

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES
UMR 7178

THÈSE présentée par :

Marine PLUMEL

soutenue le : **26 mars 2015**

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

Discipline/ Spécialité : Chimie analytique

**Optimisations des stratégies analytiques
quantitatives en protéomique : application à
l'étude des réponses adaptatives du
métabolisme chez divers organismes**

THÈSE dirigée par :

Dr. Alain Van Dorsselaer
Dr. Fabrice Bertile

Directeur de recherche, CNRS, Strasbourg
Chargé de recherche, CNRS, Strasbourg

RAPPORTEURS :

Dr. Myriam Ferro
Dr. Thierry Rabilloud

Directeur de recherche, CEA, Grenoble
Directeur de recherche, CEA, Grenoble

AUTRES MEMBRES DU JURY :

Dr. Florence Ploetze
Dr. Pierre Bize

Maitre de conférences, Université de Strasbourg
Senior lecturer, University of Aberdeen

A mes parents, mes sœurs et Mansuy,

“Bonus vir semper tiro”

Martial

(A good man is always learning)

Remerciements

En premier lieu, je souhaite remercier Alain Van Dorsselaer qui m'a accueillie au sein de son laboratoire. Merci pour votre confiance et pour les nombreux conseils que vous avez pu me donner que cela concerne le fonctionnement d'un analyseur quadripolaire ou la création d'une entreprise. Merci également de toujours avoir laissé la porte de votre bureau (grande) ouverte et ce pour tous vos étudiants. Je souhaite également remercier Sarah Cianferani pour tout le soutien que j'ai pu recevoir. Merci pour les très nombreux conseils que tu as pu me donner pour m'aider à construire mon « après-thèse » !

Je remercie Mme Myriam Ferro, Mme Florence Ploetze, M. Thierry Rabilloud et M. Pierre Bize d'avoir accepté d'évaluer ces travaux de thèse.

Ensuite, je souhaite remercier (plus que) chaleureusement Fabrice Bertile qui a encadré, tout d'abord mon stage de Master 2, puis toute ma thèse. Fabrice, merci pour tes conseils, ta rigueur, tes « ça va être superbe », ta patience devant les millions de cases des fichiers Excel du projet NOTOX et tout autant de questions que nous nous sommes posées et ta disponibilité sans limite. Tu m'as toujours soutenu quand j'ai voulu entreprendre de nouvelles choses, participer au prix Jeune Chercheur en congrès par exemple ou même lancer mon entreprise. Merci également pour les milliers de commentaires que tu as pu me faire sur tous les travaux que j'ai pu écrire et finalement sur ce manuscrit de thèse, les nombreuses discussions que nous avons eues, et pour ton soutien tout simplement pendant ces 4 années de thèse/stage.

Ensuite, je souhaite remercier tous les collaborateurs du LSMBO avec qui j'ai eu la chance de travailler : Etienne Challet, Jean-Yves Georges, Stéphane Blanc, François Criscuolo, Marie-Edith Chabouté et Morgane Batzenschlager, de même que l'ensemble des membres du consortium NOTOX. Merci de m'avoir confié une partie de vos projets et de m'avoir fait « voyager » avec toutes ces thématiques passionnantes !

Je souhaite également remercier Jean-Michel Billmann, le « super Bruker man ». Merci Jean-Michel pour ta gentillesse et le temps que tu passes à chaque intervention à transmettre tes connaissances.

Un immense merci à Christine Schaeffer pour tout le temps que nous avons passé ensemble à discuter de mes petits Q-TOFs préférés (ce n'est plus un secret, je crois). Cela a été un vrai plaisir de pouvoir travailler avec toi ! Merci pour ta gentillesse, ton soutien, ta bonne humeur, ton dynamisme et ton aide très précieuse pour la recherche d'emploi.

Laurence, merci d'avoir toujours laissé la porte de votre bureau grande ouverte aux doctorants et étudiants. Merci pour votre implication incroyable dans tout ce que vous entreprenez et merci également pour vos nombreux conseils et encouragements tout au long de mon cursus.

Merci aux « Quality girls », Véronique, Hélène et Stella ! Merci pour votre bonne humeur légendaire et pour toutes les discussions que nous avons eues pendant cette thèse (qui ne concernaient d'ailleurs pas forcément la qualité !). Merci à Fabrice (FV) pour son aide pendant ces 4 ans et pour avoir toujours pris le temps de débloquer ma session et n'importe quel logiciel que j'aurais pu bloquer ! En même temps avec tes 4 garçons tu dois bien avoir l'habitude de gérer ce genre de petits problèmes. Finalement le LSMBO, c'est un peu comme à la maison non ? Kévin, merci de prendre soin de tous les petits doctorants du laboratoire qui n'ont jamais manqué de rien grâce à toi. Merci d'être toujours arrangeant et flexible ! Merci également d'avoir pris le temps de répondre à mes

questions de comptabilité entrepreneuriale. J'ai également une petite pensée pour Bernadette, qui répond toujours avec le même sourire aux nombreuses questions que l'on peut avoir !

Merci à Alexandre B. de prendre toujours la peine d'aider tous ceux qui ne sont pas nés avec un manuel Excel dans les mains ! Merci aussi d'avoir été toujours disponible au laboratoire pour donner un petit coup de main. J'espère qu'un jour tu pourras participer à Koh-Lanta. Merci aussi à Alex 2 pour son aide et notamment concernant LA macro de traitement des données NOTOX. Merci à vous deux pour avoir toujours eu une oreille attentive aux moindres petits soucis informatiques que nous pouvons rencontrer au laboratoire. Ibrahim, bon courage pour ta nouvelle vie au LSMBO et merci pour les petites attentions dont tu peux faire preuve ! Patrick, merci pour ton aide concernant le monde merveilleux du logiciel R, quasiment incompréhensible pour les gens « normaux ».

Agnès et Allison, merci pour vos sourires et votre bonne humeur habituelle ! Charlotte, Margaux, je vous souhaite une très bonne continuation au LSMBO. Gauthier (Gauffier apparemment pour les intimes), je te souhaite également une très bonne continuation, cela devrait bien se passer maintenant que tu as fini de lire « Pubmed ». Luc, bon courage à toi pour tes multiples vies et merci de t'impliquer autant pour les autres ! Danièle, Mme gel d'électrophorèse ! Merci de m'avoir appris à faire ces fameux gels, merci pour tout le temps que nous avons passé ensemble et merci pour toutes les petites attentions dont tu peux faire preuve au quotidien, toujours une petite parole douce et encourageante... merci. Seb, ou l'homme qui murmurait à l'oreille des spectros. Merci pour tous les bons conseils que tu as pu me donner ou le SAV statistique que tu peux fournir ! Toujours une petite blague très subtile à dégainer. Bonne continuation pour ta thèse et pour ta future vie de star de la protéomique, j'en suis sûre. Nina, merci pour tous les moments hors LSMBO autour de ces petites planchettes que je suis bien sûr la seule à dévorer ! Bon courage pour la suite de ta thèse, je suis très contente de savoir que « mes » petits spectros sont entre de bonnes mains avec toi. Merci d'être très vite devenue une bonne copine.

CC, je ne sais pas définir combien de fois te dire merci mais j'ai perdu le compte depuis longtemps ! Merci pour tous les conseils que tu as pu me donner et les discussions autour de la protéomique que nous avons eues. Merci de toujours prendre du temps d'aider tout le monde au LSMBO, merci de nous avoir toujours présenté tous les gens que tu pouvais connaître, merci aussi pour ta gentillesse légendaire et la fameuse petite boule de poil que tu m'as aidé à adopter. Félicitations à la toute jeune maman que tu es devenue. Je te souhaite une très bonne continuation professionnelle comme personnelle.

J'ai également une pensée très spéciale pour Jean-Marc, un des sacrés piliers du LSMBO, sur qui on peut toujours compter. Merci pour les innombrables petits tuyaux ou coups de main de la vie de tous les jours, le temps que tu as toujours pris pour transmettre tes innombrables connaissances sur le monde de la LC, de la spectrométrie de masse et du fonctionnement des grues. Merci pour ta gentillesse tout simplement, ta disponibilité sans limite et la découverte des beignets Christian à la vanille. François (FD), je t'ai connu, c'était le bazar sur ton bureau, tu essayais de trouver du temps machine la nuit et tu cherchais des solutions pour automatiser la réalisation de cette étape si sympathique qu'est l'extraction sur sep-packs... Aujourd'hui, quelques années plus tard, rien n'a changé MAIS tu es devenu l'heureux papa de ce petit Aron (ceci dit, il a beau être mignon, il ne sait toujours pas dire protéomique je crois)... Merci FD pour ta disponibilité et ton sourire, les randos dans la forêt, tous les coups de main bricolo ou professionnels que tu as pu me donner et tout le temps que tu consacres dans ta vie pour les autres !

Georg, affectueusement dit Giorgio, merci pour toutes les discussions que nous avons eues ! Les heures passées devant les 30 millions de cases de ces fameux fichiers Excel (et on sait tous les deux qu'il ne s'agit pas d'une façon de parler !), les craquages et fous rires devant ce traitement des

données complètement dingue... Je ne peux m'empêcher de sourire en pensant à toutes les publis de laboratoires allemands que tu as pu me donner. Bonne continuation pour la suite de tes aventures et que tout aille bien pour toi et ton (grand) garçon.

Gilles, aujourd'hui tu es passé du côté obscur de la force avec les Orbitraps (ok, je reconnais que les Orbitraps, c'est un peu la force, l'aveu après 4 ans !) mais je garde un excellent souvenir de toutes les discussions que nous avons pu avoir ensemble au LSMBO et de tous ces moments incroyables de préparation de la fête de la science (avec nos 10 kgs de M&M's) ou encore les réunions du bureau des doctorants... inoubliables. Bon courage à toi pour ta nouvelle vie et à bientôt j'espère en Alsace. Magali, tu es celle qui m'a formé au LSMBO, je n'oublierai pas les très bons moments que nous avons pu avoir ensemble au LSMBO ou ... ailleurs ! Solaris, en russe, pendant trois heures... un sacré souvenir ! Merci pour tous les petits tuyaux que tu as pu me transmettre sur ces sacrés Q-TOFs qui ont bien rempli ma thèse pendant ces quelques années mais surtout merci d'avoir toujours été plus qu'une collègue.

Et puis, il y a l'équipe de joyeux lurons du R2, oui Johann et Guillaume, je parle bien de vous. Merci de m'avoir permis de manger avec François Damiens tous les jours pendant la rédaction de ma thèse. Vous m'avez fait beaucoup rire, toujours un vrai plaisir de discuter avec vous deux, je vous souhaite sincèrement bonne chance pour la fin de vos thèses respectives mais surtout pour tous les projets personnels que pourrez entreprendre à côté ! Guillaume, t'entendre dire « A demain les knäcques » avec ton petit accent du sud, cela n'a juste pas de prix ! J'en profite pour glisser un petit merci aux deux autres « supramol boys » seniors, François (FDE) et Julien, qui ont croisé ma route au LSMBO. Marion et Leslie, merci les filles pour vos sourires, vos encouragements et toutes les sucreries qui ont défilées dans notre bureau (bon mon bureau 2 en fait) ! Marion, je n'oublierai pas tes très fameuses « danses de la victoire », j'espère que tu pourras continuer à bosser dans le domaine merveilleux de la spectrométrie de masse, je te souhaite une très bonne continuation avec ta petite famille. Leslie, Nestlie pour les intimes, bonne chance pour ta thèse. Je suis sûre que tu feras des très belles choses ! Bon courage avec Benoît derrière ton dos. Justine, bonne continuation au LSMBO !

Benoît, mille mercis pour ta gentillesse, ton énergie incroyable (s'il y avait plus de gens comme toi, je pense qu'on pourrait régler le problème des ressources énergétiques sur terre), les mots gentils que tu sais toujours trouver si nécessaire, ta disponibilité sans limite pour filer un coup de main ou tes conseils de nutrition bien avisés pendant la période de la fin de la thèse. Je te souhaite bonne chance pour la suite. A vrai dire, je ne m'en fais pas trop pour toi, bien au contraire.

Et je n'oublie pas les anciens ! Amandine, Cyril, Nicolas, Alaye, Anna, Stéphanie, Cédric. Merci d'avoir été là pour discuter des merveilles de la protéomique ou rire tout simplement. Amandine, merci d'avoir toujours eu un petit mot gentil d'encouragement malgré la distance. Je vous souhaite à tous une très belle continuation, professionnelle certes, mais surtout, surtout pour tout le reste. Parmi les « anciens », il y a aussi mes deux petits stagiaires Batshéva et Quentin ! Merci d'avoir été motivés et souriants. Cela a été un réel plaisir de travailler avec vous ! J'espère que vous trouverez votre bonheur professionnel ! Batshéva, j'espère que tu m'enverras une photo le jour où tu croiseras un de tes fameux pingouins (manchots oui je sais).

Et puis, j'en arrive à Diego et Sarah... Quand vous êtes partis, vous m'avez terriblement manqué. Notre bureau, l'ASMS, les heures passées devant les machines à s'arracher les cheveux, les heures passées à vanter les mérites des Q-TOFs Bruker par rapport au G1 (ça, c'était plutôt juste avec Diego, la plus belle « Bruker team » de tous les temps), les heures à rire pour... rien probablement, toutes les discussions scientifiques que nous avons pu avoir, les coups de main, toutes ces « bullshits » dont nous avons pu rire ou discuter ensemble... Une des plus belle parties de ma thèse ! Sarah, mille mercis d'avoir relu plusieurs de mes chapitres, de m'avoir coaché comme tu l'as fait pour des

entretiens... Tes nouveaux collègues sont de sacrés veinards, d'une part pour le boulot parce que tu es juste incroyable, mais aussi parce que tu es toujours partante pour sortir du laboratoire ou faire des petites blagues à quelqu'un ! Diego (et Federica !) merci d'avoir pris la peine de faire plusieurs centaines de kilomètres pour venir juste quelques heures me surprendre pour mon anniversaire pendant cette démentielle période rédactionnelle... Bref merci à vous deux pour tous ces moments, votre aide, mais surtout pour votre belle amitié.

Un immense merci à tous mes amis fidèles Camille, Sophie, Julien, Gallou (désolée même dans ma thèse tu portes ce pseudonyme, c'est bien simple quand on parle de Thomas, je ne sais pas de qui on parle), Flo, Cécile, Claude, Lisa pour vos soutiens, vos encouragements, tous les moments que nous avons pu passer ensemble et tous ceux à venir ! Merci de contribuer à nous faire nous sentir chez nous aujourd'hui en Alsace. Merci à Gallou et Flo pour leur énergie incroyable, le coaching « anti-pistache mode » et tous les bons moments que nous avons pu vivre ensemble (même s'il nous arrive de nous retrouver pour chercher des clés de voiture dans un mètre de neige). Merci à Julien et Sophie, les [chatons d'origine] (et oui dans la thèse, il faut mettre des références), appellation d'origine contrôlée qui a désormais fait son entrée au LSMBO ! Merci pour les conseils, repas, soutiens, encouragements, desserts, et tout simplement d'être là, avec nous, tout le temps. Cécile et Claude, merci pour vos sourires, et votre présence à nos côtés ! Lisa, Camille, quelque part, pour moi, vous n'avez jamais quitté Strasbourg. Merci pour toutes les petites pensées encourageantes et d'être toujours là, dans la vie, pour tout, pour rien, pour moi.

Merci à mes parents pour leur soutien fidèle, merci de nous avoir rendues si indépendantes et de nous avoir offert l'opportunité de voyager. Merci à mes deux sœurs si chères à mon cœur. Une petite pensée pour Mathieu qui prend soin de notre Doudou maintenant. J'ai également une petite pensée pour la famille Rocquin-Pruneta. Merci pour tous vos encouragements !

Et enfin, Mansuy. Merci pour ton soutien sans limite, merci d'avoir accepté que le LSMBO devienne un peu notre « famille » pendant toutes ces années, de m'avoir emmenée sur le petit caillou de l'océan indien où tu as grandi à la fin de ma 3^{ème} année de thèse alors que c'était une petite folie, de m'avoir rappelé qu'il n'y a pas que la thèse dans la vie quand j'avais tendance à l'oublier, et de tout simplement être toujours là pour moi. Je ne sais pas si j'aurai pu réussir sans toi...si avec tout ça, nous ne sommes pas prêts pour le meilleur et pour le pire...

Alors voilà, une dernière fois, à vous tous qui avez contribué à cette thèse ou qui contribuez à ce que la vie soit toujours plus belle, du fond du cœur, **merci.**

Publications, communications orales et par affiche

Publications :

Plumel, M. I., A. Stier, et al. (2014). "Litter size manipulation in laboratory mice: an example of how proteomic analysis can uncover new mechanisms underlying the cost of reproduction." Front Zool 11: 41.

Plumel, M. I., T. Wasselin, et al. (2013). "Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (*Dermochelys coriacea*)." J Proteome Res 12(9): 4122-4135.

Communications orales :

Plumel, M., S. Blanc, et al. (2012). Muscle metabolism of the hibernating bear: a proteomic approach. First Nordforsk Symposium on translational physiology. From bear to man and back again., Skagen, Denmark.

Plumel, M., S. Blanc, et al. (2012). Muscle metabolism of the hibernating bear: a proteomic approach. Scandinavian brown bear project PhD meeting, Trondheim, Norway.

Plumel, M., D. Thiersé, et al. (2012). Mécanismes écophysiologiques: un nouveau champ d'application des stratégies protéomiques. 8èmes Journées du Club jeune SFEAP, Arras (France).

Plumel, M., T. Wasselin, et al. (2013). LC-SRM-based absolute quantitation of two novel vitellogenin isoforms in a non-sequenced organism, the leatherback turtle. 7th annual conference of the European Proteomics Association (EuPA), Young Investigator Prize, St Malo, France.

Communications par affiches :

Bertaccini, D., M. Plumel, et al. (2014). A comparison of label free DDA-HrXIC and DIA acquisition on a benchtop UHR-Q-TOF. Congrès Français de Spectrométrie de Masse et d'Analyse Protéomique (SMAP2014), Lyon, France.

Bertaccini, D., M. Plumel, et al. (2014). A comparison of label free DDA-HrXIC and DIA acquisition on a benchtop UHR-Q-TOF. 13th HUPO World Congress: The proteome quest to understand biology and disease, Madrid, Spain.

Guillaumot, N., M. Plumel, et al. (2014). Increase of sensitivity with the nano-ESI CaptiveSpray source associated to nanoBOOster (Bruker). Congrès MSMB « Mass Spectrometry of Medicine and Biotechnology (Summer School) Dubrovnik, Croatie.

Plumel, M., D. Bertaccini, et al. (2014). Increase of sensitivity with the nano-ESI CaptiveSpray source associated to nanoBooster (Bruker). Congrès Français de Spectrométrie de Masse et d'Analyse Protéomique (SMAP2014), Lyon, France.

Plumel, M., D. Bertaccini, et al. (2014). Increase of sensitivity with the nano-ESI CaptiveSpray source associated to nanoBOOster (Bruker). 13th HUPO World Congress: The proteome quest to understand biology and disease, Madrid, Spain.

Plumel, M., V. Plot, et al. (2013). Development of quantitative proteomics and application to biological questions. Absolute quantitation of vitellogenin using a microLC-SRM to determine reproductive status in female leatherback turtles, a non-sequenced organism. PRIME-XS Workshop on Quantitative proteomics of biomarkers for medical applications and basic research, Annecy.

Plumel, M., G. Tascher, et al. (2014). Investigation of drug induced steatosis in HepaRG cells after long-term treatment with valproic acid using LC/MS-based quantitative proteomics. Congrès Français de Spectrométrie de Masse et d'Analyse Protéomique (SMAP2014), Lyon, France.

Plumel, M., D. Thierse, et al. (2012). Proteomics provides mechanistic insights into ecophysiological theories. 29ème congrès de la SFSM, Orléans.

Plumel, M., D. Thierse, et al. (2012). Proteomics provides mechanistic insights into ecophysiological theories. 29ème Congrès de la SFEAP, Rouen.

Plumel, M., D. Thiersé, et al. (2012). Proteomics provides mechanistic insights into ecophysiological theories. 60th ASMS Conference on Mass Spectrometry Vancouver, Canada.

Plumel, M., A. Van Dorsselaer, et al. (2014). Expression circadienne des protéines cérébelleuses chez la souris et anticipation alimentaire: approche 2D-DIGE/MS. Congrès Français de Spectrométrie de Masse et d'Analyse Protéomique (SMAP2014), Lyon, France.

Plumel, M., T. Wasselin, et al. (2013). Absolute quantification of vitellogenin using a LC-SRM approach in a non-sequenced organism. 61st ASMS Conference on Mass Spectrometry and Allied Topics, Minneapolis, MN, USA.

Plumel, M., T. Wasselin, et al. (2013). LC-SRM-based absolute quantitation of two novel vitellogenin isoforms in a non-sequenced organism, the leatherback turtle. 7th annual conference of the European Proteomics Association (EuPA), St Malo, France.

Tascher, G., A. Hovasse, et al. (2013). Proteomics of the HepaRG cell line for studying drug-induced liver toxicity. 7th annual conference of the European Proteomics Association (EuPA), St Malo, France.

Tascher, G., A. Hovasse, et al. (2013). Proteomics for detection of drug-induced toxicity using HepaRG cells. 3rd SEURAT-1 Annual meeting, Lisbon, Portugal.

Tascher, G., A. Hovasse, et al. (2013). Proteomics for detection of drug-induced toxicity using HepaRG cells. Proteomic Forum, Berlin, Germany.

Tascher, G., M. Plumel, et al. (2014). Investigation of drug induced steatosis in HepaRG cells after long-term treatment with valproic acid using LC/MS-based quantitative proteomics. European Society of Toxicology In Vitro 2014 International Conference (ESTIV2014), Egmond aan Zee, The Netherlands.

Tascher, G., M. Plumel, et al. (2014). Investigation of drug-induced steatosis in HepaRG cells after long-term treatment with valproic acid using LC/MS-based quantitative proteomics. 13th HUPO World Congress: The proteome quest to understand biology and disease, Madrid, Spain.

Tascher, G., M. Plumel, et al. (2013). A proteomic road map of the HepaRG cell line: metabolic pathways present and their response to acetaminophen treatment. EASL - The International Liver Congress, Amsterdam, The Netherlands.

Tascher, G., M. Plumel, et al. (2012). Proteomic characterization of the HepaRG cell line for the use in pharmacology and toxicology. 29ème Congrès de la SFEAP, Rouen.

Tascher, G., M. Plumel, et al. (2015). Proteomics strategies for characterization and relative quantification of in vitro cell culture samples within the NOTOX project. SEURAT-1 5th annual meeting, Barcelona, Spain.

Tascher, G., M. Plumel, et al. (2014). HepaRG cells for detection of drug-induced liver toxicity: a proteomics perspective. 4th SEURAT-1 Annual meeting, Barcelona, Spain.

Ziemianin, A., M. Plumel, et al. (2013). Responses to physical inactivity in the hibernating brown bear (Ursus arctos). Scandinavian Bear Project PhD meeting, Orsa Grönklitt, Suède.

Ziemianin Gourdet, A., M. Plumel, et al. (2014). Réponses musculaires à l'inactivité physique chez l'Ours brun: approches protéomiques quantitatives. Congrès Français de Spectrométrie de Masse et d'Analyse Protéomique (SMAP2014), Lyon, France.

Ziemianin-Gourdet, A., M. Plumel, et al. (2013). A proteomic approach to understand muscle atrophy/preservation during physical inactivity in the hibernating bear (Ursus arctos). 13ème Edition des Journées CNES Jeunes Chercheurs, Toulouse.

Ziemianin-Gourdet, A., M. Plumel, et al. (2013). Muscle and adipose tissue responses to physical inactivity: a proteomic approach in the hibernating brown bear (Ursus arctos). 7th annual conference of the European Proteomics Association (EuPA), St Malo, France.

Liste des principales abréviations

ADN : Acide désoxyribonucléique
VTG : Vitellogénine
DEPE : Département Ecologie, Physiologie, Ethologie
IPHC : Institut pluridisciplinaire Hubert Curien
XIC : eXtracted Ion Chromatogram
2D-DIGE : Differential In-Gel Electrophoresis
INCI : Institut des neurosciences cellulaires et intégratives
MALDI : Matrix Assisted Laser Desorption Ionisation
E2D : Electrophorèse bidimensionnelle
Da : Dalton
HILIC : Hydrophilic Interaction Liquid Chromatography
FT-ICR : Analyseur à transformée de Fourier (Fourier transform ion cyclotron resonance)
SDS : Sodium dodecyl sulfate
pI : point isoélectrique
HPLC : High Pressure Liquid Chromatography
IEX : Ion Exchange Chromatography
IMAC : Immobilized Metal Affinity Chromatography
UPLC : Ultra-High Pressure Chromatography
FWHM : Full Width at Half Maximum
Q : Quadripole
IT : Ion Trap (piège à ions)
TOF : Time of flight (analyseur à temps de vol)
RF : Radiofréquence
CID : Collision Induced Dissociation
QQQ : Triple quadripôle
SIM : Single Ion Monitoring
SRM : Selected Reaction Monitoring
MRM : Multiple Reaction Monitoring
MCP : Multi Channel Plate
ETD : Electron Transfer Dissociation
ECD : Electron Capture Dissociation
DDA : Data Dependant Acquisition
DIA : Data Independant Acquisition
PMF : Peptide Mass Fingerprint
PFF : Peptide Fragmentation Fingerprint
NCBI : National Center for Biotechnology Information
FDR : False Discovery rate
BLAST : Basic Local Alignment Search Tool
LSMBO : Laboratoire de spectrométrie de masse bio-organique
LC-MS : couplage chromatographie liquide- spectrométrie de masse
QconCAT : Quantification conCATamer
PSAQ : Protein Standard Absolute Quantification
PRM : Parallel Reaction Monitoring
SILAC : Stable isotope labeling by amino acids in cell culture
ICAT : Isotope-Coded Affinity Tags
iTRAQ® : isobaric Tag for Relative and Absolute Quantitation
TMT : Tandem Mass Tag
AUC : Area under curve

eV : electron-volt

PBS : Phosphate Buffered Saline

LPS : lipopolysaccharide

B-OG : octyl glucoside

DTT : Dithiothreitol

IAA : Iodoacetamide

Sommaire

Introduction générale 2

Chapitre I : Principe général de l'analyse protéomique par spectrométrie de masse 11

I.	Les différentes approches en analyse protéomique	12
A.	L'approche « Bottom-up ».....	12
B.	L'approche « Middle-down »	13
C.	L'approche « Top-down ».....	13
II.	L'analyse protéomique selon l'approche « Bottom-up »	14
A.	Extraction des protéines.....	14
B.	Séparation des protéines.....	15
B.1.	Electrophorèse monodimensionnelle	15
B.2.	Electrophorèse bidimensionnelle (E2D).....	15
B.3.	La chromatographie liquide	16
C.	Analyse des peptides par spectrométrie de masse.....	19
C.1.	Les sources	19
C.2.	Les analyseurs.....	21
C.2.1.	L'analyseur quadripolaire (Q).....	22
C.2.2.	Le piège à ions (IT/ Ion Trap)	23
C.2.3.	L'analyseur à temps de vol (TOF/Time-Of-Flight)	25
C.3.	Les analyseurs hybrides.....	27
C.3.1.	Le triple quadripôle (QQQ).....	28
C.3.2.	L'analyseur hybride quadripôle/Temps de vol (Q-TOF)	29
C.4.	Comparaison des différentes performances d'analyseurs.....	30
D.	Les détecteurs	31
E.	Les modes de fragmentation et l'interprétation des spectres MS/MS.....	31
F.	La complémentarité des enzymes et des modes de fragmentation	32
G.	Les modes d'acquisition des données	33
G.1.	Le mode « Data Dependant Acquisition » (DDA).....	34
G.2.	Le mode « Data Independant Acquisition » (DIA).....	34

Chapitre II : Stratégies d'analyses protéomiques qualitatives 37

I.	Les stratégies d'identifications chez les organismes séquencés	37
A.	L'empreinte peptidique massique.....	37
B.	L'empreinte de fragments peptidiques.....	38
C.	Les banques de données	39
D.	Validation.....	40
II.	Les stratégies d'identifications chez les organismes dont le génome est non séquencé ou non annoté	41
A.	L'empreinte de fragments peptidiques.....	41
B.	Séquençage <i>de novo</i>	41
C.	Protéogénomique.....	42
III.	Qualité des assignations chez les espèces séquencées.....	42
A.	Le cas des peptides monochargés.....	43
B.	Les recherches dans les banques de données.....	44
C.	La qualité des spectres MS/MS	44
IV.	Conclusion et perspectives.....	45

Chapitre III : Stratégies d'analyses protéomiques quantitatives 49

I.	Les stratégies ciblées	50
A.	La stratégie SRM (Selected Reaction Monitoring).....	50
A.1.	Principe de la stratégie SRM	50
A.2.	Avantages et limites de la stratégie	51
A.3.	Méthodologie pour la mise au point d'une stratégie SRM	53
A.3.1.	Sélection des peptides	53
A.3.2.	Les standards.....	55
A.3.2.1.	Les standards peptidiques.....	55
A.3.2.2.	Les standards QconCAT (Quantification conCATamer).....	56
A.3.2.3.	Les standards PSAQ (Protein Standard Absolute Quantification).....	56
A.3.3.	Traitement des données et validation	57
B.	La stratégie PRM.....	57
II.	Les stratégies d'analyse quantitative globales	58
A.	Les stratégies d'analyse quantitative basées sur le marquage des protéines ou des peptides 58	
A.1.	Les stratégies par marquage métabolique.....	59
A.2.	Les stratégies par marquage enzymatique	60
A.3.	Les stratégies par marquage chimique des protéines.....	61
A.3.1.	La quantification sur gel d'électrophorèse	61
A.3.1.1.	Méthodologie de mise au point d'une stratégie 2D-DIGE	61
A.3.1.2.	Avantages et limites de la stratégie par gel	64
A.3.2.	La stratégie ICAT (Isotope-Coded Affinity Tag)	65
A.3.3.	Les stratégies TMT™ (Tandem Mass Tag, Thermo Fisher Scientific) et iTRAQ® (isobaric Tag for Relative and Absolute Quantitation).....	65
A.3.4.	Marquage par méthylation « Dimethyl labeling »	66
B.	Les approches sans marquage dites « label free ».....	67
B.1.	Spectral counting.....	67
B.2.	Extraction des courants d'ions	67
B.2.1.	Extraction des courants d'ions MS (XIC, eXtracted Ion Chromatogram) suite à la génération de données spectrales en mode « Data Dependant Acquisition » (DDA)	68
B.2.1.1.	Solutions logicielles pour l'extraction des courants d'ions	68
B.2.1.2.	Méthodologie de mise au point d'une stratégie XIC.....	71
B.2.1.3.	Avantages et limites de la stratégie	74
B.2.2.	Extraction des courants d'ions suite à la génération de données spectrales en mode « Data Independant Acquisition » (DIA).....	75
B.2.2.1.	Modèle du SWATH™ (AB SCIEX Technologies)	75
B.2.2.2.	Modèle MS ^e (WATERS)	77
III.	Quels points sont à considérer lors du choix d'une stratégie protéomique quantitative ?..	78
A.	Le contexte : approche globale ou ciblée ?.....	79
B.	Le type de quantification pour les stratégies ciblées : absolu/relatif	79
C.	Le parc instrumental.....	79
D.	Gestion du nombre d'échantillons/conditions à considérer.....	79

Chapitre IV : Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes 85

I.	Optimisations des méthodes d'acquisition en mode DDA pour l'analyse d'échantillons complexes sur le maXis 4G (Bruker Daltonique)	86
A.	Optimisations des paramètres d'acquisition des méthodes nanoLC-MS/MS en mode DDA	88
A.1.	Le temps de cycle MS et le nombre de sommations des spectres MS	88
A.2.	Le temps d'exclusion dynamique et le seuil de déclenchement minimal pour la MS/MS	90
A.3.	Le nombre de sommation des spectres MS/MS ainsi que les seuils d'intensité correspondants.....	93
A.3.1.	Vitesse d'acquisitions MS/MS : Principe et paramétrage	93
A.3.2.	Evaluation du seuil d'intensité « B »	94
A.4.	Optimisations des énergies de collision	97
B.	Bilan des résultats obtenus suite à l'optimisation de méthodes d'acquisition nanoLC-MS/MS pour des échantillons de levure (<i>Saccharomyces cerevisiae</i>)	101
II.	Evaluation et comparaison de deux sources Electrospray sur l'Impact HD (Bruker Daltonique)	103
A.	Contexte	103
B.	Description des sources.....	104
C.	Evaluation de la source « CaptiveSpray » équipée d'un « NanoBooster » pour l'analyse protéomique quantitative par rapport à la source « nanoESI ».....	105
C.1.	Description des échantillons de référence.....	105
C.2.	Evaluation des sources en terme de sensibilité	106
C.3.	Evaluation des sources en terme de linéarité et de justesse	108
C.3.1.	Evaluation des sources en termes de linéarité.....	108
C.3.2.	Evaluation des sources en termes de justesse.....	111
C.3.2.1.	Choix du modèle de normalisation sur les données générées avec la source « CaptiveSpray »	112
C.3.2.2.	Comparaison des sources.....	117
III.	Conclusion	119

Chapitre V : Applications et développement de la stratégie quantitative globale XIC à des échantillons biologiques complexes 123

I.	Implication dans le consortium NOTOX	123
A.	Contexte	123
B.	Schéma expérimental.....	124
II.	Mise au point de l'expérience	125
A.	Choix de la stratégie analytique quantitative.....	125
B.	Choix du protocole de traitement des échantillons pour l'analyse protéomique quantitative	126
B.1.	Points à considérer	126
B.2.	Comparaison de deux protocoles de traitement des échantillons	126
B.2.1.	Description des protocoles de traitements testés	126
B.2.2.	Analyses des échantillons tests sur le couplage LC-MS.....	129
B.2.3.	Mise au point des conditions d'analyses chromatographiques.....	129
B.2.4.	Comparaison des deux stratégies évaluées en termes de nombre de protéines identifiées	132

B.2.5. Comparaison des intensités des courants d'ions extraits pour les deux traitements d'échantillons en termes de reproductibilité	133
B.2.6. Bilan des tests réalisés.....	137
III. Etude de la réponse des cellules HepaRG à une exposition à l'acide valproïque	137
A. Préparation des échantillons.....	137
B. Identifications des protéines et extraction des courants d'ions dans le logiciel Skyline	138
C. Suivi du couplage instrumental.....	139
C.1. Suivi des « échantillons contrôles »	140
C.2. Contrôle sur l'ensemble des échantillons	141
C.3. Conclusion sur la stabilité du couplage	142
D. Traitement des données.....	143
D.1. Normalisation des données	144
D.2. Rendu des résultats par protéine.....	145
D.3. Analyse statistique	148
D.4. Interprétation préliminaire des données.....	149
IV. Conclusion	150

Chapitre VI : Application de la stratégie quantitative ciblée SRM pour la quantification de la vitellogénine dans le plasma de tortue Luth 153

I. Contexte de l'étude	153
II. Choix de la stratégie analytique	154
III. Les différentes étapes du développement de la quantification SRM des VTG de tortue Luth	155
IV. Résultats et discussion	157
A. Evaluation de la linéarité de réponse des peptides quantifiés dans des échantillons de plasma	157
B. Evaluation de l'abondance des formes oxydées du peptide MTPVLLPEAVPDIMK	159
C. Suivi de la stabilité du système pendant les analyses des échantillons de plasma des deux tortues suivies au cours de la saison de ponte.....	159
D. Concentrations des isoformes Dc-VTG1 et Dc-VTG2 dans les échantillons de plasma de tortue Luth collectés pendant la saison de ponte.....	160
V. Conclusion et perspectives.....	161

Chapitre VII : Applications et optimisations de la stratégie quantitative 2D-DIGE (Differential In-Gel Electrophoresis) à des échantillons biologiques complexes 165

I. Etude des compromis évolutifs	165
A. Etude du coût de la reproduction chez la souris	165
A.1. Contexte	165
A.2. Schéma expérimental.....	166
A.3. Analyse des échantillons	167
A.4. Résultats	168
A.5. Conclusion	170
B. Etude du coût de l'activation du système immunitaire chez la souris.....	170
B.1. Contexte	170
B.2. Schéma expérimental.....	170

B.3. Analyse des échantillons	170
B.4. Résultats	171
B.5. Conclusion	173
II. Etude des rythmes circadiens chez la souris	173
A. Contexte	173
B. Optimisation du traitement des échantillons.....	177
C. Analyse des échantillons	179
D. Résultats	179
D.1. Identification des protéines cérébelleuses régulées par les rythmes circadiens	180
D.2. Analyse des protéines impliquées dans les phénomènes d'anticipations des repas	181
E. Validation.....	182
III. Etude des mécanismes d'adaptation de l'ours brun à l'inactivité	183
A. Contexte	183
B. Schéma expérimental.....	183
C. Analyse des échantillons	183
D. Résultats	184
E. Conclusion	185
IV. Discussion méthodologique pour les quatre applications présentées dans ce chapitre	186
A. Choix de la stratégie 2D-DIGE.....	186
B. Validation.....	188
V. Conclusion	188
Travaux supplémentaires	193
Bilan de la thèse	197
Perspectives	199
Bibliographie	203
Partie expérimentale.....	215

Introduction générale

Introduction générale

Dans le domaine des Sciences de la Vie, pour pouvoir explorer de manière plus complète les processus physiologiques et leurs réponses à des situations particulières, les stratégies analytiques de criblage global (-omiques) offrent une excellente alternative aux mesures dirigées sur seulement quelques cibles prédéfinies [1]. Les progrès instrumentaux en spectrométrie de masse et les avancées en ce qui concerne la préparation des échantillons, de même que l'enrichissement extraordinaire des banques de données protéiques et les développements récents en bio-informatique pour le traitement des données ont notamment permis à la protéomique de se hisser sur le devant de la scène pour le criblage des changements moléculaires caractérisant le fonctionnement physiopathologique des organismes, tissus ou cellules [2].

