le PSAQ<sup>+1</sup> (P1) en même temps dans une fenêtre de 0,4 m/z, mais on démontrera plus loin qu'il est possible de déduire P0 à partir des données MS/MS du pic composite en P1 (forme non-marquée lourde/PSAQ<sup>+1</sup>) (Figure 41. A). Pour ce qui est de la fenêtre à 0,8 m/z, il semble possible de co-analyser les deux espèces si la fenêtre est située à mi-chemin entre la masse de la forme non-marquée et du PSAQ<sup>+1</sup> comme montré dans la Figure 39. B. La fenêtre à 1,6 m/z permet de co-analyser 3 isotopes chargés 2+ avec 1Da de différence entre chacun en même temps (Figure 41. C).

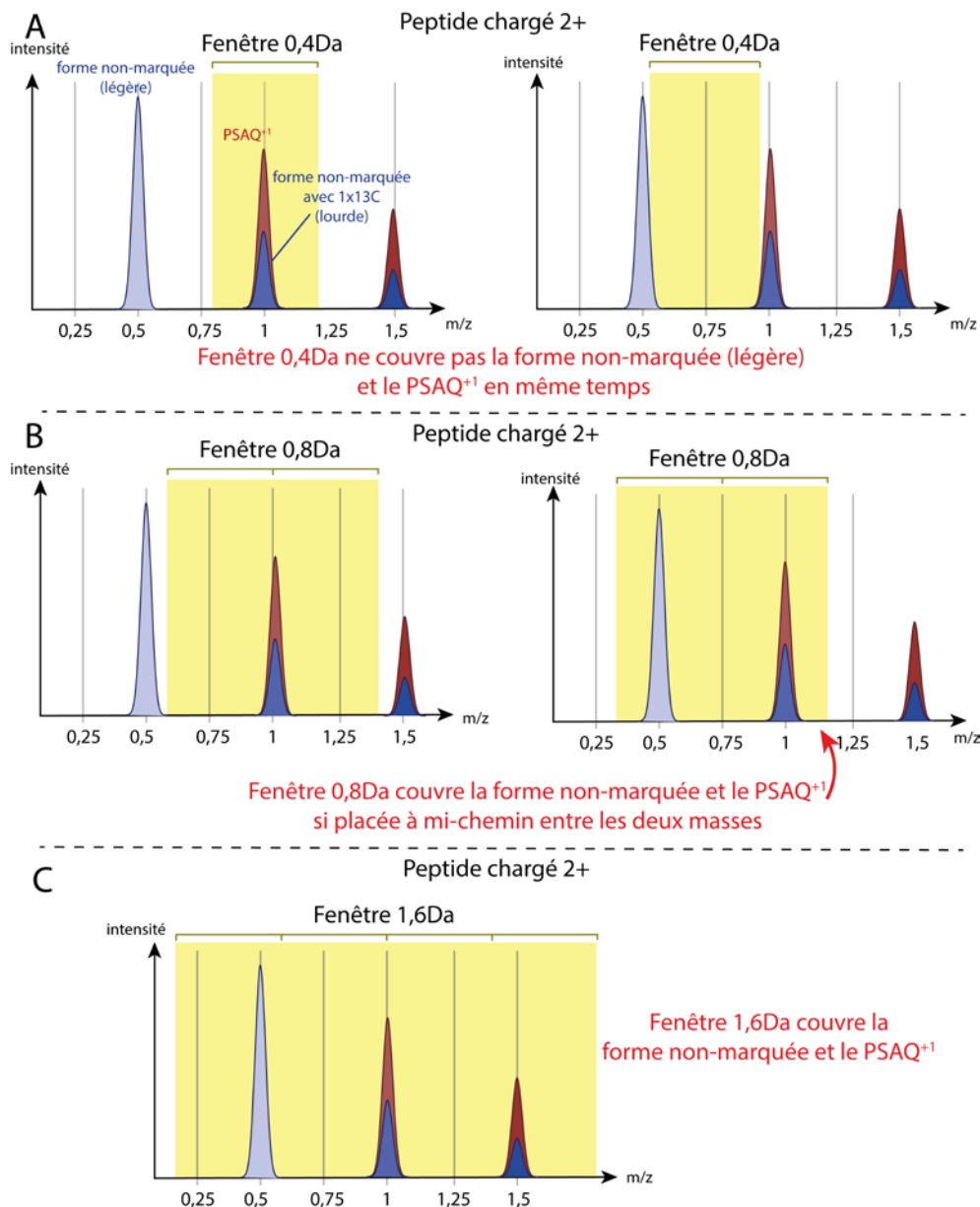


Figure 41 : Schématisation des pics inclus dans l'analyse en fonction de différentes configurations de fenêtres de sélection. Ici sont représentés les distributions isotopiques pour des peptides chargés 2+. Les différentes fenêtres de sélection représentées sont m/z 0,4 (A), 0,8 (B) et 1,6 (C).

#### Analyse des résultats (injection ALDOB non-marquée)

Les résultats analysés par la suite correspondent à une injection d'ALDOB sous forme légère pour vérifier s'il est possible de calculer l'abondance de peptides sous forme lourde provenant d'une protéine non-marquée. Le calcul de l'abondance se fait par la formule  $A / (A+B)$  où A correspond à l'abondance de forme lourde et B l'abondance de forme légère<sup>238</sup>. Cette formule permet de ne tenir compte que de ces deux formes. En effet, nous nous intéressons uniquement au précurseur non-marqué léger, au précurseur non-marqué lourd avec un incrément de masse de 1Da et au PSAQ<sup>+1</sup>. Or il existe des formes plus lourdes contenant deux isotopes (ex : 2x <sup>13</sup>C ou 1x<sup>13</sup>C/ 1x<sup>15</sup>N) qui ne vont pas rentrer dans les considérations des calculs dans un premier temps (Figure 37).

A partir des données expérimentales MS, que nous considérons l'intensité maximale ou l'aire du précurseur non-marqué léger analysé seul, l'abondance de forme alourdie en P1 peut être établie. L'abondance théorique des peptides d'ALDOB sous forme alourdie provenant de l'échantillon non-

marqué a été établie grâce au script donné en Figure 38. La Figure 42 ci-dessous montre en détail le calcul pour établir la concordance entre les modèles théorique et expérimental. Le peptide ALQASALAAWGGK est donné en exemple.

Exemple de résultats obtenus (peptide ALQASALAAWGGK) :	
Théorie :	
A = Probabilité d'obtention d'un peptide avec 1 seul isotope :	0,348
B = Probabilité d'obtention d'un peptide sans isotopes :	0,500
A/(A+B) = abondance peptide avec 1 seul isotope :	$0,348 / (0,348 + 0,500) = 0,410$
Expérimentation :	
Intensité maximale peptide avec 1 seul isotope :	1,42E+09
Intensité maximale peptide léger :	1,99E+09
A/(A+B) = abondance peptide avec 1 seul isotope :	$1,42E+09 / (1,42E+09 + 1,99E+09) = 0,416$
Différence entre les modèles théorique et expérimental : $(0,416 / 0,410) - 1 = 1,5\%$	

Figure 42 : Détail du calcul pour obtenir le pourcentage de différence entre le modèle théorique et expérimental (analyse MS<sub>1</sub>, fenêtre de sélection m/z 1,6) pour le cas du peptide ALQASALAAWGGK.

La différence entre la quantité théorique de peptide ALQASALAAWGGK lourde calculée et la quantité obtenue expérimentalement est de 1,5% d'après les données obtenues en MS<sub>1</sub> pour une fenêtre de sélection m/z 1,6.

Les valeurs des quantités théoriques et expérimentales obtenues pour tous les autres peptides lourds analysés en MS<sub>1</sub> sont données dans le Tableau 14. Les différences observées entre les valeurs théoriques et expérimentales sont généralement comprises dans une fourchette de 1-2% quelle que soit la fenêtre de sélection, à l'exception du peptide LFSYGR (4-6%) qui s'avère être également le peptide avec la séquence d'acides aminés la plus courte.

	Théorique	fenêtre m/z 0,4	% de différence	fenêtre m/z 0,8	% de différence	fenêtre m/z 1,6	% de différence
ALQASALAAWGGK	0,410	0,416	1,53%	0,408	0,40%	0,412	0,47%
ELSEIAQSI VANGK	0,435	0,433	0,43%	0,437	0,45%	0,430	1,04%
ETTIQGLDGLSER	0,420	0,424	1,01%	0,41724	0,70%	0,424	0,98%
GILAADESVGTMGNR	0,433	0,422	2,62%	0,425	1,91%	0,428	1,19%
LFSYGR	0,318	0,297	6,71%	0,302	4,94%	0,299	5,82%
YTPEQVAMATVTALHR	0,493	0,494	0,08%	0,490	0,57%	0,498	0,93%

Tableau 14 : Abondance de forme non-marquée lourde dans échantillon de forme non-marquée calculée à partir de données théoriques (abondances isotopes C, H, O, N, S) et de données expérimentales (apex intensité LC-MS/MS) et pourcentage de différence entre les deux pour des fenêtres de sélection m/z de 0,4, 0,8 et 1,6.

Le même calcul à partir des fragments issus de la MS/MS a été réalisé pour déterminer si les quantités de fragments lourds théoriques et expérimentaux correspondaient. La composition atomique des fragments y pouvant être déduite à partir de la séquence des fragments théoriques, le script utilisé précédemment pour calculer l'abondance de la forme lourde provenant de la forme non-marquée pour un fragment donné a été utilisé à nouveau. Les données MS/MS devraient permettre d'établir la proportion entre fragment non-marqué léger et le fragment non-marqué avec un  $1x^{13}C$  de la même façon qu'avec les précurseurs.

La partie A du Tableau 15 montre la moyenne de la différence entre l'abondance théorique et expérimentale pour tous les fragments de chaque peptide et cela pour les 3 fenêtres de sélection (m/z 0,4 ; 0,8 et 1,6). La partie B du Tableau 15 donne le détail obtenu pour les fragments du peptide ALQASALAAWGGK. Les résultats obtenus à partir des données MS/MS sont plus disparates que pour la MS (Tableau 15). Les meilleurs résultats sont ceux obtenus avec une fenêtre de sélection m/z de 1,6. La moyenne des différences établie entre la théorie et l'expérimental est plus basse sur cette sélection de précurseur et ceci pour tous les peptides analysés. Les différences sont comprises entre 42% (y2) et 0,75% (y9) pour la fenêtre de sélection m/z de 1,6. Les trois fragments y8-y9-y10 entourés en rouge sur la partie B du Tableau 15, donnent des différences entre théorie et expérience de l'ordre de 1%.

A				B						
Données MS2										
	0,4Da	0,8Da	1,6Da	seq	fragment	Théorique	léger	lourd	Expérimental	% de différence
ALQASALAAWGGK	Intensity	Intensity	Intensity	ALQASALAAWGGK	y1	0,070	3,02E+06	3,51E+05	0,104	48,20%
	90,62%	36,01%	12,37%	ALQASALAAWGGK	y2	0,093	5,94E+06	7,55E+05	0,113	21,24%
	area	area	area	ALQASALAAWGGK	y3	0,115	1,91E+07	4,01E+06	0,174	51,72%
	90,96%	45,63%	10,54%	ALQASALAAWGGK	y4	0,205	2,36E+07	1,35E+07	0,364	77,02%
ELSEIAQSIVANGK	Intensity	Intensity	Intensity	ALQASALAAWGGK	y5	0,229	1,79E+07	1,25E+07	0,412	80,21%
	80,30%	42,20%	13,73%	ALQASALAAWGGK	y6	0,251	2,08E+07	1,98E+07	0,488	94,46%
	area	area	area	ALQASALAAWGGK	y7	0,289	1,31E+07	1,76E+07	0,573	98,28%
	84,97%	59,42%	21,86%	ALQASALAAWGGK	y8	0,308	8,36E+06	1,54E+07	0,648	110,54%
ETTIQGLDGLSER	Intensity	Intensity	Intensity	ALQASALAAWGGK	y9	0,326	1,70E+07	3,71E+07	0,688	111,22%
	75,92%	24,44%	10,82%	ALQASALAAWGGK	y10	0,343	1,54E+07	4,11E+07	0,727	112,31%
	area	area	area	ALQASALAAWGGK	y11	0,369	2,21E+06	1,16E+07	0,840	127,58%
	82,12%	36,92%	18,03%	ALQASALAAWGGK	y12	0,393	0,00E+00	1,90E+05	1,000	154,69%
GILAADESVMGNTMGNR	Intensity	Intensity	Intensity	ALQASALAAWGGK	y1	0,070	8,80E+06	6,07E+05	0,065	8,03%
	71,26%	12,04%	20,56%	ALQASALAAWGGK	y2	0,093	1,65E+07	1,47E+06	0,082	12,01%
	area	area	area	ALQASALAAWGGK	y3	0,115	5,64E+07	6,15E+06	0,098	14,29%
	75,51%	19,84%	14,81%	ALQASALAAWGGK	y4	0,205	7,76E+07	2,33E+07	0,231	12,56%
LSFSYGR	Intensity	Intensity	Intensity	ALQASALAAWGGK	y5	0,229	5,84E+07	2,11E+07	0,266	16,04%
	144,91%	113,50%	10,43%	ALQASALAAWGGK	y6	0,251	7,79E+07	3,41E+07	0,305	21,43%
	area	area	area	ALQASALAAWGGK	y7	0,289	5,26E+07	3,26E+07	0,382	32,35%
	149,92%	122,11%	11,56%	ALQASALAAWGGK	y8	0,308	4,00E+07	2,70E+07	0,403	31,00%
YTPEQVAMATVTALHR	Intensity	Intensity	Intensity	ALQASALAAWGGK	y9	0,326	8,33E+07	6,94E+07	0,455	39,62%
	34,64%	16,45%	9,09%	ALQASALAAWGGK	y10	0,343	7,99E+07	7,70E+07	0,491	43,31%
	area	area	area	ALQASALAAWGGK	y11	0,369	1,82E+07	2,15E+07	0,542	46,77%
	33,90%	15,83%	7,61%	ALQASALAAWGGK	y12	0,393	0,00E+00	3,34E+05	1,000	154,69%

Tableau 15 : Pourcentage de différence obtenu entre abondance théorique et expérimentale (intensité et aire) sur 3 fenêtres de m/z (0,4-0,8-1,6) pour chaque peptide (A) et pour chaque fragment du peptide ALQASALAAWGGK (intensité) (B).

Certains fragments donnent des résultats expérimentaux très proches de la théorie. Il est important de déterminer quels sont les paramètres ou caractéristiques des fragments permettant d'obtenir ces résultats. La comparaison des résultats obtenus sur la base des aires ou des intensités maximales, montre qu'en général, l'intensité donne de meilleurs résultats. Une explication, à cette observation, serait que les bornes qui définissent le pic ne sont pas assez serrées et comprennent donc une étendue trop large en dehors du pic. Les bornes sont définies de façon automatique par le logiciel de traitement du signal MS, elles peuvent être définies manuellement pour obtenir de meilleurs résultats.

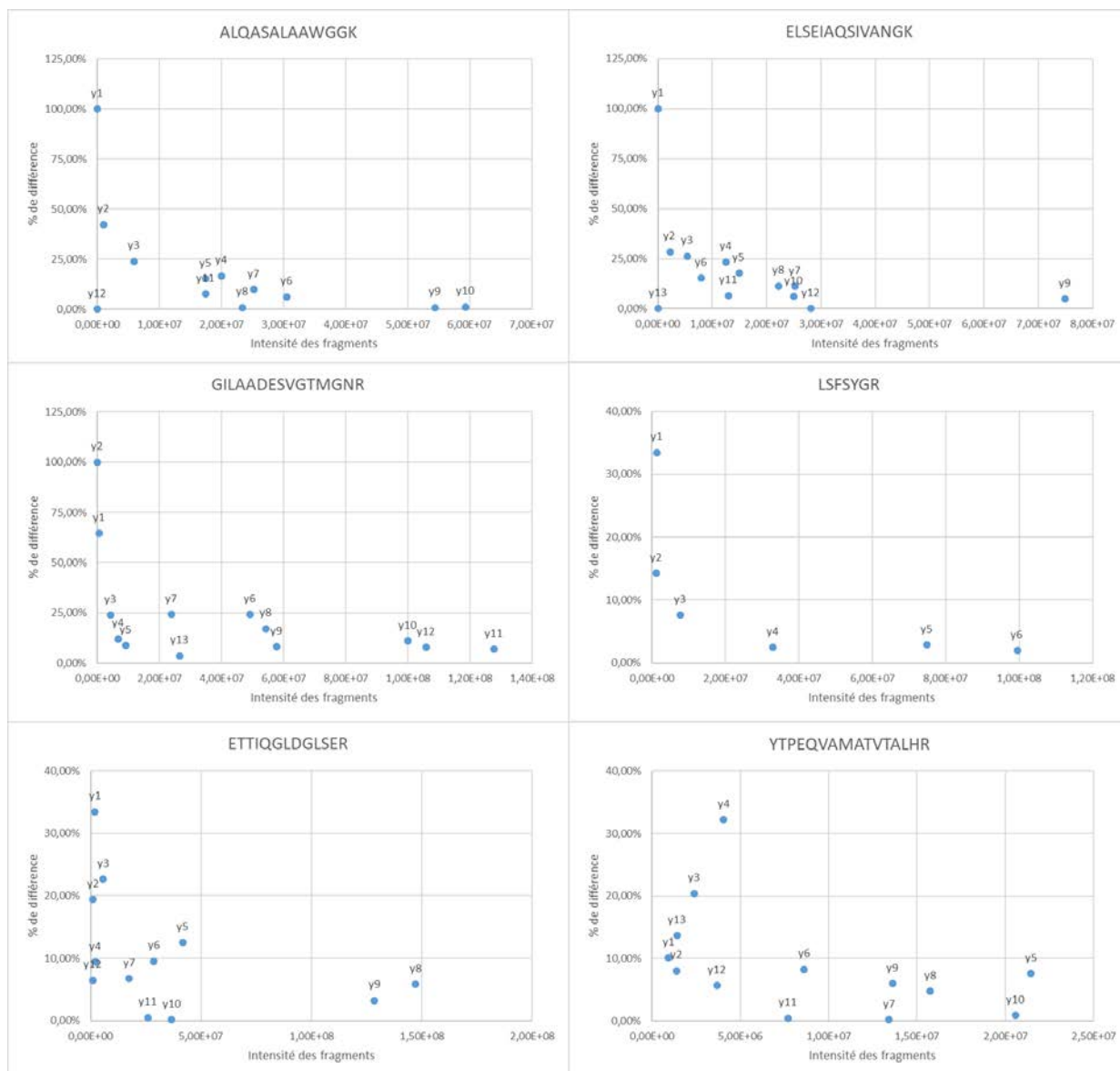


Figure 43 : Intensité de chaque fragment des 6 peptides suivis pour l'ALDOB par rapport à son pourcentage de différence entre quantité théorique et expérimentale pour une fenêtre de sélection du précurseur de m/z 1,6.

La Figure 43, présentant la comparaison des quantités théorique et expérimentale (intensité) de chaque fragment de 6 peptides de l'ALDOB, montre que de manière générale, les rapports entre forme marquée et non-marquée sont plus prévisibles pour les fragments présentant une grande intensité. Une tendance semble également se dessiner quant à la longueur des fragments car la différence entre le modèle théorique et expérimental est d'autant plus faible que le peptide est long. L'exemple le plus clair correspondant à cette explication est celui du peptide LSFSFYG, le pourcentage de différence entre le modèle théorique et expérimental diminue avec la longueur du fragment. Une raison plausible expliquant ce phénomène résiderait dans le fait que des contaminations avec des masses similaires à celle du fragment sont beaucoup plus probables pour des petits fragments que pour des plus grands fragments<sup>126</sup>. Une molécule présentant des caractéristiques similaires à celle du fragment (contamination) peut mener à l'obtention d'un signal plus élevé que ce qui est attendu.

Une telle disparité dans les résultats pourrait aussi s'expliquer par un bruit chimique important. En premier lieu, pour minimiser l'effet observé, le bruit sous chaque fragment a été estimé en utilisant le logiciel Excalibur. En retranchant l'intensité liée au bruit à l'intensité totale, les différences observées

sont significativement abaissées entre les deux conditions (Tableau 16 et Annexe K). Cependant les différences entre l'abondance théorique et expérimentale restent élevées, cette piste pourra donc être retenue pour optimiser les résultats mais des corrections majeures sont à prévoir. Deux pistes d'amélioration seront évaluées par la suite : (1) une correction du calcul théorique pour la MS/MS et (2) une correction due à un phénomène expérimental non tenu en compte jusqu'à cet instant.

*Tableau 16 : Comparaison de deux conditions lors du calcul du pourcentage de différence entre abondance théorique et expérimentale pour la moyenne des fragments de chaque peptide de l'ALDOB. Condition 1 : pas de retranchement du bruit. Condition 2 : retranchement du bruit.*

Peptide (moyenne des fragments)	Condition 1 : % de différence sans retrancher le bruit	Condition 2 : % de différence en retranchant le bruit
ALQASALAAWGGK	15,05%	12.37%
ELSEIAQSIVANGK	24.36%	20.92%
ETTIQGLDGLSER	19.45%	10.82%
GILAADESVGTMGNR	36.58%	24.06%
LSFSYGR	19.45%	10.43%
YTPEQVAMATVTALHR	16.69%	9.09%

### 3. Corrections modèle théorique :

Fenêtre de sélection des précurseurs de m/z de 1,6 :

En première approche, le modèle mathématique théorique utilisé pour l'extrapolation des massifs MS et MS/MS, est identique. Cependant, nos résultats indiquent qu'il n'est pas ajusté aux données MS/MS, contrairement aux données acquises en MS. Un modèle prenant en compte les spécificités de la MS/MS est donc indispensable. Prenons l'exemple de la Figure 44 pour laquelle la forme P0 et P1 correspondant à la combinaison des formes légère avec  $1x^{13}C$  et  $PSAQ^{+1}$  d'un précurseur doublement chargé sont ciblées avec une fenêtre de sélection m/z de 1,6 en MS centrée sur le m/z du peptide non-marqué (Figure 44. A).

Après fragmentation, plusieurs origines de fragments y sont à prendre en compte (Figure 44.B) :

- (1) des fragments y légers provenant du précurseur **non-marqué léger**.
- (2) des fragments y légers provenant du précurseur **non-marqué lourd**.
- (3) des fragments y lourds provenant du précurseur **non-marqué lourd**.
- (4) des fragments y lourds provenant de **PSAQ<sup>+1</sup>**.

Lorsque ces fragments y sont analysés en MS/MS, deux pics composés des types de fragments y et de précurseurs isotopologues différents seront observés (Figure 44. C). Le défi consiste à estimer la proportion des fragments (1) et (2) d'un côté et la part des fragments (3) et (4) de l'autre côté.

Analyse PRM  
(Orbitrap)

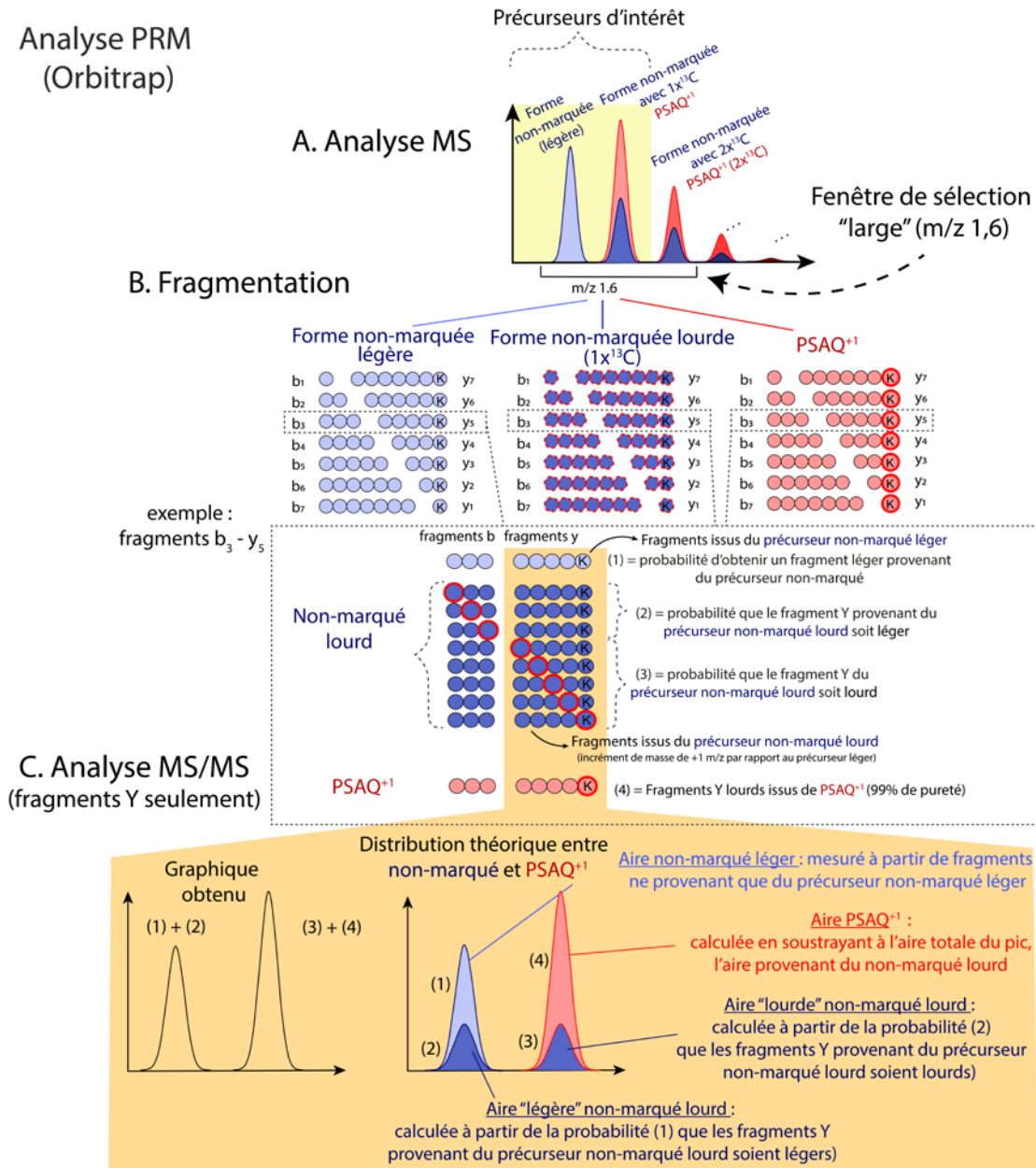


Figure 44 : Schéma du pipeline de quantification par PSAQ+1 lors de l'utilisation d'une fenêtre de sélection m/z de 1,6

Pour corriger le modèle théorique, il faut donc calculer un ratio entre précurseur non-marqué et PSAQ $^{+1}$  tenant compte du fait qu'une partie de la forme non-marquée sera sous forme lourde. Le calcul de ce ratio varie si l'on prend les données MS (Figure 44. A) plutôt que les données MS/MS (Figure 44.B). La différence réside dans le fait que pour les données MS/MS, une partie des fragments y provenant de la forme non-marquée lourde seront sous forme légère car la partie lourde se trouvera sur les fragments « b ».

La Figure 44 n'est pas exhaustive du fait qu'elle prend en compte dans la superposition uniquement les massifs isotopiques correspondant à la forme non-marquée puis à la combinaison non-marquée avec  $1 \times ^{13}\text{C}$  et PSAQ $^{+1}$ . Or il ne faut pas négliger, que lors de l'utilisation d'une fenêtre de sélection de m/z 1,6, des précurseurs avec 2 isotopes ou plus peuvent être sélectionnés et donner des fragments avec des masses similaires à celles des fragments (1-2) et (3-4) (Figure 45). Leurs probabilités respectives peuvent être calculées par une loi hypergéométrique. Il s'agit d'une loi de probabilités décrivant un modèle mathématique utilisé pour calculer la probabilité de sélectionner un certain

nombre d'éléments d'une population sans remise. Cette loi permet, par exemple, de modéliser la proportion de fragments légers provenant d'un précurseur non-marqué avec  $1 \times {}^{13}\text{C}$ . Comme montré à la Figure 45, les fragments légers peuvent également provenir de précurseurs non-marqués plus lourds ( $2 \times {}^{13}\text{C}$ ;  $3 \times {}^{13}\text{C}$ ,  $1 \times {}^{15}\text{N}$ ) faisant varier la proportion de fragments légers par rapport aux fragments lourds, tous ces événements peuvent être calculés avec la loi hypergéométrique.

Bien qu'il existe d'autres isotopes que le  ${}^{13}\text{C}$  dans les acides aminés (Tableau 11), pour la simplicité de l'argumentation et des calculs, seul le carbone sera considéré dans l'explication qui suit. Le choix du carbone se justifie également car sa probabilité de fournir un isotope apportant 1Da par rapport à la forme légère est significativement plus élevée que pour les autres atomes (HONS). La proportion de fragments légers provenant de la forme non-marquée avec  $1 \times {}^{13}\text{C}$  peut être exprimée comme la probabilité qu'il n'y ait pas de  ${}^{13}\text{C}$  dans un fragment provenant d'un peptide avec 1 seul  ${}^{13}\text{C}$ . Par ailleurs, la loi hypergéométrique permet d'établir la probabilité d'avoir un fragment avec 1 seul  ${}^{13}\text{C}$  provenant d'un peptide avec 2, 3 ou n  ${}^{13}\text{C}$ . Elle permet aussi de calculer la probabilité que le fragment soit composé entièrement de  ${}^{12}\text{C}$  (léger) alors qu'il provient d'un peptide avec 1 ou plusieurs  ${}^{13}\text{C}$ .

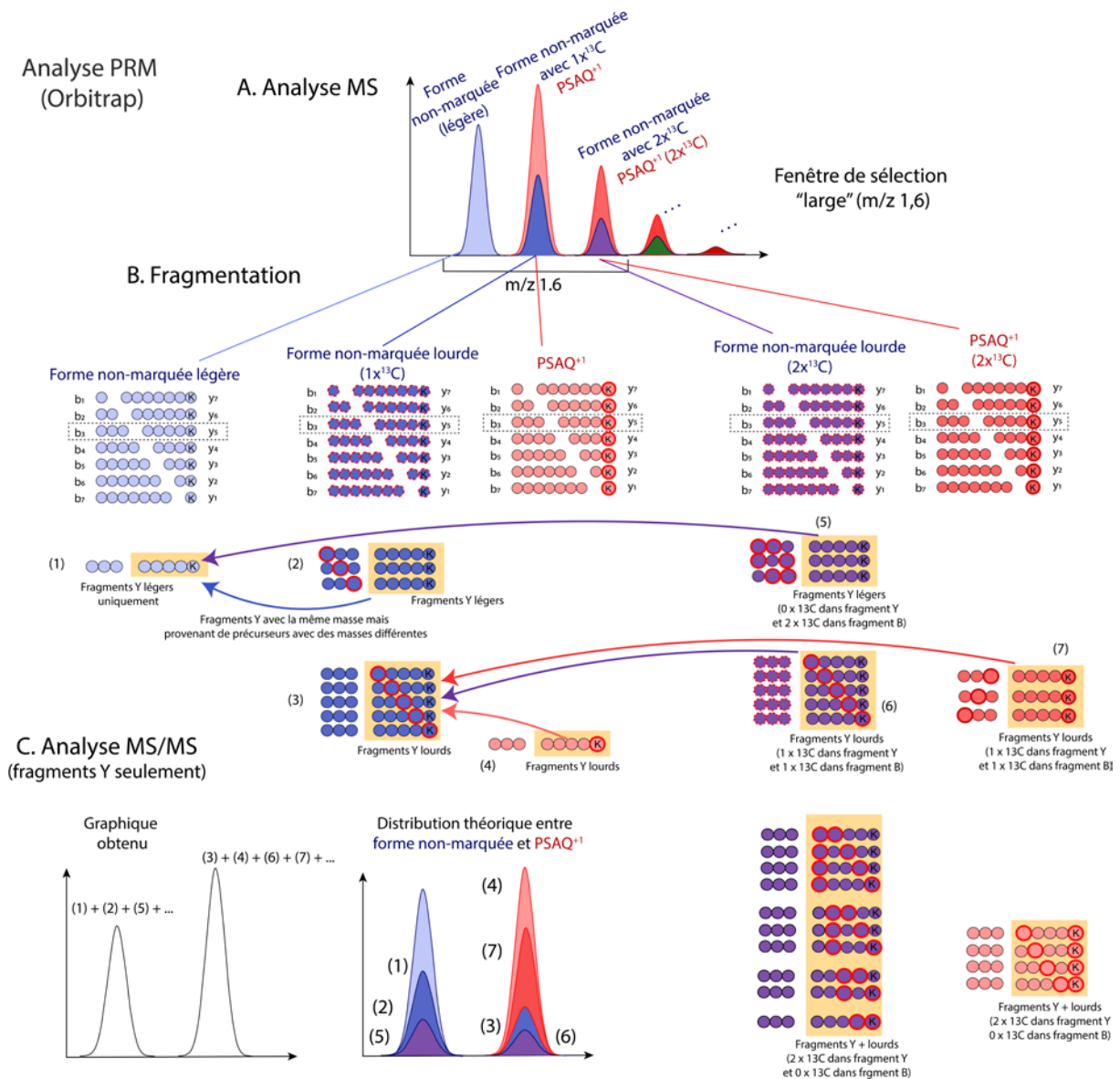


Figure 45 : Schéma du pipeline de quantification par PSAQ+1 lors de l'utilisation d'une fenêtre de sélection m/z de 1,6 tenant en compte un massif isotopique supplémentaire par rapport à la Figure 44



Une fenêtre de sélection du précurseur de  $m/z$  1,6 ciblera au moins 3 pics (ou isotopes) dans un massif isotopique. Il faut donc tenir compte de la participation de chaque isotope à l'obtention de chaque type de fragment « y ». L'intensité ou l'aire de l'isotope en MS déterminera l'importance de son apport en différents fragments pour la MS/MS. La proportion de fragment léger en MS/MS peut être modélisée par le calcul proposé en Figure 46.

