Développement de statégies analytiques quantitatives pour l'étude des protéines, de leurs phosphorylations et glycations

Charlotte Brun

▶ **To cite this version:**

Charlotte Brun. Développement de statégies analytiques quantitatives pour l'étude des protéines, de leurs phosphorylations et glycations. Chimie analytique. Université de Strasbourg, 2022. Français. NNT : 2022STRAF038 . tel-04398190

HAL Id: tel-04398190
https://theses.hal.science/tel-04398190

Submitted on 16 Jan 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.

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

IPHC – UMR 7178

THÈSE présentée par :

Charlotte BRUN

soutenue le : **5 octobre 2022**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie Analytique

Développement de stratégies analytiques quantitatives pour l'étude des protéines, de leurs phosphorylations et glycations

THÈSE dirigée par :

Dr. BERTILE Fabrice	Chargé de recherche, Université de Strasbourg
Dr. SCHAEFFER Christine	Ingénieure de recherche, Université de Strasbourg

RAPPORTEURS :

Dr. FERRO Myriam	Directrice de recherche, CEA Grenoble
Dr. MARIN Philippe	Directeur de recherche, CNRS, Montpellier

À mes amis,

À ma Mamie,

La connaissance scientifique possède en quelque sorte des propriétés fractales : nous aurons beau accroître notre savoir, le reste - si infime soit-il - sera toujours aussi infiniment complexe que l'ensemble de départ.

Isaac Asimov

Remerciements

En premier lieu, je tiens à remercier Sarah Cianferani pour son accueil au sein du laboratoire de Spectrométrie de Masse Bio-Organique de l'Institut Pluridisciplinaire Hubert Curien (IPHC, UMR7178). Je veux aussi remercier l'Université de Strasbourg qui a financé ma thèse durant ces trois années. Fabrice et Christine, je vous suis tellement reconnaissante de m'avoir fait confiance et de m'avoir guidé tout au long de ce travail. Par votre encadrement et votre soutien, vous m'avez fait grandir et je ne l'oublierais jamais. Fabrice j'espère que tu pourras partir pour ce voyage mystérieux, ainsi que pour pleins d'autres à la découverte d'espèces toujours plus farfelues, et que ce sera « superbe » ! Christine, je tiens à te remercier pour tous tes conseils toujours judicieux, ta bienveillance et ton humanité.

Je veux aussi remercier Philippe Marin et Myriam Ferro pour avoir eu la gentillesse d'accepter de relire mon manuscrit et de prendre le temps d'évaluer mon travail.

Je remercie sincèrement l'ensemble des collaborateurs avec qui j'ai eu la chance de travailler : Christiane Guillouzo, Pierre-Olivier Schmit, Serge Desmoulins, Shuai Wu, Primoz Koselj, Mario Cindrić, Sylvain Giroud, Etienne Lefai, Audrey Bergouignan, Jérémy Terrien, Cédric Moro, Martin Quque et Quentin Schull. Je tiens à remercier particulièrement François Criscuolo avec qui j'ai eu la chance de collaborer sur des sujets très variés, toujours dans la bonne humeur.

Je veux ensuite adresser un immense merci à l'ensemble du LSMBO pour les bons moments que j'ai passé en votre compagnie, pour votre accueil et votre gentillesse ainsi que pour les nombreux pots (miam !). Je tiens à remercier particulièrement tous les U30 pour les activités et les soirées qu'on a passé ensemble !

Merci Laurence de m'avoir fait découvrir le monde merveilleux de la spectrométrie de masse, je mesure la chance que j'ai eu de suivre tes cours. Je pense également à Hélène, Stella, François (sans rancunes pour le Bravo), Magali (merci pour ton aide avec Proline) et Agnès. Agnès, c'est grâce à toi que je suis venue faire ma thèse au labo, que j'ai découvert la protéomique et que je me suis passionnée pour cette technique, pour tout ça merci ! Charline et Delphine merci pour beaucoup pour votre aide avec les phosphos. Merci aux personnes avec qui j'ai partagé le bureau : Corentin, Alex, Reiko, Cyrielle (on s'en souviendra de ce road-trip à la Palmyre) et Sarahi. Sarahi, I wish you luck on the glycation project, I hope everything goes smooth! Je n'oublie pas non plus nos voisines d'Inoviem : Justine et Leslie.

Merci à la team bioinfo/info Alex et Fabrice V. pour votre aide très précieuse ! Adrien, surtout continue à avoir des idées aussi rocambolesques et à transmettre ta bonne humeur ! Je serais très heureuse que tu me succèdes à la lourde tâche de webmaster de la FPS.

Martine, merci pour ton aide pour toutes les démarches administratives et promis, la prochaine fois je ferais attention à la voiture !

Merci Christine, CC, pour tes conseils que soit pour le HF-X ou pour les manips phospho mais aussi pour ta bienveillance et ton implication dans tout ce que tu fais. Je suis heureuse que tu ais accepté de faire partie de mon jury.

Jean-Marc, dit JMS ou Junior, comment ne pas te remercier pour tout ce que tu fais au laboratoire ? En plus de savoir réparer n'importe quelle machine avec « tes doigts magiques », tu nous nourris

tout au long de l'année avec de délicieux plats fait maison (galette des rois, agneau pascal, marrons glacés, charlottes au citron vert et j'en passe) ou non (les tant attendu beignets de chez Christian).

Bonne chance et bon courage aux nouveaux : Pauline (un p'tit verre ?) et Alex. Alex, je suis heureuse que tu reprennes le projet « ours ». Je sais que tu prendras beaucoup de plaisir à travailler dessus, c'est un très beau projet, mais penses quand même à dormir de temps en temps.

Un grand merci à la team Supra, Stéphane, Oscar, Evo, Marie L., Jérôme, Rania et Hugo, pour m'avoir laissé faire quelques manips sur vos instruments. Merci Oscar et Stéphane pour vos conseils. Marie L. bon courage pour les quelques mois qu'il te reste, je te laisse la bibliothèque avec plaisir ! Evo, je n'oublie pas nos soirées à discuter, je suis sûre que tu vas trouver ta place à Utrecht et surtout n'hésite pas si tu as besoin de parler à 2h du mat' ! Rania, mais comment est-ce que je vais faire sans toi pour prendre des photos à toutes les occasions, en tout cas, garde ta bonne humeur et ton enthousiasme ! Merci Jérôme, je pense que tu gagnes la palme du meilleur Airbnb :p Bon courage Hugo pour la suite !

Je suis aussi reconnaissante aux jeunes du R2 (même si tout le monde sait que le R5 c'est mieux :p) : Jeewan, Valeriane (tu vas y arriver ne t'inquiète pas), Bastien et Nathan. Nathan, souris et arrête un peu de râler ;)

Je n'oublie pas ceux qui ont déjà quitté le laboratoire, je pense notamment à Joanna, Maxime, Steve, Chloé, Blandine, Paola, Marziyeh, Kevin et Arnaud. Nico, je te remercie pour la formation et la bonne humeur que tu répandais au laboratoire. Marie C, merci pour tes conseils toujours très utiles et désolée de t'avoir embêtée avec toutes mes questions.

Vous remarquerez que j'ai gardé les meilleurs pour la fin : Aurélie, Corentin et Marie sans vous ces 3 années de thèse n'auraient clairement pas eu la même saveur ! Je ne garde que de beaux souvenirs avec vous et nos soirées vont me manquer ! Aurélie la prochaine fois, j'te jure, on ne t'attendra pas ! Enfin si, on râlera juste comme d'hab ! Merci pour toutes les fois où tu as sauvé mes manips mais aussi et surtout pour ta gentillesse, tes goûters et les belles discussions du soir. Tu es un peu la maman de notre groupe. Corentin, je suis tellement heureuse que tu sois arrivé dans le bureau. N'en déplaise à d'autres, j'adore notre manière de travailler en discutant, c'est très enrichissant (hâte d'avoir mon Tribrid à moi :p). Tu as beaucoup râlé mais finalement Leipzig et Prague c'était trop cool ! Marie, on a partagé tellement de choses ensemble, pro comme perso, je n'ai même pas les mots pour te dire à quel point je suis heureuse de te connaitre. Tu mets des paillettes dans ma vie (littéralement…). Je serais toujours là si tu as besoin de quoi que ce soit ! Prend soin de notre bébé Bravo et du HF-X, je ne me fais aucun souci. J'espère que je pourrais venir te voir quand tu seras en train de faire ton post-doc de rêve outre-Atlantique !

Je remercie aussi chaleureusement toutes les personnes qui m'ont soutenue, mes amis, la famille Vanulli (je n'oublierais jamais votre accueil) et ma famille : mes grands-parents, mes oncles et tantes, ma sœur, mon frère et évidemment mes parents.

Un merci un peu particulier à ce petit cheval rouquin, Rose, qui m'a vu débarquer à des heures improbables parfois et qui a toujours réussi à me changer les idées. Pauline, j'espère que tu en prendras soin ! Merci Inès de m'avoir fait autant confiance avec ton cheval, je vous souhaite tout le meilleur. Merci aussi à Shaïna pour les cours de voltige à pas d'heure, c'était trop cool !

Et enfin un immense merci à tous mes amis qui m'ont épaulée durant ces trois ans. À commencer par les « real friends for real » (il faut vraiment qu'on fasse quelque chose pour ce nom de groupe) : Mathieu, Claire, Anaëlle, Yanis et Apo. On ne s'est pas beaucoup vu ces derniers temps, mais on

va essayer de changer ça ! Merci à toi, Eliot (et toute la famille Carrasco, Looping inclue), je suis toujours partante pour une soirée jeux de société, un Marvel au ciné ou n'importe quoi ! Désolée de ne pas faire partie de la Cruz cette année Clara, mais ce n'est que partie remise. Merci de m'avoir fait voyager et d'être toujours là pour moi. Et bien sûr un énorme merci à toi Hermine, qui même à plusieurs centaines de kilomètres a suivi toutes les étapes de ce travail avec ses joies et ses peines. Je ne sais pas comment j'aurais fait sans toi. J'ai tellement hâte d'ouvrir le prochain chapitre avec toi ! On va bien s'amuser !

Table des matières

LISTE DES ABREVIATIONS 15

LISTE DES PUBLICATIONS 21

INTRODUCTION GENERALE 23

PARTIE I : INTRODUCTION A LA PROTEOMIQUE QUANTITATIVE 29

Chapitre 1 : Analyse des protéines par spectrométrie de masse 31

I. Deux approches différentes : analyse "*bottom-up*" et analyse des protéines intactes 31

A. Analyse structurale des protéines intactes 32

B. Analyse « *bottom-up* » 33

II. Les étapes de la préparation d'échantillon d'une analyse « *bottom-up* » 34

A. L'extraction des protéines 34

B. La digestion enzymatique 35

1. Réduction et alkylation 35
2. Protéases utilisées 36
3. Digestion en solution 36
4. Digestion en gel 37
5. Digestion sur membrane (Suspension Trap – S-Trap) 38
6. Digestion sur billes fonctionnalisées (SP3) 39
7. Décomplexification du mélange protéique 40

a. Fractionnement au niveau protéique 40

b. Fractionnement au niveau peptidique 42

8. Automatisation de la préparation d'échantillons en protéomique 42

III. Instrumentation : la chromatographie liquide couplée à la spectrométrie de masse 43

A. Séparation peptidique par chromatographie liquide 43

B. Spectrométrie de masse en tandem 44

C. Fragmentation des peptides 48

D. L'acquisition en mode DDA 50

IV. Traitement de données : de l'identification à l'interprétation biologique 50

A. Identification et quantification 51

1. L'importance de la banque de données 51
2. Moteurs de recherche 52
3. Validation des résultats 53

B. Quantification par courant d'ion extrait 54

C. Informations fonctionnelles 55

1. Banques de données fonctionnelles 55
2. Calcul de l'enrichissement en annotations fonctionnelles 57
3. Identifier une protéine homologue *via* un algorithme FASTA 58

V. Conclusion 58

Chapitre 2 : Une modification majeure pour la régulation du vivant : la phosphorylation 59

I. Intérêt de l'étude des phosphorylations 59

II. L'étude des phosphorylations par spectrométrie de masse rencontre de nombreux défis analytiques 61

A. La préparation d'échantillons 62
B. Analyse par spectrométrie de masse des peptides phosphorylés 64
C. Stratégies de quantification des peptides phosphorylés 67

III. Outils bio-informatiques pour l'étude des phosphorylations 70

A. Identification et localisation des phosphorylations 70
B. Validation et estimation du *False Localization Rate* [FLR] 71
C. Interprétation biologique 72

IV. Conclusion 72

Chapitre 3 : Une modification non-enzymatique aux multiples facettes : la glycation 73

I. La réaction de glycation ou réaction de Maillard 73

II. Défis analytiques de l'analyse *bottom-up* des protéines glyquées 75

A. Faible représentation des peptides glyqués par rapport aux peptides non modifiés 76
B. Digestion trypsique impactée par l'encombrement stérique des lysines et arginines 76
C. Difficultés de l'analyse en spectrométrie de masse 76
D. Manque d'outils bio-informatiques dédiés 79

III. Enrichissement par chromatographie d'affinité au boronate (*boronate affinity chromatography* [BAC]) 81

IV. Conclusion 82

PARTIE II : MISE AU POINT DE PROTOCOLES D'ANALYSES PROTEOMIQUES QUANTITATIVES POUR ETUDIER L'HIBERNATION CHEZ DES ESPECES NON MODELES 83

Chapitre 1 : Analyse du protéome des mitochondries musculaires de l'ours brun pour mieux comprendre le ralentissement du métabolisme durant l'hibernation 85

I. L'hibernation de l'ours 85

II. Problématique : quels mécanismes pour le ralentissement métabolique chez l'ours brun hibernant (*Ursus arctos*) ? 86

III. Stratégie analytique 87

IV. Résultats de l'analyse protéomique 88

A. Contrôles qualité 88
B. Résultats et interprétations 89
C. Validation de l'étude protéomique par une méthode orthogonale 91

V. Conclusion et perspectives 91

Chapitre 2 : Développement d'un protocole pour étudier les phosphorylations dans le muscle de l'ours brun pour mieux comprendre les adaptations à l'hibernation 93

I. Maintien de la masse musculaire et importance des phosphorylations 93

II. Étude des phosphorylations dans le tissu musculaire 93
III. Développement d'une méthode analytique pour l'analyse des phosphorylations dans le muscle 95
A. Sélection de la préparation d'échantillon la plus adaptée à l'analyse phosphoprotéomique du muscle 95
1. Optimisation des étapes d'extraction et de digestion des protéines à partir de tissus musculaires avant enrichissement 95
2. Évaluation de l'étape d'enrichissement des phosphopeptides 99
a. Efficacité de l'enrichissement 99
b. Nombre d'identifications 101
c. Répétabilité de la préparation d'échantillon 102
d. Répétabilité des analyses LC-MS/MS 103
e. Répartition des phosphorylations sur les acides aminés 103
B. Choix de la plateforme nanoLC-MS/MS adaptée à l'analyse des phosphopeptides 104
C. Évaluation de plusieurs logiciels de traitement de données pour l'identification et la quantification des phosphorylations 106
1. Définition d'un site de phosphorylation par les différents logiciels de traitement 106
2. Nombre d'identifications obtenus par les différents logiciels 108
3. Localisation des sites de phosphorylation 110
4. Quantification des phosphorylations 112
a. Biais de la quantification 112
b. Mise en place de filtres de qualité 113
D. Optimisation des méthodes de fragmentation sur Eclipse Tribrid et timsTOF Pro 114
1. Comparaison des méthodes ETD et HCD sur un instrument de type Orbitrap 114
2. Optimisation de l'énergie de collision sur le timsTOF Pro 116
E. L'apport d'une étape de fractionnement peptidique 118
IV. Application de cette méthodologie pour étudier la préservation du muscle durant l'hibernation de l'ours 121
A. Résultats et contrôles de la qualité des analyses des échantillons avant enrichissement 122
B. Résultats et contrôles de la qualité des analyses des échantillons enrichis en peptides phosphorylés 123
C. Premières interprétations des données phosphoprotéomiques obtenues 124
1. Alignement de séquence pour identifier le site humain homologue au site chez l'ours 125
2. Utilisation des banques fonctionnelles avec les données humaines 126
3. Analyse statistique du jeu de données 126
4. Analyse d'enrichissement en annotations fonctionnelles 127
V. Conclusion 128
Chapitre 3 : Évaluation d'un protocole expérimental adapté à l'analyse de cœurs de lérots pour mieux comprendre le rôle des acides gras polyinsaturés dans les mécanismes de l'hibernation 131
I. Les acides gras polyinsaturés et la SERCA influencent la réponse cardiaque à l'hibernation mais les voies intracellulaires impliquées restent inconnues 131

II. Développement d'un protocole de préparation d'échantillons adapté à l'analyse protéomique de cœurs de lérots 132

A. L'importance de la préparation d'échantillons 132

B. Description des protocoles de digestion et de la stratégie analytique 133

C. Comparaison des 4 protocoles de préparation d'échantillons pour l'étude du protéome cardiaque 135

1. Nombre de protéines/peptides obtenus 135

2. Répétabilité de la quantification des protéines 137

3. Comparaison des populations protéiques et peptidiques obtenues par les différents protocoles 138

a. Propriétés chimiques des peptides identifiés 138

b. Taux de clivages manqués 139

D. Choix du protocole le plus adapté 140

III. Application du protocole choisi pour étudier l'influence des acides gras polyinsaturés sur le protéome du cœur de lérot au cours de leur hibernation 141

A. Contrôles qualité internes et externes 142

B. Résultats et interprétations biologiques 143

IV. Conclusions et perspectives 145

PARTIE III : DEVELOPPMENT DE NOUVELLES APPROCHES POUR ETUDIER LES GLYCATIONS DES PROTEINES SANGUINES 147

Chapitre 1: Glycation des protéines sanguines chez le diamant mandarin : analyse LC-MS 149

I. Le paradoxe longévité/forte glycémie chez les oiseaux 149

II. Stratégie analytique 150

III. Analyse des résultats obtenus 151

A. Glycémie et mesure de la glycation des protéines du plasma 151

B. Glycation de l'hémoglobine 153

C. Détermination de la structure quaternaire du complexe d'hémoglobine 155

IV. Conclusion et perspectives 157

Chapitre 2 : Développement d'une méthode analytique permettant une étude approfondie des glycations du plasma par une approche « *bottom-up* » 160

I. Le plasma : un fluide particulièrement complexe avec une large gamme dynamique en concentration 160

II. Les glycations dans le plasma 162

III. Développement de différentes stratégies analytiques pour l'identification du protéome plasmatique glyqué 163

IV. Mise en place d'un enrichissement par chromatographie d'affinité au boronate pour identifier les peptides glyqués 165

A. Étude d'une protéine modèle : la BSA 165

1. Production de BSA glyquée par incubation *in vitro* avec du glucose 165

2. Enrichissement BAC de la BSA glyquée *in vitro* 166

3. Conclusion sur le premier test de l'enrichissement BAC 170

B. Optimisation du protocole sur un anticorps monoclonal de référence : le trastuzumab 170
1. L'étude des glycations dans le contexte des mAbs 170
2. Comparaison entre échantillon glyqué in vitro et échantillon contrôle 171
3. Conclusion sur l'apport de l'enrichissement BAC pour l'étude des glycations des mAbs 175
C. Enrichissement BAC d'un mélange complexe : le plasma 175
1. Comparaison des résultats obtenus à partir de plasma incubé *in vitro* avec ou sans glucose 175
2. Optimisation de la quantité de digest peptidique de plasma natif à déposer 177
3. Comparaison des données obtenues avec la littérature 179
4. Conclusion sur l'efficacité de la BAC à enrichir un échantillon de plasma en peptides glyqués 180
V. Mise en place d'une étape de fractionnement peptidique pour décomplexifier le plasma 180
A. Optimisation de la méthode de fractionnement peptidique hpH RP 181
1. Fractionnement hpH RP 1 : premières conditions d'élution 182
2. Fractionnement hpH RP 2 : modification des conditions d'élution 183
B. Optimisation de la méthode de fractionnement peptidique CIF 184
1. Fractionnement CIF 1 : premières conditions d'élution 184
2. Fractionnement CIF 2 : modification des conditions d'élution 185
C. Analyses des digests de plasma sur différents couplages LC-MS/MS 187
1. Analyse du plasma natif sans fractionnement sur différents couplages LC-MS/MS 187
2. Analyse du plasma fractionné sur différents couplages LC-MS/MS 189
D. Comparaison des deux méthodes de fractionnement développées 190
1. Profondeur de l'analyse 190
2. Nature des peptides identifiés par les deux fractionnements 191
3. Quantification des protéines 193
4. Conclusion 194
E. Identification des peptides glyqués dans les données issues du fractionnement hpH RP....195
VI. Couplage du fractionnement peptidique et de l'enrichissement BAC des peptides glyqués..... 195
VII. Conclusions et perspectives 199
CONCLUSION GENERALE 202
PARTIE EXPERIMENTALE 208
I. Analyse du protéome des mitochondries musculaires de l'ours brun 210
1. Échantillonnage 210
2. Dosage, lyse, dosage et précipitation 210
3. Gel SDS-PAGE « Stacking » 210
4. Analyse nanoLC-MS/MS 211
5. Identification et quantification des protéines à partir des spectres obtenus.....212

6. Traitement statistique des données quantitatives....213

II. Développement d'un protocole pour étudier les phosphorylations dans le muscle de l'ours brun 213

A. Développement d'une méthode analytique pour l'analyse des phosphorylations dans le muscle 213

1. Lyse, Dosage et précipitation 213
2. Réduction alkylation....214
3. Digestion liquide et SPE (conditions 1-4, tampons « Urée » et « Urée/Thio-urée) 214
4. Digestion SP3 (condition 5, tampon "Laemmli-like") 214
5. Analyse nanoLC-MS/MS des échantillons avant enrichissement 215
6. Identification et quantification des protéines à partir des spectres obtenus....215
7. Enrichissement IMAC des conditions 3 et 5 216
8. Analyse LC-MS/MS des échantillons après enrichissement IMAC....216
   a. Q-Exactive HF-X....216
   b. TimsTOf Pro....217
9. Évaluation de différents traitements de données pour l'identification, la localisation et la quantification des sites de phosphorylation 218
   a. Mascot-Proline 218
   b. Mascot-Proteome discoverer 219
   c. MS Amanda/Mascot-Proteome Discoverer....219
10. Comparaison de la fragmentation HCD et ETD sur Eclipse Tribrid....219
    a. HCD....220
    b. ETD 220
11. Optimisation des méthodes de fragmentation sur timsTOF Pro....220
12. L'apport d'une étape de fractionnement peptidique....221
    a. Lyse et dosage 221
    b. Réduction, alkylation, digestion liquide, SPE 221
    c. Fractionnement sur phase inverse à pH basique....221
    d. Enrichissement 222
    e. Analyse nanoLC-MS/MS 222
    f. Identification et quantification des phosphopeptides à partir des spectres obtenus ..222

B. Application de cette méthodologie pour étudier la préservation du muscle durant l'hibernation de l'ours....222

1. Lyse et dosage 222
2. Réduction, alkylation, digestion liquide, SPE....222
3. Fractionnement sur phase inverse à pH basique....222
4. Enrichissement 223
5. Analyse LC-MS/MS....223
6. Identification et quantification des protéines à partir des échantillons non enrichis 223

7. Identification et quantification des phosphopeptides à partir des échantillons enrichis en phosphopeptides....223
a. Alignement de séquence pour identifier le site humain homologue au site chez l'ours 224
b. Utilisation des banques fonctionnelles avec les données humaines....224
c. Analyse statistique du jeu de données....224
d. Analyse d'enrichissement en annotations fonctionnelles....224

III. Développement d'un protocole de préparation d'échantillon adapté à l'analyse protéomique de cœurs de lérots 225

A. Développement d'un protocole de préparation d'échantillon adapté à l'analyse protéomique de cœurs de lérots....225

1. Broyage lyse dosage....225
2. Gel SDS-PAGE « Stacking »....225
3. Protocole S-trap....225
4. Protocole SP3....225
5. Protocole Auto SP3....226
6. Analyse nanoLC-MS/MS....226
7. Identification et quantification des protéines à partir des spectres obtenus....227

B. Application du protocole choisi pour étudier l'influence des acides gras polyinsaturés sur le protéome du cœur de lérot au cours de leur hibernation....227

1. Broyage lyse dosage....227
2. Réduction, alkylation et digestion par protocole S-trap....228
3. Analyse nanoLC-MS/MS....228
4. Identification et quantification des protéines à partir des spectres obtenus....228
5. Analyse statistique des données quantitatives....228

IV. Développement d'une méthode analytique permettant une étude approfondie de la glycation du plasma par une approche « bottom-up »....229

A. Mise en place d'un enrichissement par chromatographie d'affinité au boronate pour identifier les peptides glyqués....229

1. Préparation d'un échantillon de BSA glyquée....229
a. Incubation in vitro de la BSA....229
b. Analyse LC-MS....229
2. Enrichissement *Boronate Affinity Chromatography* (BAC) d'un échantillon de BSA en peptides glyqués....229
a. Réduction, alkylation, digestion liquide, SPE....230
b. Enrichissement en peptides glyqués sur cartouches de boronate....230
c. Analyse nanoLC-MS/MS....230
d. Identification des peptides de la BSA....231
3. Enrichissement BAC d'un échantillon de trastuzumab en peptides glyqués....231

a. Incubation 231
b. Réduction, alkylation, digestion SP3 231
c. Enrichissement en peptides glyqués sur cartouches de boronate 231
d. Analyse nanoLC-MS/MS 232
e. Identification et quantification des peptides du trastuzumab 232
4. Enrichissement BAC d'un échantillon de plasma en peptides glyqués 233
a. Incubation 233
b. Réduction, alkylation, digestion SP3 et SPE 233
c. Enrichissement en peptides glyqués sur cartouches de boronate 233
d. Analyse nanoLC-MS/MS 233
e. Identification et quantification des peptides glyqués 234
B. Mise en place d'une étape de fractionnement peptidique pour décomplexifier le plasma 234
1. Préparation d'un digest de plasma 234
2. Dessalage des peptides 234
3. Fractionnement peptidique à pH basique 235
4. Fractionnement peptidique sur billes carboxylates 235
5. Analyse nanoLC-MS/MS 236
a. Q-Exactive HF-X 236
b. Q-Exactive Plus 236
c. TimsTOF Pro 237
6. Analyse nanoLC-MS/MS des échantillons non fractionnés 237
7. Recherche des fructosamines dans les échantillons après fractionnement 238
C. Couplage du fractionnement peptidique et de l'enrichissement BAC des peptides glyqués 238
1. Fractionnement peptidique à pH basique sur phase inverse 238
2. Enrichissement BAC des fractions 238
3. Analyse nanoLC-MS/MS 238
4. Traitement des données 239
TRAVAUX ADDITIONNELS 240
BIBLIOGRAPHIE 245

LISTE DES ABREVIATIONS

1/ K_0	mobilité ionique réduite
ABC	hydrogénocarbonate d'ammonium
ACN	Acétonitrile
ADN	acide désoxyribonucléique
AFA	*Adaptive Focused Acoustics*
AGEs	*Advanced Glycation End-products*
AGPI	Acides Gras PolyInsaturés
ANOVA	*Analysis of variance*
ARN	Acide ribonucléique
ATP	Adénosine-TriPhosphate
BAC	*boronate affinity chromatography*
BLAST	*Basic Local Alignment Search Tool*
BSA	*Bovine Serum Albumin*
CCS	*Collision Cross Section*
CEL	Carboxyethyllysine
CHAPS	3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate
CID	*Collision Induced Dissociation*
CIF	*CMMB-Based Isopropanol Gradient Peptide Fractionation*
CML	Carboxymethyllysine
CMMB	*Carboxylate-Modified Magnetic Beads*
COX	cytochrome c oxydase
CS	citrate synthase
CSF	*CerebroSpinal Fluid*

CV	coefficient de variation
Da	Dalton
DDA	*Data Dependant Acquisition*
DIA	*Data Independant Acquisition*
DTT	Dithiothreitol
EDTA	acide éthylène diamine tétra-acétique
ERLIC	*Electrostatic Repulsion-hydrophilic interaction LIquid Chromatography*
ESI	*Electron Spray Ionization*
ETcaD	*Electron-Transfer/Collisionally Activated Dissociation*
ETD	*Electron Transfer Dissociation*
EThcD	*Electron-Transfer/Higher-Energy Collision Dissociation*
FA	acide formique
FAIMS	*Field Asymmetric Ion Mobility Spectrometry*
FASP	*Filter Aided Sample Preparation*
FASTA	*Fast-All*
FDA	*Food and Drug Administration*
FDR	*False Discovery Rate*
FFPE	*Formalin-Fixed Paraffin Embedded*
FL	fructosamine
FLR	*False Localization Rate*
FT	*Flow through*
FTL	*Flow Through Loading*
FTW	*Flow Through Washing*
GO	*Gene Ontology*

Liste des abréviations

GOLD	*Glyoxal-derivated lysine dimer*
HCD	*Higher Energy Collisional Dissociation*
HDX-MS	analyses spectrométriques d'échanges hydrogène-deutérium
HILIC	*Hydrophilic Interaction Liquid Chromatography*
hpH RP	fractionnement sur phase inverse à pH basique
HPLC	*High Pressure Liquid Chromatography*
HPPP	*Human Plasma Proteome Project*
HSA	*Human Serum Albumin*
IAA	acide iodoacétique
IAM	iodoacétamide
IDA	iminodiacetique
IMAC	*Immobilizied Metal Affinity Chromatography*
iRT	*indexed Retention Time*
iST	*in-StageTip*
iTRAQ	*Isobaric Tag for Relative and Absolute Quantitation*
K_0	mobilité ionique
K_a	constante d'association
K_d	constante de dissociation
KEGG	*Kyoto Encyclopedia of Genes and Genomes*
LA	acide linoléique
LC	Chromatographie Liquide
LC-MS/MS	Chromatographie liquide couplée spectrométrie de masse en tandem
LFQ	*Label-Free Quantification*
M	Mol/L

Liste des abréviations

m/z	Rapport masse sur charge
mAb	*anticorps monoclonal*
MALDI	*Matrix Assisted Laser Desorption Ionization*
MeOH	Méthanol
MG-H1	5-methyl-4-oxo-5-hydroimifazo-linone-2-ylornithine
MOAC	*Metal Oxide Affinity Chromatography*
MRM	*Multiple Reaction Monitoring*
MS	Spectrométrie de masse
NCBI	Centre National pour l'information Biotechnologique
nLC	nanoLC
NTA	nitrilotriacetique
PAC	*Protein Aggregation Capture*
PASEF	*Parallel Accumulation Serial Fragmentation*
PBA	acide phenylboronique
PBS	*Phosphate Buffer Saline*
PFF	*Peptide Fragmentation Fingerprint*
PME11	*ProteoRed-ISCIII and EuPA launched the Proteomics Multicentric Experiment 11*
ppm	partie par million
PRM	*Parallel Reaction Monitoring*
PSM	*peptides spectrum match*
PTMs	*Post-Translational Modifications*
pTyr	phosphotyrosines
Q-Orbitrap	quadripôle-Orbitrap
Q-TOF	quadripôle-*Time Of* Flight

Liste des abréviations

RAGE	*Receptor for Advanced Glycation End-products*
RP	*Reverse Phase*
RPS	Reverse Phase S
RT	température ambiante
RYR2	récepteur à la ryanodine
SAX	*Strong Anion Exchange*
SCX	*Strong Cation Exchange*
SDC	sodium deoxycholate
SDS	sodium dodecyl-sulfate
SDS-PAGE	*Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis*
SEC	Chromatographie d'exclusion stérique
SERCA2a	*sarcoplasmic reticulum calcium ATPase*
SIB	institut Suisse de Bioinformatique
SP3	*Single Pot Solid Phase enhanced Sample Preparation*
SPE	Extraction sur phase solide
SRM	*Single Reaction Monitoring*
S-Trap	Suspension Trap
TCA	Acide trichloracétique
TCEP	(2-carboxyethyl) phosphine
TEAB	bicarbonate de triéthylammonium
TFA	acide trifluoroacétique
TIC	*Total Ion Current*
TIMS	*Trapped Ion Mobility Spectrometry*
TOF	*Time Of Flight*

Liste des abréviations

TMT	*Tandem Mass Tag*
Tris	tris(hydroxyméthyl)aminométhane
UHPLC	*Ultra High Pressure Liquid Chromatography*
UV	ultraviolet
UVPD	*UltraViolet Photo-Dissociation*
XIC	Extraction des courants d'ions
ARN	Acide ribonucléique
PTMs	*Post-Translational Modifications*
ESI	*Electron Spray Ionization*
MALDI	*Matrix Assisted Laser Desorption Ionization*

LISTE DES PUBLICATIONS

Publications

Brun, C., Hernandez-Alba, O., Hovasse, A., Criscuolo, F., Schaeffer-Reiss, C. and Bertile, F. Resistance to Glycation in the Zebra Finch: Mass Spectrometry-Based Analysis and its Perspectives for Evolutionary Studies of Aging" *Exp Gerontol,* 164, 111811. **(2022)**

Chazarin, B., Benhaim-Delarbre, M., **Brun, C.**, Anzeraey, A., Bertile, F. and Terrien, J. Molecular Liver Fingerprint Reflects the Seasonal Physiology of the Grey Mouse Lemur (*Microcebus murinus*) during Winter. *Int. J. Mol. Sci.* 23, no. 8, 4254. (**2022**).

Cussonneau, L., Boyer, C., **Brun, C.**, Deval, C., Loizon, E., Meugnier, E., Gueret, E., Dubois, E., Taillandier, D., Polge, C., Béchet, D., Gauquelin-Koch, G., Evans, A.L., Arnemo, J.M., Swenson, J.E., Blanc, S., Simon, C., Lefai, E., Bertile, F. and Combaret, L. Concurrent BMP Signaling Maintenance and TGF-B Signaling Inhibition Is a Hallmark of Natural Resistance to Muscle Atrophy in the Hibernating Bear." *Cells* 10, no. 8, 1873. (**2021**).

Boyer, C., Cussonneau, L**., Brun, C**., Deval, C., Pais de Barros, J.P., Chanon, S., Bernoud-Hubac, N., Daira, P., Evans, A.L., Arnemo, J.M., Swenson, J.E., Gauquelin-Koch, G., Simon, C., Blanc, S., Combaret, L., Bertile, F., and Lefai. E. Specific Shifts in the Endocannabinoid System in Hibernating Brown Bears. *Front. Zool.* 17, no. 1 (**2020)**

Communications orales

Brun C., Noone J., O'Gorman D., Bergouignan A., Lefai E., and Bertile F., "*Muscle mitochondrial proteome in hibernating bears to understand metabolic rate depression*", Journée des doctorants; 12 Novembre 2020, online.

Brun C., Criscuolo F., Bertile F., Schaeffer C., "*LC-MS analysis to determine glycation rates of intact blood proteins in a bird, the zebra finch"*, Meeting of the Young Scientists in Proteomics; 7 Juin 2021, online.

Gébelin M., **Brun C**., Bertile F., Carapito C., Schaeffer-Reiss C., *"Elucidation of a phosphoproteomic workflow: from sample preparation to data treatment"***, presentation** VisioFPS en duo organisée par la French Proteomic Society (FPS); 30 septembre 2021, online.

Brun C., Guguen-Guillouzo C., Bertile F., *"Proteome analysis reveals pathways of corticoid- and shape constraint-induced transdifferentiation of HepaRG cells"*, Webinar Bruker; 2 Décembre 2021, online.

Brun C., Quque, M., Villette C., Sueur, C., Heintz D., Criscuolo, F., Bertile F., *"Proteomics and metabolomics reveal how behaviour and age influence the physiology of ants."* Journée Métabolomique; 15 Juin 2022, Strasbourg (France).

Brun C., Gébelin, M., Lefai, E., Carapito, C., Schaeffer-Reiss, C., Bertile F., *"A comprehensive quantitative phosphoproteomics workflow adapted to muscle tissue and its application to the study of bear hibernation."* Analytics; 5-8 Septembre 2022, Nantes (France).

Brun C., Quque, M., Villette C., Sueur, C., Heintz D., Criscuolo, F., Bertile F. *"Proteomics and metabolomics reveal how behaviour and age influence the physiology of ants."* Analytics; 5-8 Septembre 2022, Nantes (France).

Communications par affiches

Rolland D., **Brun C**., Cianferani S., Desmoulins S., Wu S., Schaeffer-Reiss C. "*Fully automated sequential high-pH fractionation and phosphopeptide enrichment improves coverage and reproducibility of phosphoproteomics analysis.",* American Society for Mass Spectrometry (ASMS) ; 2 Novembre 2021, online.

Brun C., Guguen-Guillouzo C., Bertile F. *"Proteome analysis reveals pathways of corticoid- and shape constraint-induced transdifferentiation of HepaRG cells",* Human Proteome Organization (HUPO); 5 Décembre 2021, online.

Brun C., Hernandez-Alba, O., Schaeffer-Reiss C., Criscuolo F., Bertile F., "*Determination of glycation rates of blood proteins in an exotic model: the zebra finch."* European Proteomics Association (EuPA); 5-7 Avril 2022, Leipzig (Allemagne).

Brun C., Hernandez-Alba, O., Schaeffer-Reiss C., Criscuolo F., Bertile F., "*Determination of glycation rates of blood proteins in an exotic model: the zebra finch."* Analytics; 5-8 Septembre 2022, Nantes (France).

INTRODUCTION GENERALE

Les protéines sont des macromolécules biologiques composées d'une succession d'acides aminés liés entre eux par des liaisons peptidiques. Dans l'organisme, les protéines, résultant de la traduction des acides ribonucléiques [ARN] messagers eux-mêmes issus de la transcription des gènes, remplissent et régulent de nombreuses fonctions biologiques. En suivant l'hypothèse d'un gène codant pour une protéine, on recenserait environ 20 000 protéines humaines canoniques, *i.e.* non modifiées[1]. Mais ce serait occulter de nombreux phénomènes tels que les variants de séquence protéique, l'épissage alternatif et les modifications post-traductionnelles (*Post-Translational Modifications*, [PTM])qui conduisent à la formation de plus de six millions de formes protéiques différentes, appelées protéoformes, chez l'Homme[2]. Plus de 400 PTMs sont connues à ce jour, dont les phosphorylations, les glycosylations, les acétylations ou les glycations[3]. Ces modifications peuvent grandement influer sur la fonction, l'adressage ou l'activité des protéines qui les portent[4].

Le terme « protéomique » a été introduit dans les années 1990[5], par analogie au terme « génomique ». La protéomique englobe les analyses visant à identifier et à quantifier toutes les protéines d'un protéome donné, à partir de matrices diverses (tissus, fluides, cellules, organites…). Outre les niveaux d'expression protéique, la protéomique permet de caractériser leurs PTMs ou de définir les molécules qui interagissent avec elles (interactome). Il est nécessaire de souligner que le protéome est bien plus dynamique que le génome et varie dans le temps, en réponse à divers stimuli, et entre les différents types de cellules[6]. Au-delà du nombre important de protéines, leur large gamme dynamique de concentration est dans certains cas (plasma/sérum), un facteur augmentant la difficulté de l'analyse des protéomes des échantillons biologiques complexes[7].

Les progrès de l'analyse protéomique par spectrométrie de masse ont suivi le développement de nouvelles technologies adaptées à la séparation et à l'analyse des peptides/protéines. Ainsi, dans les années 1980, les analyses étaient réalisées sur des gels d'électrophorèse bidimensionnelle et par séquençage Edman pour déterminer la séquence primaire d'acides aminés[7-8]. L'invention des sources d'ionisation *Electron Spray Ionization* [ESI][9] et *Matrix Assisted Laser Desorption Ionization* [MALDI][10] conjointement avec des avancées sur le traitement des données et la préparation des échantillons ont permis à la spectrométrie de masse de s'imposer au sein de la discipline[8]. Ces 20 dernières années ont vu une progression rapide de la résolution, de la précision, de la sensibilité et de la vitesse d'acquisition des spectromètres de masse ainsi que le remplissage des banques de données protéiques et des outils d'interprétation des données de MS toujours plus performants. L'analyse protéomique quantitative sans marquage préalable des protéines (quantification relative globale à haut-débit selon une stratégie dite « *label-free* ») est aujourd'hui mature et permet l'identification et la quantification de plusieurs centaines à milliers de protéines en quelques heures[11-12]. Elle comporte 3 étapes clés: la préparation d'échantillon, l'analyse par chromatographie liquide couplée à la spectrométrie de masse [LC-MS/MS] et le traitement des données. Cette stratégie peut également être utilisée pour l'étude des PTMs, qui représente un niveau supplémentaire de complexité du protéome[13-14]. Cependant, malgré les avancées technologiques de ces dernières années dans ce domaine, l'étude des PTMs reste un défi analytique. En effet, on rencontre alors des difficultés spécifiques ayant trait à la préparation d'échantillon (conservation des PTMs labiles, digestion incomplète…), à l'analyse LC-MS/MS (difficulté d'ionisation et/ou de

fragmentation) et au traitement des données (assignation difficile du site modifié, quantification sur un seul peptide…). De plus, la faible stœchiométrie existant entre les peptides modifiés et non modifiés rend leur détection extrêmement difficile[15-16]. Enfin, l'interprétation biologique des résultats est plus délicate car il y a peu d'informations concernant l'action des PTMs, *a fortiori* pour les espèces dites non modèles.

Mes travaux de thèses s'inscrivent dans ce contexte. Plus précisément, j'ai cherché à développer des protocoles expérimentaux pour étudier des systèmes biologiques complexes par protéomique quantitative (« *label-free* »), ainsi qu'en explorant deux familles de PTMs : les phosphorylations et les glycations. Les différents objectifs de ce travail de doctorat étaient les suivants :

- Adaptation de protocoles d'analyse protéomique permettant l'étude des mécanismes biologiques liés à l'hibernation d'espèces non modèles : l'ours brun (*Ursus Arctos*) et le lérot (*Eliomys Quercinus*).
- Développement d'un protocole complet allant de la préparation des échantillons à l'analyse des données et à l'interprétation biologique pour l'étude des phosphorylations à partir de tissus musculaires d'ours brun.
- Développement de stratégies analytiques permettant l'étude des glycations des protéines plasmatiques et de l'hémoglobine.

Ce manuscrit est divisé en trois parties qui sont brièvement décrites ci-après :

- La **partie I** propose un état de l'art de la stratégie de protéomique quantitative sans marquage (« *label-free* »). Cette partie inclut la description des trois étapes clés de l'analyse protéomique : la préparation des échantillons, l'analyse LC-MS/MS avec l'instrumentation dédiée et le traitement des données de l'identification des protéines aux outils d'interprétation biologique des résultats. Dans cette partie sont aussi présentés les spécificités, les stratégies existantes et les défis inhérents à l'analyse de deux familles de PTMs : les phosphorylations et les glycations.
- La **partie II** regroupe plusieurs développements mis en place pour l'étude des mécanismes biologiques liés à l'hibernation.
    - Dans le **chapitre I**, une approche globale de protéomique quantitative est appliquée pour étudier les régulations au niveau de la mitochondrie musculaire chez *Ursus Arctos* et mieux comprendre le ralentissement du métabolisme durant l'hibernation.
    - Le **chapitre II** décrit l'ensemble des développements analytiques réalisés pour mettre en place une méthode complète d'analyse globale et quantitative des phosphorylations à partir d'échantillons de muscles d'*Ursus Arctos*. Cette méthode regroupe une préparation d'échantillon semi-automatisée permettant un enrichissement en peptides phosphorylés, une méthode MS adaptée à l'étude de cette modification et une évaluation des outils de traitement des données

disponibles. La méthode développée a ensuite été appliquée pour étudier les régulations fines du phosphoprotéome musculaire durant l'hibernation.

- Dans le **chapitre III**, plusieurs protocoles de préparations d'échantillon sont comparés pour déterminer celui qui est le plus adapté pour l'analyse quantitative et globale du protéome cardiaque d'*Eliomys Quercinus.* Le protocole choisi a ensuite été appliqué afin de déterminer l'influence de la composition de la diète en lipides et de l'hibernation sur le protéome cardiaque de ces petits mammifères.

- La **partie III** explore deux stratégies différentes pour l'étude des glycations des protéines sanguines. Ces études ont pour objectif de comprendre l'influence de la glycémie élevée observée chez les oiseaux et, à plus long terme, d'identifier de potentiels mécanismes de résistance à la glycation chez ces espèces.

  - Dans le **chapitre I**, les protéines intactes majoritaires du plasma et de l'hémolysât du diamant mandarin (*Taeniopygia guttata*) sont analysées par LC-MS pour évaluer des taux de glycation chez cet oiseau dont la glycémie est très élevée par rapport à la valeur normale observée chez l'Humain.

  - Le **chapitre II** est dédié aux optimisations des protocoles de préparation d'échantillons pour l'analyse des glycations dans un mélange hautement complexe : le plasma. Des méthodes automatisées permettant, d'une part, un enrichissement des échantillons en peptides glyqués et, d'autre part, un fractionnement peptidique dans le but de décomplexifier le plasma ont été développées.

- La **partie expérimentale** décrit en détail les protocoles utilisés pour ces travaux de thèse.

En parallèle des travaux présentés dans ce manuscrit, j'ai réalisé des analyses globales de protéomique quantitative dans divers contextes biologiques. Les publications issues de ces travaux sont regroupées à la fin de ce manuscrit.

PARTIE I : INTRODUCTION A LA PROTEOMIQUE QUANTITATIVE

Chapitre 1
Analyse des protéines par spectrométrie de masse

Les protéines sont des molécules qui peuvent avoir différentes fonctions telles que catalyser les réactions biochimiques ou avoir un rôle structural au sein des cellules d'un organisme. Elles sont codées à partir de l'information génétique. La notion de protéome, introduite en 1994 par Marc Wilkins par analogie à celle de génome, définit l'ensemble des protéines contenues dans un système biologique à un instant précis et dans des conditions données[17]. Contrairement au génome, le protéome est dynamique et le niveau d'expression des protéines varie dans le temps, selon leur localisation tissulaire ou subcellulaire, et en réponse à divers stimuli[5]. Cette adaptabilité du protéome face aux stimuli environnementaux rend aujourd'hui l'étude des protéines incontournable dans des contextes divers, du domaine médical[18-21] à l'alimentaire[22-23] en passant par les adaptations physiologiques[24-25].

L'étude des protéomes, regroupant l'identification, la quantification, l'analyse structurale et/ou fonctionnelles des protéines, fait référence à l'analyse protéomique, terme introduit par Peter James en 1997[26]. L'importance de la spectrométrie de masse pour l'analyse de macromolécules biologiques, dont les protéines, est due à de nombreuses avancées technologiques, notamment l'apparition de nouvelles sources d'ionisations douces l'ESI[9] et le MALDI[10].

I. Deux approches différentes : analyse "*bottom-up*" et analyse des protéines intactes

L'analyse des protéines par spectrométrie de masse peut être réalisée soit en conservant les protéines intactes, soit après digestion des protéines en peptides, comme illustré en figure 1.

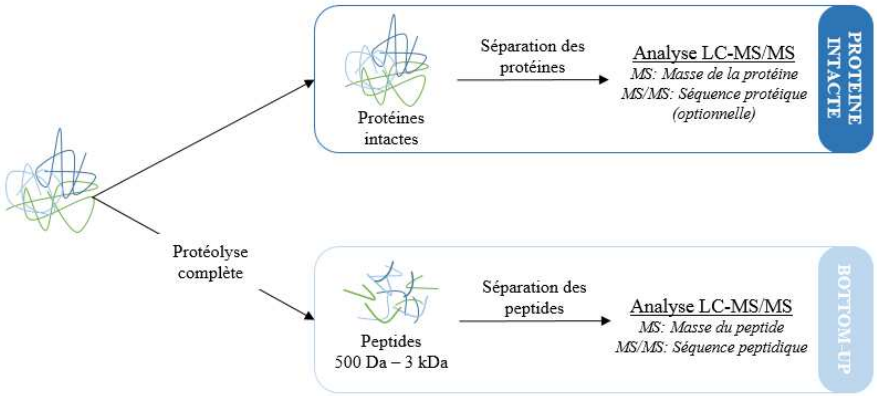

Figure 1 : Représentation schématique des approches en analyse protéomique par spectrométrie de masse : « protéine intacte » et « *bottom-up* ».

L'analyse des **protéines intactes** ou de complexes protéiques permet d'obtenir de nombreuses informations telles que la structure quaternaire des protéines. Ce pan de la protéomique est appelé protéomique structurale. Au contraire, la stratégie « ***bottom-up*** » consiste à analyser les

peptides obtenus après digestion protéolytique des protéines et permet l'identification de la séquence primaire des peptides et des protéines.

A. Analyse structurale des protéines intactes

Les protéines sont constituées d'un enchainement d'acides aminés qui définissent sa séquence ou structure primaire. Elles s'organisent ensuite selon une structure tridimensionnelle stable appelée structure tertiaire ou conformation. Les atomes s'organisent pour minimiser l'encombrement stérique, les répulsions coulombiennes entre groupements de même charge et les répulsions liées aux interactions hydrophiles/hydrophobes avec le milieu. La structure tertiaire de la protéine *in vivo* est appelée « conformation native ». La conformation d'une protéine donnée joue un rôle prépondérant dans sa fonction biologique[27].

La structure des protéines peut être analysée par spectrométrie de masse en conditions natives ou dénaturantes (Figure 2).

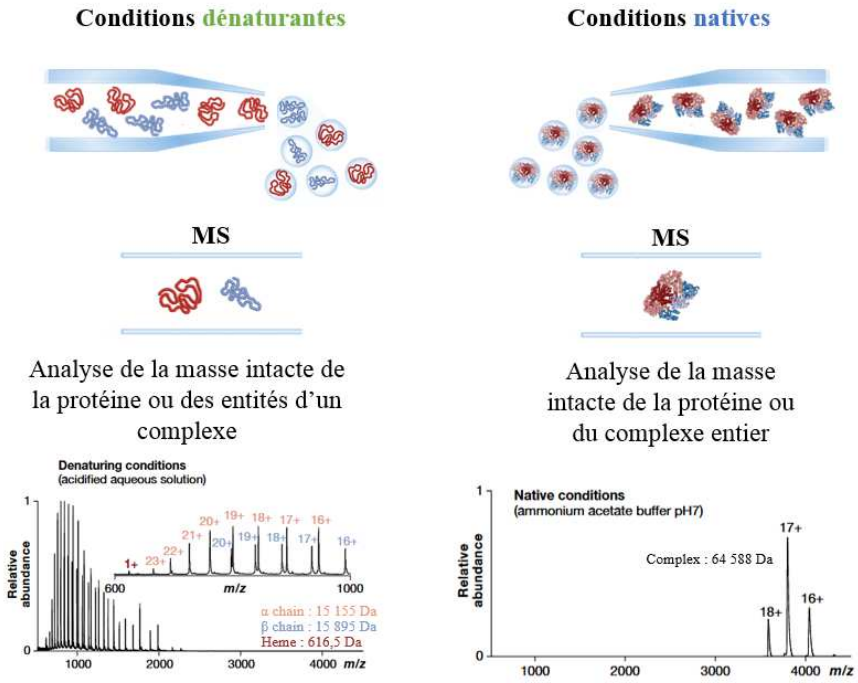

Figure 2 : Représentations schématiques des analyses de protéines intactes et spectres de l'hémoglobine humaine en conditions natives et dénaturantes (d'après Lössl *et al.*, 2016)[28].

En conditions dénaturantes, les protéines peuvent être mises en solution dans des tampons contenant des solvants organiques à des pH acides. Dans de telles conditions, les protéines perdent leur conformation native et se déploient ; on dit qu'elles sont dénaturées. Dans ces conditions, la spectrométrie de masse permet d'observer une large distribution des états de charge d'une protéine donnée à des rapports masse sur charge [m/z] assez faibles (500-3500 m/z) pour des états de charge élevés. Ces conditions vont aussi rompre les liaisons non covalentes existant entre les différentes entités d'un complexe (structure quaternaire), comme c'est le cas pour l'hémoglobine de la figure 2 où on peut observer les masses des chaines α et β et de l'hème[28].

Au contraire, en conditions natives, le milieu doit préserver la conformation native des protéines. C'est pourquoi des conditions expérimentales proches des conditions physiologiques sont utilisées : température ambiante, force ionique (utilisation de bicarbonate ou d'acétate d'ammonium, compatible avec des analyses MS, entre 10 et 50 mM) et pH physiologique (proche de 7 ou 8)[29]. Dans ces conditions, le complexe tétramérique d'hémoglobine constitué de deux chaines α et de deux chaines β ainsi que de 4 hèmes est observé (figure 2)[28].

Dans les deux cas, les échantillons analysés ne peuvent pas être très complexes, il s'agit souvent d'une protéine/un complexe purifié. Lorsque les protéines sont fragmentées dans le spectromètre de masse (analyse MS/MS), on parle alors d'analyse « *top-down* »[30-31].

B. Analyse « *bottom-up* »

L'approche « *bottom-up* » consiste à analyser les peptides obtenus par digestion enzymatique des protéines. Les peptides générés, d'une masse inférieure à 3 kDa, sont généralement séparés par chromatographie liquide [LC] puis ionisés et analysés à l'aide d'un spectromètre de masse afin d'obtenir leur masse (MS1). Ces ions sont ensuite sélectionnés et fragmentés dans la cellule de collision. Les masses des ions fragments sont mesurées (MS2 ou MS/MS) et confrontées aux masses théoriques générées par digestion *in silico* des séquences protéiques issues de banques de données de l'organisme étudié. Ce processus permet l'identification des peptides puis des protéines par inférence. Cette approche permet également la quantification relative (entre différents échantillons) par extraction des courants d'ions [XIC] correspondants aux peptides identifiés[5]. L'abondance relative des peptides mesurée par XIC permet ensuite de quantifier de façon relative les protéines correspondantes à ces peptides. Cette stratégie de quantification est appelée *label-free*.

L'approche « *bottom-up* » comporte donc trois étapes clés récapitulées sur la figure 3 : la préparation d'échantillon, l'analyse LC-MS/MS et le traitement des données.

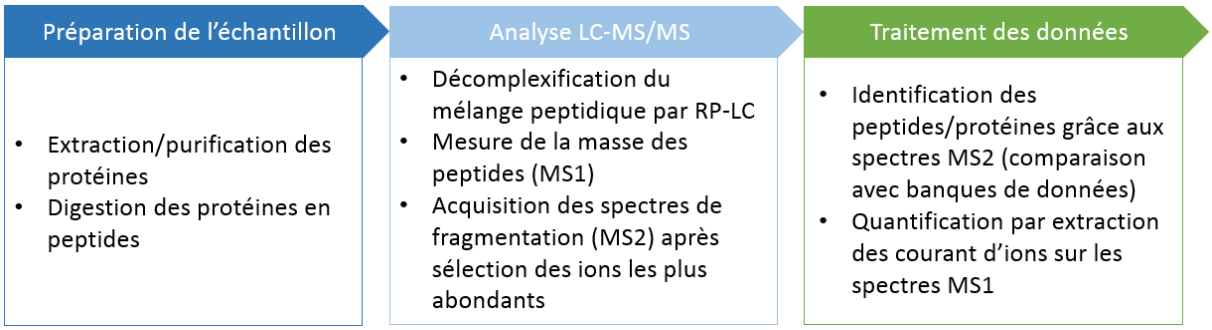

Figure 3 : Principales étapes d'une analyse de protéomique « *bottom-up* » *label-free*.

La stratégie « *bottom-up* » s'est imposée comme une méthode de référence pour l'identification et la quantification de plusieurs milliers de protéines en quelques heures dans des mélanges complexes[32]. Il s'agit de l'approche la plus utilisée dans les travaux rapportés dans ce manuscrit, c'est pourquoi elle sera détaillée ci-après.

II. Les étapes de la préparation d'échantillon d'une analyse « *bottom-up* »

La qualité et la répétabilité des analyses « bottom-up » reposent en grande partie sur la préparation des échantillons. Chaque étape, de l'extraction des protéines depuis l'obtention des peptides avant injection sur le système LC-MS/MS, doit être adaptée au type d'échantillon.

A. L'extraction des protéines

L'une des forces de l'analyse protéomique est de pouvoir être appliquée à des échantillons très divers allant de protéines purifiées dans le cas des protéines thérapeutiques[33] à des organismes entiers[24] en passant par des bio-fluides (plasma[20], *CerebroSpinal Fluid* [CSF][34]...), des cellules cultivées *in-vitro*[35], des tissus[36] ou des organites[37]. Extraire les protéines de ces divers types d'échantillons est un enjeu majeur pour rendre les protéines accessibles aux protéases lors de la digestion enzymatique. L'objectif est de solubiliser un maximum de protéines sans les dégrader, ni les modifier ou introduire de biais. La lyse des cellules et l'extraction des protéines peut être réalisée par deux approches : mécanique ou chimique.

Les approches mécaniques sont généralement utilisées sur des tissus ou des cellules. Les échantillons sont broyés, soit manuellement avec un mortier et un pilon soit à l'aide d'un broyeur à billes sous azote liquide[38]. De plus, les échantillons peuvent aussi être soniqués ou ultra-soniqués pour lyser les cellules et/ou aider à la dénaturation des protéines. La sonication est également utile pour désorber les protéines adsorbées sur les parois des contenants. Des développements récents ont été faits dans ce domaine pour améliorer la fiabilité de cette étape, par exemple avec l'*Adaptive Focused Acoustics*[39] (AFA, Covaris) ou avec le Bioruptor[40] (Diagenode). Cependant, il est important d'être vigilant à la température des échantillons durant les étapes de sonication afin de ne pas dégrader thermiquement les protéines[39].

La lyse des cellules et la mise en solution des protéines par voie chimique utilise des **détergents** dénaturants tels que les détergents ioniques (sodium dodecyl-sulfate, SDS). Il existe également des détergents non ioniques (Triton) ou zwitterioniques (3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate [CHAPS]) ainsi que des sels d'acides biliaires (sodium deoxycholate, SDC)[38] qui sont moins dénaturants. On peut également citer des détergents ioniques compatibles avec la digestion enzymatique comme le RapiGest (Waters)[41] ou l'Azo[11]. Les détergents améliorent l'efficacité de l'extraction et participent à la solubilisation des protéines, en lysant les cellules puis en formant des micelles. Des **agents chaotropiques** tels que l'urée ou la thio-urée, permettent la dénaturation des protéines, de même que des **solvants organiques** (acétonitrile [ACN], méthanol [MeOH]...). Les conditions telles que la concentration en sel, le pH et la température doivent également être adaptées lors de l'extraction et adaptées aux échantillons étudiés.

Il est parfois nécessaire d'isoler les protéines pour retirer certaines impuretés présentes dans le milieu (lipides, acides nucléiques, sels…) et qui pourraient nuire aux analyses en interférant avec les peptides durant l'analyse par exemple en induisant la co-élution des peptides dans les pics chromatographiques ou en supprimant le signal de spectrométrie de masse [MS]. Pour ce faire, il est possible de précipiter les protéines avec de l'acétone ou de l'éthanol glacial, de

l'acide trichloracétique [TCA] ou encore avec un mélange de méthanol et de chloroforme. Les culots protéiques obtenus peuvent être plus ou moins difficiles à resolubiliser en fonction du mode de précipitation choisie. Une autre méthode consiste à dialyser ou filtrer les protéines par ultrafiltration. Cependant, il a été montré que ces étapes peuvent induire des pertes de matériel[42].

B. La digestion enzymatique

La digestion enzymatique est l'étape centrale de l'approche « *bottom-up* » car c'est elle qui génère les peptides qui seront analysés par la suite en spectrométrie de masse. Classiquement, les ponts disulfures des protéines sont réduits et alkylés avant de digérer les protéines Afin d'assurer l'accessibilité des sites de clivage aux protéases utilisées durant la digestion.

1. Réduction et alkylation

Les ponts disulfures sont des liaisons covalentes entre deux résidus cystéines. Au sein d'une protéine, ils participent à la régulation de sa structure tertiaire[43]. Ces liaisons sont rompues par une réaction de réduction pour assurer l'accessibilité des sites de clivages aux protéase durant la digestion. Les groupements sulfures sont ensuite alkylés pour empêcher que les ponts disulfures ne se reforment à nouveau.

Pour la réduction, le dithiothreitol [DTT] ou le β-mercaptoéthanol (composant essentiel du tampon de Laemmli très utilisé en protéomique) sont les agents les plus utilisés, mais le Tris-(2-carboxyethyl) phosphine [TCEP] peut aussi être utilisé. Le TCEP et le DTT sont de plus en plus utilisés car le β-mercaptoéthanol est plus nocif pour la santé de l'expérimentateur[44].

Pour l'alkylation, différents agents existent : l'iodoacétamide [IAM], l'acide iodoacétique [IAA], l'acrylamide et le chloroacétamide. Le plus couramment utilisé en protéomique est l'IAM.

Ces différents réactifs sont représentés en figure 4.

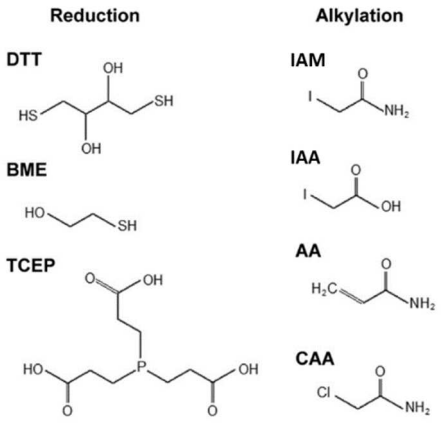

Figure 4 : Réactifs utilisés pour la réduction : DTT, BME = β-mercaptoéthanol, TCEP. Réactifs utilisés pour l'alkylation : IAM, IAA, AA = acrylamide et CAA = chloroacétamide.

Ces différents réactifs ont été comparés sur un lysat de cellules HeLa[45]. Cette étude a montré que le choix des agents réducteurs et alkylants avaient une forte influence sur le nombre de

peptides spectrum match [PSM] et donc sur le nombre de peptides identifiés. Au terme de cette étude, les auteurs préconisent l'abandon des agents alkylants iodés (IAM et IAA) car ils sont sources de modifications non spécifiques (alkylation des résidus méthionines par exemple)[45].

2. Protéases utilisées

Après réduction et alkylation des protéines, celles-ci sont digérées, le plus souvent avec la trypsine. En effet, la trypsine est facile d'utilisation et d'accès. Elle est spécifique car elle clive les peptides à l'extrémité C-terminale des lysines et des arginines[46-47]. Elle génère alors des peptides dont la masse est généralement comprise entre 500 et 3000 Dalton [Da] (entre 7 et 40 acides aminés en moyenne) couvrant de façon satisfaisante les séquences protéiques grâce à l'abondance naturelle des lysines et des arginines[46]. De plus, la présence d'une charge positive à l'extrémité C-terminale du peptide favorise son ionisation et sa fragmentation selon le modèle du proton labile rendant ainsi les peptides trypsiques adaptés aux analyses LC-MS/MS[48-49]. La digestion par la trypsine est effectuée à 37°C dans des conditions de pH physiologique, le plus souvent sur la nuit (sur une durée de 12 à 18h). Toutefois, si un excès d'enzyme est introduit ou si la digestion est trop longue, la trypsine peut s'autolyser générant des peptides qui vont entrer en compétition avec les peptides de l'échantillon lors de l'analyse LC-MS/MS[50].

Ces dernières années, il est devenu évident que la seule utilisation de la trypsine en protéomique « *bottom-up* » ne permet pas d'obtenir une couverture de séquence complète, notamment pour des peptides portant des sites particuliers de modifications post-traductionnelles. Pour accéder à un séquençage plus complet, la communauté scientifique a commencé à explorer des protéases alternatives: la chymotrypsine, la LysN, l'AspN, la GluC, la LysC ou encore l'ArgC[47, 51]. L'utilisation combinée de deux protéases permet souvent d'identifier plus de peptides et de protéines, comme cela a été montré par exemple en combinant trypsine et AspN sur des échantillons protéiques issus de levures[51]. Par ailleurs, l'association de la trypsine à la Lys-C semble également particulièrement efficace et son utilisation se répand au sein de la communauté scientifique[20, 52]. La Lys-C présente en effet l'avantage d'être résistante aux conditions dénaturantes, permet de générer de longs peptides qui sont ensuite mieux digérés par la trypsine[52-53].

Il existe plusieurs approches pour réaliser la digestion enzymatique : la digestion en solution, la digestion sur gel, la digestion sur filtres et sur billes.

3. Digestion en solution

Dans ce type de protocole, l'enzyme est ajoutée directement dans le milieu contenant les protéines. Cette approche est particulièrement simple[42], comme illustré sur la figure 5, mais elle nécessite l'utilisation d'un tampon compatible avec la digestion enzymatique et l'analyse LC-MS/MS. C'est pourquoi, pour dénaturer les protéines, en plus de la réduction et de l'alkylation des cystéines, des tampons contenant de l'urée sont souvent utilisés car ils sont compatibles avec la digestion enzymatique une fois dilués[54]. Cependant, l'urée et les autres sels doivent être éliminés avant l'analyse LC-MS/MS. Pour ce faire, une étape d'extraction sur phase solide [SPE] de type C18 est couramment utilisée[55] (Figure 5). Une alternative à la SPE a été développée récemment : la SP2[56]. Ce protocole utilise des billes paramagnétiques

fonctionnalisées avec des groupements carboxylates pour retenir les peptides et ainsi les purifier.

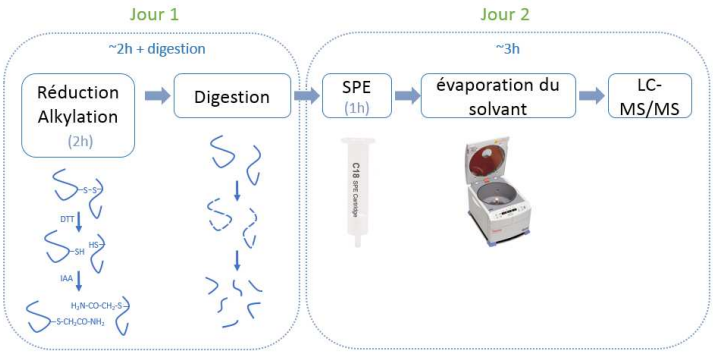

Figure 5 : Schéma analytique d'une digestion liquide.

Si ce protocole de digestion est rapide et facile à mettre en œuvre, il reste cependant limité par sa faible compatibilité avec les réactifs couramment utilisés pour réaliser la lyse des cellules et notamment le SDS.

4. Digestion en gel

Le SDS, bien qu'il soit très utile pour lyser les cellules et solubiliser les protéines, même les plus hydrophobes, n'est compatible ni avec la digestion enzymatique liquide, ni avec l'analyse MS au cours de laquelle il peut causer une suppression complète du signal à des concentrations aussi basses que 0,01%[7]. C'est pourquoi une approche compatible avec le SDS a été développée : le protocole *Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis* [SDS-PAGE], illustré sur la figure 6.

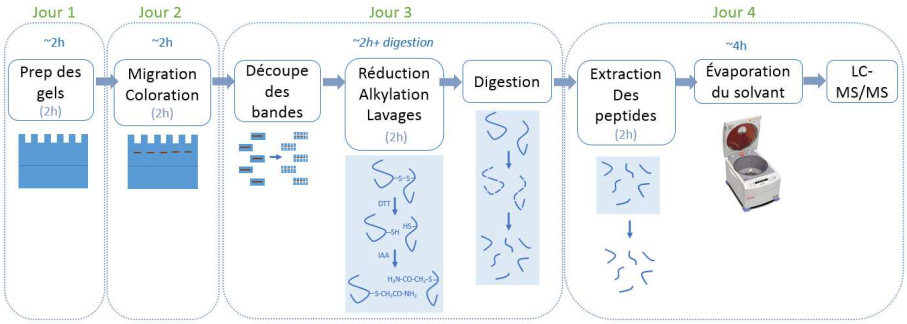

Figure 6 : Schéma analytique d'une digestion en gel SDS-PAGE.

Les protéines extraites sont tout d'abord linéarisées et chargées négativement sous l'action du SDS. Elles sont ensuite déposées sur le gel et la migration des protéines dans le gel est réalisée grâce à l'application d'un champ électrique. Grâce au SDS qui uniformise la charge des protéines, elles sont séparées uniquement suivant leur masse moléculaire par migration dans le gel d'acrylamide qui agit comme un tamis moléculaire. La séparation est contrôlée par la réticulation du gel, qui est régie par la concentration en acrylamide et bis-acrylamide. Le gel est composé de deux parties : une première portion nommée **gel de concentration** contient peu d'acrylamide/bis-acrylamide (4-5%) où les protéines sont concentrées en une seule bande (« stacking ») avant de pénétrer dans le **gel de séparation** dans lequel les protéines seront

séparées en fonction de leur masse moléculaire grâce à un réseau plus étroit (entre 8 et 15% d'acrylamide/bis-acrylamide). Une fois les protéines séparées, leur présence est révélée par coloration au bleu de Coomasie. Les bandes protéiques sont ensuite excisées et les protéines piégées dans le gel peuvent être lavées pour retirer le SDS et les autres composés non protéiques. Par la suite, les protéines sont réduites et alkylées puis digérées directement dans le gel. En effet, la trypsine est suffisamment petite pour entrer dans le gel grâce à des cycles de déshydratation/hydratation. La taille des peptides obtenus leur permet ensuite d'être élués du gel de façon passive, et de manière plus active grâce à des cycles de déshydratation/hydratation[42, 57]. Ce protocole est longtemps resté une référence dans le domaine de la protéomique, mais il est relativement long à mettre en œuvre et difficilement automatisable. C'est pourquoi, ces 10 dernières années, de nouveaux protocoles, plus adaptés aux analyses haut débit et compatibles avec l'utilisation de détergents, ont été développés.

5. Digestion sur membrane (Suspension Trap – S-Trap)

Le protocole S-Trap utilise une solution commercialisée par Protifi (Farmingdale, NY, USA). Quatre formats sont disponibles, dont trois sous forme de cartouches : micro, mini et midi qui permettent théoriquement de travailler respectivement de 1 µg à 100 µg, de 100 µg à 300 µg et jusqu'à 10 mg de protéines. Le quatrième format est une plaque à 96 puits qui est adapté aux plateformes automatisées. Comme la plupart des solutions commerciales, c'est une technique plutôt coûteuse mais elle est compatible avec de fortes concentrations de SDS. Pour ce type de protocole, la digestion par la trypsine peut être réalisée en quelques heures car les protéines sont très dénaturées et les sites de clivages sont donc très accessibles. Ce protocole est illustré en figure 7.

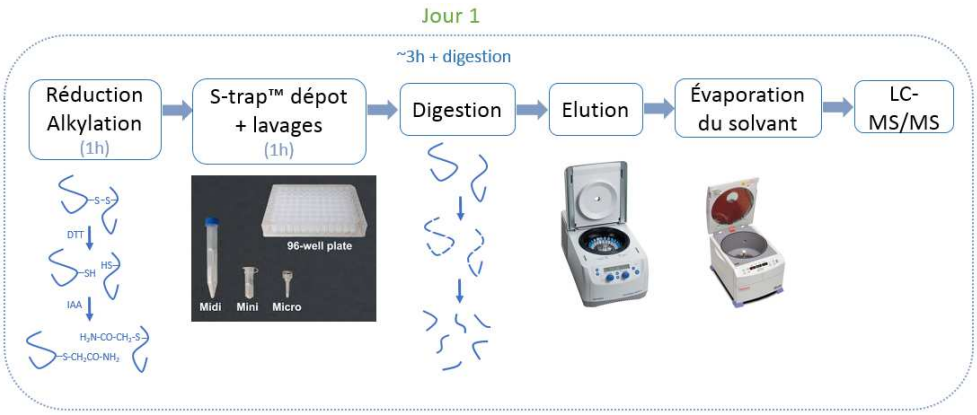

Figure 7 : Schéma analytique d'une digestion avec le protocole S-Trap.

Les cellules doivent être lysées avec au minimum 5% de SDS. Après réduction et alkylation des protéines, une émulsion est créée en ajoutant un tampon de liaison contenant 90% de MeOH et 10% de bicarbonate de triéthylammonium [TEAB] 1 M, à pH 7,1. Ensuite, les protéines dénaturées sont déposées sur la membrane dans la cartouche et lavées plusieurs fois. Après la digestion des protéines sur la membrane, les peptides sont élués de la cartouche par centrifugation. Les protocoles *Filter Aided Sample Preparation* [FASP][58-59], *in-StageTip* [iST] (PreOmics)[60], MStern[61], Sample Preparation Kit (Biognosis) et Pierce™ Mass Spec Sample Prep Kit (Thermo Fisher Scientific) fonctionnent sur un principe similaire : les protéines sont

piégées sur un filtre puis lavées et digérées sur le filtre avant que les peptides ne soient récupérés par centrifugation[62].

Le protocole S-Trap est une technique relativement récente (initialement décrite en 2014[62]) mais qui est de plus en plus répandue. Elle a été appliqué à une très large variété d'échantillons : des levures[62-63], des bactéries[64-65], de l'urine[61], des tissus *Formalin-Fixed Paraffin Embedded* [FFPE][66-67], des protéines immunoprécipitées[62] ou encore des protéines membranaires[62, 68-69] qui sont particulièrement hydrophobes et pour lesquelles l'utilisation de SDS est nécessaire. Ce protocole donne des résultats probants dans les différentes publications qui la comparent à d'autres préparations d'échantillons comme la digestion en solution[70], la digestion sur gel[68], le protocole iST[61, 71-73], le protocole FASP[12, 68, 70] et le protocole *Single Pot Solid Phase enhanced Sample Preparation* [SP3][74].

6. Digestion sur billes fonctionnalisées (SP3)

Le protocole SP3 utilise des billes paramagnétiques fonctionnalisées avec des groupements carboxylates. Les protéines vont se lier aux billes dans des conditions particulières, typiquement à pH physiologique avec 50% d'acétonitrile. Les mécanismes avancés pour expliquer ces liaisons sont basés sur des mécanismes similaires à celles mises en jeu dans chromatographies de type *Hydrophilic Interaction Liquid Chromatography* [HILIC] et *Electrostatic Repulsion-hydrophilic interaction LIquid Chromatography* [ERLIC][75] mais également sur l'agrégation des protéines sur les microsphères (*Protein Aggregation Capture* [PAC]) due aux fortes concentrations en solvants organiques[40, 56, 76]. Les billes portant les protéines sont ensuite retenues par l'utilisation d'un rack magnétique et le milieu peut être retiré. Les billes sont alors rincées plusieurs fois avec une combinaison de solutions qui permettent d'éliminer efficacement les « contaminants » avant d'ajouter l'enzyme pour réaliser la digestion. Les peptides sont enfin élués des billes et ces dernières sont séparées des peptides grâce au rack magnétique[39-40, 77]. Ce protocole est illustré en figure 8.

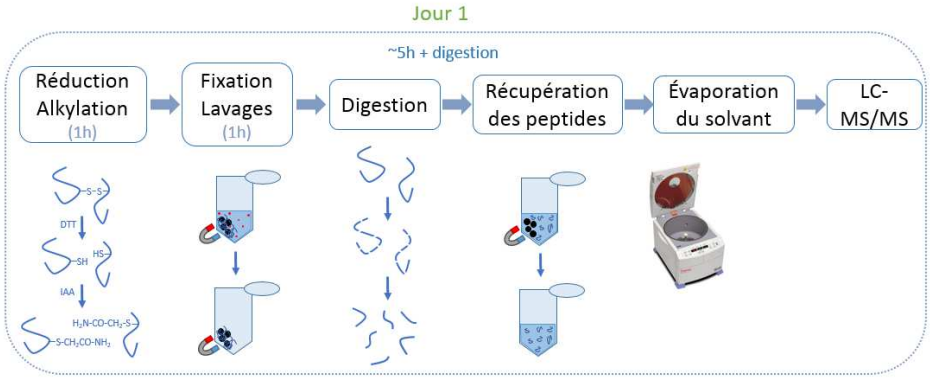

Figure 8 : Schéma analytique d'une digestion avec le protocole SP3.

Comme pour le protocole S-Trap, le protocole SP3 a été appliqué avec succès à une très large gamme d'échantillons : des levures[78-79], des tissus FFPE[80-81], de la peau[82], des protéines immunoprécipitées[83], des échantillons de paléoprotéomique[84-85]. Le protocole SP3 a été comparé à d'autres méthodes de préparations d'échantillons telles que FASP et iST ; les performances obtenues sur les différentes quantités de matières sont identiques de 1 et 20 µg

de protéines issues de cellules HeLa. Sur les faibles quantités de matériel (1-5 µg), le nombre de protéines identifiées est supérieur au FASP et équivalent à l'iST[71]. Le protocole SP3 a également été comparé au S-Trap[74, 86-88], à la digestion en gel[74]. Sur des extraits protéiques issus de différentes bactéries, le protocole SP3 a donné de meilleurs résultats en termes d'identifications que le protocole S-Trap et que la digestion en gel[74]. De même sur des extraits de plantes, il a été montré que le protocole SP3 a permis une augmentation des identifications par rapport au protocole S-trap[87]. Sur des échantillons de moules et de foie de poisson, les performances des deux protocole (SP3 et S-trap) se sont avérés similaires[86]. Enfin, sur des extraits de pathogènes de plante, le protocole S-Trap a permis d'augmenter le nombre de protéine identifiées et la répétabilité par rapport au protocole SP3[88]. Ces comparaisons montrent bien qu'il n'existe pas de protocole qui soit adapté à tous les échantillons.

Ce protocole présente de multiples avantages, notamment celui d'être compatible avec du SDS (jusqu'à 10%)[77] et d'être facilement adaptable à la quantité de matériel disponible en ajustant la quantité de billes. C'est une technique très versatile qui a déjà été adaptée pour purifier des peptides (SP2[56]), ou pour réaliser des analyses « *top-down* »[89-90]. De plus, deux groupes ont déjà automatisé ce protocole en utilisant une plateforme de préparation d'échantillons Bravo AssayMAP[39, 91]. Récemment un protocole analogue utilisant des billes inertes non magnétiques et une étape de précipitation, appelé SP4, a montré des résultats équivalents en terme d'identifications à partir de cellules HEK293 à la SP3[92]. De plus, des billes paramagnétiques fonctionnalisées avec différents groupements (HILIC, -OH…) sont commercialisées[93].

C. Décomplexification du mélange protéique

Les échantillons biologiques sont souvent hautement complexes de par le grand nombre de protéines qui les composent et leur large gamme dynamique de concentration. Ainsi il existerait plus de 6 millions de protéoformes différentes dans le protéome humain[4]. Le terme « protéoforme » renvoi à l'ensemble des formes que peut prendre une protéine issue d'un seul et même gène, cette définition englobe à la fois les variations épigénétiques et les modifications post-traductionnelles[94]. De plus, la gamme dynamique de concentration peut atteindre jusqu'à 11 ordres de grandeurs[95] tandis que les spectromètres de masse ne parviennent à couvrir que 5 ordres de magnitude[96-97]. Afin d'augmenter la couverture du protéome analysé, le mélange protéique peut être décompléxifié en amont de l'analyse LC-MS/MS par fractionnement au niveau protéique ou peptidique. Ces deux niveaux de fractionnement permettent l'analyse du protéome mais augmentent le nombre d'analyses en amont.

1. Fractionnement au niveau protéique

Il existe plusieurs approches pour décompléxifier un échantillon au niveau protéique. La plus répandue est le fractionnement sur gel SDS-PAGE. Les protéines sont séparées selon leur masse moléculaire et les bandes protéiques sont ensuite excisées et les protéines digérées par un protocole de digestion sur gel comme décrit dans le paragraphe précédent (II.B.4). Un exemple de gel d'électrophorèse SDS-PAGE sur lequel des protéines ont migré jusqu'à l'obtention de 5 fractions (bandes protéiques) est illustré en figure 9.

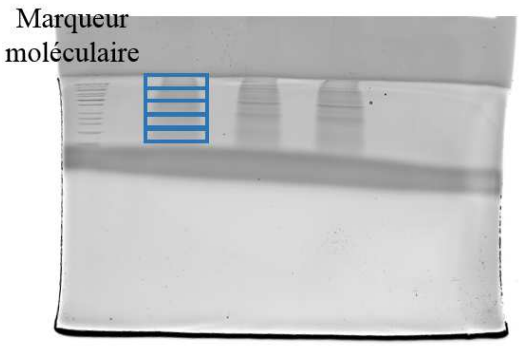

Figure 9 : Gel SDS-PAGE utilisé pour fractionner l'échantillon en 5 fractions (bandes de 2mm).

Dans le cas de l'analyse de fluides biologiques tels que le plasma, le sérum ou le CSF, la gamme dynamique des concentrations protéiques est extrême et la plus grande partie du contenu protéique de l'échantillon correspond à un nombre très restreint de protéines. Pour le plasma, 99% du protéome est représenté par 20 protéines majoritaires[98]. Une stratégie, très utilisée pour augmenter la couverture du protéome analysé dans ce type d'échantillon, est de retirer les protéines majoritaires. Cette technique, appelée déplétion, est le plus souvent réalisée par immunoaffinité[95]. Pour le plasma ou le sérum, il existe de nombreux kits commerciaux sous forme de colonnes où sont greffés des anticorps et qui permettent de retirer les protéines majoritaires (entre les 6 et 20 protéines les plus abondantes) : Multiple Affinity Removal System (MARS, Agilent), ProteoPrep kit (Sigma), Proteome Purify (R&D System)[99]. Ces kits, outre leur prix élevé, sont limités par la capacité de la colonne. La déplétion peut également être réalisée par précipitation des protéines majoritaires comme l'albumine et les immunoglobulines du sérum/plasma à l'éthanol[99]. Cependant, il faut souligner qu'aucune de ces méthodes n'est parfaitement spécifique des protéines déplétées, même l'immunodéplétion est susceptible de ne retirer qu'une fraction de ces protéines majoritaires. De plus, en éliminant les protéines majoritaires, il est possible de retirer les protéines moins abondantes qui interagissent avec elles, c'est le cas notamment avec l'albumine qui est un très bon transporter[98]. L'immunodéplétion est d'autant plus compliquée à mettre en place pour des espèces non modèles (autres que l'homme ou la souris) pour lesquelles les épitopes ciblés par les anticorps ont pu être modifiés au cours de l'évolution. Ce type de fractionnement peut donc induire un biais de reproductibilité important nuisible dans le cas d'une analyse quantitative des protéines[98].

La chromatographie d'exclusion stérique [SEC] permet de séparer les protéines selon leur volume hydrodynamique. La phase stationnaire de la colonne est constituée de particules sphériques poreuses. Les analytes avec le volume hydrodynamique le plus élevé circulent entre les particules et sont peu retenus par la colonne tandis que les plus petits analytes entrent dans les pores des billes, ce qui retarde leur élution. Cette technique peut être utilisée en amont d'analyses « *bottom-up* » pour produire des fractions contenant des protéines de taille différentes[100-101]. Elle peut être utilisée en conditions natives car elle ne repose pas sur une séparation par affinité entre les analytes et la phase stationnaire[102].

2. Fractionnement au niveau peptidique

Le fractionnement peut également être effectué au niveau peptidique, c'est à dire après la digestion des protéines. Il est alors nécessaire d'utiliser une technique orthogonale à la chromatographie en phase inverse à pH acide (typiquement inférieur ou égal à 2) utilisée pour l'analyse LC-MS/MS. Actuellement, l'une des techniques les plus utilisées est le fractionnement sur phase inverse à pH basique (pH ≈ 10, [hpH RP]). La différence de sélectivité observée entre le fractionnement et la séparation de l'analyse LC-MS/MS sur phase inverse est due aux différentes valeurs de pH qui modifient les états de charge des chaines latérales des acides aminés des peptides[103-104]. Le fractionnement hpH RP peut être réalisé sur différents supports chromatographiques : colonne HPLC[105-106], kits commerciaux (Thermo - 12 fractions[107] ou PreOmics - 3 fractions), cartouches pour micro-chromatographie[108-109] ou tips[110]. Cette méthode de fractionnement peptidique est particulièrement utilisé en phosphoprotéomique avant ou après l'enrichissement pour augmenter la couverture du phosphoprotéome analysé[106, 111-112]. Il a récemment été automatisé sur la plateforme Bravo AssayMAP[108-109].

La chromatographie échangeuse d'ions, surtout par *Strong Cation Exchange* [SCX][113-114] mais aussi par *Strong Anion Exchange* [SAX][45, 115], permet également de fractionner des extraits peptiques. Il est admis que l'orthogonalité entre la SCX et la phase inverse est plus importante que celle entre le fractionnement hpH RP et la phase inverse en milieu acide de l'analyse LC-MS/MS, car le mécanisme de séparation entre ces deux dimensions reste dépendant de l'hydrophobicité des peptides malgré la différence de forme des peptides induite par la différence de pH de la phase mobile[116-118]. Cependant, la SCX souffre d'une mauvaise résolution qui conduit à une séparation limitée des peptides tandis que le fractionnement hpH RP est plus résolutif[117, 119].

Récemment un protocole utilisant des billes paramagnétiques fonctionnalisées avec des groupements carboxylates, identiques à celles utilisées pour réaliser la SP3, a été développé pour réaliser un fractionnement peptidique, *Carboxylate-Modified Magnetic Bead [CMMB]-Based Isopropanol Gradient Peptide Fractionation* [CIF]. Les peptides sont fixés sur les billes avec un très haut pourcentage en acétonitrile (90%) puis élués séquentiellement avec un pourcentage décroissant en isopropanol[120].

D. Automatisation de la préparation d'échantillons en protéomique

Ces dernières années, la protéomique a évolué pour devenir une méthode utilisée pour des études quantitatives à large échelle dans l'espoir découvrir de nouveaux biomarqueurs de maladies ou des nouveaux traitements[18, 95]. La préparation d'échantillon étant une étape longue et source de nombreuses variations, il existe un réel besoin d'automatisation pour augmenter non seulement le débit mais aussi et surtout la répétabilité/reproductibilité des analyses. L'automatisation permet également de travailler sur des quantités de matériel toujours plus faible (< 1 µg de protéines)[121].

De nombreuses étapes de l'analyse protéomique, voir des protocoles complets, ont été automatisés à ce jour. La digestion liquide ainsi que les protocoles de S-trap ou de SP3[39] sont

automatisables. Les étapes de réduction/alkylation, SPE[122], fractionnement sur phase inverse[109] ou SCX peuvent également être réalisées par de manière automatisée. Pour l'analyse des phosphorylations ou d'autres PTMs, l'étape d'enrichissement peut également être intégrée dans un protocole automatisée[108, 123].

Il existe de nombreux robots pour la manipulation des liquides. Les deux plus répandus en protéomique sont le Bravo AssayMap (Agilent) et la plateforme KingFisher Flex (Thermo) équipée d'une tête magnétique 96 broches, mais on peut également citer le MicroLab Star (Hamilton), le Resolvex A200 (Tecan), la série de stations de travail Biomek (Beckman Coulter life science) et PreON (PreOmics)[124]. Il est intéressant de noter le développement de la possibilité d'interfaçage entre la "préparation de l'échantillon" et le système LC-MS/MS, comme c'est le cas avec le robot cellenONE[125] (CELLENION) pour les analyses protéomiques d'une cellule isolée (*single cell*) ou avec le système ADE-OPI-MS développé par AB Sciex[126]. Ceci permet de réduire drastiquement la perte de matériel et le temps nécessaire à l'analyse

III. Instrumentation : la chromatographie liquide couplée à la spectrométrie de masse

A. Séparation peptidique par chromatographie liquide

Les mélanges obtenus après digestion protéolytique sont extrêmement complexes et contiennent plusieurs dizaines de milliers de peptides. C'est pourquoi, la chromatographie liquide est systématiquement utilisée pour décomplexifier le mélange peptidique avant l'analyse MS. Ainsi, elle diminue la compétition entre les analytes à l'ionisation et permet d'augmenter la sensibilité, la sélectivité et la couverture du protéome analysé.

Pour les analyses protéomiques, le système le plus répandu est la chromatographie liquide en phase inverse en conditions acides (pH < 2). Ce type de chromatographie sépare les peptides selon leur hydrophobicité en diminuant progressivement la polarité de la phase mobile composée généralement d'un mélange d'eau acidifié et d'acétonitrile. Trois systèmes chromatographiques ont été utilisés pour réaliser des analyses « *bottom-up* » durant ce travail de thèse et sont détaillés en tableau 1.

Tableau 1 : Systèmes chromatographiques utilisés pour les travaux de ce manuscrit.UHPLC = *Ultra High Pressure Liquid Chromatography.*

Système LC	NanoAcquity UHPLC	NanoElute UHPLC	Ultimate 3000 UHPLC
Fabricant du système	Waters	Bruker Daltoniks GmbH	Thermo Fisher Scientific
Fabricant de la colonne	Waters	Ion Optiks	Ion Optiks
Phase stationnaire	C18	C18	C18
Longueur de la colonne (mm)	250	250	250
Diamètre interne (µm)	75	75	75
Taille des particules (µm)	1,7	1,6	1,6
Taille des pores (Å)	130	120	120
Débit analytique (nL/min)	350 ou 400	300	300

Différents critères influent sur la séparation des peptides en chromatographie liquide comme la composition des solvants, le débit ou le gradient utilisé. Les paramètres de la colonne (longueur, diamètre interne, taille des pores, taille des particules) jouent un rôle crucial dans la qualité de la séparation[9, 127]. Les systèmes chromatographiques (*Ultra High Pressure Liquid Chromatography* [UHPLC]) qui permettent de travailler avec ces nanocolonnes fonctionnant à des débits inférieurs au microlitre par minute sont appelés nanoLC (nLC) et fonctionnent à ultra haute pression (> 400 bars). Les systèmes nano-chromatographiques, bien qu'ils soient source de problèmes récurrents (instabilité du spray, fuites difficilement détectables, volumes morts), permettent d'augmenter la résolution, la capacité des pics et la sensibilité des analyses. De plus, ces systèmes sont tout à fait adaptés aux échantillons souvent précieux et en quantités limitées qui sont analysés en protéomique car ils ne nécessitent qu'une faible quantité de matériel (100-350 ng de peptides).

B. Spectrométrie de masse en tandem

À la sortie de la colonne, les peptides sont ionisés *via* une ESI avant de pénétrer dans le spectromètre de masse. Les instruments utilisés permettent de réaliser des analyses en tandem. En mode *Data Dependant Acquisition* [DDA], un premier spectre MS (ou MS1) est acquis : l'instrument mesure le ratio m/z et l'intensité de chaque ion. Les ions sont ensuite isolés puis fragmentés et un deuxième spectre MS (ou MS2) est acquis[127].

Au cours des travaux présentés dans ce manuscrit, j'ai eu l'occasion de travailler avec 3 types de configurations de spectromètres : quadripôle-Orbitrap [Q-Orbitrap], quadripôle-*Time Of* Flight [Q-TOF] et Tribrid (Q-Orbitrap-IT). Leurs principales caractéristiques sont détaillées en tableau 2.

Tableau 2 : Fiche technique des spectromètres de masses utilisés pour les travaux de ce manuscrit.

OT = Orbitrap, IT = *Ion Trap*, HCD= *Higher-Energy Collisional Dissociation*, CID = *Collision-Induced Dissociation*, UVPD = *UltraViolet Photo-Dissociation*, ETD = *Electron Transfer Dissociation*, TIMS = *Trapped Ion Mobility Spectrometry*, FAIMS = *Field Asymmetric Ion Mobility Spectrometry,* PTCR = *Proton Transfer Charge Reaction, ppm = partie par million.*

Système MS	Q-Exactive Plus	Q-Exactive HF-X	timsTOF Pro	Eclipse
Fabricant du système	Thermo Fisher Scientific	Thermo Fisher Scientific	Bruker Daltoniks GmbH	Thermo Fisher Scientific
Analyseur	Q-Orbitrap	Q-Orbitrap	Q-TOF	Instrument Tribrid (Q-Orbitrap-IT)
Résolution	De 17 500 à 140 000 (à 200 m/z)	De 7 500 à 240 000 (à 200 m/z)	40 000 (à 622 m/z)	OT: 7 500-1M (à 200 m/z) IT : 3 – 0.3 FWMH
Précision de masse	5 ppm	5 ppm	10 ppm	OT : 5 ppm
Vitesse d'acquisition	Jusqu'à 12 Hz	Jusqu'à 40 Hz	> 100 Hz	OT : Jusqu'à 40 Hz IT : 2200-125 000 scans/seconde
Gamme de masse	6 000 m/z	8 000 m/z	-	8 000 m/z
Type de fragmentation	HCD	HCD	CID	HCD , CID, ETD, UVPD, PTCR
Mobilité ionique	-	-	TIMS	Option FAIMS
Année d'installation	2014	2017	2019	2020

Les deux instruments les plus récents sont équipés d'une dimension de séparation supplémentaire : la mobilité ionique. Depuis le développement du *Field Asymmetric Ion Mobility Spectrometry* [FAIMS] (Thermo)[128] et du *Trapped Ion Mobility Spectrometry* (TIMS, Bruker)[129], la mobilité ionique se répand dans le monde de la protéomique. La double cellule TIMS, implémentée dans le timsTOF Pro avant le quadripôle permet de séparer les ions en fonction de leur taille et charge en phase gazeuse et d'accéder à un niveau supplémentaire d'information : les valeurs de mobilité ionique [K_0]. Plus généralement, on parle de mobilité

réduite [$1/K_0$] ou de section efficace de collision moyenne en rotation (*Collision Cross Section* [CCS]) qui correspond à la projection 2D de la structure 3D d'un ion. Le CCS peut être calculé à partir de la mobilité réduite. Pour tirer pleinement avantage de cette séparation supplémentaire, le mode *Parallel Accumulation Serial Fragmentation* [PASEF] a été développé[129]. Il se décompose en deux étapes :

- Séparation des ions dans la double cellule TIMS : La première partie de la cellule TIMS est utilisée pour accumuler les ions. Les ions sont entraînés dans la cellule TIMS par un flux constant de gaz inerte et sont retenus par l'application d'un champ électrique statique. Lorsque le temps d'accumulation est atteint, les ions accumulés sont transférés vers la deuxième partie de la cellule TIMS. Dans cette deuxième partie, les ions sont séparés en fonction de leur charge et de leur forme en phase gazeuse. A état de charge égal, les ions les plus gros seront entraînés par le flux gazeux plus loin dans le tunnel et seront plus proches de sa sortie. Les ions sont ensuite élués séquentiellement par ordre de CCS décroissant dans la partie suivante du spectromètre de masse, à savoir le quadripôle, par diminution du champ électrique. Au même moment, la première partie de la cellule TIMS a accumulé les ions suivants et le cycle recommence permettant un taux d'utilisation des ions d'environ 100% (Figure 10A). C'est donc que 100% des ions qui arrivent à la source sont transmis à l'analyseur.

- Isolation ciblée des ions élués de la cellule TIMS par le quadripôle : L'élution des précurseurs de la deuxième partie de la cellule TIMS est synchronisée avec leur sélection par le quadripôle en utilisant le traitement en temps réel des informations MS1. Cela permet au quadripôle de sélectionner les ions de manière ciblée plutôt que de balayer toute la gamme m/z, ce qui accélère encore la vitesse d'acquisition du spectromètre de masse (Figure 10B). Chaque spectre de masse MS/MS correspond à un voltage d'élution donné et une valeur de mobilité ionique donnée.

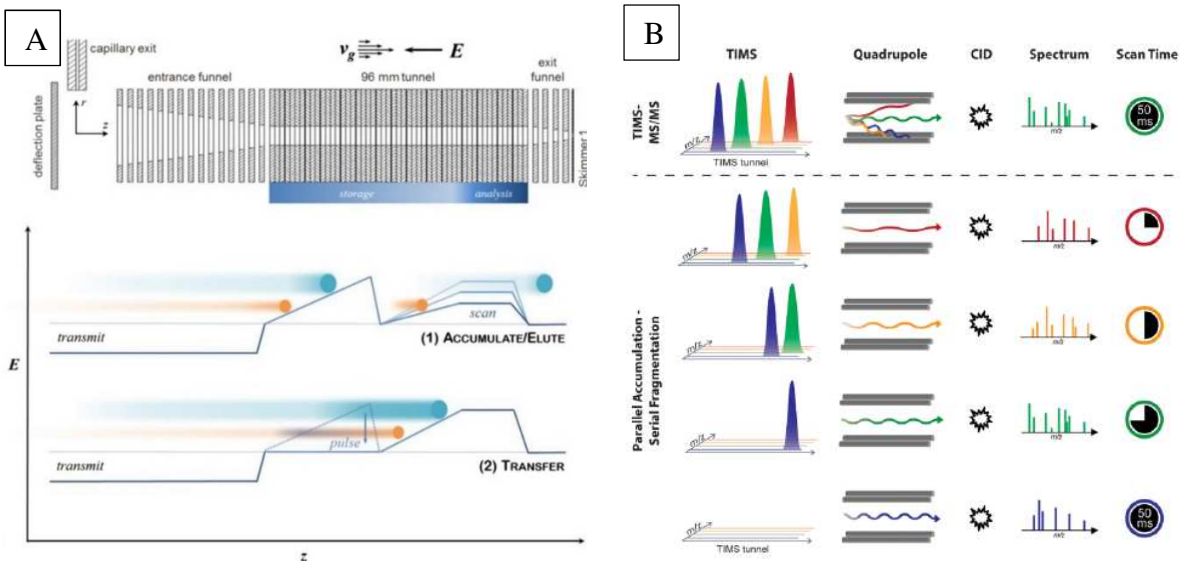

Figure 10 : A : Utilisation parallèle des deux cellules TIMS du timsTOF : accumulation des ions dans la première cellule de mobilité et elution séquentielle des ions simultanément (1) dans la seconde puis transfert des ions accumulés (2). B : Principe du mode PASEF : en fonction de la valeur de mobilité ionique, le ratio m/z est selectionné de manière ciblée par le quadripole[129].

Cette séparation supplémentaire permet d'isoler des peptides qui co-éluent dans un pic chromatographique, comme par exemple des peptides isobares phosphorylés[130] (qui portent la même modification localisée sur des acides aminés différents). De plus, le mode d'acquisition PASEF confère au timsTOF Pro une acquisition extrêmement rapide et sensible : plus de 100 spectres MS/MS peuvent être acquis en environ 1 seconde. Un schéma général du mode PASEF est illustré en figure 11. Cet appareil permet d'ajouter une dimension supplémentaire dans l'analyse, on parle d'analyse à quatre dimensions : les temps de rétention, les masses des peptides sur les spectres MS1, les masses des fragments sur les spectres MS2 et la valeur de mobilité ionique.

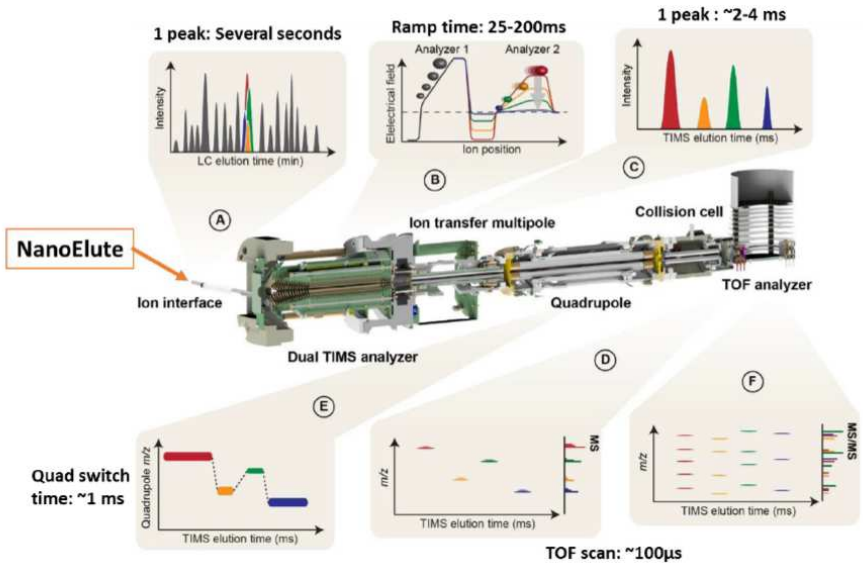

Figure 11 : Principe général du mode PASEF.
A : Les peptides élués de la colonne chromatographique sont ionisés et entrent dans l'interface à travers un capillaire de transfert en verre. B: La première section TIMS accumule les ions, et la seconde les sépare selon leur valeur de mobilité. C et D : les ions séparés par mobilité ionique sont élués séquentiellement du second analyseur TIMS en fonction de la diminution de l'intensité du champ électrique et analysés par le TOF pour produire des spectres MS à mobilité résolue. E : Dans les scans MS/MS PASEF, l'analyseur TIMS et le quadripôle sont synchronisés et la fenêtre d'isolation du quadripôle passe en quelques millisecondes d'un ion précurseur à mobilité résolue à un autre. F : Des spectres MS/MS à mobilité résolue sont obtenus à partir d'un seul balayage TIMS. Les spectres MS et MS/MS non résolus en mobilité sont projetés sur les axes de droite en (D) et (F) pour comparaison.

C. Fragmentation des peptides

Dans les approches « *bottom-up* », la fragmentation des peptides vise à rompre la liaison peptidique entre chaque acide aminé pour permettre l'identification de la séquence du peptide. Cette fragmentation peut générer différents types d'ions qui ont été classifiés par Biemman[131], voir figure 12.

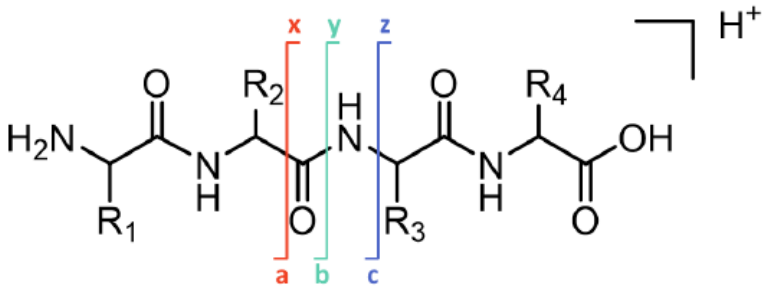

Figure 12 : Ions obtenus après fragmentation comme définis par la nomenclature de Biemann. Il existe six types d'ions : a, b, c en N-ter et x, y, z en C-ter.

La fragmentation peut être réalisée de différentes manières. Elle peut être obtenue par collision (*Collision Induced Dissociation* [CID] et *Higher Energy Collisional Dissociation* [HCD]) ou

par transfert d'électrons (*Electron Transfer Dissociation* [ETD]). Les fragmentations CID et HCD sont les méthodes de dissociation les plus utilisées. Dans les deux cas, les ions sont accélérés et entrent en collision avec des atomes de gaz neutre (argon, hélium, azote) dans la chambre de collision. L'énergie cinétique est convertie en énergie interne induisant la rupture de la liaison peptidique selon le modèle du proton mobile : la migration d'un proton mobile vers les liaisons amides induit le clivage[48-49, 132]. Pour la fragmentation HCD, les ions sont accélérés ; de fait l'énergie produisant la rupture est plus importante avec ce type de fragmentation qu'avec une fragmentation CID. La fragmentation HCD est spécifique aux instruments de type Orbitrap, dont un exemple est représenté sur la figure 13A. Sur ce type d'instrument, le transport des ions vers la cellule HCD ou vers l'analyseur est entièrement coordonné par la C-trap. Les fragmentations CID et HCD conduisent à la formation d'ions b et y comme représenté sur le spectre de la figure 13B.

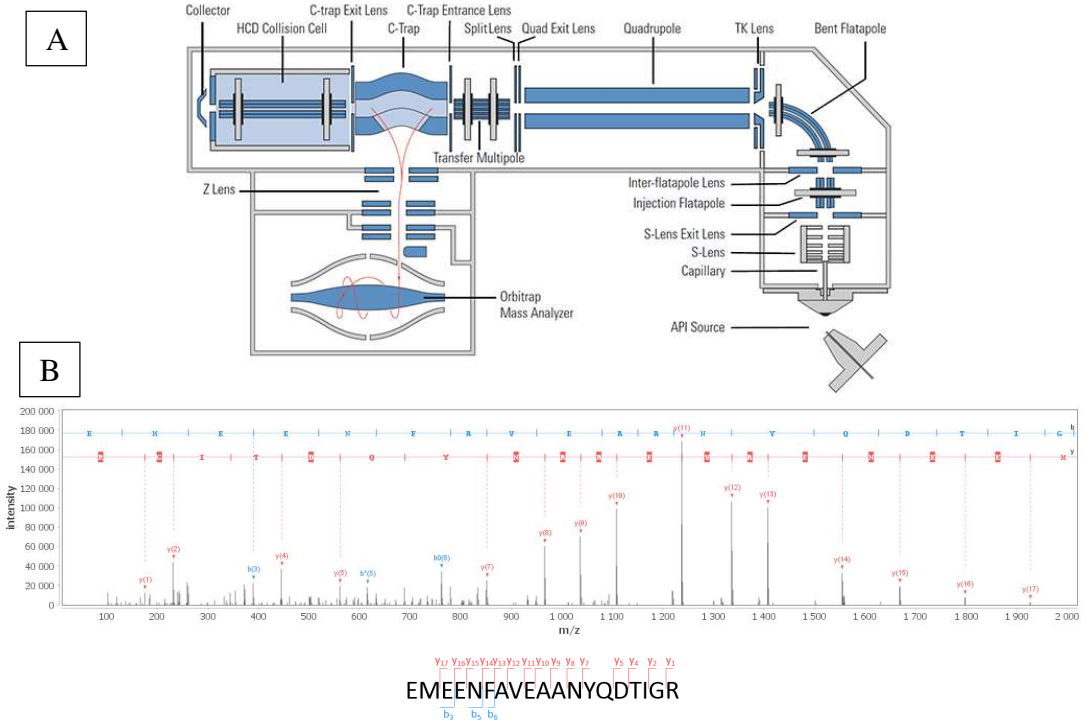

Figure 13 : Architecture d'un Q-Orbitrap (Q-Exactive) (A). Spectre MS/MS du peptide EMEENFAVEAANYQDTYGR obtenu sur un Q-Exactive Plus équippé d'une cellule de fragmentation HCD. Des ions b et y sont générés et permettent l'identification des acides aminés constituant le peptide (B).

La fragmentation ETD est obtenue grâce à la réaction du peptide avec un anion (souvent du fluoranthène généré par une source d'ionisation chimique) *via* un transfert d'électrons provoquant la fragmentation le long de la chaine peptidique. La fragmentation ETD génère surtout des ions c et z. L'ETD est une fragmentation plus douce qui peut être utile pour l'analyse des PTM labiles[133].

Il existe également d'autres méthodes de fragmentations telles que la fragmentation *Electron Capture Dissociation* [ECD][134], *UltraViolet Photo-Dissociation* [UVPD][135], *Electron-Transfer/Higher-Energy Collision Dissociation* [EThcD][136-137] ou *Electron-*

Transfer/Collisionally Activated Dissociation [ETcaD][138], mais elles ne seront pas utilisées pour les travaux détaillés dans ce manuscrit.

D. L'acquisition en mode DDA

Le mode d'acquisition DDA est le plus répandu actuellement en analyse « *bottom-up* ». Des cycles d'acquisition de spectres MS1 et MS2 s'enchainent tout au long du gradient. La sélection des ions à fragmenter est effectuée sur la base de leur intensité. Ainsi, les N ions précurseurs (aussi appelé Top N) les plus intenses d'un spectre MS1 sont isolés puis fragmentés avant d'être analysés et de générer N spectres MS2 suivant le schéma décrit sur la figure 14.

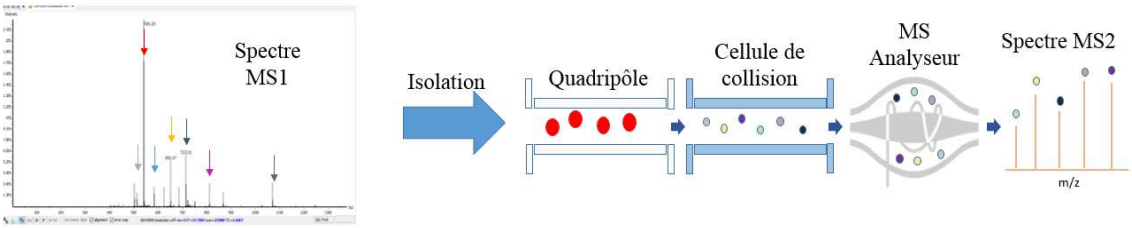

Figure 14 : Schéma d'une analyse DDA. Un spectre MS1 est acquis puis les N ions les plus intenses sont isolées 1 à 1 par ordre décroissant d'intensité. Une fois isolé dans le quadripôle, l'ion est fragmenté par collision avec du gaz puis les fragments sont analysés générant un spectre MS2.

Avec ce mode d'acquisition, l'échantillonnage stochastique des peptides pour la fragmentation entraîne un biais en privilégiant ceux dont le signal est le plus intense, ce qui nuit à l'identification/quantification répétable des peptides moins abondants. Bien que les instruments soient toujours plus sensibles et rapides, il existe néanmoins un effet de sous échantillonnage[139]. À titre d'exemple, avec trois réplicas d'injection d'un protéome de cellules HeLa réalisées sur une NanoAcquity (Waters) couplée à un Q-Exactive HF-X (Thermo), un recouvrement protéique de 75% et un recouvrement peptidique de 50% ont été obtenus au laboratoire. Pour accroitre le nombre d'identification, l'exclusion dynamique des ions les plus abondants permet de limiter le biais vers les ions majoritaires et réduit la redondance des spectres obtenus[140]. Des listes d'exclusion ou d'inclusion peuvent aussi être utilisées dans le même but[141].

Par ailleurs, il existe d'autre modes de fonctionnement qui permettent une quantification absolue tels que les modes *Single Reaction Monitoring* [SRM], *Parallel Reaction Monitoring* [PRM] et *Multiple Reaction Monitoring* [MRM].

IV. Traitement de données : de l'identification à l'interprétation biologique

Les informations acquises par le spectromètre de masse, à savoir les valeurs de m/z (rapport masse/charge) des ions d'après les spectres MS1 et MS2 ainsi que les intensités des ions mesurés et leur temps de rétention sont compilés dans une « *peaklist* ». Ces données expérimentales sont confrontées à une banque de données théoriques issues de la digestion *in silico* des séquences protéiques de l'espèce étudiée avec l'enzyme utilisée lors de la digestion. Cette approche porte le nom de *Peptide Fragmentation Fingerprint* [PFF][142]. L'identification des peptides conduit à celle des protéines par inférence. Les peptides sont soit uniques à une protéine de la banque soit partagés entre plusieurs protéines ; ces dernières peuvent alors être

regroupées dans un « *protein group* ». L'inférence protéique est un processus complexe qui respecte le principe de parcimonie afin que la liste « *protein group* » soit la plus restreinte possible[143].

A. Identification et quantification

1. L'importance de la banque de données

Le choix de la banque de données contenant les séquences protéiques de l'espèce étudiée est capital pour une identification correcte des peptides et des protéines puisqu'elle constitue l'espace de recherche. Une banque de données protéiques contenant des séquences redondantes ou erronées nuira grandement aux analyses.

Il existe plusieurs banques de données disponibles :

- UniprotKB[144] a été créée par une collaboration entre l'institut Européen de Bio-informatique (EMBL-EBI), l'institut Suisse de Bioinformatique [SIB] et le consortium « *Protein Information Ressource* (PIR) ». Elle regroupe deux banques différentes :
    - SwissProt est la banque de données protéiques de référence utilisée en protéomique lorsqu'il existe suffisamment d'entrées pour l'organisme étudié. Toutes les entrées qui la composent ont été manuellement vérifiées pour ne retenir que les informations dite de haute qualité. En plus de la séquence protéique, de nombreuses autres informations sont disponibles comme sa fonction, sa localisation sub-cellulaire, les PTMs connues, les variants de séquence, les interactions avec d'autres protéines ou sa structure tridimensionnelle. A la date du 12 août 2022, cette banque contenait 568 002 entrées.
    - TrEMBL contient toutes les séquences protéiques automatiquement traduites, annotées et classées provenant de GenBank ainsi que des séquences provenant de la littérature. Ces séquences sont en attente de validation pour être intégrées à SwissProt. La qualité de cette banque de données est assez faible, il existe beaucoup de redondances et d'erreurs. A la date du 12 août 2022, elle contenait 226 771 948 entrées.
- Refseq[145], est produite par le Centre National pour l'information Biotechnologique (NBCI). Les données sont issues de l'annotation automatisée des données génomiques. Une partie de ces données a été validée manuellement et dans ce cas l'identifiant des protéines porte le préfixe NP, les autres protéines portant le préfixe XP. Pour toutes les protéines, le lien entre le gène, le transcrit et la protéine est disponible. Au 15 juillet 2022 (Release 213), cette banque contient 234 520 053 protéines pour 121 461 organismes différents. Cependant, des erreurs lors du séquençage ou lors de la traduction des séquences nucléiques en séquences protéiques peuvent impacter négativement l'interprétation des données MS[146].

Les banques de données sont biaisées vers les organismes les plus étudiés (Homme, rat, souris) pour lesquels des informations de qualité sont disponibles car des efforts conséquents ont été

fournis. Pour les organismes moins étudiés, les informations disponibles peuvent être très lacunaires surtout si l'organisme n'a pas été séquencé.

2. Moteurs de recherche

Il existe plusieurs moteurs de recherche permettant de comparer les données de MS expérimentales et théoriques pour identifier les peptides et les protéines. On peut citer Andromeda[147], Mascot (Matrix Science), Pulsar (Spectronaut), Sequest (Thermo) ou encore MSAmanda[148].

Des informations sont nécessaires à ces logiciels pour prendre en compte les conditions expérimentales :

- La banque de séquences protéiques définissant l'espace de recherche
- La protéase utilisée lors de la digestion enzymatique et le nombre de clivages manqués autorisé
- Les modifications de certains acides aminés à rechercher, soit celles qui sont permanentes (ou fixes ; comme l'alkylation des cystéines), soit celles qui sont variables (par exemple les oxydations des méthionines ou les phosphorylations de STY)
- La tolérance de masse acceptée sur les ions parents (spectres MS1) et ions fragments (MS2)
- Le type de fragmentation utilisé

Pour les travaux présentés dans ce manuscrit, seuls Mascot, Andromeda et MS Amanda ont été utilisés :

Mascot est un moteur de recherche commercialisé par Matrix Science (London, UK). Son algorithme n'est pas open-source. Chaque spectre MS2 lui permet de calculer un score nommé « *ion score* ». Celui-ci a pour but d'évaluer la probabilité que la correspondance entre le spectre et une séquence de la base de données théoriques soit un faux positif. Plus le score est élevé, plus l'identification est robuste. Le score est donné au niveau du PSM qui correspond à l'ensemble des identifications associées à un spectre MS.

Andromeda est un moteur de recherche libre de droit mais non open-source développé par Cox *et al*. au Max Planck Institut[147]. Il est intégré au logiciel MaxQuant qui permet de recalibrer les spectres et les temps de rétention. Son fonctionnement est similaire à celui de Mascot : un score est attribué aux identifications *via* un système de notation basé sur les probabilités. Les échelles de score de Mascot et Andromeda sont différentes mais les résultats obtenus par ces deux moteurs de recherche sont comparables[147].

MSAmanda est également libre de droit mais non open-source. Il est développé par le groupe de Karl Mechtler (*Institute of Molecular Pathology* (IMP), Vienne)[148]. Il peut être utilisé seul ou au sein de Proteome Discoverer (Thermo). Son principe est analogue à celui de Mascot et Andromeda mais, contrairement à Mascot, il a été développé spécifiquement pour les analyses en haute résolution.

En moyenne, entre 60 et 75% des spectres MS2 ne sont pas assignés par ces moteurs de recherches[149], plusieurs facteurs pouvant expliquer ceci :

- La qualité des spectres MS2 peut être insuffisante si l'intensité globale est trop faible ou si le nombre de fragment est insuffisant
- La qualité de la banque est également primordiale comme mentionné dans le paragraphe précédent (IV.A.1).

- Des spectres chimériques peuvent être générés si plusieurs précurseurs sont co-isolés et co-fragmentés malgré la fenêtre d'isolation restreinte généralement utilisée (entre 1 et 3 m/z). Depuis la version 2.5, Mascot propose d'identifier tous les peptides possibles à partir de spectres chimériques. Andromeda propose une seconde recherche (« *second peptide search* ») sur les spectres MS2, après retrait des fragments ayant servi à l'identification du premier peptide, si des ions parents avec des masses proches ont été identifiés. Depuis 2018, charmeRT est implémenté avec MSAmanda pour traiter spécifiquement ces spectres chimériques sur un principe similaire au « *second peptide search* » de MaxQuant[150].

- Les peptides peuvent porter des modifications nombreuses et variées qui induise des différences de masses par rapport à la séquence non modifiée du peptide. Si les incréments de masse ne sont pas spécifiés pour la recherche, les peptides modifiés ne seront pas identifiés. Un tiers des spectres non assignés pourrait être dû à ce phénomène[151-152]. En effet, les PTMs recherchées comme modifications variables font exploser les besoins en ressources informatiques et en temps de calcul et elles sont donc recherchées avec parcimonie. Cependant, de nouveaux moteurs de recherche plus rapides et plus adaptés aux recherches sans enzymes ou avec de nombreuses PTMs (*open search*), apparaissent, tels que MSFragger[153] ou ionbot[154].

- Des erreurs dans le traitement des données, telles que l'extraction incorrecte des pics, l'attribution incorrecte du pic monoisotopique ou l'attribution d'un mauvais état de charge peuvent avoir lieu.

3. Validation des résultats

Les logiciels utilisés pour l'interprétation des données MS ne sont pas infaillibles et risquent d'introduire de fausses assignations de peptides/protéines dans les résultats. Le score assigné aux PSMs n'est pas suffisant pour juger de la véracité de l'assignation et il n'est pas envisageable de vérifier manuellement les milliers de spectres d'une analyse. C'est pourquoi des méthodes de validation automatiques ont été développées. La plus répandue est la stratégie « cible/leurre » (« *target/decoy* ») qui permet d'évaluer le taux de faux positifs dans un jeu de données[155]. Pour le calculer, des données « leurre » ou « *decoy* » sont ajoutées à la base de données théorique (espace de recherche). Chaque identification de séquence *decoy* est par définition fausse, ce qui permet de calculer le nombre de faux positifs du jeu de données. Pour créer les séquences leurres, une stratégie communément utilisée est d'inverser les séquences des protéines de la banque initiale, on obtient ainsi le même nombre de séquence leurres, qui ont une probabilité très faible de se retrouver dans les données expérimentales[155]. Le calcul du *False Discovery Rate* [FDR] est effectué comme décrit par la relation 1[156].

$$FDR = \frac{Nombre\ de\ séquence\ leurre\ assignées}{nombre\ de\ séquences\ leurres\ assignées + nombre\ de\ séquences\ cibles\ assignées} * 100$$

Relation 1 : Calcul du FDR par l'approche *target/decoy*

Le FDR peut être appliqué au niveau des PSMs, des identifications peptidiques et/ou protéiques. Les assignations sont classées selon leur score puis filtrées jusqu'à l'obtention d'un FDR inférieur au seuil accepté (généralement 1%). Plusieurs logiciels permettent de filtrer les identifications grâce à la stratégie « cible/leurre », parmi lesquels on peut citer Proline[157], Andromeda ou Proteome Discoverer. Proteome Discoverer permet aussi l'utilisation d'une autre méthode de validation basée sur une intelligence artificielle : Percolator. Percolator est un algorithme permettant la validation des PSMs grâce à un apprentissage semi-supervisé (*machine learning*)[158].

B. Quantification par courant d'ion extrait

Il existe plusieurs stratégies pour la quantification des protéines. Elles peuvent être réalisées avec (marquage métabolique ou chimique) ou sans (« *label-free* ») marquage préalable des protéines de l'échantillon à analyser. Les stratégies sans marquage sont plus faciles à mettre en place et moins onéreuses. De plus, elles sont utilisables sur tous les types d'échantillons et le nombre d'échantillons n'est pas limité car ils ne sont pas multiplexés[159]. Pour l'ensemble de ces raisons, ces approches sans marquage sont souvent préférées. Néanmoins, chaque étape de l'analyse doit être parfaitement répétable car les échantillons sont préparés et analysés séparément. Dans ce manuscrit, la quantification des peptides et des protéines a été réalisée exclusivement selon une stratégie globale sans marquage par extraction des courant d'ions. Il s'agit de la stratégie la plus couramment utilisée de nos jours en protéomique.

Dans cette approche, l'intensité du signal chromatographique obtenu en MS1 est considérée comme proportionnelle à l'abondance du peptide. Le courant d'ion (XIC) correspondant au peptide est extrait puis l'aire sous la courbe est mesurée pour obtenir les valeurs quantitatives. Pour cette méthode, les paramètres d'acquisition doivent être optimisés afin d'obtenir suffisamment de spectres MS1 pour définir un pic chromatographique et pour générer suffisamment de spectres MS2 de bonne qualité pour identifier les peptides[160]. Pour des résultats de qualité, un système chromatographique robuste et reproductible est nécessaire[159].

Pour pallier d'éventuelles variations induites tout au long du protocole d'analyse protéomique (préparation d'échantillons, fluctuations chromatographiques, instabilité du signal), des recalibrations sont nécessaires au niveau des temps de rétentions et des intensités observées (normalisation des données). Plusieurs solutions logicielles ont été développées et sont proposées dans les logiciels Proline (Profi), MaxQuant (Max Plank Institut), Skyline (MacCoss Lab) ou Proteome Discoverer (Thermo).

Pour limiter l'impact de la stochasticité du mode DDA, si un peptide est identifié à un temps de rétention et un rapport m/z donné, il est possible de reporter l'information de l'identification dans d'autres analyses au niveau des signaux MS1, et ce même si le peptide n'a pas toujours été isolé et fragmenté pour être analysé en MS/MS. Cette option est appelée « *Match Between Runs* »[161] dans le logiciel MaxQuant et « *Cross Alignment* » dans Proline. Cette technique

permet d'obtenir une valeur quantitative même dans des échantillons où les peptides n'ont pas pu être correctement identifiés.

Pour assurer la qualité des résultats et la robustesse de la méthode avec un protocole « *label-free* », des contrôles qualité doivent être mis en place. En effet, le nombre d'échantillons analysés est souvent important et il n'est pas rare que les injections durent plus d'une semaine. Il est alors nécessaire de s'assurer de la stabilité du système chromatographique et du spectromètre de masse. Dans le cadre des analyses « *label-free* » réalisées dans les travaux de ce manuscrit, deux méthodes ont été utilisées :

- Des standards internes ont été ajoutés à tous les échantillons : les peptides « *indexed Retention Time* [iRT] » sont 11 peptides synthétiques commercialisés par Biognosys (Schlieren, Suisse). Ils sont ajoutés en même quantité dans tous les échantillons puis les temps de rétention et les intensités sont suivies tout au long des analyses.
- Des standards externes sont analysés régulièrement. Ces standards doivent refléter au mieux les échantillons analysés, c'est pourquoi ils sont souvent constitués d'une quantité équivalente de chaque échantillon, cet assemblage étant réalisé avant la digestion enzymatique. Cet échantillon de référence (appelé pool) ainsi crée est ensuite préparé comme les autres échantillons. Il est injecté à intervalle de temps régulier tout au long de l'analyse des échantillons. Les variations des intensités des protéines au cours de la séquence d'analyses sont ensuite suivies.

Une fois les intensités des protéines calculées, des tests statistiques sont réalisés pour déterminer les protéines dont l'abondance varie de façon significative entre les conditions étudiées.

C. Informations fonctionnelles

L'analyse protéomique quantitative globale permet l'identification et la quantification relative de plusieurs centaines à milliers de protéines simultanément. Avec autant de données, l'interprétation biologique peut être compliquée.

1. Banques de données fonctionnelles

Il existe plusieurs banques de données regroupant des informations fonctionnelles ou spatiales (localisation sub-cellulaire) :

- Uniprot[144] : cette banque de séquences protéiques intègre de nombreuses informations relatives à la fonction, la localisation des protéines. Cependant, les données ne sont pas mises en forme de façon systématique ce qui complique leur traitement.
- *Gene Ontology* [GO][162-163] au contraire a classifié de nombreuses annotations fonctionnelles. Trois types d'informations sont disponibles : la fonction moléculaire, le processus biologique dans lequel la protéine est impliquée et sa localisation subcellulaire (Figure 15A). Dans ces trois catégories, des termes GO définissent des fonctions ou localisation précises. Les termes GO sont soigneusement organisés pour former un réseau dans lequel on peut naviguer comme illustré par un exemple en figure 15B.

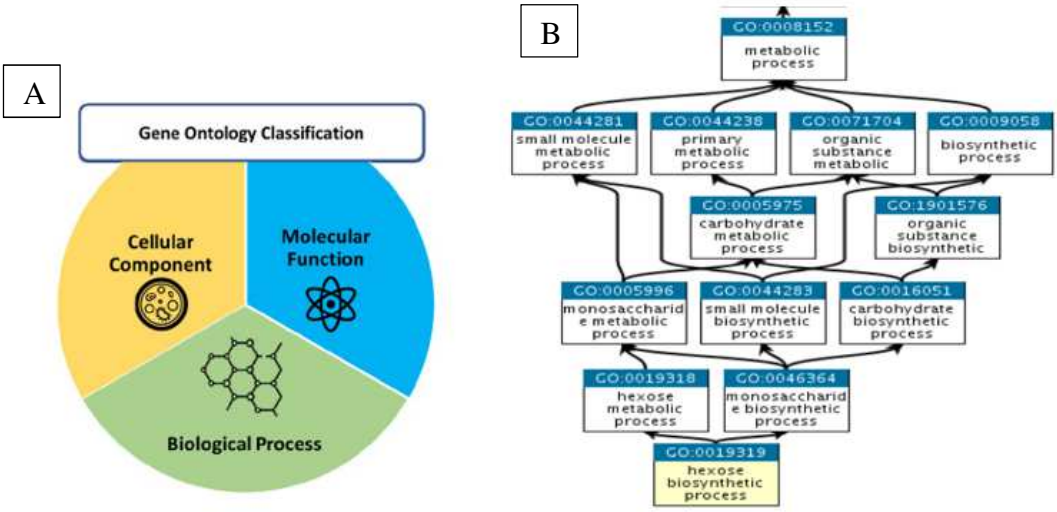

Figure 15 : A : Les trois catégories de Gene Ontology : Cellular Component, Molecular Function et Biological Process. B : Exemple d'un arbre de classification de termes GO.

Les termes GO sont régulièrement revus manuellement par des bio-informaticiens pour limiter la redondance et améliorer les relations entre eux tout en les attribuant correctement aux différentes protéines. Au vu du travail considérable nécessaire pour créer et maintenir ce genre banque de données, on comprend que ces informations ne soient disponibles que pour une poignée d'espèces : seules 185 espèces sont annotées avec plus de 1000 annotations en juillet 2022. Afin d'extraire les termes GO de la base de données Gene Ontology pour un grand nombre de protéines, un algorithme a été implémenté sur la plateforme Galaxy du laboratoire (https://iphc-galaxy.u-strasbg.fr/).

- *Kyoto Encyclopedia of Genes and Genomes* [KEGG][164] est une base de données japonaise permettant de visualiser les voies métaboliques et les cascades de signalisation dans lesquelles une protéine est impliquée (Figure 16).

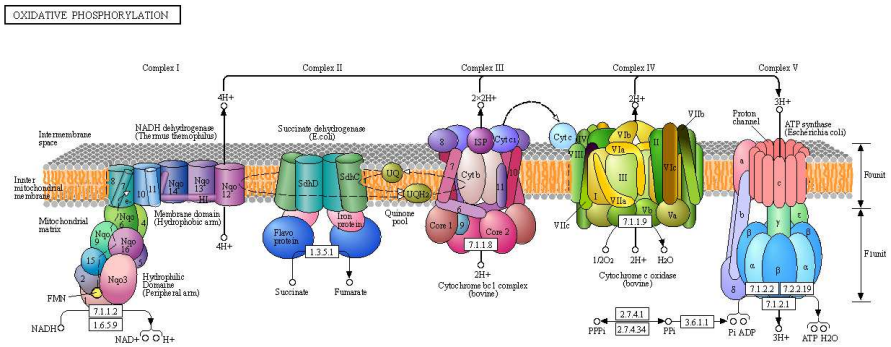

Figure 16 : Exemple d'une Carte KEGG « 190 : oxidative phosphorylation » représente la chaine mitochondriale composée de différents complexes ».

Un algorithme permettant l'identification des cartes de la base de données KEGG qui contiennent les protéines recherchées et la visualisation des protéines sur les cartes avec un code couleur reflétant les statistiques et les changements d'abondance des protéines

selon des seuils définis par l'utilisateur est intégré à la plateforme Galaxy du laboratoire (https://iphc-galaxy.u-strasbg.fr/).

- neXtProt[165] est une base de données du SIB regroupant de nombreuses annotations fonctionnelles pour les protéines humaines. Les informations sont issues principalement d'Uniprot, de Peptide Atlas[166-167] et du Human Protein Atlas[168]. Elles sont manuellement vérifiées et catégorisées avant d'être intégrées à NeXtProt. Les informations de très bonne qualité sont catégorisées « *Gold* » et celles de moins bonne qualité « *Silver* ». Dans cette banque, on trouve des informations sur la localisation subcellulaire des protéines, leur fonction, leur implication dans certaines maladies, les variants de séquences ou encore la structure et les PTMs associées.

À partir de certaines des informations qui peuvent être récupérées dans ces banques de données, des calculs statistiques peuvent être réalisés pour mettre en évidence les fonctions biologiques les plus impactées dans un jeu de données.

2. Calcul de l'enrichissement en annotations fonctionnelles

Pour déterminer quelles sont les voies biologiques les plus représentées par une liste de protéines, typiquement la liste des protéines différentiellement exprimées entre plusieurs conditions, l'algorithme EASE David[169] peut être utilisé. Cet algorithme, contrairement à l'utilisation de sa version online (https://david.ncifcrf.gov/), permet de prendre en compte les mises à jour des banques de données d'annotations fonctionnelles utilisées. Différentes annotations peuvent être prises en compte par l'algorithme, notamment les termes GO et les cartes KEGG. L'algorithme EASE David permet de comparer la fréquence d'occurrence d'une fonction biologique dans un jeu de données expérimental par rapport à la fréquence d'occurrence de cette fonction dans la banque de donnée entière de l'espèce considérée. La différence de fréquence d'occurrence de l'annotation fonctionnelle considéré entre le données expérimentales et théoriques est statistiquement évaluée par un test de Fisher et un test de Benjamini-Hochberg. Un calcul de facteur d'enrichissement est ensuite réalisé selon la relation 2.

$$\text{Facteur d 'enrichissement (FE) de l'annotation X} = \frac{\left(\frac{\text{Nombre de protéines avec l'annotation X}}{\text{Nombre de protéines dans le système considéré}}\right)_{\textit{Dans le jeu de données}}}{\left(\frac{\text{Nombre de protéines avec l'annotation X}}{\text{Nombre de protéines dans le système considéré}}\right)_{\textit{Dans la population entière}}}$$

Relation 2 : Calcul du facteur d'enrichissement d'une annotation X issue d'un système d'annotation (KEGG, GO Cellular component…) dans un jeu de données expérimentales par à l'ensemble de la banque de séquences théoriques (population entière).

Les tests statistiques et le facteur d'enrichissement permettent d'évaluer l'importance avec laquelle une fonction biologique donnée est statistiquement très représentée pour une liste de protéines données. Lorsqu'on effectue ce calcul sur les protéines différentiellement exprimées entre les conditions, cela révèle les fonctions biologiques les plus régulées.

Il existe également d'autres ressources qui permettent d'aider à l'interprétation biologique tels que STRING qui permet la visualisation des interactions entre les protéines ou Panther[170]. On peut également citer IntAct, MINT, BIOGRid et Reactome.

3. Identifier une protéine homologue *via* un algorithme FASTA

C'est d'autant plus le cas pour des espèces non modèles pour lesquelles les banques de données fonctionnelles qui permettent de connaitre les voies biologiques dans lesquelles les protéines sont impliquées sont peu renseignées. Pour surmonter ces problèmes, des solutions bio-informatiques ont été mises en place.

Dans un premier temps, la séquence d'une protéine x d'une espèce donnée est utilisée pour identifier son homologue chez un organisme modèle, cet homologue permettant ensuite d'interroger des banques de données fonctionnelles telles que Gene Ontology, Kegg ou même Uniprot (voir paragraphe suivant). Les protéines homologues sont des protéines dont les séquences sont très proches puisqu'elles sont le produit d'un gène ancestral commun, et leur identification fait généralement appel à des recherches par similarité de séquence, par exemple *via* un algorithme *Fast-All* [FASTA]. Un score, un pourcentage de similarité et un pourcentage d'identité sont calculés pour chaque alignement. Seule la correspondance ayant le score le plus élevé est généralement retenue.

À la différence de l'algorithme *Basic Local Alignment Search Tool*[171] [BLAST], le programme FASTA prend en compte la globalité de l'alignement et non seulement de petites portions de l'alignement, il est capable de mieux prendre en compte des lacunes (*gap*), ce qui permet un meilleur alignement. De ce fait, il est plus adapté pour la comparaison de séquences protéiques qui sont moins similaires. Au laboratoire, l'algorithme fasta v36[172-173] est implémenté sur une plateforme Galaxy et permet la recherche simultanée des protéines homologues pour un grand nombre de protéines contre la banque SwissProt d'une taxonomie donnée. Les annotations fonctionnelles des protéines homologues identifiées sont ensuite récupérées pour poursuivre l'interprétation des données.

V. Conclusion

L'analyse protéomique quantitative *label-free* des protéines est aujourd'hui mature et permet l'identification et la quantification de centaines/milliers de protéines simultanément. Elle comporte 3 étapes clé (la préparation d'échantillon, l'analyse LC-MS/MS et le traitement des données) qui doivent être soigneusement réfléchies et optimisées pour être adaptées aux divers type d'échantillons, à la quantité de matériel disponible et à l'objectif du projet. Cette stratégie peut aussi être adaptée pour étudier les PTMs.

Chapitre 2
Une modification majeure pour la régulation du vivant : la phosphorylation

I. Intérêt de l'étude des phosphorylations

Les PTM sont les modifications apportées aux protéines après leur biosynthèse. L'une des PTMs les plus étudiées est la phosphorylation. Elle correspond à l'ajout d'un groupement phosphate issu de la molécule d'Adénosine-TriPhosphate [ATP] sur une protéine. Cet ajout est effectué par les kinases. Cette réaction est réversible : la réaction inverse est effectuée par les phosphatases[174]. Ce processus est illustré en figure 17.