L'analyse protéomique est apparue à la suite et en complément de la génomique. Si la génomique étudie le fonctionnement physiopathologique des organismes, tissus ou cellules à l'échelle du génome, l'ensemble des gènes [3], la protéomique permet des études à l'échelle des produits de l'expression des gènes : les protéines dont l'ensemble constitue le protéome [3]. Le séquençage du génome d'un nombre toujours plus important d'espèces, puis leur annotation à l'aide d'algorithmes dédiés ont permis le développement des bases de données protéiques qui représentent aujourd'hui un outil majeur en analyse protéomique [4, 5]. En effet, le séquençage du génome de tout type d'espèce peut aujourd'hui être réalisé rapidement et à moindre coût [4]. Par contre, si le séquençage d'un génome comme celui de l'Homme semble aujourd'hui exhaustif (environ 20300 gènes ont été reportés [6]), la caractérisation complète de son protéome, c'est-à-dire plusieurs milliers de protéines, par l'analyse expérimentale, n'est pas encore d'actualité [7]. Outre la complexité en termes de nombre de protéines exprimées par un organisme, un tissu ou une cellule à un moment donné, l'amplitude de la gamme dynamique des concentrations protéiques qui peuvent être retrouvées au sein d'un même type d'échantillon [8] constitue un frein à la caractérisation complète du protéome. De plus, selon les conditions expérimentales ou le moment de l'analyse, les populations protéiques effectivement exprimées dans un échantillon donné peuvent varier (régulations, turn-over). L'existence d'un grand nombre de protéoformes, générées par les phénomènes d'épissage alternatif pendant la transcription de l'ADN, mais également une multitude de modifications post-traductionnelles [7], vient encore compliquer et s'opposer à l'exhaustivité des analyses protéomiques. Pour les espèces dont le génome n'est pas encore séquencé, ces difficultés sont encore exacerbées. La protéomique permet d'analyser les protéines issues d'organismes « exotiques », c'est-à-dire pas ou mal représentés dans les banques de données protéiques, soit en

se reposant uniquement sur les portions de séquences protéiques qui ont été strictement conservées au cours de l'évolution et qui peuvent être peu nombreuses, soit en mettant en place un séquençage *de novo* des peptides détectés par spectrométrie de masse [9]. L'identification des portions de séquences jusqu'alors inconnues reste malgré tout limitée à la tolérance de quelques mutations seulement et de nombreux peptides ne seront donc jamais identifiés par cette stratégie.

Aujourd'hui, l'identification d'un nombre toujours plus grand de protéines n'est plus suffisante et répondre à des questions de Biologie impose souvent de pouvoir quantifier leurs niveaux d'expressions [3] pour définir la signature d'un état pathologique donné (diagnostic) ou encore permettre de comprendre les mécanismes d'adaptation/réponse à des situations environnementales particulières. L'analyse protéomique peut aujourd'hui fournir des informations quantitatives [10]. Bien sûr, nous sommes encore loin de la quantification absolue de l'ensemble des protéines qui peuvent être identifiées au sein d'un échantillon. Si nombre de stratégies ont été développées ces dernières années en protéomique quantitative [11-13], il reste nécessaire de les adapter/optimiser à chaque problématique et à chaque type d'échantillon.

L'objectif de cette thèse a donc été d'appliquer et d'adapter les stratégies protéomiques quantitatives à diverses problématiques biologiques. Ainsi, il a fallu, pour chacune des problématiques posées, déterminer si la quantification devait concerner certaines protéines cibles ou, au contraire, viser une couverture maximale du protéome. Il a également été nécessaire d'évaluer la nécessité de mettre en place des méthodes de quantification absolues ou relatives. La connaissance des échantillons a été d'importance majeure tant en terme de complexité que de gamme dynamique de concentrations protéiques. Ceci a permis à chaque problématique de bénéficier d'une préparation adaptée aux échantillons et à la question posée. La disponibilité des séquences protéiques des diverses espèces étudiées a également dû être prise en compte. Enfin, le schéma expérimental (le nombre de répliques biologiques et techniques à considérer et la randomisation de leur analyse), le type d'instrument requis, l'optimisation de son paramétrage, la réalisation de contrôles (pour vérifier la qualité des données et/ou le comportement des spectromètres de masse sur toute la durée d'une expérience), puis le traitement des données générées (normalisations, gestion de données massives, statistiques) ont encore été d'autres points à considérer pour que les méthodes développées aboutissent effectivement à des informations biologiques nombreuses, pertinentes et bien sûr interprétables par nos collaborateurs biologistes.

Au travers des problématiques biologiques présentées dans ce manuscrit, l'ensemble des étapes de développement d'une stratégie d'analyse protéomique quantitative a été évalué dont notamment, l'optimisation des protocoles d'extraction des protéines, l'évaluation de différents protocoles de

décomplexification des extraits protéiques, le développement des méthodes d'analyses sur les couplages utilisés combinant chromatographie liquide et spectrométrie de masse et la caractérisation de certains de ces systèmes, l'adaptation des paramètres de recherche dans les banques de données protéiques à chaque thématique donnée, la mise en place de contrôle qualité permettant d'assurer la fiabilité des données générées et la génération d'un protocole de traitement des données quantitatives permettant de lier les données spectrales aux données protéiques.

Ce manuscrit s'articule en deux parties. Dans la première, je présenterai un état de l'art des stratégies protéomiques qualitatives et quantitatives. Je décrirai tout d'abord l'analyse protéomique dans son ensemble ainsi que les principales étapes de réalisation de ce type d'analyse. Ensuite, je présenterai l'ensemble des stratégies d'analyse en protéomique qualitative qui permettent d'identifier et/ou de caractériser les protéines dans des échantillons plus ou moins complexes puis l'ensemble des stratégies quantitatives, globales, qui permettent de quantifier plusieurs centaines de protéines simultanément, ou ciblées, portant sur quelques dizaines de protéines au maximum.

Dans la seconde partie, je rapporterai les résultats obtenus pour chacune des problématiques sur lesquelles j'ai travaillées pendant ma thèse concernant des espèces dont les génomes étaient séquencés ou non. Chacune de ces thématiques a nécessité la mise en place de protocoles adaptés qui seront présentés dans chacun des chapitres.

Le premier chapitre de la partie « Résultats » traite de l'optimisation des méthodes d'acquisition sur un spectromètre de masse de type Q-TOF. Il présente l'ensemble des critères à considérer lors de la mise au point des méthodes nanoLC-MS/MS pour l'analyse qualitative et/ou quantitative d'échantillons biologiques complexes sur ce type d'instrument. Une comparaison de deux types de configuration de source nanoElectrospray est également présentée dans ce chapitre.

Le second chapitre de la partie « Résultats » décrit le développement d'une méthode d'analyse protéomique globale basée sur l'extraction des courants d'ions en mode MS (XIC) et son application pour une étude de toxicoprotéomique. Ce projet est intégré à un programme de recherche Européen (NOTOX ; <http://www.notox-sb.eu/>) qui a pour but de développer de nouvelles méthodes d'évaluation de toxicité chronique en combinant de multiples sciences « omiques » à des modèles de prédiction mathématique. Il s'agissait de caractériser par protéomique la réponse de cellules de la lignée HepaRG exposées de manière chronique à l'acide valproïque. Pour ce projet, il a été nécessaire de mettre en place un traitement des données spécifique pour permettre l'extrapolation des valeurs

d'intensité des signaux MS à des valeurs d'abondances « peptidiques », puis des valeurs « peptidiques » à des niveaux d'expression protéiques.

Le troisième chapitre de la partie « Résultats » décrit le développement d'une méthode d'analyse quantitative ciblée de la vitellogénine (VTG) dans des échantillons de plasma collectés chez la tortue Luth. Ce projet a été réalisé en collaboration avec le Dr Jean-Yves Georges (DEPE, IPHC, Strasbourg) pour qui la quantification de la vitellogénine représente un moyen indispensable de connaître le statut reproducteur des tortues. La connaissance de ce statut reproducteur permettra par la suite une analyse plus précise des déterminants de la reproduction chez ces espèces. La difficulté de ce projet résidait dans le fait que les séquences protéiques de cette espèce n'étaient pas connues. De plus, ce projet a nécessité la mise en place d'un contrôle qualité rigoureux de manière à assurer la qualité des données quantitatives générées au cours de l'expérience.

Le quatrième chapitre de la partie « Résultats » présente l'application de la stratégie quantitative 2D-DIGE à quatre projets différents.

- Le premier projet consiste en l'étude du coût de la reproduction en collaboration avec le Dr François Criscuolo (DEPE, IPHC, Strasbourg). Il s'agissait d'identifier et de quantifier de façon relative les protéines hépatiques chez des souris chez qui l'effort reproducteur avait été manipulé.
- Le second projet concerne l'étude du coût de l'activation du système immunitaire en fonction de l'âge au niveau hépatique chez la souris. Il s'agissait d'identifier et de quantifier de façon relative les protéines hépatiques en réponse à l'activation du système immunitaire chez des souris plus ou moins âgées.
- Le troisième projet s'inscrit dans le contexte d'études plus vastes des effets de l'altération des rythmes circadiens sur le métabolisme des animaux en collaboration avec le Dr Etienne Challet (INCI, Strasbourg). L'objectif de ce projet était de mieux caractériser l'horloge biologique du cervelet en identifiant et quantifiant de façon relative les protéines cérébelleuses dont l'abondance varie selon un rythme circadien et/ou après un décalage de l'heure des repas pour mieux comprendre les mécanismes d'anticipation.
- Le quatrième projet s'inscrit dans le contexte d'une étude en collaboration avec le Dr Stéphane Blanc (DEPE, IPHC, Strasbourg) dans laquelle nous étudions les mécanismes

physiologiques qui permettent à l'ours hibernant, donc inactif pendant plusieurs mois, de préserver sa masse musculaire [14]. Il s'agissait d'identifier et de quantifier de façon relative les protéines musculaires de l'ours en dormance par rapport à celles des mêmes ours en période d'activité. La difficulté de ce projet due à l'absence de séquences protéiques pour cette espèce, a nécessité la mise en place d'un traitement des données sur la base des portions de séquences protéiques de l'ours qui sont conservées chez d'autres espèces pour lesquelles le protéome est connu.

En conclusion, je résumerai les étapes de développement des analyses protéomiques quantitatives, les optimisations réalisées pour l'ensemble des projets présentés et l'apport de chacune d'elle pour la génération de données fiables et pertinentes.

Chapitre I. Principe général de l'analyse protéomique par spectrométrie de masse

Chapitre I : Principe général de l'analyse protéomique par spectrométrie de masse

La protéomique consiste à analyser l'ensemble des protéines (protéome) exprimées dans une cellule à un moment et dans une situation donnée. Aujourd'hui, c'est l'usage de la spectrométrie de masse (MS) qui est plébiscité pour l'analyse du protéome en raison des multiples avancées techniques apparues ces 25 dernières années [5]. Parmi ces avancées, on peut citer les sources d'ionisation de type « Electrospray » et « MALDI » (Matrix Assisted Laser Desorption Ionisation) qui sont adaptées au traitement des protéines [15, 16], et dont le développement a donné lieu à un prix Nobel de chimie (J. Fenn et K. Tanaka, 2002). L'amélioration constante des performances des différentes géométries d'analyseur (sensibilité, vitesse, précision et résolution), le développement des modes de fragmentation, des techniques séparatives pour la décomplexification des échantillons biologiques (électrophorèses sur gel, chromatographie liquide), l'enrichissement considérable des banques de données protéiques issues des banques de données génomique, le développement d'algorithmes dédiés à l'annotation de ces dernières et le développement des outils bio-informatiques de traitement des données ont également largement contribué à promouvoir la spectrométrie de masse dans le domaine de la protéomique [5]. La protéomique s'articule autour de 4 grands axes présentés en figure I-1.

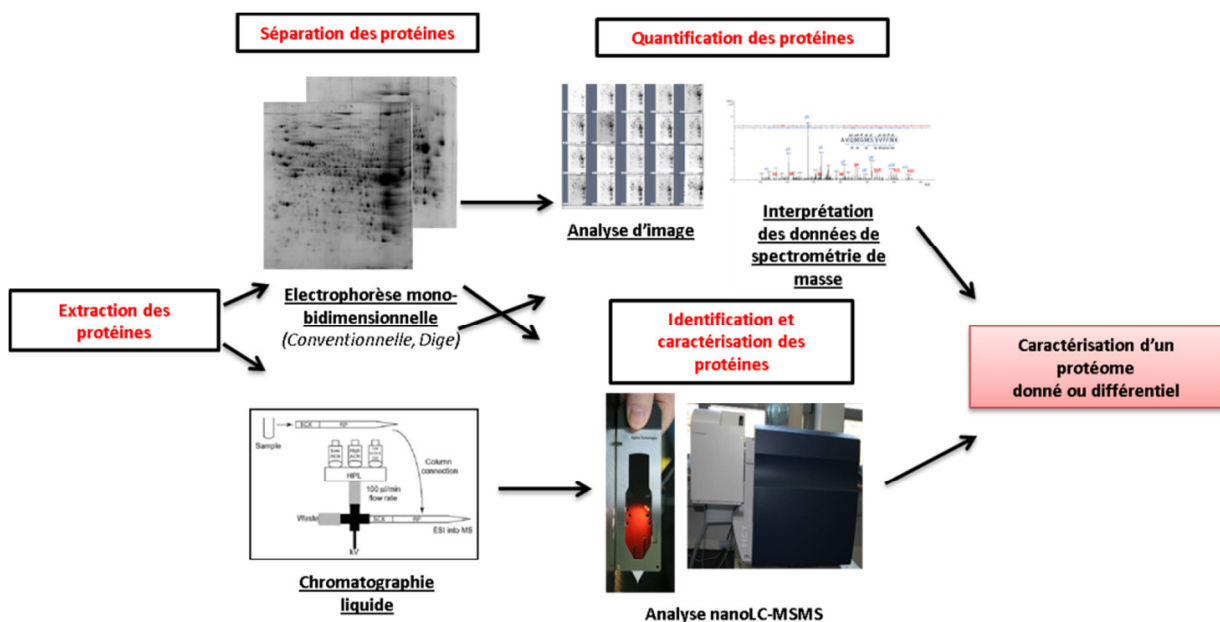


Figure I-1 Les quatre axes majeurs de la protéomique

Ces axes consistent en : l'extraction des protéines à partir de sa matrice d'origine qui est souvent un milieu biologique, la séparation des protéines extraites par chromatographie liquide ou par électrophorèse mono- ou bidimensionnelle (E2D), leur identification (et leur caractérisation structurale) par spectrométrie de masse (au travers de la génération de spectres de masse (MS), et de spectres de fragmentation (MS/MS)), et enfin leur quantification par analyse d'image ou par interprétation des données spectrales. L'ensemble de ces étapes permettent de définir un protéome différentiel.

La protéomique permet donc non seulement d'inventorier les protéines qui composent un protéome donné, mais aussi de déterminer les différences qualitatives et quantitatives induites par une situation physiopathologique particulière (analyse protéomique différentielle) [17].

I. Les différentes approches en analyse protéomique

On distingue 3 types d'approche protéomique : l'approche « Bottom-up », « Middle-down » et « Top-down ». Ces trois approches dépendent essentiellement de la masse des composés analysés, ces approches sont illustrées en figure I-2 (extrait de [18]) et seront décrites dans la suite du chapitre.

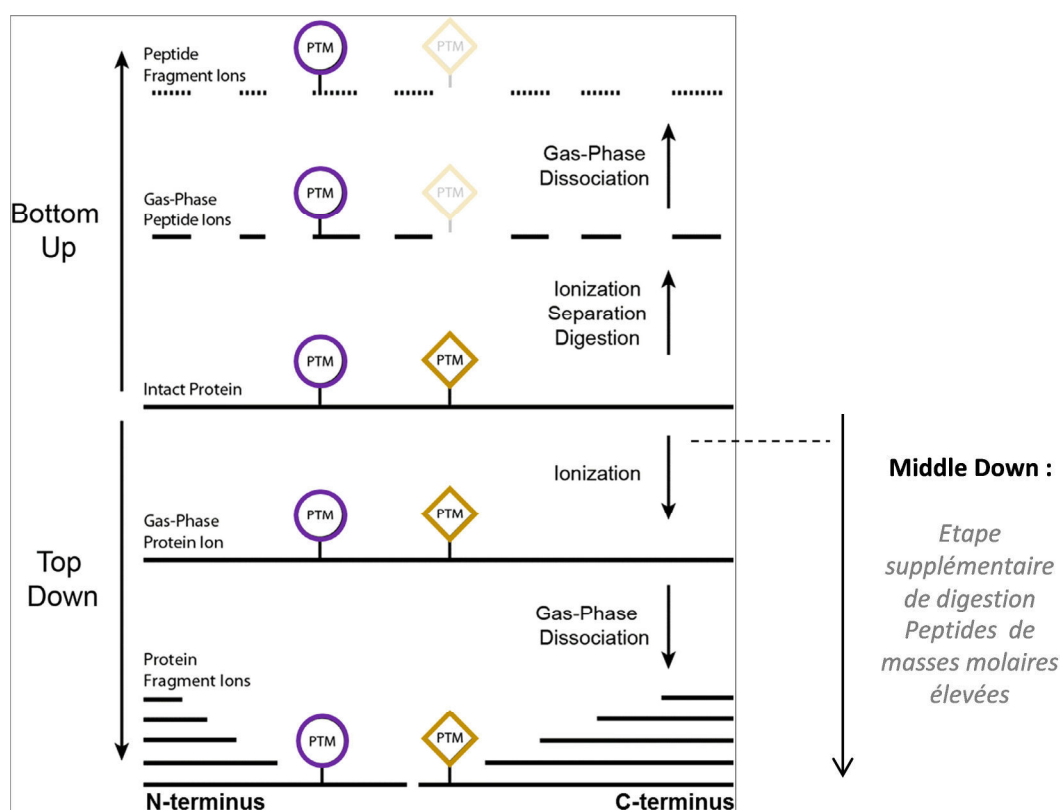


Figure I-2 Principales approches en analyse protéomique, extrait de [18]

- 1) Les approches "Bottom-up", basées sur l'analyse des protéines à partir de mélanges peptidiques issus de digestion enzymatique.
- 2) Les approches « Top-Down » basées sur l'étude des protéines entières.

L'approche Middle-down se situe entre ces deux approches, elle correspond à l'analyse de mélanges peptidiques de hautes masses moléculaires. L'approche « Middle Down » a été ajoutée sur la figure d'origine extraite de [18].

A. L'approche « Bottom-up »

L'approche « Bottom-up » consiste à analyser des protéines sous forme de digestat obtenu à la suite d'une digestion enzymatique qui génère des peptides dont la masse ne dépasse pas, en général, 3000 Da environ. Les stratégies d'identification par approche « Bottom-up » sont présentées en chapitre II « Stratégies d'analyses protéomiques qualitatives ». Cette approche est la plus largement

utilisée dans le domaine de l'analyse protéomique par spectrométrie de masse [19]. Plus spécifiquement, c'est l'utilisation de la trypsine comme protéase, qui permet la génération de peptides basiques de séquences relativement courtes (8-9 acides aminés en moyenne) et facilement ionisables en source, qui ont rendu cette approche très populaire. L'identification des peptides est basée sur l'interprétation des spectres MS, qui permettent d'accéder à la masse des peptides générés après digestion enzymatique et des spectres de fragmentation (MS/MS) de ces peptides, issus de la spectrométrie de masse en tandem [20].

Cette approche est largement utilisée aujourd'hui en haut-débit pour des échantillons très complexes contenant plusieurs milliers de protéines combinée à une séparation des mélanges peptidiques réalisée en amont du spectromètre de masse par chromatographie liquide [19]. Les différentes étapes de cette approche sont plus largement décrites dans la partie II de ce chapitre en page 14.

B. L'approche « Middle-down »

L'approche « Middle-down », ou « extended Bottom-up », est une approche également basée sur la digestion enzymatique des protéines à analyser comme l'approche « Bottom-up ». Cependant, les enzymes utilisées au sein de cette approche permettent la génération de polypeptides dont la masse est comprise entre 3 et 20 kDa. Cette stratégie se trouve à l'interface entre les stratégies « Bottom-up » et « Top-down » [20]. Les enzymes LysC, GluC ou AspN peuvent permettre la génération de peptides compris entre 3 et 7 kDa. Pour ces enzymes, on parle plutôt d'approches « extended Bottom-up » [21]. La génération de longs polypeptides est plus efficace avec des enzymes de type *OmpT* qui permet la génération de peptides de masse supérieure à 6 kDa [22].

C. L'approche « Top-down »

L'approche « Top-down » consiste en l'étude des protéines entières par spectrométrie de masse [18]. L'information qualitative est déduite des spectres de masse acquis à partir des protéines entières (MS), mais aussi à partir de ceux générés à partir des fragments générés au sein des spectromètres de masse (MS/MS). Etant donné la complexité des spectres de masse issus de l'analyse de protéines entières ou de leurs fragments (due aux multiples états de charge qu'ils peuvent prendre), cette approche est réservée à des échantillons hautement purifiés. En plus d'une purification des échantillons, il est nécessaire, comme avec l'approche « Bottom-up », de réaliser une étape de chromatographie liquide en amont de l'analyse par spectrométrie de masse. Le plus souvent, cette

étape de séparation utilise un système de chromatographie liquide sur phase inverse (silice greffée avec de courtes chaînes alkyles C4), par interaction hydrophobe (HILIC) ou par échange d'ions [23].

En plus de ces étapes préparatoires, cette technique nécessite l'utilisation d'un spectromètre de masse à haute résolution de façon à pouvoir interpréter les massifs isotopiques de ces protéines présentes sous la forme de multiples états de charge (analyseur à temps de vol ou analyseur à transformée de Fourier : FT-ICR, Orbitrap™) [18]. Par ailleurs, les temps d'acquisition nécessaires à la génération de ce type de spectre sont longs. Par exemple, l'Orbitrap™ Exactive™ Plus EMR (Thermo Scientific) génère un ou deux spectres par seconde à la résolution 140 000 (m/z 200). Cette approche ne permet donc pas de travailler à haut-débit.

Cette stratégie est plutôt plébiscitée pour l'étude de variants de séquences ou protéoformes [18], qu'il s'agisse de variants d'épissage ou de protéines modifiées (modifications post-traductionnelles par exemple).

II. L'analyse protéomique selon l'approche « Bottom-up »

Les travaux présentés dans ce manuscrit sont basés sur l'analyse d'échantillons complexes et ont donc fait l'objet d'approches « Bottom-up » exclusivement.

A. Extraction des protéines

Les échantillons biologiques sont par nature complexes. Ils peuvent contenir plusieurs milliers de protéines présentes sur une large gamme de concentrations (jusqu'à 10 ordres de grandeur pour le plasma par exemple) [8]. L'étape d'extraction et de séparation des protéines est une étape primordiale de décomplexification des échantillons. A partir des échantillons de tissus, il existe divers types d'extraction mécanique (presse de French, broyage, sonication) ou chimiques (choc osmotique, lyse enzymatique) [24] plus ou moins brutaux qui provoquent la lyse des cellules et la libération des protéines. Les échantillons ainsi obtenus peuvent être traités par des inhibiteurs de protéases, libérés pendant la lyse des cellules, afin de préserver l'intégrité des protéines. Les échantillons biologiques contiennent de nombreux interférents (sels, lipides, etc.) qui peuvent être éliminés par précipitation des protéines (acétone, acide trichloracétique, etc.) [25] ou par des méthodes de dialyse et d'ultrafiltration. Finalement, afin d'enrichir les échantillons en protéines minoritaires, l'élimination ciblée des plus abondantes peut faire appel par exemple à des techniques d'immunodéplétion ou à des préparations spécifiques d'organites sub-cellulaires [26]. La séparation

des protéines extraites peut ensuite être obtenue par des techniques électrophorétiques ou chromatographiques.

B. Séparation des protéines

B.1. Electrophorèse monodimensionnelle

L'électrophorèse monodimensionnelle (gel SDS-Page) [27] est une technique séparative permettant de décomplexifier le contenu protéique d'un échantillon au sein d'un gel d'acrylamide selon la masse moléculaire des protéines. Après extraction, les protéines sont solubilisées dans un tampon à base de SDS (Sodium Dodecyl Sulfate). Ce dernier est un détergent puissant, qui aide à la solubilisation des protéines en leur conférant des charges négatives. Sous l'effet d'un champ électrique, les protéines, qui portent toutes un ratio masse/charge équivalent après traitement au SDS, migrent dans le gel réticulé en direction de l'anode et sont ainsi séparées selon leur masse moléculaire dans les mailles du gel d'acrylamide. Les protéines sont ensuite communément révélées au bleu de Coumassie [28]. Selon le niveau de décomplexification souhaité, on pourra limiter la séparation à l'obtention d'une seule bande protéique (le gel « stacking ») ou au contraire l'étendre jusqu'à obtention de plus de 25 bandes de 2 mm environ (soit 5 cm environ). Il est possible de faire varier la réticulation du gel (la taille des mailles) afin d'augmenter la capacité résolutive sur des gammes de masses moléculaires spécifiques.

B.2. Electrophorèse bidimensionnelle (E2D)

L'électrophorèse bidimensionnelle [29] consiste à séparer des protéines sur deux dimensions. Dans une première dimension, les protéines sont séparées dans une bande de gel d'acrylamide fonctionnalisée avec des immobilines (acides et bases faibles créant un gradient de pH au sein de la bande) selon leur valeur de point isoélectrique (pI), c'est à dire jusqu'à une position dans le gel où leur charge nette est globalement neutre ($\text{pH} = \text{pI}$) [30]. Dans une seconde dimension, les protéines sont séparées sous l'effet d'un champ électrique, selon le même procédé que le gel SDS-Page, au sein d'un gel d'acrylamide (environ 20 x 20 cm le plus souvent) plus ou moins réticulé selon leur masse moléculaire. Le principe de l'électrophorèse bidimensionnelle est présentée en figure I-3.

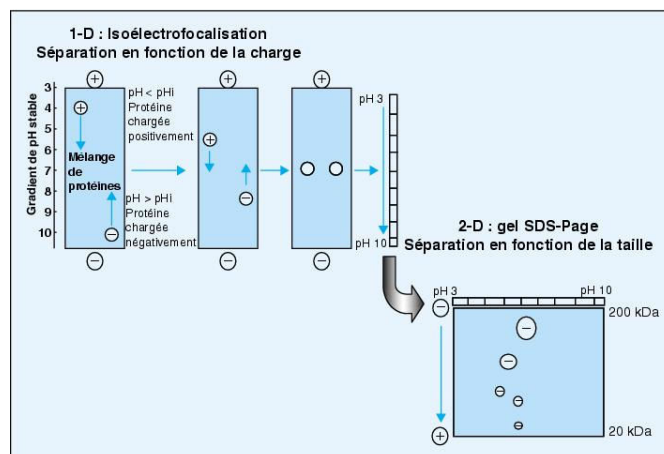


Figure I-3 Principe de l'électrophorèse bidimensionnelle

Dans une première dimension, les protéines sont séparées selon leur point isoélectrique dans une bande de gel à gradient de pH immobilisé, puis dans une seconde dimension selon leur masse moléculaire dans un gel de polyacrylamide plus ou moins réticulé.

L'électrophorèse bidimensionnelle permet la séparation en routine de 2000 protéines environ. Elle offre également l'avantage d'être compatible avec la spectrométrie de masse, de permettre l'observation de la présence de certaines modifications post-traductionnelles et de conserver des informations relatives aux propriétés physico-chimiques des protéines comme leur masse moléculaire ou la valeur de leur point isoélectrique [31]. La révélation des spots protéiques sur les gels 2D offre l'avantage de permettre la quantification relative des protéines contenues dans ces spots entre divers échantillons. Les stratégies quantitatives basées sur l'utilisation de gels d'électrophorèse bidimensionnelle sont présentées au sein du chapitre III « Stratégies d'analyses protéomiques quantitatives » en page 61. Il existe de nombreux types de révélations qui diffèrent de par leur sensibilité et gamme de linéarité de réponse [32] : la coloration au bleu de Coumassie (sensibilité 50-100 ng environ, linéarité : 2-3 ordres de grandeur), au nitrate d'argent (sensibilité de l'ordre du ng ; gamme de linéarité < 2 ordres de grandeur) [33] et le marquage des protéines avec des espèces fluorophores (sensibilité < 0,1 ng; gamme de linéarité : 4-5 ordres de grandeur) [34].

B.3. La chromatographie liquide

La chromatographie en phase liquide à haute performance (HPLC, High Performance Liquid Chromatography ou High-pressure Liquid Chromatography) est une technique séparative basée sur des interactions (hydrophobes, Van der Waals) entre les molécules à analyser qui circulent dans une phase mobile (généralement un solvant organique de composition variable) et les molécules qui composent une phase stationnaire solide (en général, silice greffée avec diverses fonctions chimiques). Cette technique de séparation est devenue incontournable dans le domaine de l'analyse

protéomique pour décomplexifier les extraits peptidiques avant l'analyse par spectrométrie de masse [35].

La phase stationnaire la plus couramment utilisée au sein des approches « Bottom-up » est une phase inverse (apolaire), C18 (octadecyl silane), qui est adaptée à l'hydrophobicité des peptides à séparer. D'autres phases peuvent également être utilisées en protéomique, et même être combinées [36], comme les phases échangeuses d'ions (IEX, Ion eXchange Chromatography) qui séparent les peptides sur la base de leur charge, la phase HILIC (Hydrophilic Interaction Liquid Chromatography), adaptée pour la séparation des peptides polaires comme les peptides glycosylés, ou les phases permettant de retenir spécifiquement certains acides aminés par chélation de métaux comme les phases IMAC (Immobilized Metal Affinity Chromatography) utilisées pour l'enrichissement des peptides phosphorylés.

Les peptides sont élués avec une combinaison de solvants compatibles avec la phase stationnaire utilisée (qui ne la dénaturent pas). Dans le cas de l'utilisation de la phase inverse C18, qui est la plus couramment utilisée, la phase mobile est communément composée d'eau et d'acétonitrile, le pourcentage d'acétonitrile augmentant au cours du temps dans le but d'éluier progressivement les peptides selon leur hydrophobicité : les plus hydrophobes étant les plus « retenus » au sein de la phase stationnaire apolaire. Un agent d'appariement d'ions est souvent ajouté aux solvants dans le but d'aider à la rétention des peptides polaires ioniques qui seraient difficilement retenus sur phase inverse [37]. Les agents d'appariement d'ions les plus utilisés en protéomique sont l'acide trifluoroacétique et l'acide formique (0,1 %).

Deux caractéristiques principales définissent la qualité de la séparation des molécules par chromatographie liquide :

- La résolution qui décrit la capacité à différencier deux pics A et B.
- Le nombre de plateaux qui décrit l'efficacité d'une colonne à séparer des molécules. Plus le nombre de plateaux théoriques est élevé, plus la résolution de la séparation chromatographique sera élevée.

Plusieurs paramètres chromatographiques sont à sélectionner en adéquation avec la complexité du mélange de peptides à décomplexifier:

- **le diamètre interne de la colonne** : Le diamètre de la colonne chromatographique est à adapter en fonction de la quantité disponible d'échantillon. La réduction des diamètres internes permet de gagner en sensibilité par la réduction des volumes de solvants en sortie de colonne [38]. Dans le cas de faible disponibilité en termes de quantité de matériel biologique, ce qui est souvent le cas dans le domaine de l'analyse protéomique, les peptides sont communément séparés grâce à des systèmes de type nano-LC (nano-colonne avec des diamètres internes de 75 micromètres).

Le facteur de concentration lorsqu'on réduit le diamètre interne de la colonne peut être calculé à l'aide de la formule suivante :

$$\left(\frac{\text{Diamètre colonne standard}}{\text{Diamètre colonne réduit}}\right)^2.$$

Dans le domaine de la protéomique quantitative, si la quantité d'échantillon biologique disponible le permet, il est préférable de travailler sur des systèmes de type micro-LC (avec des microcolonnes de diamètre interne d'environ 0,3 mm) qui sont plus robustes notamment en termes de reproductibilité de temps de rétention [39]).