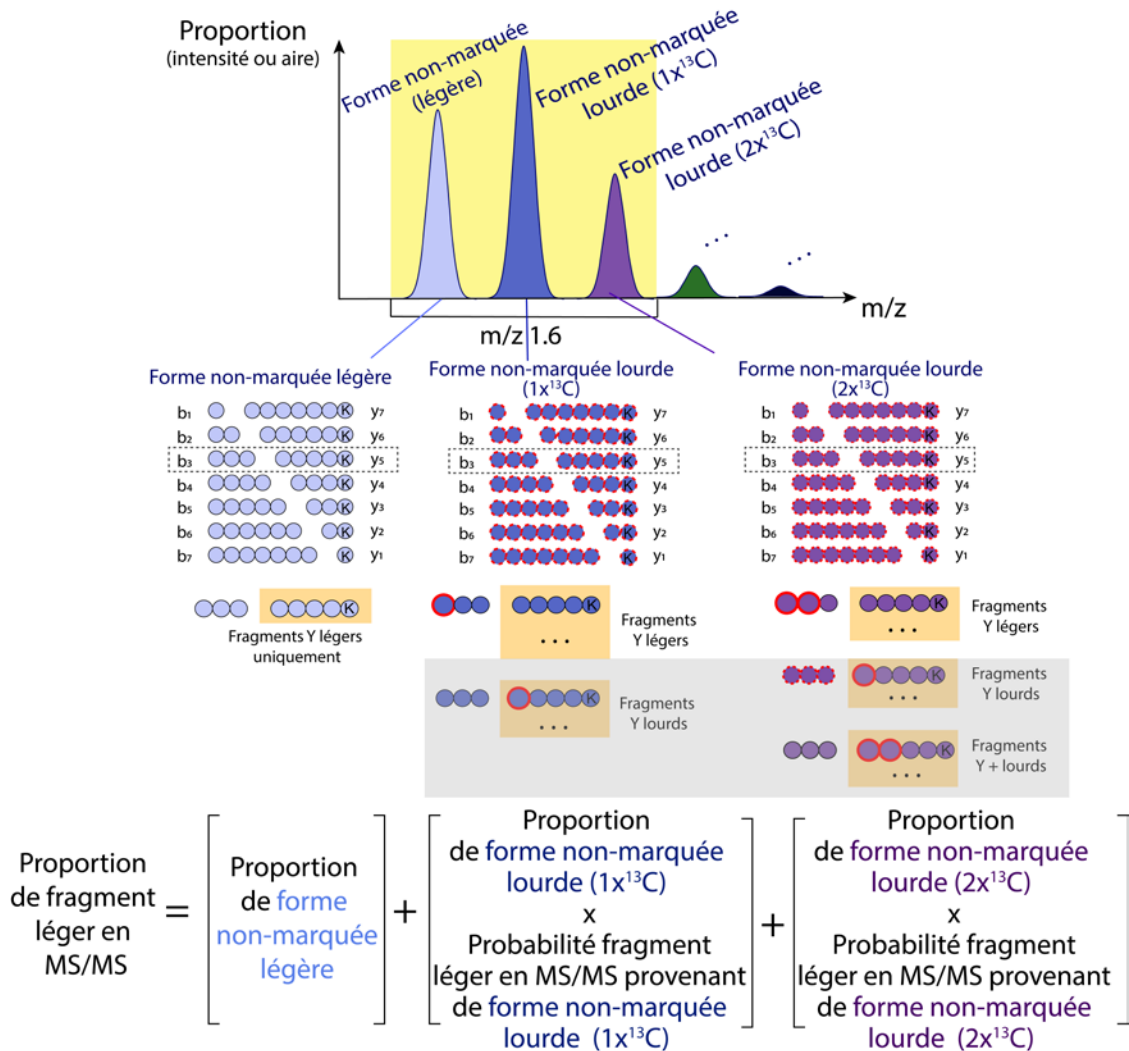


Figure 46 : Détails de la formule pour calculer la quantité de fragments légers venant de précurseurs non-marqués avec différentes quantités de  $^{13}\text{C}$ .

Un calcul similaire pour estimer la proportion de fragments y alourdis avec un atome  $^{13}\text{C}$  peut être fait. La différence avec le calcul précédent reposera sur le fait que la probabilité d'obtenir ce fragment sera différente en fonction de l'isotope pris en compte. A partir des proportions de fragments légers et lourds provenant de la forme non-marquée il est possible d'isoler les parties du pic composite P1 formé par de la forme non-marquée avec  $1 \times ^{13}\text{C}$  et du PSAQ $^{+1}$ . Avant de démontrer la faisabilité de ce calcul, nous allons nous pencher sur des corrections des modèles expérimental et théorique pour maximiser les probabilités d'obtenir des résultats cohérents.

Fenêtre de sélection  $m/z$  0,4 :

Sachant que la plus petite largeur pour une fenêtre de sélection pouvant être utilisée avec l'instrumentation MS disponible (Orbitrap Exploris) est de  $m/z$  0,4, nous avons essayé de vérifier dans ces conditions s'il était possible de séparer le signal provenant de l'espèce PSAQ $^{+1}$  de celui provenant de la forme non-marquée avec  $1 \times ^{13}\text{C}$  dans le pic composite. Lorsque les ions chargés 2+ sont analysés

avec une fenêtre aussi restreinte, il ne sera possible de cibler qu'un seul pic, en l'occurrence P1. Le deuxième isotope peut fournir suffisamment d'informations pour différencier le précurseur non-marqué avec  $1x^{13}C$  du PSAQ<sup>+1</sup>. Pour cela l'interprétation des données MS/MS est nécessaire. En effet, comme le montre la Figure 47, lorsque le deuxième pic du massif isotopique (P1) est ciblé, des fragments y légers et lourds provenant de la forme non-marquée lourde sont produits ainsi que des fragments provenant du PSAQ<sup>+1</sup>. A partir de l'intensité ou l'aire obtenue pour les fragments légers et grâce à la loi hypergéométrique, il sera possible :

- de calculer la proportion de fragment lourd issue du précurseur non-marqué avec  $1x^{13}C$ .
- d'extrapoler l'intensité ou l'aire du fragment lourd provenant du précurseur lourd
- de déduire la proportion de signal provenant du PSAQ<sup>+1</sup> en soustrayant à l'intensité ou à l'aire maximale, l'intensité ou l'aire du fragment lourd issue du précurseur non-marqué avec  $1x^{13}C$  (Figure 47.C).

L'avantage d'utiliser une fenêtre de sélection plus étroite est que la co-sélection de contaminations ou de molécules avec des temps de rétention et des masses similaires à celle de la molécule d'intérêt est limitée.

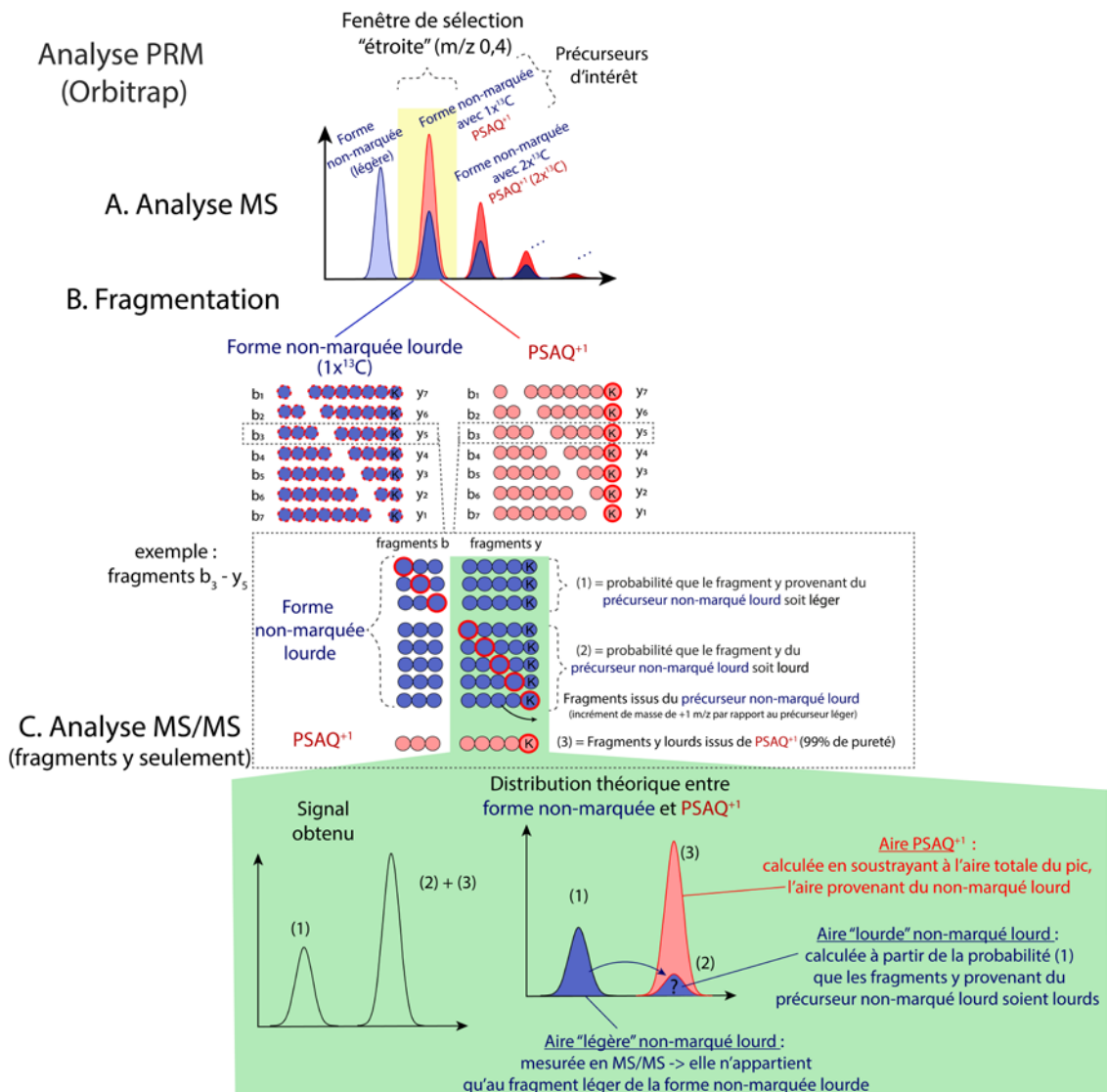


Figure 47 : Schéma du pipeline de quantification par PSAQ<sup>+1</sup> lors de l'utilisation d'une fenêtre de sélection m/z de 0,4

Un autre avantage inhérent à l'utilisation d'une fenêtre de sélection étroite est que contrairement à une fenêtre de sélection large (Figure 44), seuls les fragments provenant du précurseur d'intérêt seront présents. Par exemple, il n'y aura pas de fragment avec un  $^{13}\text{C}$  provenant d'un précurseur non-marqué lourd avec  $2\times^{13}\text{C}$ . Ceci simplifie les calculs car la proportion de fragments provenant de précurseurs « non-souhaités » ne doit plus être établie.

#### Application des corrections aux résultats MS/MS :

Dans un premier temps, les corrections ont été réalisées pour une fenêtre de sélection de m/z 0,4. Comme expliqué auparavant, les corrections consistaient essentiellement à tenir compte des fragments y légers provenant du précurseur non-marqué avec  $1\times^{13}\text{C}$ . Pour des fenêtres de sélection incluant des pics isotopiques plus lourds, le calcul a pris en compte la possibilité que des précurseurs avec 2 atomes  $^{13}\text{C}$  ou plus apportent aussi des fragments y légers. Le ratio entre précurseur non-marqué et  $\text{PSAQ}^{+1}$  a été calculé à partir de chacun des fragments des peptides suivis. La Figure 48.A correspond au ratio théorique sans correction en fonction du ratio expérimental pour une fenêtre de sélection m/z 0,4. Sans correction, le  $R^2$ , indicateur de qualité de la prédiction d'une régression linéaire, est de 0,83. La Figure 48.B correspond au ratio théorique corrigé en fonction du ratio expérimental de l'intensité et montre une amélioration du  $R^2$  (0,99) pour le même jeu de données. Les p-values ( $< 2,2\times 10^{-16}$ ) associées au  $R^2$  montrent que la confiance que l'on peut avoir sur ces mesures est élevée. Que ce soit visuellement ou par le moyen d'une régression linéaire, la Figure 48 montre que les corrections implémentées nous permettent d'obtenir des résultats théoriques et expérimentaux sensiblement plus proches pour une fenêtre de sélection de m/z 0,4 ; et ceci pour l'ensemble des fragments considérés.

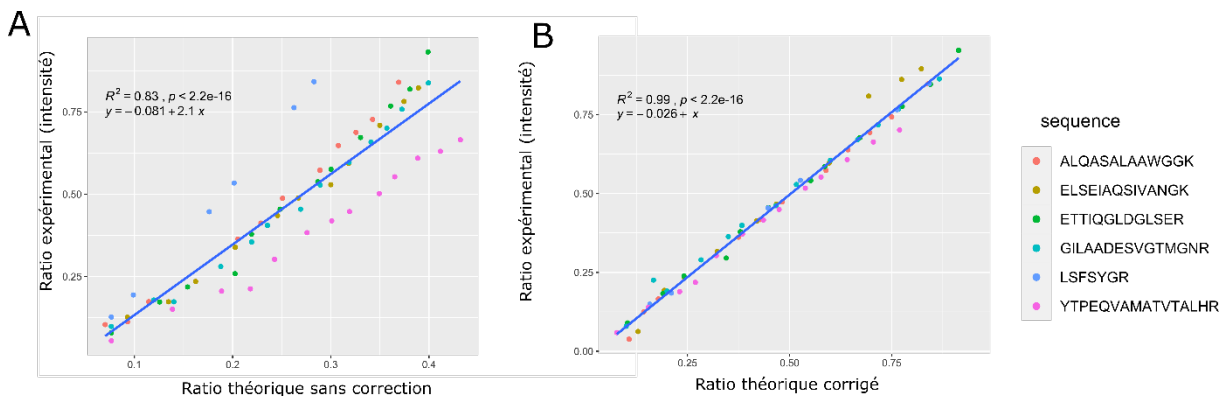


Figure 48 : Ratios théoriques en fonction des ratios expérimentaux de tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection m/z 0,4. A. Ratios théoriques non corrigés en fonction des ratios expérimentaux de l'intensité. B. Ratios théoriques corrigés en fonction des ratios expérimentaux de l'intensité.

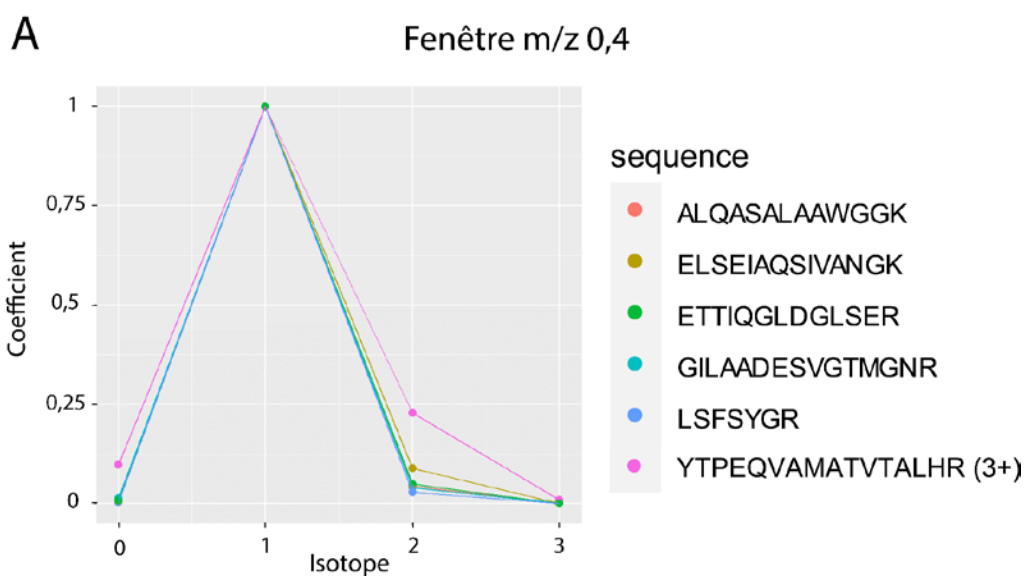
Les fenêtres de m/z 0,8 et 1,6 ont également été corrigées, des résultats complémentaires pour les 3 fenêtres peuvent être retrouvés en Annexe L, M et N. La correction s'est avérée moins performante sur les fenêtres plus larges, mais l'erreur semble stable dans le sens où tous les ratios des fragments corrigés ont un pourcentage d'erreur entre ratio expérimental et théorique similaire. Les résultats de la Figure 48 montrent que la correction du calcul théorique pour l'analyse de fragments permet d'obtenir des ratios significativement plus proches de ceux obtenus expérimentalement, notamment sur la fenêtre de sélection de m/z 0,4. Des corrections supplémentaires pourraient permettre d'affiner les résultats, par exemple en supprimant la partie du signal considéré comme du bruit comme cela a été montré dans le Tableau 16. Dans ce sens, nous avons démontré sur des données  $\text{MS}_1$ , que dans la plupart des cas, la prise en compte de tous les types d'atomes présentait des différences entre la théorie et l'expérience plus faibles et donc des résultats

plus précis sur la plupart des peptides (Annexe O). Cependant, à ce stade, les résultats obtenus par la MS/MS pour une fenêtre de sélection de  $m/z$  0,4 sont suffisamment bons pour réaliser une quantification de qualité tout en gardant en tête que celui-ci pourrait être optimisé. Nous allons tout de même nous pencher sur le modèle expérimental pour essayer de réduire cette différence sur toutes les conditions de fenêtres de sélection confondues.

#### 4. Corrections modèle expérimental :

Des différences importantes dans les résultats ont été observées entre les différentes fenêtres de sélection  $m/z$  0,4 ; 0,8 et 1,6 (Tableau 14 et Tableau 15). Nous allons d'abord essayer de vérifier si la largeur « numérique » des fenêtres de sélection correspond à la largeur réelle de celles-ci. Pour cela, un échantillon contenant uniquement du précurseur non-marqué a été analysé sans utiliser d'énergie de collision avec des fenêtres de sélection de  $m/z$  de 0,4 ; 0,8 et 1,6 de largeur centrées sur le pic composite (non-marqué avec  $1x^{13}C + PSAQ^{+1}$ ). Pour rappel, l'énergie de collision est ce qui permet aux peptides ayant passé le premier quadrupole d'être fragmentés avant de passer au 3<sup>e</sup> quadrupole pour l'analyse en MS/MS. L'idée est de vérifier l'absence de fragmentation avant la MS/MS. L'effet attendu est que l'intensité et les aires des pics observés en MS soient équivalents à ceux observés MS/MS. En principe, une fenêtre de sélection de  $m/z$  1,6 centrée sur le 1<sup>er</sup> isotope devrait permettre de voir les isotopes P0, P1 et P2 pour des espèces doublement chargées (2+) sans perte de signal (voir Figure 41.C). De même, une fenêtre de sélection de  $m/z$  0,4 centrée sur un pic en particulier devrait uniquement inclure le pic ciblé lorsqu'on a affaire à une molécule chargée 2+ (Figure 41.A).

Cependant, la Figure 49 montre des résultats s'opposant à ce principe. En effet, ici sont présentés les résultats obtenus en MS/MS lorsqu'une fenêtre de sélection de  $m/z$  0,4 est centrée sur le pic composite P1 pour chacun des peptides. En théorie, il s'agit d'une fenêtre de sélection suffisamment étroite pour ne détecter que l'isotope ciblé lorsqu'on a affaire à une molécule doublement chargée (2+). Cependant, même si la plupart du signal provient du pic isotopique ciblé, un signal parasite appartenant aux pics isotopiques voisins est également observé. Ainsi, d'un côté les pics isotopiques sans isotopes alourdis renvoient en moyenne du signal correspondant à 0,6% du signal des pics isotopiques ciblés et de l'autre côté, les pics isotopiques avec deux ou trois isotopes alourdis renvoient en moyenne du signal correspondant respectivement à 5% et 0,04% du pic isotopique ciblé. Il est important de noter que le peptide YTPEQVAMATVTALHR est celui pour lequel on détecte le plus de signal pour les pics isotopiques voisins de celui ciblé. La particularité de ce peptide c'est qu'il est chargé 3+ et donc la répartition des rapports  $m/z$  de ces différents pics isotopiques est plus rapprochée que pour les peptides chargés 2+. Ceci dit, les pics P0 et P2 d'une molécule chargée 3+, ne sont pas inclus dans la fenêtre de sélection  $m/z$  0,4 si elle est centrée sur P1. Ce résultat peut laisser penser que la fenêtre de sélection de  $m/z$  0,4 n'est pas suffisamment bien définie et que ses bornes sont plus larges qu'annoncé. Pour illustrer ce propos, on peut dire que la fenêtre de sélection n'est pas « carrée », elle aurait plutôt une forme s'apparentant à celle d'une courbe gaussienne dont l'apex se situerait sur le  $m/z$  ciblé (Figure 50).



B		C	
Isotope	Coefficient moyen	Isotope	Coefficient moyen
$^{12}\text{C}$ full	0.0067	$^{12}\text{C}$ full	0,0973
1x $^{13}\text{C}$	1.0000	1x $^{13}\text{C}$	1,0000
2x $^{13}\text{C}$	0.0503	2x $^{13}\text{C}$	0,2281
3x $^{13}\text{C}$	0.0004	3x $^{13}\text{C}$	0,0090

Figure 49 : Comparaison des intensités obtenues en MS/MS lors d'une analyse de peptides de l'ALDOB non-marquée avec une fenêtre de sélection m/z 0,4 ciblée sur le P1 du massif isotopique. (A) Représentation graphique de la moyenne des intensités obtenues pour les fragments de tous les pics isotopiques de chaque peptide. (B) Tableau des moyennes des intensités obtenues pour les fragments de tous les pics isotopiques de chaque peptide chargé 2+ et coefficients par rapport au pic le plus intense. (C) Tableau des moyennes des intensités obtenues pour les fragments des pics isotopiques du peptide chargé 3+ (YTPEQVAMATVTALHR) et coefficients par rapport au pic le plus intense.

Cette même expérimentation a été réalisée sur des fenêtres m/z 0,8 et 1,6 (Annexe P et Q). Les résultats obtenus pour la fenêtre m/z 0,4 sont schématisés sur la Figure 50. Les lignes rouges sous les pics représentent l'intensité des pics en MS/MS, tandis que les pics observés sont ceux vus en MS. Sur cette Figure, on voit qu'avec une fenêtre de m/z 0,4 ciblée sur P1, les pics P0 et P2 ont une intensité réduite en MS/MS par rapport à la MS. De plus, comme cela a été observé dans la Figure 49, du signal appartenant à un pic isotopique qui ne devrait pas être inclus dans la fenêtre est quand même détecté. La partie C de la Figure 50 schématise la forme théorique du signal obtenu avec une fenêtre de sélection de m/z 0,4. En Annexe R est schématisée la forme de la fenêtre de sélection pour un m/z de 1,6 en tenant compte des résultats obtenus pour cette largeur de fenêtre de sélection.

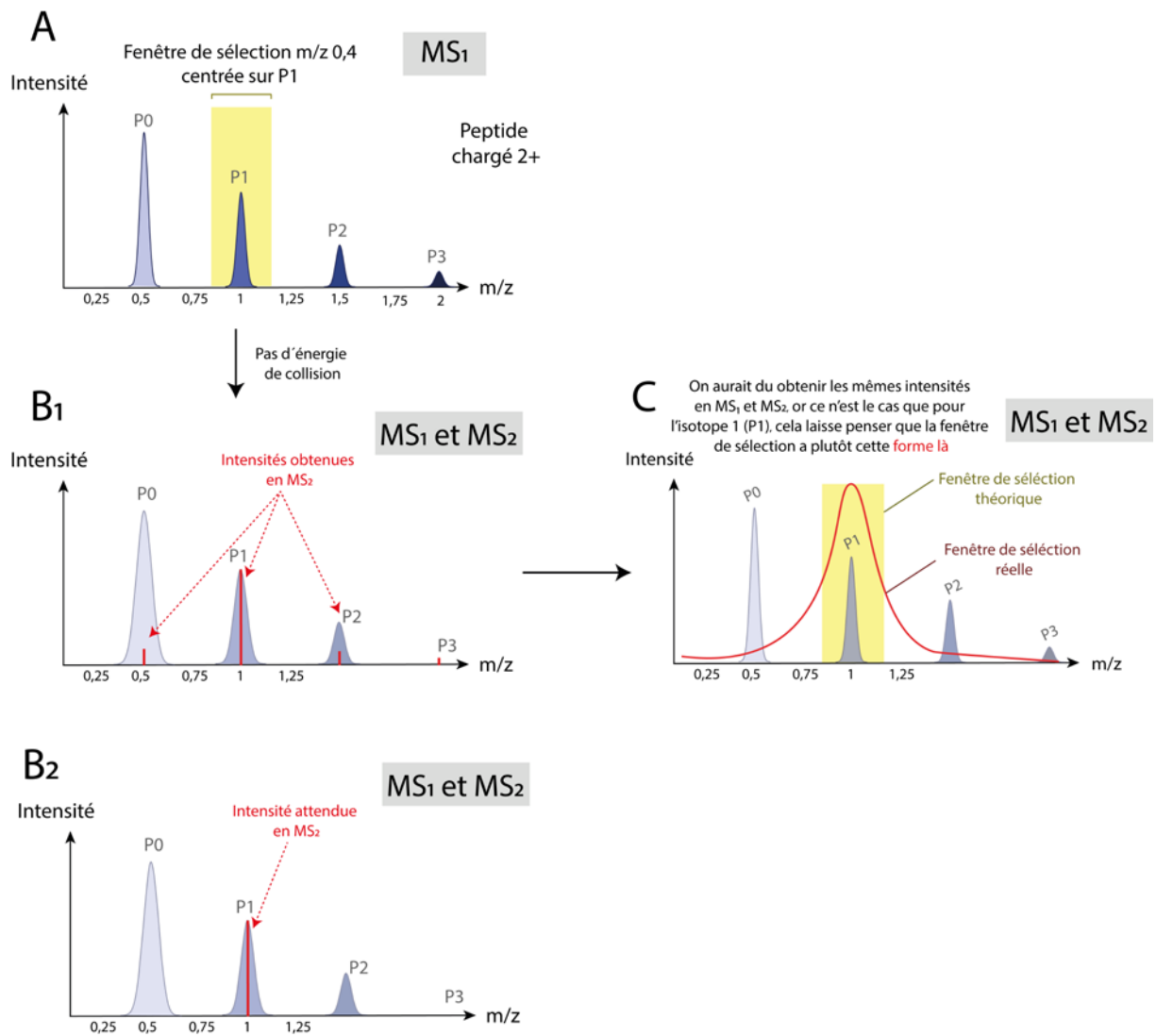


Figure 50 : Représentation des pics obtenus lors d'une analyse d'un peptide chargé 2+ sans énergie de collision en  $MS_1$  et en  $MS_2$  avec une fenêtre de sélection  $m/z$  0,4 centrée sur P1 et représentation théorique de la forme « réelle » de la fenêtre de sélection.

Certes, la fenêtre de sélection n'est pas parfaitement définie mais leur forme n'est pas tout à fait celle d'une courbe gaussienne non plus. En effet, les résultats obtenus sur la Figure 49 et les Figures en Annexe P et Q montrent que la partie du signal précédant l'apex est plus étroite que la partie du signal situé après l'apex. Une forme similaire à cette courbe peut être retrouvée lors de l'enregistrement du voltage issu de la foudre (Figure 51)<sup>239</sup>. Ceci dit, malgré cette similarité, nous en sommes restés ici quant à la compréhension de ce phénomène.

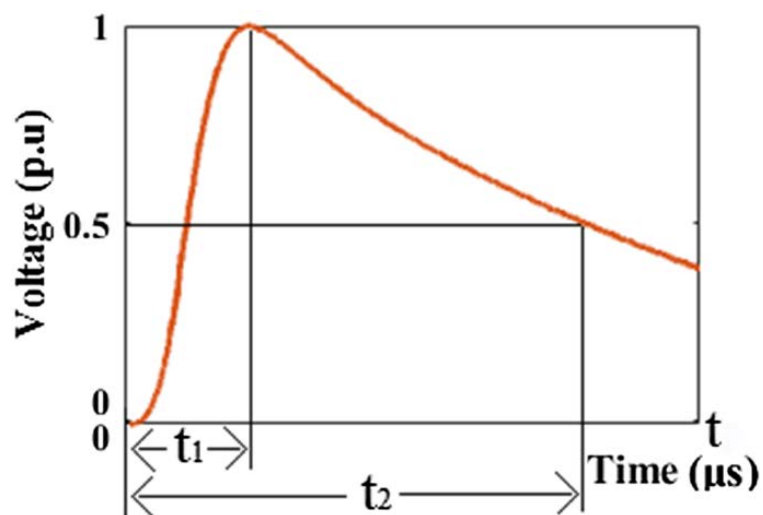


Figure 51 : Forme d'onde de la tension d'impulsion de la foudre d'après Ghosh, P. et al<sup>239</sup>

Maintenant que l'on sait que la sélection d'ions n'est pas « homogène », les intensités et aires correspondant aux fragments situés sur les extrémités de la fenêtre de sélection peuvent être corrigées (ex : pic monoisotopiques (sans isotopes alourdis) ou avec deux isotopes alourdis lorsque le pic isotopique P1 est ciblé).

La correction réalisée consiste à déterminer dans un premier temps la proportion de chaque pic du massif dans la distribution isotopique de la molécule ciblée. Ensuite, pour une fenêtre de sélection m/z donnée et incluant plusieurs pics isotopiques, on prendra comme référence le pic isotopique ciblé. La moyenne de l'intensité des peptides sous le pic ciblé sera comparée à la moyenne de l'intensité des peptides des autres pics inclus dans la fenêtre de sélection. La différence entre la distribution théorique attendue et celle réellement obtenue sera calculée pour corriger l'intensité et l'aire des massifs isotopiques situés sur les extrémités. Cette correction a été réalisée sur les résultats de la MS pour les différentes fenêtres de sélection m/z (Tableau 17).

Tableau 17 : Comparaison du pourcentage de différence entre théorique et expérimental pour des intensités et des aires obtenues en MS<sub>1</sub> sur des fenêtres de sélection m/z de 1,6 et 0,4.

Peptides	Fenêtre de sélection m/z 1,6 différence en % entre théorie et expérimentation			Fenêtre de sélection m/z 0,4 différence en % entre théorie et expérimentation		
	Sans correction	Correction par la moyenne des peptides sur l'intensité	Correction par la moyenne des peptides sur l'aire	Sans correction	Correction par la moyenne des peptides sur l'intensité	Correction par la moyenne des peptides sur l'aire
ALQASALAAWGK	3,86	3,88	3,88	9,36	7,60	10,00
ELSEIAQSIVANGK	3,26	3,54	3,54	8,34	5,48	9,22
ETTIQGLDGLSER	6,55	6,53	6,53	5,48	2,99	6,53
GILAADESVMGMR	3,34	3,15	3,15	11,90	10,70	12,80
LSFSYGR	4,81	5,41	5,41	5,67	5,15	6,72
YTPEQVAMATVTALHR	6,02	6,02	6,89	9,31	8,38	7,92

La correction par la « forme » de la fenêtre de sélection ne semble pas très utile d'après les résultats obtenus dans le Tableau 17, elle ne sera pas utilisée pour affiner les résultats par la suite.

#### 5. Calculs de ratios forme non-marquée / PSAQ<sup>+1</sup> :

L'objectif du projet PSAQ<sup>+1</sup> est d'arriver à estimer une quantité de molécule non-marquée dans un biofluide à partir d'une quantité connue de molécule marquée introduite dans un échantillon. Les étapes précédentes ont permis d'épurer le modèle permettant de calculer un ratio entre ces deux molécules. A présent, il est temps de vérifier si le modèle peut effectivement être utilisé pour réaliser une quantification « absolue ».

Pour cela, un mélange d'ALDOB recombinantes correspondant aux formes non-marquée et PSAQ<sup>+1</sup> a été réalisé dans un ratio de 0,8. Les mélanges peptidiques obtenus après dénaturation, réduction, alkylation, digestion trypsique et dessalage ont été dilués en cascade pour réaliser une gamme de concentration. Quatre dilutions en cascade avec un facteur 10 entre chacune avec une concentration initiale solution à 10 ng/ml ont été réalisées avec quatre répliques par dilution (Tableau 18). Les échantillons ont été injectés graduellement, en commençant par la concentration la plus faible et en augmentant progressivement, avec 4 répétitions à chaque fois. Cela évite tout risque de contamination en n'injectant jamais un échantillon plus concentré avant un moins concentré. Le protocole détaillé ainsi que les précisions sur la méthode LC-MS utilisée peuvent être retrouvés en Annexe I et J.

*Tableau 18 : Dilutions et répliques réalisés pour le mélange d'ALDOB légère et d'ALDOB sous forme PSAQ<sup>+1</sup>*

Identifiant	Concentration (volume)	Répliques
Dilution n°1	10 ng/μl (3 μl)	4
Dilution n°2	1 ng/μl (3 μl)	4
Dilution n°3	100 pg/μl (3 μl)	4
Dilution n°4	10 pg/μl (3 μl)	4

Les dilutions ont été analysées par HPLC-MS/MS dans l'ordre croissant de concentrations en quadrupla. Les fichiers bruts obtenus ont été traités avec le logiciel MZscope pour extraire des données MS/MS l'intensité et l'aire des fragments. Ces données permettent de calculer un ratio entre fragments légers et marqués. Sur la base du modèle théorique et compte tenu des corrections apportées, nos résultats montrent qu'il est possible d'estimer la concentration d'une protéine endogène (non-marquée) de quantités inconnues à partir de la concentration de protéine PSAQ<sup>+1</sup> introduite dans l'échantillon. Les concentrations expérimentales obtenues pour chacune des dilutions en regard des concentrations théoriques sont données dans la Figure 52 pour chacun des peptides suivis.