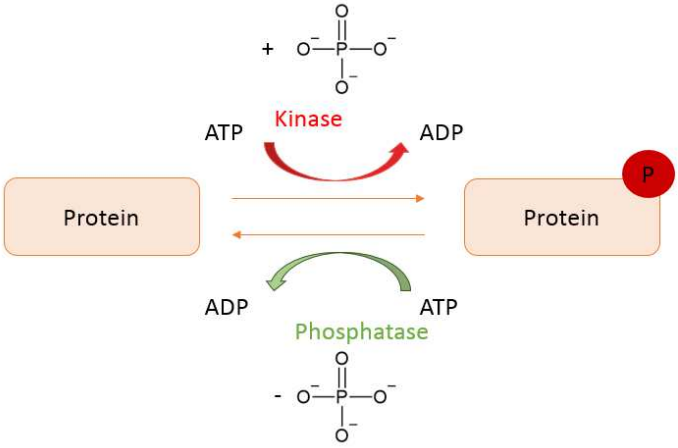

Figure 17 : Mécanismes de phosphorylation *via* les kinases et de déphosphorylation *via* les phosphatases.

La phosphorylation touche particulièrement les acides aminés portant des groupements hydroxyles libres : les sérines (86,4%), les thréonines (11,8%) et les tyrosines (1,8%)[175]. D'autres acides aminés peuvent être phosphorylés mais ils sont extrêmement minoritaires : cystéine, lysine, arginine, histidine, acides aspartique et glutamique[176]. La phosphorylation se produit sur un seul (mono-) ou plusieurs (multi-) sites et peut coexister avec d'autres types de PTM pour générer différentes protéoformes[177]. Il existerait plus de 6 millions de protéoformes différentes chez l'homme[2] (Figure 18A). Les dérégulations des protéoformes sont impliquées dans différentes maladies comme illustré sur la figure 18B.

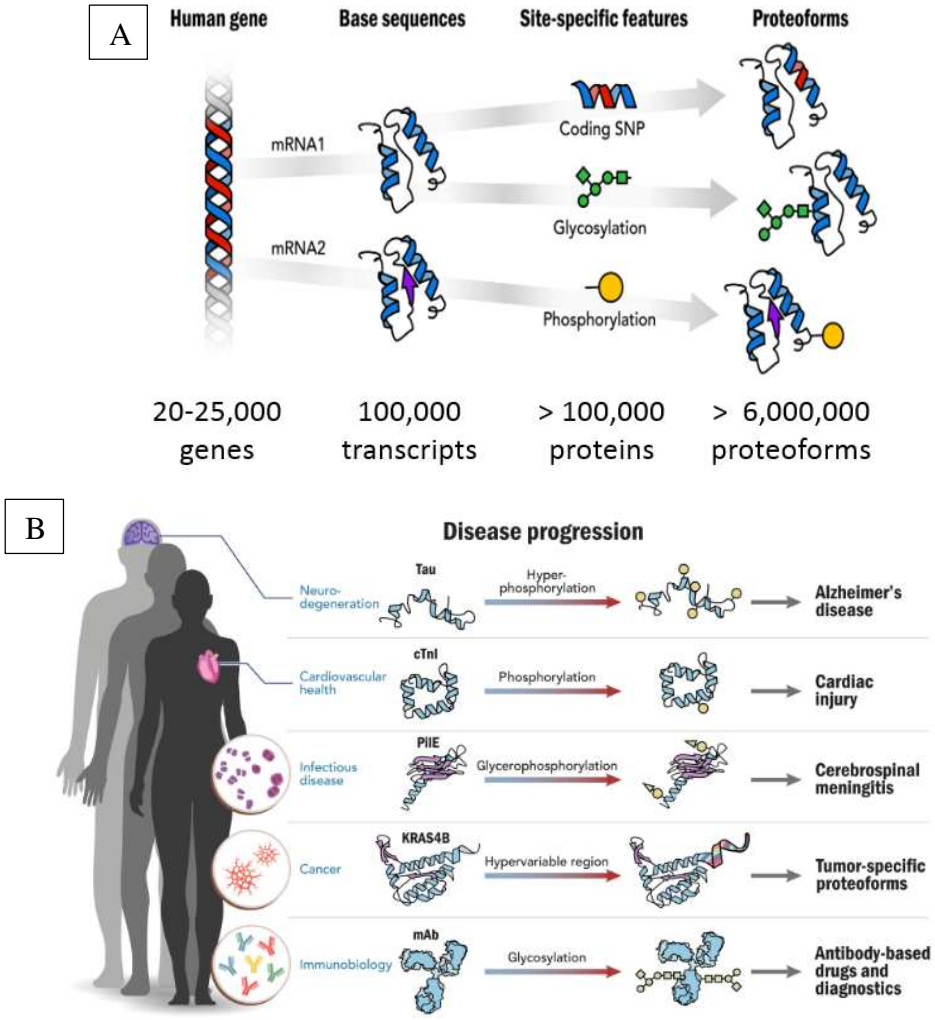

Figure 18 : A : Différentes protéoforme ont issues d'un même gène[178]. B : exemples de rôle de protéoformes dans la progression de maladies chez l'homme[178].

Chez l'homme, il existe 568 kinases et 156 phosphatases qui régulent les événements de phosphorylations[175]. Cette réaction est rapide (de l'ordre de la seconde[179]) et réversible, le coût métabolique est relativement faible et elle possède la capacité de moduler profondément la fonction de la protéine cible. L'ensemble de ces éléments a fait des phosphorylations l'une des modifications les plus importantes du vivant[179-180]. 30-75% des protéines peuvent être phosphorylées mais leur fonction précise n'est pas toujours connue[179, 181-182]. Les phosphorylations des protéines sont impliquées dans de nombreux processus cellulaires comme illustré en figure 19.

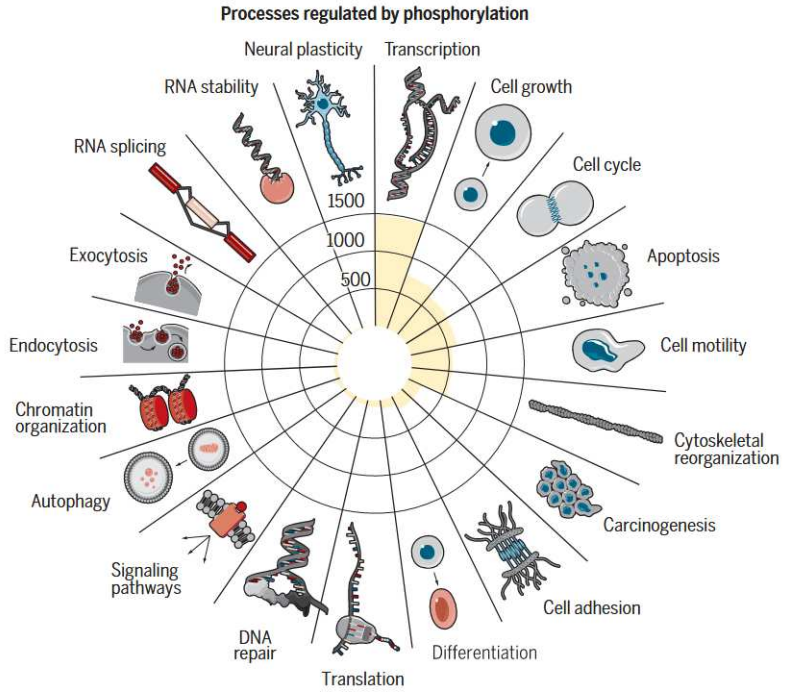

Figure 19 : Processus régulés par les phosphorylations[183].

Les dérégulations des mécanismes de phosphorylation jouent un rôle clé dans les développements du cancer[175], notamment la formation et la croissance des tumeurs, et sont aussi impliquées dans les maladies neurodégénératives telles que la maladie d'Alzheimer[112]. Par exemple, l'activité de la protéine p53, impliquée dans le cycle cellulaire et l'apoptose, est régie par des phosphorylations[175]. Il a aussi été démontré que les phosphotyrosines [pTyr] régulent les facteurs de croissance épidermiques et présentent donc un intérêt majeur pour la compréhension des mécanismes de développement et d'apparition du cancer[175]. Ces 10 dernières années, 40 médicaments basés sur la régulation de kinases ont été approuvés par la *Food and Drug Administration* [FDA][182]. La phosphorylation des protéines est aussi impliquée par exemple dans certaines pathologies sévères caractérisées par une perte de masse musculaire[184].

L'étude fonctionnelle des phosphorylations est très prometteuse pour la découverte de nouveaux traitements, c'est pourquoi, il est nécessaire de déterminer les sites, l'abondance et le rôle des phosphorylations pour mieux comprendre les signaux cellulaires et leur dérégulation. De par son caractère non ciblé et le haut débit d'analyse qu'elle propose, la spectrométrie de masse a pris une place importante dans l'analyse des phosphorylations[179, 185-187].

II. L'étude des phosphorylations par spectrométrie de masse rencontre de nombreux défis analytiques

La spectrométrie de masse s'est imposée comme une technique de choix pour l'étude des phosphorylations permettant l'identification, la localisation et la quantification d'un grand nombre de phosphorylations[21, 186-188]. Pourtant, l'analyse de la phosphorylation des protéines

pose de nombreux défis techniques à chaque étape de l'analyse protéomique comme indiqué en figure 20[15, 179].

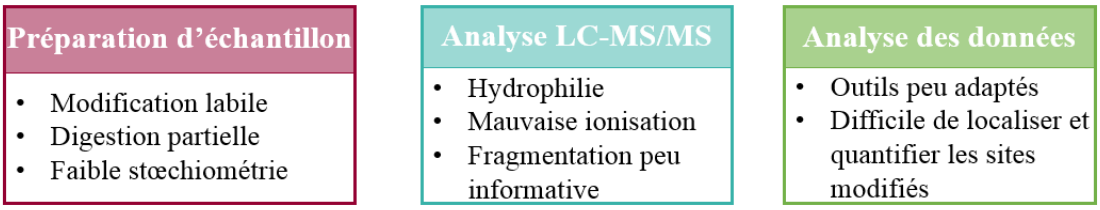

Figure 20 : Difficultés rencontrées pour l'étude des phosphorylations au cours des trois grandes étapes de l'analyse protéomique : la préparation d'échantillons, l'analyse LC-MS/MS et le traitement des données.

A. La préparation d'échantillons

Les phosphorylations sont des modifications labiles, c'est pourquoi, les étapes de préparation sont, dans la mesure du possible, réalisées à froid[176]. Le statut de phosphorylation initial doit être conservé, par exemple en ajoutant des inhibiteurs de phosphatases durant la lyse des cellules.[176] Les phosphorylations peuvent aussi impacter la digestion par la trypsine, étape clé de l'analyse protéomique, par la formation de liaisons hydrogène et de pont salins entre les groupements phosphates et les sites de clivage K/R[179]. Une digestion multi-enzymatique permet d'augmenter l'efficacité de la digestion[189].

Les peptides phosphorylés représentent en moyenne moins de 3 % des peptides d'un échantillon[182]. Cette faible stœchiométrie compromet leur détection lors de l'analyse LC-MS/MS[190] car on peut rencontrer des suppressions de signal MS si les peptides sont co-élués. De même, les ions minoritaires ne seront pas fragmentés puisque, en mode DDA, seuls les N ions les plus intenses sont sélectionnés pour la fragmentation. C'est pourquoi, une étape d'enrichissement en protéines[191]/peptides phosphorylés est ajoutée au protocole classique de protéomique. L'enrichissement au niveau peptidique est quasi systématiquement préféré[190]. Plusieurs techniques existent pour enrichir un échantillon en peptides phosphorylés, les 3 principales stratégies utilisées sont représentées en figure 21.

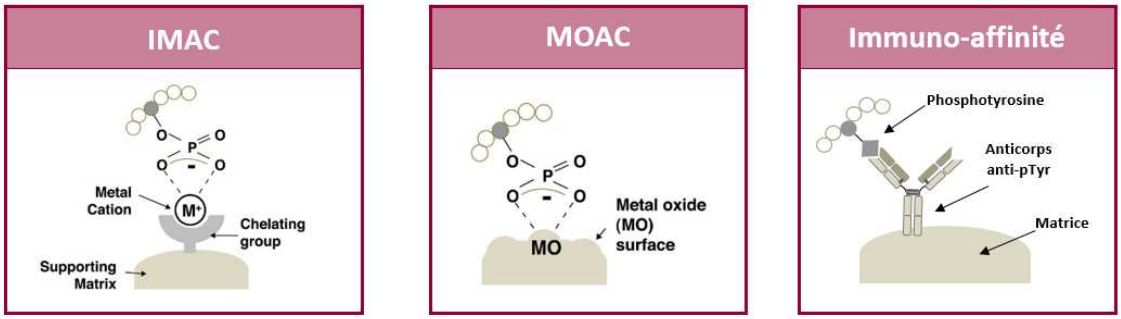

Figure 21 : Principe des trois principales stratégies pour l'enrichissement en phosphopeptides : interactions électrostatiques entre un ion métal immobilisé et le groupement phosphate (IMAC), affinité de l'oxygène de groupement phosphate pour des oxydes de métal (MOAC) et enrichissement par immuno-affinité utilisant des anticorps anti-pTyr (d'après Low *et al.* 2021)[176].

Les phases *Immobilizied Metal Affinity Chromatography* [IMAC] sont composées d'un métal de transition (Fe^{3+}, Ga^{3+}, Zr^{4+} [192], Ti^{4+}[193]) lié à une phase solide. Les deux phases les plus

utilisées sont la nitrilotriacetique [NTA] ou l'iminodiacetique [IDA][179, 190]. La phase IMAC la plus largement répandue utilise les ions ferreux[174]. Le groupement phosphate de la phosphorylation est chargé négativement et va interagir par interactions électrostatiques avec l'ion métal[176]. Ainsi les peptides phosphorylés seront retenus par la phase en condition acide et élués en condition basique[177].

Les phases *Metal Oxide Affinity Chromatography* [MOAC] utilisent l'affinité de l'oxygène du phosphate pour le métal de l'oxyde de métal (TiO_2, ZrO_2 et Fe_3O_4)[189]. Les phosphopeptides sont déposés à pH acide, souvent avec un additif (glycérol, acide lactique, acide citrique…) puis élués en augmentant le pH au-dessus de 10[182]. Cet additif permet d'augmenter l'efficacité de l'enrichissement MOAC. La phase MOAC permet une meilleure rétention des phosphotyrosines que la phase IMAC[179].

Il a été montré que ces deux techniques sont complémentaires : à partir de cellules HeLa, seuls 42% de phosphopeptides communs ont été observés entre les enrichissements réalisés sur les phases MOAC et IMAC[189]. La phase IMAC est particulièrement efficace sur les phosphosérines et sur les peptides portant plusieurs phosphorylations[179]. Elle est moins sélective que la phase MOAC car elle retient également les acides aminés très acides[194]. Il n'existe pas à ce jour de consensus sur la phase optimale pour l'analyse des phosphorylations.

Différentes stratégies existent pour augmenter l'efficacité de ces enrichissements. C'est ainsi que le *Sequential elution from IMAC* (SIMAC) a vu le jour. Trois fractions sont éluées de la phase IMAC : la fraction non retenue lors du chargement de l'échantillon, une fraction éluée à pH acide et une fraction éluée à pH basique. Les fractions non retenues et acides peuvent ensuite être regroupées et une étape additionnelle d'enrichissement sur une phase MOAC peut être réalisée. La fraction basique contient des peptides multi-phosphorylés tandis que les fractions enrichies par MOAC contiennent plutôt des peptides mono-phosphorylés. Une autre stratégie consiste à décompléxifier les échantillons, soit avant soit après l'enrichissement, par fractionnement HILIC[21], ERLIC[195], SCX[196] ou par phase RP à pH basique[108].

Les pTyr sont les phosphorylations avec la plus faible occurrence mais elles ont un intérêt thérapeutique particulier dans la lutte contre différentes maladies, notamment les cancers. C'est pourquoi, l'enrichissement par immuno-affinité, *via* des anticorps spécifiques, est utilisé préférentiellement pour caractériser/étudier ce type de phosphorylations. Un enrichissement IMAC préalable à l'enrichissement par immuno-affinité a également été décrit sur différentes cellules (HCT116 et PC3)[179].

Ces différentes phases (IMAC, MOAC et anticorps anti-pTyr) peuvent être utilisées sous différentes formes : colonne *High Pressure Liquid Chromatography* [HPLC][197], cartouches[108, 198], billes magnétiques[182, 199]. L'enrichissement en phosphopeptides exige de grandes quantité de matériel de départ (souvent > 1 mg de peptides) car seule une petite fraction de l'échantillon d'origine est conservée. Cette contrainte est particulièrement limitante pour l'étude des phosphorylations dans les tissus où la taille de l'échantillon est souvent restreinte[174].

Pour augmenter le débit et la répétabilité des analyses phosphoprotéomiques, des solutions automatisées pour l'enrichissement en phosphopeptides ont été développées ces dernières années. Un premier pas vers une intégration plus facile de l'enrichissement dans le protocole

protéomique a été le développement de la préparation « EasyPhos ». Ce protocole permet la digestion et l'enrichissement sur billes magnétiques fonctionnalisées d'1 mg de protéines dans un même contenant. Il peut être réalisé dans des plaques 96 puits facilitant son implémentation sur des systèmes automatisés[181]. La société Agilent Technologies a ensuite mis sur le marché des protocoles et des cartouches TiO_2 ou NTA-Fe^{3+} afin d'enrichir des échantillons en phosphopeptides avec le Bravo AssayMAP[108, 198]. Ce protocole permet également de diminuer la quantité de matériel de départ (100 µg pour des cellules HeLa)[200].

B. Analyse par spectrométrie de masse des peptides phosphorylés

Deux facteurs rendent l'analyse des phosphopeptides par spectrométrie de masse difficile : la plus faible ionisation par rapport aux peptides non phosphorylés et la perte du groupement phosphate labile lors de la fragmentation produisant des spectres peu informatifs. Il a en effet été montré que l'ionisation des peptides est impactée par la phosphorylation[201] car les groupements phosphates ont tendance à perdre des protons pour porter des charges négatives[177]. Par ailleurs, le groupement phosphate est labile et peut être perdu lors de la fragmentation. C'est notamment le cas avec l'une des technique la plus utilisée en protéomique : la fragmentation CID[179]. Cette fragmentation entraine des pertes de neutre suivant l'un des mécanismes décrit en figure 22. Deux pertes de neutre caractéristiques sont possibles, soit la perte du groupement phosphate (H_3PO_4, - 98 Da) pour former un cycle à cinq atomes, soit la perte de l'acide métaphosphorique (HPO_3, - 98 Da)[202].

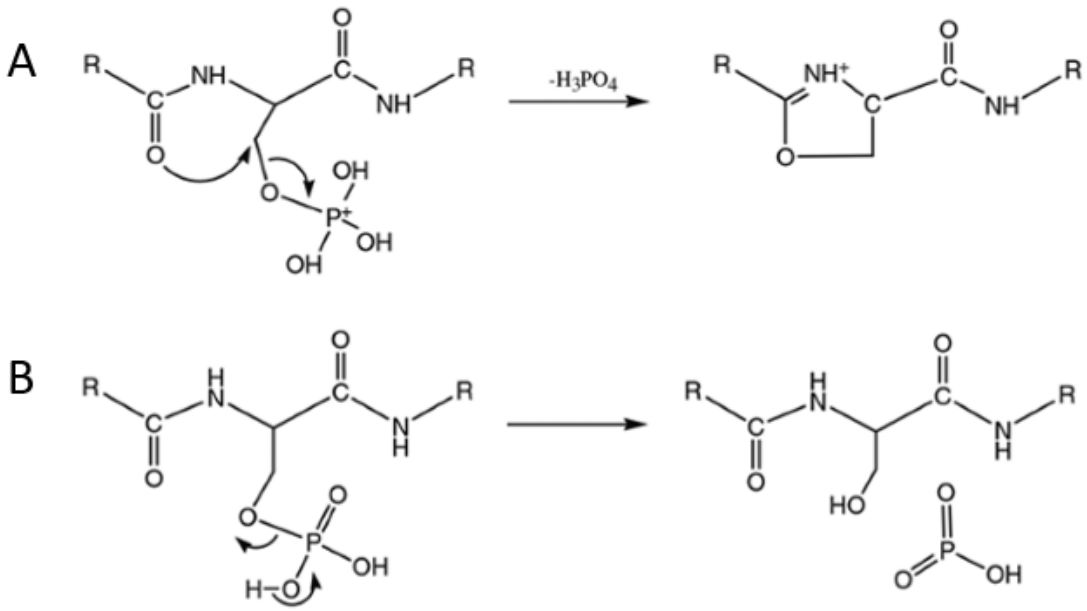

Figure 22 : A : Mécanisme de la perte de neutre H_3PO_4 des phosphopeptides. D'autres mécanismes SN2 ont été proposés, en fonction de la nature du groupe d'attaque nucléophile. Ici, la formation énergétiquement favorisée de l'ion oxazoline avec un cycle à cinq atomes est présentée. B : Mécanisme de perte de neutre HPO_3[203].

Par rapport à la fragmentation CID, la fragmentation HCD fournit plus d'énergie lors de la collision, et il en résulte des spectres plus informatifs et une moindre perte de neutres[182]. Des

méthodes de fragmentations alternatives sont également utilisées pour l'étude des phosphorylations comme l'ETD[133], l'EThcD[137] ou l'UVPD[204].

L'ETD génère des ions de type c et z, contrairement à la fragmentation CID ou HCD qui génèrent des ions y et b. L'ETD souffre d'un temps de réaction long (entre 1 et 100 ms contre 0,1 ms pour la fragmentation CID[202, 204]) entre l'anion portant l'électron et le peptide cation[204]. Elle est plus efficace sur les ions très chargés (charge supérieure à +2)[202]. C'est pourquoi, bien que les spectres obtenus soient très informatifs par rapport à la fragmentation CID, comme illustré en figure 23, cette technique ne permet pas l'identification d'un grand nombre de phosphopeptides en comparaison à la fragmentation CID, et les deux méthodes de fragmentation sont complémentaires[133].

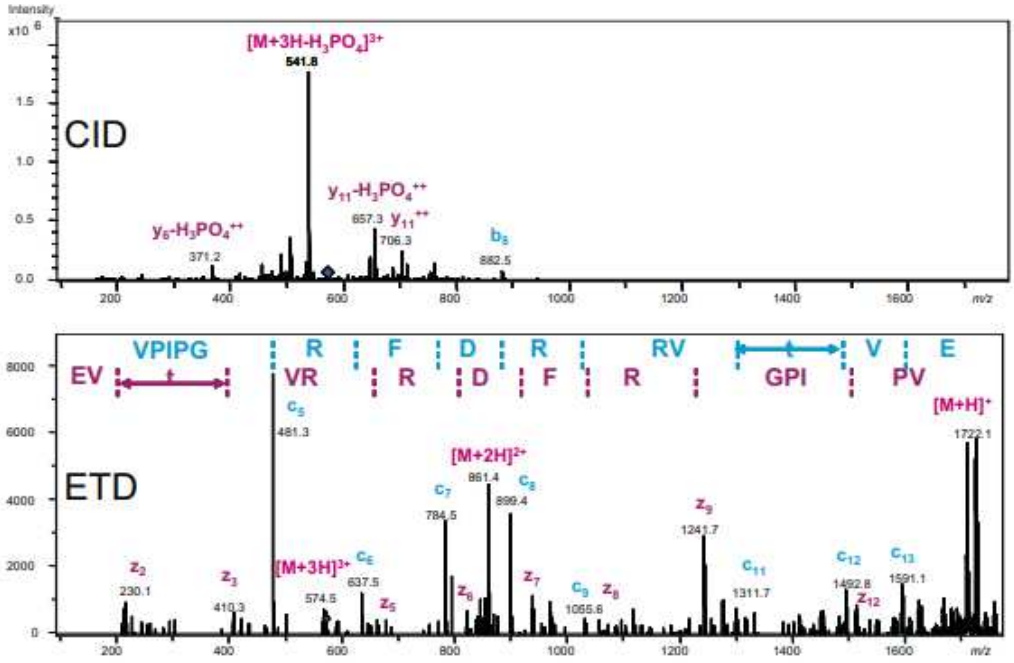

Figure 23 : Spectres MS/MS du phosphopeptide VPIPGRFDRRVtVE par CID et par ETD. Le pic le plus intense sur le spectre CID correspond à la perte de groupe phosphate (- 98 Da). Sur le spectre ETD, les différences de masse entre c11 et c12, z2 et z3 sont de 181 Da, ce qui indique la localisation de la phosphorylation sur la thréonine[205].

L'EThcD cherche à coupler l'efficacité de la fragmentation ETD avec la vitesse de la fragmentation HCD. Les peptides sont fragmentés deux fois et des spectres hybrides contenant à la fois les ions –y et –b de la fragmentation HCD et les ions –z et –c de la fragmentation ETD sont obtenus, comme illustré en figure 24[137].

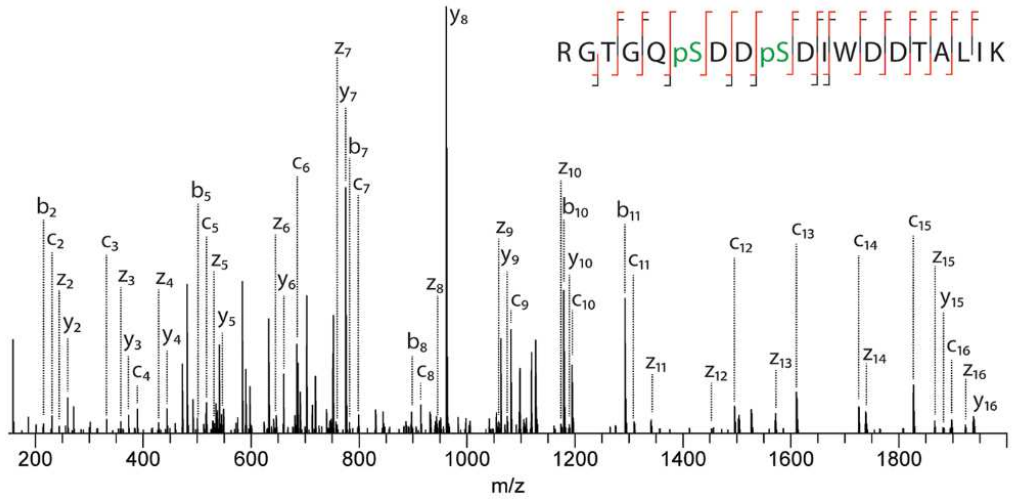

Figure 24: Spectre MS/MS EThcD d'un peptide doublement phosphorylé. RGTGQ**S**DD**S**DIWDDTALIK est doublement phosphorylé et contient au total quatre sites de phosphorylation potentiels (probabilités de sites phosphoRS : T(3), 0,0% ; S(6), 100,0% ; S(9), 100,0% ; T(15), 0,0%)[137].

Dans une étude de Freese *et al.,* ce mode de fragmentation a permis l'identification de 3679 PSM, sur des lysats de cellules HeLa enrichis en phosphopeptides par IMAC, contre 4282 avec une fragmentation HCD[137]. **95**% (3942 PSMs) des sites de phosphorylation ont été identifiés avec une probabilité de localisation supérieure ou égale à 99% contre seulement **89**% des sites (4291 PSMs) en fragmentation HCD[137]. Ainsi, il a été montré que bien qu'identifiant moins de phosphopeptides que la fragmentation HCD, l'EThcD améliore la couverture de la séquence peptidique et le pourcentage de phosphosites localisés par rapport aux fragmentations HCD et ETD[189].

L'UVPD délivre un flux élevé de protons à l'aide d'un laser pulsé durant un court laps de temps[135] (< 1 µs[204]). Le nombre d'impulsions, l'énergie du laser et la longueur d'onde sont des paramètres clés qui modulent l'énergie délivrée, affectent l'efficacité de fragmentation et joue sur les types et la distribution des fragments obtenus. Les mécanismes détaillés de cette fragmentation ne sont pas encore entièrement élucidés[135]. Cette technique génère des spectres complexes avec des ions a, b, c, x, y et z, comme illustré en figure 25[204].

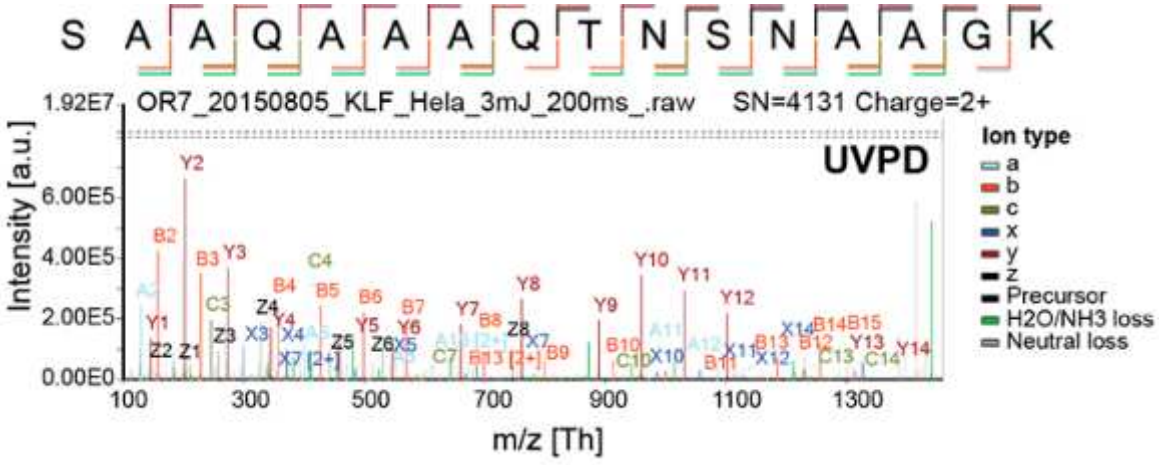

Figure 25 : Spectre MS/MS UVPD d'un peptide portant deux phosphorylations, la quasi-totalité de la séquence est couverte par les ions de type a, b, c, x, y et z[204].

C. Stratégies de quantification des peptides phosphorylés

Deux principales stratégies de quantification des peptides phosphorylés coexistent : la stratégie « *label-free* » sans marquage[108, 188] et par marquage isotopique avec les réactifs *Tandem Mass Tag* [TMT][206-207].

Dans la stratégie « *label-free* », les peptides sont identifiés à l'aide des spectres de fragmentation puis les courants d'ion associés en MS1 sont extraits pour intégrer les pics et ainsi mesurer l'aire sous la courbe correspondante. En protéomique classique, en général, une protéine est identifiée avec plusieurs peptides, les intensités issues des différents peptides sont alors sommées pour obtenir l'intensité de la protéine. Pour la phosphoprotéomique et l'étude des peptides modifiés en général, la quantification doit se faire au niveau peptidique, ce qui signifie qu'un seul peptide est utilisé pour la quantification d'un événement de phosphorylation. Sur une même protéine, les valeurs quantitatives des différents phosphopeptides peuvent être complètement différentes[189]. Il est donc nécessaire d'identifier tous les peptides dans chaque analyse et ce malgré la stochasticité de l'analyse DDA, la labilité de la modification et la reproductibilité parfois limitée de l'enrichissement. De fait, la quantification des peptides phosphorylé est plus difficile que la quantification au niveau protéique, comme illustré en figure 26.

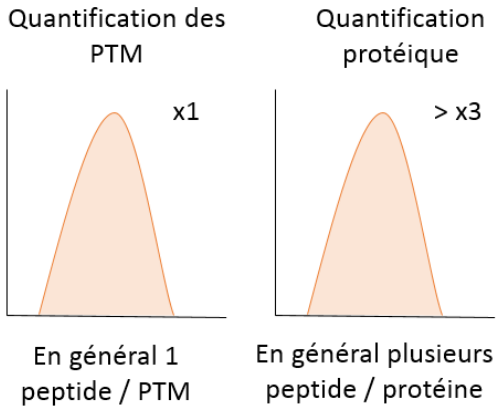

Figure 26 : Contrainte de la quantification des PTMs : pour identifier et quantifier une modification, un seul peptide peut être utilisé et doit donc être détecté dans chaque échantillon. Au contraire pour la quantification protéique, plusieurs peptides sont utilisés pour l'identification et la quantification, souvent plus de trois peptides sont utilisés, ce qui laisse plus de chance pour identifier la protéine.

Pour contourner ces problèmes, la quantification par marquage isobarique TMT est très utilisée aujourd'hui pour les études de phosphoprotéomique[19, 111, 207]. Cette stratégie est plus coûteuse que la stratégie « *label-free* » mais diminue le temps d'analyse sur l'instrument car les échantillons sont multiplexés. Cette stratégie est aussi moins flexible sur le nombre d'échantillons et, à ce jour, il existe différents kits de TMT pour analyser jusqu'à 18 échantillons[182]. La stratégie de marquage isobarique TMT est illustrée en figure 27.

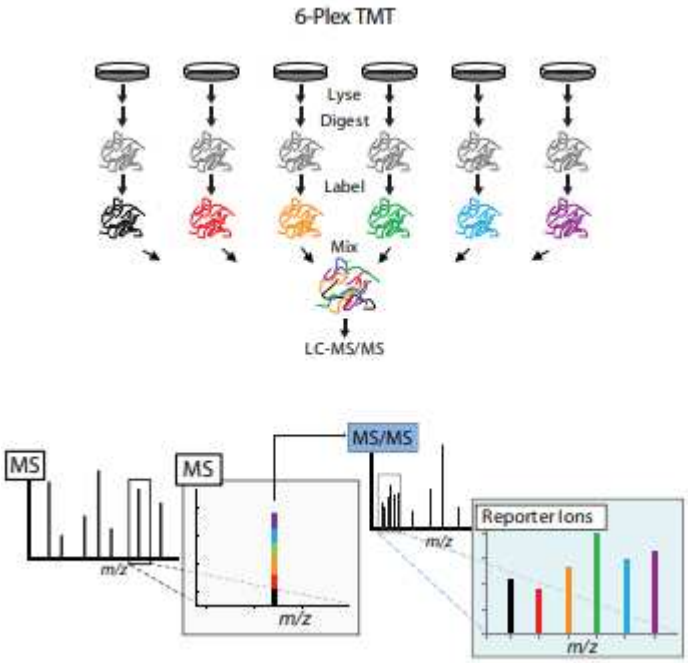

Figure 27 : Les marqueurs isobares, tels que le TMT 6-plex représenté, permettent une analyse quantitative multiplexée. Ils sont incorporés chimiquement après la digestion des peptides en se liant à l'extrémité N-terminale des peptides ou à une lysine, avant le mélange et l'analyse. Les différents marqueurs dans chaque ensemble de marqueurs isobares ont des masses identiques, et donc dans le scan complet, chaque pic est en fait un composite de peptides de chaque échantillon. Cependant, lors de la fragmentation, chaque marqueur libère un ion rapporteur unique qui peut être détecté dans un balayage MS/MS. Les hauteurs des pics fournissent une quantification relative des six échantillons[208].

Les échantillons sont marqués de façon isobarique avant[209] ou après[210] l'enrichissement en phosphopeptides. Réaliser le marquage avant l'enrichissement est très couteux car plus de 60% des peptides seront éliminés durant l'enrichissement mais il est plus utilisé car il réduit la variabilité induite par la préparation d'échantillon[182]. Puis les différents échantillons sont regroupés et analysés par LC-MS/MS. Les différents marqueurs d'un ensemble ont des masses identiques, et donc, en MS1, chaque pic est issu d'un même peptide extrait des différents échantillons. Cependant, lors de la fragmentation, chaque marqueur libère un ion rapporteur unique qui peut être détecté lors d'un balayage MS/MS. Les hauteurs des pics fournissent une quantification relative des différents échantillons. L'identification des peptides se fait sur les spectres de fragmentation et la quantification sur les ions reporters également présents sur le spectre MS2[208].

Une comparaison des deux stratégies (TMT et *label-free*) a été conduite sur des lysats de cellules DiFi enrichis en phosphopeptides par MOAC. Cette étude, réalisée par Stepath *et al.,* a montré que bien que la stratégie *label-free* permettait l'identification d'un plus grand nombre de phosphopeptides, la quantification des phosphopeptides était limitée par rapport à la stratégie TMT[211]. Cette conclusion est partagée par une autre étude réalisée sur des cellules SKOV3 et des tissus issus d'une tumeur ovarienne[209].

Une autre approche permettant d'augmenter le nombre d'événements phosphorylés quantifiés sans recourir à un marquage isobarique couteux, est l'analyse des phosphopeptides en mode *Data Independant Acquisition* [DIA]. Dans ce mode d'acquisition, ce ne sont pas les N ions les plus intenses qui sont fragmentés mais tous les ions précurseurs sur une plage de masse (fenêtre

d'isolation m/z) donnée[212] (Figure 28). La technique DIA s'affranchie donc du caractère stochastique et du problème de sous-échantillonnage des peptides inhérents au mode DDA[212]. Les données obtenues par ce mode d'acquisition sont plus complexes à analyser car les spectres sont hautement multiplexés et la masse du précurseur n'est pas connue[213].

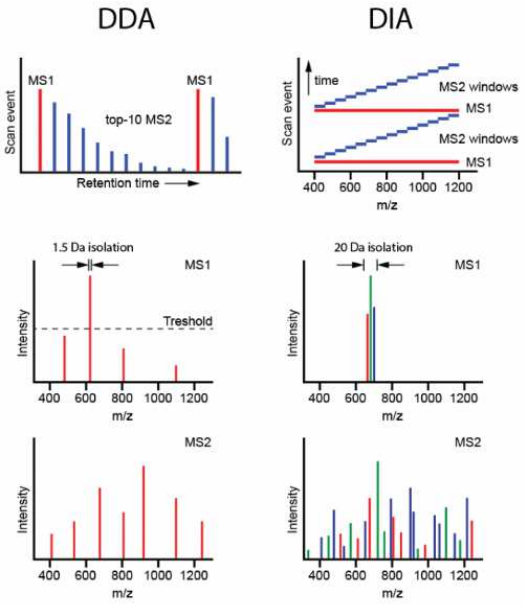

Figure 28 : Principe du mode DDA et du mode DIA. En DDA, après un scan MS1, les N (ici 10) ions les plus intenses sont fragmentés puis analysés. En DIA, après un scan MS1, tous les ions d'une fenêtre d'isolation donnée sont co-fragmentés puis analysés[214].

En mode DIA, la quantification est effectuée sur les ions fragments, ce qui est particulièrement intéressant pour l'étude des peptides phosphorylés isobares. En effet, ces peptides sont souvent co-élués lors de la séparation chromatographique car leurs propriétés chimiques sont très semblables. En quantifiant les fragments permettant l'identification précise des peptides isobares, la DIA permet une quantification plus robuste de ces peptides[213] : en comparant les approches DDA et DIA sur un jeu de données issus de de cellules humaines U2OS dopées avec des phosphopeptides synthétiques, Srinivisan *et al.* ont montré que la DIA permet une identification plus robuste des phosphopeptides (66% des phosphopeptides sont identifiés dans au moins 5/10 réplicas contre 32% en DDA).

Pour améliorer encore l'identification et la quantification des peptides phosphorylés co-élués, une dimension de séparation supplémentaire peut être ajoutée : la mobilité ionique. Il existe plusieurs technologies pour réaliser cette séparation : le TIMS[215-216] et le FAIMS[13, 217] sont les plus utilisés. En utilisant la technologie FAIMS, Zhao *et al.* ont augmenté le nombre de peptides phosphorylés détectés de 37% par rapport à des analyses LC-MS/MS classiques[218]. En couplant la DIA et la cellule TIMS, Skowronek *et al.* ont identifié 35 000 sites de phosphorylations en trois injections à partir de 100 µg de cellules HeLa stimulées par des facteurs de croissance épidermiques[215]. Un algorithme, TIMScore, a été développé pour prédire les valeurs CCS des peptides trypsiques et phosphorylés. Il permet d'augmenter le nombre moyen de phosphopeptides identifiés entre 10 et 25% par rapport à des analyses sans son utilisation[219].

Outre l'abondance relative des peptides phosphorylés dans les différents échantillons analysés, une autre donnée est particulièrement importante pour l'étude de la phosphorylation : la stœchiométrie des peptides phosphorylés par rapport aux peptides non modifiés. C'est une donnée difficile à évaluer mais qui est cruciale pour l'interprétation biologique car elle peut permettre de distinguer les événements de phosphorylation pertinents[15]. La stœchiométrie peut être mesurée par déphosphorylation enzymatique et marquage isobarique (TMT), puis comparaison des résultats avec et sans déphosphorylation[15]. Mais elle reste compliquée à évaluer car la gamme dynamique et la sensibilité du spectromètre de masse limitent l'analyse.

III. Outils bio-informatiques pour l'étude des phosphorylations

Pour l'analyse des phosphorylations et des PTMs en général, une étape s'ajoute au traitement des données par rapport à l'analyse protéomique classique. La localisation des modifications doit être déterminée et les peptides modifiés doivent être validés. De plus, des outils spécifiques à l'interprétation biologique des variations d'abondance des peptides phosphorylés peuvent être utilisés. Un schéma général des étapes du traitement des données pour des analyses de phosphoprotéomique est illustré en figure 29.

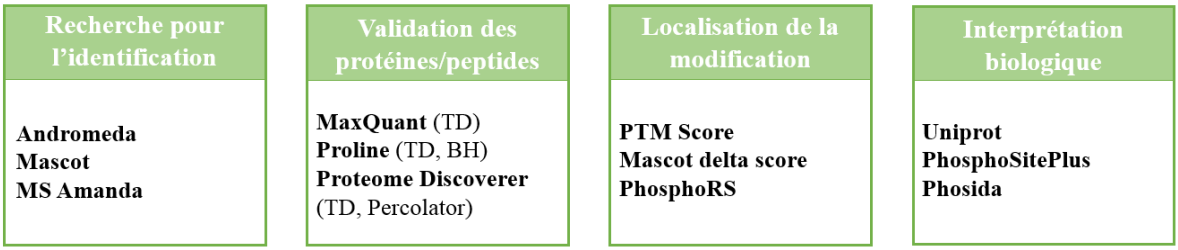

Figure 29 : Vue d'ensemble des étapes de traitement de données pour des analyses de phosphoprotéomique.

A. Identification et localisation des phosphorylations

Pour l'identification des peptides modifiés, la plupart des algorithmes de recherche (Andromeda[147], Mascot, MS Amanda[148]) permettent la recherche de modifications variables grâce à l'addition d'un delta de masse, cependant, si un trop grand nombre de modifications est recherché, le temps de recherche augmente exponentiellement. C'est pourquoi de nouveaux moteurs de recherche, qui permettent des recherches ouvertes à toutes les modifications, voient le jour comme MSFragger[153] ou ionbot[182]. Ces logiciels permettent d'étudier les phosphorylations et un phénomène appelé cross-talk, par lequel différents sites de PTM coopèrent les uns avec les autres pour remplir une fonction[13].

La localisation exacte du site de phosphorylation sur le peptide est cruciale afin de caractériser sa fonction[189]. Or 20 à 40% des phosphopeptides identifiés sont perdus car la localisation de la modification est trop ambiguë pour statuer sur la position exacte[189]. Une fois la phosphorylation et le peptide qui la porte identifiés, un score calculant la probabilité de présence de la modification doit être calculé. Différents algorithmes ont été développés pour évaluer cette probabilité comme Mascot delta score[220], PTM score[147] ou PhosphoRS[221]. Ils sont basés sur l'une des deux approches suivantes : des algorithmes calculant la probabilité d'un match

incorrect pour chaque isobare (phosphorylation placée sur les différents acides aminés phosphorylables du peptide) des phosphopeptides (PTM Score ou PhosphoRS) ou calculant la différence de score entre les différents phosphopeptides possibles (Mascot delta score)[202]. Plusieurs facteurs peuvent expliquer les différents résultats obtenus avec ces algorithmes dans des études comparatives [222-223] : (i) le type de fragment pris en compte, (ii) le seuil de détection d'un pic, (iii) l'utilisation ou non des pertes neutres du groupement phosphate, (iv) le fait que l'algorithme ait été conçu pour des mesures de résolution et de précision de masse basses ou élevées. La prise en compte des pertes de neutres est un paramètre par défaut et non modifiable avec PTM Score. Au contraire, c'est un paramètre ajustable par l'utilisateur avec PhosphoRS. Pour le calcul du PTM Score (Andromeda), le seuil de détection d'un pic est uniforme pour toutes les fenêtres m/z, alors qu'elle est déterminée individuellement pour chaque fenêtre de 100 m/z avec PhosphoRS. Le schéma général des deux algorithmes les plus utilisés, PhosphoRS et PTM Score, est présenté en figure 30.

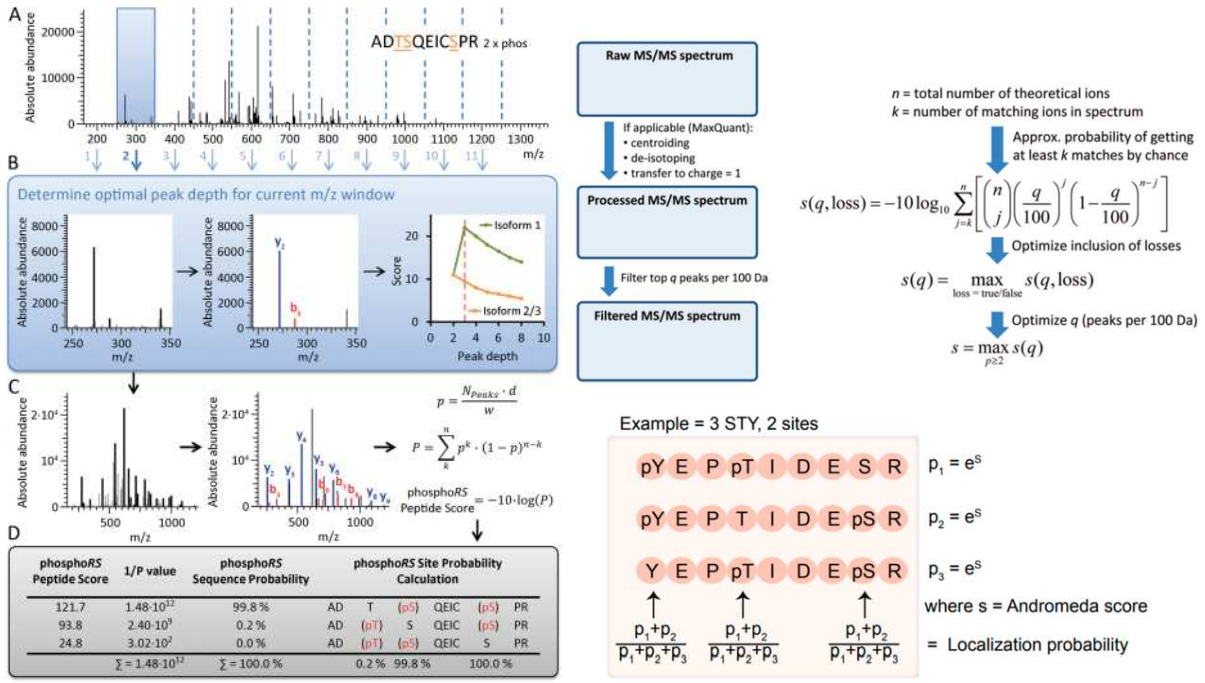

Figure 30 : Vue d'ensemble des étapes de traitement effectuées par phosphoRS (à gauche) et par PTM score (à droite)[202, 221].

En comparant différents algorithmes d'identification et de « *scoring* », Locard *et al.* ont montré que seuls 50% des phosphopeptides étaient identifiés à la fois par un traitement Andromeda-PTM Score et par Mascot-PhosphoRS. La source la plus importante des différences était dûe à l'identification des peptides, bien que le « *scoring* » joue également un rôle. Ainsi, Mascot delta score s'est révélé assez peu performant par rapport à PTM score et à PhosphoRS[222].

B. Validation et estimation du *False Localization Rate* [FLR]

Dans la plupart des études, la validation des phosphopeptides est réalisée en appliquant un filtre arbitraire sur la valeur de la probabilité de localisation. La valeur seuil est communément placée à 75% définissant ainsi les phosphopeptides qualifiés de « classe I »[202]. Il existe également une notion similaire au FDR, appelée *False Localization Rate* [FLR], elle permet d'estimer le

nombre de mauvaises localisations des modifications. Malheureusement, à ce jour, il est impossible d'obtenir des leurres qui permettent d'assurer que la modification identifiée est fausse. Il faut néanmoins souligner certaines tentatives, comme celle de Ramsbottom *et al*. qui ont exploré la possibilité d'utiliser des acides aminés leurres (par exemple l'alanine) pour l'estimation des statistqiues liées à la localisation des sites (le FLR)[224] ou le logiciel LuciPhor crée par Fermin *et al.* qui créé des leurres en rajoutant des phosphorylations sur tous les acides aminés non phosphorylables dans les peptides identifiés[225]. Cette dernière stratégie est jugée trop conservatrice car le nombre de cibles est beaucoup plus restreint que le nombre de leurres[202]. Ces approches ne font pas l'unanimité au sein de la communauté scientifique et pour l'instant, l'évaluation du FLR la plus acceptée est effectuée sur un jeu de données de peptides synthétiques où les localisations des phosphorylations sont connues[202]. Il est toutefois important de noter que de tels jeux de données ne reflètent pas la complexité de vrais échantillons biologiques[202].

C. Interprétation biologique

Pour l'interprétation des résultats obtenus après identification, localisation, validation et quantification des phosphopeptides, des solutions informatiques utilisant les langages de programmation python (PaDua)[226] ou R (PhosR)[227] ont été développées mais restent assez confidentielles. Scop3P est une ressource qui permet de visualiser la phosphorylation sur la structure tridimensionnelle de la protéine[228-229].

L'interprétation biologique de données de phosphoprotéomique passe par la consultation de bases de données répertoriant les sites de phosphorylations connus, les informations relatives aux kinases et phosphatases associées ainsi que la fonction affectée par la phosphorylation d'un acide aminé donné. PhosphositePlus[230] s'est imposé au cours du temps, regroupant près de 300 000 sites de phosphorylations décrits, mais d'autres ressources existent telles que Phosida[231] ou UniprotKB. Comme toutes les banques de données, elles ne sont pas sans défaut et contiennent certainement des redondances et des erreurs. Par ailleurs, elles sont très lacunaires puisque 97% des sites répertoriés n'ont pas de kinase connue pour les réguler et ne concerne que des espèces modèles telles que la souris ou l'homme[182].

IV. Conclusion

La spectrométrie de masse est un outil très performant pour l'analyse du phosphoprotéome. Elle permet de décrire les voies de communications intra- et extra-cellulaires dans lesquels les phosphorylations jouent un rôle crucial. Cependant, chaque étape de l'analyse phosphoprotéomique doit être réalisée avec le plus grand soin afin de réduire au maximum la variabilité. Ainsi, les études phosphoprotéomiques nécessitent un enrichissement hautement spécifique, une détection sensible et une localisation précise des sites de phosphorylation grâce à des techniques LC-MS/MS avancées et un traitement de données robuste.

Chapitre 3
Une modification non-enzymatique aux multiples facettes : la glycation

I. La réaction de glycation ou réaction de Maillard

La réaction de glycation est une réaction non enzymatique, au cours de laquelle un sucre réagit avec une fonction amine portée majoritairement par une lysine, ou plus rarement par une arginine ou l'extrémité N-terminale de la protéine[232]. Les glycations sont issues d'une réaction **spontanée** qui peut se produire indépendamment de tout contexte biologique. Ce processus, aussi appelé réaction de Maillard, a été découvert au début du XXe siècle en observant la formation de composés jaunes/bruns lorsque des acides aminés étaient chauffés en présence de différents sucres[233]. La glycation ne doit pas être confondue avec la glycosylation qui est une réaction **enzymatique**, *via* des glycosyl-transferases, au cours de laquelle des oligosaccharides (polymères formés de monosaccharides) sont ajoutés à certains acides aminés : asparagine (N-glycosylations) ou sérines et thréonines (O-glycosylations)[234-235] . Les glycosylations jouent un rôle sur la structure quaternaire, le contrôle de la qualité, la stabilité, le transport et la fonction des protéines[235].

La réaction de Maillard est illustrée en figure 31.

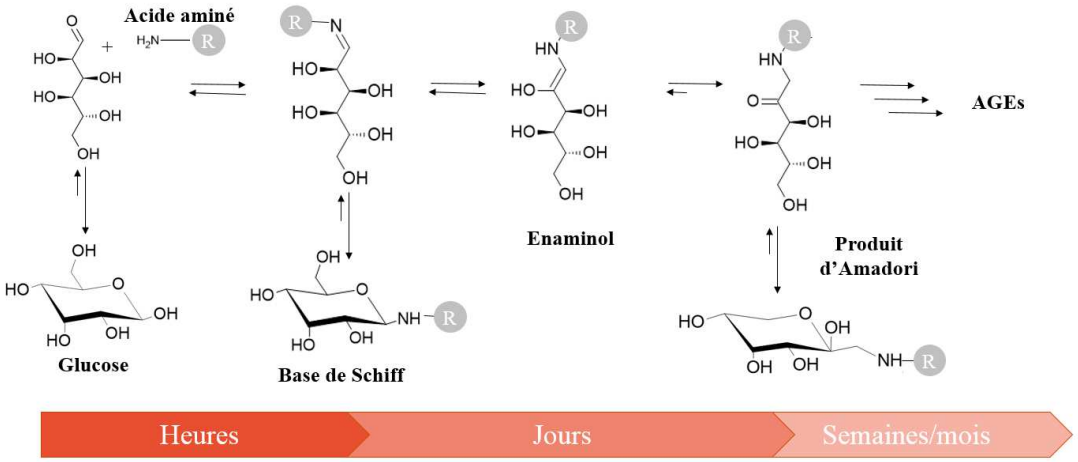

Figure 31 : Réaction entre un sucre et une amine primaire portée par un acide aminé pour former le produit d'Amadori (d'après Priego Capote *et al.* 2009 et Kim *et al.* 2017)[234, 236].

La réaction de glycation est initiée par une attaque nucléophile d'une amine primaire par un sucre réducteur (glucose, fructose, ribose…)[234]. Cette réaction réversible rapide forme une base de Schiff qui se réarrange *via* un composé énaminol pour former une cétoamine plus stable appelée produit d'Amadori. Ce produit porte plusieurs noms : (N-1-deoxy-D-fructos-1-syl)-lysine (fructosyl-lysine ou fructoselysine) lorsque l'amine est portée par une lysine ou (N-1-

deoxy-D-froctos-1-syl)acide aminé (**fructosamine**) dans le cas plus général[237]. Ce sont les produits précoces de glycation[237]. Ces composés vont ensuite subir une série de réarrangements (déshydratation et fragmentation) pour former des composés aux structures diverses, nommés produits avancés de glycations (*Advanced Glycation End-products* [AGEs]) telles que la carboxymethyllysine [CML] et la carboxyethyllysine [CEL]. De plus, ces réactions vont également générer différents composés carbonylés, généralement plus réactifs que le carbohydrate initial : le méthylglyoxal, le glyoxal, et les glucosones par exemple[234, 238]. Ces dérivés carbonylés peuvent réagir directement avec les lysines et les arginines des protéines et ainsi former des espèces telles que 5-methyl-4-oxo-5-hydroimifazo-linone-2-ylornithine [MG-H1], l'argpyrimidine ou la pyrraline[238]. Ces composés étant souvent très réactifs, il n'est pas rare que des réactions de pontage aient lieu entre deux protéines ou au sein de la protéine elle-même : *glyoxal-derivated lysine dimer* [GOLD], pentosidine[239]… Des exemples de structures d'AGEs sont représentés en figure 32.

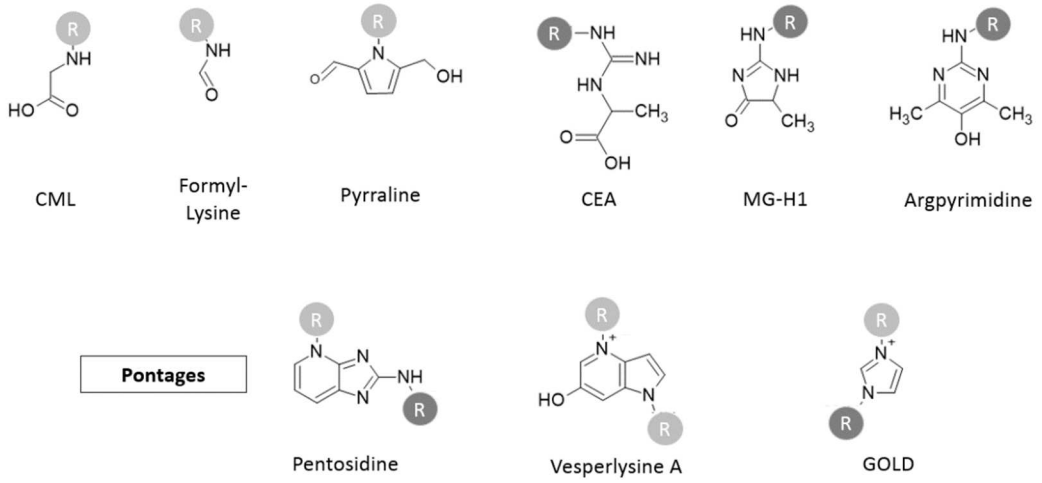

Figure 32 : Structure de différents AGEs issus de la dégradation des produits d'Amadori ou de la récation avec des dérivés carbonylés ; R (gris clair) = Lysine, R (gris foncé) = Arginine (d'après Soboleva *et al.* 2017)[239].

La réaction de glycation est spontanée et sa sélectivité est déterminée par la réactivité des amines. Plusieurs paramètres influent sur cette réactivité tels que l'accessibilité de l'amine, le pKa microscopique, la proximité d'une base conjuguée pour catalyser la déshydratation ou encore l'environnement proche de l'amine. Ainsi, les acides aminés acides se trouvant dans l'environnement proche des lysines et des arginines favorisent la réaction tandis qu'il a été montré que les liaisons hydrogène pouvaient protéger les lysines de la glycation.[240] La réaction de glycation dépend aussi du sucre considéré et de sa concentration. Le fructose a une réactivité différente du glucose[241]. Ce ne sera pas le produit d'Amadori qui sera formé à l'issue de cette réaction mais le produit d'Heyns.[242] Il a été montré que le fructose a une réactivité 8 à 10 fois supérieure au glucose du fait de la stabilité de sa conformation chaine ouverte et de son groupement cétone[241].

Cette réaction complexe se produit non seulement dans pratiquement tous les aliments chauffés[243-244], mais également sur les anticorps monoclonaux dans les industries biopharmaceutiques[245-247] et bien sûr *in vivo*[248-251]. Les protéines portant des glycations (produit

d'Amadori ou AGEs) voient leur structure[252], leur demi-vie[237] et leur fonction affectés[253]. De ce fait, il a été montré qu'un taux anormal de glycation, par exemple dû à une exposition prolongée à de fortes concentrations de sucres dans l'organisme, a des effets délétères pour la santé.[244, 254] Les glycations jouent un rôle dans les complications du diabète[255-256], mais également dans l'apparition et le développement des insuffisances renales[256], des maladies neuro-dégénératives[257] ainsi que sur le vieillissement prématuré[243]. Les AGEs sont également connus pour activer des récepteurs spécifiques (*Receptor for Advanced Glycation End-products* [RAGE]) engendrant une réponse inflammatoire[241, 258].

Par ailleurs, depuis la fin des années 1970, l'hémoglobine glyquée, dont la forme majoritaire est l'hémoglobine HbA1c portant une fructosamine sur la valine N-ter, est un marqueur clinique utilisé pour le suivi de l'équilibre glycémique chez les patients diabétiques[249, 259-260]. Les fructosamines plasmatiques et l'albumine glyquée sont également utilisées en clinique comme biomarqueurs du contrôle de la glycémie[248, 259].

Pour étudier les glycations, différentes stratégies analytiques ont été mises en place comme le suivi des AGEs par fluorescence[252, 261-262] ou par Western Blot[263], l'analyse des protéines glyquées par spectrométrie de masse sans[264-266] ou avec digestion (approche *bottom-up*)[123, 267-269]. Cependant, il est important de noter que la fluorescence est peu spécifique et ne permet pas d'obtenir des informations sur la nature des produits de glycation[259]. De même, le faible débit d'analyse et les problèmes liés à l'utilisation des anticorps, tels que leur potentielle non-spécificité et leur coût élevé, limitent les applications du Western blot[270]. Avec le développement des instruments et des techniques ces dernières années, la spectrométrie de masse est devenue un outil très puissant et a fourni une opportunité unique d'étudier les PTMs des protéines à grande échelle, bien que cette technique rencontre de nombreux défis analytiques.

II. Défis analytiques de l'analyse *bottom-up* des protéines glyquées

L'approche *bottom-up* est particulièrement adaptée à l'analyse des glycations dans des mélanges complexes puisqu'elle permet une analyse non ciblée et à haut débit. De plus, il est possible à la fois d'identifier les protéines glyquées et d'obtenir la localisation des lysines et arginines modifiées. Cependant, en plus de la multitude de formes que peuvent prendre les glycations, d'autres défis analytiques entrent en jeu tout au long de l'analyse protéomique, comme illustré en figure 33.

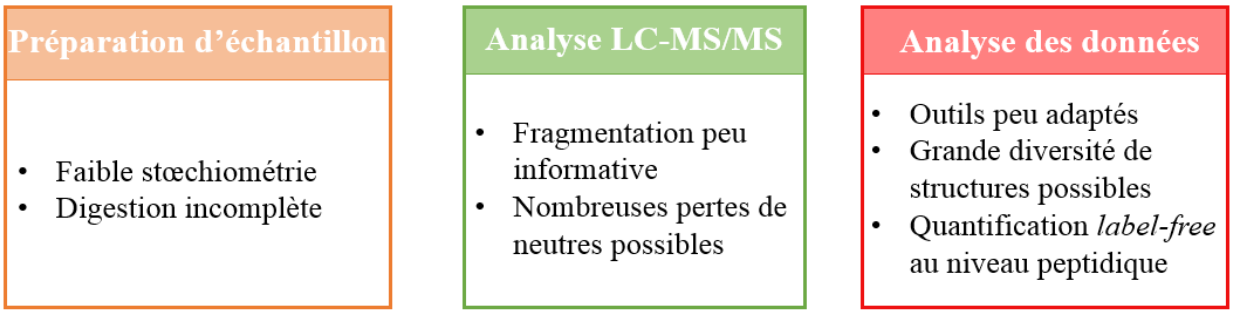

Figure 33 : Difficultés rencontrées pour l'étude des glycations au cours des trois grandes étapes de l'analyse protéomique : la préparation d'échantillons, l'analyse LC-MS/MS et le traitement des données.

A. Faible représentation des peptides glyqués par rapport aux peptides non modifiés

Tout d'abord, les peptides glyqués sont des espèces minoritaires[232] : les peptides non glyqués sont en moyenne 20 à 100 fois plus abondants[237]. Dans de nombreuses études, les échantillons sont incubés *in vitro* avec différents sucres (glucose[271-274], fructose[271, 273-274], ribose[271, 274]...) ou avec des dérivés carbonylés (méthylglyoxal[274]) pour augmenter artificiellement le nombre de peptides glyqués. Les échantillons peuvent aussi être enrichis en protéines/peptides glyqués[275-277]. L'enrichissement en protéines/peptides portant des fructosamines est le plus souvent réalisé par chromatographie d'affinité au boronate. Le principe de cette chromatographie est décrit dans le paragraphe III. Pour les autres formes de glycations, il n'existe pas d'enrichissement spécifiques préalable à des analyses protéomiques mais des anticorps anti-CML ou anti-MG-H1 ont été produits pour réaliser des tests immunologiques[278].

B. Digestion trypsique impactée par l'encombrement stérique des lysines et arginines

La digestion des protéines glyquées par la trypsine est impactée par les glycations qui sont situées pour l'immense majorité sur les lysines et les arginines. La plupart des peptides glyqués ne sont donc pas clivés, ce qui engendre des peptides plus longs et souvent difficiles à identifier et quantifier car ils présentent des profils de fragmentation inhabituels[279]. Plusieurs études ont donc proposé l'utilisation d'enzymes alternatives telles que la GluC, AspN, la pepsine, la protéinase K ou la chymotrypsine[239, 269, 280].

C. Difficultés de l'analyse en spectrométrie de masse

L'ionisation des peptides glyqués n'a été décrit comme l'un des défis majeurs pour l'analyse des peptides glyqués que par une publication de Priego-Capote *et al.*[234]. De plus, ils présentent des profils de fragmentation particuliers avec les méthodes de fragmentation classique de type CID ou HCD. Différentes pertes de neutres sont observées : perte de molécules d'eau, du composé d'Amadori (- 162 Da) ou de formaldéhyde (- 30 Da)[271-272]. Keilhauer *et al.* ont montré que la perte de neutre la plus observée pour le produit d'Amadori est la perte de 3 molécules d'eau (- 54 Da)[272]. Sur la figure 4 est représenté le spectre de fragmentation d'un peptide glyqué présenté dans l'étude de Frolov *et al.*[271].

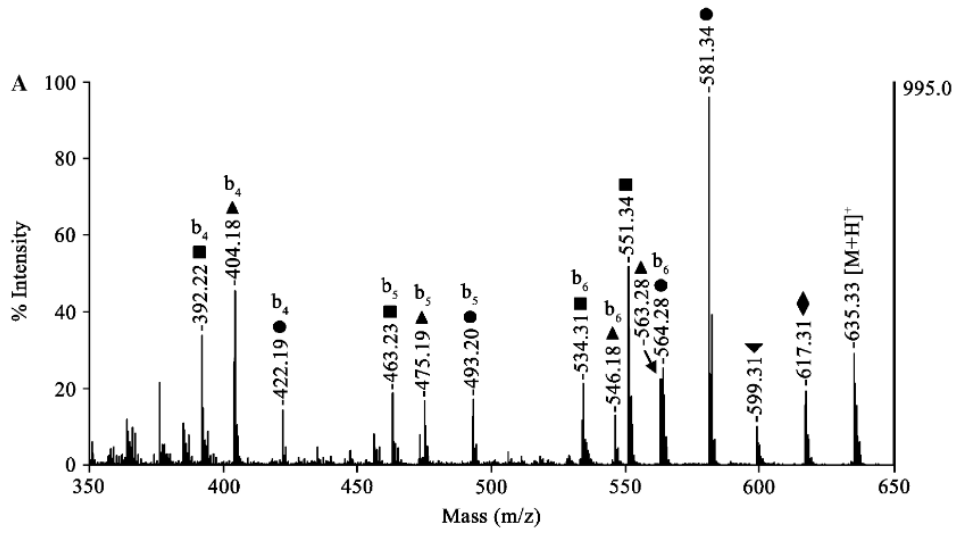

Figure 34: Spectre de fragmentation nanoESI-QqTOF-MS/MS enregistré sur un QSTAR Pulsar I (en mode ion positif), du peptide HAGG**K**AA-NH2 (m/z 635.3) portant une glycation sur la lysine. Les ions fragments sont marquées par un losange plein (♦ ,- H2O), un triangle plein (▼, - 2 x H2O), un cercle plein (●, - 3 x H2O), un triangle plein (▲, - 4 x H2O), un carré plein (■, - 3 x H2O - HCHO)[271].

Les ions issus des pertes de neutre dominent les spectres de fragmentation des peptides glyqués en CID et la fragmentation de la séquence peptidique est très limitée et peu abondante. Il est donc difficile d'obtenir des informations concernant la séquence peptidique portant la glycation[276]. C'est pourquoi des modes de fragmentation alternatifs tels que l'ETD ont été utilisés pour étudier les peptides portant des fructosamines.[276, 281] Deux exemples de spectres du même peptide fragmenté par CID et ETD sont présentés en figure 35. Sur le spectre CID, les ions obtenus correspondent à différentes pertes de neutres tandis que sur le spectre ETD, de nombreux fragments (ions c et z) informatifs sont obtenus, permettant ainsi l'identification du peptide et la localisation de la modification.

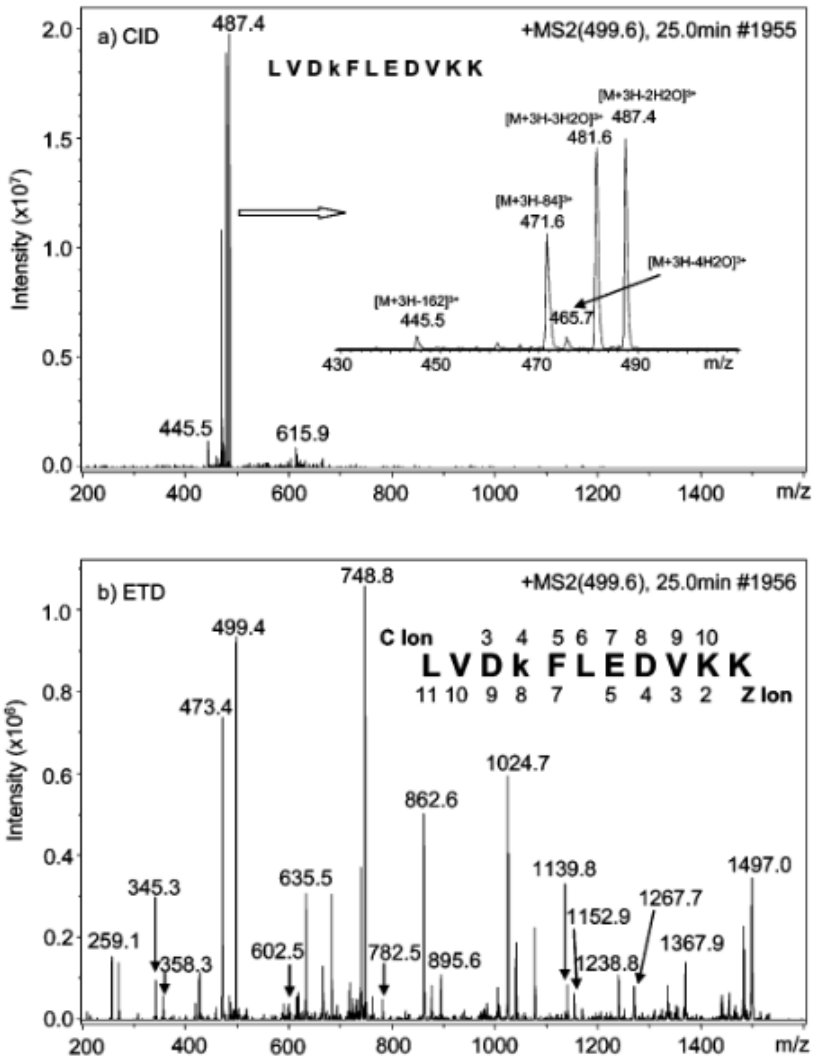

Figure 35 : Spectres de fragmentation du peptide LV**K**FLEDVKK après fragmentation par CID ou par ETD[276].

Cependant, il faut noter que la fragmentation ETD reste une technique moins sensible que la fragmentation CID ou HCD car l'efficacité de la fragmentation est plus faible : les intensités des fragments obtenus par ETD ne représentent que 20 à 30 % de celles obtenues par CID[276, 282]. De plus, l'un des défauts de cette technique est le temps réaction entre l'anion portant l'électron et le peptide cation. Pour raccourcir ce temps de réaction, une double fragmentation par EThcD, peut être utilisée en combinant une première fragmentation rapide et incomplète par ETD puis une seconde par HCD. Cette technique a été appliquée sur les peptides glyqués du NISTmab par Wang *et al.*[136]. Grâce à cette technique, 27 peptides glyquées ont été identifiés soit un de plus qu'en fragmentation HCD, tout en augmentant le nombre et l'intensité des fragments informatifs. Pour optimiser encore la fragmentation des peptides glyqués, Wang *et al.* ont développé une méthode de fragmentation selon le schéma de la figure 36.

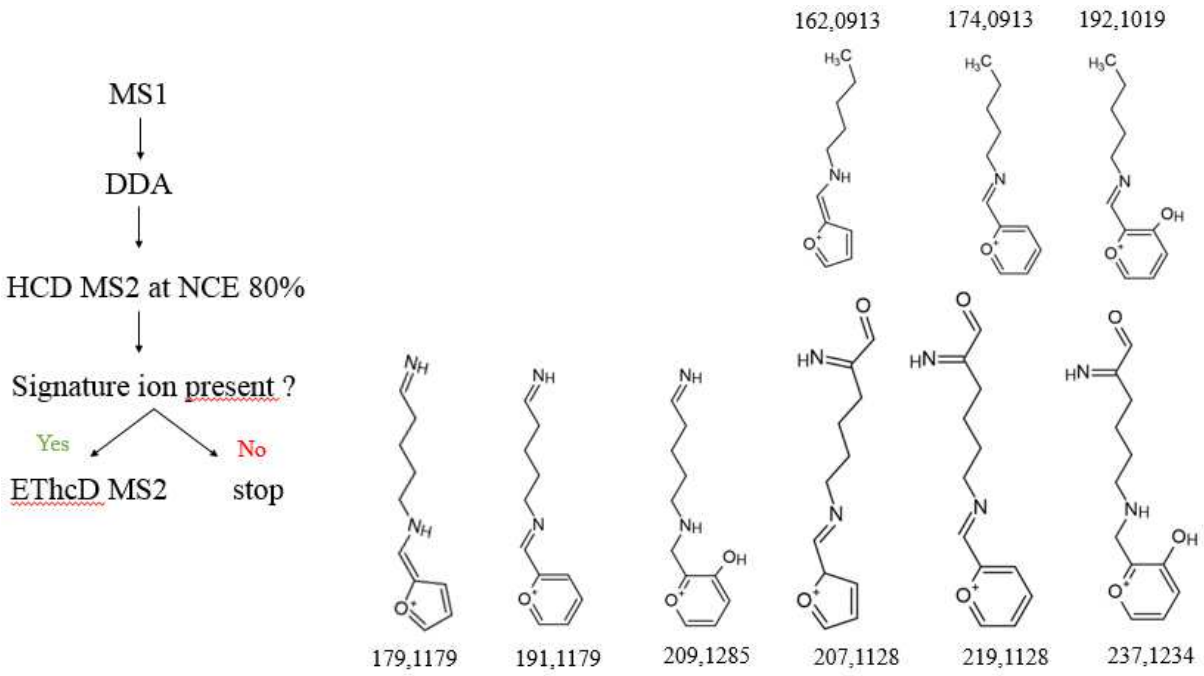

Figure 36 : Arbre décisionnel mis en place par Wang *et al.* pour la fragmentation des peptides glyqués : l'ion précurseur n'est fragmenté par EThcD que si une première fragmentation par HCD a permis l'identification d'un des 9 ions signatures des glycations[136].

Après l'acquisition d'un spectre MS, un des ions les plus intenses est fragmenté en mode DDA par HCD. Grâce à une recherche en temps réel, les ions du spectre MS/MS peuvent être attribués. Si l'un des ions signatures des glycations est détecté, l'ion précurseur est fragmenté une seconde fois par EThcD. Cette technique génère plus de spectres de peptides glyqués que l'EThcD seul et permet l'identification de 32 peptides glyqués sur le NISTmab sans enrichissement préalable.

D. Manque d'outils bio-informatiques dédiés

L'un des défis de l'analyse des glycations repose dans la difficulté à interpréter les données acquises. Lorsque les fragmentations CID ou HCD sont utilisées pour l'analyse des peptides portant une fructosamine, les fragments obtenus sont très divers. De ce fait, les logiciels utilisés classiquement en protéomique sont souvent peu performants pour l'analyse des glycations[283]. La méthode la plus courante pour interpréter les spectres de masse des peptides glyqués est similaire au traitement des autres modifications post-traductionnelles dans les recherches de bases de données protéomiques, c'est-à-dire qu'elle consiste à rechercher toutes les glycations potentielles en tant que modifications variables sur tous les sites de glycation possibles. Cette approche est limitée par le potentiel grand nombre de protéoformes glyquées, les combinaisons multiples possibles entraînant la multiplication des configurations possibles pour les peptides qui contiennent plusieurs sites potentiels de glycation. De plus, la plupart des logiciels cherchent des fragments où la modification est restée liée au peptide, or les glycations sont labiles et sont le plus souvent perdues durant la fragmentation. Certains logiciels permettent la détection de pertes de neutre, mais, encore une fois, il existe une multitude de pertes de neutre possibles pour une forme glyquée, ce qui augmente encore le nombre de combinaisons possibles. Des approches *open-search* dédiées à l'étude des glycations[284] ou des

glycosylations[283] ont été développées récemment pour interpréter au mieux les spectres de fragmentation complexes. Dans les recherches *open-search*, la masse du peptide est déterminée par l'appariement des ions fragments sans connaissance du précurseur. La différence entre la masse de la séquence déterminée par l'appariement des fragments et la masse du précurseur observée en MS1, appelée delta de masse ou décalage de masse, est la masse de toute modification non spécifiée de la séquence. Pour les modifications qui sont labiles, cette stratégie tient compte de leur présence sur le précurseur *via* la masse delta sans exiger leur présence sur les ions fragments, ce qui permet de faire correspondre une plus grande proportion des ions fragments observés.

Un autre point critique est la quantification des peptides glyqués. Comme pour tous les autres peptides modifiés, la quantification *label-free* s'effectue sur un peptide et non sur plusieurs peptides comme pour la quantification au niveau protéique. Le nombre de valeurs manquantes est de fait plus élevé et la quantification moins robuste. Pour remédier à ce problème, l'analyse ciblée par MRM[251] des peptides glyqués est utilisée afin d'augmenter la sensibilité du spectromètre de masse. Avec cette approche, les avantages de la spectrométrie de masse (analyse globale et à haut débit) sont perdus. Une autre technique repose sur l'incubation des protéines avec du glucose[$^{13}C_6$] dans des conditions « imitant » la glycation *in vivo*. Dans le cadre de cette approche, la quantification relative repose sur des signaux doublets en MS1 représentant la glycation *in vivo* avec le glucose[$^{12}C_6$] et l'incorporation *in vitro* du glucose[$^{13}C_6$], comme illustré en figure 37[38].

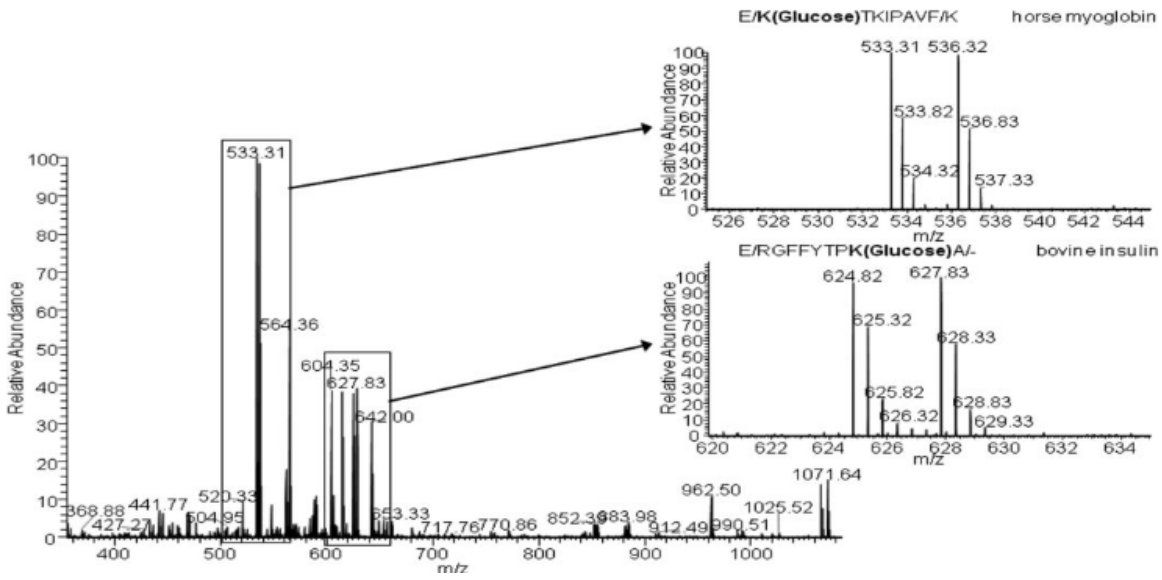

Figure 37 : Spectre MS/MS dans lequel deux peptides glyqués (E2K(glucose)TKIPAVF2K et E2RGFFYTPK(glucose)site A2C-terminal pour la myoglobine de cheval et l'insuline bovine, respectivement) sont co-élués. Les signaux du doublet sont agrandis pour montrer l'efficacité de la glycation du glucose [$^{12}C_6$] *in vivo* et du glucose [$^{13}C_6$] *in vitro*[268].

À partir de l'intensité de chacun des signaux, le ratio [$^{12}C_6$] *in vivo* / [$^{13}C_6$] *in vitro* est calculé. Il a été montré que ce ratio est proportionnel au taux d'hémoglobine glyquée pour différents sites de glycation situés sur des protéines de l'hémolysat[285].

Le dernier défi de l'analyse des glycations est le manque de ressources regroupant les informations connues sur les sites de glycations et leurs conséquences comme ce peut être le cas pour les phosphorylations avec PhosphoSitePlus. Uniprot rassemble certaines informations mais il est parfois difficile de faire le tri entre les études où les protéines ont été incubées *in vitro* dans des conditions parfois très éloignées des conditions physiologiques et les études *in vivo*. De même, un outil permettant la prédiction des sites de glycations à partir de la séquence peptidique serait également très utile pour confirmer certains résultats[237].

III. Enrichissement par chromatographie d'affinité au boronate (*boronate affinity chromatography* [BAC])

L'enrichissement BAC a pour objectif de retenir sélectivement les peptides portant une fructosamine. La phase boronate, généralement sous forme d'acide phenylboronique [PBA], réagit à pH basique avec les cis-1,2-diol portés par les fructosamines *via* une réaction d'estérification pour former un cycle à 5 atomes, stabilisé par des interactions électrostatiques entre l'atome de bore et la fonction amine[286-287]. La réaction est illustrée en figure 38.

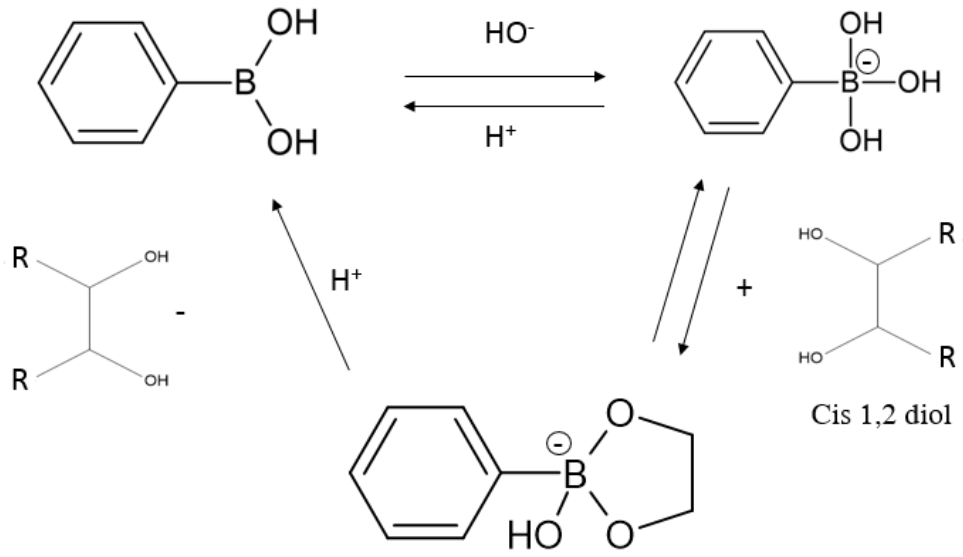

Figure 38 : Réaction entre les cis diol et l'acide phenylboronique (d'après Espina-Benitez *et al.* 2018)[287].

Pour éluer les peptides glyqués retenus sur la phase, l'ester formé est hydrolysé par des conditions acides. En dehors du pH, d'autres paramètres influent sur cette réaction tels que la force ionique, la température, la présence de certains ions $H_2PO_4^-$, Mg^{2+}, Na^+[287].

Pour la BAC, l'affinité est définie par la constante d'association [K_a] et sa réciproque la constante de dissociation [K_d]. La K_d est dépendante des conditions expérimentales (pH, température…). La K_d de la BAC se situe entre 10^{-1} et 10^{-4} M tandis que l'affinité (K_a) d'une séparation par immuno-affinité se situe typiquement en dessous de 10^{-7} M[287]. La réaction d'estérification peut aussi avoir lieu avec les fonctions Cis-1,3-diol bien que le cycle formé soit

moins stable[287]. Des interactions secondaires, comme les interactions hydrophobes et Π-Π, électrostatiques, *via* des liaisons hydrogènes ou par coordination, peuvent également piéger des composés non désirés[286-287]. Un schéma récapitulant les différentes interactions possibles avec la phase PBA est présenté en figure 39.

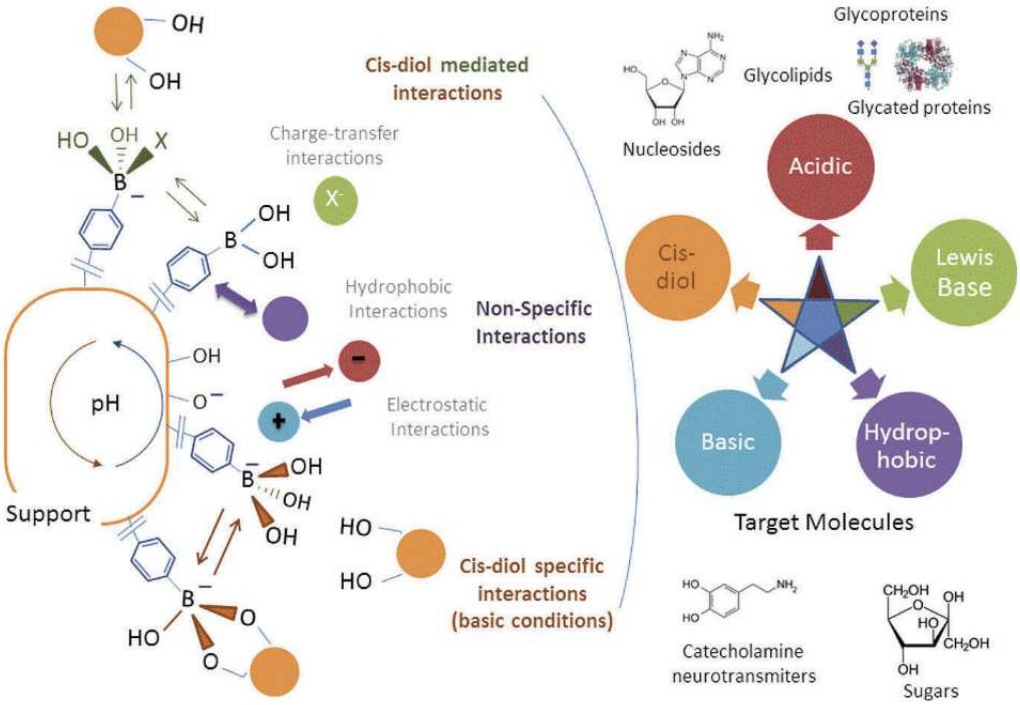

Figure 39 : Interactions possibles avec la phase PBA (d'après Espina-Benitez *et al.* 2018)[287].

Ce type d'enrichissement peut être utilisé pour isoler les protéines glyquées intactes, bien qu'il soit moins efficace que l'enrichissement au niveau peptidique[232, 282]. Différents supports d'affinité boronate existent. Le plus répandu se présente sous forme de colonne HPLC utilisée avant analyse LC-MS/MS pour isoler la fraction protéique ou peptidique glyquée.[268, 282] Ainsi, l'enrichissement BAC a été appliqué à l'étude de protéines isolées (RNase[282], protéines thérapeutiques[246]) ou de mélanges plus complexes comme du plasma[232, 288] ou du serum[289]. Ce type de colonne a aussi été monté avec succès en série (2D LC-MS/MS) avec une colonne *Reverse Phase* [RP] C18 classique afin d'augmenter le débit d'analyse et de réduire le temps nécessaire à la préparation d'échantillon[275, 290]. Cette phase a également été utilisée sous forme de billes fonctionnalisées pour analyser de la *Bovine Serum Albumin* [BSA] purifiée mais également les protéines glyquées issues du plasma de patients diabétiques[123].

IV. Conclusion

En conclusion, la réaction non enzymatique de glycation induit des modifications aux structures très variées. Ce sont des modifications qui sont beaucoup moins étudiées que les phosphorylations et des efforts importants restent à fournir à tous les niveaux de l'analyse pour parvenir à quantifier à large échelle ces modifications.

PARTIE II : MISE AU POINT DE PROTOCOLES D'ANALYSES PROTEOMIQUES QUANTITATIVES POUR ETUDIER L'HIBERNATION CHEZ DES ESPECES NON MODELES

Chapitre 1
Analyse du protéome des mitochondries musculaires de l'ours brun pour mieux comprendre le ralentissement du métabolisme durant l'hibernation

Cette étude a été réalisée en collaboration avec les Dr Etienne Lefai (INRA Centre Clermont-Ferrand) et Dr Audrey Bergouignan (IPHC DEPE, Strasbourg).

I. L'hibernation de l'ours

L'hibernation est une stratégie d'économie d'énergie qui permet à différentes espèces d'animaux de faire face au froid et/ou à un accès limité à la nourriture. Cette stratégie est adoptée par de nombreux mammifères des régions arctiques mais a également été observée chez une espèce d'oiseau et chez des mammifères originaires des tropiques et de l'hémisphère sud[291-292]. L'état de torpeur dans lequel entrent les animaux hibernants peut durer plusieurs jours à plusieurs semaines, voire plusieurs mois. Il est caractérisé par une diminution contrôlée de la température corporelle et du taux métabolique (couramment mesuré par la consommation d'oxygène de l'organisme), un ralentissement des fonctions vitales (respiratoires (7 respirations /minutes versus 12 en période active chez l'ours[293]), cardiovasculaires (32 versus 79 en période active chez l'ours[293]), rénales), une diminution de la vigilance, ainsi qu'une prise alimentaire et une activité physique fortement diminuées[294]. Durant la torpeur, le taux métabolique des petits mammifères chute à 1-10% du taux basal et leur température corporelle se situe alors généralement entre 0 et 10 °C[292]. Il existe plusieurs schémas d'hibernation : chez certaines espèces de rongeurs hibernants, les épisodes de torpeur sont interrompus par des phases de réveil où les constantes vitales retrouvent leur niveau habituel[295].

L'hibernation de l'ours est très particulière. Elle peut durer entre 4 et 7 mois, au cours desquels l'ours ne connait aucun épisode de réveil. Sa température corporelle n'est diminuée que de quelques degrés Celsius, pour atteindre 33 °C (vs 37 °C en période active), et son taux métabolique est abaissé jusqu'à 25% du taux mesuré en période active[294]. Durant son hibernation, l'ours brun (*Ursus arctos*) reste complètement inactif, sans boire, ni manger, ni éliminer ses déchets[296-297]. Malgré ces conditions extrêmes, la force et la masse musculaire de l'ours restent bien conservées : -30% de force musculaire après 110 jours d'inactivité et de jeune[298-299]. À titre de comparaison, un humain alité mais nourri perd plus de 50% de sa force musculaire en 90 jours[300]. La préservation des organes de l'ours en torpeur est, en partie, due à l'épargne d'énergie durant l'hibernation. Comprendre les mécanismes qui sous-tendent cette flexibilité métabolique autorisant la forte réduction du taux métabolique aiderait à mieux appréhender les mécanismes vitaux de préservation des organes.

II. Problématique : quels mécanismes pour le ralentissement métabolique chez l'ours brun hibernant (*Ursus arctos*) ?

Chez les petits hibernants, certaines études semblent montrer que la baisse du taux métabolique est liée à la baisse de la température corporelle, selon les règles de la loi d'Arrhenius. Aussi appelée Q10, elle décrit que la cinétique d'une réaction enzymatique est divisée par 2 à la suite d'une baisse de 10 °C[301]. Durant l'hibernation, le métabolisme de l'ours est diminué de 75% par rapport à son taux basal, indépendamment de la faible baisse de température corporelle[302]. On s'attend donc à ce que des mécanismes actifs (par opposition à un mécanisme passif entrainé par la température corporelle) soient présents chez l'ours pour contrôler l'épargne d'énergie [303].

À l'intérieur des cellules, les mitochondries sont à la fois les sites principaux de la consommation d'oxygène (90% de l'oxygène est utilisé dans les mitochondries) et de la production de chaleur[304]. C'est pourquoi, leur analyse revêt un intérêt particulier pour comprendre à la fois la dépression du métabolisme et le maintien de la température corporelle proche de la normale durant l'hibernation de l'ours brun.

Lors d'une précédente analyse protéomique différentielle (N=7/période, hibernation vs période active), réalisée sur tissu musculaire entier, nous avions montré que les sous-unités appartenant à chacun des cinq complexes de la chaine respiratoire mitochondriale sont systématiquement moins exprimées durant l'hibernation[36]. L'abondance de certaines de ces sous-unités est représentée en figure 40.

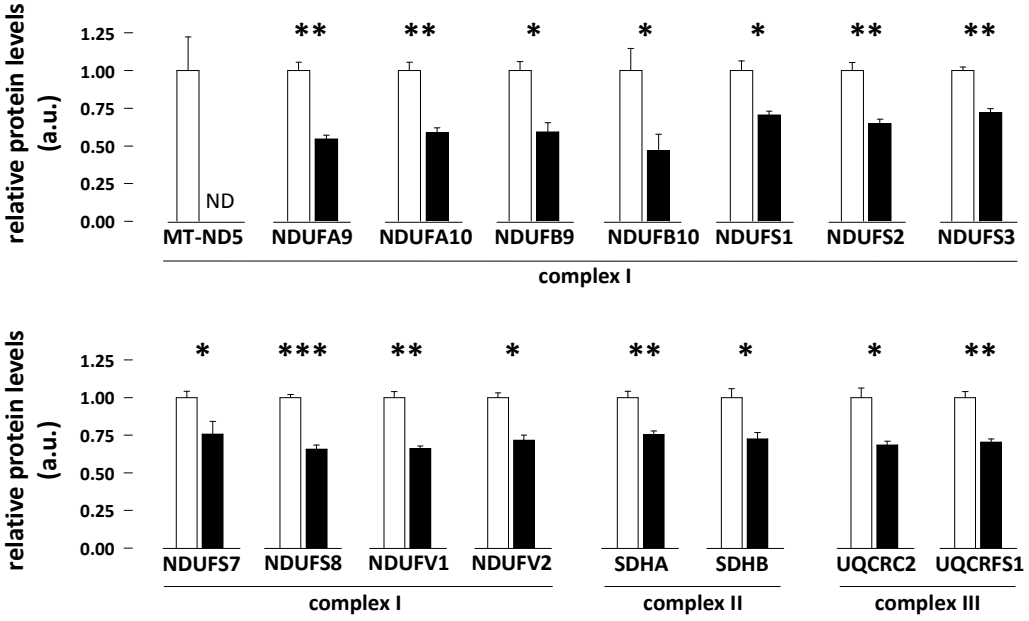

Figure 40 : Niveau d'expression des sous-unités de la chaine mitochondriale en hiver (noir) et en été (blanc).

L'abondance diminuée de la citrate synthase [CS] (enzyme du cycle de Krebs) lors de l'hibernation reflète une réduction du nombre de mitochondries en hiver. Des mesures

fonctionnelles mitochondriales, présentées en figure 41, avaient aussi été effectuées pour évaluer l'activité de la CS, et celle de la cytochrome c oxydase [COX] (complexe IV de la chaine respiratoire mitochondriale)[36].

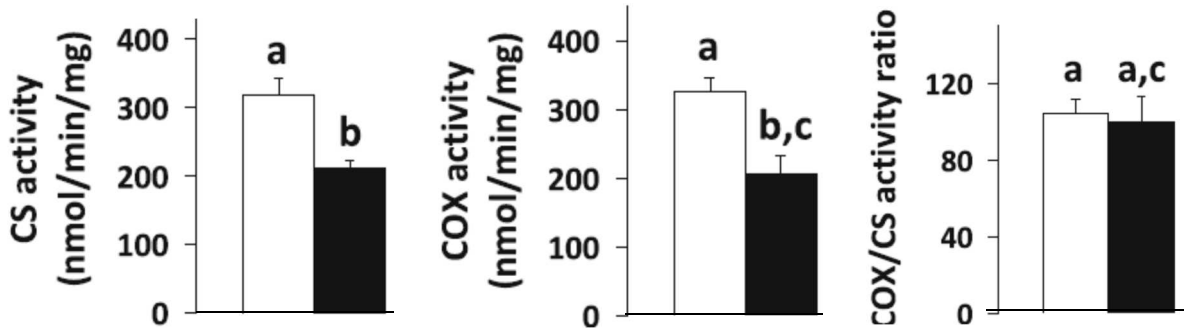

Figure 41 : Activité du cycle de Krebs (CS) et de la respiration mitochondriale (COX) et du ratio COX/CS en été (blanc) et en hiver (noir).

L'activité de ces deux enzymes apparaissait diminuée en hiver par rapport à la période active. Cependant, le ratio des activités COX et CS restait stable entre les deux saisons, suggérant que la capacité oxydative maximale des mitochondries musculaires ne diffère pas entre les deux saisons. Ainsi, une modification à l'échelle de la mitochondrie (en termes d'équipement enzymatique) ne semblait pas intervenir entre l'hibernation et l'état actif dans le muscle de l'ours brun. L'objectif de ce projet était de confirmer cette hypothèse en produisant des fractions musculaires enrichies en mitochondries puis de les analyser selon une approche protéomique quantitative « *label-free* ».

III. Stratégie analytique

La mitochondrie est un organite qui possède une double membrane très riche en protéines. D'après Gomez-Serrano *et al*., il existerait entre 1100 et 1900 protéines mitochondriales, seules 13 d'entre elles étant codées par le génome mitochondrial, tandis que la plus grande majorité est codée par le génome nucléaire[37]. Cependant ce paradigme serait remis en question récemment avec la découverte de nouveaux peptides issus du génome mitochondrial[305]. L'isolation des mitochondries peut être réalisée par centrifugation différentielle après homogénéisation mécanique[306], ou directement à partir des fibres musculaires par perméabilisées[307].

L'analyse protéomique des protéines mitochondriales représente un défi du fait de la gamme dynamique en concentration étendue (6 ordres de grandeur)[37] et de l'hydrophobicité des protéines membranaires. Il est donc important d'adapter la préparation de l'échantillon afin de ne pas perdre ces protéines au cours du processus, par exemple en utilisant des détergents facilitant la mise en solution de ces protéines. C'est pourquoi, nous avons ajouté du SDS au tampon urée/thio-urée/ tris(hydroxyméthyl)aminométhane [Tris] utilisé. La stratégie analytique mise en place pour l'étude du protéome mitochondrial est illustrée en figure 42.

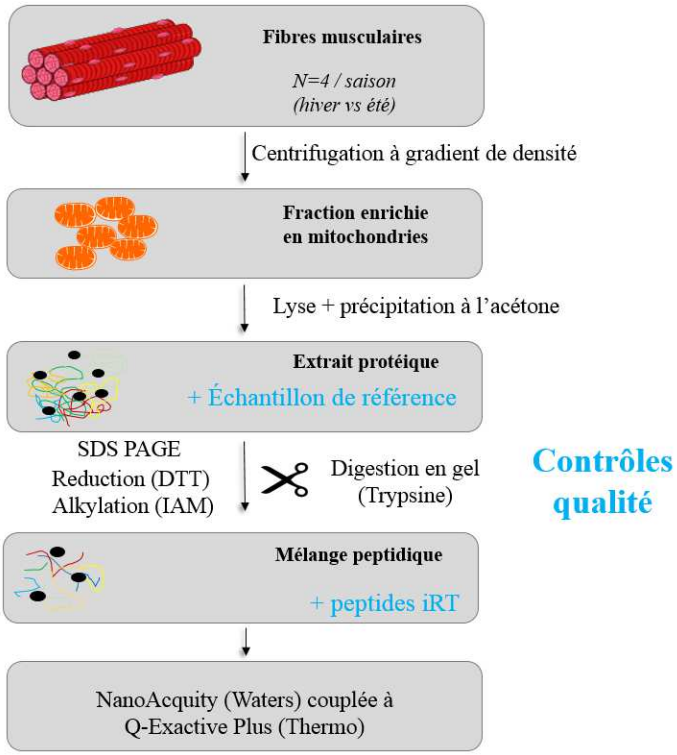

Figure 42 : Analyse des mitochondries issues des fibres musculaires d'ours en hiver et en été (N=4/saison).

Des biopsies musculaires ont été réalisées sur les mêmes ours (N=4) à deux périodes de l'année : durant l'hibernation (février) et durant la phase active (juin). Les fractions enrichies en mitochondries ont été ensuite préparées directement après prélèvement, par centrifugation différentielle dans un tampon contenant du mannitol et du sucrose. Les mitochondries ont été lysées et 6 µg de protéines ont été déposés sur un gel SDS-PAGE, puis les protéines ont migré par électrophorèse dans le gel de concentration afin de former une bande protéique concentrée. Les protéines ont ensuite été réduites, alkylées, et digérées par la trypsine dans le gel. Les peptides ont été extraits du gel et injectés sur un couplage nanoLC-MS/MS constitué d'une NanoAcquity (Waters) et d'un Q-Exactive Plus (Thermo Fisher Scientific). Les données générées ont ensuite été traitées avec MaxQuant (v 1.6.11.0). Les détails du protocole sont donnés en partie expérimentale I. Pour le suivi de la stabilité du système, des contrôles qualité ont été effectués à l'aide de standards internes (11 peptides synthétiques iRT (Biognosys)) et d'un standard externe (échantillon de référence).

IV. Résultats de l'analyse protéomique

A. Contrôles qualité

Des peptides synthétiques iRT ont été ajoutés à chaque échantillon avant injection pour servir de standards internes. Les temps de rétention de ces peptides ont été mesurés tout au long des analyses, un exemple de ce suivi est illustré en figure 43 pour le peptide de séquence

DGLDAASYYAPVR. Un coefficient de variation [CV] moyen de 2,6 % a été calculé sur l'ensemble des injections, en considérant tous les peptides iRT.

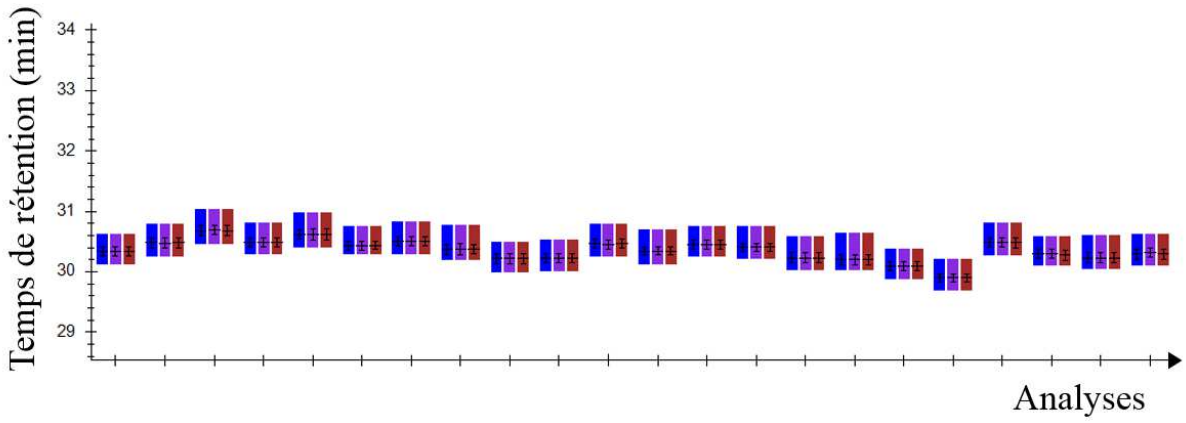

Figure 43 : Temps de rétention du peptide DGLDAASYYAPVR au cours des analyses.

Un échantillon de référence, provenant d'un échantillon utilisé lors de tests, a été injecté tous les quatre échantillons au cours de la séquence d'injection. Ce contrôle qualité externe, représentatif de nos échantillons, a permis de suivre la stabilité du système au cours des analyses. Le CV médian des valeurs d'intensités normalisées par MaxQuant (*Label-Free Quantification* [LFQ]) dans l'échantillon de référence était de 18 %. Les mesures liées aux contrôles qualité ont donc indiqué des performances stables du couplage nanoLC-MS/MS tout au long de l'expérience.

B. Résultats et interprétations

En utilisant des critères rigoureux établis pour la validation des données de spectrométrie de masse (voir partie expérimentale I.6), 967 protéines ont été identifiées et 720 d'entre elles ont été retenues pour la quantification. Par recherche de similarité de séquence à l'aide de l'algorithme fasta v36 implémenté sur une platefome Galaxy, les protéines humaines homologues des protéines d'ours ont été identifiées. À partir de ces données, les ressources bio-informatiques d'UniprotKB et de Nextprot, ont permis d'obtenir des données relatives à la localisation subcellulaire des protéines. Parmi les 720 protéines retenues pour la quantification, 329 avaient déjà été annotées comme étant mitochondriales (*i.e.* 45,7%). À titre de comparaison, lorsque des extraits protéiques de muscles entiers ont été analysés en utilisant la même méthode, 120 des 538 protéines quantifiées étaient des protéines mitochondriales (*i.e.* 22.3%)[36]. Ceci suggère que la préparation des échantillons sur le terrain a bien permis un enrichissement en mitochondries.

Dans un premier temps, un t-test apparié a été réalisé pour comparer les valeurs obtenues pour chaque protéine quantifiée. Cependant, l'analyse DDA étant stochastique, de nombreuses valeurs d'intensité manquent et ce test n'a pu être effectué que sur 600 protéines en conservant les critères requis pour la validation des données MS. Cette analyse statistique a mis en évidence 18 protéines avec une p-value inférieure à 0,05, dont seulement 7 étaient mitochondriales. Sept autres protéines mitochondriales présentaient une p-value supérieure à 0,05 mais inférieure à

0,07. Dans un deuxième temps, un t-test de Welch (adaptation du test t de Student) a également été réalisé pour pouvoir considérer davantage de protéines. Il a permis de mettre en évidence que les protéines BCL2L13 (p-value < 0,05 ; p = 0,11 avec t-test apparié), NDUFA2 (p-value = 0,056 ; p = 0,15 avec t-test apparié) et SLC27A1 (p-value = 0,053 ; p = 0,123 avec t-test apparié) présentaient une p-value inférieure à 0,06. En revanche, aucune différence significative n'a été retrouvée avec le t-test de Welch pour les protéines non testées avec le t-test apparié du fait des valeurs manquantes. Les changements d'abondance entre les deux saisons des protéines mitochondriales évoquées précédemment peuvent être observés sur la figure 44.

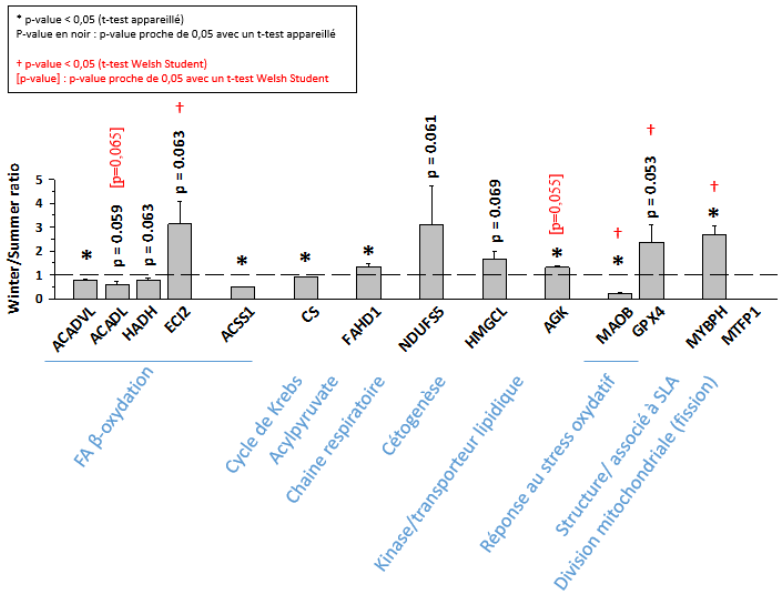

Figure 44 : Ratio de l'intensité des protéines mitochondriales en hiver par rapport à l'été.

Bien que nous n'ayons pas un enrichissement de 100% en protéines mitochondriales, on peut considérer que la comparaison concerne un contenu comparable en mitochondries pour les 2 saisons. Il est frappant de constater qu'un nombre très faible de protéines apparaissent exprimées de manière différentielle dans les extraits mitochondriaux par rapport à ce qui avait été observé lors de l'étude précédente sur des extraits de muscles entiers[36]. Ceci conforte l'idée d'une diminution du contenu mitochondrial pendant l'hibernation, avec très peu de régulations au niveau des mitochondries elles-mêmes.

En bref, les protéines impliquées dans l'oxydation des acides gras (ACADVL, ACADL, HADH, ACSS1) et le cycle de Krebs (CS) sont moins abondantes durant l'hibernation, ce qui est en accord avec les résultats de l'étude précédente[36]. ECI2, aussi impliquée dans l'oxydation des acides gras dans les mitochondries et les peroxysomes, est plus abondante en hiver. En dehors de l'abondance, des régulations spécifiques liée à sa localisation ou à des modifications spécifiques (comme des phosphorylations sur les sérines 101 et 119)[21, 308] pourraient influer sur son activité. Les données suggèrent également un stress oxydant limité (MAOB, GPX4) pendant l'hibernation, comme cela a été montré précédemment[309].

Concernant la chaîne respiratoire, 75 sous-unités ont été quantifiées, parmi lesquelles une seule (NDUFS5 ; complexe I) était plus abondante pendant l'hibernation. Enfin, MTFP1, une protéine impliquée dans la division mitochondriale (fission) est restée indétectable chez les ours en hibernation. Les interprétations biologiques plus poussées de ces résultats sont encore en cours.

C. Validation de l'étude protéomique par une méthode orthogonale

Des données supplémentaires sont venues appuyer ces observations. Les mitochondries possèdent leur propre acide désoxyribonucléique [ADN] codant pour 13 protéines. Le ratio ADN mitochondrial sur ADN nucléaire a été mesuré sur 8 ours pour les deux saisons. Une chute de ce ratio est observée en hiver par rapport à l'été (-60%, Figure 45).

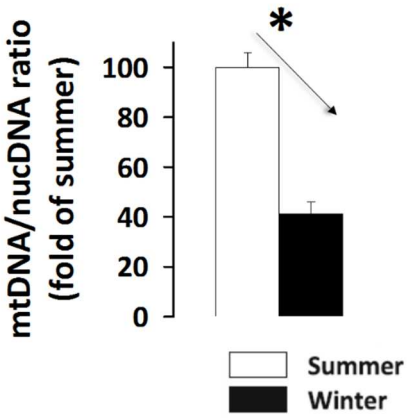

Figure 45 : Ratio ADN mitochondrial sur ADN nucléaire normalisé par rapport au ratio en été.

Ces résultats concordent avec les données présentées plus haut. Le nombre de mitochondries semble donc bien diminué durant l'hibernation de l'ours brun, l'équipement enzymatique restant très peu affecté à l'échelle de la mitochondrie elle-même. Chez le loir commun, il a également été montré que la morphologie des mitochondries restait inchangée durant l'hibernation[310], suggérant également une conservation de l'équipement enzymatique durant l'hibernation chez cette espèce.

V. Conclusion et perspectives

L'analyse protéomique des mitochondries musculaires d'ours brun a permis de montrer que si le contenu mitochondrial musculaire est diminué lors de l'hibernation, aucune modification majeure ne semble intervenir à l'échelle de la mitochondrie elle-même. Cette étude montre donc l'importance de l'analyse protéomique des sous-protéomes. Les mécanismes responsables de la diminution du nombre de mitochondries dans le muscle d'ours avant/pendant l'hibernation restent à élucider. S'agit-il d'une biogénèse durant la période active induite par l'activité physique[311], d'une fission ou fusion des mitochondries ou d'un processus de mitophagie au cours de l'hibernation ?

En 2003, Barger *et al.* ont montré que la respiration des mitochondries musculaires n'était pas affectée par l'hibernation chez *Spermophilus parryii*, l'écureuil arctique[312]. Au contraire, Staples *et al.* ont montré que la respiration mitochondriale chutait durant l'hibernation chez différents petits mammifères hibernants[313]. Afin de compléter nos analyses durant l'hibernation de l'ours brun, des mesures de respiration des mitochondries sous différentes conditions (substrats, inhibiteurs…) à deux températures (33 °C et 37 °C, température corporelle moyenne des ours durant l'hibernation et la phase active) ont été réalisées. Ces résultats sont en cours d'interprétation en vue d'une publication.

Chapitre 2
Développement d'un protocole pour étudier les phosphorylations dans le muscle de l'ours brun pour mieux comprendre les adaptations à l'hibernation

I. Maintien de la masse musculaire et importance des phosphorylations

Chez l'homme, l'inactivité physique prolongée entraine une perte importe du volume et de la force musculaire, allant jusqu'à la moitié de la force initiale en 3 mois d'immobilisation[300]. La réduction de la masse musculaire a des implications sévères sur la santé, en augmentant les risques de morbidité et de mortalité[314]. Au contraire, la force et la masse musculaire sont très bien conservées chez plusieurs espèces d'animaux hibernants malgré plusieurs mois d'inactivité et de jeûne[294]. Parmi eux, l'ours hiberne jusqu'à 7 mois sans manger, boire, éliminer ses déchets ni se réveiller, ne connait que des pertes très limitées de sa force musculaire (- 23%) et de la concentration protéique intramusculaire (- 20% en 6 mois) [298-299, 315]. La conservation de la masse musculaire est régulée par l'équilibre entre la synthèse et la dégradation protéique[316]. Lohuis *et al.* ont montré que durant l'hibernation de l'ours, la synthèse et la dégradation protéique du muscle étaient toutes deux diminuées d'où un relatif maintien de l'équilibre[298]. Des études complémentaires doivent encore être réalisées pour comprendre les bases moléculaires de la régulation de cette balance protéique durant l'hibernation de l'ours.

La synthèse protéique dépend en grande partie des processus de transcription et de traduction[207]. Les protéines impliquées dans l'efficacité de la traduction sont majoritairement des protéines phosphorylables et certains sites de phosphorylation sont connus pour réguler la vitesse d'initiation et d'élongation de la traduction chez l'homme[317-318]. C'est le cas par exemple de la sérine 235 de la protéine ribosomale S6 (RPS6) qui, phosphorylée, augmente l'initiation de la traduction[319]. De même, la phosphorylation du facteur de transcription FOXO, impliqué dans la dégradation protéique, induit une séquestration de ce facteur dans le cytoplasme inhibant ainsi son action[320].

II. Étude des phosphorylations dans le tissu musculaire

La phosphorylation est une modification post-traductionnelle qui régule un grand nombre de processus dans l'organisme (voir Partie I, chapitre 2). C'est une réaction rapide et réversible qui module la conformation de la protéine, activant ou inhibant ainsi la fonction de la protéine, l'activité de l'enzyme et influe sur les interactions entre protéines[183]. Malgré les évolutions

technologiques et instrumentales de cette dernière décade, l'identification et quantification des phosphorylations restent un défi analytique[174, 321].

Les études de phosphoprotéomique quantitative sont pour la plupart basées sur des approches de marquage isotopique (TMT ou *Isobaric Tag for Relative and Absolute Quantitation* [iTRAQ]) et combinent un enrichissement en peptides phosphorylés et un fractionnement au niveau peptidique (SCX ou hpH RP, plus de 10 fractions)[186, 207, 322-323]. Ces étapes sont très coûteuses et allongent significativement les temps de préparation d'échantillon et d'analyse. C'est pourquoi, j'ai cherché à développer une méthode robuste sans marquage qui permette l'identification et la quantification d'un grand nombre de phosphorylations en un temps d'analyse réduit.

Cette méthode devait être automatisable pour augmenter le débit d'analyse et la répétabilité tout en étant compatible avec des quantités de matériel initial relativement restreintes pour une analyse phosphoprotéomique (< 1 mg de protéines). Pour développer cette méthode, je me suis concentrée sur certains points clés de l'analyse : l'extraction des protéines, la digestion, l'analyse LC-MS/MS et le traitement des données. Le schéma analytique est illustré en figure 46.

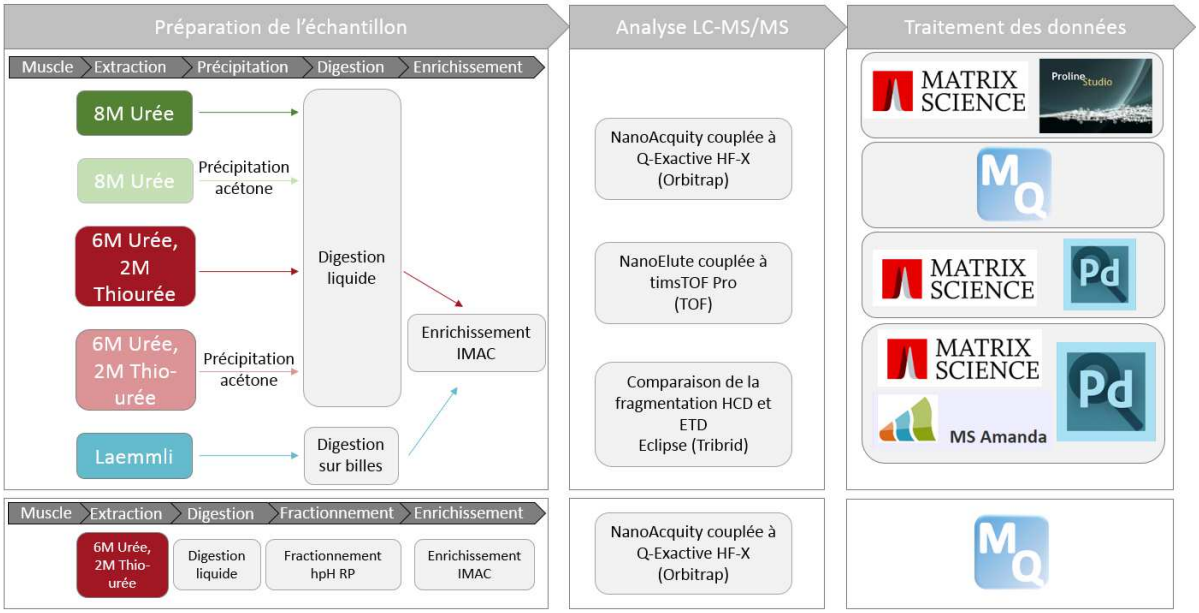

Figure 46 : Représentation schématique du développement de la méthode pour l'analyse des phosphorylations dans le muscle.

Dans un premier temps, l'extraction des protéines à partir du tissu musculaire a été optimisée en évaluant différents tampons de lyse. Après digestion des protéines, les échantillons ont été enrichis en phosphopeptides par IMAC. Cette étape avait été optimisée et automatisée sur un Bravo AssayMAP par Marziyeh Komeili au laboratoire à partir d'extraits protéiques de cellules *human Burkitt's lymphoma (BL)2 B cell line* (BL2). Dans mon cas, il s'agissait d'éprouver cette méthode sur des extraits protéiques préparés à partir de tissus musculaires. Dans le cadre de ce projet, j'ai choisi de conserver les différents paramètres mis au point sur des extraits cellulaires

(débits, conditions de chargement et d'élution, choix de la phase). L'ajout d'une étape de fractionnement peptidique, sur phase inverse à pH basique, avant l'enrichissement a aussi été évaluée pour augmenter le nombre de phosphorylations détectées.

Au niveau de l'analyse LC-MS/MS, deux systèmes ont été comparés. Il s'agit de deux configurations de spectromètres de masse différentes : Q-Exactive HF-X (Thermo, Q-Orbitrap) et timsTOF Pro (Bruker, Q-TOF), qui possède une dimension de séparation supplémentaire (la mobilité ionique). Une méthode de fragmentation alternative à la dissociation par collision a été explorée : l'ETD avec un instrument Eclipse Tribrid (Thermo).

Enfin pour l'identification et la quantification des phosphorylations, différents logiciels pour la recherche (Mascot (Matrix Sciences), Andromeda[147], MS Amanda[148]), la localisation et validation (Mascot, Andromeda[147], PhosphoRS[221]) et la quantification (Proline[157], MaxQuant[324], Proteome Discoverer (Thermo)) des événements de phosphorylation ont été confrontés.

Après développement de la méthode, elle a été appliquée à l'analyse d'échantillons de muscles d'ours prélevés durant les périodes d'hibernation et d'activité, afin de mieux comprendre les régulations moléculaires intervenant dans les muscles en réponse à l'hibernation. Outre le défi analytique que représente l'analyse des phosphorylations par spectrométrie de masse, l'interprétation biologique des variations de leur abondance reste compliquée chez certaines espèces (incluant l'ours) du fait de l'absence de bases de données d'annotations fonctionnelles. Pour pallier ce manque de données et faciliter l'interprétation des résultats qui seront obtenus, une stratégie bio-informatique adaptée a été développée pour pouvoir bénéficier des informations disponibles pour les protéines phosphorylées humaines.

L'ensemble des protocoles expérimentaux sont détaillés dans la partie expérimentale II.

III. Développement d'une méthode analytique pour l'analyse des phosphorylations dans le muscle

A. Sélection de la préparation d'échantillon la plus adaptée à l'analyse phosphoprotéomique du muscle

1. Optimisation des étapes d'extraction et de digestion des protéines à partir de tissus musculaires avant enrichissement

L'extraction des protéines à partir des tissus est une étape critique pour l'analyse protéomique où l'objectif est de lyser les cellules pour libérer les protéines qu'elles contiennent. Le tampon utilisé pour réaliser cette lyse doit également permettre de solubiliser un maximum de protéines sans les dégrader, les modifier ou introduire un biais de sur- ou sous-représentation d'une catégorie de protéines. Différents protocoles ont donc été comparés pour l'extraction et la digestion des protéines musculaires. Chaque extraction a été réalisée en triplicat à partir

d'environ 15 mg de muscle d'ours, un récapitulatif des protocoles utilisés est représenté en figure 47.

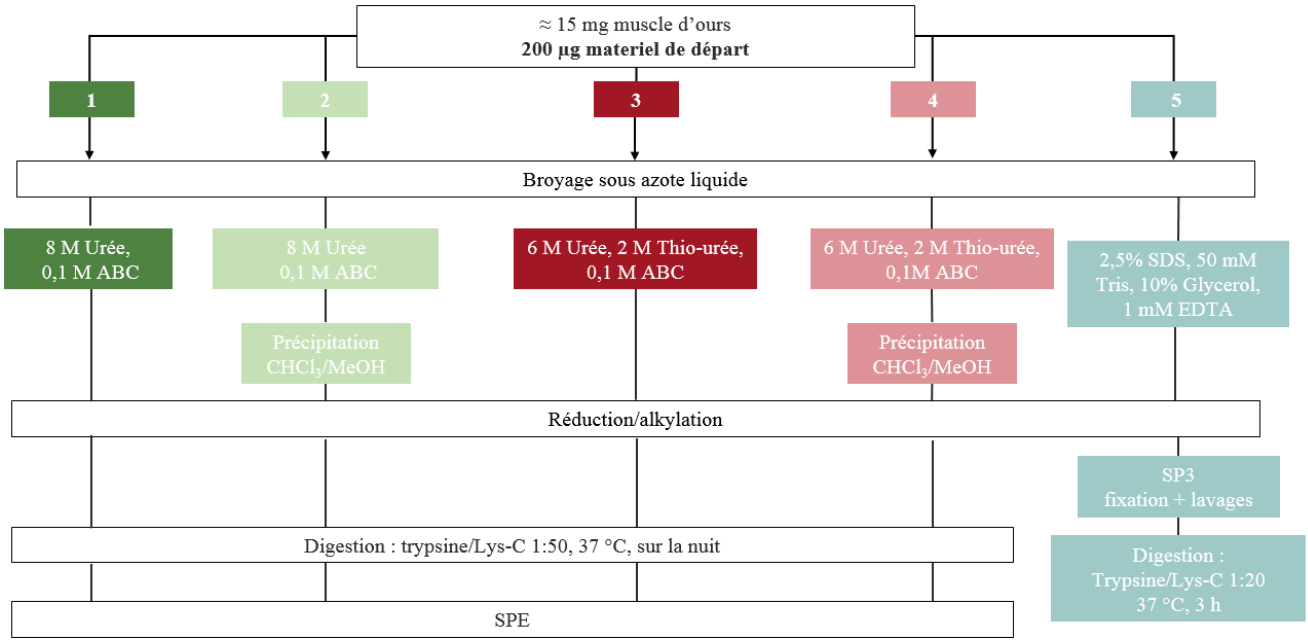

Figure 47 : Développement de la méthode pour l'obtention des peptides avant enrichissement. 5 conditions sont comparées : 1 = urée, 2 = urée avec précipitation, 3 = urée/thio-urée, 4 = urée/thio-urée avec précipitation et 5 = Laemmli-like. ABC = ammonium bicarbonate, $CHCl_3$= chloroforme, MeOH = méthanol, SDS= dodecyl sulfate de sodium, EDTA = acide éthylène éiamine Tétra-acétique.

Trois tampons d'extractions couramment décrits dans littérature pour extraire les protéines de ce type d'échantillon ont été utilisés : une solution d'urée 8 M tamponnée avec 0,1 M de bicarbonate d'ammonium [108], une solution d'urée 6 M et thio-urée 2 M tamponnée avec 0,1 M bicarbonate d'ammonium [323] et enfin un tampon dérivé du tampon de Laemmli contenant du SDS 2,5% et tamponné avec 50 mM de Tris-HCl (Laemmli-like). Chaque tampon contient également des inhibiteurs de protéases et de phosphatases pour limiter la dégradation des protéines et des phosphorylations. Pour les deux tampons contenant de l'urée, une extraction suivie par une précipitation au chloroforme et au méthanol ($CHCl_3$/MeOH) est réalisée. Cette étape permet de retirer les impuretés (sels, polysaccharides, acides nucléiques ou lipides) qui pourraient interférer avec la digestion enzymatique et l'analyse LC-MS/MS. Les rendements d'extraction calculés à partir du ratio de la quantité de protéines déterminée par dosage protéique (RC-DC, BioRad) sur la masse de tissu musculaire initial sont représentés en figure 48.

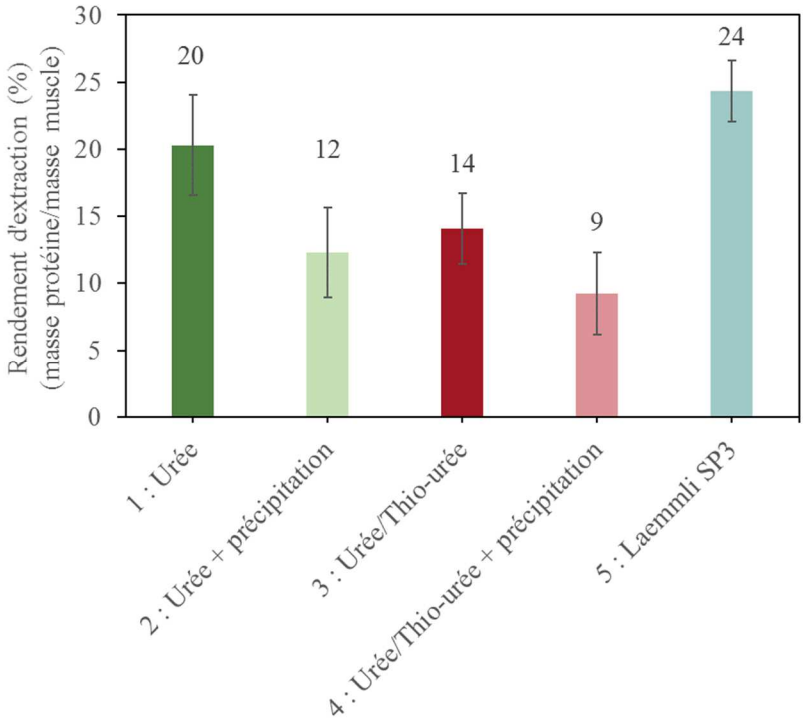

Figure 48 : Rendement d'extraction (en %) calculé à partir de la masse du tissu initial et de la quantité de protéines mesurée par dosage protéique pour chacune des extractions : 1 = Urée, 2 = Urée avec précipitation, 3 = Urée/thio-urée, 4 = Urée/thio-urée avec précipitation et 5 = Laemmli-like (SP3).

Les rendements d'extraction, calculés à partir de la quantité de protéines mesurées dans les échantillons, les plus faibles sont observés pour les conditions comportant une précipitation (conditions 2 et 4). La perte de matériel lors d'une étape de précipitation est décrite dans la littérature, mais cette étape peut être utile pour éliminer des contaminants qui pourraient entrer en compétition avec les peptides lors de l'analyse MS[6]. Par ailleurs, un meilleur rendement est obtenu avec l'extraction contenant du SDS (condition 5) par rapport à celles contenant de l'urée (conditions 1 et 3). Le SDS étant un détergent contrairement à l'urée qui est un agent chaotropique, il est possible qu'il facilite la mise en solution d'un plus grand nombre de protéines. Entre les deux tampons contenant de l'urée, un meilleur rendement est obtenu avec un fort pourcentage en urée sans thio-urée (condition 1).

Les extraits ont ensuite été digérés directement à la trypsine, à l'exception de l'extrait « Laemmli-like » pour lequel une méthode compatible avec la présence de SDS dans le tampon a été employée : la digestion SP3[40]. Pour les conditions utilisant la digestion liquide, une étape de SPE est nécessaire après digestion pour éliminer les sels (notamment l'urée). Cette étape est inutile pour la SP3 car des lavages sont réalisés avant l'ajout de la trypsine.

Les peptides obtenus ont été injectés sur un couplage nanoAcquity/Q-Exactive Plus (Thermo) puis les protéines et peptides ont été identifiés par Mascot et validés avec Proline. Le nombre de peptides et de protéines identifiés dans les différentes conditions sont illustrés en figure 49.

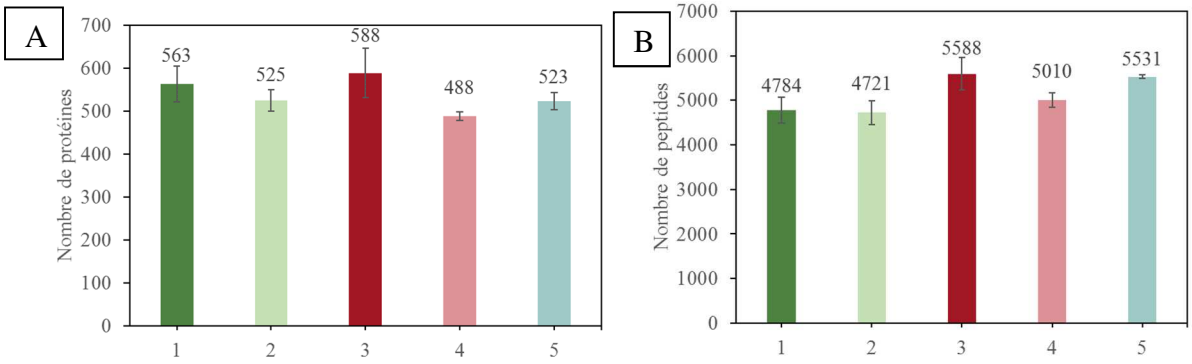

Figure 49 : Nombre moyen de protéines (A) et de peptides (B) identifiés par analyse LC-MS/MS à partir des différentes extractions : 1 = Urée, 2 = Urée avec précipitation, 3 = Urée/thio-urée, 4 = Urée/thio-urée avec précipitation et 5 = Laemmli-like.

On observe que contrairement aux résultats obtenus par le dosage protéique, le nombre d'identifications est maximal pour la condition 3 aussi bien au niveau peptidique (5 588 peptides) que protéique (588 protéines). Il semble donc que l'ajout de thio-urée permette la mise en solution ou l'extraction de protéines qui ne le sont pas en utilisant uniquement de l'urée. La condition 5 contenant du SDS, qui permettait d'extraire la plus grande quantité de protéines, conduit à l'identification d'un grand nombre de peptides mais reste limitée au final moins performante avec près de 500 protéines identifiées. Tout comme pour la quantité de protéines, le nombre de protéines et peptides identifiés après précipitation est plus faible, ce qui suggère donc que cette étape de précipitation induit une perte de matériel et n'est donc pas adaptée n'est à nos analyses.

Les valeurs quantitatives ont été extraites par Proline. Un filtre strict est appliqué pour la quantification: aucune valeur manquante dans les 3 réplicas. Le CV est calculé sur les intensités des 3 réplicas. La variabilité des différents protocoles est relativement équivalente (CV médian compris entre 16 et 22%). Les protocoles permettant la quantification du plus grand nombre de protéines sont les extractions à l'urée (conditions 1 et 3). Le tableau 3 résume les données de quantification obtenues pour chaque méthode d'extraction.

Tableau 3 : Nombre de protéines quantifiées et CV médian sur les intensités des protéines entre les 3 réplicas techniques.

Protocole	Nombre de protéines quantifiées	CV médian sur les intensités
1	472	16,0
2	387	17,9
3	401	22,0
4	374	18,7
5	454	17,5

Finalement, les conditions d'extraction 3 et 5 permettent l'identification d'un nombre maximal de peptides et c'est pourquoi elles ont été sélectionnées pour évaluer l'enrichissement en phosphopeptides.

2. Évaluation de l'étape d'enrichissement des phosphopeptides

150 µg de peptides dans une solution acide (0,1% TFA) avec 80% ACN sont déposés sur des cartouches IMAC contenant 5 µL de phase NTA-Fe(III). Les peptides retenus sont élués par augmentation du pH avec une solution à 1% NH_4OH. Les échantillons enrichis en phosphopeptides ont ensuite été injectés sur une NanoElute (Bruker) couplée à un timsTOF Pro (Bruker).

a. Efficacité de l'enrichissement

Pour évaluer l'efficacité du protocole d'enrichissement, des peptides phosphorylés synthétiques et marqués isotopiquement (heavy) ou non (light), peuvent être ajoutés au mélange peptidique. Ils sont appelés « phosphomix » et sont commercialisés par Thermo Fisher Scientific. Plusieurs mix existent et chacun contient 10 peptides synthétiques mono- ou bi-phosphorylés. Ils sont utilisés pour évaluer l'efficacité d'un protocole d'enrichissement. Ils sont assez peu utilisés dans la littérature où le rapport peptides phosphorylés identifiés sur peptides totaux est plus commun. Certains peptides issus de ce phosphomix sont difficiles à détecter comme décrit dans la fiche technique du produit (tableau 4).

Tableau 4 : Fiche technique du phosphomix 1 issue du site de Merck: séquence des peptides, masses associées, intensité du signal déclarée par le fournisseur.