- **La longueur de la colonne** : Plus la longueur d'une colonne chromatographique augmente, plus son nombre de plateaux théoriques augmente et ainsi sa capacité à séparer des molécules [35]. Les colonnes communément utilisées pour l'approche « Bottom-up » ont des longueurs comprises entre 15 et 50 cm. Il est nécessaire de considérer les temps d'allongement des analyses proportionnels à la longueur des colonnes. Des compromis entre une résolution suffisante des pics chromatographiques et un temps de séparation raisonnable seront donc à réaliser pour optimiser la qualité de la séparation.
- **la granulométrie de la phase stationnaire** : La chromatographie liquide ultra performante (UPLC, Ultra Performance Liquid Chromatography) [40] utilise des phases stationnaires de très faible granulométrie (< 2 micromètres de diamètre, en général 1,7 µm) qui permettent de gagner en résolution chromatographique, ce qui est particulièrement d'intérêt pour l'analyse de milieux très complexes comme des extraits peptidiques. La réduction du diamètre des particules entraîne des conditions de pression très élevées (> 600 bars avec une

colonne de 25 cm de longueur, 75 μm de diamètre interne et 1,7 μm de diamètre de particules).

En plus de la réduction de la granulométrie de la phase stationnaire, un gain en résolution peut également être obtenu grâce à la réduction des volumes morts (post-colonne) en utilisant un système microfluidique (comme le système nanoHPLC-chip proposé par la société Agilent technologies) qui combine une pré-colonne (qui sert de colonne de garde), la colonne chromatographique et une pointe de spray. La société Waters commercialise également ce type de système microfluidique avec une granulométrie inférieure à 2 μm de diamètre. En revanche, ces systèmes « chips » sont limités aujourd'hui dans la mesure où la longueur des colonnes disponibles ne dépasse pas 15 cm, ce qui est insuffisant pour une décomplexification suffisante de mélanges peptidiques complexes.

C. Analyse des peptides par spectrométrie de masse

Un spectromètre de masse permet la mesure du rapport masse sur charge (m/z) des peptides issus de la digestion enzymatique des protéines de l'échantillon à analyser, ainsi que de leurs fragments qui sont générés au sein des instruments. Les stratégies d'identification des protéines à partir des mesures de ratio m/z des peptides et de leurs fragments associés seront présentées au chapitre II.

Les spectromètres de masse combinent 4 éléments : la source, qui permet l'ionisation et la volatilisation des molécules, une interface qui permet la transmission et la focalisation des ions vers l'analyseur, un analyseur qui « filtre » (sépare) les ions selon leur rapport m/z et un détecteur contre lequel les ions viennent frapper permettant la détection du signal.

C.1. Les sources

Les sources qui permettent d'ioniser les protéines sont dites « douces » puisque l'énergie qu'elles leur transmettent ne risque pas de les détruire. Il s'agit des sources MALDI (Matrix-Assisted Laser Désorption/Ionisation) [15] et électrospray [16]. Le développement de ces sources a été primé par un prix Nobel en chimie 2002 (J. Fenn et K. Tanaka). Seule la source électrospray utilisée pendant ce travail de thèse sera présentée dans ce manuscrit.

La source électrospray permet l'ionisation de molécules en solution à pression atmosphérique sous l'effet d'un champ électrique. Le champ électrique est appliqué entre l'extrémité d'un capillaire

métallique (qui forme le nébuliseur) connecté à un système de chromatographie liquide (ou d'infusion directe) et l'entrée du spectromètre de masse. Le principe de l'électrospray est présenté en figure I-4.

Lors de l'application du champ électrique, la goutte présente à l'extrémité du capillaire s'allonge et prend la forme d'un cône appelé « cône de Taylor ». Quand les forces coulombiennes à la surface du cône deviennent égales à la tension de surface, appelée limite de Rayleigh, un spray de gouttelettes chargées se forme [41]. La source étant chauffée, l'évaporation du solvant va provoquer une augmentation de la densité de charge au niveau des gouttelettes formées qui, atteignant une nouvelle fois la limite de Rayleigh, génèrent de multiples gouttelettes filles [41]. Deux modèles ont été proposés pour expliquer le transfert des ions en phase gazeuse à partir des gouttelettes chargées. Le modèle de Dole propose que ce mécanisme de génération de gouttelettes se poursuive jusqu'à atteindre un seuil « ion par gouttelette » [42], tandis que le modèle d'Iribarne et Thomson stipule que les ions peuvent être désorbés directement à partir des gouttelettes chargées contenant plusieurs ions en raison des forces coulombiennes exercées sur ces gouttelettes sous l'effet de l'évaporation du solvant [43].

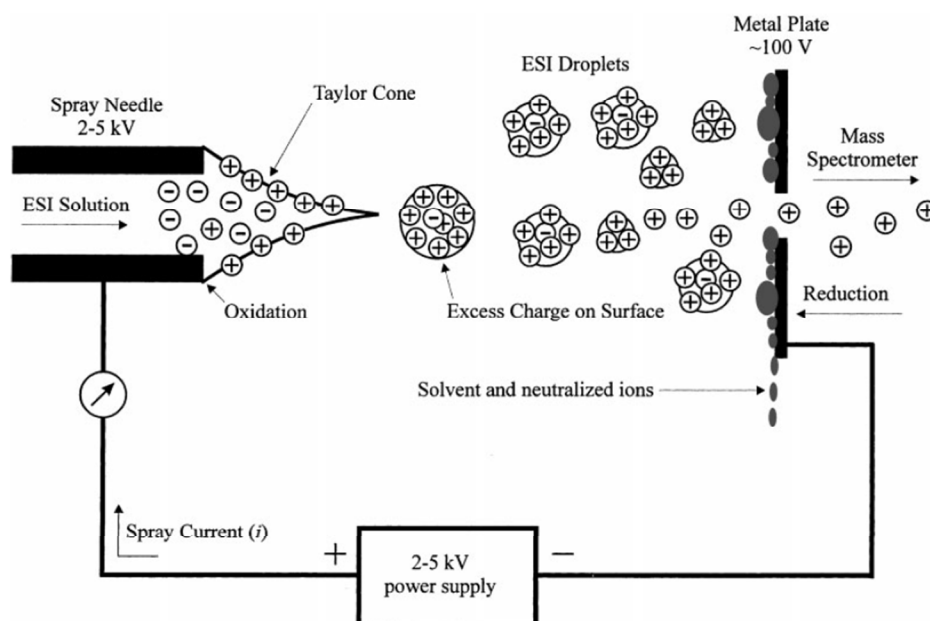


Figure I-4 Principe de formation des ions avec une source électrospray (extrait de [41])

Sous l'effet d'un champ électrique, la goutte à l'extrémité du capillaire métallique prend la forme d'un cône de Taylor. Lorsque les forces coulombiennes à la surface de ce cône égalent la tension de surface, un spray de gouttelettes chargées se forme. Sous l'effet de la chaleur, le solvant s'évapore donnant lieu à la génération de gouttelettes filles de densité de charge plus élevée. Les ions désolvatés sont finalement transférés en phase gazeuse selon le modèle de Dole ou d'Iribarne et Thomson.

C.2. Les analyseurs

Il existe différentes géométries d'analyseurs qui peuvent être caractérisés par la gamme de masses analysables, leur précision de masse, leur sensibilité et leur résolution. Le choix de l'analyseur doit donc être adapté à chaque type d'analyse, qu'elle soit qualitative ou quantitative.

La résolution est la capacité d'un analyseur à séparer deux pics sur un spectre de masse. On la calcule généralement d'après $M/\Delta M$ avec ΔM = la largeur à mi-hauteur notée FWHM (Full Width at Half Maximum). Ainsi, un analyseur ayant une résolution de 1000 pour un m/z égal à 100, aura la capacité de séparer les ions de m/z 100 et 100,1.

L'existence de différents isotopes stables pour la plupart des éléments chimiques conduit à la détection en spectrométrie de masse non pas d'un unique pic par ion détecté mais d'un massif isotopique. Les différents pics de ce massif sont annotés P, P+1, P+2, etc. P correspond au pic monoisotopique, son ratio m/z est calculé à partir de la masse molaire de l'ion qui ne tient compte que des isotopes stables les plus légers pour chaque élément (^1H , ^{12}C , ^{14}N , ^{16}O , etc.). Les ratios m/z des pics P+1, P+2, etc., sont calculés sur la base de l'incorporation d'au moins un isotope « lourd » (par exemple un ^{13}C). Les abondances relatives des différents pics du massif peuvent être calculées d'après l'abondance naturelle des différents isotopes. Un massif isotopique schématisé est présenté en figure I-5.

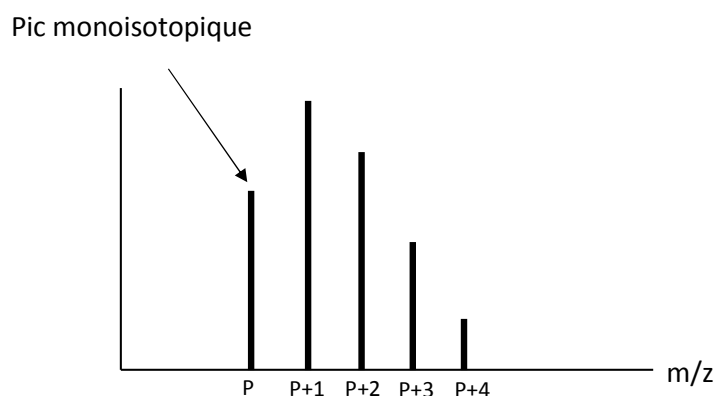


Figure I-5 Représentation schématisée d'un massif isotopique sur un spectre de masse

Si l'analyseur est de trop basse résolution, il se peut que la définition du massif isotopique soit assez limitée, dans ce cas, on parle de masse molaire moyenne.

Les protéines ou peptides analysés dans ce manuscrit peuvent être présents dans les échantillons sous de multiples états de charge. Cette information peut facilement être extraite des profils

isotopiques. Pour un peptide portant z charges, les pics P , $P+1$, $P+2$, etc. seront séparés de $(1/z) m/z$. Lorsqu'on calcule un état de charge à partir d'un massif isotopique, on dit qu'on effectue une déconvolution.

Les analyseurs principaux sont :

- L'analyseur quadripolaire (Q)
- Le piège à ions quadripolaire (IT)
- L'analyseur à temps de vol (TOF)
- L'analyseur à transformée de Fourier (FT-ICR)
- L'Orbitrap™

Seuls l'analyseur quadripolaire, le piège à ions quadripolaire et l'analyseur à temps de vol seront présentés dans ce manuscrit.

C.2.1. L'analyseur quadripolaire (Q)

Le principe de l'analyseur quadripolaire a été décrit par Paul et Steinwedel [44]. Il est constitué de 4 électrodes cylindriques connectées deux à deux (figure I-6). Un potentiel continu U ainsi qu'un potentiel alternatif de radiofréquence $V \cos \omega t$ sont appliqués sur ces deux paires d'électrodes de façon à ce que deux barres adjacentes soient de potentiels opposés. Les ions, accélérés selon l'axe z (en direction du détecteur), qui pénètrent dans le quadripôle subissent l'application du champ électrique oscillant. Les trajectoires des ions pénétrant dans le quadripôle sont décrites par les équations de Mathieu. Les ions stables, pour des tensions U et V données, sont transmis au détecteur tandis que les autres ions sont déviés [45].

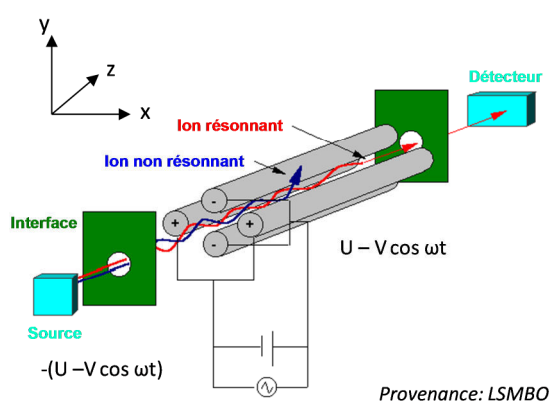


Figure I-6 Schéma d'un analyseur quadripolaire

Deux paires d'électrodes sur lesquelles s'appliquent un potentiel continu U et un potentiel alternatif $V \cos \omega t$ permettent la génération d'un champ électrique oscillant. En fonction des tensions U et V appliquées, les ions possèdent une trajectoire stable jusqu'au détecteur ou sont déviés.

Le diagramme de stabilité d'un ion donné (en fonction de sa valeur de m/z) au sein de l'analyseur en fonction des tensions U et V est représenté en figure I-7.

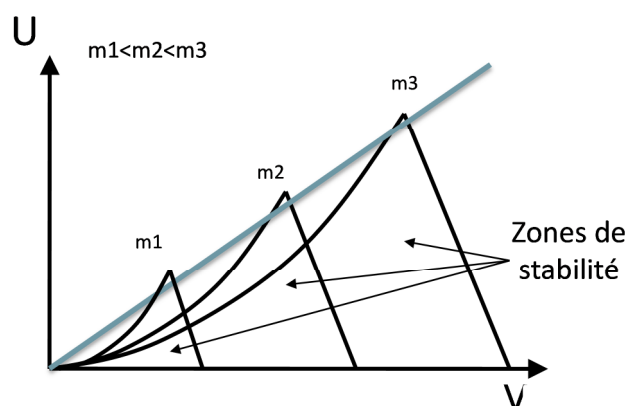


Figure I-7 Diagramme de stabilité d'un ion dans un analyseur quadripolaire en fonction des tensions U et V
La droite bleue représente l'augmentation des potentiels U et V avec U/V constant. Il s'agit de la représentation du fonctionnement du quadripôle lors de la génération d'un spectre de masse sur une gamme m/z donnée.

L'augmentation des potentiels U et V , représentée sous la forme d'une droite de fonctionnement en bleue sur la figure I-7, tout en maintenant un rapport U/V constant, permet de stabiliser successivement, et ainsi de transmettre au détecteur des masses de plus en plus élevées. La pente de la droite (le rapport U/V) définira la résolution avec laquelle les ions seront détectés ainsi que leur sensibilité. Plus la résolution sera élevée, plus la sensibilité diminuera. En général, le rapport U/V est paramétré de façon à obtenir une résolution unitaire.

En fonction des valeurs U et V , cet analyseur a la capacité d'être utilisé comme un « filtre » pour ne transmettre qu'un seul type d'ion, c'est d'ailleurs pour cette raison qu'il est généralement utilisé dans une configuration en tandem avec un second analyseur. Couplé à un autre analyseur, il peut également être utilisé en mode « RF Only », c'est-à-dire avec une tension U nulle, qui permet de transmettre tous les ions (dans une gamme de masse limitée) au second analyseur. Dans cette configuration, le quadripôle a un rôle uniquement focalisateur.

C.2.2. Le piège à ions (IT/ Ion Trap)

L'analyseur de type piège à ions, également basé sur les travaux de Paul et Steinwedel, est constitué d'une électrode annulaire et deux électrodes chapeaux au sein duquel une faible pression d'hélium est maintenue. Le schéma d'un analyseur de type piège à ions est présenté en figure I-8.

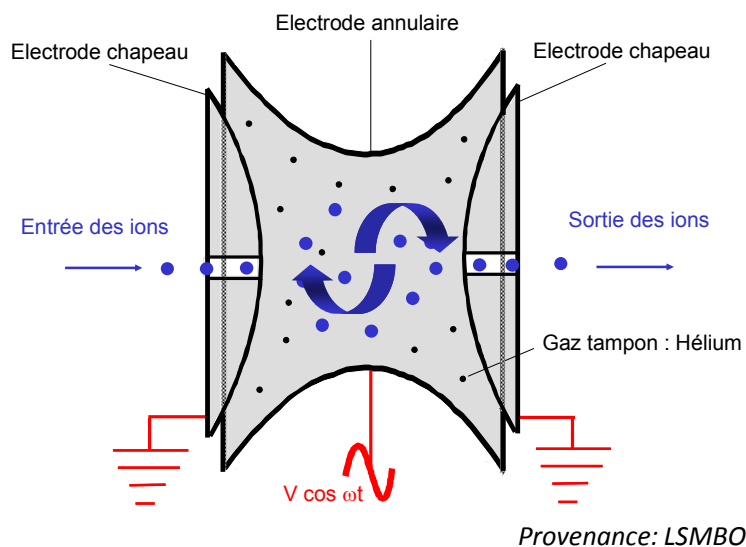


Figure I-8 Description d'un analyseur de type piège à ions

Les ions qui pénètrent dans un piège à ions sont confinés par application d'un potentiel alternatif de radiofréquence $V \cos \omega t$ à l'électrode annulaire. Comme l'analyseur quadripolaire, les ions suivent alors une trajectoire qui obéit aux équations de Mathieu. En fonction de la radiofréquence établie, une certaine gamme de masse m/z , qui aura une trajectoire stable dans le piège à ions, va pouvoir être confinée.

La détection des ions pour générer un spectre de masse se fait en éjectant les ions vers le détecteur. Par augmentation graduelle de l'amplitude de la radiofréquence sur l'électrode annulaire, la trajectoire des ions est déstabilisée et les ions sont éjectés dans l'ordre de leur m/z croissants pour être détectés. Cette éjection, dite simple, est limitée dans la gamme de masses d'ions pouvant être éjectés. Une deuxième stratégie est plus communément utilisée : l'éjection résonnante. Celle-ci consiste en l'ajout d'une tension alternative sur les électrodes chapeaux. En augmentant progressivement l'amplitude de cette tension, les ions vont être successivement éjectés du piège à ions en fonction de leur m/z .

Cet analyseur a la capacité de générer, à lui seul, des spectres de fragmentation. En mode MS/MS, les ions sont accumulés dans le piège à ions selon leur rapport m/z par application d'une radiofréquence adaptée à la gamme de masses à stabiliser. Ensuite, l'ion à fragmenter va être isolé dans le piège en déstabilisant tous les ions qui auront un rapport m/z inférieur ou supérieur à celui-ci. Les ions de plus faible m/z sont expulsés par éjection simple ou résonnante tandis que les ions de m/z supérieurs sont éjectés par éjection résonnante. Les ions isolés vont ensuite être dopés en énergie cinétique par ajustement de la fréquence de la tension alternative des électrodes chapeaux. Les ions

dopés rentrent en collision avec les molécules d'hélium, ce qui provoque leur fragmentation selon le modèle CID décrit dans ce chapitre en page 31. Les ions fils sont ensuite éjectés du piège à ions selon leur rapport m/z (éjection simple ou résonante). Il est possible de répéter ces opérations n fois de façon à générer des spectres MS^n . Le principe de génération des spectres MS^n est résumé en figure I-9.

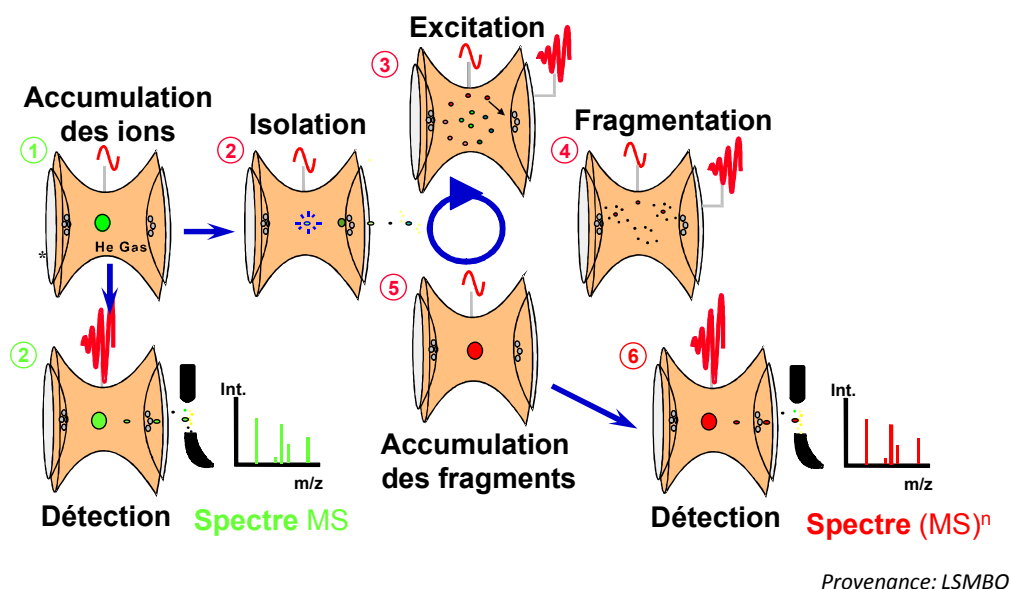


Figure I-9 Principe de génération d'un spectre MS^n à partir d'un analyseur de type piège à ions
 Les ions sont tout d'abord confinés dans l'analyseur. Les ions de m/z inférieurs ou supérieurs à l'ion à fragmenter sont ensuite expulsés. Les ions isolés sont ensuite dopés en énergie cinétique et se fragmentent en rentrant en collision avec les molécules d'hélium. Il est possible de poursuivre ce cycle de façon à générer un spectre MS^n .

C.2.3. L'analyseur à temps de vol (TOF/Time-Of-Flight)

L'analyseur à temps de vol a été décrit dans les années 1950 [46]. Il est composé d'un tube de vol, au sein duquel règne un vide poussé (1.10^{-7} mbar), dans lequel les ions sont séparés selon leur célérité sous l'effet d'une accélération électrique. Le schéma d'un analyseur à temps de vol (linéaire) est présenté en figure I-10.

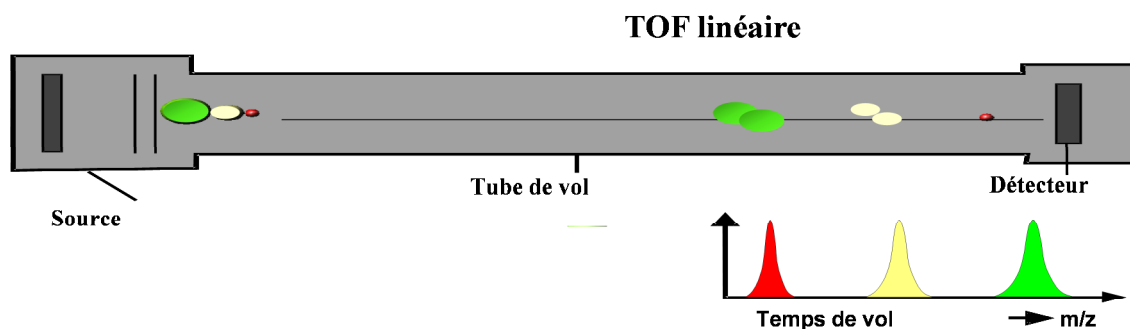


Figure I-10 Principe de fonctionnement d'un analyseur à temps de vol (linéaire) (provenance Bruker Daltonique)
 Sous l'effet d'un pulse électrique, des paquets d'ions sont propulsés dans un tube de vol au sein duquel règne un vide poussé.
 Les ions sont séparés selon leur célérité, plus un ion possèdera un ratio m/z faible, plus celui-ci traversera le tube de vol rapidement.

Les rapports m/z sont déterminés sur la base du temps nécessaire aux ions pour atteindre le détecteur après accélération selon l'équation :

$$t^2 = \left(\frac{m}{z}\right) \times \left(\frac{d^2}{2eV}\right)$$

Avec m = la masse du peptide, z = sa charge, d = la distance de vol entre la zone d'accélération et le détecteur, e = la charge d'un électron et V le potentiel auquel il est soumis. Plus l'ion sera de m/z faible, plus son temps de vol sera court.

La résolution d'un spectromètre de masse TOF est égale à :

$$R = \frac{m}{\Delta m} = \frac{d}{2\Delta z}$$

Avec m = la masse du peptide, Δm = la largeur à mi-hauteur du pic sur le spectre de masse, d = la distance de vol et $2\Delta z$ = l'épaisseur du paquet d'ions s'approchant du détecteur. La résolution étant ainsi proportionnelle à la longueur du tube de vol, il est possible de gagner en résolution si celui-ci est allongé.

L'énergie cinétique des ions de même m/z peut être différente au stade de l'accélération, ce qui peut amoindrir la résolution de l'analyseur. Deux stratégies ont été développées pour palier à cette limite [45]:

- l'extraction retardée qui consiste à refocaliser les ions avant leur entrée dans le tube de vol.
- l'utilisation d'un réflectron : celui-ci est constitué d'un ensemble de lentilles portées à des potentiels croissants créant un champ électrique opposé à celui au sein duquel les ions sont accélérés. Il est placé à l'extrémité du tube de vol alors que le détecteur est aligné sur la zone

d'accélération. Les ions propulsés vont pénétrer dans le réflecteur, de façon plus ou moins profonde en fonction de leur énergie cinétique. Plus celle-ci sera élevée, plus ils pénétreront profondément. Les ions vont ensuite être « réfléchis » (le réflecteur est aussi appelé miroir électrostatique) vers le détecteur en refocalisant des ions de même m/z qui possédaient des énergies cinétiques différentes au moment de l'accélération électrique. Un schéma d'un analyseur temps de vol possédant un réflecteur est présenté en figure I-11.

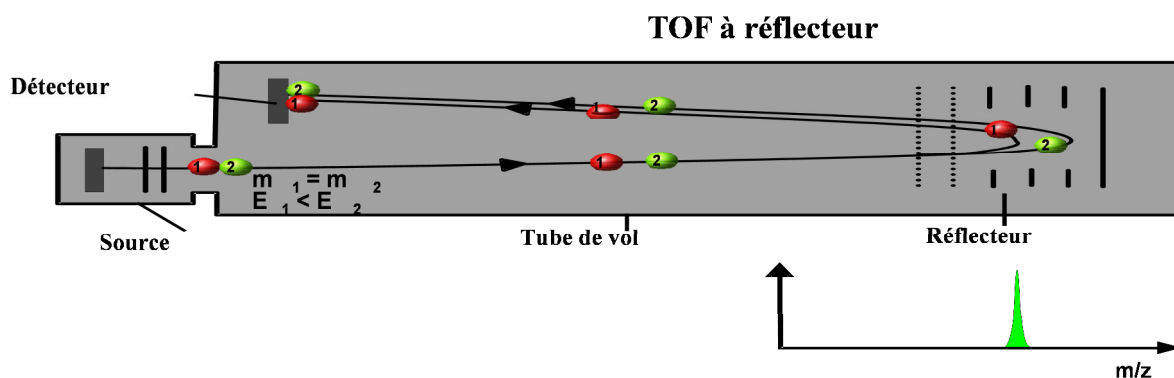


Figure I-11 Principe de fonctionnement d'un analyseur de type temps de vol possédant un réflecteur (provenance Bruker Daltonique)

Le réflecteur agit comme un miroir électrostatique de façon à refocaliser des ions de même m/z possédant des énergies cinétiques différentes après accélération.

L'analyseur à temps de vol est le plus sensible aux variations de température qui ont tendance à faire fluctuer la taille du tube de vol. Une calibration en masse régulière ainsi que l'utilisation de calibrant interne est indispensable avec ce type d'analyseur.

La génération des spectres au sein d'un TOF correspond à la sommation des spectres générés pour de multiples paquets d'ions propulsés consécutivement à l'entrée du tube de vol. Plus le nombre de sommations sera élevé, meilleur sera le ratio signal sur bruit du spectre final généré.

C.3. Les analyseurs hybrides

Les analyseurs hybrides permettent de combiner les propriétés de plusieurs analyseurs et ainsi de procéder à la fragmentation des peptides. Cette configuration est appelée spectrométrie de masse en tandem. Elle permet notamment la génération de spectres de fragmentation. Étant donné la capacité du quadripôle à isoler spécifiquement des ions, il est couramment combiné à d'autres analyseurs : quadripôle, analyseur à temps de vol, Orbitrap™, piège à ions quadripolaire. Il est également utilisé comme cellule de collision. Les deux types de spectromètres de masse hybrides présentés dans ce manuscrit sont : le triple quadripôle (QQQ) et le Q-TOF.

C.3.1. Le triple quadripôle (QQQ)

Trois quadripôles peuvent être combinés pour former un analyseur de type triple quadripôle. Le second quadripôle a une fonction de cellule de collision selon le mode CID décrit dans ce chapitre en page 31, tandis que les deux autres peuvent être utilisés en mode balayage (analyse d'une gamme m/z donnée) ou en mode « filtre ». Cette combinaison en fait un appareil offrant de multiples possibilités d'analyse. Les différents modes d'utilisation de ce type de spectromètre hybride sont décrits ci-dessous [47] et illustrés en figure I-12:

- **Le mode « Single Ion Monitoring mode/ Product Ion Scanning »** : le premier quadripôle sert à isoler un ion de m/z spécifique, le second sert de cellule de collision au sein duquel l'ion est fragmenté, et le troisième quadripôle sert à balayer l'ensemble de la gamme de masses de façon à générer un spectre de fragmentation pour l'ion isolé puis fragmenté.
- **Le mode « Precursor Ion Scanning »** : le premier quadripôle balaie la gamme de masses étudiée, le second sert de cellule de collision tandis que le troisième quadripôle « filtre » sur un seul type de fragment donné. Ce mode permet de détecter des ions qui auraient un type de fragment commun.
- **Le mode « Neutral loss scanning »** : le premier et le troisième quadripôle balaient la gamme de masses étudiée avec un décalage de masse imposé égal à la masse d'un fragment neutre perdu pendant la fragmentation. Ce mode permet de détecter des ions qui perdent un même fragment neutre donné.
- **Le mode « Selected Reaction Monitoring » (SRM/MRM)** : Le premier quadripôle « filtre » sur un ion donné, le second permet la fragmentation de celui-ci et le troisième « filtre » également sur un ion fragment spécifique. Ce mode permet d'analyser très précisément une transition ion-parent/ion-fils et est notamment utilisé pour la quantification ciblée de protéines.

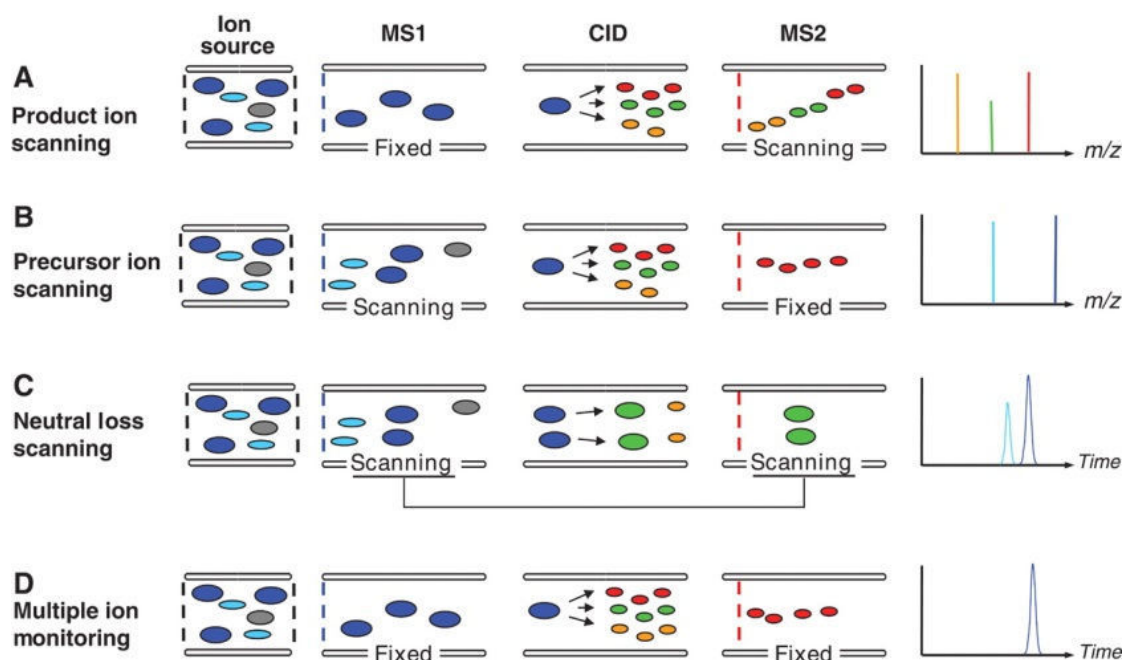


Figure I-12 Modes de fonctionnement d'un analyseur hybride possédant trois quadripôles (extrait de [47])

La configuration hybride triple quadripôle permet la réalisation de plusieurs modes d'acquisition et notamment le mode SRM (Multiple ion monitoring/Selected Reaction Monitoring/Multiple Reaction Monitoring) très spécifique qui est plébiscité pour l'analyse quantitative ciblée de protéines

C.3.2. L'analyseur hybride quadripôle/Temps de vol (Q-TOF)

Cet instrument hybride combine deux quadripôles, un utilisé comme premier analyseur et un second utilisé comme cellule de collision, et un analyseur à temps de vol. L'intérêt de l'implémentation d'un quadripôle en amont est de pouvoir isoler successivement certains ions de façon à générer des spectres de fragmentation spécifiques à un ion (peptide) donné.

Le spectromètre de masse permet ainsi de générer :

- Un spectre de masse (MS) : les deux quadripôles sont alors utilisés en mode « RF Only », ils focalisent et transmettent l'ensemble des ions sur une gamme de masses donnée jusqu'au tube de vol qui est utilisé pour générer un spectre de masse haute résolution.
- Des spectres de fragmentation (MS/MS) : le premier quadripôle est utilisé comme « filtre » pour isoler un peptide donné (ou une gamme de masse pour les stratégies « Data Independent Acquisition » présentées à la fin de ce chapitre), celui-ci sera fragmenté dans la cellule de collision, le tube de vol permet ensuite la génération d'un spectre haute résolution à partir des fragments du peptide isolé.

Le chapitre IV « Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes » (en page 86) est consacré à l'optimisation des méthodes nanoLC-MS/MS sur ce type de spectromètre de masse hybride et le principe de fonctionnement de cet analyseur hybride y sera plus amplement détaillé.

C.4. Comparaison des différentes performances d'analyseurs

Le tableau 1 résume les performances des différents types d'analyseurs utilisés au cours de cette thèse.

Tableau 1 : Résumé des performances des spectromètres de masse utilisés.

	Analyseur hybride Q-TOF		Analyseur hybride QQQ		Piège à ions (ion trap)
Instrument	maXis 4G Bruker Daltonique	Impact-CaptiveSpray Bruker Daltonique	6410 Agilent Technologies	TSQ Vantage Thermo	HCT Plus Bruker Daltonique
Résolution (m/z 922)	50000	40000	1000	1000	2000
Précision de masse	5 ppm	10 ppm	0,7 Da	0,7 Da	0,3 Da
Gamme de masse	100-2500	100-2500	100-1000	100-3000	250-2000
Temps de scan (minimum)	≈ 100 ms	≈ 50 ms	≈ 150 ms	≈ 20 ms	1 s
Gamme dynamique	+	+	+++	+++	++
Sensibilité	++	+++	++++	++++	+

Les différents analyseurs utilisés sont tout à fait complémentaires. Les appareils les moins résolutifs (de type QQQ) sont les instruments les plus sensibles et possèdent la gamme dynamique analysable la plus étendue grâce notamment à l'utilisation de mode d'acquisition très spécifique comme le mode SRM. Leur utilisation est extrêmement répandue pour la quantification ciblée de protéine. Les analyseurs Q-TOF sont moins sensibles et possèdent une gamme dynamique plus réduite, en revanche, ils sont très résolutifs et très précis en masse, ce qui les rend très utiles pour la caractérisation structurale ou la réalisation de stratégie quantitative basée sur l'extraction des courants d'ions (XIC). Les analyseurs de type pièges à ions présentent l'intérêt majeur de pouvoir générer des spectres MSⁿ qui permettent l'analyse très spécifique de molécules données mais leur intérêt a un peu faibli ces dernières années devant le développement des spectromètres Q-TOFs et OrbitrapTM (non décrit) de dernière génération qui sont précis en masse, résolutifs et rapides.