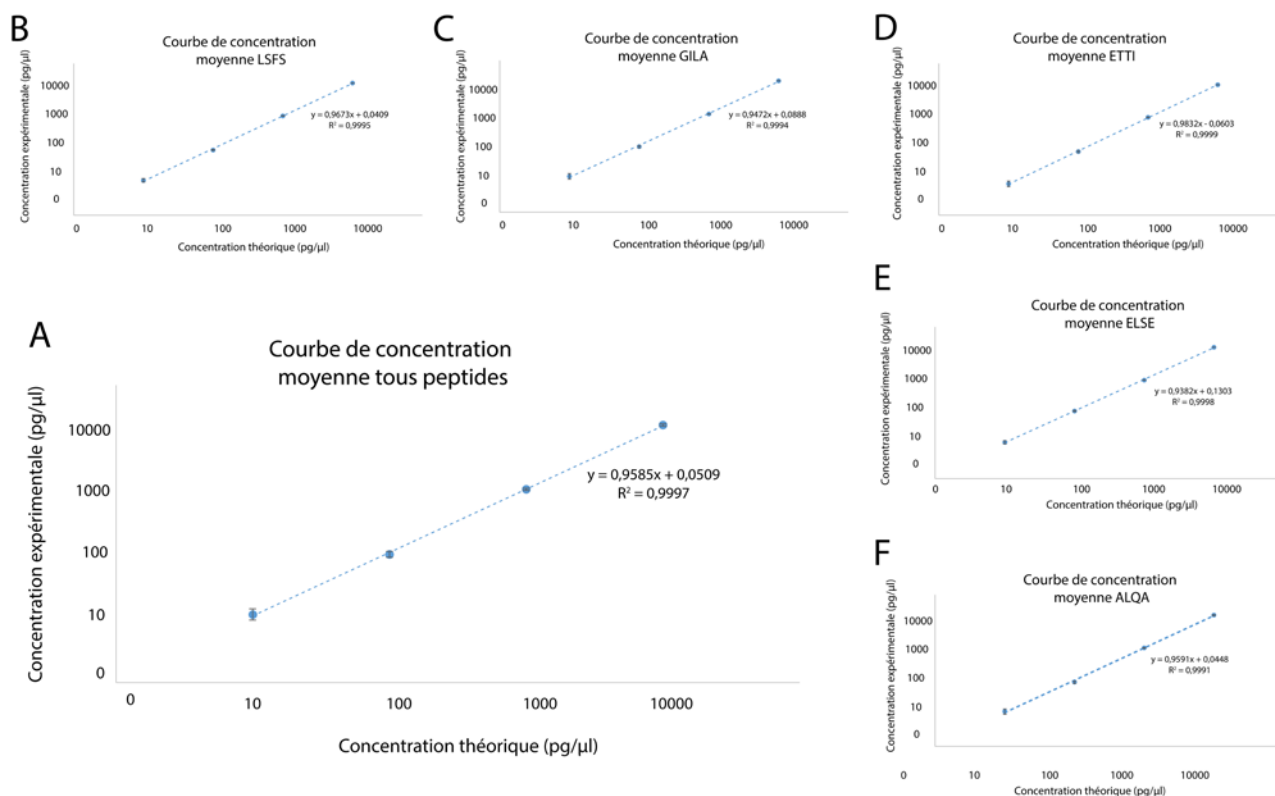


Figure 52 : Courbe de concentration de la moyenne des peptides de l'ALDOB (ALQASALAAWGGK, ELSEIAQSIVANGK, ETTIQGLDGLSER, GILAADESVGTMGNR et LFSYGR) et courbe de concentration de chacun des peptides de l'ALDOB indépendamment. Les courbes ont été réalisées à partir des données sur l'intensité des fragments de chaque peptide de l'ALDOB. Les barres d'erreur pour la courbe A ont été calculées sur la moyenne des répliques de chaque peptide et les barres d'erreur pour les courbes B, C, D, E et F ont été calculées sur l'intensité de tous les fragments.

Les ratios obtenus pour chaque fragment ont été comparés au ratio entre protéine non-marquée et PSAQ<sup>+1</sup> introduits au début de l'expérimentation dans un ratio de 0,8 dans la Figure 53.

Dans les Figure 52 et 53, les intensités associées à tous les fragments de chaque précurseur ont été rassemblées. Nous pouvons noter que pour des concentrations basses, certains fragments peuvent parfois donner des résultats aberrants. Lors d'analyses ciblées de type SRM, il est commun de choisir les fragments (transitions) qui donneront les meilleurs résultats pour éviter des contaminations. Dans ce même but, nous avons essayé de déterminer une méthodologie permettant de trier les fragments offrant le plus de cohérence possible. Pour cela, le coefficient de variation (CV) de plusieurs conditions a été calculé. Les différentes conditions sont décrites dans le Tableau 19 qui décompte également le nombre de précurseurs avec un CV inférieur à 20% pour chacune des conditions.

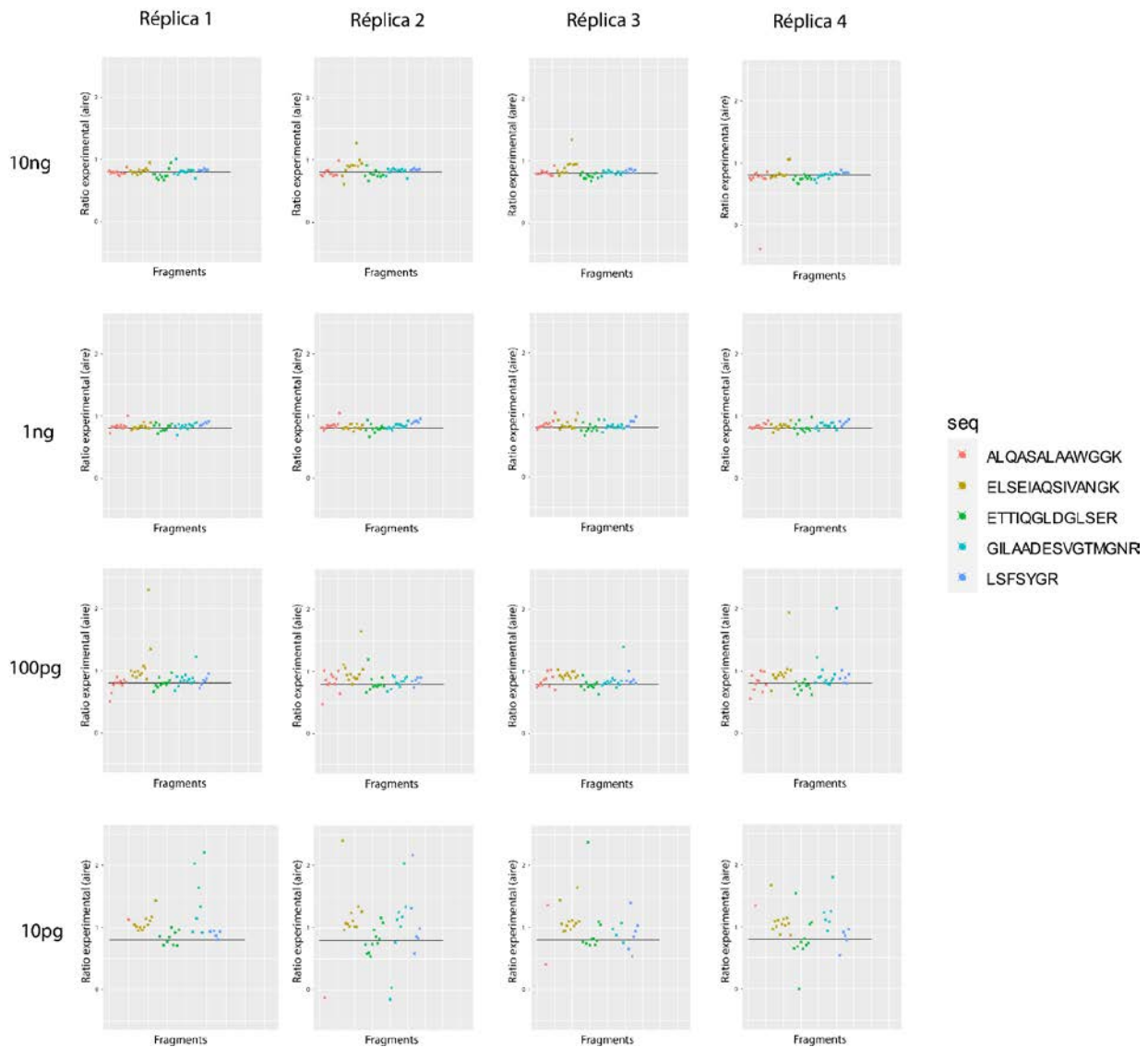


Figure 53 : Ratios entre fragment non-marqué et marqué pour chacun des fragments des peptides de l'ALDOB (ALQASALAAWGGK, ELSEIAQSIVANGK, ETTIQGLDGLSER, GILAADESVGTMGNR et LFSYGR). Quatre concentrations différentes ont été testées : 10 ng/ $\mu$ l, 1 ng/ $\mu$ l, 100 pg/ $\mu$ l et 10 pg/ $\mu$ l. Pour chacune de ces concentrations, 4 répliques ont été réalisées. La ligne noire représente le ratio entre ALDOB non-marquée et marquée (=0,8) introduites en début d'expérimentation.

Les conditions C et E du Tableau 18 permettent d'obtenir le plus grand nombre de précurseurs avec un CV de l'intensité entre fragments inférieur à 20%. La condition C consiste à prendre les 3 fragments les plus intenses dont le m/z est supérieur à celui du précurseur ; et la condition E consiste à prendre les 3 fragments les plus intenses parmi tous les fragments. L'Annexe S montre que les fragments inclus dans les conditions C et E sont rarement les mêmes. Le tableau en Annexe T donne des résultats sur les coefficients de variation associés aux conditions C et E. Bien que la condition E semble présenter un léger avantage numérique, la condition C est plus intéressante, car les fragments de la condition C ont été triés de façon à ce que leur m/z soit plus important que le m/z du précurseur. L'intérêt de ces fragments réside dans le fait que les possibles contaminations avec des m/z similaires est plus faible<sup>126</sup>. Ceci est d'autant plus vrai que l'échantillon est complexe, dans les expériences réalisées jusqu'ici les échantillons ne contenaient que des protéines recombinantes ce qui limitait forcément les problèmes de contaminations. Pour ces raisons, la condition retenue est la C.

Tableau 19 : Nombre de peptides avec un coefficient de variation des intensités inférieur à 20% pour différentes conditions de tri. Totalité : tous les peptides sont pris en compte. m/z fragment > m/z précurseur : uniquement les fragments dont le m/z est supérieur au m/z du peptide sont pris en compte. 3 fragments les plus intenses dont m/z > m/z précurseur : uniquement les 3 fragments les plus intenses et dont le m/z est supérieur au m/z du précurseur sont pris en compte. 5 fragments les plus intenses : uniquement les 5 fragments les plus intenses sont pris en compte. 3 fragments les plus intenses : uniquement les 3 fragments les plus intenses sont pris en compte.

Conditions	A Totalité des fragments	B m/z fragment > m/z précurseur	C 3 fragments les plus intenses dont m/z > m/z précurseur	D 5 fragments les plus intenses	E 3 fragments les plus intenses
Index	Nombre de peptides avec CV < 20%				
10 pg/μl - Réplica 1	2	3	4	3	4
10 pg/μl - Réplica 2	0	2	3	2	3
10 pg/μl - Réplica 3	2	3	4	3	4
10 pg/μl - Réplica 4	1	3	3	2	3
100 pg/μl - Réplica 1	3	3	5	5	5
100 pg/μl - Réplica 2	4	4	5	5	5
100 pg/μl - Réplica 3	5	4	5	5	5
100 pg/μl - Réplica 4	3	3	5	5	5
1 ng/μl - Réplica 1	5	5	5	5	5
1 ng/μl - Réplica 2	5	5	5	5	5
1 ng/μl - Réplica 3	5	5	5	5	5
1 ng/μl - Réplica 4	5	5	5	5	5
10 ng/μl - Réplica 1	5	5	5	5	5
10 ng/μl - Réplica 2	5	5	5	5	5
10 ng/μl - Réplica 3	5	5	5	5	5
10 ng/μl - Réplica 4	5	5	5	5	5

### C. Conclusion projet PSAQ<sup>+1</sup> :

En résumé, une méthode de quantification absolue utilisant des protéines isotopiquement alourdis a été optimisée et testée pendant ma thèse. Cette méthode se base sur l'utilisation du PSAQ<sup>+1</sup>, une protéine recombinante pour laquelle les arginines et les lysines possèdent un seul carbone isotopiquement alourdi. Le principe de cette méthode consiste à introduire la protéine PSAQ<sup>+1</sup> dans un échantillon pour quantifier une protéine endogène homologue par HPLC-MS/MS. La particularité des protéines PSAQ, réside dans le fait qu'elles sont introduites dans l'échantillon avant toute étape de préparation, ainsi les deux protéines, endogène et marquée, seront dénaturées, réduites, alkylées et digérées en même temps. En comparaison avec la méthode PSAQ classique, la méthode PSAQ<sup>+1</sup> permet d'analyser les peptides provenant des protéines, non-marquée et marquée, dans une même fenêtre de sélection du précurseur très réduite. Cela a pour avantage de pouvoir réduire le temps de cycle lié à une analyse.

Le travail réalisé a permis de démontrer la faisabilité de la méthode. Ainsi, en introduisant une quantité connue de protéine PSAQ<sup>+1</sup> dans un échantillon, il est possible de connaître la quantité de protéine endogène homologue. Nous avons démontré que les résultats obtenus entre les différents fragments analysés en MS/MS sont reproductibles. Nous avons également démontré qu'il est possible de réaliser une quantification fiable avec des concentrations faibles de l'ordre de la dizaine de picogrammes par microlitre. Nous avons également pu démontrer que la fenêtre de sélection m/z 0,4 était la plus adaptée pour réaliser des analyses avec du PSAQ<sup>+1</sup> car d'un côté elle permet de sélectionner le pic isotopique P1 avec une correction minimale quant à la forme de la fenêtre de masse ; et de l'autre, elle permet d'obtenir une spécificité importante. Ce dernier point reste à démontrer dans un échantillon complexe (ex : plasma).

Ces résultats, très encourageants, ont permis de développer plusieurs hypothèses qui restent à démontrer. Tout d'abord, la réduction du temps de cycle pourrait permettre d'augmenter la sensibilité des analyses par le moyen d'une acquisition de signal plus prolongée. De même, l'utilisation d'une fenêtre de sélection des précurseurs réduite, devrait permettre d'améliorer la spécificité de l'analyse car la quantité de contaminants avec un rapport m/z et un temps de rétention similaire à celui du peptide ciblé s'en trouverait réduite. Un autre avantage, non démontré à ce jour de l'utilisation du PSAQ<sup>+1</sup>, est l'obtention d'une précision accrue lors de la quantification absolue, en effet, les deux précurseurs (endogène et marqué) étant analysés dans la même fenêtre de sélection cela permet de réaliser une comparaison plus « directe » entre les deux formes. Enfin, il faudra également démontrer que la méthode PSAQ<sup>+1</sup> est utilisable dans un échantillon complexe comme le plasma. Le modèle PSAQ<sup>+1</sup> pourrait encore être affiné notamment en intégrant dans les calculs la pureté du PSAQ+1 ou encore en retranchant les signaux parasites qui se superposent au signal du précurseur ciblé. Des tests complémentaires doivent encore être réalisés pour que la méthode du PSAQ<sup>+1</sup> puisse être globalement adoptée en protéomique. Néanmoins, les perspectives quant à cette méthode sont prometteuses et ont le mérite de présenter une amélioration originale et innovante pour la quantification absolue de protéines par spectrométrie de masse.

En effet, d'autres moyens pour réaliser des quantifications absolues existent tels que les tests ELISA ou plus récemment l'utilisation d'aptamères<sup>240</sup>. Une comparaison des performances entre ces différentes méthodes pourrait être très intéressante. La LC-MS offre l'avantage de pouvoir réaliser des multiplexages très importants par rapport aux tests ELISA. Par contre, les tests ELISA ne nécessitent pas d'équipements de laboratoire très coûteux et spécialisés contrairement à la LC-MS/MS qui requiert aussi d'une expertise importante tant pour la préparation des échantillons, comme pour l'acquisition des données et l'analyse des résultats. Un dernier avantage de la méthode PSAQ<sup>+1</sup> c'est son potentiel de multiplexage, en effet plusieurs standards protéiques pourraient être analysés en même temps

dans un échantillon complexe, ce qui n'est pas le cas pour les tests ELISA dont le multiplexage est très limité pour l'instant<sup>241</sup>.

D'un autre côté, les tests utilisant des aptamères permettent quant à eux un multiplexage extrêmement important (>7000 protéines de façon simultanée)<sup>240</sup> et une spécificité et sensibilité considérables. Ils représentent une alternative très intéressante par rapport à la LC-MS ou l'ELISA mais ont aussi quelques limitations. En effet, les aptamères sont utilisés pour des quantifications de protéines intactes, il est donc possible que leur spécificité varie en fonction de la conformation de la protéine mais aussi de la présence de modifications post-traductionnelles et d'isoformes. De plus leur stabilité peut être variable et surtout leur utilisation pour la quantification de protéines est pour l'instant contrainte à un envoi des échantillons dans un centre spécialisé ce qui limite la généralisation de la technique.

Plusieurs techniques existent pour réaliser une quantification absolue de protéines, le choix de la technique à utiliser dépendra des critères recherchés par l'expérimentateur. Grâce à la méthodologie PSAQ<sup>+</sup>, la quantification absolue en LC-MS a été optimisée et peut répondre à des besoins que d'autres techniques ne peuvent pour l'instant pas satisfaire ce qui laisse penser que cette méthodologie a un futur prometteur.

## Conclusion générale :

Le mot protéome est introduit en 1995 par Marc Wilkins et Keith Williams pour désigner l'ensemble des protéines codées par un génome. L'essor de la spectrométrie de masse a notamment permis de réaliser des analyses de protéines à haut débit permettant d'étudier le protéome de l'organisme entier jusqu'à la cellule unique. A partir de ce moment, des recherches pour caractériser un protéome différentiel ont concentré un grand nombre d'efforts. La recherche de biomarqueurs dans le domaine clinique s'est largement reposé sur la puissance de la protéomique liée à la LC-MS pour établir des diagnostics, dépister des maladies ou prédire des événements pathologiques. Dans le cadre de ma thèse, j'ai pu travailler dans la recherche de biomarqueurs pour une maladie du foie (MASLD) mais j'ai également participé au développement d'une technique permettant de mieux quantifier des biomarqueurs.

Le projet relatif à la recherche de biomarqueurs était focalisé sur une maladie métabolique du foie, la MASLD. Cette atteinte hépatique touche 25% de la population mondiale. L'examen diagnostique de référence est la biopsie du foie, examen invasif, coûteux et sujet à des erreurs. En l'absence d'une alternative fiable il existe un réel besoin pour élaborer un test contournant les désavantages de la biopsie tout en présentant une performance similaire. Nous avons pour cela analysé deux cohortes de 158 et 200 patients avec plusieurs techniques de protéomique (LC-MS DDA/SRM et ELISA) pour identifier, vérifier et valider deux nouveaux biomarqueurs de fibrose que sont l'Insulin like growth factor binding protein (ALS) et la Galectin 3 binding protein (LG3BP). La performance de l'ALS et la LG3BP prises de façon individuelle est similaire à celle du FibroTest pour différencier une fibrose avancée dans le foie. Ces deux biomarqueurs ont été introduits dans un panel dérivé du FibroTest afin de mieux différencier une fibrose précoce d'une fibrose avancée. Du fait de l'utilité évidente de ce panel par rapport au FibroTest, des analyses sur un plus grand nombre de patients sont nécessaires pour évaluer à plus grande échelle le bénéfice du panel développé afin qu'il soit introduit rapidement en milieu hospitalier. On peut imaginer que des futurs progrès techniques en protéomique quantitative permettront de découvrir de nouveaux biomarqueurs de façon plus fiable et précise.

C'est en gardant à l'esprit cette perspective, que j'ai été amené à travailler sur le projet PSAQ<sup>+1</sup>. En effet, la quantification de protéines se doit d'être précise, fiable et reproductible afin d'estimer les concentrations réelles de protéines dans un échantillon à un temps donné. Plusieurs méthodes de quantification existent parmi lesquelles on peut compter l'introduction de standards isotopiques. Cette méthode consiste à introduire un isotopologue de la protéine ciblée dans l'échantillon qui permettra par comparaison des signaux LC-MS, de remonter à une concentration de protéine endogène. Ce projet se base sur une amélioration du PSAQ, une protéine marquée isotopiquement et introduite dans l'échantillon avant toute étape de préparation à l'analyse. L'originalité du PSAQ<sup>+1</sup> était de ne présenter qu'un seul carbone marqué isotopiquement sur les arginines et les lysines contrairement au PSAQ « classique » qui présente un marquage sur tous les carbones et azotes de ces acides aminés. Le défi principal était de montrer que cette méthode était utilisable en PRM sur une seule fenêtre de sélection restreinte. Pour cela, il a fallu démontrer qu'il était possible de distinguer un pic hybride composé de signal provenant des peptides du standard et de signal provenant d'un isotope naturel de peptides endogènes. A partir de ces résultats, j'ai établi qu'il était possible de remonter à une quantité de protéine endogène à partir du signal fourni par le standard isotopique. Cette méthode permettrait potentiellement d'améliorer la sensibilité, la spécificité et la précision d'une quantification absolue mais cela reste à démontrer par des expériences complémentaires. De même, l'utilisation de cette méthode en échantillon complexe doit encore être testée avant de pouvoir réaliser des analyses sur des cohortes cliniques. PSAQ<sup>+1</sup> ouvrirait de nouvelles perspectives en protéomique quantitative en levant certaines limitations grâce à l'analyse de la protéine endogène et

la protéine marquée isotopiquement dans une même fenêtre de masse. Le projet PSAQ<sup>+1</sup> permettrait notamment de réaliser un multiplexage important tout en conservant la qualité des quantifications.

La découverte de biomarqueurs reste un sujet complexe, des améliorations au niveau de l'instrumentation permettront peut-être d'obtenir de meilleures performances dans le futur mais des limitations inhérentes aux biomarqueurs pourront difficilement être dépassées. En effet, et comme cela a été dit lors de l'introduction, l'objectif d'un biomarqueur est d'être un indicateur d'un état biologique. La difficulté principale, c'est que les états biologiques ont été définis de façon binaire (ex : malade/non-malade) ou segmentée (ex : Fibrose 0-1-2-3-4) par les humains, or la biologie est beaucoup plus complexe. Cette complexité peut être observée à différents niveaux. Tout d'abord, les organismes vivants, y compris les humains, présentent une grande hétérogénéité au niveau des facteurs génétiques, environnementaux ou du mode de vie ce qui fait que trouver des biomarqueurs fiables pour tout le monde malgré les variations de ces paramètres représente un défi majeur. De plus, les processus biologiques impliquent souvent une multitude de molécules, de voies métaboliques et de réseaux de régulation qui sont souvent interconnectés et peuvent varier dans le temps ce qui rend la tâche de trouver un biomarqueur fiable d'autant plus ardue. Et finalement, il existe aussi une complexité importante au niveau d'une population car chaque individu est différent. Tous ces facteurs font qu'il est difficile de trouver des biomarqueurs avec une utilité clinique parfaitement fiable. Même si les marqueurs les plus utilisés en ce moment sont basés sur des études avec plusieurs cohortes indépendantes, représentatives et avec un nombre d'individus élevé, leur performance pour déceler un état biologique ou pathologique reste sous-optimale.

En résumé, la complexité inhérente à la biologie, associée à la variabilité des individus et aux limitations des technologies de détection crée des défis importants dans la recherche, la validation et l'application des biomarqueurs. Malgré ces difficultés, les avancées continues dans la technologie et les méthodes de recherche permettent d'aborder ces obstacles et d'améliorer notre compréhension des biomarqueurs pour le diagnostic et le traitement des maladies.

## Annexes :

Annexe A : Résumé des différents panels décrits dans l'étude bibliographique et de leur AUROC

Panel	Paramètres inclus	ROC	Référence
NAFLD ridge score <sup>®</sup>	ALT, HDL, HbA1c, cholestérol, triglycérides, comptage de leucocytes, hypertension	0,88 → détection NAFLD	Yip et al. 2017 <sup>242</sup>
NAFLD Liver Fat Score (NLFS) <sup>®</sup>	AST, AST/ALT, niveau d'insuline, quantité de gras hépatique, syndrome métabolique, diabète de type 2	0,86 → détection NAFLD	Kotronen et al. 2009 <sup>243</sup>
Hepatic Steatosis Index (HIS) <sup>®</sup>	AST/ALT, IMC, diabète, sexe	0,812 → détection NAFLD	J.-H. Lee et al. 2010 <sup>244</sup>
Fatty liver index (FLI) <sup>®</sup>	IMC, circonférence de la taille, triglycérides, GGT	0,84 → détection foie gras	Bedogni et al. 2006 <sup>245</sup>
Lipid accumulation product (LAP) index	Sexe, triglycérides, circonférence de la taille	0,79 → détection stéatose hépatique	Bedogni et al. 2010 <sup>246</sup>
SteatoTest <sup>®</sup>	ALT, $\alpha$ 2M, ApoA1, Hp, bilirubine totale, GGT, cholestérol, triglycérides, glucose, âge, sexe, IMC	0,72 - 0,86 → détection stéatose	Poynard, Ratziu, et al. 2005 <sup>247</sup>
ActiTest <sup>®</sup>		0,846 → détection stéatose	Munteanu et al. 2016 <sup>248</sup>
	Ferritine, collagène IV 7s, insuline	0,851 → différenciation SS et NASH	Yoneda et al. 2010 <sup>249</sup>
CHeK score <sup>®</sup>	age, HbA1c, GGT, adiponectine, M30	0,73 → différenciation NAFLD et NASH	Canbay et al. 2019 <sup>250</sup>
NASHTest <sup>®</sup>	Âge, sexe, hauteur, poids, triglycérides, cholestérol, $\alpha$ 2M, ApoA1, Hp, GGT, ALT, AST, bilirubine totale	0,79 → détection NASH 0,69 → détection NASH récente	Poynard et al. 2007 <sup>251</sup>
Nash Apoptosis Score <sup>®</sup>	sFasL, âge, ferritine, transferrine, triglycérides	0,78 → détection NASH	Alkhouri et al. 2015 <sup>252</sup>
	CK18, sFas	0,79 → détection NASH	Tamimi et al. 2011 <sup>253</sup>
NAFLD Fibrosis Score (NFS) <sup>®</sup>	AST/ALT, albumine, comptage de plaquettes, âge, IMC, hyperglycémie	0,82 → détection fibrose avancée	Angulo et al. 2007 <sup>82</sup>
Fibrosis-4 Score (FIB-4) <sup>®</sup>	AST, ALT, comptage de plaquettes, âge	0,86 → détection fibrose avancée	McPherson et al. 2010 <sup>254</sup>



AST to Platelet Ratio Index (APRI) <sup>®</sup>	AST, comptage de plaquettes	0,85 → détection fibrose avancée	Kruger et al. 2011 <sup>255</sup>
BARD Score <sup>®</sup>	IMC, AST/ALT, diabète	0,81 → détection fibrose avancée	S. A. Harrison et al. 2008 <sup>256</sup>
European liver fibrosis panel (ELF) <sup>®</sup>	TIMP-1, PIIINP, HA, âge,	0,782 → différenciation fibrose faible/modérée et avancée 0,804 → différenciation fibrose faible et modérée 0,887 → différenciation fibrose modérée et avancée 0,87 → détection NAFLD	Rosenberg et al. 2004 <sup>257</sup>
Hepascore <sup>®</sup>	α2M, bilirubine, GGT, HA, âge, sexe	0.85 → détection fibrose 0.96 → détection fibrose avancée 0.94 → cirrhose	Adams et al. 2005 <sup>258</sup>
FibroTest <sup>®</sup>	α2M, Hp, GGT, total bilirubine totale, ApoA1, ALT, âge, sexe	0,73 à 0,87 → détection fibrose	Poynard, Imbert-Bismut, et al. 2005 <sup>202</sup>
FibroMeter NAFLD index (FibroMeterVCTE)	Comptage de plaquettes, index prothrombine, ferritine, AST, ALT, poids, âge, élasticité du foie	0,883 → détection des stades F2 à F4	Calès et al. 2005 <sup>201</sup>
NIS4	HbA1c, α2M, YKL-40, miR-34a	0,82 → détection fibrose avancée	S. Harrison et al. 2017 <sup>259</sup>
oxNASH <sup>®</sup>	13 HODE:LA, IMC, AST	0,730 → détection inflammation 0,723 → détection ballonnement 0,705 → détection stéatose 0,673 → détection fibrose	Alkhoury et al. 2014 <sup>260</sup>
ABC3D <sup>®</sup>	Pro-C3, IMC, plaquettes, T2DM (HbA1c)	0,83 → détection Fibrose avancée (F3 - F4)	Boyle et al. 2019 <sup>261</sup>
FIB-3C <sup>®</sup>	Pro-C3, IMC	0,85 → détection Fibrose avancée (F3 - F4)	Boyle et al. 2019 <sup>261</sup>
ADAPT <sup>®</sup>	Pro-C3, plaquettes, âge,	0.86–0.87 → fibrose avancée	Daniels et al. 2019 <sup>84</sup>
FibroSpect	HA, TIMP-1, α2M et protéines impliquées dans la MEC	0,86-0,87 → détection fibrose avancée (F2 À F4)	Christensen et al. 2006 <sup>262</sup>
	MMP-1, MMP-7, HA, AFP et AST MMP-7, MMP-1, AFP, PDGF -AA, APRI	0,938 → détection 0,928 → fibrose avancée (F3F4)	Irvine et al. 2016 <sup>263</sup>

Annexe B : Résumé des différents biomarqueurs décrits dans l'étude bibliographique, de leur AUROC et caractéristiques.

Biomarqueur	ROC (ou P value si pas de ROC disponible)	ROC stratification (ou P value si pas de ROC disponible)	Caractéristiques	Références
PIIINP	0,85 - 0,87 → différenciation ss et NASH. 0,77 - 0,89 → détection inflammation lobulaire 0,76 - 0,80 → détection ballonnement	0,79 - 0,88 → corrélation score NAS	Présent dans le panel OELF®	Tanwar et al. 2013 <sup>264</sup>
Pro-C3	0,78 → détection fibrose avancée	P <<0.0001 → stéatohépatite histologique p <<0.0001 → grade de fibrose	Inclus dans les panels : FIB-C3®, FIB-4® et ABC3D®	(Boyle et al. 2019) <sup>261</sup>
Pro-C1 Pro-C4, Pro-C5, Pro-C6I	/	/	ARNm de Pro-C1, Pro-C3I et Pro-C4 augmentés si foie fibrotique Résultats non-significatifs pour ces procollagènes	Milani et al. 1990 <sup>265</sup> Bril et al. 2019 <sup>266</sup> Luo et al. 2018 <sup>267</sup>
Collagène I	/	/	Augmenté lors de stade NASH	Zheng et al. 2017 <sup>268</sup>
Collagène IV 7s	0,8458 → diagnostic NAFLD 0.8763 → diagnostic NAFLD (collagène IV 7s + YKL-40) 0.769 → détection NASH (collagène IV 7s + AST) 0.842 → détection fibrose NASH (collagène IV 7s + AST)	0,778 → corrélation avec NAS	Détection NAFLD, NASH et fibrose chez NASH Combiné avec AST et YKL-40, il donne de meilleurs résultats	Mizuno et al. 2017 <sup>269</sup> Kumagai et al. 2016 <sup>270</sup> Okanoue et al. 2018 <sup>271</sup>
MMP-1	/	0,64 → détection F1 0,90 → détection F2 0,88 → détection F3	Marqueur utile uniquement pour les fibroses	Mahmoud et al. 2012 <sup>272</sup>
TIMP-1	P = 0,03 → différenciation F2 et patient sain	P = 0,007 → prédiction stade de fibrose	Marqueur utile uniquement pour les fibroses	Yilmaz and Eren 2019 <sup>186</sup>
MMP-2, MMP-7, MMP-9, TIMP-1 TIMP-2	/	/	Concentrations significativement différentes entre NAFLD et contrôles (mais ne différencient pas une stéatohépatite d'une simple stéatose). Panel MMP-1, MMP-7, HA, AFP et AST (ROC = 0,938) Panel MMP-7, MMP-1, AFP, PDGF-AA, APRI® (ROC =	Yilmaz and Eren 2019 <sup>186</sup> Irvine et al. 2016 <sup>273</sup>

			0,928) pour détection de fibroses avancées (F3-F4).	
C1m, C3m, C4m12a1	/	/	Concentrations significativement supérieures lors de fibroses hépatiques chez les rats	Leeming et al. 2011 <sup>274</sup> Barascuk et al. 2010 <sup>275</sup> Sand et al. 2013 <sup>276</sup>
C6m + Pro-C3	0,91 → détection fibrose hépatique	/	Faible sensibilité (0,54) mais grande spécificité (0,95)	Karsdal et al. 2019 <sup>277</sup>
YKL-40	0,7638 → détection NASH	/	Peu spécifique des tissus hépatiques	Kumagai et al. 2016 <sup>270</sup>
α2M	/	/	α2M incluse dans panel NashTest <sup>®</sup> : (1) présence de NASH (ROC = 0,79), (2) NASH récente (ROC = 0,69), (3) absence de NASH (ROC = 0,83) α2M incluse dans le panel FibroTest <sup>®</sup> : ROC = 0,84 pour détecter une fibrose chez un patient NAFLD.	Poynard et al. 2006 <sup>81</sup> Poynard et al. 2007 <sup>251</sup>
laminine	0,87 → détection fibrose	/	Association laminine + HA (ROC = 0,87), laminine + collagène type IV (ROC = 0,87), laminine + AST (ROC = 0,80).	Dos Santos et al. 2005 <sup>278</sup>
FGF21	0,661 → prédiction simple stéatose	/	Augmentation de l'expression du gène FGF21 si NAFLD ou NASH Utile pour des stades précoces (stéatose)	Dushay et al. 2010 <sup>279</sup> G. Wu et al. 2017 <sup>280</sup>
ALT	0,93 → détection NAFLD	/	La plupart des études ne montrent pas de corrélation entre ALT et NAFLD, mais avec seuil à 23 UI/L il est possible de détecter efficacement NAFLD	Martin-Rodriguez et al. 2017 <sup>281</sup>
P-selectine E-selectine	/ 0,73 → différentiation NASH et non-NASH	/ /	Études contradictoires concernant la P-selectine Concentration et expression de E-selectine augmentés significativement avec le degré d'atteinte NAFLD	Simons et al. 2020 <sup>282</sup>
Tβ4	0,916 → différenciation NAFLD - NASH	/	Corrélation négative avec score NAS et conditions de stress oxydatif	Jiang et al. 2018 <sup>283</sup>
Vitronectine v10	/	p<0,001 → corrélation score NAS	Concentrations significativement augmentés sur des études de protéomique	Del Ben et al. 2018 <sup>284</sup> Paradis et al. 2005 <sup>285</sup>