Peptide	Nombre de groupement phosphate	Masse mono isotopique light	Masse mono isotopique heavy	Intensité relative du signal
VLHSG**S**	1	834.37	844.38	Faible
R**SY**SRS	2	1070.41	1080.41	Faible
RDSLG**T**YSS	1	1220.52	1230.53	Moyen
TKLTQLRDA	2	1445.70	1453.72	Fort
EVQAEQPSS**S**SP	1	1480.62	1490.63	Moyen
ADEPS**S**EESDLEID	1	1742.68	1750.69	Fort
ADEPS**S**EE**S**DLEID	2	1822.64	1830.66	Moyen
FEDEGAGFEES**S**ETGDYEE	1	2333.84	2341.85	Fort
ELSN**S**PLRENSFG**S**PLEF	2	2338.00	2348.01	Moyen
SPTEYHEPV**Y**ANPFYRPT**T**PQ	2	2809.19	2819.20	Fort

Les mêmes quantités de peptides *light* et de peptides *heavy* ont été ajoutés dans les échantillons avant et après l'enrichissement, respectivement. Le ratio des intensités des peptides *light* sur *heavy* correspond à l'efficacité de l'enrichissement. Dans une étude inter-laboratoires (*ProteoRed-ISCIII and EuPA launched the Proteomics Multicentric Experiment 11* [PME11]), les participants ont dopé des échantillons de lysat issus de levures avec ces peptides synthétiques (mix 1 et 2) puis ont procédé à l'enrichissement en phosphopeptides selon leur

protocole (IMAC ou MOAC)[325]. Cette étude a mis en évidence la variabilité des enrichissements qui existent entre les différents laboratoires. Dans une note d'application d'Agilent, les phosphomix 1 et 2 ont également été utilisés pour démontrer l'efficacité de l'enrichissement sur un Bravo AssayMap avec des cartouches de phase NTA-Fe(III)[198]. C'est la seule étude, à notre connaissance, qui décrit l'utilisation de cet enrichissement IMAC avec des échantillons contenant des phosphomix mais ils avaient été dopés dans des extraits cellulaires et non sur des extraits issus de tissus. À titre de comparaison, seuls 6 sur 10 peptides du phosphomix sont identifiés dans cette note d'application.

Les mélanges peptidiques issus de l'extraction urée/thio-urée ont été dopés avec le phosphomix 1. Les courants d'ions qui correspondent aux différents peptides du phosphomix ont ensuite été extraits avec Skyline[326] (v20.2.0). Seuls trois signaux sont exploitables : les signaux des peptides ADEPS**S**EE**S**DLEIDK, FEDEGAGFEES**S**ETGDYEEK, EVQAEQPSS**S**SPR. Les autres peptides ne sont pas détectés. Ces peptides font partie de ceux qui sont les plus fréquemment observés par des différents laboratoires dans l'étude PME11. Ils sont indiqués dans le tableau 5 ainsi que les données issues de la littérature (étude PME11 et note d'application Agilent).

Tableau 5 : Ratio *light/heavy* des 3 peptides du phosphomix observés et données issues de la littérature

Séquence des peptides du phosphomix	Ratio *light/heavy* moyen	Ratio *light/heavy* moyen dans la note d'application Agilent[198]	Fréquence d'observation dans l'étude PME11[325]
ADEPS**S**EE**S**DLEIDK	25 ± 5	95.1 ± 0.9%	65
FEDEGAGFEES**S**ETGDYEEK	26 ± 2	97.1 ± 4%	85
EVQAEQPSS**S**SPR	67 ± 16	82.5 ± 6.5%	100

On constate que l'efficacité de l'enrichissement est très faible et que le ratio est très dépendant du peptide considéré. Le peptide ADEPS**S**EE**S**DLEIDK est détecté, ce qui n'est pas le cas de ADEPS**S**EESDLEID qui possède la même séquence peptidique et qui devrait avoir un signal intense d'après la fiche produit, ce qui peut être dû au fait que les peptides portant plusieurs phosphorylations sont mieux retenus par la phase stationnaire IMAC[327]. Les ratios obtenus par notre protocole sont bien plus faibles que ce qui a été observé dans la note d'application Agilent. Ceci pourrait s'expliquer par le fait que les données issues de la note d'application ont été obtenues avec des lignées cellulaires, qui sont des échantillons moins complexes que les échantillons issus de tissus. De plus, les peptides synthétiques avaient été suivis en mode ciblé MRM, dopant ainsi la sensibilité du spectromètre de masse par rapport à ces peptides. De plus, au vu des ratios observés et du faible nombre de peptides identifiés sur nos échantillons, la qualité du lot utilisé peut être remise en question car il avait été préparé un ou deux ans plus tôt.

Une autre façon d'évaluer l'efficacité de l'enrichissement, plus communément utilisée par la communauté scientifique, est le rapport du nombre de peptides phosphorylés sur le nombre de peptides identifiés au total. Ce rapport est de 17 ± 2% pour l'extraction à l'urée/thio-urée et de

13 ± 2% pour l'extraction avec le tampon Laemmli-like. La valeur obtenue pour l'extraction à l'urée/thio-urée corrobore les ratios obtenus avec les deux premiers peptides du phosphomix et indique que le rendement de l'enrichissement est d'environ 20% pour l'extraction à l'urée. L'enrichissement est donc moins efficace sur l'extraction au Laemmli-like. L'écart-type restreint montre que l'efficacité de l'enrichissement est répétable entre les réplicas techniques.

b. Nombre d'identifications

Le nombre de protéines et de peptides phosphorylés sont présentés en figure 50.

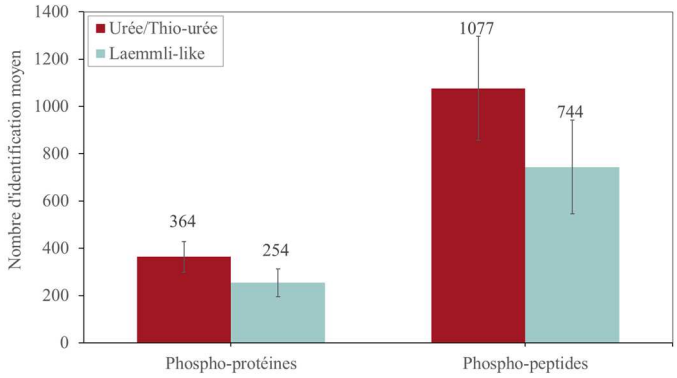

Figure 50 : Nombre d'identification moyen des phospho-protéines et des phospho-peptides pour les deux extractions considérées (Urée/Thio-urée et Laemmli-like).

Dans le muscle d'ours, 1077 phosphopeptides sur 364 protéines ont été identifiés en moyenne avec l'extraction à l'urée/thio-urée. Ces données sont difficilement comparables à la littérature (entre 3380 et plus de 15 000 phosphopeptides décrits pour ce type d'échantillon) car des stratégies de marquage isotopique (qui permettent d'augmenter le signal en MS1) et de fractionnement plus ou moins poussé avec un minimum de 12 fractions sont systématiquement utilisées[207, 322-323, 328]. Une autre étude, avec un fractionnement plus limité (6 fractions) couplé à un enrichissement TiO_2, identifie 1607 phosphopeptides dans du muscle de carpe[329]. Il n'existe, à ma connaissance, qu'une étude du phosphoprotéome musculaire sans fractionnement et qui serait comparable à nos expériences, où 3412 phosphopeptides ont été identifiés au total sur 10 échantillons avec une stratégie TMT 10-plex[206].

La figure 51 illustre les phosphoprotéines et les phosphopeptides communs aux deux extractions.

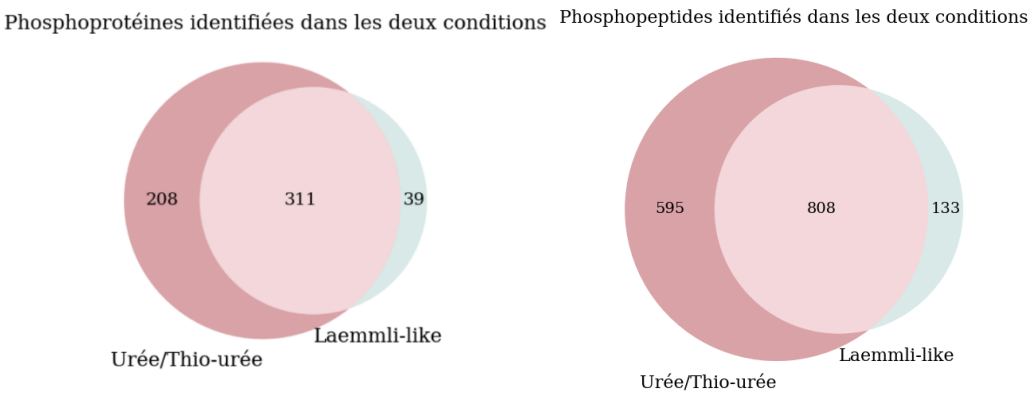

Figure 51 : Diagramme de Venn des phosphoprotéines et des phosphopeptides identifiés à partir des deux extractions réalisées : urée/thio-urée et Laemmli-like.

Près de 40% des protéines et des phosphopeptides sont identifiés uniquement par extraction urée/thio-urée, ce qui montre encore une fois la meilleure efficacité de celle-ci par rapport à l'extraction Laemmli-like.

a. Répétabilité de la préparation d'échantillon

Les peptides identifiés dans 3/3 réplicas techniques représentent environ 30% de tous les phosphopeptides observés, soit 451 peptides pour l'extraction urée/thio-urée et 339 peptides pour la deuxième condition, comme illustré sur la figure 52.

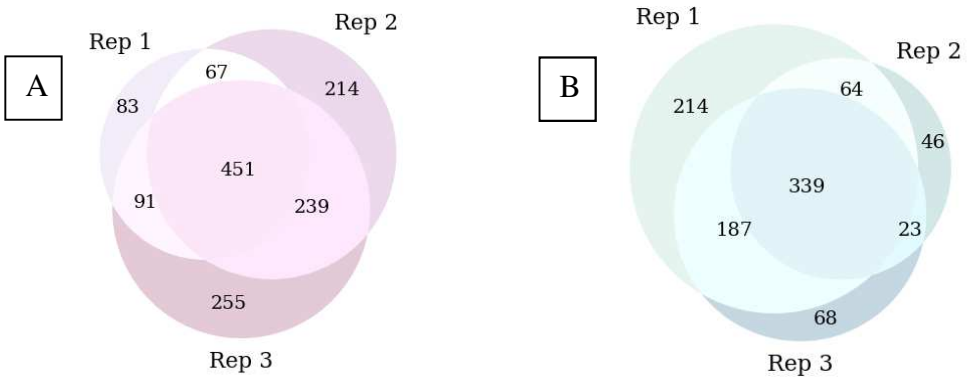

Figure 52 : Recouvrement des phosphopeptides entre les 3 réplicas techniques des extractions urée/thio-urée (A) ou Laemmli-like (B).

Le tampon utilisé pour l'extraction des protéines ne semble pas avoir d'influence sur la répétabilité du protocole. Par ailleurs, le pourcentage de peptides identifiés dans les 3 réplicas est faible (30%). Il faut cependant garder en mémoire que dans le cadre d'une analyse protéomique « classique », il a été observé au laboratoire qu'en moyenne seuls 50% des peptides sont communs à trois réplicas d'injection du fait de la stochasticité de l'analyse en mode DDA. Dans une analyse phosphoprotéomique, cette variabilité est encore exacerbée par la variabilité induite par l'étape supplémentaire d'enrichissement et la labilité de la modification. En effet, la modification peut être perdue au cours de la fragmentation de

l'analyse LC-MS/MS, les fragments seront alors peu informatifs car le spectre est dominé par l'ion précurseur qui a perdu H_3PO_4.

b. Répétabilité des analyses LC-MS/MS

Afin de tester la répétabilité de l'analyse LC-MS/MS, des réplicas d'injections ont été réalisés (N=3) durant la séquence d'analyse. Le recouvrement des phosphopeptides identifiés est illustré en figure 53.

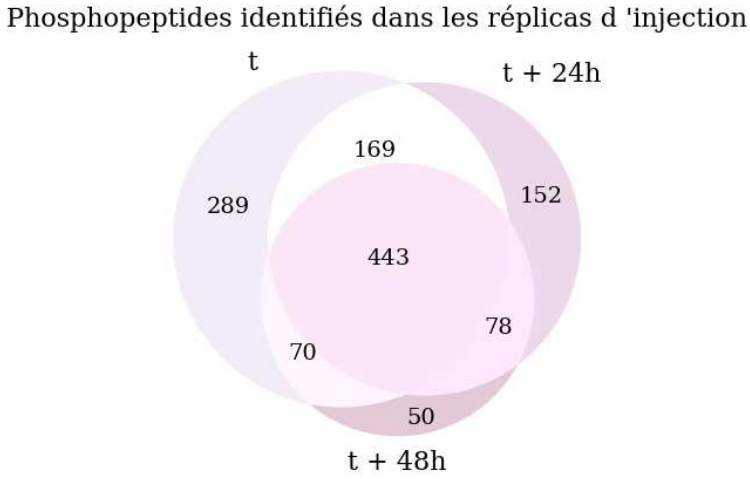

Figure 53 : Recouvrement des phosphopeptides entre les 3 réplicas d'injection.

La deuxième injection de l'échantillon a eu lieu 24 h après la première et la troisième 2 jours après la première. Le nombre de phosphopeptides identifiés diminue au cours du temps (-33% entre le premier et le dernier réplica). Cette diminution peut être due soit à la dégradation thermique des phosphopeptides, soit à leur adsorption sur les parois du vial. Afin de prévenir cette dégradation/adsorption à l'avenir, les échantillons seront congelés puis injectés par tranche de 24 h maximum.

c. Répartition des phosphorylations sur les acides aminés

La répartition des phosphorylations sur les acides aminés Sérine (S), thréonine (T) et tyrosine (Y) est illustrée pour les deux jeux de données en figure 54. La majorité des phosphorylations ont été observées sur les sérines (75%), viennent ensuite les thréonines avec 20% puis les tyrosines (5%). Cette répartition est comparable aux proportions décrites dans la littérature[187].

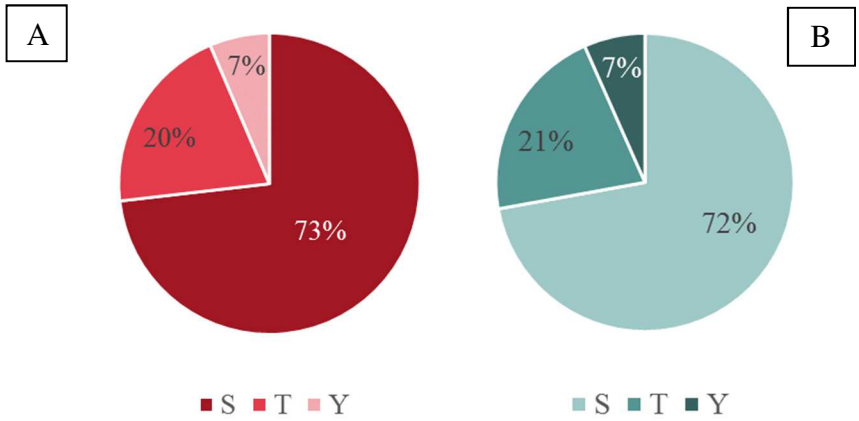

Figure 54 : Proportion des phosphorylations sur les acides aminés sérines (S), thréonines (T) et tyrosines (Y) pour les extractions Urée/thio-urée (A) ou Laemmli-like (B).

En conclusion, l'extraction à l'urée/thio-urée (condition 2) apparait plus adaptée à cette étude puisqu'elle permet un enrichissement plus efficace (rendement plus élevé) conjointement avec une augmentation du nombre d'identification d'événements de phosphorylation par rapport à l'extraction avec le tampon « Laemmli-like ».

B. Choix de la plateforme nanoLC-MS/MS adaptée à l'analyse des phosphopeptides

Nous avons voulu comparer un instrument de type Orbitrap bien établi au laboratoire et dans la littérature pour l'analyse des phosphorylations avec un spectromètre de masse de dernière génération équipé d'une cellule de mobilité ionique : un timsTOF Pro pour lequel, nous n'avons pas de recul pour l'analyse des phosphorylations. Au début de ma thèse, rien n'avait été décrit dans la littérature pour ce type d'analyse avec cet instrument.

Les échantillons issus de l'extraction urée/thio-urée enrichis en phosphopeptides ont été injectés sur le timsTOF Pro pour évaluer ses performances et les comparer aux résultats obtenus précédemment avec le Q-Exactive HF-X. Les deux spectromètres de masses se distinguent sur de nombreux points. Premièrement, le type d'analyseur est dans un cas un *Time Of Flight* [TOF] et dans l'autre une Orbitrap. De plus, le timsTOF Pro possède une cellule de mobilité avec un fonctionnement unique permettant d'augmenter grandement la sensibilité de l'appareil, grâce à son mode PASEF qui synchronise la séparation par mobilité ionique et l'élution des ions précurseurs avec le quadripôle. Les deux plateformes utilisent des modes de fragmentation différents : l'énergie apportée lors de la collision est plus élevée dans la cellule HCD du Q-Exactive HF-X que dans la cellule CID du timsTOF Pro. Cette énergie supplémentaire apportée lors la collision permet dans certains cas une meilleure fragmentation des peptides phosphorylés en réduisant les pertes du groupement phosphate et des spectres plus informatifs sont donc obtenus[174].

Le nombre d'identifications obtenues est d'environ 10% plus élevé sur le Q-Exactive HF-X par rapport au timsTOF Pro, comme indiqué en figure 55.

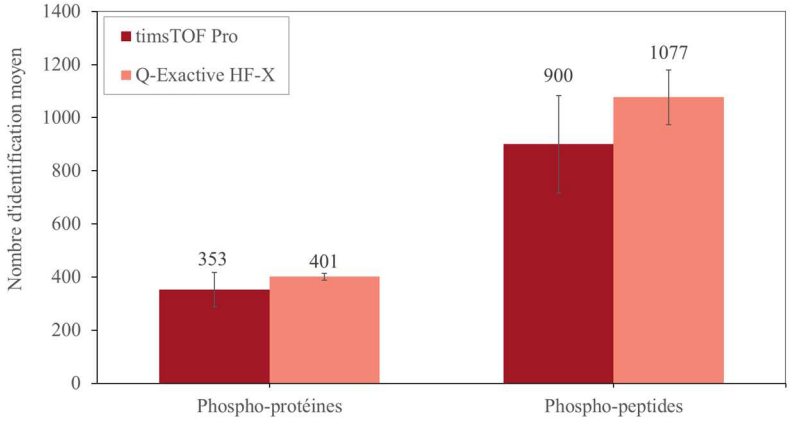

Figure 55 : Nombre moyen de phosphoprotéines et de phosphopeptides identifiés sur les deux instruments.

Le nombre de phosphopeptides observés dans les 3 réplicas sur les deux plateformes est illustré en figure 56.

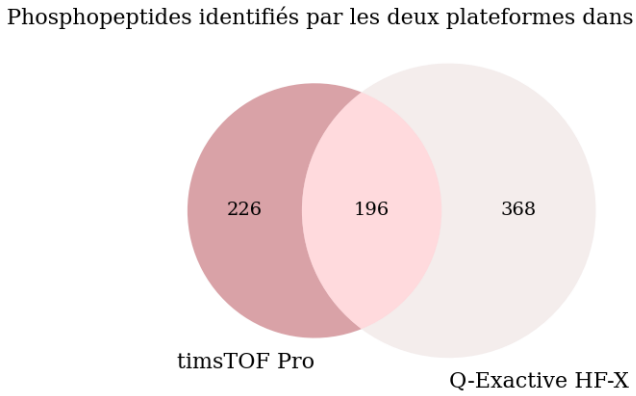

Figure 56 : Nombre de phosphopeptides identifiés dans 3/3 réplicas par les deux instruments.

Il est intéressant de noter que seuls 24% des phosphopeptides avec une identification robuste (observés dans tous les réplicas) sont identifiés par les deux plateformes. Un plus grand nombre de phosphopeptides est identifié dans tous les réplicas avec le Q-Exactive HF-X (564 vs 422 valeurs avec le timsTOF Pro). Cela montre que les identifications effectuées sur le spectromètre de type Orbitrap sont plus répétables que celles du timsTOF Pro. Ces résultats sont en accord avec une fragmentation plus efficace par la cellule HCD[202]. Les paramètres d'acquisition du timsTOF, notamment la fragmentation dans la cellule CID, Pro nécessiteraient des optimisations pour tirer entièrement parti de la dimension supplémentaire de mobilité ionique et du mode PASEF.

Les CV médians sur l'intensité brute des phosphopeptides identifiés dans 3/3 réplicas ont été calculés : 28% pour les données issues du Q-Exactive HF-X (564 valeurs) et 36% pour les données issues du timsTOF Pro (422 valeurs). La robustesse de l'analyse sur le Q-Exactive HF-X se retrouve également au niveau de l'intensité des peptides phosphorylés observés.

En conclusion, l'analyse par Orbitrap avec une fragmentation HCD se traduit par une robustesse accrue par rapport au timsTOF Pro. Les deux plateformes apparaissent complémentaires puisque les populations de phosphopeptides identifiées sont très différentes. Contrairement au Q-Exactive HF-X, le timsTOF Pro est une nouvelle instrumentation pour ce type d'analyse. De fait, des optimisations supplémentaires, notamment au niveau de la fragmentation des peptides, sont nécessaires et devraient permettre de gagner en reproductibilité et en identifications. Ces optimisations seront réalisées dans un second temps.

C. Évaluation de plusieurs logiciels de traitement de données pour l'identification et la quantification des phosphorylations

L'analyse des données obtenues est une étape clé dans une analyse protéomique et c'est d'autant plus le cas lors de l'étude des phosphorylations. En effet, outre l'identification des peptides et protéines grâce aux spectres de fragmentation, le logiciel doit être capable de localiser la ou les positions qui portent une phosphorylation. Pour cela, il doit être capable d'attribuer un score à cette localisation reflétant la probabilité qu'elle soit correcte. Différents logiciels de traitement de données ont donc été comparés. Les combinaisons étudiées sont récapitulées en tableau 6.

Tableau 6 : Combinaisons d'algorithmes de recherche, validation et localisation étudiées.

	Proline TD	MaxQuant	Mascot PD	Mascot + MS Amanda PD
Recherche	Mascot	Andromeda	Mascot	Mascot + MS Amanda
Validation	Target Decoy Proline	Target Decoy MaxQuant	Percolator PD	Percolator PD
Localisation	MD-Score	PTM-score	PhosphoRS	PhosphoRS

Les données utilisées ici sont celles obtenues par l'analyse des triplicats d'extraction à l'urée/thio-urée sur le Q-Exactive HF-X.

1. Définition d'un site de phosphorylation par les différents logiciels de traitement

La définition du site de phosphorylation et la gestion des peptides multi-phosphorylés est propre à chaque logiciel Au sens généralement utilisé, un site de phosphorylation correspond à la localisation de l'acide aminé portant une phosphorylation dans la séquence peptidique. La notion de « sites » est très utile puisque plusieurs peptides peuvent porter le même site de phosphorylation. C'est le cas par exemple d'un peptide avec une coupure manquée ou une autre modification comme illustré sur la figure 57.

Figure 57 : Illustration de la notion de site de phosphorylation sur quatre peptides portant deux sites de phosphorylations (P), les deux premiers peptides ont la même séquence en AA mais le deuxième peptide porte également une oxydation (O), le quatrième peptide est mal clivé.

La régulation des différentes voies de signalisation peut être modulée selon la localisation d'une phosphorylation, et parfois même par la coexistence de plusieurs événements de phosphorylation sur une même protéine.[180] On peut se retrouver avec plusieurs sites phosphorylés sur le même peptide (notion de multiplicité), et il est donc important de pouvoir accéder à ce type d'information.

MaxQuant, Proline et Proteome Discoverer permettent l'identification des sites de phosphorylations : Proteome Discoverer au travers de son onglet « *Modification Sites* », Proline par ses onglets « *modification sites* » ou « *modification cluster* » et MaxQuant grâce au fichier « PhosphoSites » qui peut ensuite être traité dans Perseus avec l'option « *expand site tables* ». Cependant la définition du site de phosphorylation varie en fonction du logiciel. La définition du site pour chaque outil informatique est illustrée en figure 58.

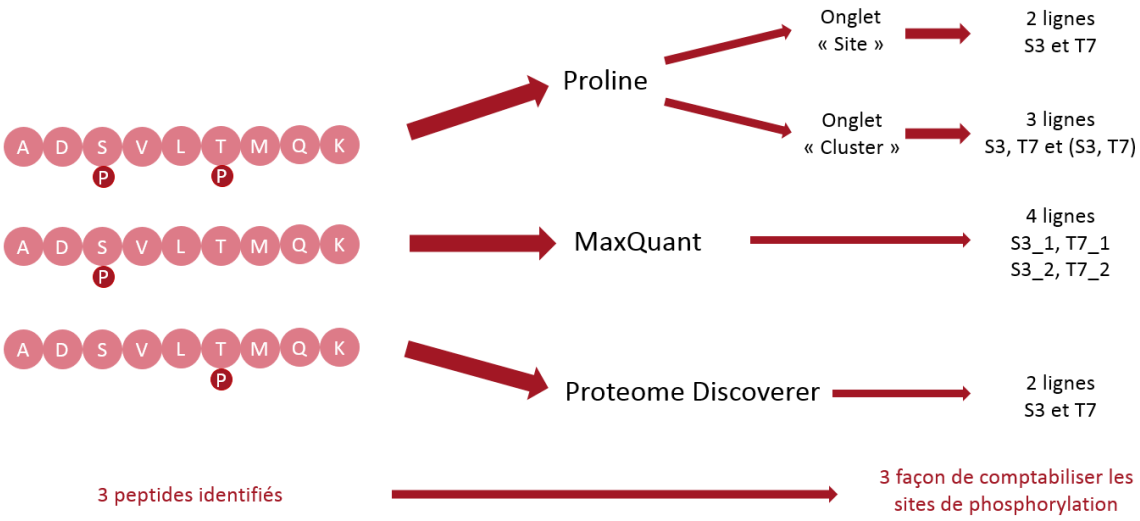

Figure 58 : Illustration de la définition des sites de phosphorylation par les différents logiciels sur 3 peptides de même séquence portant 2 sites de phosphorylation différents. Proline (onglet « site ») et protéome Discoverer comptent le nombre de séquence peptidique identique. Proline compte 3 sites de phosphorylation car il y a trois peptides différents. MaxQuant compte quatre sites car la sérine phosphorylée peut appartenir soit à un peptide mono soit à un peptide bi-phosphorylé et ces deux situations donneront lieu à deux lignes, il en va de même pour la thréonine.

La façon dont sont traités les peptides multi-phosphorylés est propre à chaque logiciel. La multiplicité, c'est-à-dire le fait qu'un peptide soit mono-, bi- ou tri-phosphorylé, est conservée par le logiciel MaxQuant, ce qui entraine une certaine redondance de l'information. Proteome Discoverer fait le choix d'éliminer cette redondance et ne donne donc pas accès à cette information. Proline permet de choisir si l'on souhaite conserver ou non cette information en choisissant soit l'onglet « Site » soit l'onglet « Cluster ». Pour comparer les résultats obtenus au niveau des sites de phosphorylation sur notre jeu de données, le plus simple est d'éliminer la multiplicité et donc la redondance dans MaxQuant pour se ramener à la définition de Proteome Discoverer. Cependant sur de véritables échantillons, la possibilité de conserver la multiplicité pourrait être un critère de choix entre les différents logiciels.

2. Nombre d'identifications obtenus par les différents logiciels

La gestion des peptides partagés par plusieurs protéines est également différente entre ces logiciels. Proline et Proteome Discoverer associent les peptides/sites phosphorylés aux différentes protéines produisant ainsi une redondance inutile et un biais d'analyse puisque pour un PSM donné, plusieurs sites peuvent être identifiés dans différentes protéines. Ainsi le peptide ADIAESQVNKLR est associé à 8 protéines différentes (différentes isoformes de myosines), voir figure 59. MaxQuant au contraire regroupe les protéines potentielles dans un « *protein group* ». Pour pallier ce problème, j'ai développé un script en utilisant le langage de programmation Python 3.8 permettant d'éliminer la redondance observée dans Proline et Proteome Discoverer en attribuant le peptide phosphorylé à la protéine ayant le score Mascot/MSAmanda le plus élevé.

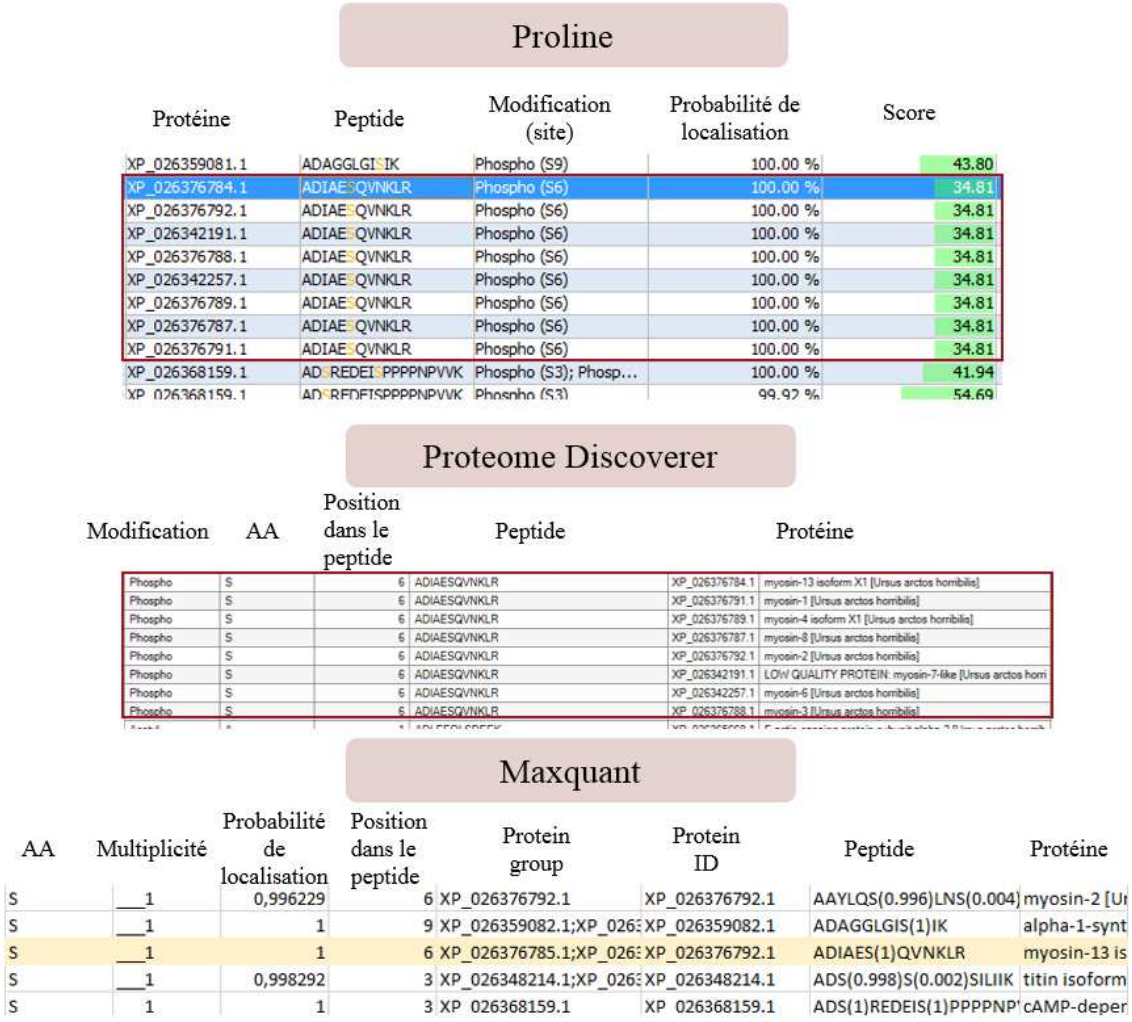

Figure 59 : Gestion des peptides partagés par les différents logiciels : sur Proline et Proteome Discoverer, un peptide/PSM est attribué à 8 protéines différentes tandis que sur MaxQuant, ces protéines sont regroupées dans le « *protein group* ».

Le décompte des identifications après élimination de la redondance est illustré en figure 60.

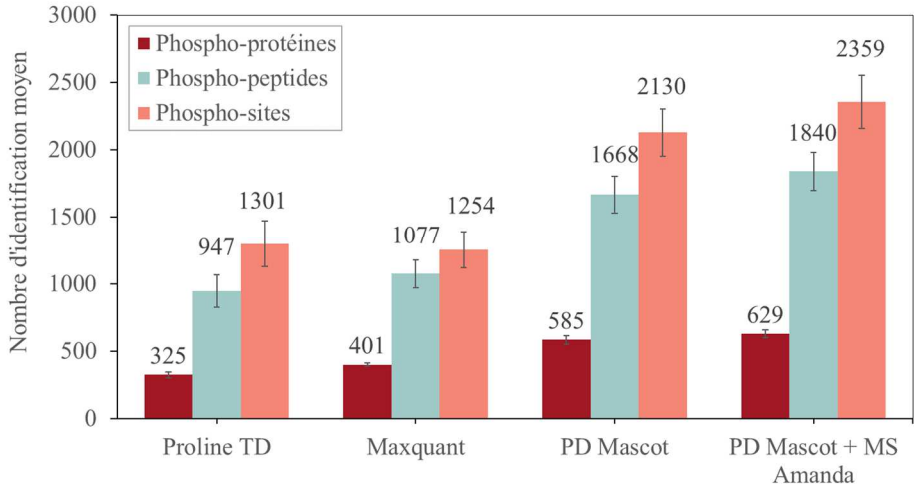

Figure 60 : Nombre moyen d'identifications en fonction des différents traitements de données.

Les 4 traitements de données identifient des nombres très différents de protéines, peptides et sites phosphorylés. Proline TD et Mascot PD partagent le même moteur de recherche Mascot, une augmentation de plus de 50% est pourtant observée grâce à l'utilisation de Percolator. Percolator est un algorithme permettant la validation des PSMs grâce à un apprentissage semi-supervisé (*machine learning*)[158]. Cette augmentation liée à l'utilisation de Percolator par rapport à une approche *target-decoy* avait également été observée lors de l'étude de Locard-Paulet *et al*. en 2020[222].

Plusieurs paramètres influent sur les identifications. Avec les mêmes critères de validation, la combinaison de MS Amanda et Mascot entraine une augmentation de 10% du nombre de phosphopeptides identifiés par rapport à Mascot seul. L'algorithme utilisé pour l'identification des peptides joue donc un rôle. Cependant, après vérification manuelle des spectres uniquement identifiés avec MS Amanda, il apparait que ces spectres sont souvent peu informatifs ou de mauvaise qualité comme illustré sur la figure 61.

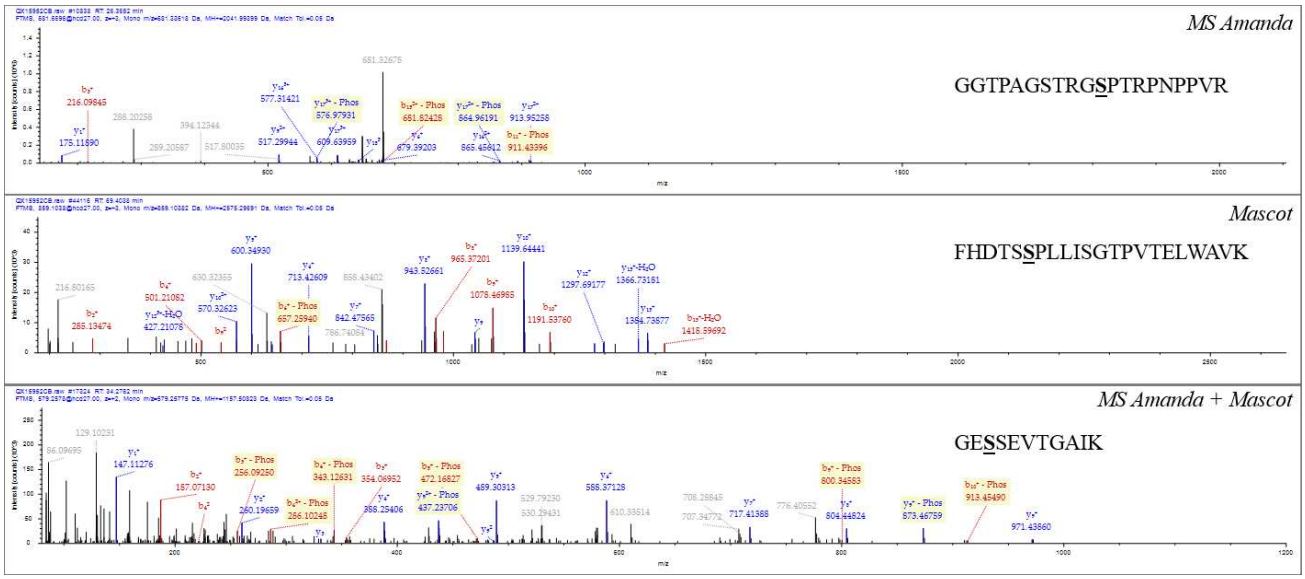

Figure 61 : Spectres de peptides phosphorylés identifiés par MS Amanda seul, Mascot seul et identifiés par les deux logiciels.

3. Localisation des sites de phosphorylation

Les logiciels considérés utilisent des algorithmes d'attribution du score de localisation différents : Mascot MD-Score, MaxQuant PTM-Score et Proteome Discoverer PhosphoRS. Ces différents algorithmes permettent de calculer la probabilité de localisation de la phosphorylation sur le peptide dans le cas où plusieurs sites peuvent accueillir une modification, dans le cas contraire la probabilité est de 1. Classiquement, le seuil accepté sur le score donné par MaxQuant est de 0,75. Les résultats obtenus avec les différents algorithmes, illustrés en figure 62, sont très différents les uns par rapport aux autres.

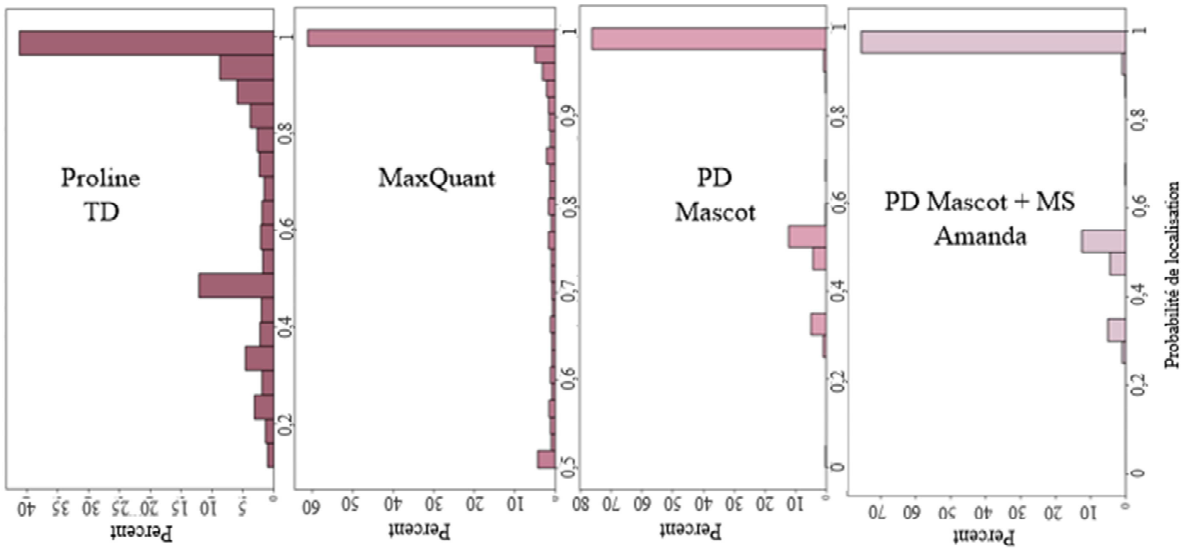

Figure 62 : Fréquence (en %) des scores de localisation des sites de phosphorylations des différents traitements de données.

MD-Score (Proline) est l'algorithme qui localise le plus de sites de phosphorylation avec une faible probabilité. Pour PTM-Score de MaxQuant, au contraire, plus de 60% des sites ont une probabilité de localisation proche de 1. Le profil obtenu avec PhosphoRS est particulier, une très forte proportion de sites sont identifiés avec certitude, près de 80% des sites sont localisés avec une probabilité proche de 1. Les autres sites sont localisés avec une probabilité de 0,5 et 0,3 correspondant à deux ou 3 sites de localisations probables. Ces observations rejoignent celles décrites en 2020 par Locard-Paulet *et al.*[222].

L'index de Jaccard permet d'évaluer la ressemblance entre deux jeux de données. Il est utilisé dans la publication de Locard-Paulet *et al*. L'indice de Jaccard est alors défini comme le rapport entre le nombre d'identifiants communs aux deux traitements sur le nombre total d'identifiants[222]. Plus l'indice est proche de 1, plus la ressemblance entre les jeux de données est forte. Les indices de Jaccard calculés sur nos données sont représentés sur la figure 63.

Proline TD	Maxquant	Mascot PD	Mascot + MS Amanda PD	
	49	51	46	Proline TD
		47	44	Maxquant
			87	Mascot PD
				Mascot + MS Amanda PD

Figure 63 : Indices de Jaccard entre les différents traitements.

L'utilisation du même logiciel pour la recherche (identification) aboutit à des jeux de données très similaires comme c'est le cas entre Mascot PD et Mascot + MS Amanda PD (indice de Jaccard = 87). Pour les autres jeux de données, environ la moitié des peptides phosphorylés sont communs entre des différents traitements. Des résultats similaires sont obtenus dans l'étude de Locard-Paulet *et al.*, il faut cependant noter que les comparaisons entre MaxQuant et Protéome Discoverer identifiaient plus de peptides communs (65%)[222].

À ce stade, on ne peut pas déterminer avec certitude le logiciel à utiliser et chaque traitement présente des avantages et des inconvénients, qui sont récapitulés dans le tableau 7.

Tableau 7 : Principaux avantages et inconvénients des différents traitements comparés.

	Proline	MaxQuant	Mascot PD	Mascot + MS Amanda PD
Avantage	Possibilité de choisir de travailler sur les sites avec ou sans multiplicité	Possibilité de travailler sur les sites en conservant l'information de multiplicité	- Plus d'identifications grâce à l'utilisation de Percolator - Bon *scoring*	-Les deux moteurs de recherches permettent d'identifier plus de phosphopeptides - Bon *scoring*
Inconvénient	- Algorithme de *scoring* peu efficace - Validation trop strict (TD)	Le *scoring* des phosphorylations n'est peut-être pas assez strict	Impossible de conserver l'information multiplicité au niveau du site	Les identifications supplémentaires semble provenir de spectres de mauvaise qualité

4. Quantification des phosphorylations

a. Biais de la quantification

Comme vu précédemment, la notion de site est cruciale dans les analyses phosphoprotéomiques. C'est d'autant plus le cas pour les analyses quantitatives car si la quantification est effectuée sur les peptides et non sur les sites, un biais est introduit. En effet, les coupures manquées ou des modifications variables peuvent générer des peptides différents qui risquent d'être quantifiés séparément alors qu'ils reflètent le même événement de phosphorylation. Malheureusement, si Proline et Proteome Discoverer permettent l'identification des sites de phosphorylation, ils ne permettent pas d'extraire les valeurs quantitatives associées. À ce stade, MaxQuant est le seul logiciel à proposer une quantification au niveau du site en sommant les intensités des peptides présentant une coupure manquée ou une oxydation. Cette quantification n'est pas exempte de défauts puisqu'un même peptide multi-phosphorylé va aboutir à deux valeurs de quantification, ce qui peut fausser les statistiques qui seront réalisées par la suite.

Dans la suite de notre étude, nous avons donc choisi MaxQuant pour l'identification, la validation, la localisation et la quantification des sites de phosphorylation.

b. Mise en place de filtres de qualité

Afin de s'assurer de la qualité des identifications retenues pour la quantification, un certain nombre de filtres doivent être mis en place. Ces filtres dépendent grandement de l'objectif final de l'étude considérée et du type d'échantillon considéré. Dans le cas présent, l'objectif est d'identifier les voies biologiques impactées lors de l'hibernation de l'ours par rapport au statut actif. Il faut donc trouver un compromis acceptable afin de ne pas être trop strict pour ne pas perdre certaines infos mais tout de même assurer la validité des hypothèses proposées en obtenant des informations de qualité. Tout d'abord, afin d'assurer l'appartenance du phosphopeptide à la protéine auquel il est rattaché, seuls les peptides uniques au « *protein group* » sont retenus.

Une notion similaire au FDR existe pour les modifications post-traductionnelles des protéines, appelée FLR pour *False Localization Rate* (voir partie I chapitre 2), elle permet d'estimer le nombre de mauvaises localisations des modifications. Malheureusement, à ce jour il n'existe pas de leurres qui permettent d'assurer d'estimer le FLR dans des analyses phosphoprotéomiques à large échelle. Pour l'instant, le calcul du FLR ne peut être réalisé que sur un jeu de données de peptides synthétiques pour lesquels les localisations des phosphorylations sont connues. Sur la base d'un tel calcul, Locard *et al.* ont montré que le score de PhosphoRS reflète plus la réalité que le PTM-score de MaxQuant, c'est pourquoi, ils suggèrent d'appliquer un filtre plus strict sur le PTM-score que celui qui est généralement utilisé (0,9 contre 0,75). L'augmentation de ce seuil permettrait de ramener le FLR à un taux comparable (1%) à celui obtenu avec une localisation minimale de 0,75 en utilisant PhosphoRS pour attribuer le score de localisation. Cependant, la plupart des études phosphoprotéomiques qui utilisent MaxQuant se concentrent sur les phosphosites avec probabilité de localisation supérieure à 0,75 (souvent appelées phosphosites de classe I)[181].

Le projet sur lequel je travaille est une étude très exploratoire et des analyses complémentaires devront être menées pour appuyer les conclusions biologiques. C'est pourquoi, il est acceptable de ne pas appliquer des critères trop stricts et un seuil à 0,75 sera appliqué sur la probabilité de localisation.

Enfin, pour augmenter la robustesse de la quantification et limiter le nombre de valeurs manquantes, un filtre sur le taux d'observation des protéines dans les réplicas biologiques est couramment appliqué en protéomique. Pour la quantification au niveau des protéines, plusieurs peptides sont pris en compte, tandis que pour la quantification des PTMs, la quantification ne se fait que sur un peptide. La probabilité d'avoir une valeur pour chaque échantillon dans le cadre de l'étude des PTMs est donc plus faible que pour les protéines. C'est pourquoi, il est acceptable d'appliquer des filtres moins stricts, par exemple en acceptant 50% de valeurs manquantes par groupe.

En résumé, pour l'analyse phosphoprotéomique du muscle d'ours, des filtres sur la probabilité de localisation (supérieure à 0,75) et sur le nombre de valeurs manquantes (maximum 50%) seront appliquées avant l'analyse statistiques des données quantitatives après n'avoir sélectionné que les peptides protéotypiques.

D. Optimisation des méthodes de fragmentation sur Eclipse Tribrid et timsTOF Pro

1. Comparaison des méthodes ETD et HCD sur un instrument de type Orbitrap

Dans de nombreuses publications, l'utilisation de méthodes de fragmentation alternatives à la fragmentation induite par collision est recommandée pour améliorer la localisation des sites de modifications et notamment des phosphorylations. En effet, la fragmentation par collision (CID ou HCD) induit souvent la perte du groupement phosphate et non la fragmentation des liaisons peptidiques. De fait, aucune information relative à la séquence du peptide phosphorylé et encore moins sur la position de la modification ne peut être obtenu.[202] Des méthodes de fragmentations plus douces permettent de conserver le groupement phosphate sur l'acide aminé modifié tout en fragmentant le peptide afin d'en connaitre la séquence. C'est le cas de la dissociation induite par transfert d'électron (ETD)[330]. Nous avons voulu tester ce mode de fragmentation disponible au laboratoire sur l'Orbitrap Eclipse Tribid (Thermo) équipé de techniques de fragmentation multiples: HCD, CID, ETD, EThcD et UVPD. Les échantillons ont donc été injectés et deux modes de fragmentation ont été comparés : ETD et HCD. Les données ont ensuite été traitées avec Proteome Discoverer.

Les spectres de fragmentation du peptide VIMR**T**SLEVK obtenus par ETD et par HCD sont illustrés en figure 64. Par fragmentation HCD, le pic le plus intense correspond à la masse du peptide non fragmenté et sans groupement phosphate (perte de - 98 Da). Très peu de fragments sont détectés. Tandis que sur le spectre obtenu par fragmentation ETD, de nombreux fragments sont identifiés qui permettent de reconstituer la séquence peptidique tout en localisant le groupement phosphate.

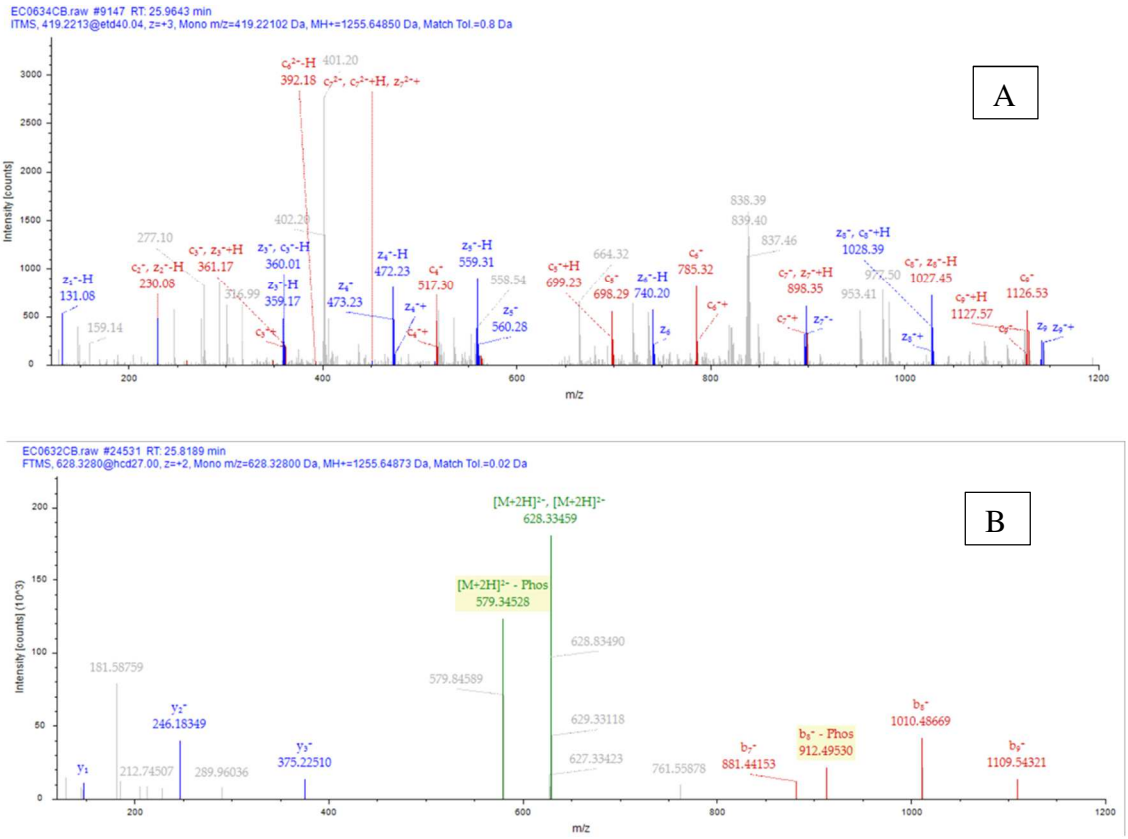

Figure 64 : Spectres du peptide VIMR**T**SLEVK obtenus par fragmentation ETD (A) et HCD (B).

La comparaison entre les phosphopeptides identifiés par chaque technique de fragmentation est présentée en figure 65. Malgré une identification facilitée, beaucoup plus de phosphopeptides sont identifiés par fragmentation HCD. L'ETD est une technique très lente car elle nécessite un temps de réaction long entre les peptides positivement chargés et l'anion fluoranthène. Ceci rallonge le temps de cycle de l'analyse MS et est difficilement compatible avec la largeur des pics chromatographiques. C'est pourquoi il est difficile d'identifier un grand nombre de phosphopeptides. Par ailleurs, ce ne sont pas les mêmes peptides qui sont observés par les deux techniques. Dans la littérature, il est également décrit que ces techniques sont complémentaires et permettent d'identifier des populations différentes de phosphopeptides[133].

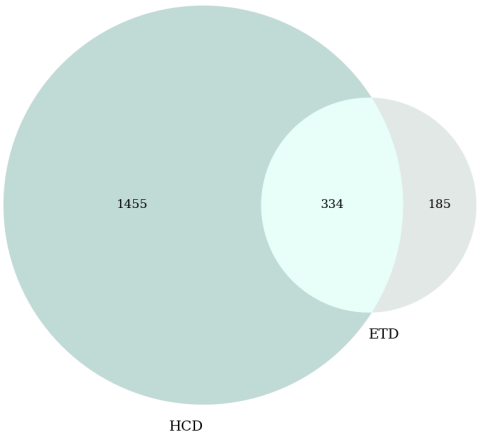

Figure 65 : Recouvrement entre les sites de classe I identifiés avec chacune des méthodes de fragmentation : HCD et ETD.

Bien que prometteuse en théorie pour l'identification des phosphorylations, la fragmentation par ETD reste trop lente pour une identification à large échelle des phosphopeptides.

2. Optimisation de l'énergie de collision sur le timsTOF Pro

Afin d'améliorer les performances du timsTOF Pro, une optimisation de l'énergie de collision a été effectuée sur des échantillons de cerveaux de souris enrichis en phosphopeptides. La valeur de mobilité ionique est fonction de la masse, de la charge et de la forme de l'ion considéré. Elle est aussi affectée par la présence de modifications comme les phosphorylations. Ogata *et al.* ont ainsi montré que le coefficient de mobilité réduit ($1/K_0$) des peptides phosphorylés différait lorsque le peptide portait une phosphorylation par rapport au peptide non modifié[130]. L'énergie de collision délivrée est fonction de la valeur de mobilité mesurée dans la cellule de mobilité ionique. Les échantillons ont été injectés avec la rampe de collision utilisée jusqu'à présent et une rampe adaptée aux phosphorylations, ces données sont tracées en figure 66. Les données ont ensuite été traitées avec MaxQuant.

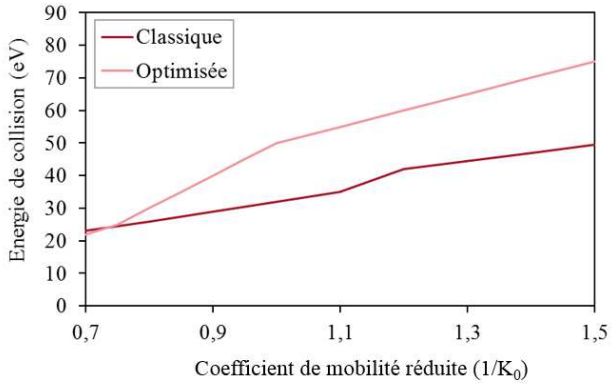

Figure 66 : Rampe de l'énergie de collision délivrée à l'ion en fonction de la valeur de mobilité ionique mesurée pour les deux méthodes considérées (classique et optimisée).

La méthode optimisée délivre plus d'énergie aux ions que la méthode classique.

Comme illustré sur la figure 67, près de 2000 phosphosites classe I supplémentaires sont identifiés avec la méthode optimisée.

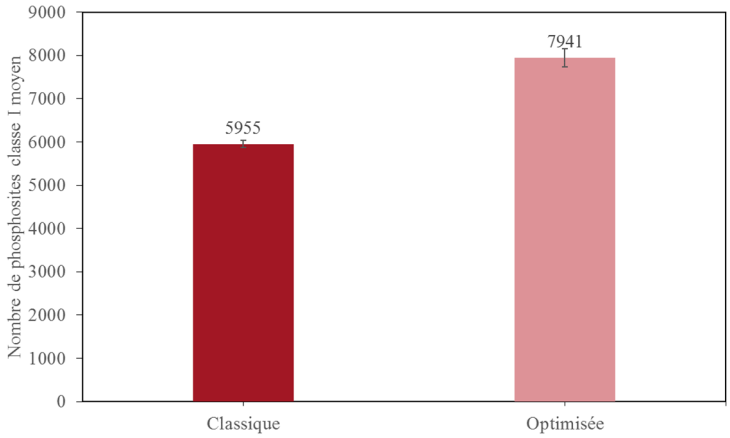

Figure 67 : Nombre de sites de phosphorylations de classe I moyen identifié avec les deux méthodes MS.

La fragmentation semble donc plus efficace avec la méthode optimisée. Étant donné que l'énergie fournie aux ions pour la fragmentation est plus importante que dans la méthode classique, il est possible que l'efficacité de la fragmentation se rapproche de celle de l'HCD qui délivre plus d'énergie que la fragmentation CID.

La longueur des peptides phosphorylés et le nombre de modifications par peptides sont représentés en figure 68. En comparant les populations de phosphopeptides obtenus avec les deux méthodes, on observe que les peptides identifiés par la méthode optimisée sont légèrement plus longs. Cette augmentation de la longueur semble également se traduire par un plus grand nombre de peptides multi-phosphorylés (25% contre 19%).

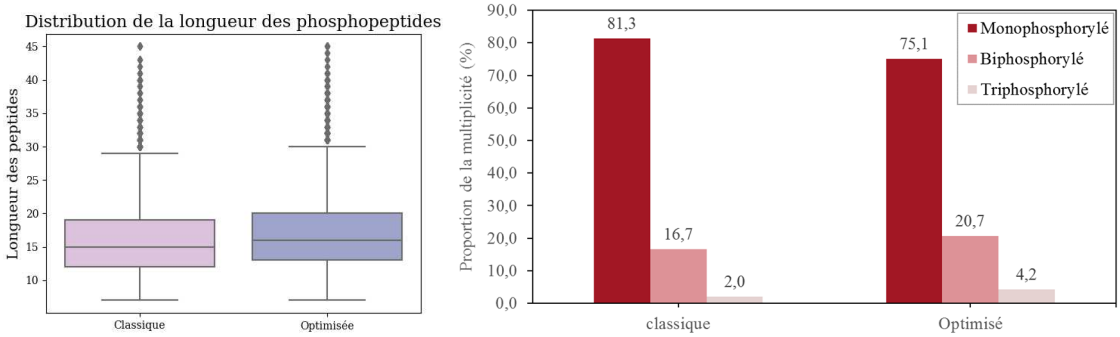

Figure 68 : Distribution de la longueur des peptides phosphorylés et du nombre de phosphorylations portés par peptides issus des identifications avec les deux méthodes MS.

Les phosphosites identifiés dans au moins 3/4 réplicas sont retenus pour évaluer la qualité de la quantification avec les deux méthodes MS, les résultats obtenus ainsi que le CV médian calculé sur l'intensité des phosphopeptides sont renseignés en tableau 8.

Tableau 8 : Nombres de sites de phosphorylation quantifiés et CV médian calculé sur l'intensité brutes des phosphosites entre les réplicas.

Méthode	CV médian sur l'intensité des phosphosites (%)	Nombre de phosphosites quantifiés
Classique	8,7	5233
Optimisée	11,8	6892

Près de 1700 sites de phosphorylations supplémentaires sont quantifiés avec la méthode optimisée. La répétabilité des deux méthodes est similaire puisque le CV médian de l'intensité des phosphosites sur les réplicas d'injection reste faible (11,8% et 8,7%).

Ces résultats montrent bien l'importance de l'optimisation des méthode MS pour l'analyse des phosphorylations. C'est pourquoi des développements sont actuellement en cours pour optimiser d'autres paramètres d'acquisition sur le timsTOF Pro tels que la gamme de mobilité ou les temps de d'accumulation et d'élution des ions dans les cellules TIMS ainsi que le gradient chromatographique utilisé.

E. L'apport d'une étape de fractionnement peptidique

Un fractionnement des échantillons avant l'enrichissement IMAC a été mis en place pour décomplexifier l'échantillon et ainsi vérifier si cela permettait d'augmenter la couverture de phosphoprotéome.

La capacité des cartouches IMAC (5 µL) est limitée à 150 µg de peptides, afin d'augmenter la quantité de départ et donc le nombre de phosphorylations identifiées, nous avons cherché à fractionner l'échantillon. Le fractionnement peptidique sur phase inverse à haut pH est le plus répandu ; il est généralement utilisé pour générer un grand nombre de fractions (>10). Mon objectif était de trouver un compromis entre garder un temps d'analyse raisonnable et augmenter la profondeur de l'analyse. L'apport d'une étape automatisée de fractionnement peptidique limitée à trois fractions a été donc évaluée. Le protocole utilisé est basé sur une étude de Murillo *et al.* utilisant des cartouches Reverse Phase S [RPS] pour un fractionnement hpH RP sur un Bravo AssayMAP à partir de tissus issus de tumeurs.[108] La figure 69 illustre le schéma analytique suivi.

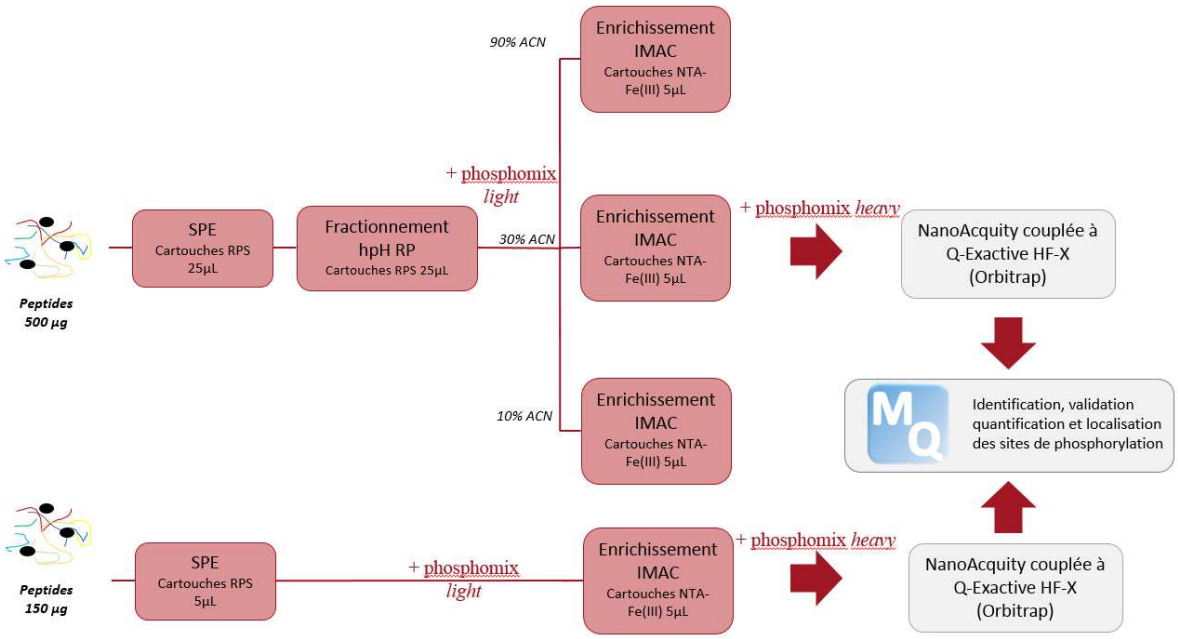

Figure 69 : Schéma analytique comprenant le fractionnement hpH RP suivi par l'enrichissement IMAC des fractions obtenues.

Pour obtenir suffisamment de matériel après fractionnement pour réaliser un enrichissement IMAC, une quantité plus importante de matériel de départ doit être utilisée. Les cartouches RPS qui sont utilisées pour le fractionnement hpH RP sur un BravoAssay MAP contiennent classiquement 5 µL de phase. Elles arrivent à saturation à partir de 200µg. Agilent a produit récemment des cartouches RPS à haute capacité. Ces cartouches contiennent 25 µL et auraient une limite théorique de saturation à 800µg. Il a donc fallu adapter le protocole de fractionnement utilisé dans l'étude de Murillo *et al.* à ces nouvelles cartouches pour lesquelles, rien n'était encore décrit.

Premièrement, une SPE est réalisée sur 500 µg de peptides avec des cartouches RPS (25 µL). Cette étape est nécessaire pour éliminer les sels qui pourraient impacter négativement la séparation des peptides lors du fractionnement. Pour limiter au maximum la perte d'échantillon, les même cartouches RPS sont ensuite réutilisées pour le fractionnement hpH RP. Les peptides sont élués séquentiellement avec 10, 30 et 90% d'acétonitrile. Les fractions obtenues contiennent théoriquement (si la répartition des peptides est équilibrée) 166 µg de peptides et sont ensuite enrichies sur cartouches IMAC 5 µL. En parallèle, un enrichissement dans les mêmes conditions que précédemment (sans fractionnement) est effectué. Les échantillons enrichis ont été injectés sur un Q-Exactive HF-X. Les données sont ensuite traitées avec MaxQuant.

Comme précédemment, les échantillons ont été dopés avec les peptides phosphorylés du phosphomix 1. Un plus grand nombre de peptides du phosphomix sont identifiés (5/10) par rapport aux résultats du paragraphe III.A.2, voir tableau 9, peut-être grâce à l'utilisation d'un nouveau lot.

Tableau 9 : Ratio *light/heavy* des phosphopeptides synthétiques observés du phosphomix 1 dans les échantillons fractionnés et non fractionnés.

Peptide phosphorylé	Ratio *light/heavy* des échantillons fractionnés	Ratio *light/heavy* des échantillons non fractionnés
RDSLG**T**YSSR	41 ± 10	56 ±3
TKLI**T**QLRDAK	42 ±10	60 ± 3
EVQAEQPSS**S**SPR	27 ±15	22 ±2
ADEP**S**SEESDLEIDK	45 ±10	76 ±9
SPTE**Y**HEPVYANPFYRPT**T**PQR	38 ±11	59 ±2

Le ratio *light/heavy* est plus faible dans les échantillons fractionnés (55) par rapport aux échantillons sans fractionnement (39). Il semble donc que l'enrichissement soit plus efficace après fractionnement des échantillons.

Le ratio des peptides phosphorylés sur l'ensemble des peptides identifiés est calculé : pour les échantillons non fractionnés, il est de 44 ± 4% soit plus que précédemment (paragraphe III.A.2, 17 ± 2%). Le fractionnement semble permettre d'augmenter l'efficacité de l'enrichissement car 52 ± 3% des peptides identifiés sont phosphorylés. Le rendement de l'enrichissement reste très peu élevé par rapport à celui obtenu sur des cultures cellulaires (en moyenne 90%) dans la note d'application d'Agilent évaluant l'enrichissement sur le Bravo AssayMAP avec les cartouches Fe(III)-NTA [198]. Il est possible que l'enrichissement soit moins efficace car l'échantillon (tissus musculaire) est plus complexe que les lysats de culture cellulaire. Cependant, même sur les tissus musculaires, le rendement demeure constant entre les échantillons. Ainsi même si la sélectivité de l'enrichissement semble plutôt mauvaise, l'enrichissement est reproductible, que ce soit pour les échantillons fractionnés ou non fractionnés.

Les nombres de phosphoprotéines et de phosphosites classe I sont détaillés sur la figure 70.

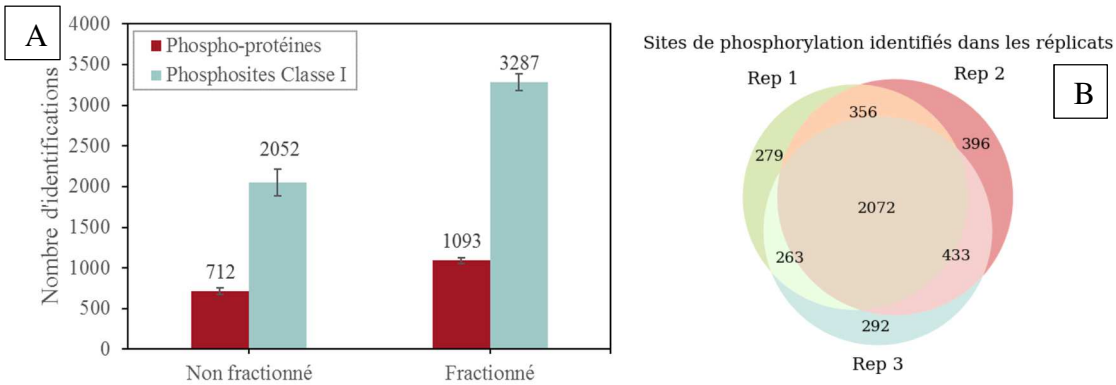

Figure 70 : Nombre moyen de phosphoprotéines et de phosphosites classe I identifiés à partir des échantillons fractionnés ou non fractionnés (A). Recouvrement des sites de phosphorylations identifiés entre les trois réplicas (B).

Tout d'abord, près du double de phosphosites est identifié en moyenne par rapport au premier test sur les échantillons sans fractionnement (paragraphe III.A.2). Cette augmentation peut s'expliquer par un meilleur rendement d'enrichissement et par un recours limité au concentrateur sous vide. Après le dessalage, le solvant n'a pas été évaporé mais de l'acétonitrile a été ajouté à l'échantillon pour atteindre les conditions requises pour l'enrichissement en phosphorylations, ce qui a pu limiter la dégradation des échantillons. De plus, 27% de l'échantillon ont été injectés contre 20% précédemment.

Bien que limité à 3 fractions, ce fractionnement a permis d'augmenter de 60% le nombre de phosphosites classe I identifiés démontrant ainsi l'efficacité de cette stratégie. Dans l'échantillon fractionné, 3047 phosphopeptides sont identifiés en combinant les 3 fractions analysées, soit une valeur proche de celle obtenue par Potts *et al.* lors de l'enrichissement de tissus musculaires[328]. Ce résultat est d'autant plus satisfaisant que l'expérience de Potts *et al.* avait été réalisée à partir d'un milligramme de protéines (contre 500 µg dans nos analyses)[328].

De plus, le recouvrement entre les sites de phosphorylations identifiés dans les réplicas des échantillons fractionnés est de 51%, ce qui montre une excellente répétabilité à la fois du fractionnement, de l'enrichissement et de l'analyse. Même avec une étape supplémentaire, la répétabilité des analyses est supérieure à celle observée précédemment. Cette étape supplémentaire a permis de décomplexifier l'échantillon en amont et d'augmenter ainsi la rétention des phosphopeptides sur la cartouche IMAC.

Les proportions des acides aminés STY et la multiplicité des phosphopeptides (mono-, bi- et tri-phosphorylés) restent inchangées par rapport aux premiers tests et entre les échantillons fractionnés et non fractionnés, ce qui signifie que le fractionnement se semble pas introduire de biais quant à la nature des phosphopeptides identifiés.

Le CV médian sur l'intensité brute des phosphosites observés dans les 3 réplicas est de 12% (1308 sites) pour les échantillons non fractionnés et légèrement plus élevé pour les échantillons fractionnés : 17% (2250 sites). L'ajout d'une étape supplémentaire augmente la variabilité de la méthode comme attendu. Mais la variabilité observée reste en deçà de 20%, seuil habituel pour la quantification au niveau protéique. De plus, l'addition d'une étape de fractionnement permet de quantifier 72% de sites supplémentaires.

Pour conclure, ces résultats montrent bien l'apport d'un fractionnement, même limité, avant l'enrichissement sur cartouches IMAC puisque non seulement un plus grand nombre de phosphosites sont identifiés et quantifiés mais l'analyse reste très reproductible. Le fractionnement automatisé sur cartouches RPS à grande capacité a donc toute sa place dans une méthode pour l'identification et la quantification des phosphorylations à large échelle.

IV. Application de cette méthodologie pour étudier la préservation du muscle durant l'hibernation de l'ours

Le protocole retenu est illustré en figure 71. Cette méthode est semi-automatisée et chaque étape de l'analyse a été optimisée, de la préparation d'échantillon au traitement des données, comme

présenté dans les parties précédentes. Les protéines sont extraites par un tampon urée/thio-urée puis digérées en solution. Les peptides sont ensuite fractionnés en 3 par fractionnement hpH RP. Ces fractions sont ensuite enrichies en phosphopeptides par IMAC et injectées sur un spectromètre Q-Exactive HF-X. L'identification et la quantification au niveau des sites de phosphorylation sont effectués avec MaxQuant.

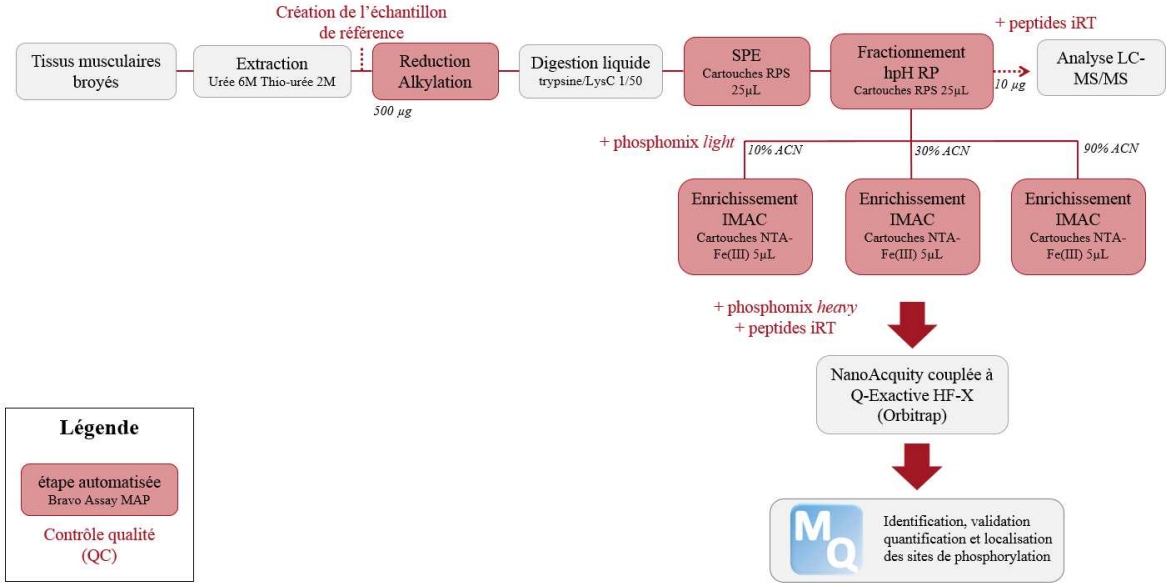

Figure 71 : Stratégie analytique développée pour l'étude des phosphorylations dans le muscle d'ours.

Cette méthode a été appliquée à des biopsies musculaires d'ours (N=11/groupe) pour étudier les mécanismes à l'œuvre dans l'adaptation du muscle de l'ours au cours de l'hibernation. Les échantillons ont été prélevés durant la période d'hibernation (hiver) et la période active (été). En plus de l'analyse phosphoprotéomique, le protéome musculaire complet est analysé afin de comparer les variations d'abondance au niveau protéique et au niveau des sites de phosphorylation. Pour ce faire, 10 µg des échantillons fractionnés avant enrichissement sont prélevés et analysés pour accéder au protéome musculaire complet.

A. Résultats et contrôles de la qualité des analyses des échantillons avant enrichissement

Dans les échantillons avant enrichissement, 2228 protéines ont été identifiées au total dont 1775 ont pu être quantifiées (au minimum 5 valeurs sur 11 dans chaque groupe (hiver et été) et 1 peptide unique). 586 protéines se sont révélées statistiquement différentielles.

Des peptides synthétiques iRT ont été ajoutés dans tous les échantillons. Le CV médian sur les temps de rétention est de 0,47% et sur les aires mesurées après extraction des courants d'ion de 31%.

Un échantillon de référence constitué d'une fraction de chacun des autres échantillons est traité en parallèle des autres échantillons. Comme pour les autres échantillons, une portion est

prélevée après fractionnement et avant enrichissement. Avant injection, les différentes fractions de ces échantillons sont regroupées pour former un échantillon qui est ensuite injecté à intervalles réguliers à plusieurs reprises (5) durant la séquence d'injection. Le CV médian des valeurs LFQ des protéines quantifiées dans au moins 3/5 échantillon est de 12%.

Ces résultats démontrent la très bonne stabilité de l'analyse LC-MS/MS à la fois au niveau des temps de rétention et de la détection et de la quantification des analytes.

B. Résultats et contrôles de la qualité des analyses des échantillons enrichis en peptides phosphorylés

Le nombre de protéines phosphorylées, peptides protéotypiques phosphorylés et phosphosites de classe I avant et après l'application des filtres définis dans le paragraphe III.C.4, à savoir au minimum 5 valeurs sur 11 dans chaque groupe (hiver et été) et une probabilité de localisation supérieure à 0,75 est présenté en figure 72.

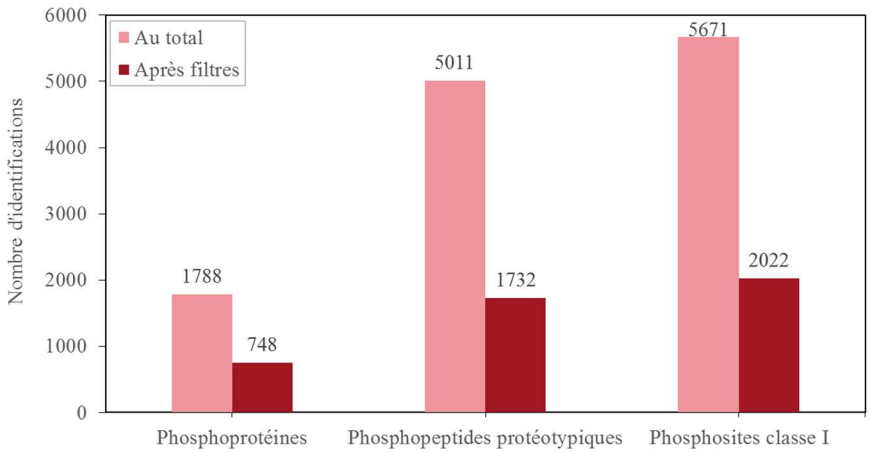

Figure 72 : Nombre de protéines phosphorylées, peptides protéotypiques phosphorylés et phosphosites classe I identifiés sur l'ensemble des échantillons avant et après l'application des filtres.

Pour les échantillons enrichis en phosphopeptides, 1788 phosphoprotéines, 5011 phosphopeptides protéotypiques et 5671 phosphosites de classe I sont identifiés au total. Après application des filtres, 2022 phosphosites localisés sur 1732 peptides issus de 748 protéines sont retenus pour la quantification.

Pour s'assurer de la stabilité des analyses tout au long des injections, des contrôles de la qualité des injections doivent être réalisés. Comme décrit par Bittremieux *et al.*[331], ces contrôles doivent être adaptés à la spécificité de l'analyse phosphoprotéomique : contrôle de l'enrichissement et échantillon de référence représentatif injecté au cours de la séquence.

Pour évaluer l'enrichissement en phosphopeptides, les peptides phosphorylés synthétiques du phosphomix 1 sont ajoutés avant (*light*) et après (*heavy*) l'enrichissement IMAC. Les courants d'ions correspondant aux peptides du mix sont extraits et l'aire sous la courbe des chromatogrammes obtenus est mesurée. 3 peptides sont détectés dans tous les échantillons, les ratios *heavy/light* d'enrichissement et les CV calculés sur l'ensemble des échantillons sont

présentés en tableau 10. En moyenne, le rendement de l'enrichissement se situe entre 56 et 65%. De plus, il apparait très reproductible puisque le CV médian sur les ratios est faible – entre 9 et 13%.

Tableau 10 : Ratio *light/heavy* des phosphopeptides synthétiques observés du phosphomix 1 dans les échantillons.

	Ratio *light/heavy* moyen (%)	**CV médian sur les ratios *light/heavy* (%)**
RDSLG**T**YSSR	64 ± 6	9,3
TKLI**T**QLRDAK	65 ± 8	11,8
EVQAEQPSS**S**SPR	56 ± 7	13,1

L'échantillon de référence décrit au paragraphe précédent a été enrichi sur cartouches NTA-Fe(III) avec les autres des échantillons. Les différentes fractions enrichies de ces échantillons ont ensuite été regroupées pour former le mélange peptidique de référence enrichi en phosphopeptides qui a été injecté à de multiples reprises (5) durant la séquence d'injection. Le CV médian sur les intensités brutes des phosphosites observés dans au moins 3 des 5 réplicas a été calculé. Sur 1764 phosphosites, il était de 18,3%, ce qui montre une bonne répétabilité de l'analyse phosphoprotéomique puisque ce résultat est comparable à ce qui est obtenu pour la quantification à l'échelle protéique avec des analyses *label-free* (CV inférieur à 20%).

Par ailleurs dans tous les échantillons, des peptides synthétiques iRT ont été ajoutés. Le CV médian sur les temps de rétention de ces peptides était de 0,45% et sur les aires est de 30%. Ce sont des valeurs extrêmement similaires à celles obtenues sur les échantillons avant enrichissement, montrant encore une fois la stabilité du système chromatographique et du spectromètre de masse.

Ces différents contrôles de la qualité montrent la robustesse de la méthode avec une excellente répétabilité de la chromatographie, un enrichissement robuste et une quantification reproductible.

C. Premières interprétations des données phosphoprotéomiques obtenues

Pour faciliter l'interprétation biologique des données de phosphoprotéomique par les biologistes, des outils bio-informatiques sont disponibles. Ils recensent les informations fonctionnelles relatives aux sites de phosphorylation ainsi que les voies de kinases qui induisent l'apparition de la phosphorylation sur les protéines. Mais la plupart ont été développés pour des organismes très étudiés comme l'homme et la souris. Une stratégie permettant d'obtenir la

correspondance entre les sites humains et ours a donc été développée en collaboration avec Alexandre Burel et Adrien Brown au sein du laboratoire.

1. Alignement de séquence pour identifier le site humain homologue au site chez l'ours

Cette stratégie comporte plusieurs étapes illustrées en figure 73.

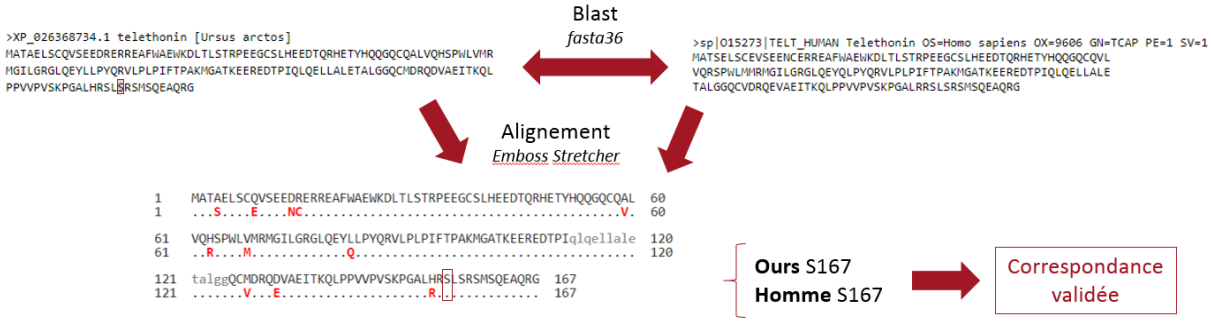

Figure 73 : Correspondance entre les sites « ours » et « humains »

Afin d'obtenir la correspondance entre les sites de phosphorylation chez l'ours et chez l'humain, une recherche par homologie de séquence de la protéine phosphorylée ours a été réalisée grâce à l'algorithme fasta v36[172-173] contre la banque SwissProt humaine. Si l'homologie de séquence était supérieure à 80% ou si les descriptions des deux protéines (humaine et ours) concordaient, la correspondance était validée.

Lorsque le blast était validé, les séquences ours et humaines ont été alignées par l'outil Emboss Stretcher[332], l'acide aminé en face du site de phosphorylation de la protéine « ours » a alors été identifié. La correspondance devait ensuite être validée manuellement. La validation de ces correspondances a été effectuée comme suit :

- Si l'acide aminé « humain » correspondant était un acide aminé phosphorylable (S, T et Y) et que sa position dans la protéine est proche (10 acides aminés près) de la position occupée par la phosphorylation chez l'ours après correction du plus grand gap introduit entre les deux protéines : la correspondance entre les sites était validée. Dans le cas contraire on ne pouvait pas conclure et des investigations plus poussées devaient être réalisées.

- Si l'acide aminé humain n'était pas une sérine, une thréonine ou une tyrosine : la correspondance était rejetée.

Après application de ces deux filtres sur la nature et la position du site humain correspondant, il restait les cas pour lesquels l'acide aminé était phosphorylable mais la position dans la protéine était très éloignée de la protéine ours. Un premier tri a pu être fait en s'intéressant au cas de la titine. Cette protéine musculaire de très haut poids moléculaire (> 3 000 000 Da) portait à elle seule 965 sites identifiés et 300 quantifiés, une observation proche de ce qui avait été décrit par Potts *et al.*[328]. La quasi-totalité de ces sites n'a pas pu être validée par les critères précédents car il existe un très grand gap de 977 positions entrecoupées de 39 acides aminés

répartis régulièrement au niveau du 3437e acide aminé de la protéine ours. En corrigeant la position des sites humains avec cet incrément de 938 positions, la correspondance des sites de la titine ont pu être validés.

Il restait alors les correspondances qui n'ont pas pu être validées car bien qu'étant sur un acide aminé phosphorylable, leur position est trop éloignée de la position ours. Notre objectif a alors été de comparer l'environnement proche du site. Une fenêtre de 10 acides aminées de part et d'autre du site de phosphorylation a été créé pour les sites humains et ours. Un blast a ensuite été réalisé sur chaque paire pour obtenir le pourcentage d'homologie entre les séquences. Ce blast a été effectué avec ncbi-blast-2.12.0+, spécialisé dans la comparaison de séquences courtes. Si l'identité du blast était supérieure à 80%, la correspondance est validée, sinon elle était rejetée.

Au total, 1840 sites quantifiés de l'ours ont pu être mis en correspondance avec un site humain.

2. Utilisation des banques fonctionnelles avec les données humaines

La banque PhosphoSitePlus fait référence pour l'analyse fonctionnelle des sites de phosphorylations avec près de 300 000 sites de phosphorylations répertoriés en 2018[230].

En utilisant les sites de phosphorylation humains correspondant aux sites de phosphorylations observés chez l'ours et l'outil Perseus, les informations issues de PhosphositePlus ont pu être extraites pour nos données. 1331 sites phosphorylés chez l'ours sont connus pour être phosphorylés chez l'homme, soit 72% des sites pour lesquels une correspondance a été validée. Il est également possible avec cet outil de connaitre la kinase associée à la phosphorylation, c'est le cas pour 155 sites identifiés dans nos données.

Par ailleurs des informations relatives aux phosphorylations sont également présentes dans la banque de données UniprotKB. Bien que cette banque soit moins complète que PhosphoSitePlus, des informations relatives à 200 de sites de phosphorylation ont été reportés uniquement avec UniprotKB. Dans cette banque, l'information fonctionnelle est reportée au niveau de la protéine ou au niveau du site. Cependant, ces informations doivent être traitées manuellement car il n'y a pas de formatage spécifique des annotations fonctionnelles.

3. Analyse statistique du jeu de données

Pour élargir le jeu de données, nous avons décidé d'ajouter les sites de phosphorylations dont la localisation n'est pas certaine (probabilité de localisation < 0,75) mais qui sont connus pour être phosphorylés chez l'homme. Au final, un total de 2094 sites de phosphorylations quantifiés est considéré.

Un t-test est réalisé pour déterminer les sites dont l'abondance varie significativement entre les groupes (p-value < 0,05) : 986 sites situés sur 416 protéines.

L'origine de la variation d'abondance des sites de phosphorylation peut être liée soit à la variation de la régulation du site sur une protéine dont l'abondance ne varie pas ou alors la proportion de la phosphorylation reste identique mais la protéine varie en abondance. Le

recouvrement entre les protéines quantifiées avant enrichissement et des protéines portant des sites de phosphorylation dont l'abondance varie est présentée en figure 74.

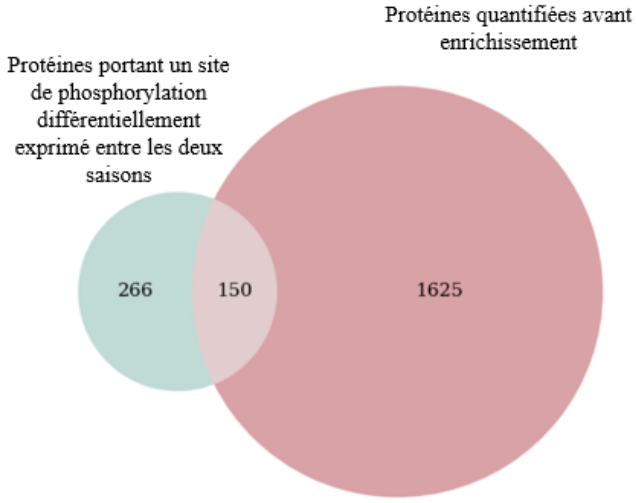

Figure 74 : Recouvrement entre les protéines quantifiées avant enrichissement et les protéines portant des sites de phosphorylation dont l'abondance varie

Dans la suite de ce projet, les variations des sites de phosphorylations et des protéines dans le jeu de données avant enrichissement seront comparées sur les 150 protéines qui ont été quantifiées avant enrichissement et dont l'un des sites de phosphorylation est différentiellement abondant dans le jeu de donnée enrichi par IMAC. L'objectif étant d'avoir un regard critique sur les variations des sites de phosphorylations à défaut de pouvoir discriminer l'effet de la variation d'abondance de la protéine et du site de phosphorylation.

4. Analyse d'enrichissement en annotations fonctionnelles

Pour obtenir un premier aperçu des fonctions impactées par l'hibernation à partir des données de phosphorylations, un enrichissement en annotation fonctionnelle a été réalisé. Pour ce faire, le logiciel EASE s'appuyant sur l'algorithme DAVID a été utilisé. Son principe est présenté en partie I chapitre 1. L'enrichissement en annotations fonctionnelles est réalisé sur les protéines pour lesquelles l'abondance d'au moins un site de phosphorylation est statistiquement différent entre le statut hibernant et actif. Les formes phosphorylées des protéines impliquées dans la contraction musculaire, le métabolisme des protéines, le cytosquelette et la croissance sont les plus impactées par l'hibernation, comme illustré en figure 75.

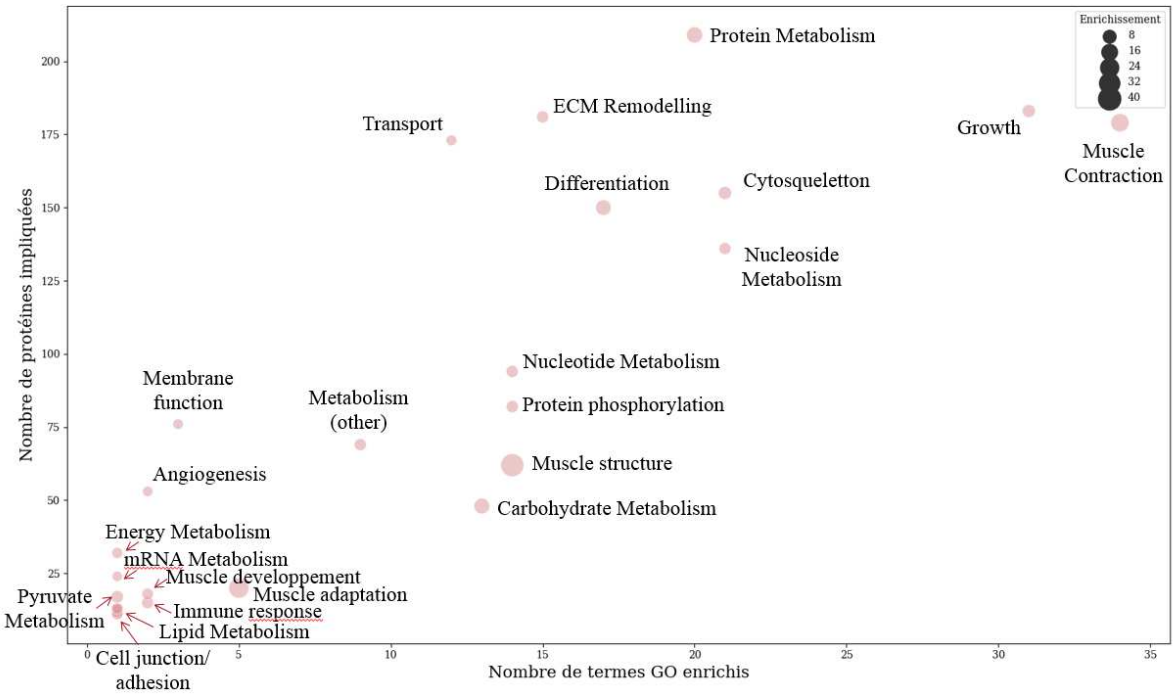

Figure 75 : Résultats du calcul d'enrichissement en annotations fonctionnelles sur les protéines portant un phosphosite dont l'abondance varie différentiellement entre la phase hibernante et la phase active ; ECM = matrice extracellulaire.

Dans la suite de ce projet, ces données seront analysées pour interpréter fonctionnellement les modifications d'abondance observées au niveau des sites de phosphorylations. Les voies de kinases privilégiées en fonction du statut hibernant ou actif seront également identifiées grâce aux informations présentes sur PhosphoSite Plus.

V. Conclusion

Au terme de nombreuses optimisations des différentes étapes de l'analyse protéomique, une méthode complète et robuste pour l'analyse quantitative sans marquage des phosphorylations à partir de tissus musculaires a été développée. La préparation des échantillons a été réalisée de façon semi-automatique sur une quantité de matériel de départ restreinte (500 µg) à l'aide d'un robot Bravo AssayMAP. Après extraction des protéines avec un tampon contenant de l'urée et du thio-urée, les protéines ont été digérées en solution. Les peptides ont été fractionnés par fractionnement hpH RP sur des cartouches RPS puis les fractions ont été enrichies en phosphopeptides. Ces peptides ont ensuite été injectés sur un spectromètre de type Orbitrap et fragmentés par HCD. Une quantification fine au niveau du site de phosphorylation a été rendue possible par l'utilisation de MaxQuant pour le traitement des données. Enfin, des outils bio-informatiques (fasta v36, Emboss Stretcher) ont permis d'utiliser les banques de données existantes (GO, PhosphoSite Plus, UniprotKB) pour une espèce non modèle comme l'ours *via* la correspondance avec un site humain. Ainsi, des informations fonctionnelles ont été obtenues. Cette méthodologie a ensuite été appliquée à des échantillons de muscle issus d'ours hibernants ou actifs (N=11 par groupe), ce qui a permis l'identification de 5011 peptides phosphorylés

protéotypiques. Les différents contrôles qualité utilisés lors de cette série d'analyses ont démontré la robustesse et la répétabilité de la méthode employée.

Durant la mise au point de cette méthode, je me suis focalisée en premier lieu sur la préparation d'échantillons. Cependant, les méthodes LC et MS gagneraient certainement à être optimisées, notamment sur le timsTOF Pro. Des optimisations supplémentaires sont en cours sur cet appareil notamment au niveau des temps d'accumulation et de la gamme de mobilité ionique. De plus, bien que plusieurs traitements de données aient été comparés, il serait intéressant d'évaluer de nouveaux algorithmes basés sur la prédiction des spectres de fragmentation et des temps de rétention, comme peut le faire ionbot[154].

Un autre aspect qui mériterait d'être développé est l'utilisation du mode DIA pour l'analyse des phosphorylations. Il permettrait une quantification plus robuste en réduisant le nombre de valeurs manquantes et les problèmes liés au sous échantillonnage des peptides phosphorylés[333].

Par ailleurs, les études les plus récentes pointent les deux grands défis auxquels la phosphoprotéomique doit encore s'atteler : le calcul du taux de phosphorylation au niveau protéique ou au niveau du site et l'interprétation biologique des données[182]. En effet, un effort doit être fourni afin de remplir les bases de données comme PhosphoSite Plus en identifiant le rôle spécifique des phosphorylations pour faciliter l'interprétation des données.

Partie II : Mise au point de protocoles d'analyses protéomiques quantitatives pour étudier l'hibernation chez des espèces non modèles

Chapitre 3

Évaluation d'un protocole expérimental adapté à l'analyse de cœurs de lérots pour mieux comprendre le rôle des acides gras polyinsaturés dans les mécanismes de l'hibernation

Cette étude a été réalisée en collaboration avec le Dr Sylvain Giroud (University of Veterinary Medicine, Vienne, Autriche).

I. Les acides gras polyinsaturés et la SERCA influencent la réponse cardiaque à l'hibernation mais les voies intracellulaires impliquées restent inconnues

Malgré deux siècles de recherches sur l'hibernation, les mécanismes physiologiques qui régulent cette stratégie de survie hivernale sont encore mal compris. Les hibernants économisent de l'énergie en diminuant considérablement le taux métabolique et la température corporelle pendant les épisodes de torpeur. Cependant, l'hibernation est interrompue à plusieurs reprises par le réchauffement des individus et ces périodes d'euthermie sont responsables d'environ 80% de la dépense énergétique d'un hibernant pendant l'hiver. Ces réveils périodiques sont communs à de nombreux petits mammifères hibernants[295] comme par exemple le lérot. Cependant, leurs fonctions restent un mystère. Il n'y a que deux facteurs externes connus pour affecter fortement la durée de la torpeur et donc la fréquence des réveils périodiques : la température corporelle (qui peut être contrainte par la température ambiante)[334], et l'absorption alimentaire d'acides gras polyinsaturés essentiels [AGPI].

L'acide linoléique [LA] (C18:2), l'acide gras majeur de la famille n-6, a de forts effets positifs sur la torpeur, en particulier sur sa durée.[335] Cela suggère que les AGPI agissent sur le temps où les individus peuvent rester torpides jusqu'à ce qu'ils aient besoin de réchauffer. Il a déjà été suggéré que les AGPI n-6 exercent leur effet sur le maintien de la fonction cardiaque pendant la torpeur profonde[336]. Plus précisément, il semblerait qu'une teneur élevée en AGPI n-6 dans les phospholipides de la membrane cellulaire (PL) atténue les effets de la température sur l'activité de la protéine cardiaque *sarcoplasmic reticulum calcium ATPase* [SERCA2a]. Cette pompe transmembranaire est responsable de l'élimination du calcium dans le réticulum sarcoplasmique et donc du maintien de la fonction cardiaque même à une température corporelle extrêmement faible. Une hypothèse stipule qu'une teneur élevée en LA pourrait compenser partiellement une réduction de l'activité de SERCA2a résultant de la protéolyse au fil du temps, prolongeant le temps pendant lequel les hibernants peuvent rester torpides jusqu'à ce que la nécessité de synthétiser la protéine SERCA2a les oblige à se réchauffer.

Cependant, la consommation de LA ne doit pas être trop importante car des niveaux élevés risquent de déclencher une diminution de la capacité à entrer en hibernation, qui pourrait être due à une augmentation des dommages oxydatifs liés aux AGPI pendant le cycle torpeur-excitation[337] qui impacte négativement l'activité de la SERCA2a[338-340] et le maintien somatique en général. Par ailleurs, les effets des acides gras n-3 sont moins bien connus. Deux études ont notamment montré qu'un régime spécifiquement enrichi en acide gras n-3 réduit la capacité à hiberner[341-342]. Ainsi, ces données suggèrent des effets opposés des AGPI n-6 et n-3 sur l'hibernation et la torpeur.

Au-delà des hypothèses autour de la SERCA2a, les voies moléculaires spécifiques par lesquelles les acides gras insaturés affectent l'hibernation sont aujourd'hui entièrement inconnues. Ce projet visait donc à déterminer les mécanismes intracellulaires affectés dans le tissu cardiaque selon que le régime alimentaire soit enrichi en AGPI n-6 ou n-3, et selon le statut physiologique des lérots (torpeur, réveil, période active).

II. Développement d'un protocole de préparation d'échantillons adapté à l'analyse protéomique de cœurs de lérots

A. L'importance de la préparation d'échantillons

L'analyse protéomique selon une stratégie « *bottom-up* » sans marquage des protéines est aujourd'hui une méthode de choix permettant l'identification et la quantification de milliers/centaines de protéines simultanément. Pour l'approche « *label-free* », chaque échantillon est soumis indépendamment aux différentes étapes de l'analyse protéomique qui sont source de nombreuses variations. Cette stratégie est utilisée dans de nombreux contextes, notamment pour mieux comprendre comment un tissu répond à une situation physiologique particulière, ce qui permet de mieux comprendre les mécanismes biologiques. Bien que chaque étape de l'analyse protéomique soit importante, un soin particulier doit être apporté au traitement des échantillons car il peut être source de nombreuses variations.

La préparation des échantillons a beaucoup évolué ces dernières années avec l'apparition de nouveaux protocoles, souvent automatisables, visant à augmenter la répétabilité du processus[39-40, 60, 62]. Malgré les efforts fournis par la communauté scientifique, il n'existe pas à ce jour de protocole universel adapté à tous les échantillons. C'est pourquoi, cette étape doit être adaptée à chaque type d'échantillon et à la question biologique posée.

Ce projet s'intéresse à l'influence de la diète et/ou du statut physiologique (torpeur, réveil, période active) sur le protéome du muscle cardiaque de lérots. Pour un tel projet, deux étapes de la préparation des échantillons sont particulièrement importantes : l'extraction des protéines et leur digestion en peptides avant analyse. En effet, comme décrit par Rodrigues-Ribeiro *et al.*, l'extraction des protéines du tissu cardiaque est particulièrement difficile. D'une part, c'est un tissu très hétérogène à l'échelle microscopique (différentes cellules) et à l'échelle macroscopique (ventricules vs oreillettes) et d'autre part, la gamme dynamique de son protéome est élevée (6 ordres de magnitudes avec une surreprésentation des myosines, troponines et

actines)[343]. Dans la chapitre précédent, l'extraction des protéines a été optimisée sur un tissu très similaire, le muscle squelettique, qui possède une gamme dynamique équivalente. Le tampon d'extraction contenant de l'urée et du thio-urée avait permis d'identifier davantage de protéines qu'avec un tampon contenant de l'urée ou un « Laemmli-like ». Au vu de ces résultats, conforté par l'utilisation de ce tampon pour l'analyse protéomique des cœurs de primates dans l'étude de Hu *et al.*[344], l'extraction des protéines de cœur de lérot a donc été réalisée avec le tampon contenant de l'urée et du thio-urée.

Au laboratoire, le protocole couramment utilisé pour ce type d'analyses (stratégie « *label-free* » à partir de tissu) est le gel « Stacking ». Il est compatible avec de nombreux tampons d'extraction, qui peuvent être éliminés avant digestion par lavage du gel. Deux méthodes plus récentes ont été introduites au laboratoire et testées avec succès sur différents types d'échantillons[62, 77]. Ces nouveaux protocoles ont l'avantage d'être automatisables, de réduire le temps de préparation et le nombre d'étapes nécessaires par rapport au protocole de gel « Stacking ». L'objectif est de comparer ces nouveaux protocoles (S-Trap et SP3) au gel « Stacking » préexistant sur des échantillons cardiaques.

B. Description des protocoles de digestion et de la stratégie analytique

Les trois protocoles étudiés reposent sur différentes procédures de digestion et ne nécessitent pas le même temps de préparation. Les différentes étapes des protocoles sont récapitulées en figure 76.

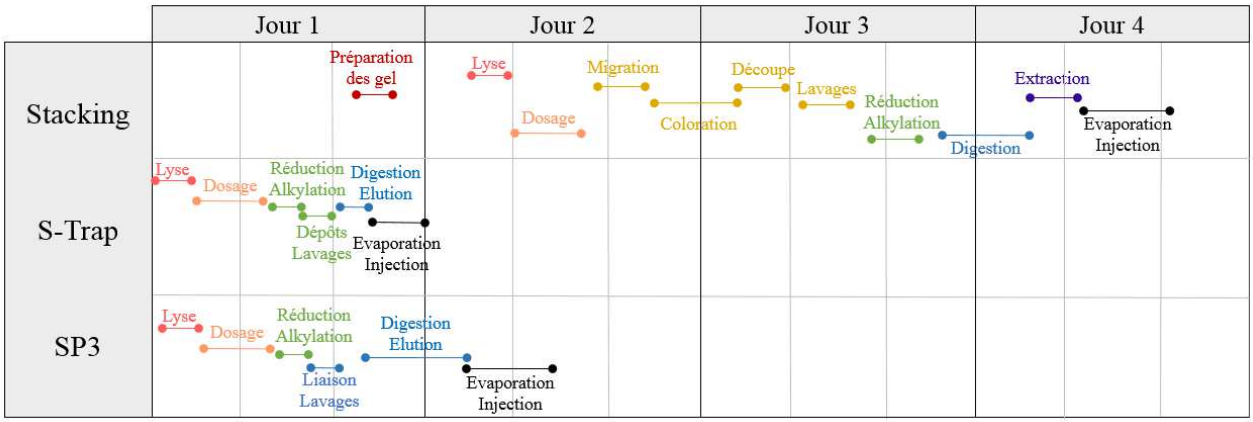

Figure 76 : Description des étapes de préparation d'échantillons des trois méthodes employées.

Le protocole « Stacking » consiste à déposer les protéines dans un tampon à forte concentration en SDS (2,5%) sur un gel SDS-PAGE. Le SDS se fixe sur les protéines par interactions hydrophobes avec la chaine peptidique[345]. Les complexes détergent-protéines négativement chargés migrent dans un gel faiblement réticulé sous l'impulsion d'un champ électrostatique. Lorsque les protéines ont pénétré le gel et ont été concentrées selon le principe décrit en partie I chapitre 1, la migration est stoppée et les protéines, piégées dans le gel, sont réduites, alkylées et digérées dans le gel. Après digestion, les peptides issus de la digestion sont extraits du gel[346]. C'est le protocole le plus long car il est nécessaire de laisser les gels polymériser une nuit avant

le dépôt des échantillons pour éviter la propionamidation des cystéines, des lysines et des N-ter par réaction avec l'acrylamide[347]. Cette modification peut en effet entrainer une digestion limitée des protéines et empêcher l'alkylation complète de toutes les cystéines, ce qui peut nuire à l'identification des peptides[68]. Il faut ensuite colorer les protéines dans le gel au cours d'une deuxième nuit et enfin la digestion à la trypsine dure également une nuit. Au total, il faut donc 4 jours pour réaliser ce protocole. Pendant longtemps, cette technique de digestion était la seule à être compatible avec des tampons contenant du SDS qui permet la mise en solution de protéines hydrophobes[348].

Le protocole de digestion sur membrane, nommé « S-Trap », repose sur la mise en suspension des protéines par ajout d'une solution méthanolique aux protéines solubilisées dans un milieu acide avec 5% de SDS. Ces protéines en suspension sont ensuite déposées sur une membrane formée par des fibres de quartz. Les protéines sont piégées dans des pores (< 1 µm) où elles sont ensuite digérées. Les peptides ne sont pas retenus par la membrane et sont élués[12, 62]. Il a été montré précédemment au laboratoire que la durée de la digestion optimale était de 3h seulement pour ce protocole. Ceci permet de réaliser la totalité du protocole en une journée.

Les protocoles de digestion sur billes paramagnétiques fonctionnalisées par des groupements carboxylates (SP3) sont basés sur la fixation des protéines aux billes. Les sels et les impuretés du milieu peuvent alors être retirés durant les lavages. Les protéines sont ensuite digérées sur billes et les peptides obtenus sont élués[77]. La totalité du protocole peut être réalisée dans le même contenant, ce qui diminue les pertes dues aux transvasements et à l'adsorption sur les parois des tubes. La durée nécessaire pour ce protocole est de 2 jours puisque la digestion s'effectue sur la nuit. Les protéines solvatées par l'acétonitrile sont retenues sur les billes par des interactions hydrophiles entre les protéines et les groupements carboxylates qui fonctionnalisent les billes. Le mécanisme à l'œuvre est analogue à celui de la chromatographie d'interaction hydrophile (HILIC) où les analytes sont séparés par séparation liquide/liquide entre une fine couche partiellement immobilisée sur la phase stationnaire et l'éluant principal[40, 77, 349].

Les conditions de digestion pour chacun des 3 protocoles (protéase, ratio protéase/protéine, durée et température) pour chacune de ces techniques ont été optimisées précédemment au laboratoire sur des échantillons standard de type cellules HeLa ou de levure (*S.cerevisiae*) et j'ai donc utilisé ces conditions. Ainsi, la digestion en gel est réalisée sur la nuit à 37°C avec un ratio 1/100 (enzyme/protéines) de trypsine. La digestion sur membrane est effectuée à un ratio élevé de 1/10 (enzyme/protéines) avec un mélange 1/1 de trypsine et Lys-C pendant 3h à 37°C et enfin la digestion sur billes paramagnétiques nécessite un ratio 1/20 avec de la trypsine/Lys-C (1:1) sur la nuit à 37°C. La digestion sur gel nécessite un ratio plus faible afin de limiter l'autolyse de la trypsine et favoriser la pénétration de la trypsine dans le gel. Au contraire, la digestion sur membrane est facilitée par la mise en suspension des protéines rendant les sites de clivage très accessibles car les protéines sont entièrement dépliées sous l'effet du SDS. C'est pourquoi le temps de digestion peut être fortement réduit.

Afin de comparer les différents protocoles de digestion, un demi-cœur de Lérot (*Eliomys quercinus),* a été broyé et les protéines ont été extraites par un tampon « urée/thio-urée ». 30 µg

de protéines ont été soumis aux trois protocoles de digestion (figure 2). Ce dernier protocole de digestion a été réalisé de manière manuelle (SP3) ou automatisée (AutoSP3) pour évaluer l'apport de l'automatisation. Chaque digestion a été réalisée en quadruplicat à partir d'un seul et même échantillon d'extrait protéique. Les peptides obtenus ont ensuite été injectés sur une nanoAcquity couplée à un Q-Exactive HF-X. Les données générées ont été analysées par le logiciel MaxQuant (v2.0.3.0) (identification, validation et quantification) avec une banque créée avec les séquences d'*Ictidomis tridecemlineatus* issues de RefSeq et des séquences générées à partir du génome mitochondrial d'*Eliomys quercinus*. Le schéma analytique est présenté en figure 77. Les détails de chaque protocole utilisé sont décrits en partie expérimentale (III.1).

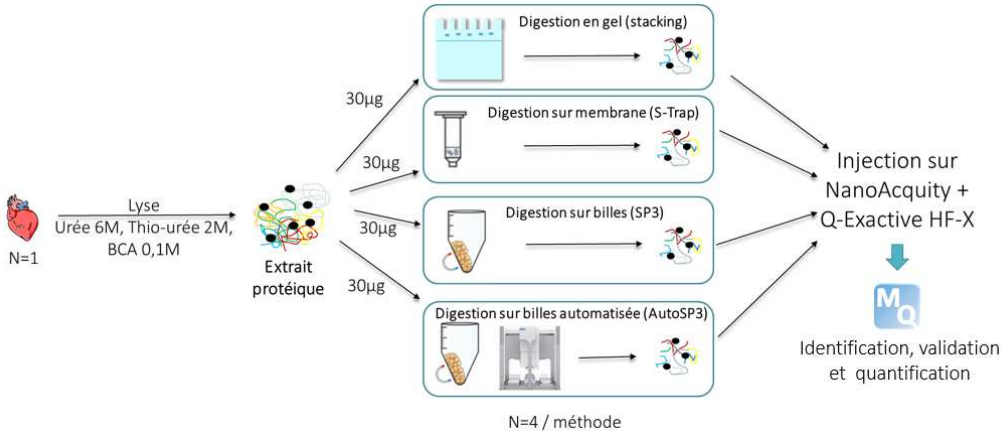

Figure 77 : Schéma analytique mis en œuvre pour comparer la digestion en gel (Stacking) aux digestions sur membrane (S-Trap), manuelle sur billes paramagnétiques (SP3) et automatisée sur bille (Auto SP3).

C. Comparaison des 4 protocoles de préparation d'échantillons pour l'étude du protéome cardiaque

Pour évaluer ces différentes méthodes de digestion, plusieurs critères ont été pris en compte : le nombre de protéines et de peptides identifiés, la répétabilité du protocole en termes d'identification et de quantification, le taux de peptides portant des clivages manqués et enfin la durée du protocole.

1. Nombre de protéines/peptides obtenus

Dans un premier temps, les nombres de protéines et de peptides identifiés ont été comparés pour les différents protocoles et les résultats sont présentés sur la figure 78.

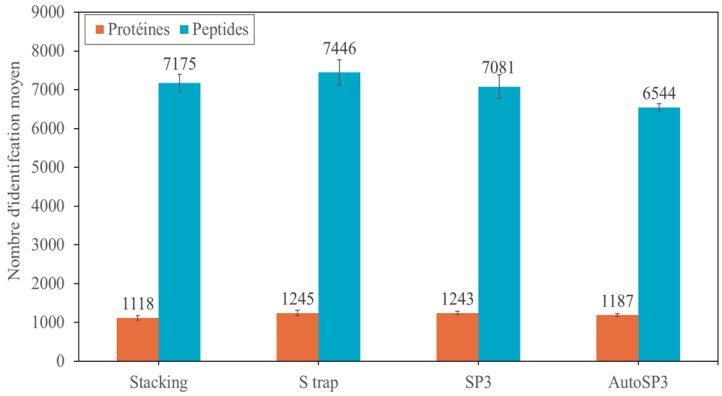

Figure 78 : Nombre de protéines et de peptides identifiés par les différents protocoles de digestion (Stacking, S-Trap, SP3 et AutoSP3).

Le protocole de digestion sur membrane permet d'identifier le plus grand nombre de peptides (entre 4 et 14% d'augmentation par rapport aux autres protocoles) et de protéines (entre 2 et 11% d'augmentation par rapport aux autres protocoles). Au contraire, le protocole « Stacking » identifie le moins de protéines par rapport aux autres, cette observation avait déjà été reportée par Yang *et al.,* qui avaient comparé ce protocole à 5 autres dont le protocole « S-Trap »[68]. Le protocole AutoSP3 identifie le moins les peptides. Les meilleures performances en terme d'identification avaient été obtenues avec le protocole « S-Trap »[68]. Cependant, il est important de souligner que le nombre d'identification est quasi identique et ne permet donc pas de conclure qu'un protocole est plus efficace qu'un autre.

Les études protéomiques du muscle cardiaque décrites dans la littérature reportent des nombres plus élevés (entre 3200 et 11 200 protéines) que les nôtres mais ces résultats sont obtenus grâce à un fractionnement peptidique poussé (environ 10 fractions) rallongeant considérablement les temps d'analyse[11, 107, 344, 350-351]. Une étude sans fractionnement du protéome cardiaque de souris identifie un nombre de protéines similaire au notre (1145 protéines) par digestion liquide et marquage iTRAQ[352]. C'est donc que les protocoles que nous avons utilisés permettent d'atteindre des performances au moins similaires à ce qui a été observé par d'autres équipes. Il faut cependant souligner une étude récente qui a permis l'identification de 4000 protéines à partir de faibles quantités de tissus cardiaque grâce à l'utilisation d'un spectromètre de dernière génération, le timsTOF Pro, et d'un nouveau détergent (Azo) compatible avec la digestion et l'analyse MS[11].

Le recouvrement est identique pour les 4 digestions évaluées (≈ 65%, données non représentées). Habituellement, il se situe entre 70 et 75% sur 3 réplicas d'injections du fait de l'aspect stochastique de l'analyse en mode DDA. Ici quatre réplicas de préparation sont comparés, donc un recouvrement légèrement moins élevé est attendu.

Afin de ne retenir que les identifications les plus robustes, un filtre sur le nombre de valeurs manquantes observées dans les quatre réplicas d'un protocole est appliqué : 1 valeur manquante est acceptée au maximum. Le tableau 11 récapitule le nombre de protéines retenues ainsi que le pourcentage de protéines éliminées lors de l'application du filtre.

Table 11 : Nombre de protéines retenues après application du filtre « 1 valeur manquante maximum » en fonction du protocole de digestion.

Protocole de digestion	Nombre de protéines dans au moins ¾ réplicas	Pourcentage de protéines éliminées par le filtre ¾ valeurs
Stacking	1052	- 6%
S-Trap	1165	- 7%
SP3	1168	- 6%
Auto SP3	1150	- 3%

Environ 100 protéines de moins sont retenues pour le protocole « Stacking » par rapport aux trois autres. Pour chaque protocole, environ 6% des protéines identifiées sont éliminées par ce filtre, à l'exception du protocole « AutoSP3 » pour lequel seules 3% des protéines sont retirées. Grâce à l'automatisation du protocole, les identifications sont plus robustes.

2. Répétabilité de la quantification des protéines

Pour évaluer la répétabilité de la quantification, le CV de l'intensité de chacune des protéines retenues après filtre sur le nombre de valeurs manquantes est calculé. La distribution des CVs est représentée sur la figure 5.

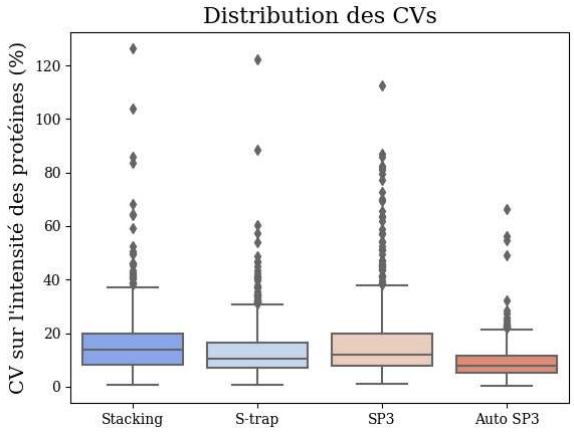

Figure 79 : Distribution des coefficients de variation (CVs) sur l'intensité des protéines identifiées par les différents protocoles de digestion (Stacking, S-Trap, SP3 et AutoSP3).

Il est intéressant de noter que tous les protocoles présentent une répétabilité inférieure à 20% avec une médiane des CVs sur l'intensité des protéines entre 7 et 14 %. La digestion « S-Trap » est le protocole manuel le plus répétable avec une valeur médiane des CV de 11%. L'automatisation du protocole de digestion SP3 réduit la valeur médiane des CV de 12 à 7% par rapport au protocole manuel tout en limitant le nombre de valeurs extrêmes. L'automatisation permet donc d'augmenter la répétabilité, au niveau qualitatif et quantitatif, du protocole de digestion sur billes paramagnétiques en limitant la variabilité induite par l'opérateur.

3. Comparaison des populations protéiques et peptidiques obtenues par les différents protocoles

Les diagrammes de Venn montrent que 65% des protéines sont identifiées par les 4 réplicas (Figure 80A). Ce recouvrement est élevé, notamment en comparaison avec le recouvrement entre 3 réplicas techniques qui se situe habituellement entre 70 et 75% du fait de l'aspect stochastique de l'analyse DDA. Dans une étude de Ding *et al.* où 4 protocoles (dont la méthode S-Trap) étaient comparés, seules 34% des protéines étaient communes aux 4 protocoles[61].

Par contre, seuls 30% des peptides totaux sont communs aux 4 protocoles (Figure 80B). Le protocole « Stacking » identifie 16% de peptides qui lui sont uniques, bien plus que les autres protocoles (7% pour S-Trap, 4% pour SP3 et 6% pour AutoSP3).

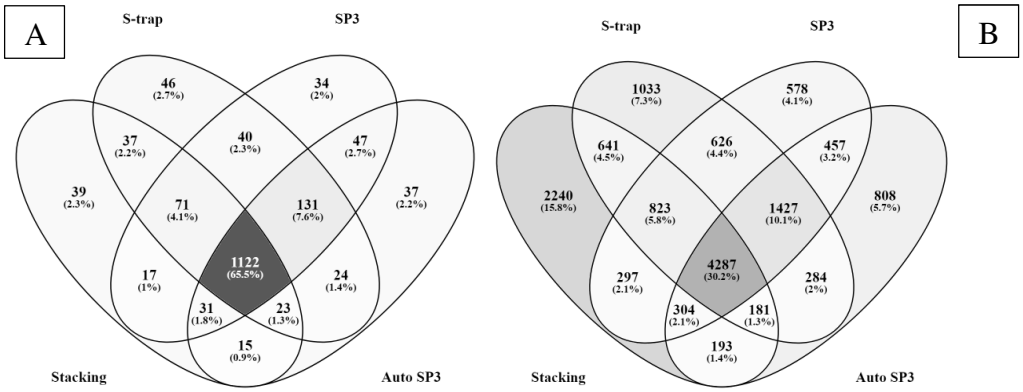

Figure 80 : Recouvrement protéique (A) et de peptidique (B) entre les différents protocoles de digestion (Stacking, S-Trap, SP3 et AutoSP3).

Les protéines identifiées avec les différents protocoles sont assez similaires (65% sont identiques entre les 4 protocoles) mais le recouvrement peptidique est bien plus faible (30%). Les protocoles de digestion étant différents, il est possible que les peptides générés aient des propriétés physico-chimiques différentes.

a. Propriétés chimiques des peptides identifiés

L'hydrophobicité et la longueur des peptides identifiés par les différents protocoles ont été comparés. Bien que le recouvrement peptidique entre les protocoles (Figure 80B) soit faible (30%), les populations extraites sont très semblables comme illustré en figure 81.

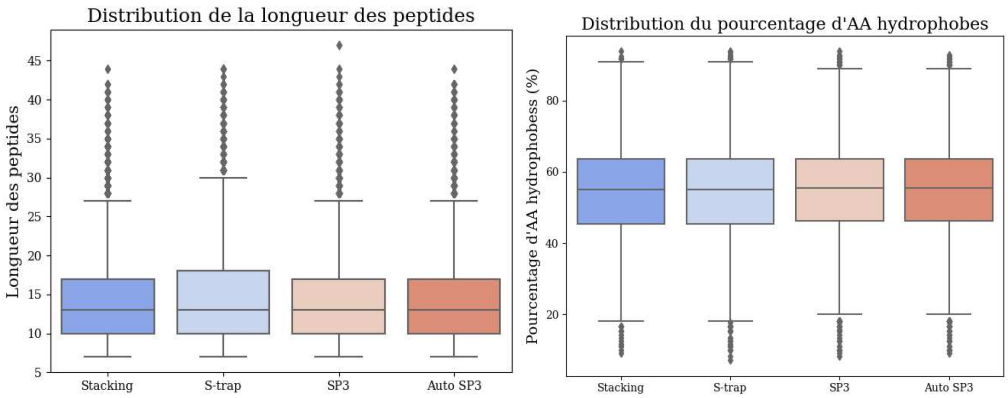

Figure 81 : Distribution de la longueur et du pourcentage d'acide aminés hydrophobes des peptides identifiés avec les différents protocoles de digestion (Stacking, S-Trap, SP3 et AutoSP3).

Le protocole choisi ne semble pas avoir d'influence sur la longueur ou l'hydrophobicité des peptides identifiés. Yang *et al.* avaient aussi reporté que l'utilisation d'un protocole « Stacking » ou d'un protocole « S-Trap » n'influait pas sur l'hydrophobicité des peptides obtenus[68].

b. Taux de clivages manqués

Le taux de clivages manqués, calculé à partir du nombre de peptides identifiés avec un clivage manqué par rapport au nombre total de peptides identifiés, est un bon indicateur de la qualité de la digestion, il est illustré en figure 82 pour chacun des protocoles.

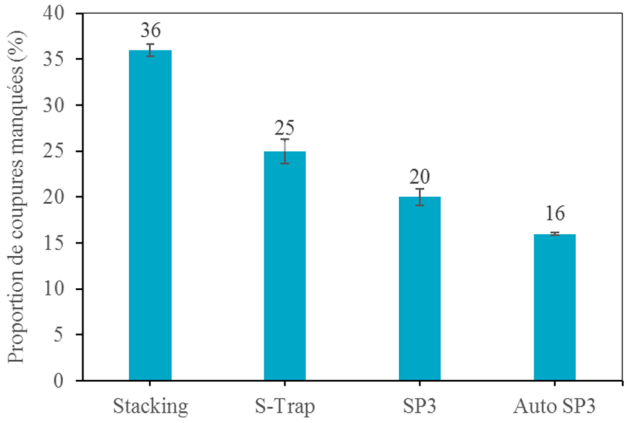

Figure 82 : Proportion de clivages manqués par la trypsine dans les différents protocoles de digestion (Stacking, S-Trap, SP3 et AutoSP3).

La proportion de clivages manqués varie beaucoup entre les protocoles, allant de 36% pour la digestion en gel à 16% pour la digestion automatisée sur billes fonctionnalisées. Le fort pourcentage de clivages manqués du protocole « Stacking » pourrait expliquer le grand nombre de peptides identifiés uniquement par le protocole « Stacking » et le faible recouvrement des peptides identifiés par les différents protocoles. Le nombre de clivages manqués plus élevé obtenu lors de la digestion en gel, par rapport à une digestion sur membrane, avait déjà été observé par Yang *et al.*[68]. L'utilisation de trypsine seule, contre un mélange de trypsine/Lys-C pour les autres digestions, pourrait expliquer le fort taux de clivages manqués pour ce

protocole[53]. Pourtant, ce phénomène n'avait pas été observé lors des optimisations des conditions de digestions sur des échantillons standards (cellules HeLa ou levure). Au contraire, les pourcentages de clivages manqués observés étaient plus faibles lors de ces optimisations par rapport à nos données dans les mêmes conditions de digestion (durée, enzymes, température). Lors des optimisations sur des échantillons standards, 20% de peptides avec un clivage manqué au moins pour le protocole « Stacking » et 17% pour le protocole S-Trap avaient été observés. L'utilisation d'urée est connue pour favoriser les clivages manqués comme décrit par Leon *et al.* et pourrait expliquer l'augmentation générale du nombre de clivages manqués[41].

Il est intéressant d'observer que la digestion est plus efficace avec le protocole automatisé par rapport à la digestion sur billes paramagnétiques en mode manuel, ce qui peut être dû à un ratio protéines/billes plus faible. En diminuant la quantité de protéines par billes, la digestion serait plus efficace. Cette observation doit encore être vérifiée.

En conclusion, il est possible que la variabilité entre les différents protocoles puisse être expliquée par le nombre de clivages manqués engendrant des peptides différents entre les protocoles, mais cette hypothèse reste à prouver.

D. Choix du protocole le plus adapté

Cette étude de comparaison de 4 protocoles de digestion a permis de montrer les avantages et inconvénients de chaque protocole, récapitulés dans la tableau 12. Il est important de garder en mémoire que ces conclusions ne doivent pas être généralisées à tous les types d'échantillons et qu'il est important d'adapter le protocole de préparation d'échantillon en fonction du type d'échantillons.

Tableau 12 : Récapitulatif des avantages et inconvénients des différents protocoles suivant les résultats de cette étude sur le tissu du muscle cardiaque de lérot.

	Stacking	**S-Trap**	**SP3**	**AutoSP3**
Nombre d'identifications (protéines)	+	+	+	+
Nombre d'identifications (peptides)	+	++	+	-
Clivages manqués	+++	++	+	-
Recouvrement protéique entre les réplicas	++	++	++	++
Nombre de protéines quantifiées	+	++	++	++
Répétabilité de la quantification	+	++	++	+++
Durée du protocole	+++	-	+	+

Le protocole « Stacking » est celui qui présente le plus de clivages manqués et reste le plus long à réaliser. Les trois autres protocoles offrent des performances relativement similaires. L'automatisation de la digestion sur billes paramagnétiques doit encore être optimisée pour

l'analyse de tissus afin d'atteindre un nombre d'identifications équivalent à celui obtenu en mode SP3 manuel.

Par cette étude, nous avons donc pu montrer que les nouveaux protocoles développés au laboratoire sont plus adaptés à l'étude du protéome cardiaque que le protocole couramment utilisé ces dernières années car ils permettent d'obtenir des résultats meilleurs ou équivalents dans un temps plus restreint et ceci en augmentant la répétabilité des analyses.

Au final, le protocole retenu pour la suite des analyses du protéome cardiaque est le protocole de digestion sur membrane « S-Trap » car il offre le meilleur compromis. En effet, il allie un nombre important de protéines identifiées et quantifiées et une bonne répétabilité avec un temps de préparation très court. Dans la suite de ce projet, ce protocole va être appliqué pour comprendre les voies biologiques influencées par les acides gras polyinsaturés lors de l'hibernation des lérots.

III. Application du protocole choisi pour étudier l'influence des acides gras polyinsaturés sur le protéome du cœur de lérot au cours de leur hibernation

Les cœurs ont été prélevés à 4 stades physiologiques différents (la période active (S), au début (ET) et à la fin (LT) d'épisodes de torpeur, et durant une phase de réveil (I-IBE)). Les lérots étaient nourris avec deux régimes distincts : soit un régime riche en AGPI n-6 par rapport aux AGPI n-3 (HLA), soit un régime pauvre en AGPI n-6 par rapport aux AGPI n-3 (LLA). Au total, 8 conditions (N=5/groupe) sont comparées entre elles : S-LLA, S-HLA, ET-LLA, ET-HLA, LT-LLA, LT-HLA, I-IBE-LLA, I-IBE-HLA. Une figure décrivant les différents groupes est présentée en figure 83.

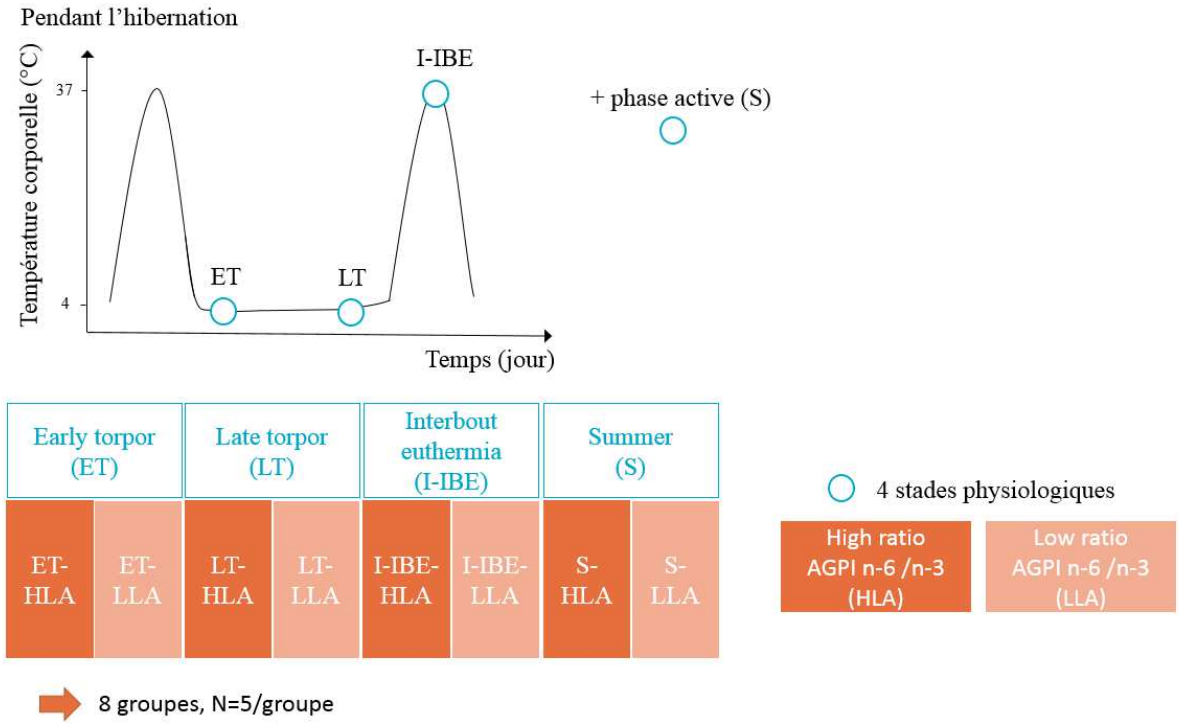

Figure 83 : Description des différents groupes considérés dans cette étude. Prélèvement des échantillons à différents stades biologiques : durant la phase active, au début (ET) et à la fin (LT) d'épisodes de torpeur, durant une phase de réveil (I-IBE) définis en fonction de la température corporelle durant l'hibernation couplé à deux régimes alimentaires (fort (HLA) ou faible (LLA) ratio AGPI n-6/n-3).

Le protocole développé précédemment a été appliqué à ces 40 échantillons. Les détails du protocole sont donnés en partie expérimentale (III.2). Brièvement, entre 55 et 200 mg de cœur de Lérots ont été broyés et lysés avec un tampon contenant de l'urée et du thio-urée. Les protéines ont ensuite été digérées sur membrane suivant le protocole S-Trap. Les peptides obtenus ont été analysés sur un système NanoAcquity couplée à un Q-Exactive HF-X. Les données ont été traitées par le logiciel MaxQuant (v2.0.3.0) avec la banque créée avec les séquences d'*Ictidomis tridecemlineatus* issues de RefSeq et des séquences générées à partir du génome mitochondrial d'*Eliomys quercinus*.

A. Contrôles qualité internes et externes

Pour suivre les performances et la stabilité du couplage pendant la semaine d'analyses, deux contrôles ont été effectués en utilisant soit un standard externe, soit des standards internes.

Des peptides synthétiques iRT ont été ajoutés à chaque échantillon analysé. Puis, les temps de rétention et l'intensité des peptides iRT ont été suivis sur l'ensemble des analyses, des CV médian de 0,64% et 15,8%, respectivement ont été obtenus. Ces valeurs démontrent une excellente stabilité du système chromatographique et spectrométrique. En figure 84 sont représentés les évolutions des temps de rétention et des intensités d'un de ces peptides sur l'ensemble des analyses.

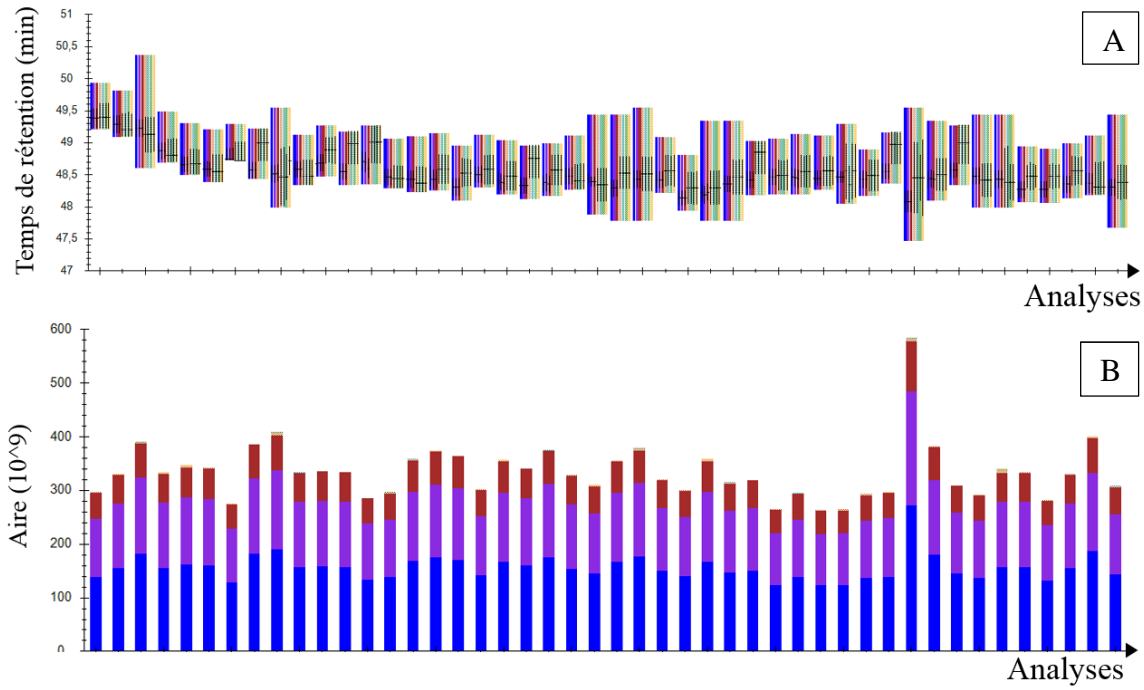

Figure 84 : Suivi du temps de rétention (A) et de l'intensité (B) du peptide ADVTPADFSEWSK au cours des analyses.