D. Les détecteurs

Le rôle du détecteur dans un spectromètre de masse est de convertir le signal des ions qui lui parviennent en un signal électrique et de l'amplifier. Les détecteurs de type multiplicateurs d'électrons sont les plus communs pour les spectromètres de masse de type quadripôle ou à temps de vol. On distingue les multiplicateurs d'électrons à dynodes discrètes (séparées) ou continues (Channeltron, Multi Channel plate, MCP). Le principe général de ces détecteurs consiste à utiliser des dynodes (séparées ou continues) qui permettent de générer des électrons secondaires à partir de l'impact d'un ion sur une surface semi-conductrice. Ces électrons secondaires, vont au contact d'une nouvelle surface semi-conductrice (la même dans le cas des dynodes continues, une autre dans le cas des dynodes discrètes) générer un plus grand nombre d'électrons de manière répétitive. Ces processus de générations sont amplifiés jusqu'à pouvoir récupérer un courant électrique mesurable [45].

E. Les modes de fragmentation et l'interprétation des spectres MS/MS

Il existe trois modes de fragmentation en analyse protéomique : CID (Collision Induced Dissociation) [48], ETD (Electron Transfer Dissociation) [49] et ECD (Electron Capture Dissociation) [50]. En analyse protéomique, le mode de fragmentation le plus couramment utilisé est le mode CID qui est le plus adapté à la fragmentation des peptides tryptiques. Au sein de ce mode, les peptides sont dopés en énergie cinétique (soit par l'accélération des ions en entrée de cellule de collision pour le quadripôle, soit par ajustement de la radiofréquence appliquée sur l'électrode chapeau dans les analyseurs de type piège à ion quadripolaire). Les peptides ainsi accélérés rentrent alors en collision avec des molécules de gaz inerte (Hélium, Azote, Argon, etc.) qui remplissent la cellule de collision. La fragmentation de ces peptides suit le modèle du proton mobile [51]. Ce type de fragmentation est parfaitement adaptée aux peptides tryptiques qui portent majoritairement deux charges (en position N-ter et sur la chaîne latérale de l'arginine ou de la lysine, la trypsine clivant spécifiquement en position C-ter de ces deux acides aminés). Ce mode de fragmentation provoque la rupture de la liaison peptidique.

Afin d'assigner les spectres MS/MS à un peptide donné, il est nécessaire de bien connaître les règles de fragmentation, c'est-à-dire les types d'ions générés majoritairement par le mode de fragmentation utilisé pour générer les données MS ou MS/MS [52]. La nomenclature de Biemann [53], utilisée pour la qualification des ions générés, est présentée sur la figure I-13. Ainsi en mode CID, on génère plutôt des ions de type b et y.

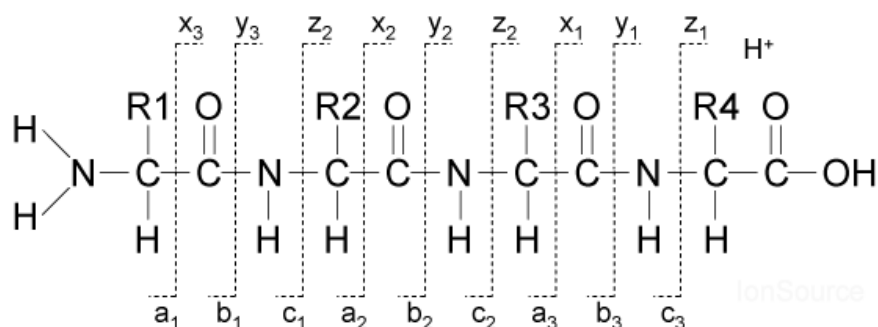


Figure I-13 Nomenclature de Biemann [53]

Les mécanismes de fragmentation les plus couramment utilisés correspondent à la génération des ions *b* et *y* (CID) et *c* et *z* (ETD, ECD)

Les différences de masse entre les différents ions détectés sur le spectre MS/MS permet de déterminer la séquence en acide aminé d'un peptide. Cette identification est illustrée en figure I-14.

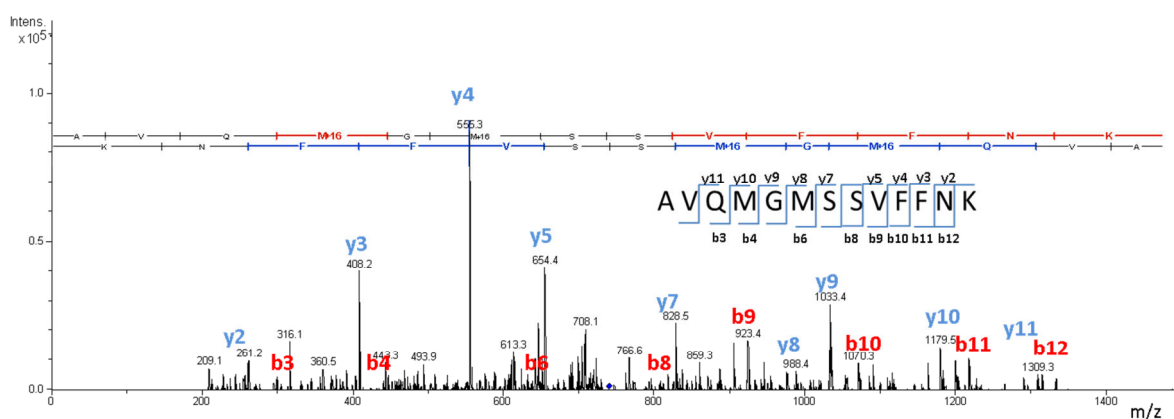


Figure I-14 Exemple de spectre annoté selon la nomenclature de Biemann [53]

La différence de masse entre les ions consécutifs d'une même série permet de définir la séquence en acide aminé du peptide fragmenté.

F. La complémentarité des enzymes et des modes de fragmentation

Afin d'analyser un protéome donné plus en profondeur, il est possible de multiplier les conditions de digestions enzymatiques et d'utiliser d'autres protéases que la trypsine, qui est la plus couramment utilisée, comme la LysC, l'ArgC, l'AspN et la GluC. Celles-ci induiront des sites de clivages différents (éventuellement aspécifiques) de ceux induits par la trypsine qui clive les protéines en C-ter des lysines et arginines. La trypsine génère de nombreux peptides qui possèdent moins de 6 acides aminés (environ 50 % pour le protéome de levure *Saccharomyces cerevisiae*) [54]. Ces peptides sont trop courts pour pouvoir être détectés en LC-MS en raison d'une trop faible rétention sur les systèmes chromatographiques et de leur spectre de fragmentation qui sont peu informatifs [55]. Ainsi, l'utilisation d'enzymes permettant de générer des peptides plus longs va permettre d'obtenir une meilleure couverture de séquence protéique [56]. Il est également intéressant de multiplier les

modes de fragmentation, d'autant que les peptides non tryptiques ne se fragmentent pas toujours efficacement en CID (en raison de leurs état de charge élevés par exemple) [54, 57].

G. Les modes d'acquisition des données

Deux modes d'acquisition existent aujourd'hui en protéomique. Ces modes concernent la sélection des ions pour la génération de spectres de fragmentation : soit quelques ions sont sélectionnés, isolés puis fragmentés successivement, en général sur la base de leur intensité sur le spectre MS précédent, soit l'intégralité des ions élués au cours d'une analyse seront fragmentés. Ces deux modes, « Data Dependant Acquisition » (DDA), et « Data Independant Acquisition » (DIA) décrits plus en détails ci-dessous, sont illustrés en figure I-15 (respectivement à gauche et à droite).

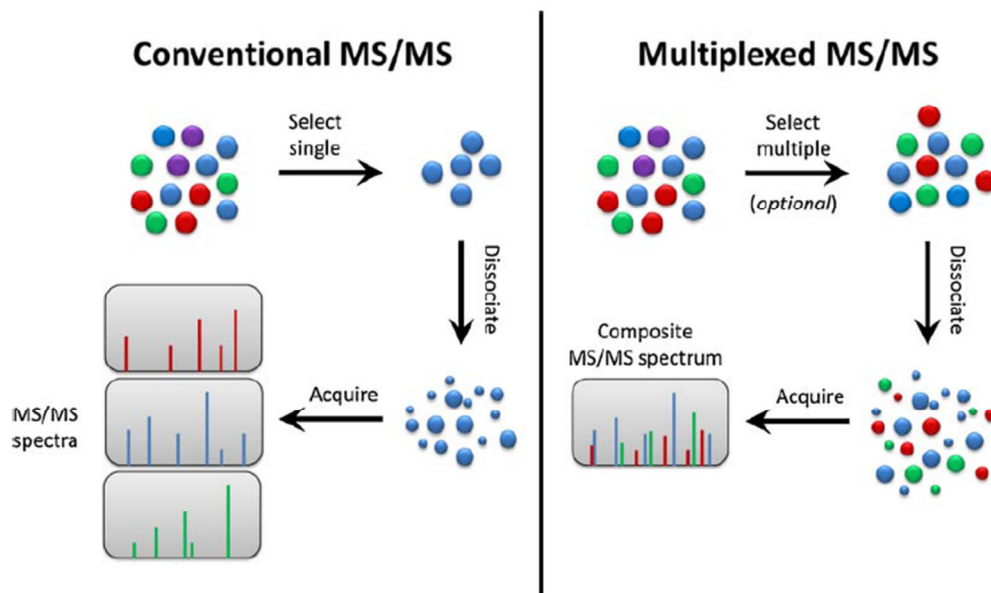


Figure I-15 Illustration des deux modes d'acquisition des données MS/MS (extrait de [58])

Les peptides sont soit sélectionnés successivement sur la base de leur intensité pour être isolés puis fragmentés, soit tous fragmentés de façon plus ou moins simultanée.

G.1. Le mode « Data Dependant Acquisition » (DDA)

Le déroulement des analyses réalisées en mode DDA est présenté en figure I-16.

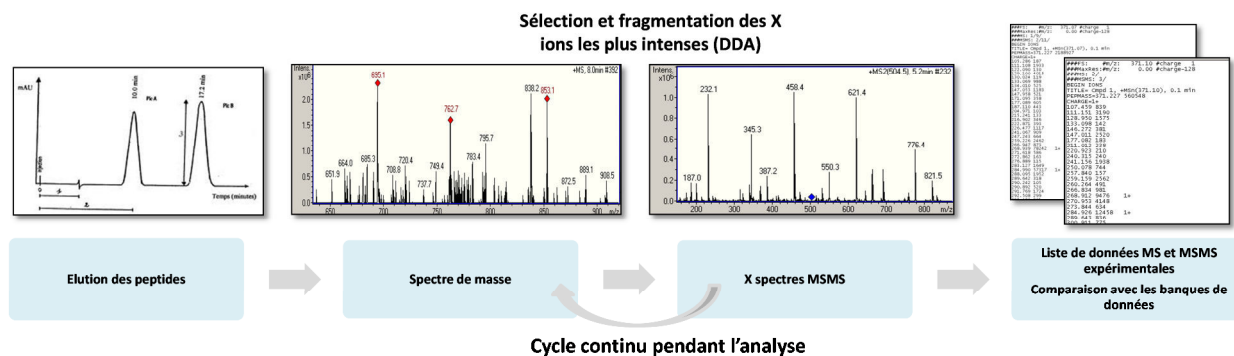


Figure I-16 Principe de la génération des spectres MS et MS/MS au sein d'une analyse nanoLC-MS/MS réalisée en mode DDA

La génération des spectres MS et MS/MS est réalisée de façon cyclique : au cours de l'élution des peptides, une succession de scan MS et MS/MS est réalisée. En général, un scan MS est suivi de plusieurs scans MS/MS sous la forme d'un « TOP N », c'est-à-dire 1 MS, N MS/MS. A la suite de chaque spectre MS généré, N peptides sont sélectionnés en fonction de leur intensité pour être isolés puis fragmentés successivement dans la cellule de collision [60]. Les spectres de fragmentation sont ainsi générés successivement jusqu'à la prochaine génération de spectre MS où le cycle reprendra. Plus le spectromètre de masse est rapide, plus il est possible de générer un nombre important de spectre MS/MS entre chaque spectre MS. Le paramétrage de ces méthodes sera détaillé plus amplement dans le chapitre IV « Optimisations instrumentales pour l'analyse d'échantillons biologiques complexes » en page 88.

G.2. Le mode « Data Independent Acquisition » (DIA)

De nouveaux modes d'acquisition ont été développés ces dernières années notamment pour l'analyse protéomique quantitative. Contrairement au mode DDA, ces modes consistent en la fragmentation de l'ensemble des peptides élués au cours d'une analyse soit par la fragmentation simultanée de l'ensemble des ions transmis à l'analyseur à chaque instant, soit par la fragmentation de peptides isolés par fenêtres d'environ 25 Daltons [59]. Aujourd'hui, il n'existe pas d'algorithme permettant d'identifier des peptides à partir de spectres de fragmentation générés par ce mode d'acquisition. Ils ne sont utilisés qu'en analyse protéomique quantitative et seront décrit plus amplement dans le chapitre III « Stratégies d'analyses protéomiques quantitatives » en page 75.

Chapitre II : Stratégies d'analyses protéomiques qualitatives

Chapitre II : Stratégies d'analyses protéomiques qualitatives

Les spectromètres de masse permettent de générer des spectres de masse MS ou MS/MS à partir desquels il va être possible d'extraire des informations qualitatives (identifications des peptides détectés) et/ou quantitatives (abondance relative d'un peptide entre deux conditions données). De nombreuses stratégies qualitatives et quantitatives ont été développées de façon à répondre à la diversité des thématiques biologiques en fonction de la disponibilité du protéome de l'espèce étudiée, du type de quantification globale ou ciblée à réaliser, etc. Les stratégies majeures existantes sont présentées dans ces deux prochains chapitres.

I. Les stratégies d'identifications chez les organismes séquencés

Il existe 2 stratégies majeures d'identification des protéines par spectrométrie de masse. Celles-ci sont basées soit sur l'interprétation des spectres MS, soit sur l'interprétation des spectres de fragmentations MS/MS.

A. L'empreinte peptidique massique

L'« empreinte peptidique massique » (PMF, Peptide Mass Fingerprint) identifie les protéines au travers de la comparaison des listes de rapports m/z expérimentales (extraits à partir des spectres MS) aux m/z théoriques déduits de la digestion *in silico* des protéines contenues dans des banques de données protéiques (décrites dans la suite du chapitre en page 39) à partir d'algorithmes de recherche [60]. Il est nécessaire de soumettre à ces algorithmes les valeurs m/z détectées ainsi que les états de charge respectifs. Il est également nécessaire de paramétrer en amont de la recherche dans les banques de données :

- La tolérance accordée sur la masse du peptide
- La banque protéique dans laquelle soumettre la recherche
- Le nombre de site de clivage enzymatique, qui peuvent être manqués, toléré
- Les modifications post-traductionnelles fixes ou variables susceptibles d'être détectées

Cette approche est rapide et a été largement utilisée pour des échantillons de faible complexité analysés par MALDI-MS [61]. En revanche, le volume des banques de données protéiques ne cesse d'augmenter et cet enrichissement exponentiel augmente la probabilité de retrouver des peptides de

séquences différentes ayant pourtant un même rapport m/z , ce qui risque de conduire à des identifications par PMF ambiguës voire erronées [62]. La stratégie PMF a donc été un peu délaissée au profit des stratégies par empreinte de fragments peptidiques qui sont plus spécifiques.

B. L'empreinte de fragments peptidiques

L'« empreinte de fragments peptidiques » (PFF), identifiée comme telle en 2004 [63] par analogie à la stratégie « empreinte peptidique massique », identifie les peptides en comparant les masses expérimentales issues de l'analyse des peptides, et de leurs fragments (extraits à partir des spectres MS/MS), à des listes déduites à partir de la digestion et fragmentation *in silico* des séquences contenues dans les banques de données protéiques à l'aide d'algorithmes de recherche [64] [65].

Les informations suivantes doivent être extraites des spectres MS et MS/MS pour soumission aux algorithmes de recherche :

- les rapports m/z des ions parents isolés puis fragmentés
- les rapports m/z des fragments générés dans la cellule de collision
- les états de charge des ions parents et fragments (si disponibles)
- les intensités de ces ions parents et fragments, ces intensités vont notamment permettre d'extraire les pics d'intérêt du bruit de fond [66]

Les paramétrages nécessaires à la comparaison des masses expérimentales et théoriques à l'aide des algorithmes de recherche sont :

- La tolérance accordée sur la masse du parent
- La tolérance accordée sur la masse des fragments
- La banque protéique dans laquelle soumettre la recherche
- L'enzyme utilisée pour la digestion
- Le nombre de site de clivage enzymatique manqué toléré
- Le type de fragments générés en fonction du mode de fragmentation utilisé.

Parmi les algorithmes de recherche principaux, on trouve Sequest [65], Mascot [67], Omssa [66], et X!Tandem [68]. Ces algorithmes attribuent un score statistique (score d'ion) à chaque assignation de séquence pour un spectre MS/MS donné. Pour l'algorithme Mascot, ce score est égal à « $-\log(p)$ »

avec p la probabilité que l'assignation soit due à un résultat « faux-positif ». Plus le score est élevé, plus la confiance dans cette assignation sera élevée. L'algorithme de recherche Mascot attribue également un score « d'identité » relatif à l'espace de recherche (la taille de la banque protéique dans laquelle est soumise la recherche). Plus l'espace de recherche est grand, plus le score d'identité sera élevée, de façon à ne pas augmenter le pourcentage de résultats « faux-positifs ». Si le score « $-\log(p)$ » attribué au spectre MS/MS est supérieur au score d'identité, alors la probabilité que l'assignation du spectre soit due au hasard est inférieure à 5% [67].

Le principe de fonctionnement de l'ensemble de ces algorithmes n'est pas toujours connu précisément, même si on en trouve des descriptions [52]. Les algorithmes de ces moteurs de recherche n'étant pas strictement identiques, utiliser plusieurs algorithmes de recherche pour interpréter un même ensemble de données présente un intérêt en termes de validation de résultat [69]. La confiance dans les spectres assignés de façon similaire par plusieurs moteurs de recherche simultanément sera augmentée. En revanche, le fonctionnement différent de ces algorithmes peut évidemment conduire à des assignations spécifiques par certains moteurs de recherche [52], ce qui présente un intérêt dans le but d'augmenter les couvertures de protéomes étudiés.

Cette stratégie qualitative sera appliquée à l'ensemble des projets décrits dans ce manuscrit.

C. Les banques de données

Les banques de données protéiques utilisées pour ces recherches représentent l'outil de travail de base pour l'assignation de spectres MS et/ou MS/MS. Les banques protéiques utilisées sont issues de l'annotation des banques génomiques. Il se peut qu'elles comportent des erreurs car la plupart de ces banques de données sont générées de façon automatique sans validation manuelle [4]. Il est donc nécessaire de connaître la qualité du contenu des banques protéiques utilisées.

La banque NCBI nr [70] est une banque protéique qui regroupe les séquences issues des annotations réalisées à partir de multiples banques nucléotidiques telles que « GenBank/ EMDL/ DDBL » ou d'autres bases de données protéiques telle que Swiss-Prot (présentée ci-dessous). Ces annotations ont été générées automatiquement et cette banque contient aujourd'hui de nombreuses redondances, des erreurs non corrigées et est très volumineuse.

La banque Swiss-Prot [71] est aujourd'hui la banque de données protéiques de référence. Cette banque est issue de l'annotation des banques de séquences nucléotidiques provenant du consortium

« GenBank/EMDL/DDBL » [72]. Les annotations de chacune des entrées ont été validées par une équipe d'experts en annotation. La banque TrEMBL contient les séquences annotées *in silico* des séquences nucléotidiques « GenBank/EMDL/DDBL » qui n'ont pas encore fait l'objet d'une validation manuelle. Les deux sont représentées sous le consortium Uniprot/KB depuis 2003.

La plupart des banques de données protéiques utilisées dans ce manuscrit sont issues de la banque Swiss-Prot qui fait référence dans le domaine de la protéomique [73]. En général, les recherches sont effectuées dans des banques restreintes à la taxonomie étudiée dans le but de réduire les temps d'interrogations dans les banques de données et de limiter la probabilité d'obtenir des résultats « faux-positifs » qui croit avec la taille de l'espace de recherche.

D. Validation

Les identifications réalisées à l'aide des algorithmes de recherche sont ensuite soumises à des logiciels de validation et de filtre des assignations tels que le logiciel Scaffold (Proteome software).

Afin de renforcer la confiance dans les identifications obtenues, des critères de filtres restrictifs sur la valeur du score minimal accepté pour une identification sont appliqués aux résultats ainsi qu'une stratégie dite « target-decoy » qui permet l'estimation du taux de « faux-positifs » (FDR, False Discovery Rate) parmi les identifications [74]. Cette stratégie consiste à créer une banque de données qui contient à la fois la banque de données cible (« target ») et les séquences inversées de toutes les protéines contenues dans la banque cible (« decoy »). C'est l'identification de ces séquences inversées (donc fausses) qui permettra de déterminer le taux de « faux-positif » dans les résultats.

Le calcul du taux de « faux-positif » se fait d'après l'équation suivante :

$$\frac{\text{Nombre de séquences "Decoys"}}{\text{Nombre de séquences ("Decoys" + "Target")}} \times 100$$

II. Les stratégies d'identifications chez les organismes dont le génome est non séquencé ou non annoté

A. L'empreinte de fragments peptidiques

Chez les espèces dont le génome n'est pas disponible ou annoté, il est possible d'appliquer les stratégies PFF et d'utiliser, pour la soumission des recherches, les banques de données protéiques d'espèces proches phylogénétiquement qui partageront des homologues de séquences peptidiques complètes avec l'espèce étudiée. Le choix de la banque sera donc très important et devra privilégier les taxonomies les plus proches de l'espèce étudiée afin de maximiser la probabilité de disposer de séquences les plus conservées possibles. Pour augmenter la probabilité d'analyser une séquence peptidique effectivement conservée, il pourra aussi être utile d'analyser le plus grand nombre de peptides possible par protéines [75]. Pour cela, une décomplexification poussée des échantillons pourra être réalisée, avec par exemple l'utilisation d'une séparation sur gel d'électrophorèse bidimensionnelle.

Ce type d'approche sera appliqué à la quantification de la vitellogénine dans le plasma de tortue Luth décrit dans le chapitre VI en page 153 et à la caractérisation du protéome musculaire de l'ours brun décrit dans le chapitre VII en page 183, deux espèces, pour lesquelles les séquences protéiques ne sont que partiellement connues.

B. Séquençage *de novo*

Les stratégies PMF et PFF ne sont possibles que dans le cas d'études d'organismes séquencés ou chez des espèces non séquencées pour lesquelles des peptides issus de la digestion enzymatique sont conservés chez d'autres espèces. Dans le cas contraire, l'identification des peptides peut être basée sur le séquençage *de novo*, c'est-à-dire l'interprétation experte des spectres MS/MS [76]. Le but de cette interprétation est de déterminer la séquence primaire de fragments peptidiques d'après les spectres MS/MS. Cette détermination peut être réalisée manuellement mais elle requiert le plus souvent l'utilisation d'algorithmes de traitement des données automatisés tels que PEAKS Studio [77] ou PepNovo [78] du fait de la quantité de données générées par analyse (plusieurs milliers de spectres). La séquence déterminée est ensuite comparée aux séquences existantes dans les banques de données protéiques par des programmes de type BLAST (Basic Local Alignment Search Tool) [9]. Comme pour l'approche précédente, le choix de la banque est très important puisqu'elle doit

contenir le protéome d'espèces suffisamment proche phylogénétiquement pour pouvoir contenir des homologies de séquences avec l'espèce étudiée.

Cette stratégie peut également être utilisée afin d'identifier de nouveaux variants de séquences chez des espèces dont le protéome est connu. En effet, les banques de données actuelles ne sont pas exhaustives et ne contiennent pas l'ensemble des variants existants [9].

Afin de ne pas introduire d'erreurs dans les identifications réalisées, il est préférable d'appliquer l'approche « *de novo* » à l'interprétation des spectres de très haute qualité (c'est-à-dire très informatifs et avec de multiples fragments dont le ratio signal sur bruit est largement supérieur au bruit de fond) non assignés par la recherche classique. Le logiciel RECOVER (<https://msda.unistra.fr>), développé par l'équipe bio-informatique du LSMBO, permet de réaliser cette étape de tri [79].

De même, l'utilisation de spectromètres de masse résolutifs et possédant une précision de masse élevée est indispensable pour limiter le risque de commettre des erreurs [76].

C. Protéogénomique

Bien que le génome de certaines espèces soit séquencé, il n'est pas obligatoirement annoté. Dans ce cas, malgré l'absence de banque de données protéiques, il est néanmoins possible de réaliser des analyses protéomiques et d'interpréter les données de spectrométrie de masse à l'aide des banques génomiques grâce à des algorithmes qui permettent, par exemple, de les traduire automatiquement dans les 6 cadres de lecture possibles et qui se basent alors sur des annotations déjà existantes et validées chez d'autres espèces [4]. Les données protéiques obtenues expérimentalement peuvent, en retour, permettre de corriger les annotations génomiques [4].

III. Qualité des assignations chez les espèces séquencées

Nous avons pu constater au LSMBO que, sur un échantillon complexe, seuls 30 à 50% des spectres MS/MS donnent lieu à une assignation avec une confiance élevée (FDR < 1 % et score d'ion Mascot ≥ 30). Les multiples raisons susceptibles d'expliquer ce fort pourcentage de spectres non assignés ou possédant une faible confiance d'assignation sont présentées en tableau 1 (liste non exhaustive).

Tableau 1 : Résumé des principales sources de non-assignation des spectres générés sur un spectromètre de masse Q-TOF (cette liste n'est pas exhaustive).

Possibles sources de non-assignation ou d'assignation de faible confiance des spectres MS/MS par l'algorithme Mascot	
Etats de charge	<ul style="list-style-type: none"> - Sélection de peptides monochargés qui ne donnent pas lieu à des spectres MS/MS correctement interprétables - Erreur de déconvolution qui entraîne une erreur dans la détermination de la masse d'un peptide
Recherche dans les banques de données	<ul style="list-style-type: none"> - Substitutions d'acides aminés - Présence de modifications post-traductionnelles - Site de clivage aspécifique
Qualité des spectres MS/MS	<ul style="list-style-type: none"> - Faible qualité de spectres (intensité insuffisante ou fragmentation inefficace) - Co isolement de plusieurs peptides au moment de la fragmentation, ce qui est nuisible pour l'interprétation du spectre MS/MS - Spectres peu informatifs générés pour des peptides très courts

A. Le cas des peptides monochargés

Les peptides monochargés donnent lieu à très peu d'identifications (selon l'analyse de données générées au LSMBO, moins de 5% des spectres monochargés générés sont assignés). Ce très faible taux d'assignations est lié aux difficultés de fragmentation de ces peptides. Ce phénomène est connu sous le terme de « séquestration du proton ». Un seul proton est séquestré au sein de la fonction basique des peptides tryptiques (lysine et arginine), ce qui les rend plus difficiles à fragmenter [51].

Les logiciels d'acquisition permettent d'exclure ce type de peptides pour la sélection en MS/MS mais il se peut qu'ils soient tout de même sélectionnés pour la fragmentation en raison d'une erreur de déconvolution réalisée par le logiciel d'acquisition (souvent due à une trop faible intensité). Le

nombre de spectres monochargés peut varier en fonction du paramétrage des méthodes nanoLC-MS/MS : si les seuils d'intensité pour la sélection des peptides est trop bas, les erreurs de ce type augmentent.

B. Les recherches dans les banques de données

Un certain nombre de spectres sont également non assignés à cause de la présence de modifications post-traductionnelles ou de sites de clivage aspécifiques qui n'étaient pas recherchés. Il existe également probablement de nombreux variants de séquences (résultant d'épissages par exemple) qui ne sont pas présents dans les banques de données protéiques. Rechercher systématiquement l'ensemble des modifications existantes, des sites de clivages aspécifiques ou la substitution d'acides aminés et donc augmenter de façon conséquente la taille des espaces de recherche, donnerait lieu à trop de résultats « faux-positifs » en termes d'identifications. De plus, le temps de recherche croît avec l'augmentation des espaces de recherche. Ces paramètres doivent donc être finement optimisés en fonction de chaque thématique.

C. La qualité des spectres MS/MS

L'évolution des spectromètres de masse doit encore évoluer de façon à obtenir des appareils toujours plus résolutifs, plus rapides et possédant des sensibilités suffisamment élevées pour pouvoir générer des spectres de fragmentation informatifs, et ce, même pour des protéines très faiblement exprimées. Une certaine proportion des spectres non assignés est liée à un manque d'intensité ou au co-isolément de plusieurs peptides précurseurs, dont les rapports m/z sont très proches, en amont de l'étape de fragmentation. Ainsi, le spectre de fragmentation contient les fragments de plusieurs précurseurs différents ce qui nuit à l'identification de la séquence par les algorithmes de recherche actuels qui ne savent pas pour le moment gérer ce type de spectres. Un parallèle peut être réalisé entre ces spectres de fragmentation et le mode d'acquisition de type DIA pour lesquels aucun algorithme d'identification n'a encore été développé. Le développement de quadripôles et de systèmes électroniques permettant la transmission très spécifique d'ion de rapport m/z donné sans perdre en sensibilité ou le développement de nouveaux types d'algorithmes de recherche permettrait de palier à ces difficultés.

Ainsi, il existe de nombreuses perspectives d'amélioration concernant les stratégies d'analyses qualitatives, et ce, notamment en complétant les banques de données protéiques existantes avec les multiples variants de séquence existants, en développant de nouveaux algorithmes de traitement des

données (qui pourraient être implémentés au sein des stratégies DIA) et en poursuivant le développement instrumental des spectromètres de masse en termes de sensibilité et de résolution.

IV. Conclusion et perspectives

La multitude de stratégies existantes en analyse protéomique qualitative permet aujourd'hui de travailler sur presque n'importe quelle espèce avec un degré plus ou moins élevé d'exhaustivité en fonction de la disponibilité du protéome de l'espèce étudiée. Cette puissance d'analyse est mise au service de la biologie dans le but de décrire, à l'échelle moléculaire, les mécanismes sous-jacents de certaines (patho) physiologies d'intérêt. Un certain nombre de ces stratégies seront donc directement appliquées dans le cadre des problématiques biologiques présentées au sein de ce manuscrit.

Chapitre III : Stratégies d'analyses protéomiques quantitatives

Chapitre III : Stratégies d'analyses protéomiques quantitatives

Pour améliorer la compréhension des mécanismes biologiques, l'analyse des protéomes n'est plus aujourd'hui seulement qualitative : la protéomique quantitative est devenue incontournable [12, 80]. De nombreuses stratégies d'analyse protéomique quantitative ont été développées, et chaque méthode possède ses spécificités, ses avantages et ses limites [11]. Le choix de la stratégie/méthode quantitative à appliquer dépend directement des caractéristiques propres à chaque thématique biologique et à chaque type d'échantillon.

On peut dissocier les stratégies quantitatives en deux grands types représentés en figure III-1:

- **les stratégies d'analyses ciblées** qui permettent de quantifier quelques protéines (jusque plusieurs dizaines) environ simultanément de façon absolue ou relative basée sur l'exploitation du spectromètre de masse hybride de type triple quadripôle et de son mode d'acquisition très spécifique SRM décrit dans ce chapitre en page 50.
- **les stratégies globales** qui permettent de faire de la quantification relative à très haut débit (jusque plusieurs milliers de protéines simultanément). Ces stratégies globales regroupent les stratégies basées sur des techniques de marquage des protéines ou des peptides (chimique, enzymatique et/ou isotopique) ou sur des techniques sans marquage (« label free »), qui ont vu le jour ces dernières années.

Le terme multiplexage correspond à la quantification simultanée de plusieurs protéines par échantillon et par analyse dans la suite du manuscrit.

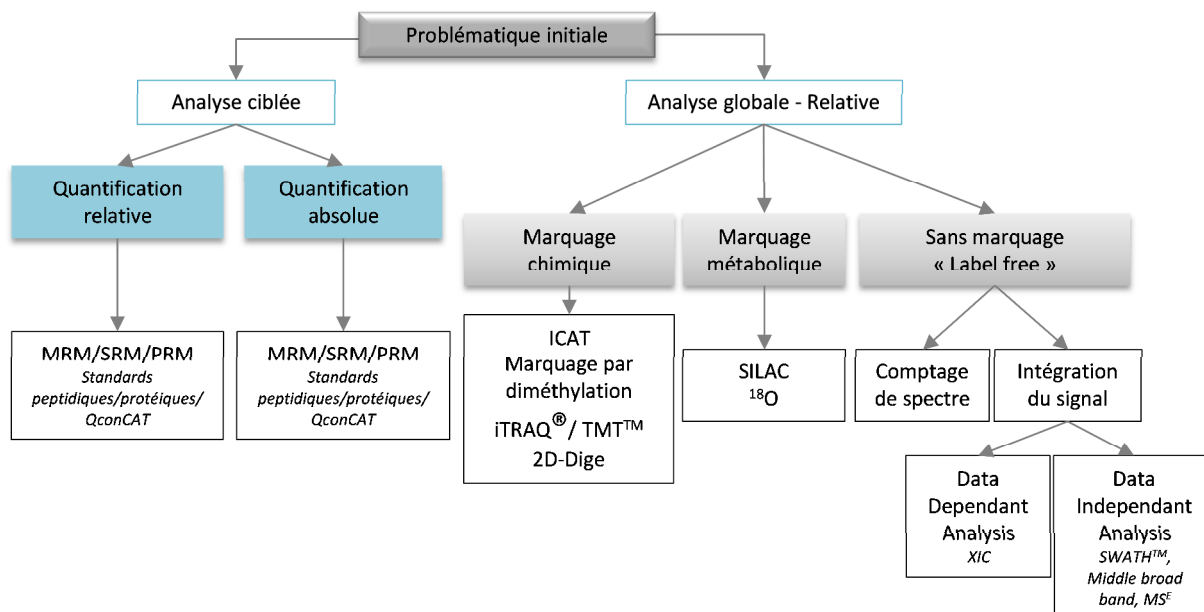


Figure III-1 Diversité des stratégies existantes dans le domaine de la protéomique quantitative

Les stratégies sont divisées en deux grands axes majeurs : les stratégies ciblées, qui peuvent être absolues ou relatives, et les stratégies globales qui sont éventuellement basées sur le marquage chimique ou enzymatique des molécules mais il en existe également différents types sans marquage.