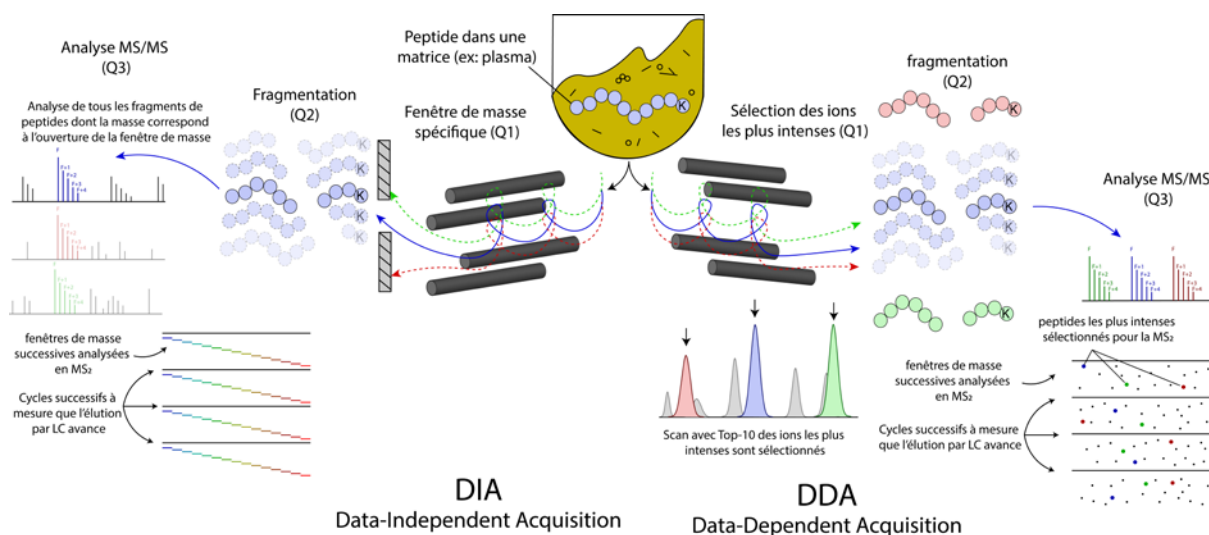
M2BP	0,748 → différenciation SS et NASH 0,816 → détection NASH	0,750 → F1 0,809 → F2 0,738 → F3	M2BP liée à la fibrose et présence de NASH	Atsukawa et al. 2018 <sup>286</sup> Kamada et al. 2017 <sup>187</sup>
Hs-CRP	0,833 → différenciation SS et NASH	P = 0,0002 → F1 P = 0.0001 → F2 P = 0,0009 → F3 P = 0,0046 → F4	Hs-CRP liée au stade de fibrose ARNm augmenté en cas de NASH Biomarqueur testé sur grande cohorte pour prédire NAFLD	Yoneda et al. 2007 <sup>287</sup> J. Lee et al. 2017 <sup>288</sup>
PTX3	0,904 → différenciation NAFLD et contrôles 0,955 → différenciation NASH et contrôles 0,702 → différenciation non-NASH et contrôles 0,858 → différenciation NASH et non-NASH 0,779 → détection NAFLD chez les jeunes	P < 0,001 → corrélation score d'activité NAFLD P < 0,01 → stade de fibrose P < 0,001 → stade de stéatose	Corrélation avec le score d'activité NAFLD, le stade de fibrose et le grade de stéatose (P < 0,001, P < 0,01 et P < 0,001 respectivement)	Boga et al. 2015 <sup>289</sup> Hamza et al. 2016 <sup>290</sup>
IL-6	0,81 → présence stéatose sévère	IL-6 : /	IL-6 utile pour détecter une stéatose sévère	Mahmoud, Bakir, and Shabana 2012 <sup>272</sup>
TGF-1β	/	0,94 → F1 0,96 → F2 0,97 → F3	TGF-1β utile pour stratifier une fibrose	Younossi et al. 2008 <sup>291</sup> Wandrer et al. 2020 <sup>292</sup>
TNFα	/	TNFα : /	TNFα utile si utilisé en panel et utile pour cible médicamenteuse et pour prédire une apparition NAFLD	Y. Y. Seo et al. 2013 <sup>293</sup>
Visfatine	P < 0,001 entre NASH et NAFLD P < 0,001 entre NAFLD et simple stéatose	/	Résultats contradictoires en fonction des études	Gaddipati et al. 2010 <sup>294</sup> Qiu et al. 2019 <sup>295</sup>
Ferritine	0,732 → différenciation NASH et SS	/	Incluse dans panel avec collagène IV 7s + insuline (ROC = 0,851) Ferritine permet de prédire atteintes plus sévères chez NAFLD	Yoneda et al. 2010 <sup>249</sup> Sumida et al. 2011 <sup>296</sup>
PEDF	p < 0,001 entre NAFLD et contrôle	/	Niveaux de PEDF corrélés indépendamment avec Pro-C3	Yilmaz et al. 2011 <sup>297</sup>

IL-1Ra	p < 0,05 → détection NASH et inflammation lobulaire	/	IL-1Ra corrélé indépendamment avec niveaux d'ALT	Pihlajamäki et al. 2012 <sup>298</sup>
OPG	0,82 → différenciation NASH et contrôle 0,963 → différenciation NASH et contrôle	/	OPG utile pour des stades avancés. Négalement corrélé avec stade NASH	Yilmaz et al. 2010 <sup>299</sup> Yang M. et al. 2016 <sup>300</sup>
RBP4	/	/	Études contradictoires : certaines trouvent une corrélation entre NAFLD et RBP4 et d'autres ne trouvent aucun résultat significatif	Zhou et al. 2017 <sup>301</sup>
CK18  CK18 M30, M65, M65ED	0,93 → détection NASH 0,70 et 0,91 → détection NASH et NAFLD respectivement  > 0,9 → détection NAFLD > 0,7 → détection NASH	/	Les concentrations en CK18 M30, M65, M65ED sont corrélées avec le degré d'atteinte hépatique  CK18 inclus dans le score FICK-3 permettant de prédire NASH (ROC = 0,765), diagnostiquer NASH (ROC > 0,8), prédire une fibrose (ROC > 0,7) et détecter une fibrose (ROC = 0,954).	Wieckowska et al. 2006 <sup>302</sup> J. Shen et al. 2012 <sup>303</sup> Mohammed et al. 2019 <sup>304</sup>
sFas  sFasL	0,668 → détection NASH 0,714 → détection NASH p = 0,0015 → détection fibrose	/	sFasL incluse dans panel Nash apoptosis score <sup>®</sup> (ROC = 0,78) pour prédire NASH et dans un autre panel contenant CK18 (ROC = 0,79) Corrélation sFasL et degré d'atteinte histopathologique	Alkhoury et al. 2015 <sup>252</sup> Tamimi et al. 2011 <sup>253</sup> Page et al. 2011 <sup>305</sup>
GGT b-GGT s-GGT b-GGT/s-GGT	0,74 → fibrose avancée 0,85 → détection NAFLD 0,853 → Détection CHC 0,93 → différenciation NAFLD et CHC.	/	GGT incluse dans les panels FibroTest <sup>®</sup> et FLI <sup>®</sup>	Franzini et al. 2012 <sup>306</sup> Ratziu et al. 2006 <sup>81</sup> Bedogni et al. 2006 <sup>245</sup>
Lp-PLA2	P < 0,001 → détection NASH P = 0,001 → détection NASH limite P < 0,001 → Détection stéatose	P = 0,0016 → scores de stéatose histologique	Lp-PLA2 corrélé avec stéatose histologique Études contradictoires sur le fait que Lp-PLA2 est plutôt augmentée ou diminuée lors d'une atteinte NAFLD	Colak et al. 2012 <sup>307</sup> Liu et al. 2018 <sup>308</sup>
Gas6	0,734 → détection fibrose 0,788 → détection maladie hépatique sévère.	/	Gas6 serait utile pour détecter des stades très avancés (cirrhose). Expression des gènes GAS6 et sAXL augmentée en cas de cirrhose. Gas6 et sAXL possibles cibles thérapeutiques.	Bellan et al. 2016 <sup>309</sup>

PLGF	/	/	Corrélation entre niveaux de PLGF et images histologiques du foie et rigidité du foie Implication de PLGF dans fibrogenèse hépatique	Krawczyk et al. 2017 <sup>310</sup>
Hp-Fuc	0,734 → différenciation NASH et non- NASH 0,759 → présence gonflement hépatocyte	0,724 → différenciation stades de fibrose	Marqueur de fibrose, donc de stades NAFLD avancés Protéine liée à l'IL-6	Kamada et al. 2013 <sup>311</sup>
AAT - A3F	0,687 → différenciation NAFLD et NASH 0,696 → détection stades NASH précoces	/	Augmentation de l'expression des gènes FUT6 et de l'IL-6 corrélée avec augmentation en AAT - A3F AAT - A3F reflète l'inflammation, la fibrose hépatique et peut différencier les degrés de fibrose chez des patients NASH	Ogawa et al. 2020 <sup>312</sup>
HbA1/HbA1c	HbA1 : sensibilité = 89% et spécificité de 83% pour prédire NAFLD	/	HbA1c corrélée avec ALT, indice de foie gras et indicateur de stéatose hépatique. Des taux élevés en HbA1c favorisent la progression de NAFLD	Xu et al. 2009 <sup>313</sup> Fiorentino et al. 2017 <sup>314</sup> C. Chen et al. 2020 <sup>315</sup>
WFA +- M2BP	0,832 → détection cirrhose <sup>249</sup>	0,70 → F2 0,71 → F3 0,75 → F4 0,77 → cirrhose 0,77 → CHC	WFA+-M2BP liée aux stades avancés de fibrose	Atsukawa et al. 2018 <sup>286</sup> Feng et al. 2020 <sup>316</sup>
Protéome N-glyqué	0,87 → détection cirrhose	/	GlycoCirrhoTest® utile uniquement pour stades NAFLD avancés	Verhelst et al. 2015 <sup>317</sup>
Ubiquitinations	/	/	NEDD8 et SUMO impliquées dans fibrose hépatique	Lachiondo-Ortega et al. 2019 <sup>318</sup>

## Modes d'acquisition en protéomique de découverte :

La protéomique en mode découverte peut utiliser différents modes d'acquisition, parmi les plus utilisés on retrouve la Data dependent acquisition (DDA) et la Data independent acquisition (DIA)(Annexe C). Ces deux modes d'acquisition varient au niveau du fonctionnement de la MS pour acquérir des données liées à la protéomique. Pour être plus précis, la DDA permet de sélectionner les ions les plus intenses (Top-10 par exemple) lors du passage dans le premier quadrupole du spectromètre de masse pour les fragmenter et les analyser dans les deuxième et troisième quadrupoles. Au contraire la DIA, permet de définir des fenêtres de masse successives pour lesquelles tous les ions correspondant à cette fenêtre de masse seront fragmentés et analysés en MS/MS. La DDA a pour avantage de donner des données très précises pour les ions les plus intenses et la DIA permet quant à elle d'atteindre une profondeur dans le protéome plus importante. L'inconvénient de la DIA par rapport à la DDA c'est que les spectres obtenus en MS/MS sont multiplexés, c'est-à-dire qu'ils n'appartiennent pas uniquement à un peptide sélectionné lors de la MS. L'interprétation de ces spectres peut s'avérer complexe, cependant certaines méthodes de démultiplexage ont permis de rendre l'analyse de données DIA plus accessible <sup>319,320</sup>.



Annexe C : Schéma résumant les principes des modes d'acquisition DDA et DIA

Concernant le prétraitement des données, le logiciel Mascot (version 2.7.1.0, Matrix Science) a été utilisé pour identifier les protéines. Proline a été le logiciel choisi pour réaliser la quantification relative des protéines calculée à partir de l'addition des abondances des peptides spécifiques.

Les échantillons ont été analysés par microLC couplée à un spectromètre de masse en tandem (LC-MS/MS) (Ultimate 3000 et Qtrap6500+, Thermo Fisher Scientific et Ab Sciex). Les peptides ont été injectés et séparés sur une colonne C18 de 75  $\mu$ m 250 mm (Aurora Ultimate, 1,7  $\mu$ m, IonOpticks). La méthode micro-LC consistait en un gradient multilinéaire de 1 h à un débit de 300 nl/min, allant de 2 à 41 % d'acétonitrile dans 0,1 % d'acide formique. La température de la colonne a été maintenue à 40 °C à l'aide d'un four à colonne intégré (PRSO-V2, Sonation). Le voltage du spray été réglé à 2kV et le capillaire chauffé à 285°C. La méthode d'acquisition SRM ciblait les ions et les peptides dont les masses ont préalablement été renseignées. Pour la méthode en scheduled SRM, les temps de rétention ont également été renseignés afin d'optimiser l'analyse. La SRM a été acquise avec une résolution de 30 000 après l'accumulation de 5x10<sup>5</sup> ions (temps de remplissage maximal de 100 ms). La largeur d'isolation a été fixée à  $\pm 0,2$  m/z et l'énergie de collision normalisée à 30 %.

Annexe D : Détails de l'injection LC-MS (SRM) pour la quantification de peptides du projet NAFLD.

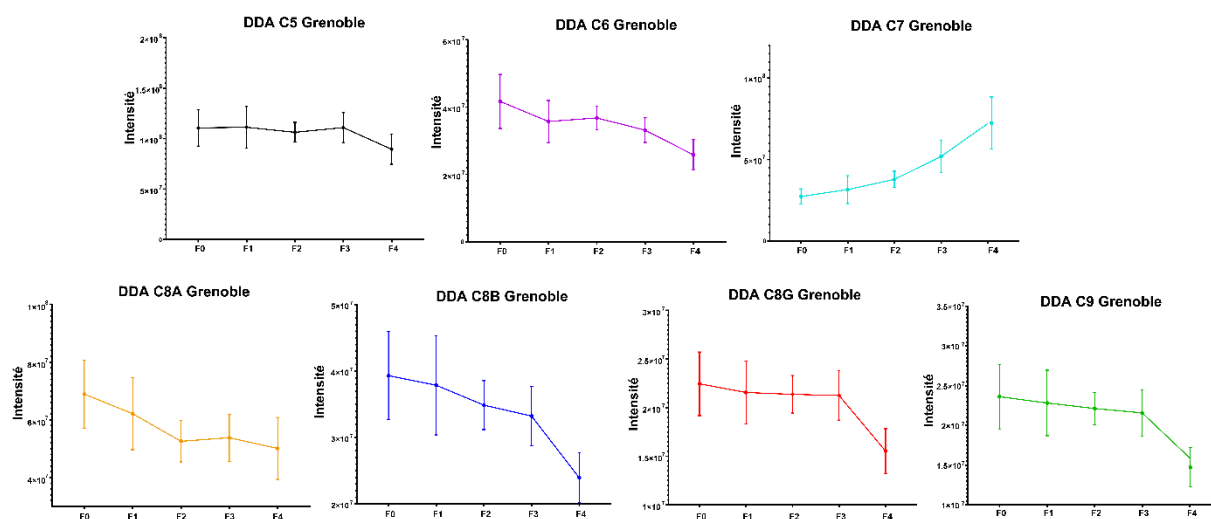
Annexe E : Tableau avec la masse et le temps de rétention moyen de chaque protéine, peptide et fragment choisis pour l'étude SRM (en jaune la transition utilisée pour la quantification).

Protein_Name (Uniprot)	MRM targetted peptides	Fragment ions					
		light			heavy		
		transitions	Mz	Average Retention Time	transitions	Mz	Average Retention Time
C7_HUMAN	ENPLTQAVPK	P - y8+	853.514179	15.09	P - y8+	861.528378	15.09
		T - y6+	643.377351	15.09	T - y6+	651.39155	15.09
		P - y2+	244.165568	15.09	P - y2+	252.179767	15.09
	SSGWHFVVK	G - y7+	872.477734	17.22	G - y7+	880.491933	17.22
		H - y5+	629.376957	17.22	H - y5+	637.391156	17.22
		F - y4+	492.318046	17.22	F - y4+	500.332245	17.22
	YSAWAESVTNLPQVIK	T - y8+	912.551293	31.68	T - y8+	920.565492	31.68
		N - y7+	811.503615	31.68	N - y7+	819.517814	31.68
		P - y5+	584.376623	31.68	P - y5+	592.390822	31.68
LAMP2_HUMAN	IPLNDLFR	P - y7+	874.478128	28.39	P - y7+	884.486397	28.39
		L - y6+	777.425364	28.39	L - y6+	787.433633	28.39
		N - y5+	664.3413	28.39	N - y5+	674.349569	28.39
HPTR_HUMAN	SDLGAVISLLLWGR	L - y5+	644.387856	41.16	L - y5+	654.396125	41.16
		L - y4+	531.303792	41.16	L - y4+	541.312061	41.16
		W - y3+	418.219728	41.16	W - y3+	428.227997	41.16
	VTSIQHWVQK	S - y8+	1025.55269	13.03	S - y8+	1033.56688	13.03
		Q - y6+	825.436598	13.03	Q - y6+	833.450797	13.03
		H - y5+	697.37802	13.03	H - y5+	705.392219	13.03
APOD_HUMAN	NILTSNNIDVK	L - y9+	1003.541851	17.52	L - y9+	1011.55605	17.52
		T - y8+	890.457787	17.52	T - y8+	898.471986	17.52
		S - y7+	789.410108	17.52	S - y7+	797.424307	17.52
A POM_HUMAN	AFLLTPR	L - y5+	599.387522	20.6	L - y5+	609.395791	20.6
		L - y4+	486.303458	20.6	L - y4+	496.311727	20.6
		T - y3+	373.219394	20.6	T - y3+	383.227663	20.6
	WIYHLTEGSTDLR	G - y6+	648.33113	21.1	G - y6+	658.339399	21.1
		S - y5+	591.309666	21.1	S - y5+	601.317935	21.1
		L - y2+	288.203016	21.1	L - y2+	298.211285	21.1
APOL1_HUMAN	VTEPISAESGEQVER	S - y7+	804.384622	12.13	S - y7+	814.392891	12.13
		V - y3+	403.229959	12.13	V - y3+	413.238228	12.13
	WWTQAQAHDLVIK	A - y7+	795.472314	22.88	A - y7+	803.486513	22.88
		D - y5+	587.376289	22.88	D - y5+	595.390488	22.88
		L - y4+	472.349346	22.88	L - y4+	480.363545	22.88
ALS_HUMAN	AFWLDVSHNR	L - y7+	840.432241	22.05	L - y7+	850.44051	22.05
		D - y6+	727.348177	22.05	D - y6+	737.356446	22.05
		V - y5+	612.321233	22.05	V - y5+	622.329502	22.05
	ANVFVQLPR	F - y6+	759.451185	22.11	F - y6+	769.459454	22.11
		V - y5+	612.382771	22.11	V - y5+	622.39104	22.11
		L - y3+	385.25578	22.11	L - y3+	395.264049	22.11
FA5_HUMAN	EDGILGPIIR	L - y6+	668.445371	26.09	L - y6+	678.45364	26.09
		G - y5+	555.361307	26.09	G - y5+	565.369576	26.09
		P - y4+	498.339844	26.09	P - y4+	508.348113	26.09
	SWWGDYWEFPR	G - y8+	1069.47377	39.58	G - y8+	1079.48204	39.58
		E - y4+	548.282723	39.58	E - y4+	558.290992	39.58
		P - y3+	419.24013	39.58	P - y3+	429.248399	39.58
	WIISLTPK	I - y7+	745.445431	27.07	I - y7+	753.45963	27.07



		S - y6+	632.361367	27.07	S - y6+	640.375566	27.07
		S - y5+	545.329339	27.07	S - y5+	553.343538	27.07
PI16_HUMAN	AQVSPTASDMLHMR	S - y7+	889.401868	17.12	S - y7+	899.410137	17.12
		M - y5+	687.342897	17.12	M - y5+	697.351166	17.12
	WDEELAFAK	E - y8+	878.461809	25.57	E - y8+	886.476008	25.57
		A - y5+	507.292559	25.57	A - y5+	515.306758	25.57
A - y4+		436.255445	25.57	A - y4+	444.269644	25.57	
ENPP2_HUMAN	WWGGQPLWITATK	G - y11+	1171.64698	35.69	G - y11+	1179.66118	35.69
		P - y8+	929.545479	35.69	P - y8+	937.559678	35.69
		T - y4+	420.245275	35.69	T - y4+	428.259474	35.69
KLKB1_HUMAN	IAYGTQGSSGYSLR	G - y11+	1112.53307	14.35	G - y11+	1122.54134	14.35
		Q - y9+	954.463935	14.35	Q - y9+	964.472204	14.35
		G - y8+	826.405357	14.35	G - y8+	836.413626	14.35
	IYSGILNLSDITK	S - y11+	1160.65212	31.64	S - y11+	1168.66632	31.64
		L - y8+	903.514573	31.64	L - y8+	911.528772	31.64
		N - y7+	790.430509	31.64	N - y7+	798.444708	31.64
	YSPGGTPTAIK	P - y9+	841.477794	11.34	P - y9+	849.491993	11.34
		G - y8+	744.42503	11.34	G - y8+	752.439229	11.34
		G - y7+	687.403566	11.34	G - y7+	695.417765	11.34
QSOX1_HUMAN	LEEIDGFFAR	E - y8+	954.467957	27.85	E - y8+	964.476226	27.85
		I - y7+	825.425364	27.85	I - y7+	835.433633	27.85
		D - y6+	712.3413	27.85	D - y6+	722.349569	27.85
	SFYTAYLQR	Y - y7+	914.473043	22.68	Y - y7+	924.481312	22.68
		T - y6+	751.409714	22.68	T - y6+	761.417983	22.68
A - y5+	650.362036	22.68	A - y5+	660.370305	22.68		
TTHY_HUMAN	AADDTWEPFASGK	W - y8+	921.446494	22.13	W - y8+	929.460693	22.13
		E - y7+	735.367181	22.13	E - y7+	743.38138	22.13
		P - y6+	606.324588	22.13	P - y6+	614.338787	22.13
	GSPAINVAVHVFR	A - y6+	728.420219	22.3	A - y6+	738.428488	22.3
		V - y5+	657.383105	22.3	V - y5+	667.391374	22.3
		H - y4+	558.314692	22.3	H - y4+	568.322961	22.3
ZPI_HUMAN	ETSNFGFSLLR	F - y7+	839.4774	30.35	F - y7+	849.485669	30.35
		G - y6+	692.408986	30.35	G - y6+	702.417255	30.35
		S - y4+	488.319108	30.35	S - y4+	498.327377	30.35
	IFSPFADLSATSATGR	L - y6+	604.3413	38.25	L - y6+	614.349569	38.25
		S - y5+	491.257236	38.25	S - y5+	501.265505	38.25
		A - y4+	404.225208	38.25	A - y4+	414.233477	38.25
	NLELGLTQGSFAFIHK	T - y10+	1135.58947	31.06	T - y10+	1143.60366	31.06
		G - y8+	906.483214	31.06	G - y8+	914.497413	31.06
		A - y5+	615.361307	31.06	A - y5+	623.375506	31.06

Annexe F : Distribution des coefficients de variation de l'intensité pour les données SRM (étude de vérification).



Annexe G : Courbes par stade de fibrose des protéines C5, C6, C7, C8A, C8B, C8G et C9 obtenues par DDA lors de l'étude de découverte sur la cohorte de Grenoble.

Annexe H : Tableau de comparaison de l'AUROC obtenue sur plusieurs tests de 1ère ligne, de panels issus de tests de 1ère ligne, sur des tests de 2ème ligne et sur les biomarqueurs et panels développés dans le projet MASLD. L'AUROC a été calculée sur la cohorte de validation (cohorte d'Angers) et indique la performance de chaque tests pour prédire une fibrose activité, inflammation lobulaire, ballonnement des hépatocytes et MASH précoces ou avancée chez des patients diagnostiqués MASLD. HA : acide hyaluronique; INR : international normalized ratio; NFS (NAFLD fibrosis score) : age + IMC + diabète + ASAT/ALAT + plaquettes + albumine; APRI : AST to platelet ratio index; AAR : AST to ALT ratio index; FibroQ : INR + ASAT + ALAT + plaquettes + âge; Forns index : plaquettes + âge + GGT + cholestérol, HUI score : IMC + bilirubine + albumine + plaquettes; LOK index : plaquettes + AST + INR ; F : Fibrose ; S : Stéatose ; A : Activité ; Inf. lob. : Inflammation lobulaire ; Ball : ballonnement.

	F0-2 vs F3-4	S0-1 vs S2-3	A0-1 vs A2-3	Inf. lob. 0-1 vs 2	Ball 0-1 vs 2	NASH
ALS	0,742	0,561	0,685	0,697	0,629	0,669
LG3BP	0,745	0,612	0,628	0,635	0,62	0,614
ALS;LG3BP	0,804	0,632	0,707	0,723	0,668	0,693
A2M	0,741	0,481	0,641	0,639	0,609	0,615
GGT	0,663	0,525	0,517	0,58	0,506	0,531
A2M;GGT	0,768	0,485	0,619	0,644	0,583	0,599
FibroTest	0,758	0,632	0,658	0,646	0,583	0,631
A2M;GGT;ALS.LG3BP	0,835	0,57	0,693	0,719	0,65	0,677
ASAT	0,723	0,654	0,71	0,731	0,677	0,724
ALAT	0,628	0,638	0,676	0,65	0,655	0,691
Plaquettes	0,492	0,458	0,569	0,565	0,594	0,562
Fib-4	0,641	0,515	0,626	0,655	0,635	0,62
Fibroscan médiane	0,856	0,646	0,8	0,796	0,781	0,787
Stéatose	0,602		0,75	0,739	0,753	0,73
Ballonnement	0,737	0,639	0,941	0,797		0,891
Inflammation	0,66	0,659	0,864		0,771	0,871
Activité	0,764	0,684		0,957	0,98	0,98
Fibrose		0,631	0,81	0,784	0,804	0,782
HA	0,738	0,539	0,643	0,627	0,616	0,624
INR	0,726	0,504	0,703	0,632	0,721	0,656
NFSA	0,676	0,516	0,696	0,715	0,691	0,711
APRI	0,696	0,606	0,538	0,531	0,469	0,533
AAR	0,581	0,538	0,541	0,478	0,54	0,515
FibroQ	0,631	0,574	0,531	0,571	0,542	0,52
Forns Index	0,624	0,539	0,662	0,617	0,696	0,631
HUI score	0,577	0,511	0,548	0,604	0,551	0,549
LOK Index	0,677	0,538	0,598	0,618	0,573	0,597

Protocole Digestion Aldolase B Light et PSAQ<sup>+1</sup> en solution :

Réduction/dénaturation :

Mix 2 (ALDOB light / PSAQ <sup>+1</sup> )	Quantité / concentration finale	Volume à prélever (→ Volume final = 100µl)
---	---------------------------------	---

Aldolase-B Light (~ 0,27µg/µl)	1,48µg (estimation)	5,5µl
PSAQ <sup>+1</sup> (~ 1,47µg/µl → AAA)	1,47µg (estimation)	1µl (Objectif ratio ~ 1:1)
TCEP (100mM) <i>réducteur</i>	10mM (il faut une concentration 10x plus élevée que de protéine)	10µL
Urée (8M) (guanidine hydrochloride pourrait remplacer mais les spectros n'aiment pas) <i>dénaturant</i>	2M	25µl
Bicarbonate (50mM) (il faut travailler en pH neutre pour que tout se déroule correctement)	50mM	67,5 → 58,5 µl Mix 2

- 30 minutes en agitation à 20°C 600 rpm.

#### Alkylation:

Solution mère	Concentration finale	Volume à prélever (→ Volume final = 125µl)
Chloroacétamide (500mM) <i>Alkylant</i>	100mM (il faut une concentration 10x plus élevée que de réducteur)	25µl à ajouter dans les solutions précédentes

- 30 minutes en agitation à 20°C 600 rpm.

#### Digestion:

Solution mère	Quantité finale	
	Protéines (Mix 2) : 2,98µg/100µl	Enzyme : 3,75µl
Bicarbonate (50mM) (il faut travailler en pH neutre pour que tout se déroule correctement)	50mM (conc° finale TCEP : 4,72mM) (conc° finale urée : 0,94mM < 1mM) (conc° finale CAA : 47,2mM)	140µL

- 1) Digestion overnight à 37° à l'étuve sous agitation ou 2h si rapport enzyme/prot est de 1 :20
- 2) Centrifugation 14000g /12' /4°C (que si manip dans tubes FASP pour laisser passer peptides).
- 3) Dessalage :

#### Matériel :

- Harvard Ultra-Macro Spin columns (référence : 74-4101)
- Tampon A (5% ACN, 0,1% FA, eau milliQ Q.S.P)
- Tampon B1 (50% ACN, 0,1% FA, eau milliQ Q.S.P)
- Tampon B2 (75% ACN, 0,1% FA, eau milliQ Q.S.P)

*Annexe I : Protocole de réduction, dénaturation, alkylation, digestion et dessalage pour le mix ALDOB sans marquage et PSAQ<sup>+1</sup>*

Les échantillons ont été analysés par nanoLC couplée à un spectromètre de masse en tandem (LC-MS/MS) (Ultimate 3000 RSLCnano and the Exploris 480, Thermo Fisher Scientific). Les peptides ont été injectés sur une précolonne PepMap C18 Precolumn (Thermo Fisher Scientific) et séparés sur une colonne C18 de 75 µm 250 mm (Aurora Ultimate, 1,7 µm, IonOpticks). La méthode nano-LC consistait en un gradient multilinéaire de 30 min à un débit de 300 nl/min, allant de 2 à 41 % d'acétonitrile dans

0,1 % d'acide formique. La température de la colonne a été maintenue à 40 °C à l'aide d'un four à colonne intégré (PRSO-V2, Sonation). Le voltage du spray été réglé a 2kV et le capillaire chauffé à 285°C. La méthode d'acquisition ciblée a combiné deux « scan events » correspondant à un « full scan » et une PRM ciblant le deuxième isotope des peptides endogènes. Les spectres en full-scan ont été acquis avec une résolution de 60 000 après l'accumulation de  $10^6$  ions (temps de remplissage maximal de 200 ms). L'événement PRM a été acquis avec une résolution de 30 000 après l'accumulation de  $5 \times 10^5$  ions (temps de remplissage maximal de 100 ms). La largeur d'isolation a été fixée à  $\pm 0,2$  m/z et l'énergie de collision normalisée à 30 %.

En fonction des manip, différents paramètres ont été modifiés comme l'énergie d'ionisation, l'isotope ciblé ou la largeur de la fenêtre de masse.