Le standard externe est un échantillon de référence contenant la même quantité de protéines issues de chacun des extraits protéiques. Cet échantillon a été soumis au même protocole de digestion que les échantillons puis est injecté à 5 reprises à intervalle régulier au cours de la séquence d'analyses. Le CV médian de l'intensité LFQ des protéines identifiées sur les 5 injections est de 12,1%, attestant de la bonne stabilité du système.

B. Résultats et interprétations biologiques

Sur l'ensemble des 40 analyses, **1626** protéines ont été identifiées. Seules les protéines identifiées dans 4 réplicas pour chacun des groupes (soit une valeur manquante au plus par groupe) ont été retenues pour la quantification, soit **1168** protéines.

Un test *Analysis of variance* [ANOVA] multifactoriel a permis de montrer un effet significatif du régime alimentaire (HLA vs LLA, $p<0,05$) pour 48 protéines et du statut physiologique (S vs ET vs LT vs IBE, $p<0,05$) pour 164 protéines. Il apparait donc que l'effet du régime alimentaire est plus limité que celui du statut physiologique sur le protéome cardiaque. Par ailleurs, des tests statistiques post hoc (TukeyHSD, $p<0,05$) ont montré que les différences majeures se situent entre les échantillons « phase active » et les échantillons issus des animaux hibernants.

Concernant le statut physiologique, un enrichissement en annotations fonctionnelles a été réalisé sur ces 164 protéines différentielles en utilisant l'algorithme DAVID. Le résultat de cet enrichissement est illustré en figure 85.

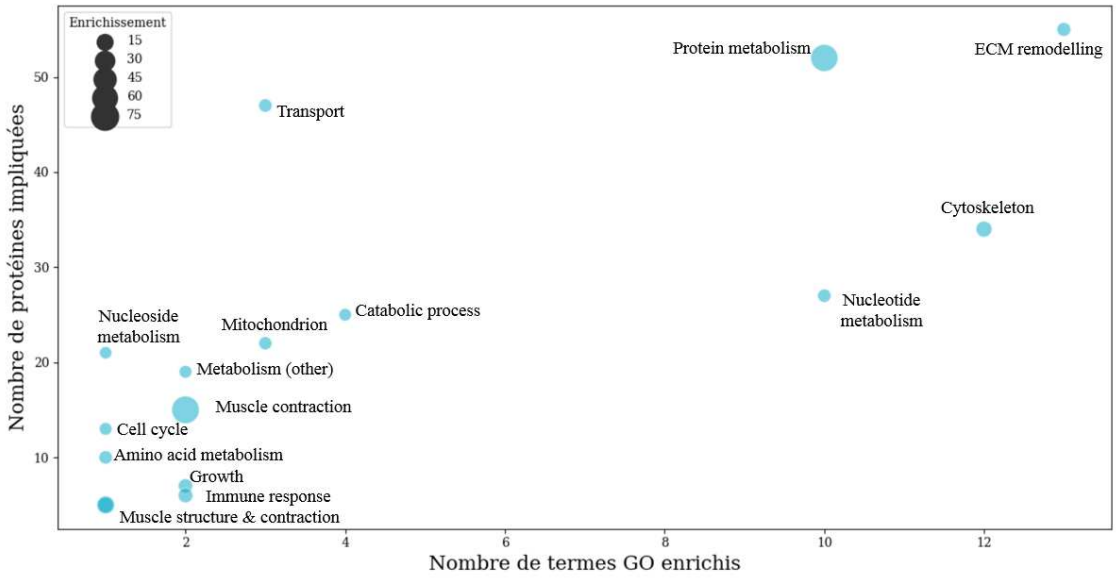

Figure 85 : Résultat du calcul d'enrichissement en annotations fonctionnelles sur les 164 protéines dont l'abondance est impactée par le statut physiologique (ANOVA multi-factorielle, pvalue<0.05). ECM= Matrice extracellulaire

Il apparait que les fonctions les plus impactées par la différence de statut physiologique concernent le métabolisme des protéines, des nucléosides/nucléotides ainsi que les protéines mitochondriales et celles impliquées dans la contraction des muscles. Ces fonctions étaient aussi impactées dans le muscle d'ours brun durant l'hibernation, suggérant des régulations communes non seulement entre le muscle cardiaque et squelettique mais aussi entre ces différentes espèces[36].

Dans un deuxième temps, nous avons donc restreint l'analyse statistique en comparant les statuts I-IBE et « torpeur » composé des groupes ET et LT réunis et en éliminant les valeurs des échantillons « phase active ». Avec ce jeu de données, un effet significatif du régime alimentaire (HLA vs LLA, $p<0.05$) a été mis en évidence pour 47 protéines par analyse ANOVA multifactorielle (dont 26 étaient déjà impactées par le régime alimentaire dans la première analyse statistique avec tous les échantillons). Malheureusement, aucune fonction biologique pertinente n'a été retrouvée significativement enrichie lors du calcul d'enrichissement en annotations fonctionnelles. Une approche plus poussée nécessitant d'examiner dans le détail le rôle connu de ces 47 protéines doit donc être réalisée pour comprendre les voies impactées par les acides gras polyinsaturés durant l'hibernation. Il se pourrait que ces AGPI augmentent l'activité de la SERCA2 à basse température[336].

Concernant la protéine SERCA2, celle-ci a été identifiée et quantifiée dans tous les échantillons, cependant son abondance relative est équivalente dans tous les groupes. Il en est de même pour récepteur à la ryanodine [RYR2], une autre protéine impliquée dans la contraction musculaire *via* le transport d'ion calcium Ca^{2+}. Il se peut que le niveau d'expression de ces protéines ne soit pas le facteur influant le plus sur leur activité. La composition lipidique des membranes, mais aussi l'abondance de molécules inhibitrices (*e.g.* phospholamban et sarcolipine), la présence de modifications post-traductionnelles et la localisation de ces protéines dans la

membrane du réticulum sarcoplasmique devraient être étudiée[353-354]. En effet, comme pour la mitochondrie (voir partie II chapitre1), le message biologique tiré des analyses protéomiques peut être différent entre l'organite et le tissu entier.

IV. Conclusions et perspectives

Grâce à cette approche de protéomique quantitative différentielle, nous avons démontré que le protéome cardiaque du lérot est modifié pendant l'hibernation par rapport à la période active. Un examen plus poussé du jeu de données obtenu doit être mené pour identifier les voies biologiques les plus impactées. De surcroît, l'influence du régime alimentaire doit également être examinée, notamment en investiguant les protéines qui sont différentiellement exprimées à partir des résultats obtenus dans le second jeu de données (après retrait des échantillons « phase active » et de la fusion des deux états de torpeur du jeu de données).

Pour investiguer le protéome du réticulum endoplasmique cardiaque, et notamment la protéine SERCA, une analyse de protéomique différentielle sur des échantillons enrichis en réticulum endoplasmique a été réalisée et les résultats sont encore en cours d'interprétation. Ces conclusions seront mises en lien avec les données obtenues sur le tissu total.

PARTIE III : DEVELOPPMENT DE NOUVELLES APPROCHES POUR ETUDIER LES GLYCATIONS DES PROTEINES SANGUINES

Chapitre 1
Glycation des protéines sanguines chez le diamant mandarin : analyse LC-MS

Ce projet a été réalisé en collaboration avec le Dr. François Criscuolo (IPHC DEPE, Strasbourg).

I. Le paradoxe longévité/forte glycémie chez les oiseaux

Chez l'humain, le maintien d'une forte concentration sanguine en glucose sur le long terme est associé à de nombreuses complications, telles que diabète, l'insuffisance rénale, des maladies inflammatoires et/ou neurodégénératives ainsi qu'à un vieillissement accéléré. Ces complications sont en partie dues à l'accumulation de glycations et de leurs produits de dégradations avancés (AGEs) dans l'organisme[243, 257-258]. Les oiseaux sont considérés comme les vertébrés avec la glycémie la plus élevée[355]. Pourtant, à quelques exceptions près, ils vivent 2 à 3 fois plus longtemps que les mammifères de masse similaire[356]. Ainsi un canari (*Serinus canaria*, 22 g) peut vivre jusqu'à 24 ans lorsqu'une souris (*Mus musculus*, 20 g) ne vivra que 4 ans maximum. Il semblerait donc que les oiseaux soient exempts, du moins en partie, des conséquences négatives liées à des glycémies aussi élevées[356].

L'étude des glycations des protéines chez les oiseaux pourraient apporter une partie de la réponse à ce paradoxe « forte glycémie – faible taux de glycation – longévité ». Plusieurs hypothèses ont été avancées pour expliquer ce paradoxe : une moindre perméabilité et/ou un temps de demi vie plus court de certaines cellules telles que les globules rouges. Il est également possible que les oiseaux aient une plus faible propension à accumuler les AGEs au cours du temps[357-358] ou encore qu'une spécificité des séquences protéiques aviaires limite la réaction de glycation (accessibilité des sites de glycation)[356, 359]. Malgré cela, les glycations sur les protéines aviaires sont peu étudiées.

Chez l'humain, le taux d'hémoglobine glyquée (HbA1c) est le biomarqueur par excellence du diabète car il existe une corrélation positive entre la glycémie sur le long terme (8-12 semaines) et le taux d'HbA1c. Il est admis qu'un taux de plus de 6,5% permet de diagnostiquer le diabète[360] Rendell *et al*. (1985), Miksik *et al*. (1992), Beuchat *et al*. (1998), Ardia *et al.* (2006) et plus récemment Récapet *et al.* (2016) se sont intéressés au taux d'hémoglobine glyquée chez diverses espèces aviaires (poulet, cygne, corbeau, colibri, crécerelle et gobemouche). Les techniques utilisées sont basées sur une séparation de l'hémoglobine glyquée par chromatographie d'affinité au boronate ou échangeuse d'ion suivie par une détection par ultraviolet [UV][361-365]. Dans ces études, les taux de glycation observés sont plus faibles qu'attendus au vu de la glycémie de ces animaux. Ainsi, par exemple, chez le gobemouche ce taux se situe entre 0,7 et 3,7%. Cependant, ces méthodes de dosage non spécifiques présentent de nombreux biais analytiques qui peuvent fausser les résultats, comme les interférences avec d'autres protéines qui portent des amines glyquées.

La spectrométrie de masse est une des techniques de référence permettant l'analyse des protéines glyquées et notamment de l'hémoglobine glyquée chez l'homme[238, 255, 366-367]. En effet, l'addition d'une molécule de sucre (hexose sur une fonction amine conduit à la formation d'un produit d'Amadori, appelée fructosamine , et se traduit par une addition de masse de + 162 Da[239]. Cependant, la mesure des glycations chez les espèces aviaires par spectrométrie de masse a rarement été employée et il n'existe à ce jour aucune mesure de l'hémoglobine glyquée reportée dans la littérature. À notre connaissance, il n'existe qu'une seule étude utilisant la spectrométrie de masse pour mesurer le taux de glycation de protéines *in vivo* chez les oiseaux, mais elle s'intéresse au plasma. Le taux d'albumine glyquée a ainsi été mesuré (entre 14 et 24% selon l'espèce) chez différents oiseaux de proie par LC-MS[266]. Par ailleurs, le taux de glycation de l'albumine purifiée de poulet a été mesuré à environ 40%[359].

Notre objectif était d'identifier et de quantifier les protéines glyquées, notamment la protéine homologue de l'HbA1c, dans le sang de diamant mandarin. Ces mesures réalisées par spectrométrie de masse seront ensuite mises en relation avec des paramètres tels que la glycémie, la masse corporelle, l'âge, ou encore le sexe des individus.

II. Stratégie analytique

Un schéma de la stratégie expérimentale utilisée est représenté en figure 86.

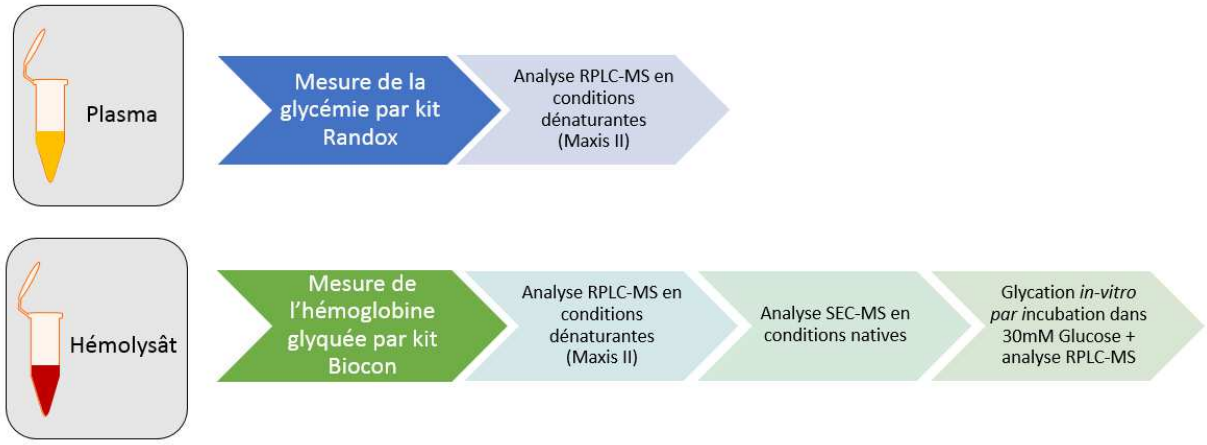

Figure 86 : Représentation schématique des analyses réalisées sur les échantillons de plasma et d'hémolysats.

Le sang de 11 diamants mandarin, *Taeniopygia guttata* (6 mâles, 5 femelles, âgés de 48 à 60 mois) a été prélevé, puis le plasma a été séparé de l'hémolysat par centrifugation. La glycémie de chaque individu a été mesurée dans le plasma par un kit de détection du glucose (Randox Laboratories) suivant la méthode de Trinder permettant de former un complexe détectable par spectroscopie UV-visible à 500 nm. Par ailleurs, le plasma des 11 individus a été analysé, sans digestion préalable des protéines, après dilution dans de l'eau acide (conditions dénaturantes) par chromatographie en phase inverse (silice greffée C8, HPLC 1200 Series (Agilent Technologies) couplée à un spectromètre de masse de type Q-TOF (maXis II, Bruker Daltonik). Pour les protéines identifiées, les courants d'ions correspondant aux 10 états de charge les plus intenses de chaque protéoforme d'intérêt (correspondant à des protéines avec différentes

glycations) ont été extraits et les aires sous la courbe ont été mesurées afin de calculer le taux de glycation de protéines majeures du plasma.

La même stratégie a été appliquée à l'étude de l'hémolysat de ces 11 diamants mandarin et en parallèle de 5 humains.

L'hémoglobine glyquée a été mesurée dans l'hémolysat avec un kit de détection de l'hémoglobine HbA1c (Biocon Biologics, Bangalore, Inde) qui utilise une résine échangeuse d'ions pour isoler l'hémoglobine glyquée suivie d'un dosage par spectroscopie UV-visible. Il s'agit du kit utilisé dans la publication la plus récente mesurant l'hémoglobine glyquée chez les oiseaux (Recapet *et al*., 2016)[361].

De plus, 4 hémolysats aviaires et 4 humains ont été incubés *in vitro* dans un milieu riche en glucose dans le but de forcer la glycation de l'hémoglobine et ainsi suivre la cinétique de la glycation de cette protéine. Par la suite, la structure oligomérique des hémoglobines aviaires et humaines ont été comparées par analyse SEC-MS (NanoAcquity couplée à un Q-TOF Synapt G2) en conditions natives.

Les protocoles détaillés sont présentés dans la partie expérimentale de l'étude publiée dans Experimental Gérontology[368].

III. Analyse des résultats obtenus

A. Glycémie et mesure de la glycation des protéines du plasma

Les glycémies mesurées s'élevaient de 146,5 à 417,5 mg/dL chez les diamants mandarins. Comme observé chez de nombreux oiseaux, ces valeurs sont 2 à 3 fois plus élevées que celles d'un humain sain[369].

Par analyse LC-MS, trois protéines majeures du plasma ont pu être identifiées : la sérotransferrine, l'anhydrase carbonique 2 et l'albumine (protéine la plus abondante dans le plasma). Pour ces trois protéines, différentes protéoformes portant de 0 à 3 molécules de glucose ont été identifiées. Les spectres déconvolués de ces protéines sont présentés en figure 87A. La proportion de chacune de ces protéoformes a été évaluée (87B).

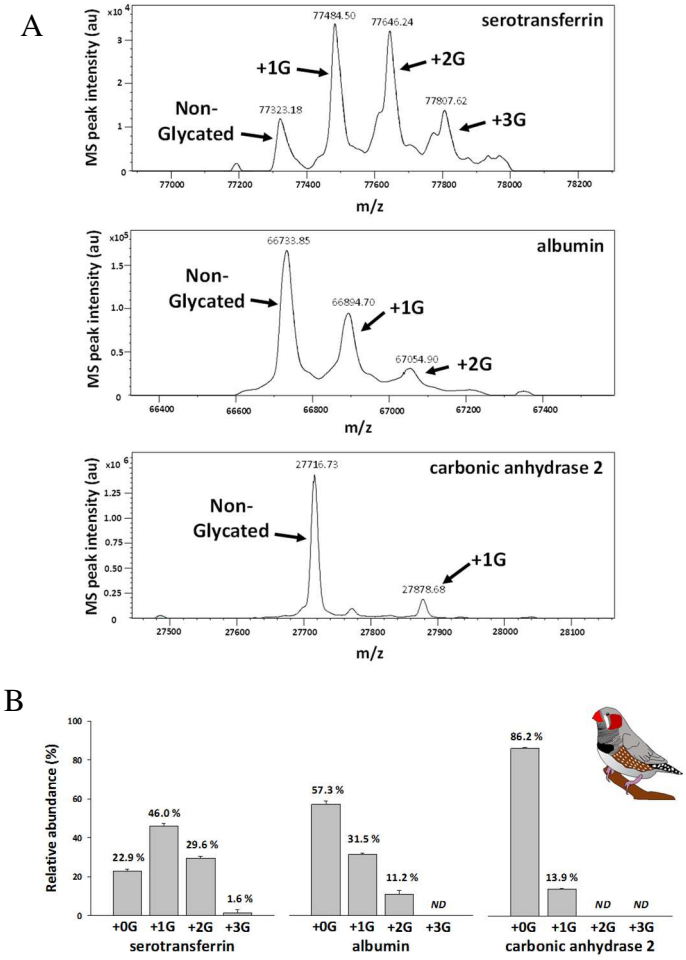

Figure 87 : Glycation des protéines plasmatiques du diamant mandarin ; spectre déconvolué de la sérotransferrine, l'albumine et l'anhydrase carbonique, le nombre de fructosamine portée par chaque protéoforme est indiqué sur la figure (A), proportion de chaque protéoforme pour chacune des protéines (B).

La forme majoritaire de la sérotransferrine est la protéoforme portant une seule glycation (46 ± 1,5%), suivie de la forme doublement glyquée (29,6 ± 1%). La sérotransferrine non modifiée représente moins d'un quart des 4 protéoformes détectées (22,9 ± 1,1%). Le nombre de glycations est plus limité sur les deux autres protéines. La forme non glyquée de l'albumine représentait 57,3 ± 1,6%, la forme mono glyquée 31,5 ± 0,4% et la forme doublement glyquée 11,2 ± 1,8%. Seules deux protéoformes de l'anhydrase carbonique ont été identifiées : la forme non glyquée (86,2 ± 0,4%) et la forme modifiée avec 1 glucose (13,9 ± 0,4%).

En comparant les taux de glycation total de chaque protéine avec la glycémie, le sexe, l'âge et la masse corporelle, une corrélation s'est avérée significative (sérotransferrine et glycémie, r = 0,610, t = 2,311, df = 9, P = 0,046). Par ailleurs, l'état de glycation de l'anhydrase carbonique est positivement corrélé avec l'âge des individus de façon significative (r = 0,699, t = 2,938, df = 9, P = 0.017).

La très forte proportion de sérotransferrine glyquée observée (77%) est bien plus élevée que celle décrite chez des patients diabétiques (5% contre 2% pour les patients sains)[250]. Il en est de même pour le taux d'albumine glyquée (43%), qui est décrit entre 10 et 15% pour les patients

sains, entre 12 et 25% pour les patients diabétiques. Ces taux peuvent atteindre jusqu'à 94% dans des cas extrêmes où la glycémie est très mal contrôlée[252, 370-371]. Cette valeur est également plus élevée que celle observée avec une technique similaire chez les oiseaux de proie (14 - 24%)[266] mais elle est similaire à la valeur observée *in vitro* par Zuck *et al.*(2017) chez le poulet[359].

B. Glycation de l'hémoglobine

Dans la littérature, le taux d'hémoglobine glyquée avait été mesuré grâce à des kits, pour lesquels la détection est effectuée par lecture UV. Afin de pouvoir comparer nos résultats aux données de la littérature, nous avons utilisé le kit Biocon utilisé par Recapet *et al*.[361]. Grâce à cette méthode non spécifique, nous avons mesuré un taux d'hémoglobine glyquée entre 2,6 et 4,0%, soit un taux à peine plus élevé que ce qu'avait mesuré Recapet *et al*. chez le gobemouche (0,7-3,7%)[361]. Cependant, l'analyse par LC-MS d'hémolysats n'a pas permis de détecter une masse moléculaire à +162 sur la chaine beta, chaine portant la modification sur la valine N-terminale chez l'homme. Les spectres de masse déconvolués de la chaine bêta de l'hémoglobine du diamant mandarin et de l'homme sont présentés en figure 88 (panneaux A et B). Chez l'humain, la MS avait pourtant permis de détecter des taux d'HbA1c de 3,4% chez des individus sains comme représenté sur le spectre déconvolué de la figure 3 (panneau C). À ce stade, l'absence de détection d'hémoglobine glyquée chez l'oiseau peut être due à trois facteurs : soit la concentration en hémoglobine glyquée est en dessous de la limite de détection du spectromètre de masse, soit aucune forme glyquée de l'hémoglobine n'existe chez le diamant mandarin, soit il existe des interférences analytiques nous empêchant d'identifier la forme glyquée de la protéine.

Pour apporter des preuves expérimentales à ces hypothèses, nous avons cherché à forcer la glycation de l'hémoglobine en incubant les hémolysats dans une solution riche en glucose pendant plusieurs jours et mesurer ensuite l'évolution du taux de glycation. Les résultats sont présentés sur la figure 88 (panneau D).

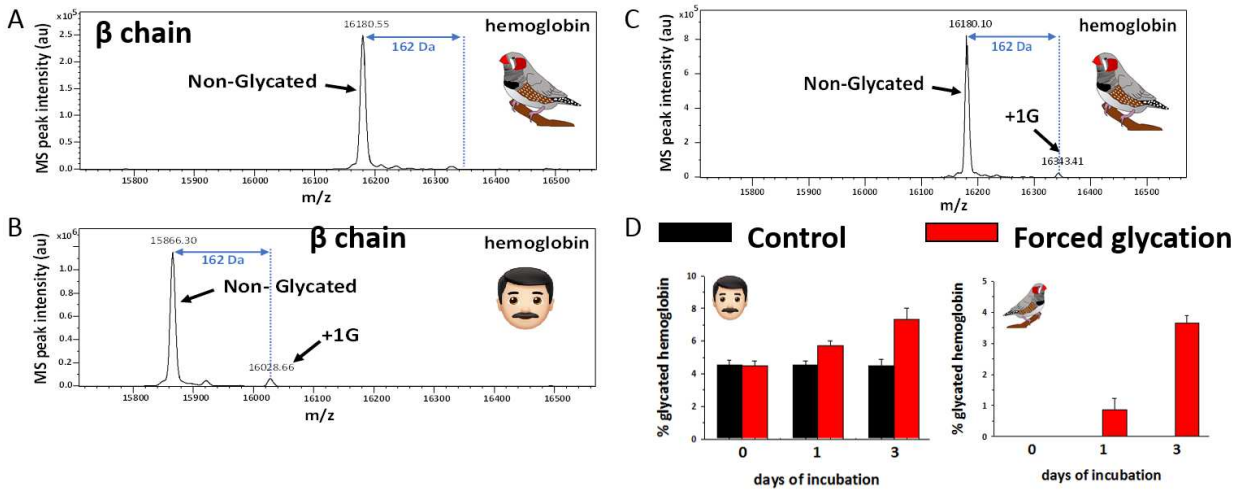

Figure 88 : Glycation de l'hémoglobine du diamant mandarin et de l'humain : spectre déconvolué de la chaine bêta de l'hémoglobine du diamant mandarin *in vivo* (A) ; spectre déconvolué de la chaine bêta de l'hémoglobine humaine *in vivo* (B) ; spectre déconvolué de la chaine bêta de l'hémoglobine du diamant mandarin glyquée *in vitro* (C) ; proportion d'hémoglobine glyquée au cours du temps chez l'homme et le diamant mandarin dans un milieu avec (forced glycation) ou sans (control) glucose (D).

Dans les échantillons incubés dans un milieu « contrôle » sans glucose, aucune évolution du taux d'hémoglobine glyquée initial n'est détectée chez l'humain sur les 3 jours d'incubation (0,36% de coefficient de variation moyen chez les individus sains) et aucune glycation n'est observée chez le diamant mandarin. Au contraire, dans les échantillons incubés dans un milieu riche en glucose, de très faibles taux d'hémoglobine glyquée ont été mesurés après un jour d'incubation (1,44 ± 0,03%) pour les hémolysats de 3 oiseaux seulement sur les 5. Au bout de 3 jours, une forme glyquée de l'hémoglobine a été détectée dans les échantillons issus de tous les oiseaux mais les taux restaient relativement faibles (3,7 ± 0,3%). Concernant l'hémoglobine humaine, une augmentation de son taux de glycation a également été observée, passant de 4,5 ± 0,3% au jour 0 à 7,3 ± 0,7% au jour 3 chez les individus sains. Cette expérience a par ailleurs permis de montrer que notre méthode LC-MS était suffisamment sensible pour détecter l'hémoglobine glyquée même pour des taux aussi bas que 1,5%, confirmant ainsi que les valeurs obtenues par l'utilisation du kit Biocon sont surestimées et que la concentration d'hémoglobine glyquée chez les oiseaux est très faible voire qu'elle en est absente.

Cette absence ou la très faible occurrence de la glycation sur l'hémoglobine du diamant mandarin pourrait être expliquée par différents facteurs. Premièrement, le temps de demi-vie des érythrocytes aviaires est bien plus court (30-42 jours) que chez l'homme (60-120 jours), ce qui les protégerait de l'accumulation d'hémoglobine glyquée. Deuxièmement, les globules rouges aviaires contiennent des noyaux contrairement aux globules des mammifères. Ce noyau pourrait permettre aux globules d'exprimer des enzymes permettant d'éliminer les glycations telles que la kinase fructosamine 3, connue pour déglyquer les protéines[372-373]. Cette hypothèse est toutefois remise en cause par l'étude de Delplanque *et al.*, selon laquelle, cette enzyme, bien que très active dans les globules humains, ne présente qu'une activité très réduite dans les érythrocytes de poulet[374]. D'après Swergold *et al.* (2003), il serait même possible qu'il existe

chez les oiseaux un autre mécanisme permettant de retirer les fructosamines qui seraient indépendants de cette enzyme[375]. Troisièmement, il a été montré chez le poulet que la concentration en glucose dans les globules rouges est plus faible[376]. Enfin, l'hémoglobine de diamant mandarin pourrait présenter des particularités structurales la protégeant du phénomène de glycation. Il se pourrait, par exemple que les sites de glycation potentiels (lysines, arginines, N-ter) ne soient pas accessibles pour réagir avec les molécules de glucose. C'est pourquoi, nous avons mené des études de MS native (SEC-MS) pour tenter d'élucider cette structure.

C. Détermination de la structure quaternaire du complexe d'hémoglobine

Ces analyses sont réalisées en conditions natives (voir partie I chapitre 1 I.a.) par SEC-MS. Cette technique sépare les composés en fonction de leur volume hydrodynamique avant qu'ils soient analysés par spectrométrie de masse. Chez l'homme, l'hémoglobine est un tétramère constitué de 2 chaines alpha et de 2 chaines bêta et de quatre hèmes (de masse moléculaire 616 Da) pour une masse totale de 64 454 Da[377]. Lors de l'analyse par SEC-MS de l'hémoglobine humaine, nous avons bien détecté ce tétramère ainsi qu'une faible population de dimère. Tandis que pour le diamant mandarin, 3 populations de tétramères de masses différentes ont été identifiées ainsi qu'une population de dimères (12,5%) et de monomères (3,1%). En réalisant un ajustement gaussien sur les profils UV obtenus en utilisant l'outil IGOR Pro (WaveMetrics, Oregon, USA), la proportion de chaque tétramère a pu être évaluée. Les profils UV ainsi que les proportions de chaque complexe identifié sont présentés en figure 89.

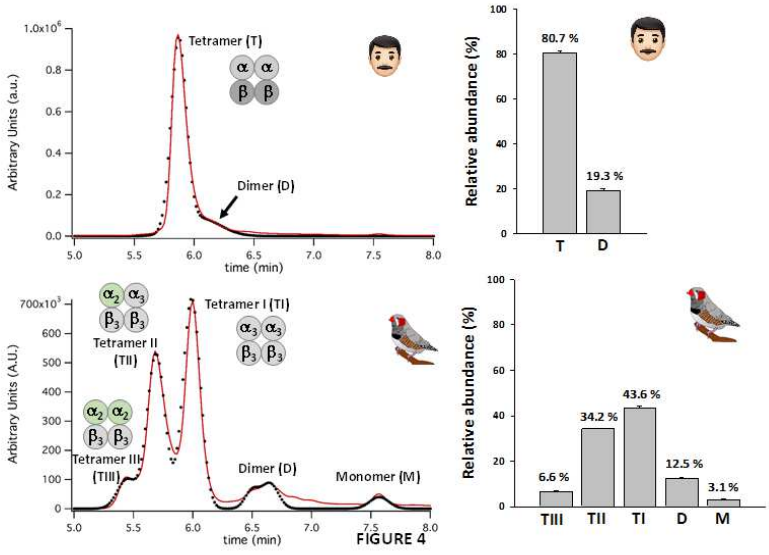

Figure 89 : Ajustement gaussien sur les profils UV de l'hémolysat du diamant mandarin et de l'humain ainsi que les proportions relatives de chaque complexe identifié chez les deux espèces.

En conditions dénaturantes, le complexe est dissocié avant l'injection sur la colonne et ce sont les différentes chaines qui sont séparées par chromatographie et non les différents complexes. Ici, en conditions natives, l'application de hautes énergies en source induit la dissociation des complexes, ce qui permet l'identification des chaines constituant les tétramères. Les masses identifiées sont corrélées avec les masses des chaines déterminées en conditions dénaturantes.

Les différentes chaines constituant le complexe d'hémoglobine chez le diamant mandarin ont été étudiées par Alex *et al.* : 3 chaines alpha (alpha 1, 2, 3) et 3 chaines beta (beta 1, 2, 3) ont été identifiées. Seules les chaines alpha2, alpha3 et beta3 ont été observées chez les diamants mandarins adultes[378]. Dans nos données, le tétramère (TI) le plus abondant (43,6%) est constitué de deux chaines alpha3 et de deux chaines beta3. Le tétramère (TII) (34,2%) a subi une substitution d'une des chaines alpha3 par une chaine alpha2 et sur le dernier (TIII, 6,6%), les deux chaines alpha3 ont été substituées en alpha2. Ces tétramères portent comme chez l'homme 4 hèmes (616 Da) ainsi qu'un cofacteur de masse moléculaire 580 Da. Des analyses MS/MS de ce cofacteur ont été réalisées et les fragments obtenus pourraient correspondre à des fragments issus de l'inositol pentaphosphate, une molécule connue pour moduler l'affinité de l'oxygène pour le complexe d'hémoglobine chez les oiseaux (figure 90)[379]. Ce cofacteur pourrait également permettre de le protéger du stress oxydatif, comme c'est le cas pour le 2,3-diphosphoglycerate chez les mammifères[380].

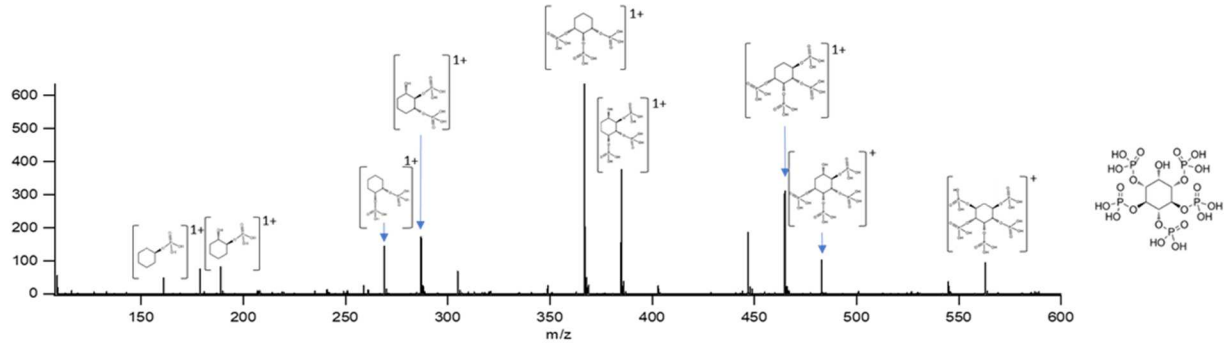

Figure 90 : Spectre de fragmentation de l'inositol pentaphosphate obtenu sur un Q-TOF Synapt G2 après fragmentation CID (les structures proposées sont putatives).

Bien que la SEC soit une méthode peu résolutive, les trois tétramères, qui ont des masses très similaires, ont été séparés par la colonne. Ceci signifie que les tétramères ont des conformations très différentes. Plus un composé a un volume hydrodynamique important (structure plus large), plus il va être élué rapidement. De plus, l'intensité des monomères issus de la dissociation en source permet également de formuler des hypothèses quant à la conformation spatiale des tétramères : plus cette intensité est élevée, plus il est aisé de fragmenter le composé en source et moins le complexe est compact. Ainsi, il semblerait que le tétramère TIII soit moins compact que les 2 autres tétramères car l'intensité du monomère après dissociation en source est plus élevée et c'est le premier tétramère à être élué en SEC (5,4 min (TIII) vs 5,6 (TII) vs 6 min(TI)). Les spectres non déconvolués de chaque tétramère sont présentés en figure 91.

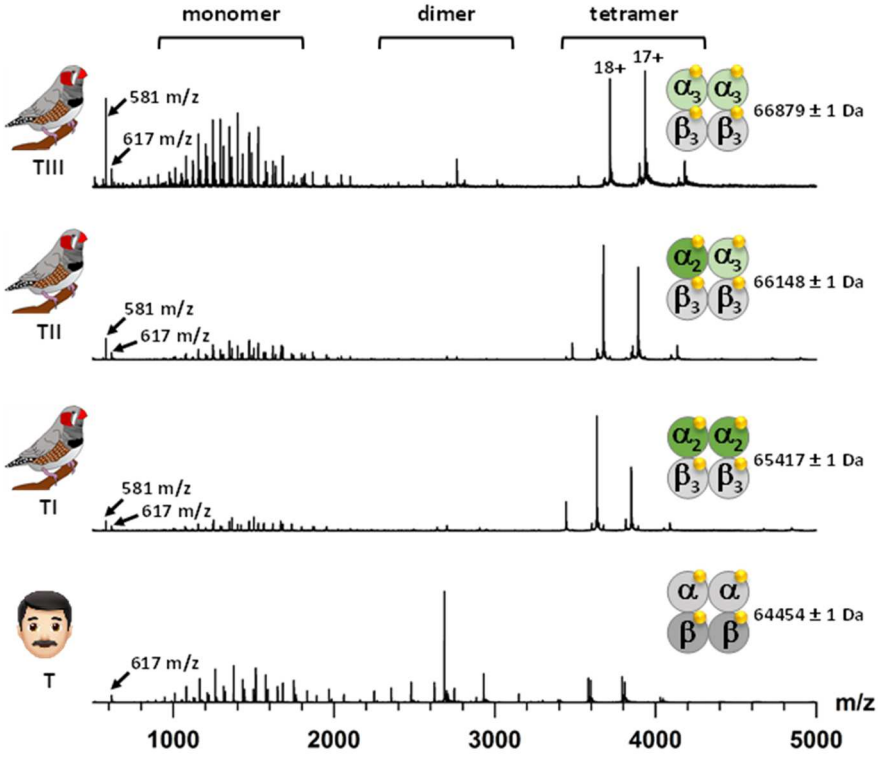

Figure 91 : Spectres bruts des 3 tétramères du diamant mandarin (TI, TII et TIII) ainsi que du tétramère unique de l'homme.

Il existe une corrélation entre la masse des individus et les proportions en tétramères et dimères (pearson's correlation, r=-0,752, p=0,031) : les individus les plus lourds sont ceux qui présentent les proportions en tétramères les plus élevées par rapport aux dimères. Plusieurs études montrent que l'affinité pour l'oxygène de l'hémoglobine dépend de sa structure oligomérique[381-383]. Ceci pourrait suggérer que les individus avec une meilleure condition corporelle pourraient délivrer de l'oxygène aux organes de manière optimisée. De plus, l'existence de plusieurs tétramères chez le diamant mandarin pourrait aussi jouer un rôle dans la régulation fine de l'apport d'oxygène aux différents organes. Cette particularité permettrait à l'oiseau de s'adapter à des changements environnementaux (accessibilité de la nourriture, maladie…) ou à des changements physiologiques liés aux différentes étapes de la vie de l'individu.

IV. Conclusion et perspectives

En utilisant pour la première fois à notre connaissance une méthode d'analyse spécifique par MS pour mesurer l'hémoglobine glyquée *in vivo* chez une espèce aviaire, nous avons pu montrer que l'hémoglobine du diamant mandarin n'était pas glyquée, ou alors en quantité négligeable (< 1%), compte tenu de la forte glycémie de ces oiseaux. Ceci pourrait s'expliquer par une structure particulière des hémoglobines aviaires où les sites de glycation (lysine, arginine, N-ter) ne seraient pas exposés. En parallèle, nous avons déjà pu déterminer qu'il existe 3 structures oligomériques différentes, les deux structures majoritaires sont les plus compactes par rapport à la troisième qui est très minoritaire. Ceci pourrait expliquer en partie la résistance de l'hémoglobine à la glycation. Une caractérisation structurale plus poussée du complexe de

l'hémoglobine pourrait également apporter des réponses plus précises sur l'absence d'hémoglobine glyquée chez les oiseaux. Des analyses spectrométriques d'échanges hydrogène-deutérium [HDX-MS][384], de top-down en conditions natives[30] ou dénaturantes[385] (TD-MS) permettraient d'obtenir des informations sur la structure tridimensionnelle de l'hémoglobine, notamment concernant l'accessibilité des sites de glycation et leur localisation sur la séquence. Ceci permettrait ensuite d'élucider si les régions les plus exposées contiennent ou non des sites de glycation en fonction des complexes formés et/ou de leurs conformations.

Les méthodes de LC-MS (RP-MS et SEC-MS) utilisées dans ce projet vont maintenant permettre d'analyser les hémoglobines de nombreuses espèces aviaires afin de tenter d'établir des liens entre la phylogénie des espèces et les potentielles différences de structure de l'hémoglobine des oiseaux. Ce travail devrait aussi permettre de généraliser l'observation de l'absence de glycation sur l'hémoglobine chez les oiseaux.

Si nous avons montré l'absence de glycation sur l'hémoglobine de diamant mandarins *in vivo*, l'albumine et la sérotransferrine plasmatiques sont quant à elles très fortement glyquées par rapport à ce qui est connu chez l'homme. Des analyses protéomiques « *bottom-up* » devraient permettre de localiser les sites de glycation et de mesurer leur abondance relative, tout en augmentant la profondeur de l'analyse, qui pour l'instant reste limitée aux protéines majoritaires et à la seule forme fructosamine, et d'accéder ainsi une caractérisation plus poussée des protéines glyquées du plasma[269]. Toutefois, pour réaliser ce type d'analyse, des développements analytiques sont nécessaires, qui sont décrits dans le chapitre suivant. À terme, cette méthode « *bottom-up* » pourrait apporter des connaissances sur les glycations chez les oiseaux et leur possibles conséquences, notamment au niveau du métabolisme des sucres.

Les résultats de cette étude ont été publiés dans Experimental Gérontology[368].

Experimental Gerontology 164 (2022) 111811

Contents lists available at ScienceDirect

Experimental Gerontology

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

Resistance to glycation in the zebra finch: Mass spectrometry-based analysis and its perspectives for evolutionary studies of aging

Charlotte Brun [a,b], Oscar Hernandez-Alba [a,b], Agnès Hovasse [a,b], François Criscuolo [a,*], Christine Schaeffer-Reiss [a,b,1], Fabrice Bertile [a,b,**,1]

[a] Université de Strasbourg, CNRS, IPHC UMR7178, Strasbourg, France
[b] Infrastructure Nationale de Protéomique ProFi, FR2048 Strasbourg, France

ARTICLE INFO

Section Editor: Michal Masternak

Keywords:
Glycation
Glucose
Hemoglobin
Albumin
Mass spectrometry
Bird

ABSTRACT

In humans, hyperglycemia is associated with protein glycation, which may contribute to aging. Strikingly, birds usually outlive mammals of the same body mass, while exhibiting high plasma glucose levels. However, how birds succeed in escaping pro-aging effects of glycation remains unknown. Using a specific mass spectrometry-based approach in captive zebra finches of known age, we recorded high glycaemia values but no glycated hemoglobin form was found. Still, we showed that zebra finch hemoglobin can be glycated *in vitro*, albeit only to a limited extent compared to its human homologue. This may be due to peculiar structural features, as supported by the unusual presence of three different tetramer populations with balanced proportions and a still bound cofactor that could be inositol pentaphosphate. High levels of the glycated forms of zebra finch plasma serotransferrin, carbonic anhydrase 2, and albumin were measured. Glucose, age or body mass correlations with either plasma glycated proteins or hemoglobin isoforms suggest that those variables may be future molecular tools of choice to monitor glycation and its link with individual fitness. Our molecular advance may help determine how evolution succeeded in associating flying ability, high blood glucose and long lifespan in birds.

1. Introduction

The survival of living systems is conditioned by the maintenance of many variables in the internal milieu within a limited range of values (Bernard, 1879), this steady state being called homeostasis (Calow, 1982; Cannon, 1929). One key aspect of homeostasis is the regulation at a stable level of blood glucose (Herman and Kahn, 2006), any drift from the species-specific set-points leading to deleterious consequences for health (Polakof et al., 2011). Glucose is indeed the main circulating energy fuel in organisms (Frayn, 2010) and the main energy source for key organs like the brain (Butterfield and Halliwell, 2019). Because dysregulation of glucose metabolism is more frequent as organisms age, glucose homeostasis has been highlighted as part of the deregulated nutrient-sensing, one of the hallmarks of aging (Lopez-Otin et al., 2013). This physiological decline is related to a loss of insulin sensitivity and to an increase in circulating glucose that give rise to insulin resistance (Gluckman and Hanson, 2004). Metabolic inflexibility corresponds to a loss of organism's ability to adapt substrate oxidation rates to variations in fuel availability, *e.g.* to switch from carbohydrate to lipid metabolism in fasting *vs.* fed nutritional conditions (Rynders et al., 2018). In mammals, metabolic inflexibility is mainly involved in the development of insulin resistance (Bergouignan et al., 2011; Goodpaster and Sparks, 2017), leading to a reduction in life expectancy. Similarly in birds, baseline glucose level has been reported to be higher in response to adverse environmental conditions, and it was negatively associated with lifespan of zebra finches (*Taeniopygia guttata*) (Montoya et al., 2018). Therefore, understanding the mechanisms linking variations in glucose metabolism with individual fitness (*e.g.* lifespan) takes on a particular importance in evolutionary biology.

One astonishing biological observation is the remarkable slow aging rate of many bird species in comparison to similar-sized mammals (Holmes and Harper, 2018), with as an example the 10 g hummingbirds living up to 10–15 years in the wild (Calder, 1990). Interestingly, birds are often presented as the animals with the highest blood glucose levels

* Corresponding author.
** Correspondence to: F. Bertile, Université de Strasbourg, CNRS, IPHC UMR7178, Strasbourg, France.
E-mail addresses: francois.criscuolo@iphc.cnrs.fr (F. Criscuolo), fbertile@unistra.fr (F. Bertile).
[1] These authors share senior authorship.

https://doi.org/10.1016/j.exger.2022.111811
Received 30 November 2021; Received in revised form 15 March 2022; Accepted 18 April 2022
Available online 25 April 2022
0531-5565/© 2022 Published by Elsevier Inc.

among vertebrates (Polakof et al., 2011). Despite this and their high metabolic rate (Prinzinger, 1993), their lifespans are proportionally 2–3 times greater than in comparable sized-mammals (Holmes et al., 2001). This paradoxical observation actually raises a large number of yet unresolved questions about the exact health-related impact of elevated plasma glucose concentrations in birds, and their possible mechanisms of resilience.

The ability of reducing sugars to chemically react with protein amino groups is a phenomenon called glycation (Tessier, 2010). It corresponds to the non-enzymatic covalent bonding of sugar molecules, *e.g.* glucose or fructose, to a protein molecule to form a Schiff base, which then undergoes an irreversible Amadori rearrangement in the case of glucose or Heyns rearrangement in the case of fructose. Further reactions yield a highly heterogeneous group of advanced glycation end products (AGEs, (Poulsen et al., 2013; Ramasamy et al., 2010)). Abnormal glycation levels in humans, *e.g.* when glycaemia is high, promotes oxidative stress and diabetes, cardiovascular and neurodegenerative diseases, and may ultimately contribute to aging (Masoro et al., 1989; O'Brien and Timmins, 1994; Sell et al., 1996; Semba et al., 2010; Ulrich and Cerami, 2001). Recent hypotheses put forward to explain the bird paradox "high glucose – low glycation – long lifespan" included a lower propensity to accumulate AGEs over time (Klandorf et al., 1999; Rattiste et al., 2015), a lower permeability to glucose and/or a shorter half-life of certain cell types that are sensitive to AGEs accumulation (*i.e.* red blood cells) or specificities of avian protein sequences that may limit potential for Amadori reactions (Holmes et al., 2001; Zuck et al., 2017). However, only very few studies have investigated protein glycation in birds.

To address the glycation – longevity bird's paradox, one remaining challenge is our ability to analyze protein glycation. To date, glycated proteins have been measured using chromatography methods (Andersson and Gustafsson, 1995) or using commercial kits implementing a step of glycated protein capture based on boronate affinity (Ardia, 2006; Beuchat and Chong, 1998; Miksik and Hodny, 1992; Rendell et al., 1985) or cation exchange chromatography (Récapet et al., 2016) and a detection using spectrophotometry. However, these methods clearly lack specificity and sensitivity to detect protein-specific modifications, as shown recently by Suo and collaborators (Suo et al., 2019). Moreover, the measure of glycated hemoglobin has sometimes been performed from the whole blood (Ardia, 2006; Beuchat and Chong, 1998), when it should have been done from red blood cells only (Récapet et al., 2016). Therefore, it is highly probable that interferences had come from other blood proteins, glycated or not, thus introducing measurement biases. One of the gold standard methods to specifically analyze glycation events is mass spectrometry (Jeppsson et al., 2002; Priego-Capote et al., 2014; Soboleva et al., 2017), which enables addressing protein site-specificity of glycation (and thereby the intimate mechanisms of AGE formation). To our knowledge, only one bird study used liquid chromatography coupled to mass spectrometry to measure albumin glycation rates *in vivo* (Ingram et al., 2017). In the present paper, our main objective was to develop the use of mass spectrometry-based methods for the non-invasive study, from blood sampling in captive zebra finches, of glycation levels of several proteins. To do so, we used known-age individuals in which glucose and glycated forms of albumin, serotransferrin and carbonic anhydrase 2 from plasma and hemoglobin from hemolysates were accurately measured.

2. Materials and methods

2.1. Bird and human samples

We used blood sample aliquots of two sub-groups of captive zebra finches (*Taeniopygia guttata*) that were part of an experiment on social stress conducted in our laboratory (Quque et al., unpublished). Our individuals all belong to the experimental control group (not socially stressed). This study complied with the 'Principles of Animal Care' publication no. 86-23, revised 1985 of the National Institutes of Health, and with current legislation (L87-848) on animal experimentation in France. It was approved by an independent ethical committee and authorized by the French Ministry of Research (authorization reference APAFIS#12019-2018012511525879).

All birds, kept in unisex cages (0.57 × 0.31 × 0.39 m) and housed in standard conditions (24 °C, 13:11 h light-dark light cycle), came from an in-house colony. They were fed *ad-libitum* with a commercial mix of seeds for exotic birds (tropical finches Prestige Premium Versele laga©), completed weekly with vitamins and proteins. The first group was of 11 individuals (6 males, 5 females, age 48–60 months) and was used to access glycation proteins proportion in whole blood (hemoglobin and plasma proteins). Because we wanted to avoid freezing-melting cycles that may alter proteins structures, a second group of 8 birds was later used to assess hemoglobin isoform proportions (4 males, 4 females, age 19–60 months). In both cases, whole blood was collected (originally for the social stress experiment) in the early morning (50 μL from the brachial vein using heparinized syringe) then immediately centrifuged at 4 °C (10 min, 2000 ×*g*), and aliquots of red blood cells and plasma were stored at −80 °C until utilization.

In parallel, self-collecting blood from a finger prick enabled us to obtain specimens from three healthy and two diabetic human individuals. Briefly, blood was collected using heparinized capillaries and immediately centrifuged (10 min, 2000 ×*g*, 4 °C) for preparing red blood cells before they were stored at −70 °C until utilization.

2.2. Plasma glucose measurement

Plasma glucose levels were measured enzymatically for all zebra finch individuals using the GLUC-PAP kit from Randox (UK). The intra-individual variability was assessed using the intra-class coefficient (ICC value = 0.849; (Cicchetti, 1994)).

2.3. Protein glycation measurement in zebra finch red blood cells using a commercial kit

The fraction of glycated hemoglobin was measured using the Biocon® Diagnostik HbA_1 kit (Biocon Diagnostik, Germany) in red blood cells from a sub-sample (due to blood volume constraints) of our first group of zebra finches ($n = 9$). In a first step, glycohemoglobin (HbA1) has been separated from non-glycated hemoglobin using a cation-exchanged resin. It is worth noting here that this cation-exchanged resin is not specific to hemoglobin as any other positively charged molecule can also interact with it. Then, the percent of glycohemoglobin was calculated, as indicated by the manufacturer, from the measurement of the absorbance (415 nm) of the two fractions and the kit calibrator. Again, the specificity can be disputed here, because although the absorption of the heme molecule is maximum at 415 nm, one cannot exclude the possible absorption of other molecules at this wavelength. Because of volume constraint, samples have not been run in duplicate. Intra-assay coefficient of variation of the HbA_1 kit used in birds has already been reported (Récapet et al., 2016).

2.4. Protein glycation measurement in zebra finch and human samples using mass spectrometry

Unless otherwise specified, all chemicals and reagents were purchased from ThermoFisher Scientific (Waltham, MA, USA).

Protein glycation was assessed in red blood cells of one human individual taken as a positive control, and in the plasma and red blood cells of each zebra finch individuals from our first group ($n = 11$). Twenty-two and 199 μL of water containing 0.1% of formic acid were used to dilute 3 μL of plasma and 1 μL of red blood cells, respectively. Then, 5 μL of each sample were injected on an Agilent 1200 Series HPLC system (Agilent Technologies, Paolo Alto, USA) coupled to a quadrupole-time-of-flight (Q-TOF) mass spectrometer equipped with an electrospray source (maXis II, Bruker Daltonik GmbH, Bremen,

Germany). This mass spectrometer was selected here because of its high resolution (*e.g.* twice that of the Synapt G2), which allows accurate determination of *m/z* values for intact proteins in denaturing conditions. Protein separation was achieved using a Vydac 208TP C8 HPLC column (i.d. 2.1 × 250 mm, 300 Å, 5 μm particle size, Grace, Columbia, MD, USA) maintained at 60 °C. The solvent system consisted of 0.1% trifluoroacetic acid in water (solvent A) and 0.1% trifluoroacetic acid in acetonitrile (solvent B). The elution was performed at a flow rate of 250 μL/min, using a 40-minute linear gradient from 5% to 40% of solvent B. The mass spectrometer was operated in positive mode, with the source temperature set to 220 °C, the dry gas flow to 4 L/min, and the electrospray voltage to 4500 V. Mass calibration of the TOF was achieved using ESI-L low concentration Tuning Mix (Agilent Technologies) in the mass range of 322–2722 *m/z*. Full scan MS spectra (500–3000 *m/z*) were acquired while the focus option was activated. The system was fully controlled by Hystar v3.2 (Bruker Daltonik GmbH).

MS raw data were processed using DataAnalysis v4.3 (Bruker Daltonik GmbH). Before deconvolution, MS spectra were smoothed using a Gaussian algorithm. All studied proteins eluted at different times (serotransferrin: 31 min, albumin: 34 min, carbonic anhydrase 2: 39 min), each within a narrow period of less than 1 min. From the analysis of red blood cell samples, charge states ranging from +9 to +18 were consistently detected with a high intensity for hemoglobin in both species. For plasma samples, the charge state of the 10 most intense ions ranged from +47 to +56, +41 to +50, and + 21 to +30, for serotransferrin, albumin, and carbonic anhydrase, respectively. For each non-glycated, mono- or multi-glycated protein forms, an extracted ion chromatogram was obtained by summing the signals from the ten most abundant charge state ions. Areas under curves were then used to obtain the relative abundances of each protein form of interest. The proportion of a given protein form was calculated as the ratio, multiplied by 100, of its abundance to the sum of abundances of all its forms.

2.5. In vitro forced glycation of zebra finch and human hemoglobin

Unless otherwise specified, all chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO, USA).

Hemoglobin glycation (intact β-chain) was assessed *in vitro* under forced conditions, in both a sub-sample of zebra finches from our first group (again because of volume constraints, $n = 5$) and humans ($n = 2$ diabetic and $n = 3$ non-diabetic). Hemolysates were obtained after one freeze-thaw cycle, and then they were centrifuged at 12000 ×*g* for 10 min at room temperature. Hemoglobin-containing supernatants (1 μL) were diluted (1/100, V/V) either in the control buffer (50 mM Tris, pH = 7.5) or glucose buffer (50 mM Tris, 30 mM glucose, pH = 7.5). Forced glycation was then performed by incubation at 37 °C. From twenty-five μL of sample, collected either immediately (d0), after 24 h (d1) or 72 h (d3) of incubation, proteins were reduced by addition of 10 μL of denaturing solution (56 mM tris(2-carboxyethyl)phosphine, 50 mM Tris, pH = 7.5) and 65 μL of guanidine buffer (3.5 M guanidine, 50 mM Tris, pH = 7.5). After incubation at 57 °C for 60 min, quenching was performed by addition of 1 μL of trifluoroacetic acid. Samples were then stored at −80 °C until analysis. Mass spectrometry analysis was performed following the same procedure as for native bird and human samples (see above), except for the chromatographic gradient. Here, elution was performed using the following steps: 5% of solvent B for 7 min, from 5% to 33% of solvent B in 1 min, and finally a 20-min linear gradient from 33% to 70% of solvent B.

2.6. Determination of the oligomeric state of zebra finch and human hemoglobin using nMS and nTD-MS

Hemoglobin quaternary structure was assessed in our second group of zebra finch ($n = 8$) and human ($n = 2$ diabetic and $n = 3$ non-diabetic) red blood cells diluted (1/50, V/V) in a 10 mM ammonium acetate buffer (pH 6.8). After a 5-minute centrifugation (3000 ×*g*), supernatants (5 μL for human and 10 μL for zebra finch samples) were injected on an Acquity UPLC H-Class system equipped with an UV detector operating at 280 nm and 214 nm (Waters, Manchester, UK), and coupled to a Q-TOF mass spectrometer equipped with an electrospray source (Synapt G2 HDMS mass spectrometer, Waters, Manchester, UK). This mass spectrometer is particularly adapted and tuned to conduct native MS analyses, allowing the transmission of molecular ions of proteins with a high molecular weight and reduced net charge state without inducing the dissociation of the non-covalent interactions of the analytes. Hemoglobin oligomers were separated using an Acquity BEH SEC column (i.d. 4.6 × 150 mm, 200 Å, 1.7 μm particle size; Waters) maintained at room temperature. The mobile phase was composed of 10 mM ammonium acetate (pH 6.8). The elution was performed at a flow rate of 250 μL/min, according to an isocratic gradient. Native mass spectrometry (nMS) was performed to identify the different oligomeric populations with the mass spectrometer operating in the sensitivity mode and positive polarity with a capillary voltage of 3.0 kV. Desolvation and source temperatures were set to 200 °C and 70 °C, respectively. Desolvation and cone gas flow rates were 750 L/h and 60 L/h, respectively. The cone voltage was set to 60 V and the backing pressure of the *Z*-spray source to 6 mbar. Spectra (500–10,000 *m/z*) were acquired without ion selection in the quadrupole. In a second step, native top-down mass spectrometry (nTD-MS) was used to fragment the hemoglobin tetramers and further identify tetramer subunits. For this purpose, the mass spectrometer source was operating under harsh conditions, *i.e.* with increased source and desolvation temperatures (90 °C and 450 °C, respectively), giving rise to the dissociation of oligomeric hemoglobin. Thereby, the experimental masses of the obtained monomers along with the experimental masses of the intact scaffolds were used to determine the composition of the different hemoglobin populations. The system was controlled by MassLynx v4.1 software (Waters, Manchester, UK). A multiple Gaussian fitting of the 280 nm chromatographic profiles was used to determine the proportions of the different hemoglobin populations (IGOR Pro, WaveMetrics, Oregon, USA).

2.7. Statistics

All statistics were done using R v 4.0.2. We followed five complementary analytic steps to analyze the mass spectrometry glycation measurements in birds. First, using Chi-square proportion comparisons (Pearson's Chi-squared test *from chisq.test* function), we tested for each protein target whether the proportions of non-glycated *versus* glycated forms were differing from an even distribution. Secondly, using pairwise comparisons of proportions (*prop.test* R function), we tested whether the proportion of glycated forms differed among proteins, by running successive two by two comparisons (*i.e.* glycated serotransferrin *vs.* glycated albumin, glycated serotransferrin *vs.* glycated carbonic anhydrase and glycated albumin *vs.* glycated carbonic anhydrase). To control for multiple testing, we applied Bonferroni correction to all tests. Thirdly, using standardized data of gathered glycation levels (thereafter Total Glycation Level [TGL], as the summed weighted proportion of multiple glycated protein forms, *i.e.* glycated serotransferrin, albumin and carbonic anhydrase, see Jeffs et al. (Jeffs et al., 2017)), we tested whether individual levels of glycation were predicted by plasma glucose concentration, sex or body mass. While doing this, we also checked for existing correlations between plasma glucose levels and the glycated proportion of each protein, using Pearson's correlation coefficients. Using then TGL as the independent variable, we ran a General Linear Model (GLM) using the *glm2* package, with glucose levels, age, sex and body mass as fixed factors. TGL normal distribution was checked using the *shapiro-Wilk normality.test* and *qqnorm* R functions (*fitdistrplus* R package, (Delignette-Muller and Dutang, 2015)). Model selection was based on Akaike Information Criterion (AIC) corrected for small sample sizes (*Mumin* R package, (Barton, 2020)). The same procedure was used to check for any significant correlation between the glycated hemoglobin levels measured using the commercial kit and individual age, sex,

body mass or glucose levels. Fourthly, we checked for differences in the dynamic of glycation rates in our experiment of forced glycation of human and bird hemoglobin. Because of the small sample size for the human part of the experiment, we could not test for the effect of diabetic ($n = 2$) compared to non-diabetic status ($n = 3$). We restricted our analyses to distinguish human ($n = 5$) and bird (n = 5) dynamics of hemoglobin glycation. Distribution of data satisfied with normality (*qqplot*, Shapiro-Wilk normality test, see results). We then used a General Linear Mixed Model with individual glycation levels as the response variable and time of exposure to glucose (day 0, 1 and 3) and species (Human and Bird) as fixed factors, with their interaction. Repeated measures on the same sample at different times (d0, d1, d3) were controlled using sample identity as random factor. Posthoc comparisons were conducted using *Emmeans* R package (Lenth et al., 2018), with Tukey's correction for multiple testing. As a last step, we used the same PCA and GLM – *Mumin* approach to test for any correlational link between hemoglobin isoforms' proportion and individual sex, body mass, age and plasma glucose levels. After checking for normality of the isoforms distributions (*i.e.* monomer, dimer, tetramer I, II and III) using the *shapiro-Wilk normality. test* and *qqnorm* R functions, we merged those variables using a Principal component analysis (PCA; (Le et al., 2008)) that extracted two principal axes (eigenvalue >1, explaining 75.8% of the total variance,

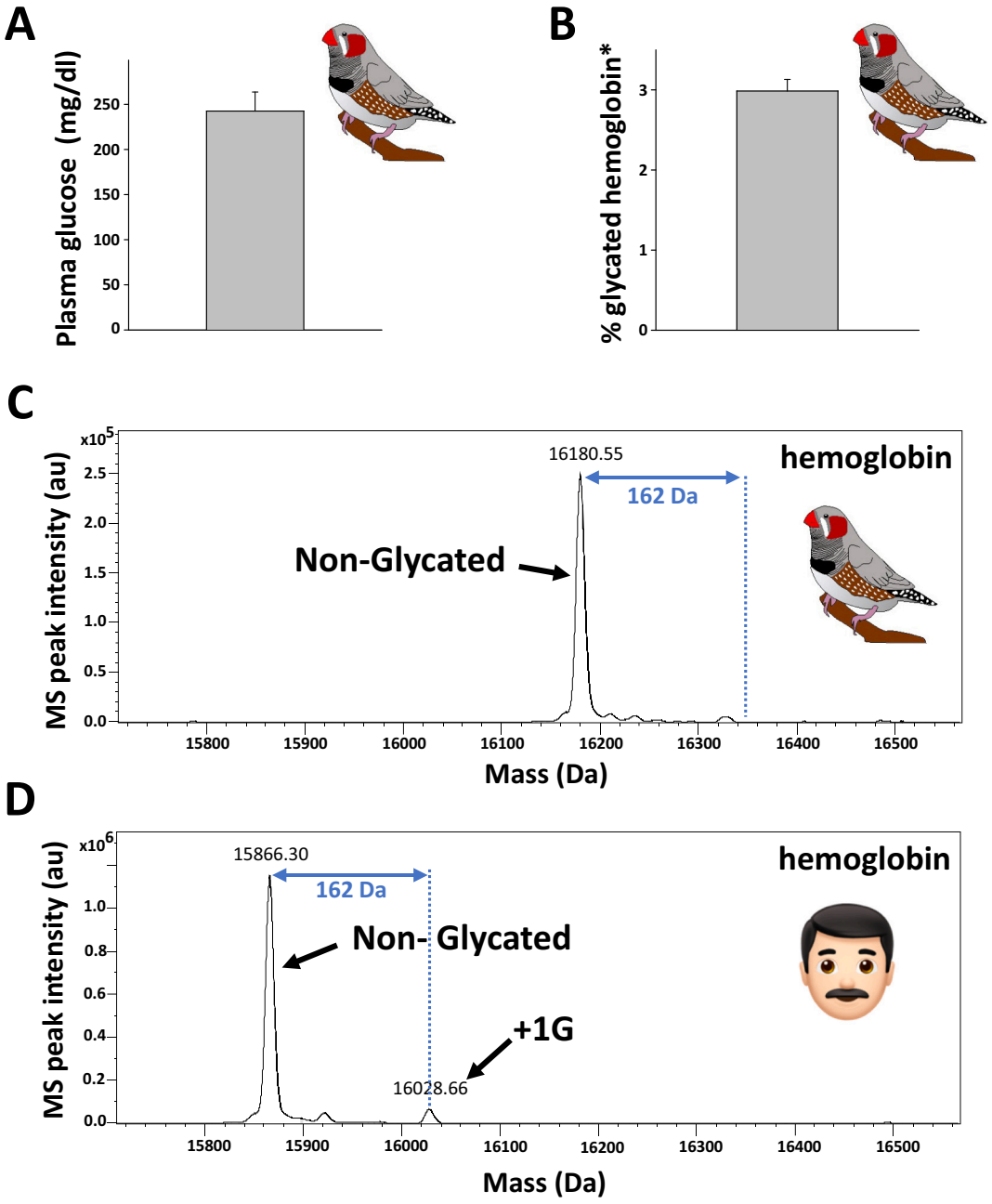

Fig. 1. Glycemia and hemoglobin glycation in zebra finch and human samples
Plasma glucose concentration in zebra finches is expressed as the mean ± sem of 11 independent determinations (panel A). The fraction of glycated hemoglobin (mean ± sem of 9 independent determinations) was measured using the Biocon® Diagnostik HbA_1 kit (Biocon Diagnostik, Germany) (panel B; * this method lacks specificity; because hemoglobin is by far the most abundant protein in red blood cells, it is assumed that the signal is actually mostly due to glycated hemoglobin). Examples of deconvoluted mass spectra are shown for zebra finch (panel C) and human (panel D) hemoglobin β-chains. The number of glucose molecules bound to proteins is indicated for each peak.

Supplementary Fig. 1a): PC1 positively loaded with hemoglobin dimer proportion (0,857) and negatively loaded with tetramer III proportion (−0,943) and PC2 positively loaded with hemoglobin monomer (0.858). In fact, tetramer III and dimer isoforms' proportions are significantly negatively correlated (Supplementary Fig. 1b, Pearson's correlation coefficient, $r = -0{,}815$, $t = -3449$, $P = 0{,}014$). GLM models included either PC1 or PC2 as response variables and individual sex, body mass, age and plasma glucose levels as fixed factors. In all models, collinearity among fixed factors were checked to be <5 using the *vif* function of the *car* R package (Fox and Weiberg, 2019).

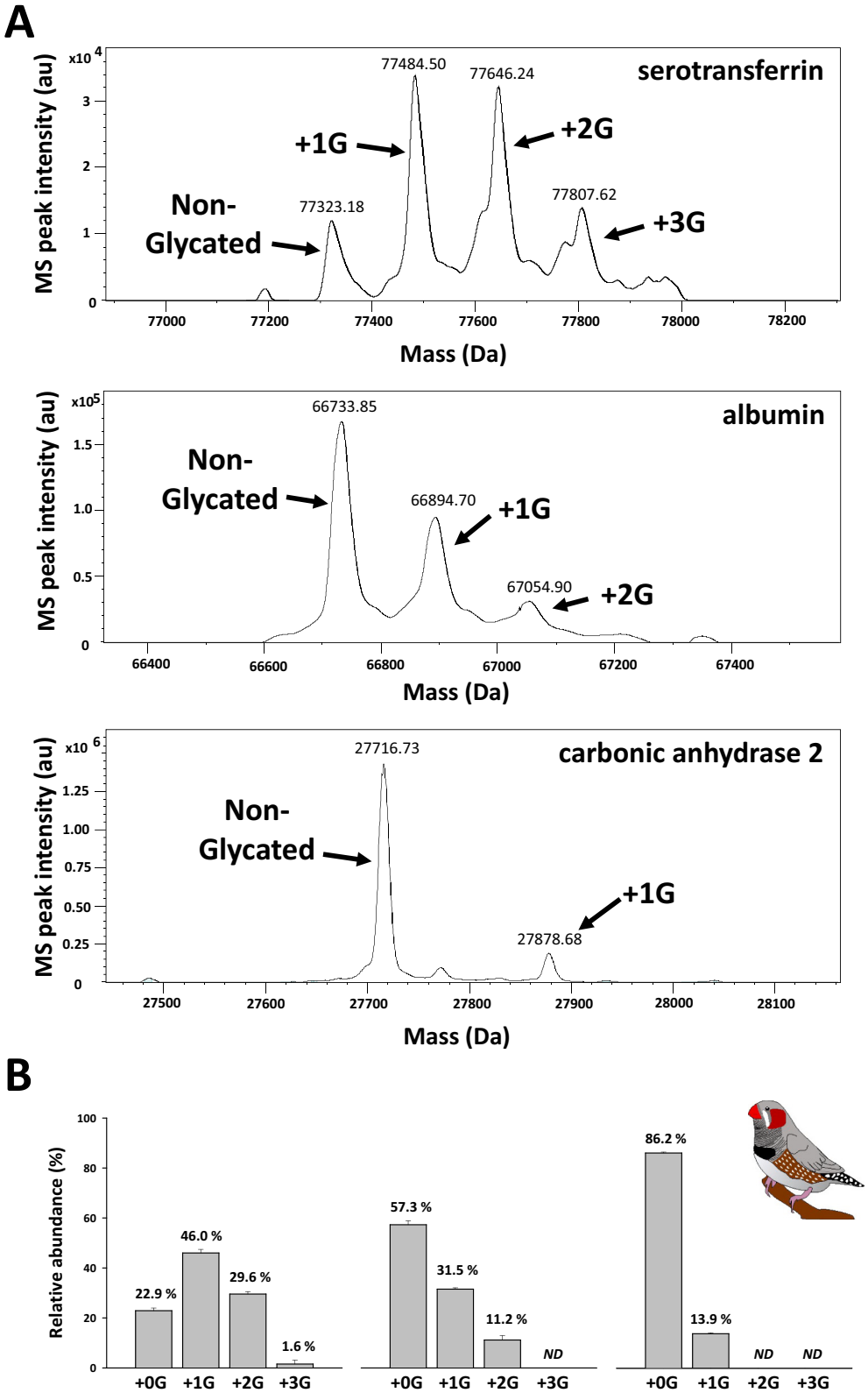

Fig. 2. Protein glycation in zebra finch plasma
Examples of deconvoluted mass spectra, obtained through liquid chromatography online with mass spectrometry, are shown for serotransferrin, albumin, and carbonic anhydrase 2 (panel A). The number of glucose molecules bound to proteins is indicated for each peak. For a given protein, the proportion of each of its detected forms is expressed as the mean ± sem of 11 independent determinations (panel B).

3. Results

3.1. Glycaemia and protein glycation in zebra finch and human samples

Detailed raw mass spectra along with LC chromatograms are given in Supplementary Figs. 2–4.

Plasma glucose concentrations in zebra finches ranged from 146.5 to 417.5 mg/dl (mean ± sem: 242.5 ± 21.1 mg/dl; Fig. 1A). Concerning glycation, isolating the glycohemoglobin fraction from red blood cells using a cation-exchanged resin then performing spectrophotometric measurements enabled us to calculate the percent of glycated hemoglobin in zebra finches at values of 2.99 ± 0.14% (Fig. 1B). Those glycated hemoglobin levels were not significantly associated with individual body mass, age, sex or glucose levels (GLM, $t = 1459$, $P = 0.149$; $t = 1648$, $P = 0.099$; $t = 0.565$, $P = 0.572$, t = 0,355, P = 0,723, respectively). However, only the non-glycated form of hemoglobin was detected using mass spectrometry (Fig. 1C). Indeed, no peak was detected with a mass shift of +162 Da, and the three small peaks that were detected (16,210.98 Da, 16,236.60 Da, and 16,331.49 Da) exhibit mass shifts of, respectively, +30.43 Da, +56.05 Da, and + 150.94 Da. Such mass shifts do not correspond to glyoxal-modified (carboxymethylamine +58.04 Da) or methylglyoxal-modified (carboxyethylamine +72.06 Da) hemoglobin forms. Either they may correspond to minor proteins other than hemoglobin or they could just reflect deconvolution artefacts due to their low intensity. Conversely, we were able to easily detect both the non-glycated and glycated (HbA1c) hemoglobin forms from human hemolysates, the latter representing 3.4% of total hemoglobin (Fig. 1D).

Mass spectrometry analysis enabled us to visualize all the zebra finch plasma proteins we decided to focus on, both in their non-modified form (*i.e.* non-glycated) and in their mono- or multi-glycated forms (Fig. 2A). Hence, proteoforms bearing up to one, two, and three glucose molecules were observed for plasma carbonic anhydrase 2, albumin, and serotransferrin, respectively. The proportions of each plasma protein form was calculated (Fig. 2B). For serotransferrin, the higher proportion was for the mono-glycated form (46.0 ± 1.5%), the doubly glycated form (29.6 ± 1.0%), and the non-modified form (22.9 ± 1.1%). It should be noted that we have detected the serotransferrin triply glycated form only from one individual, where balanced proportions of the non-glycated and triply glycated forms (15% and 17%, respectively), and of the doubly and triply glycated forms (33% and 34%, respectively) were observed. For albumin and carbonic anhydrase 2, the proportion of the non-glycated form was of 57.3 ± 1.6% and 86.2 ± 0.4%, respectively. Mono-glycated forms were detected for these two plasma proteins (31.5 ± 0.5% and 13.8 ± 0.4%, respectively), and a doubly glycated form was finally also detected for albumin (11.2 ± 1.8%). For those three targeted proteins, all proportions were found to significantly differ from even distributions (serotransferrin, Chi-square = 39.6; albumin, Chi-square = 31.8; carbonic anhydrase, Chi-square = 51.8; all P ⟨0,001). It should be noted here that the small peak that can be seen for carbonic anhydrase 2 at 27772.18 Da (mass shift +55.45 Da) could correspond to a deconvolution artefact, or to an eventual sequence variant where an Arginine was replaced a Threonine. However, no such sequence variant is known, *e.g.* in the sequence of the human chicken carbonic anhydrase 2.

When comparing the proportion of glycated *versus* non-glycated forms (*i.e.* irrespective of the number of glycated sites) for given proteins, we found that there was a higher proportion of glycated serotransferrin than of glycated albumin (77% *vs.* 43%; Two-sample test for equality of proportions, chi-square = 22.69, 95% confidence interval (0.203–0.477); $p < 0.001$). The proportion of glycated serotransferrin also turned out to be higher than the proportion of glycated carbonic anhydrase (77% *vs.* 14%; chi-square = 77.51, 95% CI (0.513–0.765); $p < 0.001$). Finally, the proportion of glycated albumin was higher than of glycated carbonic anhydrase (43% *vs.* 14%; chi-square = 19.24, 95% CI (−0.419 to −0.162); $p < 0.001$).

Proportion of total glycated levels (weighted for multiple glycations) were found to follow a normal distribution (Shapiro-Wilk normality test, w = 0.998, $p = 0.231$). The output of the GLM analysis testing for relationships between individual TGL and age, sex, glucose or body mass are given in Table 1. Selection model did not find any model significantly better than the null model. None of the fixed factors was significantly linked to TGL, despite the fact that TGL and glucose levels or age tended to be positively related. When we checked for correlations between plasma glucose levels and the proportions of each glycated protein, we only found a significant correlation for glycated serotransferrin ($r = 0.731$, $t = 3.211$, df = 9, $P = 0.011$; glycated albumin, $r = 0.07$, $t = 0.211$, df = 9, $P = 0.837$; glycated carbonic anhydrase 2, $r = 0.393$, $t = 1.281$, df = 9, $P = 0.232$). When checking for age correlations, we only found a positive linear link with glycated carbonic anhydrase 2 to be significantly affected ($r = 0.700$, $t = 2.938$, df = 9, $P = 0.017$; glycated albumin, $r = 0.329$, $t = 1.045$, df = 9, $P = 0.323$; glycated serotransferrin, $r = 0.500$, $t = 1.732$, df = 9, $P = 0.118$). Very similar results were obtained in the case where multiple glycations were not weighted (data not shown).

3.2. Protein glycation in forced glycated samples

To go further in the understanding of bird protein glycation, we next focused on hemoglobin modification dynamics under *in vitro* conditions of experimentally forced glycation. The experiment compared mass spectrometry-based measurements for human and zebra finch beta-globins. Sequences of human beta-globin (or hemoglobin subunit beta; P68871 in UniprotKb database) and zebra finch beta-globin (B5G106 in UniproKb or NP_001232396 in NCBI database) are both of same length (147 amino acids), and their very similar amino acid composition leads them to share 70.7% sequence identity (83.7% sequence similarity). Finally, all known glycation sites in the human protein are conserved in the bird homologue (Fig. 3A).

Mass spectrometry analysis enabled us to show that percent of glycated human beta-globin (HbA1c) remained at stable levels (mean Coefficient of Variation (CV) of 0.36% for humans and 1.97% for diabetic humans) in the *in vitro* control conditions over the three days of the experiment (Fig. 3B). In the presence of glucose, the percent of glycated hemoglobin increased from 4.5 ± 0.3% at day 0 to 7.3 ± 0.7% at day 3 for humans, and from 9.1 ± 2.1% at day 0 to 11.4 ± 2.2% at day 3 for diabetic humans. Concerning zebra finch hemoglobin, no glycated form was detected in the control condition from day 0 to day 3, whereas detection was recorded in the presence of glucose (Fig. 3C) at day 1 for only 3 birds of 5 (1.44 ± 0.03%) and at higher levels for the five birds at day 3 (3.7 ± 0.3%). Table 2 shows the output of the GLMM analysis: the *Day* x *Species* interaction was significant, indicating that the dynamics of

Table 1
General linear model estimates for the linear relationship between total glycation levels (summed proportion of glycated serotranferrin, albumin and carbonic anhydrase) and plasma glucose levels, sex, age (months) and body mass (g). The model presented here is the full model, as the selection process did not indicate that any model was significantly better than the null model (AIC < 2). Nagelkerke R^2 indicates the power of explanation of the selected model (total variance).

	Total glycation level				
Predictors	Estimates	Std. error	CI	Statistic	P
Intercept	−0.22	0.63	−1.46–1.02	−035	0.739
Plasma glucose (mg/dL)	0.5	0.34	−0.16–1.17	1.48	0.189
Sex [M]	0.41	1.04	−1.64–2.45	0.39	0.711
Body mass (g)	0.31	0.45	−0.57–1.2	0.7	0.513
Age	0.43	0.57	−0.68–1.54	0.76	0.477
Observations	11				
R^2 Nagelkerke	0.602				

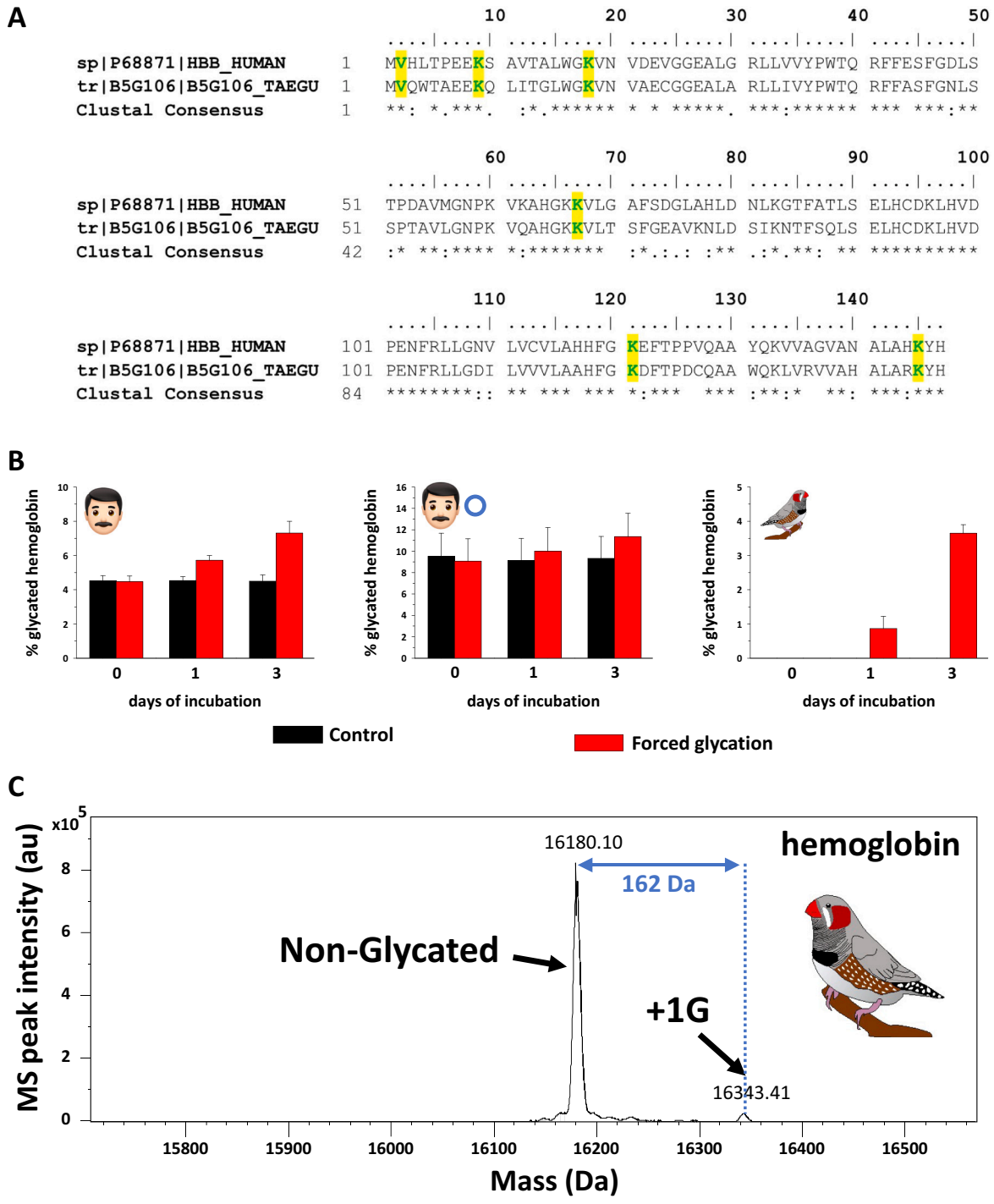

Fig. 3. Human and zebra finch hemoglobin sequences and their glycation under forced *in vitro* conditions
Amino acid sequences of human and zebra finch β-globins were aligned using Clustal W program (panel A). An asterisk (*) indicates positions that have a single and fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties with a score higher than 0.5 in the Gonnet PAM 250 matrix. A period (.) indicates conservation between groups of weakly similar properties with a score lower or equal to 0.5 in the Gonnet PAM 250 matrix. The amino acids in green highlight the known glycation sites of human hemoglobin and whether they appear to be present in the avian counterpart. The percent of glycated hemoglobin is expressed as the mean ± sem of 3 independent determinations for humans, 2 for diabetic humans (blue circle), and 5 for zebra finches (panel B). An example of deconvoluted mass spectrum is shown for zebra finch hemoglobin (panel C).

glycation reactions differed between human and bird hemoglobin. Posthoc Tukey tests indicated that birds had lower levels of glycated hemoglobin than humans did at day 0 (estimate −6.310 ± 1308, $t = -4825$, $P = 0.017$), day 1 (−6.566 ± 1308, $t = -5021$, $P = 0.013$) and day 3 (estimate −5.278 ± 1308, $t = -4825$, $P = 0.050$). In humans, the level of glycated hemoglobin increased steadily at each time (day 0-day1, −1.122 ± 0,258, $t = -4344$, $P = 0.008$; day 1 – day 3, −1.500 ± 0,258, $t = -5808$, $P < 0.001$) while it only increased significant at day 3 in birds (day 0-day1, −0.866 ± 0,258, $t = -3353$, $P = 0.061$; day 1 – day 3, −2.788 ± 0,258, $t = -10{,}795$, $P < 0.001$).

Table 2
General linear mixed model estimates testing the differences in the dynamic of glycation of human and bird (zebra finch) hemoglobin (Species effect) over time (Day effect) when experimentally exposed to high glucose concentrations. σ^2 and τ00 values are used to calculate the intraclass correlation coefficient of the random factor (ICC). N_{Id} corresponds to the number of levels for the random factor, while the total number of observations are also given. Marginal and Conditional R^2 indicates the percentage of variance explained without and with the random factor (sample Id). P-values below 0.05 are highlighted in bold.

	Glycated hemoglobin (%)		
Predictors	Estimates	CI	P
Intercept	−0.00	−1.81–1.81	1.000
Day [d1]	0.87	0.36–1.37	**0.001**
Day [d3]	3.65	3.15–4.16	**<0.001**
Species [human]	6.31	3.75–8.87	**<0.001**
Day [d1] * Species [human]	0.26	−0.46 to −0.32	0.483
Day [d3] * Species [human]	−1.03		**0.005**
Random effects			
σ^2	0.17		
τ00 Id	4.11		
ICC	0.96		
N_{Id}	10		
Observations	30		
Marginal R^2 / Conditional R^2	0.726 / 0.989		

3.3. Oligomeric states of zebra finch and human hemoglobin

Using nMS in combination with nTD-MS, information about the quaternary structure of zebra finch *versus* human hemoglobin was obtained. In the case of the human homologue, no difference was seen between plasma samples from diabetic and non-diabetic patients, and we therefore present here combined data for all human samples. Tetramers' composition was determined using the experimental masses of the intact oligomers along with the observed monomers upon dissociation of the tetrameric hemoglobin populations (Fig. 4, Supplementary Figs. 5–6, and Supplementary Table 1). For zebra finches, the experimental masses of the subunits obtained during tetramer fragmentation were also compared to the relative circulating abundance of their six known different globin chains (Alev et al., 2009). Hence, a tetramer population, identified as composed of two α-chains and two β-chains, and a dimer population were found to represent 80.7 ± 0.8% and 19.3 ± 0.8% of total human hemoglobin, respectively (Fig. 4). Hemoglobin profiles in zebra finches were much more complex, as three different hemoglobin tetramer populations were detected, representing 43.6 ± 0.6% (two alpha3 and two beta3 chains, tetramer III), 34.2 ± 0.2% (one alpha2, one alpha3, and two beta3 chains, tetramer II), and 6.6 ± 0.2% (two alpha2 chains and two beta3 chains, tetramer I) of total hemoglobin, respectively (Fig. 4). Additional signals corresponding to populations of dimers (12.5 ± 0.5%) and monomers (3.1 ± 0.1%) were also observed in zebra finch samples (Fig. 4). Table 3 gives the results of the GLM testing for relationship between hemoglobin isoforms (PC1 and PC2 axes) and individual phenotypes. Model selection indicated that only PC1 was significantly related to individual body mass (Pearson's correlation, $r = -0.752$, $t = -2.795$, df = 7, $P = 0.031$): heavier individuals present lower PC1 values, *i.e.* higher hemoglobin tetramer and lower dimer proportions. There was no significant effect of age, plasma glucose levels or sex on PC1. PC2 (*i.e.* hemoglobin monomer proportion) was not significantly related to any of the fixed factors.

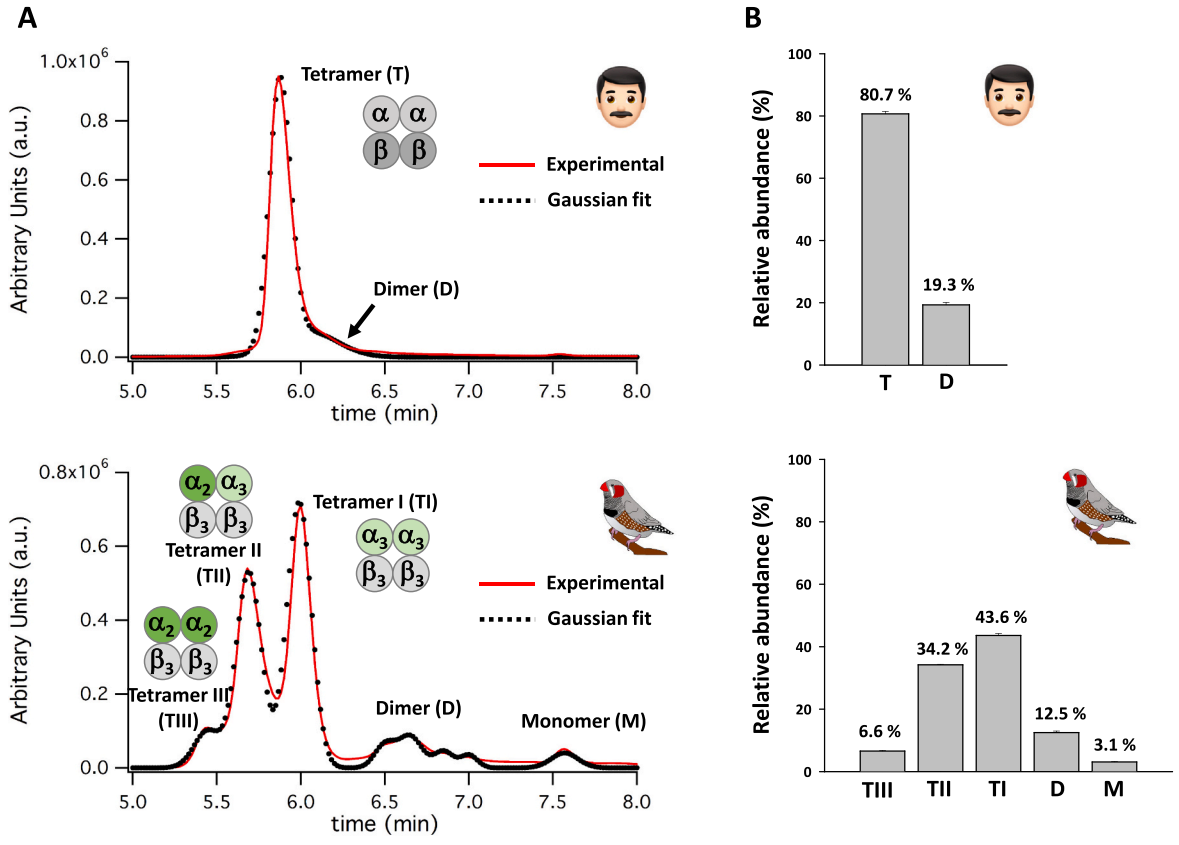

Fig. 4. Oligomeric states and proportions of human and zebra finch hemoglobins
LC-UV-MS analysis of whole hemoglobins allowed visualizing different oligomer states (see the representative chromatograms in panel A) and determining their relative proportions (mean ± sem of 8 independent determinations, panel B). Mass spectra obtained after dissociation of zebra finch tetramers inside the mass spectrometer are shown in Supplementary Figs. 5–6.

Table 3
General linear model output testing for effects of phenotypic (individual Body mass in g, Age in months, sex) and physiological (plasma glucose levels in mg/dL) variables on hemoglobin isoforms' proportion (tetramer I, II and III, dimer and monomer) measured from whole blood of 8 captive zebra finches. PC1 encompasses dimer and tetramer III proportions (loaded positively and negatively, respectively), while PC2 reflects only monomer proportion (positively loaded). Selection of models has been made using the *Mumin* R package. R^2 Nagelkerke indicates the percentage of total variance explained by the model.

A	PC1		
Predictors	Estimates	CI	P
Intercept	0.00	−0.49–0.49	1.000
Body mass (g)	−0.75	−1.28 to −0.22	0.031
Observations	8		
R^2 Nagelkerke	0.669		

B	PC2		
Predictors	Estimates	CI	P
Intercept	−0.06	−1.31–1.19	0.936
Plasma glucose (mg / dL)	−0.37	−1.25–0.51	0.471
Age	−0.22	−1.28–0.83	0.708
Sex [M]	0.11	−1.86–2.09	0.919
Body mass (g)	0.49	−0.38–1.35	0.351
Observations	8		
R^2 Nagelkerke	0.584		

4. Discussion

To gain insight into the extent protein glycation may affect birds, this study investigated structural features of several proteins in the zebra finch, and how they are related with circulating glucose levels or body mass. The first observation confirmed that, as in other birds (Polakof et al., 2011), glycaemia is high in zebra finches, reaching values that are usually recorded in human diabetic patients (Rhee et al., 2019). In humans, blood glucose and glycated hemoglobin (HbA1c) levels are clearly linked together and a positive relationship is observed (Beltran Del Rio et al., 2016; Lledo-Garcia et al., 2013; Lo et al., 2014; Rohlfing et al., 2002). Conversely in birds, blood glucose and blood hemoglobin concentrations have been shown to negatively correlate in adult great (*Parus major*) and blue tits (*Cyanistes caeruleus*) (Kalinski et al., 2016), and the use of chromatography- or affinity-based methods has revealed low values for glycated hemoglobin despite elevated circulating glucose concentrations in few bird species. For instance, the high glycaemia (17 mM) measured in hummingbirds was associated with levels of glycated hemoglobin (*Calypte spe.*, *Archilocus colubris*, 3.7–4.5%) close to those in non-diabetic humans (Beuchat and Chong, 1998). Low percentages of glycated hemoglobin have also been reported in mute swans and rooks (*Cygnus olor*, *Corvus frugilegus*, 1.4–1.8%) (Miksik and Hodny, 1992), as well as chickens, turkeys and ducks (0.5–0.9%) (Rendell et al., 1985), American kestrels (*Falco sparverius*, 0.6–2.7%) (Ardia, 2006), and collared flycatchers (*Ficedula albicollis*, 0.7–3.7%) (Récapet et al., 2016). Using the same type of method, we found low percentages (2.6–4.0%) of glycated hemoglobin in zebra finches as well. Glycated hemoglobin has been proposed as a measure of condition in flycatchers, its percentage (0.9–2.1%) being not significantly influenced by age, sex or body mass but positively correlated with reproductive performance proxies, suggesting that it reflects pre-breeding condition in migrating birds (Andersson and Gustafsson, 1995). Our data do not support a link between glycated hemoglobin levels in zebra finches and individual body mass, sex or glucose levels.

It is worth reminding at this stage that all of the above-mentioned assays in birds were performed using non-specific methods (see Introduction). Therefore, to get rid of those potentially erroneous measures, we set up a specific method based on the use of mass spectrometry, which did not detect any signal corresponding to any form of glycated hemoglobin in zebra finches. It would be surprising this is due to a lack of sensitivity and a matter of limit of detection. Indeed, the glycated hemoglobin form we targeted is the major one with the higher abundant glycation adduct, *i.e.*, the β-chain N-terminal fructosyl-valine residue (Wang et al., 2014; Zhang et al., 2001). Although of lower abundance, one cannot exclude that the β-chain of bird hemoglobin bears other modifications such as carboxymethyl-valine or carboxyethyl-valine (Jagadeeshaprasad et al., 2016; Lee et al., 2021). However, we could not detect any of these forms suggesting they may be absent or below our detection threshold.

Our method allowed detection of glycated hemoglobin in one human sample, with a proportion as low as 3.4% of total hemoglobin. Very low levels or the absence of glycated hemoglobin in zebra finches may reflect the existence of *in vivo* mechanisms of protection peculiar to bird species. First, the lower half-life of bird *vs.* human red blood cells (30–42 and 60–120 days, respectively; (Beltran Del Rio et al., 2016; Ottesen, 1948; Reddy et al., 1975; Rodnan et al., 1957) likely do not favor accumulation of glycated hemoglobin in bird's erythrocytes. Secondly, in sharp contrast with most mammal red blood cells including human ones, avian erythrocytes retain a nucleus (Jones, 2015). They might therefore have been hypothesized to express enzymes able to remove fructosamine glycation adducts, like fructosamine-3-kinase (Delpierre et al., 2002; Szwergold and Beisswenger, 2003), at higher levels than their mammalian counterparts do. Contradictory to this hypothesis, fructosamine-3-kinase appears active in human erythrocytes but only at very low levels in chicken erythrocytes (Delplanque et al., 2004). It is not known if this is a peculiarity of chickens or non-flying (captive laboratory strains) birds, or if this is a general feature in all birds. Also, it cannot be ruled out that other deglycation enzymes may exist and be active in bird red blood cells. This hypothesis may be supported by an earlier report that suggested the existence of a still unknown fructosamine deglycation mechanism independent of fructosamine-3-kinase (Szwergold and Beisswenger, 2003). Thirdly, reduced deglycation in bird erythrocytes may be linked to the near absent transport of glucose across the plasma membrane, as shown for chicken red blood cells (Johnstone et al., 1998). Less glucose in red cells would be in line with lower glycation levels. Furthermore, avian erythrocytes are equipped with functional mitochondria, including those of the zebra finch (Stier et al., 2013), and their survival can consequently depend on oxidizable substrates other than glucose. Nucleosides and glutamine have notably been reported as main substrates for chicken erythrocytes (Mathew et al., 1993).

The evolutionary origin of the differences between red blood cells of birds and mammals appears linked to the independent emergence of homeothermy in different geological ages in these two groups (Gavrilov, 2013). Among the differences, lower affinity for oxygen has been reported for hemoglobin in birds compared to mammals (Ajiloo et al., 2002), with the exception of diving (Meir and Ponganis, 2009) and high-altitude flying bird species (Natarajan et al., 2018). Therefore, the high-energy demand in birds, which turns into high-oxygen demand, is likely satisfied not only by the high efficiency of the avian respiratory system in uptaking oxygen, but also by facilitating the release of oxygen from hemoglobin into the tissues. Structural characteristics may be responsible for such peculiarities of bird hemoglobin, which have notably a more tense conformation than mammalian ones (Ajiloo et al., 2002; Natarajan et al., 2018). A different conformation may indicate that, although hemoglobin β-chain glycation sites appear conserved, *e.g.*, in the zebra finch compared to the human (Fig. 3A), they could remain less accessible in the avian protein. This hypothesis is reinforced by the results we obtained in our *in vitro* experiment. The gradual increase in the levels of human glycated hemoglobin (HbA1c) upon *in vitro* exposure to glucose was observed for both the samples from non-diabetic and diabetic individuals, however, the magnitude of variation was less pronounced for the latter and the levels did not exceed 13.5%. This may suggest a limitation, at least a slowdown, in the glycation rate of hemoglobin beyond about 10–12%. In zebra finch samples, we show here

that exposure to glucose allows mass spectrometry-based detection of glycated hemoglobin *in vitro* at a proportion lower than 4%, which confirms that there was no lack of sensitivity in our *in vivo* assays (see above). However, we did not detect glycated hemoglobin *in vitro* in bird control samples, and the increased levels in bird glycated hemoglobin upon glucose exposure was delayed and not very pronounced. Overall, these data, which support the absence of hemoglobin glycation *in vivo*, may be explained by structural peculiarities of bird hemoglobin that interfere with its glycation.

Structural differences between birds and humans regarding the quaternary structure of hemoglobin was determined in our study. In its mature form, hemoglobin is generally reported to be made up of four subunits (two α- and two β-chains in adults), including in the few birds that were studied (Abbasi and Lutfullah, 2002; Grispo et al., 2012; Manconi et al., 2007; Mohamed Abubakkar et al., 2014). From these previous works, it is also known that two tetramer isoforms differing in their alpha chains, namely HbA and HbD, may coexist in birds. Whereas we observed that the expected tetramer composed of two α- and two β-chains was dominating the human hemoglobin signal, we made an unprecedented observation of the presence of three different tetramers along with dimer and monomer populations from the analysis of zebra finch hemoglobin. The expected tetramer with two α_3- and two β_3-chains was identified as the most abundant (tetramer I, TI), the two α_3-chains being successively replaced by a α_2-chain in the two other tetramers (tetramers II and III, TII and TIII). It is possible that β-chain glycation sites in these bird tetramers are less accessible than in the human tetramer. The α-chain substitution we suspect from one to another tetramer (Fig. 4) seems to correspond to transitions from the most abundant α-globin in adults (alpha3) to a more primitive one (alpha2). Moreover, while using the masses of monomers and that of the heme molecule to reconstitute the composition of each tetramer, we still could not explain an additional mass of 580 Da per tetramer (see Supplementary Fig. 5). This could correspond to the presence of a cofactor, which association with globin chains could remain highly stable in zebra finches due to a putatively highly tense conformation of hemoglobin tetramers. It could be a phosphate compound like inositol pentaphosphate, already found in few bird species (Isaacks and Harkness, 1980; Riera et al., 1991). Inositol pentaphosphate modulates hemoglobin affinity for oxygen and possibly protects it from oxidative stress, as do 2,3-diphosphoglycerate in mammals (Tellone et al., 2019). The relative intensities of the monomeric populations obtained during nTD-MS analysis (Supplementary Fig. 5) suggest that TIII of bird hemoglobin is more prone to dissociation compared to TII and TI. This experimental evidence could be tentatively correlated with the hydrodynamic volume, and hence to the retention time (Fig. 4), of each individual tetrameric population. Thereby, TIII, with greater hydrodynamic volume (more extended structure and lower retention time) is more easily fragmented (higher monomer/tetramer ratio) in comparison with the more compact TII and TI tetramers.

The 'modern' hemoglobin tetramer has evolved from a noncooperative homodimer (itself evolved from an ancient monomer) with high oxygen affinity that existed before the gene duplication that generated distinct α- and β-subunits (Pillai et al., 2020). As suggested elsewhere for few birds species about the coexistence of hemoglobin A and D isoforms (Storz, 2016a, 2016b), and reinforced here by our detailed data about three hemoglobin isoforms in zebra finches, the multiplicity of these isoforms and adjustments in their relative abundance may allow to fine tune oxygen carrying capacity to match changing environmental and physiological conditions (as suggested by our significant link between body mass and hemoglobin tetramer and dimer proportions), and/or energy expenditure levels. Our results also raise the possibility for the hemoglobin cofactor inositol pentaphosphate to add another controlling mechanism. Overall, these mechanisms should be of key importance to ensure that bird physiology remains optimal when facing any challenge to energy expenditure levels imposed by, *e.g.*, changes in food availability or an immune challenge, or simply due to life stage transitions (from the inactive nestling to the active fledgling) (Cornell et al., 2017). In addition, inter-specific differences in hemoglobin isoforms may also be expected, as birds live in much contrasted environments in terms of, *e.g.*, the partial pressure of oxygen in the atmospheric air, and they do not have all the same expected lifespan or ability to fly or migrate (Minias et al., 2013). A phylogenetic and comparative analysis of hemoglobin isoform structures, relative abundances, and oxygen binding capacity in many bird species should help better understand how such major traits supporting aerobic metabolism have coevolved with flying performances or if they are related to particular environmental constraints only. Further, because glucose levels are negatively related to individual lifespan (Montoya et al., 2018) and glycation, involved in the loss of proteostasis (Lopez-Otin and Kroemer, 2021), is one of the hallmarks of aging (Chaudhuri et al., 2018), the evaluation of whether glycation eventually occurs in any bird species hence disturbing potentially the functionality of hemoglobin isoforms (or any other plasma protein) will add insight into how glycation impacts on bird's fitness and life expectancy. Despite our small sample size, we notably could see that PC1 was significantly related to individual body mass: heavier individuals seemed to present a larger proportion of hemoglobin tetramer III and a lower proportion of dimer, independently of sex, age or glycaemia. This may suggest that individuals in better body condition may also have a higher capacity of oxygen delivery to the organs. Exploring further how individual differences in hemoglobin isoform proportions is related to the organism's oxygen carrying and delivering capacities may be of key importance in the evolutionary context of individual quality (Wilson and Nussey, 2010) (*i.e.* why some individual perform better than others). This may reveal itself particularly important in determining how the transition from nestling to fledging is done successfully because: (i) hemoglobin concentration in whole blood was highlighted as a proxy of fledging physiological maturity defining survival chances (Cornell et al., 2017), (ii) glycated hemoglobin levels was previously found to be higher in nestlings that grew faster (Ardia, 2006), and (iii) we hypothesized that glycation levels were low in zebra finches, maybe due to specific structural particularities (see above). Whether surviving fledging present a higher proportion of tetramer III hemoglobin, which may be more resistant to glycation, is a tempting hypothesis.

Unlike the analysis of zebra finch hemoglobin, the application of our mass spectrometry-based method to plasma samples has shown that these birds are not free of glycation. Indeed, glycation events were detected, at single or multiple sites, for select highly abundant plasma proteins. Of these, high levels were measured for mono- (32–46%) and doubly-glycated (11–30%) forms of albumin and serotransferrin, and only a mono-glycated form (14%) was detected for carbonic anhydrase 2. In humans, non-enzymatic glycation has been reported as one of the most frequent post-translational modification occurring in the carbonic anhydrase family, with adverse consequences on enzymatic activity that can be considered among the initial steps of altered metabolism in diabetes (Di Fiore et al., 2020). Of the six glycation sites that were described for human carbonic anhydrase 2, only two appear to be conserved in the zebra finch counterpart (Supplementary Fig. 7). This is consistent with the fact that, in contrast to mammals, diabetes is rarely diagnosed in birds. However, we did not find any value in the literature for the proportion of plasma glycated carbonic anhydrase 2 to be compared to the 14% we measured in zebra finches. For the two other plasma proteins we examined, the proportion of glycated forms in zebra finches was higher than that in healthy persons. The highest extent of glycation (all forms together) we observed was 77% for zebra finch glycated serotransferrin, whereas the percentage of glycated serotransferrin was of 5% in diabetic patients and only 2% in healthy people if one takes Lys-534, the major site of glycation, as a surrogate estimate of the glycated protein form (Golizeh et al., 2017). Five of the six known glycation sites in human serotransferrin are conserved in the zebra finch (Supplementary Fig. 8). It could be that more lysine residues are easily accessible to glycation in the avian counterpart. Not only the zebra finch

serotransferrin contains a high number of lysines (64) but its structure may be significantly different from that of the human protein, therefore favouring glycation. This is supported by the low percentage (only 66%) of amino acid identities between the human and avian sequences.

The value of 43% found here for glycated albumin in zebra finches is higher than in humans, except from severe diabetic patients. Indeed, albumin is usually reported as being glycated at about 1–10% and up to 15% in healthy people, 12–25% in diabetic patients, and up to 94% in severe diabetic patients with poor glycemic control (Anguizola et al., 2013; Kisugi et al., 2007; Rondeau and Bourdon, 2011), thus being an alternative biomarker to glycated hemoglobin for monitoring glycaemia (Soboleva et al., 2019). The eleven sites known to be possibly glycated in the human albumin sequence are also present in the avian sequence (Supplementary Fig. 9). As argued above for serotransferrin, a higher number of glycation sites may exist in the zebra finch albumin, which contains numerous lysines (66) and shares only 64% of amino acid identities with the human protein. However, a recent study reported contrasting results where resistance of recombinant chicken albumin compared to recombinant bovine albumin was imputed to fewer exposed lysine residues (Anthony-Regnitz et al., 2020). Glycated albumin values were even higher abundant in zebra finches than in other bird species, like the bald eagle (24%), barred owl (24%), great horned owl (19%), and red-tailed hawk (14%), their glycaemia being also high (20–22 mM) putatively due to a diet promoting gluconeogenesis (Ingram et al., 2017). Despite less sophisticated methods were used, an earlier study also found glycated albumin at higher levels in the plasma from chickens, turkeys and ducks compared to humans (Rendell et al., 1985). In addition to possible structural features peculiar to zebra finches, differences between species may be imputed to a different half-life of this protein or more global differences in protein turnover rates.

In humans, albumin levels are positively correlated with glycated albumin levels (Mo et al., 2016) but negatively with glycated hemoglobin (Rodriguez-Segade et al., 2005; Tiwari et al., 2015). This link between hemoglobin glycation and the abundance and glycation status of albumin has further been demonstrated *in vitro* (Jagadeeshaprasad et al., 2018). Therefore, through mechanisms that remain to be determine, high levels of glycated albumin in zebra finches might contribute to the absence of hemoglobin glycation we depicted (see above).

A relationship between glycated albumin levels and circulating glucose has been reported in humans (Inaba et al., 2007). Conversely in our study, neither glycated albumin nor TGL were related to plasma glucose levels, in contrast to glycated serotransferrin levels. Therefore, serotransferrin appears as a possible better target to monitor glycation events and the response to elevated plasma glucose concentrations in birds. However, only glycated carbonic anhydrase 2 seemed to be higher in older individuals. Again, a phylogenetic and comparative future work should help better conclude on how glycaemia and glycation interact and relate to aging in more bird species.

5. Conclusion

The formation of a glycated protein is usually considered to depend on the amount of glucose it has been in contact with over the last months in humans. Using a specific mass spectrometry-based method, our study clearly shows that hemoglobin in the zebra finch is not glycated *in vivo*, and only at very low levels after an *in vitro* treatment with glucose. Our results then support the hypothesis that structural avian hemoglobin characteristics may limit or prevent glycation, a possible prerequisite for the ability to fly, and they may help adjust the aerobic capacity of birds. Thus, they may have played a key role in the evolution of slow pace of life in flying birds. Concerning glycation of bird plasma proteins, differences appear to exist between species, which calls for a deeper analysis of their structure in the future. The identification of all glycation sites as well as the determination of the extent to which they are glycated and the evaluation of the links between inter-specific structural differences and contrasting life history traits in many bird species should help better understand the determinants of avian protein glycation.

Data availability statement

Electronic supplementary materials are provided for the principal component analysis (PCA) of hemoglobin isoforms (Supplementary Fig. 1), Chromatogram and representative raw mass spectra for zebra finch (Supplementary Fig. 2) and human (Supplementary Fig. 3) hemoglobin chains, and for zebra finch plasma proteins (Supplementary Fig. 4), representative nTD-MS mass spectra of zebra finch and human hemoglobin (Supplementary Figs. 5–6), human and zebra finch carbonic anhydrase 2 (Supplementary Fig. 7), serotransferrin (Supplementary Fig. 8) and albumin (Supplementary Fig. 9) sequences, and for the calculation of the mass of hemoglobin tetrameric forms according to experimental subunit masses (Supplementary Table 1), and bird individual data (Supplementary Table 2).

Ethics statement

Our experimental study was conducted according to EU regulation (Directive 2010/63/EU). The protocol was approved by the ethical committee CREMEAS Strasbourg (#AL/02/02/01/13).

CRediT authorship contribution statement

FB and FC designed the study. MS-based glycation analysis was performed by CB and AH under the supervision of CS-R, and structural analyses were performed by CB and OH-A. Statistical analyses were generated by FC. FB and FC wrote the manuscript with the input of all authors.

Funding

This work was supported by the Mission pour l'Interdisciplinarité of the CNRS (PEPS, Exomod 2014), the French Proteomic Infrastructure (ProFI; ANR-10-INSB-08-03), and the the Agence Nationale de la Recherche (Programme AGEs; ANR-21-CE02-0009-02). CB was supported by a grant from the French Ministry of Higher Education, Research and Innovation.

Declaration of competing interest

None.

Data availability

We have shared all data in supplementary material. Any raw data can be made available upon request.

Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.exger.2022.111811.

References

Abbasi, A., Lutfullah, G., 2002. Molecular basis of bird respiration: primary hemoglobin structure component from tufted duck (Aythya fuligula, Anseriformes)–role of alpha99Arg in formation of a complex salt bridge network. Biochem. Biophys. Res. Commun. 291, 176–184.

Ajiloo, D., Moosavi-Movahedi, A.A., Sadeghi, M., Gharibi, H., 2002. Comparative structural and functional studies of avian and mammalian hemoglobins. Acta Biochim. Pol. 49, 459470.
Alev, C., Shinmyozu, K., McIntyre, B.A.S., Sheng, G.J., 2009. Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. Dev. Genes Evol. 219, 353–360.
Andersson, M.S., Gustafsson, L., 1995. Glycosylated hemoglobin - a new measure of condition in birds. Proc. R. Soc. B Biol. Sci. 260, 299–303.
Anguizola, J., Matsuda, R., Barnaby, O.S., Hoy, K.S., Wa, C., DeBolt, E., Koke, M., Hage, D.S., 2013. Review: glycation of human serum albumin. Clin. Chim. Acta 425, 64–76.
Anthony-Regnitz, C.M., Wilson, A.E., Sweazea, K.L., Braun, E.J., 2020. Fewer exposed lysine residues may explain relative resistance of chicken serum albumin to in vitro protein glycation in comparison to bovine serum albumin. J. Mol. Evol. 88, 653–661.
Ardia, D.R., 2006. Glycated hemoglobin and albumin reflect nestling growth and condition in American kestrels. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 143, 62–66.
Barton, K., 2020. MuMIn: multi-model inference [R package version 1.43.17]. Retrieved from. https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf.
Beltran Del Rio, M., Tiwari, M., Amodu, L.I., Cagliani, J., Rodriguez Rilo, H.L., 2016. Glycated hemoglobin, plasma glucose, and erythrocyte aging. J. Diabetes Sci. Technol. 10, 1303–1307.
Bergouignan, A., Rudwill, F., Simon, C., Blanc, S., 2011. Physical inactivity as the culprit of metabolic inflexibility: evidence from bed-rest studies. J. Appl. Physiol. 1985 (111), 1201–1210.
Bernard, C., 1879. Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux, Vol. 2. J.-B. Baillère et Fils, Paris.
Beuchat, C.A., Chong, C.R., 1998. Hyperglycemia in hummingbirds and its consequences for hemoglobin glycation. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 120, 409–416.
Butterfield, D.A., Halliwell, B., 2019. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. Nat. Rev. Neurosci. 20, 148–160.
Calder, W.A., 1990. Avian longevity and aging. In: Harrison, D. (Ed.), Genetic Effects on Aging. The Telford Press, Inc, Caldwell, New Jersey, pp. 185–204.
Calow, P., 1982. Homeostasis and fitness. Am. Nat. 120, 416–419.
Cannon, W.B., 1929. Organization for physiological homeostasis. Physiol. Rev. 9, 399–431.
Chaudhuri, J., Bains, Y., Guha, S., Kahn, A., Hall, D., Bose, N., Gugliucci, A., Kapahi, P., 2018. The role of advanced glycation end products in aging and metabolic diseases: bridging association and causality. Cell Metab. 28, 337–352.
Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290.
Cornell, A., Gibson, K.F., Williams, T.D., 2017. Physiological maturity at a critical life-history transition and flight ability at fledging. Funct. Ecol. 31, 662–670.
Delignette-Muller, M.L., Dutang, C., 2015. Fitdistrplus: an R package for fitting distributions. J. Stat. Softw. 64, 1–34.
Delpierre, G., Collard, F., Fortpied, J., Van Schaftingen, E., 2002. Fructosamine 3-kinase is involved in an intracellular deglycation pathway in human erythrocytes. Biochem. J. 365, 801–808.
Delplanque, J., Delpierre, G., Opperdoes, F.R., Van Schaftingen, E., 2004. Tissue distribution and evolution of fructosamine 3-kinase and fructosamine 3-kinase-related protein. J. Biol. Chem. 279, 46606–46613.
Di Fiore, A., Supuran, C.T., Scaloni, A., De Simone, G., 2020. Human carbonic anhydrases and post-translational modifications: a hidden world possibly affecting protein properties and functions. J. Enzyme Inhib. Med. Chem. 35, 1450–1461.
Fox, J., Weiberg, S., 2019. An R Companion to Applied Regression, 3rd ed. Sage, Thousands Oaks, CA.
Frayn, N.F., 2010. Metabolic Regulation: A Human Perspective, 3rd ed. Wiley–Blackwell, Oxford.
Gavrilov, V.M., 2013. Origin and development of homoiothermy: a case study of avian energetics. Adv. Biosci. Biotechnol. 4, 1–17.
Gluckman, P.D., Hanson, M.A., 2004. The developmental origins of the metabolic syndrome. Trends Endocrinol. Metab. 15, 183–187.
Golizeh, M., Lee, K., Ilchenko, S., Osme, A., Bena, J., Sadygov, R.G., Kashyap, S.R., Kasumov, T., 2017. Increased serotransferrin and ceruloplasmin turnover in diet-controlled patients with type 2 diabetes. Free Radic. Biol. Med. 113, 461–469.
Goodpaster, B.H., Sparks, L.M., 2017. Metabolic flexibility in health and disease. Cell Metab. 25, 1027–1036.
Grispo, M.T., Natarajan, C., Projecto-Garcia, J., Moriyama, H., Weber, R.E., Storz, J.F., 2012. Gene duplication and the evolution of hemoglobin isoform differentiation in birds. J. Biol. Chem. 287, 37647–37658.
Herman, M.A., Kahn, B.B., 2006. Glucose transport and sensing in the maintenance of glucose homeostasis and metabolic harmony. J. Clin. Invest. 116, 1767–1775.
Holmes, J., Harper, J.M., 2018. Birds as models for the biology of aging and aging-related disease: an update. In: Conn's Handbook of Models for Human Aging. Academic Press, pp. 301–312.
Holmes, D.J., Fluckiger, R., Austad, S.N., 2001. Comparative biology of aging in birds: an update. Exp. Gerontol. 36, 869–883.
Inaba, M., Okuno, S., Kumeda, Y., Yamada, S., Imanishi, Y., Tabata, T., Okamura, M., Okada, S., Yamakawa, T., Ishimura, E., Nishizawa, Y., Grp, O.C.E.R., 2007. Glycated albumin is a better glycemic indicator than glycated hemoglobin values in hemodialysis patients with diabetes: effect of anemia and erythropoietin injection. J. Am. Soc. Nephrol. 18, 896–903.
Ingram, T., Zuck, J., Borges, C.R., Redig, P., Sweazea, K.L., 2017. Variations in native protein glycation and plasma antioxidants in several birds of prey. Comp. Biochem. Physiol. B 210, 18–28.
Isaacks, R.E., Harkness, D.R., 1980. Erythrocyte organic phosphates and hemoglobin function in birds, reptiles, and fishes. Am. Zool. 20, 115–129.
Jagadeeshaprasad, M.G., Batkulwar, K.B., Meshram, N.N., Tiwari, S., Korwar, A.M., Unnikrishnan, A.G., Kulkarni, M.J., 2016. Targeted quantification of N-1-(carboxymethyl) valine and N-1-(carboxyethyl) valine peptides of beta-hemoglobin for better diagnostics in diabetes. Clin. Proteomics 13, 7.
Jagadeeshaprasad, M.G., Venkatasubramani, V., Unnikrishnan, A.G., Kulkarni, M.J., 2018. Albumin abundance and its glycation status determine hemoglobin glycation. ACS Omega 3, 12999–13008.
Jeffs, J.W., Ferdosi, S., Yassine, H.N., Borges, C.R., 2017. Ex vivo instability of glycated albumin: a role for autoxidative glycation. Arch. Biochem. Biophys. 629, 36–42.
Jeppsson, J.O., Kobold, U., Barr, J., Finke, A., Hoelzel, W., Hoshino, T., Miedema, K., Mosca, A., Mauri, P., Paroni, R., Thienpont, L., Umemoto, M., Weykamp, C., 2002. Approved IFCC reference method for the measurement of HbA1c in human blood. Clin. Chem. Lab. Med. 40, 78–89.
Johnstone, R.M., Mathew, A., Setchenska, M.S., Grdisa, M., White, M.K., 1998. Loss of glucose transport in developing avian red cells. Eur. J. Cell Biol. 75, 66–77.
Jones, M.P., 2015. Avian hematology. Clin. Lab. Med. 35, 649–659.
Kalinski, A., Banbura, M., Gladalski, M., Markowski, M., Skwarska, J., Wawrzyniak, J., Zielnski, P., Cyzewska, I., Banbura, J., 2016. Relationship between blood haemoglobin and glucose concentrations in adult great tits (Parus major) and blue tits (Cyanistes caeruleus). Avian Biol. Res. 9, 152–158.
Kisugi, R., Kouzuma, T., Yamamoto, T., Akizuki, S., Miyamoto, H., Someya, Y., Yokoyama, J., Abe, I., Hirai, N., Ohnishi, A., 2007. Structural and glycation site changes of albumin in diabetic patient with very high glycated albumin. Clin. Chim. Acta 382, 59–64.
Klandorf, H., Probert, I.L., Iqbal, M., 1999. In the defense against hyperglycemiea: an avain strategy. Worlds Poult. Sci. J. 55, 251–268.
Le, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. J. Stat. Softw. 25, 1–18.
Lee, J.H., Samsuzzaman, M., Park, M.G., Park, S.J., Kim, S.Y., 2021. Methylglyoxal-derived hemoglobin advanced glycation end products induce apoptosis and oxidative stress in human umbilical vein endothelial cells. Int. J. Biol. Macromol. 187, 409–421.
Lenth, R., Singmann, H., Love, J., Buerkner, P., Herve, M., 2018. Emmeans: estimated marginal means, aka least-squares means [R Package version 1.2. 4]. Retrieved from. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf.
Lledo-Garcia, R., Mazer, N.A., Karlsson, M.O., 2013. A semi-mechanistic model of the relationship between average glucose and HbA1c in healthy and diabetic subjects. J. Pharmacokinet. Pharmacodyn. 40, 129–142.
Lo, C., Lui, M., Ranasinha, S., Teede, H.J., Kerr, P.G., Polkinghorne, K.R., Nathan, D.M., Zheng, H., Zoungas, S., 2014. Defining the relationship between average glucose and HbA1c in patients with type 2 diabetes and chronic kidney disease. Diabetes Res. Clin. Pract. 104, 84–91.
Lopez-Otin, C., Kroemer, G., 2021. Hallmarks of health. Cell 184, 33–63.
Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G., 2013. The hallmarks of aging. Cell 153, 1194–1217.
Manconi, B., Olianas, A., Sanna, M.T., Messana, I., Demurtas, L., Castagnola, M., Giardina, B., Pellegrini, M., 2007. Functional characterization of the single hemoglobin of the migratory bird Ciconia ciconia. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 147, 242–249.
Masoro, E.J., Katz, M.S., Mcmahan, C.A., 1989. Evidence for the glycation hypothesis of aging from the food-restricted rodent model. J. Gerontol. 44. B20-B22.
Mathew, A., Grdisa, M., Johnstone, R.M., 1993. Nucleosides and glutamine are primary energy substrates for embryonic and adult chicken red cells. Biochem. Cell Biol. 71, 288–295.
Meir, J.U., Ponganis, P.J., 2009. High-affinity hemoglobin and blood oxygen saturation in diving emperor penguins. J. Exp. Biol. 212, 3330–3338.
Miksik, I., Hodny, Z., 1992. Glycated hemoglobin in mute swan (Cygnus-Olor) and rook (Corvus-Frugilegus). Comp. Biochem. Physiol. B 103, 553–555.
Minias, P., Kaczmarek, K., Wlodarczyk, R., Janiszewski, T., 2013. Hemoglobin concentrations in waders vary with their strategies of migration: a comparative analysis. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 165, 7–12.
Mo, Y.F., Ma, X.J., Li, H., Ran, X.W., Yang, W.Y., Li, Q., Peng, Y.D., Li, Y.B., Gao, X., Luan, X.J., Wang, W.Q., Xie, Y., Zhou, J., Jia, W.P., 2016. Relationship between glycated albumin and glycated hemoglobin according to glucose tolerance status: a multicenter study. Diabetes Res. Clin. Pract. 115, 17–23.
Mohamed Abubakkar, M.H., Saraboji, K., Ponnuswamy, M.N., 2014. The crystal structure of oxy hemoglobin from high oxygen affinity bird emu (Dromaius novaehollandiae). Gen. Physiol. Biophys. 33, 373–382.
Montoya, B., Briga, M., Jimeno, B., Moonen, S., Verhulst, S., 2018. Baseline glucose level is an individual trait that is negatively associated with lifespan and increases due to adverse environmental conditions during development and adulthood. J. Comp. Physiol. B. 188, 517–526.
Natarajan, C., Jendroszek, A., Kumar, A., Weber, R.E., Tame, J.R.H., Fago, A., Storz, J.F., 2018. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose. PLoS Genet. 14, e1007331.
O'Brien, R., Timmins, K., 1994. The role of oxidation and glycation in the pathogenesis of diabetic atherosclerosis. Trends Endocrinol. Metab. 5, 329–334.
Ottesen, J., 1948. Life-span of red and white blood corpuscles of the hen. Nature 162, 730.

Pillai, A.S., Chandler, S.A., Liu, Y., Signore, A.V., Cortez-Romero, C.R., Benesch, J.L.P., Laganowsky, A., Storz, J.F., Hochberg, G.K.A., Thornton, J.W., 2020. Origin of complexity in haemoglobin evolution. Nature 581, 480–485.
Polakof, S., Mommsen, T.P., Soengas, J.L., 2011. Glucosensing and glucose homeostasis: from fish to mammals. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 160, 123–149.
Poulsen, M.W., Hedegaard, R.V., Andersen, J.M., de Courten, B., Bugel, S., Nielsen, J., Skibsted, L.H., Dragsted, L.O., 2013. Advanced glycation endproducts in food and their effects on health. Food Chem. Toxicol. 60, 10–37.
Priego-Capote, F., Ramirez-Boo, M., Finamore, F., Gluck, F., Sanchez, J.C., 2014. Quantitative analysis of glycated proteins. J. Proteome Res. 13, 336–347.
Prinzinger, R., 1993. Life-span in birds and the aging theory of absolute metabolic scope. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 105, 609–615.
Ramasamy, R., Bucciarelli, L.G., Yan, S.F., Schmidt, A.M., 2010. Advanced glycation end products, RAGE, and aging. Oxid Stress Appl. Bas. 79–90.
Rattiste, K., Klandorf, H., Urvik, J., Sepp, T., Asghar, M., Hasselquist, D., Cooey, C., Horak, P., 2015. Skin pentosidine and telomere length do not covary with age in a long-lived seabird. Biogerontology 16, 435–441.
Récapet, C., Sibeaux, A., Cauchard, L., Doligez, B., Bize, P., 2016. Selective disappearance of individuals with high levels of glycated haemoglobin in a free-living bird. Biol. Lett. 12.
Reddy, P.R., Van Krey, H.P., Gross, W.B., Siegel, P.B., 1975. Erythrocyte lifespan in dwarf and normal pullets from growth selected lines of chickens. Poult. Sci. 54, 1301–1303.
Rendell, M., Stephen, P.M., Paulsen, R., Valentine, J.L., Rasbold, K., Hestorff, T., Eastberg, S., Shint, D.C., 1985. An interspecies comparison of Normal levels of glycosylated hemoglobin and glycosylated albumin. Comp. Biochem. Physiol. B 81, 819–822.
Rhee, M.K., Ho, Y.L., Raghavan, S., Vassy, J.L., Cho, K., Gagnon, D., Staimez, L.R., Ford, C.N., Wilson, P.W.F., Phillips, L.S., 2019. Random plasma glucose predicts the diagnosis of diabetes. PLoS One 14, e0219964.
Riera, M., Fuster, J.F., Palacios, L., 1991. Role of erythrocyte organic phosphates in blood oxygen transport in anemic quail. Am. J. Phys. 260. R798-803.
Rodnan, G.P., Ebaugh Jr., F.G., Fox, M.R., 1957. The life span of the red blood cell and the red blood cell volume in the chicken, pigeon and duck as estimated by the use of Na2Cr51O4, with observations on red cell turnover rate in the mammal, bird and reptile. Blood 12, 355–366.
Rodriguez-Segade, S., Rodriguez, J., Mayan, D., Camina, F., 2005. Plasma albumin concentration is a predictor of HbA(1c) among type 2 diabetic patients, independently of fasting plasma glucose and fructosamine. Diabetes Care 28, 437–439.
Rohlfing, C.L., England, J.D., Wiedmeyer, H.M., Tennill, A., Little, R.R., Goldstein, D.E., 2002. Defining the relationship between plasma glucose and HbA(1c) - analysis of glucose profiles and HbA(1c) in the diabetes control and complications trial. Diabetes Care 25, 275–278.
Rondeau, P., Bourdon, E., 2011. The glycation of albumin: structural and functional impacts. Biochimie 93, 645–658.
Rynders, C.A., Blanc, S., DeJong, N., Bessesen, D.H., Bergouignan, A., 2018. Sedentary behaviour is a key determinant of metabolic inflexibility. J. Physiol. Lond. 596, 1319–1330.
Sell, D.R., Lane, M.A., Johnson, W.A., Masoro, E.J., Mock, O.B., Reiser, K.M., Fogarty, J. F., Cutler, R.G., Ingram, D.K., Roth, G.S., Monnier, V.M., 1996. Longevity and the genetic determination of collagen glycoxidation kinetics in mammalian senescence. Proc. Natl. Acad. Sci. U. S. A. 93, 485–490.
Semba, R.D., Nicklett, E.J., Ferrucci, L., 2010. Does accumulation of advanced glycation end products contribute to the aging phenotype? J. Gerontol. A Biol. Sci. Med. Sci. 65, 963–975.
Soboleva, A., Vikhnina, M., Grishina, T., Frolov, A., 2017. Probing protein glycation by chromatography and mass spectrometry: analysis of glycation adducts. Int. J. Mol. Sci. 18.
Soboleva, A., Mavropulo-Stolyarenko, G., Karonova, T., Thieme, D., Hoehenwarter, W., Ihling, C., Stefanov, V., Grishina, T., Frolov, A., 2019. Multiple glycation sites in blood plasma proteins as an integrated biomarker of type 2 diabetes mellitus. Int. J. Mol. Sci. 20.
Stier, A., Bize, P., Schull, Q., Zoll, J., Singh, F., Geny, B., Gros, F., Royer, C., Massemin, S., Criscuolo, F., 2013. Avian erythrocytes have functional mitochondria, opening novel perspectives for birds as animal models in the study of ageing. Front. Zool. 10.
Storz, J.F., 2016a. Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. Physiology (Bethesda) 31, 223–232.
Storz, J.F., 2016b. Hemoglobin-oxygen affinity in high-altitude vertebrates: is there evidence for an adaptive trend? J. Exp. Biol. 219, 3190–3203.
Suo, M.H., Wen, D.M., Wang, W.J., Zhang, D.C., Xu, S.N., Wang, X., Hu, T., 2019. False measurement of glycated hemoglobin in patients without hemoglobin a. Biosci. Rep. 39.
Szwergold, B.S., Beisswenger, P.J., 2003. Enzymatic deglycation–a new paradigm or an epiphenomenon? Biochem. Soc. Trans. 31, 1428–1432.
Tellone, E., Barreca, D., Russo, A., Galtieri, A., Ficarra, S., 2019. New role for an old molecule: the 2,3-diphosphoglycerate case. BBA-Gen. Subjects 1863, 1602–1607.
Tessier, F.J., 2010. The maillard reaction in the human body. The main discoveries and factors that affect glycation. Pathol. Biol. 58, 214–219.
Tiwari, S., Bothale, M., Hasan, I., Kulkarni, M.J., Sayyad, M.G., Basu, R., Basu, A., Unnikrishnan, A.G., 2015. Association between serum albumin and glycated hemoglobin in Asian Indian subjects. Indian J. Endocrinol. Metab. 19, 52–55.
Ulrich, P., Cerami, A., 2001. Protein glycation, diabetes, and aging. Recent Prog. Horm. Res. 56, 1–21.
Wang, S.H., Wang, T.F., Wu, C.H., Chen, S.H., 2014. In-depth comparative characterization of hemoglobin glycation in normal and diabetic bloods by LC-MSMS. J. Am. Soc. Mass Spectrom. 25, 758–766.
Wilson, A.J., Nussey, D.H., 2010. What is individual quality? An evolutionary perspective. Trends Ecol. Evol. 25, 207–214.
Zhang, X., Medzihradszky, K.F., Cunningham, J., Lee, P.D., Rognerud, C.L., Ou, C.N., Harmatz, P., Witkowska, H.E., 2001. Characterization of glycated hemoglobin in diabetic patients: usefulness of electrospray mass spectrometry in monitoring the extent and distribution of glycation. J. Chromatogr. B Biomed. Sci. Appl. 759, 1–15.
Zuck, J., Borges, C.R., Braun, E.J., Sweazea, K.L., 2017. Chicken albumin exhibits natural resistance to glycation. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 203, 108–114.

Chapitre 2

Développement d'une méthode analytique permettant une étude approfondie des glycations du plasma par une approche « *bottom-up* »

I. Le plasma : un fluide particulièrement complexe avec une large gamme dynamique en concentration

Le plasma correspond à la partie liquide du sang obtenue par centrifugation en présence d'anticoagulants tels que l'acide éthylène diamine tétra-acétique [EDTA], l'héparine ou le citrate de sodium. Il ne contient pas de cellules mais de nombreuses protéines (≈ 50 µg/µL), des métabolites (sucres, lipides...) et des ions (Na^+, Ca^{2+}…), qui sont transportés en suspension dans l'eau[386]. Il circule entre les différents organes, leur apportant oxygène et nutriments ; en retour, certains organes secrètent des molécules comme des hormones ou des cytokines, ou encore certaines protéines libérées en cas de maladie.

Le plasma est l'un des échantillons les plus utilisés pour le diagnostic, le pronostic et les suivis thérapeutiques de nombreuses maladies (cancers[387], maladies cardiovasculaires[388], diabètes[389]…). D'une part, l'échantillonnage est non invasif, contrairement à des tissus ou à d'autres types de fluides biologiques. D'autre part, les molécules plasmatiques se trouvent en solution et leur extraction/séparation s'en trouve grandement facilitée. C'est le cas d'un grand nombre de biomarqueurs connus tels que la myoglobine La concentration de cette protéine chez un patient sain est de 1-85 ng/mL mais peut atteindre 200-1100 ng/mL en cas d'infarctus du myocarde[390]. Cependant, le plasma est aussi l'un des échantillons les plus difficiles à analyser en protéomique. L'une des difficultés majeures repose sur sa très large gamme dynamique en termes d'abondances protéiques comme présenté en figure 92. Les concentrations plasmatiques de l'albumine, la protéine majoritaire du plasma, et des protéines faiblement abondantes (interleukine 6 par exemple) sont séparées par 11 ordres de grandeur[95]. À titre de comparaison, la gamme dynamique d'un spectromètre de masse ne présente que 4 à 5 ordres de grandeurs[391-392].

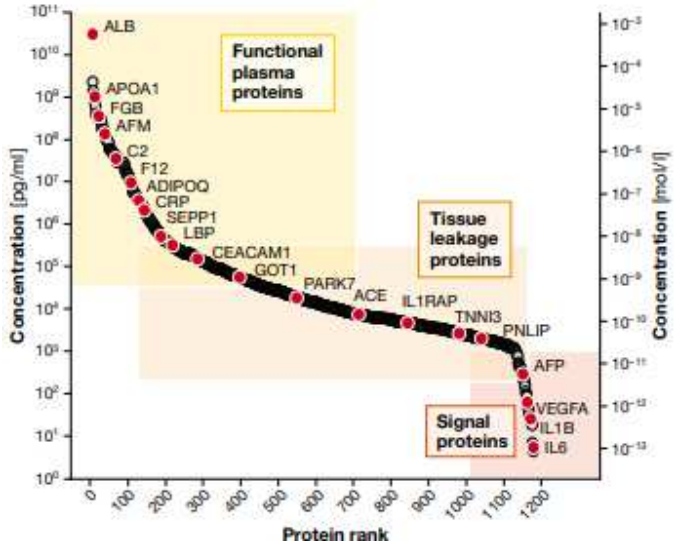

Figure 92 : Plage de concentration des protéines plasmatiques avec les noms de gènes de plusieurs protéines sanguines (points rouges)[95].