Toutes ces stratégies sont présentées dans les pages suivantes, de façon plus détaillée pour la 2D-DIGE, la SRM et les stratégies sans marquage (XIC), qui sont les principales approches qui ont été appliquées aux différents projets présentés dans ce manuscrit.

I. Les stratégies ciblées

A. La stratégie SRM (Selected Reaction Monitoring)

A.1. Principe de la stratégie SRM

La stratégie SRM est une stratégie quantitative ciblée qui consiste en l'analyse exclusive d'une combinaison de peptides donnés. L'analyse spécifique de ces peptides s'appuie sur l'utilisation d'un spectromètre de masse hybride comprenant trois quadripôles et opérant en mode SRM (qui a donné le nom à la stratégie) décrit dans le chapitre I en page 28 [81]. Ce mode permet l'analyse très spécifique de transitions pour chaque peptide : ion parent/ion fragment. Le spectromètre de masse scanne l'ensemble des transitions sélectionnées de façon cyclique pendant une analyse entière ou sur une période donnée correspondant au temps d'élution d'un peptide à quantifier.

Plus le nombre de transitions (ion parent/ion fragment) sera élevé pour chacun des peptides à quantifier, plus la spécificité de l'analyse sera élevée. En effet, la co-élution parfaite de multiples transitions pour un seul peptide au sein d'une même analyse garantit la présence du peptide à quantifier et l'absence d'interférence (peptides co-éluant et possédant des valeurs de m/z très similaires au peptide d'intérêt) qui pourraient nuire à la justesse de la quantification.

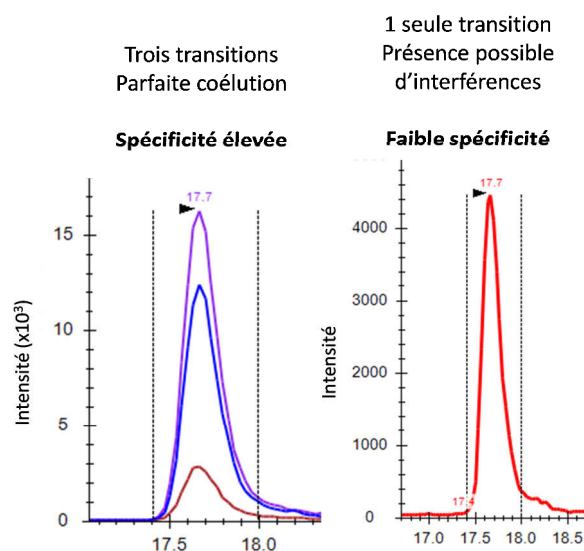


Figure III-2 Illustration du gain en spécificité acquis par l'augmentation du nombre de transitions analysées par peptide
Analyser systématiquement plusieurs transitions par peptide permet de gagner en spécificité.

A.2. Avantages et limites de la stratégie

La précision et la force de cette technique résident dans l'utilisation de standards internes : des peptides ou des protéines marqués aux isotopes stables, (plus ou moins) précisément quantifiés, et possédant, en couplage LC-MS, un comportement similaire (parfaite co-élution et ionisation) à celui des peptides endogènes dont le niveau de concentration est à évaluer. Les transitions spécifiques aux peptides standards seront analysées simultanément par le QQQ avec celles du peptide endogène. L'information quantitative par peptide sera déduite sur la base de la comparaison des intensités des signaux « peptide marqué/peptide endogène » acquis simultanément. L'utilisation de ces standards permet de compenser les éventuels biais techniques dus à la préparation d'échantillon, il est donc nécessaire de les intégrer le plus tôt possible au cours de celle-ci.

Ce type de stratégie quantitative s'appuyant sur l'utilisation d'un spectromètre de masse hybride de type triple quadripôle en mode SRM a d'abord été utilisé pour la quantification de petites molécules puis est devenue depuis une dizaine d'année la méthode de référence d'analyse ciblée des protéines. La spécificité de ce type d'acquisition permet de quantifier de très faible quantité de matériel

protéique (jusqu'à l'attomole en fonction de l'échantillon) et ce, sur des gammes dynamiques de concentration très importantes (de 4 à 5 ordres de grandeur) [82].

La stratégie SRM est souvent comparée aux techniques basées sur l'utilisation d'anticorps telles que le Western-Blot, la cytométrie en flux ou le dosage ELISA, reconnues dans le domaine de l'analyse protéique ciblée [81]. En plus de son extrême sensibilité, elle a l'avantage, par rapport à ces méthodes, de permettre la quantification simultanée de plusieurs peptides par protéine, ce qui équivaut à la quantification de plusieurs épitopes (soit plusieurs analyses en Western-Blot) en parallèle pour la quantification d'une seule et même protéine. La spécificité et la sensibilité des méthodes SRM en font une stratégie de choix pour la recherche ou la validation de biomarqueurs [83, 84].

La stratégie SRM permet également de quantifier plusieurs protéines simultanément. Cette possibilité de suivre spécifiquement plusieurs peptides par protéine et plusieurs protéines au sein d'une seule analyse s'est accrue ces dernières années avec l'évolution du parc instrumental. Les spectromètres de masse sont toujours plus rapides et plus sensibles, ce qui a permis la réduction des temps de scan dédiés à chacune des transitions (appelés « dwell time »), et a ainsi permis l'augmentation du nombre de transitions (et donc de peptides) analysées. L'implémentation du mode « scheduled » qui permet de suivre certaines transitions uniquement à un temps de rétention donné (autour du temps de rétention du peptide à analyser) a également permis d'augmenter le nombre de transitions suivies par analyse [85]. Aujourd'hui, on utilise la SRM pour quantifier simultanément plusieurs dizaines de protéines en routine (environ 50) sur la base de 20 ms de temps de scan par transition, 2 peptides par protéine et 2 transitions par peptide (ces valeurs pouvant varier d'une expérience à l'autre) [86].

La technologie SRM offre également la possibilité d'apporter des réponses quantitatives quant aux isoformes et modifications post-traductionnelles des protéines (impossible par Western-Blot) en analysant spécifiquement des peptides modifiés ou uniques entre différentes isoformes à quantifier.

La limite de cette technique réside dans l'impossibilité d'analyser simultanément plusieurs échantillons [81].

A.3. Méthodologie pour la mise au point d'une stratégie SRM

Les différentes étapes à considérer pour le développement d'une stratégie SRM, de la mise au point à l'analyse des échantillons, sont présentées en figure III-3 :

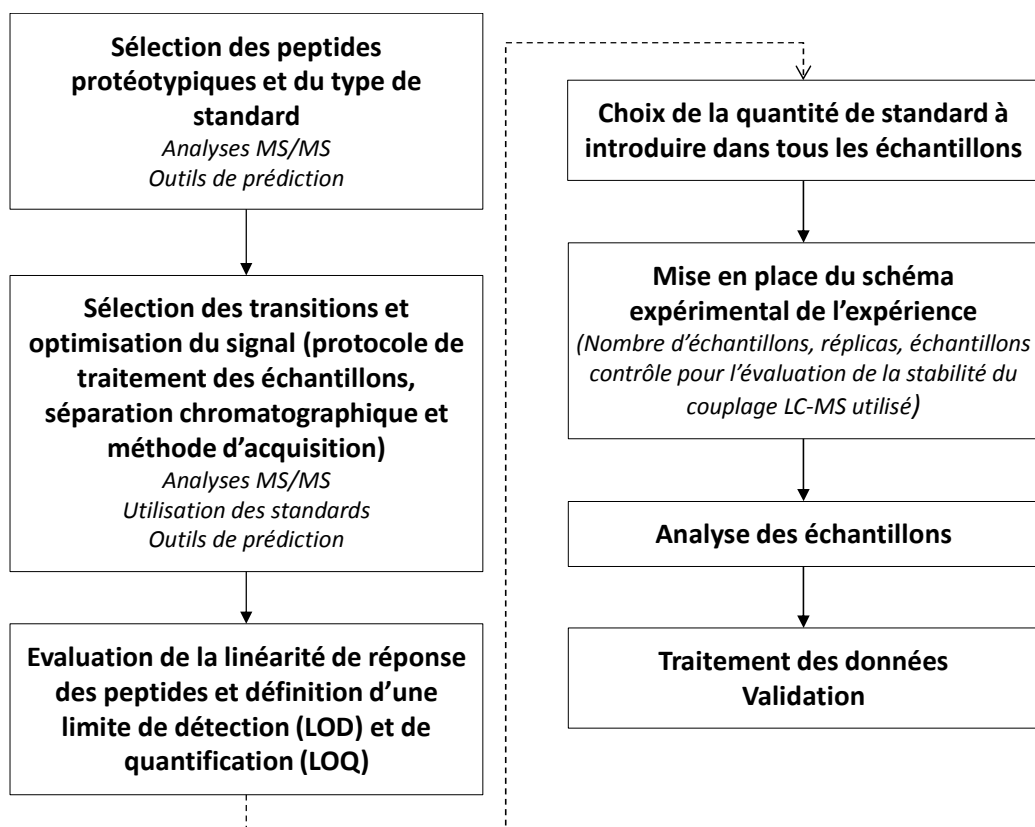


Figure III-3 Etapes de mise au point d'une méthode SRM pour la quantification ciblée de protéines

Les étapes de sélection des peptides, de choix du type de standard, de traitement des données et de leur validation sont décrites plus en détails dans la suite du chapitre.

A.3.1. Sélection des peptides

L'étape de développement majeure de cette stratégie réside dans le choix des peptides à sélectionner (puis quantifier) spécifiquement pour chaque protéine. Il n'y a pas de nombre spécifique de peptides à quantifier par protéine étant donné la variabilité des problématiques et des protéines à quantifier, de leur taille, du nombre de peptides tryptiques détectables pour chacune d'elles, etc. Idéalement, il est bénéfique de couvrir une fraction importante des séquences protéiques. La concordance des réponses de différents peptides pour une protéine donnée permettra de valider la

quantification de celle-ci. Il s'agit là de trouver le meilleur compromis entre le nombre de protéines à quantifier simultanément, leur recouvrement et le coût des standards utilisés.

Certaines règles sont désormais unanimement admises au sein de la communauté pour le choix des peptides à quantifier : [86]

- les peptides doivent être protéotypiques (uniques à la séquence de la protéine à quantifier et détectables en spectrométrie de masse).
- Ils ne doivent pas porter (si possible) de modifications post-traductionnelles (si elles ne sont pas l'objet de l'étude).
- Ils ne doivent préférentiellement pas contenir de site de clivage enzymatique susceptible d'être manqué.

De nombreux outils et bases de données partagées au sein de la communauté se sont développés ces dernières années de manière à guider le choix des peptides protéotypiques à suivre en SRM comme, par exemple, PASSEL [87, 88], Pride [89] ou PeptideAtlas (www.peptideatlas.org) [90], etc. pour les bases de données et Peptide Picker [91] pour les outils de prédiction.

Les peptides standards permettent d'optimiser les paramètres instrumentaux sur le spectromètre de masse (QQQ) de façon à obtenir la meilleure sensibilité possible pour chacun d'eux et leur homologue endogène. Il faut choisir les transitions à suivre spécifiquement par peptide et ajuster l'énergie nécessaire à leur fragmentation. Ces étapes peuvent être réalisées en infusant directement le standard sur le spectromètre de masse (s'il est disponible pur) ou en l'introduisant dans la matrice correspondant à l'échantillon à analyser. En général, trois transitions sont utilisées pour monitorer un peptide mais une seule est suffisante pour apporter une information quantitative. L'utilisation de multiples transitions est cependant recommandée pour s'affranchir de tout biais dû à la présence d'interférences [86, 92]. De même, il existe des outils d'aide aux choix des transitions à suivre spécifiquement pour chacun des peptides choisis comme par exemple SRMATlas (www.srmatlas.org) au sein du PeptideAtlas. En revanche, ces outils concernent l'étude d'espèces séquencées telles que l'homme, la levure, certaines bactéries et la souris. Le gradient chromatographique utilisé est également à optimiser de façon à ce que le moins de peptides possibles à quantifier ne co-éluent. Finalement, les paramètres d'acquisition du QQQ doivent être également optimisés de façon à avoir les temps de scans par transition les plus élevés possibles.

Cette stratégie quantitative est préférentiellement utilisée pour des organismes dont le protéome est connu mais elle peut également être appliquée chez des espèces « exotiques » en basant la sélection

des peptides à quantifier sur des morceaux de séquence des protéines d'intérêt qui ont été conservées au cours de l'évolution chez d'autres espèces dont le protéome est connu. Ce type d'application sera présenté dans ce manuscrit au cours du chapitre VI de la partie résultats concernant la quantification de la vitellogénine dans le plasma de tortue Luth en page 152.

A.3.2. Les standards

Les standards peuvent être incorporés aux échantillons sous différentes formes : peptidique [93], protéique [94] ou sous la forme de protéines recombinées en une succession de peptides d'intérêt (QconCAT) [95]. L'incorporation de ces différents types de standard doit donc être réalisée en adéquation avec le type de standard utilisé (figure III-4).

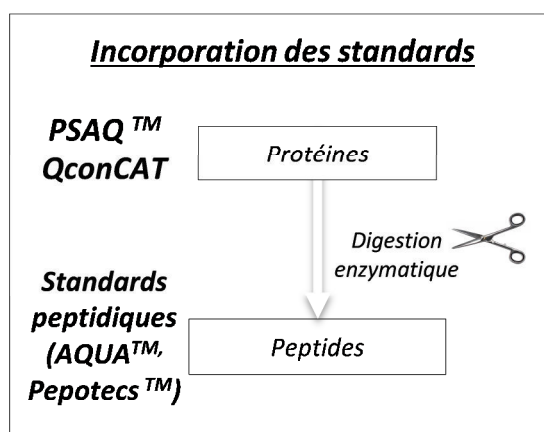


Figure III-4 Étapes d'incorporation des standards aux échantillons pour l'analyse SRM

Les standards protéiques (PSAQ™, QconCAT) doivent être incorporés dès le début de la préparation d'échantillons tandis que les standards peptidiques peuvent être intégrés après digestion enzymatique.

A.3.2.1. Les standards peptidiques

L'utilisation des standards peptidiques dans le domaine de la stratégie quantitative ciblée a été présentée en 2003 [93]. Il existe plusieurs types de standards peptidiques : certains sont hautement purifiés et précisément quantifiés (5 pmol/μL ± 5 %, AQUA™, Thermo Fisher Scientific), et permettent ainsi de faire de la quantification très précise, voire absolue, d'autres ne sont ni purifiés ni quantifiés précisément (Pepotecs™, Thermo Fisher Scientific) et ne permettent donc de ne réaliser que de la quantification relative. La gamme de prix de ces standards est assez large : quelques dizaines d'euros pour les peptides non purifiés et/ou non quantifiés (≈ 100 ng) à plusieurs centaines d'euros (pour quelques nanomoles) pour les peptides hautement purifiés et quantifiés.

Cette technique présente l'inconvénient de ne pas refléter le comportement protéique avant l'étape de digestion enzymatique puisque les standards introduits sont sous forme peptidique et que de ce fait, leur comportement en amont de l'étape de digestion, risque d'être différent de celui des protéines. Ainsi, ce type de quantification basée sur l'utilisation des peptides AQUA™ (Thermo Fisher Scientific) est parfois qualifiée de « quantification relative précise » [81].

A.3.2.2. Les standards QconCAT (Quantification conCATamer)

L'utilisation de protéines synthétiques combinant une succession de peptides d'intérêt « QconCAT » (*Quantification conCATamer*) a été présentée en 2006 [95]. Il s'agit d'une stratégie alternative à la stratégie basée sur l'utilisation des peptides AQUA™. Après sélection de peptides d'intérêt, un gène codant pour une protéine combinant l'ensemble des peptides choisis (il est possible d'intégrer des peptides de plusieurs protéines au sein d'un même standard) est synthétisé. Il est ensuite exprimé dans un milieu enrichi en acides aminés marqués, conduisant ainsi, après purification, à la synthèse d'une protéine marquée, combinant l'ensemble des peptides d'intérêt marqués. Cette stratégie aurait pu lever le biais évoqué pour les standards peptidiques au sujet de leur comportement pendant l'étape de digestion enzymatique. En revanche, la faible similarité de séquences avec les protéines endogènes à quantifier ne peut pas conduire à un comportement similaire à celui des protéines endogènes au cours des étapes en amont de la digestion enzymatique des protéines [81]. De ce fait, l'utilisation de ce type de standard ne permet pas de réaliser une quantification plus précise qu'avec les standards peptidiques.

A.3.2.3. Les standards PSAQ (Protein Standard Absolute Quantification)

Etant donné les limites concernant l'utilisation des standards peptidiques ou QconCAT, le standard idéal serait la protéine d'intérêt, marquée aux isotopes stables. Ces standards pourraient ainsi être incorporés au stade le plus précoce de la préparation d'échantillons et possèderaient le même comportement que la protéine à quantifier (notamment au stade de la digestion enzymatique par exemple). L'utilisation de ce type de standard a été présentée en 2007 [94]. Cette stratégie est nommée « PSAQ™ » (Protein Standard Absolute Quantification). L'ensemble des lysines et arginines est marqué aux isotopes stables ^{13}C et ^{15}N . Ainsi, la digestion enzymatique trypsique permet de générer des peptides qui contiennent un acide aminé marqué pour chaque peptide analysé. La force de cette stratégie réside dans la parfaite correspondance de séquences entre les protéines marquées et endogènes. Cette stratégie présente également le nombre le plus élevé de séquences peptidiques quantifiables par protéine puisque tous les peptides trypsiques ionisables uniques à la séquence de la

protéine posséderont un homologue marqué et pourraient être utilisés. En revanche, ce nombre de peptides doit être suffisamment restreint pour pouvoir réaliser la quantification d'autres protéines simultanément. Cette stratégie est cependant limitée par le coût et le temps de développement nécessaire à la préparation de tels standards notamment concernant les étapes de synthèse, de purification et de quantification de ces protéines. De plus, les protéines endogènes peuvent présenter des modifications post-traductionnelles qui risquent d'induire une différence de comportement entre la protéine standard et la protéine endogène [81].

A.3.3. Traitement des données et validation

Les courants d'ions sont extraits, puis intégrés, pour l'ensemble des peptides standard et endogène à l'aide de logiciel de traitement des données. Pour une quantification plus fiable, il est nécessaire d'évaluer manuellement la qualité de l'intégration des courants d'ions par le logiciel de traitement automatisé. De plus, il est nécessaire de vérifier la parfaite co-élution des peptides standards et endogènes. De même que les proportions relatives des transitions monitorées pour le peptide standard doivent être similaires à celles du peptide endogène. Les concentrations endogènes des protéines à quantifier sont ensuite calculées d'après les abondances relatives mesurées entre les peptides standards et endogènes. Les concentrations des peptides standard sont connues et permettent d'extrapoler les concentrations des peptides endogènes dans les échantillons.

Les premières règles de « bonnes pratiques de publication » pour la quantification ciblée, « guidelines » en anglais, commencent à émerger au sein de la communauté [92]. Ces règles pourraient être adaptées en fonction du contexte des études quantitatives allant du diagnostic clinique à la recherche exploratoire, pour laquelle les critères de validation pourraient être, respectivement, très ou peu élevés. Par exemple, la validation en milieu clinique nécessite l'utilisation d'un standard marqué pour l'ensemble des peptides quantifiés, ce qui pourrait ne pas être systématiquement requis pour une expérience exploratoire [92]. De plus, un coefficient de variation de 20 % entre deux répliques (biologiques) semble accepté comme valeur référence pour cette stratégie [92].

B. La stratégie PRM

Une stratégie parallèle à la stratégie SRM est apparue ces dernières années sur un autre type de spectromètre de masse hybride. La stratégie PRM (Parallel Reaction Monitoring) exploite les capacités de haute résolution et de haute précision en masse d'un système combinant un analyseur

de type Orbitrap™ et d'un analyseur quadripolaire. Cette stratégie quantitative est également basée sur le suivi spécifique d'une combinaison de peptides mais au lieu de suivre des transitions, c'est un spectre de fragmentation MS/MS qui est généré pour chaque peptide. L'information quantitative sera déduite de l'extraction des courants d'ions à partir des fragments extraits des spectres MS/MS. L'utilisation d'un analyseur de haute résolution, et précis en masse comme l'analyseur Orbitrap™ en fait une stratégie très spécifique qui lui vaut d'être comparé à la stratégie SRM, et ce, notamment en termes de sensibilité [96, 97]. En effet, il est possible d'accumuler les ions filtrés par le quadripôle en amont de la génération du spectre de fragmentation dans l'Orbitrap™ : cette possibilité permet d'abaisser les limites de détection et de quantification. En revanche, ce type d'analyseur permet de générer au maximum une dizaine de spectres MS/MS par seconde. Ainsi si cette technique est spécifique, elle ne permet pas d'égaliser le nombre de protéines qu'il est possible de quantifier simultanément avec les derniers spectromètres de masse triple quadripôle (environ 50 protéines) [98].

II. Les stratégies d'analyse quantitative globales

En général, ce type de stratégie est utilisé comme première approche quantitative globale au sein d'une thématique biologique et ne permet pas une quantification précise des protéines. Les protéines identifiées comme différentielles doivent ensuite faire l'objet d'une confirmation ciblée à l'aide de stratégie comme la SRM, le test ELISA, le Western-Blot, etc.

Quelques considérations sont communes à l'ensemble des stratégies décrites dans la suite du manuscrit : à l'exception de la technique 2D-DIGE et de l'approche par comptage de spectres, l'information quantitative est basée sur l'extraction de courants d'ions en MS (SILAC, ICAT, ¹⁸O, marquage par diméthylation, XIC, MS^e) ou MS/MS (iTRAQ®, TMT™, SWATH). L'extraction de courant d'ions consiste à reconstituer une trace au cours du temps correspondant à l'aire sous la courbe du rapport m/z à extraire dans l'ensemble des spectres MS ou MS/MS acquis au cours d'une analyse. L'identification des peptides (puis des protéines) se fera par analyse du spectre de fragmentation selon les stratégies décrites dans le chapitre II.

A. Les stratégies d'analyse quantitative basées sur le marquage des protéines ou des peptides

La diversité des méthodes quantitatives basées sur le marquage des protéines (ou des peptides) est très importante. Ces stratégies diffèrent selon l'étape de marquage des protéines, ou des peptides,

dans le processus de préparation des échantillons (culture cellulaire, protéines entières, peptides) et le type de signal quantifié (MS ou MS/MS). Certaines stratégies permettent, en plus de quantifier plusieurs milliers de protéines simultanément, de quantifier plusieurs échantillons au sein d'une unique analyse. Les différentes stratégies présentées dans la suite de ce manuscrit sont positionnées en figure III-5 selon leur étape de marquage.

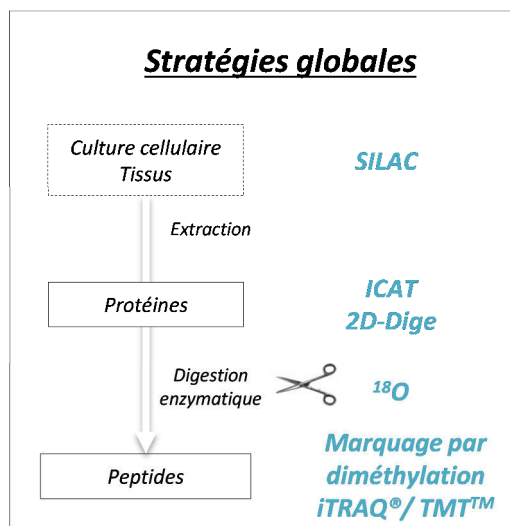


Figure III-5 Positionnement des différentes stratégies globales en fonction des étapes de marquage des protéines (ou des peptides)

A.1. Les stratégies par marquage métabolique

La stratégie SILAC (Stable Isotope Labeling by Amino acids in Cell culture) [99] est une stratégie de marquage basée sur l'incorporation métabolique d'isotopes stables à des protéines nouvellement synthétisées. Elle se destine donc, en général, aux échantillons de type culture cellulaire même si des exceptions, décrites dans la suite du manuscrit, ont été reportées. Les acides aminés marqués les plus couramment utilisés sont la lysine et l'arginine ($^{13}\text{C}/^{15}\text{N}$) de façon à ce que les peptides issus d'une digestion trypsique possèdent au moins un acide aminé marqué par peptide. En général, deux échantillons sont comparés : un « léger » non marqué et un autre au sein duquel ont été incorporés les acides aminés marqués arginine ($^{13}\text{C}_6\text{-}^{14}\text{N}_4$) et lysine ($^{13}\text{C}_6\text{-}^{14}\text{N}_2$). Jusque trois échantillons peuvent être comparés simultanément avec cette approche en incorporant les acides aminés marqués arginine ($^{13}\text{C}_6\text{-}^{15}\text{N}_4$) et lysine ($^{13}\text{C}_6\text{-}^{15}\text{N}_2$) à un troisième échantillon [80]. La force de cette technique réside dans la précocité de l'étape de marquage, permettant une quantification plus juste de l'ensemble des protéines d'un échantillon et la rendant ainsi moins sensible aux biais liés au cumul d'étapes préparatoires [99].

De nouvelles variantes de cette stratégie sont apparues ces dernières années avec le « Pulsed SILAC » [100] qui consiste à incorporer des acides aminés marqués dans un échantillon pendant une période définie, le ratio protéine marquée/protéine non marquée permet ensuite d'évaluer les cycles de renouvellement protéique (balance synthèse /dégradation) dans un type de cellule donnée dans des conditions données.

Cependant, la limitation majeure de ces approches SILAC réside dans le type d'échantillons restreint pour lesquels la stratégie peut être appliquée (cultures cellulaires). La possibilité d'utiliser un mammifère marqué (une souris) a été présentée en 2008 [101] mais ce type de protocole est très difficilement applicable chez l'Homme puisqu'elle suppose l'incorporation d'isotopes marqués et que ce type de recherche clinique doit faire l'objet d'une validation experte et fastidieuse par les agences de sécurité du médicament [102]. Le « super-SILAC » a néanmoins été développé dans le but d'étendre le champ d'application de cette stratégie à des échantillons de tissus humains ou provenant d'animaux. Le « super-SILAC mix » est un mélange de plusieurs types de cultures cellulaires marqués (incorporation d'acides aminés marqués sur le modèle SILAC) qui peuvent être comparés deux à deux avec des échantillons de tissus humains ou provenant d'animaux non marqués. Au travers de cette comparaison avec le standard marqué, il est ainsi possible de comparer plusieurs échantillons humains ou provenant d'animaux entre eux indirectement [103]. En revanche, les « super-SILAC mix », provenant de différents types de culture cellulaire, ne reflètent pas précisément la composition protéique des échantillons à analyser, ce qui rend cette stratégie assez imprécise [80].

A.2. Les stratégies par marquage enzymatique

La stratégie par marquage enzymatique consiste à introduire un isotope lourd (^{18}O) au sein du groupe carboxylique C-ter des peptides grâce à la réalisation de la digestion enzymatique dans un solvant enrichi (H_2^{18}O) [104]. Cette stratégie permet de comparer deux échantillons simultanément : l'un marqué, l'autre non. La différence de masse induite est de 2 Da par ^{18}O incorporé. Il est donc préférable d'utiliser une enzyme comme la trypsine ou la lysC qui permet d'incorporer 2 atomes d'oxygène, ce qui facilitera la différenciation des peptides marqués et non marqués sur le spectre de masse [10]. Cette stratégie permet de réduire la variabilité due au traitement des échantillons en combinant l'étape de marquage et l'étape de digestion. En revanche, les degrés d'incorporation variables du ^{18}O ont limité l'expansion de cette stratégie même si différents protocoles de digestion ont été développés pour palier à ces limites [104].

A.3. Les stratégies par marquage chimique des protéines

A.3.1. La quantification par gel d'électrophorèse

Les techniques basées sur l'utilisation des gels d'électrophorèse ont constitué les premières stratégies quantitatives dans le domaine de la protéomique [31]. Le principe des gels d'électrophorèse bidimensionnelle est présenté en chapitre I page 15. L'information quantitative sera déduite d'une analyse par densitométrie des spots entre les différents échantillons après révélation des spots protéiques. L'identification des protéines contenues dans les spots possédant une intensité différentielle (ou non) sera réalisée par spectrométrie de masse.

Cette stratégie est basée sur la quantification des protéines entières. La forte résolution obtenue sur les gels d'électrophorèse bidimensionnelle permet la différenciation de formes modifiées telles que les glycosylations ou phosphorylations.

A.3.1.1. Méthodologie de mise au point d'une stratégie 2D-DIGE

Les différentes étapes inhérentes à la mise au point de cette stratégie de la mise au point à l'analyse des échantillons sont présentées en figure III-6.

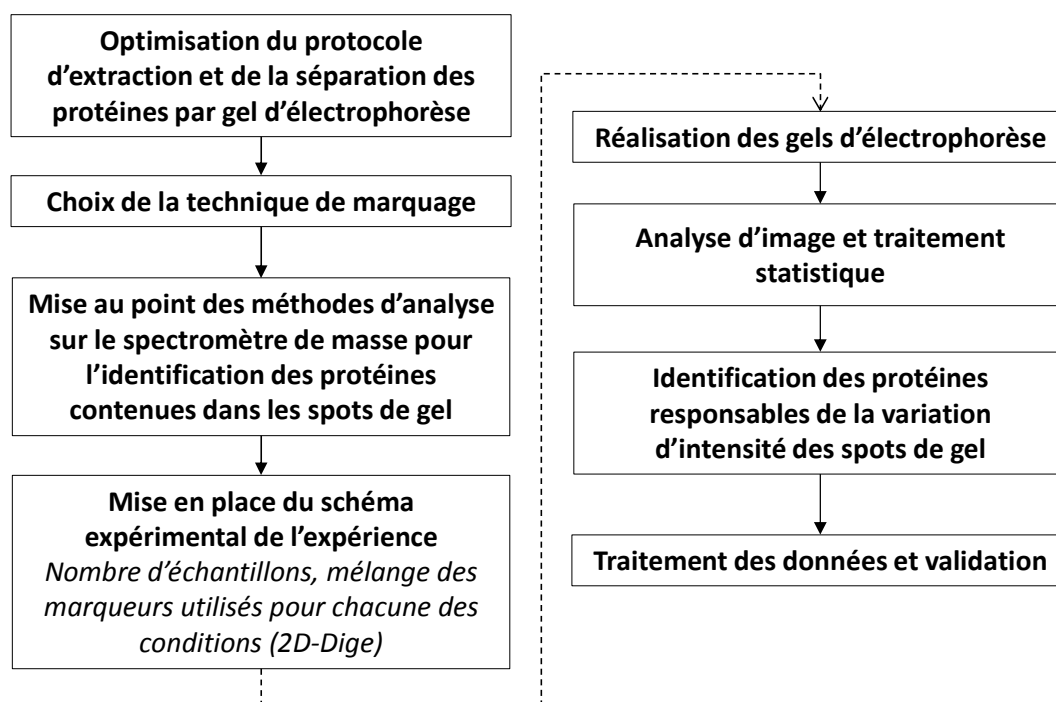


Figure III-6 Etapes inhérentes à la quantification de protéines sur la base de gels d'électrophorèse bidimensionnelle

Les étapes de choix de la technique de marquage, d'analyse d'image et d'identification des protéines responsables de la variation d'intensité des spots différentiels sont décrites dans la suite du chapitre.

A.3.1.1.1. Les techniques de marquage

Il existe différents types de marquages chimiques utilisés pour la révélation des spots protéiques. Ils peuvent être réalisés en amont ou en aval de la décomplexification par gel.

Marquage réalisé avant l'étape d'électrophorèse bidimensionnelle :

La stratégie quantitative basée sur l'utilisation de cyanines (CyDyes, Amersham Biosciences) s'intitule « 2D-DIGE » pour « Difference In-Gel Electrophoresis » [105]. Il existe deux générations de CyDyes. Les premières générations de CyDyes sont pourvues d'un groupement N-hydroxysuccinimide permettant le marquage des amines primaires des protéines (ϵ -amines des lysines ou extrémités N-ter). Dans cette stratégie intitulée « marquage minimal », le ratio fluorochrome/protéine est maintenu très bas, ainsi, moins de 3 % des amines primaires des protéines d'un échantillon sont marquées. Cette stratégie présente une linéarité de réponse supérieure à trois ordres de grandeur et présente une forte sensibilité ($\approx 0,1$ ng) [34].

La seconde génération de CyDyes comprend un groupement maléimide permettant le marquage (à saturation) des groupements thiols des cystéines. Cette stratégie, présentant une très forte sensibilité ($\approx 0,02$ ng), est destinée aux échantillons disponibles en très faible quantité [34].

Trois types de cyanines (Cy2, Cy3 et Cy5) sont disponibles pour chacune des générations de CyDyes (marquage minimal ou à saturation). Elles diffèrent par leur longueur d'onde d'absorption et d'émission. Deux de ces cyanines sont dédiées au marquage des échantillons qui peuvent ainsi être mélangés sur chaque gel, la troisième est utilisée pour le marquage d'un standard interne (c'est-à-dire un mélange de tous les échantillons), également déposé sur chacun des gels, qui permettra de réaliser une normalisation inter-gel [106].

Marquage réalisé après l'étape d'électrophorèse bidimensionnelle :

- La coloration au bleu de Coumassie est une technique de révélation peu onéreuse mais qui présente une linéarité et une sensibilité faible (1 ordre de grandeur, environ 10 ng) [107].

- Les espèces fluorochromes « Deep-purple » et « Sypro®Ruby » sont des agents fluorochromes permettant de révéler les protéines jusque l'ordre du nanogramme. Ces marqueurs présentent une bonne linéarité de réponse (jusque 3-4 ordres de grandeur) [107].
- Les espèces fluorophores spécifiques tels que Pro-Q Diamond® et Pro-Q Emerald® (« Molecular probes ») permettent la révélation spécifique de formes respectivement phosphorylées ou glycosylées [107].

A.3.1.1.2. Analyse d'image

Il existe de nombreux logiciels permettant de traiter cette étape (*Decyder*, *Decodon*; *PD-Quest*, *Bio-Rad*; *Samespots Nonlinear Dynamics*). Les logiciels de traitement d'analyse d'image permettent de réaligner les images de gel entre elles puis de déterminer l'abondance relative des différentes intensités détectées pour chaque spot protéique entre les différentes conditions à comparer. Il est nécessaire de vérifier manuellement le réaligement des images de gel de façon à éviter la comparaison erronée de spots protéiques différents entre les gels.

A.3.1.1.3. Identification des protéines responsables de la variation d'intensité des spots différentiels

Les protéines contenues dans les spots de gel sont identifiées grâce à la réalisation d'analyses nanoLC-MS/MS acquises en mode DDA à partir de la digestion enzymatique in-gel des protéines contenues dans les spots à analyser. Les identifications sont réalisées d'après les procédures présentées dans le chapitre II « stratégies d'analyse protéomique qualitatives » en page 37. Plusieurs algorithmes de recherche peuvent être utilisés dans le but de renforcer la confiance dans les identifications réalisées comme décrit dans le chapitre II. La concordance entre les valeurs expérimentales et théoriques des masses moléculaires et points isoélectriques des protéines identifiées doivent également être vérifiées.