*Annexe J : Détails de l'injection LC-MS du mix ALDOB sans marquage et PSAQ<sup>+1</sup>.*

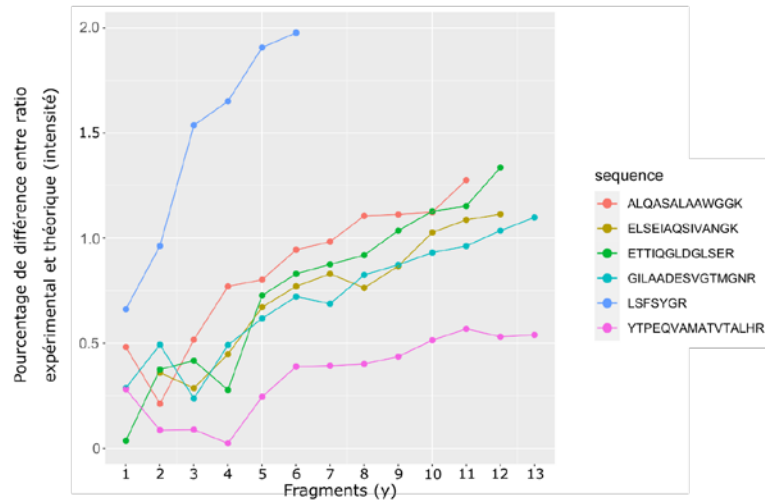
*Annexe K : Comparaison du pourcentage de différence pour chaque fragment des peptides ALQASALAAWGGK, ELSEIAQSIVANGK, ETTIQGLDGLSER, GILAADESVGTMGNR, LFSYGR, YTPEQVAMATVTALHR sans et avec retranchement du bruit de chaque fragment sur l'intensité maximale de chaque fragment.*

Peptide	Fragment	% de différence sans retrancher le bruit	% de différence en retranchant le bruit
---------	----------	--	---

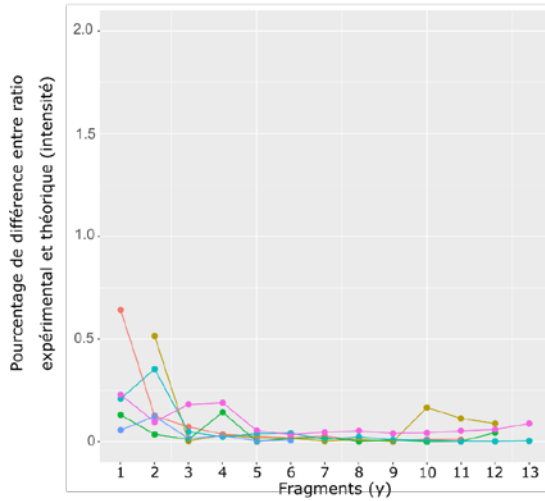
ALQASALAAWGGK	y1	/	/
	y2	61.22%	42.22%
	y3	27.76%	23.82%
	y4	17.67%	16.53%
	y5	16.33%	15.11%
	y6	6.80%	6.12%
	y7	10.53%	9.80%
	y8	1.50%	0.80%
	y9	0.48%	0.75%
	y10	1.26%	1.03%
	y11	6.96%	7.50%
	y12	/	/
	<b>Total</b>	<b>15,05%</b>	<b>12,37%</b>
ELSEIAQSIVANGK	y1	100.00%	100.00%
	y2	48.14%	28.41%
	y3	33.66%	26.15%
	y4	26.53%	23.39%
	y5	20.44%	17.81%
	y6	19.99%	15.28%
	y7	12.50%	11.10%
	y8	12.75%	11.31%
	y9	5.24%	4.86%
	y10	7.09%	6.10%
	y11	5.28%	6.55%
	y12	0.68%	0.10%
	<b>Total</b>	<b>24,36%</b>	<b>20,92%</b>
ETTIQGLDGLSER	y1	33.44%	33.44%
	y2	72.40%	19.38%
	y3	30.40%	22.63%
	y4	33.68%	9.35%
	y5	13.47%	12.54%
	y6	10.81%	9.52%
	y7	8.61%	6.74%
	y8	6.02%	5.82%
	y9	3.45%	3.26%
	y10	0.43%	0.12%
	y11	1.20%	0.51%
	y12	/	/
	<b>Total</b>	<b>19,45%</b>	<b>10,82%</b>
GILAADESVGTMGNR	y1	/	/
	y2	64.51%	64.51%
	y3	235.03%	100.00%
	y4	35.69%	23.92%
	y5	19.43%	12.15%
	y6	13.75%	8.78%
	y7	25.06%	24.25%
	y8	25.84%	24.26%
	y9	17.86%	17.20%
	y10	8.80%	8.25%
	y11	11.31%	11.01%
	y12	7.19%	6.98%
	y13	8.11%	7.88%
	y14	3.02%	3.61%
	<b>Total</b>	<b>36,58%</b>	<b>24,06%</b>
LSFSYGR	y1	33.45%	33.45%
	y2	33.78%	14.23%
	y3	10.32%	7.59%
	y4	3.09%	2.46%
	y5	2.64%	2.86%
	y6	1.84%	1.99%
	<b>Total</b>	<b>19,45%</b>	<b>10,43%</b>
YTPEQVAMATVTALHR	y1	55.26%	10.09%

	y2	36.54%	8.05%
	y3	33.95%	20.41%
	y4	39.15%	32.20%
	y5	8.96%	7.59%
	y6	11.34%	8.25%
	y7	1.54%	0.18%
	y8	3.48%	4.74%
	y9	7.39%	5.97%
	y10	1.72%	0.92%
	y11	2.23%	0.42%
	y12	3.16%	5.64%
	y13	12.30%	13.66%
	<b>Total</b>	<b>16.69%</b>	<b>9.09%</b>

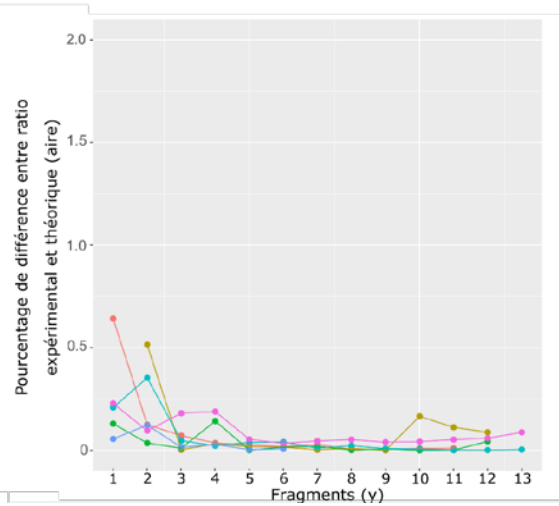
A



B<sub>1</sub>

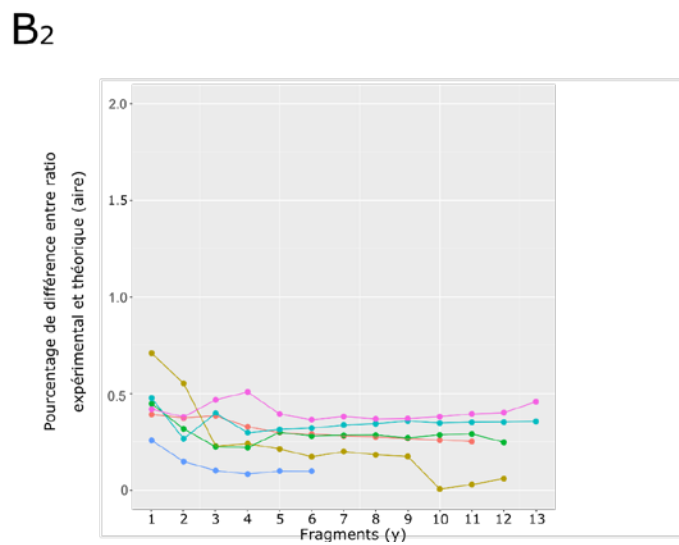
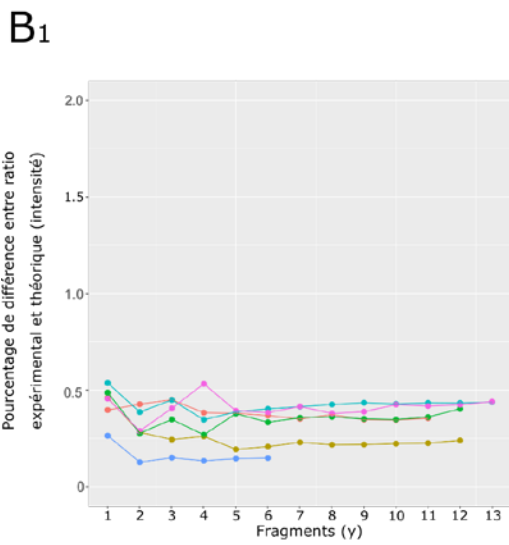
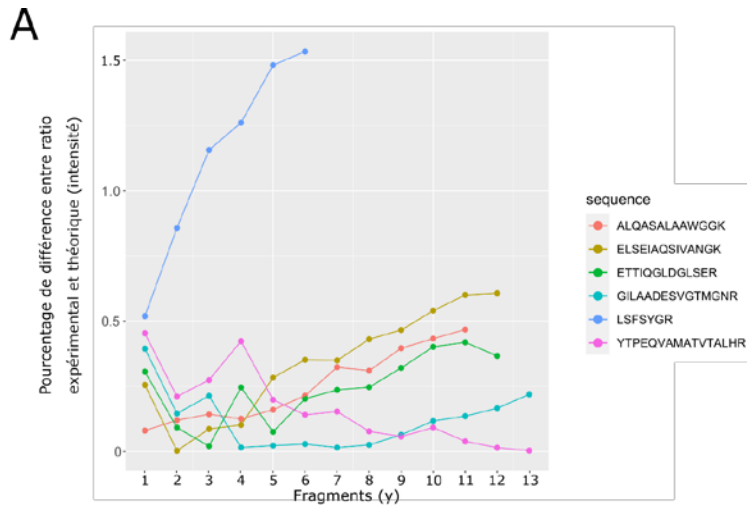


B<sub>2</sub>



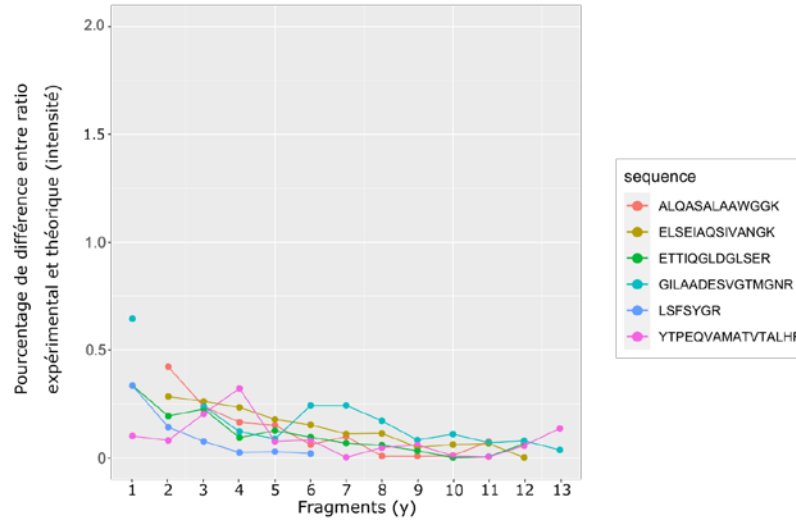
Annexe L : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 0,4m/z. A. Pourcentages de différence entre ratios théoriques non corrigés et ratios expérimentaux de l'intensité. B<sub>1</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'intensité. B<sub>2</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'aire.



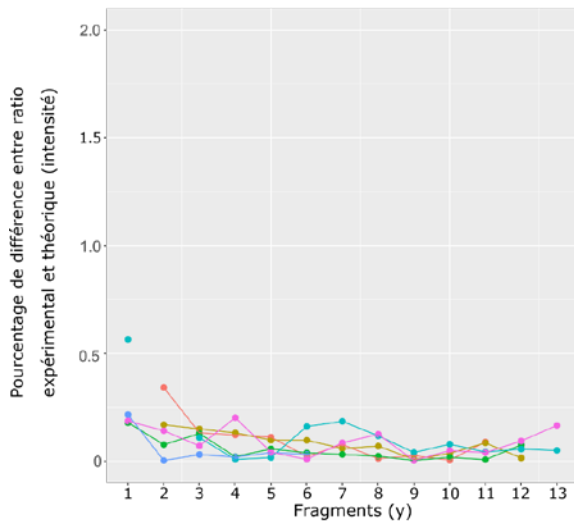


*Annexe M* : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 0,8 m/z. A. Pourcentages de différence entre ratios théoriques non corrigés et ratios expérimentaux de l'intensité. B<sub>1</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'intensité. B<sub>2</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'aire.

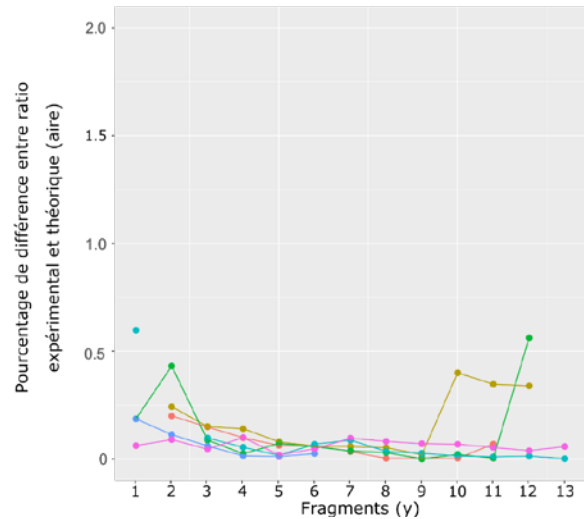
A



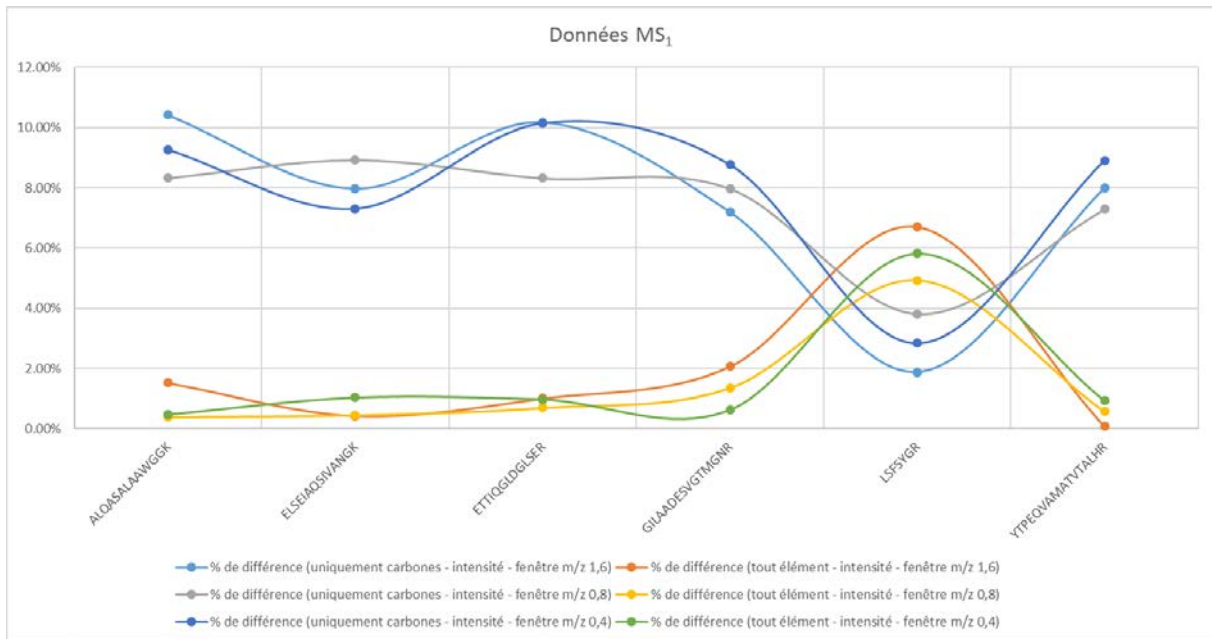
B<sub>1</sub>



B<sub>2</sub>

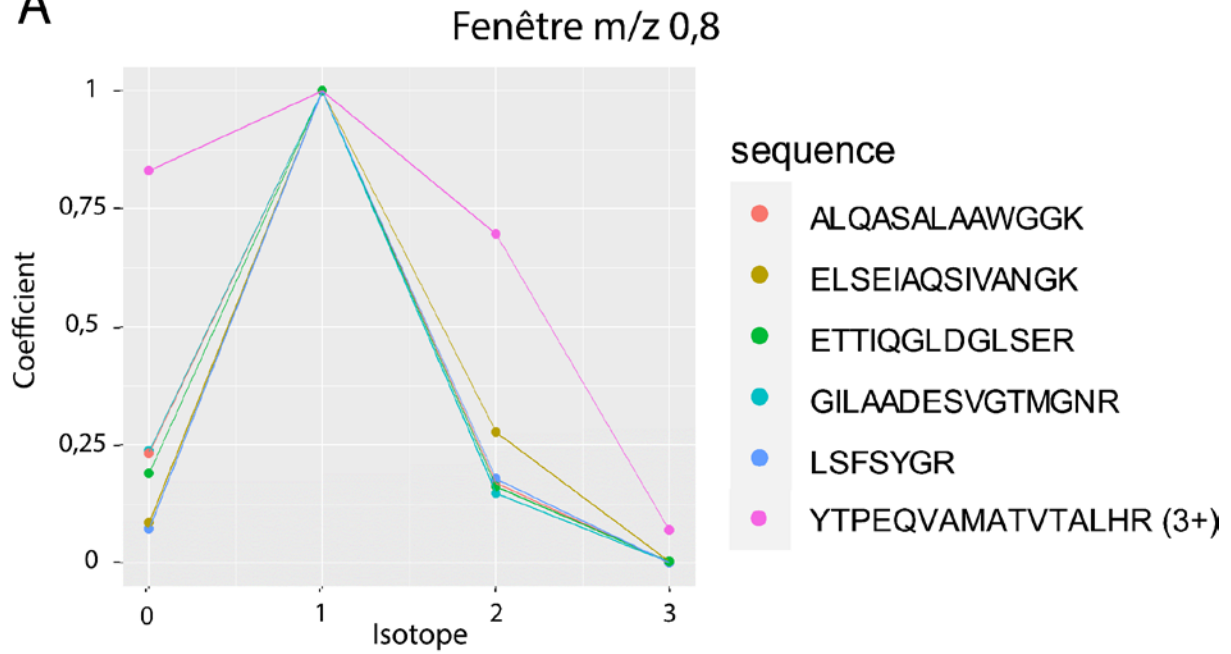


Annexe N : Pourcentages de différence entre ratios théoriques et expérimentaux pour tous les fragments des peptides de l'ALDOB pour une fenêtre de sélection 1,6 m/z. A. Pourcentages de différence entre ratios théoriques non corrigés et ratios expérimentaux de l'intensité. B<sub>1</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'intensité. B<sub>2</sub>. Pourcentages de différence entre ratios théoriques corrigés et ratios expérimentaux de l'aire.



Annexe O : Pourcentage de différence entre ratio expérimental et théorique sur les données d'intensité obtenues en MS<sub>1</sub> pour chacun des peptides de l'ALDOB sur 3 fenêtres de sélection m/z (0,4 – 0,8 – 1,6).

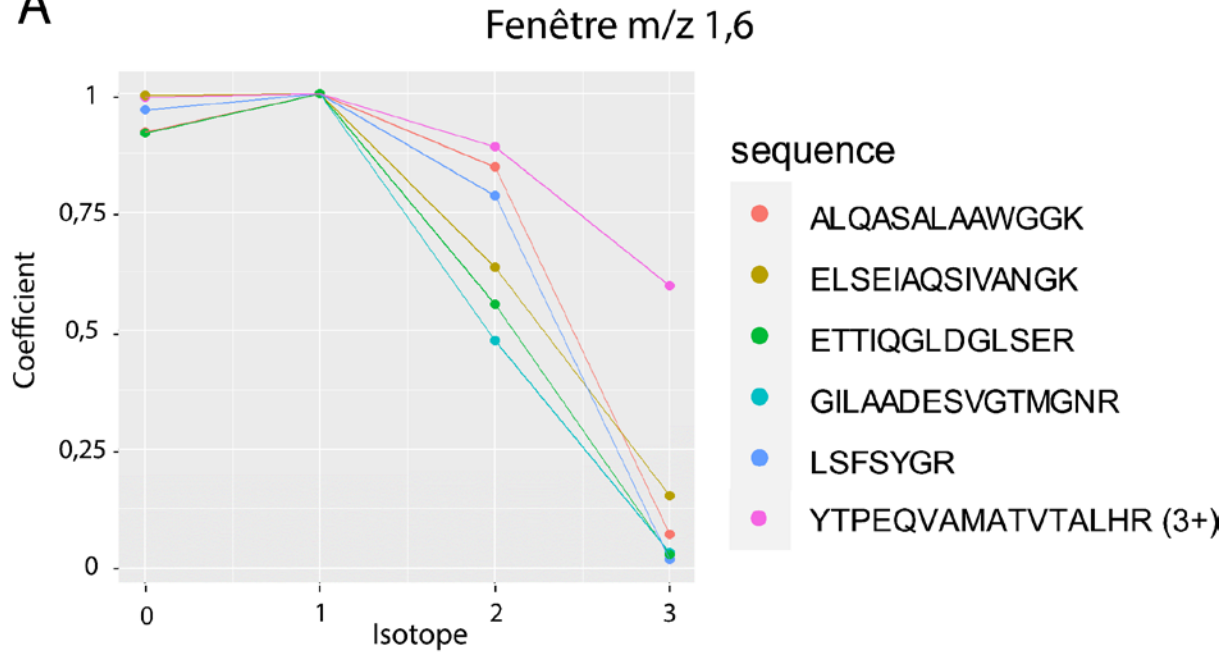
**A**



<b>B</b>		<b>C</b>	
Isotope	Coefficient moyen	Isotope	Coefficient moyen
$^{12}\text{C}$ full	0.1634	$^{12}\text{C}$ full	0,8311
1x $^{13}\text{C}$	1.0000	1x $^{13}\text{C}$	1,0000
2x $^{13}\text{C}$	0.1870	2x $^{13}\text{C}$	0,6964
3x $^{13}\text{C}$	0.0018	3x $^{13}\text{C}$	0,0691

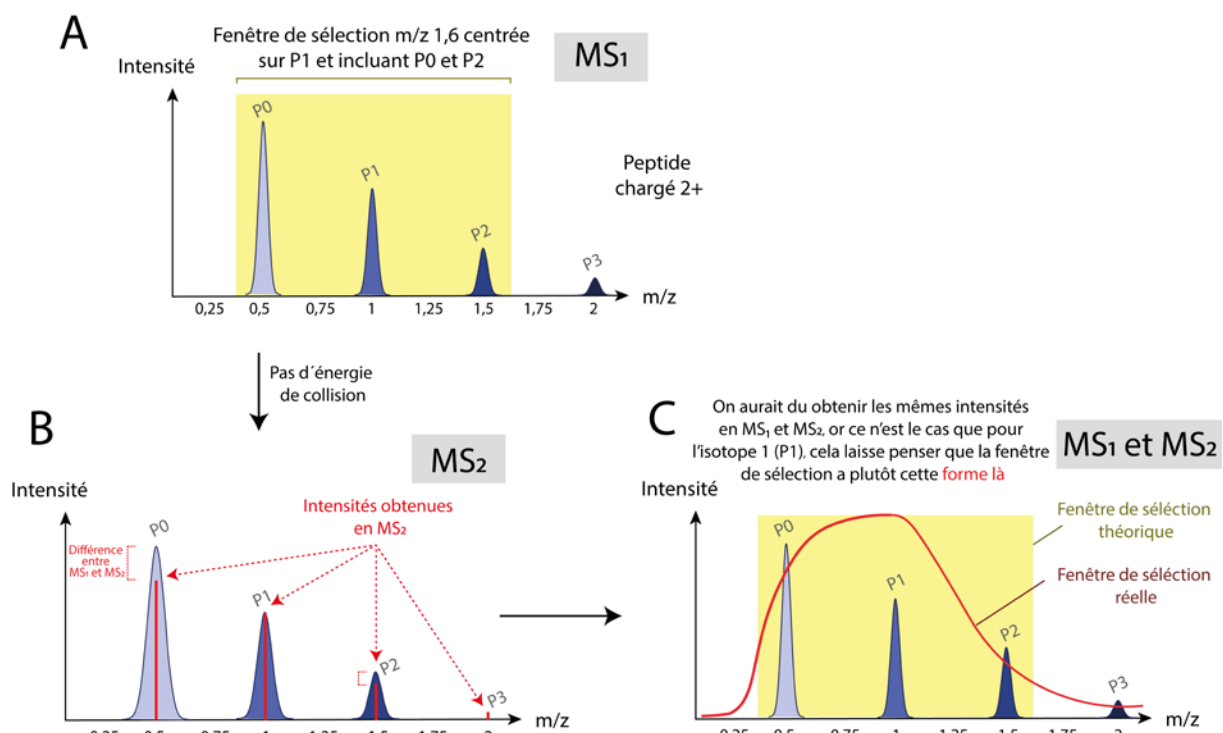
*Annexe P* : Comparaison des intensités obtenues en MS<sub>2</sub> lors d'une analyse de peptides de l'ALDOB en PRM et sans énergie de collision avec une sélection de l'isotope 1 et avec une fenêtre de sélection de 0,8Da. (A) Représentation graphique de la moyenne des intensités obtenues pour les fragments de tous les isotopes de chaque peptide. (B) Tableau des moyennes des intensités obtenues pour les fragments de tous les isotopes de chaque peptide chargé 2+ et coefficients par rapport au pic le plus intense. (C) Tableau des moyennes des intensités obtenues pour les fragments des isotopes du peptide chargé 3+ (YTPEQVAMATVTALHR) et coefficients par rapport au pic le plus intense.

**A**



<b>B</b>		<b>C</b>	
Isotope	Coefficient moyen	Isotope	Coefficient moyen
$^{12}\text{C}$ full	0.9530	$^{12}\text{C}$ full	0,9924
1x $^{13}\text{C}$	1.0000	1x $^{13}\text{C}$	1,0000
2x $^{13}\text{C}$	0.6597	2x $^{13}\text{C}$	0,8882
3x $^{13}\text{C}$	0.0596	3x $^{13}\text{C}$	0,5943

*Annexe Q* : Comparaison des intensités obtenues en  $\text{MS}_2$  lors d'une analyse de peptides de l'ALDOB en PRM et sans énergie de collision avec une sélection de l'isotope 1 et avec une fenêtre de sélection de 1,6Da. (A) Représentation graphique de la moyenne des intensités obtenues pour les fragments de tous les isotopes de chaque peptide. (B) Tableau des moyennes des intensités obtenues pour les fragments de tous les isotopes de chaque peptide chargé 2+ et coefficients par rapport au pic le plus intense. (C) Tableau des moyennes des intensités obtenues pour les fragments des isotopes du peptide chargé 3+ (YTPEQVAMATVTALHR) et coefficients par rapport au pic le plus intense.



*Annexe R* : Représentation des pics obtenus lors d'une analyse d'un peptide chargé 2+ sans énergie de collision en MS<sub>1</sub> et en MS<sub>2</sub> avec une fenêtre de sélection  $m/z$  1,6 centrée sur P1 et représentation théorique de la forme « réelle » de la fenêtre de sélection.

Annexe S : m/z des fragments inclus dans différentes conditions de sélection.

Conditions	A Totalité des fragments	B m/z fragment > m/z précurseur	C 3 fragments les plus intenses dont m/z > m/z précurseur	D 5 fragments les plus intenses	E 3 fragments les plus intenses
ALQASALAAWGGK	148.1162 205.1376 262.1591 448.2384 519.2755 590.3126 703.3967 774.4338 861.4658 932.5029 1060.5615 1173.6456	703.3967 774.4338 861.4658 932.5029 1060.5615 1173.6456	703.3967 861.4658 932.5029	262.1591 448.2384 519.2755 590.3126 861.4658	262.1591 448.2384 590.3126
ELSEIAQSIVANGK	148.1162 205.1376 319.1805 390.2177 489.2861 602.3701 689.4022 817.4607 888.4979 1001.5819 1130.6245 1217.6565 1330.7406	817.4607 888.4979 1001.5819 1130.6245 1217.6565 1330.7406	817.4607 888.4979 1001.5819	319.1805 390.2177 489.2861 689.4022 888.4979	390.2177 489.2861 888.4979
ETTIQGLDGLSER	176.1223 305.1649 392.1969 505.281 562.3025 677.3294 790.4135 847.4349 975.4935 1088.5776 1189.6252 1290.6729	790.4135 847.4349 975.4935 1088.5776 1189.6252 1290.6729	790.4135 847.4349 975.4935	176.1223 562.3025 677.3294 847.4349 975.4935	562.3025 847.4349 975.4935
GILAADESVGTMGNR	176.1223 290.1652 347.1867 478.2272 579.2749 636.2963 735.3647 822.3968 951.4394 1066.4663 1137.5034 1208.5405 1321.6246 1434.7087	822.3968 951.4394 1066.4663 1137.5034 1208.5405 1321.6246 1434.7087	822.3968 1066.4663 1137.5034	636.2963 822.3968 951.4394 1066.4663 1137.5034	636.2963 822.3968 1137.5034
LSFSYGR	176.1223 233.1438 396.2071 483.2391 630.3075 717.3396	483.2391 630.3075 717.3396	483.2391 630.3075 717.3396	176.1223 396.2071 483.2391 630.3075 717.3396	483.2391 630.3075 717.3396

Annexe T : Coefficients de variation moyens des conditions C et E pour les différentes concentrations et répliques de la courbe de calibration.

Index	E 3 fragments les plus intenses	
	CV moyen	
10 pg/μl - Réplica 1	11.95	10.12
10 pg/μl - Réplica 2	15.83	15.87
10 pg/μl - Réplica 3	8.41	7.22
10 pg/μl - Réplica 4	13.98	12.92
100 pg/μl - Réplica 1	5.45	4.32
100 pg/μl - Réplica 2	5.46	6.17
100 pg/μl - Réplica 3	4.71	4.75
100 pg/μl - Réplica 4	8.15	6.75
1 ng/μl - Réplica 1	3.35	2.48
1 ng/μl - Réplica 2	2.87	2.78
1 ng/μl - Réplica 3	3.81	2.79
1 ng/μl - Réplica 4	3.70	2.67
10 ng/μl - Réplica 1	3.41	2.74
10 ng/μl - Réplica 2	3.41	2.74
10 ng/μl - Réplica 3	3.41	2.74
10 ng/μl - Réplica 4	2.10	2.02



## Références :

1. Asrani, S. K., Devarbhavi, H., Eaton, J. & Kamath, P. S. Burden of liver diseases in the world. *J Hepatol* **70**, 151–171 (2019).
2. Kiernan, F. & Green, J. H. XXIX. The anatomy and physiology of the liver. *Philos. Trans. R. Soc. Lond.* **123**, 711–770 (1833).
3. Cunningham, R. P. & Porat-Shliom, N. Liver Zonation - Revisiting Old Questions With New Technologies. *Front Physiol* **12**, 732929 (2021).
4. Roberts, H. R. & Cederbaum, A. I. The liver and blood coagulation: physiology and pathology. *Gastroenterology* **63**, 297–320 (1972).
5. Gille, C. *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol* **6**, 411 (2010).
6. Kalra, A., Yetiskul, E., Wehrle, C. J. & Tuma, F. Physiology, Liver. in *StatPearls* (StatPearls Publishing, 2022).
7. Hao, Z., Jin, D.-Y., Stafford, D. W. & Tie, J.-K. Vitamin K-dependent carboxylation of coagulation factors: insights from a cell-based functional study. *Haematologica* **105**, 2164–2173 (2020).
8. Almazroo, O. A., Miah, M. K. & Venkataramanan, R. Drug metabolism in the liver. *Clin Liver Dis* **21**, 1–20 (2017).
9. Di Ciaula, A. *et al.* Bile Acid Physiology. *Ann Hepatol* **16 Suppl 1**, S4–S14 (2017).
10. Campbell, I. Liver: metabolic functions. *Anaesth Intensive Care Med* **7**, 51–54 (2006).
11. Smith, G. S., Walter, G. L. & Walker, R. M. Chapter 18 - Clinical Pathology in Non-Clinical Toxicology Testing. in *Haschek and Rousseaux's Handbook of Toxicologic Pathology (Third Edition)* (eds. Haschek, W. M., Rousseaux, C. G. & Wallig, M. A.) 565–594 (Academic Press, 2013).
12. Amitrano, L., Guardascione, M. A., Brancaccio, V. & Balzano, A. Coagulation disorders in liver disease. *Semin Liver Dis* **22**, 83–96 (2002).

13. Ekstedt, M. *et al.* Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology* **44**, 865–873 (2006).
14. Michalopoulos, G. K. & Bhushan, B. Liver regeneration: biological and pathological mechanisms and implications. *Nat Rev Gastroenterol Hepatol* **18**, 40–55 (2021).
15. Yuen, M.-F. *et al.* Hepatitis B virus infection. *Nat Rev Primer* **4**, 18035 (2018).
16. Theise, N., Bodenheimer, H. & Feerel, L. Acute and chronic viral hepatitis. *MacSweens Pathol. Liver* 361–401 (2011).
17. Cauza, E. *et al.* Screening for Wilson’s disease in patients with liver diseases by serum ceruloplasmin. *J Hepatol* **27**, 358–362 (1997).
18. Hanson, E. H., Imperatore, G. & Burke, W. HFE Gene and Hereditary Hemochromatosis: A HuGE Review. *Am J Epidemiol* **154**, 193–206 (2001).
19. Fromme, M., Schneider, C. V., Trautwein, C., Brunetti-Pierri, N. & Strnad, P. Alpha-1 antitrypsin deficiency: A re-surfacing adult liver disorder. *J Hepatol* **76**, 946–958 (2022).
20. Trail, F., Mahanti, N. & Linz, J. *Molecular biology of aflatoxin biosynthesis. Microbiology* vol. 141 (1995).
21. Altamirano, J. & Bataller, R. Alcoholic liver disease: pathogenesis and new targets for therapy. *Nat Rev Gastroenterol Hepatol* **8**, 491–501 (2011).
22. Kaplowitz, N. Drug-induced liver injury. *Clin Infect Dis* **38 Suppl 2**, S44-8 (2004).
23. El-Serag, H. B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264-1273.e1 (2012).
24. Heidelbaugh, J. J. & Bruderly, M. Cirrhosis and chronic liver failure: part I. Diagnosis and evaluation. *Am Fam Physician* **74**, 756–762 (2006).
25. White, D. L., Kanwal, F. & El-Serag, H. B. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. *Clin Gastroenterol Hepatol* **10**, 1342-1359.e2 (2012).

26. Global Burden of Disease Cancer Collaboration *et al.* Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* **3**, 524–548 (2017).
27. Rinella, M. E. *et al.* A multi-society Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* (2023).
28. Younossi, Z. M. *et al.* Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 73–84 (2016).
29. Malaguarnera, M., Di Rosa, M., Nicoletti, F. & Malaguarnera, L. Molecular mechanisms involved in NAFLD progression. *J Mol Med* **87**, 679–695 (2009).
30. Zakaria, Z. A. *et al.* Hepatoprotective activity of dried- and fermented-processed virgin coconut oil. *Evid Based Complement Altern. Med* **2011**, 142739 (2011).
31. Hübscher, S. G. Histological assessment of non-alcoholic fatty liver disease. *Histopathology* **49**, 450–465 (2006).
32. Cirrhose. *Wikipédia* (2021).
33. El Jabbour, T., Lagana, S. M. & Lee, H. Update on hepatocellular carcinoma: Pathologists' review. *World J Gastroenterol* **25**, 1653–1665 (2019).
34. Younossi, Z. M. *et al.* The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatol. Baltim. Md* **77**, 1335–1347 (2023).
35. Giugliano, D., Ceriello, A. & Esposito, K. Glucose metabolism and hyperglycemia. *Am J Clin Nutr* **87**, 217S-222S (2008).
36. Yuan, G., Al-Shali, K. Z. & Hegele, R. A. Hypertriglyceridemia: its etiology, effects and treatment. *CMAJ* **176**, 1113–1120 (2007).
37. Lazar, M. A. How obesity causes diabetes: not a tall tale. *Science* **307**, 373–375 (2005).

38. Parlati, L., Régnier, M., Postic, C. & Guillou, H. Physiopathologie de la stéatose hépatique non-alcoolique (NAFLD). *Médecine Mal. Métaboliques* (2023).
39. Mitra, S., De, A. & Chowdhury, A. Epidemiology of non-alcoholic and alcoholic fatty liver diseases. *Transl Gastroenterol Hepatol* **5**, 16 (2020).
40. Trépo, E. & Valenti, L. Update on NAFLD genetics: From new variants to the clinic. *J Hepatol* **72**, 1196–1209 (2020).
41. Maitiabola, G. *et al.* Proteome characteristics of liver tissue from patients with parenteral nutrition-associated liver disease. *Nutr Metab* **17**, 43 (2020).
42. Di Rosa, M. & Malaguarnera, L. Genetic variants in candidate genes influencing NAFLD progression. *J Mol Med* **90**, 105–118 (2012).
43. Gouda, W., Ashour, E., Shaker, Y. & Ezzat, W. MTP genetic variants associated with non-alcoholic fatty liver in metabolic syndrome patients. *Genes Dis* **4**, 222–228 (2017).
44. Anstee, Q. M. & Day, C. P. The Genetics of Nonalcoholic Fatty Liver Disease: Spotlight on PNPLA3 and TM6SF2. *Semin Liver Dis* **35**, 270–290 (2015).
45. Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M. & Sanyal, A. J. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med* **24**, 908–922 (2018).
46. Trauner, M., Arrese, M. & Wagner, M. Fatty liver and lipotoxicity. *Biochim Biophys Acta* **1801**, 299–310 (2010).
47. Sanyal, A. J. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* **16**, 377–386 (2019).
48. Parola, M. & Pinzani, M. Liver fibrosis: Pathophysiology, pathogenetic targets and clinical issues. *Mol Asp. Med* **65**, 37–55 (2019).
49. Guo, Y.-C. & Lu, L.-G. Antihepatic Fibrosis Drugs in Clinical Trials. *J Clin Transl Hepatol* **8**, 304–312 (2020).
50. Lemoine, S. & Friedman, S. L. New and emerging anti-fibrotic therapeutics entering or already in clinical trials in chronic liver diseases. *Curr Opin Pharmacol* **49**, 60–70 (2019).