De plus, une dizaine de protéines : (albumine, IgG, IgA, haptoglobine, transferrine et antitrypsine alpha 1) représentent à elles seules plus de 85% du protéome total du plasma[98]. Cette répartition complique grandement l'analyse protéomique, le signal des peptides des protéines les plus abondantes masquant celui des peptides des protéines minoritaires.

Pour surmonter ces défis, différentes stratégies de préparation d'échantillons plasmatiques sont utilisées pour l'analyse protéomique. Les protéines majoritaires peuvent être déplétées, généralement par immunoaffinité avec des anticorps dirigés contre les 1 à 20 protéines les plus abondantes. C'est une stratégie utilisée dans la majorité des études protéomiques sur le plasma (Figure 93). Une autre approche, moins répandue (Figure 93), consiste à fractionner le plasma au niveau protéique ou peptidique afin de le décomplexifier. Cette méthode augmente le nombre d'analyses par échantillon. Par ailleurs, l'ajout d'une étape supplémentaire dans un protocole d'analyse protéomique induit toujours le risque de rencontrer des problèmes de reproductibilité.

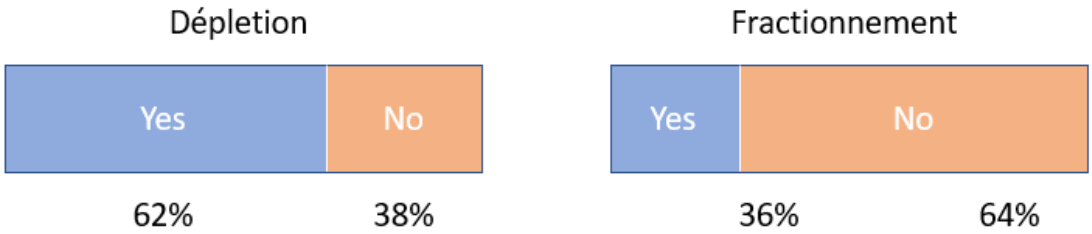

Figure 93 : Proportion des études protéomiques du plasma utilisant la déplétion et le fractionnement (d'après Geyer *et al*)[95].

Ces deux approches peuvent bien sûr être combinées afin d'augmenter la profondeur d'analyse du protéome plasmatique. Ainsi, en 2016, Geyer *et al.* sont parvenus à identifier environ un millier de protéines par échantillon de plasma humain en combinant l'immunodéplétion des 20 protéines majoritaires à un fractionnement hpH RP générant 46 fractions, qui ont ensuite été regroupées en 8 fractions avant analyse LC-MS/MS[20]. En 2002, de nombreuses équipes de

recherche se sont associées au sein du *Human Plasma Proteome Project* [HPPP] afin de cartographier le protéome du plasma humain. En 2017, 3500 protéines ont pu être identifiées dans le plasma au travers de 180 études différentes, illustrant encore une fois la complexité du protéome plasmatique[393].

Dans le cadre de l'étude des glycations sur les protéines plasmatiques aviaires, dans la continuité du projet rapporté dans le chapitre précédant, notre objectif est d'analyser le plasma suivant une stratégie « *bottom-up* » *label-free*. Cependant, il n'est pas possible de dépléter nos échantillons des protéines majoritaires du plasma car la déplétion est majoritairement réalisée par immuno-affinité et aucun anticorps spécifique n'est commercialisé pour ces espèces.

II. Les glycations dans le plasma

La réaction de glycation est une réaction non enzymatique entre un sucre réducteur et un acide aminé (le plus souvent une lysine ou une arginine en raison de leur groupement amine). Dans un premier temps, un produit d'Amadori ou fructosamine [FL] est formé *via* une base de Schiff. Ce composé va ensuite se dégrader en une multitude de produits aux structures très diverses (carboxyethyllysine (CEL), carboxymethyllysine (CML), 5-methyl-4-oxo-5-hydroimifazo-linone-2-ylornithine (MG-H1), glyoxal-derivated lysine dimer (GOLD)…). Ces produits sont appelés AGEs pour *advanced glycation endproducts*[238]. Une description détaillée de cette réaction est faite dans le chapitre 3 de la partie I.

Chez l'homme, parmi les protéines sanguines retrouvées glyquées, on peut citer l'albumine, l'hémoglobine, l'insuline, les immunoglobulines, les lipoprotéines ou encore le fibrinogène. Depuis 2009, l'hémoglobine glyquée est le marqueur par excellence du diabète, bien que la relation entre une glycémie élevée et un taux d'hémoglobine glyquée ait été mise en lumière bien avant[394-395]. Elle est utilisée pour le suivi du contrôle de la glycémie chez les diabétiques sur le long terme (1-2 mois). Mais ce n'est pas un marqueur pertinent dans le cas de changements rapides dans le contrôle de la glycémie ou pour le suivi de patients qui présentent des troubles hématologiques tels que l'anémie ou qui ont des séquences d'hémoglobine mutées[248, 396]. C'est pourquoi, depuis plusieurs années, les glycations des protéines plasmatiques sont de plus en plus étudiées dans l'objectif de trouver des biomarqueurs plus pertinents[397]. C'est, par exemple, le cas de l'albumine glyquée, qui est utilisée notamment dans le cadre du suivi des diabètes gestationnels. La demi-vie de l'albumine est plus courte que celle de l'hémoglobine (18 jours vs 120 jours), ce qui permet d'obtenir des informations à plus court terme[398]. L'étude des glycations peut aussi aider à la compréhension de différents mécanismes biologiques impliqués dans des maladies diverses comme les maladies neurodégénératives[257] ou dans le vieillissement[243].

Lors de l'analyse des glycations dans le plasma, on combine deux difficultés majeures : un échantillon très « difficile » à analyser et des protéoformes très diverses et minoritaires. De plus, à cela s'ajoute les difficultés inhérentes à l'analyse des glycations : mauvaise digestion par la trypsine car encombrement stérique des lysines et des arginines, faible efficacité d'ionisation des peptides modifiés, schéma de fragmentation particuliers et manque d'outils bio-informatiques dédiés à cette modification[234].

L'enrichissement des peptides glyqués est nécessaire pour faire face à la sensibilité limitée et la gamme dynamique restreinte de l'analyse LC-MS/MS. La principale stratégie utilisée dans la littérature consiste à enrichir les échantillons en peptides glyqués en réalisant un enrichissement hors ligne sur colonne chromatographique de phase boronate[269]. Cette stratégie n'est pas optimale pour l'analyse d'un grand nombre d'échantillons. À ce jour, deux études ont proposé des protocoles permettant une analyse quantitative des glycations des protéines plasmatiques à large échelle. Jeon *et al.* ont utilisé, en 2018, une plaque filtrante à 96 puits contenant des billes fonctionnalisées avec du boronate sur un système robotique de manipulation des liquides[123]. Cet enrichissement a été décrit sur des échantillons de BSA ainsi que sur du sérum issu de patients sains et diabétiques. Il a permis d'identifier au total 982 peptides glyqués sur 524 protéines sériques différentes[123]. Un autre protocole a été mis en place en 2018 par Zhang *et al.* : un système de chromatographie à deux dimensions (LC 2D) avec une colonne d'enrichissement par affinité au boronate « en ligne » suivie d'une colonne RP C18. Avec cette approche, 376 peptides glyqués sur 64 protéines différentes ont pu être identifiés dans du plasma exposé à du glucose *in vitro* pour induire une glycation marquée des protéines[275].

Notre objectif à long terme est d'obtenir une vision la plus exhaustive possible du protéome plasmatique glyqué avec identification et quantification des produits d'Amadori mais aussi des AGEs sur un grand nombre de protéines. Cette méthode pourra ensuite être appliquée aussi bien à l'étude d'échantillons humains qu'à celle d'échantillons d'espèces aviaires dans le cadre de la poursuite du projet rapporté dans le chapitre précédant.

III. Développement de différentes stratégies analytiques pour l'identification du protéome plasmatique glyqué

Dans un premier temps et afin de réduire la complexité de l'analyse, j'ai décidé de me focaliser sur l'étude des fructosamines. Ces modifications correspondent à la forme précoce de la réaction de glycation. Elles seront appelées « glycations » dans la suite de ce travail. Les produits issus de la dégradation des fructosamines ont été étudiés dans un second temps. Afin d'aller vers une étude étendue des fructosamines sur les protéines plasmatiques, trois stratégies analytiques, décrites en figure 94, ont été envisagées.

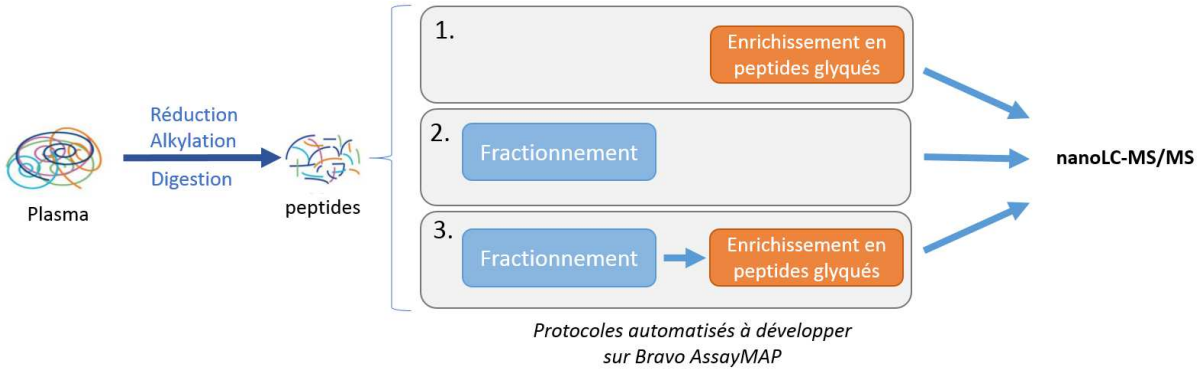

Figure 94 : Trois différentes stratégies analytiques pour l'étude des fructosamines dans le plasma : 1. l'enrichissement en peptides glyqués, 2. le fractionnement du mélange peptidique et 3. le fractionnement du mélange peptidique suivi d'un enrichissement des fractions en peptides glyqués.

Pour améliorer la reproductibilité, la robustesse et le débit d'analyses, ces trois stratégies ont été entièrement développées sur une plateforme de préparation d'échantillons, le Bravo AssayMAP (Agilent Technologies).

- Enrichissement BAC des peptides glyqués

La première stratégie consiste à enrichir les échantillons en peptides glyqués puisque ce sont des espèces très minoritaires[287]. L'enrichissement par affinité au boronate (BAC) permet de piéger les fonctions cis diol présentes sur les fructosamines, selon le principe décrit dans le chapitre 3 de la partie I. C'est une stratégie d'enrichissement classiquement utilisée dans les études sur les glycations, mais il n'existe pas de méthode décrite pour un enrichissement par micro-chromatographie automatisé sur un Bravo AssayMAP et adapté à de faibles quantités de matériel biologique. Les cartouches pour Bravo AssayMAP, conditionnées avec de la phase stationnaire fonctionnalisée avec des groupements boronates, seront appelées par la suite « cartouches BAC ». Elles ont été confectionnées à façon en collaboration avec Primoz Koselj et Mario Cindrić de BeneLab (Zagreb, Croatie). Le remplissage de la cartouche a été réalisé avec l'Affi-Gel Boronate Affinity (Bio-Rad), habituellement utilisé dans des colonnes HPLC. Ici, le remplissage est très différent puisqu'il s'agit de micro chromatographie avec seulement 5 µL de phase. De plus, cette phase est très différente des phase RP habituellement utilisées par BeneLab. Le remplissage a été évalué sur la Bravo AssayMAP avec un sucre, le sorbitol par BeneLab. Cet échantillon n'est absolument pas représentatif de nos échantillons puisque le sucre pur ne reflète pas le comportement des peptides glyqués, ni l'influence de la matrice complexe d'un mélange peptidique plasmatique. Aucun test n'a été effectué sur des mélanges peptidiques complexes. C'est pourquoi, ces cartouches ont tout d'abord été évaluées avec des protéines modèles (BSA, trastuzumab), avant de véritablement passer à des échantillons plasmatiques.

- Décomplexification du plasma par fractionnement peptidique

La deuxième stratégie vise à décomplexifier le plasma par fractionnement avec une méthode orthogonale à la chromatographie en phase inverse à pH acide utilisée pour l'analyse LC-MS.

Partie III : Développement de nouvelles approches pour étudier les glycations des protéines sanguines

J'ai optimisé et comparé deux approches de fractionnement différentes : le fractionnement hpH RP et le fractionnement sur billes fonctionnalisées avec des groupement carboxylates.

- Couplage du fractionnement peptidique et de l'enrichissement BAC

La troisième stratégie consiste à combiner le protocole développé de fractionnement peptidiques et celui d'enrichissement BAC et ainsi évaluer l'apport du couplage en terme de nombre de fructosamines observées et de profondeur d'analyse.

IV. Mise en place d'un enrichissement par chromatographie d'affinité au boronate pour identifier les peptides glyqués

A. Étude d'une protéine modèle : la BSA

La BSA est une protéine dont la séquence est proche de celle de l'albumine humaine (77% d'identité). Elle présente un fort taux de glycation du fait de sa forte concentration (elle représente 60% du plasma) et de la présence de glucose dans le sang. Ainsi, chez un individu sain, entre 1 et 10% de l'albumine est glyquée[399]. Pour des patients atteints de diabète, ce taux est généralement situé entre 20 et 30% et jusqu'à 90% dans les cas les plus extrêmes[252].

Certaines lysines de l'albumine sont des sites de glycation privilégiés du fait de leur accessibilité ou des acides aminés composant leur environnement proche. C'est le cas de la lysine K548 de la BSA (K549 ou K525 si le peptide signal n'est pas pris en compte chez l'homme)[370, 400-401]. D'autres sites sont souvent glyqués tels que K36 et K256[370]. Certains sites, tels que K235, K304 ou K559 ne sont glyqués que lorsque les valeurs de glycémie sont supérieures à la normale ou dans des expériences d'incubation forcée où les protéines sont placées dans un milieu riche en sucre. La glycation des arginines reste limitée en conditions physiologiques[370].

1. Production de BSA glyquée par incubation *in vitro* avec du glucose

Dans un premier temps, afin de faciliter l'identification des glycations, j'ai cherché à créer un échantillon protéique simple avec lequel tester les cartouches BAC. Un échantillon de BSA (N=1) a été incubé dans un milieu riche en glucose (0,4 M) afin de « forcer » la glycation et ainsi augmenter le taux de fructosamines (condition « glucose »). Un deuxième échantillon de BSA (N=1) a été incubé dans des conditions identiques mais sans glucose (condition « contrôle »). Cette première expérience permettait également d'évaluer grossièrement la cinétique (sur 14 jours) de la réaction de glycation sur la BSA grâce à un suivi par LC-MS de la protéine intacte. Le détail du protocole utilisé est décrit dans la partie expérimentale.

Les spectres MS de la BSA obtenus au jour 1 et au jour 15 de chacune des conditions (« contrôle » et « glucose ») sont représentés en figure 95.

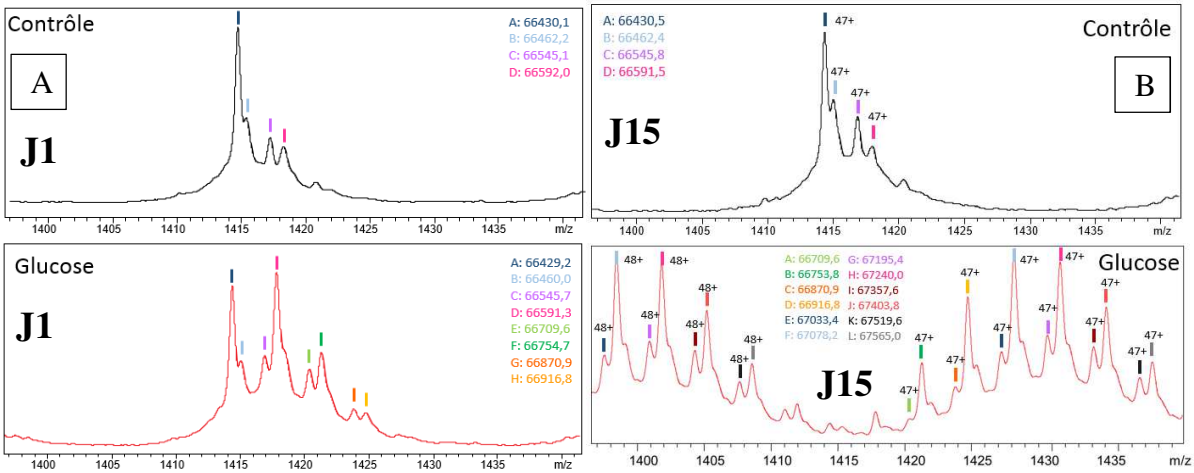

Figure 95 : Spectres LC-MS de l'état de charge 47+ de la BSA incubée dans le milieu « contrôle » sans sucre et dans le milieu contenant du glucose au jour 1 (A) et au jour 15 (B)

Sur la forme native de la protéine, (condition « contrôle », figure 4), 4 protéoformes sont identifiées, en accord avec les observations effectuées par Leblanc *et al.* sur la HSA[265]. Pour la forme majoritaire de la protéine (66 430 Da), la cystéine 34 (seule cystéine non impliquée dans des liaisons intramoléculaires) est libre. La masse 66 462 Da correspond à l'addition de deux oxydations (delta = 32 Da = 2*16 Da). La protéoforme avec une cystéine additionnée sur la cystéine 34 est identifiée à 66 545 Da (delta=115 Da). Enfin on détecte une forme glyquée à 66 592 Da (delta =162 Da). L'albumine est une protéine hétérogène car elle possède différentes protéoformes, dont une portant une glycation, naturellement. Il est intéressant de noter que le profil MS n'a pas évolué pendant les 15 jours d'incubation dans le milieu sans glucose.

Au bout de 24h d'incubation dans un milieu riche en glucose (Figure 95A, condition « glucose »), la forme majoritaire de la BSA correspond déjà à la protéine liée à une molécule de glucose. On voit également apparaitre des protéoformes portant 2 et 3 molécules de glucose (66 754 Da, 66 954 Da) ainsi que les formes liées à des cystéines libres et des molécules de glucose (66 710 Da et 66 871 Da). Ces résultats semblent être en accord avec ceux obtenus par Zuck *et al.* qui montrent un taux de glycation de la *Human Serum Albumin* [HSA] de 35% après incubation pendant 1 jour à 37 °C dans un milieu contenant 0,22 M de glucose (soit la moitié de la concentration utilisée pour notre étude)[359].

Après 15 jours d'incubation (Figure 95B, condition « glucose »), le profil de la BSA incubée avec du glucose est bien plus complexe et présente une hétérogénéité élevée, due à des dégradations et des modifications. La forme native n'est plus détectable car les espèces plus glyquées sont trop intenses, et toutes les formes détectées portent entre 2 et 7 molécules de glucose. Ces premiers résultats montrent qu'une durée d'incubation de 24 h est suffisante pour obtenir une protéine modèle glyquée pour la mise au point des méthodes d'enrichissement et de fractionnement peptidique.

2. Enrichissement BAC de la BSA glyquée *in vitro*

La BSA incubée dans 0,4 M de glucose pendant 24 h a été utilisée pour réaliser un premier enrichissement avec les cartouches BAC. Le débit délivré sur les cartouches est strictement contrôlé par la plateforme robotisée Bravo AssayMAP.

Je voulais d'une part évaluer l'efficacité des cartouches à enrichir les échantillons en peptides glyqués et d'autre part avoir un premier aperçu de la capacité de ces cartouches de 5 µL de phase stationnaire car les seules données disponibles avaient été réalisées avec un sucre, le sorbitol, soit un échantillon très éloigné des protéines/peptides que nous étudions. Le schéma analytique suivi est illustré en figure 96 et les détails du protocole sont donnés en partie expérimentale IV.A.2.

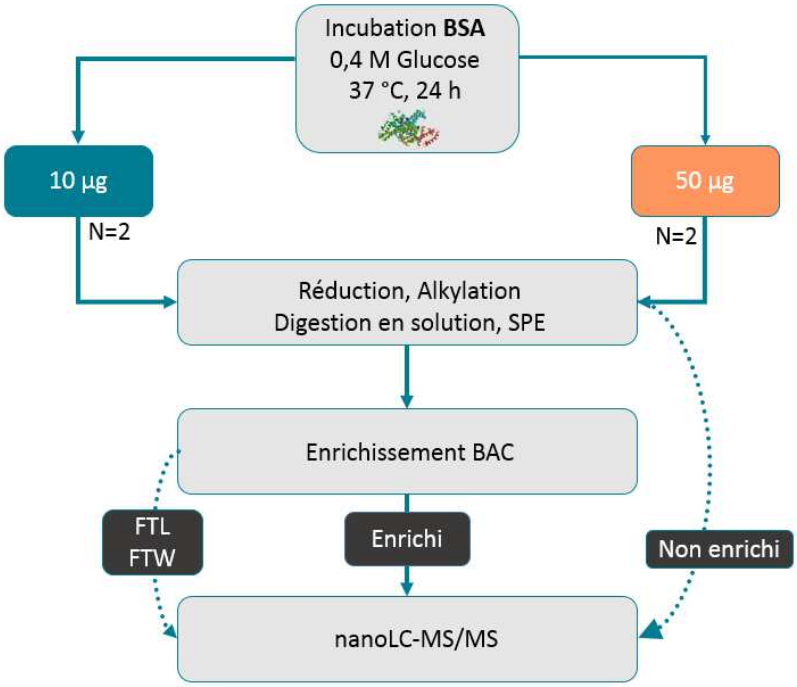

Figure 96 : Schéma analytique pour évaluer la capacité des cartouches et leur efficacité (BSA = *Bovin Serum Albumine*, SPE = *Solid Phase Extraction*, BAC = *Boronate affinity Chromatography*, FTL = *Flow Through Loading*, FTW = *Flow Through Washing*).

10 µg (N=2) et 50 µg (N=2) de BSA, préalablement incubée pendant 24 h dans un milieu riche en glucose, ont été soumis à une digestion trypsique en solution. Les peptides obtenus ont été dessalés sur cartouche C18. Le solvant a été évaporé et les peptides ont été repris dans une solution d'acétate d'ammonium. Ce mélange peptidique a ensuite été déposé sur les cartouches BAC et la fraction non retenue a été collectée (*Flow Through Loading* [FTL]). De même lors du lavage par une solution d'acétate d'ammonium, la fraction non retenue a été collectée (*Flow Through Washing* [FTW]). Enfin, après élution par une solution d'acide acétique 0,1 M, la fraction enrichie en peptides glyqués a été collectée. Les peptides obtenus ont été repris dans des volumes proportionnels à la quantité de BSA de départ (10 ou 50 µg) puis le même volume a été injecté pour l'analyse LC-MS/MS pour les 4 échantillons. Ainsi bien qu'on ne connaisse pas avec précision la quantité injectée, les résultats obtenus peuvent être comparés. Le nombre de peptides glyqués identifiés dans chaque échantillon est répertorié dans le tableau 13.

Tableau 13 : Nombre de peptides glyqués identifiés dans chaque échantillon.

	Rep 1 "10 µg"	Rep 2 "10 µg"	Rep 1 "50 µg"	Rep 2 "50 µg"
Avant enrichissement	0	0	3	2
FTL	0	0	2	0
FTW	3	1	2	2
Enrichie	6	17	22	13

On remarque que le nombre de peptides identifiés avec une fructosamine est plus élevé dans les échantillons enrichis mais il est très variable entre les deux réplicas d'une condition. L'enrichissement semble donc manquer de répétabilité. Évidemment, cette comparaison est très limitée puisqu'il n'y a que deux réplicas. Nous n'avions pas la possibilité de réaliser un plus grand nombre de réplicas car nous étions limités par le nombre de cartouches prototypes dans cette première phase de l'étude.

Dans le tableau 13, on observe également que certains peptides glyqués sont identifiés dans les fractions non retenues (FTL et FTW). Leur nombre est légèrement plus important lorsque 50 µg de peptides sont déposés sur la cartouche BAC par rapport à 10 µg. Une majorité (55%) de sites identifiés après enrichissement sont communs aux deux quantités déposées (Figure 97A). 10 sites ne sont identifiés que dans la condition « 50 µg ». Partir d'une plus grande quantité de matériel permet donc d'identifier plus de peptides glyqués. Au vu des peptides identifiés dans les fractions non retenues, la capacité de la cartouche semble proche de 50 µg, il faut donc ajuster au mieux la quantité déposée.

Les lysines K36, K498 et K548 sont les sites de glycation les plus fréquemment identifiés, aussi bien dans les fractions non retenues (FTL et FTW) que dans les fractions enrichies. Ils sont décrits à de nombreuses reprises dans la littérature dans des conditions plus ou moins riches en glucose favorisant la réaction de glycation[123, 272-273, 290]. Dans certaines de ces études, la numérotation des acides aminées est différente car le peptide signal de la BSA n'est pas pris en compte, et celle-ci est donc décalée de 24 résidus par rapport à nos données. Pour apporter une dimension quantitative à ces identifications, l'intensité observée pour les peptides glyqués portant ces modifications (K36, K498 et K548) a été comparée entre les deux conditions (10 µg et 50 µg) sur la figure 97B.

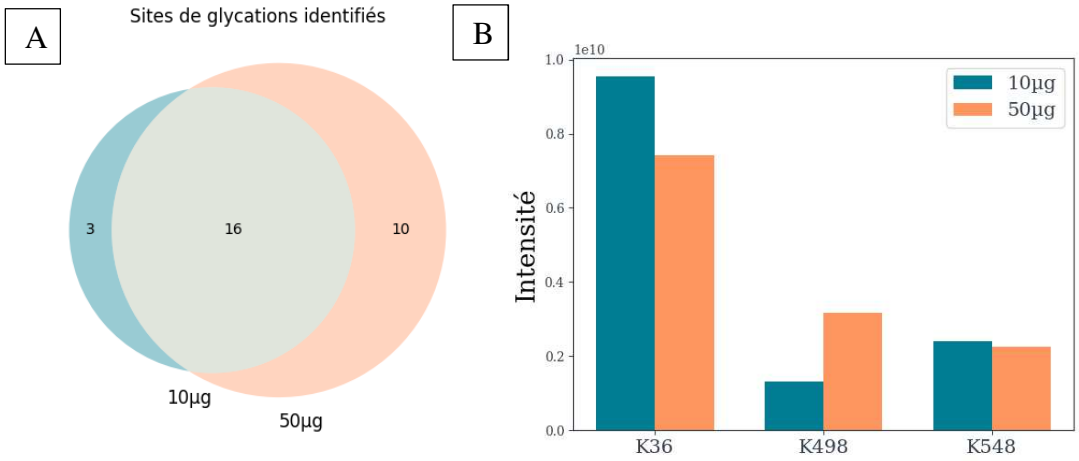

Figure 97 : Recouvrement entre les sites de glycations identifiés entre les deux conditions (A), intensité des peptides glyqués portant les lysines K36, K498 et K548 sans les échantillons enrichis après dépôt de 50 ou 10µg sur la cartouche (N=2) (B).

L'intensité des trois peptides glyqués est relativement similaire entre les deux conditions (10 et 50 µg) après enrichissement comme attendu puisque les peptides ont été repris au prorata de la quantité initiale déposée sur la cartouche et que les mêmes volumes ont été injectés pour l'analyse LC-MS/MS.

Sur la figure 98, l'intensité des 3 peptides glyqués contenant K36, K498 et K 548 est représentée pour les échantillons avant, après enrichissement et dans les fractions non retenues par la cartouche boronate (FTL, FTW) pour les deux conditions (10 µg et 50 µg).

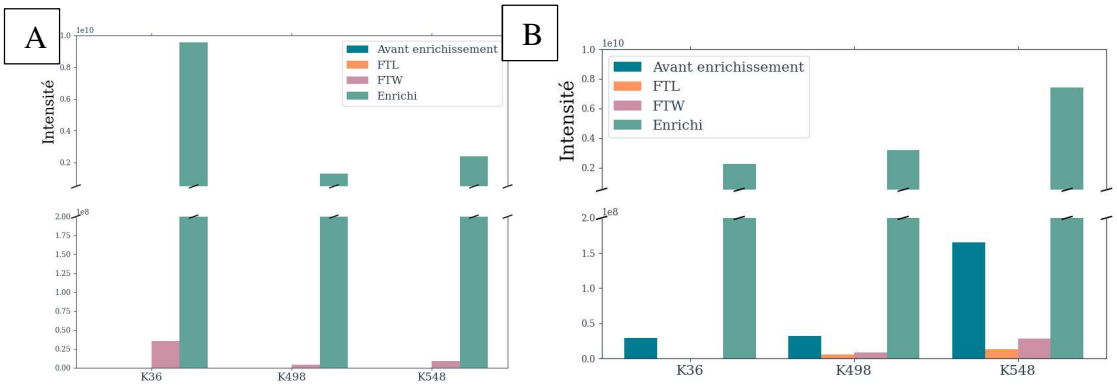

Figure 98 : Intensité moyenne (N=2) des sites de glycation dans les échantillons avant enrichissement, FTL, FTW et après enrichissement pour les dépôts de 10 µg (A) ou 50 µg (B) sur la cartouche.

Cette comparaison est limitée car la quantité exacte de peptides injectée après enrichissement ou dans les fractions non retenues ne peut pas être évaluée sans dosage peptidique. Mais il apparait tout de même que l'intensité des sites identifiés dans les échantillons enrichis est bien plus élevée. Ce résultat, ainsi que le nombre plus élevé de sites de glycation observés, montre une certaine efficacité de l'enrichissement des peptides glyqués. On remarque que de faibles quantités de peptides glyqués, à la fois en termes de nombre de peptides et d'intensité, sont présentes dans les fractions non retenues du lavage (FTW), ce qui signifie que la rétention des cartouches lors du lavage n'est pas complète. Des peptides glyqués sont également identifiés, en faible quantité, dans les fractions non retenues du chargement (FTL)

uniquement lorsque l'on dépose 50 µg. Il semble donc que la limite de capacité de la cartouche soit proche de 50 µg.

Au vu de ces premiers résultats, charger plus de quantité de matériel permet d'identifier plus de sites mais diminue la sélectivité de la cartouche. L'identification de peptides glyqués dans le FTL posera problème pour la quantification de ces peptides à l'avenir. Ce phénomène est certainement dû à un début de saturation de la cartouche, la quantité de gel est très faible (5 µL) et rien n'est connu sur la capacité de ce gel. Il se peut également que la mauvaise rétention de certains peptides glyqués soit dûe à une sélectivité limitée de la phase boronate liée à de mauvaises conditions de liaison. Pour rappel, une réaction réversible a lieu pour piéger les cis-diol à la phase boronate créant ainsi une liaison covalente. La cinétique de cette réaction est très dépendante des conditions expérimentales (pH, force ionique, temps de contact entre les peptides glyqués et la phase boronate), ce qui impacte, de fait, l'affinité et la sélectivité de l'enrichissement[287]. Pour tenter de palier à ce problème, dans les mises au point suivantes des ions Mg^{2+}, sous forme de sel $MgCl_2$ (0,5 M), seront ajoutés dans la solution de reprise de l'échantillon. Ces ions sont connus pour stabiliser le complexe anionique boronate, augmentant la sélectivité de l'enrichissement[232, 287].

3. Conclusion sur le premier test de l'enrichissement BAC

En conclusion, ce premier test semble très encourageant : l'enrichissement parait bien effectif (plus de sites identifiés et peptides glyqués plus abondants dans la fraction enrichie). Toutefois, il semble légèrement manquer de répétabilité (disparité entre les 2 réplicas) et de sélectivité (présence de peptides glyqués dans les fractions non retenues) notamment lorsque l'on augmente la quantité déposée. En déposant 50 µg, on semble se situer à la limite de la saturation de la cartouche. Après ces premiers résultats, j'ai voulu poursuivre les développements sur un autre échantillon peu complexe : le trastuzumab, un anticorps monoclonal très souvent utilisé comme protéine modèle pour la mise au point de méthodes analytiques d'étude des anticorps monoclonaux [mAb].

B. Optimisation du protocole sur un anticorps monoclonal de référence : le trastuzumab

1. L'étude des glycations dans le contexte des mAbs

Les anticorps monoclonaux (mAbs) sont utilisés pour mimer l'action des anticorps d'un patient et d'augmenter ou de restaurer la réponse immunitaire[402]. Ces molécules, composées de deux chaines lourdes et deux chaines légères, comme illustré en figure 99, ont une affinité très spécifique pour des antigènes cibles.

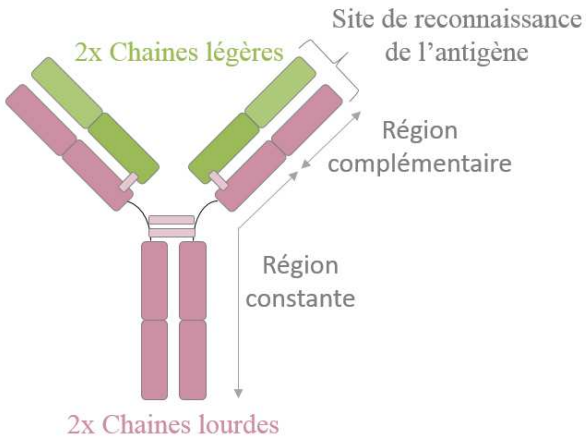

Figure 99 : Un mAb est composé de 2 chaines lourdes et deux chaines légères.

Ce sont des protéines produites en laboratoire par des cellules en cultures (microbiennes ou animales)[403]. Elles sont notamment utilisées ces dernières années dans la lutte contre le cancer. Le premier mAb a été approuvé par la US FDA en 1986 pour lutter contre le rejet lors de greffes. C'était un anticorps murin engendrant des risques immunogènes. Il a fallu attendre 2004 pour que le premier anticorps humain soit approuvé pour le traitement de l'arthrite rhumatoïde[404]. La qualité de ce produit thérapeutique est cruciale, la structure tertiaire (agrégation, dénaturation) ou secondaire (modification chimiques, troncature, dégradation…) peut être modifiée lors de la production et de la purification que subit la protéine avant de pouvoir être proposée à la vente. Les glycations sont l'une des modifications qui peuvent avoir un impact sur la qualité et l'efficacité du mAb, notamment au niveau de sa structure, de sa fonction ou de sa stabilité.[240] En effet, durant la production de ces molécules, le milieu de culture des cellules contient des concentrations de sucre en excès (concentration molaire). Durant les dernières étapes de purification, le taux de glycation sur le produit final doit être évalué. C'est pourquoi des méthodes analytiques permettant l'identification et la quantification des sites de glycation sont nécessaires. La localisation de la glycation est cruciale car les glycations présentes sur les régions constantes auront peu d'impact sur la fonction du mAb tandis qu'une glycation dans la région complémentaire est déterminante et peut affecter la reconnaissance de l'antigène. La chromatographie par affinité au boronate[246, 405] est utilisée ainsi que d'autres techniques telles que l'analyse top-down, ou par HDX-MS[246]. Aucune méthode d'analyse Aucune méthode d'analyse protéomique entièrement automatisée et dédiée à la quantification des glycations des protéines thérapeutiques n'est décrite à notre connaissance. C'est pourquoi, nous avons voulu évaluer notre protocole sur un mAb. Le trastuzumab est un mAb très utilisé dans les études de développements méthodologiques et de caractérisation d'anticorps. Aujourd'hui, il est prescrit pour traiter les cancers du sein et de l'estomac. C'est un anticorps dirigé contre la protéine *epidermal growth factor receptor 2* (HER2). Il est commercialisé par de nombreux laboratoires (Roche, Pfizer…).

2. Comparaison entre échantillon glyqué in vitro et échantillon contrôle

Après analyse par LC-MS, Lee *et al.* ont déterminés qu'entre 2 et 3% des chaines légères et lourdes du trastuzumab seraient glyquées[406]. Nous avons voulu tester les capacités des

cartouches boronate avec un échantillon contenant un nombre de peptides glyqués élevés, c'est pourquoi nous avons incubé le trastuzumab avec du glucose (condition « glucose ») plus longtemps que la BSA (72 h). Il faut également mentionner que les protéines thérapeutiques portent des glycosylations qui peuvent entrer en compétition avec les peptides glyqués lors de l'enrichissement BAC. Nous voulions comparer les résultats obtenus sur un échantillon incubé avec du glucose in vitro avec un contrôle, des échantillons « contrôle » incubés sans glucose ont été donc préparés. Après incubation, le trastuzumab a été digéré par un protocole SP3[40]. Ce protocole de digestion sur billes paramagnétiques fonctionnalisées permet d'éliminer les sels et le glucose des échantillons avant la digestion grâce aux lavages, de s'affranchir de l'étape chronophage de SPE qui était nécessaire avant l'enrichissement BAC. À la suite de cette digestion, 40 µg théoriques de peptides (N=3/ condition) ont été enrichis sur les cartouches BAC. Le schéma analytique de ces expériences est présenté en figure 100.

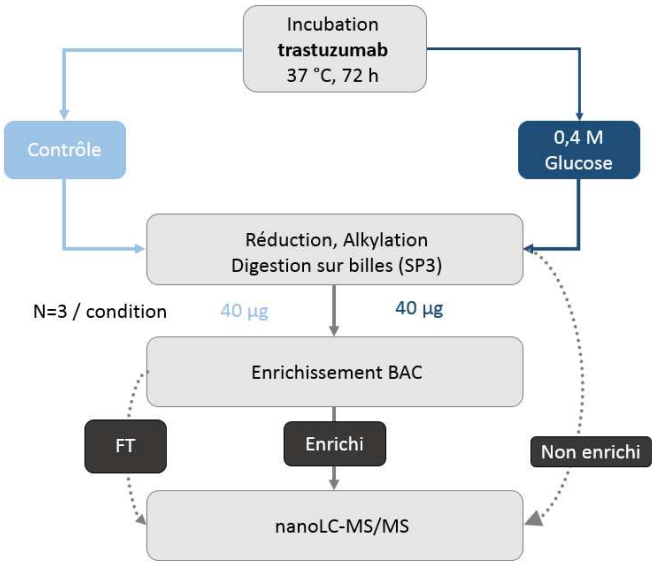

Figure 100 : Schéma analytique pour évaluer l'efficacité de l'enrichissement BAC pour l'analyse des glycations du trastuzumab ; FT = *Flow Through*.

Les sites de glycations identifiés dans les différents réplicas des conditions « Contrôle » (C) et « Glucose » (G) sont représentés sous formes de diagrammes de Venn en figure 101.

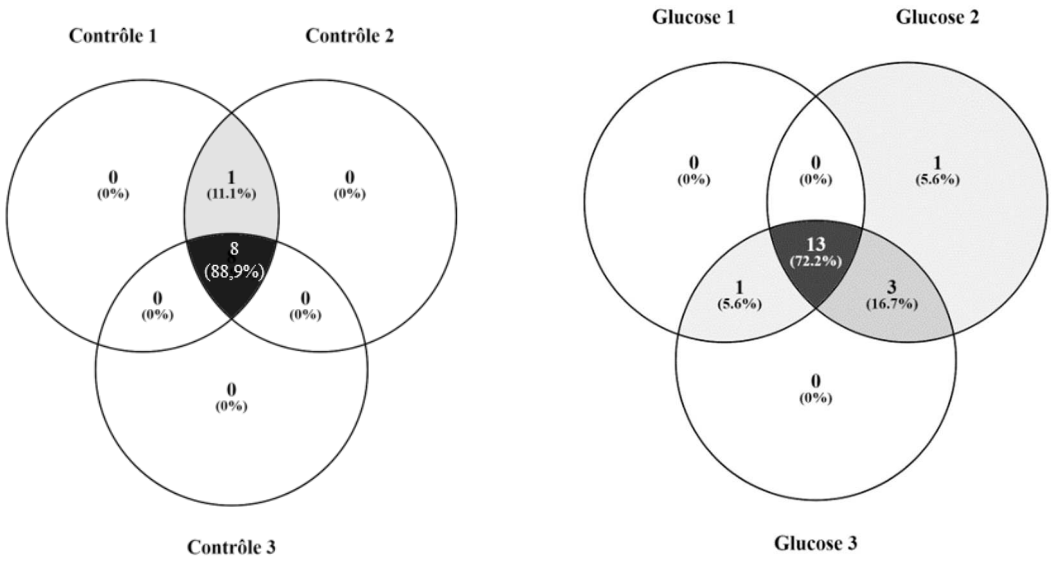

Figure 101 : Diagramme de Venn des sites de glycations identifiés dans les différents réplicas pour les conditions « Contrôle » et « Glucose ».

Sur l'ensemble des trois réplicas, 9 sites de glycation ont été identifiés sur le trastuzumab incubé dans le milieu contrôle et 18 dans ceux incubés avec du glucose. Pour la condition « Contrôle », les mêmes sites de glycation ont été identifiés dans les différents réplicas. Pour la condition « Glucose », 70% des 18 sites sont identifiés dans les 3 réplicas. Ces résultats dénotent d'une excellente répétabilité des identifications et du protocole. On peut supposer que l'addition de $MgCl_2$ dans la solution de binding, évoquée dans le paragraphe IV.A, a permis d'augmenter la stabilité du complexe boronate-cis diol et ainsi d'augmenter la sélectivité et la répétabilité de l'enrichissement BAC.

Les sites de glycations identifiés par les conditions « Contrôle » et « Glucose » sont comparés sur la figure 102.

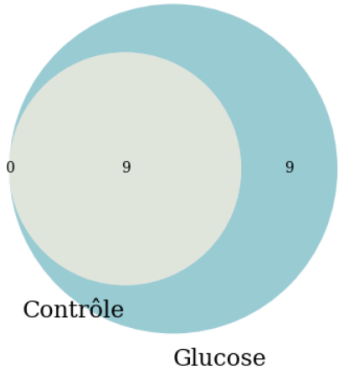

Figure 102 : Diagramme de Venn des sites de glycations identifiés dans les conditions « Contrôle » et « Glucose ».

Les sites de glycation qui sont observés dans la condition contrôle ont tous été observés après incubation au glucose. Ces résultats suggèrent que la glycation naturelle cible quelques hot spot

préférentiels, certainement les lysines les plus exposées et accessibles, puis que la glycation forcée par incubation in vitro va entrainer la glycation de nouveaux sites.

La figure 103 compare les sites de glycation identifiés dans nos données avec une note d'application qui décrit les glycations du trastuzumab[407].

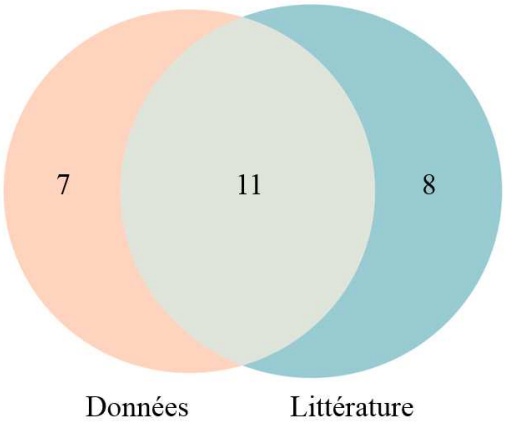

Figure 103 : Comparaison entre les sites de glycation observés dans nos données (orange, les deux conditions (« contrôle » et « glucose ») sont regroupées) et ceux listés dans la note d'application de Millán-Martín *et al*[407].

Une majorité de sites de glycation sont détectés à la fois dans nos données et celles de la littérature, ce sont les hot spot du trastuzumab. Il est important de noter que sur l'ensemble de nos échantillons, 14 sites sont identifiés sur la chaine lourde et seulement 4 sur la chaine légère. Seule une lysine de la chaine légère porte une glycation dans les échantillons « contrôle ». Ce n'est pas le cas dans l'étude de Millán-Martín *et al.* qui reporte 12 sites de glycations portés par la chaine lourde et 7 par la chaine légère. Il est difficile de comparer ces chiffres puisque les analyses n'ont pas été effectuer dans les mêmes conditions mais les tendances sont similaires.

Pour aller plus loin, nous avons voulu quantifier certains peptides glyqués avec le logiciel BioPharmaFinder (v3.2, Thermo Fisher Scientific). Il permet une quantification fine par intégration manuelle des courant d'ions extraits correspondant aux peptides d'intérêt. Il est très utilisé pour l'étude des protéines thérapeutiques[33]. L'intensité de deux peptides glyqués détectés communs aux deux conditions (« contrôle » et « glucose ») est représentée avant et après enrichissement sur la figure 104.

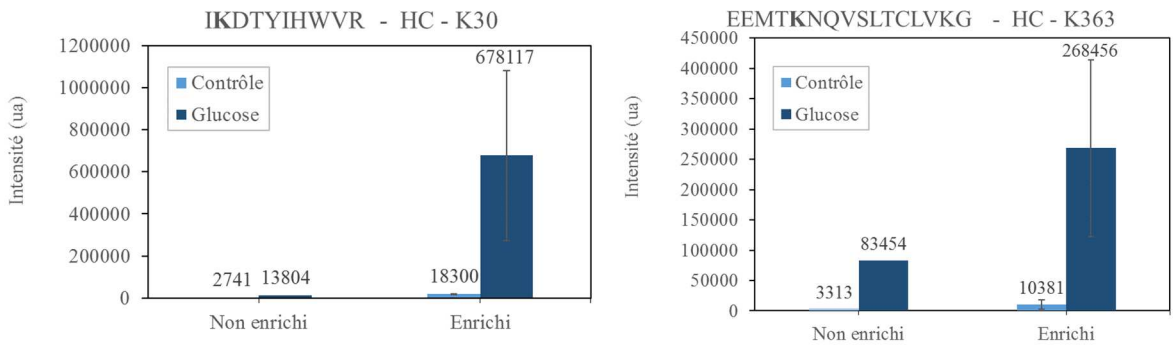

Figure 104 : Intensités de deux peptides glyqués de la chaine lourde avant et après enrichissement pour les conditions « Contrôle » et « Glucose ».

Comme attendu, l'intensité de ces deux peptides glyqués est bien plus élevée pour le mAb soumis à une glycation forcée. L'intensité observée après enrichissement pour ces peptides est multipliée au moins par 3 par rapport à la condition sans enrichissement. Cette étape d'enrichissement permet donc d'obtenir des fraction plus riche en peptides glyqués et ainsi d'augmenter la quantité de peptides glyqués à la fois en termes d'identification et d'abondance.

3. Conclusion sur l'apport de l'enrichissement BAC pour l'étude des glycations des mAbs

L'enrichissement BAC automatisé semble adapté à l'identification des glycations dans le cadre de l'analyse du trastuzumab. Des efforts doivent à présent être fournis pour évaluer la répétabilité de la quantification grâce à cette méthode. Il serait également intéressant de vérifier si les glycopeptides interfèrent lors de l'enrichissement BAC et diminuent la capacité de liaison en éliminant les glycosylations grâce à des enzymes de type PNGase avant la digestion du mAb. En effet, ces glycopeptides sont également retenus par la phase boronate et peuvent donc interférer dans la réaction de liaison entre les peptides glyqués et la phase boronate. Une autre perspective de ce travail serait l'étude d'autres mAbs tels que le NISTmAb considéré comme le matériel de référence afin de généraliser cette méthode à différentes protéines thérapeutiques. À présent, nous allons nous intéresser à un mélange beaucoup plus complexe de par la variété et la gamme dynamique de ses protéines : le plasma.

C. Enrichissement BAC d'un mélange complexe : le plasma

1. Comparaison des résultats obtenus à partir de plasma incubé *in vitro* avec ou sans glucose

Nous avons voulu tester l'efficacité des cartouches BAC à enrichir un échantillon de plasma plus représentatif des études de protéomique que les protéines modèles étudiées précédemment. Pour cette phase de mise au point du protocole, nous avons à nouveau préparé un échantillon incubé *in-vitro*. En bref, 400 µg de protéiques plasmatiques ont été incubés 3 jours avec (condition « glucose ») ou sans (condition « contrôle ») glucose puis ont été soumis à un protocole de digestion SP3. 100 µg de ces mélanges peptidiques ont été chargés sur les cartouches BAC (N=3/condition). La démarche analytique est retracée en figure 105.

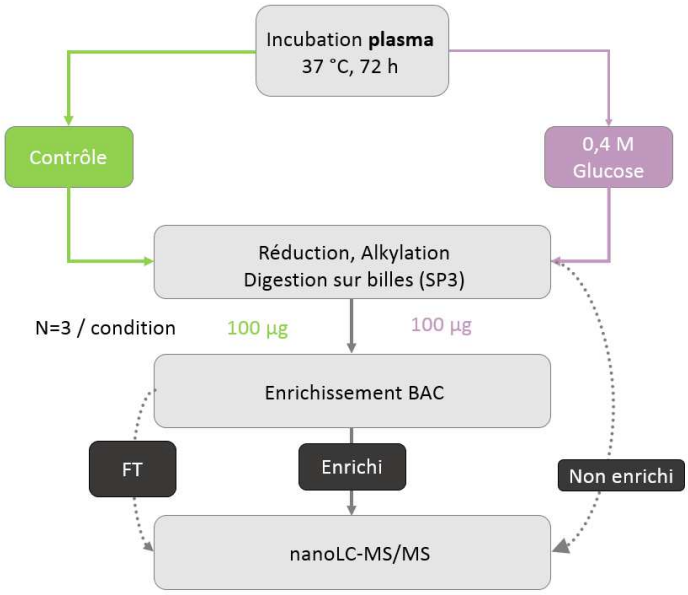

Figure 105 : Démarche analytique pour la comparaison de l'enrichissement BAC pour des échantillons de plasma incubés ou non avec du glucose.

En moyenne, un plus grand nombre de sites de glycation a été identifié dans les échantillons incubés au glucose (112 sites) par rapport à ceux incubés sans sucre (73 sites), comme illustré sur la figure 106A. C'est un premier résultat encourageant sur ce mélange complexe.

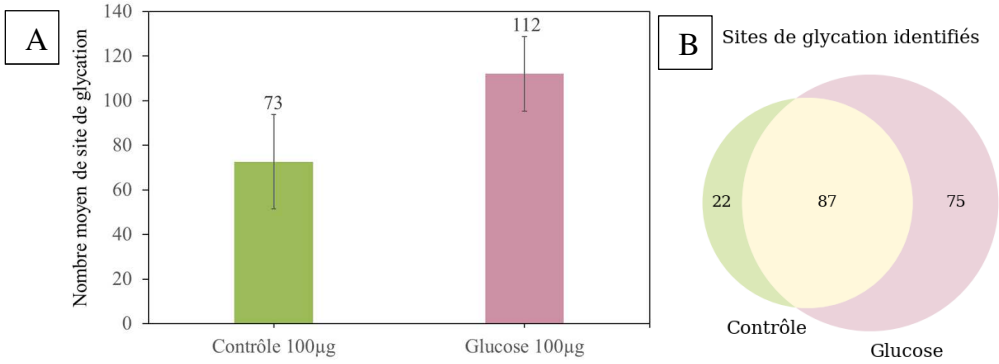

Figure 106 : Nombre de sites de glycation identifiés dans les échantillons incubés avec ou sans glucose (A), recouvrement des sites de glycation identifiées dans les deux conditions (B).

De plus, la majeure partie des sites identifiés dans les échantillons « contrôle » le sont également dans les échantillons incubés au glucose (Figure 106B).

L'intensité moyenne de trois sites de glycations majoritaires[370] : K36, K402 et K549 est illustré en figure 107A. Elle est légèrement supérieure dans les échantillons « glucose » par rapport aux échantillons « contrôle ». Mais, l'importante variabilité ne permet pas de conclure sur ce point. L'enrichissement BAC semble manquer de répétabilité pour la quantification des peptides glyqués, notamment pour les échantillons « contrôle ».

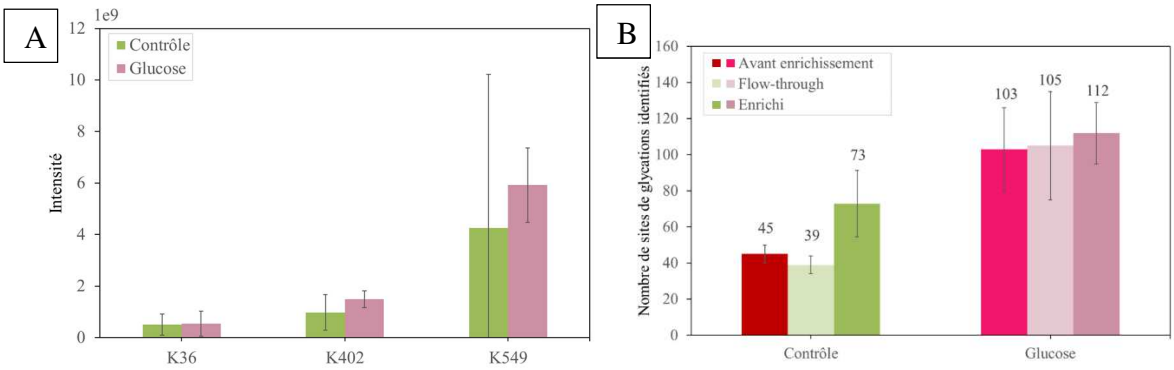

Figure 107 : Intensité moyenne des sites K36, K402 et K549 dans les échantillons enrichis par BAC des conditions « contrôle » et « glucose » (A) ; Nombre de sites de glycations observés dans les échantillons avant/après enrichissement et dans les fractions non retenues (*Flow through*) pour les conditions « contrôle » et « glucose » (B).

Pour essayer de comprendre d'où peut venir le manque de répétabilité observé dans les échantillons enrichis, nous avons cherché à savoir si tous les peptides glyqués étaient retenus sur la cartouche. C'est pourquoi le nombre de sites de glycations observés dans les échantillons avant enrichissement et dans les fractions non retenues (*Flow through* [FT]) pour les conditions « contrôle » et « glucose » a été comparé aux échantillons après enrichissement en figure 107B. On constate que différents sites de glycation sont identifiés dans les fractions non retenues (39 pour la condition « contrôle » et 105 pour la condition « glucose »). Il apparait donc que tous les peptides glyqués ne sont pas retenus par la cartouche, ce qui peut expliquer les problèmes de répétabilité observés précédemment. Étrangement, la répétabilité de la condition « contrôle » était plus limitée que celle de la condition « glucose » alors que plus de sites sont détectés dans la fraction non retenue par la cartouche pour la condition glucose. Plusieurs phénomènes pourraient contribuer à cette faible efficacité : la présence de glucose résiduelle dans les échantillons incubés au glucose, les protéines glycosylées présentes dans le plasma[408], les conditions expérimentales, la quantité limité de gel boronate qui limite la capacité de la cartouche. À ce stade, aucune hypothèse n'a été vérifiée mais il serait possible de doser le glucose par un kit tel que le kit commercialisé par Randox utilisé dans le chapitre précédent pour mesurer la glycémie des diamants mandarins. On peut toutefois s'interroger sur la sensibilité de ce genre de tests. De même, les glycosylations pourraient être retirées avant digestion pour éliminer la compétition entre ces peptides

2. Optimisation de la quantité de digest peptidique de plasma natif à déposer

Nous allons à présent évaluer la quantité de matériel de départ optimale pour l'enrichissement BAC avec un plasma incubé sans glucose. Notre objectif est de déterminer la quantité de peptides nécessaire pour maximiser le nombre de sites de glycation identifiés tout en conservant une bonne répétabilité entre les analyses aussi bien en termes d'identification que de quantification. Trois quantités ont été déposées sur les cartouches BAC : 10, 100, 150 µg (N=3 / condition). Je voulais couvrir la gamme couramment utilisée pour les analyses protéomiques, et générer assez de matériel enrichi pour l'analyse LC-MS/MS. Au vu des résultats précédents, nous avons estimé que plus de la moitié de la fraction enrichie obtenue après dépôt de 10 µg devait être injectée. Suite à un problème technique d'injection, nous

n'avons que 2 valeurs disponibles pour la condition « 10 µg ». Le nombre de peptides glyqués identifiés avec les différentes quantités de matériel de départ est présenté en figure 108A.

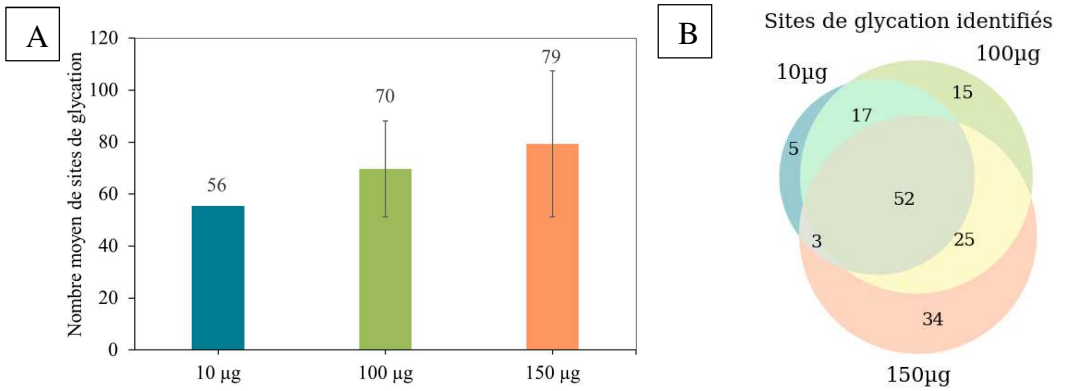

Figure 108 : Nombre moyen de sites de glycation identifiés avec le dépôt de 10 µg (N=2), 100µg (N=3) et 150µg (N=3) (A) ; Recouvrement des sites identifiés dans les 3 conditions (B).

À première vue, plus on dépose de peptides sur la cartouche, plus on identifie de peptides glyqués passant ainsi de 56 sites en moyenne pour 10 µg à 79 avec 150 µg en accord avec l'étude de Jeon *et al*[123]. Cependant la variabilité est élevée, ce qui peut indiquer un manque de répétabilité de l'enrichissement dû à une saturation de la cartouche ou une capacité de liaison à augmenter pour ce mélange complexe (conditions expérimentales, cartouches de plus large capacité…).

Sur la figure 108B, on constate que 40% (52 sites) des sites sont communs aux différentes conditions tandis que 30% sont identifiés uniquement pour la quantité de matériel la plus élevée. Par ailleurs, on identifie au total 77 sites sur 19 protéines à partir de 10 µg, 109 sites sur 24 protéines à partir de 100 µg et 114 sites sur 22 protéines à partir de 150 µg. On pourrait s'attendre à identifier beaucoup plus de sites dans les conditions avec beaucoup de matériel de départ (100 et 150 µg) par rapport aux dépôts de 10 µg. Encore une fois, il est possible que ceci reflète une saturation de la cartouche.

La figure 109 montre l'intensité de trois des sites de l'HSA les plus décrits dans la littérature (K36, K402 et K549) pour les différentes conditions (10 µg, 100 µg et 150 µg).

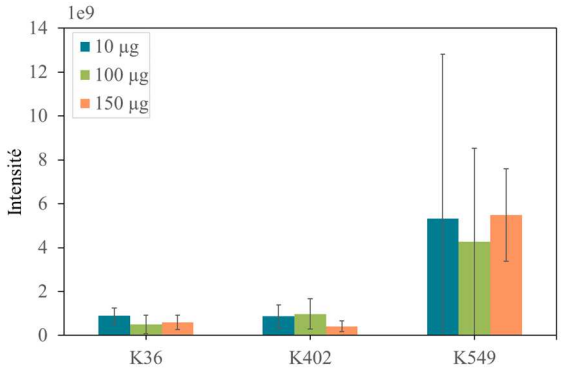

Figure 109 : Intensité moyenne dans les échantillons enrichis avec les dépôts de 10 µg (N=2, les barres d'erreurs ont été illustrées grâce à un écart-type standard calculé sur Excel), 100 µg (N=3) et 150 µg (N=3) sur les lysines 36, 402 et 549.

La quantité déposée ne semble pas influer sur l'intensité des sites. On constate aussi une grande variabilité et ce même pour la condition où le moins de matériel a été déposé (10 µg), reflétant une mauvaise répétabilité de l'enrichissement. De plus, le CV médian sur l'intensité des peptides portant les 34 sites identifiés dans 3/3 réplicas de la condition « 100 µg » est très élevé (58%). Des valeurs similaires sont obtenues pour la conditions « 150 µg » (43 sites identifiés dans tous les réplicas ; CV médian = 59%). Ces résultats démontrent le manque de répétabilité de l'enrichissement BAC sur un mélange peptidique complexe qui peut être lié à une saturation des cartouches ou à d'autres limitations discutées précédemment. Des optimisations supplémentaires sont nécessaires pour réussir à enrichir spécifiquement les peptides glyqués très minoritaires dans cet échantillon complexe. Des cartouches avec une capacité plus élevée (25 µL de phase vs 5 µL actuellement) sont en cours de développement. Elles pourraient permettre d'éviter la saturation de la cartouche. Les conditions de réactions peuvent aussi être optimisées. Le pH de la solution pour le dépôt des peptides par exemple est l'un des paramètres influant le plus sur la réaction entre les cis diol et le boronate[287].

3. Comparaison des données obtenues avec la littérature

Les sites de glycation de la HSA identifiés dans nos analyses ont été comparés à ceux décrits dans la littérature. De nombreuses études ont en effet déjà décrit les différentes lysines portant une fructosamine sur la HSA. Ces études ont été conduites dans des conditions normales de glycémie ou dans des cas d'hyperglycémie *in vivo* (situation de diabète) ou *in vitro* (incubation), et en employant différentes techniques analytiques (enrichissement BAC par colonne HPLC, fractionnement après enrichissement). Au total, les données issues de 18 études ont été utilisées et représentées sur la figure 110[252, 254, 268, 272, 275, 288-289, 370, 409-418].

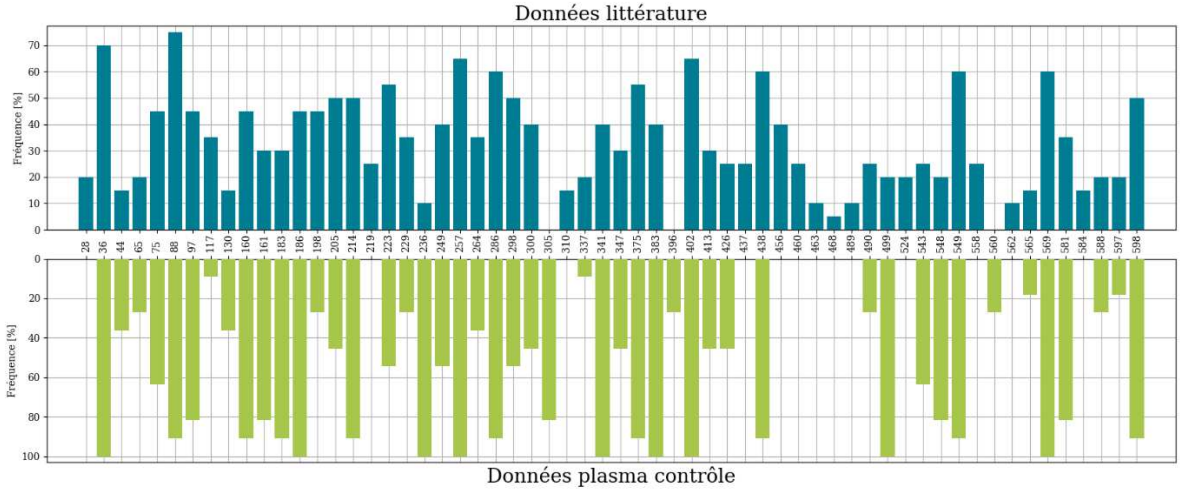

Figure 110 : La position des lysines dans la séquence de la HSA est écrite au milieu des deux graphes sur l'axe des abscisses. La fréquence d'observation de la lysine glyquée dans la littérature (N=18, bleu) et dans nos données issues du plasma natif (N=14, vert) est illustrée.

Une des premières observations qui peut être faite dans nos données issues de la condition « contrôle » (graphe vert) et dans celles de la littérature (graphe bleu) est qu'une très grande majorité des lysines de la HSA a été observée au moins une fois portant une glycation. Une conclusion partagée par Anguizola *et al*[370]. Avec cette représentation des données de la littérature (Figure 20, graphe bleu), il apparait que certaines lysines de la HSA sont

particulièrement sujettes à la glycation (K36, K88, K257, K402, K438, K549, K569) tandis que d'autres n'ont jamais été détectées portant une glycation (K305, K396, K560). Ceci reflète l'influence de l'environnement direct de la lysine ainsi que la structure de la protéine sur la réaction de glycation. La région étendue entre les lysines K456 et K548 semble moins propice aux glycations, notamment dans les données de la littérature. D'après la structure tridimensionnelle de la HSA, cette région est moins accessible[371]. Certains sites sont très fréquemment glyqués à la fois dans nos données (graphe vert) et dans les données de la littérature (graphe bleu). C'est le cas des lysines K36, K88, K186, K257, K402, K549, K569 et K598. L'objectif était de comparer si les sites détectés dans nos analyses étaient les plus enclins à être modifiés et donc ceux qui sont majoritairement détectés dans les études publiées. Les peptides portant ces sites majoritaires pourraient être utilisés par la suite comme contrôle qualité pour des analyses quantitatives.

4. Conclusion sur l'efficacité de la BAC à enrichir un échantillon de plasma en peptides glyqués

En conclusion, ces premiers résultats sur l'enrichissement BAC sont encourageants pour l'étude de peptides portant des fructosamines dans une matrice extrêmement complexe comme le plasma : l'enrichissement du plasma non incubé est effectif et il est possible de distinguer un échantillon incubé au glucose d'un non incubé. Mais il reste de nombreux problèmes à résoudre, notamment au niveau de la répétabilité de la méthode aussi bien au niveau des identifications que de la quantification. Ce manque de répétabilité peut être dû soit à la saturation de la cartouche soit à des conditions réactionnelles non optimales de liaison entre les peptides glyqués et la phase boronate. Concernant ce dernier point, des optimisations doivent être menées sur différents paramètres (pH, force ionique, vitesse de chargement de la cartouche…). Pour remédier à la saturation de la cartouche, de nouvelles cartouches de plus haute capacité sont en cours de développement. Dans la suite de ce travail, nous allons chercher à décomplexifier le mélange peptidique en amont de l'enrichissement, ce qui pourrait résoudre ces problèmes de saturation.

V. Mise en place d'une étape de fractionnement peptidique pour décomplexifier le plasma

En parallèle de l'approche par enrichissement BAC, j'ai exploré une approche de fractionnement peptidique pour décomplexifier les échantillons. En m'appuyant sur des protocoles récemment décrits, j'ai choisi de comparer deux types de fractionnements : le fractionnement sur billes magnétiques fonctionnalisées avec des groupements carboxylates (CMMB-based Isopropal Gradient peptide fractionation (**CIF**)) et sur cartouches de phase inverse à pH basique (**hpH RP**)[109, 120]. Ces deux techniques s'appuient soit sur une modification de la sélectivité de la phase RP par un changement de pH soit sur des interactions différentes à la chromatographie en phase inverse à pH acide classiquement utilisée pour l'analyse LC-MS/MS, ce qui permet de decomplexifier les mélanges peptidiques deux fois consécutives. La démarche expérimentale est représentée en figure 111.

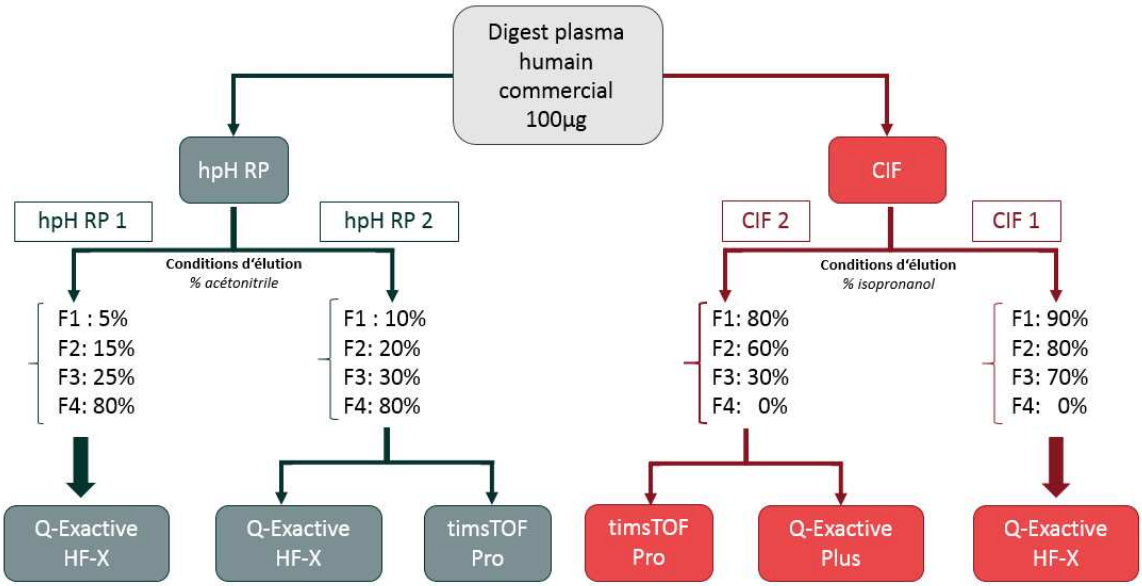

Figure 111 : Démarche expérimentale de l'optimisation du fractionnement peptidique (N=3 / fractionnement).

Dans un premier temps, les conditions d'élution ont été optimisées pour chacune des méthodes de fractionnement. Une fois ces optimisations réalisées, différentes plateformes instrumentales LC-MS/MS ont été comparées, notamment un timsTOF Pro. Enfin, les résultats des 2 méthodes ont été comparés en termes d'identification, de profondeur d'analyse et de répétabilité.

A. Optimisation de la méthode de fractionnement peptidique hpH RP

Le fractionnement hpH RP a été développé sur la plateforme de préparation d'échantillon Bravo AssayMAP. Ce fractionnement nécessite l'utilisation de phase stationnaire polymérique compatible avec le pH élevé utilisé (pH = 10). Nous avons donc utilisé les cartouches standard (5 µL) avec une phase inverse polymérique RPS. Leur capacité est donnée de 10 à 150 µg de peptides. Le protocole de fractionnement utilisant ces cartouches RPS a été décrit sur 75 µg de protéines issues de tissus congelés tels que de la rate de rat ou des biopsies de cancer pancréatique humain par Kuras *et al.* en 2019, donc des échantillons très différents du plasma en termes de composition protéique (concentration et nature)[109]. Aucune étude, à notre connaissance, n'a été publiée sur le plasma avec ce mode de fractionnement automatisé sur cartouche RP. Dans l'étude de Kuras *et al*, les peptides, obtenus après digestion des protéines, ont été piégés sur les cartouches puis élués séquentiellement en augmentant la concentration en acétonitrile. Les peptides avaient été élués par 6 fractions : 4, 10, 15, 20, 25 et 80% d'acétonitrile dans 20 mM de formate d'ammonium[109].

Un nombre important de fractions augmente drastiquement le temps d'analyse. Notre objectif a donc été de trouver un compromis entre les capacités de séparation du fractionnement et le temps d'analyse, en gardant à l'esprit qu'on envisage de réaliser un enrichissement BAC sur chaque fraction par la suite.

1. Fractionnement hpH RP 1 : premières conditions d'élution

J'ai débuté avec 4 fractions composées respectivement de 5, 15, 25 et 80% d'ACN. Les identifications obtenues sont présentées sur la figure 112.

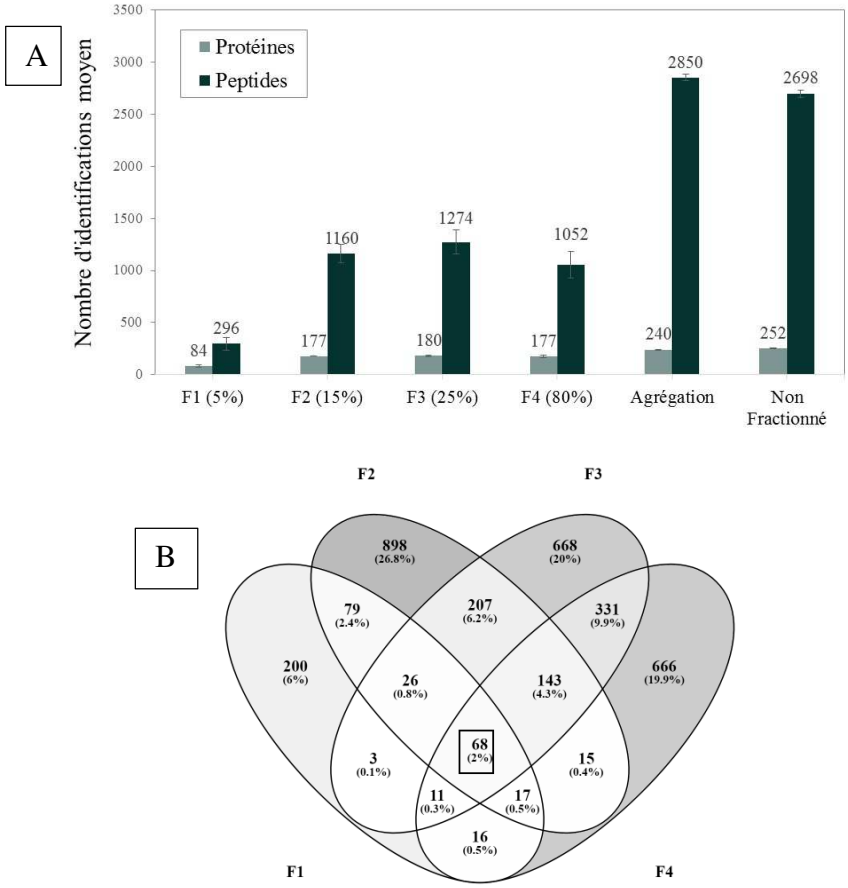

Figure 112 : Nombre de peptides et protéines identifiées dans les différentes fractions, du fractionnement hpH RP, dans l'échantillon (après agrégation des données issues des fractions) et dans l'échantillon non fractionné (A), diagramme de Venn des peptides identifiés dans les différentes fractions (B).

L'augmentation attendue du nombre d'identifications dans les échantillons fractionnés par rapport aux non fractionnés reste très faible (+ 5% de peptides identifiés pour un nombre de protéines équivalent (240 et 252 protéines respectivement avec et sans fractionnement). Pour ce premier test, nous nous sommes focalisés sur la répartition des peptides au sein des différentes fractions afin de pouvoir optimiser les conditions d'élution. Les quantités de fractions injectées n'ont pas été ajustées avec un dosage peptidique, de fait, il semblerait au vu de l'intensité moyenne du *Total Ion Current* [TIC] qu'une quantité injectée plus faible (TIC sur les fractions 2, 3 et 4 = $1,8^{E}9$) après fractionnement a été injectée par rapport aux échantillons non fractionnés (TIC= $1,07^{E}10$). Ceci pourrait expliquer la faible augmentation d'identifications après fractionnement. Par ailleurs, la fraction 1 (5% ACN) ne contient que très peu de peptides par rapport aux autres fractions dans lesquelles les peptides semblent plus équitablement répartis. Le diagramme de Venn (Figure 112B) montre le recouvrement des peptides identifiés dans les différentes fractions : il semble y avoir assez peu de recouvrement

même entre des fractions adjacentes. Ceci signifie que les pourcentages d'acétonitrile utilisés permettent bien d'éluer des populations différentes de peptides.

Pour augmenter le nombre de protéines/peptides identifiés après fractionnement, les conditions d'élution ont ensuite été modifiées.

2. Fractionnement hpH RP 2 : modification des conditions d'élution

La fraction 1 du fractionnement hpH RP 1 était la plus pauvre en peptides, c'est donc que trop peu de peptides sont élués avec 5% d'ACN. Le pourcentage d'acétonitrile dans la fraction 1 est légèrement augmenté (de 5 à 10%) tout en conservant l'écart de 10% entre chaque fractions (sauf pour la dernière fraction qui reste à 80% pour éluer les peptides les plus hydrophobes).

Le deuxième fractionnement (hpH RP 2) a donc été effectué avec les concentrations suivantes (10%, 20%, 30% et 80% en acétonitrile). Les identifications des peptides et des protéines dans les différentes fractions sont illustrées en figure 113A.

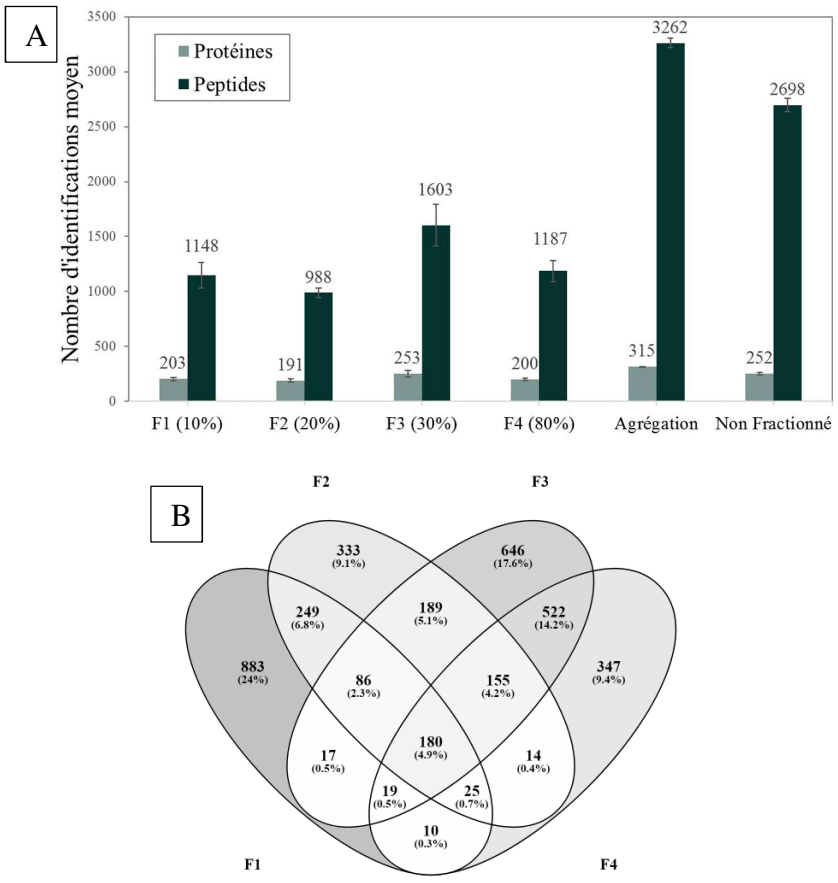

Figure 113 : Nombre de peptides et protéines identifiés dans les différentes fractions du fractionnement hpH RP, dans l'échantillon (après agrégation des données issues des fractions) et dans l'échantillon non fractionné après modification des conditions d'élution (A), diagramme de Venn des peptides identifiés dans les différentes fractions (B).

Pour ce deuxième fractionnement, bien que les peptides n'aient pas été dosés, la quantité injectée pour les fractions a été adaptée pour obtenir un TIC comparable à celui observé sur les échantillons non fractionnés (le même volume a été injecté pour chaque fraction). Ces

conditions d'élution ont permis d'augmenter significativement le nombre de peptides (+36%) et de protéines (+20%) identifiés par rapport à un échantillon de plasma non fractionné. La répartition des peptides au sein des différentes fractions est bien plus homogène que lors du premier test de fractionnement : plus de 200 protéines et 1000 peptides sont identifiées dans la fraction 1. Ces conditions d'élution ont permis d'identifier 30% de protéines et 15% de peptides de plus qu'avec le fractionnement hpH RP1. Sur le diagramme de Venn en figure 113B, on constate que peu de peptides sont communs aux différentes fractions, bien qu'ils soient plus nombreux que pour le premier fractionnement.

De meilleurs résultats sont donc obtenus avec ces nouvelles conditions d'élution. La répartition des peptides étant homogène entre les fractions, il sera difficile d'obtenir plus d'identifications en conservant ce nombre restreint de fractions. Il reste néanmoins possible d'améliorer les performances en optimisant le gradient chromatographique ou les paramètres MS (notamment, le temps d'exclusion dynamique et/ou l'accumulation des ions par l'*AGC target*) lors des analyses suivant le fractionnement ou en utilisant un spectromètre plus performant, ce que nous ferons dans un second temps.

B. Optimisation de la méthode de fractionnement peptidique CIF

En parallèle de de la méthode de fractionnement hpH RP, j'ai exploré une autre approche pour réaliser le fractionnement peptidique des échantillons. Les billes paramagnétiques couvertes de groupements carboxylates et utilisées au laboratoire dans nos protocoles automatisés (ou non) de digestion (SP3 et Auto SP3) permettent de fractionner des échantillons complexes au niveau peptidique. Ce protocole de fractionnement sur billes CMMB (CIF) a été décrit en 2021 par Deng *et al.*[120]. Pour éluer séquentiellement 100 µg de peptides issus de cellules HEK293 en interaction avec la couche fonctionnalisée des billes, des concentrations décroissantes en isopropanol ont été utilisées (90, 85, 80, 75, 70 et 0%)[120]. Tout comme pour le fractionnement hpH RP automatisé, rien n'a été décrit pour le fractionnement du plasma ou d'un échantillon semblable *via* cette méthode. Ce protocole présente l'avantage de pouvoir être facilement adapté à la quantité de matériel disponible car il suffit de modifier la quantité de billes utilisée. Il pourra être automatisé par la suite s'il se révèle performant, comme cela a été fait pour la digestion sur billes SP3.

1. Fractionnement CIF 1 : premières conditions d'élution

Cherchant un compromis entre le nombre d'identifications et le nombre de fractions, j'ai débuté avec 4 fractions contenant respectivement 90, 80, 70 et 0% d'isopropanol. Le nombre de peptides et de protéines identifiés dans chaque fraction est rapporté en figure 114A.

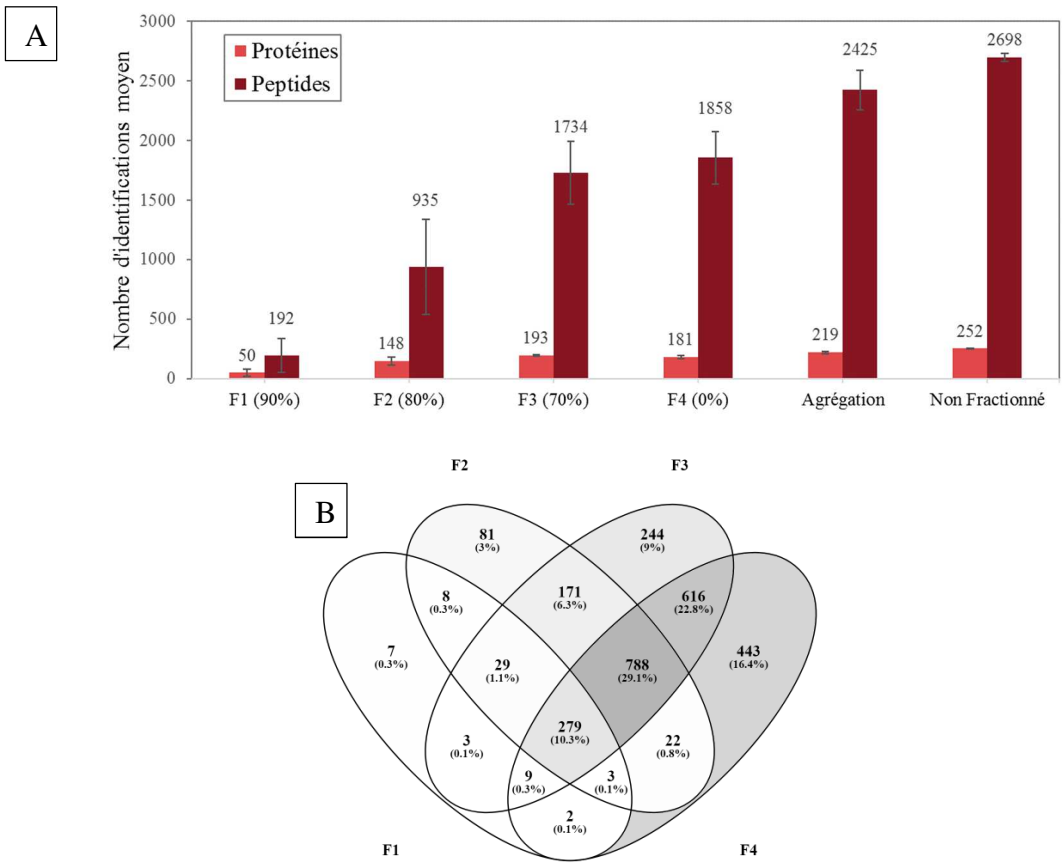

Figure 114 : Nombre de peptides et protéines identifiés dans les différentes fractions du fractionnement CIF, dans l'échantillon (après agrégation des données issues des fractions) et dans l'échantillon non fractionné (A), diagramme de Venn des peptides identifiés dans les différentes fractions (B).

Il apparait que la première fraction, et dans une moindre mesure la deuxième fraction, est plus pauvre en peptides/protéines que les autres fractions. Au niveau du recouvrement des peptides entre les fractions, illustré en figure 114B par un diagramme de Venn, il apparait que les fractions 2, 3 et 4 contiennent un grand nombre de peptides communs.

Comme lors du fractionnement hpH RP 1 (premières conditions d'élution), moins de peptides et de protéines ont été identifiés après fractionnement para rapport à l'échantillon non fractionné car l'intensité du TIC des analyses des échantillons fractionnés est plus faible que celui des échantillons non fractionnés. Il semble donc que la quantité de peptides injectée sur le spectromètre pour le fractionnement CIF est inférieure à la quantité injectée de l'échantillon non fractionné, ce qui pourrait expliquer le nombre d'identification plus faible dans l'échantillon fractionné par rapport à l'échantillon natif (sans fractionnement).

Les conditions d'élution doivent donc être revues afin d'augmenter la capacité de séparation du fractionnement CIF.

2. Fractionnement CIF 2 : modification des conditions d'élution

Afin d'augmenter le nombre de protéines/peptides identifiés, un deuxième fractionnement a été effectué avec les proportions en isopropanol suivantes : 80, 60, 30 et 0%. Pour des raisons de disponibilité instrumentales, ces injections ont été effectuées sur un Q-Exactive Plus qui est

légérement moins sensible et moins rapide que Q-Exactive HF-X utilisé précedemment. La répartition des peptides identifiés entre les différentes fractions est illustrée en figure 115.

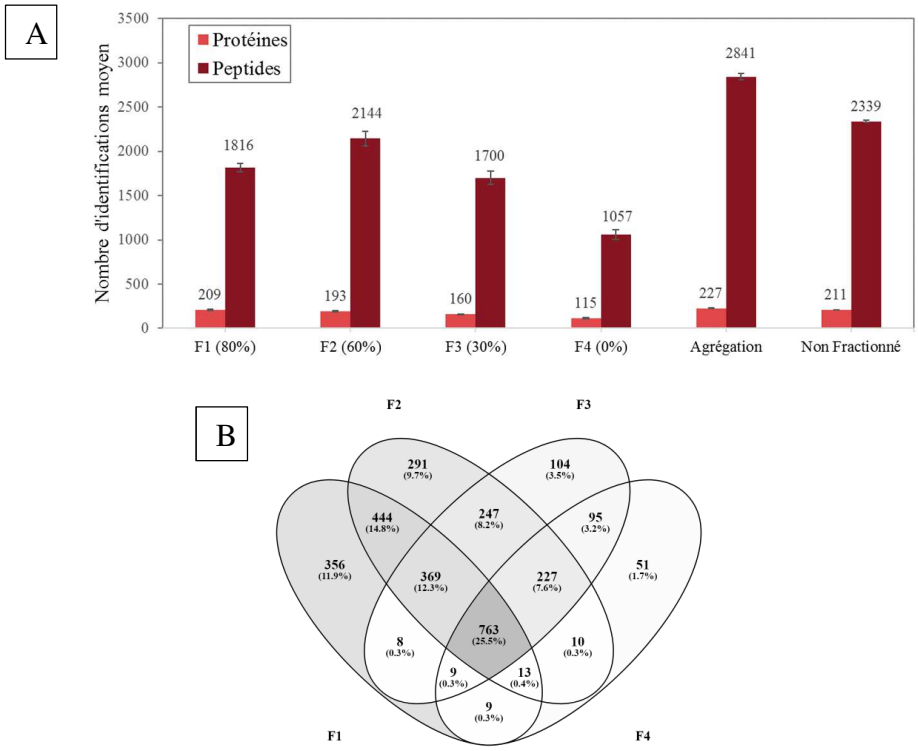

Figure 115 : Nombre de peptides et protéines identifiés dans les différentes fractions du fractionnement CIF, dans l'échantillon total (après agrégation des données issues des fractions) et dans l'échantillon non fractionné après modification des conditions d'élution (A), diagramme de Venn des peptides identifiés dans les différentes fractions (B).

Le nombre de peptides/protéines identifiés dans les différentes fractions est plus équilibré qu'avec les conditions d'élution précédentes. On observe une augmentation du nombre de peptides (2841 contre 2425 précédemment) et de protéines (227 contre 219 précédemment) identifiés avec ce fractionnement optimisé, et ce malgré des performances plus limitées de l'instrument utilisé (moins rapide et moins sensible). De plus, lorsque l'on compare l'échantillon fractionné avec l'échantillon natif (les deux conditions ont été analysées sur le Q-Exactive Plus), on observe également une augmentation du nombre de peptides (+18%) et de protéines observées (+8%). Avec ces conditions d'élution, le fractionnement CIF permet une légère augmentation du nombre d'identification. Au niveau du recouvrement des peptides entre les fractions, illustré en figure 25B par un diagramme de Venn, il apparait que les fractions 1, 2 et 3 contiennent un grand nombre de peptides communs, il semble donc que l'élution des fractions sur les billes CMMB manque un peu de sélectivité ou que les pourcentages utilisés soient trop proches.

Si l'on compare le nombre d'identifications obtenues sur l'échantillon non fractionné injecté sur le Q-Exactive HF-X et sur le Q-Exactive Plus, on constate une diminution de 400 peptides et 40 protéines avec le Q-Exactive Plus à quantité théoriques égales injectées. Cela montre bien l'importance de la sensibilité du spectromètre de masse pour augmenter la profondeur d'analyse dans l'échantillon complexe qu'est le plasma. C'est pourquoi, une fois les conditions d'élution

choisies, les deux protocoles de fractionnement ont été comparés sur un couplage LC-MS/MS de dernière génération plus performant : le timsTOF Pro.

C. Analyses des digests de plasma sur différents couplages LC-MS/MS

1. Analyse du plasma natif sans fractionnement sur différents couplages LC-MS/MS

Dans un premier temps, les résultats obtenus sur les différentes plateformes instrumentales (Q-Exactive Plus, Q-Exactive HF-X et timsTOF Pro) avec du plasma natif non fractionné ont été comparés (figure 116). L'idée est d'évaluer l'utilisation d'un timsTOF Pro qui en théorie offre une sensibilité plus importante ainsi qu'une dimension supplémentaire de mobilité ionique permettant de maximiser le nombre d'identifications.

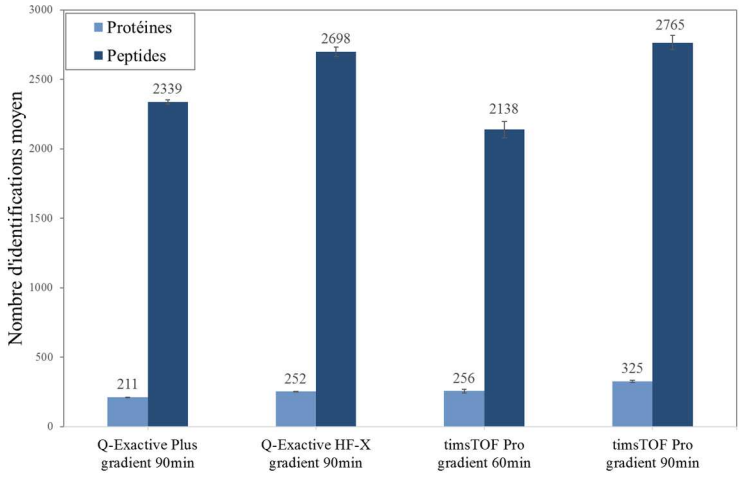

Figure 116 : Nombre de peptides et protéines identifiés dans des échantillons de plasma natif non fractionné sur les différentes plateformes.

Avec une même durée de gradient, on observe bien une augmentation du nombre de peptides et de protéines identifiés avec la sensibilité et la vitesse d'acquisition de l'instrument. Par ailleurs, il apparait que le timsTOF Pro semble identifier moins de peptides par protéines : à nombre de protéines identifiées égal (gradient 90min sur Q-Exactive HF-X / gradient 60 min sur timsTOF Pro), le nombre de peptide est en effet 26% plus faible sur le timsTOF Pro par rapport au Q-Exactive HF-X. Il est possible que la colonne NanoEasy m/z utilisée (C18, 250 mm × 75 µm, 1,7 µm; Waters) utilisé sur le système couplé au Q-Exactive HF-X soit plus performante par rapport à la colonne Aurora (C18 25 cm*75 µm, 1,6 µm ; IonOpticks) utilisée avec le timsTOF Pro ou que le gradient utilisé soit plus performant.

Sur le Q-Exactive Plus, l'analyse de 94 plasmas non déplétés de patients humains a permis l'identification de 336 protéines au total[419]. Ce chiffre élevé, obtenu avec l'instrument le moins performant comparé ici, est dû au grand nombre d'échantillons analysés. Sur le timsTOF Pro, un article en pré-print depuis juin 2019 écrit par Sandow *et al.* fait état de 238 protéines identifiées par échantillon après fractionnement en 12 fractions avec un gradient de 8 min en utilisant un couplage NanoElute - timsTOF Pro, soit des performances inférieures à ce que nous observons avec le gradient de 60 minutes sans fractionnement[420].

Les spectromètres de type Orbitrap présentent une caractéristique très intéressante dans le cadre de l'analyse du plasma : le détecteur ne sature pas. Ainsi, il a été possible d'injecter 450 ng sur les spectromètres de type Orbitrap. Au-delà de cette valeur, le nombre d'identification stagne, la colonne doit être saturée, comme présenté pour le Q-Exactive HF-X sur la figure 117. Sur le timsTOF Pro, la saturation du système (colonne, cellule de mobilité et/ou détecteur) est atteinte à 300 ng (données non présentées). Or pour l'analyse du plasma, il est intéressant de pouvoir injecter de grandes quantité de matériel afin de faciliter l'identification des espèces minoritaires (dans la limite de la gamme dynamique de l'instrument).

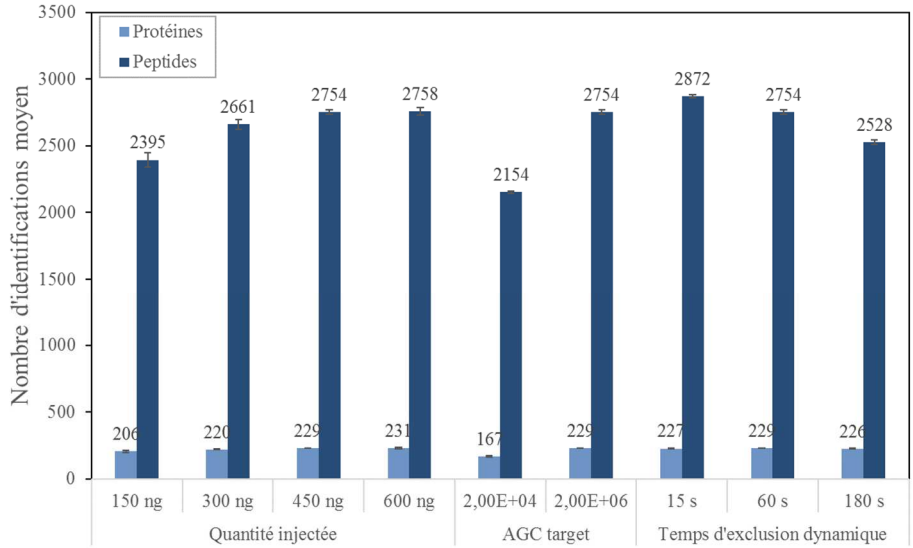

Figure 117 : Nombre de peptides et protéines identifiés dans des échantillons de plasma natif non fractionné sur un Q-Exactive HF-X en modifiant différents paramètres (quantité injectée, *AGC target* et temps d'exclusion dynamique).

D'autres paramètres ont également été modifiés : l'*AGC target* et le temps d'exclusion dynamique. L'*AGC target* correspond au nombre d'ions qui doivent être accumulés dans l'orbitrap avant que le spectre MS ne soit acquis. Plus l'*AGC target* est élevé plus un nombre important d'ions seront analysés, cette valeur influe sur la qualité des spectres MS obtenus. Comme attendu, la diminution de l'*AGC target* de 2^E6 (valeur standard) à 2^E4 impacte négativement le nombre d'identifications (-27% de peptides identifiés). Le temps d'exclusion dynamique définit, quant à lui, le temps durant lequel un ion déjà fragmenté ne sera plus re-sélectionné pour être fragmenté. L'augmentation du temps d'exclusion dynamique diminue le nombre de peptides identifiés (-14% entre un temps d'exclusion de 15 s et de 180 s). Un temps d'exclusion trop long ne permet pas de reprendre un spectre de meilleure qualité sur des ions minoritaires.

Les conditions optimales obtenues avec le Q-Exactive HF-X sont donc : une quantité théorique de 450 ng de peptides injectés avec un *AGC target* de 2^E6 et un temps d'exclusion dynamique de 15 s. Ces optimisations pourront être utilisées par la suite lorsque les méthodes de fractionnement seront mises au point. Concernant l'utilisation du timsTOF Pro, les résultats obtenus sont aussi bons qu'avec le Q-Exactive HF-X mais nous n'avons pas eu l'occasion d'optimiser plus en avant les méthodes chromatographiques et d'acquisition. Il serait

intéressant d'optimiser des paramètres tels que le gradient, la gamme de masse et de mobilité ionique. Pour la suite des injections (échantillons fractionnés) sur le timsTOF Pro, un gradient de 60 min a été utilisé afin de limiter le temps d'analyse.

2. Analyse du plasma fractionné sur différents couplages LC-MS/MS

Les échantillons obtenus précédemment avec les conditions d'élution modifiées (fractionnement hpH RP 2 (10, 20, 30, 80% ACN) et CIF 2 (80, 60, 30, 0% isopropanol) ont été injectés sur le timsTOF Pro et sur des instruments de type Q-Orbitrap.

Le profil de répartition des peptides dans les fractions reste semblable à ce qui a été présenté précédemment pour les méthodes de fractionnement (données non représentées). Le nombre de peptides et de protéines identifiés après regroupement des données des différentes fractions issues des fractionnements CIF et hpH RP sur les différentes plateformes expérimentales est représenté en figure 118.

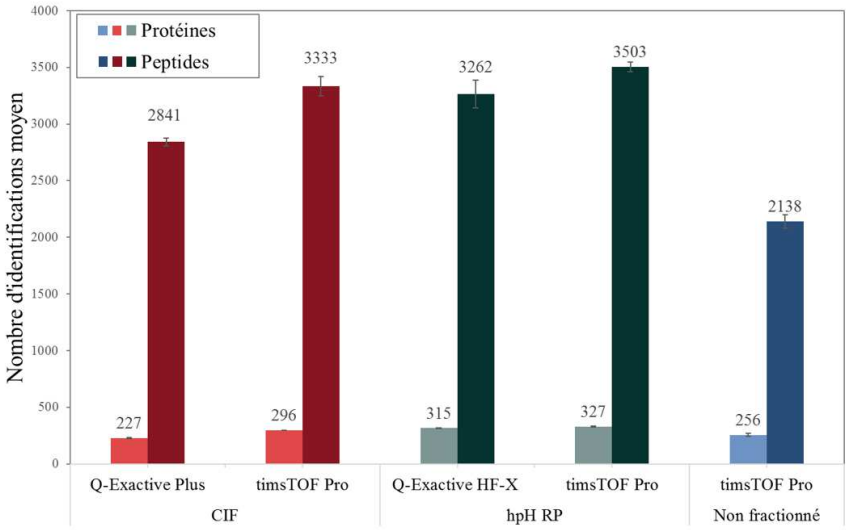

Figure 118 : Nombre moyen de peptides et protéines identifiés dans les échantillons fractionnés après agrégation des fractions (CIF : rouge, hpH RP : vert) et les échantillons non fractionnés (bleu) sur les différentes plateformes instrumentales.

Avec l'utilisation du timsTOF Pro, on observe une augmentation du nombre d'identifications par rapport aux deux plateformes utilisées jusqu'alors, notamment en comparaison avec le Q-Exactive Plus. Le fractionnement permet d'augmenter le nombre de protéines identifiées de15% (CIF) et de 27% (hpH RP) par rapport à un échantillon non fractionné analysé avec un gradient chromatographique identique (60min). Le nombre de peptides identifiés est également augmenté de 55% (CIF) et de 64% (hpH RP). L'apport le plus marqué du fractionnement dans ce cas est l'augmentation du nombre de peptides identifiés, ce qui permet une meilleure couverture de séquence. Les deux protocoles de fractionnement permettent d'aboutir à des performances similaires en termes d'identification, ce qui avait déjà été montré par Deng *et al.* sur des cellules HK2n[120]. Nous allons à présent comparer les résultats obtenus entre les deux méthodes de fractionnement de façon plus détaillée.

D. Comparaison des deux méthodes de fractionnement développées

Pour la comparaison des méthodes de fractionnement (hpH RP et CIF), seules les données acquises sur le timsTOF Pro ont été utilisées.

1. Profondeur de l'analyse

Un des premiers critères pour évaluer la performance du fractionnement est de mesurer la profondeur de l'analyse du protéome. Pour cela, deux métriques peuvent être utilisées : la couverture de séquence des protéines identifiées et la concentration estimée des protéines identifiées dans le plasma. Les concentrations de 4072 protéines plasmatiques ont été mesurées par diverses études regroupées dans le cadre du HPPP[421]. Sur la figure 119A, la distribution de la concentration estimée des protéines plasmatiques identifiées avec les deux méthodes de fractionnement hpH RP et CIF ainsi que dans les échantillons non fractionnés est représentée après leur transformation en log.

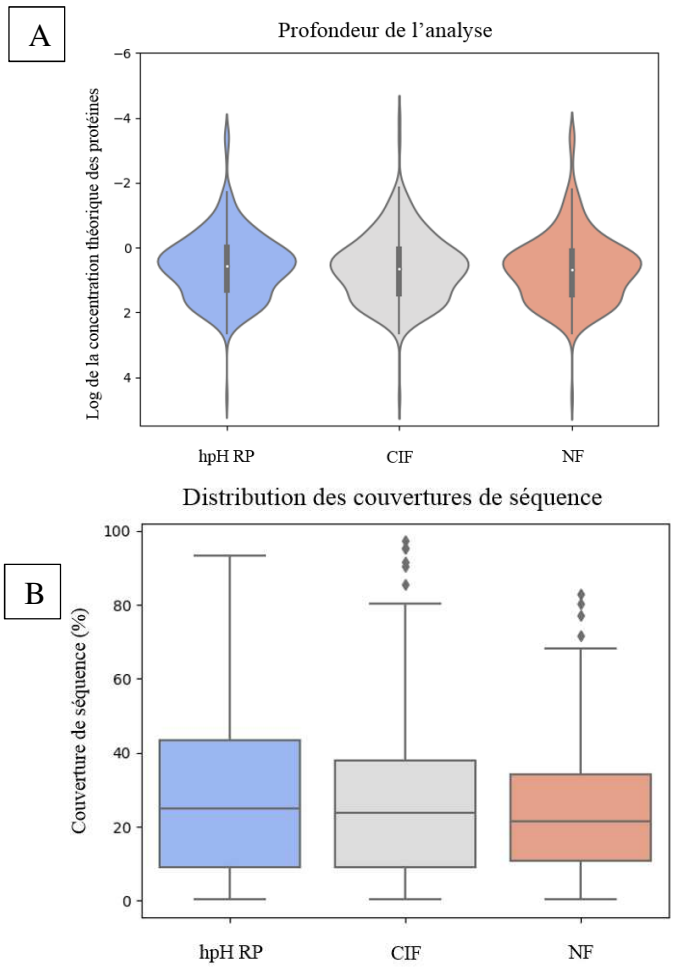

Figure 119 : Distribution de la concentration théorique des protéines identifiées (les valeurs mesurées dans les études HPPP ont été correlées avec nos données grâce à l'identifiant protéique d'Uniprot) avec les deux méthodes de fractionnement hpH RP et CIF ainsi que dans les échantillons non fractionnés (NF) (A) et de la couverture de séquence (B) dans les échantillons fractionnés (hpH RP, CIF) et non fractionnés.

Concernant la concentration théorique des protéines identifiées dans nos analyses, les trois distributions sont quasiment identiques et reflètent donc une profondeur d'analyse similaire. Une grande partie de la gamme dynamique du plasma est couverte par chacune des analyses (10 ordres de grandeurs sont représentés). La majorité des protéines identifiées se situe à une concentration d'environ 0,3 mg/L (300 000 pg/mL). Les protéines les plus faiblement abondantes ne sont pas identifiées malgré le fractionnement.

La distribution des couvertures de séquences sur l'ensemble des protéines identifiées avec les trois méthodes est présentée en figure 119B. Les valeurs obtenues avec fractionnement (CIF ou hpH RP) sont plus dispersées que celles obtenues sans fractionnement, mais la médiane reste stable entre les analyses. Il n'est donc pas possible de conclure que le fractionnement augmente significativement la couverture de séquence des protéines bien que plus de peptides soient identifiés après fractionnement.

2. Nature des peptides identifiés par les deux fractionnements

Les deux fractionnements reposent sur des interactions différentes (interactions hydrophobes pour hpH RP, interactions de type HILIC pour CIF), c'est pourquoi on s'attend à obtenir des populations de peptides différentes après fractionnement. Si on observe un bon recouvrement au niveau protéique entre les deux approches (64%), seuls 45% des peptides identifiés sont communs aux deux fractionnements (figure 120). Ils permettent donc d'obtenir des populations peptidiques complémentaires. Un plus grand nombre de peptides et de protéines sont identifiés par fractionnement hpH RP que par CIF.

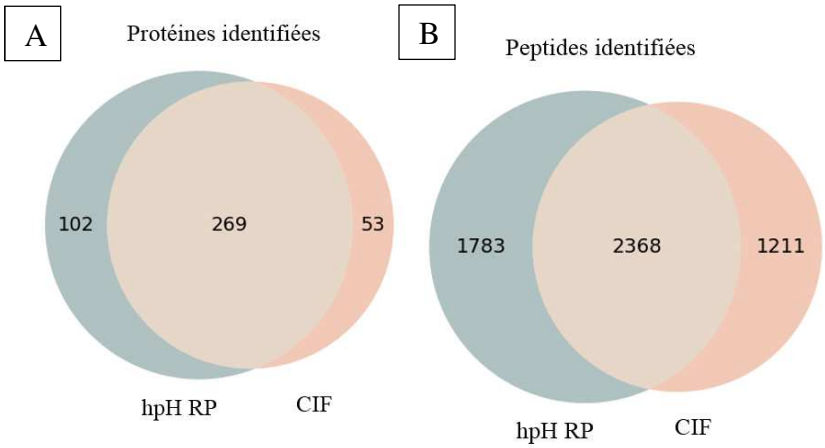

Figure 120 : Recouvrement des protéines (A) et les peptides (B) identifiés entre les données des deux protocoles de fractionnement (hpH RP et CIF).

La longueur et le pourcentage en acides aminés hydrophobes des peptides obtenus ont été comparés pour identifier si une certaine population peptidique est mieux retenue sur l'une des deux phases utilisées (RP ou fonction carboxylates) par un des deux fractionnements (figure 121).

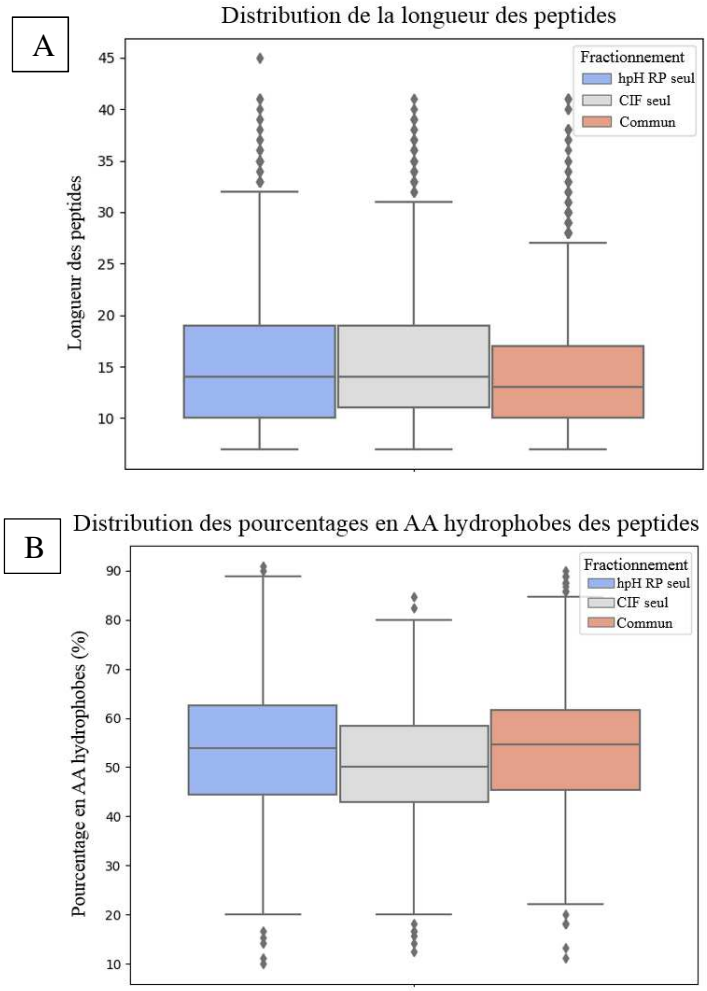

Figure 121 : Distribution de la longueur des peptides (A) et du pourcentage d'acides aminés hydrophobes (B) pour les peptides identifiés uniquement dans les données de l'un ou l'autre des deux protocoles de fractionnement (hpH RP 2 : bleu et CIF 2 : gris) ou dans les deux protocoles (rouge). AA = acide aminé.

Il semblerait que les peptides les plus courts sont un peu mieux retenus et identifiés par les deux fractionnements (hpH RP et CIF), au contraire, ce sont des peptides plus longs (en médiane 2 acides aminés de plus) qui sont retenus par l'une des deux méthodes uniquement. Ceci suggère que l'impact de la composition en acide aminés sur les interactions spécifiques à chacune des deux méthodes est plus important si le peptide est plus long. Par ailleurs, le fractionnement CIF a tendance à identifier des peptides moins hydrophobes que ceux identifiés uniquement avec le fractionnement hpH RP. Ceci est une propriété connue des phases HILIC utilisées fréquemment pour retenir des composés trop polaires ou hydrophiles pour être retenus sur des phases RP[349].

Si l'on compare la distribution des temps de rétention de tous les peptides identifiés pour les deux fractionnements dans chacune des quatres fractions, on obtient des profils très différents entre les deux méthodes comme représenté sur les diagrammes en violons en figure 122. La répartition des peptides sur le gradient est importante, surtout avec des analyses DDA car il faut éviter la co-élution des peptides pour maximiser les identifications.

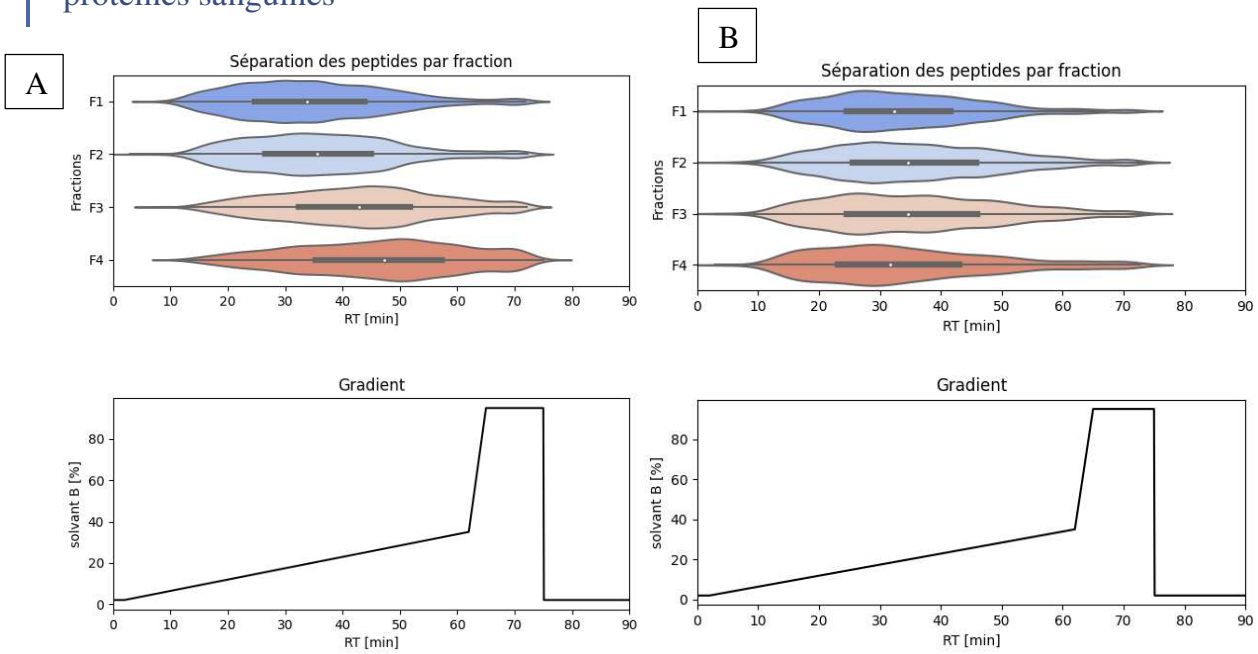

Figure 122 : Distribution des temps de rétention des peptides de chaque fraction pour le fractionnement hpH RP (A) et pour le fractionnement CIF (B) ; le gradient est tracé sous la distribution.

Pour le fractionnement CIF, la mediane (représentée par un point blanc) est alignée dans les 4 fractions et la distribution des peptides sur le gradient semble très homogène dans chaque fraction. La distribution correspond à une analyse protéomique classique avec une médiane au milieu du gradient, peu de peptides sont élués en début ou en fin de gradient. Au contraire, avec le fractionnement hpH RP, la mediane se décale vers des temps de rétention plus élevés avec le pourcentage croissant en acétonitrile. Ces résultats montrent que le fractionnement hpH RP est moins orthogonal à la séparation chromatographique à pH acide de l'analyse LC-MS/MS que le fractionnement CIF. Ces résultats sont similaires à ceux montrés par Deng *et al.*[120]. Malgré une orthogonalité moins marquée avec l'analyse LC-MS/MS, plus de protéines/peptides sont identifiés avec le fractionnement hpH RP.

3. Quantification des protéines

Pour réaliser la quantification des protéines détectées par chaque protocole, les protéines avec une valeur manquante ou plus sur les trois réplicas ont été éliminées, les valeurs obtenues sont données en tableau 14.

Tableau 14 : Nombre de protéines identifiées dans 3/3 réplicas.

	hpH RP	CIF	NF
Nombre de protéines avant filtre	371	323	296
Nombre de protéines identifiées dans 3/3 réplicas	294	269	215

La distribution des CVs sur les intensités brutes est donnée en figure 123.

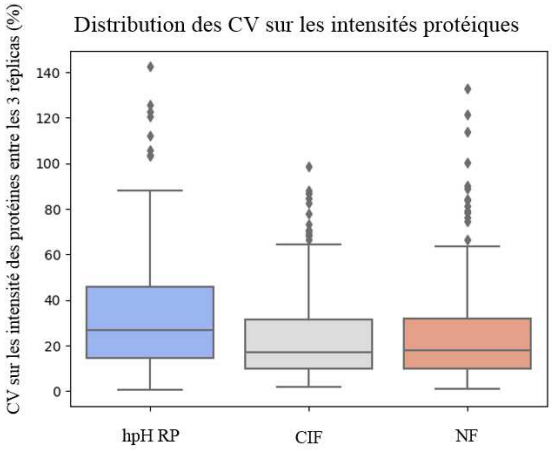

Figure 123 : Distribution des CVs sur les intensités des protéines identifiées dans les échantillons fractionnés (hpH RP 2, CIF 2) et non fractionnés.

Les CVs du fractionnement hpH RP sont légérement supérieurs à ceux sans fractionnement (CV médian : 28% contre 21%). Le fractionnement CIF permet une quantification aussi répétable que sans fractionnement tout en apportant 25% de protéines supplémentaires, il apparait donc comme un compromis très intéressant. Il est possible que la répétabilité accrue du fractionnement CIF par rapport au fractionnement hpH RP soit due au fait que chaque fraction est éluée avec la solution éluante deux fois consécutives.

4. Conclusion

En conclusion, nous avons bien montré qu'un fractionnement, même limité en termes de nombre de fractions, permet d'augmenter le nombre de protéines et de peptides identifiés dans des échantillons de plasma par rapport à une analyse sans fractionnement préalable. Le nombre de protéines identifiées dans le plasma est plus faible que ce qui est communément rapporté dans la littérature, notamment dans la comparaison bibliographique réalisée par Kaur *et al.,* où plus de 1000 protéines sont détectées dans la majorité des études. Cependant, il faut garder à l'esprit que toutes ces études contiennent une étape de déplétion des protéines majoritaires ainsi que des fractionnements très poussés (entre 18 et 100 fractions par échantillon)[422]. Il est important de ne pas perdre de vue notre objectif qui reste l'analyse du protéome plasmatique glyqué, y compris chez des espèces pour lesquelles aucune solution technologique n'a été développée pour dépléter les protéines plasmatiques majoritaires.

À l'avenir, les fractions peptidiques seront dosées au préalable pour ajuster la quantité de matériel injecté, ce qui permettra de gagner en répétabilité. Un plus grand nombre de peptides/protéines a été identifié avec le fractionnement hpH RP par rapport au fractionnement CIF. Pourtant, ce dernier est une technique plus orthogonale à l'analyse chromatographique RP à pH acide que le fractionnement hpH RP. Le fractionnement CIF permet aussi une quantification plus reproductible. Il présente aussi d'autres avantages, en termes de coût, de flexibilité sur la quantité de matériel de départ, sur la nature des greffages utilisés pour fonctionnaliser les billes (HILIC, SAX...). Ce fractionnement devrait être implémenté sur la plateforme automatisée Bravo AssayMAP dans un avenir proche, comme c'est déjà le cas pour la digestion des protéines sur billes CMMB (protocole SP3). Au contraire, le fractionnement

HpH RP est déjà automatisé sur le Bravo AssayMAP. Ce fractionnement a été utilisé, dans un premier temps, pour coupler l'étape de fractionnement à l'étape d'enrichissement BAC.

Pour aller plus loin dans la couverture du protéome plasmatique, des optimisations doivent encore être réalisées sur les méthodes LC et MS utilisées et notamment pour le timsTOF Pro qui permet déjà l'identification d'un plus grand nombre de peptides/protéines sur les échantillons fractionnés.

E. Identification des peptides glyqués dans les données issues du fractionnement hpH RP

Nous avons profité de l'opportunité d'avoir des jeux de données générés par les optimisations du paragraphe précédent pour chercher les peptides glyqués identifiés dans le plasma natif fractionné. C'est pourquoi, les données générées sur le timsTOF Pro avec le fractionnement hpH RP 2 (paragraphe V.A.2) ont été soumises à une recherche des fructosamines.

À partir de ces données, 702 sites de glycation, répartis sur 162 protéines différentes, ont été identifiés. Ce chiffre est bien plus élevé que celui obtenu après enrichissement BAC du plasma (natif : 73 sites ou incubé : 112 sites) dans le paragraphe IV.C. La décomplexification de l'échantillon semble donc plus favorable à l'identification des peptides glyqués que l'enrichissement.

Pour la quantification des peptides glyqués, seuls les sites sans valeur manquante sur les trois réplicas techniques ont été conservés. Après ce filtre, 158 sites ont été retenus sur 39 protéines différentes, ce qui est très limité par rapport au nombre de sites identifiés initialement. La majeure partie de ces 158 sites appartient à l'albumine (42 sites), ce qui est en accord avec le fait que l'albumine est une protéine naturellement fortement glyquée[370]. De plus, c'est la protéine la plus abondante du plasma, ce qui montre les limites du mode DDA pour l'identification des peptides glyqués minoritaires. Le CV médian obtenu sur l'intensité brute des peptides glyqués est de 38%. Cette valeur est plus élevée que le CV médian de l'intensité des protéines obtenu sur ce jeu de données (28%). Ces résultats mettent en lumière l'un des aspects contraignant de l'analyse quantitative des peptides modifiés : la quantification est réalisée sur un seul peptide. L'analyse doit donc être très robuste car la quantification ne peut pas être lissée par l'identification/quantification de plusieurs peptides palliant aux valeurs manquantes ou aux variations observées sur l'abondance d'un peptide comme discuté dans la chapitre 2 de la partie II pour les phosphorylations. De plus, l'analyse des peptides glyqués est sujette au sous échantillonnage en mode DDA lié à la faible abondance de ces derniers. C'est pourquoi, les deux stratégies développées précédemment (fractionnement hpH RP et enrichissement BAC) vont être couplées afin de tenter d'enrichir les échantillons en peptides glyqués.

VI. Couplage du fractionnement peptidique et de l'enrichissement BAC des peptides glyqués

Les deux approches de fractionnement peptidique et enrichissement BAC développées précédemment ont ensuite été couplées selon le schéma expérimental en figure 124.

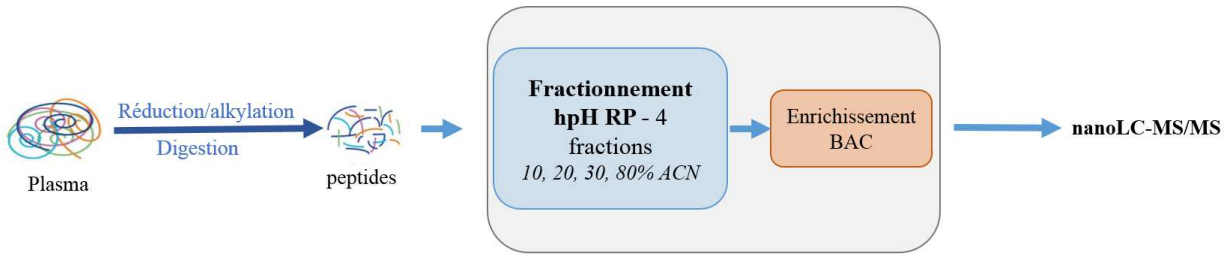

Figure 124 : Schéma expérimental du couplage entre le fractionnement hpH RP et l'enrichissement BAC.

100 µg de peptides obtenus après digestion trypsique des protéines du plasma natif commercial ont été soumis à un fractionnement hpH RP, où 4 fractions ont été éluées (10, 20, 30 et 80% en ACN) sur cartouches RPS. Ces fractions ont ensuite été enrichies sur des cartouches BAC (5 µL de phase). Les fractions enrichies ont été analysées sur un couplage nanoLC-Q-Exactive Plus. Les fractions non enrichies ont également été analysées pour évaluer l'apport de l'enrichissement. Le nombre de sites de glycation et le nombre de protéines associées est représenté en figure 35.

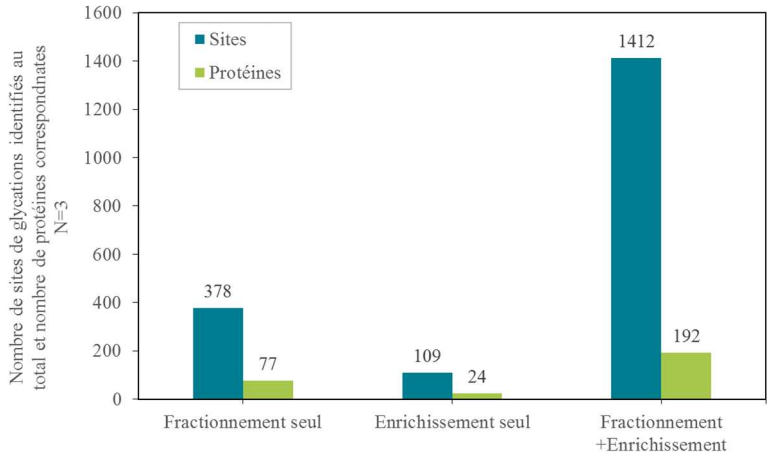

Figure 125 : Nombre total de sites de glycation identifiés dans les trois réplicas des échantillons uniquement fractionnés, uniquement enrichis et fractionnés + enrichis analysés sur le Q-Exactive Plus et le nombre de protéines glyquées correspondantes.

En combinant le fractionnement peptidique et l'enrichissement en peptides glyqués, 1412 sites de glycation ont été identifiés sur 192 protéines différentes. Avec ce protocole, le nombre de sites de glycation identifiés est donc multiplié par quatre par rapport au fractionnement seul et par 24 par rapport au à l'enrichissement seul (paragraphe IV.C.). Ces résultats montrent que la combinaison du fractionnement et de l'enrichissement apporte une réelle amélioration en termes d'identification par rapport aux stratégies utilisées précédemment.

La figure 126A représente les sites de glycation identifiés dans les différents réplicas. 65% des sites de glycation identifiés sont communs aux trois réplicas techniques. Le recouvrement des identifications est proche d'une analyse de protéomique classique (75% sur les peptides identifiés dans 3 réplicas d'injections). La figure 126B illustre le nombre de sites de glycations identifiés dans 3/3 réplicas et le nombre de protéines glyquées correspondantes

issus de 3 jeux de données : enrichissement seul (plasma natif avec dépôt de 100 µg, paragraphe IV.C) ; fractionnement seul et couplage du fractionnement et de l'enrichissement.

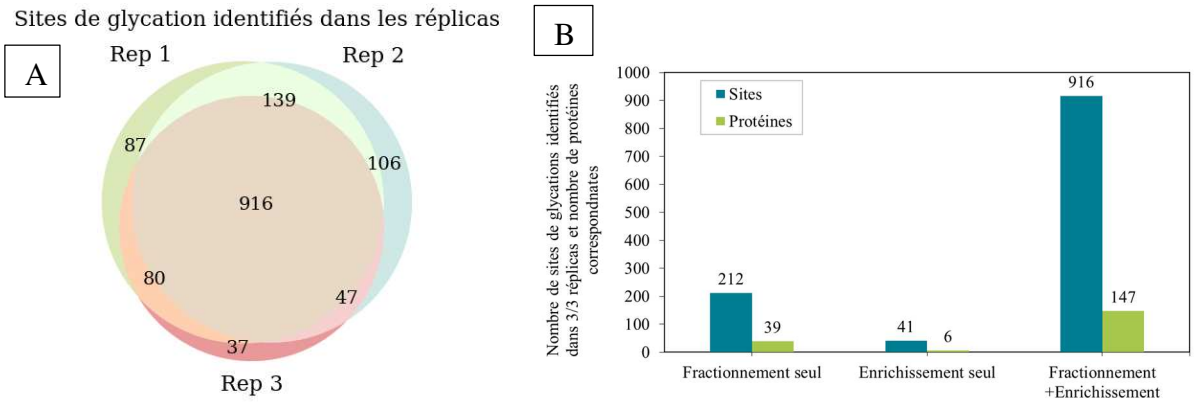

Figure 126 : Recouvrement des sites de glycations identifiés dans les différents réplicas (A) et nombre de sites de glycation identifiés dans 3/3 réplicas dans les échantillons uniquement fractionnés, uniquement enrichis et fractionnés + enrichis (B).

En appliquant un filtre de 3/3 valeurs observées dans les réplicas techniques, 916 sites ont pu être quantifiés sur 147 protéines différentes avec la combinaison du fractionnement et de l'enrichissement. C'est bien plus que ce qui avait été quantifié jusqu'à présent car les méthodes prises individuellement (enrichissement BAC et fractionnement hpH RP) manquaient de répétabilité au niveau de l'identification des peptides glyqués. Par ailleurs, pour la combinaison des deux méthodes, le CV médian sur les intensités brutes des peptides glyqués est de 24%. C'est une valeur parfaitement acceptable pour une analyse de peptides modifiés avant normalisation. Ainsi, bien que le fractionnement ou l'enrichissement BAC seuls ne permettent pas une quantification robuste à large échelle des sites de glycation, le couplage de ces deux méthodes est plus répétable et peut donc être envisagé pour l'étude des glycations plasmatiques d'échantillons biologiques.

Les résultats obtenus avec cette approche (fractionnement + enrichissement BAC) sont à comparer avec ceux obtenus par Zang *et al*. où 376 peptides glyqués répartis sur 64 protéines différentes avaient été identifiés dans du plasma glyqué *in vitro* par un système 2D LC (colonne BAC x colonne C18)[275]. Dans notre cas, nous utilisons un plasma natif et détectons 3 fois plus de sites de glycations, reflétant une profondeur d'analyse plus importante. Dans une seconde étude réalisée par Jeon *et al.*, sur 10 analyses (5 plasma d'individus sains et 5 de patients diabétiques), 982 peptides glyqués avaient été identifiés au total après enrichissement sur un système automatisé utilisant des billes fonctionnalisées au boronate retenues dans une cartouche par une membrane. Cependant, il existait une très forte variabilité entre les différents individus qui pouvait être liée soit à la méthode analytique soit à une forte disparité entre les échantillons. C'est pourquoi, il est plus rigoureux de comparer nos résultats avec les triplicats d'enrichissement effectués sur du sérum humain effectués dans cette même étude de Jeon *et al*. Dans ces triplicats, 50 peptides glyqués sont identifiés en moyenne et seuls 34 sites sont identifiés dans les 3 réplicas[123]. La combinaison d'un fractionnement hpH RP et d'un enrichissement boronate surpasse donc en termes d'identifications et de répétabilité, les

méthodes utilisées dans ces deux études tout en conservant un temps d'analyse raisonnable et une capacité à traiter de larges cohortes grâce à l'automatisation du protocole.

À ce stade, nous avons mis au point les méthodes de préparation d'échantillons, il s'agit maintenant de se focaliser sur l'analyse LC-MS/MS et sur le traitement bio-informatique des données. En effet, les glycations induisent des spectres de fragmentation particuliers. Ces profils particuliers rendent plus complexe l'identification des peptides glyqués par les logiciels de traitement. Un exemple de spectre de fragmentation d'un peptide glyqué est présenté en figure 127.

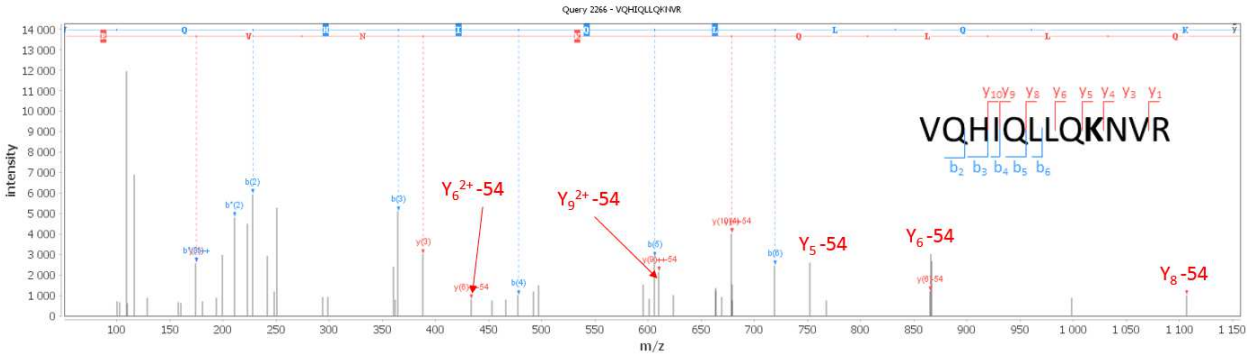

Figure 127 : Spectre de fragmentation effectué sur un Q-Orbitrap du peptide VQHIQLLQ**K**NVR, des pertes de neutres (3*H_2O) sont observées sur les fragments y portant la lysine glyquée.

Sur ce spectre, on observe les pertes de neutre de 3 molécules d'eau issues de la déshydratation des produits d'Amadori (m = 54,03 Da) décrites par Keilhauer *et al.*[272] et Zang *et al.*[423]. Dans ces études, d'autres pertes de neutres (pertes de 3 molécules d'eau et d'une molécule de formaldéhyde (m = 84,04 Da), perte du composé d'Amadori (m = 162,05 Da)), plus minoritaires sont décrites. Une comparaison poussée des performances des logiciels en termes d'identifications obtenues en prenant en compte ces différentes pertes de neutre doit être réalisée. De même, différents logiciels (MaxQuant[424], MS Amanda[148], Mascot (Matrix Science), MSFragger[153]) de traitement de données pourraient être comparés afin d'identifier le logiciel le plus adapté à l'analyse des glycations car les algorithmes utilisés ne sont pas les mêmes. Par ailleurs, des méthodes de fragmentations alternatives pourraient être considérées afin de réduire ces réactions de déshydratation provoquant les pertes de neutres comme c'est le cas pour la fragmentation ETD[281-282].

De plus, l'accent doit également être mis sur la validation des sites de glycation afin de pouvoir garantir la qualité des résultats obtenus. Un premier critère de qualité relativement simple à mettre en place est un filtre sur la probabilité de localisation de la glycation. Les glycations ne sont pas des modifications aussi labiles que les phosphorylations. De plus, elles se trouvent systématiquement sur un point de clivage si on utilise la trypsine comme protéase (du fait de leur localisation sur les lysines et les arginines), on pourrait donc envisager un filtre plus strict sur la probabilité de localisation. Pour les phosphorylations, seules les phosphorylations avec une probabilité de localisation supérieure à 0,75 sont classiquement retenues. Pour l'analyse des glycations on pourrait envisager une valeur seuil à 0,9. Pour l'étude des phosphorylations, seuls les peptides modifiés protéotypiques sont gardés. Ce serait également un contrôle qualité qui pourrait être appliqué facilement pour étudier les glycations.

La répétabilité de l'enrichissement pourrait également être évaluée, par exemple en calculant le nombre de peptides glyqués identifiés sur le nombre total de peptides identifiés après enrichissement sur le modèle de ce qui est fait pour les phosphorylations. On pourrait aussi imaginer introduire un glycomix sur le modèle du phosphomix. Il faudrait alors faire réaliser à façon des peptides synthétiques glyqués.