Concernant la non-ambiguïté des quantifications, bien que les gels 2D offrent une importante résolution, de multiples protéines peuvent être détectées et identifiées sans ambiguïté au sein d'un même spot protéique [108]. Plusieurs types de stratégies ont été décrits dans le but de comparer les abondances des protéines au sein d'un même spot. Elles sont basées sur le marquage isotopique des protéines [109, 110] ou sur la combinaison du comptage de spectres et de l'extraction des courants d'ions [111]. Il est également possible de réaliser une analyse quantitative « grossière » des

différentes protéines présentes dans les spots de gel 2D sur la base du nombre de séquences peptidiques qui leur sont assignées par spot [111]. Le nombre de peptides identifiés par protéine n'est pas normalisé par la taille de chacune des protéines puisque les protéines ayant migré au sein d'un même spot sont de même masse moléculaire.

A.3.1.2. Avantages et limites de la stratégie par gel

Ces stratégies présentent l'intérêt d'apporter une forte décomplexification des échantillons (jusque quelques protéines par spot) qui peut être particulièrement d'intérêt pour l'étude d'organismes non séquencés [75]. En effet, la forte couverture de séquence obtenue après analyse par spectrométrie de masse des protéines issues des spots de gel permet d'envisager la réalisation d'identifications par stratégie *de novo*). Ce type de stratégie est plus facilement conciliable avec des stratégies quantitatives où le recouvrement de séquence protéique est plus élevé (contrairement aux stratégies sans marquage présentées dans la suite du chapitre par exemple).

La stratégie 2D-DIGE, la plus puissante de ces techniques basée sur l'utilisation de gel d'électrophorèse bidimensionnelle grâce à l'analyse simultanée de plusieurs échantillons, a été un peu délaissée au profit d'autres stratégies d'analyses quantitatives en spectrométrie de masse en raison de plusieurs facteurs : le manque d'automatisation requérant un haut savoir-faire technique et l'impossibilité d'évaluer les niveaux d'expressions protéiques de certaines protéines à savoir de masses moléculaires faibles ou très élevées, très hydrophiles, très hydrophobes ou possédant un point isoélectrique extrême [31]. Pour pallier à ces difficultés, de nouveaux systèmes d'électrophorèse horizontaux où les manipulations de l'opérateur sont simplifiées, notamment grâce à l'utilisation de gels précoulés, ont été développés (HPE Tower, Serva). Ce système se compose de quatre chambres d'électrophorèse horizontales qui permettent de réaliser 4 gels 2D simultanément. Dans ce système, les gels sont isolés électriquement, de grande taille (strips de 24 cm), et placés horizontalement. L'optimisation de la conduction thermique de ce système et la faible épaisseur des gels précoulés utilisés (commerciaux donc produisant plus de reproductibilité) favorisant un meilleur refroidissement, permettent de travailler à des potentiels élevés qui permettent d'accélérer les temps de migration sans perte de résolution. Un autre avantage réside dans l'obtention de migrations plus reproductibles ce qui facilite le traitement des images de gels par la suite. Les gels précoulés commerciaux utilisés sont plus onéreux, mais ils présentent l'avantage de pouvoir être manipulés beaucoup plus facilement étant donné leur rigidité (ils sont fixés sur un film en plastique). La quantité de matériel pouvant être chargée sur les gels précoulés est limitée à environ 300 µg. En

revanche, la présence du film plastique au dos de ces gels ne permet pas l'utilisation de robots pour l'excision des spots protéiques et cette étape redevient manuelle, donc chronophage et fastidieuse.

Ces nouveaux systèmes ont été utilisés pour deux projets présentés dans le chapitre VII de ce manuscrit.

A.3.2. La stratégie ICAT (Isotope-Coded Affinity Tag)

La stratégie ICAT [112] est une stratégie quantitative basée sur le marquage des cystéines au stade protéique. Le réactif ICAT (Isotope Coded Affinity Tag) est composé d'un groupement réactif avec la fonction thiol (souvent N-hydroxysuccinimide), d'un « bras espaceur » ainsi que d'un groupement « biotine ». Cette stratégie permet de comparer deux conditions car il existe deux types de « bras espaceurs » marqués isotopiquement (^2H) ou non [112]. Après digestion enzymatique, les peptides portant un marqueur sont ensuite isolés par affinité grâce à la présence du groupement biotine (sur une colonne greffée avec de la streptavidine par exemple) puis injectés sur un système LC-MS [112].

Les peptides marqués au deutérium (^2H) peuvent avoir un comportement chromatographique différent de leur homologue endogène sur une colonne C18 [113]. L'utilisation de ce type d'isotopes risque d'introduire un biais dans la quantification réalisée. En effet, si l'environnement du peptide en source est différent du peptide de référence, alors il est possible que leur capacité d'ionisation diffère et que leurs intensités respectives ne se corrélient pas. Pour palier à cette difficulté, de nouveaux marqueurs ICAT ont été développés utilisant préférentiellement des marqueurs ^{13}C [113]. La faible occurrence naturelle des cystéines (environ 8% des protéines n'en contiennent pas) [10] est la principale limite de cette technique.

A.3.3. Les stratégies TMTTM (Tandem Mass Tag, Thermo Fisher Scientific) et iTRAQ[®] (isobaric Tag for Relative and Absolute Quantitation)

La stratégie iTRAQ[®] [114] est une stratégie quantitative basée sur le marquage des peptides après digestion enzymatique. Les marqueurs iTRAQ[®] sont constitués d'une combinaison de deux molécules (un ion « rapporteur » et un ion « balance »), le tout isobare : ces ions « rapporteur » et « balance » sont marqués isotopiquement (^{13}C , ^{15}N , ^{18}O) de façon à ce que la masse totale du marqueur soit toujours égale à 145 Daltons et qu'ils ne puissent être différenciés qu'après fragmentation. Les peptides marqués avec des standards iTRAQ[®] différents, présentent les mêmes propriétés physicochimiques et seront parfaitement co-élus puis co-fragmentés. L'information quantitative

sera extraite à partir des intensités des ions rapporteurs au sein des spectres MS/MS. Aujourd'hui, il existe jusque 8 marqueurs iTRAQ[®] avec des masses d'ions rapporteurs allant de 113 à 121 Daltons.

Un autre type de marqueurs basés sur cette technologie de peptides isobares a été développé : TMT[™] [115]. Cette technologie est très similaire à la stratégie iTRAQ[®]. Elle propose aujourd'hui jusque 10 marqueurs différents avec des ions rapporteurs allant de 126 à 131 Daltons.

Ces deux stratégies présentent l'avantage de ne pas complexifier les chromatogrammes et les spectres MS étant donné la parfaite co-élution des peptides marqués isobares. Sa principale limite réside dans la gamme de masse des ions rapporteurs qui est sensible aux interférences et dans l'impossibilité d'utiliser des analyseurs de type piège à ions qui ne permettent pas l'analyse d'ions de très bas m/z [80].

Ces dernières années, de nouvelles stratégies sont apparues combinant les avantages des stratégies iTRAQ[®] (et TMT[™]) et SILAC, NeuCode –*Neutron enCoding*, cumulant jusque 18 marqueurs différents [116]. Ces nouvelles stratégies nécessitent des appareils hautement résolutifs tels que l'Orbitrap[™] Fusion[™] Tribrid[™], Thermo Scientific (résolution de 480 000 pour un m/z de 200) pour permettre de différencier toutes les formes de peptides marqués.

A.3.4. Marquage par méthylation « Dimethyl labeling »

La stratégie d'incorporation d'isotopes stables (deutérium) par méthylation des amines primaires (N-ter et ϵ -amines des lysines) existe depuis plus de 30 ans mais son application quantitative en protéomique n'est apparue qu'en 2003 [117]. Cette stratégie peu onéreuse vise à marquer les protéines après digestion enzymatique par méthylation des amines primaires en utilisant des réactifs marqués isotopiquement. La stratégie originale permettait de comparer deux échantillons, un échantillon marqué ^2H et un non marqué. Un troisième échantillon à comparer a pu être introduit avec l'utilisation d'un second marqueur composé des isotopes ^{13}C et ^2H . Les limites de cette technique résident notamment dans la séparation chromatographique des peptides marqués et endogènes [118].

B. Les approches sans marquage dites « label free »

Depuis quelques années, les méthodes sans marquage se sont beaucoup développées [80]. Moins onéreuses que les stratégies avec marquage, elles ne sont pas limitées, en théorie, par le nombre de comparaisons possibles à effectuer. De plus, elles sont faciles à mettre en œuvre et sont applicables à tous type d'échantillon contrairement aux stratégies SILAC par exemple.

B.1. Spectral counting

Les méthodes dites « spectral counting » sont basées sur l'hypothèse que, en mode DDA, plus une protéine est abondante, plus la probabilité d'acquérir un spectre MS/MS pour cette protéine est importante, le tout à pondérer par la masse de la protéine. Cette stratégie est donc basée sur le comptage des spectres MS/MS générés par protéine et par analyse [119].

Afin de faire ressortir des différences significatives entre les conditions à comparer, il est nécessaire d'avoir suffisamment de spectres pour chaque protéine à quantifier [80]. Dans une analyse DDA classique, non destinée à l'application de ce type de stratégie quantitative, des temps d'exclusion dynamique sont paramétrés de telle sorte qu'un peptide ne doit être sélectionné qu'une ou deux fois dans le but de pouvoir générer des spectres MS/MS pour le plus grand nombre de protéines possibles. Dans ce contexte quantitatif de stratégie par comptage de spectres, il est nécessaire de réduire ces temps d'exclusion dynamique de façon à sélectionner les peptides un nombre de fois suffisamment élevé pour pouvoir observer des différences significatives. Le nombre de protéines quantifiées s'en trouve donc immédiatement diminué. Cette stratégie ne peut donc être conçue que sur des spectromètres assez rapides pour pouvoir générer beaucoup d'identifications par analyse, et beaucoup de redondances pour chacune d'elle. Ce dernier élément est à pondérer par la complexité de l'échantillon.

Le traitement des données liées à cette méthode est en revanche beaucoup plus simple et plus court à réaliser que pour d'autres stratégies quantitatives basées sur les extractions de courant d'ions [120].

B.2. Extraction des courants d'ions

Ces stratégies consistent à intégrer l'aire sous la courbe (AUC) de courants d'ions extraits de peptides pour un rapport m/z et un temps de rétention donnés sans utiliser de références marquées

contrairement aux stratégies présentées précédemment. Ces stratégies requièrent l'utilisation de spectromètres de masse très résolutifs et précis en masse afin de ne pas risquer d'erreur d'attribution des pics à intégrer [121]. Elles ont pris de l'ampleur grâce au développement des Orbitrap™ (Thermo Fisher Scientific) et des Q-TOFs dernière génération, très résolutifs, très rapides (jusqu'à 100 Hz pour le dernier TTOF 6600 d'ABSCIEX) et toujours plus précis en masse (< 1 ppm en calibration interne), qui promettent à ces stratégies sans marquage un avenir prometteur. Le développement de logiciels de traitement spécifiques a également participé au regain d'intérêt pour ces techniques (Skyline [122], MaxquantLFQ [123]).

B.2.1. Extraction des courants d'ions MS (XIC, eXtracted Ion Chromatogram) suite à la génération de données spectrales en mode « Data Dependant Acquisition » (DDA)

La stratégie quantitative XIC est basée sur l'extraction de courants d'ions MS dans une analyse générée en mode DDA. La comparaison des aires sous les courbes des ions P, P+1 et P+2 entre toutes les analyses permet d'apporter une information quantitative d'un peptide donné entre différents échantillons [121].

B.2.1.1. Solutions logicielles pour l'extraction des courants d'ions

Deux types d'approches permettent de traiter les données quantitatives basées sur l'extraction des courants d'ions [124] :

- Approche 1 : Extraction des courants de tous les ions détectés

Dans un premier temps, la détection et l'intégration des aires sous la courbe des courants induits par tous les ions se fait seulement à partir des spectres MS. Ainsi, tous les composés détectés se retrouvent caractérisés par une valeur de m/z, par une valeur d'intensité et par un temps de rétention. Une information quantitative sera donc extraite pour chaque composé détecté. L'identification des composés détectés et quantifiés, se fera dans un second temps d'après l'interprétation des spectres MS/MS selon les stratégies présentées au chapitre I. Ce type d'approche peut être mis en place à l'aide de différents logiciels, tels Progenesis LC-MS (*Nonlinear Dynamics*) et MaxQuantLFQ [123].

- Approche 2 : Extraction des courants d'ions des peptides dont l'assignation a été validée

La sélection des courants d'ions à extraire est réalisée à partir des identifications MS/MS (en général d'après une librairie de spectres filtrée <1% FDR). Seuls les courants d'ions des peptides dont l'identification aura été validée seront extraits. Avant intégration des aires sous la courbe des courants d'ions, ceux-ci sont repérés sur la base, entre autre, du temps de rétention enregistré au moment de l'acquisition de chaque spectre MS/MS. Le logiciel Skyline fonctionne notamment avec ce type d'approche [122].

Le traitement des données XIC avec l'utilisation du logiciel Skyline est présenté en figure III-7.

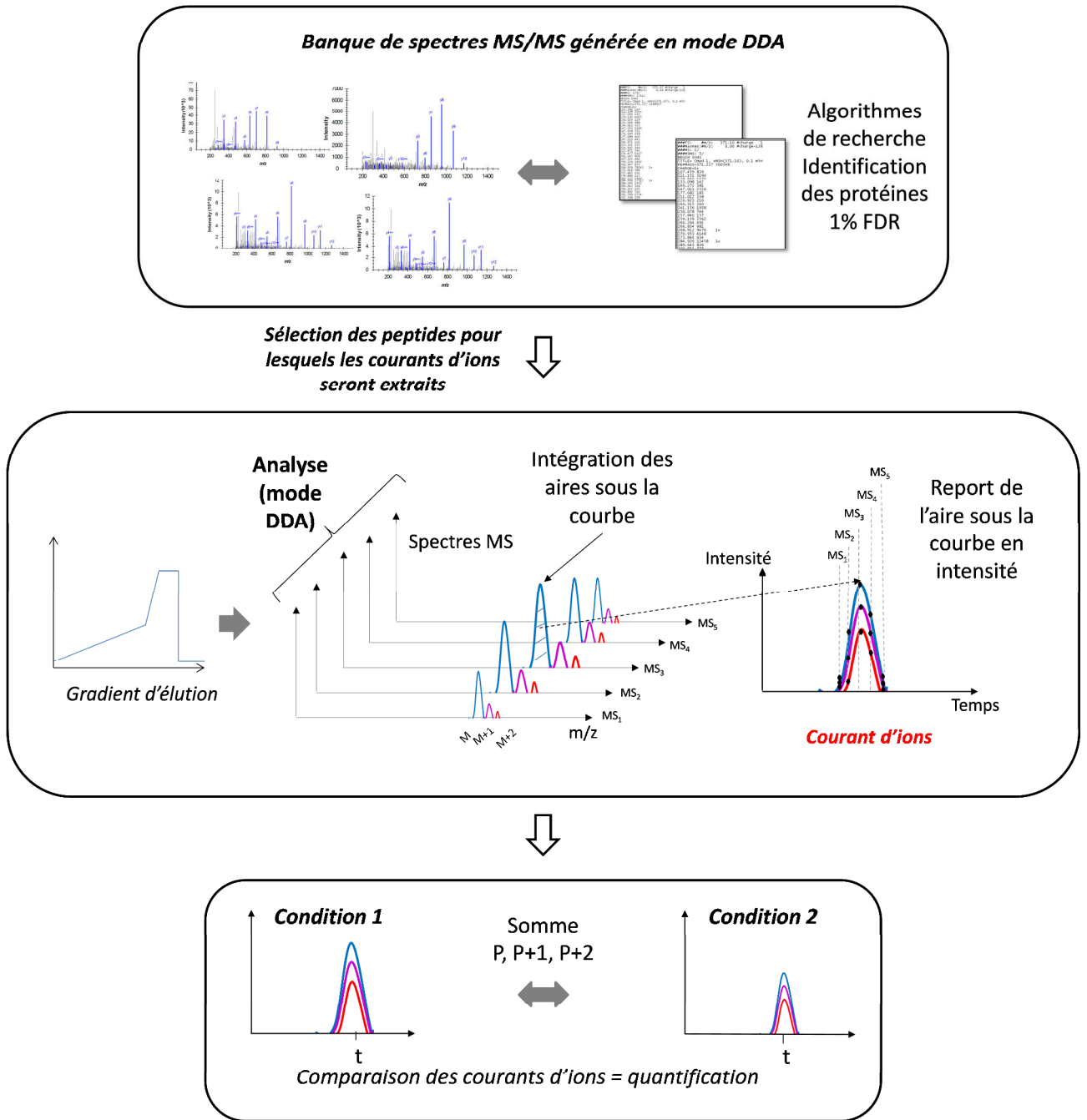


Figure III-7 Principe de la stratégie d'analyse quantitative XIC basée sur l'utilisation du logiciel Skyline

Une librairie de spectres est tout d'abord réalisée à partir des spectres de fragmentation générés pendant l'analyse acquise en mode DDA. Seuls les spectres dont l'identification est de confiance élevée seront utilisés. Le logiciel Skyline extrait ensuite les courants d'ions (P, P+1 et P+2) de l'ensemble des peptides sélectionnés. La comparaison des courants d'ions entre les différentes conditions permet d'établir un protéome différentiel.

B.2.1.2. Méthodologie de mise au point d'une stratégie XIC

Les différentes étapes de développement d'une stratégie quantitative XIC, de la mise au point au traitement des données en passant par l'analyse des échantillons, sont illustrées en figure III-8.

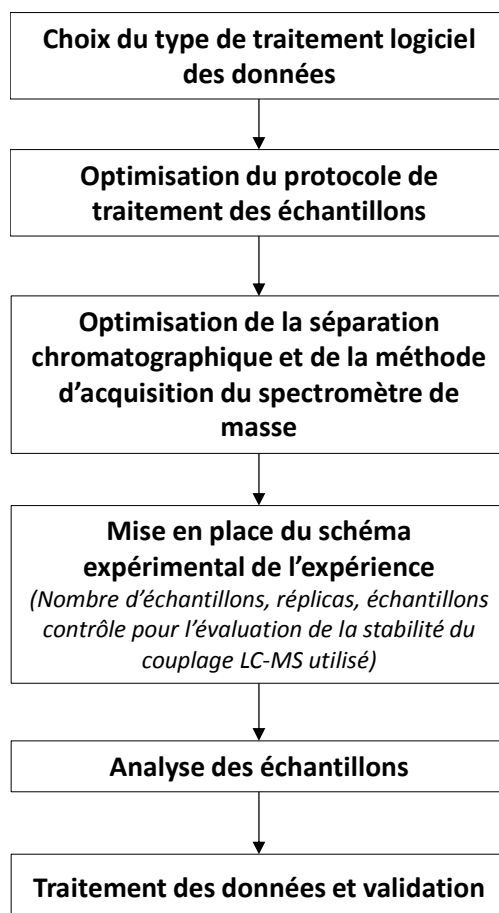


Figure III-8 Etapes de développement d'une stratégie XIC

La partie traitement des données est un des « nœuds » de ce type d'analyse quantitative puisque cette partie consiste à traiter les données spectrales de milliers de peptides de façon à aboutir à une information quantitative unique par protéine. Cette partie semble aujourd'hui être la problématique majeure de ce type de quantification et est décrite en détail dans la suite de ce chapitre. Le traitement des données présenté est adapté pour l'utilisation du logiciel Skyline.

Le logiciel SKYline ne permet l'extraction de courants d'ions que de peptides pour lesquels une assignation a été validée (FDR < 1 %) dans les analyses nanoLC-MS/MS réalisées. Ces peptides vont être regroupés sous la forme d'une librairie de spectre. Pour la quantification, il est préférable de ne sélectionner que les peptides uniques aux protéines à quantifier [19]. Les peptides modifiés, présents

donc sous de multiples formes dans les échantillons, risquent de présenter des courants d'ions dont les intensités seront moins reproductibles et peuvent également être éliminés des données.

Les extractions de courant d'ions de tous les peptides représentés dans la librairie de spectres sont réalisées dans une fenêtre de temps restreinte autour du temps de rétention auquel les spectres MS/MS de chaque peptide présents dans la librairie de spectres ont été générés. Cette stratégie permet de réduire des possibles erreurs d'intégration des courants d'ions et de raccourcir le temps nécessaire pour l'extraction des courants d'ions de l'ensemble des peptides dans un ensemble d'analyses. En général, ces fenêtres sont réduites de l'ordre de 3 à 5 minutes autour du temps de rétention du peptide présent dans la librairie de spectres. **Les aires sous la courbe des courants d'ions sont ensuite intégrées** de façon automatisée par le logiciel de traitement. La validation manuelle de ces données permet de vérifier la justesse des intégrations réalisées par le logiciel (et éventuellement de les corriger). Cette étape, chronophage puisqu'elle consiste en la vérification manuelle des intégrations de courants d'ions réalisées pour plusieurs milliers de peptides, permet de garantir davantage de fiabilité. La valeur d' « Idot product » (produit scalaire) calculée par le logiciel Skyline permet d'aider à la réalisation de cette étape de validation. Il s'agit d'une valeur numérique comprise entre 0 et 1. Cette valeur caractérise la proportion relative des aires P, P+1 et P+2 par rapport à la distribution théorique attendue pour un peptide donné (décrit dans le chapitre I en page 21). Plus elle est proche de 1, plus le profil des aires P, P+1, P+2 est proche de la distribution isotopique théorique du peptide à quantifier. Les erreurs d'intégration ne sont pas spécifiques au logiciel Skyline et la plupart des logiciels d'extraction des courants d'ions proposent des outils de visualisation des intégrations réalisées [123].

La transformation logarithmique des données (aires sous la courbe de tous les peptides dont les courants d'ions ont été extraits dans l'ensemble des analyse) permet de distribuer ces valeurs selon une loi normale, ce qui est nécessaire pour l'application de certains tests statistiques [125]. En général, cette étape est réalisée au stade de la normalisation des données, de ce fait, la plupart des outils de normalisation l'incluent automatiquement. De plus, cette transformation permet de donner un poids équivalent à l'ensemble des peptides quantifiés pour chaque protéine quel que soit l'intensité de leur courant d'ion respectif.

La normalisation des données entre les analyses permet d'éliminer tout biais technique introduisant des différences mineures d'une analyse à l'autre. Ces biais peuvent survenir au moment de la préparation d'échantillon et/ou être dus à une variabilité éventuelle des performances du spectromètre de masse.

Un graphique (« MA Plot », « minus versus average ») permet d'évaluer l'importance des biais techniques et permet d'aider au choix du modèle de normalisation [125, 126]. Ce type de graphique est présenté en figure III-9. Il représente en ordonnée le rapport des aires sous la courbe des courants d'ions extraits pour un peptide donné entre deux analyses et en abscisse la moyenne de ces aires. L'hypothèse de base est que la plupart des peptides ne voient pas leur abondance modifiée entre les différentes conditions à comparer, ainsi l'ensemble des points devrait se trouver sur une droite d'équation $y = ax + b$ avec a qui tend vers « 0 » et b qui tend vers « 1 » en cas d'absence de biais technique. Ces graphiques peuvent être générés manuellement mais des outils comme Normalizer [127] permettent de les visualiser de façon automatisée.

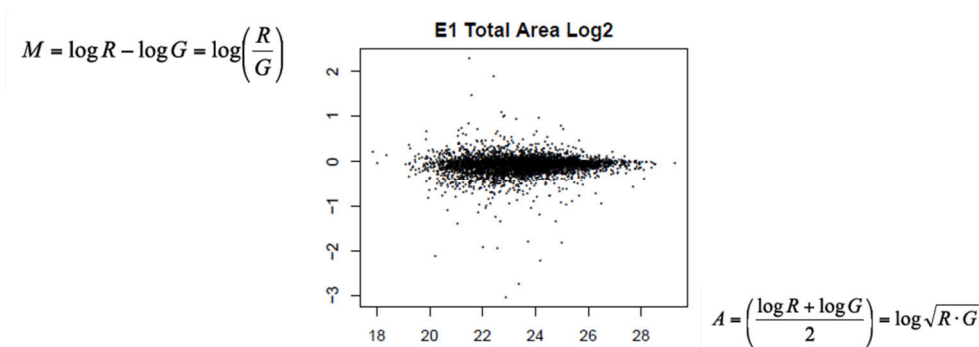


Figure III-9 Graphique schématisé permettant d'évaluer le besoin de normaliser les données

Les graphiques de type « MA Plot » représentent en ordonnée la valeur « M », c'est-à-dire l'abondance relative du peptide entre les deux analyses à comparer avec R et G les aires sous la courbe des courants d'ions extraits de ce peptide dans ces deux analyses et en abscisse la valeur « A » c'est-à-dire la moyenne des aires sous la courbe des courants d'ions extraits du peptide dans les deux analyses à comparer.

Il existe différents modèles de normalisation. Nombre de ces modèles ont été développés pour la génomique puis ont été appliqués aux données de protéomique [126]. Les deux modèles couramment utilisés pour la normalisation de ce type de données sont présentés dans la suite du chapitre.

La normalisation globale consiste à appliquer un facteur correcteur constant généralement calculé sur la base des médianes des intensités de l'ensemble des peptides au sein des analyses à comparer. A l'issue de la normalisation, les médianes des intensités des différents groupes doivent être égales.

La normalisation par quantile consiste à appliquer un modèle non paramétrique qui consiste à normaliser des données selon leur niveau d'intensité. Les différentes intensités (ou aires) acquises pour l'ensemble des peptides de chacune des analyses vont être nivelées : les intensités pour lesquelles le nombre de peptides est le maximum dans chacune des analyses vont être réajustées pour se correspondre et ainsi de suite pour tous les niveaux d'intensité. Une représentation

graphique de l'effet de la normalisation par quantile sur une population est représentée en figure III-10. Les différentes intensités relevées pour l'ensemble des données sont présentées en abscisse tandis que les effectifs correspondants (le pourcentage de peptides qui présentent ces intensités) sont présentés en ordonnée.

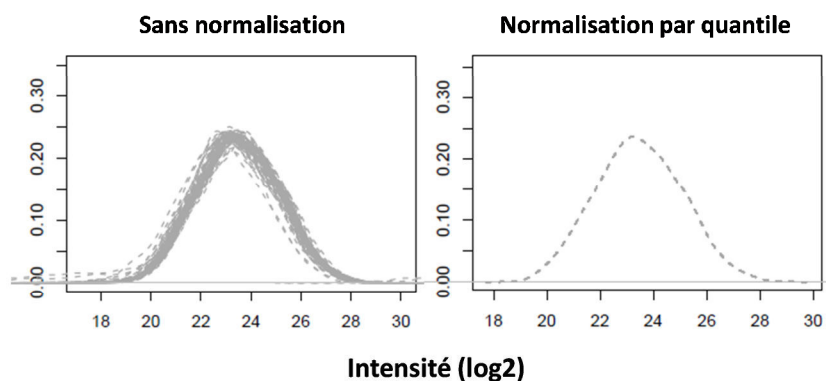


Figure III-10 Représentation de l'application du modèle de normalisation par quantile

Ces figures ont été réalisées grâce au logiciel Normalyzer. Elles présentent en abscisse les intensités à normaliser et en ordonnée, les effectifs correspondant à chacune des intensités.

A l'issue de l'étape de normalisation, il est possible de vérifier la correcte réalisation de cette étape : les boîtes à moustache représentant les intensités des analyses normalisées pour tous les peptides par ce modèle doivent être parfaitement identiques.

Différentes stratégies de traitement des données ont été développées **afin d'approximer des valeurs quantitatives protéiques à partir des valeurs peptidiques quantitatives**. Les aires sous les courbes des courants d'ions extraits pour chaque peptide peuvent être sommées, ou moyennées, par protéine. Il est également possible d'utiliser la valeur médiane des aires de l'ensemble des peptides quantifiés de façon à aboutir à une valeur quantitative unique par protéine [128-130]. La médiane ou les moyennes (éventuellement pondérées) sont souvent utilisées car elles sont moins sensibles aux valeurs aberrantes ou extrêmes (« outliers ») [125].

B.2.1.3. Avantages et limites de la stratégie

En théorie, ces stratégies permettent d'analyser un nombre illimité d'échantillons et sont très faciles à mettre en œuvre. En revanche, puisqu'aucun marqueur n'est utilisé, tous les échantillons sont analysés les uns à la suite des autres. Les temps d'analyses globaux pour une cohorte d'échantillons peuvent donc être très longs (plusieurs semaines consécutives). Ces stratégies nécessitent donc impérativement l'utilisation d'un système chromatographique reproductible et robuste de façon à pouvoir faciliter la comparaison des aires intégrées sur la base des temps de rétention des peptides

entre les différentes conditions à évaluer. De plus, l'utilisation de spectromètres de masse de très haute résolution et précis en masse est indispensable.

B.2.2. Extraction des courants d'ions suite à la génération de données spectrales en mode « Data Independent Acquisition » (DIA)

Le dernier type de stratégie sans marquage présenté dans ce manuscrit est le mode d'acquisition « Data Independent Acquisition ». Comme décrit en chapitre I « Principe général de l'analyse protéomique par spectrométrie de masse » en page 34, le mode d'acquisition DIA vise à fragmenter l'ensemble des peptides élués au cours d'une analyse, soit en se basant sur l'utilisation d'une succession de fenêtres d'isolement de façon à couvrir l'ensemble de la gamme de masse de façon cyclique au cours d'une analyse, soit en fragmentant simultanément l'ensemble des peptides élués à chaque instant (la gamme de masses entière).

B.2.2.1. Modèle du SWATH™ (AB SCIEX Technologies)

Le principe de cette stratégie a été présenté pour la première fois en 2004 [131]. Plusieurs évolutions de ce type d'acquisition sont apparues ces dernières années. La plus récente, basée sur le modèle développé par Gillet et al. [59], est nommée SWATH™ (AB SCIEX Technologies). L'évolution de cette technique au cours des dernières années réside notamment dans l'évolution du traitement des données.

Cette stratégie vise à isoler, par fenêtre de 25 Da, puis fragmenter, toute la gamme de masse à analyser. Ainsi, toute la gamme de masse est balayée par des cycles de quelques secondes. Un schéma récapitulant le type de balayage effectué sur l'ensemble de la gamme à analyser est présenté en figure III-11.

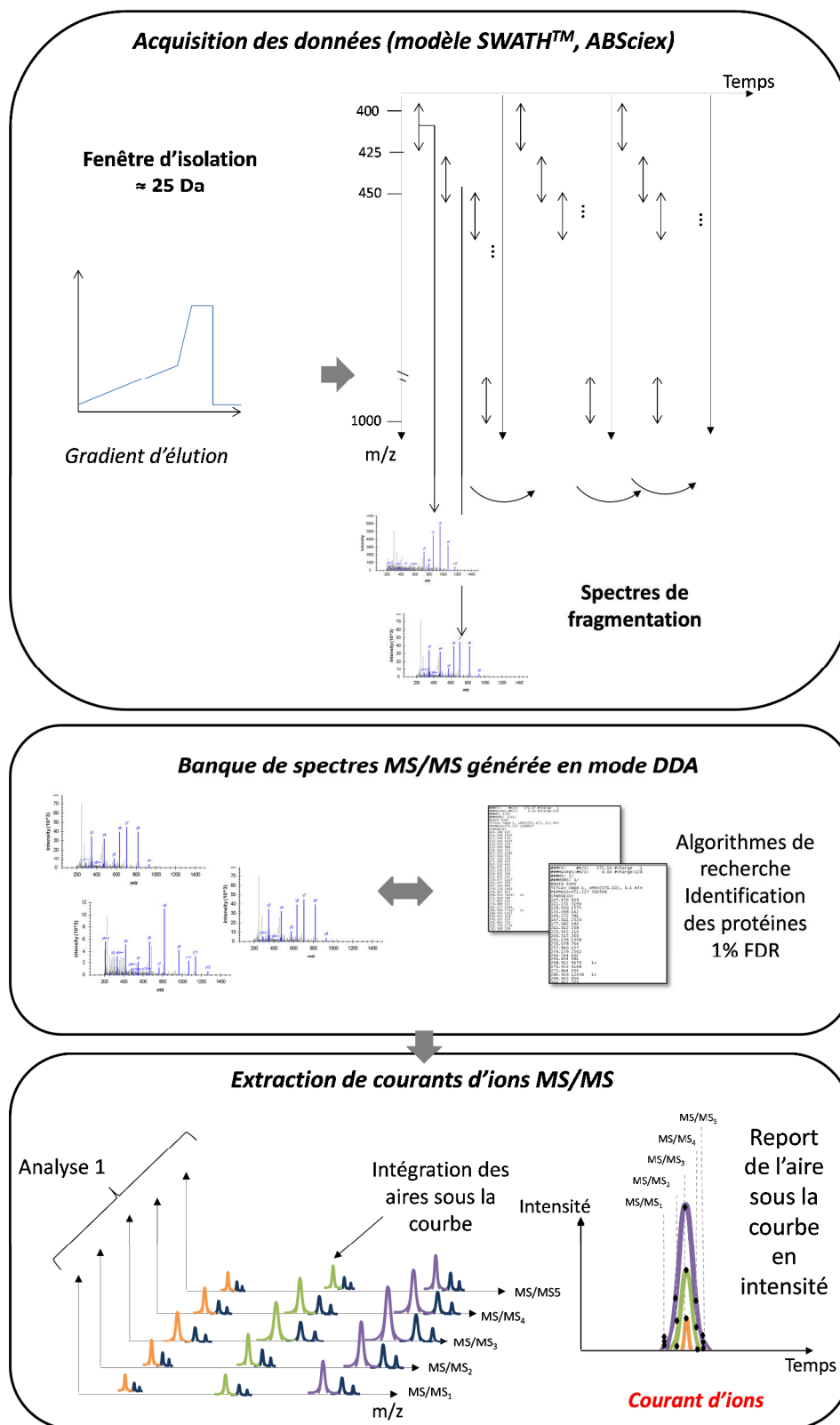


Figure III-11 Principe de la stratégie SWATH™

Les spectres MS/MS permettent l'extraction des courants d'ions. Il n'existe pas d'algorithme de recherche pour ce type de données. Une librairie spectrale générée en mode DDA est requise pour sélectionner les ions pour lesquels extraire un courant d'ions.

La quantification est ensuite basée sur l'intégration de l'aire sous la courbe des courants d'ions fragments extraits des spectres générés de façon cyclique (pseudo MRM). Il est nécessaire d'avoir des cycles totaux assez courts (inférieurs à quelques secondes) de façon à avoir un minimum 8 points par pic à intégrer. Il est donc nécessaire de trouver un compromis correct entre la largeur des fenêtres d'isolement, la gamme de masse à analyser et la vitesse d'acquisition utilisée pour générer les spectres, ce qui est souvent réalisé au détriment de la gamme de masse totale analysée (en général réduite de 400 à 1000 m/z).