51. Romero-Gómez, M., Zelber-Sagi, S. & Trenell, M. Treatment of NAFLD with diet, physical activity and exercise. *J Hepatol* **67**, 829–846 (2017).
52. Gadiparthi, C. *et al.* NAFLD Epidemiology, Emerging Pharmacotherapy, Liver Transplantation Implications and the Trends in the United States. *J Clin Transl Hepatol* **8**, 215–221 (2020).
53. Mundi, M. S. *et al.* Evolution of NAFLD and Its Management. *Nutr Clin Pr.* **35**, 72–84 (2020).
54. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69**, 89–95 (2001).
55. Mayeux, R. Biomarkers: potential uses and limitations. *NeuroRx* **1**, 182–188 (2004).
56. Vockley, J. *et al.* Phenylalanine hydroxylase deficiency: diagnosis and management guideline. *Genet Med* **16**, 188–200 (2014).
57. Chalasani, N. *et al.* The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* **67**, 328–357 (2018).
58. Festi, D. *et al.* Review article: the diagnosis of non-alcoholic fatty liver disease – availability and accuracy of non-invasive methods. *Aliment Pharmacol Ther* **37**, 392–400 (2013).
59. Hashimoto, E., Taniai, M. & Tokushige, K. Characteristics and diagnosis of NAFLD/NASH. *J Gastroenterol Hepatol* **28 Suppl 4**, 64–70 (2013).
60. Piccinino, F., Sagnelli, E., Pasquale, G. & Giusti, G. Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. *J Hepatol* **2**, 165–173 (1986).
61. Sumida, Y., Nakajima, A. & Itoh, Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol* **20**, 475–485 (2014).
62. Bedossa, P., Dargère, D. & Paradis, V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* **38**, 1449–1457 (2003).
63. Rousselet, M.-C. *et al.* Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* **41**, 257–264 (2005).

64. European Association for Study of Liver & Asociacion Latinoamericana para el Estudio del Higado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* **63**, 237–264 (2015).
65. Hernaez, R. *et al.* Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: A meta-analysis. *Hepatology* **54**, 1082–1090 (2011).
66. Armstrong, M. J. *et al.* Severe asymptomatic non-alcoholic fatty liver disease in routine diabetes care; a multi-disciplinary team approach to diagnosis and management. *QJM* **107**, 33–41 (2014).
67. Xu, R., Pan, J., Zhou, W., Ji, G. & Dang, Y. Recent advances in lean NAFLD. *Biomed Pharmacother* **153**, 113331 (2022).
68. Saadeh, S. *et al.* The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* **123**, 745–750 (2002).
69. Petzold, G. Role of Ultrasound Methods for the Assessment of NAFLD. *J Clin Med Res* **11**, (2022).
70. Dasarathy, S. *et al.* Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. *J Hepatol* **51**, 1061–1067 (2009).
71. Younossi, Z. M. *et al.* Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. *Mod Pathol* **11**, 560–565 (1998).
72. Ratziu, V. *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* **128**, 1898–1906 (2005).
73. Hernandez Roman, J. & Siddiqui, M. S. The role of noninvasive biomarkers in diagnosis and risk stratification in nonalcoholic fatty liver disease. *Endocrinol Diab Metab* **101**, 477 (2020).
74. Brunt, E. M., Janney, C. G., Di Bisceglie, A. M., Neuschwander-Tetri, B. A. & Bacon, B. R. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol* **94**, 2467–2474 (1999).

75. Nassir, F., Rector, R. S., Hammoud, G. M. & Ibdah, J. A. Pathogenesis and Prevention of Hepatic Steatosis. *Gastroenterol Hepatol* **11**, 167–175 (2015).
76. Caldwell, S. *et al.* Hepatocellular ballooning in NASH. *J Hepatol* **53**, 719–723 (2010).
77. El Ray, A. *et al.* Usefulness of the SAF score to characterize NAFLD/NASH in non-cirrhotic HCV patients. *Egypt. Liver J.* **12**, 1–7 (2022).
78. Dulai, P. S. *et al.* Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatol. Baltim. Md* **65**, 1557–1565 (2017).
79. Younossi, Z. M. *et al.* Pathologic criteria for nonalcoholic steatohepatitis: Interprotocol agreement and ability to predict liver-related mortality. *Hepatology* vol. 53 (2011).
80. Lambrecht, J., van Grunsven, L. A. & Tacke, F. Current and emerging pharmacotherapeutic interventions for the treatment of liver fibrosis. *Expert Opin Pharmacother* **21**, 1637–1650 (2020).
81. Ratziu, V. *et al.* Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol* **6**, 6 (2006).
82. Angulo, P. *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* **45**, 846–854 (2007).
83. Ampuero, J. *et al.* Development and Validation of Hepamet Fibrosis Scoring System-A Simple, Noninvasive Test to Identify Patients With Nonalcoholic Fatty Liver Disease With Advanced Fibrosis. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **18**, 216-225.e5 (2020).
84. Daniels, S. J. *et al.* ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. *Hepatology* vol. 69 (2019).
85. Vallet-Pichard, A. *et al.* FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology* **46**, 32–36 (2007).

86. Adams, L. A. & Chan, W.-K. Noninvasive Tests in the Assessment of NASH and NAFLD Fibrosis: Now and Into the Future. *Semin. Liver Dis.* **40**, 331–338 (2020).
87. Tyers, M. & Mann, M. From genomics to proteomics. *Nature* **422**, 193–197 (2003).
88. Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol* **13**, 942 (2017).
89. Dayon, L., Cominetti, O. & Affolter, M. Proteomics of human biological fluids for biomarker discoveries: technical advances and recent applications. *Expert Rev Proteomics* **19**, 131–151 (2022).
90. Belczacka, I. *et al.* Proteomics biomarkers for solid tumors: Current status and future prospects. *Mass Spectrom Rev* **38**, 49–78 (2019).
91. Bataller, R. & Brenner, D. A. Liver fibrosis. *J Clin Invest* **115**, 209–218 (2005).
92. Bellan, M., Castello, L. M. & Pirisi, M. Candidate Biomarkers of Liver Fibrosis: A Concise, Pathophysiology-oriented Review. *J Clin Transl Hepatol* **6**, 317–325 (2018).
93. Pernemalm, M. *et al.* In-depth human plasma proteome analysis captures tissue proteins and transfer of protein variants across the placenta. *eLife* **8**, e41608 (2019).
94. Li, Z. *et al.* Prevalence of nonalcoholic fatty liver disease in mainland of China: A meta-analysis of published studies. *Journal of Gastroenterology and Hepatology* vol. 29 (2014).
95. Rich, N. E. *et al.* Racial and Ethnic Disparities in Nonalcoholic Fatty Liver Disease Prevalence, Severity, and Outcomes in the United States: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol* **16**, 198-210.e2 (2018).
96. Nabi, O. La stéatose hépatique non alcoolique en France, épidémiologie et pronostic : analyse des données de la cohorte Constances et du Système National des Données de Santé (SNDS). (Sorbonne Université, 2022).
97. Accueil - Insee - Institut national de la statistique et des études économiques.  
<https://www.insee.fr/fr/accueil>.
98. Demographics of Europe. *Wikipedia* (2023).



99. Median age. <https://www.cia.gov/the-world-factbook/field/median-age/country-comparison/>.
100. EU's median age increased to 44.1 years in 2021. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220228-1>.
101. Overweight and obesity - BMI statistics. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Overweight\\_and\\_obesity\\_-\\_BMI\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Overweight_and_obesity_-_BMI_statistics).
102. Loomba, R. & Sanyal, A. J. The global NAFLD epidemic. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 686–690 (2013).
103. Lonardo, A. *et al.* Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps. *Hepatology* **70**, 1457 (2019).
104. World population. *Wikipedia* (2023).
105. Human sex ratio. *Wikipedia* (2023).
106. Ritchie, H. & Roser, M. Age Structure. *Our World Data* (2019).
107. Ritchie, H. & Roser, M. Obesity. *Our World Data* (2017).
108. Sebastiani, G., Gkouvatsos, K. & Pantopoulos, K. Chronic hepatitis C and liver fibrosis. *World J Gastroenterol* **20**, 11033–11053 (2014).
109. Croquet, V. *et al.* Prothrombin index is an indirect marker of severe liver fibrosis. *Eur. J. Gastroenterol. Hepatol.* **14**, 1133 (2002).
110. Farrell, G. C., Chitturi, S., Lau, G. K. K., Sollano, J. D., & Asia-Pacific Working Party on NAFLD. Guidelines for the assessment and management of non-alcoholic fatty liver disease in the Asia-Pacific region: executive summary. *J Gastroenterol Hepatol* **22**, 775–777 (2007).
111. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), & European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* **64**, 1388–1402 (2016).
112. Borges, H. *et al.* Well Plate Maker: A user-friendly randomized block design application to limit batch effects in largescale biomedical studies. *Bioinformatics* (2021).

113. Borges, H. Méthodes statistiques pour le traitement des données de protéomique quantitative longitudinale. (Université Grenoble Alpes, 2021).
114. Wieczorek, S. *et al.* DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**, 135–136 (2017).
115. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
116. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
117. Välikangas, T., Suomi, T. & Elo, L. L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform* **19**, 1–11 (2018).
118. Giai Gianetto, Q. *et al.* Calibration plot for proteomics: A graphical tool to visually check the assumptions underlying FDR control in quantitative experiments. *Proteomics* **16**, 29–32 (2016).
119. Giai Gianetto, Q., Wieczorek, S., Couté, Y. & Burger, T. A peptide-level multiple imputation strategy accounting for the different natures of missing values in proteomics data. *bioRxiv* (2020).
120. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
121. Wieczorek, S., Giai Gianetto, Q. & Burger, T. Five simple yet essential steps to correctly estimate the rate of false differentially abundant proteins in mass spectrometry analyses. *J Proteomics* **207**, 103441 (2019).
122. van Ooijen, M. P. *et al.* Identification of differentially expressed peptides in high-throughput proteomics data. *Brief Bioinform* **19**, 971–981 (2018).
123. Pascovici, D., Handler, D. C. L., Wu, J. X. & Haynes, P. A. Multiple testing corrections in quantitative proteomics: A useful but blunt tool. *Proteomics* **16**, 2448–2453 (2016).

124. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol Cell Proteomics* **22**, 100577 (2023).
125. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **27**, 633–641 (2009).
126. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4**, 222 (2008).
127. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806 (2009).
128. The Human Protein Atlas. <https://www.proteinatlas.org/>.
129. Brønstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Rev Proteomics* **1**, 503–512 (2004).
130. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol* **6**, 577–583 (2005).
131. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **25**, 125–131 (2007).
132. Fusaro, V. A., Mani, D. R., Mesirov, J. P. & Carr, S. A. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol* **27**, 190–198 (2009).
133. Colangelo, C. M., Chung, L., Bruce, C. & Cheung, K.-H. Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* **61**, 287–298 (2013).
134. Stahl-Zeng, J. *et al.* High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* **6**, 1809–1817 (2007).
135. Gallien, S., Duriez, E. & Domon, B. Selected reaction monitoring applied to proteomics. *J Mass Spectrom* **46**, 298–312 (2011).
136. Sacks, F. M. The apolipoprotein story. *Atheroscler Suppl* **7**, 23–27 (2006).

137. Lupsa, N. *et al.* Skin-homing CD8<sup>+</sup> T cells preferentially express GPI-anchored peptidase inhibitor 16, an inhibitor of cathepsin K. *Eur J Immunol* **48**, 1944–1957 (2018).
138. Regn, M. *et al.* Peptidase inhibitor 16 is a membrane-tethered regulator of chemerin processing in the myocardium. *J Mol Cell Cardiol* **99**, 57–64 (2016).
139. Reeves, J. R., Dulude, H., Panchal, C., Daigneault, L. & Ramnani, D. M. Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. *Clin Cancer Res* **12**, 6018–6022 (2006).
140. Wang, P. *et al.* PI16 attenuates response to sorafenib and represents a predictive biomarker in hepatocellular carcinoma. *Cancer Med* **9**, 6972–6983 (2020).
141. Ilani, T. *et al.* A secreted disulfide catalyst controls extracellular matrix composition and function. *Science* **341**, 74–76 (2013).
142. Niu, L. *et al.* Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat Med* **28**, 1277–1287 (2022).
143. Kang, S. W. *et al.* Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide generated in response to growth factors and tumor necrosis factor- $\alpha$ . *J Biol Chem* **273**, 6297–6302 (1998).
144. Lee, D. H. *et al.* Peroxiredoxin 6 Confers Protection Against Nonalcoholic Fatty Liver Disease Through Maintaining Mitochondrial Function. *Antioxid Redox Signal* **31**, 387–402 (2019).
145. Nabeshima, A. *et al.* Peroxiredoxin 4 protects against nonalcoholic steatohepatitis and type 2 diabetes in a nongenetic mouse model. *Antioxid Redox Signal* **19**, 1983–1998 (2013).
146. Stein, A. J. *et al.* Structural Basis for Inhibition of Human Autotaxin by Four Potent Compounds with Distinct Modes of Binding. *Mol Pharmacol* **88**, 982–992 (2015).
147. Ikeda, H. *et al.* Performance of autotaxin as a serum marker for liver fibrosis. *Ann Clin Biochem* **55**, 469–477 (2018).
148. Fujimori, N. *et al.* Serum autotaxin levels are correlated with hepatic fibrosis and ballooning in patients with non-alcoholic fatty liver disease. *World J Gastroenterol* **24**, 1239–1249 (2018).

149. Bain, G. *et al.* Selective Inhibition of Autotaxin Is Efficacious in Mouse Models of Liver Fibrosis. *J Pharmacol Exp Ther* **360**, 1–13 (2017).
150. Honda, Y. *et al.* Autotaxin is a valuable biomarker for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *Hepatol Res* **49**, 1136–1146 (2019).
151. Firth, S. M. & Baxter, R. C. Cellular actions of the insulin-like growth factor binding proteins. *Endocr Rev* **23**, 824–854 (2002).
152. Domené, H. M. *et al.* Phenotypic effects of null and haploinsufficiency of acid-labile subunit in a family with two novel IGFALS gene mutations. *J Clin Endocrinol Metab* **92**, 4444–4450 (2007).
153. Heath, K. E. *et al.* Primary acid-labile subunit deficiency due to recessive IGFALS mutations results in postnatal growth deficit associated with low circulating insulin growth factor (IGF)-I, IGF binding protein-3 levels, and hyperinsulinemia. *J Clin Endocrinol Metab* **93**, 1616–1624 (2008).
154. Nogueira, V. C. *et al.* UPLC-HDMSE to discover serum biomarkers in adults with type 1 diabetes. *Int J Biol Macromol* **221**, 1161–1170 (2022).
155. Qin, S. *et al.* SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients. *Proteomics* **12**, 1244–1252 (2012).
156. Huang, J. *et al.* Novel potential biomarkers for severe alcoholic liver disease. *Front Immunol* **13**, 1051353 (2022).
157. Han, X., Fiehler, R. & Broze, G. J., Jr. Characterization of the protein Z-dependent protease inhibitor. *Blood* **96**, 3049–3055 (2000).
158. Nielsen, M. J. *et al.* Haptoglobin-related protein is a high-affinity hemoglobin-binding plasma protein. *Blood* **108**, 2846–2849 (2006).
159. Glycobiomarkers for precise diagnosis of liver diseases. *Precis Med* **3**, (2018).
160. Fuentes-Prior, P., Fujikawa, K. & Pratt, K. P. New insights into binding interfaces of coagulation factors V and VIII and their homologues lessons from high resolution crystal structures. *Curr Protein Pept Sci* **3**, 313–339 (2002).

161. Kotronen, A. *et al.* Increased coagulation factor VIII, IX, XI and XII activities in non-alcoholic fatty liver disease. *Liver Int* **31**, 176–183 (2011).
162. Herbert, J. *et al.* Transthyretin: a choroid plexus-specific transport protein in human brain. The 1986 S. Weir Mitchell award. *Neurology* **36**, 900–911 (1986).
163. Steinhoff, J. S., Lass, A. & Schupp, M. Biological Functions of RBP4 and Its Relevance for Human Diseases. *Front Physiol* **12**, 659977 (2021).
164. Saki, F. *et al.* Association of Plasma Retinol Binding Protein-4 (RBP4) and Sonographic Grading of Fatty Liver in Obese Iranian Children. *Hepat Mon* **12**, e7103 (2012).
165. Wu, H. *et al.* Serum retinol binding protein 4 and nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus. *Diabetes Res Clin Pr.* **79**, 185–190 (2008).
166. Huang, S.-C. & Yang, Y.-J. Serum retinol-binding protein 4 is independently associated with pediatric NAFLD and fasting triglyceride level. *J Pediatr Gastroenterol Nutr* **56**, 145–150 (2013).
167. Campbell, D. J. The kallikrein-kinin system in humans. *Clin Exp Pharmacol Physiol* **28**, 1060–1065 (2001).
168. Zhang, W. *et al.* Theaflavin TF3 Relieves Hepatocyte Lipid Deposition through Activating an AMPK Signaling Pathway by targeting Plasma Kallikrein. *J Agric Food Chem* **68**, 2673–2683 (2020).
169. Cuervo, A. M. & Dice, J. F. A receptor for the selective uptake and degradation of proteins by lysosomes. *Science* **273**, 501–503 (1996).
170. Cuervo, A. M. & Dice, J. F. Unique properties of lamp2a compared to other lamp2 isoforms. *J Cell Sci* **113 Pt 24**, 4441–4450 (2000).
171. Ghareghani, P. *et al.* Aerobic endurance training improves nonalcoholic fatty liver disease (NAFLD) features via miR-33 dependent autophagy induction in high fat diet fed mice. *Obes Res Clin Pr.* **12**, 80–89 (2018).
172. Yang, Y. *et al.* Moderate Treadmill Exercise Alleviates NAFLD by Regulating the Biogenesis and Autophagy of Lipid Droplet. *Nutrients* **14**, (2022).

173. DiScipio, R. G., Chakravarti, D. N., Muller-Eberhard, H. J. & Fey, G. H. The structure of human complement component C7 and the C5b-7 complex. *J Biol Chem* **263**, 549–560 (1988).
174. Cheng, Q. *et al.* Serum proteome profiling reveals differentially expressed proteins between subjects with metabolically healthy obesity and nonalcoholic fatty liver disease. *J Proteomics* **260**, 104556 (2022).
175. Bell, L. N. *et al.* Serum proteomics and biomarker discovery across the spectrum of nonalcoholic fatty liver disease. *Hepatology* **51**, 111–120 (2010).
176. Hou, W. *et al.* Proteomic screening of plasma identifies potential noninvasive biomarkers associated with significant/advanced fibrosis in patients with nonalcoholic fatty liver disease. *Biosci Rep* **40**, (2020).
177. Lădaru, A., Bălănescu, P., Stan, M., Codreanu, I. & Anca, I. A. Candidate proteomic biomarkers for non-alcoholic fatty liver disease (steatosis and non-alcoholic steatohepatitis) discovered with mass-spectrometry: a systematic review. *Biomarkers* **21**, 102–114 (2016).
178. Niu, L. *et al.* Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol Syst Biol* **15**, e8793 (2019).
179. Gold, L., Walker, J. J., Wilcox, S. K. & Williams, S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. *N Biotechnol* **29**, 543–549 (2012).
180. Petretera, A. *et al.* Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *J Proteome Res* **20**, 751–762 (2021).
181. Billing, A. M. *et al.* Complementarity of SOMAscan to LC-MS/MS and RNA-seq for quantitative profiling of human embryonic and mesenchymal stem cells. *J Proteomics* **150**, 86–97 (2017).
182. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
183. Rensen, S. S. *et al.* Activation of the complement system in human nonalcoholic fatty liver disease. *Hepatology* **50**, 1809–1817 (2009).

184. Han, J. & Zhang, X. Complement Component C3: A Novel Biomarker Participating in the Pathogenesis of Non-alcoholic Fatty Liver Disease. *Front Med* **8**, 653293 (2021).
185. Li, L. *et al.* Proteome Dynamics Reveals Pro-Inflammatory Remodeling of Plasma Proteome in a Mouse Model of NAFLD. *J Proteome Res* **15**, 3388–3404 (2016).
186. Yilmaz, Y. & Eren, F. Serum biomarkers of fibrosis and extracellular matrix remodeling in patients with nonalcoholic fatty liver disease: association with liver histology. *Eur J Gastroenterol Hepatol* **31**, 43–46 (2019).
187. Kamada, Y. *et al.* Use of Mac-2 binding protein as a biomarker for nonalcoholic fatty liver disease diagnosis. *Hepatol Commun* **1**, 780–791 (2017).
188. Podack, E. R., Tschoop, J. & Müller-Eberhard, H. J. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. *J Exp Med* **156**, 268–282 (1982).
189. Schuppan, D. & Afdhal, N. H. Liver cirrhosis. *Lancet* **371**, 838–851 (2008).
190. Barnum, S. R., Bubeck, D. & Schein, T. N. Soluble Membrane Attack Complex: Biochemistry and Immunobiology. *Front Immunol* **11**, 585108 (2020).
191. Sveinbjornsson, G. *et al.* Multiomics study of nonalcoholic fatty liver disease. *Nat Genet* **54**, 1652–1663 (2022).
192. Christin, C. *et al.* A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Mol Cell Proteomics* **12**, 263–276 (2013).
193. Cook, N. R. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* **54**, 17–23 (2008).
194. Cook, N. R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115**, 928–935 (2007).
195. Pencina, M. J., D’Agostino, R. B. & Vasan, R. S. Statistical methods for assessment of added usefulness of new biomarkers. *Clin Chem Lab Med* **48**, 1703–1711 (2010).



196. Kampfrath, T. & Levinson, S. S. Brief critical review: Statistical assessment of biomarker performance. *Clin Chim Acta* **419**, 102–107 (2013).
197. Søreide, K. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *J Clin Pathol* **62**, 1–5 (2009).
198. Alemayehu, D. & Zou, K. H. Applications of ROC analysis in medical research: recent developments and future directions. *Acad Radiol* **19**, 1457–1464 (2012).
199. Sterling, R. K. *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
200. Guha, I. N. *et al.* Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology* **47**, 455–460 (2008).
201. Calès, P. *et al.* A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* **42**, 1373–1381 (2005).
202. Poynard, T., Imbert-Bismut, F., Munteanu, M. & Ratziu, V. FibroTest-FibroSURE: towards a universal biomarker of liver fibrosis? *Expert Rev Mol Diagn* **5**, 15–21 (2005).
203. Desaire, H. How (Not) to Generate a Highly Predictive Biomarker Panel Using Machine Learning. *J Proteome Res* **21**, 2071–2074 (2022).
204. Giovannetti, G. The linear regression model under test, by W. Krämer and H. Sonnberger. Physica Verlag, Heidelberg, 1986, ISBN 3-7908-0356-1 cloth, D.M. 98.00, pp. i-ix + 189. *J Appl Econ* **4**, 209–211 (1989).
205. Domené, H. M., Hwa, V., Jasper, H. G. & Rosenfeld, R. G. Acid-labile subunit (ALS) deficiency. *Best Pr. Res Clin Endocrinol Metab* **25**, 101–113 (2011).
206. Hellstern, S. *et al.* Functional studies on recombinant domains of Mac-2-binding protein. *J Biol Chem* **277**, 15690–15696 (2002).
207. Vuppalanchi, R. *et al.* Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol* **7**, 481–486 (2009).

208. Mehta, S. H., Lau, B., Afdhal, N. H. & Thomas, D. L. Exceeding the limits of liver histology markers. *J. Hepatol.* **50**, 36–41 (2009).
209. Nakayasu, E. S. *et al.* Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat Protoc* **16**, 3737–3760 (2021).
210. Pan, S. *et al.* Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* **8**, 787–797 (2009).
211. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24**, 971–983 (2006).
212. Megger, D. A., Bracht, T., Meyer, H. E. & Sitek, B. Label-free quantification in clinical proteomics. *Biochim Biophys Acta* **1834**, 1581–1590 (2013).
213. Bouyssié, D. *et al.* Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **36**, 3148–3155 (2020).
214. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).
215. Pailleux, F. *et al.* Mass Spectrometry-Based Proteomics Reveal Alcohol Dehydrogenase 1B as a Blood Biomarker Candidate to Monitor Acetaminophen-Induced Liver Injury. *Int J Mol Sci* **22**, (2021).
216. Kettenbach, A. N., Rush, J. & Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. *Nat Protoc* **6**, 175–186 (2011).
217. Ahsan, N., Rao, R. S. P., Gruppuso, P. A., Ramratnam, B. & Salomon, A. R. Targeted proteomics: Current status and future perspectives for quantification of food allergens. *J Proteomics* **143**, 15–23 (2016).
218. Brun, V., Masselon, C., Garin, J. & Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics* **72**, 740–749 (2009).

219. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **100**, 6940–6945 (2003).
220. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2**, 587–589 (2005).
221. Courçon, M. *et al.* Targeted Proteomics Analysis of Staphylococcal Superantigenic Toxins in Menstrual Fluid from Women with Menstrual Toxic Shock Syndrome (mTSS). *Toxins* **14**, (2022).
222. Pan, S. *et al.* High throughput proteome screening for biomarker detection. *Mol Cell Proteomics* **4**, 182–190 (2005).
223. Kirkpatrick, D. S., Gerber, S. A. & Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265–273 (2005).
224. Rozanova, S. *et al.* Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol Biol* **2228**, 85–116 (2021).
225. Brun, V. *et al.* Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics \*. *Mol Cell Proteomics* **6**, 2139–2149 (2007).
226. Thiede, B. *et al.* Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun Mass Spectrom* **14**, 496–502 (2000).
227. Siepen, J. A., Keevil, E.-J., Knight, D. & Hubbard, S. J. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J Proteome Res* **6**, 399–408 (2007).
228. Fu, Q. *et al.* A Plasma Sample Preparation for Mass Spectrometry using an Automated Workstation. *J Vis Exp* (2020).

229. Yang, Z. & Sun, L. Recent technical progress in sample preparation and liquid-phase separation-mass spectrometry for proteomic analysis of mass-limited samples. *Anal Methods* **13**, 1214–1225 (2021).
230. Stejskal, K., Potěšil, D. & Zdráhal, Z. Suppression of peptide sample losses in autosampler vials. *J Proteome Res* **12**, 3057–3062 (2013).
231. Liu, B. & Quertermous, T. Approximating the Sum of Independent Non-Identical Binomial Random Variables. *R J.* **10**, 472 (2018).
232. Peng, S. Y. & Hsu, H. C. ALDOB (aldolase B, fructose-bisphosphate). *Atlas Genet Cytogenet Oncol Haematol* (2011).
233. Asaka, M., Miyazaki, T., Hollinger, F. B. & Alpert, E. Human aldolase B serum levels: a marker of liver injury. *Hepatology* **4**, 531–535 (1984).
234. Tao, Q.-F. *et al.* Aldolase B inhibits metastasis through Ten-Eleven Translocation 1 and serves as a prognostic biomarker in hepatocellular carcinoma. *Mol Cancer* **14**, 170 (2015).
235. Li, H. *et al.* Acute liver failure in neonates with undiagnosed hereditary fructose intolerance due to exposure from widely available infant formulas. *Mol Genet Metab* **123**, 428–432 (2018).
236. Qin, S. *et al.* Identification of Organ-Enriched Protein Biomarkers of Acute Liver Injury by Targeted Quantitative Proteomics of Blood in Acetaminophen- and Carbon-Tetrachloride-Treated Mouse Models and Acetaminophen Overdose Patients. *J Proteome Res* **15**, 3724–3740 (2016).
237. Bourmaud, A., Gallien, S. & Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics* **16**, 2146–2159 (2016).
238. Hayes, J. M. An introduction to isotopic calculations.
239. Ghosh, P., Das, A. K., Dalai, S. & Chatterjee, S. The effects of non-standard lightning impulse on electrical insulation: a review. *Univ Polytechnica Buchar. Sci Bull Ser C Electr Eng* **104**, 4239–4254 (2022).

240. Haslam, D. E. *et al.* Stability and reproducibility of proteomic profiles in epidemiological studies: comparing the Olink and SOMAscan platforms. *Proteomics* **22**, e2100170 (2022).
241. ELISA in the multiplex era: Potentials and pitfalls - Tighe - 2015 - PROTEOMICS – Clinical Applications - Wiley Online Library.  
<https://onlinelibrary.wiley.com/doi/10.1002/prca.201400130>.
242. Yip, T. C.-F. *et al.* Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther* **46**, 447–456 (2017).
243. Kotronen, A. *et al.* Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* **137**, 865–872 (2009).
244. Lee, J.-H. *et al.* Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis* **42**, 503–508 (2010).
245. Bedogni, G. *et al.* *The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population.* *BMC Gastroenterology* vol. 6 (2006).
246. Bedogni, G., Kahn, H. S., Bellentani, S. & Tiribelli, C. A simple index of lipid overaccumulation is a good marker of liver steatosis. *BMC Gastroenterol* **10**, 98 (2010).
247. Poynard, T. *et al.* The diagnostic value of biomarkers (SteatoTest) for the prediction of liver steatosis. *Comp. Hepatol.* **4**, 10 (2005).
248. Munteanu, M. *et al.* Diagnostic performance of FibroTest, SteatoTest and ActiTest in patients with NAFLD using the SAF score as histological reference. *Aliment Pharmacol Ther* **44**, 877–889 (2016).
249. Yoneda, M. *et al.* Serum ferritin is a clinical biomarker in Japanese patients with nonalcoholic steatohepatitis (NASH) independent of HFE gene mutation. *Dig Sci* **55**, 808–814 (2010).
250. Canbay, A. *et al.* Non-invasive assessment of NAFLD as systemic disease-A machine learning perspective. *PLoS One* **14**, e0214436 (2019).

251. Poynard, T. *et al.* Meta-analyses of FibroTest diagnostic value in chronic liver disease. *BMC Gastroenterol* **7**, 40 (2007).
252. Alkhoury, N. *et al.* Circulating Soluble Fas and Fas Ligand Levels Are Elevated in Children with Nonalcoholic Steatohepatitis. *Dig Sci* **60**, 2353–2359 (2015).
253. Tamimi, T. I. A.-R. *et al.* An apoptosis panel for nonalcoholic steatohepatitis diagnosis. *J Hepatol* **54**, 1224–1229 (2011).
254. McPherson, S., Stewart, S. F., Henderson, E., Burt, A. D. & Day, C. P. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut* **59**, 1265–1269 (2010).
255. Kruger, F. C. *et al.* APRI: a simple bedside marker for advanced fibrosis that can avoid liver biopsy in patients with NAFLD/NASH. *Afr Med J* **101**, 477–480 (2011).
256. Harrison, S. A., Oliver, D., Arnold, H. L., Gogia, S. & Neuschwander-Tetri, B. A. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut* **57**, 1441–1447 (2008).
257. Rosenberg, W. M. C. *et al.* Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* **127**, 1704–1713 (2004).
258. Adams, L. A. *et al.* Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* **51**, 1867–1873 (2005).
259. Harrison, S. *et al.* A new non-invasive diagnostic score to monitor change in disease activity and predict fibrosis evolution in patients with NASH. *J Hepatol* **66**, S110 (2017).
260. Alkhoury, N. *et al.* OxNASH score correlates with histologic features and severity of nonalcoholic fatty liver disease. *Dig Sci* **59**, 1617–1624 (2014).
261. Boyle, M. *et al.* Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. *JHEP Rep* **1**, 188–198 (2019).

262. Christensen, C. *et al.* Diagnostic accuracy of a fibrosis serum panel (FIBROSpect II) compared with Knodell and Ishak liver biopsy scores in chronic hepatitis C patients. *J Viral Hepat* **13**, 652–658 (2006).
263. Irvine, K. M. *et al.* Multiplex Serum Protein Analysis Identifies Novel Biomarkers of Advanced Fibrosis in Patients with Chronic Liver Disease with the Potential to Improve Diagnostic Accuracy of Established Biomarkers. *PLoS One* **11**, e0167001 (2016).
264. Tanwar, S. *et al.* Validation of terminal peptide of procollagen III for the detection and assessment of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. *Hepatology* **57**, 103–111 (2013).
265. Milani, S. *et al.* Cellular localization of type I, III and IV procollagen gene transcripts in normal and fibrotic human liver. *Am J Pathol* **137**, 59–70 (1990).
266. Bril, F. *et al.* Use of Plasma Fragments of Propeptides of Type III, V, and VI Procollagen for the Detection of Liver Fibrosis in Type 2 Diabetes. *Diabetes Care* **42**, 1348–1351 (2019).
267. Luo, Y. *et al.* An Evaluation of the Collagen Fragments Related to Fibrogenesis and Fibrolysis in Nonalcoholic Steatohepatitis. *Sci Rep* **8**, 12414 (2018).
268. Zheng, X. *et al.* Collagen I promotes hepatocellular carcinoma cell proliferation by regulating integrin  $\beta$ 1/FAK signaling pathway in nonalcoholic fatty liver. *Oncotarget* **8**, 95586–95595 (2017).
269. Mizuno, M. *et al.* Classification of patients with non-alcoholic fatty liver disease using rapid immunoassay of serum type IV collagen compared with liver histology and other fibrosis markers. *Hepatol Res* **47**, 216–225 (2017).
270. Kumagai, E. *et al.* Serum YKL-40 as a marker of liver fibrosis in patients with non-alcoholic fatty liver disease. *Sci Rep* **6**, 35282 (2016).
271. Okanoue, T. *et al.* A simple scoring system using type IV collagen 7S and aspartate aminotransferase for diagnosing nonalcoholic steatohepatitis and related fibrosis. *J Gastroenterol* **53**, 129–139 (2018).

272. Mahmoud, A. A., Bakir, A. S. & Shabana, S. S. Serum TGF- $\beta$ , Serum MMP-1, and HOMA-IR as non-invasive predictors of fibrosis in Egyptian patients with NAFLD. *Saudi J Gastroenterol* **18**, 327–333 (2012).
273. Irvine, K. M. *et al.* Serum matrix metalloproteinase 7 (MMP7) is a biomarker of fibrosis in patients with non-alcoholic fatty liver disease. *Sci Rep* **11**, 2858 (2021).
274. Leeming, D. *et al.* A novel marker for assessment of liver matrix remodeling: an enzyme-linked immunosorbent assay (ELISA) detecting a MMP generated type I collagen neo-epitope (C1M). *Biomarkers* **16**, 616–628 (2011).
275. Barascuk, N. *et al.* A novel assay for extracellular matrix remodeling associated with liver fibrosis: An enzyme-linked immunosorbent assay (ELISA) for a MMP-9 proteolytically revealed neo-epitope of type III collagen. *Clin Biochem* **43**, 899–904 (2010).
276. Sand, J. M. *et al.* MMP mediated degradation of type IV collagen alpha 1 and alpha 3 chains reflects basement membrane remodeling in experimental and clinical fibrosis—validation of two novel biomarker assays. *PLoS One* **8**, e84934 (2013).
277. Karsdal, M. A. *et al.* Collagen biology and non-invasive biomarkers of liver fibrosis. *Liver International* vol. 40 (2020).
278. Santos, V. N. D. *et al.* Serum laminin, type IV collagen and hyaluronan as fibrosis markers in non-alcoholic fatty liver disease. *Braz J Med Biol Res* **38**, 747–753 (2005).
279. Dushay, J. *et al.* Increased fibroblast growth factor 21 in obesity and nonalcoholic fatty liver disease. *Gastroenterology* **139**, 456–463 (2010).
280. Wu, G. *et al.* Complementary Role of Fibroblast Growth Factor 21 and Cytokeratin 18 in Monitoring the Different Stages of Nonalcoholic Fatty Liver Disease. *Sci Rep* **7**, 5095 (2017).
281. Martin-Rodriguez, J. L., Gonzalez-Cantero, J., Gonzalez-Cantero, A., Arrebola, J. P. & Gonzalez-Calvin, J. L. Diagnostic accuracy of serum alanine aminotransferase as biomarker for nonalcoholic fatty liver disease and insulin resistance in healthy subjects, using 3T MR spectroscopy. *Medicine (Baltimore)* **96**, e6770 (2017).