VII. Conclusions et perspectives

Au cours de ces développements, deux techniques de fractionnement ont été optimisées sur du plasma: le fractionnement CIF sur billes CMMB et le fractionnement hpH RP sur cartouches RPS. Les résultats obtenus par ces deux approches sont relativement similaires en termes de nombre d'identification. Les résultats préliminaires obtenus avec le fractionnement CIF sont très encourageants et il doit encore être adapté sur le Bravo AssayMAP afin d'être automatisé, contrairement au fractionnement hpH RP. Durant ma thèse, je me suis concentrée sur l'automatisation et la préparation des échantillons. À présent, les méthodes LC et MS doivent être optimisées sur le timsTOF Pro afin de bénéficier de sa sensibilité et de la dimension supplémentaire de séparation de mobilité ionique. Des analyses en mode DIA[425] devraient également permettre d'augmenter à la fois le nombre de d'identifications en identifiant des peptides minoritaires et la répétabilité des identifications des analyses. De plus, le développement de ces méthodes peut avoir d'autres applications : l'étude d'autres espèces minoritaires (autres PTMs, peptides pontés (*cross-link*)) ou pour des études très exhaustives tels que la constitution d'une librairie spectrale pour faire des études en DIA.

Un enrichissement BAC miniaturisé et automatisé a été entièrement développé avec des cartouches remplies à façon de gel de boronate (5 µL). Cet enrichissement des peptides glyqués s'est révélé effectif sur des mélanges peu complexes (BSA et trastuzumab). Cependant, il reste relativement peu efficace sur le plasma dû à une saturation rapide de la cartouche même en déposant de petites quantités de matériel (10 µg). Or si la quantité de matériel de départ est trop faible, il peut ne pas rester assez de matériel pour l'analyse LC-MS/MS après enrichissement. Une des pistes qui reste à explorer pour remédier à ce problème est l'utilisation de cartouches boronate de plus grande capacité (25 µL de phase). Ces cartouches, réalisées en collaboration avec Bene Lab (Zagreb, Croatie), sont en cours d'évaluation.

Une approche prometteuse consiste à fractionner les échantillons avant l'enrichissement pour les décompléxifier. Cette approche s'est révélée très fructueuse, permettant l'identification d'un grand nombre de peptides glyqués avec une bonne répétabilité entre les réplicas techniques. Des optimisations sont encore à réaliser, notamment au niveau de la fragmentation des peptides glyqués afin de générer des spectres de fragmentation de bonne qualité et donc plus informatifs. De même, au niveau de l'identification des peptides glyqués, des optimisations restent nécessaires : différents logiciels peuvent être comparés et différentes pertes de neutres peuvent être prises en compte[284, 423]. La qualité des résultats obtenus devrait être contrôlée par la mise en place de filtres stricts tels qu'un seuil de probabilité de localisation, un nombre maximum de valeurs manquantes acceptées ou même un seuil sur le CV des intensités.

Au cours de ma thèse je me suis focalisée sur l'étude des produits d'Amadori mais dans le contexte d'une étude biologique, les produits issus de la dégradation oxydative des fructosamines pourraient aussi apporter des informations intéressantes sur les interactions AGEs/RAGE ainsi que sur l'accumulation des AGEs dans l'organisme et ses conséquences[244, 261]. Les données générées dans ce chapitre pourront être traitées à nouveau en recherchant d'autres modifications : des produits de dégradation du produit d'Amadori comme les CML ou les CEL. Ces peptides glyqués ne sont pas retenus par la phase boronate mais devraient être détectés dans les *flow through* des analyses ou dans les échantillons avant enrichissement. Dans cette optique, des moteurs de recherche de dernière génération basés sur la prédiction des spectres et des temps de rétention pourraient être utilisés (ionbot[154] ou MSfragger[153]). Contrairement aux recherches classiques, ces logiciels permettent des recherches ouvertes (*open search*) à un très grand nombre de modifications, donc sans à priori, sans que le temps de recherche n'augmente de façon exponentielle. Les développeurs de MSfragger proposent un module dédié à l'étude des glycosylations qui serait également adapté à l'étude des glycations, ce qui reste à tester[283]. À l'avenir, ces logiciels devraient aussi pouvoir intégrer les données de mobilité ionique générées par les spectromètres de masse de type timsTOF.

CONCLUSION GENERALE

Mes travaux de thèse avaient pour objectifs, d'une part, d'accroitre les connaissances des mécanismes biologiques liés au phénotype hibernant grâce à une stratégie de protéomique quantitative « *label-free* » et, d'autre part, de développer des méthodes analytiques adaptées à l'étude globale et quantitative de deux types de PTMs : les phosphorylations et les glycations.

Une analyse protéomique quantitative *label-free* a été réalisée à partir d'échantillons enrichis en mitochondries musculaires d'ours bruns. Ceci a permis de montrer que l'équipement enzymatique mitochondrial reste stable durant l'hibernation par rapport à la période active. Ces données viennent préciser des conclusions préalables issues d'une analyse au niveau des fibres musculaires totales qui ne permettait que de suggérer une réduction du contenu mitochondrial musculaire lors de l'hibernation. Notre étude illustre donc l'intérêt de l'analyse de sous protéomes pour mieux comprendre des mécanismes physiologiques et moléculaires d'adaptation à des contraintes environnementales. On peut également noter que ces résultats ont été obtenus chez un organisme non modèle grâce au séquençage et à l'annotation de son génome pour la première fois seulement en 2018. Ceci souligne l'importance à poursuivre les efforts de développements bioinformatiques pour accélérer l'annotation des génomes.

Une analyse protéomique quantitative *label-free* a été réalisée à partir d'échantillons de cœur de lérots. Des optimisations de l'étape de préparation de ces tissus cardiaques ont été tout d'abord réalisées. Plusieurs protocoles de digestions ont ainsi été comparés, montrant que des résultats au moins équivalents en termes de nombre d'identification et de répétabilité pouvaient être obtenus en utilisant un protocole rapide et simple (S-trap) par rapport à un protocole classiquement utilisé pour l'analyse des tissus au laboratoire (gel Stacking). Une autre méthode de digestion, sur billes paramagnétiques fonctionnalisées, nécessitant un temps de préparation plus court, a été également évalué et s'est montré performante en termes de nombre de d'identification. Les résultats obtenus avec le protocole automatisé de SP3 montrent clairement l'intérêt d'automatiser l'étape de préparation d'échantillons pour gagner en répétabilité et robustesse L'application du protocole S-Trap à l'étude de l'influence de la diète et du statut physiologique des lérots sur leur protéome cardiaque a produit de nombreux résultats, encore en cours de discussion avec notre collaborateur. Les données suggèrent déjà une influence de la diète et du statut physiologique sur la contraction musculaire cardiaque et les flux calciques.

Une méthode semi-automatisée complète a été développée pour analyser les phosphorylations à large échelle à partir de tissu musculaire d'ours brun. Les différentes étapes de l'analyse protéomique ont été optimisées : la préparation d'échantillon, l'analyse LC-MS/MS et le traitement des données. L'étape automatisée de fractionnement hpH RP au niveau peptidique avant l'enrichissement IMAC ainsi que l'utilisation d'une fragmentation HCD et du logiciel MaxQuant ont permis une identification et quantification robuste d'un grand nombre de phosphopeptides. Une fois cette méthode développée, elle a été appliquée pour étudier les régulations ayant lieu dans le muscle durant l'hibernation de l'ours. Dans cet optique, des outils permettant l'interprétation biologique des résultats sur des espèces non modèles ont été mis en place grâce à une recherche des sites de phosphorylations humains homologues afin de

bénéficier des informations contenues dans les banques de données fonctionnelles. Ces données restent très complexes et doivent être analysées dans le détail pour faire ressortir les messages biologiques.

La dernière partie de ce manuscrit était consacrée à l'étude des glycations, une famille de PTMs hautement complexes et assez peu étudiés par analyse protéomique *label-free*, dans des fluides biologiques. Dans un premier temps, une méthode LC-MS a été développée pour l'analyse des glycations dans plasma et de l'hémolysat de diamants mandarins. Cette méthode a permis de déterminer des taux de glycation sur les protéines majoritaires. Nous avons aussi analysé la structure du complexe hémoglobine par la mise en place d'une méthode SEC-MS. Les techniques mises au point pour cette étude permettront à l'avenir l'analyse du plasma et de l'hémolysats de différentes espèces d'oiseaux mais également de chauve-souris. L'objectif à long terme étant d'étudier l'étendue du phénomène de glycation chez ces espèces volantes qui présentent de fortes concentrations sanguines en sucres sans en subir les conséquences habituellement observées chez le mammifère (*e.g.* diabète). Cependant, l'analyse de protéines entières ne permet pas d'avoir la vision globale d'un protéome modifié (seulement quelques protéines majoritaires glyquées analysées, pas de localisation des évènements de glycation sur la séquences des protéines). C'est pourquoi, dans un second temps, une méthode d'analyse protéomique quantitative *label-free* a été développée pour analyser les glycations dans le plasma en implémentant des protocoles automatisés d'enrichissement par BAC sur cartouches et de fractionnement peptidique. Avec les prototypes de cartouches testés à ce jour, l'enrichissement BAC automatisé s'est avéré peu efficace sur les échantillons plasmatiques, mais nous espérons parvenir à de meilleures performances avec des cartouches de plus grande capacité. Mais il pourrait être envisagé pour l'étude des glycations des protéines thérapeutiques comme une alternative miniaturisée et automatisée à l'enrichissement BAC des protéines intactes sur colonne HPLC couramment utilisé. L'ajout d'une étape de fractionnement peptidique hpH RP avant l'enrichissement BAC des peptides plasmatiques a permis d'augmenter la répétabilité de la méthode et permet à présent d'envisager des études globales à haut débit des glycations dans un fluide aussi complexe que le plasma.

Tout au long de ce manuscrit, des méthodes automatisées ont été mises au point pour l'analyse du protéome global ou de peptides modifiés. Au vu des avantages apportés par l'automatisation, notamment la répétabilité et le débit d'analyse[426], il parait évident que la préparation d'échantillon automatisée sera à l'avenir incontournable pour la préparation d'échantillon.

Ces travaux de thèse ont également mis en évidence qu'aucun protocole de préparation d'échantillon n'est universel et qu'il est important d'adapter les différentes étapes de la préparation d'échantillon au type d'échantillon, à la quantité disponible et à l'analyse souhaitée (globale ou plutôt ciblée sur une famille de protéines ou de modifications).

Par ailleurs, des développements doivent encore être réalisés au niveau de l'analyse LC-MS des peptides modifiés. À ce jour, des optimisations au niveau du gradient, de la fragmentation, de la gamme de mobilité ionique et des temps de cycle, sont en cours pour tirer

entièrement parti de la sensibilité accrue et de la dimension de séparation supplémentaire du timsTOF Pro. Des premières optimisations ont montré l'impact positif des adaptations des énergies de collision en fonction de la mobilité ionique, sur le nombre de phosphopeptides identifiés mais doivent être poursuivis pour l'étude d'autres PTMs comme les glycations.

Dans l'optique d'améliorer la répétabilité, la reproductibilité et la robustesse des analyses, le mode d'acquisition DIA va certainement se démocratiser pour toutes sortes d'applications et notamment la protéomique quantitative *label-free*. En effet, ce mode d'acquisition permet théoriquement d'identifier et quantifier tous les peptides d'un échantillon et non plus seulement les plus intenses. Grâce à des avancées, notamment au niveau de l'instrumentation et du traitement des données, la DIA devient plus accessible avec la possibilité de travailler sans librairie spectrale dont la génération était fastidieuse et consommatrice en matériel. Cependant, comme pour l'analyse DDA, il est primordial de rester vigilant quant à la qualité des données et c'est pourquoi, des critères de validation doivent être mis en place au sein de la communauté scientifique pour garantir la robustesse des résultats. La DIA pourra théoriquement être appliquée à tout système biologique comme ce peut être le cas pour la DDA, mais elle peut également être tout à fait adaptée à l'étude des PTMs[213].

A l'heure actuelle, l'une des faiblesses de l'analyse protéomique reste le traitement bio-informatique permettant une interprétation quantitative fiable et robuste des données de spectrométrie de masse, aussi bien pour les acquisitions en DDA qu'en DIA, et ce d'autant plus pour les PTMs. L'implémentation de l'intelligence artificielle pourrait améliorer cette étape critique en permettant d'accéder aux informations présentes dans les nombreux spectres qui n'ont pas conduit à des identifications. Elle pourrait prédire les spectres de fragmentations, et ce en prenant en compte les fragmentations spécifiques aux différentes modifications, ce qui permettrait de différencier plus facilement le bruit de fond des signaux d'intérêt. Par ailleurs, suivant les avancées technologiques sur les spectromètres de masse, la plupart des logiciels devrait pleinement intégrer les données de mobilité ionique pour l'identification des peptides.

De plus en plus de logiciels peuvent réaliser des recherches « *open search* », ce qui va permettre, en plus des avancées décrites précédemment, l'analyse des PTMs de façon plus globale, sans forcément se concentrer sur un type de modifications. Avec ces approches, il sera possible d'étudier les cross-talk entre PTMs et d'obtenir une vision encore plus globale de la régulation des cellules.

Enfin, l'un des plus grands défis concernant l'analyse des PTMs reste la validation des résultats obtenus pour garantir la fiabilité des résultats. Des approches robustes et non biaisées de calcul de FLR, similaire au FDR, doivent être mises en place afin d'assurer la localisation des sites modifiés.

PARTIE EXPERIMENTALE

Sauf indication contraire, tous les produits ont été achetés chez Sigma Aldrich (Saint-Louis, MO, USA). Les solvants (acétonitrile, méthanol, éthanol, chloroforme) et les acides (acide formique [FA], acide trifluoroacétique [TFA]) proviennent de Fisher Scientific (Hampton, NH, USA)

I. Analyse du protéome des mitochondries musculaires de l'ours brun

1. Échantillonnage

Dans le cadre de notre implication dans le consortium international « *Scandinavian Brown Bear Research Project* » depuis plus de 10 ans, environ 100 mg de muscle (*vastus lateralis*) ont été prélevés sur 4 ours bruns (*Ursus arctos*, 2 femelles de 2 et une de 3 ans, 1 mâle de 2 ans) à deux périodes de l'année (février et juin) en Suède dans le Comté de Dalarna. Cette étude a été approuvée par le Comité d'éthique Suédois sur l'Expérimentation Animale (applications Dnr C3/2016 and Dnr C18/2015), l'Agence Suédoise de Protection de l'Environnement (NV-0741-18), et l'Office Suédois de l'Agriculture Agriculture (Dnr 5.2.18–3060/17). Toutes les procédures étaient conformes aux lois et à la réglementation suédoises. Les captures, anesthésies et échantillonnages ont été réalisées selon un protocole bien établi[427].
Immédiatement après la collecte, le muscle a été découpé aux ciseaux et incubé 30 min à 37°C dans 1 mL de tampon (mannitol 225 mM, sucrose 75 mM, Tris 10 mM, EDTA 0,1 mM, pH 7.4) contenant de la trypsine (0,5 mg/g de tissu). Après 3 lavages dans le même tampon (3 mL) sans trypsine, les tissus ont été broyés (tissue ruptor II, Qiagen) et centrifugés une première fois pour retirer les débris cellulaires (700 g, 10 min, 4°C). Le surnageant (fraction cytosolique) a alors été centrifugé (7000 g, 10 min, 4 °C) pour culoter les mitochondries. Cette étape a été répétée, puis le culot de mitochondrie a été repris dans 100 µL du même tampon sans trypsine. Les échantillons ont ensuite été congelés à -80 °C jusqu'à analyse.

2. Dosage, lyse, dosage et précipitation

Les protéines ont été dosées à l'aide du kit RC-DC selon les instructions du fournisseur (Bio-Rad, Hercules, CA, USA) en ajoutant une solution de SDS 10% dans les échantillons lors du dosage pour obtenir une concentration finale en SDS de 0,6%. Un tampon de lyse (urée 8 M, thio-urée 2 M, DTT 1%, inhibiteurs de protéase 1/50, 50 mM Tris pH 7.4) a été ajouté à 30µg de protéines mitochondriales. Après incubation pendant 45 min à température ambiante [RT], huit volumes d'acétone froide ont été ajoutés aux échantillons qui ont été conservés à -20°C pendant 2 heures. Les protéines précipitées ont été centrifugées (20 minutes, 4°C, 13500 x g), lavées une fois avec de l'acétone puis reprises dans un tampon Laemmli (10 mM Tris pH 6.8, 1 mM EDTA, 5% β- mercaptoéthanol, 5% SDS, 10% glycérol). La concentration en protéines totales a été déterminée à l'aide du kit RC-DC. À ce stade, un échantillon de référence comprenant des quantités égales de tous les extraits de protéines a été constitué, afin d'être analysé régulièrement pendant toute l'expérience et de permettre des mesures liées au contrôle qualité.

3. Gel SDS-PAGE « Stacking »

6 µg de chaque échantillon ont été chargés sur des gels SDS-PAGE (polyacrylamide à 4 %) et l'électrophorèse a été réalisée pendant 20 minutes à 35 V afin d'obtenir une bande dans le gel de concentration. Après fixation des protéines (EtOH à 50 %, acide phosphorique à 3% (VWR

International, Radnor, AR, USA)) pendant 15 min, la coloration au bleu de Coomassie colloïdal (30 min) a permis la visualisation d'une bande protéique. Elle a été excisée du gel, ainsi que la partie du gel située au-dessus de cette bande. Après décoloration à l'aide d'ACN/hydrogénocarbonate d'ammonium [ABC] 25 mM (75/25) et déshydratation à l'aide d'ACN pur, les protéines ont été réduites à l'aide de DTT 10 mM dans un tampon ABC 25 mM (30 min à 60°C puis 30 min à RT) et alkylées dans l'IAM 55 mM dans un tampon ABC 25 mM (20 min à RT dans l'obscurité). Ensuite, après plusieurs lavages du gel à l'aide du tampon ABC (5 min à RT) et d'ACN (5 min à RT), ainsi qu'une déshydratation à l'aide d'ACN (2*5 min à RT), la digestion des protéines dans le gel a été réalisée pendant une nuit à 37 °C à l'aide de la trypsine (Mass Spec Grade, Promega, Madison, WI, USA ; 60 ng par bande). Après digestion par la trypsine, les peptides résultants ont été extraits deux fois à 450 rpm sur un agitateur orbital pendant 45 min en utilisant 60 % ACN/ 0,5 % FA dans l'eau, puis pendant 15 min en utilisant 100 % d'ACN.

4. Analyse nanoLC-MS/MS

Avant analyse, un ensemble de peptides de référence (kit iRT ; Biognosys AG, Schlieren, Suisse) a été ajouté aux extraits peptidiques (3 µL/échantillon après resuspension dans 500 mL 20% ACN/1% FA) pour les mesures liées au contrôle qualité. Le volume des extraits peptidiques a d'abord été réduit à l'aide d'une centrifugeuse sous vide (SpeedVac, Savant, Thermo Fisher Scientific, San Jose, CA, USA) puis ajusté à 20 µL en utilisant une solution 1% ACN/ 0,1% FA. Les échantillons ont été analysés sur un système nanoUPLC (nanoAcquity, Waters, Milford, MA, USA) couplé à un spectromètre de masse hybride quadripole-Orbitrap (Q-Exactive Plus, Thermo Fisher Scientific). Le système était entièrement contrôlé par le logiciel XCalibur (v3.0.63 ; Thermo Fisher Scientific).

- Conditions chromatographiques

L'étape de concentration/dessalage a d'abord été effectuée par chargement de 2 µL (300 ng) par échantillon sur une précolonne Symmetry (C18, 180 µm × 20 mm, 5 µm ; Waters) en utilisant de l'acide formique à 0,1 % dans l'eau (solvant A, 99 %) et de l'acide formique à 0,1 % dans l'acétonitrile (solvant B, 1 %) à un débit de 5 µl/min pendant 3 min. L'élution des peptides a ensuite été réalisée à un débit de 400 nL/min en utilisant une colonne NanoEasy m/z (C18, 250 mm × 75 µm, 1,7 µm, 130 Å ; Waters) maintenue à 60 °C avec le gradient détaillé dans le tableau 16. Un blanc de solvant a été réalisé entre chaque échantillon.

Tableau 15 : Gradient chromatographique utilisé pour l'analyse des échantillons enrichis en mitochondries.

Temps (min)	Pourcentage en solvant B (%)
0	1
2.5	8
78	40
79	90
85	90
86	1
110	1

- Paramètres MS

Le Q-Exactive Plus fonctionnait en mode ESI positif avec une température de source réglée à 250 °C et une tension de spray de 2,0 kV. Les spectres MS Full Scan (300-1800 m/z) ont été acquis avec une résolution de 70 000 à 200 m/z, un temps d'injection maximal de 50 ms et une *AGC target* de 3×10^6 charges. L'option de verrouillage de la masse (*lock-mass*) était activée (445,12002 m/z). Pour chaque spectre MS Full scan, les 10 précurseurs les plus intenses ont été isolés avec une largeur de fenêtre de 2 m/z et fragmentés en utilisant la dissociation par collision à haute énergie (HCD, énergie de collision normalisée à 27%), et l'exclusion dynamique des précurseurs déjà fragmentés a été fixée à 60 s. Les spectres MS/MS ont été acquis avec une résolution de 17 500 à 200 m/z, un temps d'injection maximal de 100 ms et une AGC *target* de 1×10^5.

5. Identification et quantification des protéines à partir des spectres obtenus

Les données brutes ont été traitées avec MaxQuant (version 1.6.11.0)[424]. Des listes de pics ont été créées à l'aide de paramètres par défaut et recherchées à l'aide du moteur de recherche Andromeda implémenté dans MaxQuant par rapport à une base de données de protéines créée à l'aide du logiciel MSDA[428]. La base de données contenait des séquences de protéines RefSeq (33 139 entrées) dérivées du dernier assemblage complet et annoté (Annotation release ID : 100) du génome d'*Ursus arctos* (ASM358476v1), qui ont été téléchargées en février-mars 2020. Dans le moteur de recherche Andromeda, les séquences de contaminants comme les kératines et la trypsine (247 entrées) ont été ajoutées à la base de données. Pour la première recherche, la tolérance de masse des précurseurs a été fixée à 20 partie par million [ppm] et à 4,5 ppm pour la recherche principale après recalibrage. La tolérance de masse des ions fragments a été fixée à 20 ppm. La carbamidométhylation des cystéines a été considérée comme fixe et l'oxydation des méthionines et l'acétylation des extrémités N-terminales des protéines comme des modifications variables pendant la recherche. Un nombre maximum d'un clivage manqué a été toléré, et un FDR de 1% pour PSM (longueur minimale de sept acides aminés) et les protéines a été accepté pendant l'identification. À des fins quantitatives, la normalisation des données et l'estimation de l'abondance des protéines ont été effectuées à l'aide de l'option MaxLFQ (quantification sans marquage) mise en œuvre dans MaxQuant[324] en utilisant un " *minimal ratio count* " de 1. L'option "*Match between runs*" a été activée en utilisant une fenêtre temporelle de 0,7 min après l'alignement des temps de rétention. Seuls les peptides spécifiques (modifiés inclus) ont été pris en compte pour la quantification. Tous les autres paramètres de MaxQuant

ont été définis par défaut. Les protéines retenues pour la suite de l'analyse étaient celles pour lesquelles des valeurs quantitatives étaient obtenues à partir d'un minimum de deux peptides uniques, et ne présentant pas plus d'une valeur absente par groupe. Les protéines absentes dans un groupe donné (c'est-à-dire pas du tout détectées) ont également été retenues.

6. Traitement statistique des données quantitatives

En utilisant le logiciel R v3.4.0 (34), les tests de Shapiro-Wilk et de Bartlett ont été utilisés pour vérifier la distribution normale et l'homoscédasticité des données ($p > 0,01$). Un T-test a été effectué pour ici pour comparer les deux moyennes issues des deux conditions (été et hiver). Un T-test apparié a également été effectué. Pour ces deux tests, le seuil de significativité a été fixé à $p < 0,05$.

II. Développement d'un protocole pour étudier les phosphorylations dans le muscle de l'ours brun

A. Développement d'une méthode analytique pour l'analyse des phosphorylations dans le muscle

1. Lyse, Dosage et précipitation

Trois tampons de lyse ont été comparés :

- « Urée »: urée 8M, ABC 0,1 M, inhibiteur de protéase 1/50 (v/v) et sodium ortho vanadate 1 mM
- « Urée/Thio-urée »: urée 6 M, thio-urée 2 M, ABC 0,1 M, inhibiteur de protéase 1/50 et sodium ortho vanadate 1 mM
- « Laemmli-like »: Tris-HCl 50 mM pH= 6.8, glycérol 10%, SDS 2,5%, EDTA 1 mM, inhibiteur de protéase 1/50 et sodium ortho vanadate 1 mM.

Les tampons « Urée » et « Urée/Thio-urée » ont été utilisés avec ou sans précipitation au méthanol/chloroforme. Au total, 5 conditions ont donc comparé, comme récapitulé en figure 128.

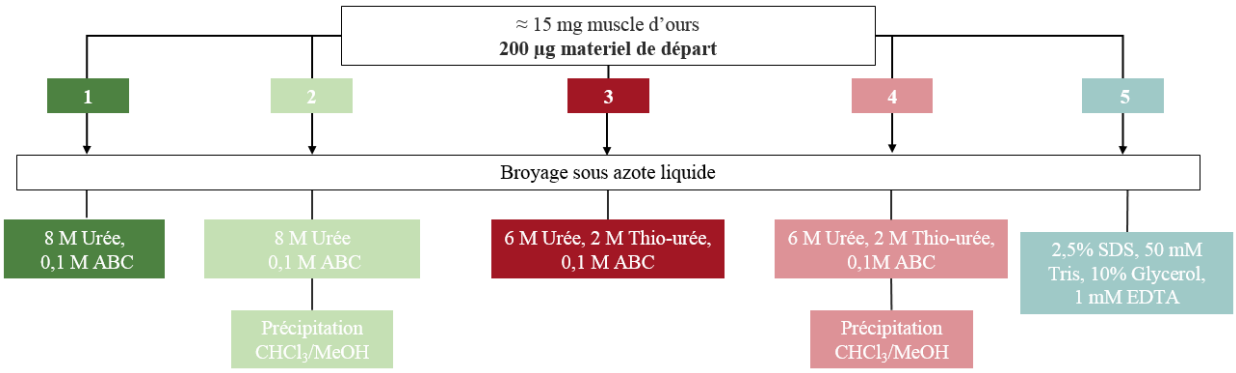

Figure 128 : Cinq conditions d'extractions sont comparées : « Urée » sans (1) et avec (2) précipitation méthanol/chloroforme, « Urée/Thio-urée » sans (3) et avec (4) précipitation méthanol/chloroforme et « Laemmli-like » (5).

Entre 13 et 28 mg de muscle congelé d'ours (N=3/condition) ont été broyés (3*30s à 30Hz) sous azote liquide à l'aide d'un broyeur à billes (MM400, Retsch, Eragny sur Oise, France) puis un volume du tampon de lyse correspondant calculé au prorata 1/10 (masse (mg) / volume (µL)) de la masse de l'échantillon a été ajouté. Les protéines des conditions 2 et 4 ont été isolées en ajoutant 4 volumes de méthanol froid, 1 volume de chloroforme puis 3 volume d'eau. Après centrifugation (30 min, 4 °C, 5000 g), les protéines étaient situées à l'interface entre les deux phases. La phase supérieure a été retirée et 3 volumes de méthanol froid ont été ajoutés. Les protéines ont été culotées par centrifugation (10 min, 4 °C, 5000 g) puis lavées avec 1 mL de méthanol froid avant d'être centrifugées à nouveau (5 min, 4 °C, 5000 g). Les protéines ont ensuite été reprises dans le volume initial avec le tampon initial correspondant. La concentration en protéines de tous les échantillons a ensuite été évaluée par dosage RC-DC (Bio-Rad, Hercules, CA, USA).

2. Réduction alkylation

200 µg de protéines pour chaque réplica ont été réduits pendant 30 min à 37°C avec une solution de DTT 20 mM, ABC 0,1 M pour atteindre une concentration finale en DTT de 12 mM puis alkylées pendant 1 h à RT à l'obscurité avec une solution d'IAM 700 mM, ABC 0,1 M pour atteindre une concentration finale en IAM de 40 mM. Pour les conditions 1, 2, 3 et 4, les échantillons ont été dilués avec un tampon ABC 0,1 M pour diminuer la concentration d'urée en dessous de 1M.

3. Digestion liquide et SPE (conditions 1-4, tampons « Urée » et « Urée/Thio-urée)

Pour les conditions d'extraction 1 à 4, les protéines ont été digérées sur la nuit à 37 °C par ajout d'un mélange trypsine/Lys-C (Mass Spec Grade, Promega, Madison, WI, USA) avec un ratio de 1/50 (enzyme/protéine, m/m). La digestion a été stoppée par ajout de FA pour obtenir un pH inférieur à 3.

Les peptides obtenus ont été purifiés par SPE avec un protocole automatisé. Le protocole « peptide clean-up » v2.0 a été utilisé sur une station Bravo AssayMAP (Agilent Technologies, Santa Clara, CA, USA). En bref, les cartouches remplies de 5 µL de phase C18 ont été lavées et conditionnées avec 50% ACN, 0,1% TFA, puis équilibrées avec une solution acide 0,1% TFA. Les échantillons ont été chargés à 5 µL/min sur la phase solide puis lavés avec une solution acide 0,1% TFA avant que les peptides aient été élués dans 40 µL de 70% ACN, 0.1% FA à 5 µL/min.

4. Digestion SP3 (condition 5, tampon "Laemmli-like")

Le tampon de la condition 5 contenant un détergent non compatible avec la digestion en solution, un protocole de digestion sur billes (SP3) a été utilisé. Les billes A (billes Sera-Mag Speed, Thermo Fisher Scientific, 45152105050250) et les billes B (billes Sera-Mag Speed, Thermo Fisher Scientific, 65152105050250) ont été combinées (ratio 1:1) et, après 3 étapes de lavage à l'eau ultra-pure, une solution à 20 µg/µL a été préparée. Ce mélange de billes a ensuite été ajouté aux échantillons (ratio billes:protéines de 10:1 pour chaque type de billes, soit un ratio 20:1 pour la combinaison de billes). Après fixation des protéines sur les billes avec 50% d'ACN pendant 18 minutes à RT, les tubes sont incubés 2 min sur le rack magnétique et le surnageant est retiré. Suivant ce principe, les protéines ont été lavées deux fois avec 80% d'EtOH (200 µL) et une fois avec 100% d'ACN (180 µL) en incubant à RT pendant 5 min et 2

min sur le rack magnétique. Les protéines ont ensuite été remises en suspension dans 45 µl de solution ABC 0,1 M avant d'être digérées pendant 3 h sur les billes (rapport enzyme:protéine de 1:20) à 1 000 tr/min à 37 °C en utilisant une combinaison de trypsine/Lys-C Mass Spec Grade, Promega, Madison, WI, USA). La digestion a été arrêtée à l'aide de TFA (pH final < 2). Après sonication pendant 1 min et incubation sur rack magnétique pendant 2 min, le surnageant contenant les peptides a été récupéré.

5. Analyse nanoLC-MS/MS des échantillons avant enrichissement

Dans chaque réplica, 2µL d'échantillon ont été prélevés et repris dans une solution de 2% ACN et 0,1% FA pour analyse sur le couplage NanoAcquity Q-Exactive Plus.

- Conditions chromatographiques

Les conditions utilisées étaient identiques à celles du paragraphe I.4 à l'exception du gradient, qui est présenté sur la tableau 16.

Tableau 16 : Gradient chromatographique utilisé pour l'analyse des échantillons avant enrichissement IMAC.

Temps (min)	Pourcentage en solvant B (%)
0	1
2	8
79	35
80	90
85	90
87	1
105	1

- Paramètres MS

Les paramètres utilisés étaient identiques à ceux du paragraphe I.4 sauf la tension du spray qui était de 2.1 kV.

6. Identification et quantification des protéines à partir des spectres obtenus

Les listes de pics (fichiers mgf) ont été créées à partir des données brutes par ProteoWizard MS Convert (v 3.0.11417). La recherche a été effectuée à l'aide de Mascot (v 2.6.2) en interrogeant une base de données de protéines créée à l'aide du logiciel MSDA[428]. La base de données est identifique à celle utilisée au paragraphe I.5 à laquelle a été ajoutée par MSDA les séquences de contaminants comme les kératines et la trypsine (118 entrées) et les séquences leurres obtenues par inversion des séquences protéiques. La tolérance de masse des précurseurs a été fixée à 5 ppm. La tolérance de masse des ions fragments a été fixée à 0,07 Da. La carbamidométhylation des cystéines a été considérée comme fixe et l'oxydation des méthionines et l'acétylation des extrémités N-terminales des protéines comme des modifications variables pendant la recherche. Un nombre maximum d'un clivage manqué a été toléré. Les données obtenues ont été validées avec la suite logicielle Proline[157] (v 1.0) en appliquant un FDR de 1% sur les PSM et les protéines ainsi qu'un filtre sur le « *pretty rank* » des protéines (supérieur à 1), sur le score des protéines (supérieur à 20) et sur la longueur minimale des peptides (7 acides aminés). La normalisation des données et l'estimation de

l'abondance des protéines ont aussi été réalisée par Proline (v 1.0). L'option « cross alignement » était désactivée, tous les autres paramètres étaient définis par défaut.

7. Enrichissement IMAC des conditions 3 et 5

Les extraits peptidiques issus des conditions 3 et 5 ont été évaporés à sec à l'aide d'une centrifugeuse sous vide (SpeedVac, Savant, Thermo Fisher Scientific) puis repris dans 170 µL 80% ACN, 0,1% FA. Un mélange de phosphoppeitdes synthétiques (solution de phosphomix I light, Thermo Fisher Scientific) (rapport peptide (µg)/mix(fmol) = 1,6) a été ajouté dans chaque échantillons. Les échantillons ont ensuite été enrichis en phosphopeptides en utilisant des cartouches de 5 µL remplies de phase Fe(III)-NTA sur une plateforme Bravo AssayMAP suivant le protocole « Phosphopeptide Enrichment » (v2.0). En bref, les cartouches ont été lavées et conditionnées avec 50% ACN 0,1% TFA et conditionnées avec 80% ACN, 0,1% TFA puis 160 µL d'échantillons ont été chargés à 2 µL/min sur la phase, lavés avec 80% ACN, 0,1% TFA avant d'être élués dans 20 µL de solution d'ammoniaque (NH_4OH) 1% à 5 µL/min. Après l'enrichissement, du FA a été ajouté à chaque échantillon pour acidifier le milieu ainsi que du mélange de phosphomix I marqués isotopiquement (Thermo Fisher Scientific) (rapport peptide (µg)/mix(fmol) = 1,6).

8. Analyse LC-MS/MS des échantillons après enrichissement IMAC

Les échantillons enrichis en phosphopeptides ont été évaporés à sec à l'aide d'une centrifugation sous vide (SpeedVac, Savant, Thermo Fisher Scientific) et remis en suspension dans 40 µL d'une solution d'ACN 2%, 0,1% FA, puis analysés sur un système nanoUPLC (nanoAcquity (Waters)) couplé à un Q-Exactive HF-X (Thermo Fisher Scientific). Le système était contrôlé par le logiciel XCalibur (v4.0.27.19, Thermo Fisher Scientific). Ces mêmes échantillons ont également été analysés sur un système nanoUPLC (nanoElute, Bruker Daltoniks GmbH, Brême, Allemagne) couplé à un spectromètre de masse hybride à mobilité ionique - quadripôle et TOF (timsTOF Pro, Bruker Daltonics) équipé d'une source d'ion CaptiveSpray (Bruker Daltonics). Le système était contrôlé par OtofControl (v6.0, Bruker Daltonics).

a. Q-Exactive HF-X

- Conditions chromatographiques

La concentration/dessalage a d'abord été effectuée par chargement de 8 µL par échantillon sur une précolonne Symmetry (C18, 180 µm × 20 mm, 5 µm ; Waters) en utilisant 0,1 % FA dans l'eau (solvant A, 99 %) et du FA à 0,1 % dans l'ACN (solvant B, 1 %) à un débit de 5 µl/min pendant 3 min. L'élution des peptides a ensuite été réalisée à un débit de 400 nL/min en utilisant une colonne NanoEasy m/z (C18, 250 mm × 75 µm, 1,7 µm ; 130 Å, Waters) maintenue à 60 °C avec le gradient appliqué détaillé en tableau 17.

Tableau 17 : Gradient chromatographique utilisé pour l'analyse des échantillons enrichis en phosphopeptides.

Temps (min)	Pourcentage en solvant B (%)
0	1
2	2
79	35
80	90
85	90
87	1
105	1

- Paramètres MS

Le Q-Exactive HF-X fonctionnait en mode ion positif avec une température de source réglée à 250 °C et une tension de spray de 2,0 kV. Les spectres MS Full Scan (375-1500 m/z) ont été acquis avec une résolution de 120 000 à 200 m/z, un temps d'injection maximal de 60 ms et une *AGC target* de 3×10^6 charges. Pour chaque spectre MS Full scan, les 20 précurseurs les plus intenses ont été isolés en utilisant une fenêtre de 2 m/z et fragmentés en utilisant la dissociation par collision à haute énergie (HCD, énergie de collision normalisée à 27%), et l'exclusion dynamique des précurseurs déjà fragmentés a été fixée à 40 s. Les spectres MS/MS (200-2000 m/z) ont été acquis avec une résolution de 15 000 à 200 m/z, un temps d'injection maximal de 60 ms et une *AGC target* de 1×10^5.

- Identification et quantification des phosphopeptides à partir des spectres obtenus

Les données brutes de spectrométrie de masse ont été traitées à l'aide de MaxQuant v1.6.14.0, en utilisant les paramètres du paragraphe I.5 en ajoutant la recherche de phosphorylation comme modification variable sur les sérines, thréonines et tyrosines. Le nombre de clivages manqués maximal autorisés était fixé à 2. L'option « *Match between runs* », ainsi que la normalisation des données d'intensité par MaxLFQ, était désactivée. Le fichier « Phospho (STY)Sites » a été utilisé. Les identifications et quantifications des sites de phosphorylation ont été obtenues par l'utilisation de la fonction « *expand site tables* » de Perseus (v1.6.14.0).

b. TimsTOf Pro

- Conditions chromatographiques

La chromatographie liquide a été réalisée avec le solvant A (eau avec 0,1% de FA) et le solvant B (acétonitrile avec 0,1% de FA). Les peptides (5 µL) ont d'abord été chargés sur une précolonne Acclaim™ PepMap™ 100 C18 (100 Å, 5 µm, 100 µm × 20 mm ; Thermo Fisher Scientific) à 2 % de B à 120 bars. Ils ont ensuite été séparés sur une colonne Aurora C18 (120 Å, 1. 6 µm, 75 µm × 250 mm ; IonOpticks, Melbourne, Australie) maintenue à 50 °C avec un gradient allant de 2% à 50% de solvant B en 60 min à un débit de 300 nL/min. La colonne a ensuite été lavée avec 95% de B pendant 10 min et rééquilibrée avec 2 % de solvant B pendant 10 minutes.

- Paramètres MS

Le spectromètre de masse timsTOF Pro, équipé d'une cellule de mobilité ionique TIMS et d'une source captive Spray (Bruker Daltoniks GmbH), a été utilisé en mode DDA-PASEF. La température de la source a été réglée à 180 °C et la tension du capillaire de transfert à 1,6 kV. Les ions ont été désolvatés en utilisant de l'azote à un débit de 3 L/min. Dans le TIMS, les temps d'accumulation et d'élution ont tous deux été réglés à 166 ms, ce qui permet de réaliser des cycles d'utilisation proches de 100 %. La plage de balayage de la cellule de mobilité ionique a été fixée à de 0,7 à 1,25 V/s/cm^2. Le temps du cycle d'acquisition était de 1,88 sec et comprenait 1 MS et 10 balayages MS/MS PASEF. Pour la sélection des précurseurs, les ions les plus chargés (jusqu'à 5 charges) ont été privilégiés, les ions chargé 1+ étant exclus après évaluation de leurs valeurs de m/z et de mobilité ionique (filtre polygonal). Pour la sélection des ions précurseurs, une "valeur minimale de 1000 (unités arbitraires) a été fixée, et les précurseurs atteignant la valeur cible de 17000 (unités arbitraires) ont été exclus dynamiquement pendant 0,4 min. Les spectres MS et MS/MS (100-1700 m/z) ont été acquis avec une résolution de 40000 à 622 m/z. L'isolement des précurseurs au niveau du quadripôle a été effectué en utilisant une fenêtre de 2-3 m/z selon le m/z de l'ion, et les ions ont été fragmentés en utilisant la dissociation induite par collision avec l'azote. L'énergie de collision était comprise entre 20 eV à 52 eV en fonction de la valeur de mobilité des ions.

- Identification et quantification des phosphopeptides à partir des spectres obtenus

Les données brutes de spectrométrie de masse ont été traitées à l'aide de MaxQuant v1.6.14.0, en utilisant la banque de données protéiques décrite dans le paragraphe I.5.

Les tolérances de masse pour les ions précurseurs ont été fixées à 20 ppm pour la première recherche et à 10 ppm pour la recherche principale. La tolérance pour les ions fragments a été fixée à 40 ppm. La trypsine\P a été sélectionnée comme enzyme et un maximum de deux clivages manqués a été accepté. La carbamidométhylation des résidus cystéine a été considérée comme une modification fixe. L'oxydation des résidus méthionine, l'acétylation des amines N-terminales des protéines et la phosphorylation des sérines, thréonines et tyrosines ont été paramétrées comme modifications variables. La longueur minimale des peptides était de sept acides aminés. Le FDR a été contrôlé à moins de 1% au niveau des PSM et des protéines. L'option "*Match between runs*" a été désactivée. Tous les autres paramètres de MaxQuant ont été définis par défaut. Le fichier « Phospho (STY)Sites» a été utilisé.

9. Évaluation de différents traitements de données pour l'identification, la localisation et la quantification des sites de phosphorylation

Les données issues du Q-Exactive HF-X ont été traitées par différents logiciels pour identifier et localiser les sites de phosphorylation.

a. Mascot-Proline

Les listes de pics (fichiers mgf) ont été créées à partir des données brutes par ProteoWizard MS Convert (v 3.0.11417). La recherche a été effectuée par Mascot (v 2.6.2) avec la banque de données protéiques utilisée dans le paragraphe II.A.6. La tolérance de masse des précurseurs a été fixée à 5 ppm. La tolérance de masse des ions fragments a été fixée à 0,07 Da. La carbamidométhylation des cystéines a été considérée comme fixe. La phosphorylation des

sérines, thréonines et tyrosines ainsi que l'oxydation des méthionines et l'acétylation des extrémités N-terminales des protéines ont été considérées comme des modifications variables pendant la recherche. Un nombre maximum de deux clivages manqués a été toléré. Les données obtenues ont été validées avec Proline[157] (v 1.0) en appliquant un taux de faux positifs (FDR) de 1% aux niveaux PSM et protéiques, ainsi qu'un un filtre sur le « *pretty rank* » (supérieur à 1), sur le score (supérieur à 20) et sur la longueur minimale (7 acides aminés).

b. Mascot-Proteome discoverer

Une recherche Mascot a été effectuée à partir des données brutes en utilisant Proteome Discoverer (v2.5, Thermo Fisher Scientific). La banque de données protéiques et les paramètres de recherche utilisés étaient les mêmes que dans le paragraphe précédent (II.A.9.a). La probabilité de localisation des sites de phosphorylation a été évaluée par phosphoRS[221]. Les données obtenues ont été validées en appliquant un FDR de 1% pour les PSM, les peptides et les protéines avec Percolator[158] et seuls les sites modifiés avec une probabilité de localisation supérieure à 25% ont été conservés.

c. MS Amanda/Mascot-Proteome Discoverer

Protéome Discoverer offre la possibilité d'effectuer deux recherches en parallèle. Ainsi une recherche Mascot avec les mêmes paramètres de recherche que dans le paragraphe précédent (II.A.9.a) a été effectuée en ajoutant une recherche avec MS Amanda 2.0[148]. La même banque de données protéiques et les mêmes paramètres ont été utilisés que pour la recherche Mascot. La probabilité de localisation des sites de phosphorylation issus des deux recherches a été évaluée par phosphoRS[221]. Les données obtenues ont également été validées en appliquant un FDR de 1% pour les PSM, les peptides et les protéines avec Percolator[158] et seuls les sites modifiés avec une probabilité de localisation supérieure à 25% sont conservés.

10. Comparaison de la fragmentation HCD et ETD sur Eclipse Tribrid

Les échantillons préparés précédemment (condition 3 après enrichissement IMAC) ont été utilisés pour comparer la fragmentation HCD à la fragmentation ETD sur un instrument de type Tribrid Eclipse (Thermo Fisher Scientific).

- Conditions chromatographiques

La concentration/dessalage a d'abord été effectuée par chargement de 6 µL par échantillon sur une précolonne Acclaim™ PepMap™ (C18, 100 µm × 20 mm, 5 µm ; Thermo Fisher Scientific) en utilisant de l'acide formique à 0,1 % dans l'eau (solvant A, 99 %) et de l'acide formique à 0,1 % dans l'acétonitrile (solvant B, 1 %) à un débit de 10 µl/min pendant 3 min. L'élution des peptides a ensuite été réalisée à un débit de 300 nL/min en utilisant une colonne Aurora (C18, 250 mm × 75 µm, 1,6 µm ; 120 Å, Ion Optics) maintenue à 60 °C avec le gradient appliqué détaillé en tableau 18.

Tableau 18 : Gradient chromatographique utilisé pour l'analyse des échantillons enrichis en phosphopeptides sur l'Eclipse.

Temps (min)	Pourcentage en solvant B (%)
0	2
65	35
66	90
71	90
72	2
90	2

a. HCD

L'Eclipse fonctionnait avec une source nanoSpray en mode ion positif avec une température de source réglée à 250 °C et une tension de spray de 2,0 kV. Les spectres MS Full Scan (375-1500 m/z) ont été acquis sur l'analyseur Orbitrap avec une résolution de 120 000 à 200 m/z, un temps d'injection maximal de 50 ms et une *AGC target* de 4×10^6 charges. Pour chaque spectre MS Full scan, les 20 précurseurs les plus intenses ont été isolés en utilisant une fenêtre de 2 m/z et fragmentés en utilisant la dissociation par collision à haute énergie (HCD, énergie de collision normalisée à 27%), et l'exclusion dynamique des précurseurs déjà fragmentés a été fixée à 40 s. Les spectres MS/MS (120-1200 m/z) ont été acquis sur l'analyseur Orbitrap avec une résolution de 15 000 à 200 m/z, un temps d'injection maximal de 22 ms et une *AGC target* de 5×10^5.

b. ETD

L'Eclipse fonctionnait en mode ion positif avec une température de source réglée à 250 °C et une tension de spray de 2,0 kV. Les spectres MS Full Scan (375-1500 m/z) ont été acquis sur le détecteur Orbitrap avec une résolution de 120 000 à 200 m/z, un temps d'injection maximal de 50 ms et une *AGC target* de 4×10^6 charges. Pour chaque spectre MS Full scan, les 20 précurseurs les plus intenses ont été isolés en utilisant une fenêtre de 2 m/z et fragmentés en utilisant la dissociation par transfert d'électron (ETD, option « *use calibrated charge ETD parameters* » activée), et l'exclusion dynamique des précurseurs déjà fragmentés a été fixée à 40 s. Les spectres MS/MS (120-1200 m/z) ont été acquis dans la trappe ionique en scan « Rapid », avec un temps d'injection maximal de 35 ms et une *AGC target* de 1×10^5.

11. Optimisation des méthodes de fragmentation sur timsTOF Pro

Des échantillons de cerveaux de souris enrichis en phosphopeptides par IMAC ont été utilisés pour optimiser les paramètres de fragmentation du timsTOF Pro (Bruker Daltoniks). Les paramètres LC et MS étaient identiques à ceux utilisés dans le paragraphe II.A.8.b. L'énergie de collision appliquée est fonction de la valeur de mobilité de l'ion. Deux rampes d'énergies de collision ont été comparées et les valeurs sont détaillées dans la figure 129. La méthode « classique » correspond à l'énergie de collision appliquée dans le paragraphe II.A.8.b.

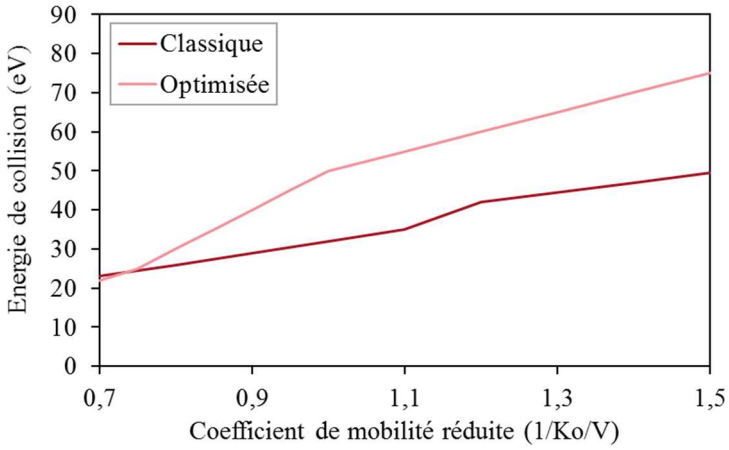

Figure 129 : Énergie de collision en fonction du coefficient de mobilité ionique.

Les données obtenues ont ensuite été traitées avec MaxQuant (v 2.0.3.0) selon les paramètres décrits dans le paragraphe II.A.8.b.

12. L'apport d'une étape de fractionnement peptidique

a. Lyse et dosage

Un échantillon de muscle d'ours (54 mg) a été préparé selon le protocole « Urée/Thio-urée » sans précipitation comme décrit dans le paragraphe II.A.1. À partir de cet extrait, trois aliquots, contenant chacun 520 µg de protéines, sont destinés à être digérés puis fractionnés et trois autres aliquots de 170 µg ont été préparés pour une digestion sans fractionnement.

b. Réduction, alkylation, digestion liquide, SPE

Les protéines ont été réduites et alkylées sur la plateforme Bravo AssayMAP (Agilent Technologies) en utilisant le protocole « In-solution Digestion : Single plate » (v1.2). Brièvement, les protéines ont été réduites pendant 30 min à 37 °C avec une solution de DTT 20 mM ABC 0,1 M pour atteindre une concentration finale en DTT de 12 mM puis alkylées pendant 1 h à RT avec une solution d'IAM 700 mM, ABC 0,1 M pour atteindre une concentration finale en IAM de 130 mM. Un tampon ABC 0,1 M a été ajouté pour diminuer la concentration d'urée en dessous de 1 M.

Les protéines ont été digérées pendant une nuit à 37°C par ajout manuel d'un mélange trypsine/Lys C (Mass Spec Grade, Promega, Madison, WI, USA) avec un ratio de 1/50 (enzyme/protéine). La digestion a été stoppée par ajout de FA pour obtenir un pH inférieur à 3.

Les peptides obtenus ont été purifiés par SPE avec le protocole « peptide clean-up » (v3.0) sur un Bravo AssayMAP (Agilent Technologies). En bref, les cartouches de 5µL de phase C18 pour les échantillons de 170 µg et de 25µL de phase RPS pour les échantillons de 550 µg ont été lavées et conditionnées avec 50% d'ACN, 0,1% de TFA, puis équilibrées avec H2O, 0,1% de TFA. Les échantillons ont été chargés à 5 µL/min sur la phase solide puis lavés avec H2O, 0,1% TFA avant d'être élués dans 120 µL de 80% ACN, 0,1% FA à 5 µL/min.

c. Fractionnement sur phase inverse à pH basique

Les échantillons de 550 µg ont été évaporés à sec à l'aide d'une centrifugeuse à vide (SpeedVac, Thermo Fisher Scientific) et repris dans 260 µL d'une solution de formate d'ammonium 10 mM, pH = 10. Les échantillons ont été fractionnés sur la plateforme Bravo AssayMAP (Agilent

Technologies) en utilisant un protocole customisé « Fractionation » (v1.1) en réutilisant les cartouches RPS 25 µL utilisées pour la SPE au paragraphe précédent (II.A.12.b). Les cartouches ont été lavées avec 250 µL d'une solution à 90% d'ACN à 300 µL/min puis conditionnées avec 250 µL d'une solution de formate d'ammonium 10 mM, pH = 10 à 10 µL/min. Les échantillons ont été déposés sur la cartouche avec un débit de 5µL/min puis lavés avec 250 µL d'une solution de formate d'ammonium 10 mM, pH = 10 à 10 µL/min. Les fractions ont été éluées séquentiellement en augmentant le pourcentage en acétonitrile de la solution de formate d'ammonium 10 mM, pH = 10 : F1 (20% acétonitrile), F2 (30% acétonitrile) et F3 (90% acétonitrile). Chaque fraction a été éluée avec 120 µL de solution à 5 µL/min. Les échantillons ont été évaporés à sec à l'aide d'une centrifugeuse à vide (SpeedVac, Thermo Fisher Scientific) et repris dans 170 µL d'une solution à 80% ACN 0,1% FA.

d. Enrichissement

Tous les échantillons (550 µg après fractionnement et 170 µg) ont été dopés avec le phosphomix I (Thermo Fisher Scientific) et enrichis comme indiqué dans le paragraphe II.A.7.

e. Analyse nanoLC-MS/MS

Les phosphopeptides ont été analysés sur le système NanoAcquity couplée à un Q-Exactive HF-X. L'analyse nanoLC-MS/MS a été réalisée avec les mêmes paramètres que ceux décrits dans le paragraphe II.A.8.a, seul le débit de la phase mobile a été réduit à 350 nL/min.

f. Identification et quantification des phosphopeptides à partir des spectres obtenus

Les données ont été traitées avec MaxQuant v1.6.17.0 selon les paramètres décrits dans le paragraphe II.A.8.a.

B. Application de cette méthodologie pour étudier la préservation du muscle durant l'hibernation de l'ours

1. Lyse et dosage

Entre 37 et 66 mg de muscle d'ours (N=11/ condition, hiver + été) ont été broyés, lysés et dosés selon le protocole « Urée/Thio-urée » sans précipitation comme décrit dans le paragraphe II.A.1. À ce stade, un échantillon de référence a été préparé en mélangeant une fraction égale de chaque extrait protéique. Cet échantillon a ensuite été traité de la même façon que les autres.

2. Réduction, alkylation, digestion liquide, SPE

550 µg de protéines de chaque échantillon ont été réduites, alkylées, digérées avec un mélange trypsine/Lys-C (Promega) selon les conditions décrites dans le paragraphes II.A.12.b. Les sels ont été éliminés par une étape de SPE selon les conditions décrites dans le paragraphe II.A.12.b.

3. Fractionnement sur phase inverse à pH basique

Le fractionnement sur phase inverse à pH basique a été effectué avec le protocole classique « Fractionation » (v1.1) sur le Bravo AssayMAP (Agilent Technologies). Les conditions étaient les mêmes que celles décrites dans le paragraphe II.A.12.c. Les fractions obtenues ont été évaporées à sec par centrifugeuse à vide (SpeedVac, Thermo Fisher Scientific) et reprises dans 175µL d'une solution à 80% ACN 0,1% TFA. À ce stade, 10 µL (≈ 10 µg) ont été prélevés pour réaliser l'analyse du protéome total.

4. Enrichissement

Les échantillons ont été dopés avec le mélange de phosphomix I et soumis à un enrichissement Fe(III)-NTA sur la plateforme Bravo AssayMAP suivant le protocole « Phosphopeptide Enrichment » (v2.0) selon les paramètres décrits dans le paragraphe II.A.12.d. Les différentes fractions enrichies de l'échantillon de référence ont été regroupées pour former un mélange enrichi en phosphopeptides qui a été injecté à plusieurs reprises au cours de la séquence d'analyse afin d'effectuer un suivi de contrôle qualité du système.

5. Analyse LC-MS/MS

Après évaporation à sec, les échantillons enrichis en phosphopeptides ont été repris dans 30 µL d'une solution 2% ACN 0,1% FA et congelés à -80°C. Ils ont été décongelés et placés dans le passeur automatisé d'échnatillons pour une durée maximale de 24h afin d'éviter la dégradation thermique des phosphopeptides. Avant injection, le mélange de peptides iRT a été ajouté (1/10 v/v) dans chaque échantillon. L'échantillon de référence a été injecté tous les 6 échantillons. Les phosphopeptides (9 µL) ont été analysés sur le système NanoAcquity couplée à un Q-Exactive HF-X. Les trois fractions enrichies d'un échantillon ont été injectées à la suite puis deux analyses de solvant (blanc) utilisant un gradient de 30 min ont été réalisées. L'injection des échantillons a été randomisée. L'analyse nanoLC-MS/MS a été réalisée avec les même paramètres que ceux décrits dans le paragraphe II.A.8.a.

Les échantillons pour l'analyse du protéome total ont été évaporés à sec et repris dans 67 µL d'une solution 2% ACN 0,1% FA. Les différentes fractions non enrichies de l'échantillon de référence ont été regroupées et cet échantillon a été injecté tous les 6 échantillons. Les peptides iRT (Biognosis) ont été ajoutés (1/10, v/v) dans tous les échantillons et 340 ng de peptides ont été injectés avec les mêmes paramètres LC et MS que pour l'analyse des échantillons enrichis en phosphopeptides.

6. Identification et quantification des protéines à partir des échantillons non enrichis

Les données brutes issues de l'analyse des échantillons non enrichis ont été traitées avec MaxQuant (version 1.6.17.0)[424] selon les paramètres décrits dans le paragraphe I.5 en utilisant une base de données de protéines créée à l'aide de notre plateforme Galaxy (https://iphc-galaxy.u-strasbg.fr/). La base de données contenait les séquences de protéines RefSeq (35 421 entrées) dérivées du dernier assemblage complet et annoté (Annotation release ID : 100) du génome d'*Ursus arctos* (ASM358476v2), qui ont été téléchargées en août 2021. Les protéines retenues pour la suite de l'analyse étaient celles pour lesquelles les valeurs quantitatives ne présentaient pas plus de six valeurs absentes par groupe et Les protéines absentes dans un groupe donné (c'est-à-dire pas du tout détectées).

7. Identification et quantification des phosphopeptides à partir des échantillons enrichis en phosphopeptides

Les données brutes issues de l'analyse des échantillons enrichis en phosphopeptides ont été traitées avec MaxQuant (version 1.6.17.0)[424] avec les mêmes paramètres que ceux décrits dans le paragraphe précédent (II.A.8.a.) à quelques modifications près. Les phosphorylations ont été paramétrées comme modifications variables sur les sérines, thréonine et tyrosines. Les options « *Match betwenn runs* » et « MaxLFQ » ont été désactivées. Les identifications et intensités

des phosphopeptides se trouvaient dans le fichier texte « Phospho (STY)Sites ». Ce fichier a ensuite été traité avec l'option « *expand site tables* » de Perseus (v1.6.14.0) afin d'obtenir les identifications et les données quantitatives au niveau des sites de phosphorylation. Les sites situés sur des peptides partagés ont été exclus du jeu de données. Seuls les sites avec une probabilité de localisation supérieure à 0,75 et pour lesquels les valeurs quantitatives ne présentaient pas plus de six valeurs absentes par groupe ont été retenus pour la suite de l'analyse. Les sites non détectés dans un groupe donné (c'est-à-dire sans valeurs quantitatives) ont également été retenus.

a. Alignement de séquence pour identifier le site humain homologue au site chez l'ours

La correspondance entre les sites de phosphorylation sur les protéines d'ours et les sites correspondants sur les protéines humaines a été réalisée avec l'outil Emboss Stretcher[332], la validation manuelle des correspondances est expliquée en détails dans le paragraphe dédié de la partie résultats (Partie II, chapitre 2, IV.C.1).

b. Utilisation des banques fonctionnelles avec les données humaines

À partir des positions des phosphorylations sur les protéines humaines déterminées dans le paragraphe précédent, les données disponibles sur ces sites de phosphorylations ont été récupérées dans différentes banques. À partir de Perseus (v1.6.14.0), la présence du site dans la banque PhosphoSitePlus ainsi que les informations relatives aux kinases et des informations fonctionnelles ont ainsi été collectées. De plus, des informations complémentaires ont été recherchées sur UniprotKB, telles que la description des sites de phosphorylation et, lorsque disponible, leur rôle biologique.

c. Analyse statistique du jeu de données

En utilisant le logiciel R (v3.4.0), les tests de Shapiro-Wilk et de Bartlett ont été utilisés pour vérifier la distribution normale et l'homoscédasticité des données ($p > 0,01$). Un T-test a été effectué pour ici pour comparer les deux moyennes issues des deux conditions (été et hiver). Le seuil de significativité a été fixé à $p < 0,05$.

d. Analyse d'enrichissement en annotations fonctionnelles

L'analyse d'enrichissement en annotations fonctionnelles des données nanoLC-MS/MS issues des échantillons enrichis en phosphopeptides a été réalisée à l'aide de la version téléchargeable de DAVID (Ease v2.1, https://david.ncifcrf.gov/) et d'une version actualisée des bases de données Gene Ontology (GO) (juillet 2022). Les termes GO ont été considérés comme significativement enrichis lorsqu'ils présentaient un score Ease inférieur à 0,0001, une p-value de Benjamini inférieure à 0,001 et un *« fold enrichment »* supérieur à 2. Enfin, les fonctions altérées entre les deux saisons ont été déterminées après que les termes GO enrichis aient été regroupés en grandes catégories fonctionnelles.

III. Développement d'un protocole de préparation d'échantillon adapté à l'analyse protéomique de cœurs de lérots

A. Développement d'un protocole de préparation d'échantillon adapté à l'analyse protéomique de cœurs de lérots

Les échantillons ont été fournis par notre collaborateur Sylvain Giroud (University of Veterinary Medicine, Vienne, Autriche).

1. Broyage lyse dosage

Un demi cœur de lérot (*Eliomys Quercinus*) congelé a été lavé au *Phosphate Buffer Saline* [PBS] afin d'éliminer les traces résiduelles de sang risquant de contaminer les échantillons, puis environ 20 mg ont été broyés sous azote liquide à l'aide d'un broyeur à billes (MM400, Retsch, Eragny sur Oise, France) (3*30 s à 30 Hz). 1,5mL de tampon de lyse (urée 6 M, thio-urée 2 M, ABC 0,1 M, inhibiteurs de protéases 1/50) a été ajouté et l'échantillon a été incubé 20 min à RT. Les protéines ont été dosées par dosage Bradford (Bio-Rad, Hercules, CA, USA).

2. Gel SDS-PAGE « Stacking »

4 réplicas techniques de 30 µg de protéines ont été préparés selon le protocole décrit dans le paragraphe I.3. Les peptides ont ensuite été évaporés à sec.

3. Protocole S-trap

4 réplicas techniques de 30 µg de protéines ont été préparés. Les protéines ont été réduites et alkylées sur la plateforme Bravo AssayMAP (Agilent Technologies) en utilisant le protocole « In-solution Digestion : Single plate » (v1.2). Les protéines ont été réduites pendant 30 min à 37 °C avec une solution de DTT 20 mM ABC 0,1 M pour atteindre une concentration finale en DTT de 12 mM puis alkylées 1 h à RT avec une solution d'IAM 210 mM ABC 0,1 M pour atteindre une concentration finale en IAM de 40 mM. La digestion des protéines a été réalisée à l'aide du protocole de traitement des échantillons S-Trap™ (Protifi, Huntington, NY, États-Unis). Brièvement, les échantillons ont d'abord été mélangés avec une solution de SDS (10%) pour atteindre une concentration finale de 2,3 %, puis une acidification a été réalisée à l'aide d'une solution d'acide phosphorique (VWR International, Radnor, AR, USA) à 27,5 % (1/10 ; v/v). Après l'ajout de six volumes de tampon de fixation (MeOH 90%, TEAB 100 mM, pH=7,1 ; Fisher Scientific, Hampton, NH, USA), les échantillons ont été chargés sur un filtre S-Trap™ micro et centrifugés (1 min, 4000 x g, RT), et trois lavages ont été effectués avec 150 µL du tampon de fixation. Les protéines ont été digérées par incubation pendant 3 heures à 37 °C dans une solution d'enzymes trypsine/Lys-C (Promega, Madison, WI, USA) diluée dans d'ABC 25 mM à un ratio protéine/enzyme de 10/1. Les peptides ont été élués par étapes avec 40 µL de trois tampons différents (ABC 25 mM, puis 0,2% FA dans l'eau, puis 50% ACN dans l'eau). Les trois solutions ont été regroupées et évaporées à sec.

4. Protocole SP3

4 réplicas techniques de 30 µg de protéines ont été préparés. Les protéines ont été réduites et alkylées sur la plateforme Bravo AssayMAP en utilisant le protocole « In-solution Digestion : Single plate » (v1.2) avec les mêmes paramètres dans le paragraphe précédent (III.A.3). Les protéines ont ensuite été digérés selon le protocole SP3 décrit dans le paragraphe II.A.4. Les

peptides ont ensuite été évaporés à sec à l'aide d'un Speedvac (Savant, Thermo Fisher Scientific).

5. Protocole Auto SP3

4 réplicas techniques de 30 µg de protéines ont été préparés. Les billes A (billes Sera-Mag Speed, Thermo Fisher Scientific, 45152105050250) et les billes B (billes Sera-Mag Speed, Thermo Fisher Scientific, 65152105050250) ont été combinées (ratio 1:1) et ont été lavées à l'eau ultra-pure à trois reprises avant d'être reprises dans 600 µL d'eau ultrapure pour aboutir à une concentration de 50 µg/µL. Le protocole « SP3 protocol » (v0.9b.5) a été réalisé sur la plateforme Bravo (Agilent Technologies) équipée d'une tête 96LT. Les protéines ont été réduites à 37°C pendant 30 min avec une solution de DTT 36 mM ABC 0,1 M pour atteindre une concentration finale en DTT de 12 mM puis alkylées à RT pendant 30 min avec une solution d'IAM 160 mM ABC 0,1 M pour atteindre une concentration finale de 40 mM. 5µL de la solution de billes a été ajoutée aux échantillons. Après fixation des protéines aux billes avec 50% ACN pendant 10 minutes, les protéines ont été lavées deux fois avec 80% d'EtOH (200µL) et une fois avec 100% d'ACN (180µL). Les protéines ont ensuite été remises en suspension dans 45 µl de solution ABC 0,1 M avant d'être digérées pendant sur la nuit sur les billes (rapport enzyme:protéine de 1:20) à 1 000 tr/min à 37 °C en utilisant une combinaison de trypsine/lys-C (Promega, Madison, USA). Durant la nuit, la plaque a été recouverte par un film pour éviter l'évaporation. La digestion a été arrêtée en ajoutant une solution de TFA à 5% pour atteindre 1% de concentration finale. Après incubation sur le rack magnétique, le surnageant contenant les peptides a été récupéré. Les échantillons ont ensuite été évaporés à sec.

6. Analyse nanoLC-MS/MS

Tous les échantillons ont été repris dans 200 µL et 2 µL (300 ng) de cette solution ont été injectés le couplage NanoAcquity Q-Exactive HF-X

- Conditions chromatographiques

Les paramètres LC sont identiques au paragraphe II.8.a avec le gradient détaillé en tableau 19 à un débit de 350 nL/min.

Tableau 19 : Gradient chromatographique utilisé pour l'analyse des échantillons de cœur de lérot.

Temps (min)	Pourcentage en solvant B (%)
0	2
3	2
80	40
81	90
89	90
90	2
110	2

- Paramètres MS

Le Q-Exactive HF-X est utilisé en mode ion positif avec la température de la source réglée à 250 °C et la tension de spray à 1,9 kV. Les spectres MS à balayage complet (300-1800 m/z) ont été acquis à une résolution de 120 000 à 200 m/z, avec un temps d'injection maximal fixé à 50

ms et une *AGC target* fixée à $3x10^6$ charges. L'option de verrouillage de la masse a été activée (polysiloxane, 445.12002 m/z). Jusqu'à 20 peptides les plus intenses (au moins doublement chargés) par balayage complet ont été isolés en utilisant une fenêtre de 2 m/z et ils ont été fragmentés en utilisant la dissociation par collision à haute énergie (énergie de collision normalisée fixée à 27 et exclusion dynamique des précurseurs déjà fragmentés fixée à 60 s). Les spectres MS/MS (300-2000 m/z) ont été acquis avec une résolution de 15 000 à 200 m/z, avec un temps d'injection maximum de 100 ms et une *AGC target* fixée à $1x10^5$ et l'option de sélection de correspondance peptidique était activée.

7. Identification et quantification des protéines à partir des spectres obtenus

Le traitement des données brutes MS a été effectué dans MaxQuant (v2.0.3.0). Les listes de pics ont été analysées à l'aide du moteur de recherche Andromeda implémenté dans MaxQuant. La base de données de protéines que nous avons utilisée contenait les séquences de protéines connues dérivées de la dernière annotation du génome d'*Ictidomys tridecemlineatus* (TaxID 43179) dans Refseq (Refseq Assembly accession GCF_016881025.1 ; Assembly Name HiC_Itri_2). La base de données des protéines contenait également des séquences de protéines dérivées de l'annotation du génome complet de la mitochondrie d'*E. quercinus* (TaxID 53277 ; Genbank MN935777.1) réalise par nos soins. En bref, les recherches de cadres de lecture ouverts (ORF) et la traduction subséquente ont été effectuées à l'aide de ORFfinder (www.ncbi.nlm.nih.gov/orffinder/) en novembre 2021. Les paramètres par défaut ont été utilisés. Seules les protéines les plus longues par gène (séquences d'ADN codantes) ont été retenues. Les protéines de moins de 40 acides aminés n'ont pas été prises en compte. Pour attribuer un nom à ces séquences traduites, des recherches de type blast ont été effectuées à l'aide de Blast+2.12.0 (NCBI) contre des séquences protéiques UniprotKB (consultées en mars 2022) provenant d'espèces phylogénétiquement proches de *E. quercinus*. En utilisant un seuil de bit-score minimum fixé à 50. Seul le meilleur résultat de blast a été retenu pour une protéine d'*E. quercinus* donnée, permettant la propagation du nom de la protéine. Si aucune correspondance satisfaisante n'était trouvée, une autre recherche était effectuée sur des séquences protéiques UniprotKB d'espèces plus éloignées phylogénétiquement. Les correspondances ont été validées manuellement. Après élimination de la redondance, la base de données contenait 40 831 séquences de protéines. Les paramètres utilisés sont identiques à ceux décrit au paragraphe I.5.

B. Application du protocole choisi pour étudier l'influence des acides gras polyinsaturés sur le protéome du cœur de lérot au cours de leur hibernation

1. Broyage lyse dosage

Les cœurs ont été prélevés à 4 stades physiologiques différents (la période active (S), au début (ET) et à la fin (LT) d'épisodes de torpeur, et durant une phase de réveil (I-IBE)). Les lérots étaient nourris avec deux régimes distincts : soit un régime riche en acides gras polyinsaturés (AGPI) n-6 par rapport aux AGPI n-3 (HLA), soit un régime pauvre en AGPI n-6 par rapport aux AGPI n-3 (LLA). Au total, 8 conditions (N=5/groupe) ont été comparées entre elles : S-LLA, S-HLA, ET-LLA, ET-HLA, LT-LLA, LT-HLA, I-IBE-LLA, I-IBE-HLA.

Entre 55 et 200 mg de cœur de lérots ont été broyés et lysés selon le protocole décrit dans le paragraphe III.A.1. Les protéines ont été dosées par dosage Bradford (Bio-Rad, Hercules, CA,

USA). À ce stade, un échantillon de référence a été préparé en mélangeant une fraction égale de chaque échantillon. Cet échantillon a ensuite été traité de la même façon que les autres.

2. Réduction, alkylation et digestion par protocole S-trap

15 µg de protéines par échantillon ont été préparés selon le protocole décrit en III.A.3.

3. Analyse nanoLC-MS/MS

Après reprise des peptides dans une solution 2% ACN 0,1% FA, le mix de peptides iRT (Biognosis) ont été ajoutés (1/10, v/v). 300 ng de peptides ont été analysés avec les paramètres LC et MS décrits en paragraphe III.A.6. L'échantillon de référence a été injecté tous les 8 échantillons.

4. Identification et quantification des protéines à partir des spectres obtenus

Les données ont été analysées avec MaxQuant (v2.0.3.0) selon les paramètres décrits en paragraphe III.A.7. Toutes les protéines identifiées avec un minimum d'un peptide unique et quantifiées chez au moins quatre des cinq individus par groupe (c'est-à-dire qu'une seule valeur manquante a été acceptée par groupe) ont été conservées pour la suite de l'analyse.

5. Analyse statistique des données quantitatives

En utilisant le logiciel R v3.4.0, les tests de Shapiro-Wilk et de Bartlett ont été utilisés pour vérifier la distribution normale et l'homoscédasticité des données ($p > 0,01$). Les moyennes obtenues dans tous les groupes ont d'abord été comparées en utilisant l'ANOVA à 1 facteur. Lorsque l'ANOVA a montré que les moyennes n'étaient pas toutes égales ($p< 0,05$), une comparaison des conditions deux à deux a ensuite été effectuée à l'aide de la méthode HSD (*Honestly Significant Difference*) de Tukey (test posthoc). Il convient de noter que la procédure HSD de Tukey prend en compte la comparaison multiple et maintient le coefficient alpha expérimental au niveau spécifié ; ici, la valeur seuil a été fixée à $p < 0,05$. Un test ANOVA multifactoriel a également été appliqué pour évaluer les effets de la diète et du statut physiologique ainsi que les interactions de ces deux facteurs, le seuil de significativité choisi était de 0,05.

Ces différentes analyses statistiques ont été réalisées sur l'ensemble du jeu de données mais aussi sur les données restreintes aux échantillons I-IBE et « torpeur » composé des groupes ET et LT réunis (échantillons « phase active » (S) ont été retirés de l'analyse).

IV. Développement d'une méthode analytique permettant une étude approfondie de la glycation du plasma par une approche « bottom-up »

A. Mise en place d'un enrichissement par chromatographie d'affinité au boronate pour identifier les peptides glyqués

1. Préparation d'un échantillon de BSA glyquée

a. Incubation in vitro de la BSA

La BSA (Sigma Aldrich, Saint-Louis, MO, USA) a été incubée dans 2 solutions différentes (voir tableau 20) à une concentration de 25 g/L dans un volume total de 25 mL.

Tableau 20 : Composition des tampons d'incubation.

	Contrôle	Glucose
Composition du tampon	50 mM Tris-HCl pH=7.5	0,4 M Glucose 50 mM Tris-HCl pH=7.5

Ces solutions ont été incubées dans l'obscurité à 37 °C pendant 23 jours. Des aliquots de 150 µL ont été prélevés à 5 temps : 0, 1, 8 ; 16 et 23 jours.

b. Analyse LC-MS

Après dilution par 10 dans de l'eau acidifiée 0,1% FA, 5µL de ces échantillons ont été injectés sur une colonne HPLC vydac 208TP C8 (2,1 x 250 mm, 300 Å, 5 µm, Grace, Columbia, MD, USA) maintenue à 60 °C et montée sur une HPLC 1200 (Agilent Technologies, Paolo Alto, USA) couplée à un spectromètre de masse quadripôle-TOF (maXis II, Bruker Daltoniks GmbH, Bremen, Germany). Les solvants étaient composés de : A (eau acidifiée par 0,1% TFA) et B (ACN acidifiée par 0,1% TFA). La séparation a été effectuée à un débit de 0,25 mL/min selon un gradient linéaire de 25 à 70% de solvant B entre 7 et 34 min, puis 10 min à 70% de solvant B.

Le spectromètre de masse opérait en mode ESI positif avec une température en source de 220 °C, un débit de N2 de 4 L/min et le voltage à 4500V. L'acquisition des spectres MS « Full Scan » a été réalisée avec une gamme de masse de 500 à 3000 en mode « focus. Le système était contrôlé par Hystar (v3.2 ; Bruker Daltonik GmbH). Les données ont été analysées avec DataAnalysis (v4.3; Bruker Daltonik GmbH). Les spectres ont été lissés en utilisant un algorithme gaussien puis déconvolués par l'outil « *Protein/ Large Molecule Deconvolution* ».

2. Enrichissement *Boronate Affinity Chromatography* (BAC) d'un échantillon de BSA en peptides glyqués

Différentes quantités de matière ont été prélevées dans l'échantillon incubé dans une solution de glucose 0,4 M (jour 1 ; préparation décrite en paragraphe IV.A.1.a) : 2 échantillons de 50 µg et 2 échantillon de 10 µg. Ces échantillons ont été dilués afin d'obtenir un volume final de 20 µL avec une solution de 8 M urée, 0,1 M ABC

a. Réduction, alkylation, digestion liquide, SPE

Les étapes de réduction, d'alkylation et de dilution ont été réalisées de manière automatisée en suivant le protocole détaillé en paragraphe II.A.12.b. Les échantillons ont ensuite été digérés par de la trypsine (Promega) avec un ratio de 1 :50 (enzyme/protéine) pendant une nuit à 37 °C. La digestion a été stoppée en ajoutant de l'acide formique pur afin d'obtenir un pH inférieur à 3. Les peptides ont été dessalés sur des cartouches de phase C18 5 µL (Agilent Technologies) en utilisant un protocole de « peptide clean-up » (v2.0) sur un Bravo AssayMAP (Agilent Technologies) en suivant les paramètres décrits dans le paragraphe II.A.3. Le solvant a été évaporé par SpeedVac.

b. Enrichissement en peptides glyqués sur cartouches de boronate

Les échantillons ont été repris avec 60 µL d'une solution de 250mM acétate d'ammonium à un pH de 8,1. L'enrichissement a été effectué de façon automatisé sur le Bravo AssayMAP à l'aide de cartouches remplies à façon de 5 µL de phase boronate (Affi-gel, Bio-Rad) par Benelab (Zagreb, Croatie) en utilisant le protocole de « peptide clean-up » (v2.0). Tout d'abord, les cartouches ont été conditionnées avec 150 µL de solution d'acétate d'ammonium 250 mM pH 8,1 à un débit de 10 µL/min puis équilibrées avec 100 µL de cette même solution au même débit. Les échantillons ont été déposés sur les cartouches à un débit de 5 µL/min avant d'être lavés par 25 µL de solution d'acétate d'ammonium 250 mM pH 8,1 à 5 µL/min. Pour ces deux étapes, les fractions non retenues ont été collectées individuellement. Enfin, les peptides retenus ont été élués avec 20 µL d'une solution acide (0,1 M acide acétique). Le solvant a été évaporé et les échantillons ont été repris dans 50 µL pour les échantillons « 50 µg » et dans 10 µL pour les échantillons « 10 µg ».

c. Analyse nanoLC-MS/MS

3 µL des échantillons enrichis en peptides glyqués ont été injectés sur une NanoAcquity (Waters) couplée à un Q-Exactive HF-X (Thermo). Pour les échantillons non enrichis, 6 ng théoriques de BSA ont été injectés. Les premières fractions non retenues (*flow through loading*, FTL) ont été reprises afin d'obtenir une concentration théorique de 3 ng/µL (en faisant l'hypothèse que rien n'était retenu sur la cartouche) et 2 µL ont été injectés. Les fractions non retenues lors du lavage de l'échantillon (flow through washing, FTW) ont été reprises dans 25 µL pour les échantillons avec 50 µg de quantité initiale et 5 µL pour les échantillons avec 50 µg de quantité initiale. Ces solutions ont été diluées 100 x avec une solution à 2% ACN, 0,1% FA avant l'injection d'1 µL.

Les échantillons ont été analysés sur le couplage NanoAcquity Q-Exactive HF-X selon les paramètres LC et MS du paragraphe II.A.8.a avec le gradient détaillé en tableau 21.

Tableau 21 : Gradient chromatographique utilisé pour l'analyse des échantillons de BSA.

Temps (min)	Pourcentage en solvant B (%)
0	1
2	3
12	35
13	70
14	70
30	3

d. Identification des peptides de la BSA

Les données brutes issues de l'analyse LC-MS/MS ont été traitées avec MaxQuant (version 1.6.17.0)[424]. Les listes de pics ont été créées à l'aide de paramètres par défaut et analysées à l'aide du moteur de recherche Andromeda implémenté dans MaxQuant par rapport à une base de données de protéines créée à l'aide de l'outil MSDA. La base de données contenait des séquences de protéines SwissProt de *Bos Taurus* (6005 entrées) téléchargées en juillet 2020. Les fructosamines sur les lysines et les arginines sont recherchées en modifications variables. Un nombre maximum de deux clivages manqués a été toléré, et un FDR de 1% pour les PSM (longueur minimale de sept acides aminés) ont été acceptés pendant l'identification. Les options "*Match between runs*" ainsi que MaxLFQ étaient désactivées. Le fichier texte « FLsites » a été utilisé pour l'identification des peptides glyqués. Les autres paramètres ne sont pas modifiés par rapport au paragraphe I.5.

3. Enrichissement BAC d'un échantillon de trastuzumab en peptides glyqués

Des échantillons d'un mAb commercial, le trastuzumab (Roche, Bâle, Suisse), ont été utilisés pour évaluer l'efficacité d'enrichissement des cartouches boronate pour l'étude des glycations des protéines thérapeutiques.

a. Incubation

150 µg d'échantillon ont été incubés 3 jours à 37 °C dans un tampon PBS (Thermo Fisher Scientific) à pH=7,2 avec ou sans glucose 30 mM (N=1/condition, « Glucose » et « Contrôle »). Le solvant a été évaporé et l'extrait protéique a été repris dans un tampon urée 8 M ABC 0,1 M.

b. Réduction, alkylation, digestion SP3

Les protéines ont été réduites et alkylées manuellement selon le protocole détaillé dans le paragraphe II.A.2. et digérées sur billes (méthode SP3) avec les conditions détaillées dans le paragraphe III.A.4. Les protéines ont été digérées avec de la trypsine (Promega) avec un ratio enzyme/protéines (1/50). La digestion a été effectuée sur la nuit à 37 °C, puis stoppée par ajout de TFA pour obtenir pH < 2. Les échantillons ont été soumis au bain à ultra-sons 2 min puis incubés à RT 2 min sur le rack magnétique. Le surnageant contenant les peptides a été récupéré.

c. Enrichissement en peptides glyqués sur cartouches de boronate

40 µg de peptides (N=3/condition) ont été enrichis en peptides glyqués avec les cartouches de 5 µL de phase boronate sur un Bravo AssayMAP (Agilent Technologies) en suivant le protocole décrit au paragraphe précédent (IV.A.2.b) à deux exceptions près : du $MgCl_2$ (50 mM) a été

ajouté au tampon d'acétate d'ammonium pH=8.1 et le débit de chargement de l'échantillon a été diminué à 2 µL/min. Les fractions non retenues par la cartouche (*Flow Through Loading* (FTL) et *Flow Through Washing* (FTW)) ont été regroupées (sous le terme *Flow Through* (FT)).

d. Analyse nanoLC-MS/MS

Le solvant a été évaporé et les échantillons enrichis ont été repris dans 10 µL de solution 2% ACN 0,1% FA. 15 µg d'échantillon non enrichi ont été repris dans 200 µL de cette solution et les FT dans 530 µL. Pour les échantillons enrichis, 5 µL ont été injectés. Pour les échantillons non enrichis et les FT, 2 µL (soit 150 ng théoriques) ont été injectés.

Les peptides ont été analysés sur une NanoAcquity (Waters) couplée à un Q-Exactive Plus (Thermo Fisher Scientific).

- Conditions chromatographiques

Les conditions chromatographiques étaient identiques à celles utilisées dans le paragraphe I.4 à l'exception du gradient détaillé en tableau 22.

Tableau 22 : Gradient chromatographique utilisé pour l'analyse des échantillons de trastuzumab enrichis en peptide glyqués.

Temps (min)	Pourcentage en solvant B (%)
0	1
2	8
32	35
33	90
37	90
38	1
58	1

- Paramètres MS

Les analyses ont été réalisés avec les mêmes paramètres de MS que ceux décrits dans le paragraphe I.4.

e. Identification et quantification des peptides du trastuzumab

Les listes de pics (fichiers mgf) ont été créées à partir des données brutes à l'aide de ProteoWizard MS Convert (v3.0.11417). La recherche a été effectuée par Mascot (v 2.6.2) avec une banque de données protéique humaine issue de SwissProt contenant les séquences des chaines légère et lourde du trastuzumab disponibles sur DrugBank (DB00072, avril 2021). La tolérance de masse pour les précurseurs a été fixée à 5 ppm. La tolérance de masse pour les ions fragments a été fixée à 0,8 Da. La carbamidométhylation des cystéines a été considérée comme fixe. Les fructosamines (+ 162 Da) des lysines et arginines ainsi que l'oxydation des méthionines et l'acétylation des extrémités N-terminales des protéines ont été considérées comme des modifications variables. Un nombre maximum de deux clivages manqués a été toléré. Les données obtenues ont été validées avec Proline[157] (v 1.0) en appliquant un filtre sur le « *pretty rank* » (supérieur à 1) et sur le nombre de peptide d'un protéine (minimum 3).

La quantification a été réalisée avec le logiciel BioPharmaFinder (v 3.2, Thermo Fisher Scientific). Les mêmes paramètres de recherche que dans le paragraphe ci-dessus ont été utilisés pour la recherche Mascot puis les courants d'ion extraits correspondant aux peptides portant une fructosamine ont été intégrés manuellement.

4. Enrichissement BAC d'un échantillon de plasma en peptides glyqués

Du plasma commercial (Thermo Fisher Scientific) a été utilisé pour ces phases d'optimisation.

a. Incubation

Le contenu protéique du plasma a été mesuré par dosage protéique BCA (Thermo Fisher Scientific). 2 mg de protéines ont été incubés 3 jours à 37 °C dans un tampon PBS (Thermo Fisher Scientific) à pH=7,2 avec ou sans Glucose 30 mM.

b. Réduction, alkylation, digestion SP3 et SPE

Le solvant a été évaporé et les protéines reprises, réduites et alkylées ont été digérées sur billes (protocole SP3) comme décrit dans le paragraphe IV.A.3.b. Les échantillons ont été dessalés manuellement sur cartouches Sep-Pak C18 1cc, 50 mg de phase (Waters). Avant de déposer les échantillons, les cartouches ont été préparées avec 2 mL de méthanol et 3 mL d'ACN, puis conditionnées avec 3 mL d'eau acidifiée 0,1% FA. Les échantillons ont ensuite été lavés par 2 mL de cette même solution avant d'être élués avec 600 µL d'une solution 60% ACN et 0,1% FA. Le solvant a été évaporé et les peptides ont été repris dans une solution 2% ACN, 0,1% FA.

c. Enrichissement en peptides glyqués sur cartouches de boronate

Des quantités définis d'extraits peptidiques (10 µg, 100 µg et 150 µg pour les échantillons « contrôle » et 100 µg pour les échantillons « glucose », N=3 par condition) ont été enrichis sur les cartouches boronate selon le protocole décrit dans le paragraphe IV.A.3.c.

Le solvant a été évaporé et les échantillons enrichis repris dans 75 µL (150 µg), 40 µL (100 µg) et 6 µL (10 µg) de solution 2% ACN 0,1% FA. Les échantillons non enrichis ont été repris dans cette solution pour obtenir 150 ng/µL et les FT dans 500 µL (150 µg), 670 µL (100 µg) et 67 µL (10 µg).

d. Analyse nanoLC-MS/MS

Après enrichissement en peptides glyqués, 5 µL d'échantillon ont été injectés. Pour les échantillons avant enrichissement et les FT, 2 µL (300 ng théoriques) ont été injectés. Les peptides ont été analysés sur le couplage NanoAcquity Q-Exactive Plus.

- Conditions chromatographiques

Les conditions chromatographiques étaient identiques à celles utilisées dans le paragraphe I.4 à l'exception du gradient détaillé en tableau 23.

Tableau 23 : Gradient chromatographique utilisé pour l'analyse des échantillons de plasma enrichis en peptide glyqués.

Temps (min)	Pourcentage en solvant B (%)
0	1
2	8
79	35
80	90
85	90
87	1
105	1

- Paramètres MS

Les analyses ont été réalisés avec les mêmes paramètres de MS que ceux décrits dans le paragraphe I.4.

e. Identification et quantification des peptides glyqués

Les données brutes issues de l'analyse LC-MS/MS ont été traitées avec MaxQuant (version 2.0.3.0)[424]. Les listes de pics ont été créées à l'aide des paramètres par défaut et analysées à l'aide du moteur de recherche Andromeda implémenté dans MaxQuant par rapport à une base de données de protéines créée à l'aide de l'outil MSDA. La base de données contenait les séquences de protéines SwissProt d'*Homo Sapiens* (20 421 entrées) téléchargées en mars 2020. Les paramètres sont les mêmes que ceux du paragraphe IV.A.2.d.

B. Mise en place d'une étape de fractionnement peptidique pour décomplexifier le plasma

1. Préparation d'un digest de plasma

Après resuspension du plasma commercialisé sous forme de poudre par Thermo Fisher Scientific dans 1 mL d'eau ultra-pure, la concentration protéique du plasma a été évaluée par dosage BCA. 5 mg de protéines ont été diluées 10 fois dans une solution d'urée 8 M, Tris-HCl 50 mM. Après réduction pendant 30 min à 37 °C par ajout d'une solution de DTT pour atteindre 12 mM finale et alkylation pendant 1 h à l'obscurité à RT par ajout d'une solution d'IAA pour atteindre 40 mM de concentration finale, les protéines ont été digérées par ajout d'un mélange de trypsine/Lys-C (Promega) avec un ratio de 1 :50 (enzyme/protéine) pendant une nuit à 37 °C. La digestion a été stoppée en ajoutant l'acide formique pur afin d'obtenir un pH inférieur à 3. Les peptides ont ensuite été aliquotés et congelés à -80 °C avant analyse.

2. Dessalage des peptides

350 µg (N=4) de digeste de protéines plasmatiques ont été dessalés sur cartouche Sep-Pak C18 1cc, 100 mg de phase (Waters) en suivant le même protocole que dans le paragraphe IV.A.4.b.

3. Fractionnement peptidique à pH basique

Après évaporation, les peptides purifiés ont été repris dans une solution 10 mM de formate d'ammonium à pH 10 afin d'obtenir une concentration 1 µg/µL. Le fractionnement de 100 µg d'extrait (N=3) a ensuite été effectué sur un Bravo AssayMAP (Agilent) en utilisant des cartouches 5 µL de phase RPS avec un protocole customisé « Fractionation » (v1.1). Ces cartouches ont été préparées avec 100 µL d'une solution de 90% ACN à un débit de 300 µL/min puis conditionnées avec 70 µL d'une solution 10 mM de formate d'ammonium à pH 10 à 10 µL/min. 100 µg d'échantillon ont ensuite été déposés sur la cartouche à 5 µL/min puis lavés par 25 µL de la solution de reprise avant d'être élués de manière séquentielle avec 30 µL à 10 µL/min en 4 étapes avec des concentrations croissantes d'ACN dans 10 mM de formate d'ammonium à pH 10. Les pourcentages en acétonitrile pour les deux conditions d'élution testées sont présentés dans le tableau 24.

Tableau 24 : Pourcentages en acétonitrile des solutions d'élution dans 10 mM de formate d'ammonium à pH 10 pour les deux conditions d'élution testées.

Fraction	Pourcentage d'acétonitrile ; conditions hpH RP 1 (%)	Pourcentage d'acétonitrile ; conditions hpH RP 2 (%)
F1	5	10
F2	15	20
F3	25	30
F4	80	80

Les peptides ont été élués dans 50µL d'une solution 50% ACN 0,1%TFA afin d'acidifier à nouveau le milieu. Les solvants ont ensuite été évaporés et les peptides repris dans 2% ACN 0,1% FA afin d'obtenir une concentration théorique de 150 ng/µL.

4. Fractionnement peptidique sur billes carboxylates

Avant utilisation, les billes A (billes Sera-Mag Speed, Thermo Fisher Scientific, 45152105050250) et les billes B (billes Sera-Mag Speed, Thermo Fisher Scientific, 65152105050250) ont été mélangées (50/50, v/v) et lavées trois fois à l'eau ultra-pure puis suspendues dans de l'eau ultra-pure afin d'obtenir une concentration de 100 µg de billes/µL. Après évaporation, 100 µg de peptides purifiés (N=3) ont été repris dans de l'eau ultra-pure. Les billes ont été ajoutées avec un ratio de 25/1 (billes totales/µg de protéines) soit 12,5/1 par rapport à chaque type de bille. De l'ACN a ensuite été ajouté pour atteindre une proportion finale de 95% et le mélange a été incubé 10 min à RT afin de laisser les peptides se lier aux billes. Le surnageant a été retiré et les peptides ont ensuite été élués avec 30 µL de solution contenant des concentrations décroissantes d'isopropanol. Les pourcentages en isopropanol des solutions d'élution pour les deux conditions d'élution testées sont reportés en tableau 25.

Tableau 25 : Pourcentages en isopropanol des solutions d'élution pour les deux conditions d'élution testées.

Fraction	Pourcentage d'isopropanol ; conditions CIF 1 (%)	Pourcentage d'isopropanol ; conditions CIF 2 (%)
F1	90	80
F2	80	60
F3	70	30
F4	0	0

Chaque fraction a été éluée 2 fois après incubation à RT pendant 10 min sous agitation à 1200 rpm. Les solvants ont ensuite été évaporés et les peptides repris dans 2% ACN 0,1% FA afin d'obtenir une concentration théorique de 150 ng/µL.

5. Analyse nanoLC-MS/MS

a. Q-Exactive HF-X

Les peptides obtenus après le fractionnement CIF 1, hpH RP 1 et hpH RP 2 ont été analysés (300 ng théoriques injectés) sur le couplage NanoAcquity Q-Exactive HF-X selon les paramètres LC et MS du paragraphe IV.A.2.c suivant le gradient détaillé en tableau 26.

Tableau 26 : Gradient chromatographique utilisé pour l'analyse des échantillons fractionné de plasma sur la NanoAcquity.

Temps (min)	Pourcentage en solvant B (%)
0	1
1	1
90	35
92	90
97	90
97.1	1
115	2

- Identification et quantification des peptides/protéines

Les données brutes issues de l'analyse LC-MS/MS ont été traitées avec MaxQuant (version 2.0.3.0)[424] avec les paramètres du paragraphe IV.A.4.e en retirant les fructosamines des modifications recherchées. Un nombre maximal d'un clivage manqué a été toléré.

b. Q-Exactive Plus

300 ng des peptides obtenus après le fractionnement CIF 2 ont été analysés par nano-chromatographie (NanoAcquity, Waters) couplée à un spectromètre de masse quadripôle – Orbitrap (Q-Exactive Plus, Thermo Fisher Scientific). Les conditions LC et MS étaient identiques à celles du paragraphe IV.A.4.d. Les données ont été traitées avec MaxQuant selon les paramètres du paragraphe précédent.

c. TimsTOF Pro

150 ng des peptides obtenus après les fractionnement hpH RP 2 et CIF 2 ont été injectés sur un système nanoUPLC (nanoElute, Bruker Daltoniks GmbH, Brême, Allemagne) couplé à un spectromètre de masse hybride à mobilité ionique - quadripôle et TOF (timsTOF Pro, Bruker Daltonics).

- Condition chromatographiques

La chromatographie liquide a été réalisée avec le solvant A (eau avec 0,1% de FA) et le solvant B (acétonitrile avec 0,1% de FA). Les peptides (5 µL) ont d'abord été chargés sur une précolonne Acclaim™ PepMap™ 100 C18 (100 Å, 5 µm, 100 µm × 20 mm ; Thermo Fisher Scientific) à 2 % de B à 120 bar. Ils ont ensuite été séparés à l'aide d'une colonne Aurora C18 (120 Å, 1. 6 µm, 75 µm × 250 mm ; IonOpticks, Melbourne, Australie) maintenue à 50 °C à l'aide d'un gradient allant de 2% à 50% de solvant B en 60 minutes à un débit de 300 nL/min. La colonne a ensuite été lavée avec 95% de B pendant 10 minutes et rééquilibrée avec 2 % de solvant B pendant 10 minutes.

- Paramètres MS

Les paramètres MS utilisés étaient les mêmes que ceux du paragraphe II.A.8.b.

- Identification et quantification des peptides/protéines

Les données brutes issues de l'analyse LC-MS/MS ont été traitées avec MaxQuant (version 2.0.3.0)[424] selon les paramètres du paragraphe IV.B.5.a à l'exception des tolérances de masse. Pour la première recherche, la tolérance de masse pour les précurseurs a été fixée à 20 ppm et elle a été fixée à 10 ppm pour la recherche principale. La tolérance de masse pour les ions fragments a été fixée à 40 ppm.

6. Analyse nanoLC-MS/MS des échantillons non fractionnés

Les peptides non fractionnés ont été analysés sur 3 plateformes instrumentales différentes :

- NanoAcquity couplée à un Q-Exactive Plus avec les conditions décrites dans le paragraphe IV.B.5.b (450 ng de peptides injectés, N=3). Les données ont été interprétées avec MaxQuant avec les paramètres du paragraphe IV.B.5.b.
- NanoAcquity couplée à un Q-Exactive HF-X avec les conditions décrites dans le paragraphe IV.B.5.a (450 ng de peptides injectés, N=3). Les données ont été interprétées avec MaxQuant avec les paramètres du paragraphe IV.B.5.b.
- NanoElute couplée à un timsTOF Pro avec les conditions décrites dans le paragraphe IV.B.5.c (300 ng de peptides injectés, N=3). Les données ont été interprétées avec MaxQuant avec les paramètres du paragraphe IV.B.5.c.

Sur le timsTOF Pro, un deuxième gradient, décrit en tableau 27 a été utilisé (300 ng de peptides injectés, N=3). Les paramètres MS et la recherche MaxQuant restaient inchangés.

Tableau 27 : Deuxième gradient chromatographique utilisé pour l'analyse des échantillons non fractionnés sur le timsTOF Pro.

Temps (min)	Pourcentage en solvant B (%)
0	2
2	2
62	50
65	95
75	95
76	2
86	2

Sur le Q-Exactive HF-X, différentes quantités ont été injectées (150 ng, 300 ng et 600 ng ; N=3 /quantité injectée) avec les paramètres du paragraphe IV.B.5.b. Une autre valeur de l'*AGC target* a été utilisée (2 x 10^4 charges, 450 ng de peptides injectés ; N=3) dans le but d'évaluer l'influence de ce paramètre. Le temps d'exclusion dynamique a également été modifié (15 s et 180 s, 450 ng de peptides injectés, *AGC target* = 2 x 10^6 charges ; N=3).

7. Recherche des fructosamines dans les échantillons après fractionnement

Les données issues de l'analyse du fractionnement hpH RP 2 analysées sur le timsTOF Pro ont été traitées à nouveau avec MaxQuant selon les paramètres décrits plus haut (IV.2.e.iii) en ajoutant les fructosamines sur les lysines et les arginines comme modifications variables.

C. Couplage du fractionnement peptidique et de l'enrichissement BAC des peptides glyqués

Les peptides issus du plasma commercialisé par Thermo Fisher Scientific obtenus selon le protocole décrit dans le paragraphe IV.B.1 ont été utilisés.

1. Fractionnement peptidique à pH basique sur phase inverse

Le fractionnement décrit dans le paragraphe IV.B.3 (condition d'élution de hpH RP 2) a été réalisé à partir de 100 µg de peptides (N=3). Quatre fractions ont été obtenues.

2. Enrichissement BAC des fractions

L'enrichissement BAC décrit dans le paragraphe IV.A.4.c a été réalisé sur chacune des fractions obtenues.

3. Analyse nanoLC-MS/MS

Les échantillons enrichis ont été repris dans 6 µL d'une solution 2%ACN 0,1%FA et la totalité de l'échantillon a été injecté sur le couplage NanoAcquity Q-Exactive Plus. 450 ng de peptides des échantillons fractionnés avant enrichissement ont également été injectés.

- Conditions chromatographiques

Les conditions LC étaient identiques à celles du paragraphe I.4 à l'exception du gradient détaillé en tableau 28.

Tableau 28 : Gradient chromatographique utilisé pour l'analyse des échantillons fractionnés par fractionnement peptidique à pH basique sur phase inverse puis enrichis par BAC.

Temps (min)	Pourcentage en solvant B (%)
0	1
2.5	8
58	35
59	90
65	90
66	1
90	1

- Paramètres MS

Les paramètres MS étaient identiques à ceux du paragraphe I.4.

4. Traitement des données

Les données ont été traitées avec MaxQuant (version 2.0.3.0). Les paramètres décrits dans le paragraphe IV.A.4.d étaient utilisés.

TRAVAUX ADDITIONNELS

Travaux Additionnels

Durant ma thèse, j'ai également eu l'opportunité de réaliser des analyses de protéomique quantitatives globales dans divers contextes biologiques qui n'ont pas été décrites dans ce manuscrit:

- Étude du protéome des fourmis ouvrières en fonction de leur âge et de leur rôle dans la fourmilière en collaboration avec Martin Quque et François Criscuolo (IPHC, DEPE, Strasbourg) : publication en cours de soumission.
- Étude du protéome musculaire du microcèbe au cours des changements physiologiques subis par l'animal au cours de l'hiver en collaboration avec Jérémy Terrien (MECADEV, CNRS, Brunoy)
- Étude du protéome de cellules hépocytaires HepaRG durant une trans-différentiation induite par méchano-transduction en collaboration avec Christiane Guillouzo (Inserm, Université Rennes 1, Rennes)

J'ai également réalisé des analyses par Western Blot :

- Étude de marqueurs hépatiques du microcèbe au cours des changements physiologiques subis par l'animal au cours de l'hiver en collaboration avec Jérémy Terrien (MECADEV, CNRS, Brunoy) : publication dans *International Journal of Molecular Sciences.*

International Journal of
Molecular Sciences

MDPI

Article

Molecular Liver Fingerprint Reflects the Seasonal Physiology of the Grey Mouse Lemur (*Microcebus murinus*) during Winter

Blandine Chazarin [1,2,†], Margaux Benhaim-Delarbre [1,2,†], Charlotte Brun [1,2], Aude Anzeraey [3], Fabrice Bertile [1,2,*,‡] and Jérémy Terrien [3,*,‡]

[1] Laboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien, University of Strasbourg, CNRS, UMR 7178, 25 Rue Becquerel, 67087 Strasbourg, France; blandine.chazarinorgel@cshs.org (B.C.); benhaim.margaux@gmail.com (M.B.-D.); charlotte.brun@etu.unistra.fr (C.B.)
[2] Infrastructure Nationale de Protéomique ProFI—FR 2048, 25 Rue Becquerel, 67087 Strasbourg, France
[3] Unité Mécanismes Adaptatifs et Evolution (MECADEV), UMR 7179, CNRS, Muséum National d'Histoire Naturelle, 1 Avenue du Petit Château, 91800 Brunoy, France; aude.anzeraey@mnhn.fr
* Correspondence: fbertile@unistra.fr (F.B.); terrien@mnhn.fr (J.T.)
† These authors contributed equally to this work.
‡ These authors contributed equally to this work.

Abstract: Grey mouse lemurs (*Microcebus murinus*) are primates that respond to environmental energetic constraints through strong physiological seasonality. They notably fatten during early winter (EW), and mobilize their lipid reserves while developing glucose intolerance during late winter (LW), when food availability is low. To decipher how the hepatic mechanisms may support such metabolic flexibility, we analyzed the liver proteome of adult captive male mouse lemurs, whose seasonal regulations are comparable to their wild counterparts. We highlight profound hepatic changes that reflect fat accretion in EW at the whole-body level, without triggering an ectopic storage of fat in the liver, however. Moreover, molecular regulations are consistent with the decrease in liver glucose utilization in LW, and therefore with reduced tolerance to glucose. However, no major regulation was seen in insulin signaling/resistance pathways. Fat mobilization in LW appeared possibly linked to the reactivation of the reproductive system while enhanced liver detoxification may reflect an anticipation to return to summer levels of food intake. Overall, these results show that the physiology of mouse lemurs during winter relies on solid molecular foundations in liver processes to adapt fuel partitioning while opposing the development of a pathological state despite large lipid fluxes.

Keywords: season; winter; proteomics; primate; liver

check for updates

Citation: Chazarin, B.; Benhaim-Delarbre, M.; Brun, C.; Anzeraey, A.; Bertile, F.; Terrien, J. Molecular Liver Fingerprint Reflects the Seasonal Physiology of the Grey Mouse Lemur (*Microcebus murinus*) during Winter. *Int. J. Mol. Sci.* **2022**, *23*, 4254. https://doi.org/10.3390/ijms23084254

Academic Editor: Masashi Tanaka

Received: 10 March 2022
Accepted: 8 April 2022
Published: 12 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The measurement of time is a vital necessity for all living systems to be synchronized with the astronomical cycles of their environment. Each organism, from unicellular organisms to humans, harbors an endogenous biological clock, allowing it to anticipate and adapt its biological functions to daily and seasonal cycles [1]. In nature, organisms use clues from their environment as indicators of time. The variations in sunlight constitute a precise and reproducible temporal signal, which provides information on both the time of day (light intensity, light/dark cycles) and the day of year (day-length also called photoperiod). Additionally, this physical parameter was selected during evolution as the most robust external indicator of time, unlike ambient temperature and food availability, which certainly influence biological rhythms, but to a lesser extent than photoperiod [2]. Mechanistically, the integration of the photoperiod occurs through the secretion of melatonin by the pineal gland [2] and results in the generation of circannual rhythms. This feature is particularly critical in highly seasonal environments, characterized by drastic changes in

abiotic (ambient temperature, rainfall) and biotic (energy availability) factors throughout the year.

Resource/energy availability fluctuates throughout the year, alternating between periods of food abundance and scarcity. To cope with these fluctuations, seasonal organisms alternate between active and inactive reproductive states during the favorable and unfavorable periods, respectively [3]. The reproductive seasonal phenotype of alternation between periods of gonadal recrudescence and regression, to be compared to yearly puberty, is accompanied by profound changes in metabolic status, generally reinforced by the necessity to respond to harsh environmental conditions during winter. Seasonal reproduction is indeed often characterized by extreme winter metabolic phenotypes, including, e.g., migration, hibernation and daily torpor. Physiological remodeling, underlining such changing phenotypes, includes changes in lipid metabolism, cell membrane function, neural and cardiovascular functions, anorexia- and hypoxia-related responses, skeletal muscle function, immune function and non-shivering thermogenesis [4,5].

Non-model organisms or exotic species could be highly valuable, as their natural life-history traits have evolved into behavioral and physiological adjustments to repeatedly survive periods of food scarcity and abundance. Such an evolution ended with the expression of adaptive phenotypes, revealing striking features that may be of extreme importance in a clinical context [6–10]. More specifically, seasonal species undergo massive fattening to anticipate periods of partial or complete energy depletion in the field. This ability to store massive amounts of fat in anticipation of phases of poor food availability seems to be associated with the protective mechanisms against adverse effects usually associated with alterations of glucose/insulin signaling [11,12]. Although energy homeostasis relies on a complex set of mechanisms, involving multiple neuroendocrine pathways [13] and centered on hypothalamic control [14,15], it is also quite clear that peripheral tissues, including liver, pancreas, muscle and adipose tissue, play major roles in the seasonal regulation of energy homeostasis, especially at the level of nutrient partitioning [16]. However, the molecular bases of seasonal adaptation are not always fully understood, e.g., in the liver, a key tissue to metabolic flexibility [17], i.e., the capacity to switch from carbs to lipid oxidation.

The grey mouse lemur (*Microcebus murinus*) is a nocturnal lemuriform primate (Cheirogaleidae), phylogenetically close to humans and showing robust seasonal modifications in energy balance related to its endemicity to Madagascar. Relative to its body size (body mass varying between ~50 and ~90 g in wild conditions), this species has a long lifespan (half-life around 5.7 years; mean lifespan ~8–10 years in captivity [18]), which makes it a good model for ageing studies [19]. In addition, mouse lemurs undergo profound seasonal remodeling synchronized to photoperiod [20] and following annual variations in environmental conditions, including rainfalls, ambient temperature and consequently food availability [21,22]. Energy homeostasis in the mouse lemur is characterized by spontaneous transitions from high to low body mass levels throughout seasons, resulting from an exceptional capacity to adjust energy metabolism to food scarcity during winter or high food availability during summer [23]. In particular, the extensive use of daily torpor in early winter (EW) allows the mouse lemur to build up fat reserves which are later mobilized for reproduction, therefore conferring to a strong adaptive value on winter fattening. More specifically, the condition for reproductive success in the male is centered on access to females, which constitutes its major energy constraint. Therefore, male mouse lemurs must reactivate their reproductive system prior to the female receptivity period, thereby anticipating the transition from winter to summer. Male mouse lemurs' reproductive investment indeed relies on spermatogenesis and mating with competition, leading to massive energy investment in late winter (LW), when energy is not available in the environment.

The seasonal phenotype in mouse lemurs can be mimicked in captivity only with the manipulation of photoperiod (to reproduce the seasonal transitions) and is completely comparable to what is observed in wild individuals, although the environmental constraint is drastically reduced. Indeed, captive male mouse lemurs acclimated to winter-like photoperiod (10 h/day) over 6 months and maintained in constant housing conditions

(high, stable ambient temperature, food availability) exhibit deep changes in body mass explained by variations in metabolic rates, thus resulting in a biphasic phenotype with a phase of massive body mass gain in EW, followed by the spontaneous arrest of fattening in LW [11]. Strikingly, mouse lemurs are protected against glucose intolerance during fattening, but not during the second half of winter when they lose fat and exhibit increased fasting insulin levels. Most importantly, these natural transitions are fully reversible and never reach a pathological stage, as reflected by the apparent absence of an inflammatory response [11]. With regard to classical models of obesity, these results emphasize a paradox as insulin sensitivity seems to be preserved during the fattening phase. However, a similar paradox was described in hibernating bears that preserve insulin sensitivity during the fattening period through altered adipose PTEN/AKT signaling [24]. Finally, recent work has shown that the relative expression of phospho-IRS-1 was enhanced during torpor in the mouse lemur in muscle, while the phospho-insulin receptor level was decreased in white adipose, thus suggesting an inhibition of insulin/IGF-1 signaling during torpor in these tissues [25]. These results altogether show that mouse lemurs undergo profound physiological changes, underlying their ability to spontaneously reverse the fattening process. We investigated which liver pathways are activated/repressed at the protein level in the fattening (EW) vs. slimming (LW) phenotype of mouse lemurs by comparing phases of gain and loss of body mass, using a mass spectrometry-based approach. We expected major regulations of the glucose/insulin and lipid pathways, according to the physiological phenotype described in the paper by Terrien et al. [11].

2. Results

2.1. Same Body Mass but Different Dynamics

Mouse lemurs sampled during the first half of winter (Early Winter, EW) exhibited the same mean body mass as those sampled during the second half of winter (Late Winter, LW) (mean ± SEM: 104.0 ± 4.7 g vs. 101.4 ± 6.0 g in EW and LW, respectively; $t = 0.3$, $p = 0.74$; Figure 1A). However, the analysis of body mass variations revealed striking differences between EW and LW. Indeed, animals sampled in EW (i.e., 6 weeks after winter onset) were actually undergoing massive fat accretion and had gained 55 ± 11% BM during the first half of winter (Figure 1B). In comparison, mouse lemurs sampled in LW had lost 37 ± 5% BM, therefore, showing strong fat mobilization during the second phase of winter ($t = 6.8$, $p < 0.001$; Figure 1B), thus confirming that animals in EW and LW were undergoing very distinct body mass dynamics. In this line, mouse lemurs from the LW group showed lower circulating levels of triglycerides (TG; −36%, $p = 0.001$) but greater levels of total cholesterol (CHOL) and low-density lipoprotein (LDL) (+33%, $p = 0.001$ and +51%, $p = 0.0001$, respectively) and a tendency for a higher CHOL/HDL ratio (+32%, $p = 0.102$) than EW animals, while glycaemia and high-density lipoprotein (HDL) remained unchanged (Figure 1C). These changes showed interesting correlations with BM variations (Figure 1D): animals showing the highest body mass loss were those with the lowest circulating levels of TG ($r^2 = 0.31$, $p = 0.01$) but the highest levels of CHOL ($r^2 = 0.21$, $p = 0.05$) and LDL ($r^2 = 0.32$, $p = 0.02$).

2.2. Overview of Liver Proteomics Data

HPLC performances remained very good and stable throughout the whole experiment, as median coefficient of variation (CV) of retention times of all iRT peptides when considering all injections was 0.7%. The reproducibility of quantitative data was also very good, with low median CVs for the raw intensity of iRT peptides among all injections (13%) and for Maxquant-derived LFQ values of all quantified proteins within each of the two experimental groups (21.6%) and in the reference sample injected repeatedly during the course of MS-based analyses (9.5%).

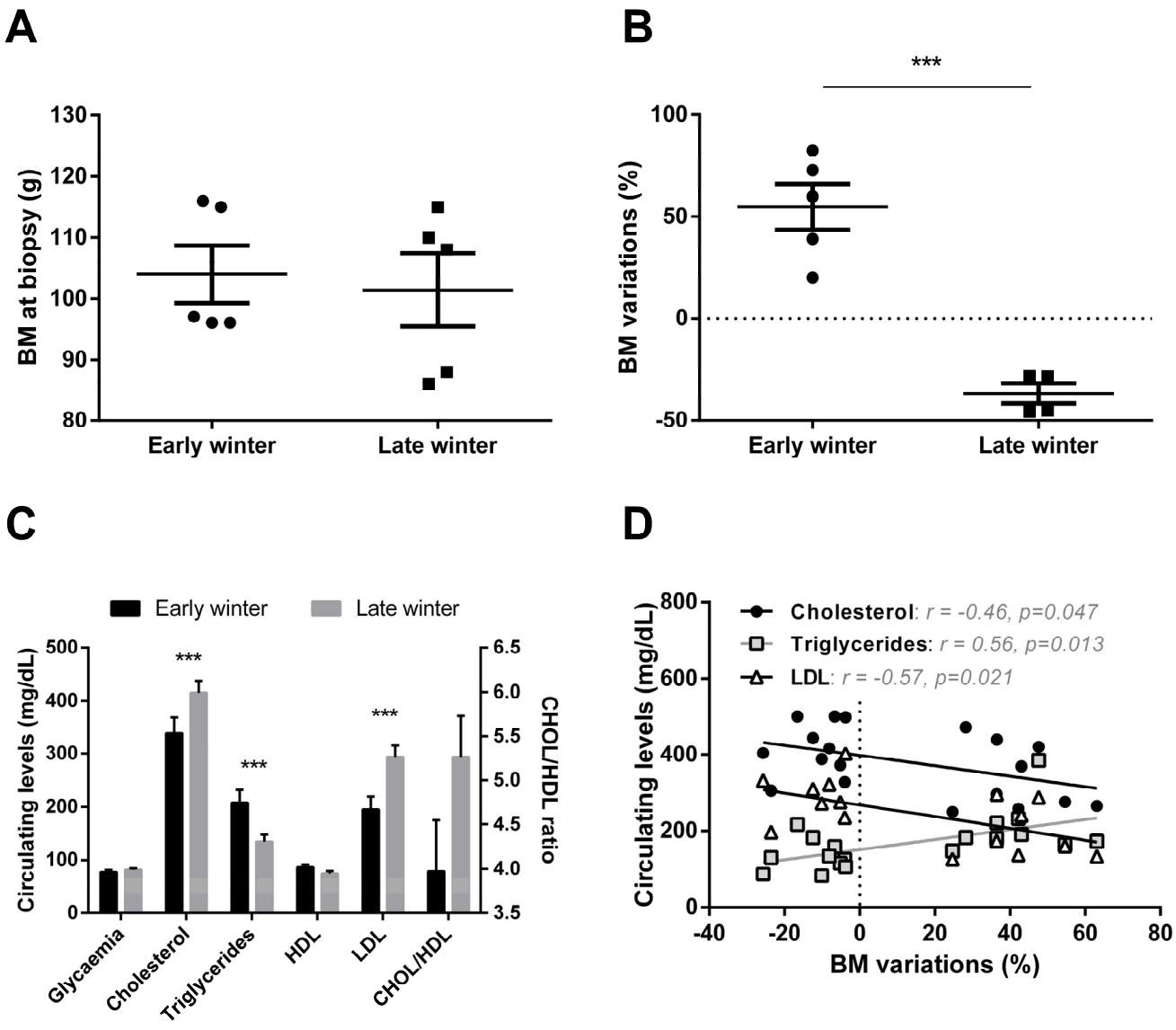

Figure 1. Body mass and plasma profiles in grey mouse lemurs during winter. Although mouse lemurs sampled in early (EW) and late (LW) winter (N = 5/group) exhibited similar body masses (BM) at the time of liver biopsy (panel **A**), variations during the first vs the second half of winter clearly showed BM gain in EW but BM loss during LW (panel **B**). Changes in BM dynamics were paralleled by changes in plasma metabolites and lipoprotein levels (panel **C**, data are presented as the means ± SEM), the relationship between these parameters being significant in some cases (panel **D**). *** indicate significance at $p < 0.0001$ between EW and LW.

Overall, the identification of 4964 proteins was robust enough to comply with the stringent filtering criteria that were applied (see above), and 4522 of these proteins were retained for MS1 intensity-based label-free quantification after the elimination of those exhibiting a single unique peptide (Table S1). Based on the expression of these 4522 proteins, very good discrimination could be observed between the two biological groups using a PCA analysis (Figure 2A). Indeed, animals from the EW and LW groups discriminated significantly around dimension 2 ($r^2 = 0.42$, $p = 0.04$), which was mostly explained by the regulations of 759 proteins, amongst which GAPDH and LDHA (involved in carbohydrate metabolism), MAN1B1 (involved in protein metabolism) and NSDHL (involved in lipid metabolism) were included in the 30 most contributory proteins (Figure 2A, Table S2).

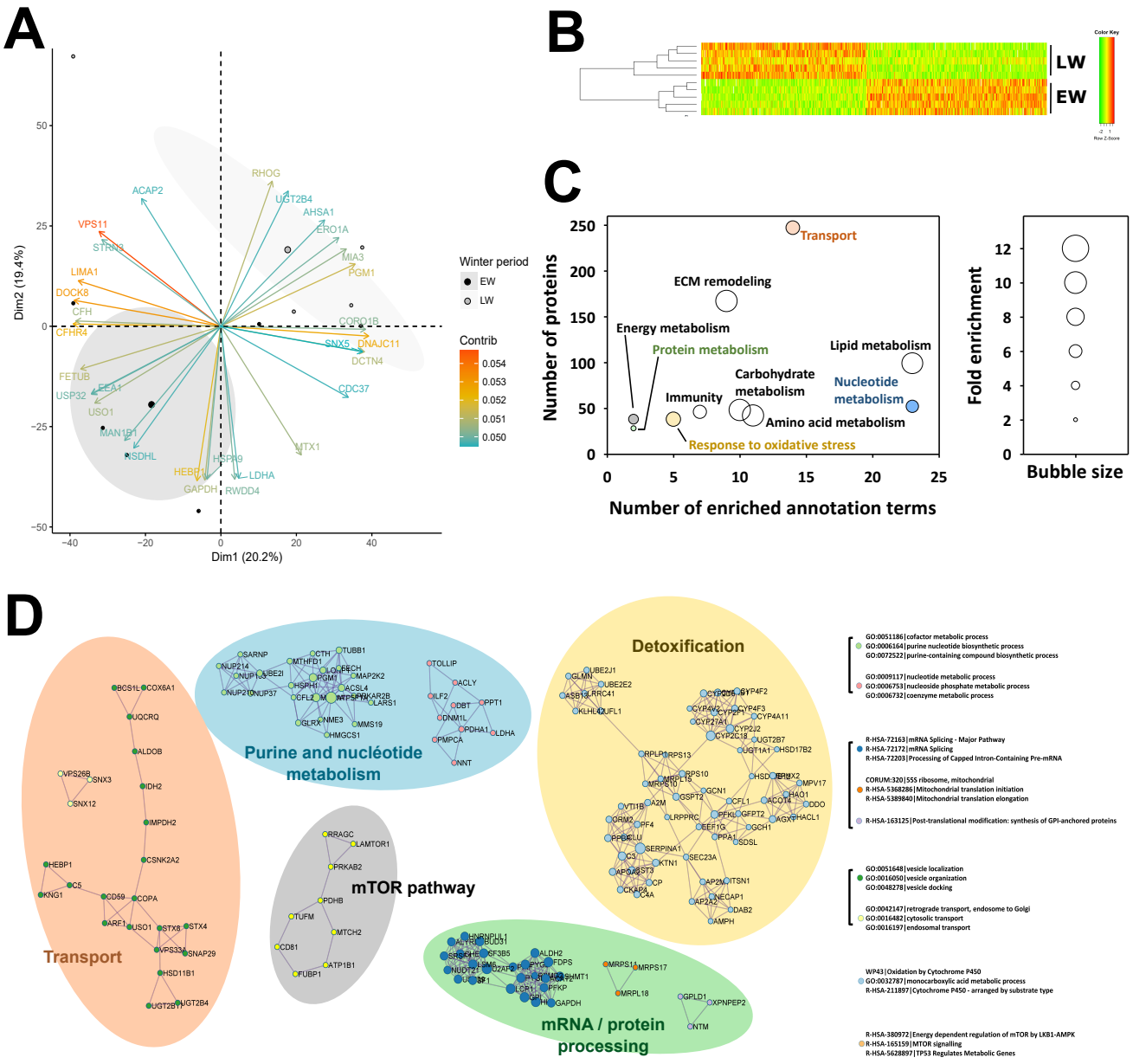

Figure 2. Overview of liver proteomics data. Principal component analysis performed from expression levels of all 4522 proteins analyzed using mass spectrometry (panel **A**; the 30 most contributing proteins are represented) and hierarchical clustering performed from expression levels of all 538 proteins found differentially expressed between early (EW) and late (LW) winter (panel **B**) allowed these two conditions to be well discriminated. Functional annotation analysis performed from the list of the 538 differential proteins revealed enriched ontology terms, which allowed the determination of broad functions significantly affected between EW and LW. Complementary "functional information" was obtained using the algorithm from Ease/DAVID (panel **C**) and that from Metascape (panel **D**). Colors in (panels **C**,**D**) indicate overlapping information. Detailed protein abundances and fold changes are given in Table S1. ECM: extracellular matrix.

The abundance of 534 proteins was significantly changed between EW and LW, 288 of them being higher in the LW vs. the EW period, while the reverse was observed for the 246 remaining proteins. Three additional proteins (NCF1, HSD11B2, TMX2) were not detected in EW animals and one (CYP2C18) in LW. Also, PCA analysis (data not shown, PC1 at 59%) and hierarchical clustering (Figure 2B) showed a distinct discrimination of the two biological groups, confirming the low variability between biological replicates.

Two different algorithms for functional annotation enrichment analysis from all differentially abundant proteins (including proteins not detected in a given phase) brought similar and complementary information. Using the EASE/DAVID algorithm, a total number of 106 GO terms/Kegg maps were found significantly enriched, which allowed highlighting ten processes, including lipid/carbohydrate/amino-acid metabolisms and the remodeling of extracellular matrix (ECM) as the most enriched (Figure 2C). Enrichment in annotation terms was also observed, to a lesser extent, for other functions like energy, nucleotide and protein metabolisms, transport, response to oxidative stress and immune response. Using Metascape/Cytoscape, 25 annotation terms were found significantly enriched, corresponding to differential proteins distributed in nine interaction networks that define five functional categories, including detoxification, nucleotide metabolism, transport, mRNA processing and protein metabolism, and, finally, the mTOR pathway (Figure 2D).

2.3. Specific Analysis of Significantly Enriched Pathways

2.3.1. Protein Metabolism

A high number of liver proteins involved in protein metabolism were detected. Comparing their abundance in LW vs. EW, a *p*-value lower than 0.05 was calculated for sixty of them. Concerning protein synthesis, a higher abundance of six proteins (SF1, SUPT4H1, SUPT5H, TCEA1, U2AF2, and USP39; +20 to +93%, median = +27%) in LW may support an overall higher transcriptional activity at this stage, while the reverse could be interpreted for lower abundance of two proteins (LSM6 and SF3B5; −4% and −13%, respectively). Protein translational process also appeared regulated during winter, with nine proteins more abundant (GCN1, MRPL15, MRPL18, MRPS10, MRPS11, MRPS17, RPLP1, RPS10 and RPS13; +9 to +62%, median = +28%) and twelve others less abundant (CARS, CSNK2A2, EEF1G, EIF5, GSPT2, GATB, LARS, MTG1, OSGEP, RPS6KA1, TUFM and UPF1; −12 to −45%, median = −20%) in LW versus EW. Finally, adjustments in the process of protein folding seems to be needed, depending on winter timing, with the abundance of one protein being higher (DNAJB11, +44%) and that of three others being lower (MAN1B1, NUDCD2, and OS9; −18 to −32%, median = −25%) in LW versus EW animals.

Concerning protein catabolism, numerous changes were recorded for the ubiquitin proteasome system, with eight proteins higher expressed (ATG3, GLMN, MARCH5, NUB1, PSMB10, PSME3, UBE2E2, UBE2J1, +29 to +90%; median = +40%) and eleven others lower expressed (ASB13, CUL4, KLHL42, LRRC41, PSMD2, PSMG2, PSMG3, PSMG4, UCHL5, UFL1, and USP32; −12 to −60%, median = −27%) in LW versus EW lemurs. With three proteins higher expressed (BNIP3, RMC1, and SNAP29; +43 to +61%, median = +60%) and only one lower expressed (NRBF2; −30%) in LW, liver autophagy may appear enhanced at this stage. Finally, expression levels of other liver intracellular proteases were found higher (DNPEP; +28%) and lower (NRDC, XPNPEP2 and XPNPEP3; −23 to −36%, median = −32%) in LW compared to EW.

Changes in the abundance of proteins controlling protein metabolism were reinforced by the observation of 46 additional regulated proteins when the significance level was set between 0.05 and 0.1 (see Table S1 for details). Among them, upregulated proteins in LW vs. EW included three proteins involved in transcription (ELP2, RTRAF and SUPT6H; +5 to +21%, median = +19%), twelve in translation (ATE1, BRIX1, EEF1E1, EIF1B, MRPL17, MRPL19, MRTO4, NOP58, PARS2, RARS2, RPS17, and WARS; +12 to +69%, median = +29%), one in protein folding (AUP1, +13%), seven in the ubiquitin proteasome system (DTX3L, OTULIN, PSMA1, PSMB9, UBE2G2, UBE3C, USP19; +15 to +128%, median = +38%) and one in autophagy (SQSTM1, +62%). Downregulated proteins in LW vs. EW grouped together one protein involved in transcription (SNU13, −11%), nine in translation (AARSD1, EIF3A, EIF3J, MRPL2, MRPS21, RPL34, RPS27L, RPS28, and SBDS; −5 to −47%, median = −20%) and twelve in the ubiquitin proteasome system (BRCC3, CDC34, FBXO21, PSMD7, PSMD9, PSMD13, RPS27A, UBAP2L, UFD1, UFM1, USP10 and ZNF598; −7 to −40%, median = −20%).

2.3.2. Immune System, Inflammatory Response and Apoptosis

Of the 538 regulated proteins between EW and LW, 26 were associated with inflammatory and immune function and 8 with apoptosis. More specifically, regulated proteins involved in the inflammatory process were very few (ITIH4, $p = 0.02$; SELENOP, $p = 0.005$; and FETUB, $p = 0.02$) but were all three downregulated between EW and LW (respectively −33, −22 and −80%). The levels of three of the eight proteins involved in apoptosis were lower in LW vs. EW (−24 to −52%; IER3IP1, $p = 0.03$; PDCD5, $p = 0.02$ and S100A14, $p = 0.001$), while the other five were upregulated between the EW and LW periods (+33, to +114%, median = +57%). Furthermore, 24 proteins involved in the immune system were significantly regulated in this dataset (12 proteins downregulated, −22 to −65%, median = −35%; 11 proteins upregulated, +21 to +119%, median = +66%; NCF1 not detected in EW). Interestingly, the complement cascade seemed to be clearly downregulated as seven proteins involved in this process were found in significantly lower abundance in LW vs. EW (−22 to −65%, median = −35%; CD59, $p = 0.005$; C4A, $p = 0.03$; C1QTNF9, $p = 0.04$; C5, $p = 0.004$; C8A, $p = 0.009$; C3, $p = 0.03$; C1RL, $p = 0.02$). This pattern was confirmed at a less stringent significance threshold ($0.05 < p < 0.1$), where five additional proteins were downregulated while none of the proteins detected were upregulated (CFH, −40%, $p = 0.087$; CFI, −42%, $p = 0.078$; CFB, −49%, $p = 0.053$; C9, −30%, $p = 0.075$).

2.3.3. Carbohydrate Metabolism and Insulin Signaling

Out of the 21 significantly regulated proteins classified in the carbohydrate metabolism pathway (Figure 3A,B), 14 were downregulated in LW vs. EW (−15 to −56%, median = −37%), suggesting an overall reduction in carbohydrate metabolism in LW. More specifically, glucose uptake in hepatocytes did not seem to be altered as SLC2A2 expression was not changed between EW and LW ($p = 0.619$). For glycolysis and gluconeogenesis processes, proteins such as GAPDH (−15%, $p = 0.024$), GPI (−27%, $p = 0.021$), MINPP1 (−30%, $p = 0.020$), ALDOB (−37%, $p = 0.001$), PFKP (−19%, $p = 0.044$), PFKL (−31%, $p = 0.022$) or FBP1 (−32% as an average of two isoforms, $p = 0.012$ and $p < 0.0001$) were all showing lower levels in LW versus EW. Conversely, HK1 appeared upregulated in LW vs. EW (+75%; $p = 0.008$). Interestingly, animals exhibiting the highest expression levels of FBP1 corresponded to those showing the greatest fattening phenotype (positive correlation between BMC and FBP1 levels, $p = 0.01$; Figure 3C). In addition, proteins involved in anaerobic glycolysis (LDHA: −51%, $p = 0.017$ for one isoform), fructose biosynthesis (SORD, three isoforms: median = −51%, $p = 0.003$ to $p = 0.017$), xylose metabolism (XYLB: −18%, $p = 0.046$) and glycerol degradation (GPD1: −38%, $p = 0.028$; GPD2: −39%, $p = 0.013$) were also downregulated between EW and LW, except for AQP9, which was upregulated (+29%, $p = 0.025$). In contrast, proteins involved in glycogen metabolism, including PGM1 (two isoforms, +85% in average, $p = 0.019$ and $p = 0.036$) and PYGB (brain form; +31%, $p = 0.046$) were upregulated in LW compared to EW (except for PYGL: liver form; −16%, $p = 0.026$). A less stringent analysis ($0.05 < p < 0.1$) confirmed regulations of proteins involved in glycolysis and gluconeogenesis processes (ALDOA, +30%; $p = 0.052$; GCK, +7%; $p = 0.089$; and PGK1, +38%, $p = 0.098$), glycerol degradation (GK, +38%, $p = 0.055$) and anaerobic glycolysis (second isoform of LDHA: −17%; $p = 0.08$).

In our dataset, no clear regulation of the insulin signaling pathway was detected in the liver between EW and LW. Protein levels of INSR remained unchanged (+17%, $p = 0.86$), although we observed a positive correlation between RNA and protein levels ($p = 0.04$, Figure 3D). On another hand, IRS1 tended to be increased in LW animals (+30%, $p = 0.079$), this upregulation being potentially sustained by reduced levels of MAPK9 (−27%, $p = 0.004$), map kinases acting as repressors of IRS1 [26]. In accordance with the decreased utilization of glucose in LW, suggested by the proteomic data, we observed a greater intolerance to glucose in LW than in EW animals, resulting from an oral glucose tolerance test (oGTT) (Figure 3E). Interestingly, while circulating basal levels of insulin remained unchanged between EW and LW (data not shown), animals exhibiting the highest insulinemia during

EW were the same animals showing the best regulation of glycaemia after oGTT (r^2 = 0.42, p = 0.057; Figure 3F); this correlation was lost in LW (p = 0.39; Figure 3F).

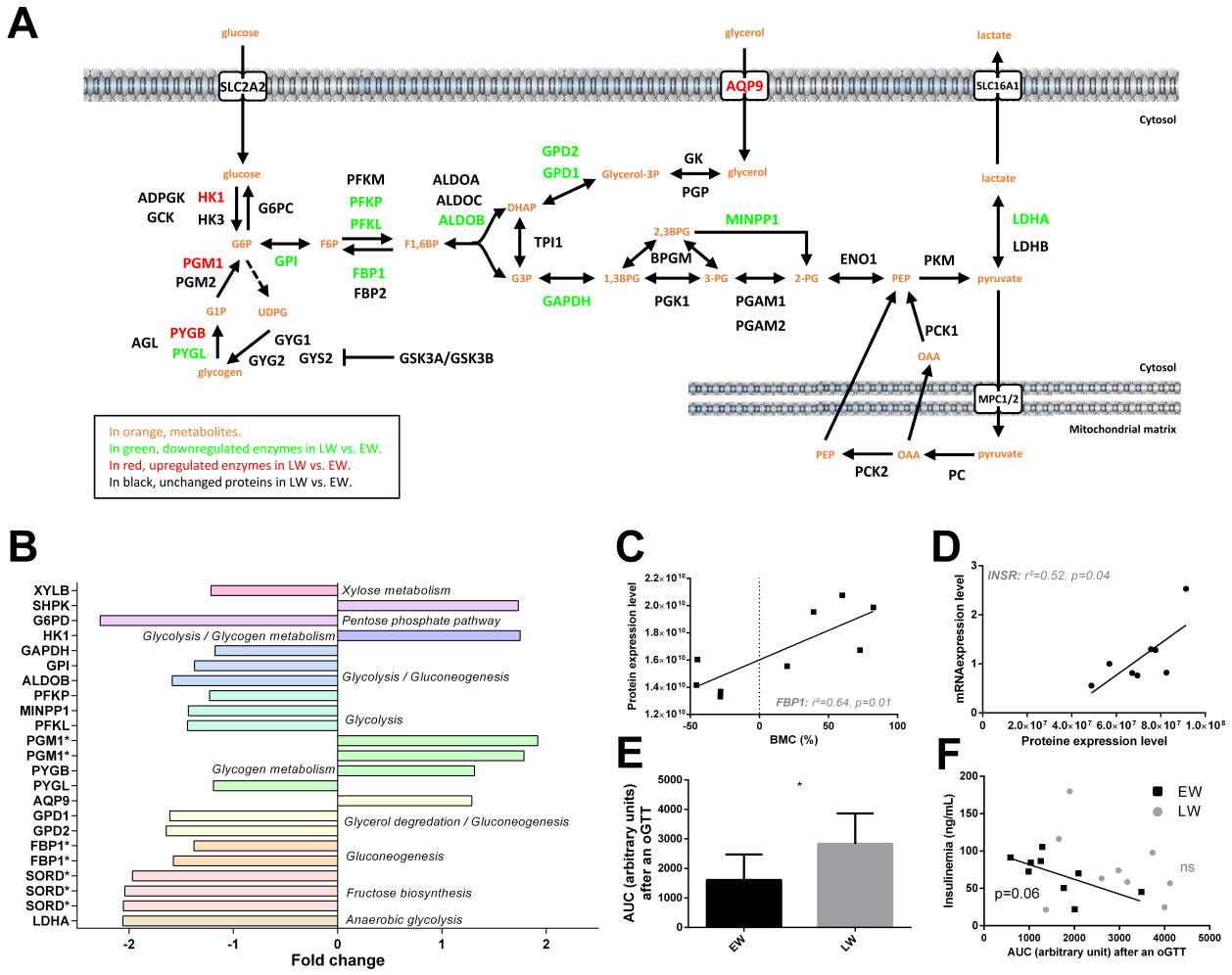

Figure 3. Regulation of liver carbohydrate metabolism and insulin signaling in grey mouse lemurs during winter. The relative abundance of liver proteins involved in carbohydrate metabolism measured in mouse lemurs (N = 5/group) in early (EW) vs late (LW) winter is shown using a color code in the corresponding pathways (panel **A**), and mean fold changes in LW vs. EW can be visualized (panel **B**) (presence of several isoforms of the same protein are indicated with asterisks (*)). Significant and positive correlations were found between protein expression level of fructose-1,6-bisphosphatase 1 (FBP1) and body mass changes (BMC, panel **C**) and between mRNA and protein levels of the insulin receptor (INSR, panel **D**). Retrieving data from a previous study showing the most extreme phenotypes (EW vs. LW) revealed a greater intolerance to glucose in LW than in EW during oral glucose tolerance tests (oGTT, panel **E**; results are presented as the means ± SEM) and a negative correlation between insulinemia and glucose intolerance (AUC, area under the curve) for EW lemurs, albeit not a significant one (panel **F**). * indicates significance at $p < 0.05$ between EW and LW.