L'isolement par fenêtre de 25 m/z confère à cette technique une spécificité élevée par rapport à la stratégie de quantification XIC du fait de la réduction des interférences sur les courants d'ions extraits. De plus, de 6 à 9 fragments (extraits sous la forme de courants d'ions) peuvent ainsi servir à la quantification d'un seul peptide ou au moins à définir correctement quel est le pic à intégrer dans le logiciel Skyline qui permet également de traiter ce type de données. Il n'existe pas d'algorithme d'identification des peptides sur la base des fragments générés avec cette méthode. Elle nécessite donc la réalisation d'une librairie de spectres en mode DDA en parallèle. Le temps d'analyse est donc nécessairement allongé pour ce type de stratégie et la quantité de matériel biologique utilisé augmentée. Il est possible d'utiliser des banques spectrales partagées entre membres de la communauté protéomique mais ces outils ne sont pas aussi puissants que la génération d'une librairie de spectre propre aux échantillons, acquise dans les mêmes conditions et notamment avec le même couplage instrumental.

Les prochaines évolutions concernant cette stratégie semblent s'orienter vers la réduction des fenêtres d'isolement, ce qui lui permettrait de gagner encore en spécificité, et donc en performance. Cette évolution passera donc par des spectromètres de masse plus rapide. Ce gain en rapidité permettra également d'élargir les gammes de masses analysées.

B.2.2.2. Modèle MS^e (WATERS)

La seconde stratégie DIA, basée sur la fragmentation de tous les peptides élués à chaque instant, est nommée MS^e et a émergé il y a quelques années sur des instruments de type TOF [132]. Aujourd'hui, cette technique est commercialisée par la société WATERS et est adaptée pour des spectromètres de masse de type Q-TOF.

L'ensemble de la gamme de masses est simultanément fragmentée en permanence. Un algorithme permet d'assigner les fragments ainsi obtenus à leurs précurseurs à partir d'un alignement

chromatographique : les profils d'intensité des ions fragments sont en effet similaires à ceux des précurseurs au cours du temps [132]. Contrairement à la stratégie présentée précédemment, la quantification MS^e est réalisée d'après les courants d'ions extraits à partir des spectres MS. Etant donné la complexité des spectres de fragmentation générés, il est indispensable d'utiliser un spectromètre de masse résolutif et précis en masse. Cette stratégie est souvent associée à une dimension supplémentaire de séparation de façon à réduire la complexité des spectres de fragmentation générés. Cette dimension supplémentaire consiste à séparer des peptides en amont de la cellule de collision en fonction de leur taille et de leur forme sous l'effet d'un champ électrique faible (mobilité ionique) [133].

En principe, la stratégie MS^e pourrait être implémenté sur tous types de spectromètre de masse de haute résolution mais l'existence d'un seul algorithme de traitement de ce type de données (implémenté au sein du logiciel ProteinLynx Global SERVERTM, PLGSTM), non open source, et la complexité des spectres de fragmentation générés, a freiné un peu l'expansion de cette technologie.

III. Quels points sont à considérer lors du choix d'une stratégie protéomique quantitative ?

De nombreux points sont à considérer et les principales questions à se poser avant d'effectuer ce choix sont:

- Quelles sont les protéines à quantifier ? Combien sont-elles ?
- Quel est le parc instrumental disponible ?
- S'il s'agit d'une approche ciblée, est-il nécessaire d'apporter une réponse « absolue » ?
- Quel est le type d'échantillon à analyser ? (culture cellulaire, tissus, etc.)
- Quel est le nombre de conditions à comparer ? Le nombre d'échantillons à analyser requiert-il l'utilisation d'une stratégie qui puisse permettre l'analyse de plusieurs échantillons simultanément ?

Etant donné le nombre (conséquent) d'approches quantitatives développées, il n'existe pas de réponse unique quant au choix d'une analyse quantitative pour une problématique biologique. Le choix de la bonne stratégie quantitative en protéomique fait souvent l'objet de compromis : un faible coût/nombre d'échantillons analysés simultanément, la sensibilité/le nombre de protéines quantifiées, etc.

A. Le contexte : approche globale ou ciblée ?

Pour chaque application biologique, il est important de savoir quelle est l'hypothèse de départ. S'agit-il d'une stratégie ciblée ou non ? La plupart des stratégies quantitatives peuvent donner une réponse globale à l'exception de la SRM (qui permet cependant aujourd'hui de quantifier environ 50 protéines simultanément grâce à l'évolution du parc instrumental). Plus l'hypothèse sera précise, plus la réponse pourra être elle-même précise, voire absolue. Il n'est toutefois pas possible, à l'heure actuelle, de donner une réponse ciblée (voire absolue) sur des milliers de protéines simultanément.

B. Le type de quantification pour les stratégies ciblées : absolu/relatif

Nous avons vu dans ce chapitre qu'il existe un grand nombre de standards qui peuvent être couplés à la technique SRM dans le but de réaliser une quantification absolue tel que des peptides standards, les protéines de type QconCAT ou PSAQTM. Il est important de considérer que la précision de la quantification est inversement proportionnelle au coût et au temps de développement nécessaire à la production de standards de haute qualité. De plus, la justesse de ces stratégies est encore à discuter en fonction des échantillons et du type de protéine à quantifier.

C. Le parc instrumental

Lorsque que la vitesse d'acquisition du spectromètre de masse est limitée, il est possible d'allonger les gradients en chromatographie liquide ou de décomplexifier les échantillons par gel d'électrophorèse (SDS-Page par exemple). En revanche, ce type de traitement allonge fortement les temps d'analyse (souvent au détriment du nombre de répliques analysés) et complique le traitement des données (i.e. les protéines présentes dans plusieurs bandes après décomplexification par gel). Dans la mesure où une dimension de séparation supplémentaire est requise, il peut être intéressant de privilégier des stratégies permettant le multiplexage des échantillons. Il est également possible de palier à des limites instrumentales en sélectionnant des stratégies quantitatives type 2D-DIGE.

D. Gestion du nombre d'échantillons/conditions à considérer

La stratégie quantitative ainsi que la préparation de l'échantillon doivent être choisies en considérant le nombre d'échantillons et de conditions à évaluer. La plupart des stratégies par marquage chimique ne permettent d'analyser simultanément que deux ou trois échantillons. Les stratégies permettant d'analyser un plus grand nombre d'échantillons (iTRAQ®/TMTTM ou combinaison

iTRAQ®/TMT™/SILAC) nécessitent l'utilisation d'un appareil hautement résolutif. Les stratégies sans marquage ou basées sur l'utilisation de gels d'électrophorèse bidimensionnelle permettent également de quantifier les protéines extraites pour un nombre élevé d'échantillons.

En guise de conclusion, un schéma résumant la diversité des approches protéomiques qualitatives et quantitatives en spectrométrie de masse ainsi que leur spécificité est représenté en figure III-12.

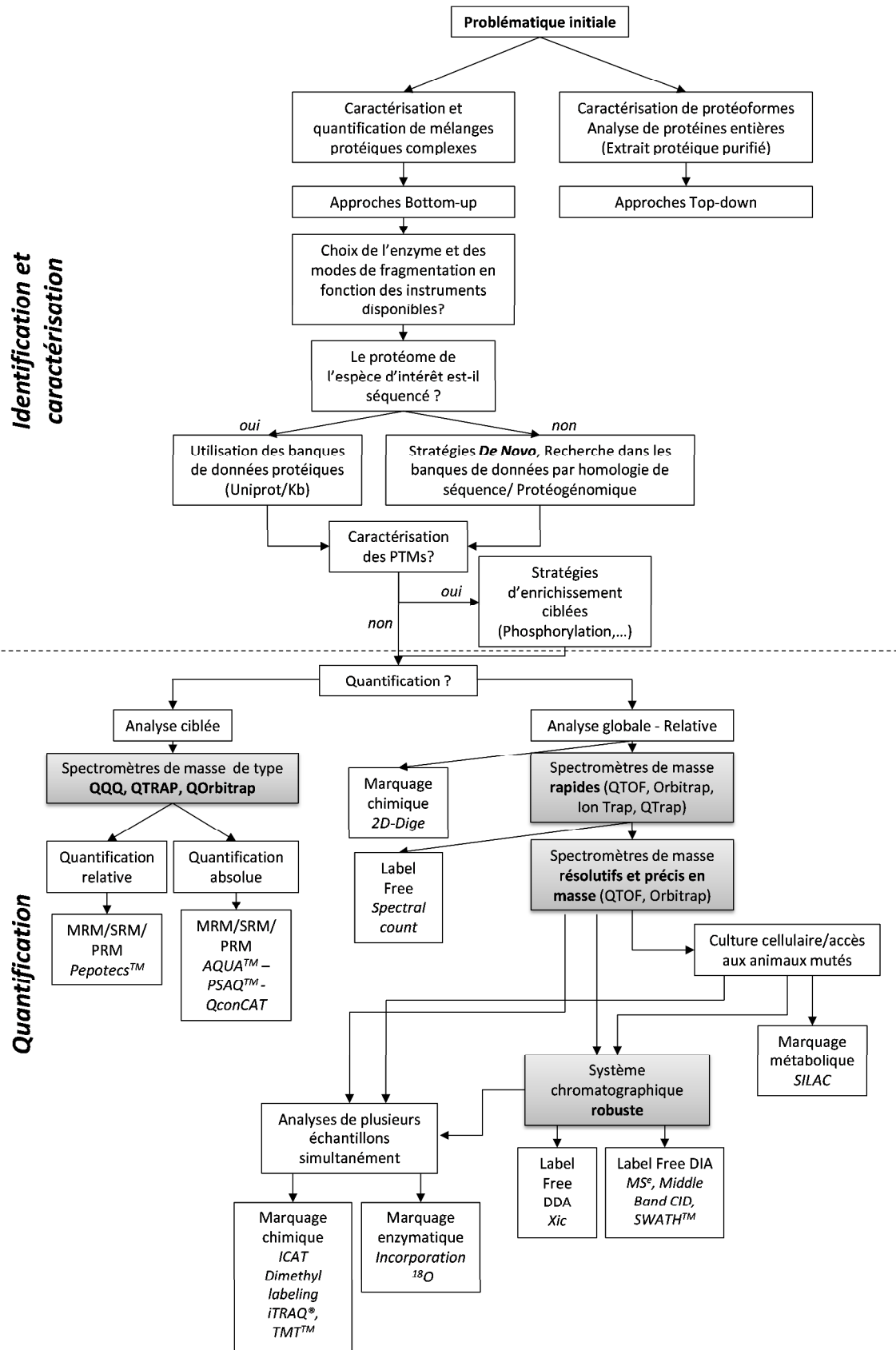


Figure III-12 Schéma décisionnel pour le développement d'une stratégie d'analyse qualitative et quantitative en protéomique

Chapitre IV : Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes

Chapitre IV : Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes

Les méthodes d'acquisition des données en spectrométrie de masse doivent être finement optimisées pour augmenter le nombre de peptides, et donc de protéines, qu'il est possible d'identifier ou de quantifier dans des matrices complexes tout en générant des données de qualité permettant d'obtenir des informations robustes et fiables.

Les instruments utilisés dans ce chapitre sont des spectromètres de masse de type Q-TOF commercialisés par la société Bruker Daltonique couplés à des systèmes de chromatographie liquide de type nanoUPLC (nanoAcquity Waters) équipés d'une colonne C18 BEH 250 × 0,075 mm (taille de particule : 1,7 µm de diamètre) : le maXis 4G et l'Impact-HD. Les spécifications de chacun de ces instruments sont présentées en tableau 1.

Tableau 1 : Spécifications des spectromètres de masse de type Q-TOF maXis 4G (*avant la mise à jour en version 4G Plus*) et Impact-HD

	maXis 4G	Impact-HD
Précision	< 5ppm	< 10ppm
Résolution (m/z 900)	≈ 60000	≈ 40000
Vitesse de scan	≈ 20Hz	≈ 50Hz

L'Impact-HD est une version du maXis 4G optimisée pour la protéomique. Le raccourcissement de son tube de vol par rapport au maXis 4G a réduit le pouvoir résolutif de l'analyseur, cependant toujours supérieur à 40 000 environ pour une valeur de m/z égale à 922. En revanche, l'Impact-HD a la capacité d'atteindre des vitesses de scan accrues (jusqu'à 50 Hz), sans perdre en qualité de spectre de fragmentation, ce qui permet de réduire les phénomènes de sous-échantillonnage (présentés dans la suite du chapitre) et ainsi de gagner en sensibilité (du simple au double en termes de protéines identifiées pour un même échantillon de levure (*Saccharomyces cerevisiae*) injecté dans les mêmes conditions. Ce gain en sensibilité permet donc d'augmenter la couverture du protéome étudié.

Ce chapitre décrit l'établissement de méthodes d'acquisition de type nanoLC-MS/MS en haute résolution générées en mode DDA (Data Dependant Acquisition) adaptées aux analyses qualitatives (carte peptidique) et quantitatives (stratégie XIC présentée dans le chapitre III « Stratégies d'analyses protéomiques quantitatives » en page 68) d'échantillons complexes. L'influence sur les acquisitions des différents paramètres à optimiser sera présentée dans la première partie de ce chapitre.

Nous avons également évalué les performances d'une toute nouvelle source de type « CaptiveSpray » équipée d'un « NanoBooster » installée sur l'Impact-HD (Bruker Daltonique). Cette source et les résultats de son évaluation sont présentés dans la deuxième partie de ce chapitre.

I. Optimisations des méthodes d'acquisition en mode DDA pour l'analyse d'échantillons complexes sur le maXis 4G (Bruker Daltonique)

Le principe d'acquisition des analyses en mode DDA est présenté au chapitre I de la partie bibliographique (en page 33). Le spectromètre de masse maXis 4G (Bruker Daltonique), pour lequel les optimisations des méthodes ont été réalisées, est présenté en figure IV-1.

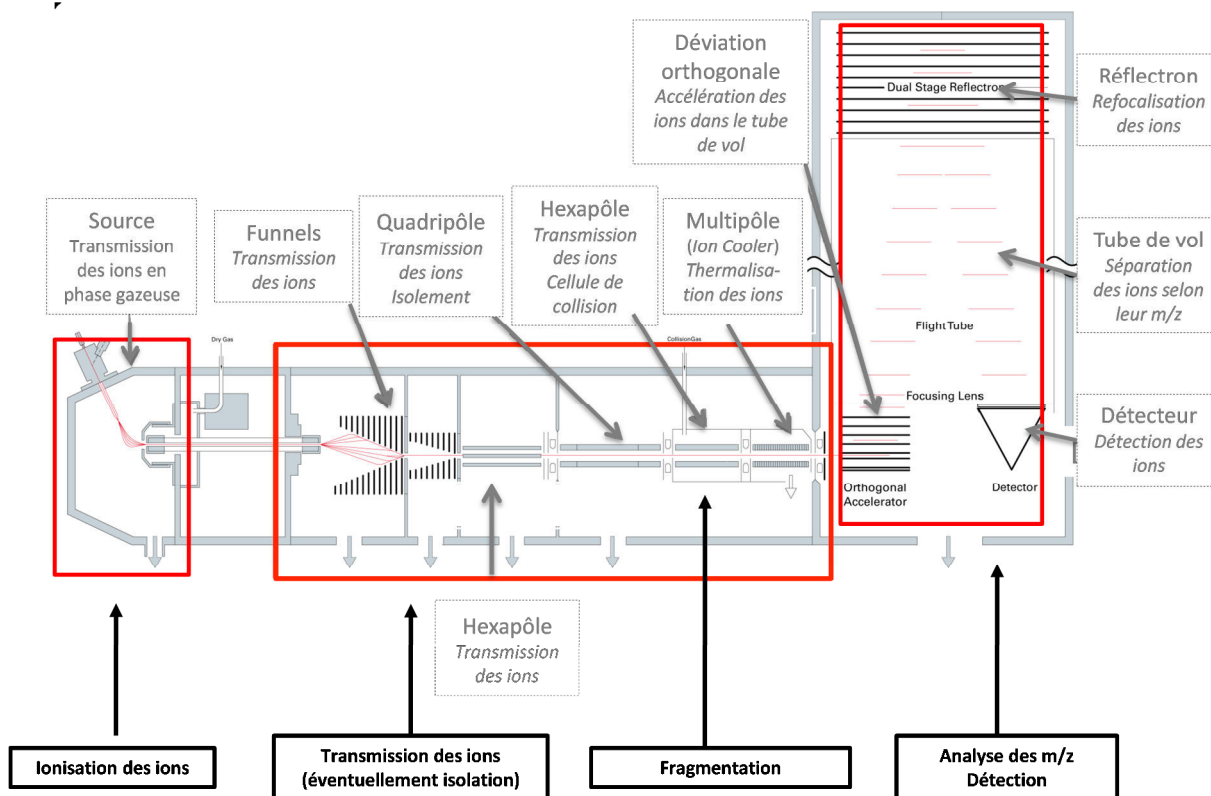


Figure IV-1 Schéma de l'analyseur Q-TOF maXis 4G (Bruker Daltonique)

Le système est composé d'une source électrospray, d'une combinaison de funnels, lentilles et multipôles pour la transmission des ions jusqu'à l'analyseur (et éventuellement l'isolement d'un peptide donné et sa fragmentation dans la cellule de collision), d'un analyseur TOF composé d'une déviation orthogonale permettant l'accélération des ions dans le tube de vol, d'un réflectron permettant de refocaliser les ions de même m/z qui auraient des énergies cinétiques différentes au moment de leur accélération et d'un détecteur qui transforme le signal de détection des ions en un courant électrique.

De nombreux paramètres doivent être considérés pour la mise au point des méthodes d'acquisition en mode DDA sur un système de type LC-Q-TOF. Ces paramètres sont présentés en figure IV-2. Ces paramètres peuvent concerner des valeurs physiques du spectromètre de masse (tension, température, débit de gaz) qui permettent de transmettre jusqu'au détecteur le plus d'ions possible sur une gamme de masses donnée, et ce avec une résolution élevée, représentés dans les cadres noirs sur la figure IV-2. Ces valeurs sont en général choisies et optimisées à l'installation d'un appareil et ne font pas l'objet d'optimisation à chaque nouveau type d'échantillon à condition de toujours travailler sur la même gamme de masses. Ces paramètres seront nommés « instrumentaux » dans la suite du chapitre. En revanche, il existe un certain nombre de paramètres d'acquisition qui eux doivent être optimisés pour chaque type d'échantillon analysé. Ils concernent les valeurs de débit ou de gradient d'éluion en chromatographie liquide, les critères de sélection des ions pour la MS/MS, les temps de cycle, ou encore le nombre de sommations réalisées pour générer les spectres MS et/ou MS/MS. Ces paramètres sont représentés dans les cadres bleu-vert sur la figure IV-2.

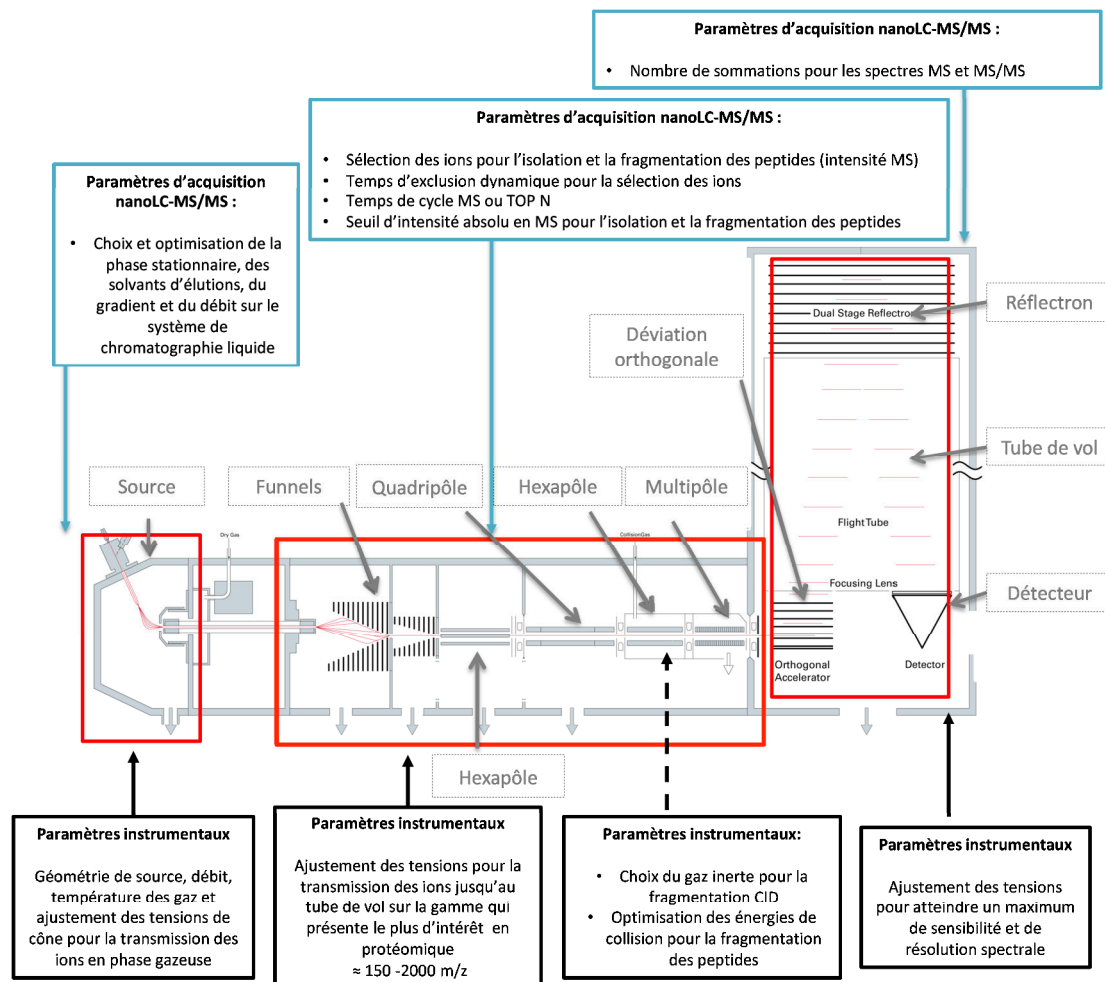


Figure IV-2 Paramètres instrumentaux et d'acquisition à optimiser pour la réalisation d'analyses nanoLC-MS/MS

Les paramètres dits « instrumentaux » concernent les débits de gaz, tensions et températures qui permettent la transmission du plus d'ions possible, et avec une résolution spectrale élevée, sur la gamme de masse d'intérêt en protéomique. Les paramètres dits d'acquisition concernent eux la sélection des ions pour la fragmentation, les temps de cycle MS ou MS/MS et le nombre de sommations réalisé pour la génération des spectres MS ou MS/MS.

Seule l'optimisation des paramètres dits d'acquisition, ainsi que celle des énergies de collision (paramètres instrumentaux) seront présentés dans ce chapitre.

A. Optimisations des paramètres d'acquisition des méthodes nanoLC-MS/MS en mode DDA

Les paramètres dits « d'acquisition » clés lors de la mise au point d'une méthode nanoLC-MS/MS sont :

- le temps de cycle MS (ou TOP N) et le nombre de sommations des spectres MS
- le temps d'exclusion dynamique pour la sélection des peptides en MS/MS
- le nombre de sommations des spectres MS/MS en fonction des intensités des ions précurseurs

A.1. Le temps de cycle MS et le nombre de sommations des spectres MS

Ces paramètres sont particulièrement importants pour les analyses quantitatives basées sur l'extraction des courants d'ions à partir des spectres MS (XIC). En effet, suffisamment de spectres MS doivent être générés de façon à pouvoir extraire un courant d'ions précis au cours du temps. Cependant, le nombre de spectres MS réalisés ne doit pas être trop important de façon à pouvoir, en parallèle de la quantification, générer suffisamment de spectres MS/MS afin d'identifier le plus grand nombre de peptides possible au sein de chaque analyse. Il est donc nécessaire de trouver un bon compromis entre le nombre de spectres MS (qui permettent de réaliser la quantification) et le nombre de spectres MS/MS (qui permettent l'identification des peptides).

Le temps de cycle MS peut être défini de plusieurs façons : les spectres MS peuvent être acquis à intervalles réguliers (toutes les trois secondes par exemple en accord avec le gradient chromatographique utilisé) ou au travers d'un nombre de spectres MS/MS fixe entre chaque spectre MS généré (stratégie souvent notée « TOP N » pour 1 acquisition MS suivie de N MS/MS).

L'effet du temps de cycle MS sur la définition des courants d'ions extraits est schématisé en figure IV-3 :

1. Le courant d'ion a été extrait à partir d'une analyse pour laquelle la méthode d'acquisition possède un temps de cycle MS trop court. Le courant d'ion est très précis mais très peu de spectres MS/MS seront générés simultanément pour l'ensemble des peptides co-élus avec ce peptide.

2. Le courant d'ion a été extrait à partir d'une analyse pour laquelle la méthode d'acquisition est optimale : 8 spectres MS environ permettent de décrire le courant d'ions, le reste du temps est alloué à la génération de spectres MS/MS pour ce peptide ou les autres peptides co-élus.
3. Le courant d'ion a été extrait à partir d'une analyse pour laquelle la méthode d'acquisition possède un temps de cycle MS trop long (la génération des spectres MS est trop espacée dans le temps). Le temps de génération de spectres MS/MS est élevé, ce qui est bien pour générer des spectres MS/MS pour un grand nombre de peptides mais le courant d'ion du peptide à quantifier n'est défini qu'avec trois points, ce qui est insuffisant pour extraire un courant d'ions précis et faire de la quantification.

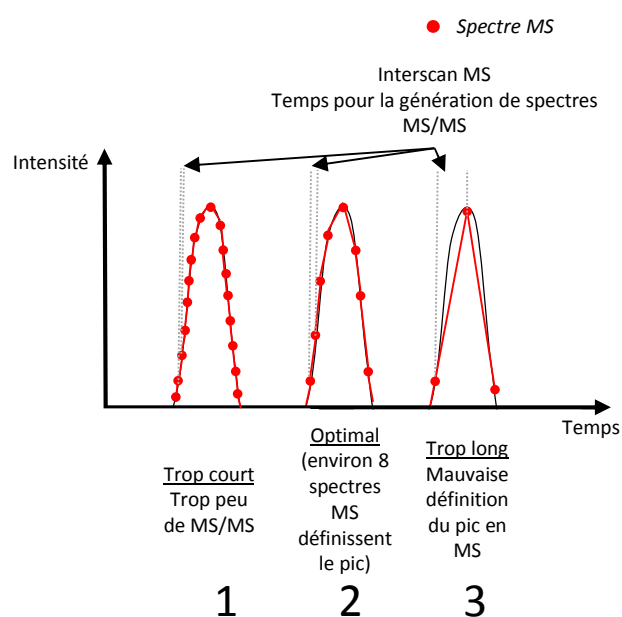


Figure IV-3 Représentation schématique de courants d'ions extraits à partir de spectres MS pour un peptide donné et pour trois configurations de temps de cycle MS et MS/MS différentes

Trois configurations de temps de cycle sont présentées : 1) le nombre de spectres MS réalisé est trop élevé 2) le nombre de spectres MS est idéal avec environ 8 points par pic 3) le nombre de spectres MS est trop faible.

Dans le contexte de l'analyse quantitative XIC, un autre paramètre à considérer particulièrement est le nombre de sommations pour la génération des spectres MS. Comme décrit au chapitre I « Principe général de l'analyse protéomique par spectrométrie de masse », en page 25, la génération des spectres MS et MS/MS sur un Q-TOF est le résultat de la sommation d'une multitude de spectres générés à la suite de l'accélération électrique des ions qui arrivent à l'entrée du tube de vol. Plus la quantité d'ions qui arrive à l'entrée du tube de vol est importante, plus le nombre de sommations nécessaire à la génération du spectre MS ou MS/MS avec un rapport signal sur bruit élevé sera réduit. Pour la stratégie XIC, les ions doivent présenter un ratio signal sur bruit élevé sur le spectre MS de façon à pouvoir extraire un courant d'ions précis au cours du temps. Plus le ratio signal sur bruit des ions pour lesquels extraire des courants d'ions est élevé, moins ces courants d'ions sont

sensibles à des interférences qui pourraient nuire à la quantification. Dans la suite du manuscrit, on parlera de vitesse d'acquisition pour la génération des spectres MS et MS/MS. Elle est couramment exprimée en Hertz (Hz). Cette valeur correspond au nombre de spectres générés par seconde. Par exemple, 2 Hz signifie que 2 spectres sont générés par seconde soit que chaque spectre est généré en 500 ms. Le temps de parcours des ions dans le tube de vol du maXis 4G est égal à 0,16 ms (pour une gamme de masse d'environ 100-2500 m/z), ce qui correspond au temps nécessaire pour générer un spectre avec 1 sommation. Un spectre généré en 500 ms correspond donc à 3141 sommations par spectre généré. Les intensités relatives des ions sur les spectres devront toujours être rapportées au nombre de sommations utilisées pour générer ces spectres. Plus le nombre de sommation est élevé, plus l'intensité de l'ion concerné sera élevée sur le spectre généré.

A.2. Le temps d'exclusion dynamique et le seuil de déclenchement minimal pour la MS/MS

En mode DDA, un certain nombre d'ions sont isolés, puis fragmentés successivement en fonction de leur intensité sur le spectre MS précédent [58]. La totalité des ions élués au sein d'une analyse ne sera donc pas fragmentée. Ceci est essentiellement dû au fait que les vitesses de scans ne sont toujours pas assez élevées pour pouvoir le permettre, bien que les développements instrumentaux aient permis de largement les augmenter ces dernières années. Ceci crée un phénomène de sous-échantillonnage. Afin de couvrir le protéome le plus largement possible, il est donc nécessaire de fixer un temps d'exclusion dynamique, de façon à ne pas toujours sélectionner et fragmenter les mêmes ions (même s'ils restent les plus intenses sur le dernier spectre MS généré du cycle). En limitant la redondance d'acquisition des spectres MS/MS acquis pour un même peptide, le temps d'exclusion dynamique permet donc d'analyser davantage de peptides (et donc de protéines) au sein d'une même analyse.

Sur les logiciels du constructeur Bruker Daltonique (MicroTOF control et OTOFcontrol), il est possible de paramétrer un seuil d'intensité absolu minimal sur le dernier spectre MS réalisé pour la sélection des peptides en mode MS/MS (ce seuil est lié au nombre de sommations utilisées pour générer ce spectre). Ainsi, un peptide présentant une valeur d'intensité trop basse (proche du bruit de fond) ne sera pas sélectionné pour la fragmentation. En effet, si l'intensité de cet ion est trop faible, il sera très difficile de pouvoir détecter suffisamment d'ions fragments pour pouvoir générer un spectre MS/MS informatif. Cette valeur seuil d'intensité absolue doit donc être correctement ajustée, pour chaque type d'échantillon, par rapport à la valeur de bruit de fond.

Le temps d'exclusion dynamique doit être choisi en fonction de :

- la largeur à la base des pics chromatographiques
- la complexité de l'échantillon
- la valeur du seuil d'intensité minimal des ions précurseurs en MS pour la sélection en MS/MS

En général, le temps d'exclusion dynamique choisi est égal à la moitié de la largeur des pics chromatographiques. Ainsi, si le seuil d'intensité absolu pour la sélection des ions est très bas et que le peptide est sélectionné très tôt lors de son élution, alors il sera sélectionné une seconde fois (ou plus) quand il sera plus proche de son extremum. De façon générale, si l'échantillon n'est pas trop complexe, le phénomène de sous-échantillonnage n'est plus limitant et il est alors préférable d'avoir un temps d'exclusion assez court de façon à pouvoir re-sélectionner le peptide et ainsi favoriser l'acquisition d'un spectre MS/MS de meilleure qualité pour un peptide donné. Le phénomène est présenté en figure IV-4.

Au contraire, dans le cas d'un échantillon très complexe, il est préférable d'allonger ce temps d'exclusion (supérieur à la largeur à la base des pics chromatographiques) et d'élever le seuil d'intensité minimum de déclenchement pour la MS/MS. Ainsi, le peptide ne sera sélectionné qu'une seule fois, ce qui limitera le phénomène de sous-échantillonnage. De plus, l'intensité des peptides sélectionnés pour la fragmentation sera, dans ces conditions, suffisamment élevée pour générer un spectre MS/MS suffisamment informatif pour être interprétable par les algorithmes de recherche de type Mascot (Matrix Science) ou les algorithmes permettant de réaliser un séquençage *de novo* (présentés dans le chapitre II « Stratégies d'analyses protéomiques qualitatives »).

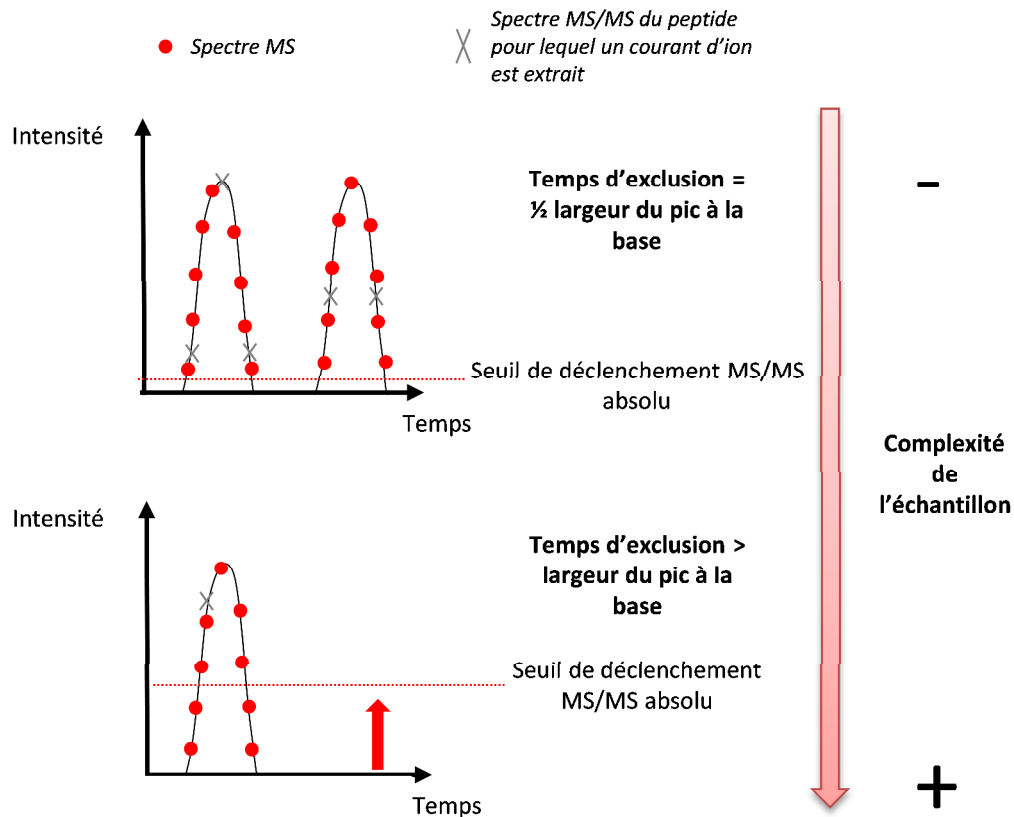


Figure IV-4 Courant d'ion schématique pour un peptide donné en fonction de différentes configurations de temps d'exclusion dynamique et de seuil de déclenchement d'intensité absolue pour la MS/MS

Si l'échantillon est peu complexe, le seuil de déclenchement d'intensité absolue pour la MS/MS peut être abaissé. L'utilisation d'un temps d'exclusion dynamique environ égal à la moitié de la largeur du pic chromatographique permettra de sélectionner deux à trois fois le peptide pour générer un spectre MS/MS au cours de son élution. Ainsi, la probabilité de générer un spectre MS/MS informatif sera plus élevée. En revanche, pour l'analyse d'un échantillon complexe, il est recommandé d'augmenter le seuil d'intensité qui permet le déclenchement MS/MS et d'allonger les temps d'exclusion dynamique, de façon à ne générer qu'un seul spectre MS/MS par peptide et ainsi pouvoir générer des spectres MS/MS pour le plus grand nombre de peptide possible.