282. Simons, N. *et al.* The endothelial function biomarker soluble E-selectin is associated with nonalcoholic fatty liver disease. *Liver International* vol. 40 (2020).
283. Jiang, Y., Han, T., Zhang, Z.-G., Zhang, Y. & Qi, F.-X. Serum thymosin beta4 as a noninvasive biomarker in patients with nonalcoholic steatohepatitis. *Rev Esp Enferm Dig* **110**, 19–24 (2018).
284. Del Ben, M. *et al.* Overexpression of the Vitronectin V10 Subunit in Patients with Nonalcoholic Steatohepatitis: Implications for Noninvasive Diagnosis of NASH. *Int J Mol Sci* **19**, (2018).
285. Paradis, V. Glycomics: a new taste of cirrhosis marker. *J Hepatol* **43**, 913–914 (2005).
286. Atsukawa, M. *et al.* Serum *Wisteria floribunda* agglutinin-positive Mac-2 binding protein more reliably distinguishes liver fibrosis stages in non-alcoholic fatty liver disease than serum Mac-2 binding protein. *Hepatology Research* vol. 48 (2018).
287. Yoneda, M. *et al.* High-sensitivity C-reactive protein is an independent clinical feature of nonalcoholic steatohepatitis (NASH) and also of the severity of fibrosis in NASH. *J Gastroenterol* **42**, 573–582 (2007).
288. Lee, J., Yoon, K., Ryu, S., Chang, Y. & Kim, H.-R. Correction: High-normal levels of hs-CRP predict the development of non-alcoholic fatty liver in healthy men. *PLoS One* **13**, e0206834 (2018).
289. Boga, S. *et al.* Plasma Pentraxin 3 Differentiates Nonalcoholic Steatohepatitis (NASH) from Non-NASH. *Metab Syndr Relat Disord* **13**, 393–399 (2015).
290. Hamza, R. T., Elfaramawy, A. A. & Mahmoud, N. H. Serum Pentraxin 3 Fragment as a Noninvasive Marker of Nonalcoholic Fatty Liver Disease in Obese Children and Adolescents. *Horm Res Paediatr* **86**, 11–20 (2016).
291. Younossi, Z. M. *et al.* A novel diagnostic biomarker panel for obesity-related nonalcoholic steatohepatitis (NASH). *Obes Surg* **18**, 1430–1437 (2008).
292. Wandrer, F. *et al.* TNF-Receptor-1 inhibition reduces liver steatosis, hepatocellular injury and fibrosis in NAFLD mice. *Cell Death Dis* **11**, 212 (2020).
293. Seo, Y. Y. *et al.* Tumor Necrosis Factor- $\alpha$  as a Predictor for the Development of Nonalcoholic Fatty Liver Disease: A 4-Year Follow-Up Study. *Endocrinol Metab Seoul* **28**, 41–45 (2013).

294. Gaddipati, R. *et al.* Visceral adipose tissue visfatin in nonalcoholic fatty liver disease. *Ann Hepatol* **9**, 266–270 (2010).
295. Qiu, Y. *et al.* Association of Circulating Adipsin, Visfatin, and Adiponectin with Nonalcoholic Fatty Liver Disease in Adults: A Case-Control Study. *Ann Nutr Metab* **74**, 44–52 (2019).
296. Sumida, Y. *et al.* A simple clinical scoring system using ferritin, fasting insulin, and type IV collagen 7S for predicting steatohepatitis in nonalcoholic fatty liver disease. *J Gastroenterol* **46**, 257–268 (2011).
297. Yilmaz, Y. *et al.* Serum pigment epithelium-derived factor levels are increased in patients with biopsy-proven nonalcoholic fatty liver disease and independently associated with liver steatosis. *Clin Chim Acta* **412**, 2296–2299 (2011).
298. Pihlajamäki, J. *et al.* Serum interleukin 1 receptor antagonist as an independent marker of non-alcoholic steatohepatitis in humans. *J Hepatol* **56**, 663–670 (2012).
299. Yilmaz, Y. *et al.* Serum levels of osteoprotegerin in the spectrum of nonalcoholic fatty liver disease. *Scand J Clin Lab Invest* **70**, 541–546 (2010).
300. Yang M. *et al.* [Value of serum osteoprotegerin in noninvasive diagnosis of nonalcoholic steatohepatitis]. *Zhonghua Gan Zang Bing Za Zhi* **24**, 96–101 (2016).
301. Zhou, Z., Chen, H., Ju, H. & Sun, M. Circulating retinol binding protein 4 levels in nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Lipids Health Dis* **16**, 180 (2017).
302. Wieckowska, A. *et al.* In vivo assessment of liver cell apoptosis as a novel biomarker of disease severity in nonalcoholic fatty liver disease. *Hepatology* **44**, 27–33 (2006).
303. Shen, J. *et al.* Non-invasive diagnosis of non-alcoholic steatohepatitis by combined serum biomarkers. *J Hepatol* **56**, 1363–1370 (2012).
304. Mohammed, A. *et al.* FICK-3 Score Combining Fibrosis-4, Insulin Resistance and Cytokeratin-18 in Predicting Non-alcoholic Steatohepatitis in NAFLD Egyptian Patients. *Pakistan Journal of Biological Sciences* vol. 22 (2019).

305. Page, S. *et al.* CCL-2 and Soluble Fas Ligand as Markers of Hepatic Fibrosis in Patients with Non-alcoholic Fatty Liver Disease (NAFLD). *American Journal of Gastroenterology* vol. 106 (2011).
306. Franzini, M. *et al.* Accuracy of b-GGT fraction for the diagnosis of non-alcoholic fatty liver disease. *Liver Int* **32**, 629–634 (2012).
307. Colak, Y. *et al.* Association of serum lipoprotein-associated phospholipase A2 level with nonalcoholic fatty liver disease. *Metab Syndr Relat Disord* **10**, 103–109 (2012).
308. Liu, Z. *et al.* Association of Lipoprotein-Associated Phospholipase A2 with the Prevalence of Nonalcoholic Fatty Liver Disease: A Result from the APAC Study. *Sci Rep* **8**, 10127 (2018).
309. Bellan, M. *et al.* Gas6 as a putative noninvasive biomarker of hepatic fibrosis. *Biomark Med* **10**, 1241–1249 (2016).
310. Krawczyk, M. *et al.* Panel of three novel serum markers predicts liver stiffness and fibrosis stages in patients with chronic liver disease. *PLoS One* **12**, e0173506 (2017).
311. Kamada, Y. *et al.* Serum Fucosylated Haptoglobin as a Novel Diagnostic Biomarker for Predicting Hepatocyte Ballooning and Nonalcoholic Steatohepatitis. *PLoS One* **8**, e66328 (2013).
312. Ogawa, K. *et al.* Tri-antennary tri-sialylated mono-fucosylated glycan of alpha-1 antitrypsin as a non-invasive biomarker for non-alcoholic steatohepatitis: a novel glycobiomarker for non-alcoholic steatohepatitis. *Sci Rep* **10**, 321 (2020).
313. Xu, L., Xu, C.-F., Yu, C.-H., Miao, M. & Li, Y.-M. Haemoglobin and non-alcoholic fatty liver disease: further evidence from a population-based study. *Gut* **58**, 1706–1707 (2009).
314. Fiorentino, T. V. *et al.* Association between hemoglobin glycation index and hepatic steatosis in non-diabetic individuals. *Diabetes Res Clin Pr.* **134**, 53–61 (2017).
315. Chen, C. *et al.* HbA1c may contribute to the development of non-alcoholic fatty liver disease even at normal-range levels. *Biosci Rep* **40**, (2020).
316. Feng, S., Wang, Z., Zhao, Y. & Tao, C. Wisteria floribunda agglutinin-positive Mac-2-binding protein as a diagnostic biomarker in liver cirrhosis: an updated meta-analysis. *Sci Rep* **10**, 10582 (2020).

317. Verhelst, X. *et al.* A Glycomics-Based Test Predicts the Development of Hepatocellular Carcinoma in Cirrhosis. *Clin Cancer Res* **23**, 2750–2758 (2017).
318. Lachiondo-Ortega, S. *et al.* Ubiquitin-Like Post-Translational Modifications (Ubl-PTMs): Small Peptides with Huge Impact in Liver Fibrosis. *Cells* **8**, (2019).
319. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111.016717 (2012).
320. Egertson, J. D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* **10**, 744–746 (2013).

**Article en cours de soumission :**

**Title: Plasma ALS and LG3BP differentiate liver fibrosis in MASLD patients and improve upon FibroTest**

Authors names:

David Pérez Compte<sup>1</sup>, Lucas Etourneau<sup>1,2</sup>, Anne-Marie Hesse<sup>1</sup>, Alexandra Kraut<sup>1</sup>, Justine Barthelon<sup>3</sup>, Nathalie Sturm<sup>3</sup>, Hélène Borges<sup>1</sup>, Salomé Biennier<sup>1</sup>, Marie Courçon<sup>1</sup>, Marc de Saint Loup<sup>5,6</sup>, Victoria Blary<sup>3,4</sup>, Charlotte Costentin<sup>3,4</sup>, Thomas Burger<sup>1</sup>, Yohann Couté<sup>1</sup>, Christophe Bruley<sup>1</sup>, Thomas Decaens<sup>3,4</sup>, Michel Jaquinod<sup>1</sup>, Jérôme Boursier<sup>5,6</sup>, Virginie Brun<sup>1,7</sup>

Authors institutions:

<sup>1</sup> Univ. Grenoble Alpes, INSERM, CEA, UA13 BGE, CNRS, CEA, FR2048 ProFI, 38000 Grenoble, France; <sup>2</sup> Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France ; <sup>3</sup> Univ. Grenoble Alpes, Clinique Universitaire d'Hépatogastroentérologie, CHU Grenoble Alpes, 38000 Grenoble, France; <sup>4</sup> Univ. Grenoble Alpes, Institute for Advanced Biosciences-INSERM U1209/CNRS UMR 5309, Grenoble, France; <sup>5</sup> Hepato-Gastroenterology Department, University Hospital, Angers, France; <sup>6</sup> HIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France; <sup>7</sup> CEA, LETI, Cinatec, F-38000 Grenoble, France.

Correspondence:

Virginie BRUN (lead author): virginie.brun@cea.fr (EDyP team – 17 Avenue des Martyrs, Grenoble, France)

Michel JAQUINOD: Michel.jaquinod@cea.fr (EDyP team – 17 Avenue des Martyrs, Grenoble, France)

## Abstract:

**Background:** Metabolic dysfunction-associated steatotic liver disease (MASLD) is estimated to affect 30% of the world's population, and its prevalence is increasing in line with obesity. Liver fibrosis is closely related to mortality, making it the most important clinical parameter for MASLD. It is currently assessed by liver biopsy – an invasive procedure that shows some limitations. There is thus an urgent need for a reliable non-invasive means to diagnose earlier MASLD stages.

**Methods:** A discovery study was performed on 158 plasma samples from histologically-characterized MASLD patients using mass spectrometry (MS)-based quantitative proteomics. Candidate differentially-expressed proteins were selected for verification by ELISA in the same cohort. They were subsequently validated in an independent MASLD cohort (n=200). An original method using a combination of a logistic regression model and likelihood ratio tests was used to develop a panel for fibrosis diagnosis.

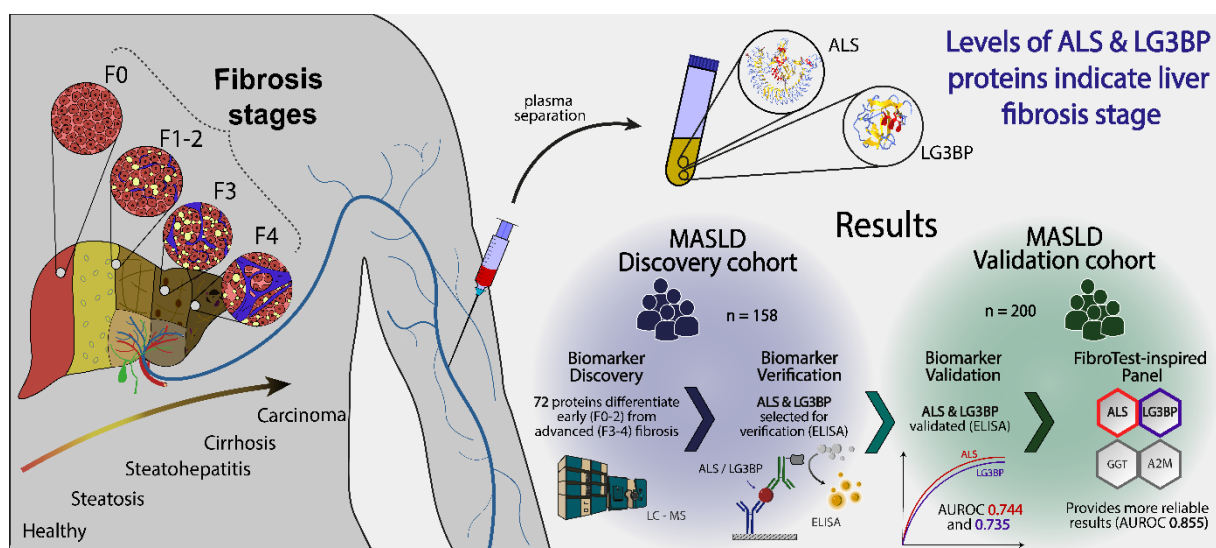
**Results:** From the 72 proteins differentially expressed between patients with early (F0-2) and advanced fibrosis (F3-4), we selected Insulin-like growth factor-binding protein complex acid labile (ALS, P35858) and Galectin-3-binding protein (LG3BP, Q08380) for further study. In our validation cohort, AUROCs with 95% CIs of 0.744 [0.673 – 0.816] and 0.735 [0.661 – 0.81] were obtained for ALS and LG3BP, respectively. Adding ALS and LG3BP to a FibroTest-inspired panel improved the assessment of advanced liver fibrosis, with an AUROC of 0.855 [0.802 – 0.908].

**Conclusions:** Further investigations with larger cohorts will be needed to confirm the usefulness of ALS, LG3BP and the panels developed for liver fibrosis diagnosis. With the availability of ELISA assays, these findings could be rapidly clinically translated, providing benefits for patients.

## Keywords:

Biomarker, MASLD, fibrosis, proteomics, liver

## Graphical abstract:



## *Background:*

Liver diseases are a major global health problem, accounting for more than 2 million deaths worldwide each year <sup>1</sup>. Linked to the spread of obesity, MASLD has become an epidemic affecting up to 1.66 billion adults in 2019, and its prevalence is increasing <sup>1-3</sup>. This disease has several stages, starting from steatosis and progressing to metabolic dysfunction-associated steatohepatitis (MASH), cirrhosis, or hepatocellular carcinoma (HCC) <sup>4</sup>. The early stages may be asymptomatic, but progress can be rapid by losing essential liver functions; for the most severe cases, liver transplantation (LT) may be required <sup>5,6</sup>. MASLD is the fastest-growing indication for LT in Western countries <sup>7</sup>. Although LT is the only curative treatment for patients with advanced MASLD, it remains a major intervention that can only be performed on patients who meet strict eligibility criteria and when donor livers are available. We lack effective pharmacological options for MASLD, although numerous clinical trials of potential treatments are in advanced phase II or III <sup>8</sup>. Among candidate therapies, those targeting fibrosis are promising as it is closely associated with MASLD progression <sup>4</sup>. Currently, only the early stages of liver fibrosis are reversible <sup>9</sup>. Consequently, robust antifibrotic treatments that could reverse the more advanced stages represent a critical target for the future management of MASLD <sup>8</sup>.

The most reliable available method to diagnose and characterize MASLD is liver biopsy, which is used to differentiate MASLD from MASH and to stratify patients with liver steatosis, fibrosis, or inflammation <sup>10</sup>. However, liver biopsy remains expensive and invasive, and it also provides suboptimal performance due to sampling variability and low histological reproducibility <sup>11</sup>. Given that more than a quarter of the world's population has MASLD, a reliable and cost-effective non-invasive screening test would greatly simplify and accelerate early diagnosis for those concerned. Over the past 15 years, several techniques have been developed as potential substitutes for liver biopsy <sup>11</sup>. However, because MASLD involves many molecular mechanisms that can lead to a range of pathophysiological states, no replacement for biopsy has yet been found that provides an equivalent level of information to clinicians. Nevertheless, some tests can guide medical decisions if they specifically target critical parameters and have acceptable performance.

Based on its link with mortality <sup>12</sup>, fibrosis is the most urgent variable to be determined in MASLD and is an essential parameter for the medical decision-making process. Liver biopsies are therefore considered when there is a significant risk of advanced liver fibrosis <sup>8</sup>. Clinicians have defined a histological SAF score (steatosis, activity and fibrosis score) to stratify fibrosis development from stage F0 (no fibrosis) to stage F4 (cirrhosis) <sup>13</sup>. The distinction between early stages (F0 to F2) and advanced stages (F3 to F4) is critical, as it determines patient management <sup>14</sup>. However, liver biopsy is not without risks and presents a certain number of limitations in this scenario <sup>11</sup>. The search for non-invasive fibrosis tests has therefore driven intense efforts. In this regard, several simple composite scores such as the FibroTest <sup>15</sup> or the MASLD Fibrosis score <sup>16</sup> have been developed. More recently, scores based on multiple parameters and including more complex algorithms incorporating MASLD-related protein biomarkers have been proposed <sup>17,18</sup>. However, although their performances are close to that of liver biopsy, none has reached the level required to replace it, and the urgent need for alternative non-invasive biomarkers remains <sup>19</sup>.

Over the last decade, clinical proteomics and associated biological discoveries have gradually been introduced into patient care <sup>20</sup>. Thanks to continuous progress, the

instrumentation used in this field has become remarkably efficient. As a result, it is now possible to characterize large cohorts of patients in a reasonable timeframe<sup>21,22</sup>. Here, an unbiased MS-based discovery study of plasma biomarkers was conducted on samples collected from histologically-characterized MASLD-suspected patients (n=158). Two candidate biomarkers were selected, and the results were verified using ELISA assays. The results from a validation study performed on an independent cohort (n=200) compared favourably with those from the widely used fibrosis identification panel FibroTest<sup>13</sup>.

### Methods:

**1. Patient cohorts:** The discovery and verification studies were conducted using a cohort of MASLD-suspected patients (n=158; 94 male; 64 female) recruited at Grenoble Hospital, France. The validation study was performed using an independent cohort of MASLD-suspected patients (n=200; 135 male; 65 female) constituted at Angers Hospital, France. Inclusion criteria for patients were suspicion of MASLD at a clinic visit and suspicion of advanced fibrosis based on clinical chemistry analysis and imaging exploration. If patients met these criteria, a liver biopsy and histological examination were performed to confirm the diagnosis (Table 1) (details about liver biopsy examination in Supp. Mat. 1). Patients were tested negative for common non-metabolic comorbidities that could induce liver diseases such as viral hepatitis, excessive alcohol consumption (male: < 210 g of alcohol per week; female: < 140 g of alcohol per week), drug-induced hepatitis or HIV.

**Table 1: Clinical, biological, and histological characteristics of patients in the Grenoble and Angers cohorts (n=158 and n=200, respectively).** BMI was calculated as weight/height<sup>2</sup> (kg/m<sup>2</sup>). A2M: alpha-2 macroglobulin; ApoA1: apolipoprotein A1; ALT and AST: alanine and aspartate aminotransferases; GGT: gamma-glutamyltransferase; CRP: C-reactive protein; LDL and HDL: low-density and high-density lipoprotein. The activity is the sum of the histological score for hepatocellular ballooning and lobular inflammation. The variable MASH corresponds to patients with at least a score of 1 on steatosis and activity assessment (ballooning + lobular inflammation). P-values for statistical significance were verified using the Mann–Whitney test.

Variable	Grenoble 158	Angers 200	p-values
Age (year)	55 ± 12	57 ± 12	0.029
Weight (kg)	89 ± 18	94 ± 19	0,023
Height (cm)	169 ± 10	167 ± 10	0,231
Waist size (cm)	108 ± 15	113 ± 14	0.0054
Body Mass Index (BMI)(kg/m <sup>2</sup> )	30.9 ± 5.5	33.4 ± 6.1	0.00014
Biological and clinical features			
Haptoglobin (g/L)	1.25 ± 0.61	1.24 ± 0.55	0.859
A2M (g/L)	2.25 ± 0.87	2.06 ± 0.82	0.027
ApoA1 (g/L)	1.46 ± 0.25	1.39 ± 0.21	0.029
Albumin (g/L)	44.8 ± 4.0	42.4 ± 3.7	9.00E-10
CRP (mg/L)	4.3 ± 5.5	4.3 ± 4.9	0.677



ALT UI/L	68.4 ± 34.3	63.7 ± 41.9	0.030
AST UI/L	41.2 ± 21.6	45.5 ± 26.3	0.099
GGT UI/L	153 ± 187	117 ± 170	0.186
Total cholesterol (g/L)	1.96 ± 0.54	1.86 ± 0.45	0.061
HDL cholesterol (g/L)	0.53 ± 0.47	0.43 ± 0.11	0.0076
LDL cholesterol (g/L)	1.16 ± 0.43	1.20 ± 0.42	0.328
Triglycerides (g/L)	1.77 ± 1.22	1.70 ± 1.12	0.976
Urea (g/L)	5.4 ± 1.5	5.6 ± 2.0	0.603
Bilirubin (µmol/L)	11.2 ± 6.1	12.8 ± 7.2	0.028
Histological features			
Steatosis grade, n (%)			
0	10 (6.3%)	15 (7.5%)	
1	31 (19.6%)	91 (45.5%)	
2	54 (34.2%)	63 (31.5%)	
3	63 (39.9%)	31 (15.5%)	
Activity			
Hepatocellular ballooning, n (%)			
0	29 (18.3%)	37 (18.5%)	
1	72 (45.6%)	83 (41.5%)	
2	57 (36.1%)	80 (40.0%)	
Lobular inflammation, n (%)			
0	49 (31.0%)	37 (18.5%)	
1	84 (53.2%)	136 (67.5%)	
2	25 (15.8%)	27 (13.5%)	
Fibrosis, n (%)			
0	24 (15.2%)	22 (11.0%)	
1	30 (19.0%)	49 (24.5%)	
2	46 (29.1%)	54 (27.0%)	
3	33 (20.9%)	61 (30.5%)	
4	25 (15.8%)	14 (7.0%)	
MASH, n (%)	102 (64.6%)	148 (74.0%)	

**2. Sample preparation – in-solution tryptic digestion:** Blood from patients was collected within minutes/hours of the biopsy in anticoagulant (EDTA)-treated tubes. Samples were then centrifuged to isolate plasma and stored at -80 °C. Plasma samples were collected between 2013 and 2021. For protein analysis, samples were randomly distributed using Well Plate Maker <sup>23</sup> and then proteins were denatured, reduced, alkylated, and digested with Trypsin/Lys-C Mix before LC–MS/MS analysis (detailed protocols in Supp. Mat. 2 and 3).

**3. MS-based proteomics analyses:** Briefly, plasma samples were analyzed by nanoLC–MS/MS. MS and MS/MS data were acquired in data-dependent acquisition mode (for further information, see Supp. Mat. 4).

**4. MS-based proteomics data processing:** Data were processed automatically using Mascot Distiller software (version 2.7.1.0, Matrix Science). Peptides and proteins were identified using Mascot (version 2.6) through concomitant searches against Swiss-Prot

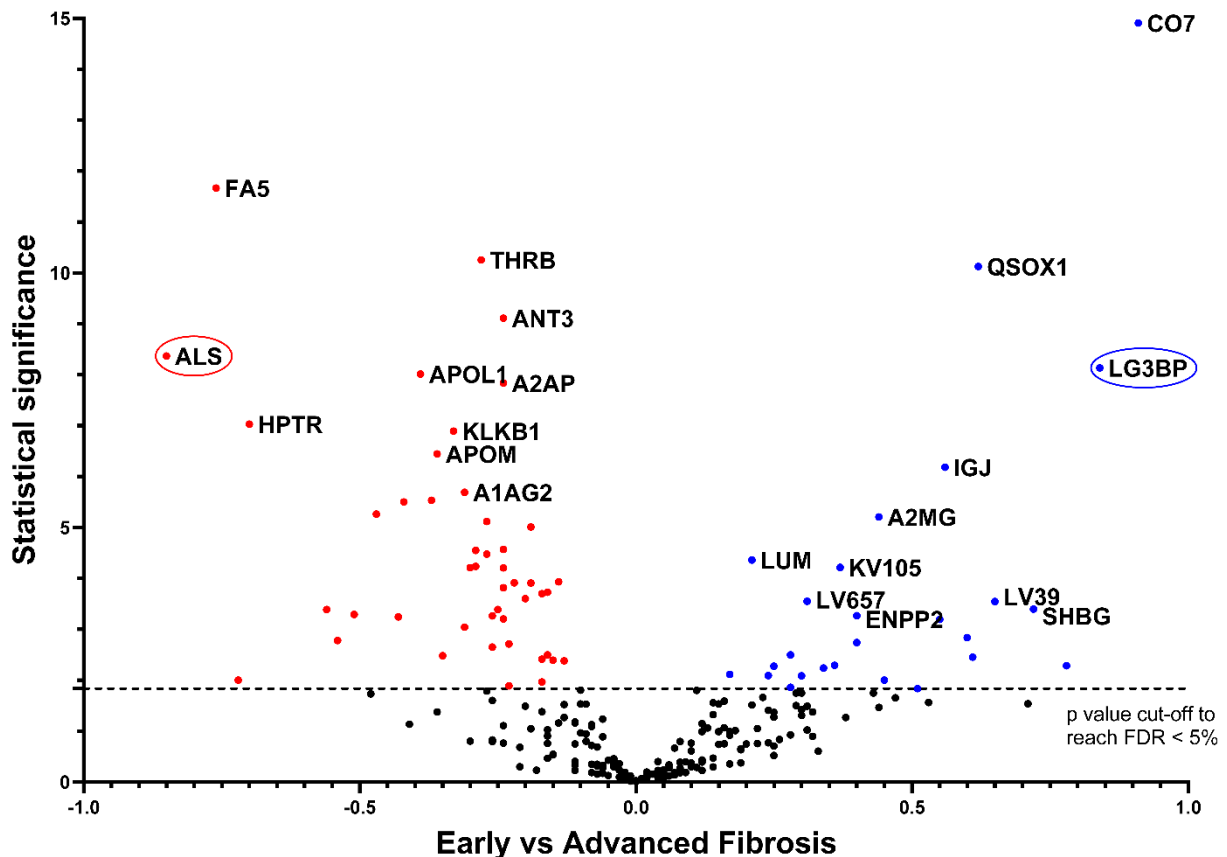
(*Homo sapiens* taxonomy, downloaded in November 2019), a classical contaminants database (homemade), and their corresponding reversed databases. Trypsin/P was chosen as the enzyme, and a maximum of two missed cleavages was allowed. Precursor and fragment mass error tolerances were set to 10 ppm and 25 mmu, respectively. Peptide modifications allowed during the search were (1) cysteine carbamidomethylation (fixed); (2) acetylation of the protein's N-terminus (variable), and (3) methionine oxidation (variable). Proline software (version 2.1)<sup>24</sup> was used to merge data for all patients. After combination, results were filtered: conserving rank 1 peptide-spectrum matches with a minimal length of 7 amino acids and a minimal Mascot peptide score of 25. With these parameters, the false discovery rate (FDR) for peptide/spectrum match identifications was below 1% according to the target-decoy approach. A minimum of two peptides was required per identified protein group. Proline was then used to perform MS1-based label-free quantification of the protein groups identified, based on their specific peptide abundances (Supp. Table 1).

5. Statistical analysis of MS-based quantitative proteomic data: Data were statistically analyzed using ProStaR (version 1.26.4)<sup>25</sup>. Protein sets were filtered out if they were not quantified across 70% of patient samples. Proteins identified in the contaminant database were discarded. After log<sub>2</sub> transformation, protein abundances were normalized by applying the variance stabilising normalisation method<sup>26</sup>. Missing values were imputed using the structured least square adaptative (SLSA) method<sup>27</sup>. The capacity to distinguish between early fibrosis (F0-2) and advanced fibrosis (F3-4) was statistically verified using the limma test<sup>28</sup>. Differentially expressed proteins were selected using a p-value cut-off, providing an FDR of less than 5% according to the Benjamini–Hochberg procedure (Supp. Table 1)<sup>29</sup>.

6. ELISA procedure (Supp. Mat. 5): Commercial ELISA assays were used to quantify ALS and LG3BP in plasma in the verification (Grenoble cohort) and validation (Angers cohort) studies. Tests were performed according to the manufacturer's instructions, as detailed in Supp. Mat. 5. Full ELISA results can be found in Supp. Table 2.

## Results:

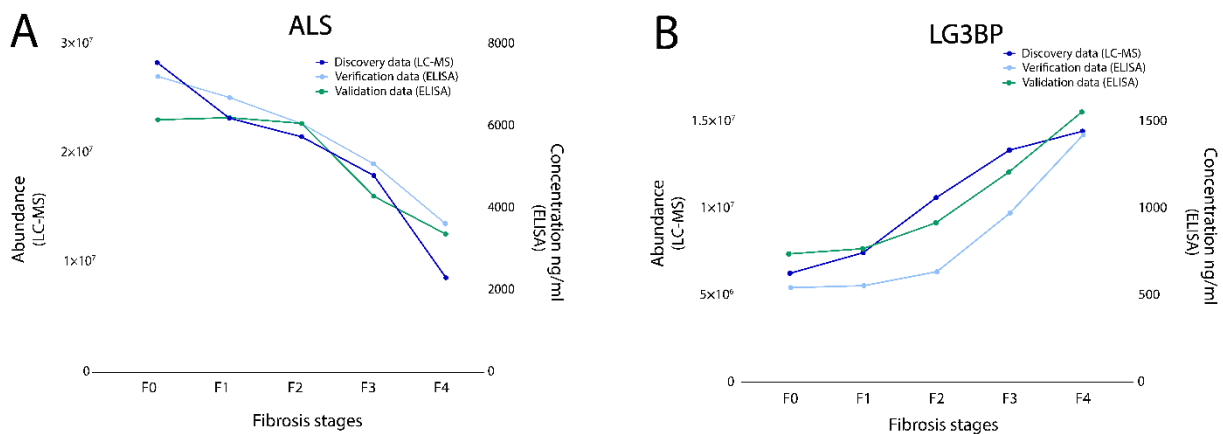
**1. Discovery Proteomics:** LC–MS/MS-based discovery proteomics analysis was performed using 158 plasma samples (Grenoble cohort) from patients with suspected MASLD. The aim was to identify candidate biomarkers with abundance levels making it possible to differentiate early fibrosis (F0-2) from advanced fibrosis (F3-4). The strategy deployed reliably identified and quantified 235 plasma proteins.



**Fig. 1: Differentially abundant proteins identified in plasma from MASLD-suspected patients with early (F0-2) and advanced (F3-4) liver fibrosis.** Plasma samples were collected from patients with suspected MASLD and early or advanced fibrosis, as determined from liver biopsies. Plasma proteins were submitted to MS-based label-free quantitative proteomics. For each protein, the Volcano plot displays the  $-\log_{10}(\text{Limma p-value})$  on the Y-axis; the X-axis corresponds to the  $\log_2$  (fold-change between early and advanced fibrosis). The cut-off for statistical significance was set for a p-value providing an FDR of less than 5% according to the Benjamini–Hochberg procedure. Red and blue dots represent less abundant and more abundant proteins, respectively, when advanced stages are compared to early stages. Only the 10 proteins with the best statistical significance in each condition are annotated in this figure (complete list can be found in Supp. Table 1).

The statistical analysis (Limma test) of the difference between relative abundances of each protein in the samples from patients with early and advanced fibrosis revealed 72 differentially abundant proteins (FDR inferior to 5% according to the Benjamini–Hochberg procedure, Fig. 1). Among these 72 proteins, when comparing advanced fibrosis to early fibrosis, 26 were up-regulated and 46 were down-regulated (Supp. Table 1).