2.3.4. Lipid Metabolism

Of the 47 significantly regulated proteins classified in the "lipid metabolism pathway", the abundance of 24 of them was increased (+20 to +923%, median = +49%) and that of 23 others was decreased (−6 to −68%, median = −42%) during the EW period compared to the LW period (Figure 4A). Overall, liver lipid or fatty acid synthesis seemed to be slightly downregulated towards the end of winter (Figure 4A) as 55% of the proteins involved in these processes were decreased (e.g., HTD2, −28%, p = 0.036; SCD, −81%, p = 0.001; MCAT, −20%, p = 0.003; HSD17B12, −33%, p = 0.006; PHOSPHO1, −68%, p = 0.008; HMGCS1,

−66%, $p = 0.026$; FDPS, −56%, $p = 0.006$); altough, ACSL5 (+153%, $p = 0.033$) and ACOT4 (+56%, $p = 0.034$), which are specifically involved in the biosynthesis of unsaturated fatty acids (FA), were strongly upregulated in LW animals (Figure 4A). In this respect, the metabolism of arachidonic acid (C20:4 n-6) seemed to be altered during the winter period as CYP4F2 (+99%, $p = 0.019$) and EPHX2 (+196%, $p < 0.0001$) were highly and significantly upregulated in LW. No regulation was observed in the alpha-linolenic acid metabolism pathway. Additionally, apart from NSDHL (−61% in LW vs. EW), proteins involved in liver CHOL (RALY, +29%, $p = 0.030$) and bile acid (SLC27A5, +108%, $p = 0.026$; CYP27A1, +38%, $p = 0.001$; and HSD17B4, +21% on average for two isoforms, $p = 0.017$ and $p = 0.029$) synthesis were upregulated from the EW to LW. A less stringent analysis confirmed the regulations of proteins involved in bile acid synthesis (ACOT8, −12%; BAAT, −22%; HSD3B7, +12%; ABCB11, +87%; SULT2A2, +205%).

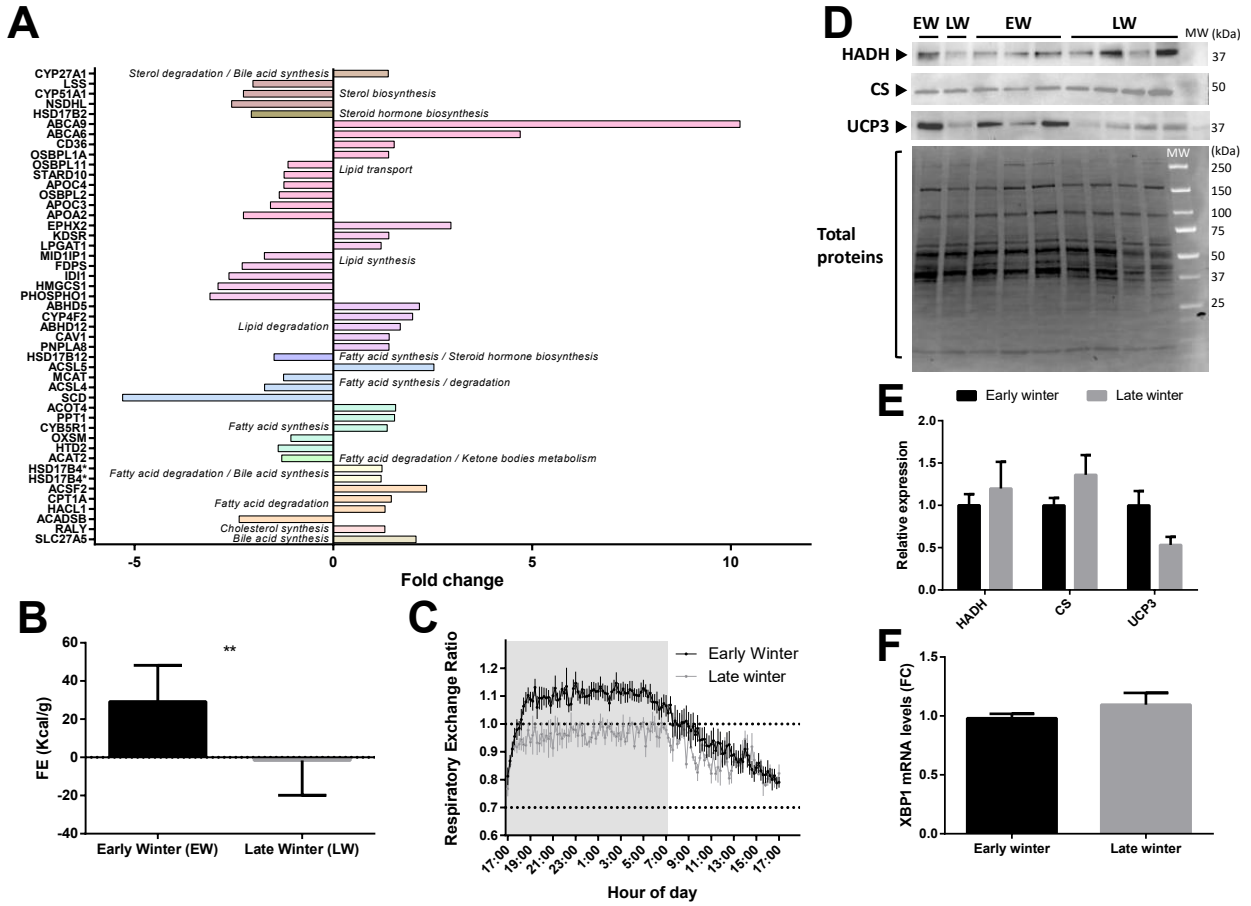

Figure 4. Regulation of liver lipid metabolism in grey mouse lemurs during winter. The relative abundance of liver proteins involved in carbohydrate metabolism in early (EW) and late (LW) winter mouse lemurs (N = 5/group) is shown using mean fold changes in LW vs EW (panel **A**; presence of several isoforms of the same protein are indicated with asterisks (*). Food efficiency (FE), calculated as the ratio between food intake (in Kcal/day) and BMC were significantly different between EW and LW animals (panel **B**), and the respiratory exchange ratio measured in EW (black line) vs. LW (grey line) reflected the main energy substrates used, depending on time-of-day (panel **C**). Protein expression levels of mitochondrial hydroxyacyl-coenzyme A dehydrogenase (HADH), citrate synthase (CS) and uncoupling protein 3 (UCP3) were measured using Western-blot analysis in liver samples from mouse lemurs in EW and LW (blots shown in panel **D** and relative expression in (panel **E**); N = 4–5/group; results are presented as the means ± SEM). XBP1 gene expression levels did not differ in the liver of adult male mouse lemurs between EW and LW (panel **F**). * and ** indicate significance at $p < 0.05$ and 0.01 between EW and LW, respectively.

In contrast, most of the proteins (77%; except ACADSB, −58%, $p = 0.003$ and ACAT2, −23%, $p = 0.008$) involved in FA or lipid degradation, including ABHD5 (+116%, $p = 0.037$), ACSF2 (+134%, $p = 0.034$), PNPLA8 (+40%, $p = 0.020$) and CPT1A (+46%, $p = 0.040$), were quantified as significantly increased in LW relative to EW. ACOX1, which is involved in the first step of beta-oxidation, was not significantly regulated at the protein level ($p = 0.348$) but showed a strong negative correlation between mRNA and protein expression levels ($r^2 = 0.71$, $p = 0.008$).

FA transport regulation was not regulated in a clear way, as some proteins (4 out of 10) were upregulated (ABCA9, +923%, $p = 0.039$; ABCA6, +370%, $p = 0.042$; CD36, +53%, $p = 0.026$; OSBPL1A, +39%, $p = 0.004$) and some others (6 out of 10) were downregulated (e.g., APOA2, −56%, $p = 0.038$; APOC3, −37%, $p = 0.035$; OSBPL2, −27%, $p = 0.001$; APOC4, −20%, $p = 0.042$).

Observations at the proteome level accords well with variations observed at the physiological level in the individuals exhibiting the most extreme fattening (EW) and leaning (LW) phenotypes (see Methods). Indeed, animals in LW actually ate less than in EW ($p = 0.0003$), and food efficiency was dramatically decreased to negative values (Figure 4B) and went along with a decrease of respiratory exchange ratio close to 0.7 (Figure 4C), therefore reflecting increased lipid mobilization from fat stores. Lipid oxidation in LW was not statistically supported by changes in liver HADH and CS (Figure 4D,E), although both proteins tended to be increased in LW vs. EW (+20%, $p = 0.61$ and +36%, $p = 0.23$, respectively). However, animals showing the greatest body mass loss (corresponding to animals in LW) tended to be those exhibiting the greatest levels of CS ($r^2 = 0.34$, $p = 0.13$). Finally, mRNA levels of the transcription factor Xbp1 were not modified between EW and LW (Figure 4F), and the protein was not detected in the proteomic analysis.

2.3.5. Energy Metabolism

Among the 538 significantly regulated proteins between EW and LW, 30 proteins were classified in the 'Energy metabolism' pathway, including 13 upregulated (+15 to +111%, median = +23%) and 17 downregulated proteins (−55 to −12%, median = −21%) in LW vs. EW. A rapid overview of the energy metabolism pathway (Figure 5A) highlighted a possible downregulation of both the tricarboxylic acid (TCA) cycle and mitochondrial respiration, in LW vs. EW. The expression of six out of eight proteins involved in TCA cycle (Figure 5B) was decreased in LW relative to EW. Indeed, in addition to ACLY (−62%, $p = 0.040$) and SDHD (−18%, $p = 0.037$), the expression levels of pyruvate dehydrogenases (PDHA1: −36%, $p = 0.001$; PDHB: −36%, $p = 0.002$; PDHX: −25%, $p = 0.002$), as well as that of PDK1 (−54%, $p < 0.0001$), which is known to inhibit pyruvate dehydrogenase and therefore to limit feeding of the TCA cycle with glycolytic intermediates, were lower in LW vs. EW. By contrast, IDH2, which is involved in the interconversion of oxalosuccinate into either isocitrate or 2-oxo-glutarate, and SDHAF1, which is involved in the assembly of succinate dehydrogenase, were upregulated between EW and LW (+59%, $p = 0.011$, and +10%, $p = 0.037$, respectively).

Proteins involved in mitochondrial oxidative phosphorylation (OXPHOS) showed a tendency to be downregulated in LW vs. EW animals (Figure 5A,B, 7 out of 13 regulated proteins at $p < 0.05$). In parallel, even if oxygen consumption remained unchanged between EW and LW (Figure 5C), heat production was significantly decreased in LW compared to EW (Figure 5D). Additionally, NDUFB7, a subunit of OXPHOS complex I, was significantly upregulated in EW animals (+27%, $p = 0.024$). As already mentioned, complex II did not seem either upregulated nor downregulated between EW and LW, as SDHD decreased (−18%), while SDHAF1 increased (+10%). Two proteins constitutive of cytochrome c reductase (Complex III: UQCRQ, −17%, $p = 0.013$) and cytochrome c oxidase (Complex IV: COX6A1, −34%, $p = 0.010$) showed significantly decreased levels in LW vs. EW. In addition, complex III assembly factors were either significantly downregulated (Complex III: BCS1L: −23%, $p = 0.036$; Complex IV: COX17, −49%, $p = 0.022$) or significantly upregulated (Complex III: UQCC2, +35%, $p = 0.003$). Finally, one catalytic subunit of ATP

synthase (Complex V) was significantly downregulated in LW vs. EW (ATP5F1A, −14%, $p = 0.028$), whereas the complex V assembly factor PPIF was upregulated (+57%, $p = 0.044$). Interestingly, PPA1, a protein involved in the fueling of ATP synthase with Pi (inorganic Phosphate) was significantly increased in LW vs. EW animals (+20%, $p = 0.048$). In addition, UCP3 was not detected in the proteomic analysis, but showed a significant downregulation in LW in the Western blot analysis (−47%, $p = 0.04$; Figure 4D,E). More generally, enzymes involved in mitochondrial processing peptidase (PMPCA: −13%, $p = 0.011$; PMPCB: −21%, $p = 0.043$) and SAMM50 (−10%, $p = 0.024$) were less abundant in LW vs. EW. Contrary to what was generally observed for mitochondrial proteins, we could also see that the expression levels of important ATP consumers located at the plasma membrane were increased in LW (ATP1B1: +50%; $p = 0.038$).

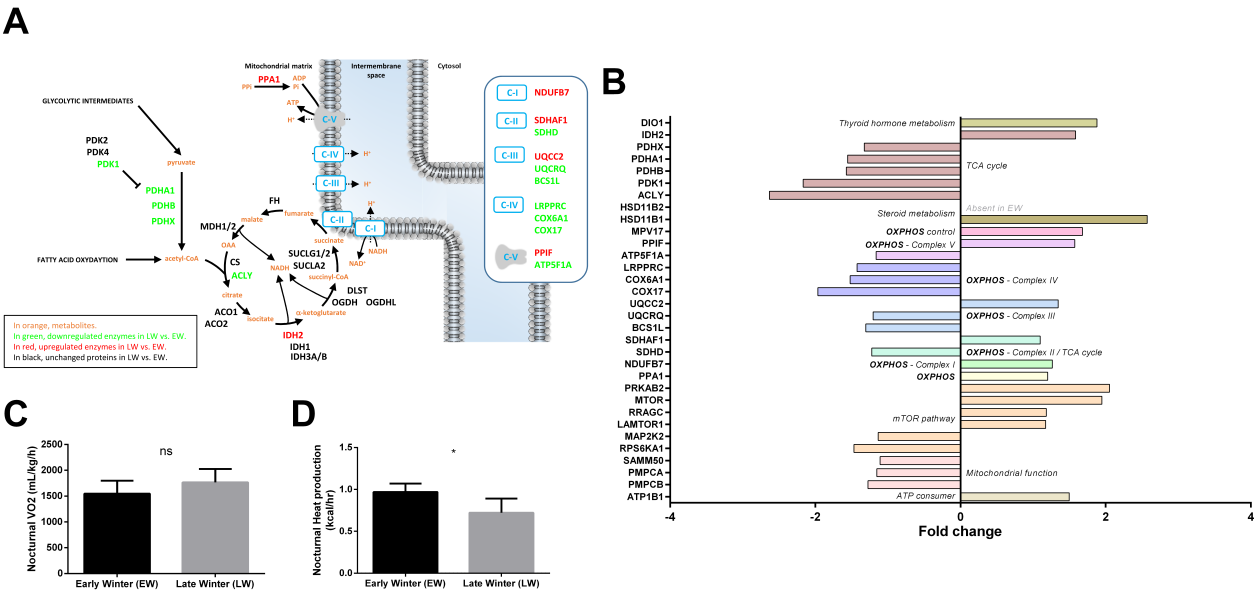

Figure 5. Regulation of liver energy metabolism in grey mouse lemurs during winter. The relative abundance of liver proteins involved in mitochondrial energy metabolism in early (EW) and late (LW) winter mouse lemurs (N = 5/group) is shown using a color code in the corresponding pathways (panel **A**), and mean fold changes in LW vs. EW can be visualized in (panel **B**) (HSD11B2 was not detected in EW). Although no change could be observed in nocturnal oxygen consumption (VO2) between EW and LW animals (panel **C**), nocturnal heat dissipation was significantly lower in LW lemurs (panel **D**). * indicates significance at $p < 0.05$ between EW and LW. ns indicates non significant.

Metabolism is regulated by key factors like FOXO, MTOR and AMPK. While FOXO was not detected in our analysis, MTOR and AMPK signaling pathways seemed to be involved. Increased levels of MTOR itself was observed in LW (+95%; $p = 0.006$). On the other hand, the mTORC1 complex could be involved as suggested by the upregulation of the AMPK subunit PRKAB2 (+105%, $p = 0.002$) and RRAGC (+18%, $p = 0.025$). The increase in RRAGC abundance in LW was paralleled by increased expression of LAMTOR1 (+17%, $p = 0.019$), which is involved in amino acid sensing.

Finally, thyroid hormones and steroids are involved in key biological processes, including energy metabolism. While thyroid hormone signaling seemed to be increased in LW relative to EW animals (DIO1, +88%, $p = 0.013$), steroid metabolism was overall enhanced at multiple levels. Indeed, HSD11B1 (+157%, $p = 0.038$) and HSD11B2 (non-detected in EW), which are involved in cortisol synthesis and inactivation, were both significantly upregulated in LW animals. Less stringent statistical analysis confirmed significant regulations of TCA cycle (ACO1: −16%, $p = 0.052$; MDH2: −17%, $p = 0.077$; and IDH3B: −21%, $p = 0.079$), OXPHOS (14 out of 21 proteins to be downregulated at $p < 0.1$ including NDUFS5, NDUFA4, NDUFA2, SDHC, UQCRB, COX6B1 and COX7A2; Table S1) or ATP use (ATP1B3: +66%; $p = 0.081$). In addition, the 30% increase in IRS1 levels ($p = 0.079$), combined with the 18%

decrease in PRKCA levels ($p = 0.089$) in LW vs. EW did not allow the determination of whether mTORC2 may play a role in mouse lemurs during winter. In contrast, regulations of RHEB (+10%, $p = 0.085$) and RRAGA (+20%, $p = 0.076$) tended to confirm a role of mTORC1 in the transition from EW to LW.

2.3.6. Amino Acid Metabolism

We observed changing levels for liver enzymes involved in the metabolism of several amino acids (Figure 6A). The liver is indeed important for directing the use of amino acids towards carbohydrate and/or lipid anabolic pathways. Notably, we observed a general lower abundance of many enzymes involved in pathways that make use of amino acids for glucose neosynthesis in LW vs. EW animals (−21 to −60%; SDSL: −35%, $p = 0.024$; AGXT, −23%, $p = 0.030$; MAT1A, −24%, $p = 0.014$; BHMT, −45%, $p = 0.025$; AMDHD1, −41%, $p = 0.0009$; HAL, −57%, $p = 0.001$; CTH, −34% for one isoform, $p = 0.032$; GLUD1, −28%, $p = 0.013$; DDO, −63%, $p = 0.002$; DMGDH, −46%, $p = 0.011$). Only one enzyme of these latter pathways was higher abundant in LW (SHMT1, +30%, $p = 0.007$). The enzymes involved in the synthesis of both carbohydrate and ketone bodies from amino acids manifested opposite regulations in LW (ACMSD, +122%, $p = 0.011$; GCDH, −30%, $p = 0.042$; ACAT2: −23%, $p = 0.008$). Several enzymes involved in ammonia detoxification and the urea cycle were found differentially expressed, notably, GLUL (+197% in LW, $p = 0.038$) and ASS1 (−24% in LW, $p = 0.015$). Finally, the significantly higher abundance of four other important proteins linked to amino acid metabolism was observed in LW (Figure 6A), including SMS (+50%, $p = 0.016$), GRHPR (+26%, $p = 0.027$), SAT2 (+41%, $p = 0.025$) and DBT (+20%, $p = 0.0009$); one additional protein was less abundant in LW (ALDH18A1, −33%, $p = 0.003$).

Additional proteins were regulated in LW compared to EW, when the p-value threshold was increased to $p < 0.1$, notably, several enzymes that regulate amino acid metabolism (CSAD, +28%, $p = 0.082$; BCAT2, +9%, $p = 0.089$; HPD, +29%, $p = 0.059$; IVD, −45%, $p = 0.053$), gluconeogenesis (PIPOX, +60%, $p = 0.073$; and one isoform of CTH, −24%, $p = 0.085$) and NAGS (−22% in LW, $p = 0.051$).

2.3.7. Response to Oxidative Stress

The regulation of protein expressions involved in the response to free radicals did not clearly indicate changes in levels of liver oxidative stress between LW and EW (see Table S1). Indeed, proteins that favor free radical or hydrogen peroxide or nitric oxide production exhibited either higher (HAO1: +31%; $p = 0.031$; and NCF1, only detected in LW) or lower (NOS3: −30%, $p = 0.037$) expression levels in LW vs. EW animals (Figure 7). Similarly, several antioxidant enzymes were found at higher expression levels in the liver during LW (VKORC1L1, +49%, $p = 0.032$; and TMX2, uniquely detected in LW), whereas other proteins were determined as less expressed (TXN, −30%, $p = 0.030$) (Figure 7). Additionally, two proteins involved in the glutathione system were less expressed in LW (GGT1, −67%, $p = 0.012$; GLRX, −49%, $p = 0.028$). Regarding heat shock proteins, which are oxidative stress-responsive proteins, higher levels during LW were recorded for two of them (AHSA1, +64%, $p = 0.001$; HSPH1, +43%, $p = 0.048$), and lower levels for three of them (ST13, −25%, $p = 0.012$; TRAP1, −20%, $p = 0.007$; HSPA9, −9%, $p = 0.038$) (Figure 7).

Additional changes were highlighted by using higher p-values ($0.1 < p < 0.05$), with higher expressed proteins in LW, including antioxidant enzymes (CAT, +24%, $p = 0.067$; PRDX6, +31%, $p = 0.058$; GCLC, +21%, $p = 0.055$) and one heat shock protein (HSPA12A, +30%, $p = 0.071$). In the same way, data were in line with a lower expression level in LW for other antioxidant enzymes (HMOX2, −20%, $p = 0.087$; PRDX4, −28%, $p = 0.052$), as well as glutathione system factors (MGST1, −15%, $p = 0.066$; GSTA5, −41%, $p = 0.088$) and one chaperone protein (P4HB, −15%, $p = 0.087$).

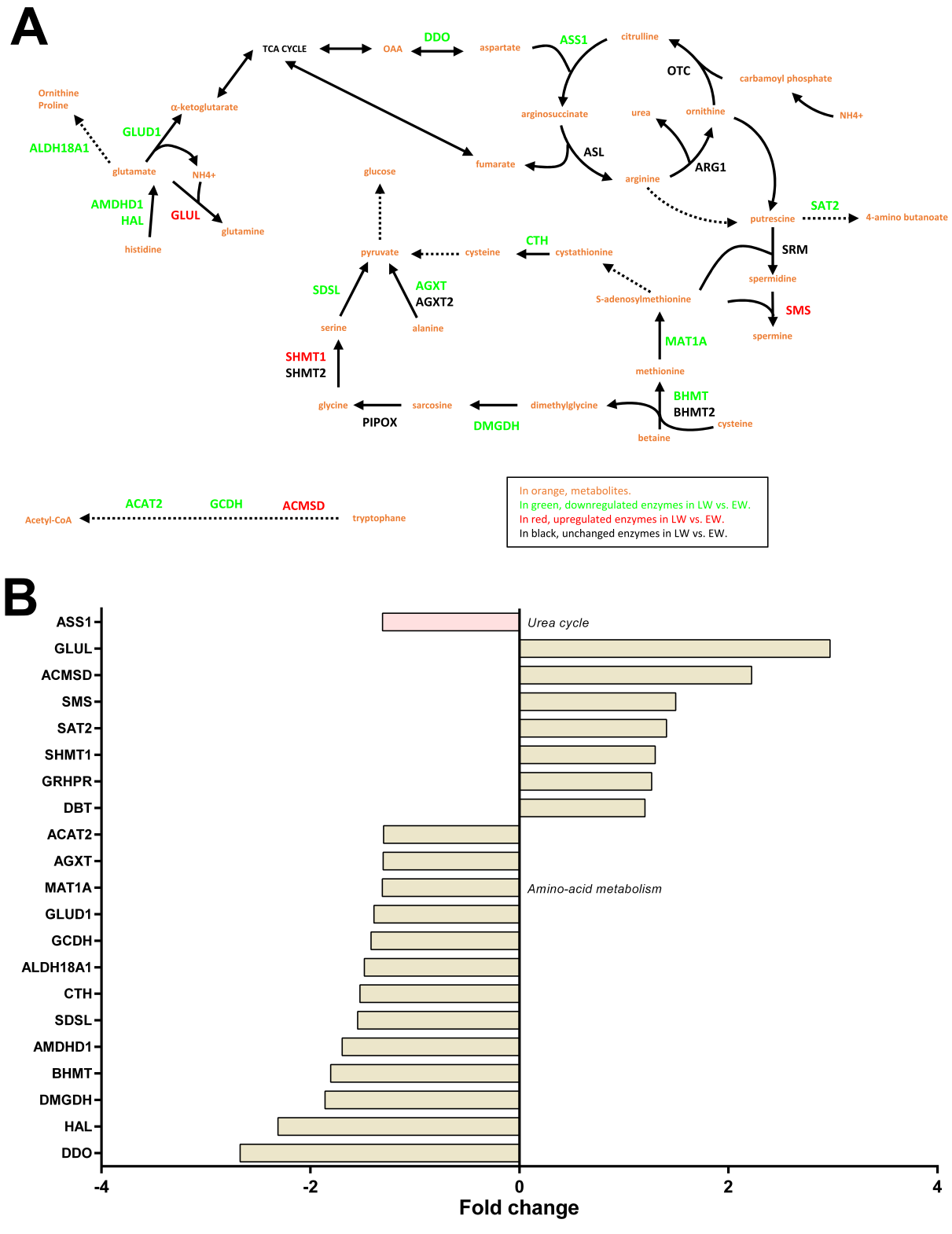

Figure 6. Regulation of liver amino acid metabolism and the urea cycle in grey mouse lemurs during winter. The relative abundance of liver proteins involved in amino acid metabolism and the urea cycle in early (EW) and late (LW) winter mouse lemurs (N = 5/group) is shown using a color code in the corresponding pathways (panel **A**), and mean fold changes in LW vs. EW can be visualized in (panel **B**).

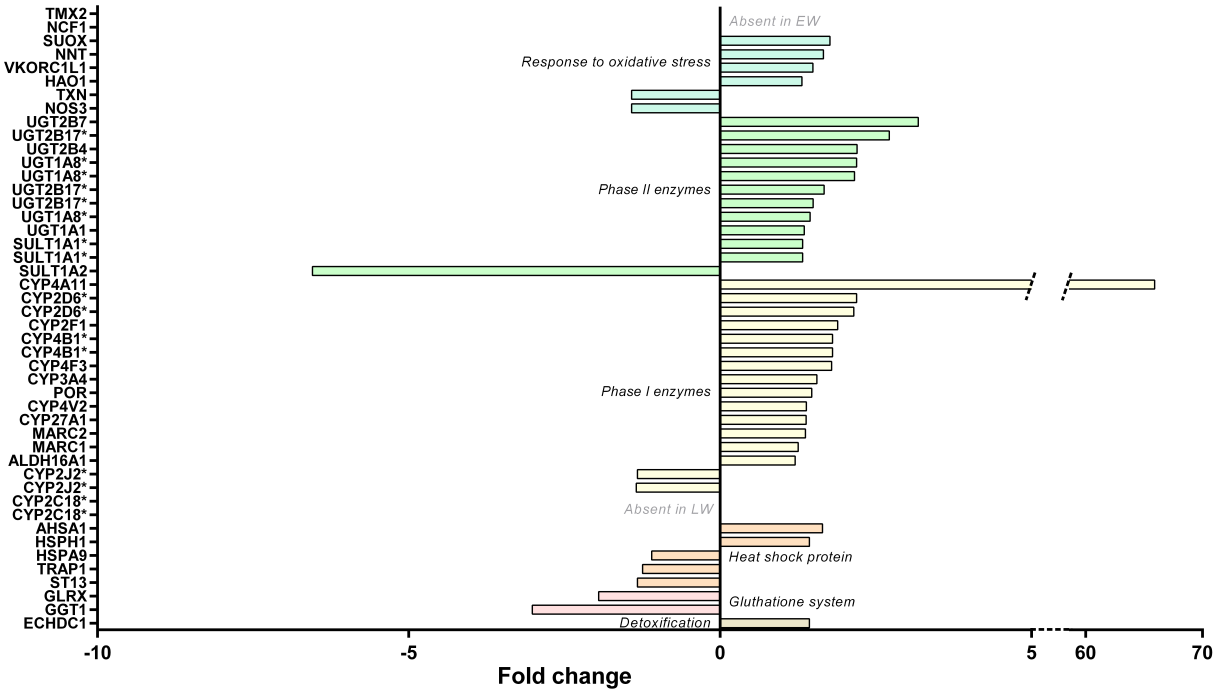

Figure 7. Regulation of liver stress responses in grey mouse lemurs during winter. The relative abundance of liver proteins involved in liver detoxification, oxidative stress and that of heat shock proteins in early (EW) and late (LW) winter mouse lemurs (N = 5/group) are shown using mean fold changes in LW vs EW. The presence of several isoforms of the same protein are indicated with asterisks (*).

2.3.8. Liver Detoxification

Examining the liver detoxification function, a majority of proteins exhibited higher abundances in LW vs. EW (Figure 7 and Table S1), including fourteen phase I enzymes (ALDH16A1, two isoforms of CYP2D6, CYP2F1, CYP27A1, CYP3A4, CYP4A11, two isoforms of CYP4B1, CYP4F3, CYP4V2, MARC1, MARC2, and POR; +21 to +6491%, median = +38%), eleven phase II enzymes (two isoforms of SULT1A1, UGT1A1, three isoforms of UGT1A8, UGT2B4, UGT2B7, three isoforms of UGT2B17; +32 to +218%, median = +49%), and two other enzymes (ECHDC1 and SUOX, +31–77%, median = +43%). Nevertheless, other phase I enzymes or isoforms were quantified and exhibited downregulations in LW (two isoforms of CYP2J2, −25 and −26%; and two isoforms of CYP2C18, not detected in LW—six other isoforms of CYP2C18 were also detected but not regulated, Figure 7 and Table S1). This was also the case for an additional phase II enzyme (SULT1A2, −85%; $p = 0.008$) with lower abundances in LW animals (Figure 6B). Again, being less stringent on the p-value threshold highlighted additional changes for phase I enzymes (CYP2C19, 2 upregulated isoforms in LW, +49% and +168%, and one downregulated isoform, −21%) and for one phase II enzyme (SULT2A1, +205% in LW) (Figure 7).

3. Discussion

Seasonal species are subject to very strong variations in environmental conditions, including energy availability. Amongst adaptive responses to such constraints, the energy-saving mechanisms, set up to survive periods of low energy availability, are coupled with preparatory phases, during which organisms store a maximum of energy in the form of fat. These transitions from high to low food availability (and the opposite) trigger strong variations in body conditions, which must be accompanied by profound molecular remodeling. Because the liver acts as a main hub in metabolic regulation, we decided to investigate how the variations of the hepatic proteome relate to the phenotypic features of the mouse lemur between early winter (EW) and late winter (LW), i.e., fattening and leaning,

respectively. Our results emphasize the regulations that operate in the liver between EW and LW in the mouse lemur, involving significant changes in glycolysis/gluconeogenesis, lipid metabolism but also detoxification. Although this study was conducted in captive animals, the regulations we identified reflected the seasonal phenotype of wild mouse lemurs and may emphasize protective mechanisms against metabolic disorders.

3.1. Massive Fattening Does Not Induce a Pathological State, as Witnessed by Liver Response

Seasonal species anticipate periods of low energy availability by storing large amounts of fat. Fattening results from the combination of food intake, though food availability begins to reduce in field conditions, and reduced metabolic rate. This process is obviously enhanced in captive conditions as animals do not experience reduction in food availability and an increase in food intake has in fact been reported previously in EW vs. LW animals [11]. Considering the known causes for hypertriglyceridemia [27], the elevated circulating levels of triglycerides we observed in captive animals during EW (Figure 1C) are probably due to a higher food intake. From body mass dynamics and food efficiency values, fattening is obvious in EW, and the activity of lipoprotein lipase is likely expected to be high at this stage for favoring triglyceride hydrolysis from chylomicrons and very low-density lipoproteins (VLDL) and their delivery to white adipose tissue. The production of chylomicrons might therefore exceed triglyceride clearance in EW animals. It is worth noting here that our macroscopic observation of the livers (not shown) suggests that this situation of enhanced lipid accretion does not become pathogenic, as no ectopic accumulation of fat was observed in the liver of EW lemurs, even with a diet that was rich in carbohydrates. This result is in line with the absence of any increase in hepatic lipid levels in rats fed a high-carb diet [28]. Conversely, hepatic lipogenesis is induced in healthy subjects in response to a high carbohydrate diet, which involves an increase in protein expression levels of transcription factor XBP1 [29]. In mouse lemurs, hepatic Xbp1 mRNA expression (Figure 4F) remained at similar levels in EW and LW, suggesting a lack of response to food intake levels (higher in EW) and a regulation uncoupled from the global orientation of lipid metabolism (lipogenesis in EW vs. lipolysis and lipid oxidation in LW). Our proteomic data in EW vs. LW, which highlight higher abundances of several hepatic proteins involved in lipid and fatty acid synthesis as well as lipid transport, do not necessarily contradict this hypothesis. Although the regulations we observed support the hypothesis that the physiology of EW animals was mainly oriented towards the constitution of fat reserves, it is unlikely that this actually reflects an increased production of VLDL in EW. Indeed, the major component of VLDL, namely, apolipoprotein B-100 [30], remained at stable levels in our two groups of lemurs (see Table S1). In addition, an increased production of VLDL would have resulted in the increased generation of LDL particles [31], and our data show the opposite. LDL particles are the main drivers for atherogenic disorders [32]; therefore, their lower levels in EW animals are another supporting element that the fattening process in EW animals does not lead to a pathogenic situation. Hence, proteome regulations for lipid metabolism should instead be considered in the light of the inhibition of lipogenesis in LW, which should contribute to promoting the mobilization and oxidation of lipids (see below).

In human adipose tissue, de novo lipogenesis is activated after high carbohydrate feeding [33], and our results support that similar regulations happen in EW animals. The obvious conversion of the excess of ingested carbs into fat reserves in captive EW mouse lemurs is testified by RER values above 1 during the night (feeding phase) [34,35]. In the liver of EW animals, the lower abundances of enzymes involved in the synthesis of unsaturated fatty acids (ACSL5 and ACOT4) and of enzymes known to convert arachidonic acid (C20:4 n-6) (CYP4F2 and EPHX2) suggest a reduced turnover for this fatty acid and a possible maintenance of its concentration at high levels. This would be consistent with a previous study in female mouse lemurs in EW, where n-6 fatty acids represented ~39% of the total lipids in the liver (vs. ~13% for n-3), with C20:4 n-6 being the main form [36]. The maintenance of high content of n-6 fatty acids (and reduced content of n-3 fatty acids) was

notably associated with deep and long torpor bouts, and this is also observed in hibernating species [37–39]. Therefore, high levels of n-6 fatty acids in EW mouse lemurs could support their extensive use of daily torpor. Conversely the higher levels of CYP4F2 in LW vs. EW animals may serve other functions (see below).

Interestingly, it has be demonstrated that captive mouse lemurs reduce their oxygen consumption during the first half of winter [11], however, we show here that this global reduction in metabolic rate does not translate into a reduction in nocturnal oxygen consumption or in liver protein levels that play a prominent role in mitochondrial processes for energy fuel oxidation(TCA cycle and OXPHOS). These last results corroborate the results of a study on golden-mantled ground squirrels [40], which highlighted that respiratory rates (all respiratory states) reached their highest levels during pre-hibernation in fattening animals. Expression of proteins involved in OXPHOS activity (Complex I and Complex I + II) was greatest during this period, suggesting a higher hepatic carbohydrate oxidation capacity, which is in line with our results (see below). Hence, the global decrease in oxygen consumption in EW animals might be mostly attributed to the drastic reduction of muscle activity and muscle metabolism instead of liver regulations. Our results, consisting of higher levels of glycolytic factors and lower levels of fatty acid oxidation-related proteins, suggest that maintenance of liver metabolic activity in EW animals and the maintained production of liver ATP largely involves carbohydrate fuels. This hypothesis is in perfect line with the preservation of insulin sensitivity during the fattening phase as it has been reported previously [11] and is further confirmed in the current study based on oGTT results and the well-regulated glycaemia. Previous studies showed an induction of insulin resistance during high-carbohydrate diets that involved not only hepatic steatosis but also the induction of *G6pc* and *Gckr* genes as a protective mechanism promoting homeostasis [41]. In our data, no such changes were observed in EW animals concerning liver lipid accumulation (see above) and regarding protein expression levels of G6PC and GCKR (see Table S1). EW mouse lemurs therefore exhibit a physiological state that appears to be safely managed without the induction of the known adverse effects that are usually triggered by high carbohydrate feeding or in relation with fat accumulation in white adipose. Also, at the systemic level, massive fattening did not induce excess of fat deposits in metabolic tissues (absence of hepatic steatosis; personal observation), and gene [11] and protein expression (present study) did not evidence any sign of hepatic inflammation. In that sense, EW animals look very similar to animals in the pre-hibernating period.

3.2. Liver Metabolism in LW Is Sustained by Mobilization of Peripheral Lipid Stores, but Independent of Glucose Metabolism

The massive fattening observed in male mouse lemurs in the first weeks of winter naturally stops in the middle of winter. Afterwards, the combination of decreased food intake (despite no change in food availability) and increased locomotor activity and metabolic rates [11] results in a decrease in body mass. Such seasonal regulation of feeding behavior and satiety signals has already been demonstrated in hibernating species and seems to be controlled through modifications in hormonal signaling [42] and/or the gut-brain axis [43]. The reduction in RER in LW (Figure 4C) testifies that oxidative metabolism in LW animals is essentially fueled by the mobilization of fat stores instead of food carbs. That being said, levels of RER do not reach a strict value of 0.7, meaning that metabolic oxidation is not purely lipidic, which is probably linked to the fact that animals are still eating high-carbs food (food intake is not null). The higher levels of CYP4F2 in LW animals may promote the conversion of C20:4 n-6 into 20-HETE, thus allowing for the adjustment of liver metabolic activity to control the fat-dependent energy supply and metabolism [44]. The action of CYP4F2 may also favor the catabolism of pro-inflammatory eicosanoids, such as leukotriene B4 [45], which could prevent the development of an inflammation due to large lipid fluxes in LW mouse lemurs as already suggested [11]. Similar regulations have already been described in hibernating animals, albeit with slight differences between animal species [46]. On the other hand, it is possible that liver n-6 fatty acids in LW are used in relation with

the reactivation of the reproductive axis in male mouse lemurs at the end of winter (see below). Indeed, C20:4 n-6 is a precursor for prostaglandins that play a role in reproduction and in thermoregulation [47,48].

At the proteomic level, the switch from carbs to lipid mobilization in LW might involve a reduction in glucose utilization in the liver, as suggested by a decreased abundance of glycolytic enzymes and of LDHA. In the same way, reduced expression levels of several isoforms of pyruvate dehydrogenase (PDHA1, PDHB and PDHX) suggest a reduction of the oxidation of glycolytic intermediates. Conversely, beta-oxidation of fatty acids is nicely reflected by an increased abundance (although not significant) of HADH and citrate synthase (CS) in LW. These protein regulations are in perfect line with observations at the systemic level in LW mouse lemurs, i.e., a lower tolerance to glucose after an oGTT, albeit basal insulin levels were higher than in EW animals [11]. On the other hand, the reduced levels of fructose-1,6-bisphosphatase 1 (FBP1) would be in line with a reduction in liver gluconeogenesis in LW compared to EW.

Glycerol is one of the main sources for liver gluconeogenesis, after entry into hepatocytes through aquaporin-9 (AQP9) known as a key rate-limited step [49]. In LW animals, glycerol is released due to white adipose tissue lipolysis and its uptake into liver cells is expected to increase. Our data showing higher expression levels of AQP9 protein during LW, support this hypothesis. However, the reduced abundances of GPD1 and GPD2 suggest that glycerol oxidation might not be increased in the liver of LW animals. Because metabolism is oriented toward fat utilization at this stage, it is unlikely that glycerol is used for liver lipogenesis and, therefore, glycerol might accumulate in LW. Strikingly, the increased expression of AQP9 at the mRNA level has been linked to insulin deficiency or to insulin resistance in mice [50]. It is therefore possible that the regulation of AQP9 in LW lemurs may be linked to glucose intolerance development (as reflected by oGTT results) and linked to a possible alteration in insulin signaling [11]. However, insulinemia and insulin receptor (INSR) expression levels remain unchanged between EW and LW in the current study, which suggests that glucose intolerance may not be due to an alteration in liver insulin signaling or that the cause of such an alteration (if it actually happens) are to be searched for elsewhere (see below).

Amino acids are another main substrate for liver gluconeogenesis. An increase in amino acid-derived hepatic gluconeogenesis was well described in response to fasting, as well as the resulting enhancement of protein protection through the maintenance of IGF-1 levels [51]. Such maintenance of IGF-1 abundance was also observed during caloric restriction in mouse lemurs [52] and let us hypothesize that this process is involved in the transition between EW and LW hepatic phenotypes. However, the reduced expression of all enzymes involved in liver gluconeogenesis from amino acids, between EW and LW, is another piece of evidence that liver glucose metabolism is not favored in LW. An association has been made between increased fasting-induced autophagy and increased hepatic gluconeogenesis [53], which contributes to the maintaining of a basal level of glycaemia. However, such mechanisms do not appear to be involved in maintaining the circulating glucose levels that were observed in mouse lemurs between EW and LW [11]. Indeed, the molecular regulations we observed point toward a reduction in gluconeogenesis in LW (see above), but contradictory data were recorded in the current study for the autophagy process. In fact, key autophagy factors (RMC1, BNIP3, and SNAP29) were more abundant in the liver of LW vs. EW animals, while at the same time the expression levels of WD repeat-containing protein 6 (WDR6), a negative regulator of amino acid starvation-induced autophagy (McKnight et al., 2012), were increased between EW and LW (Table S1; $p = 0.003$).

In accordance with the apparent repression of liver glucose metabolism in LW, glucose intolerance, associated with higher basal levels of insulin, was observed in LW in this species [11]. This suggests a possible impairment in insulin sensitivity during fat mobilization in LW. Furthermore, the expression of FBP1 protein is lower at the end of winter, when the animal is leaning, compared to the beginning of winter. According to

the literature, FBP1 is involved in the negative regulation of appetite and of adiposity [54], and it prevents the development of insulin resistance [55]. Indeed, increased expression of FBP1 in the liver after nutrient excess increases circulating satiety hormones and reduces appetite-stimulating neuropeptides and thus seems to provide a feedback mechanism to limit weight gain. In LW, animals are probably more starved than in EW, which could explain the reduced expression of FBP1 and further explain the transient settlement of insulin resistance in LW. However, pathway analysis from our proteomic dataset did not show any major regulation in the insulin signaling pathway, although the vast majority of actors were well detected. We think that the repression of liver glucose metabolism in LW without an important alteration of liver insulin sensitivity might indicate, again, that a pathological stage is not reached in the tissue. It would be interesting to pursue our investigation by focusing on proteome regulations in skeletal muscle to better identify the seasonal regulation of insulin signaling. In addition, seasonal control of this function could come from finest regulations, including phosphorylation of key proteins [56].

Despite a decrease in blood urea nitrogen during LW [11], it is difficult to conclude if the liver urea cycle is significantly affected in LW mouse lemurs. Indeed, we observed an upregulation of two key enzymes (DDO and ASS1), whereas the expression of three other enzymes (OTC, ARG1 and ASL) remained unchanged between EW and LW. As explained above, we cannot exclude that a finest regulation led by post-translational modifications and/or the modulation of the uptake of amino acids may be involved in the downregulation of the cycle, which should be explored in future studies. A possible hypothesis is that the rate of flux of amino acids toward the liver, a key regulator of ureagenesis, may be likely to be low in LW, due to a decrease in the muscle efflux of amino acids. This idea is supported by the self-imposed "caloric restriction" situation (according to the fact that food intake spontaneously decreases) in LW, which seems to resemble the low energy diets in several animals, including food-storing hibernators, which maintain their protein pool and limit muscle loss [10]. Hence, the reduction in food intake, together with increased energy expenditure (see below), is accompanied with the mobilization of fat stores in LW lemurs, which might help spare body proteins. Further, increased locomotor activity in LW animals (data not shown) might even favor physical exercise-stimulating protein synthesis [57]. Future studies to characterize the molecular events in the muscle from LW vs. EW lemurs, with an emphasis on the regulation of protein balance, should help answer this question.

Finally, although increased oxygen consumption was observed in LW [11], the regulations at the protein level observed here for the different subunits of the mitochondrial respiratory chain complexes suggest a general decrease in liver OXPHOS activity. At the systemic level, the increase in VO_2 in LW could be explained by increased locomotor activity [20] and by a higher muscle metabolic rate. We expect that muscle proteome regulations would reveal changes in OXPHOS activity and enhance the burning of fat to fuel metabolic processes. On another hand, hepatic protein regulations from our dataset suggest that mitochondrial uncoupling may be decreased in LW (UCP3 being downregulated as compared to EW), which conforms to the increase in lipid beta-oxidation (CPT1 upregulation in LW).

In this study, we identified a hepatic proteome downregulation of liver gluconeogenesis in LW that is not fully in line with regulations observed during hibernation, e.g., in hibernating bears [58]. In addition, species arousing from hibernation exhibit an increase in hepatic uncoupling [59], which was not observed in LW mouse lemurs. It is crucial to distinguish mouse lemurs from these other species knowing that they are still eating during LW and may not fully rely on their fat reserves. Such phenotypes may, however, rather resemble that of food-storing hibernators, which rely on food reserves during hibernation, therefore restoring blood glucose levels through food intake during arousals [10].

3.3. Fat Mobilization in LW Is Associated with the Reactivation of the Reproductive System and Liver Detoxification

In addition to the metabolic phenotype observed in wintering male mouse lemurs, the activity of the reproductive system is also deeply modified. Indeed, while a complete

regression of testes and a decrease in the circulating levels of testosterone are observed in EW, testicular recrudescence begins ~20 weeks after the beginning of winter time and is sustained by an increase in testosterone blood levels [20,60,61]. Body mass loss in LW is probably mostly explained by increased metabolic rates that could be a consequence of a more important locomotor activity [11]. However, we suspect that reactivation of the reproductive system (which can be likened to a yearly puberty) may contribute to body mass regulations and has been neglected so far. The net energetic cost of spermatogenesis is largely overlooked in mammals, while a few studies show that this process may explain body mass variations in reptiles [62]. In addition, male grey mouse lemurs present all the features describing sperm competition: high spermatogenesis efficacy and motility and low percentage of defect [63,64]. Most of the reproductive constraint in the male mouse lemur rely on the access to fertile females at the very beginning of the summer season. These observations may highlight an anticipatory phenomenon of sexual recrudescence in LW that would result in a significant energy expenditure. This also means that there is a strong need for the production of testosterone in LW, whereas EW animals experience a complete damping of circulating testosterone levels. Interestingly, the hepatic regulation of proteins involved in CHOL and steroid metabolism may reflect this process. Indeed, excess circulating CHOL in LW does not seem to be paralleled by an increase in its delivery to the liver by HDL. The augmented levels of circulating LDL could positively regulate CHOL delivery to the gonads for testosterone production, but it can also reflect a situation where the cardiovascular system of mouse lemurs during LW is at risk. Interestingly, female mouse lemurs have a delayed return to reproductive activity compared to males (estrus begin at the photoperiodic transition between winter and summer) and use their fat reserves later on, during gestation and lactation [20]. This temporal difference results in different patterns of seasonal variations in body mass [20] and probably interferes with the response to environmental stress [65,66]. There is, therefore, a high probability that liver remodeling will differ between males and females in winter, which should be investigated. The absence of significant variation in the ratio CHOL/HDL between EW and LW (although close to significance) and the maintenance of this ratio at a low level throughout winter [60] leads to the conclusion that the proportion of so-called "bad cholesterol" (LDL) might remain stable and low. The fact that liver low-density lipoprotein receptor (LDLR) and hepatic triacylglycerol lipase (LIPC) protein abundances remain unchanged between EW and LW does not support an increase in liver CHOL uptake. However, bile acid synthesis and secretion could be increased in LW animals, since we observed a significant upregulation for several key proteins of this pathway (CYP27A1, CYP3A4, HSD17B4, SLC27A5 and UGT2B4). It may be that increased bile acid production in LW may help remove excess CHOL, and it may also constitute an answer to the reduction of food intake for improving dietary lipid absorption, a well-known function of bile [61]. Interestingly, an activation of the bile acid receptor (NR1H4 or FXR) has been reported to repress gluconeogenic gene expressions and to enhance insulin sensitivity in diabetic mice [67]. In the same way, the possible overproduction of bile acids in LW lemurs may contribute to the global downregulation of liver gluconeogenesis we observed (see above). Alternatively, liver interconversion of cholesterol into bile acids has been linked to increased adaptive thermogenesis involving brown-adipose tissue (BAT) [68]. An increased activity in this thermogenic tissue has been observed in LW in mouse lemurs [11] and may explain the possible higher synthesis of bile acids in LW.

One last hepatic-specific process that seemed notably regulated in LW in mouse lemur is detoxification. With regards to oxidative stress, although systemic DNA damage (measuring urinary 8-OHdG levels) and circulating gamma-glutamyl transpeptidase (aka GGT) have both been demonstrated as increased in LW in mouse lemurs [11], proteome liver regulations of main antioxidant enzymes were unchanged or even reduced. Further evidence suggests that systemic DNA damage is related to metabolic rates in male mouse lemurs [66] and could not necessarily translate into a hepatic signature. Such observations were already made in the arctic ground squirrel, where oxidative stress during hibernation

is tissue-specific, being observed in BAT (a thermogenic tissue) but not in the liver [69]. The absence of liver oxidative stress upregulation in LW could indicate an absence of a pathogenic state for mouse lemurs. It is worth noting here that this may be a peculiar feature in mouse lemurs, as it was already pointed that prolonged fasting in rats triggers an increase in liver oxidative stress [70]. This appears further reinforced by the apparent correct functioning of the liver protein folding processes, as suggested by the fact that very few proteins involved in protein processing in endoplasmic reticulum were regulated in LW vs. EW animals, except for the downregulation of MAN1B1 and protein OS-9 that functioned as quality control targeting of misfolded proteins for degradation.

In our study, we successfully quantified 27 out of 30 detected proteins known to be involved in detoxification and all of them, mostly phase I and II enzymes, were upregulated between EW and LW. This may reflect an important activation of this process in LW. Aldehyde dehydrogenases are proteins that help to detoxify the accumulated derivatives of lipid peroxidation during torpor [71], and a previous study already revealed their upregulation in golden-mantled ground squirrels at the beginning of hibernation [72] and after the fattening phase [73]. Interestingly, golden-mantled ground squirrels are described as food-storing hibernators, they mostly rely on their fat reserves during hibernation [74], therefore resembling mouse lemurs' phenotype in LW [11]. On the other hand, studies in hibernating bears suggest a down-regulation of liver proteins involved in detoxification, due to their prolonged fasted state [75]. Based on these observations, we hypothesized that enhancement of liver detoxification in LW mouse lemurs could highlight their greater sensitivity to xenobiotics as compared to the EW situation. In fact, all the animals used for the proteomics analysis were anesthetized using a 2% isoflurane/98% air mixture for ~45 min (after ketamine induction). The impact of a general gaseous anesthesia was already demonstrated on hepatic gene expressions of drug metabolizing enzymes in rats after a long-lasting procedure (6 h under anesthesia) [76], although this has not yet been investigated for shorter exposures. According to the striking physiological differences that exists between EW and LW, which are supported by the hepatic proteome regulations we identified, we cannot reject the hypothesis that sensitivity to isoflurane per se may differ during winter and be greater in LW vs. EW. Alternatively, enhanced sensitivity to xenobiotics in LW might correspond to an anticipating process of the upcoming summertime with higher food intake. Anticipation processes are important during nutritional transitions. For instance, intestinal villi atrophy induced by food deprivation in the short term is restored during late fasting in laboratory rats, through cell divisions and a temporary interruption of apoptosis that correspond to anticipation of refeeding and the need for improving food assimilation and the restoration of the whole body condition [77].

4. Materials and Methods

4.1. Animals and Ethical Statement

Adult male grey mouse lemurs (mean age ± SEM: 3.7 ± 0.6 yo), all born in the captive population of Brunoy (MNHN, Brunoy, France, license approval N° E91.114.1), were randomly chosen and were in good health at the start of the experiment. General conditions of captivity were maintained at a constant: Ta (24–26 °C), relative humidity (55%). Food (including fresh fruits and a lab-made mixture) and water were available ad libitum. The lab-made mixture (1.31 kcal/g) was composed of cereals, gingerbread, condensed milk, yogurt, banana and egg and resulted in a macronutrient balance of ~60% carbohydrates, ~20% proteins and ~20% lipids. Diet fatty-acids were distributed as previously described [36]. In captivity, seasonal variations of physiological functions can be entrained by alternating 6-month periods of a summer-like long photoperiod (14 h of light/day) and a winter-like short photoperiod (10 h of light/day) under artificial light. The present study focused on the winter season, where adult (mean age ± SEM: 4.1 ± 0.1 years, range: 3.3–4.3) male mouse lemurs were maintained in constant photoperiod conditions (10 h of light/day). General conditions of captivity were applied, and animals were maintained in social groups. All experiments were carried out in accordance with the

European Communities Council Directive (86/609/EEC) and all experimental procedures were evaluated by an independent ethical council and approved for ethical contentment by the French government (authorizations n°03210.02, n°APAFIS#3004-2015111015031850 and n°APAFIS#3697-2016012111304236).

4.2. Glycaemia, Circulating Lipids and Metabolic Parameters

Fasting glycemia (mg·dL^{-1}) was measured using hand-held blood and circulating lipids (total cholesterol CHOL, high-density lipoprotein HDL and triglycerides TG, mg·dL^{-1}), using a hand-held blood OneTouch® Vita glucometer (LifeScan Inc., Milpitas, CA, USA) and Mission® Cholesterol Meter (ACON Laboratories, Inc., San Diego, CA, USA), respectively, a after blood puncture from the saphenous vein during the resting phase of the animals (Early Winter EW, N = 9; Late Winter LW, N = 10; N = 19 total) and before food became available. Low-density lipoprotein levels (LDL, mg·dL^{-1}) and CHOL/HDL ratios were calculated. In addition, metabolic data were reanalyzed from already published [details on the methods in 11] and additional unpublished works to extract data from individuals exhibiting the most extreme phenotypes between EW (fattening: highest body mass gain; N = 11, BMC range = +19–39%, median = 22%) and LW (leaning: highest body mass loss; N = 9, BMC range = −10–28%, median = −14%). Data [details of the methods in 11] on food efficiency (kcal/g), metabolic performance (nocturnal mean values of VO2 in ml/kg/h, heat production in kcal/h and respiratory exchange ratio (RER)) and glucose homeostasis (area under the curve AUC in arbitrary units during glucose tolerance test and basal insulinemia in ng/mL) were all put in the perspective of observations made in the proteomic analysis.

4.3. Body Mass and Liver Biopsies

Ten animals were randomly chosen for each critical period of the winter season: during the first half of the winter (week 6 of 26, 20% of winter period undergone, EW; N = 5) and the second half of winter (weeks 18 of 26, 66% of winter period undergone, LW; N = 5). Body mass (BM) was recorded at the time of liver biopsy. In addition, animals were continuously monitored for body mass in the captive population and data were implemented into a database. These data were used to calculate the difference between the body mass (BMC; expressed in %) recorded at the onset (for animals sampled in EW) or the offset (for animals sampled in LW) of the 6-month winter photoperiod and the maximum body mass (usually reached in mild winter).

Liver biopsies (5–10 mg samples) were sampled at the apex of liver in mouse lemurs under general anesthesia (induction with Diazepam 0.05 mg/kg, Valium®, Roche, Boulogne-Billancourt, France; Ketamine 0.2 mg/kg, Imalgen 1000, Boehringer Ingelheim, Ingelheim am Rhein, Germany; maintenance with isoflurane 1–2%, Iso-Vet, Piramal Healthcare, Bombay, India) and per-operating analgesia (before surgery: 0.05 mg/kg IM 30 min of Bupredine®, Buprenorphine, Laboratoire OSALIA, Paris, France; local cutaneous injection, around the abdominal aperture, of Lurocaïne, Vetoquinol S.A., Lure, France; after surgery: renewal of buprenorphine 4 h later, then meloxicam per os for 2 days, Metacam®, Boehringer Ingelheim, Ingelheim am Rhein, Germany). Monitoring was maintained over two weeks to ensure that the animal healed and recovered properly.

4.4. Liver Gene Expression

Hepatic gene expression was investigated using qPCR arrays designed for human (RT^2 Profiler PCR arrays—'Human Fatty Liver', Qiagen SAS, Courtaboeuf, France). Then, mRNA was extracted using trizol reagent and following the instructions of the manufacturer. RNA concentration was determined using Qubit fluorometric quantitation (ThermoFisher Scientific, Courtaboeuf, France) and the contamination of RNA was assessed using a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA). During this step, one sample was determined to be contaminated (sample from the second half of winter) and was therefore excluded from the experiment. The reverse transcription

and gene expression analysis were then processed following the instructions of the manufacturer. A pre-analysis using the Qiagen online PCR Array Data Analysis Web-based Software (https://geneglobe.qiagen.com/gb/; accessed on 14 December 2015) determined that 1 of our samples was not appropriate for gene expression analysis which meant that we had 4 liver samples left over for both the first and second half of winter. Variations between EW and LW have already been published [11] and will not be discussed here. Instead, gene expression data will be used to test the existence or none of any correlation between proteomic and transcriptomic results, as both methods were performed on the same liver samples.

4.5. Liver Proteome Analysis

Unless otherwise specified, all reagents and chemicals were purchased from Sigma Diagnostics (St. Louis, MO, USA). Frozen liver samples were weighed and grinded using a ball mill (2 × 30 s @ 25 Hz; MM400, Retsch, Eragny, France) under liquid nitrogen, and proteins were extracted using 10 volumes (i.e., 10 µL/mg of tissue) of extraction buffer (8 M urea, 2 M thiourea, 2% CHAPS, 2% DTT, 30 mM Tris pH 8.8, protease inhibitors). After sonication on ice (3 × 10 s, 135 Watts), followed by centrifugation (10 min, 2000× *g*, 4 °C) to remove possible cell debris, proteins were acetone-precipitated (6 volumes of cold acetone) overnight at −20 °C. Protein pellets were obtained by centrifugation (10 min, 13,500× *g*, 4 °C) before being washed using cold acetone and re-solubilised in Laemmli buffer (10 mM Tris pH 6.8, 1 mM EDTA, 5% β-mercaptoethanol, 5% SDS, 10% Glycerol). Protein concentrations were determined using the RC-DC Protein Assay (BioRad, Hercules, CA, USA), and similarity of protein electrophoretic profiles between all samples was then confirmed using 12% SDS-PAGE acrylamide gels (20 µg loaded) using colloidal Coomassie blue (Fluka, Buchs, Switzerland) staining. At this stage, a reference sample comprising equal amounts of all protein extracts was made to be repeatedly analysed during the whole duration of the nanoLC-MS/MS experiments, allowing for QC-related measurements.

Fifty µg of each sample were mixed with 4 × SDS sample buffer (1×: 10 mM Tris pH 6.8, 1 mM EDTA, 5% β-mercaptoethanol, 5% SDS and 10% glycerol), incubated at 95 °C for 5 min, and proteins were separated by electrophoresis using SDS-PAGE gels composed of a stacking (5% polyacrylamide) and a running (12% polyacrylamide) layer. Electrophoresis was stopped after protein migration in the running gel had reached 1.2 cm, and protein staining was performed using colloidal Coomassie Blue. Six gel bands (2 mm each) were excised, and proteins were in-gel digested with trypsin (Promega, Madison, WI, USA; ratio 1:100) at 37 °C overnight after de-staining, reduction (10 mM DTT), alkylation (55 mM iodoacetamide) and dehydration using a MassPrep station (Micromass, Waters, Milford, MA, USA). Tryptic peptides were extracted by incubation on an orbital shaker (450 rpm) for one hour using 60% acetonitrile, 0.1% formic acid in water and then for 30 min using 100% acetonitrile. A set of reference peptides (iRT kit; Biognosys AG, Schlieren, Switzerland) was finally added to all samples. Solvent was then eliminated using a vacuum centrifuge (SpeedVac, Savant, Thermoscientific, Waltham, MA, USA), and peptides were re-suspended in 30 µL of 1% acetonitrile, 0.1% formic acid in water.

Tryptic peptides were analysed on a nanoUPLC system (nanoAcquityUPLC, Waters, Milford, MA, USA) coupled with a quadrupole-Orbitrap hybrid mass spectrometer (Q-Exactive Plus, Thermo Scientific, San Jose, CA, USA). The six peptide extracts from an individual sample were injected in a row, and samples from the two animal groups were injected alternatively. To minimize carry-over, a column wash program (50% acetonitrile during 20 min) and two solvent blank injections were included in between each biological replicate. Peptide extracts from the reference sample were injected three times over the course of the nanoLC-MS/MS analyses.

Briefly, 1.5 µL of each sample were concentrated/desalted on a trap column (Symmetry C18, 180 µm × 20 mm, 5 µm; Waters, Milford, MA, USA) using 99% of solvent A (0.1% formic acid in water) and 1% of solvent B (0.1% formic acid in acetonitrile) at a flow rate of 5 µL/min for 3 min. Afterwards, peptides were eluted from the separation column

(BEH130 C18, 75 µm × 250 mm, 1.7 µm; Waters, Milford, MA, USA) maintained at 60 °C using a 65 min gradient from 1–35% of B at a flow rate of 450 nL/min.

The Q-Exactive Plus was operated in positive ion mode with source temperature set to 250 °C and spray voltage to 1.8 kV. Spectra were acquired through automatic switching between full MS and MS/MS scans in data dependent acquisition mode. Full scan MS spectra (300–1800 m/z) were acquired at a resolution of 70,000 (m/z 200) with an automatic gain control (AGC) value set to 3×10^6 ions, a maximum injection time set to 50 ms and the lock-mass option being enabled (polysiloxane, 445.12002 m/z). Up to 10 of the most intense multi-charged precursors per full MS scan were isolated using a 2 m/z window and fragmented using higher energy collisional dissociation (HCD, normalised collision energy of 27 eV). Dynamic exclusion of already fragmented precursors was set to 60 s. MS/MS spectra were acquired at a resolution of 17,000 (m/z 200) with an AGC value set to 1×10^5, a maximum injection time set to 100 ms and the peptide match selection option was turned on. The system was fully controlled by Xcalibur software (v3.0.63; Thermo Fisher Scientific, Whaltham, MA, USA). Peak intensities and retention times of reference peptides were monitored in a daily fashion.

MS raw data were processed using MaxQuant (v1.5.3.30) [78]. Peak lists were created using default parameters and searched using the Andromeda search engine implemented in MaxQuant against a protein database derived from the latest annotation of the *Microcebus murinus* (TaxID 30608) genome in Refseq (Refseq Assembly accession GCF_000165445.2; Assembly Name Mmur_3.0). Only the longest proteins per gene (coding DNA sequences) were retained, and after the elimination of any redundancy, the database contained 39,712 protein sequences to which sequences of common contaminants were added (247 entries; contaminants.fasta included in MaxQuant). The first search was performed using a precursor mass tolerance of 20 ppm and 4.5 ppm for the main search after recalibration. Fragment ion mass tolerance was set to 20 ppm. The second peptide research option was enabled. The carbamidomethylation of cysteine residues was considered as a fixed modification and oxidation of methionine residues and acetylation of protein N-termini as variable modifications during the search. A maximum number of two missed cleavages and a false discovery rate (FDR) of 1% for both peptide spectrum matches (minimum length of seven amino acids) and proteins were accepted during identification. Regarding quantification, data normalisation and protein abundance estimation were performed using the MaxLFQ (label-free quantification) option implemented in MaxQuant [78] using a "minimal ratio count" of one. "Match between runs" was enabled using a two-minute time window after retention time alignment. For quantification, we considered unmodified peptides and, if detected, also their modified counterpart (acetylation of protein N-termini and oxidation of methionine residues), while shared peptides were excluded. All other MaxQuant parameters were set as default. Proteins identified with a single peptide were not considered, and eventual contaminants and reversed proteins were removed from the dataset. Only proteins with at least four of five values per group as well as the ones "absent" (i.e., zero valid values) in samples from a given group were kept for quantification. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD030198.

To annotate identified *Microcebus murinus* proteins, we searched known homologous proteins in the human Swissprot database (accessed in April 2019). This was done by using BLAST searches (FASTA program v36; downloaded from http://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml; accessed on 23 April 2019), and only the top BLAST hit per protein was retained. To validate this procedure, we automatically extracted ortholog annotations of *Microcebus murinus* proteins and of their human homologues from the OrthoDB resource [79]. The relevance of the match between *Microcebus murinus* proteins and their human homologues was then checked manually.

4.6. Proteomic Bioinformatic Analysis

To identify functionally relevant biological processes differently regulated in EW and LW animals, enrichment and functional annotation analyses of differential proteomics data were performed using two complementary algorithms. First, we used the desktop version of DAVID (Ease v2.1) [80] and a version of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (Kegg) databases, downloaded in July 2019. Stringent criteria for considering significantly enriched GO/Kegg terms included an Ease score lower than 0.001, a Benjamini *p*-value lower than 0.01 and a fold enrichment higher than 2. The liver functions altered between EW and LW were determined after enriched GO/Kegg terms had been grouped together into broad functional categories. Secondly, we used the interactome analysis and CAME (ID Conversion, Annotation, Membership, function Enrichment) workflow for the annotation enrichment analysis applied in Metascape (https://metascape.org/; accessed on 5 May 2021). Default parameters were used. Using mainly the Biogrid database, information about physical protein—protein interaction was first extracted and a protein interaction network was built. A molecular complex detection (MCODE) algorithm was then applied to this network to identify neighborhoods where proteins were densely connected. Ontology enrichment analysis, which was performed using the most popular annotation sources, e.g., GO and Kegg, but also WikiPathways, Reactome, Corum and a series of other databases, was finally applied to each MCODE neighborhoods to assign them a biological function on the basis of the top-three functional enriched terms, when available. Networks were visualized using Cytoscape software v3.8.2 [81].

4.7. Western-Blot Analysis

The liver proteins previously solubilized (see above) in Laemmli buffer were used for Western blotting. Total protein concentration was determined using the Bio-Rad RC DC Protein Assay kit (Hercules, CA, USA). Protein electrophoresis was performed by loading 20 µg of each sample onto 4–20% Bio-Rad Mini-Protean TGX Stain-Free Precast Gels (Bio-Rad Hercules, CA, USA). Gel imaging was obtained after activation using the Bio-Rad ChemiDoc Touch Imaging System, then the proteins were transferred to 0.2 µm nitrocellulose membranes (Bio-Rad, Hercules, CA, USA) using the Trans-Blot Turbo Transfer System (Bio-Rad, Hercules, CA, USA). Blot imaging allowed the checking of the quality of the transfer. Membranes were blocked for 1 h at room temperature with a solution of TBS-T (Tris 25 mM, NaCl 137 mM, KCl 2.68 mM, 5% Tween 20) containing 4% of BSA, and they were then incubated overnight at 4 °C with primary antibodies. Primary antibodies were purchased from ThermoFisher Scientific (Waltham, MA, USA; Invitrogen manufacturer) for mitochondrial hydroxyacyl-coenzyme A dehydrogenase (HADH; PA5-31157) and citrate synthase (CS; PA5-221126) and from Abcam (Paris, France) for uncoupling protein 3 (UCP3; ab10985). They were used at 1/2000 (HADH), 1/1000 (CS) and 1/500 (UCP3) dilution in the blocking solution. Membranes were then washed three times for 10 min in TBS-T, before being incubated for 1 h at room temperature with a peroxidase-conjugated secondary antibody purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA), which was used at 1/5000 dilution in the blocking solution. Membranes were then washed three times for 10 min in TBS-T, before being incubated for 5 seconds in Luminata Classico Western HRP substrate (Merck Millipore, Molsheim, France) and then imaged for chemiluminescence using the ChemiDoc Touch Imaging System (Bio-Rad Hercules, CA, USA). Image analysis was carried out using the Bio-Rad Image Lab software v5.1. Signals were normalized to total proteins, as measured from stain-free gel images, and intensity values were expressed relative to those from EW animals, which were assigned an arbitrary value of one.

4.8. Statistical Analysis

Body mass and metabolic data are given as means ± standard error to the mean (SEM) and statistical significance was obtained using *t*-tests (p-values < 0.05). Statistical analysis of proteomics quantitative data was performed under R software environment v3.4.0 [82]. Normality of MaxLFQ values distribution and homoscedasticity were checked

using Shapiro-Wilk and Bartlett tests (p-values > 0.01), respectively. Changes in the protein abundances between animal groups were tested using Welch's t-tests, significance being set to p-values < 0.05 (given the small number of animals, we sometimes extended our data exploration to a significance level set at $p < 0.1$). A principal component analysis (PCA) was computed with all quantified proteins using the 'FactominR' 2.4 package in R. Missing values were imputed using the 'missMDA' 1.18 package. A heatmap was generated based on differentially expressed proteins between EW and LW (p-value < 0.05) using the 'heatmap.2' function in the 'gplots' 3.1.1 package. Hierarchical clustering was applied within animals only, not proteins, to attest for the difference between EW and LW animals, based on their proteomic profiles.

5. Conclusions

During winter, grey mouse lemurs (*Microcebus murinus*) exhibit an early phase of fattening followed by a late phase of fat mobilization. By investigating how their liver proteome is modulated from EW to LW, we were able to show that hepatic regulations nicely support fuel partitioning on the molecular level and really reflect the metabolic phenotype. Moreover, no sign of the typical adverse effects that are associated with fattening in humans, e.g., high levels of LDL or the absence of signs of altered insulin signaling and inflammation in the liver, were observed in EW animals. Similarly, despite the glucose intolerance indicated from an oGTT, we did not identify evidence of the alteration of insulin signaling in LW on the hepatic level. Animal species living in seasonal environments with strong energetic constraints have developed unique adaptive mechanisms conferring resistance capabilities to safely face metabolic challenges. Thus, this research perfectly illustrates that studying such species is relevant to identify targets involved in metabolic flexibility and energy homeostasis and could be an important asset to the biomedical field.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23084254/s1.

Author Contributions: J.T. conceived the study and carried out the experimental work on mouse lemurs; F.B. supervised proteomics analyses; A.A. helped with the animal work; B.C. and M.B.-D. conducted mass spectrometry-based analyses, and C.B. Western-blot analyses; J.T. and F.B. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the CNRS (2014 PEPS ExoMod program) and Strasbourg University (H2E project; IdEx Unistra) and the French Proteomic Infrastructure (ProFI; ANR-10-INSB-08-03). During the tenure of this study, M.B.-D. was the recipient of a PhD grant from the University of Strasbourg (IdEx Unistra), and C.B. from the French Ministry of Higher Education, Research and Innovation.

Institutional Review Board Statement: The animal study protocol was approved by the Ethics Committee of the Muséum national d'Histoire naturelle (CEEA Cuvier n°68) under the following protocol references: n°03210.02 (approved 11 March 2015), n°APAFIS#3004-2015111015031850 (approved 3 March 2016) and n°APAFIS#3697-2016012111304236 (approved 4 April 2016). All experiments were carried out in accordance with the European Communities Council Directive (86/609/EEC), and all experimental procedures were evaluated by an independent ethical council and approved for ethical contentment by the French government (authorizations n°03210.02, n°APAFIS#3004-2015111015031850 and n°APAFIS#3697-2016012111304236).

Data Availability Statement: The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository using the dataset identifier PXD030198. Electronic supplementary materials are provided for the detailed mass spectrometry-based identification and quantification of all proteins (Table S1) and the principal component analysis (PCA) of protein abundances (Table S2). Mouse lemur individual data remain available upon demand from the corresponding authors.

Acknowledgments: We are grateful to P. Guterl for his help in the bioinformatics analysis of proteomics data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dunlap, J.C.; Loros, J.J. Making Time: Conservation of Biological Clocks from Fungi to Animals. *Microbiol. Spectr.* **2017**, *5*, 5. [CrossRef] [PubMed]
2. Hut, R.A.; Beersma, D.G. Evolution of time-keeping mechanisms: Early emergence and adaptation to photoperiod. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2011**, *366*, 2141–2154. [CrossRef]
3. Gwinner, E. *Circannual Rhythms: Endogenous Annual Clocks in the Organization of Seasonal Processes*; Springer: Berlin/Heidelberg, Germany, 1986; Volume 18.
4. Carey, H.V.; Andrews, M.T.; Martin, S.L. Mammalian hibernation: Cellular and molecular responses to depressed metabolism and low temperature. *Physiol. Rev.* **2003**, *83*, 1153–1181. [CrossRef] [PubMed]
5. Klug, B.J.; Brigham, R.M. Changes to Metabolism and Cell Physiology that Enable Mammalian Hibernation. *Springer Sci. Rev.* **2015**, *3*, 39–56. [CrossRef]
6. Jastroch, M.; Oelkrug, R.; Keipert, S. Insights into brown adipose tissue evolution and function from non-model organisms. *J. Exp. Biol.* **2018**, *221*, jeb169425. [CrossRef] [PubMed]
7. Kinnunen, S.; Manttari, S.; Herzig, K.H.; Nieminen, P.; Mustonen, A.M.; Saarela, S. Maintenance of skeletal muscle energy homeostasis during prolonged wintertime fasting in the raccoon dog (Nyctereutes procyonoides). *J. Comp. Physiol. B* **2015**, *185*, 435–445. [CrossRef]
8. Berg von Linde, M.; Arevstrom, L.; Frobert, O. Insights from the Den: How Hibernating Bears May Help Us Understand and Treat Human Disease. *Clin. Transl. Sci.* **2015**, *8*, 601–605. [CrossRef]
9. Chanon, S.; Chazarin, B.; Toubhans, B.; Durand, C.; Chery, I.; Robert, M.; Vieille-Marchiset, A.; Swenson, J.E.; Zedrosser, A.; Evans, A.L.; et al. Proteolysis inhibition by hibernating bear serum leads to increased protein content in human muscle cells. *Sci. Rep.* **2018**, *8*, 5525. [CrossRef]
10. Bertile, F.; Habold, C.; Le Maho, Y.; Giroud, S. Body Protein Sparing in Hibernators: A Source for Biomedical Innovation. *Front. Physiol.* **2021**, *12*, 634953. [CrossRef]
11. Terrien, J.; Gaudubois, M.; Champeval, D.; Zaninotto, V.; Roger, L.; Riou, J.F.; Aujard, F. Metabolic and genomic adaptations to winter fattening in a primate species, the grey mouse lemur (*Microcebus murinus*). *Int. J. Obes.* **2017**, *42*, 221–230. [CrossRef]
12. Nelson, O.L.; Jansen, H.T.; Galbreath, E.; Morgenstern, K.; Gehring, J.L.; Rigano, K.S.; Lee, J.; Gong, J.; Shaywitz, A.J.; Vella, C.A.; et al. Grizzly bears exhibit augmented insulin sensitivity while obese prior to a reversible insulin resistance during hibernation. *Cell Metab.* **2014**, *20*, 376–382. [CrossRef] [PubMed]
13. Carrascosa, J.M.; Ros, M.; Andres, A.; Fernandez-Agullo, T.; Arribas, C. Changes in the neuroendocrine control of energy homeostasis by adiposity signals during aging. *Exp. Gerontol.* **2008**, *44*, 20–25. [CrossRef]
14. Migrenne, S.; Le Foll, C.; Levin, B.E.; Magnan, C. Brain lipid sensing and nervous control of energy balance. *Diabetes Metab.* **2011**, *37*, 83–88. [CrossRef] [PubMed]
15. Woods, S.C.; D'Alessio, D.A. Central control of body weight and appetite. *J. Clin. Endocrinol. Metab.* **2008**, *93*, S37–S50. [CrossRef] [PubMed]
16. Joly-Amado, A.; Denis, R.G.; Castel, J.; Lacombe, A.; Cansell, C.; Rouch, C.; Kassis, N.; Dairou, J.; Cani, P.D.; Ventura-Clapier, R.; et al. Hypothalamic AgRP-neurons control peripheral substrate utilization and nutrient partitioning. *EMBO J.* **2012**, *31*, 4276–4288. [CrossRef]
17. Smith, R.L.; Soeters, M.R.; Wust, R.C.I.; Houtkooper, R.H. Metabolic Flexibility as an Adaptation to Energy Resources and Requirements in Health and Disease. *Endocr. Rev.* **2018**, *39*, 489–517. [CrossRef]
18. Perret, M. Change in photoperiodic cycle affects life span in a prosimian primate (*Microcebus murinus*). *J. Biol. Rhythms* **1997**, *12*, 136–145. [CrossRef]
19. Languille, S.; Blanc, S.; Blin, O.; Canale, C.I.; Dal-Pan, A.; Devau, G.; Dhenain, M.; Dorieux, O.; Epelbaum, J.; Gomez, D.; et al. The grey mouse lemur: A non-human primate model for ageing studies. *Ageing Res. Rev.* **2012**, *11*, 150–162. [CrossRef]
20. Perret, M.; Aujard, F. Regulation by photoperiod of seasonal changes in body mass and reproductive function in gray mouse lemurs (*Microcebus murinus*): Differential responses by sex. *Int. J. Primatol.* **2001**, *22*, 5–24. [CrossRef]
21. Dewar, R.E.; Richard, A.F. Evolution in the hypervariable environment of Madagascar. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13723–13727. [CrossRef]
22. Hladik, A. *The Dry Forest of the West Coast of Madagascar: Climate, Phenology and Food Availability for Prosimians*; Charles-Dominique, P., Cooper, H.M., Hladik, A., Hladik, C.M., Pages, E., Pariente, G.F., Petter-Rousseaux, J.J., Schilling, A., Eds.; Academic Press: New York, NY, USA, 1980; pp. 1–40.
23. Schmid, J. Daily torpor in free-ranging gray mouse lemurs (*Microcebus murinus*) in Madagascar. *Int. J. Primatol.* **2001**, *22*, 1021–1031. [CrossRef]
24. Rigano, K.S.; Gehring, J.L.; Evans Hutzenbiler, B.D.; Chen, A.V.; Nelson, O.L.; Vella, C.A.; Robbins, C.T.; Jansen, H.T. Life in the fat lane: Seasonal regulation of insulin sensitivity, food intake, and adipose biology in brown bears. *J. Comp. Physiol. B* **2017**, *187*, 649–676. [CrossRef] [PubMed]

25. Tessier, S.N.; Zhang, J.; Biggar, K.K.; Wu, C.W.; Pifferi, F.; Perret, M.; Storey, K.B. Regulation of the PI3K/AKT Pathway and Fuel Utilization During Primate Torpor in the Grey Mouse Lemur, *Microcebus murinus*. *Genom. Proteom. Bioinform.* **2015**, *13*, 91–102. [CrossRef] [PubMed]
26. De Fea, K.; Roth, R.A. Modulation of insulin receptor substrate-1 tyrosine phosphorylation and function by mitogen-activated protein kinase. *J. Biol. Chem.* **1997**, *272*, 31400–31406. [CrossRef]
27. Packard, C.J.; Boren, J.; Taskinen, M.R. Causes and Consequences of Hypertriglyceridemia. *Front. Endocrinol. (Lausanne)* **2020**, *11*, 252. [CrossRef]
28. Bertram, H.C.; Larsen, L.B.; Chen, X.; Jeppesen, P.B. Impact of high-fat and high-carbohydrate diets on liver metabolism studied in a rat model with a systems biology approach. *J. Agric. Food Chem.* **2012**, *60*, 676–684. [CrossRef] [PubMed]
29. Glimcher, L.H.; Lee, A.H. From Sugar to Fat How the Transcription Factor XBP1 Regulates Hepatic Lipogenesis. *Integr. Physiol.* **2009**, *1173*, E2–E9. [CrossRef]
30. Gibbons, G.F. Assembly and secretion of hepatic very-low-density lipoprotein. *Biochem. J.* **1990**, *268*, 1–13. [CrossRef]
31. Sniderman, A.D.; Cianflone, K.; Frayn, K. The pathogenetic role of impaired fatty acid trapping by adipocytes in generating the pleiotropic features of hyperapoB. *Diabetologia* **1997**, *40* (Suppl. 2), S152–S154. [CrossRef]
32. Boren, J.; Chapman, M.J.; Krauss, R.M.; Packard, C.J.; Bentzon, J.F.; Binder, C.J.; Daemen, M.J.; Demer, L.L.; Hegele, R.A.; Nicholls, S.J.; et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease: Pathophysiological, genetic, and therapeutic insights: A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* **2020**, *41*, 2313–2330. [CrossRef]
33. Chong, M.F.; Hodson, L.; Bickerton, A.S.; Roberts, R.; Neville, M.; Karpe, F.; Frayn, K.N.; Fielding, B.A. Parallel activation of de novo lipogenesis and stearoyl-CoA desaturase activity after 3 d of high-carbohydrate feeding. *Am. J. Clin. Nutr.* **2008**, *87*, 817–823. [CrossRef] [PubMed]
34. Hellerstein, M.K.; Schwarz, J.M.; Neese, R.A. Regulation of hepatic de novo lipogenesis in humans. *Annu. Rev. Nutr.* **1996**, *16*, 523–557. [CrossRef] [PubMed]
35. Ferrannini, E. The theoretical bases of indirect calorimetry: A review. *Metabolism* **1988**, *37*, 287–301. [CrossRef]
36. Vuarin, P.; Henry, P.Y.; Guesnet, P.; Alessandri, J.M.; Aujard, F.; Perret, M.; Pifferi, F. Shallow hypothermia depends on the level of fatty acid unsaturation in adipose and liver tissues in a tropical heterothermic primate. *J. Therm. Biol.* **2014**, *43*, 81–88. [CrossRef] [PubMed]
37. Ruf, T.; Arnold, W. Effects of polyunsaturated fatty acids on hibernation and torpor: A review and hypothesis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2008**, *294*, R1044–R1052. [CrossRef] [PubMed]
38. Dark, J. Annual lipid cycles in hibernators: Integration of physiology and Behavior. *Annu. Rev. Nutr.* **2005**, *25*, 469–497. [CrossRef]
39. Giroud, S.; Chery, I.; Bertile, F.; Bertrand-Michel, J.; Tascher, G.; Gauquelin-Koch, G.; Arnemo, J.M.; Swenson, J.E.; Singh, N.J.; Lefai, E.; et al. Lipidomics Reveals Seasonal Shifts in a Large-Bodied Hibernator, the Brown Bear. *Front. Physiol.* **2019**, *10*, 389. [CrossRef]
40. Heim, A.B.; Chung, D.; Florant, G.L.; Chicco, A.J. Tissue-specific seasonal changes in mitochondrial function of a mammalian hibernator. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2017**, *313*, R180–R190. [CrossRef]
41. Agius, L. High-carbohydrate diets induce hepatic insulin resistance to protect the liver from substrate overload. *Biochem. Pharmacol.* **2013**, *85*, 306–312. [CrossRef]
42. Schwartz, C.; Hampton, M.; Andrews, M.T. Hypothalamic gene expression underlying pre-hibernation satiety. *Genes Brain Behav.* **2015**, *14*, 310–318. [CrossRef]
43. Otis, J.P.; Raybould, H.E.; Carey, H.V. Cholecystokinin activation of central satiety centers changes seasonally in a mammalian hibernator. *Gen. Comp. Endocrinol.* **2011**, *171*, 269–274. [CrossRef] [PubMed]
44. Elshenawy, O.H.; Shoieb, S.M.; Mohamed, A.; El-Kadi, A.O. Clinical Implications of 20-Hydroxyeicosatetraenoic Acid in the Kidney, Liver, Lung and Brain: An Emerging Therapeutic Target. *Pharmaceutics* **2017**, *9*, 9. [CrossRef] [PubMed]
45. Jin, R.; Koop, D.R.; Raucy, J.L.; Lasker, J.M. Role of human CYP4F2 in hepatic catabolism of the proinflammatory agent leukotriene B4. *Arch. Biochem. Biophys.* **1998**, *359*, 89–98. [CrossRef] [PubMed]
46. Giroud, S.; Evans, A.L.; Chery, I.; Bertile, F.; Tascher, G.; Bertrand-Michel, J.; Gauquelin-Koch, G.; Arnemo, J.M.; Swenson, J.E.; Lefai, E.; et al. Seasonal changes in eicosanoid metabolism in the brown bear. *Naturwissenschaften* **2018**, *105*, 58. [CrossRef]
47. Arnold, W.; Ruf, T.; Frey-Roos, F.; Bruns, U. Diet-independent remodeling of cellular membranes precedes seasonally changing body temperature in a hibernator. *PLoS ONE* **2011**, *6*, e18641. [CrossRef]
48. Arnold, W.; Kim, P.Y.; Allen, K.G.D.; Florant, G.L. Seasonal variation in brain prostaglandin D2 and E2 of marmots and n-6 fatty acid availability. In *Living in a Seasonal World: Thermoregulatory and Metabolic Adaptations*; Ruf, T., Bieber, C., Arnold, W., Millesi, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 531–542.
49. Gena, P.; Buono, N.D.; D'Abbicco, M.; Mastrodonato, M.; Berardi, M.; Svelto, M.; Lopez, L.; Calamita, G. Dynamical modeling of liver Aquaporin-9 expression and glycerol permeability in hepatic glucose metabolism. *Eur. J. Cell Biol.* **2017**, *96*, 61–69. [CrossRef]
50. Kuriyama, H.; Shimomura, I.; Kishida, K.; Kondo, H.; Furuyama, N.; Nishizawa, H.; Maeda, N.; Matsuda, M.; Nagaretani, H.; Kihara, S.; et al. Coordinated regulation of fat-specific and liver-specific glycerol channels, aquaporin adipose and aquaporin 9. *Diabetes* **2002**, *51*, 2915–2921. [CrossRef]
51. Norrelund, H.; Nair, K.S.; Jorgensen, J.O.; Christiansen, J.S.; Moller, N. The protein-retaining effects of growth hormone during fasting involve inhibition of muscle-protein breakdown. *Diabetes* **2001**, *50*, 96–104. [CrossRef]

52. Dal-Pan, A.; Terrien, J.; Pifferi, F.; Botalla, R.; Hardy, I.; Marchal, J.; Zahariev, A.; Chery, I.; Zizzari, P.; Perret, M.; et al. Caloric restriction or resveratrol supplementation and ageing in a non-human primate: First-year outcome of the RESTRIKAL study in *Microcebus murinus*. *Age (Dordr)* **2010**, *33*, 15–31. [CrossRef]
53. Ezaki, J.; Matsumoto, N.; Takeda-Ezaki, M.; Komatsu, M.; Takahashi, K.; Hiraoka, Y.; Taka, H.; Fujimura, T.; Takehana, K.; Yoshida, M.; et al. Liver autophagy contributes to the maintenance of blood glucose and amino acid levels. *Autophagy* **2011**, *7*, 727–736. [CrossRef]
54. Visinoni, S.; Khalid, N.F.; Joannides, C.N.; Shulkes, A.; Yim, M.; Whitehead, J.; Tiganis, T.; Lamont, B.J.; Favaloro, J.M.; Proietto, J.; et al. The role of liver fructose-1,6-bisphosphatase in regulating appetite and adiposity. *Diabetes* **2012**, *61*, 1122–1132. [CrossRef] [PubMed]
55. Hunter, R.W.; Hughey, C.C.; Lantier, L.; Sundelin, E.I.; Peggie, M.; Zeqiraj, E.; Sicheri, F.; Jessen, N.; Wasserman, D.H.; Sakamoto, K. Metformin reduces liver glucose production by inhibition of fructose-1-6-bisphosphatase. *Nat. Med.* **2018**, *24*, 1395–1406. [CrossRef]
56. Tanti, J.F.; Jager, J. Cellular mechanisms of insulin resistance: Role of stress-regulated serine kinases and insulin receptor substrates (IRS) serine phosphorylation. *Curr. Opin. Pharmacol.* **2009**, *9*, 753–762. [CrossRef] [PubMed]
57. Holwerda, A.M.; Paulussen, K.J.M.; Overkamp, M.; Smeets, J.S.J.; Gijsen, A.P.; Goessens, J.P.B.; Verdijk, L.B.; van Loon, L.J.C. Daily resistance-type exercise stimulates muscle protein synthesis in vivo in young men. *J. Appl. Physiol. (1985)* **2018**, *124*, 66–75. [CrossRef]
58. Shimozuru, M.; Kamine, A.; Tsubota, T. Changes in expression of hepatic genes involved in energy metabolism during hibernation in captive, adult, female Japanese black bears (Ursus thibetanus japonicus). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2012**, *163*, 254–261. [CrossRef] [PubMed]
59. Brustovetsky, N.N.; Amerkanov, Z.G.; Yegorova, M.E.; Mokhova, E.N.; Skulachev, V.P. Carboxyatractylate-sensitive uncoupling in liver mitochondria from ground squirrels during hibernation and arousal. *FEBS Lett.* **1990**, *272*, 190–192. [CrossRef]
60. Lemieux, I.; Lamarche, B.; Couillard, C.; Pascot, A.; Cantin, B.; Bergeron, J.; Dagenais, G.R.; Despres, J.P. Total cholesterol/HDL cholesterol ratio vs LDL cholesterol/HDL cholesterol ratio as indices of ischemic heart disease risk in men: The Quebec Cardiovascular Study. *Arch. Intern. Med.* **2001**, *161*, 2685–2692. [CrossRef]
61. Perino, A.; Demagny, H.; Velazquez-Villegas, L.; Schoonjans, K. Molecular Physiology of Bile Acid Signaling in Health, Disease, and Aging. *Physiol. Rev.* **2021**, *101*, 683–731. [CrossRef]
62. Olsson, M.; Madsen, T.; Shine, R. Is sperm really so cheap? Costs of reproduction in male adders, Vipera berus. *Proc. R. Soc. B-Biol. Sci.* **1997**, *264*, 455–459. [CrossRef]
63. Harcourt, A.H.; Harvey, P.H.; Larson, S.G.; Short, R.V. Testis weight, body weight and breeding system in primates. *Nature* **1981**, *293*, 55–57. [CrossRef]
64. Aslam, H.; Schneiders, A.; Perret, M.; Weinbauer, G.F.; Hodges, J.K. Quantitative assessment of testicular germ cell production and kinematic and morphometric parameters of ejaculated spermatozoa in the grey mouse lemur, *Microcebus murinus*. *Reproduction* **2002**, *123*, 323–332. [CrossRef] [PubMed]
65. Noiret, A.; Karanewsky, C.J.; Aujard, F.; Terrien, J. Why be thrifty? Sex-specific heterothermic patterns in wintering captive *Microcebus murinus* do not translate into differences in energy balance. *bioRxiv* **2021**. [CrossRef]
66. Noiret, A.; Puch, L.; Riffaud, C.; Costantini, D.; Riou, J.F.; Aujard, F.; Terrien, J. Sex-Specific Response to Caloric Restriction After Reproductive Investment in *Microcebus murinus*: An Integrative Approach. *Front. Physiol.* **2020**, *11*, 506. [CrossRef] [PubMed]
67. Zhang, Y.; Lee, F.Y.; Barrera, G.; Lee, H.; Vales, C.; Gonzalez, F.J.; Willson, T.M.; Edwards, P.A. Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1006–1011. [CrossRef]
68. Worthmann, A.; John, C.; Ruhlemann, M.C.; Baguhl, M.; Heinsen, F.A.; Schaltenberg, N.; Heine, M.; Schlein, C.; Evangelakos, I.; Mineo, C.; et al. Cold-induced conversion of cholesterol to bile acids in mice shapes the gut microbiome and promotes adaptive thermogenesis. *Nat. Med.* **2017**, *23*, 839–849. [CrossRef]
69. Orr, A.L.; Lohse, L.A.; Drew, K.L.; Hermes-Lima, M. Physiological oxidative stress after arousal from hibernation in Arctic ground squirrel. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **2009**, *153*, 213–221. [CrossRef]
70. Wasselin, T.; Zahn, S.; Maho, Y.L.; Dorsselaer, A.V.; Raclot, T.; Bertile, F. Exacerbated oxidative stress in the fasting liver according to fuel partitioning. *Proteomics* **2014**, *14*, 1905–1921. [CrossRef]
71. Carey, H.V.; Frank, C.L.; Seifert, J.P. Hibernation induces oxidative stress and activation of NK-kappaB in ground squirrel intestine. *J. Comp. Physiol. B* **2000**, *170*, 551–559. [CrossRef]
72. Epperson, L.E.; Dahl, T.A.; Martin, S.L. Quantitative analysis of liver protein expression during hibernation in the golden-mantled ground squirrel. *Mol. Cell. Proteom.* **2004**, *3*, 920–933. [CrossRef]
73. Blake, B.H. The annual cycle and fat storage in two populations of golden-mantled ground squirrels. *J. Mammal.* **1972**, *53*, 157–167. [CrossRef]
74. Pengelley, E.; Fisher, K.C. Rhythmical Arousal from Hibernation in Golden-Mantled Ground Squirrel, Citellus Lateralis Tescorum. *Can. J. Zool.* **1961**, *39*, 105–120. [CrossRef]
75. Fedorov, V.B.; Goropashnaya, A.V.; Toien, O.; Stewart, N.C.; Chang, C.; Wang, H.; Yan, J.; Showe, L.C.; Showe, M.K.; Barnes, B.M. Modulation of gene expression in heart and liver of hibernating black bears (*Ursus americanus*). *BMC Genom.* **2011**, *12*, 171. [CrossRef] [PubMed]

76. Nakazato, K.; Yoshida, Y.; Takemori, K.; Kobayashi, K.; Sakamoto, A. Expressions of genes encoding drug-metabolizing enzymes are altered after sevoflurane, isoflurane, propofol or dexmedetomidine anesthesia. *Biomed. Res.* **2009**, *30*, 17–24. [CrossRef] [PubMed]
77. Habold, C.; Chevalier, C.; Dunel-Erb, S.; Foltzer-Jourdainne, C.; Le Maho, Y.; Lignot, J.H. Effects of fasting and refeeding on jejunal morphology and cellular activity in rats in relation to depletion of body stores. *Scand. J. Gastroenterol.* **2004**, *39*, 531–539. [CrossRef] [PubMed]
78. Cox, J.; Hein, M.Y.; Luber, C.A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **2014**, *13*, 2513–2526. [CrossRef]
79. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simao, F.A.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [CrossRef]
80. Hosack, D.A.; Dennis, G., Jr.; Sherman, B.T.; Lane, H.C.; Lempicki, R.A. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **2003**, *4*, R70. [CrossRef]
81. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
82. R-Core-Team. R: A Language and Environment for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 9 March 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Both social environment and chronological age shape the physiology of ant workers.

Martin Quque [a, *], Charlotte Brun [a, b], Claire Villette [c], Cédric Sueur [a, d], François Criscuolo [a], Dimitri Heintz [c, 1], Fabrice Bertile [a, b, 1]

[a] Université de Strasbourg, CNRS, IPHC UMR 7178 – 23 rue du Loess, 67037 Strasbourg Cedex 2, France
[b] Infrastructure Nationale de Protéomique ProFI – 25 rue Becquerel, 67037 Strasbourg Cedex 2, France
[c] Plant Imaging & Mass Spectrometry (PIMS), Institut de biologie moléculaire des plantes, CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France.
[d] Institut Universitaire de France, 1 rue Descartes, 75231 Paris Cedex, France

*Corresponding author: Martin Quque

Email: martin.quque@iphc.cnrs.fr

Orcid IDs
Martin Quque 0000-0001-7386-1416
Charlotte Brun 0000-0001-8842-8420
Claire Villette 0000-0003-2564-5529
Cédric Sueur 0000-0001-8206-2739
François Criscuolo 0000-0001-8997-8184
Dimitri Heintz 0000-0001-7368-1394
Fabrice Bertile 0000-0001-5510-4868

Author Contributions: FC designed the ANT project, DH, FB, FC, CS and MQ designed the experimental protocol, MQ performed the behavioral observations and prepared samples before use in omics; CB and FB performed proteomics analysis; DH and CV performed metabolomics analysis; MQ performed the whole statistical analysis and data interpretation; MQ wrote the first draft; all authors edited the first draft.

Competing Interest Statement: The authors declare no conflict of interest.

Data statement: Should the manuscript be accepted, the data supporting the results will be archived and available in Figshare (doi activated at publication: 10.6084/m9.figshare.20013392).

Running title: Sociality shapes aging and physiology

[1] Share senior authorship of the paper.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Position within the social group has consequences on individual lifespans in diverse taxa. This is especially obvious in eusocial insects, where workers differ in both the tasks they perform and their aging rates. However, in eusocial wasps, bees and ants, the performed task usually depends strongly on age. As such, untangling the effects of social role and age on worker physiology is a key step towards understanding the coevolution of sociality and aging. We performed an experimental protocol that allowed a separate analysis of these two factors using four groups of black garden ant (*Lasius niger*) workers: young foragers, old foragers, young nest workers, and old nest workers. We highlighted age-related differences in the proteome and metabolome of workers that were primarily related to worker subcaste and only secondarily to age. The relative abundance of proteins and metabolites suggests an improved xenobiotic detoxification, and a fuel metabolism based more on lipid use than carbohydrate use in young ants, regardless of their social role. Regardless of age, proteins related to the digestive function were more abundant in nest workers than in foragers. Old foragers were mostly characterized by weak abundances of molecules with an antibiotic activity or involved in chemical communication. Finally, our results suggest that even in tiny species, extended lifespan may require to mitigate cancer risks. This is consistent with results found in eusocial rodents and thus opens up the discussion of shared mechanisms among distant taxa and the influence of sociality on life history traits such as longevity.

Keywords: behavioral ecology, senescence, division of labor, trade-off, ecophysiology, convergent evolution, *Lasius niger*

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1. Introduction

The place that an individual holds in the social network of the group has emerged as a strong explanatory factor of aging in various taxa (reviewed in 1): birds (2, 3), insects (4–6), and mammals (7) including humans (7–9). To better understand such a well conserved evolutionary relationship between sociality and aging, the use of additional key species that evolved remarkable social organizations will make it possible to test specific hypotheses experimentally and conduct studies faster than in humans. Some bee, ant, wasp, and termite species are known for their well-established social structure in which each individual performs tasks specific to the behavioral caste to which it belongs. Such a social division of labor has led to the naming of these insects as eusocial insects. In recent years, many studies have shown that eusocial insects are relevant models to understand the mechanistic basis of lifespan diversity (10, 11). For example, reproductive individuals in social insects may live up to 100 times longer than individuals of solitary insect species (12). However, the mechanisms usually known to be associated with aging in mammals or birds sometimes show either no or opposite relationships in eusocial insects. In fact, neither antioxidant enzymes (13–15), nor oxidative damage (reviewed in 16) appeared so far to be strongly associated with lifespan in social insects. Moreover, telomere length or their attrition rate are identified as reliable indicators of life expectancy in mammals and birds (*e.g.* 17–19). However, in the black garden ant (*Lasius niger*), telomeres are of similar length in workers and queens, while the latter live ten times longer (20).

Interpreting the results of such studies is sometimes complicated by the fact that the social role of a worker subcaste and its age are intimately related. The youngest workers most often perform tasks inside the nest (*e.g.*, feeding queen, building the nest) and the oldest ones perform tasks outside the nest (*e.g.*, food provisioning, waste management). Through experimental demographic manipulations, old workers can be forced to resume nest worker tasks and young workers to forage, opposing the classical pattern. By using this approach, a few studies have tried to disentangle the respective influences of chronological age and behavior on the physiology of eusocial insect workers (21–24). Some of those studies found that the social role was better than chronological age to explain the differences in gene expression or proteome (21, 22, 24). Conversely, despite significant differences in behavior, morphology, and longevity between the most extreme social castes (i.e. queen and workers), transcriptomics comparison in *Lasius niger* showed that the divergence in gene expression was better explained by age than behavior (25). Finally, some metabolic characteristics turn to be found varying in both an age- and behavior-dependent manner, like the concentration of brood pheromones involved in caste maturation (26). Thus, there is so far no consensus on the respective importance of age and social role effects on the physiology of social insects.

To address this question, we successively sampled workers from colonies of black garden ants. The first sampling occurred at the time of colony foundation where both worker subcastes had young individuals. One year later, we sampled the same colonies a second time. As the eggs were regularly removed before hatching, ants of both worker subcastes were one year older. Thanks to this unprecedented protocol, we have obtained old nest workers that were very unlikely to have ever been foragers and thus retained no potential molecular trace of past foraging activities. We then jointly analyzed the metabolome and proteome to get a comprehensive picture of physiological changes. By separating the effects of social role and age on physiology, this study aims to better understand the mechanisms of aging and to what extent the social environment can influence it.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2. Results and Discussion

The combined mass spectrometry analyses (LC-MS/MS) detected 712 metabolites and 1719 proteins (**Figure S1**). Original data sets are available in electronic supplementary material (**Table S1-S3**). The tables used for the fold-change (FC) analysis, cleaned from metabolites or proteins found in less than 3/5 samples in a group, are also available online (**Tables S4 and S5)**. Supplementary materials also encompass the statistics summary (FC, FDR, class, biological processes) of these analyses (**Tables S7 and S8**), as well as the references used to link analytes (proteins + metabolites) to the biological processes mentioned below. We found 16 proteins and 11 metabolites completely absent from at least one worker subcaste (**Table S9 and S10).** No functional annotation and no relevant studies were found to help us to attribute a clear biological meaning to them. Consequently, absent proteins and metabolites will not be further discussed thereafter.

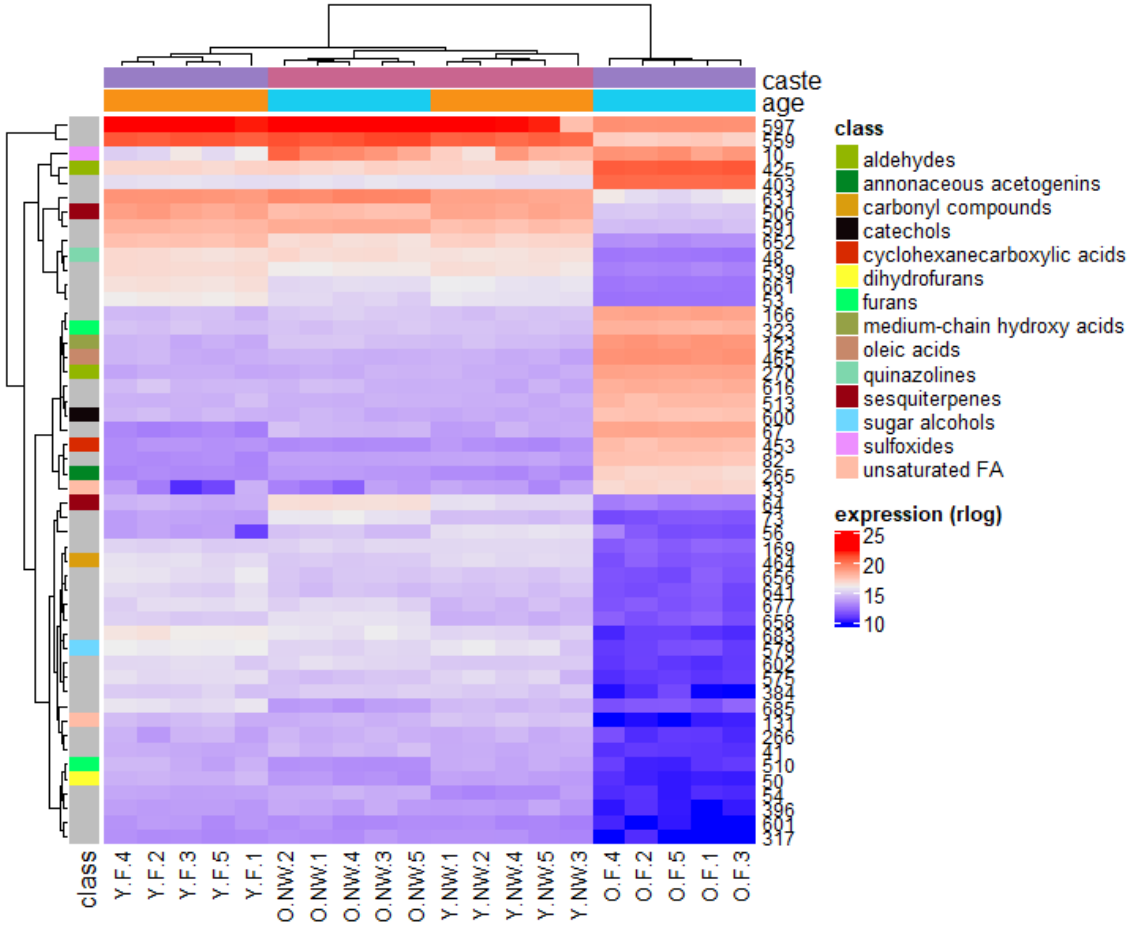

Figure 1. Heat map of the 50 most expressed metabolites amongst our four experimental groups in *L. niger*. At the top, colours indicate the age (**orange**: young, **light blue**: old) and caste (**purple**: foragers **pink**: nest-workers,). The left column indicates the metabolite class. Grey means a NA value. At the bottom are the sample identifiers (F: forager, NW: nest-worker, Y: young, O: old). For a given metabolite, the right column indicates its ID rather than the full name for legibility reasons (correspondence is indicated in every table provided). All metabolites presented here have an FDR < 0.05 and |fold-change| > 2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We drew heat maps from the metabolomics and proteomics data to picture how the four experimental groups (Y.F: young foragers, O.F: old foragers, Y.NW: young nest workers, O. NW: old nest workers) clustered together. The metabolomics-based clustering (**Figure 1**) shows that the four experimental groups have a weaker intra-group variability than inter-group variability and the O.F group differs the most when compared to the three other groups. Still, the two groups of nest workers (Y.NW and O.NW) are closer to each other than to Y.F, indicating that the behavior has a stronger impact than age on individual metabolome. On the contrary, the proteomics-based clustering (**Figure 2**) assorts samples according to the age (old or young) but independently of the worker subcaste (NW or F). As the relative importance of worker subcaste and age may depend on the biological processes studied, it is understandable that past studies found conflicting results about the role of age and social role (24, 25).

2.1. Old ants show a poorer somatic maintenance

Compared to O.NW, Y.NW had greater quantities of cytochrome P450 4g15, belonging to the CYP4 subfamily of cytochrome P450. Although potentially to a lesser extent than CYP1, 2 and 3 proteins (27), CYP4 have been found to be linked to the degradation of toxic xenobiotics (28–30). Similarly, Y.F compared to O.F had greater quantities of this same protein, as well as transferrin, which limits bacterial infection through iron-chelation (31). The metabolome analysis was consistent with the proteome analysis as it found greater quantities in Y.F than in O.F of metabolites associated with xenobiotic degradation, *e.g.*, 1-methoxyphenanthrene, cyclohexanone (KEGG maps 00930 and 00624); and immunity, *e.g.*, cyclohexane undecanoic acid, armillarin (32, 33). Besides, O.F had larger amounts of four metabolites associated with oxidative damage (*e.g,* 3,4-dimethoxybenzoic acid, robustocin), and only one of such metabolites (psilostachyin) had a larger amount in Y.F. Transferrin, which was found in greater quantities in Y.F (vs O.F) and Y.NW (vs O.NW), also has a protective role against oxidative stress (34, 35). Taken together these data depict the young ant workers as more likely to protect their organism integrity from both external (pathogens, pollutants) and internal (oxidative stress) threats. When compared to O.F, O.NW exhibit larger quantities of metabolites related to immunity (*e.g.*, thiophene, nebularine) and xenobiotic detoxification (cyclohexanone). Poorer somatic maintenance might therefore be a marker of senescence in worker ants, and old foragers might be the subcaste that suffers the most from it. This result fuels the adversarial debate between studies showing a cumulative deterioration of the organism with age in social insects and others showing the opposite, making it difficult in the current state of knowledge to describe a consistent model (16). Furthermore, these results show that both age and social role participate in the determination of the individual phenotype.

Finally, these results have to be viewed in relation to those presented below on cancer-related molecules. Indeed, it is known that immunity has anti-cancer roles (36), and several of our metabolites and proteins are known to be associated with immunity-related defense against cancer (*e.g.*, methylfurfuryl alcohol, armillarin, psilostachyin, thiophene). This exemplifies the interconnection between anti/pro-aging processes in a large variety of organisms.

2.2. Ants also experience a metabolic shift when aging

The organism's source of energy (lipids or carbohydrates) has been shown to affect longevity. For example, a previous study in black garden ants has positively associated a high body fat

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

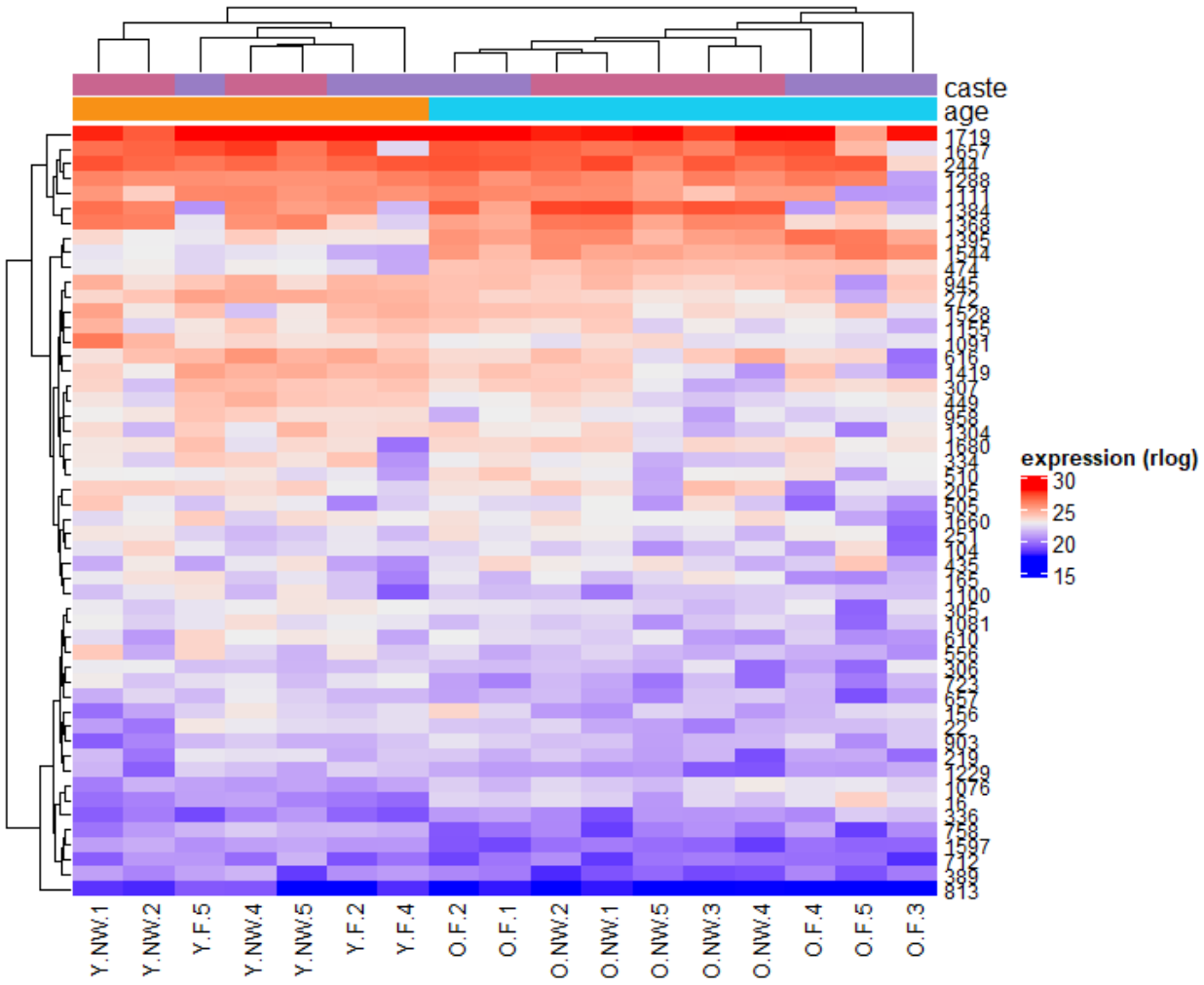

Figure 2. Heat map of the 50 most expressed proteins amongst our four experimental groups in *L. niger*. At the top, colours indicate the age (**orange**: young, **light blue**: old) and caste (**pink**: nest-workers, **purple**: foragers). At the bottom are the sample identifiers (F: forager, NW: nest-worker, Y: young, O: old). For a given protein, the right column indicates its ID rather than the full name for legibility reasons (correspondence indicated in every table provided). All proteins have an FDR< 0.05 and |fold-change| > 2.

content with a greater probability of survival of workers under stress (23). In honey bees, the greater longevity of queens has been found positively associated with a low polyunsaturated lipids/monounsaturated lipids ratio (37). A similar relationship has been reported in long-lived birds and mammals (38, 39). All these observations suggest that lipid metabolism is crucial in determining either survival or lifespan. However, our study suggests a metabolic shift with age, from lipid to carbohydrate use in ant workers. Indeed, compared to O.F, Y.F had greater amounts of two metabolites related to glycerophospholipid metabolism (lysoPE(0:0/18:1(11Z)) and lysolecithin), and eight related to unsaturated fatty acid metabolism (e.g., (1RS,2RS)-guaiacylglycerol 1-glucoside, 14,16-nonacosanedione, 6,8-tricosanedione). Other metabolites related to lipid assimilation and metabolism in general, not only unsaturated fatty acids, were found more abundant in Y.F (*e.g.*, ximenic acid, (E)-2-tetracosenoic acid, 9-Decenoylcarnitine).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Besides, we found greater abundances in Y.F vs O.F of proteins involved in lipid metabolism (e.g., acetyl-cytosolic). On the other hand, O.F (*vs* Y.F) showed greater abundances of many proteins involved in carbohydrate metabolism (*e.g.*, glucosidase, maltase 1, trehalose transporter tret1-like) but no protein related to lipid metabolism were found more abundant in this group. Similarly, we found a higher abundance for proteins involved in carbohydrate metabolism (*e.g.*, alpha-like protein) in O.NW vs Y.NW. Interestingly, we found greater abundances of metabolites in connection with lipid metabolism ((9Z,11E)-(13S)-13-hydroperoxyoctadeca-9,11-dienoate, and a C16 ceramide (d18:1/16:0) in O.NW compared to O.F. Similarly, larger amounts of proteins involved in lipid storage (lipid storage droplets surface-binding protein 2) were observed in O.NW *vs.* O.F. Hence, the apparent metabolic shift from lipid to carbohydrate metabolism as workers age appeared more pronounced in foragers, which is the worker subcaste that ages the fastest (40, 41). In humans too, when we age, lipids are less used as an energy source (42). Less efficient use of lipids might therefore be a signature of aging common to phylogenetically very far species. Still, we cannot rule out the possibility that this may also be a side-effect accompanying other functional changes with age. For instance, the relative abundance of few molecules suggested a less active immunity or chemical communication in O.F. than Y.F (see above), and these functions are lipid consuming (43–46), which could explain a slowing down of lipid use in O.F. Whether the observed metabolic shift has a causal effect on workers' aging rate, for instance *via* deleterious impact due to carbohydrate use as fuel (47–49), needs further experimental studies.

2.3. The consequences of social division of labor on digestion and communication

When comparing the proteome of nest workers with that of foragers (Y.NW *vs.* Y.F and O.NW *vs.* O.F), we systematically found the arylphorin subunit alpha in greater amount for nest workers, independently of age. This protein stimulates stem cells in the midgut and notably allows its regeneration after stress (50–52). The importance of digestive function in *Lasius niger* nest workers seems strong since we already found it in two previous studies (53, 54). To our knowledge, no other study experimentally addressed this question in adult nest workers. Two non-mutually exclusive hypotheses can be formulated to explain this phenotype: (i) the excess of food is stored and pre-digested by nest workers to make it quickly available for the rest of the colony in case of a food shortage; (ii) this pre-digested food would allow queens to process food more efficiently, by lowering energy investment in their own digestive system. The mechanism proposed in the latter hypothesis might form an energy trade-off solution explaining the unexpected equation of high reproduction and high longevity in ant queens.

When comparing O.F to Y.F (age effect) or O.NW (social role effect), among eight compounds involved in chemical communication (*e.g.,* tetracosanedioic acid, oleamide, hexadecanedioic acid, 2-[(methylthio)methyl]-2-butenal), only one (octanal) was found more abundant in O.F. This suggests that regardless of age and social role, O.F appeared to communicate much less with their congeners than do the ants in the other groups. Besides, the over-representation of oleic acid and derivatives in O.F (**Figure 1 and 3**) could be interpreted as a possible mechanism for members of the colony to detect senescent individuals. Indeed, oleic acid, linoleic acid and their derivatives are known to be secreted in a very conservative manner within insect taxa, including ants, when they die (55–57). It is known that this signal helps ants to recognize a corpse and is involved in a broad diversity of corpse management behaviors (58, 59), such as removal, cannibalism, burial (57, 58, 60, 61). It is also known that ants practice social immunity,

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

for instance, by reducing interactions with an individual recognized as sick (62, 63). Hence, oleic acid-derived signals may allow behavioral adaptions in social interactions, *e.g.,* avoiding interactions with O.F, which potentially care more germs and might have fewer antibiotic metabolites to tackle them according to our previously discussed results. It has also been shown that sick ants isolate themselves before dying (64, 65). Though causality is difficult to assess here, both cases (ignored by others or self-isolation) would lead to the social isolation of O.F and are consistent with a lower synthesis of communication molecules.

2.4. The emergence of anti-cancer mechanisms in the history of life

Smaller animals have fewer cells and therefore a lower probability of developing a tumor (as discussed in 66–68). These species therefore tend not to evolve mechanisms of cell replication blocking, found in bigger animals to prevent tumor development (66, 67). On the other hand, developing tumors over the course of life increases when the life span extends. Hence, long-lived social insects would face the situation of no replication-related protective mechanisms against cancer due to their size, combined with a higher probability of tumor development due to their longevity. However, no study describes a high prevalence of cancers among ants and eusocial insects in general. This raises the question of the mechanisms involved in the protection against cancer in those species. In our study, each group had at least one molecule potentially involved in anti-carcinogenic activity (e.g., [6]-shogaol, psilostachyin, thiophene; see **Table S8**) with a significantly higher relative abundance compared to the other groups. In other words, no clear variation due to age or social role could be identified for anti-carcinogenic molecules taken as a whole. The selection of molecular mechanisms of cancer resistance not based on replicative senescence is also found in naked mole rats, rodents that like ants are eusocial, small-sized, but long-lived (69).

Sociality, by increasing individual longevity, would therefore select for convergent anti-cancer mechanisms in distant taxa. However, only to a certain extent in ants since we found age-dependent abundances of proteins with pro-cancer activity, namely protein g12 and tetraspanin (70, 71). Old workers (O.F *vs* Y.F and O.NW *vs* Y.NW) had indeed larger amounts of these two proteins. Despite the presence of anti-cancer molecules, old ants might thus be more prone to cancer. This suggests that other age-related energetic trade-offs potentially made at the expense of anti-cancer mechanisms should also be taken into account. Still, it would also be possible that the observed amount of anti-cancer molecules co-varied in relation to other functions, which may have anti-aging effects, *e.g.*, immunity, involvement in the oxidative balance, antibiotic activity, or glycerophospholipid metabolism (see **Table S8**). The anti-cancer activity would be, in that case, only a side effect. Beyond the mechanistic questions, this would provide information on how sociality may impact cancer-related mechanisms, both at the intraspecific scale (workers vs queens, young vs old ants) and at the interspecific scale (between species of same or different level of sociality).

3. Conclusion

The purpose of this study was to disentangle the influence of social role and age on the physiology of workers from a eusocial species thanks to four experimental groups: young nest workers (Y.NW), young foragers (Y.F), old nest workers (O.NW) and old foragers (O.F). Specific functions (immunity, chemical communication, oxidative stress) and the clustering of samples showed in our study that both social role and age interact to define the individual phenotype.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Besides, our multi-omics approach highlighted evolutionary conserved mechanisms of aging (cancer, immunosenescence, poorer somatic maintenance in senescent groups, metabolic modifications). Thus, even if social insects, and more particularly ants, show distinctive features (notably the lack of a fertility/longevity trade-off), they share aging mechanisms with very distant taxa (mammals). This underlines potentially universal mechanisms, or at least shared by a majority of taxa and set-up very early in evolution of multicellular organisms (72). Furthermore, they offer us the opportunity to test how illness prevalence may be dependent on the synergy of ageing processes (73), under the influence of a strong degree of sociality. To describe more exhaustively the influence of sociality on individual physiology, further studies should include males and queens. In most ant species, males live only during the short breeding period, so it would be difficult to study the influence of age on their physiology. Conversely, because queens live up to ten times longer than workers (12), their regular monitoring over the course of their lifetime would take several years but should provide valuable information on the mechanisms slowing down aging processes. Beyond eusocial insects, uncovering mechanisms that promote a healthy lifespan could fuel progress in promoting human health.

4. Materials and Methods

The reader will find online (supplementary **Figure S1**) a graphical overview of the workflow applied in this study, from sample preparation of the four experimental groups, to the combined metabolomics and proteomics protocol and the curation of raw mass spectrometry data. Statistical analysis does not appear on this graphical overview but is detailed in the sections below.

4.1. Ant keeping and set-up of experimental groups

The black garden ant is a very common species in Western Europe where it predominantly inhabits urban habitats (74). For this species, the absence of dimorphism between workers (discussed in 75, 76) and the monogyny of the colonies reduce the potential sources of variation other than age and caste. In laboratory conditions, black garden ant workers have been shown to live 310 days on average and up to 1094 days, *ca.* 3 years (77). In our study, wild newly mated queen ants were collected in Lausanne, Switzerland (N 46.5234, E 6.5791). After being placed in individual glass tubes in the dark, 97 queens established a viable colony. Colonies were kept at a temperature of 21°C at night and 26°C during the day. To respect the biological rhythm of ants, we allowed them to enter diapause by gradually lowing the room temperature until 10°C for three months from December to March, before raising it to the usual values. The photoperiod varied throughout the year to mimic the natural photoperiod of the capture area. Relative humidity was 50-60%. Once a week, ant colonies were provided with a 0.3 M sugar water solution and mealworms (*Tenebrio molitor*; VenteInsecte, Courthézon, France).

After the first workers hatched, we let the colony develop for a month without any intervention, except feeding the ants. Due to the limited number of individuals available, young colonies at foundation sometimes have a less strict division of labour (78, 79). By waiting a month before the first collection, we allowed the colonies to grow sufficiently for the task specialization to take place. In addition, more developed colonies ensured that we could collect enough older individuals from the same colonies several months later. After a month, we ran the worker caste segregation protocol for the first time. The segregation of worker subcastes was based on their respective behaviors. Following a 48-hour period of starvation, colonies were proposed a 1 M

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sugar water. To maximize the forager recruitment, we waited five minutes after the first forager discovered the food source. Then, all the ants that came to the food source were considered as foragers, collected, and marked on the abdomen with an acrylic ink (Posca ©). We carried out this protocol three times to ensure we captured all foragers and with four-day intervals to allow the colony to rest. At the end of the procedure, not-marked ants were considered as nest workers. These young nest workers and foragers (0-1 month) were flash-frozen in liquid nitrogen and stored intact at -80°C until use. From that moment on, we followed the colonies carefully and removed the eggs in the last stages of maturation before they hatched. In this way, when we ran the worker caste segregation protocol for the second time, 11 months later, all the ants in the colonies were 11-12 months old. These two groups of old workers (nest workers and foragers) were also flash-frozen in liquid nitrogen and stored at -80°C until use. We ended up with four worker groups of both different behavior and age: young foragers (Y.F), old foragers (O.F), young nest worker (Y.NW) and old nest workers (O.NW). We sampled the experimental groups between August and September, several months before the ants enter diapause, which occurs from December to March in our laboratory conditions.

Before use for mass spectrometry analyses (*i.e.* proteomics and metabolomics), ink on foragers' abdomens was removed with acetone. We constituted pools of 16 workers for each group and used 5 pools per group (80 workers per group). Pools were made of a balanced mixture from different colonies to remove this possible confounding effect (2-3 workers of each colony). Ants were ground under liquid nitrogen for 30s followed by 15s at 30 Hz with steel beads (Ø 2 mm, Mixer Mill MM400, Retsch, Eragny Sur Oise, France). Tubes containing the resulting powder and beads were then stored at -80°C until their use for mass spectrometry.

4.2. Proteomics analysis

Unless otherwise specified, all chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA).

4.2.1. Sample Preparation

The frozen samples (powder) were suspended in 220 µL of lysis buffer (urea 8M, thiourea 2M, dithiothreitol [DTT] 1%, ammonium hydrogen carbonate [HNH_4CO_3] 0.1M, protease inhibitors 1.6 µM to 2 mM and sonicated on ice for 2 x 10 sec. at 135W, then samples were centrifuged (2000 x g, 2 min, RT) to eliminate possible cuticle remnants. Eight volumes of cold acetone (ThermoFisher Scientific, Rockford, IL, USA) were finally used for protein precipitation at -20°C overnight. Precipitated proteins were pelleted by centrifugation (13500 x g, 20 min, 4°C), washed once with cold acetone and then dissolved in Laemmli buffer (Tris 10 mM pH 7.5, EDTA 1 mM [Fluka, Buchs, Switzerland], β-mercaptoethanol 5%, SDS 5%, glycerol 10% [ThermoFisher Scientific]). Sonication and centrifugation were repeated as above to pellet and eliminate possibly remaining cell debris. Acetone fractions were all evaporated using a vacuum centrifuge (SpeedVac, Savant, Thermoscientific, Waltham, MA, USA) then kept at -20°C for further metabolomics analysis (see section '*Metabolomics preparation*' below).

Total protein concentration was determined using the RC-DC Protein Assay kit (Bio-Rad, Hercules, CA, USA). At this stage, a reference sample comprising equal amounts of all protein extracts was made, to be analysed regularly during the whole experiment and allow QC-related measurements. Twenty micrograms of proteins from each sample were loaded onto SDS-PAGE

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

gels (4% polyacrylamide for the stacking gel and 12% for the resolving gel) and electrophoresed for 20 minutes at 50 V then 20 minutes at 100 V. Proteins were thereafter fixed by a 15-minute incubation of gels in a solution composed of 50% ethanol and 3% phosphoric acid. Staining was performed using colloidal Coomassie Blue (30 min), and visualisation of proteins allowed five protein bands (2 mm each) to be excised from the gel. After destaining using acetonitrile/ammonium hydrogen carbonate 25 mM (75/25, v/v) and dehydration using pure acetonitrile, proteins were reduced and alkylated in-gel using 10 mM DTT in 25 mM ammonium hydrogen carbonate buffer (30 minutes at 60°C then 30 minutes at RT) and 55 mM iodoacetamide in 25 mM ammonium hydrogen carbonate buffer (20 min at RT in the dark), respectively. Gel slices were then washed using 25 mM ammonium hydrogen carbonate buffer (5 min, RT) and acetonitrile (5 min, RT) three times, and dehydration was finally performed using pure acetonitrile (2 x 5 min, RT). In-gel digestion of proteins was performed overnight at 37°C using trypsin (Promega Madison, WI, USA; 40 ng per band), and the resulting peptides were extracted twice (2 x 45 min) on an orbital shaker (450 rpm) using a solution composed of 60% acetonitrile and 0.5% formic acid in water. Another peptide extraction step was then performed (1 x 15 min) using 100% acetonitrile. At this stage, a set of reference peptides (iRT kit; Biognosys AG, Schlieren, Switzerland) was added to peptide extracts (6µL/sample after resuspension in 500mL of 20% acetonitrile/1% formic acid) for QC-related measurements. Organic solvent was thereafter evaporated using a vacuum centrifuge (SpeedVac) and the volume of peptide extracts was adjusted to 27 µL using a solution composed of 1% acetonitrile and 0.1% formic acid in water.

4.2.2. nanoLC-MS/MS analysis

NanoLC-MS/MS analysis was performed using a nanoUPLC-system (nanoAcquity; Waters, Milford, MA, USA) coupled to a quadrupole-Orbitrap hybrid mass spectrometer (Q-Exactive plus; Thermo Scientific, San Jose, CA, USA). The system was fully controlled by XCalibur software (v3.0.63; ThermoFisher Scientific). Samples (1 µl) were first concentrated/desalted onto a NanoEAse M/Z Symmetry precolumn (C18, 100 Å, 5 µm, 180 µm × 20 mm; Waters) using 99% of solvent A (0.1% formic acid in water) and 1% of solvent B (0.1% formic acid in acetonitrile) at a flow rate of 5 µl/min for 3 minutes. A solvent gradient from 1 to 6% of B in 0.5 minute then from 6 to 35% of B in 60 minutes was used for peptide elution, which was performed at a flow rate of 450 nL/min using a NanoEAse M/Z BEH column (C18, 130 Å, 1.7 µm, 75 µm x 250 mm; Waters) maintained at 60 °C. Samples were analysed randomly per block, each block being composed of one biological sample from each group. The reference sample was analysed six times throughout the experiment. In between each sample, washing of the column using 90% acetonitrile for 6 minutes and running of a solvent blank allowed limiting carry-over effects. Peak intensities and retention times of reference peptides were monitored in a daily fashion.

The Q-Exactive Plus was operated in positive ion mode with source temperature set to 250 °C and spray voltage to 1.8 kV. Full-scan MS spectra (300−1800 m/z) were acquired at a resolution of 70 000 at m/z 200. MS parameters were set as follows: maximum injection time of 50 ms, AGC target value of 3×10^6 ions, lock-mass option enabled (polysiloxane, 445.12002 m/z), selection of up to 10 most intense precursor ions (doubly charged or more) per full scan for subsequent isolation using a 2 m/z window, fragmentation using higher energy collisional dissociation (HCD, normalised collision energy of 27), dynamic exclusion of already fragmented precursors set to 60 seconds. MS/MS spectra (300-2000 m/z) were acquired with a resolution of

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

17500 at m/z 200. MS/MS parameters were set as follows: maximum injection time of 100 ms, AGC target value of 1×10^5 ions, peptide match selection option turned on.

4.2.3. Mass spectrometry data processing

Raw data were processed using MaxQuant v1.6.7.0 (80). Peak lists were created using default parameters. Using Andromeda search engine implemented in MaxQuant, peaklists were searched against a UniprotKb protein database (*Lasius niger*, TaxID 67767; 18217 entries) created in November 2019 with MSDA software suite (81). The database was complemented by Andromeda with the addition of the sequences of common contaminants like keratins and trypsin (247 entries) and of decoy (reverted) sequences for all *Lasius niger* proteins. Parameters were set as follows: precursor mass tolerance set to 20 ppm for the first search and to 4.5 ppm for the main search after recalibration, fragment ion mass tolerance set to 20 ppm, carbamidomethylation of cysteine residues considered as fixed modification, oxidation of methionines and acetylation of protein N-termini considered as variable modifications, peptide length of minimum 7 amino acids, maximum number of trypsin missed cleavages set to one, false discovery rate (FDR) set to 1% for both peptide spectrum matches and proteins. The proteins found with a single peptide or with a negative score were discarded from annotation data, as well as decoy hits and potential contaminants.

Protein quantification was performed using the MaxLFQ (label-free quantification) option implemented in MaxQuant. Parameters were set as follows: “minimal ratio count” of one, “match between runs” option enabled using a 0.7-minute time window after retention time alignment, consideration of both unmodified and modified (acetylation of protein N-termini and oxidation of methionines) peptides for quantitative determination, exclusion of shared peptides. All other MaxQuant parameters were set as default. Finally, criteria for retained proteins were as follows: at least two unique peptides quantified, no more than two missing values per group. Proteins absent in given groups (i.e. not detected at all) but satisfying above-mentioned criteria for the other groups were also retained. Among quantified proteins, 19 were annotated as “uncharacterized” (1.9% of all quantified proteins) for which we searched known homologous proteins in the Protostomia clade using BLAST searches (FASTA program v36; downloaded from http://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml), and only the best hits were retained. To validate this procedure, we automatically extracted orthology annotations and sequence domains of *Lasius niger* uncharacterized proteins and of their homologues from the OrthoDB (82) and InterPRO (83) resources. The relevance of the match among *Lasius niger* uncharacterized proteins and their homologues was then checked manually. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (84) partner repository with the dataset identifier PXD026565. From QC-related measurements, we could see that the whole analysis system remained stable throughout the experiment. Indeed, a median coefficient of variation (CV) of 1.3% was calculated for retention times of iRT peptides over all injections, and a median CV of 28% was computed for all LFQ values obtained from the repeated analysis of the reference sample.

4.3. Metabolomics preparation

4.3.1. Chemicals

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Deionised water was filtered through a Direct-Q UV (Millipore) station, isopropanol and methanol were purchased from Fisher Chemicals (Optima ® LC/MS grade). NaOH was obtained from Agilent Technologies, acetic acid, and formic acid from Sigma Aldrich.

4.3.2. Sample preparation

Dried acetone fractions collected during sample preparation for proteomics (see above) were rehydrated with 500µl ethyl acetate and 300µl water. The samples were vortexed for 10 seconds, and the ethyl acetate phase was harvested after phase partitioning for each sample and stored until LC-MS/MS analysis. The water phase was also collected from each sample and diluted with 1 ml of water acidified with 1% formic acid. The acidified water phase was then desalted using Solid Phase Extraction (SPE) based on HLB matrix Oasis 96-Well plate (30 µm, 5mg; Waters) coupled to a vacuum pump. Each SPE well was conditioned with 1 ml of methanol, then with 1 ml of water. The samples were then applied on the SPE and washed with 1 ml of water acidified with 0.1% formic acid. The samples were then eluted with 700 µl of methanol. The elution fractions were mixed with the ethyl acetate fractions prior to further analysis and analyzed in liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) described in the next section.

4.3.3. LC-HRMS analysis

Samples were analysed using liquid chromatography coupled to high resolution mass spectrometry on an UltiMate 3000 system (Thermo) coupled to an Impact II (Bruker) quadrupole time-of-flight (Q-TOF) spectrometer. Chromatographic separation was performed on an Acquity UPLC ® BEH C18 column (2.1x100mm, 1.7 µm, Waters) equipped with and Acquity UPLC ® BEH C18 pre-column (2.1x5mm, 1.7 µm, Waters) using a gradient of solvents A (Water, 0.1% formic acid) and B (MeOH, 0.1% formic acid). Chromatography was carried out at 35°C with a flux of 0.3mL.min^{-1}, starting with 5% B for 2 minutes, reaching 100% B at 10 minutes, holding 100% for 3 minutes and coming back to the initial condition of 5% B in 2 minutes, for a total run time of 15 minutes. Samples were kept at 4°C, 10 µL were injected in full loop mode with a washing step after sample injection with 150 µL of wash solution (H_2O/MeOH, 90/10, v/v). The spectrometer was operated in positive ion mode on a mass range of 20 to 1000 Da with a spectra rate of 2Hz in AutoMS/MS scan mode. The end plate offset was set to 500 V, capillary voltage at 2500 V, nebulizer at 2 Bar, dry gas at 8 L.min^{-1} and dry temperature at 200°C. The transfer time was set to 20-70µs and MS/MS collision energy at 80-120% with a timing of 50-50% for both parameters. The MS/MS cycle time was set to 3 seconds, absolute threshold to 816 cts and active exclusion was used with an exclusion threshold at 3 spectra, release after 1 min and precursor ion was reconsidered if the ratio current intensity/previous intensity was higher than 5. A calibration segment was included at the beginning of the runs allowing the injection of a calibration solution from 0.05 to 0.25min. The calibration solution used was a fresh mix of 50mL isopropanol/water (50/50, v/v), 500 µL NaOH 1M, 75 µL acetic acid and 25 µL formic acid. The spectrometer was calibrated in high precision calibration (HPC) mode with a standard deviation below 1 ppm before the injections, and recalibration of each raw data was performed after injection using the calibration segment.

4.3.4. Metabolite annotation

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Raw data were processed in MetaboScape 4.0 software (Bruker): molecular features were considered and grouped into buckets containing one or several adducts and isotopes from the detected ions with their retention time and MS/MS information when available. The parameters used for bucketing were a minimum intensity threshold of 10000, a minimum peak length of 4 spectra, a signal-to-noise ratio (S/N) of 3 and a correlation coefficient threshold set to 0.8. The $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$ ions were authorized as possible primary ions, the [M+H-H2O] ion was authorized as common ion. Replicate samples were grouped and only the buckets found in 80% of the samples of at least one group were extracted from the raw data. The area of the peaks was used to compare the abundance of the features between the different groups. The obtained list of buckets was annotated using SmartFormula to generate raw formula based on the exact mass of the primary ions and the isotopic pattern. The maximum allowed variation on the mass ($\Delta m/z$) was set to 3ppm, and the maximum mSigma value (assessing the good fitting of isotopic patterns) was set to 30. To put a name on the obtained formulae, metabolite lists were derived from Human Metabolite Database (HMDB, hmdb.ca), FooDB (foodb.ca), LipidMaps (lipidmaps.org) and SwissLipids (swisslipids.org). The parameters used for the annotation with the metabolite lists were the same as for SmartFormula annotation.

4.4. Statistics and biological interpretation of mass spectrometry results

Unless otherwise specified, the analysis and graphical representations were made using R software, version 4.0 (85).

4.4.1. Datasets

We only analyzed the proteins and metabolites (grouped under the term 'analytes'), which we could consider present or absent in full confidence. Analytes were considered present in a group only when present in at least 3 out of 5 samples of this group. Conversely, an analyte was considered completely absent from a group only when none of the samples contained it. Consequently, analytes present in only 1 or 2 samples of a group were discarded from the statistical analysis. For analytes present in only 3 or 4 out of 5 samples of a given experimental group, we imputed the missing values using an iterative PCA (principal component analysis) algorithm (MissMDA package v.1.17, 86). Considering only the proteins and metabolites present or absent in a group according to our criteria, missing data represented 2.3% of the proteomics data and 0.9% of the metabolomic data. In the proteomics statistical analysis, we discarded one young forager sample and one young nest-worker sample identified as outliers during the statistical workflow. In all statistical analyses, proteins and metabolites were studied separately to better underline their respective roles in the physiology of ant workers.

4.4.2. Heat maps and pairwise comparisons

First, we verified whether the different groups could be discriminated from each other using relative abundance of analytes and whether the discrimination criterion was age or behavior. For this, we used the 'rlog' function of the DESeq2 package (v.1.28, 87) to transform the proteomics and metabolomics data to the log2 scale in a way which minimizes differences among samples with small counts, and normalizes with respect to the size of the dataset (88). Based on this normalized data, we built heat maps with the ComplexHeatmap package (v.2.42, 89). Heat maps represent the most differentially expressed analytes and perform a hierarchical clustering, which reflects the proteomic or metabolomic similarities among samples. Then, we

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sought to identify exactly which proteins and which metabolites were expressed depending on the age or behavior. Comparing O. NW vs. O.F or Y.NW vs. Y.F provides information on the influence of the behavior since the age among the groups compared is identical. Conversely, keeping the behavior factor constant, by comparing Y.NW vs O.NW or Y.F vs O.F, makes it possible to assess the influence of age. We identified the analytes that differed most strongly between the groups compared two-by-two by calculating fold-changes (FC) of each analyte with the DESeq2 package. In this analysis, we retained only the analytes with a false discovery rate (FDR) lower than 0.05 and a FC higher than 2 (up-regulated) or lower than -2 (down-regulated).

4.4.3. Classification and functional annotation of analytes

In metabolomics, a molecule is stated '*identified*' only when compared with a reference standard. However, we did not use a standard for the hundreds of metabolites found in this study, but we used online databases (see sections above) to attribute names to the analytes found. This is called annotation: in this study, the presented annotations are at the level 3 of Schymanski's classification (90). Conversely, peptides do not require comparison to a standard to be formally identified. However, for consistency throughout the article, we used the term "*annotated*" to referred to *annotated* metabolites and *identified* proteins. Consequently, when an analyte (protein or metabolite) is simply referred to as '*annotated*', it means named and not functionally annotated.

To address the function of detected molecules, we ran functional enrichment analyses. Such methods evaluate the significance of a set of functionally related molecules. Regarding the proteome, functional annotation enrichment analysis was performed using the desktop version of DAVID (Ease v2.1) and versions of Gene Ontology (GO) and KEGG databases downloaded in December 2021. However, no such annotation was available for ant proteins, and no significant enrichment has been found when using homologous proteins in *Drosophila melanogaster* or *Homo sapiens*. Regarding the metabolome, we used the online platform MetaboAnalyst (v. 5.0, www.metaboanalyst.ca, 91) and performed a metabolite set enrichment analysis (aka. MSEA). However, among 196 annotated metabolites kept for the characterization of the four experimental worker groups, only 53 (27.04%) were recognized by the platform. This issue presumably came from the fact that the databases used by MetaboAnalyst are mainly derived from human studies. The analysis did not sufficiently cover our dataset to fully depict its diversity. The results of this analysis will therefore not be discussed but are available as supplementary material (**ESM2**). As automated functional enrichment was not compatible with our dataset, we investigated the biological meaning of the proteomics and metabolomics profiles by automatically retrieving metabolic maps from the KEGG database (genome.jp/kegg) when available. If not, we "manually" looked in the literature for functions fulfilled by the molecules in concern. References are provided in Tables S7 and S8.

Metabolites may encompass very distinct molecules, *e.g.,* lipids, free amino acids, free nucleic acids, carbohydrates. We automatically classified metabolites thanks to the ChemRICH online tool (chemrich.fiehnlab.ucdavis.edu; 92), which covered 79.03% of our metabolite dataset. We used these classes of metabolites in a principal component analysis (PCA, **Figure 3**) to see whether some groups of metabolites characterized rather the behavior than age, and conversely (FactoMineR package v.2.3, 93). Using the PCA coordinates, we calculated the intra-group repeatability (package "rptR" v.09.22, 94) to assess their homogeneity. The metabolite's class

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

547 was also added to heat maps to provide additional information about the most segregating
548 molecules.

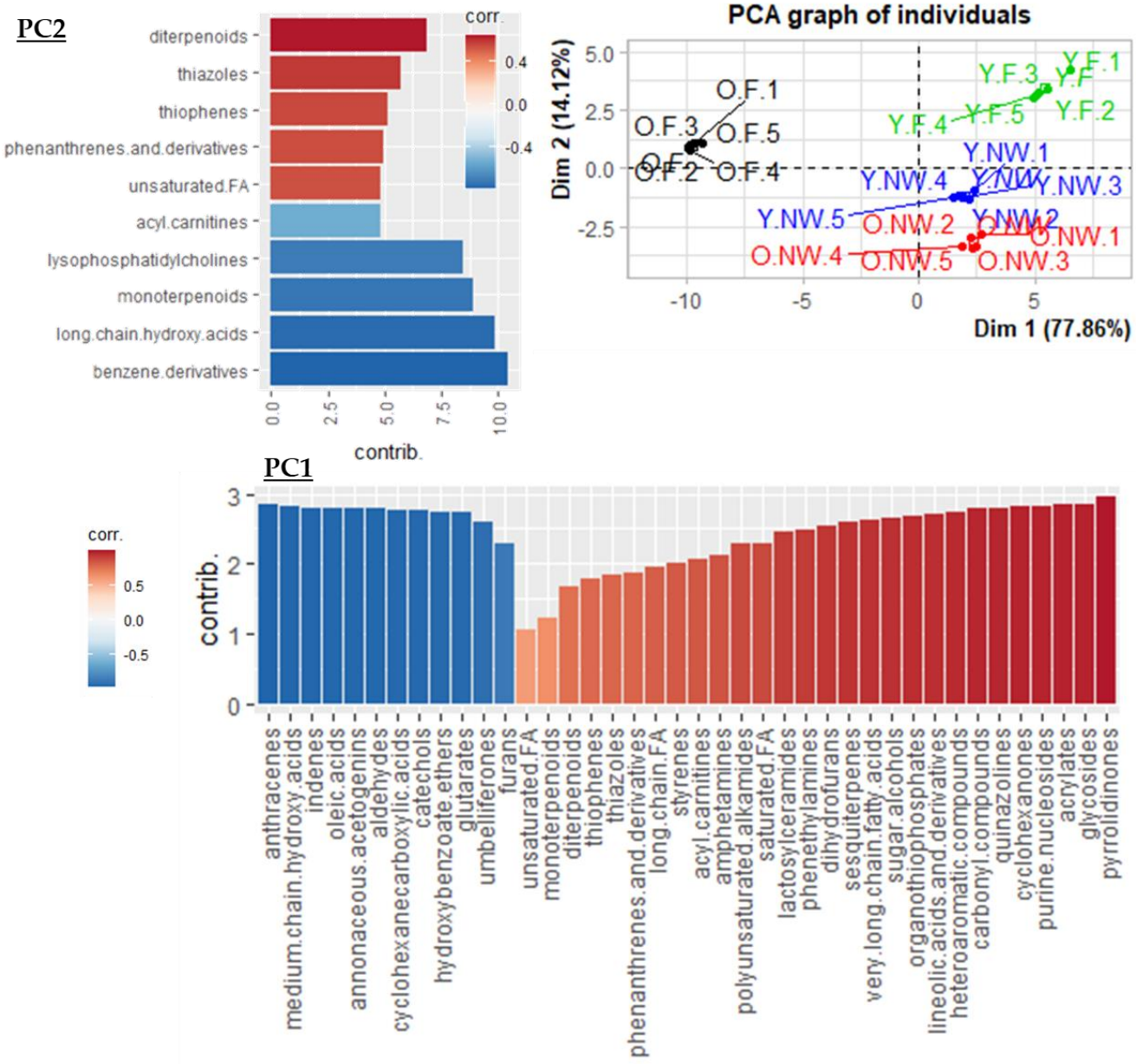

Figure 3. Classification-based PCA (PC1 and PC2). We ran a PCA with the metabolite classes as variables and explore the metabolomic differences amongst our four experimental groups in *Lasius niger* ants: **Y.F** (in green) = young foragers, **O.F** (in black) = old foragers, **Y.NW** (in blue) = young nest workers, **O.NW** (in red) = old nest-workers. The bar plots represent the classes of metabolites and their correlation with the first and second principal component (PC1 and PC2) from blue (negatively correlated) to red (positively correlated), as well as their contribution (bar length). All metabolites showed are correlated.
549

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5. Abbreviations

FC: fold change. **FDR:** false discovery rate. **PC (A):** principal component (analysis). **Y/O.NW**: young/old nest-workers. **Y/O.F:** young/old foragers.

6. Acknowledgments

We thank Nathalie Stroeymeyt for providing the newly-mated queens, Hélène Gachot-Neveu, Aurélie Kranitsky and David Bock for their precious work in the animal husbandry.

The study was supported by the Centre National de la Recherche Scientifique (CNRS) and the French Proteomic Infrastructure (ProFi; ANR-10-INSB-08-03). M. Quque and C. Brun PhDs were funded by the University of Strasbourg and the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (French Ministry of higher Education, Research and Innovation).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

560 References

561 1. C. Sueur, M. Quque, A. Naud, A. Bergouignan, F. Criscuolo, Social capital: an independent
562 dimension of healthy ageing. *Peer Community Journal* **1** (2021).

563 2. R. Covas, M. A. du Plessis, C. Doutrelant, Helpers in colonial cooperatively breeding
564 sociable weavers *Philetairus socius* contribute to buffer the effects of adverse breeding
565 conditions. *Behav Ecol Sociobiol* **63**, 103–112 (2008).

566 3. D. Aydinonat, *et al.*, Social isolation shortens telomeres in African grey parrots (*Psittacus*
567 *erithacus erithacus*). *PLoS ONE* **9**, e93839 (2014).

568 4. H. Ruan, C.-F. Wu, Social interaction-mediated lifespan extension of Drosophila Cu/Zn
569 superoxide dismutase mutants. *PNAS* **105**, 7506–7510 (2008).

570 5. E. H. Dawson, *et al.*, Social environment mediates cancer progression in Drosophila.
571 *Nature Communications* **9** (2018).

572 6. B. Wild, *et al.*, Social networks predict the life and death of honey bees. *bioRxiv*,
573 2020.05.06.076943 (2020).

574 7. N. Snyder-Mackler, *et al.*, Social determinants of health and survival in humans and other
575 animals. *Science* **368**, eaax9553 (2020).

576 8. I. Kawachi, S. V. Subramanian, D. Kim, "Social Capital and Health" in *Social Capital and*
577 *Health*, I. Kawachi, S. V. Subramanian, D. Kim, Eds. (Springer, 2008), pp. 1–26.

578 9. N. Grant, M. Hamer, A. Steptoe, Social isolation and stress-related cardiovascular, lipid,
579 and cortisol responses. *Ann Behav Med* **37**, 29–37 (2009).

580 10. L. Keller, S. Jemielity, Social insects as a model to study the molecular basis of ageing.
581 *Experimental gerontology* **41**, 553–556 (2006).

582 11. J. D. Parker, What are social insects telling us about aging? *Myrmecological News* **13**,
583 103–110 (2010).

584 12. L. Keller, M. Genoud, Extraordinary lifespans in ants: a test of evolutionary theories of
585 ageing. *Nature* **389**, 958–960 (1997).

586 13. J. D. Parker, K. M. Parker, B. H. Sohal, R. S. Sohal, L. Keller, Decreased expression of Cu-Zn
587 superoxide dismutase 1 in ants with extreme lifespan. *Proc Natl Acad Sci U S A* **101, 101**,
588 3486, 3486–3489 (2004).

589 14. M. Corona, K. A. Hughes, D. B. Weaver, G. E. Robinson, Gene expression patterns
590 associated with queen honey bee longevity. *Mechanisms of Ageing and Development*
591 **126**, 1230–1238 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

15. M. Corona, G. E. Robinson, Genes of the antioxidant system of the honey bee: annotation and phylogeny. *Insect Molecular Biology* **15**, 687–701 (2006).

16. E. R. Lucas, L. Keller, Ageing and somatic maintenance in social insects. *Current Opinion in Insect Science* **5**, 31–36 (2014).

17. B. J. Heidinger, *et al.*, Telomere length in early life predicts lifespan. *PNAS* **109**, 1743–1748 (2012).

18. J. R. Eastwood, *et al.*, Early-life telomere length predicts lifespan and lifetime reproductive success in a wild bird. *Molecular Ecology* **28**, 1127–1137 (2018).

19. K. Whittemore, E. Vera, E. Martínez-Nevado, C. Sanpera, M. A. Blasco, Telomere shortening rate predicts species life span. *PNAS*, 201902452 (2019).

20. S. Jemielity, *et al.*, Short telomeres in short-lived males: what are the molecular and evolutionary causes? *Aging Cell* **6**, 225–233 (2007).

21. I. Iovinella, *et al.*, Antennal protein profile in honeybees: caste and task matter more than age. *Front. Physiol.* **9** (2018).

22. C. W. Whitfield, A. M. Cziko, G. E. Robinson, Gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296–299 (2003).

23. A. Dussutour, L.-A. Poissonnier, J. Buhl, S. J. Simpson, Resistance to nutritional stress in ants: when being fat is advantageous. *Journal of Experimental Biology* **219**, 824–833 (2016).

24. P. Kohlmeier, A. R. Alleman, R. Libbrecht, S. Foitzik, B. Feldmeyer, Gene expression is more strongly associated with behavioural specialisation than with age or fertility in ant workers. *Molecular Ecology* (2018) https:/doi.org/10.1111/mec.14971 (December 10, 2018).

25. E. R. Lucas, J. Romiguier, L. Keller, Gene expression is more strongly influenced by age than caste in the ant *Lasius niger*. *Mol Ecol* **26**, 5058–5073 (2017).

26. C. Alaux, *et al.*, Regulation of brain gene expression in honey bees by brood pheromone. *Genes, Brain and Behavior* **8**, 309–319 (2009).

27. Y. B. Jarrar, S.-J. Lee, Molecular functionality of cytochrome P450 4 (CYP4) genetic polymorphisms and their clinical implications. *International Journal of Molecular Sciences* **20**, 4274 (2019).

28. M. J. Snyder, Cytochrome P450 enzymes belonging to the CYP4 family from marine invertebrates. *Biochemical and Biophysical Research Communications* **249**, 187–190 (1998).

29. R. Feyereisen, Insect P450 Enzymes. *Annual Review of Entomology* **44**, 507–533 (1999).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

30. E.-J. Won, *et al.*, Expression of three novel cytochrome P450 (CYP) and antioxidative genes from the polychaete, *Perinereis nuntia* exposed to water accommodated fraction (WAF) of Iranian crude oil and Benzo[α]pyrene. *Marine Environmental Research* **90**, 75–84 (2013).

31. D. L. Geiser, J. J. Winzerling, Insect transferrins: Multifunctional proteins. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 437–451 (2012).

32. T. Obuchi, *et al.*, Armillaric acid, a new antibiotic produced by *Armillaria mellea*. *Planta medica* **56**, 198–201 (1990).

33. J. Šmidrkal, T. Karlová, V. Filip, M. Zárubová, I. Hrádková, Antimicrobial properties of 11-cyclohexylundecanoic acid. *Czech Journal of Food Sciences* **27 (2009)**, 463–469 (2009).

34. G. W. Felton, C. B. Summers, Antioxidant systems in insects. *Archives of Insect Biochemistry and Physiology* **29**, 187–197 (1995).

35. K. S. Lee, *et al.*, Transferrin inhibits stress-induced apoptosis in a beetle. *Free Radical Biology and Medicine* **41**, 1151–1161 (2006).

36. H. Ikeda, Y. Togashi, Aging, cancer, and antitumor immunity. *Int J Clin Oncol* **27**, 316–322 (2022).

37. L. S. Haddad, L. Kelbert, A. J. Hulbert, Extended longevity of queen honey bees compared to workers is associated with peroxidation-resistant membranes. *Experimental Gerontology* **42**, 601–609 (2007).

38. R. Pamplona, G. Barja, M. Portero-Otín, Membrane fatty acid unsaturation, protection against oxidative stress, and maximum life span. *Annals of the New York Academy of Sciences* **959**, 475–490 (2002).

39. A. J. Hulbert, R. Pamplona, R. Buffenstein, W. A. Buttemer, Life and death: metabolic rate, membrane composition, and life span of animals. *Physiological Reviews* **87**, 1175–1213 (2007).

40. M. Chapuisat, L. Keller, Division of labour influences the rate of ageing in weaver ant workers. *Proceedings of the Royal Society of London B: Biological Sciences* **269**, 909–913 (2002).

41. P. Kohlmeier, *et al.*, Intrinsic worker mortality depends on behavioral caste and the queens' presence in a social insect. *The Science of Nature* **104** (2017).

42. M. J. Toth, A. Tchernof, Lipid metabolism in the elderly. *Eur J Clin Nutr* **54**, S121–S125 (2000).

43. L. I. Gilbert, "Lipid Metabolism and Function in Insects" in *Advances in Insect Physiology*, (Elsevier, 1967), pp. 69–211.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

44. D. W. Stanley-Samuelson, R. A. Jurenka, C. Cripps, G. J. Blomquist, M. de Renobales, Fatty acids in insects: Composition, metabolism, and biological significance. *Arch. Insect Biochem. Physiol.* **9**, 1–33 (1988).

45. K. H. Lockey, Lipids of the insect cuticle: origin, composition and function. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **89**, 595–645 (1988).

46. B. J. Sinclair, K. E. Marshall, The many roles of fats in overwintering insects. *Journal of Experimental Biology* **221**, jeb161836 (2018).

47. Q. Sun, J. Li, F. Gao, New insights into insulin: The anti-inflammatory effect and its clinical relevance. *World J Diabetes* **5**, 89–96 (2014).

48. M. A. Cooper, *et al.*, Reduced mitochondrial reactive oxygen species production in peripheral nerves of mice fed a ketogenic diet. *Experimental Physiology* **103**, 1206–1212 (2018).

49. C. Xu, *et al.*, Feeding restriction alleviates high carbohydrate diet-induced oxidative stress and inflammation of *Megalobrama amblycephala* by activating the AMPK-SIRT1 pathway. *Fish & Shellfish Immunology* **92**, 637–648 (2019).

50. M. B. Blackburn, M. J. Loeb, E. Clark, H. Jaffe, Stimulation of midgut stem cell proliferation by *Manduca sexta* α-arylphorin. *Archives of Insect Biochemistry and Physiology* **55**, 26–32 (2004).

51. R. S. Hakim, *et al.*, Growth and mitogenic effects of arylphorin in vivo and in vitro. *Archives of Insect Biochemistry and Physiology* **64**, 63–73 (2007).

52. A. Castagnola, *et al.*, Alpha-arylphorin is a mitogen in the Heliothis virescens midgut cell secretome upon Cry1Ac intoxication. *PeerJ* **5**, e3886 (2017).

53. M. Quque, *et al.*, Division of labour in the black garden ant (*Lasius niger*) leads to three distinct proteomes. *Journal of Insect Physiology* **117**, 103907 (2019).

54. M. Quque, *et al.*, Eusociality is linked to caste-specific differences in metabolism, immune system, and somatic maintenance-related processes in an ant species. *Cell. Mol. Life Sci.* **79**, 29 (2021).

55. C. Buehlmann, P. Graham, B. S. Hansson, M. Knaden, Desert ants locate food by combining high sensitivity to food odors with extensive crosswind runs. *Current Biology* **24**, 960–964 (2014).

56. H.-L. Qiu, D.-F. Cheng, A chemosensory protein gene Si-CSP1 associated with necrophoric behavior in red imported fire ants (hymenoptera: formicidae). *J Econ Entomol* **110**, 1284–1290 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

57. Q. Sun, K. F. Haynes, X. Zhou, Managing the risks and rewards of death in eusocial insects. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170258 (2018).

58. D. M. Gordon, Dependence of necrophoric response to oleic acid on social context in the ant, *Pogonomyrmex badius*. *J Chem Ecol* **9**, 105–111 (1983).

59. Q. Sun, X. Zhou, Corpse management in social insects. *Int J Biol Sci* **9**, 313–321 (2013).

60. L. Diez, L. Moquet, C. Detrain, Post-mortem changes in chemical profile and their influence on corpse removal in ants. *J Chem Ecol* **39**, 1424–1432 (2013).

61. Q. Sun, K. F. Haynes, X. Zhou, Dynamic changes in death cues modulate risks and rewards of corpse management in a social insect. *Functional Ecology* **31**, 697–706 (2017).

62. S. Cremer, S. A. O. Armitage, P. Schmid-Hempel, Social immunity. *Current Biology* **17**, R693–R702 (2007).

63. N. Stroeymeyt, *et al.*, Social network plasticity decreases disease transmission in a eusocial insect. *Science* **362**, 941–945 (2018).

64. J. Heinze, B. Walter, Moribund ants leave their nests to die in social isolation. *Current Biology* **20**, 249–252 (2010).

65. N. Bos, T. Lefèvre, A. B. Jensen, P. D'Ettorre, Sick ants become unsociable. *Journal of Evolutionary Biology* **25**, 342–351 (2012).

66. A. Seluanov, *et al.*, Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. *Aging Cell* **7**, 813–823 (2008).

67. V. Gorbunova, A. Seluanov, Coevolution of telomerase activity and body mass in mammals: From mice to beavers. *Mechanisms of Ageing and Development* **130**, 3–9 (2009).

68. X. Tian, *et al.*, Evolution of telomere maintenance and tumour suppressor mechanisms across mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20160443 (2018).

69. A. Seluanov, V. N. Gladyshev, J. Vijg, V. Gorbunova, Mechanisms of cancer resistance in long-lived mammals. *Nature Reviews Cancer* **18**, 433–441 (2018).

70. J. Juneja, P. J. Casey, Role of G12 proteins in oncogenesis and metastasis. *British Journal of Pharmacology* **158**, 32–40 (2009).

71. H.-X. Wang, Q. Li, C. Sharma, K. Knoblich, M. E. Hemler, Tetraspanin protein contributions to cancer. *Biochemical Society Transactions* **39**, 547–552 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

72. C. A. Aktipis, *et al.*, Cancer across the tree of life: cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140219 (2015).

73. J.-F. Lemaître, *et al.*, Eco-evolutionary perspectives of the dynamic relationships linking senescence and cancer. *Functional Ecology* **34**, 141–152 (2020).

74. E. A. Konorov, *et al.*, Genomic exaptation enables *Lasius niger* adaptation to urban environments. *BMC Evolutionary Biology* **17**, 39 (2017).

75. I. M. Grześ, M. Okrutniak, J. Grzegorzek, The size-dependent division of labour in monomorphic ant *Lasius niger*. *European Journal of Soil Biology* **77**, 1–3 (2016).

76. M. Okrutniak, B. Rom, F. Turza, I. M. Grześ, Body size differences between foraging and intranidal workers of the monomorphic ant *Lasius niger*. *Insects* **11**, 433 (2020).

77. B. H. Kramer, R. Schaible, A. Scheuerlein, Worker lifespan is an adaptive trait during colony establishment in the long-lived ant *Lasius niger*. *Exp. Gerontol.* **85**, 18–23 (2016).

78. C. R. Ferreira Brandao, Sequential ethograms along colony development of *Odontomachus affinis* Guérin (Hymenoptera, Formicidae, Ponerinae). *Ins. Soc* **30**, 193–203 (1983).

79. C. T. Holbrook, P. M. Barden, J. H. Fewell, Division of labor increases with colony size in the harvester ant Pogonomyrmex californicus. *Behav Ecol* **22**, 960–966 (2011).

80. J. Cox, *et al.*, Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513–2526 (2014).

81. C. Carapito, *et al.*, MSDA, a proteomics software suite for in-depth Mass Spectrometry Data Analysis using grid computing. *Proteomics* **14**, 1014–1019 (2014).

82. E. V. Kriventseva, *et al.*, OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).

83. A. L. Mitchell, *et al.*, InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).

84. J. A. Vizcaíno, *et al.*, 2016 update of the PRIDE database and its related tools. *Nucleic acids research* **44**, D447–D456 (2016).

85. R Core Team, *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2019).

86. J. Josse, F. Husson, missMDA: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software* **70**, 1–31 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.13.495886; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

87. M. Love, Assessment of DESeq2 performance through simulation. *www.huber.embl.de/DESeq2paper* (2014) (April 4, 2019).

88. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

89. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

90. E. L. Schymanski, *et al.*, Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).

91. Z. Pang, *et al.*, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **49**, W388–W396 (2021).

92. D. K. Barupal, O. Fiehn, Chemical similarity enrichment analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Scientific Reports* **7**, 14567 (2017).

93. S. Lê, J. Josse, F. Husson, FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **025**, 1–18 (2008).

94. M. A. Stoffel, S. Nakagawa, H. Schielzeth, rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* **8**, 1639–1644 (2017).

BIBLIOGRAPHIE

1. Salzberg, S. L., Open questions: How many genes do we have? *Bmc Biol* **2018,** *16*.
2. Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I., The Size of the Human Proteome: The Width and Depth. *Int J Anal Chem* **2016,** *2016*.
3. Ramazi, S.; Zahiri, J., Post-translational modifications in proteins: resources, tools and prediction methods. *Database-Oxford* **2021,** *2021*.
4. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schluter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlen, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nat Chem Biol* **2018,** *14* (3), 206-214.
5. Zhang, Y. Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev* **2013,** *113* (4), 2343-2394.
6. Bodzon-Kulakowska, A.; Bierczynska-Krzysik, A.; Dylag, T.; Drabik, A.; Suder, P.; Noga, M.; Jarzebinska, J.; Silberring, J., Methods for samples preparation in proteomic research. *J. Chromatogr. B* **2007,** *849* (1-2), 1-31.
7. Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Dark, C. C., A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020,** *8* (3).
8. Patterson, S. D.; Aebersold, R. H., Proteomics: the first decade and beyond. *Nat Methods* **2003,** *33*, 311-323.
9. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.
10. Karas, M.; Hillenkamp, F., Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Anal Chem* **1988,** *60* (20), 2299-2301.
11. Aballo, T. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Brown, K. A.; Ge, Y., Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and timsTOF Pro. *J Proteome Res* **2021,** *20* (8), 4203-4211.
12. HaileMariam, M.; Eguez, R. V.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. B., S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018,** *17* (9), 2917-2924.
13. Adoni, K. R.; Cunningham, D. L.; Heath, J. K.; Leney, A. C., FAIMS Enhances the Detection of PTM Crosstalk Sites. *J Proteome Res* **2022,** *21* (4), 930-939.
14. Leutert, M.; Entwisle, S. W.; Villen, J., Decoding Post-Translational Modification Crosstalk With Proteomics. *Mol Cell Proteomics* **2021,** *20*.
15. Solari, F. A.; Dell'Aica, M.; Sickmann, A.; Zahedi, R. P., Why phosphoproteomics is still a challenge. *Mol. BioSyst.* **2015,** *11* (6), 1487-1493.
16. Kim, M. S.; Zhong, J.; Pandey, A., Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics* **2016,** *16* (5), 700-14.
17. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; HumpherySmith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio-Technol* **1996,** *14* (1), 61-65.
18. Meissner, F.; Geddes-McAlister, J.; Mann, M.; Bantscheff, M., The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* **2022**.
19. Liu, Z. W.; Liu, Y. L.; Qian, L. L.; Jiang, S. W.; Gai, X. M.; Ye, S.; Chen, Y. H.; Wang, X. M.; Zhai, L. H.; Xu, J.; Pu, C. Y.; Li, J.; He, F. C.; Huang, M.; Tan, M. J., A proteomic and phosphoproteomic

landscape of KRAS mutant cancers identifies combination therapies. *Mol Cell Biol* **2021,** *81* (19), 4076-+.
20. Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M., Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst* **2016,** *2* (3), 185-95.
21. Zhou, H. J.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J. R.; Mohammed, S., Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. *J Proteome Res* **2013,** *12* (1), 260-271.
22. Liu, S. L.; Kerr, E. D.; Pegg, C. L.; Schulz, B. L., Proteomics and glycoproteomics of beer and wine. *Proteomics* **2022**.
23. Bräcker, J.; Brockmeyer, J., Characterization and Detection of Food Allergens Using High-Resolution Mass Spectrometry: Current Status and Future Perspective. *J. Agric. Food Chem.* **2018,** *66* (34), 8935-8940.
24. Quque, M.; Benhaim-Delarbre, M.; Deneubourg, J.-L.; Sueur, C.; Criscuolo, F.; Bertile, F., Division of labour in the black garden ant (Lasius niger) leads to three distinct proteomes. *J. Insect Physiol.* **2019,** *117*, 103907.
25. Grabek, K. R.; Martin, S. L.; Hindle, A. G., Proteomics approaches shed new light on hibernation physiology. *J Comp Physiol B* **2015,** *185* (6), 607-627.
26. James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* **1997,** *30* (4), 279-331.
27. Wu, Y.; Potempa, L. A.; El Kebir, D.; Filep, J. G., C-reactive protein and inflammation: conformational changes affect function. *Biol Chem* **2015,** *396* (11), 1181-1197.
28. Lossl, P.; van de Waterbeemd, M.; Heck, A. J. R., The diverse and expanding role of mass spectrometry in structural and molecular biology. *Embo J* **2016,** *35* (24), 2634-2657.
29. Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Sheil, M. M., Electrospray ionization mass spectrometry of oligonucleotide complexes with drugs, metals, and proteins. *Mass Spectrom Rev* **2001,** *20* (2), 61-87.
30. Zhang, J.; Malmirchegini, G. R.; Clubb, R. T.; Loo, J. A., Native top-down mass spectrometry for the structural characterization of human hemoglobin. *Eur J Mass Spectrom* **2015,** *21* (3), 221-231.
31. Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z. Q.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B. F.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Pasa-Tolic, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* **2019,** *16* (7), 587-594.
32. Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., The One Hour Yeast Proteome. *Mol Cell Proteomics* **2014,** *13* (1), 339-347.
33. Carillo, S.; Millán Martín, S.; Bones, B., High-throughput peptide mapping of trastuzumab using a tandem LC-MS workflow. Application note ThermoScientific, 2020.
34. Whelan, C. D.; Mattsson, N.; Nagle, M. W.; Vijayaraghavan, S.; Hyde, C.; Janelidze, S.; Stomrud, E.; Lee, J.; Fitz, L.; Samad, T. A.; Ramaswamy, G.; Margolin, R. A.; Malarstig, A.; Hansson, O., Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun* **2019,** *7* (1), 169.
35. Tascher, G.; Burban, A.; Camus, S.; Plumel, M.; Chanon, S.; Le Guevel, R.; Shevchenko, V.; Van Dorsselaer, A.; Lefai, E.; Guguen-Guillouzo, C.; Bertile, F., In-Depth Proteome Analysis Highlights HepaRG Cells as a Versatile Cell System Surrogate for Primary Human Hepatocytes. *Cells* **2019,** *8* (2).
36. Chazarin, B.; Storey, K. B.; Ziemianin, A.; Chanon, S.; Plumel, M.; Chery, I.; Durand, C.; Evans, A. L.; Arnemo, J. M.; Zedrosser, A.; Swenson, J. E.; Gauquelin-Koch, G.; Simon, C.; Blanc, S.; Lefai, E.; Bertile, F., Metabolic reprogramming involving glycolysis in the hibernating brown bear skeletal muscle. *Front. Zool.* **2019,** *16* (12).
37. Gomez-Serrano, M.; Camafeita, E.; Loureiro, M.; Peral, B., Mitoproteomics: Tackling Mitochondrial Dysfunction in Human Disease. *Oxid Med Cell Longev* **2018,** *2018*.

38. Canas, B.; Pineiro, C.; Calvo, E.; Lopez-Ferrer, D.; Gallardo, J. M., Trends in sample preparation for classical and second generation proteomics. *J Chromatogr A* **2007,** *1153* (1-2), 235-258.
39. Muller, T.; Kalxdorf, M.; Longuespee, R.; Kazdal, D. N.; Stenzinger, A.; Krijgsveld, J., Automated sample preparation with SP3 for low-input clinical proteomics. *Mol Syst Biol* **2020,** *16* (1), e9111.
40. Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J., Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **2014,** *10* (10).
41. Leon, I. R.; Schwammle, V.; Jensen, O. N.; Sprenger, R. R., Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol Cell Proteomics* **2013,** *12* (10), 2992-3005.
42. Feist, P.; Hummon, A. B., Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *International Journal of Molecular Sciences* **2015,** *16* (2), 3537-3563.
43. Ishikawa, H.; Kim, S.; Kwak, K.; Wakasugi, K.; Fayer, M. D., Disulfide bond influence on protein structural dynamics probed with 2D-IR vibrational echo spectroscopy. *P Natl Acad Sci USA* **2007,** *104* (49), 19309-19314.
44. Mthembu, S. N.; Sharma, A.; Albericio, F.; de la Torre, B. G., Breaking a Couple: Disulfide Reducing Agents. *Chembiochem* **2020,** *21* (14), 1947-1954.
45. Muller, T.; Winter, D., Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol Cell Proteomics* **2017,** *16* (7), 1173-1187.
46. Vandermarliere, E.; Mueller, M.; Martens, L., Getting Intimate with Trypsin, the Leading Protease in Proteomics. *Mass Spectrom Rev* **2013,** *32* (6), 453-465.
47. Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R., Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **2016,** *11* (5), 993-1006.
48. Tabb, D. L.; Huang, Y. Y.; Wysocki, V. H.; Yates, J. R., Influence of basic residue content on fragment ion peak intensities in low-energy - collision-induced dissociation spectra of peptides. *Anal. Chem.* **2004,** *76* (5), 1243-1248.
49. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000,** *35* (12), 1399-1406.
50. Naldi, M.; Tramarin, A.; Bartolini, M., Immobilized enzyme-based analytical tools in the -omics era: Recent advances. *J Pharmaceut Biomed* **2018,** *160*, 222-237.
51. Dau, T.; Bartolomucci, G.; Rappsilber, J., Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal Chem* **2020,** *92* (14), 9523-9527.
52. Hakobyan, A.; Schneider, M. B.; Liesack, W.; Glatter, T., Efficient Tandem LysC/Trypsin Digestion in Detergent Conditions. *Proteomics* **2019,** *19* (20).
53. Glatter, T.; Ludwig, C.; Ahrne, E.; Aebersold, R.; Heck, A. J.; Schmidt, A., Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* **2012,** *11* (11), 5145-56.
54. Kim, S. C.; Chen, Y.; Mirza, S.; Xu, Y. D.; Lee, J.; Liu, P. S.; Zhao, Y. M., A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea. *J Proteome Res* **2006,** *5* (12), 3446-3452.
55. Tubaon, R. M.; Haddad, P. R.; Quirino, J. P., Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics* **2017,** *17* (20).
56. Waas, M.; Pereckas, M.; Jones Lipinski, R. A.; Ashwood, C.; Gundry, R. L., SP2: Rapid and Automatable Contaminant Removal from Peptide Samples for Proteomic Analyses. *J Proteome Res* **2019,** *18* (4), 1644-1656.
57. Granvogl, B.; Ploscher, M.; Eichacker, L. A., Sample preparation by in-gel digestion for mass spectrometry-based proteomics. *Anal Bioanal Chem* **2007,** *389* (4), 991-1002.

58. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Protoc* **2009,** *6* (5), 359-362.
59. Duong, V. A.; Park, J. M.; Lee, H., A review of suspension trapping digestion method in bottom-up proteomics. *J Sep Sci* **2022**.
60. Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **2014,** *11* (3), 319-U300.
61. Ding, H.; Fazelinia, H.; Spruce, L. A.; Weiss, D. A.; Zderic, S. A.; Seeholzer, S. H., Urine Proteomics: Evaluation of Different Sample Preparation Workflows for Quantitative, Reproducible, and Improved Depth of Analysis. *J Proteome Res* **2020,** *19* (4), 1857-1862.
62. Zougman, A.; Selby, P. J.; Banks, R. E., Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **2014,** *14* (9), 1006-0.
63. Prianichnikov, N.; Koch, H.; Koch, S.; Lubeck, M.; Heilig, R.; Brehmer, S.; Fischer, R.; Cox, J., MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Mol Cell Proteomics* **2020,** *19* (6), 1058-1069.
64. Yu, Y. B.; O'Rourke, A.; Lin, Y. H.; Singh, H.; Eguez, R. V.; Beyhan, S.; Nelson, K. E., Predictive Signatures of 19 Antibiotic-Induced Escherichia coli Proteomes. *Acs Infect Dis* **2020,** *6* (8), 2120-2129.
65. Gibbs, K. D.; Washington, E. J.; Jaslow, S. L.; Bourgeois, J. S.; Foster, M. W.; Guo, R.; Brennan, R. G.; Ko, D. C., The Salmonella Secreted Effector SarA/SteE Mimics Cytokine Receptor Signaling to Activate STAT3. *Cell Host Microbe* **2020,** *27* (1), 129-+.
66. Kuras, M.; Woldmar, N.; Kim, Y.; Hefner, M.; Malm, J.; Moldvay, J.; Dome, B.; Fillinger, J.; Pizzatti, L.; Gil, J.; Marko-Varga, G.; Rezeli, M., Proteomic Workflows for High-Quality Quantitative Proteome and Post-Translational Modification Analysis of Clinically Relevant Samples from Formalin-Fixed Paraffin-Embedded Archives. *J Proteome Res* **2021,** *20* (1), 1027-1039.
67. Marchione, D. M.; Ilieva, I.; Devins, K.; Sharpe, D.; Pappin, D. J.; Garcia, B. A.; Wilson, J. P.; Wojcik, J. B., HYPERsoI: High-Quality Data from Archival FFPE Tissue for Clinical Proteomics. *J Proteome Res* **2020,** *19* (2), 973-983.
68. Yang, Y. X.; Anderson, E.; Zhang, S., Evaluation of six sample preparation procedures for qualitative and quantitative proteomics analysis of milk fat globule membrane. *Electrophoresis* **2018,** *39* (18), 2332-2339.
69. Christakopoulos, C.; Cehofski, L. J.; Christensen, S. R.; Vorum, H.; Honore, B., Proteomics reveals a set of highly enriched proteins in epiretinal membrane compared with inner limiting membrane. *Exp Eye Res* **2019,** *186*.
70. Ludwig, K. R.; Schroll, M. M.; Hummon, A. B., Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J Proteome Res* **2018,** *17* (7), 2480-2490.
71. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U., Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* **2017,** *16* (11), 4060-4072.
72. Lefeuvre, B.; Cantero, P.; Ehret-Sabatier, L.; Lenormand, C.; Barthel, C.; Po, C.; Parveen, N.; Grillon, A.; Jaulhac, B.; Boulanger, N., Effects of topical corticosteroids and lidocaine on Borrelia burgdorferi sensu lato in mouse skin: potential impact to human clinical trials. *Sci Rep-Uk* **2020,** *10* (1).
73. Zhang, X.; Sadowski, P.; Punyadeera, C., Evaluation of sample preparation methods for label-free quantitative profiling of salivary proteome. *J Proteomics* **2020,** *210*.
74. Hayoun, K.; Gouveia, D.; Grenga, L.; Pible, O.; Armengaud, J.; Alpha-Bazin, B., Evaluation of Sample Preparation Methods for Fast Proteotyping of Microorganisms by Tandem Mass Spectrometry. *Front Microbiol* **2019,** *10*.
75. Alpert, A. J., Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **2008,** *80* (1), 62-76.
76. Batth, T. S.; Tollenaere, M. A. X.; Ruther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V., Protein Aggregation Capture on Microparticles

Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics* **2019,** *18* (5), 1027-1035.
77. Hughes, C. S.; Moggridge, S.; Muller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J., Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **2019,** *14* (1), 68-+.
78. Paulo, J. A.; Navarrete-Perea, J.; Gygi, S. P., Multiplexed proteome profiling of carbon source perturbations in two yeast species with SL-SP3-TMT. *J Proteomics* **2020,** *210*.
79. Navarrete-Perea, J.; Gygi, S. P.; Paulo, J. A., Growth media selection alters the proteome profiles of three model microorganisms. *J Proteomics* **2021,** *231*.
80. Griesser, E.; Wyatt, H.; Ten Have, S.; Stierstorfer, B.; Lenter, M.; Lamond, A. I., Quantitative Profiling of the Human Substantia Nigra Proteome from Laser-capture Microdissected FFPE Tissue. *Mol Cell Proteomics* **2020,** *19* (5), 839-851.
81. Hughes, C. S.; McConechy, M. K.; Cochrane, D. R.; Nazeran, T.; Karnezis, A. N.; Huntsman, D. G.; Morin, G. B., Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. *Sci Rep-Uk* **2016,** *6*.
82. Kaleja, P.; Emmert, H.; Gerstel, U.; Weidinger, S.; Tholey, A., Evaluation and improvement of protein extraction methods for analysis of skin proteome by noninvasive tape stripping. *J Proteomics* **2020,** *217*.
83. Gonzalez-Lozano, M. A.; Koopmans, F.; Paliukhovich, I.; Smit, A. B.; Li, K. W., A Fast and Economical Sample Preparation Protocol for Interaction Proteomics Analysis. *Proteomics* **2019,** *19* (9).
84. Cleland, T. P., Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *J Proteome Res* **2018,** *17* (11), 3976-3983.
85. Naihui, W.; Samantha, B.; Peter, D.; Sandra, H.; Maxim, K.; Sindy, L.; Oshan, W.; Stefano, G.; Michael, C.; Liora, H. K.; Matthew, S.; Glenn, S.; Michael, S.; Kristine, R. K.; Katerina, D., Testing the efficacy and comparability of ZooMS protocols on archaeological bone. *J Proteomics* **2021,** *233*.
86. Araujo, M. J.; Sousa, M. L.; Felpeto, A. B.; Turkina, M. V.; Fonseca, E.; Martins, J. C.; Vasconcelos, V.; Campos, A., Comparison of Sample Preparation Methods for Shotgun Proteomic Studies in Aquaculture Species. *Proteomes* **2021,** *9* (4).
87. Mikulasek, K.; Konecna, H.; Potesil, D.; Holankova, R.; Havlis, J.; Zdrahal, Z., SP3 Protocol for Proteomic Plant Sample Preparation Prior LC-MS/MS. *Front Plant Sci* **2021,** *12*.
88. Balotf, S.; Wilson, R.; Tegg, R. S.; Nichols, D. S.; Wilson, C. R., Optimisation of Sporosori Purification and Protein Extraction Techniques for the Biotrophic Protozoan Plant Pathogen Spongospora subterranea. *Molecules* **2020,** *25* (14).
89. Dagley, L. F.; Infusini, G.; Larsen, R. H.; Sandow, J. J.; Webb, A. I., Universal Solid-Phase Protein Preparation (USP3) for Bottom-up and Top-down Proteomics. *J Proteome Res* **2019,** *18* (7), 2915-2924.
90. Yang, Z. C.; Shen, X. J.; Chen, D. Y.; Sun, L. L., Toward a Universal Sample Preparation Method for Denaturing Top-Down Proteomics of Complex Proteomes. *J Proteome Res* **2020,** *19* (8), 3315-3325.
91. Zecha, J.; Lee, C. Y.; Bayer, F. P.; Meng, C.; Vincent; Zerweck, J.; Schnatbaum, K.; Michler, T.; Pichlmair, A.; Ludwig, C.; Kuster, B., Data, Reagents, Assays and Merits of Proteomics for SARS-CoV-2 Research and Testing. *Mol Cell Proteomics* **2020,** *19* (9).
92. Johnston, H. E.; Yadav, K.; Kirkpatrick, J. M.; Biggs, G. S.; Oxley, D.; Kramer, H. B.; Samant, R. S., Solvent Precipitation SP3 (SP4) Enhances Recovery for Proteomics Sample Preparation without Magnetic Beads. *Anal. Chem.* **2022,** *94* (29), 10320-10328.
93. Moggridge, S.; Sorensen, P. H.; Morin, G. B.; Hughes, C. S., Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. *J Proteome Res* **2018,** *17* (4), 1730-1740.
94. Smith, L. M.; Kelleher, N. L.; Proteomics, C. T. D., Proteoform: a single term describing protein complexity. *Nat Protoc* **2013,** *10* (3), 186-187.
95. Geyer, P. E.; Holdt, L. M.; Teupser, D.; Mann, M., Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol* **2017,** *13* (9), 942.

96. Channaveerappa, D.; Wetie, A. G. N.; Darie, C. C., Bottlenecks in Proteomics: An Update. *Adv Exp Med Biol* **2019,** *1140*, 753-769.
97. Zubarev, R. A., The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* **2013,** *13* (5), 723-726.
98. Gianazza, E.; Miller, I.; Palazzolo, L.; Parravicini, C.; Eberini, I., With or without you — Proteomics with or without major plasma/serum proteins. *J Proteomics* **2016,** *140*, 62-80.
99. Lee, P. Y.; Osman, J.; Low, T. Y.; Jamal, R., Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery. *Bioanalysis* **2019,** *11* (19), 1799-1812.
100. Bludau, I.; Heusel, M.; Frank, M.; Rosenberger, G.; Hafen, R.; Banaei-Esfahani, A.; van Drogen, A.; Collins, B.; Gstaiger, M.; Aebersold, R., Complex-centric proteome profiling by SEC-SWATH-MS for the parallel detection of hundreds of protein complexes. *Nat Protoc* **2020,** *15* (8), 2341-+.
101. Guo, J. H.; Wu, C. H.; Lin, X. Y.; Zhou, J.; Zhang, J. Y.; Zheng, W. T.; Wang, T.; Cui, Y. Z., Establishment of a simplified dichotomic size-exclusion chromatography for isolating extracellular vesicles toward clinical applications. *J Extracell Vesicles* **2021,** *10* (11).
102. Fekete, S.; Beck, A.; Veuthey, J. L.; Guillarme, D., Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *J Pharmaceut Biomed* **2014,** *101*, 161-173.
103. Gilar, M.; Olivova, P.; Chakraborty, A. B.; Jaworski, A.; Geromanos, S. J.; Gebler, J. C., Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum. *Electrophoresis* **2009,** *30* (7), 1157-1167.
104. Kovacs, J. M.; Mant, C. T.; Hodges, R. S., Determination of intrinsic hydrophilicity/hydrophobicity of amino acid side chains in peptides in the absence of nearest-neighbor or comformational effects. *Biopolymers* **2006,** *84* (3), 283-297.
105. Hao, P. L.; Ren, Y.; Dutta, B.; Sze, S. K., Comparative evaluation of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and high-pH reversed phase (Hp-RP) chromatography in profiling of rat kidney proteome. *J Proteomics* **2013,** *82*, 254-262.
106. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* **2014,** *13* (12), 6176-86.
107. Barbarics, B.; Eildermann, K.; Kaderali, L.; Cyganek, L.; Plessmann, U.; Bodemeyer, J.; Paul, T.; Strobel, P.; Urlaub, H.; Tirilomis, T.; Lenz, C.; Bohnenberger, H., Proteomic mapping of atrial and ventricular heart tissue in patients with aortic valve stenosis. *Sci Rep* **2021,** *11* (1), 24389.
108. Murillo, J. R.; Kuras, M.; Rezeli, M.; Milliotis, T.; Betancourt, L.; Marko-Varga, G., Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. *PLOS ONE* **2018,** *13* (12), e0208562.
109. Kuras, M.; Betancourt, L. H.; Rezeli, M.; Rodriguez, J.; Szasz, M.; Zhou, Q.; Miliotis, T.; Andersson, R.; Marko-Varga, G., Assessing Automated Sample Preparation Technologies for High-Throughput Proteomics of Frozen Well Characterized Tissues from Swedish Biobanks. *J Proteome Res* **2019,** *18* (1), 548-556.
110. Kim, H.; Dan, K.; Shin, H.; Lee, J.; Wang, J. I.; Han, D., An efficient method for high-pH peptide fractionation based on C18 StageTips for in-depth proteome profiling. *Analytical Methods* **2019,** *11* (36), 4693-4698.
111. Koenig, C.; Martinez-Val, A.; Franciosa, G.; Olsen, J. V., Optimal analytical strategies for sensitive and quantitative phosphoproteomics using TMT-based multiplexing. *Proteomics* **2022**.
112. Ping, L.; Kundinger, S. R.; Duong, D. M.; Yin, L.; Gearing, M.; Lah, J. J.; Levey, A. I.; Seyfried, N. T., Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data* **2020,** *7* (1), 315.
113. Morgenstern, M.; Stiller, S. B.; Lübbert, P.; Peikert, C. D.; Dannenmaier, S.; Drepper, F.; Weill, U.; Höß, P.; Feuerstein, R.; Gebert, M.; Bohnert, M.; van der Laan, M.; Schuldiner, M.; Schütze, C.; Oeljeklaus, S.; Pfanner, N.; Wiedemann, N.; Warscheid, B., Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. *Cell Reports* **2017,** *19* (13), 2836-2852.

114. Law, H. C. H.; Kong, R. P. W.; Szeto, S. S. W.; Zhao, Y.; Zhang, Z. J.; Wang, Y. Q.; Li, G. H.; Quan, Q.; Lee, S. M. Y.; Lam, H. C.; Chu, I. K., A versatile reversed phase-strong cation exchange-reversed phase (RP-SCX-RP) multidimensional liquid chromatography platform for qualitative and quantitative shotgun proteomics. *Analyst* **2015,** *140* (4), 1237-1252.
115. Ritorto, M. S.; Cook, K.; Tyagi, K.; Pedrioli, P. G. A.; Trost, M., Hydrophilic Strong Anion Exchange (hSAX) Chromatography for Highly Orthogonal Peptide Separation of Complex Proteomes. *J Proteome Res* **2013,** *12* (6), 2449-2457.
116. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional separations-based shotgun proteomics. *Chem Rev* **2007,** *107* (8), 3654-3686.
117. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G., Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: An alternative approach to high-resolution peptide separation for shotgun proteome analysis. *J Proteome Res* **2007,** *6* (11), 4363-4373.
118. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** *19* (3), 242-247.
119. Nakamura, T.; Kuromitsu, J.; Oda, Y., Evaluation of comprehensive multidimensional separations using reversed-phase, reversed-phase liquid chromatography/mass spectrometry for shotgun proteomics. *J Proteome Res* **2008,** *7* (3), 1007-1011.
120. Deng, W.; Sha, J.; Plath, K.; Wohlschlegel, J. A., Carboxylate-Modified Magnetic Bead (CMMB)-Based Isopropanol Gradient Peptide Fractionation (CIF) Enables Rapid and Robust Off-Line Peptide Mixture Fractionation in Bottom-Up Proteomics. *Mol Cell Proteomics* **2021,** *20*, 100039.
121. Alexovic, M.; Sabo, J.; Longuespee, R., Automation of single-cell proteomic sample preparation. *Proteomics* **2021,** *21* (23-24).
122. Gallert, C.; Vorberg, E.; Roddelkopf, T.; Junginger, S.; Fleischer, H.; Thurow, K., Evaluation of an Automated Solid-Phase Extraction Method Using Positive Pressure. *Am Lab* **2015,** *47* (2), 28-33.
123. Jeon, J.; Yang, J.; Park, J. M.; Han, N. Y.; Lee, Y. B.; Lee, H., Development of an automated high-throughput sample preparation protocol for LC-MS/MS analysis of glycated peptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018,** *1092*, 88-94.
124. Alexovic, M.; Urban, P. L.; Tabani, H.; Sabo, J., Recent advances in robotic protein sample preparation for clinical analysis and other biomedical applications. *Clin Chim Acta* **2020,** *507*, 104-116.
125. Ctortecka, C.; Hartlmayr, D.; Seth, A.; Mendjan, S.; Tourniaire, G.; Mechtler, K., An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. bioRxiv, 2022.
126. Zhang, H.; Liu, C.; Hua, W. Y.; Ghislain, L. P.; Liu, J. H.; Aschenbrenner, L.; Noell, S.; Dirico, K. J.; Lanyon, L. F.; Steppan, C. M.; West, M.; Arnold, D. W.; Covey, T. R.; Datwani, S. S.; Troutman, M. D., Acoustic Ejection Mass Spectrometry for High-Throughput Analysis. *Anal. Chem.* **2021,** *93* (31), 10850-10861.
127. Zhang, Z. R.; Wu, S.; Stenoien, D. L.; Pasa-Tolic, L., High-Throughput Proteomics. *Annu Rev Anal Chem* **2014,** *7*, 427-454.
128. Bonneil, E.; Pfammatter, S.; Thibault, P., Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *J Mass Spectrom* **2015,** *50* (11), 1181-1195.
129. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Rather, O.; Mann, M., Online Parallel Accumulation Serial Fragmentation (PASEF) with a Novel Trapped on Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018,** *17* (12), 2534-2545.
130. Ogata, K.; Chang, C. H.; Ishihama, Y., Effect of Phosphorylation on the Collision Cross Sections of Peptide Ions in Ion Mobility Spectrometry. *Mass Spectrom (Tokyo)* **2021,** *10*, A0093.
131. Biemann, K., Nomenclature for Peptide Fragment Ions (Positive-Ions). *Methods in Enzymology* **1990,** *193*, 886-887.

132. Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J., Influence of secondary structure on the fragmentation of protonated peptides. *J Am Chem Soc* **1999,** *121* (22), 5142-5154.
133. Wiesner, J.; Premsler, T.; Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008,** *8* (21), 4466-4483.
134. Halim, A.; Rüetschi, U.; Larson, G.; Nilsson, J., LC–MS/MS Characterization of O-Glycosylation Sites and Glycan Structures of Human Cerebrospinal Fluid Glycoproteins. *J Proteome Res* **2013,** *12* (2), 573-584.
135. Brodbelt, J. S.; Morrison, L. J.; Santos, I., Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem Rev* **2020,** *120* (7), 3328-3380.
136. Wang, L.; Nwosu, C.; Gao, Y. F.; Zhu, M. M., Signature Ions Triggered Electron-Transfer/Higher-Energy Collisional Dissociation (EThcD) for Specific and Confident Glycation Site Mapping in Therapeutic Proteins. *J Am Soc Mass Spectrom* **2020,** *31* (3), 473-478.
137. Frese, C. K.; Zhou, H. J.; Taus, T.; Altelaar, A. F. M.; Mechter, K.; Heck, A. J. R.; Mohammed, S., Unambiguous Phosphosite Localization using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *J Proteome Res* **2013,** *12* (3), 1520-1525.
138. Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R., Strong Cation Exchange-based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-translational Modifications. *Mol Cell Proteomics* **2009,** *8* (1), 190-200.
139. Michalski, A.; Cox, J.; Mann, M., More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS. *J Proteome Res* **2011,** *10* (4), 1785-1793.
140. Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A., Fundamentals and Advances of Orbitrap Mass Spectrometry. 2019; pp 1-40.
141. Hodge, K.; Have, S. T.; Hutton, L.; Lamond, A. I., Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics* **2013,** *88*, 92-103.
142. Blueggel, M.; Chamrad, D.; Meyer, H. E., Bioinformatics in proteomics. *Curr Pharm Biotechno* **2004,** *5* (1), 79-88.
143. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data - The protein inference problem. *Mol Cell Proteomics* **2005,** *4* (10), 1419-1440.
144. Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B.; Bye-A-Jee, H.; Coetzee, R.; Cukura, A.; Da Silva, A.; Denny, P.; Dogan, T.; Ebenezer, T.; Fan, J.; Castro, L. G.; Garmiri, P.; Georghiou, G.; Gonzales, L.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Jokinen, P.; Joshi, V.; Jyothi, D.; Lock, A.; Lopez, R.; Luciani, A.; Luo, J.; Lussi, Y.; Mac-Dougall, A.; Madeira, F.; Mahmoudy, M.; Menchi, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Oliveira, C. S.; Pundir, S.; Qi, G. Y.; Raj, S.; Rice, D.; Lopez, M. R.; Saidi, R.; Sampson, J.; Sawford, T.; Speretta, E.; Turner, E.; Tyagi, N.; Vasudev, P.; Volynkin, V.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Bolleman, J.; Boutet, E.; Breuza, L.; Casals-Casas, C.; de Castro, E.; Echioukh, K. C.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Keller, G.; Kerhornou, A.; Lara, V.; Le Mercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T. B.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Pozzato, M.; Pruess, M.; Rivoire, C.; Sigrist, C.; Sonesson, K.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C. M.; Chen, Y. X.; Garavelli, J. S.; Huang, H. Z.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q. H.; Wang, Y. Q.; Yeh, L. S.; Zhang, J.; Consortium, U., UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **2021,** *49* (D1), D480-D489.
145. O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; McVeigh, D. H. R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y. M.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W. J.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.;

Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H. Z.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **2016,** *44* (D1), D733-D745.
146. Armengaud, J., A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr Opin Microbiol* **2009,** *12* (3), 292-300.
147. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* **2011,** *10* (4), 1794-1805.
148. Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K., MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J Proteome Res* **2014,** *13* (8), 3679-3684.
149. Griss, J.; Perez-Riverol, Y.; Lewis, S.; Tabb, D. L.; Dianes, J. A.; del-Toro, N.; Rurik, M.; Walzer, M.; Kohlbacher, O.; Hermjakob, H.; Wang, R.; Vizcaino, J. A., Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods* **2016,** *13* (8), 651-+.
150. Dorfer, V.; Maltsev, S.; Winkler, S.; Mechtler, K., CharmeRT: Boosting Peptide Identifications by Chimeric Spectra Identification and Retention Time Prediction. *J Proteome Res* **2018,** *17* (8), 2581-2589.
151. Skinner, O. S.; Kelleher, N. L., Illuminating the dark matter of shotgun proteomics. *Nat Biotechnol* **2015,** *33* (7), 717-718.
152. Chick, J. M.; Kolippakkam, D.; Nusinow, D. P.; Zhai, B.; Rad, R.; Huttlin, E. L.; Gygi, S. P., A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol* **2015,** *33* (7), 743-749.
153. Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I., MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **2017,** *14* (5), 513-+.
154. Degroeve, S.; Gabriels, R.; Velghe, K.; Bouwmeester, R.; Tischenko, N.; Martens, L., ionbot: a novel, innovative and sensitive machine learning approach to LC-MS/MS peptide identification. bioRxiv, 2022.
155. Elias, J. E.; Gygi, S. R., Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods Mol Biol* **2010,** *604*, 55-71.
156. Navarro, P.; Vazquez, J., A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases. *J Proteome Res* **2009,** *8* (4), 1792-1796.
157. Bouyssie, D.; Hesse, A. M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez de Peredo, A.; Coute, Y.; Dupierris, V.; Burel, A.; Menetrey, J. P.; Kalaitzakis, A.; Poisat, J.; Romdhani, A.; Burlet-Schiltz, O.; Cianferani, S.; Garin, J.; Bruley, C., Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020,** *36* (10), 3148-3155.
158. Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **2007,** *4* (11), 923-925.
159. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012,** *404* (4), 939-965.
160. Blein-Nicolas, M.; Zivy, M., Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Bba-Proteins Proteom* **2016,** *1864* (8), 883-895.
161. Tyanova, S.; Temu, T.; Cox, J., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016,** *11* (12), 2301-2319.
162. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G.; Consortium, G. O., Gene Ontology: tool for the unification of biology. *Nat Genet* **2000,** *25* (1), 25-29.
163. Carbon, S.; Douglass, E.; Good, B. M.; Unni, D. R.; Harris, N. L.; Mungall, C. J.; Basu, S.; Chisholm, R. L.; Dodson, R. J.; Hartline, E.; Fey, P.; Thomas, P. D.; Albou, L. P.; Ebert, D.; Kesling, M.

J.; Mi, H. Y.; Muruganujan, A.; Huang, X. S.; Mushayahama, T.; LaBonte, S. A.; Siegele, D. A.; Antonazzo, G.; Attrill, H.; Brown, N. H.; Garapati, P.; Marygold, S. J.; Trovisco, V.; Dos Santos, G.; Falls, K.; Tabone, C.; Zhou, P. L.; Goodman, J. L.; Strelets, V. B.; Thurmond, J.; Garmiri, P.; Ishtiaq, R.; Rodriguez-Lopez, M.; Acencio, M. L.; Kuiper, M.; Laegreid, A.; Logie, C.; Lovering, R. C.; Kramarz, B.; Saverimuttu, S. C. C.; Pinheiro, S. M.; Gunn, H.; Su, R. Z.; Thurlow, K. E.; Chibucos, M.; Giglio, M.; Nadendla, S.; Munro, J.; Jackson, R.; Duesbury, M. J.; Del-Toro, N.; Meldal, B. H. M.; Paneerselvam, K.; Perfetto, L.; Porras, P.; Orchard, S.; Shrivastava, A.; Chang, H. Y.; Finn, R. D.; Mitchell, A. L.; Rawlings, N. D.; Richardson, L.; Sangrador-Vegas, A.; Blake, J. A.; Christie, K. R.; Dolan, M. E.; Drabkin, H. J.; Hill, D. P.; Ni, L.; Sitnikov, D. M.; Harris, M. A.; Oliver, S. G.; Rutherford, K.; Wood, V.; Hayles, J.; Bahler, J.; Bolton, E. R.; De Pons, J. L.; Dwinell, M. R.; Hayman, G. T.; Kaldunski, M. L.; Kwitek, A. E.; Laulederkind, S. J. F.; Plasterer, C.; Tutaj, M. A.; Vedi, M.; Wang, S. J.; D'Eustachio, P.; Matthews, L.; Balhoff, J. P.; Aleksander, S. A.; Alexander, M. J.; Cherry, J. M.; Engel, S. R.; Gondwe, F.; Karra, K.; Miyasato, S. R.; Nash, R. S.; Simison, M.; Skrzypek, M. S.; Weng, S.; Wong, E. D.; Feuermann, M.; Gaudet, P.; Morgat, A.; Bakker, E.; Berardini, T. Z.; Reiser, L.; Subramaniam, S.; Huala, E.; Arighi, C. N.; Auchincloss, A.; Axelsen, K.; Argoud-Puy, G.; Bateman, A.; Blatter, M. C.; Boutet, E.; Bowler, E.; Breuza, L.; Bridge, A.; Britto, R.; Bye-A-Jee, H.; Casas, C. C.; Coudert, E.; Denny, P.; Estreicher, A.; Famiglietti, M. L.; Georghiou, G.; Gos, A.; Gruaz-Gumowski, N.; Hatton-Ellis, E.; Hulo, C.; Ignatchenko, A.; Jungo, F.; Laiho, K.; Le Mercier, P.; Lieberherr, D.; Lock, A.; Lussi, Y.; MacDougall, A.; Magrane, M.; Martin, M. J.; Masson, P.; Natale, D. A.; Hyka-Nouspikel, N.; Orchard, S.; Pedruzzi, I.; Pourcel, L.; Poux, S.; Pundir, S.; Rivoire, C.; Speretta, E.; Sundaram, S.; Tyagi, N.; Warner, K.; Zaru, R.; Wu, C. H.; Diehl, A. D.; Chan, J. N.; Grove, C.; Lee, R. Y. N.; Muller, H. M.; Raciti, D.; Van Auken, K.; Sternberg, P. W.; Berriman, M.; Paulini, M.; Howe, K.; Gao, S.; Wright, A.; Stein, L.; Howe, D. G.; Toro, S.; Westerfield, M.; Jaiswal, P.; Cooper, L.; Elser, J.; Consortium, G. O., The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **2021,** *49* (D1), D325-D334.

164. Kanehisa, M.; Furumichi, M.; Sato, Y.; Ishiguro-Watanabe, M.; Tanabe, M., KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* **2021,** *49* (D1), D545-D551.

165. Zahn-Zabal, M.; Michel, P. A.; Gateau, A.; Nikitin, F.; Schaeffer, M.; Audot, E.; Gaudet, P.; Duek, P. D.; Teixeira, D.; de Laval, V. R.; Samarasinghe, K.; Bairoch, A.; Lane, L., The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res* **2020,** *48* (D1), D328-D334.

166. Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R., The PeptideAtlas project. *Nucleic Acids Res* **2006,** *34*, D655-D658.

167. Deutsch, E. W., The PeptideAtlas Project. *Methods Mol Biol* **2010,** *604*, 285-296.

168. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F., Tissue-based map of the human proteome. *Science* **2015,** *347* (6220).

169. Hosack, D. A.; Dennis, G.; Sherman, B. T.; Lane, H. C.; Lempicki, R. A., Identifying biological themes within lists of genes with EASE. *Genome Biol* **2003,** *4* (10).

170. Calderon-Gonzalez K.; Hernandez-Monge J.; Herrera-Aguirre M. E.; P., L.-A. J., Bioinformatics Tools for Proteomics Data Interpretation. In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*, Adv Exp Med Biol., 2016; Vol. 919, pp 281-341.

171. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *J Mol Biol* **1990,** *215* (3), 403-410.

172. Pearson, W. R., Effective protein sequence comparison. *Method Enzymol* **1996,** *266*, 227-258.

173. Pearson, W. R.; Lipman, D. J., Improved Tools for Biological Sequence Comparison. *P Natl Acad Sci USA* **1988,** *85* (8), 2444-2448.

174. von Stechow, L.; Francavilla, C.; Olsen, J. V., Recent findings and technological advances in phosphoproteomics for cells and tissues. *Expert Rev Proteomics* **2015,** *12* (5), 469-87.

175. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **2017,** *40* (2), 271-280.
176. Low, T. Y.; Mohtar, M. A.; Lee, P. Y.; Omar, N.; Zhou, H.; Ye, M., Widening the Bottleneck of Phosphoproteomics: Evolving Strategies for Phosphopeptide Enrichment. *Mass Spectrom Rev* **2021,** *40* (4), 309-333.
177. Qiu, W.; Evans, C. A.; Landels, A.; Pham, T. K.; Wright, P. C., Phosphopeptide enrichment for phosphoproteomic analysis - A tutorial and review of novel materials. *Anal Chim Acta* **2020,** *1129*, 158-180.
178. Smith, L. M.; Agar, J. N.; Chamot-Rooke, J.; Danis, P. O.; Ge, Y.; Loo, J. A.; Pasa-Tolic, L.; Tsybin, Y. O.; Kelleher, N. L.; Proteomics, C. T.-D., The Human Proteoform Project: Defining the human proteome. *Sci Adv* **2021,** *7* (46).
179. Urban, J., A review on recent trends in the phosphoproteomics workflow. From sample preparation to data analysis. *Anal Chim Acta* **2022,** *1199*.
180. Humphrey, S. J.; James, D. E.; Mann, M., Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol Metab* **2015,** *26* (12), 676-687.
181. Humphrey, S. J.; Karayel, O.; James, D. E.; Mann, M., High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. *Nat Protoc* **2018,** *13* (9).
182. Paulo, J. A.; Schweppe, D. K., Advances in quantitative high-throughput phosphoproteomics with sample multiplexing. *Proteomics* **2021,** *21* (9).
183. Needham, E. J.; Parker, B. L.; Burykin, T.; James, D. E.; Humphrey, S. J., Illuminating the dark phosphoproteome. *Sci. Signal.* **2019,** *12* (565), eaau8645.
184. Hasselgren, P. O., Ubiquitination, phosphorylation, and acetylation--triple threat in muscle wasting. *J Cell Physiol* **2007,** *213* (3), 679-89.
185. Needham, E. J.; Hingst, J. R.; Parker, B. L.; Morrison, K. R.; Yang, G.; Onslev, J.; Kristensen, J. M.; Hojlund, K.; Ling, N. M. X. Y.; Oakhill, J. S.; Richter, E. A.; Kiens, B.; Petersen, J.; Pehmoller, C.; James, D. E.; Wojtaszewski, J. F. P.; Humphrey, S. J., Personalized phosphoproteomics identifies functional signaling. *Nat Biotechnol* **2022,** *40* (4), 576-+.
186. Bareja, A.; Draper, J. A.; Katz, L. H.; Lee, D. E.; Grimsrud, P. A.; White, J. P., Chronic caloric restriction maintains a youthful phosphoproteome in aged skeletal muscle. *Mech Ageing Dev* **2021,** *195*, 111443.
187. Martinez-Montanes, F.; Casanovas, A.; Sprenger, R. R.; Topolska, M.; Marshall, D. L.; Moreno-Torres, M.; Poad, B. L. J.; Blanksby, S. J.; Hermansson, M.; Jensen, O. N.; Ejsing, C. S., Phosphoproteomic Analysis across the Yeast Life Cycle Reveals Control of Fatty Acyl Chain Length by Phosphorylation of the Fatty Acid Synthase Complex. *Cell Rep* **2020,** *32* (6), 108024.
188. Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villen, J.; Haas, W.; Sowa, M. E.; Gygi, S. P., A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **2010,** *143* (7), 1174-89.
189. Riley, N. M.; Coon, J. J., Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016,** *88* (1), 74-94.
190. Fila, J.; Honys, D., Enrichment techniques employed in phosphoproteomics. *Amino Acids* **2012,** *43* (3), 1025-1047.
191. Nilsson, C. L.; Dillon, R.; Devakumar, A.; Shi, S. D.; Greig, M.; Rogers, J. C.; Krastins, B.; Rosenblatt, M.; Kilmer, G.; Major, M.; Kaboord, B. J.; Sarracino, D.; Rezai, T.; Prakash, A.; Lopez, M.; Ji, Y.; Priebe, W.; Lang, F. F.; Colman, H.; Conrad, C. A., Quantitative phosphoproteomic analysis of the STAT3/IL-6/HIF1alpha signaling network: an initial study in GSC11 glioblastoma stem cells. *J Proteome Res* **2010,** *9* (1), 430-43.
192. Arribas Diez, I.; Govender, I.; Naicker, P.; Stoychev, S.; Jordaan, J.; Jensen, O. N., Zirconium(IV)-IMAC Revisited: Improved Performance and Phosphoproteome Coverage by Magnetic Microparticles for Phosphopeptide Affinity Enrichment. *J Proteome Res* **2021,** *20* (1), 453-462.
193. Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S., Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* **2013,** *8* (3), 461-80.

194. Reinders, J.; Sickmann, A., State-of-the-art in phosphoproteomics. *Proteomics* **2005,** *5* (16), 4052-4061.
195. Zarei, M.; Sprenger, A.; Gretzmeier, C.; Dengjel, J., Combinatorial Use of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) and Strong Cation Exchange (SCX) Chromatography for In-Depth Phosphoproteome Analysis. *J Proteome Res* **2012,** *11* (8), 4269-4276.
196. Villen, J.; Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* **2008,** *3* (10), 1630-8.
197. Ruprecht, B.; Koch, H.; Medard, G.; Mundt, M.; Kuster, B.; Lemeer, S., Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics* **2015,** *14* (1), 205-215.
198. Wu, S.; Wu, L., Human Breast Cancer Cell Line Phosphoproteome Revealed by an Automated and Highly Selective Enrichment Workflow. Application Note Agilent Technologies, 2018.
199. von Ziegler, L. M.; Floriou-Servou, A.; Waag, R.; Das Gupta, R.; Sturman, O.; Gapp, K.; Maat, C.; Kockmann, T.; Lin, H.; Duss, S.; Privitera, M.; Hinte, L.; von Meyenne, F.; Zeilhofer, H.; Germain, P.; Bohacek, J., Molecular roadmap of the healthy stress response in the mouse hippocampus. bioRxiv, 2021.
200. Post, H.; Penning, R.; Fitzpatrick, M. A.; Garrigues, L. B.; Wu, W.; MacGillavry, H. D.; Hoogenraad, C. C.; Heck, A. J.; Altelaar, A. F., Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **2017,** *16* (2), 728-737.
201. Gao, Y.; Wang, Y. S., A method to determine the ionization efficiency change of peptides caused by phosphorylation. *J Am Soc Mass Spectrom* **2007,** *18* (11), 1973-1976.
202. Potel, C. M.; Lemeer, S.; Heck, A. J. R., Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal Chem* **2019,** *91* (1), 126-141.
203. Rozman, M., Modelling of the gas-phase phosphate group loss and rearrangement in phosphorylated peptides. *J Mass Spectrom* **2011,** *46* (9), 949-955.
204. Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R., Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal. Chem.* **2016,** *88* (4), 2303-2310.
205. Tang, N.; Goodley, P., Characterization of Protein Phosphorylation Using HPLC-Chip Electron Transfer Dissociation Ion Trap Mass Spectrometry Agilent Technologies Application Note, 2006.
206. Li, Z.; Li, M.; Li, X.; Xin, J. Z.; Wang, Y.; Shen, Q. W. W.; Zhang, D. Q., Quantitative phosphoproteomic analysis among muscles of different color stability using tandem mass tag labeling. *Food Chemistry* **2018,** *249*, 8-15.
207. Lin, K. H.; Wilson, G. M.; Blanco, R.; Steinert, N. D.; Zhu, W. G.; Coon, J. J.; Hornberger, T. A., A deep analysis of the proteomic and phosphoproteomic alterations that occur in skeletal muscle after the onset of immobilization. *J Physiol* **2021,** *599* (11), 2887-2906.
208. Dephoure, N.; Gould, K. L.; Gygi, S. P.; Kellogg, D. R., Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell* **2013,** *24* (5), 535-542.
209. Zhang, Y.; Dreyer, B.; Govorukhina, N.; Heberle, A. M.; Koncarevic, S.; Krisp, C.; Opitz, C. A.; Pfander, P.; Bischoff, R.; Schluter, H.; Kwiatkowski, M.; Thedieck, K.; Horvatovich, P. L., Comparative Assessment of Quantification Methods for Tumor Tissue Phosphoproteomics. *Anal Chem* **2022**.
210. Hogrebe, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V., Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat Commun* **2018,** *9*.
211. Stepath, M.; Zulch, B.; Maghnouj, A.; Schork, K.; Turewicz, M.; Eisenacher, M.; Hahn, S.; Sitek, B.; Bracht, T., Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi. *J Proteome Res* **2020,** *19* (2), 926-937.

212. Zhang, F. F.; Ge, W. G.; Ruan, G.; Cai, X.; Guo, T. N., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020,** *20* (17-18).
213. Srinivasan, A.; Sing, J. C.; Gingras, A. C.; Rost, H. L., Improving Phosphoproteomics Profiling Using Data-Independent Mass Spectrometry. *J Proteome Res* **2022,** *21* (8), 1789-1799.
214. Site d'OncoProteomics Laboratory. https://oncoproteomics.nl/research/ (accessed 23/08/2022).
215. Skowronek, P.; Thielert, M.; Voytik, E.; Tanzert, M.; Hansen, F.; Willems, S.; Karayel, Ö.; Brunner, A.; Meier, F.; Mann, M., Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. bioRxiv, 2022.
216. Oliinyk, D.; Meier, F., Ion mobility-resolved phosphoproteomics with dia-PASEF and short gradients. bioRxiv, 2022.
217. Xuan, Y.; Creese, A. J.; Horner, J. A.; Cooper, H. J., High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Commun Mass Sp* **2009,** *23* (13), 1963-1969.
218. Zhao, H. Y.; Cunningham, D. L.; Creese, A. J.; Heath, J. K.; Cooper, H. J., FAIMS and Phosphoproteomics of Fibroblast Growth Factor Signaling: Enhanced Identification of Multiply Phosphorylated Peptides. *J Proteome Res* **2015,** *14* (12), 5077-5087.
219. Press Realease Details du TIMScore. https://ir.bruker.com/press-releases/press-release-details/2022/Bruker-Expands-dia-PASEF-on-the-timsTOF-4D-Proteomics-and-Epiproteomics-Platform-to-Identify-up-to-13000-Protein-Groups-with-1-FDR/default.aspx (accessed 23/08/2022).
220. Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B., Confident Phosphorylation Site Localization Using the Mascot Delta Score. *Mol Cell Proteomics* **2011,** *10* (2).
221. Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K., Universal and Confident Phosphorylation Site Localization Using phosphoRS. *J Proteome Res* **2011,** *10* (12), 5354-5362.
222. Locard-Paulet, M.; Bouyssie, D.; Froment, C.; Burlet-Schiltz, O.; Jensen, L. J., Comparing 22 Popular Phosphoproteomics Pipelines for Peptide Identification and Site Localization. *J Proteome Res* **2020,** *19* (3), 1338-1345.
223. Jiang, X.; Horn, D.; Bomgarden, R.; Schroeder, T.; Viner, R.; Huhmer, A., Evaluation of search engines for phosphopeptide identification and quantitation. Poster Note 64793, Thermo Fisher Scientific, 2016.
224. Ramsbottom, K. A.; Prakash, A.; Riverol, Y. P.; Camacho, O. M.; Martin, M. J.; Vizcaino, J. A.; Deutsch, E. W.; Jones, A. R., Method for Independent Estimation of the False Localization Rate for Phosphoproteomics. *J Proteome Res* **2022**.
225. Fermin, D.; Walmsley, S. J.; Gingras, A. C.; Choi, H.; Nesvizhskii, A. I., LuciPHOr: algorithm for phosphorylation site localization with false localization rate estimation using modified target-decoy approach. *Mol Cell Proteomics* **2013,** *12* (11), 3409-19.
226. Ressa, A.; Fitzpatrick, M.; van den Toorn, H.; Heck, A. J. R.; Altelaar, M., PaDuA: A Python Library for High-Throughput (Phospho)proteomics Data Analysis. *J Proteome Res* **2019,** *18* (2), 576-584.
227. Kim, H. J.; Kim, T.; Hoffman, N. J.; Xiao, D.; James, D. E.; Humphrey, S. J.; Yang, P. Y., PhosR enables processing and functional analysis of phosphoproteomic data. *Cell Reports* **2021,** *34* (8).
228. Ramasamy, P.; Turan, D.; Tichshenko, N.; Hulstaert, N.; Vandermarliere, E.; Vranken, W.; Martens, L., Scop3P: A Comprehensive Resource of Human Phosphosites within Their Full Context. *J Proteome Res* **2020,** *19* (8), 3478-3486.
229. Ramasamy, P.; Vandermarliere, E.; Vranken, W. F.; Martens, L., Panoramic Perspective on Human Phosphosites. *J Proteome Res* **2022,** *21* (8), 1894-1915.
230. Hornbeck, P. V.; Kornhauser, J. M.; Latham, V.; Murray, B.; Nandhikonda, V.; Nord, A.; Skrzypek, E.; Wheeler, T.; Zhang, B.; Gnad, F., 15 years of PhosphoSitePlus(R): integrating post-

translationally modified sites, disease variants and isoforms. *Nucleic Acids Res* **2019,** *47* (D1), D433-D441.
231. Gnad, F.; Gunawardena, J.; Mann, M., PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res* **2011,** *39*, D253-D260.
232. Zhang, Q.; Schepmoes, A. A.; Brock, J. W. C.; Wu, S.; Moore, R. J.; Purvine, S. O.; Baynes, J. W.; Smith, R. D.; Metz, T. O., Improved Methods for the Enrichment and Analysis of Glycated Peptides. *Anal. Chem.* **2008,** *80* (24), 9822-9829.
233. Maillard, L. C., Action des Acides Amines sur les Sucres: Formation des Melanoidines par voie Methodique. **1912,** *154*.
234. Priego Capote, F.; Sanchez, J.-C., Strategies for proteomic analysis of non-enzymatically glycated proteins. *Mass Spectrom Rev* **2009,** *28* (1), 135-146.
235. Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H., Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Bio* **2020,** *21* (12), 729-749.
236. Kim, C. S.; Park, S.; Kim, J., The role of glycation in the pathogenesis of aging and its prevention through herbal products and physical exercise. *J Exerc Nutrition Biochem* **2017,** *21* (3), 55-61.
237. Rabbani, N.; Ashour, A.; Thornalley, P. J., Mass spectrometric determination of early and advanced glycation in biology. *Glycoconj. J.* **2016,** *33* (4), 553-568.
238. Soboleva, A.; Vikhnina, M.; Grishina, T.; Frolov, A., Probing Protein Glycation by Chromatography and Mass Spectrometry: Analysis of Glycation Adducts. *Int. J. Mol. Sci.* **2017,** *18* (12).
239. Soboleva, A.; Schmidt, R.; Vikhnina, M.; Grishina, T.; Frolov, A., Maillard Proteomics: Opening New Pages. *Int. J. Mol. Sci.* **2017,** *18* (12), 2677.
240. Baynes, J. W.; Watkins, N. G.; Fisher, C. I.; Hull, C. J.; Patrick, J. S.; Ahmed, M. U.; Dunn, J. A.; Thorpe, S. R., The Amadori product on protein: structure and reactions. *Prog Clin Biol Res* **1989,** *304*, 43-67.
241. Gugliucci, A., Formation of Fructose-Mediated Advanced Glycation End Products and Their Roles in Metabolic and Inflammatory Diseases. *Adv Nutr ADV NUTR* **2017,** *8* (1), 54-62.
242. Schalkwijk, C. G.; Stehouwer, C. D.; van Hinsbergh, V. W., Fructose-mediated non-enzymatic glycation: sweet coupling or bad modification. *Diabetes Metab Res Rev* **2004,** *20* (5), 369-82.
243. Luevano-Contreras, C.; Chapman-Novakofski, K., Dietary Advanced Glycation End Products and Aging. *Nutrients* **2010,** *2* (12), 1247-1265.
244. Poulsen, M. W.; Hedegaard, R. V.; Andersen, J. M.; de Courten, B.; Bügel, S.; Nielsen, J.; Skibsted, L. H.; Dragsted, L. O., Advanced glycation endproducts in food and their effects on health. *Food and Chemical Toxicology* **2013,** *60*, 10-37.
245. Wei, B.; Berning, K.; Quan, C.; Zhang, Y. T., Glycation of antibodies: Modification, methods and potential effects on biological functions. *MAbs* **2017,** *9* (4), 586-594.
246. Mo, J.; Jin, R.; Yan, Q.; Sokolowska, I.; Lewis, M. J.; Hu, P., Quantitative analysis of glycation and its impact on antigen binding. *MAbs* **2018,** *10* (3), 406-415.
247. Fischer, S.; Hoernschemeyer, J.; Mahler, H. C., Glycation during storage and administration of monoclonal antibody formulations. *Eur J Pharm Biopharm* **2008,** *70* (1), 42-50.
248. Koga, M., Glycated albumin; clinical usefulness. *Clin Chim Acta* **2014,** *433*, 96-104.
249. Beltran Del Rio, M.; Tiwari, M.; Amodu, L. I.; Cagliani, J.; Rodriguez Rilo, H. L., Glycated Hemoglobin, Plasma Glucose, and Erythrocyte Aging. *J Diabetes Sci Technol* **2016,** *10* (6), 1303-1307.
250. Golizeh, M.; Lee, K.; Ilchenko, S.; Osme, A.; Bena, J.; Sadygov, R. G.; Kashyap, S. R.; Kasumov, T., Increased serotransferrin and ceruloplasmin turnover in diet-controlled patients with type 2 diabetes. *Free Radical Bio Med* **2017,** *113*, 461-469.
251. Shin, A.; Vazmitsel, Y.; Connolly, S.; Kabytaev, K., Comprehensive profiling and kinetic studies of glycated lysine residues in human serum albumin. *Anal Bioanal Chem* **2022,** *414* (17), 4861-4875.

252. Kisugi, R.; Kouzuma, T.; Yamamoto, T.; Akizuki, S.; Miyamoto, H.; Someya, Y.; Yokoyama, J.; Abe, I.; Hirai, N.; Ohnishi, A., Structural and glycation site changes of albumin in diabetic patient with very high glycated albumin. *Clin Chim Acta* **2007,** *382* (1-2), 59-64.
253. Ulrich, P.; Cerami, A., Protein glycation, diabetes, and aging. *Recent Prog Horm Res* **2001,** *56*, 1-21.
254. Frolov, A.; Bluher, M.; Hoffmann, R., Glycation sites of human plasma proteins are affected to different extents by hyperglycemic conditions in type 2 diabetes mellitus. *Anal Bioanal Chem* **2014,** *406* (24), 5755-5763.
255. D'Aronco, S.; Crotti, S.; Agostini, M.; Traldi, P.; Chilelli, N. C.; Lapolla, A., The role of mass spectrometry in studies of glycation processes and diabetes management. *Mass Spectrom. Rev.* **2019,** *38* (1), 112-146.
256. Bohlender, J. M.; Franke, S.; Stein, G.; Wolf, G., Advanced glycation end products and the kidney. *Am J Physiol-Renal* **2005,** *289* (4), F645-F659.
257. Vicente Miranda, H.; El-Agnaf, O. M.; Outeiro, T. F., Glycation in Parkinson's disease and Alzheimer's disease. *Mov Disord* **2016,** *31* (6), 782-90.
258. Simm, A., Protein glycation during aging and in cardiovascular disease. *J. Proteom.* **2013,** *92*, 248-259.
259. Rabbani, N.; Thornalley, P. J., Protein glycation ? biomarkers of metabolic dysfunction and early-stage decline in health in the era of precision medicine. *Redox Biology* **2021,** *42*.
260. Jeppsson, J.-O.; Kobold, U.; Barr, J.; Finke, A.; Hoelzel, W.; Hoshino, T.; Miedema, K.; Mosca, A.; Mauri, P.; Paroni, R.; Thienpont, L.; Umemoto, M.; Weykamp, C., Approved IFCC Reference Method for the Measurement of HbA1c in Human Blood. *Clinical Chemistry and Laboratory Medicine* **2002,** *40* (1).
261. Sotokawauchi, A.; Matsui, T.; Higashimoto, Y.; Yamagishi, S.-i., Fructose causes endothelial cell damage via activation of advanced glycation end products–receptor system. *Diabetes and Vascular Disease Research* **2019,** *16* (6), 556-561.
262. Jagadeeshaprasad, M. G.; Venkatasubramani, V.; Unnikrishnan, A. G.; Kulkarni, M. J., Albumin Abundance and Its Glycation Status Determine Hemoglobin Glycation. *ACS Omega* **2018,** *3* (10), 12999-13008.
263. Mitsuhashi, T.; Vlassara, H.; Founds, H. W.; Li, Y. M., Standardizing the immunological measurement of advanced glycation endproducts using normal human serum. *J Immunol Methods* **1997,** *207* (1), 79-88.
264. Hempe, J. M.; McGehee, A. M.; Chalew, S. A., Two-dimensional analysis of glycated hemoglobin heterogeneity in pediatric type 1 diabetes patients. *Anal Biochem* **2013,** *442* (2), 205-212.
265. Leblanc, Y.; Bihoreau, N.; Chevreux, G., Characterization of Human Serum Albumin isoforms by ion exchange chromatography coupled on-line to native mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018,** *1095*, 87-93.
266. Ingram, T.; Zuck, J.; Borges, C. R.; Redig, P.; Sweazea, K. L., Variations in native protein glycation and plasma antioxidants in several birds of prey. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2017,** *210*, 18-28.
267. Ramirez-Boo, M.; Priego-Capote, F.; Hainard, A.; Gluck, F.; Burkhard, P.; Sanchez, J. C., Characterization of the glycated human cerebrospinal fluid proteome. *J Proteomics* **2012,** *75* (15), 4766-82.
268. Priego-Capote, F.; Scherl, A.; Muller, M.; Waridel, P.; Lisacek, F.; Sanchez, J. C., Glycation isotopic labeling with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. *Mol Cell Proteomics* **2010,** *9* (3), 579-92.
269. Priego-Capote, F.; Ramírez-Boo, M.; Finamore, F.; Gluck, F.; Sanchez, J.-C., Quantitative Analysis of Glycated Proteins. *J Proteome Res* **2014,** *13* (2), 336-347.
270. Sun, F.; Suttapitugsakul, S.; Xiao, H.; Wu, R., Comprehensive Analysis of Protein Glycation Reveals Its Potential Impacts on Protein Degradation and Gene Expression in Human Cells. *J Am Soc Mass Spectrom* **2019,** *30* (12), 2480-2490.

271. Frolov, A.; Hoffmann, P.; Hoffmann, R., Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. *J Mass Spectrom* **2006,** *41* (11), 1459-69.
272. Keilhauer, E. C.; Geyer, P. E.; Mann, M., HCD Fragmentation of Glycated Peptides. *J Proteome Res* **2016,** *15* (8), 2881-2890.
273. Hinton, D. J.; Ames, J. M., Site specificity of glycation and carboxymethylation of bovine serum albumin by fructose. *Amino Acids* **2006,** *30* (4), 425-34.
274. Tupe, R. S.; Bangar, N.; Diwan, A.; Changale, D.; Choudhary, S.; Chaware, S., Comparative study of different glycating agents on human plasma and vascular cells. *Mol Biol Rep* **2020,** *47* (1), 521-531.
275. Zhang, L.; Liu, C.-W.; Zhang, Q., Online 2D-LC-MS/MS Platform for Analysis of Glycated Proteome. *Anal. Chem.* **2018,** *90* (2), 1081-1086.
276. Zhang, Q. B.; Tang, N.; Brock, J. W. C.; Mottaz, H. M.; Ames, J. M.; Baynes, J. W.; Smith, R. D.; Metz, T. O., Enrichment and analysis of nonenzymatically glycated peptides: Boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry. *J Proteome Res* **2007,** *6* (6), 2323-2330.
277. Cho, S.; Duong, V. A.; Mok, J. H.; Joo, M.; Park, J. M.; Lee, H., Enrichment and analysis of glycated proteins. *Rev Anal Chem* **2022,** *41* (1), 83-97.
278. Kilhovd, B. K.; Giardino, I.; Torjesen, P. A.; Birkeland, K. I.; Berg, T. J.; Thornalley, P. J.; Brownlee, M.; Hanssen, K. F., Increased serum levels of the specific AGE-compound methylglyoxal-derived hydroimidazolone in patients with type 2 diabetes. *Metabolism-Clinical and Experimental* **2003,** *52* (2), 163-167.
279. Sun, B.; Smialowski, P.; Straub, T.; Imhof, A., Investigation and Highly Accurate Prediction of Missed Tryptic Cleavages by Deep Learning. *J Proteome Res* **2021,** *20* (7), 3749-3757.
280. Lapolla, A.; Fedele, D.; Reitano, R.; Arico, N. C.; Seraglia, R.; Traldi, P.; Marotta, E.; Tonani, R., Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. *J Am Soc Mass Spectrom* **2004,** *15* (4), 496-509.
281. Zhang, Q. B.; Frolov, A.; Tang, N.; Hoffmann, R.; van de Goor, T.; Metz, T. O.; Smith, R. D., Application of electron transfer dissociation mass spectrometry in analyses of non-enzymatically glycated peptides. *Rapid Commun Mass Sp* **2007,** *21* (5), 661-666.
282. Zhang, Q.; Tang, N.; Brock, J. W. C.; Mottaz, H. M.; Ames, J. M.; Baynes, J. W.; Smith, R. D.; Metz, T. O., Enrichment and Analysis of Nonenzymatically Glycated Peptides: Boronate Affinity Chromatography Coupled with Electron-Transfer Dissociation Mass Spectrometry. *J Proteome Res* **2007,** *6* (6), 2323-2330.
283. Polasky, D. A.; Yu, F. C.; Teo, G. C.; Nesvizhskii, A. I., Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat Methods* **2020,** *17* (11), 1125-+.
284. Berger, M. T.; Hemmler, D.; Diederich, P.; Rychlik, M.; Marshall, J. W.; Schmitt-Kopplin, P., Open Search of Peptide Glycation Products from Tandem MassSpectra. *Anal. Chem.* **2022,** *94* (15), 5953-5961.
285. Priego-Capote, F.; Ramirez-Boo, M.; Hoogland, C.; Scherl, A.; Mueller, M.; Lisacek, F.; Sanchez, J. C., Human Hemolysate Glycated Proteome. *Anal. Chem.* **2011,** *83* (14), 5673-5680.
286. Capote, F. P.; Sanchez, J.-C., Strategies for proteomic analysis of non-enzymatically glycated proteins. 12.
287. Espina-Benitez, M. B.; Randon, J.; Demesmay, C.; Dugas, V., Back to BAC: Insights into Boronate Affinity Chromatography Interaction Mechanisms. *Sep Purif Rev* **2018,** *47* (3), 214-228.
288. Spiller, S.; Frolov, A.; Hoffmann, R., Quantification of Specific Glycation Sites in Human Serum Albumin as Prospective Type 2 Diabetes Mellitus Biomarkers. *Protein Peptide Lett* **2017,** *24* (10), 887-896.
289. Bai, X.; Wang, Z. J.; Huang, C. C.; Wang, Z.; Chi, L. L., Investigation of Non-Enzymatic Glycosylation of Human Serum Albumin Using Ion Trap-Time of Flight Mass Spectrometry. *Molecules* **2012,** *17* (8), 8782-8794.
290. Pepaj, M.; Thorsby, P. M., Analysis of Glycated Albumin by On-Line Two-Dimensional Liquid Chromatography Mass Spectrometry. *J Liq Chromatogr R T* **2015,** *38* (1), 20-28.

291. Nowack, J.; Levesque, D. L.; Reher, S.; Dausmann, K. H., Variable Climates Lead to Varying Phenotypes: "Weird" Mammalian Torpor and Lessons From Non-Holarctic Species. *Front Ecol Evol* **2020,** *8*.
292. Ruf, T.; Geiser, F., Daily torpor and hibernation in birds and mammals. *Biol Rev* **2015,** *90* (3), 891-926.
293. Evans, A. L.; Sahlen, V.; Stoen, O. G.; Fahlman, A.; Brunberg, S.; Madslien, K.; Frobert, O.; Swenson, J. E.; Arnemo, J. M., Capture, Anesthesia, and Disturbance of Free-Ranging Brown Bears (Ursus arctos) during Hibernation. *Plos One* **2012,** *7* (7).
294. Bertile, F.; Habold, C.; Le Maho, Y.; Giroud, S., Body Protein Sparing in Hibernators: A Source for Biomedical Innovation. *Front Physiol* **2021,** *12*, 634953.
295. Heldmaier, G.; Ortmann, S.; Elvert, R., Natural hypometabolism during hibernation and daily torpor in mammals. *Respir Physiol Neurobiol* **2004,** *141* (3), 317-29.
296. Craighead, J., Studying Grizzly Habitat by Satellite. *Natl Geogr* **1976,** *150* (1), 148-158.
297. Hellgren, E. C., Physiology of hibernation in bears. *Ursus-Series* **1998,** *10*, 467-477.
298. Lohuis, T. D.; Harlow, H. J.; Beck, T. D., Hibernating black bears (Ursus americanus) experience skeletal muscle protein balance during winter anorexia. *Comp Biochem Physiol B Biochem Mol Biol* **2007,** *147* (1), 20-8.
299. Harlow, H. J.; Lohuis, T.; Beck, T. D.; Iaizzo, P. A., Muscle strength in overwintering bears. *Nature* **2001,** *409* (6823), 997.
300. Alkner, B. A.; Tesch, P. A., Knee extensor and plantar flexor muscle size and function following 90 days of bed rest with or without resistance exercise. *Eur J Appl Physiol* **2004,** *93* (3), 294-305.
301. Geiser, F., Metabolic rate and body temperature reduction during hibernation and daily torpor. *Annu Rev Physiol* **2004,** *66*, 239-74.
302. Toien, O.; Blake, J.; Edgar, D. M.; Grahn, D. A.; Heller, H. C.; Barnes, B. M., Hibernation in Black Bears: Independence of Metabolic Suppression from Body Temperature. *Science* **2011,** *331* (6019), 906-909.
303. Geiser, F., Conserving energy during hibernation. *J Exp Biol* **2016,** *219* (Pt 14), 2086-7.
304. Rolfe, D. F.; Brown, G. C., Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev* **1997,** *77* (3), 731-58.
305. Capt, C.; Passamonti, M.; Breton, S., The human mitochondrial genome may code for more than 13 proteins. *Mitochondrial DNA A* **2016,** *27* (5), 3098-3101.
306. Frezza, C.; Cipolat, S.; Scorrano, L., Organelle isolation: functional mitochondria from mouse liver, muscle and cultured filroblasts. *Nat Protoc* **2007,** *2* (2), 287-295.
307. Kuznetsov, A. V.; Veksler, V.; Gellerich, F. N.; Saks, V.; Margreiter, R.; Kunz, W. S., Analysis of mitochondrial function in situ in permeabilized muscle fibers, tissues and cells. *Nat Protoc* **2008,** *3* (6), 965-976.
308. Bian, Y.; Song, C.; Cheng, K.; Dong, M.; Wang, F.; Huang, J.; Sun, D.; Wang, L.; Ye, M.; Zou, H., An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteomics* **2014,** *96*, 253-62.
309. Chazarin, B.; Ziemianin, A.; Evans, A. L.; Meugnier, E.; Loizon, E.; Chery, I.; Arnemo, J. M.; Swenson, J. E.; Gauquelin-Koch, G.; Simon, C.; Blanc, S.; Lefai, E.; Bertile, F., Limited Oxidative Stress Favors Resistance to Skeletal Muscle Atrophy in Hibernating Brown Bears (Ursus Arctos). *Antioxidants (Basel)* **2019,** *8* (9).
310. Malatesta, M.; Perdoni, F.; Battistelli, S.; Muller, S.; Zancanaro, C., The cell nuclei of skeletal muscle cells are transcriptionally active in hibernating edible dormice. *BMC Cell Biol* **2009,** *10*, 19.
311. Bhatti, J. S.; Kumar, S.; Vijayan, M.; Bhatti, G. K.; Reddy, P. H., 3.2 Mitochondrial Biogenesis. In *Molecular Biology of Aging*, Reddy, P. H., Ed. 2017; Vol. 146, pp 13-46.
312. Barger, J. L.; Brand, M. D.; Barnes, B. M.; Boyer, B. B., Tissue-specific depression of mitochondrial proton leak and substrate oxidation in hibernating arctic ground squirrels. *Am J Physiol-Reg* **2003,** *284* (5), R1306-R1313.
313. Staples, J. F., Metabolic suppression in mammalian hibernation: the role of mitochondria. *J Exp Biol* **2014,** *217* (12), 2032-2036.

314. Bertile, F.; Habold, C.; Le Maho, Y.; Giroud, S., Body Protein Sparing in Hibernators: A Source for Biomedical Innovation. *Frontiers in Physiology* **2021,** *12*.
315. Tinker, D. B.; Harlow, H. J.; Beck, T. D. I., Protein use and muscle-fiber changes in free-ranging, hibernating black bears. *Physiol Zool* **1998,** *71* (4), 414-424.
316. Goodman, C. A.; Mayhew, D. L.; Hornberger, T. A., Recent progress toward understanding the molecular mechanisms that regulate skeletal muscle mass. *Cell Signal* **2011,** *23* (12), 1896-1906.
317. Hizli, A. A.; Chi, Y.; Swanger, J.; Carter, J. H.; Liao, Y.; Welcker, M.; Ryazanov, A. G.; Clurman, B. E., Phosphorylation of eukaryotic elongation factor 2 (eEF2) by cyclin A-cyclin-dependent kinase 2 regulates its inhibition by eEF2 kinase. *Mol Cell Biol* **2013,** *33* (3), 596-604.
318. Hershey, J. W.; Sonenberg, N.; Mathews, M. B., Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* **2012,** *4* (12).
319. Roux, P. P.; Shahbazian, D.; Vu, H.; Holz, M. K.; Cohen, M. S.; Taunton, J.; Sonenberg, N.; Blenis, J., RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J Biol Chem* **2007,** *282* (19), 14056-64.
320. Brown, A. K.; Webb, A. E., Regulation of FOXO Factors in Mammalian Cells. *Curr Top Dev Biol* **2018,** *127*, 165-192.
321. Sampadi, B.; Mullenders, L. H. F.; Vrieling, H., Phosphoproteomics Sample Preparation Impacts Biological Interpretation of Phosphorylation Signaling Outcomes. *Cells* **2021,** *10* (12).
322. Yang, Y. Y.; Han, L.; Yu, Q. L.; Gao, Y. F.; Song, R. D.; Zhao, S. N., Phosphoproteomic analysis of longissimus lumborum of different altitude yaks. *Meat Sci* **2020,** *162*.
323. Hoffman, N. J.; Parker, B. L.; Chaudhuri, R.; Fisher-Wellman, K. H.; Kleinert, M.; Humphrey, S. J.; Yang, P.; Holliday, M.; Trefely, S.; Fazakerley, D. J.; Stockli, J.; Burchfield, J. G.; Jensen, T. E.; Jothi, R.; Kiens, B.; Wojtaszewski, J. F.; Richter, E. A.; James, D. E., Global Phosphoproteomic Analysis of Human Skeletal Muscle Reveals a Network of Exercise-Regulated Kinases and AMPK Substrates. *Cell Metab* **2015,** *22* (5), 922-35.
324. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014,** *13* (9), 2513-26.
325. Colome, N.; Abian, J.; Aloria, K.; Arizmendi, J. M.; Barcelo-Batllori, S.; Braga-Lagache, S.; Burlet-Schiltz, O.; Carrascal, M.; Casal, J. I.; Chicano-Galvez, E.; Chiva, C.; Clemente, L. F.; Elortza, F.; Estanyol, J. M.; Fernandez-Irigoyen, J.; Fernandez-Puente, P.; Fidalgo, M. J.; Froment, C.; Fuentes, M.; Fuentes-Almagro, C.; Gay, M.; Hainard, A.; Heller, M.; Hernandez, M. L.; Ibarrola, N.; Iloro, I.; Kieselbach, T.; Lario, A.; Locard-Paulet, M.; Marina-Ramirez, A.; Martin, L.; Morato-Lopez, E.; Munoz, J.; Navajas, R.; Odena, M. A.; Odriozola, L.; de Oliveira, E.; Paradela, A.; Pasquarello, C.; de Los Rios, V.; Ruiz-Romero, C.; Sabido, E.; Sanchez Del Pino, M.; Sancho, J.; Santamaria, E.; Schaeffer-Reiss, C.; Schneider, J.; de la Torre, C.; Valero, M. L.; Vilaseca, M.; Wu, S.; Wu, L.; Ximenez de Embun, P.; Canals, F.; Corrales, F. J.; ProteoRed, I.; EuPa, Multi-laboratory experiment PME11 for the standardization of phosphoproteome analysis. *J Proteomics* **2022,** *251*, 104409.
326. Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J., The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* **2020,** *39* (3), 229-244.
327. Salovska, B.; Tichy, A.; Rezacova, M.; Vavrova, J.; Novotna, E., Enrichment strategies for phosphoproteomics: state-of-the-art. *Rev Anal Chem* **2012,** *31* (1), 29-41.
328. Potts, G. K.; McNally, R. M.; Blanco, R.; You, J. S.; Hebert, A. S.; Westphall, M. S.; Coon, J. J.; Hornberger, T. A., A map of the phosphoproteomic alterations that occur after a bout of maximal-intensity contractions. *J Physiol* **2017,** *595* (15), 5209-5226.
329. Chen, L.; Liu, J.; Kaneko, G.; Xie, J.; Wang, G.; Yu, D.; Li, Z.; Ma, L.; Qi, D.; Tian, J.; Gong, W.; Zhang, K.; Yu, E., Quantitative phosphoproteomic analysis of soft and firm grass carp muscle. *Food Chem* **2020,** *303*, 125367.
330. Wiese, H.; Kuhlmann, K.; Wiese, S.; Stoepel, N. S.; Pawlas, M.; Meyer, H. E.; Stephan, C.; Eisenacher, M.; Drepper, F.; Warscheid, B., Comparison of alternative MS/MS and bioinformatics approaches for confident phosphorylation site localization. *J Proteome Res* **2014,** *13* (2), 1128-37.

331. Bittremieux, W.; Tabb, D. L.; Impens, F.; Staes, A.; Timmerman, E.; Martens, L.; Laukens, K., Quality control in mass spectrometry-based proteomics. *Mass Spectrom Rev* **2018,** *37* (5), 697-711.
332. Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **2019,** *47* (W1), W636-W641.
333. Bekker-Jensen, D. B.; Bernhardt, O. M.; Hogrebe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V., Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat Commun* **2020,** *11* (1).
334. French, A. R., Effects of temperature on the duration of arousal episodes during hibernation. *J Appl Physiol Respir Environ Exerc Physiol* **1982,** *52* (1), 216-20.
335. Munro, D.; Thomas, D. W., The role of polyunsaturated fatty acids in the expression of torpor by mammals: a review. *Zoology (Jena)* **2004,** *107* (1), 29-48.
336. Ruf, T.; Arnold, W., Effects of polyunsaturated fatty acids on hibernation and torpor: a review and hypothesis. *Am J Physiol Regul Integr Comp Physiol* **2008,** *294* (3), R1044-52.
337. Frank, C. L.; Dierenfeld, E. S.; Storey, K. B., The relationship between lipid peroxidation, hibernation, and food selection in mammals. *Am Zool* **1998,** *38* (2), 341-349.
338. Scherer, N. M.; Deamer, D. W., Oxidation of thiols in the Ca2+-ATPase of sarcoplasmic reticulum microsomes. *Biochim Biophys Acta* **1986,** *862* (2), 309-17.
339. Xu, K. Y.; Zweier, J. L.; Becker, L. C., Hydroxyl radical inhibits sarcoplasmic reticulum Ca(2+)-ATPase function by direct attack on the ATP binding site. *Circ Res* **1997,** *80* (1), 76-81.
340. Scherer, N. M.; Deamer, D. W., Oxidative stress impairs the function of sarcoplasmic reticulum by oxidation of sulfhydryl groups in the Ca2+-ATPase. *Arch Biochem Biophys* **1986,** *246* (2), 589-601.
341. Hill, V. L.; Florant, G. L., The effect of a linseed oil diet on hibernation in yellow-bellied marmots (Marmota flaviventris). *Physiol Behav* **2000,** *68* (4), 431-7.
342. Frank, C. L.; Hood, W. R.; Donnelly, M. C., The role of α-linolenic acid (18:3) in mammalian torpor. . In *Life in the Cold: Evolution, Mechanisms, Adaptation, and Application*, Barnes BM, C. H., eds. Fairbanks, Alaska, USA, Ed. 2004; pp 71-80.
343. Rodrigues-Ribeiro, L.; Melo-Braga, M. N.; Kjeldsen, F.; Gomez-Mendoza, D. P.; Verano-Braga, T., Assessment of protein extraction and digestion efficiency of well-established shotgun protocols for heart proteomics. *Anal Biochem* **2019,** *578*, 51-59.
344. Hu, H. L.; Kang, Y.; Zeng, Y.; Zhang, M.; Liao, Q.; Rong, M. Q.; Zhang, Q.; Lai, R., Region-resolved proteomics profiling of monkey heart. *J Cell Physiol* **2019,** *234* (8), 13720-13734.
345. Reynolds, J. A.; Tanford, C., Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proc Natl Acad Sci U S A* **1970,** *66* (3), 1002-7.
346. Smith, B. J., SDS polyacrylamide gel electrophoresis of proteins. *Methods Mol Biol* **1994,** *32*, 23-34.
347. Balliau, T.; Blein-Nicolas, M.; Zivy, M., Evaluation of Optimized Tube-Gel Methods of Sample Preparation for Large-Scale Plant Proteomics. *Proteomes* **2018,** *6* (1).
348. Seddon, A. M.; Curnow, P.; Booth, P. J., Membrane proteins, lipids and detergents: not just a soap opera. *Bba-Biomembranes* **2004,** *1666* (1-2), 105-117.
349. Hemstrom, P.; Irgum, K., Hydrophilic interaction chromatography. *J Sep Sci* **2006,** *29* (12), 1784-821.
350. Linscheid, N.; Santos, A.; Poulsen, P. C.; Mills, R. W.; Calloe, K.; Leurs, U.; Ye, J. Z.; Stolte, C.; Thomsen, M. B.; Bentzen, B. H.; Lundegaard, P. R.; Olesen, M. S.; Jensen, L. J.; Olsen, J. V.; Lundby, A., Quantitative proteome comparison of human hearts with those of model organisms. *Plos Biol* **2021,** *19* (4).
351. Doll, S.; Dressen, M.; Geyer, P. E.; Itzhak, D. N.; Braun, C.; Doppler, S. A.; Meier, F.; Deutsch, M. A.; Lahm, H.; Lange, R.; Krane, M.; Mann, M., Region and cell-type resolved quantitative proteomic map of the human heart. *Nat Commun* **2017,** *8*.

352. Chen, A.; Chen, Z. W.; Xia, Y.; Lu, D. B.; Jia, J. G.; Hu, K.; Sun, A. J.; Zou, Y. Z.; Qian, J. Y.; Ge, J. B., Proteomics Analysis of Myocardial Tissues in a Mouse Model of Coronary Microembolization. *Frontiers in Physiology* **2018,** *9*.
353. Giroud, S.; Frare, C.; Strijkstra, A.; Boerema, A.; Arnold, W.; Ruf, T., Membrane phospholipid fatty acid composition regulates cardiac SERCA activity in a hibernator, the Syrian hamster (Mesocricetus auratus). *PLoS One* **2013,** *8* (5), e63111.
354. Stammers, A. N.; Susser, S. E.; Hamm, N. C.; Hlynsky, M. W.; Kimber, D. E.; Kehler, D. S.; Duhamel, T. A., The regulation of sarco(endo)plasmic reticulum calcium-ATPases (SERCA). *Can J Physiol Pharm* **2015,** *93* (10), 843-854.
355. Polakof, S.; Mommsen, T. P.; Soengas, J. L., Glucosensing and glucose homeostasis: from fish to mammals. *Comp Biochem Physiol B Biochem Mol Biol* **2011,** *160* (4), 123-49.
356. Holmes, D. J.; Flückiger, R.; Austad, S. N., Comparative biology of aging in birds: an update. *Exp. Gerontol.* **2001,** *36* (4-6), 59-66.
357. Klandorf, H.; Boyce, C. S.; Holt, S. B.; Iqbal, M.; Killefer, J.; Peterson, R. A.; Deaver, D. R., Effect of thyroid hormone on concentrations of plasma calcitonin in broiler chicks. *Comp Biochem Phys C* **1999,** *122* (1), 55-60.
358. Rattiste, K.; Klandorf, H.; Urvik, J.; Sepp, T.; Asghar, M.; Hasselquist, D.; Cooey, C.; Horak, P., Skin pentosidine and telomere length do not covary with age in a long-lived seabird. *Biogerontology* **2015,** *16* (4), 435-41.
359. Zuck, J.; Borges, C. R.; Braun, E. J.; Sweazea, K. L., Chicken albumin exhibits natural resistance to glycation. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2017,** *203*, 108-114.
360. Pongudom, S.; Chinthammitr, Y., Determination of Normal HbA1C Levels in Non-Diabetic Patients with Hemoglobin E. *Ann Clin Lab Sci* **2019,** *49* (6), 804-809.
361. Récapet, C.; Sibeaux, A.; Cauchard, L.; Doligez, B.; Bize, P., Selective disappearance of individuals with high levels of glycated haemoglobin in a free-living bird. *Biology Letters* **2016,** *12* (8), 20160243.
362. Rendell, M.; Stephen, P. M.; Paulsen, R.; Valentine, J. L.; Rasbold, K.; Hestorff, T.; Eastberg, S.; Shint, D. C., An interspecies comparison of normal levels of glycosylated hemoglobin and glycosylated albumin. *Comp Biochem Physiol B* **1985,** *81* (4), 819-22.
363. Miksik, I.; Hodny, Z., Glycated Hemoglobin in Mute Swan (Cygnus-Olor) and Rook (Corvus-Frugilegus). *Comp Biochem Phys B* **1992,** *103* (3), 553-555.
364. Beuchat, C. A.; Chong, C. R., Hyperglycemia in hummingbirds and its consequences for hemoglobin glycation. *Comp Biochem Phys A* **1998,** *120* (3), 409-416.
365. Ardia, D. R., Glycated hemoglobin and albumin reflect nestling growth and condition in American kestrels. *Comp Biochem Phys A* **2006,** *143* (1), 62-66.
366. Jeppsson, J. O.; Kobold, U.; Barr, J.; Finke, A.; Hoelzel, W.; Hoshino, T.; Miedema, K.; Mosca, A.; Mauri, P.; Paroni, R.; Thienpont, L.; Umemoto, M.; Weykamp, C.; Med, I. F. C. C. L., Approved IFCC reference method for the measurement of HbA(1c) in human blood. *Clin Chem Lab Med* **2002,** *40* (1), 78-89.
367. Niwa, T., Mass spectrometry for the study of protein glycation in disease. *Mass Spectrom. Rev.* **2006,** *25* (5), 713-723.
368. Brun, C.; Hernandez-Alba, O.; Hovasse, A.; Criscuolo, F.; Schaeffer-Reiss, C.; Bertile, F., Resistance to glycation in the zebra finch: Mass spectrometry-based analysis and its perspectives for evolutionary studies of aging. *Exp Gerontol* **2022,** *164*, 111811.
369. Rhee, M. K.; Ho, Y. L.; Raghavan, S.; Vassy, J. L.; Cho, K.; Gagnon, D.; Staimez, L. R.; Ford, C. N.; Wilson, P. W. F.; Phillips, L. S., Random plasma glucose predicts the diagnosis of diabetes. *PLoS One* **2019,** *14* (7), e0219964.
370. Anguizola, J.; Matsuda, R.; Barnaby, O. S.; Hoy, K. S.; Wa, C. L.; DeBolt, E.; Koke, M.; Hage, D. S., Review: Glycation of human serum albumin. *Clin Chim Acta* **2013,** *425*, 64-76.
371. Rondeau, P.; Bourdon, E., The glycation of albumin: Structural and functional impacts. *Biochimie* **2011,** *93* (4), 645-658.
372. Jones, M. P., Avian Hematology. *Clin Lab Med* **2015,** *35* (3), 649-59.

373. Delpierre, G.; Collard, F.; Fortpied, J.; Van Schaftingen, E., Fructosamine 3-kinase is involved in an intracellular deglycation pathway in human erythrocytes. *Biochem J* **2002,** *365* (Pt 3), 801-8.
374. Delplanque, J.; Delpierre, G.; Opperdoes, F. R.; Van Schaftingen, E., Tissue distribution and evolution of fructosamine 3-kinase and fructosamine 3-kinase-related protein. *J Biol Chem* **2004,** *279* (45), 46606-13.
375. Szwergold, B. S.; Beisswenger, P. J., Enzymatic deglycation--a new paradigm or an epiphenomenon? *Biochem Soc Trans* **2003,** *31* (Pt 6), 1428-32.
376. Johnstone, R. M.; Mathew, A.; Setchenska, M. S.; Grdisa, M.; White, M. K., Loss of glucose transport in developing avian red cells. *Eur J Cell Biol* **1998,** *75* (1), 66-77.
377. Ahmed, M. H.; Ghatge, M. S.; Safo, M. K., Hemoglobin: Structure, Function and Allostery. *Subcell Biochem* **2020,** *94*, 345-382.
378. Alev, C.; Shinmyozu, K.; McIntyre, B. A.; Sheng, G., Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. *Dev Genes Evol* **2009,** *219* (7), 353-60.
379. Barvitenko, N. N.; Adragna, N. C.; Weber, R. E., Erythrocyte signal transduction pathways, their oxygenation dependence and functional significance. *Cell Physiol Biochem* **2005,** *15* (1-4), 1-18.
380. Tellone, E.; Barreca, D.; Russo, A.; Galtieri, A.; Ficarra, S., New role for an old molecule: The 2,3-diphosphoglycerate case. *Biochim Biophys Acta Gen Subj* **2019,** *1863* (10), 1602-1607.
381. Mills, F. C.; Ackers, G. K., Quaternary enhancement in binding of oxygen by human hemoglobin. *Proc Natl Acad Sci U S A* **1979,** *76* (1), 273-7.
382. Tsuneshige, A.; Kanaori, K.; Samuni, U.; Danstker, D.; Friedman, J. M.; Neya, S.; Giangiacomo, L.; Yonetani, T., Semihemoglobins, high oxygen affinity dimeric forms of human hemoglobin respond efficiently to allosteric effectors without forming tetramers. *J Biol Chem* **2004,** *279* (47), 48959-67.
383. Pillai, A. S.; Chandler, S. A.; Liu, Y.; Signore, A. V.; Cortez-Romero, C. R.; Benesch, J. L. P.; Laganowsky, A.; Storz, J. F.; Hochberg, G. K. A.; Thornton, J. W., Origin of complexity in haemoglobin evolution. *Nature* **2020,** *581* (7809), 480-485.
384. Sowole, M. A.; Konermann, L., Comparative analysis of oxy-hemoglobin and aquomet-hemoglobin by hydrogen/deuterium exchange mass spectrometry. *J Am Soc Mass Spectrom* **2013,** *24* (7), 997-1005.
385. Song, Y.; Laskay, U. A.; Vilcins, I. M. E.; Barbour, A. G.; Wysocki, V. H., Top-Down-Assisted Bottom-Up Method for Homologous Protein Sequencing: Hemoglobin from 33 Bird Species. *J Am Soc Mass Spectrom* **2015,** *26* (11), 1875-1884.
386. Kiseleva, O.; Kurbatov, I.; Ilgisonis, E.; Poverennaya, E., Defining Blood Plasma and Serum Metabolome by GC-MS. *Metabolites* **2021,** *12* (1).
387. Bhawal, R.; Oberg, A. L.; Zhang, S.; Kohli, M., Challenges and Opportunities in Clinical Applications of Blood-Based Proteomics in Cancer. *Cancers (Basel)* **2020,** *12* (9).
388. Asl, S. K.; Rahimzadegan, M., The recent progress in the early diagnosis of acute myocardial infarction based on myoglobin biomarker: Nano-aptasensors approaches. *J Pharm Biomed Anal* **2022,** *211*, 114624.
389. Chen, Z. Z.; Gerszten, R. E., Metabolomics and Proteomics in Type 2 Diabetes. *Circ Res* **2020,** *126* (11), 1613-1627.
390. Cloonan, M. J.; Bishop, G. A.; Wilton-Smith, P. D.; Carter, I. W.; Allan, R. M.; Wilcken, D. E., An enzyme-immunoassay for myoglobin in human serum and urine. Method development, normal values and application to acute myocardial infarction. *Pathology* **1979,** *11* (4), 689-99.
391. Liu, Z.; Hu, Y.; Gong, Y.; Zhang, W.; Liu, C.; Wang, Q.; Deng, H., Hydrogen peroxide mediated mitochondrial UNG1-PRDX3 interaction and UNG1 degradation. *Free Radic Biol Med* **2016,** *99*, 54-62.
392. Makarov, A.; Denisov, E.; Lange, O.; Horning, S., Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer (vol 17, pg 977, 2006). *J Am Soc Mass Spectrom* **2006,** *17* (12), 1758-1758.

393. Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W., The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J Proteome Res* **2017,** *16* (12), 4299-4310.
394. International Expert, C., International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care* **2009,** *32* (7), 1327-34.
395. Diabetes, C.; Complications Trial Research, G.; Nathan, D. M.; Genuth, S.; Lachin, J.; Cleary, P.; Crofford, O.; Davis, M.; Rand, L.; Siebert, C., The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* **1993,** *329* (14), 977-86.
396. Suo, M.; Wen, D.; Wang, W.; Zhang, D.; Xu, S.; Wang, X.; Hu, T., False measurement of glycated hemoglobin in patients without hemoglobin A. *Biosci Rep* **2019,** *39* (1).
397. Soboleva, A.; Mavropulo-Stolyarenko, G.; Karonova, T.; Thieme, D.; Hoehenwarter, W.; Ihling, C.; Stefanov, V.; Grishina, T.; Frolov, A., Multiple Glycation Sites in Blood Plasma Proteins as an Integrated Biomarker of Type 2 Diabetes Mellitus. *Int. J. Mol. Sci.* **2019,** *20* (9), 2329.
398. Rabbani, G.; Ahn, S. N., Structure, enzymatic activities, glycation and therapeutic potential of human serum albumin: A natural cargo. *Int J Biol Macromol* **2019,** *123*, 979-990.
399. Shaklai, N.; Garlick, R. L.; Bunn, H. F., Nonenzymatic glycosylation of human serum albumin alters its conformation and function. *J Biol Chem* **1984,** *259* (6), 3812-7.
400. Bhat, S.; Jagadeeshaprasad, M. G.; Venkatasubramani, V.; Kulkarni, M. J., Abundance matters: role of albumin in diabetes, a proteomics perspective. *Expert Rev Proteomics* **2017,** *14* (8), 677-689.
401. Garlick, R. L.; Mazer, J. S., The principal site of nonenzymatic glycosylation of human serum albumin in vivo. *J Biol Chem* **1983,** *258* (10), 6142-6.
402. Bayer, V., An Overview of Monoclonal Antibodies. *Semin Oncol Nurs* **2019,** *35* (5), 150927.
403. Das, T. K.; Narhi, L. O.; Sreedhara, A.; Menzen, T.; Grapentin, C.; Chou, D. K.; Antochshuk, V.; Filipe, V., Stress Factors in mAb Drug Substance Production Processes: Critical Assessment of Impact on Product Quality and Control Strategy. *J Pharm Sci* **2020,** *109* (1), 116-133.
404. Singh, S.; Kumar, N. K.; Dwiwedi, P.; Charan, J.; Kaur, R.; Sidhu, P.; Chugh, V. K., Monoclonal Antibodies: A Review. *Curr Clin Pharmacol* **2018,** *13* (2), 85-99.
405. Saleem, R. A.; Affholter, B. R.; Deng, S.; Campbell, P. C.; Matthies, K.; Eakin, C. M.; Wallace, A., A chemical and computational approach to comprehensive glycation characterization on antibodies. *MAbs* **2015,** *7* (4), 719-31.
406. Lee, J.; Kang, H. A.; Bae, J. S.; Kim, K. D.; Lee, K. H.; Lim, K. J.; Choo, M. J.; Chang, S. J., Evaluation of analytical similarity between trastuzumab biosimilar CT-P6 and reference product using statistical analyses. *MAbs* **2018,** *10* (4), 547-571.
407. S. Millán-Martín; I. Zaborowska; C. Jakes; S. Carillo; Bones, J., Comparability study for the determination of posttranslational modifications of biotherapeutic drug products and biosimilars by automated peptide mapping analysis. Application Note ThermoScientific, 2018.
408. Clerc, F.; Reiding, K. R.; Jansen, B. C.; Kammeijer, G. S. M.; Bondt, A.; Wuhrer, M., Human plasma protein N-glycosylation. *Glycoconj. J.* **2016,** *33* (3), 309-343.
409. Frolov, A.; Hoffmann, R., Identification and relative quantification of specific glycation sites in human serum albumin. *Anal Bioanal Chem* **2010,** *397* (6), 2349-2356.
410. Miyamoto, H.; Kohzuma, T.; Ohnishi, A., Changes in the albumin glycation site, plasma pentosidine and esRAGE concentrations before and after intensive diabetic treatment in patients with abnormally high glycated albumin levels. *Ann Clin Biochem* **2018,** *55* (1), 84-91.
411. Zhang, Q. B.; Monroe, M. E.; Schepmoes, A. A.; Clauss, T. R. W.; Gritsenko, M. A.; Meng, D.; Petyuk, V. A.; Smith, R. D.; Metz, T. O., Comprehensive Identification of Glycated Peptides and Their Glycation Motifs in Plasma and Erythrocytes of Control and Diabetic Subjects. *J Proteome Res* **2011,** *10* (7), 3076-3088.
412. Zhang, M.; Xu, W.; Deng, Y., A new strategy for early diagnosis of type 2 diabetes by standard-free, label-free LC-MS/MS quantification of glycated peptides. *Diabetes* **2013,** *62* (11), 3936-42.

413. Zhang, Q.; Tang, N.; Schepmoes, A. A.; Phillips, L. S.; Smith, R. D.; Metz, T. O., Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes. *J Proteome Res* **2008,** *7* (5), 2025-32.
414. Korwar, A. M.; Vannuruswamy, G.; Jagadeeshaprasad, M. G.; Jayaramaiah, R. H.; Bhat, S.; Regin, B. S.; Ramaswamy, S.; Giri, A. P.; Mohan, V.; Balasubramanyam, M.; Kulkarni, M. J., Development of Diagnostic Fragment Ion Library for Glycated Peptides of Human Serum Albumin: Targeted Quantification in Prediabetic, Diabetic, and Microalbuminuria Plasma by Parallel Reaction Monitoring, SWATH, and MSE. *Mol Cell Proteomics* **2015,** *14* (8), 2150-2159.
415. Brede, C.; Hop, B.; Jorgensen, K.; Skadberg, O., Measurement of glycated albumin in serum and plasma by LC-MS/MS. *Scand J Clin Lab Invest* **2016,** *76* (3), 195-201.
416. Barzegar, A.; Moosavi-Movahedi, A. A.; Sattarahmady, N.; Hosseinpour-Faizi, M. A.; Aminbakhsh, M.; Ahmad, F.; Saboury, A. A.; Ganjali, M. R.; Norouzi, P., Spectroscopic studies of the effects of glycation of human serum albumin on L-Trp binding. *Protein Peptide Lett* **2007,** *14* (1), 13-18.
417. Greifenhagen, U.; Frolov, A.; Bluher, M.; Hoffmann, R., Site-specific analysis of advanced glycation end products in plasma proteins of type 2 diabetes mellitus patients. *Anal Bioanal Chem* **2016,** *408* (20), 5557-66.
418. Rathore, R.; Sonwane, B. P.; Jagadeeshaprasad, M. G.; Kahar, S.; Santhakumari, B.; Unnikrishnan, A. G.; Kulkarni, M. J., Glycation of glucose sensitive lysine residues K36, K438 and K549 of albumin is associated with prediabetes. *J Proteomics* **2019,** *208*, 103481.
419. Carapito, R.; Li, R.; Helms, J.; Carapito, C.; Gujja, S.; Rolli, V.; Guimaraes, R.; Malagon-Lopez, J.; Spinnhirny, P.; Lederle, A.; Mohseninia, R.; Hirschler, A.; Muller, L.; Bastard, P.; Gervais, A.; Zhang, Q.; Danion, F.; Ruch, Y.; Schenck, M.; Collange, O.; Chamaraux-Tran, T. N.; Molitor, A.; Pichot, A.; Bernard, A.; Tahar, O.; Bibi-Triki, S.; Wu, H. G.; Paul, N.; Mayeur, S.; Larnicol, A.; Laumond, G.; Frappier, J.; Schmidt, S.; Hanauer, A.; Macquin, C.; Stemmelen, T.; Simons, M.; Mariette, X.; Hermine, O.; Fafi-Kremer, S.; Goichot, B.; Drenou, B.; Kuteifan, K.; Pottecher, J.; Mertes, P. M.; Kailasan, S.; Aman, M. J.; Pin, E.; Nilsson, P.; Thomas, A.; Viari, A.; Sanlaville, D.; Schneider, F.; Sibilia, J.; Tharaux, P. L.; Casanova, J. L.; Hansmann, Y.; Lidar, D.; Radosavljevic, M.; Gulcher, J. R.; Meziani, F.; Moog, C.; Chittenden, T. W.; Bahram, S., Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. *Sci Transl Med* **2022,** *14* (628).
420. Sandow, J. J.; Infusini, G.; Dagley, L. F.; Larsen, R.; Webb, A. I., Simplified high-throughput methods for deep proteome analysis on the timsTOF Pro. bioRxiv, 2019.
421. Human Plasma Proteome Project. http://www.peptideatlas.org/hupo/hppp/ (accessed 18/07/2022).
422. Kaur, G.; Poljak, A.; Ali, S. A.; Zhong, L.; Raftery, M. J.; Sachdev, P., Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *J Proteome Res* **2021,** *20* (2), 1261-1279.
423. Zhang, Q.; Petyuk, V. A.; Schepmoes, A. A.; Orton, D. J.; Monroe, M. E.; Yang, F.; Smith, R. D.; Metz, T. O., Analysis of non-enzymatically glycated peptides: neutral-loss-triggered MS3 versus multi-stage activation tandem mass spectrometry. *Rapid Commun Mass Sp* **2008,** *22* (19), 3027-3034.
424. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008,** *26* (12), 1367-72.
425. Fossati, A.; Richards, A. L.; Chen, K. H.; Jaganath, D.; Cattamanchi, A.; Ernst, J. D.; Swaney, D. L., Toward Comprehensive Plasma Proteomics by Orthogonal Protease Digestion. *J Proteome Res* **2021,** *20* (8), 4031-4040.
426. Chen, Y.; Lease, N. K.; Gin, J. W.; Ogorzalek, T. L.; Adams, P. D.; Hillson, N. J.; Petzold, C. J., Modular automated bottom-up proteomic sample preparation for high-throughput applications. *Plos One* **2022,** *17* (2).
427. Arnemo, J. M.; Evans, A. L., Biomedical Protocols for Free-Ranging Brown Bears,Wolves,Wolverines and Lynx. 2017.

428. Carapito, C.; Burel, A.; Guterl, P.; Walter, A.; Varrier, F.; Bertile, F.; Van Dorsselaer, A., MSDA, a proteomics software suite for in-depth Mass Spectrometry Data Analysis using grid computing. *Proteomics* **2014,** *14* (9), 1014-9.

Université de Strasbourg

Charlotte BRUN

Développement de stratégies analytiques quantitatives pour l'étude des protéines, de leurs phosphorylations et glycations

Résumé

L'analyse protéomique par spectrométrie de masse permet l'étude de systèmes biologiques complexes en identifiant et quantifiant les protéines impliquées et en caractérisant leur structure. Cette stratégie peut aussi être appliquée à l'étude des modifications post-traductionnelles des protéines, niveau de complexité supplémentaire du protéome. Ce travail de thèse est axé sur la mise en place de stratégies analytiques quantitatives globales de type « *label-free* » pour explorer les mécanismes biologiques à l'œuvre chez l'ours brun et le lérot hibernants. D'autre part, des développements pour l'analyse quantitative des phosphorylations dans des tissus musculaires et des glycations dans des échantillons plasmatiques ont été réalisés, de la préparation des échantillons avec notamment des stratégies d'enrichissement par chromatographies d'affinité et de fractionnement peptidique, jusqu'au traitement des données de spectrométrie de masse par diverses solutions logicielles.

Mots-clés : Analyse protéomique quantitative, Spectrométrie de masse, Modifications Post-Traductionnelles (PTMs)

Résumé en anglais

Proteomic analysis by mass spectrometry enables the study of complex biological systems by identifying and quantifying their proteins and characterising their structure. This strategy can also be applied to the study of post-translational modifications, an additional level of complexity in the proteome. This thesis work focuses on the implementation of global quantitative label-free strategies to explore the biological mechanisms at work in the hibernating brown bear and garden dormouse. On the other hand, developments for the quantitative analysis of phosphorylations in muscle tissues and glycations in plasma samples were carried out, from the sample preparation with enrichment strategies by affinity chromatography and peptide fractionation, to the processing of mass spectrometry data by several software solutions.

Keywords: Quantitative Proteomics, Mass Spectrometry, Post-Translational Modifications (PTMs)