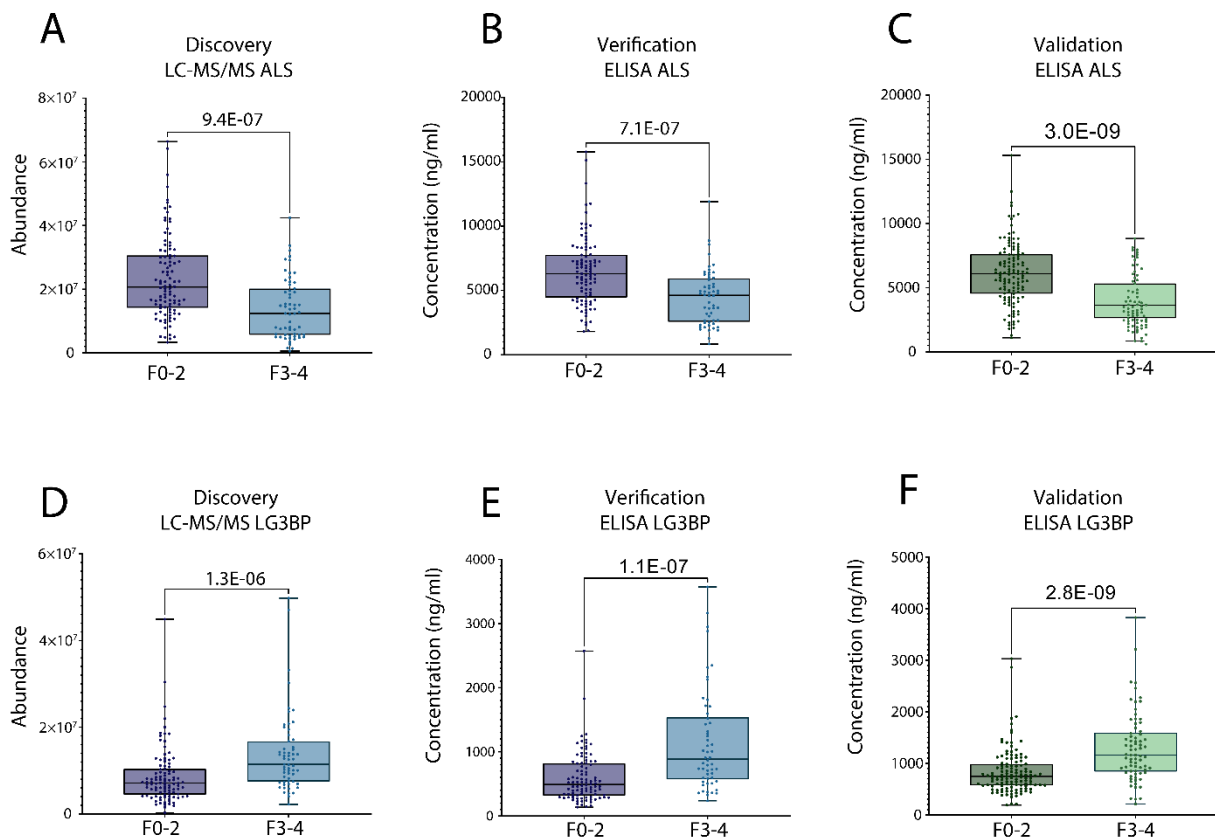
**2. Verification and validation of biomarkers:** In the verification phase, quantitative assays were used to confirm that the target protein levels differed significantly between samples from the advanced and early fibrosis groups. To select appropriate candidates for verification, several criteria were applied to the potential biomarkers identified by discovery proteomics: (1) substantial fold-change between the conditions compared and (2) availability and performance of ELISA assays. Based on these criteria, from the 72 differentially abundant proteins identified, two proteins were selected for the verification study: ALS and LG3BP. Notably, the abundance variation for C7 was 0,91 Log2FC (p-value 1.24E-15), but the corresponding ELISA assay showed poor quantification performance. ALS and LG3BP concentrations were determined in the same 158 plasma samples from the Grenoble cohort using ELISA assays. The results showed similar trends in terms of relative abundances for each protein between the discovery and verification studies for the different fibrosis stages (Fig. 2. A and B).



**Fig. 2: Mean abundances of ALS and LG3BP in plasma samples correlate with fibrosis stages and phases.** ALS (A) and LG3BP (B) levels based on MS abundance (discovery study, dark blue, left Y-axis arbitrary units) and concentrations based on ELISA measurements (verification and validation studies, light blue and green, right Y-axis ng/mL).

The validation study aimed to confirm the usefulness of the biomarkers with samples from an independent cohort. To do so, 200 samples from MASLD-suspected patients attending Angers Hospital were analyzed by ELISA. The trends for ALS and LG3BP abundances through fibrosis stages were similar for the two cohorts, regardless of the analytical method used (Fig. 2. A and B).

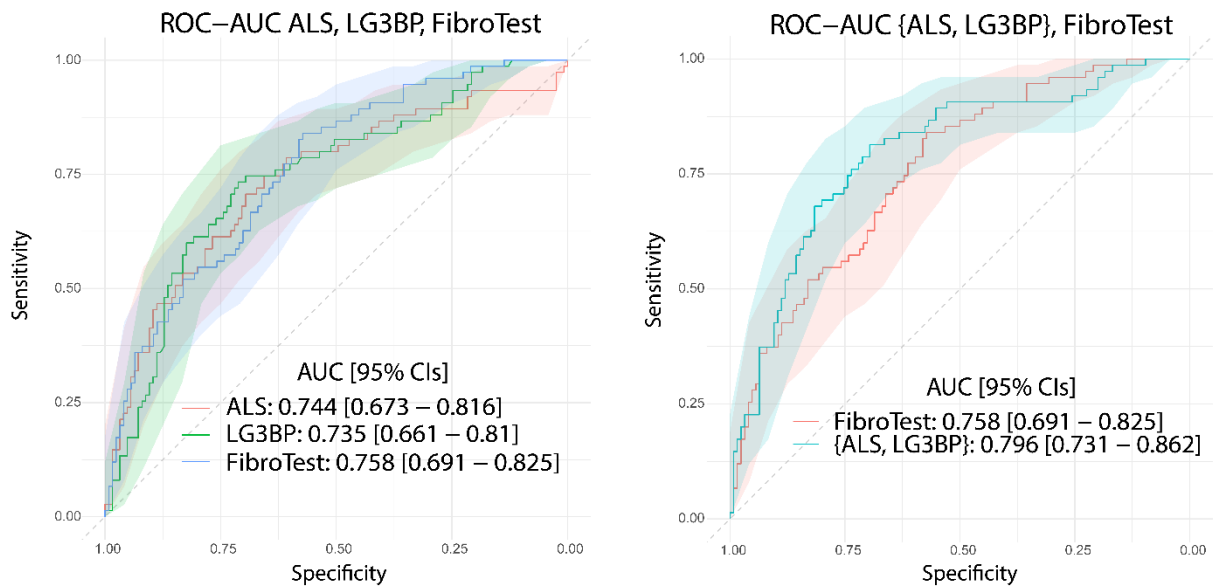
Finally, the results obtained for early (F0-2) and advanced stages (F3-4) of fibrosis in the discovery/verification and validation cohorts were statistically compared by applying a Mann–Whitney test (Fig. 3). Statistically significant levels were obtained for the discovery, verification and validation studies for ALS (9.4E-07, 4.4E-07 and 3.0E-09, respectively. Fig. 3. A, B and C) and LG3BP (1.3E-06, 4.1E-07 and 2.8E-09, respectively. Fig. 3. D, E and F). These levels of significance suggest that the abundances of both proteins can discriminate MASLD-suspected patients with early disease from those with advanced liver fibrosis.



**Fig. 3: ALS and LG3BP abundances can distinguish patients with early and advanced fibrosis.** ALS abundance in discovery (MS analysis, arbitrary unit) and verification (ELISA, test, concentration) studies – Grenoble cohort (A and B) and in validation study – Angers cohort (ELISA test, concentration, C). LG3BP abundance results from discovery/verification tests – Grenoble cohort (D and E) and validation test – Angers cohort (F). Data are represented as boxplots. Statistical significance was verified using the Mann–Whitney test. Dots represent ALS and LG3BP abundance for each individual patient.

### 3. Comparison with FibroTest for the non-invasive assessment of liver fibrosis:

We then focused on how well ALS and LG3BP differentiated early versus advanced liver fibrosis, examining the biomarkers either independently or in combination. For this analysis, we performed an AUROC calculation on the validation cohort (Angers) using (1) the concentration of ALS measured by ELISA, (2) the concentration of LG3BP measured by ELISA, and (3) the combined concentrations of LG3BP and ALS (Fig. 4). The AUROC values were compared with those obtained with the FibroTest (for the same cohort), as it is widely used and has been tested in several MASLD studies<sup>30,31</sup>. FibroTest was initially defined as a panel of three protein concentrations, one enzymatic activity, one metabolite concentration, age and sex, combined in a logistic regression model and validated on independent cohorts<sup>15</sup>. In our comparison, ALS and LG3BP provided surprisingly similar performances to the original FibroTest: ALS 0.744 [0.673 – 0.816], LG3BP 0.735 [0.661 – 0.81], FibroTest 0.758 [0.691 – 0.825]. Notably, their 95% CIs largely overlapped (Fig. 4). Using the same logistic regression methodology applied to the discovery cohort (Grenoble), we determined the weightings for ALS and LG3BP as a two-protein panel, termed {ALS, LG3BP} in Fig. 4. With an AUROC of 0.796 [0.731 – 0.862], the {ALS, LG3BP} panel improved fibrosis differentiation compared to ALS or LG3BP alone but also compared to the FibroTest panel.

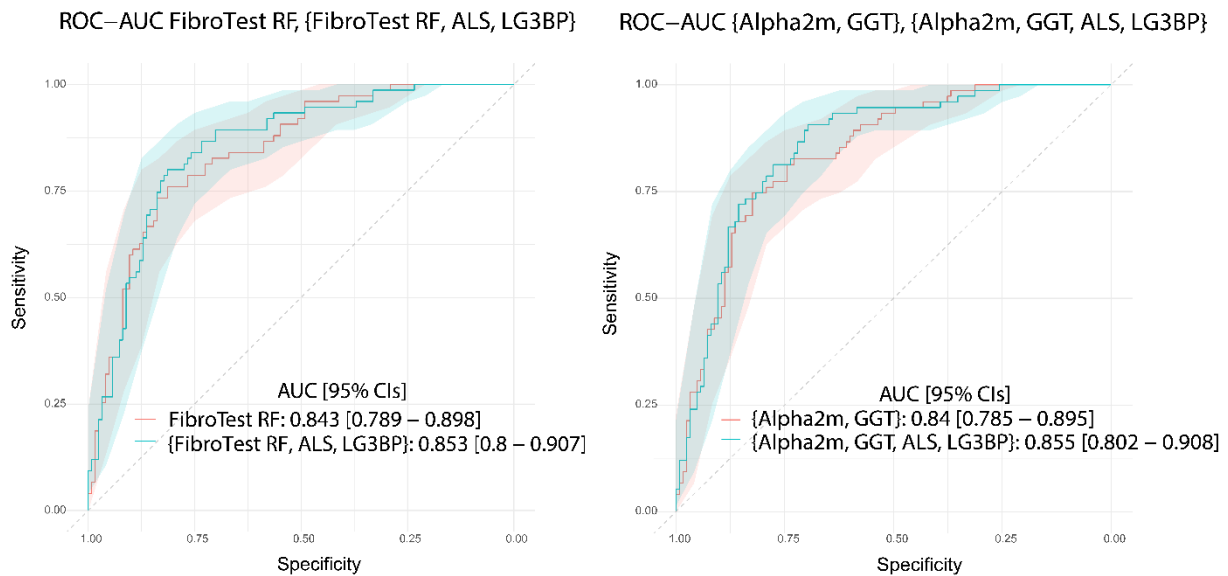


**Fig. 4: Plasma concentration of ALS and LG3BP discriminate early (F0-2) from advanced (F3-4) fibrosis as well as the FibroTest panel; as a 2-protein panel they outperform FibroTest.** ROC curves and AUROCs are shown with their respective 95% CIs for ALS/LG3BP ELISA quantifications and the original FibroTest score (left). Combined concentrations of ALS and LG3BP compared to the original FibroTest (right). The data presented corresponds to the Angers cohort. (Details about the CI calculation can be found in Supp. Mat. 6; the curves are displayed on two distinct plots for the sake of clarity only as a result of the extensive overlaps in CIs).

Based on the results obtained, we hypothesized that the combination of ALS and LG3BP with FibroTest (or with at least some of the FibroTest variables) in a multivariate model could improve the overall stratification performance. However, this hypothesis requires careful validation, as modifying a pre-existing panel by adding (or removing) variables must be supported by a precise assessment of each variable's contribution to the overall performance. For this type of validation, nested hypothesis testing has long been relied upon<sup>32</sup>; however, it could not be directly used here because FibroTest was trained on a distinct cohort. To address this difficulty, we computed an intermediate mathematical object, called FibroTest RF (re-fitted), which basically consisted in retraining the weightings of the original FibroTest variables using the same original logistic regression, but with data for our discovery cohort (Grenoble). The associated AUROC on the Angers cohort was much higher (0.843 [0.789 – 0.898]) (Fig. 5) than that obtained with the original FibroTest (0.758 [0.691 – 0.825]) (Fig. 4). Despite this apparently dramatic difference, we cannot claim that the FibroTest RF represents an improvement upon the original FibroTest as the two were assessed on distinct cohorts. Therefore, FibroTest RF should essentially be interpreted as an abstract mathematical model, providing a baseline for fair comparisons: any modification of the panel yielding a larger AUROC than FibroTest RF can be considered to improve upon the original FibroTest. This conclusion will apply even in the worst-case scenario where the performance difference due to the change of cohort cannot be accounted for. In this context, the results presented in Fig. 5 support the conclusion that combining previously identified biomarkers with ALS and LG3BP should yield a more powerful panel, better distinguishing between early and advanced stages of fibrosis.

To explain the performance gap between the original and the re-fitted FibroTest, we hypothesized that some of its variables were not true markers of disease severity

in the Grenoble and Angers cohorts and that they could therefore be safely removed from the re-fitted version without loss of performance. We verified this hypothesis by calculating iterative likelihood ratio tests for the Grenoble cohort: at each iteration, we removed the variable least likely to contribute to the model's performance (Supp. Mat. 7) until only significant variables remained. From this process, we concluded that five variables could be safely removed from the model without decreasing performance and hence retained only A2M and GGT. This result is supported by the fact that there is almost no change in AUROC between the re-fitted FibroTest (0.843 [0.789 – 0.898]) and its reduced version {A2M, GGT} (0.84 [0.785 – 0.895]) (Fig. 5).



**Fig. 5: Combination of re-fitted FibroTest with ALS and LG3BP improves performance, even with FibroTest reduced to two variables {A2M, GGT}.** ROC curves and AUROCs with their respective 95% CIs are shown for ALS and LG3BP combined with the re-fitted FibroTest (FibroTest RF) (left) and {A2M, GGT} (right). Models were fitted to data from the Grenoble cohort, and ROC-AUROC were measured using data from the Angers cohort.

We then examined whether adding ALS and LG3BP to FibroTest RF or to data related to {A2M, GGT} improved discrimination between early and advanced fibrosis stages. From the data presented in Fig. 5, we can see that adding ALS and LG3BP increased the overall AUROCs for these two panels by 0.01 and 0.015, respectively. However, it is difficult to estimate the true gain as CIs considerably overlap. To further assess the comparison, we performed likelihood ratio tests for these models fitted to the Angers cohort alone (Table 2). Specifically, we tested whether the models including ALS, LG3BP, or both fit the data significantly better than the models lacking these markers (*i.e.*, FibroTest RF and {A2M, GGT} models). We obtained p-values of 0.007 and 0.009, respectively, when both ALS and LG3BP were added to these two models. When ALS was added alone, p-values of 0.004 and 0.011 were obtained, respectively, whereas with LG3BP, p-values of 0.066 and 0.028 were obtained.

**Table 2: P-values from likelihood ratio tests assessing whether {A2M, GGT} and FibroTest RF improved when ALS, LG3BP, or both were added to the test panel.** These tests were performed on the Angers cohort only. The null hypothesis for a given p-value in the i-th row and j-th column is that variable(s) in the j-th column do not contribute when added to the i-th model. The alternative hypothesis is that the j-th variable(s) make a significant improvement to the i-th model. Hence, the lower the p-value is, and the more the added variable significantly improves the default model.

		Variable(s) added		
		ALS	LG3BP	ALS & LG3BP
Default model	{A2M, GGT}	0.011	0.028	0.009
	FibroTest RF	0.004	0.066	0.007

By combining ROC curves with nested significance testing, we found that using both previously discovered biomarkers from the FibroTest alongside ALS and LG3BP produced a more powerful panel to distinguish between early (F0-2) and advanced (F3-4) fibrosis stages. Relying on Ockham's razor principle and as we observed no significant effect of the other variables included in FibroTest (*i.e.*, haptoglobin, age, bilirubin, apoA1 and sex) on our cohorts, we assessed new biomarker panels for liver fibrosis: (1) a first combination including {A2M, GGT} and (2) a second one composed of {A2M, GGT, ALS, LG3BP}. These panels provided AUROCs of 0.84 [0.785 – 0.895] and 0.855 [0.802 – 0.908], respectively, with data from the Angers cohort. Details of weightings and thresholds, with associated specificity and sensitivity can be found in Supp. Mat. 8.

#### Discussion:

MASLD is a global health problem because of its prevalence and its consequences. We currently lack validated therapies and reliable non-invasive early diagnostic methods. Fibrosis remains the most important parameter to clarify, allowing clinicians to assess liver-related mortality and to adapt their prescription to provide more effective treatment<sup>12,33</sup>. It is therefore urgent to identify biomarkers that can differentiate early (F0-2) and advanced (F3-4) stages of fibrosis to better monitor the disease. To this end, we designed a discovery proteomics study to detect variations of individual protein abundances in the plasma of MASLD-suspected patients. We identified ALS and LG3BP for further study among a group of potential biomarkers with differential abundances in the plasma of patients with early and advanced fibrosis. These proteins were selected based on their FC difference between fibrosis conditions and the commercial availability of effective ELISA assays. These two proteins are involved in biological mechanisms linked to the pathophysiological processes behind MASLD, such as insulin regulation<sup>34</sup>, activation of the immune system<sup>35</sup>, or interactions with the extracellular matrix<sup>35</sup>. Decreased ALS levels have previously been associated with the progression of fibrosis in patients with hepatitis C virus<sup>36</sup> or with alcoholic hepatitis<sup>37</sup>. Two proteomic discovery studies have evidenced a decrease in ALS plasma abundance in MASLD patients with advanced fibrosis<sup>38,39</sup>. However, ALS has



never been validated at a large scale using independent cohorts. Based on a validation study in a large cohort (n=200), our study demonstrates the potential of ALS to decrease the need for liver biopsy in MASLD-suspected patients. For LG3BP, other authors have already linked it to MASLD and fibrosis using immunostaining techniques<sup>40</sup>, LC-MS proteomics approaches<sup>38</sup> or aptamers<sup>41,42</sup>. Ongoing research is focused on the development of anti-fibrotic treatments targeting galectin-3, the ligand of LG3BP<sup>43</sup>.

Many panels have already been developed in attempts to better differentiate liver fibrosis in MASLD<sup>15,16,44</sup>. Some of the most recent panel developments are based on proteomic studies using aptamers. For example, a panel of 8 proteins, including ADAMTSL2, obtained AUROCs of 0.87 and 0.89 in validation cohorts for differentiating F0-1 from F2-4 in MASLD patients<sup>45</sup>, and the FIBC3 panel obtained an AUROC of 0.83 for detecting advanced fibrosis stages ( $F \geq 3$ )<sup>46</sup>. Our results are comparable to those of these studies, with the advantage of presenting a panel that can be rapidly introduced into clinical practice thanks to the availability of ELISA tests. Indeed, one of the disadvantages of aptamers is the difficulty of introducing them into the hospital environment. Here, we demonstrated that ALS (0.744 [0.673 – 0.816]) and LG3BP (0.735 [0.661 – 0.81]) taken independently, performed similarly to the FibroTest panel (0.758 [0.691 – 0.825]) for the identification of advanced fibrosis in our validation cohort (Fig. 4). However, the combination of ALS and LG3BP (0.796 [0.731 – 0.862]) significantly improved the performance of the test compared to the FibroTest panel. Furthermore, by combining ALS and LG3BP with the FibroTest panel, advanced fibrosis was even more clearly evidenced. To achieve these results, we re-trained the weightings of the FibroTest variables to fit the characteristics of the Grenoble cohort. This resulted in a FibroTest RF providing clearly improved performance compared to the other models (0.843 [0.789 – 0.898]) (Fig. 5). We then demonstrated that five of the original FibroTest variables could be safely removed without significant loss of performance. When combined, the two remaining variables {A2M; GGT} showed an excellent AUROC performance of 0.84 [0.785 – 0.895]. The addition of ALS and LG3BP to this model provided an AUROC of 0.855 [0.802 – 0.908].

Despite the differences shown in Tab. 1 for some variables, the biomarkers and the panels were effective in differentiating advanced fibrosis in both cohorts. This suggests a strong potential for generalization to other cohorts and strengthens the validity of our results. The two cohorts have quite different male to female ratios, but the likelihood test in Supp. Mat. 7 shows that there is no significant change in the model when the sex variable is removed. Although our study was performed on two independent cohorts and the results were confirmed by two distinct analysis methods, it does have some limitations. First, the stage of fibrosis was determined by liver biopsy, which was only performed on patients with suspicion of MASLD and advanced fibrosis. Indeed, the selection of patients for liver biopsy is often based on clinical criteria such as elevated liver enzymes, serum markers of fibrosis or other non-invasive tests for liver fibrosis. These criteria can introduce a bias in the composition of the cohort by primarily including patients with advanced disease progression or, in some cases, patients without steatosis. While 6.3% and 7.5% of patients in the discovery and validation cohorts, respectively, did not have steatosis, it is important to keep these patients in the analysis, as they meet all the criteria established by clinicians for suspected advanced liver fibrosis in the context of MASLD. In addition, the biomarkers and panels developed could provide clinicians with additional information about the likelihood of

the patient having advanced fibrosis and potentially avoid biopsies in patients with early or no fibrosis.

Second, because histological lesions are unevenly distributed throughout the liver parenchyma, sampling errors in liver biopsies may lead to misdiagnosis and staging for MASLD patients, as reported elsewhere <sup>11</sup>. Liver biopsy results may also be more or less accurate depending on the length of the sample <sup>47</sup>, the distribution of histological lesions in the liver <sup>48</sup>, or inter- and intra-observer variability <sup>11</sup>. These parameters could lead to significant misdiagnosis and erroneous staging of MASLD. Nevertheless, the fact that the two distinct quantitative techniques used here – MS and ELISA – provided equivalent results in terms of performance lends support to the validity of the biomarkers identified while also eliminating any risk of instrumental bias. The advantage of ELISA over LC–MS when analyzing large cohorts is that it can be performed in a shorter timeframe and that it is more suitable for use in a hospital setting or clinical laboratory environment. In contrast, LC–MS analysis requires tight instrumental control due to its inherent complexity and sensitivity to technical variations <sup>49</sup>.

#### Conclusion:

In summary, the MS discovery study identified numerous proteins that could differentiate the stages of liver fibrosis in MASLD-suspected patients. Among them, two proteins (ALS and LG3BP) were of particular interest. Their respective performances were verified with a second analytical technique and validated on samples from an independent cohort. A panel have been created to improve the diagnostic performance of fibrosis. We therefore propose a new biomarker panel for liver fibrosis composed of {A2M, GGT, ALS, LG3BP}. This panel provided an AUROC of 0.855 [0.802 – 0.908]. Given these promising results, further investigations should now be undertaken on larger cohorts. Longitudinal studies could also be performed to better understand the behaviour of these proteins as fibrosis progresses.

Abbreviations:

A2M: Alpha-2 macroglobulin

ALT: alanine aminotransferases

ALS: Insulin-like growth factor-binding protein complex acid labile

ApoA1: apolipoprotein A1

AST: aspartate aminotransferase

AUROC: Area under the receiver operating characteristic curve

CI: Confidence interval

CRP: C-reactive protein;

LDL and HDL: low-density and high-density lipoprotein

FDR: False discovery rate

GGT: Gamma-glutamyl transferase

HCC: Hepatocellular carcinoma

LG3BP: Galectin-3-binding protein

LT: Liver transplantation

MS: Mass spectrometry

MASLD: Metabolic dysfunction-associated steatotic liver disease

MASH: Metabolic dysfunction-associated steatohepatitis

SAF score: Steatosis, Activity and Fibrosis score

SLSA: Structured Least Square Adaptive

Declarations:

Ethics approval and consent to participate:

Both parts of the study were approved by the respective Institutional Review Boards of the participating centres, and patients gave written informed consent for participation before samples were collected.

Consent for publication:

Availability of data and materials:

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (<https://www.proteomexchange.org/>) under dataset identifier PXD043340. Username: reviewer\_pxd043340@ebi.ac.uk Password: jM8845UU

Competing interests:

The authors who took part in this study declare that they have nothing to disclose in terms of funding or conflict of interest in relation to this manuscript.

#### Funding:

This work was supported by grants from the “Fonds de dotation – AGIR pour les maladies chroniques”, the “Association Française pour l’Étude du Foie”, and several programmes funded by the French National Research Agency (ANR). (1) ProFI project [ANR-10-INBS-08], GRAL, (2) a programme administered by the Chemistry Biology Health (CBH) Graduate School at University Grenoble Alpes [ANR-17-EURE-0003], (3) the LIFE project [ANR-15-IDEX-02], and (4) MIAI @ Grenoble Alpes [ANR-19-P3IA-0003].

#### Author’s contributions:

The authors listed above all contributed to this manuscript and approved the final submitted version. DPC conceptualised and designed the analysis, collected the data, performed the analysis and wrote part of the draft. LE conceptualised and designed the analysis, performed the analysis and wrote part of the draft. AMH conceptualised and designed the initial project and the analysis, performed the analysis, wrote part of the draft and contributed to critical revision of the paper. AK and MC collected the data. JBa, NS, MSL, CC, TD and JBo helped with plasma collection and cohort characterisation. HB and SB performed the analyses. VM contributed to the critical revision of the paper. TB, YC, CB, MJ, VB conceptualised and designed the initial project and the analysis and contributed to critical revision of the paper.

#### Acknowledgements:

The authors would like to thank Axelle Moulart for handling the samples at Grenoble CHU and Maighread Gallagher for her help in revising the script. The authors would also like to thank the entire team at EDyP for their support throughout the project.

#### References:

1. Asrani, S. K., Devarbhavi, H., Eaton, J. & Kamath, P. S. Burden of liver diseases in the world. *J Hepatol* **70**, 151–171 (2019).
2. Younossi, Z. M. *et al.* The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology* **77**, 1335 (2023).
3. Rinella, M. E. *et al.* A multi-society Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* (2023).
4. Malaguarnera, M., Di Rosa, M., Nicoletti, F. & Malaguarnera, L. Molecular mechanisms involved in NAFLD progression. *J Mol Med* **87**, 679–695 (2009).

5. Gadiparthi, C. *et al.* NAFLD Epidemiology, Emerging Pharmacotherapy, Liver Transplantation Implications and the Trends in the United States. *J Clin Transl Hepatol* **8**, 215–221 (2020).
6. Berasain, C., Arechederra, M., Argemí, J., Fernández-Barrena, M. G. & Avila, M. A. Loss of liver function in chronic liver disease: An identity crisis. *J. Hepatol.* **78**, 401–414 (2023).
7. Battistella, S. *et al.* Liver transplantation for non-alcoholic fatty liver disease: indications and post-transplant management. *Clin. Mol. Hepatol.* **29**, S286–S301 (2023).
8. Paternostro, R. & Trauner, M. Current treatment of non-alcoholic fatty liver disease. *J Intern Med* **292**, 190–204 (2022).
9. Trovato, F. M., Castrogiovanni, P., Malatino, L. & Musumeci, G. Nonalcoholic fatty liver disease (NAFLD) prevention: role of Mediterranean diet and physical activity. *Hepatobiliary Surg Nutr* **8**, 167–169 (2019).
10. Nalbantoglu, I. L. K. & Brunt, E. M. Role of liver biopsy in nonalcoholic fatty liver disease. *World J. Gastroenterol.* **20**, 9026–9037 (2014).
11. Sumida, Y., Nakajima, A. & Itoh, Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol* **20**, 475–485 (2014).
12. Dulai, P. S. *et al.* Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatology* **65**, 1557 (2017).
13. Bedossa, P. *et al.* Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* **56**, 1751–1759 (2012).
14. Maya-Miles, D., Ampuero, J., Gallego-Durán, R., Dingianna, P. & Romero-Gómez, M. Management of NAFLD patients with advanced fibrosis. *Liver Int.* **41**, 95–104 (2021).
15. Poynard, T., Imbert-Bismut, F., Munteanu, M. & Ratziu, V. FibroTest-FibroSURE: towards a universal biomarker of liver fibrosis? *Expert Rev Mol Diagn* **5**, 15–21 (2005).
16. Angulo, P. *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* **45**, 846–854 (2007).

17. Daniels, S. J. *et al.* ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. *Hepatology* **69**, 1075–1086 (2019).
18. Sanyal, A. J. *et al.* Enhanced diagnosis of advanced fibrosis and cirrhosis in individuals with NAFLD using FibroScan-based Agile scores. *J Hepatol* **78**, 247–259 (2023).
19. Adams, L. A. & Chan, W.-K. Noninvasive Tests in the Assessment of NASH and NAFLD Fibrosis: Now and Into the Future. *Semin. Liver Dis.* **40**, 331–338 (2020).
20. He, T. Implementation of Proteomics in Clinical Trials. *Proteomics Clin Appl* **13**, e1800198 (2019).
21. Meissner, F., Geddes-McAlister, J., Mann, M. & Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discov.* **21**, 637–654 (2022).
22. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24**, 971–983 (2006).
23. Borges, H. *et al.* Well Plate Maker: A user-friendly randomized block design application to limit batch effects in largescale biomedical studies. *Bioinformatics* (2021).
24. Bouyssié, D. *et al.* Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **36**, 3148–3155 (2020).
25. Wieczorek, S. *et al.* DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**, 135–136 (2017).
26. Huber, W., von Heydebreck, A., Sülthmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
27. Giai Gianetto, Q., Wieczorek, S., Couté, Y. & Burger, T. A peptide-level multiple imputation strategy accounting for the different natures of missing values in proteomics data. *bioRxiv* (2020).
28. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

29. Wieczorek, S., Giai Gianetto, Q. & Burger, T. Five simple yet essential steps to correctly estimate the rate of false differentially abundant proteins in mass spectrometry analyses. *J Proteomics* **207**, 103441 (2019).
30. Lassailly, G. *et al.* Validation of noninvasive biomarkers (FibroTest, SteatoTest, and NashTest) for prediction of liver injury in patients with morbid obesity. *Eur J Gastroenterol Hepatol* **23**, 499–506 (2011).
31. Bril, F. *et al.* Performance of Plasma Biomarkers and Diagnostic Panels for Nonalcoholic Steatohepatitis and Advanced Fibrosis in Patients With Type 2 Diabetes. *Diabetes Care* **43**, 290–297 (2020).
32. Krämer, W. & Sonnberger, H. *The Linear Regression Model Under Test*. (Physica-Verlag, 1986).
33. Ginès, P. *et al.* Population screening for liver fibrosis: Toward early diagnosis and intervention for chronic liver diseases. *Hepatology* **75**, 219–228 (2022).
34. Domené, H. M., Hwa, V., Jasper, H. G. & Rosenfeld, R. G. Acid-labile subunit (ALS) deficiency. *Best Pr. Res Clin Endocrinol Metab* **25**, 101–113 (2011).
35. Hellstern, S. *et al.* Functional studies on recombinant domains of Mac-2-binding protein. *J Biol Chem* **277**, 15690–15696 (2002).
36. Qin, S. *et al.* SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients. *Proteomics* **12**, 1244–1252 (2012).
37. Huang, J. *et al.* Novel potential biomarkers for severe alcoholic liver disease. *Front Immunol* **13**, 1051353 (2022).
38. Niu, L. *et al.* Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol Syst Biol* **15**, e8793 (2019).
39. Govaere, O. *et al.* A proteo-transcriptomic map of non-alcoholic fatty liver disease signatures. *Nat. Metab.* **5**, 572–578 (2023).
40. Kamada, Y. *et al.* Use of Mac-2 binding protein as a biomarker for nonalcoholic fatty liver disease diagnosis. *Hepatol Commun* **1**, 780–791 (2017).

41. Luo, Y. *et al.* SOMAscan Proteomics Identifies Serum Biomarkers Associated With Liver Fibrosis in Patients With NASH. *Hepatol Commun* **5**, 760–773 (2021).
42. Sveinbjornsson, G. *et al.* Multiomics study of nonalcoholic fatty liver disease. *Nat Genet* **54**, 1652–1663 (2022).
43. Chalasani, N. *et al.* Effects of Belapectin, an Inhibitor of Galectin-3, in Patients With Nonalcoholic Steatohepatitis With Cirrhosis and Portal Hypertension. *Gastroenterology* **158**, 1334-1345.e5 (2020).
44. Guha, I. N. *et al.* Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology* **47**, 455–460 (2008).
45. Corey, K. E. *et al.* ADAMTSL2 protein and a soluble biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in adults with NAFLD. *J. Hepatol.* **76**, 25–33 (2022).
46. Boyle, M. *et al.* Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. *JHEP Rep* **1**, 188–198 (2019).
47. Vuppalanchi, R. *et al.* Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol* **7**, 481–486 (2009).
48. Ooi, G. J. *et al.* Evaluation of the histological variability of core and wedge biopsies in nonalcoholic fatty liver disease in bariatric surgical patients. *Surg. Endosc.* **35**, 1210–1218 (2021).
49. Nakayasu, E. S. *et al.* Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat Protoc* **16**, 3737–3760 (2021).





## **Abstract :**

### Résumé:

L'objectif de mes travaux de thèse était de développer et mettre en application des méthodes de quantification en protéomique. Dans un premier volet, chez des patients atteints de MASLD une maladie hépatique métabolique touchant 25% de la population mondiale, des biomarqueurs plasmatiques indicateur de l'état de fibrose hépatique ont été découverts et validés. L'ALS et la LG3BP, deux de ces marqueurs, ont été validés sur deux cohortes indépendantes puis ont été introduites dans un nouveau panel dérivé du FibroTest. Le deuxième volet de cette thèse consistait à optimiser les quantifications absolues en LC-MS basées sur l'utilisation de protéines standards isotopiquement alourdis (PSAQ). L'originalité du projet PSAQ<sup>+1</sup> consistait à introduire la protéine standard assimilée à la protéine endogène dont toutes les arginines et lysines sont marquées uniquement avec un <sup>13</sup>C au début du pipeline de protéomique quantitative. Les résultats ont montré la fiabilité et la robustesse d'une telle approche apportant plus de spécificité, sensibilité et précision lors des quantifications absolues par LC-MS/MS utilisant des standards isotopiques.

Mots clés: Protéomique ; biomarqueurs ; MASLD ; quantification par dilution isotopique

### Summary:

The aim of my thesis was to develop and apply proteomic quantification methods. Firstly, plasma biomarkers were discovered and validated in patients suffering from MASLD, a metabolic liver disease that affects 25% of the world's population. ALS and LG3BP, two of these markers, were validated in two independent cohorts and then incorporated into a new panel derived from FibroTest. The second part of this work involved the optimisation of absolute LC-MS quantifications based on the use of isotopically enriched standard proteins (PSAQ). The originality of the PSAQ+1 project was to introduce the standard protein assimilated to the endogenous protein with all arginines and lysines labelled only with <sup>13</sup>C at the beginning of the quantitative proteomics pipeline. The results demonstrated the reliability and robustness of such approach, providing greater specificity, sensitivity and precision for absolute quantification by LC-MS/MS using PSAQ<sup>+1</sup>.

Keywords : Proteomics; biomarkers; MASLD; quantification by isotope dilution