Certains constructeurs (Bruker Daltonique) se tournent vers de nouvelles configurations logicielles. Bien que l'exclusion dynamique soit activée, il est possible de re-déclencher l'acquisition d'un spectre MS/MS si l'intensité du précurseur évolue au-delà d'une certaine valeur. Ce nouveau type de paramétrage limite le besoin d'optimiser finement le paramètre de temps d'exclusion dynamique pour chaque nouveau type d'échantillon, ce qui permet de gagner du temps.

Cas particulier : les redondances de génération de spectres MS/MS sont requises pour le cas particulier des stratégies protéomiques quantitatives basées sur une méthode de comptage de spectre (« Spectral Counting »). En effet, comme présenté dans le chapitre III « stratégies d'analyses protéomiques quantitatives » en page 48, cette méthode nécessite une certaine redondance de spectres par protéine de façon à générer des données statistiquement fiables. Ces méthodes nécessitent un paramétrage fin de façon à tout de même permettre une caractérisation du protéome importante.

A.3. Le nombre de sommations des spectres MS/MS ainsi que les seuils d'intensité correspondants

A.3.1. Vitesse d'acquisitions MS/MS : Principe et paramétrage

Sur les instruments QTOF Bruker Daltonique (comme sur le dernier Q-TOF 6600 commercialisé par la société ABSCIEX), le nombre de sommations pour les spectres MS/MS peut être ajusté en fonction de l'intensité du peptide sur le dernier spectre MS (à pondérer par le nombre de sommations réalisées pour générer ce spectre). De ce fait, plus les intensités des ions précurseurs sont élevées, plus les temps de génération des spectres MS/MS peuvent être courts, et ainsi plus la vitesse de génération des spectres MS/MS peut être rapide.

Le paramétrage du nombre de sommations à effectuer est dépendant :

- Des propriétés physiques du spectromètre de masse : le nombre d'ions parvenant au détecteur doit être suffisamment élevé pour réaliser des spectres de qualité. Par exemple, dépasser 50 spectres par seconde (soit 50 Hz) pour l'Impact-HD (Bruker Daltonique) donnerait lieu à des spectres avec des ratios signal sur bruit trop bas. L'appareil n'est pas assez sensible pour générer des spectres aussi rapidement.
- De l'intensité des précurseurs : plus la quantité de peptide à isoler, puis fragmenter, est élevée, plus le nombre de sommations peut être réduit. La figure IV-5 représente de façon schématique l'intensité d'un ion au cours de son élution, plus l'intensité est élevée (plus le cadre est clair), plus le nombre de sommations nécessaire pour un générer un spectre MS/MS informatif est faible.

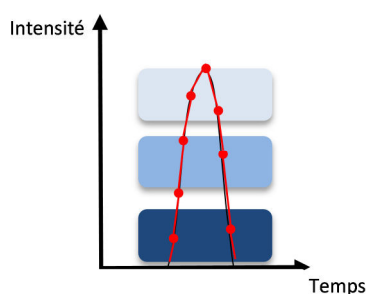


Figure IV-5 Représentation schématique de l'intensité d'un ion au cours de son élution

- De la complexité de l'échantillon : si l'échantillon est très complexe, ces vitesses doivent être paramétrées de façon à générer des spectres informatifs le plus rapidement possible, en revanche, si l'échantillon est très peu complexe, il est possible de générer les spectres de façon plus lente afin d'obtenir un meilleur signal sur bruit.

Le principe de paramétrage du nombre de sommations pour chaque spectre MS/MS sur les spectromètres de masse de type Q-TOF commercialisés par la société Bruker Daltonique est illustré en figure IV-6. Il est à noter que, sur ce type d'instruments, ce paramétrage peut être affiné en fonction de la complexité des zones d'éluion des peptides (grâce à l'utilisation de segments réalisés par l'opérateur). Un nombre de sommations maximum et minimum est fixé par segment. Le nombre de sommations réalisées pour chaque spectre MS/MS dépendra de l'intensité du précurseur sur le dernier spectre MS et oscillera entre le nombre de sommations maximum et minimum. Les seuils d'intensité (A) et (B) sont les intensités clés de ce paramétrage : pour une intensité de précurseur inférieure à la valeur A, le nombre de sommations réalisées sera automatiquement égal au nombre de sommations maximal paramétré et pour une intensité supérieure à B, le nombre de sommations sera forcément égal au nombre de sommations minimum imposé. Le nombre de sommations pour chaque peptide donné dépendra ensuite de son intensité et sera compris entre les nombres de sommations minimum et maximum qui sont fixés par l'opérateur selon la courbe présentée en figure IV-6.

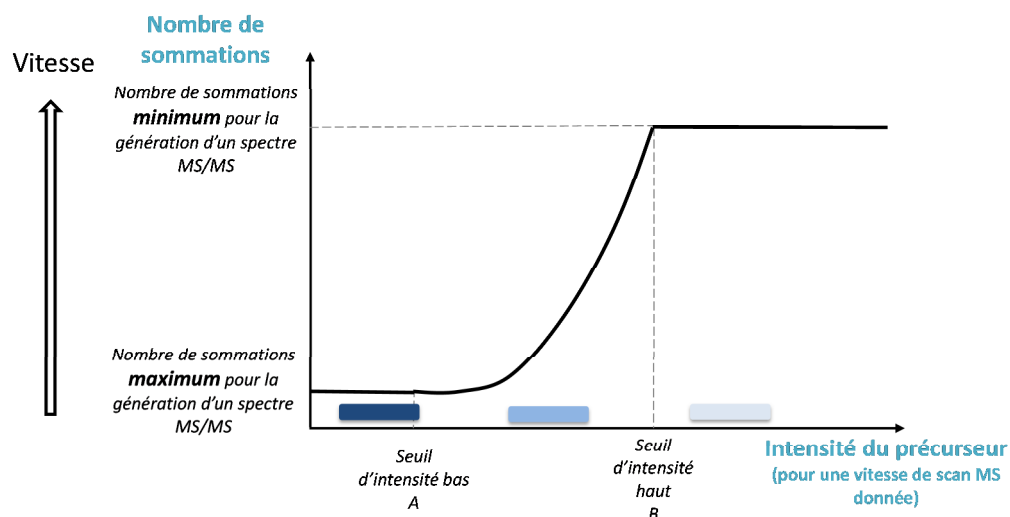


Figure IV-6 Principe de paramétrage du nombre de sommations pour chaque spectre MS/MS avec les logiciels d'acquisition des spectromètres de masse commercialisés par la société Bruker Daltonique

Les seuils d'intensité A et B sont les seuils d'intensité clés à paramétrer dans le logiciel d'acquisition des données. Pour un précurseur dont l'intensité est inférieure à A, le nombre de sommations réalisées sera égal au nombre de sommations maximum paramétré, tandis que pour une intensité supérieure à B, le nombre de sommations sera égal au nombre de sommations minimum paramétré. Entre ces deux intensités A et B, le nombre de sommations réalisées pour chaque précurseur dépendra de son intensité sur le dernier spectre MS réalisé et oscillera donc entre les nombres de sommations minimum et maximum imposés par l'opérateur selon la courbe présentée.

A.3.2. Evaluation du seuil d'intensité « B »

Les analyses nanoLC-MS/MS génèrent un très grand nombre de spectres. La plupart de ces spectres ne sont pas assignés ou sont assignés à des séquences peptidiques avec une faible confiance (30-50% des spectres en moyenne). Afin de réduire la proportion des spectres MS/MS non assignés du fait de

leur faible qualité, nous avons évalué la possibilité de réduire la quantité de spectres générés par analyse au profit de leur qualité. Cette évaluation a été réalisée à partir d'un digestat de levure (1 µg élué en 80 minutes).

Pour ce faire, nous avons augmenté le seuil d'intensité de B à B', ce qui conduit à l'aplatissement de la courbe et à l'augmentation du nombre de sommations global pour l'ensemble des spectres MS/MS générés (figure IV-7).

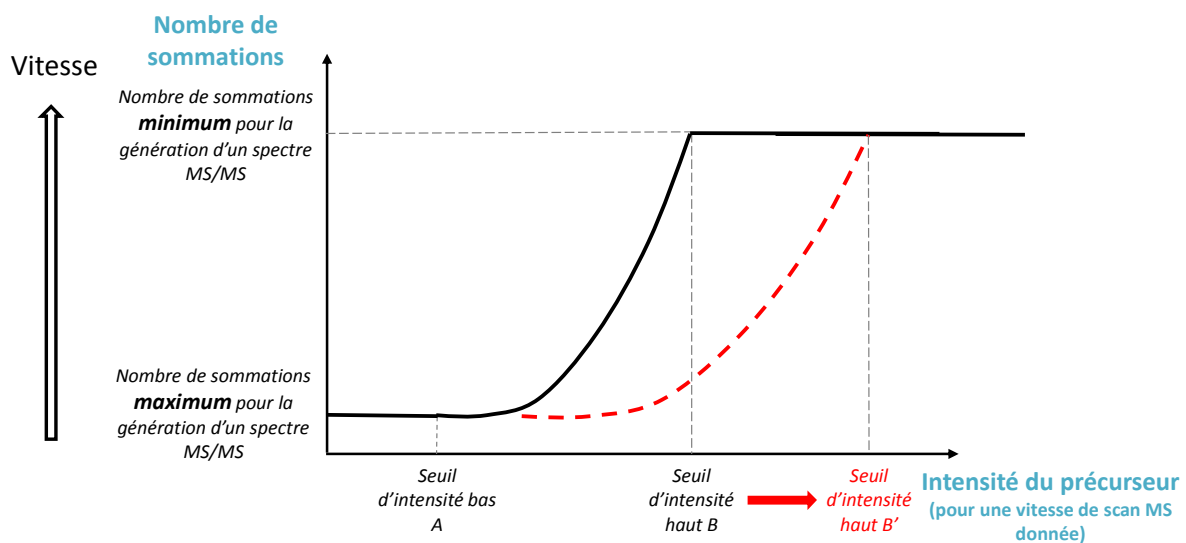


Figure IV-7 Principe de paramétrage du nombre de sommations pour chaque spectre MS/MS avec les logiciels d'acquisition des spectromètres de masse commercialisés par la société Bruker Daltonique (augmentation du seuil B à B')
En augmentant le seuil B', on augmente le nombre de sommations global réalisées pour l'ensemble des précurseurs sélectionnés.

Une illustration de l'impact de la modification du seuil B à B' (pour B'= 60000, 70000, 80000, 100000 et 150000) sur les données générées est présentée en figure IV-8. Cette figure présente, en fonction des différents seuils d'intensité B' testés, le nombre total de protéines identifiées (C), le nombre total de spectres générés (D), le nombre de spectres assignés avec un score supérieur ou égal à 60 d'après l'algorithme de recherche Mascot (E) considérés comme très informatifs ainsi que le nombre total de spectres assignés (F).

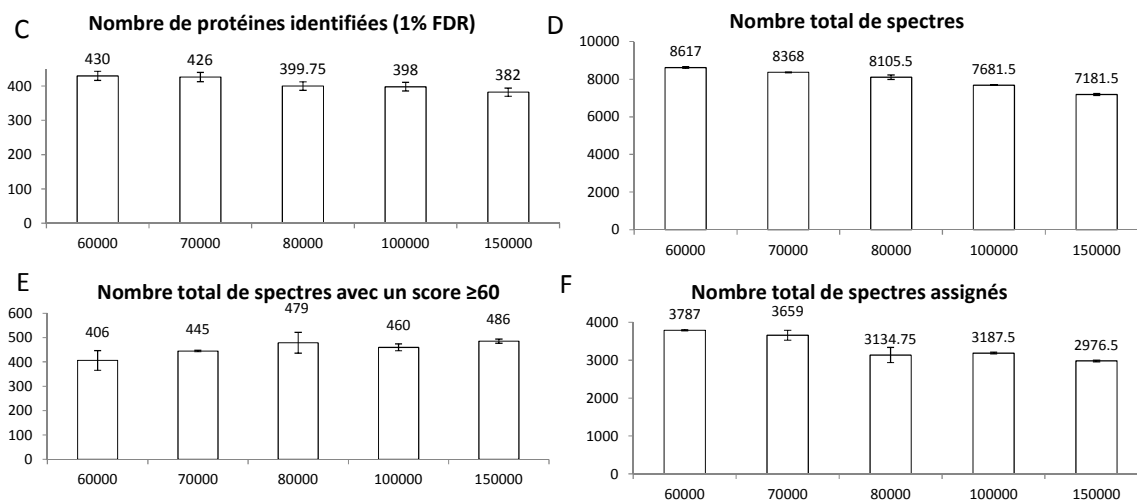


Figure IV-8 Illustration de l'influence de la modification du seuil B sur le nombre de protéines identifiées, le nombre total de spectres, le nombre total de spectres avec un score supérieur à 60 et le nombre total de spectres assignés
 Ces données ont été générées à partir d'un microgramme de digestat protéique de levure. Le temps de gradient utilisé était de 80 minutes.

Pour ce test, plus la valeur de B' est élevée (de 60000 à 150 000), plus les nombres de sommations réalisées pour l'ensemble des spectres MS/MS seront élevés. Plus le seuil d'intensité « B' » est élevé, plus le nombre de spectres générés, de même que le nombre de spectres assignés (figure IV-8D et F) est réduit car les temps d'acquisition sont plus longs pour l'ensemble des spectres. En revanche, l'augmentation du nombre de sommations des spectres MS/MS a une influence sur la qualité des spectres générés puisqu'on constate que le nombre total de spectres avec un score supérieur à 60 augmente (figure IV-8E). En revanche, il est possible de constater que le nombre total de protéines identifiées a tendance à diminuer avec l'augmentation du seuil B'. Ainsi nous avons pu observer qu'il était préférable d'augmenter la cadence de génération des spectres MS/MS de façon à générer des spectres pour un plus grand nombre de peptides plutôt que d'augmenter le nombre de sommations pour chacun d'eux dans le but d'améliorer leur qualité. Dans le cadre d'une analyse *de novo*, il serait en revanche nécessaire de travailler avec des spectres de la plus haute qualité possible de façon à ne pas introduire d'erreur d'interprétation.

Ralentir la vitesse de génération de spectres MS/MS réduit le nombre total de spectres et donc, en conséquence, la fraction du protéome étudié. Lorsque l'on augmente le temps d'acquisition moyen pour l'ensemble des spectres, on l'augmente également pour des peptides possédant des sites de clivage enzymatique manqués, des substitutions d'acides aminés, des coupures enzymatiques aspécifiques, etc., qui ne seront pas identifiés avec des paramètres de recherche classique. Augmenter le nombre de sommations pour ces peptides ne permettra pas d'augmenter le nombre

de protéines identifiées puisque ces peptides ne seront tout simplement pas recherchés avec les critères de recherche classiques.

Pour les futures méthodes d'acquisition développées sur ce spectromètre de masse, la valeur B paramétrée a donc été fixée à 80 000 qui représente un bon compromis entre le nombre total des spectres générés (qui permet d'obtenir des pourcentages de recouvrements importants) et leur qualité.

A.4. Optimisations des énergies de collision

Le mode de fragmentation utilisé sur les Q-TOFs Bruker est le mode CID (décrit dans le chapitre « Principe général de l'analyse protéomique par spectrométrie de masse » en page 31). L'énergie de collision à apporter pour la fragmentation dépend de l'état de charge des peptides concernés, de la longueur de leur séquence, de leur composition en acides aminés, etc. [51]. Il n'est pas possible d'optimiser les énergies de collision pour chaque peptide étant donné le millier de peptides qui peuvent être générés avec l'approche « Bottom-up » (contrairement à la stratégie SRM). En revanche, il est possible de réaliser une optimisation globale des énergies de collision en fonction des masses et des charges des peptides dans un mélange donné et de réutiliser ces valeurs optimisées pour l'analyse de peptides dans les mêmes gammes de masses et pour la même gamme d'états de charge. L'énergie de collision est contrôlée dans ce type d'analyseur par la différence de potentiel appliquée entre les lentilles et multipôles placés en amont de la cellule de collision et cette même cellule, de façon à accélérer les ions qui vont ensuite rentrer en collision avec des molécules de gaz inerte qui remplissent la cellule de collision, ce qui provoquera leur fragmentation. Si ces énergies sont trop basses, la proportion du précurseur restera très élevée sur le spectre de fragmentation et le nombre de fragments générés, ainsi que leur intensité, seront trop faibles pour que les spectres MS/MS puissent être correctement assignés à des séquences peptidiques. En revanche, si ces énergies sont trop élevées, on risque de générer de plus petits fragments secondaires, par rapport aux fragments b et y générés, qui compliqueront l'interprétation de ces spectres. Les énergies de collision (exprimées en eV) peuvent en théorie être transposées d'un instrument à l'autre (s'ils présentent exactement la même configuration) mais il est préférable de les optimiser plus finement sur chaque spectromètre de masse, à leur installation, afin de générer des spectres de qualité élevée.

Ces optimisations sont plutôt chronophages puisqu'elles nécessitent l'analyse d'échantillon sur une gamme d'énergies de collision assez importante. Nous avons réalisé cette optimisation en deux temps. Nous avons, tout d'abord, réalisé une première série d'optimisations à partir d'un échantillon peu complexe (composé de quelques protéines) qui va générer des peptides de longueurs et d'états

de charge variables. Ce mélange simple nous a permis d'optimiser les énergies de collision sur la gamme de 300 à 900 m/z avec des gradients chromatographiques courts (inférieurs à 20 minutes). La première série d'optimisation des énergies de collision du maXis 4G (Bruker Daltonique) a été réalisée sur un mélange de 4 protéines (commercialisé par la société Waters). La même quantité de ce mélange a été injectée de façon successive pour des énergies de collision fixes allant de 21 à 57 eV. Les scores des spectres assignés pour les 4 protéines ont ensuite été sommés par analyse, par état de charge et par gamme de m/z. Ces données sont présentées en figure IV-9. Etant donné que le logiciel d'acquisition, MlcrOTOF control ou OTOF control, est paramétré de façon à exclure les peptides monochargés pour lesquels les spectres de fragmentation sont peu informatifs, leurs énergies de collision n'ont pas été optimisées. De plus, la proportion de peptides 4 ou 5 fois chargés n'est pas assez représentative (moins de 10 % des spectres générés) pour pouvoir optimiser finement leurs énergies de collision. L'optimisation a donc porté sur les peptides di- et tri-chargés.

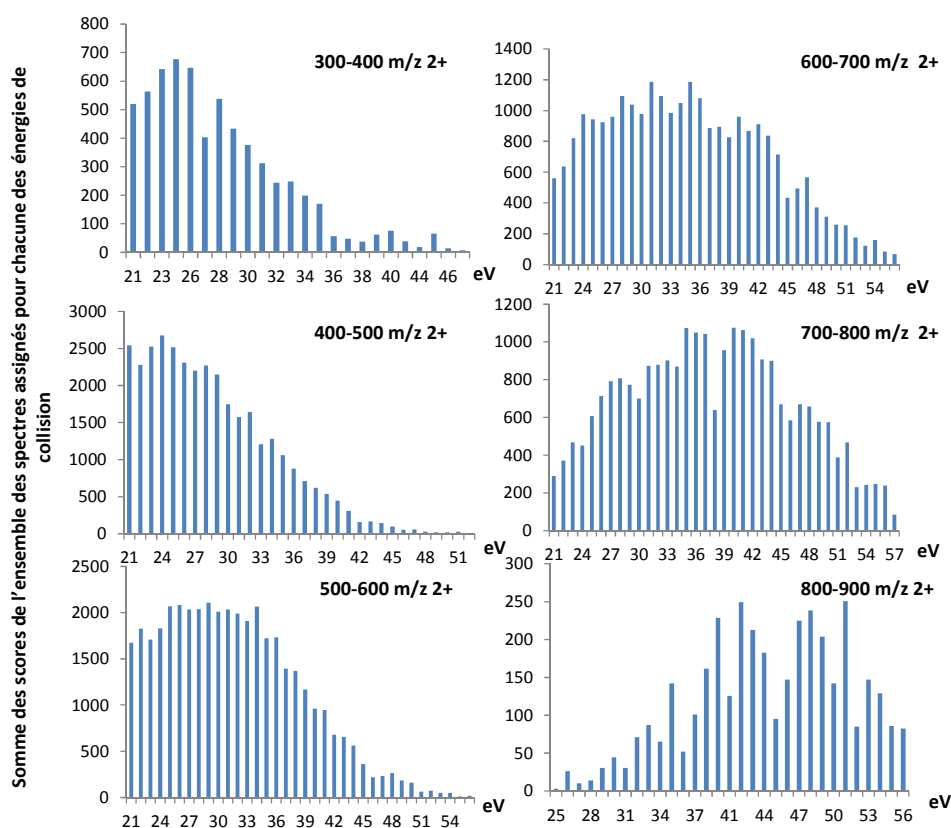


Figure IV-9 Somme des scores de l'ensemble des spectres assignés pour chacune des énergies de collision testées pour les peptides dichargés en fonction de la gamme de m/z

Plus la somme des scores est élevée pour l'ensemble des protéines, plus l'énergie de collision est adaptée pour la fragmentation des peptides pour une gamme de m/z et un état de charge donné.

On constate d'après la figure IV-9 que les énergies de collision optimales (celles pour lesquelles la somme des scores est la plus importante) varient en fonction du rapport m/z des peptides. Pour les peptides di-chargés, plus la masse des peptides est élevée, plus l'énergie requise pour leur fragmentation est élevée, ce qui était attendu [51]. Les énergies de collision finalement retenues

étaient celles pour lesquelles la somme des scores pour les 4 protéines confondues était la plus importante pour des gammes de masses de 100 m/z. Les énergies de collision retenues pour les gammes 300-400, 400-500, 500-600, 600-700, 700-800, 800-900 sont respectivement 25, 25, 30, 34, 42 et 52 eV.

Dans un deuxième temps, nous avons utilisé un digestat de levure (*Saccharomyces cerevisiae*), contenant plusieurs milliers de protéines, de façon à générer des peptides de masses molaires plus élevées pour pouvoir optimiser les énergies de collision des peptides possédant une valeur m/z supérieur à 900. Réaliser l'ensemble des optimisations sur cet échantillon complexe aurait nécessité plusieurs jours consécutifs d'analyse.

Le même traitement (somme des scores des spectres pour toutes les protéines identifiées) a été réalisé sur ces données. Les sommes des scores obtenues pour l'état de charge 2+ pour les ions possédant des rapports m/z compris entre 800 et 1100 sont présentées en figure IV-10. Les énergies de collision retenues pour les gammes 800-900, 900-1000, 1000-1100 sont respectivement 55, 58 et 59 eV.

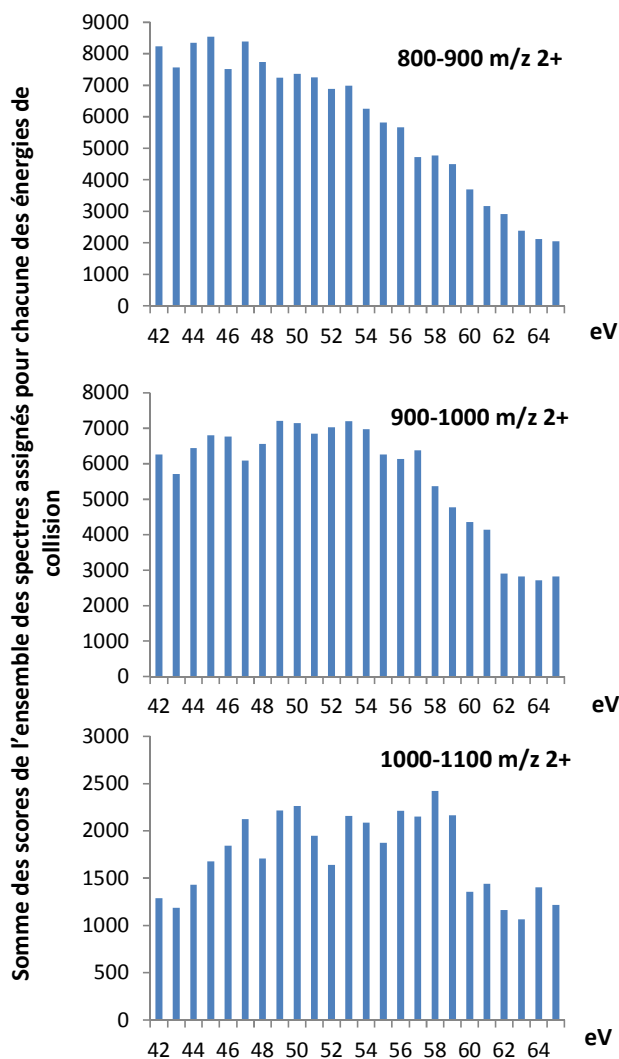


Figure IV-10 Somme des scores de l'ensemble des spectres assignés pour chacune des énergies de collision testées pour les peptides dichargés en fonction de la gamme de m/z
Plus la somme des scores est élevée, plus l'énergie de collision est adaptée pour la fragmentation des peptides pour une gamme de m/z et un état de charge donnés.

Les valeurs d'énergies de collision optimisées obtenues par état de charge (2+, 3+) sont résumées en figure IV-11. On constate que les énergies de collision sont légèrement inférieures pour les états de charge 3+ par rapport aux états de charge 2+, signifiant que les peptides de plus forte densité de charge nécessitent des énergies de collisions inférieures.

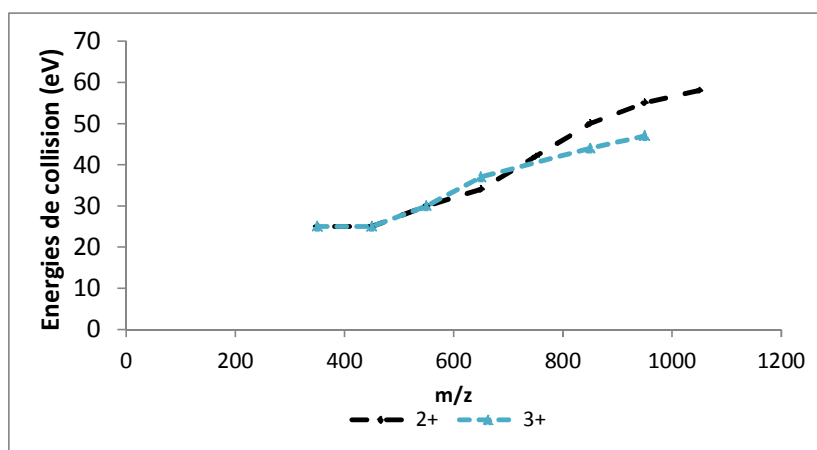


Figure IV-11 Valeurs d'énergies de collision optimisées pour les états de charge 2+ et 3+

Ces énergies de collision optimisées ont été intégrées aux futures méthodes d'acquisition nanoLC-MS/MS développées au laboratoire pour cet instrument.

B. Bilan des résultats obtenus suite à l'optimisation de méthodes d'acquisition nanoLC-MS/MS à partir d'échantillons de levure (*Saccharomyces cerevisiae*)

L'ensemble des paramètres présentés dans ce chapitre a été optimisé pour un maXis 4G (avant mise à jour en 4G Plus) pour un échantillon de digestat de levure (*Saccharomyces cerevisiae*, 1µg) élué sur un gradient de 80 minutes (1 à 35 % d'acétonitrile). Les paramètres des méthodes d'origine fournies par le constructeur à l'installation du spectromètre de masse comparés aux paramètres de la méthode à l'issue de nos optimisations sont présentés en tableau 2.

Tableau 2 : Paramètres d'origine et optimisés des méthodes d'acquisition nanoLC-MS/MS pour un échantillon constitué d'un microgramme de digestat de levure élué en 80 minutes, et analysé avec le couplage LC-MS nanoAcquity-Maxis 4G.

	Méthode constructeur	Méthode optimisée
Nombre de sommation MS	2500 (2,5 Hz)	2500 (2,5 Hz)
Nombre de spectres MS/MS générés à l'issue de chaque spectre MS (Top N)	6	6
Seuil d'intensité absolue de déclenchement de la fragmentation des ions	1500	5000
Temps d'exclusion	0,3 min	1 min
Nombre de sommations maximum pour les spectres MS/MS	7500 sommations (0,8Hz)	4688 sommations (1,5 Hz)
Seuil d'intensité correspondant au nombre de sommations maximum (A)	5000	5000
Nombre de sommations minimum pour les spectres MS/MS	312 (20 Hz)	694 (9 Hz)
Seuil d'intensité correspondant au nombre de sommations minimum (B)	250 000	80 000
Nombre de protéines identifiées (≤1% FDR)	358 ± 7	550 ± 1

Grâce à l'optimisation de l'ensemble des paramètres d'acquisition des méthodes nanoLC-MS/MS présentés dans ce chapitre, nous avons été capables d'identifier 50% de protéines supplémentaires à partir du même échantillon. On peut donc constater que les méthodes fournies par le constructeur n'étaient pas optimisées pour l'analyse d'échantillons très complexes comme l'est un microgramme de digestat de levure élué sur un système chromatographique en 80 minutes. L'ensemble de ces paramètres sont étroitement liés et devraient être évalués pour chaque nouvel échantillon et/ou problématique. En revanche, le temps alloué à ces optimisations, ainsi que la quantité de matériel requise, ne permet pas de l'envisager de façon aussi exhaustive et systématique. Un certain nombre de paramètres peuvent être établis à l'installation de l'instrument comme les énergies de collision

par exemple, ou les vitesses de scan MS ou MS/MS globales qui pourront servir de base ensuite pour le développement des prochaines méthodes d'acquisition. En revanche, le temps d'exclusion dynamique et le seuil d'intensité absolue pour le déclenchement de la MS/MS doivent être adaptés à chaque échantillon en fonction des conditions chromatographiques utilisées. Le gain obtenu pour le digestat de levure par rapport aux méthodes d'acquisition standards délivrées avec l'instrument illustre bien la nécessité de paramétrer finement les méthodes d'acquisition pour chaque type d'échantillon. On constate cependant que les vitesses d'acquisition du maXis 4G sont assez limitées puisque le nombre de protéines identifiées avec ces échantillons est loin d'être représentatif du protéome complet de la levure.

II. Evaluation et comparaison de deux sources Electrospray sur l'Impact HD (Bruker Daltonique)

A. Contexte

Une nouvelle source, développée par Microm Biosciences et aujourd'hui commercialisée par Bruker Daltonique, a été développée pour le spectromètre de masse Impact-HD (Bruker Daltonique) acquis au cours de ma thèse. Nous l'avons comparée à l'ancien type de source utilisée, nommé « nanoESI » dans la suite du chapitre, de façon à déterminer quelles étaient les avantages liés à son utilisation et ce notamment pour les stratégies d'analyse quantitatives de type XIC (extraction des courants d'ions MS).

L'instrument utilisé dans cette partie est le Q-TOF Impact-HD, commercialisé par la société Bruker Daltonique, couplé à un système de chromatographie liquide de type nanoUPLC (nanoAcquity Waters) équipé d'une colonne C18 BEH 250 × 0,075 mm (taille de particules : 1,7 µm de diamètre).

B. Description des sources

L'ancien type de source « nanoESI » et la « CaptiveSpray » sont représentées en figure IV-12.

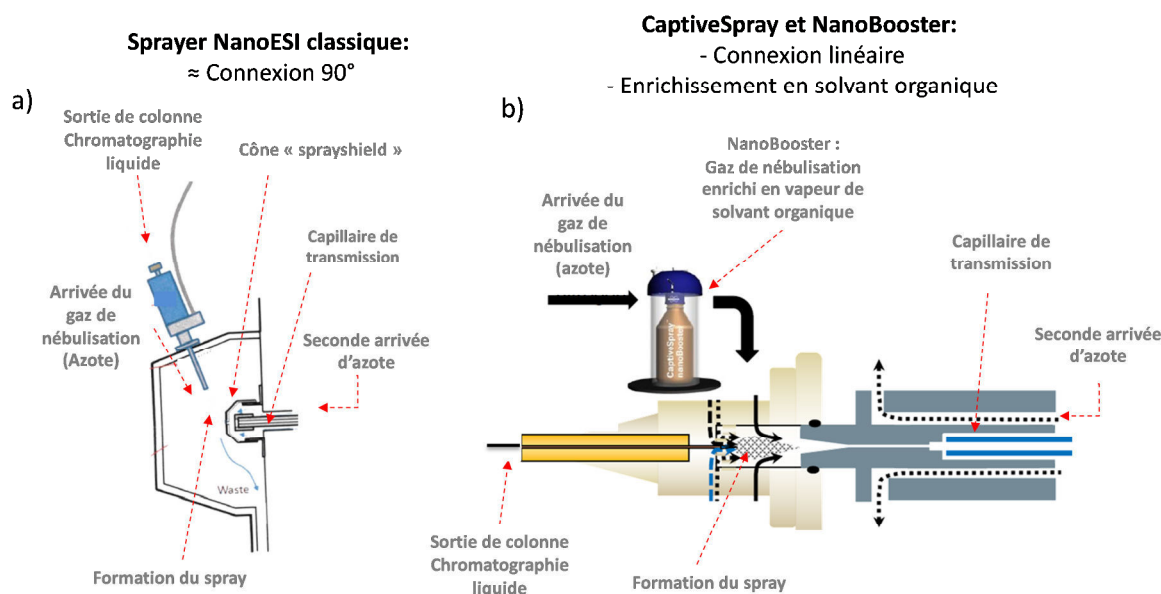


Figure IV-12 Schémas des sources "nanoESI" (a) et "CaptiveSpray" associée à un "NanoBooster" (b)

Les différences majeures entre ces deux sources résident dans l'angle formé entre le nébuliseur et l'entrée du spectromètre de masse ($\approx 90^\circ$ et 180° respectivement pour les sources « nanoESI » et « CaptiveSpray ») et dans l'implémentation d'un système d'enrichissement du gaz de nébulisation en vapeur de solvant organique appelé NanoBooster (Bruker Daltonique) pour la source « CaptiveSpray ».

Ces 2 types de sources fonctionnent sur le principe de l'électrospray décrit dans le chapitre I « Principe général de l'analyse protéomique par spectrométrie de masse » en page 19. L'ancien modèle de source, la source « nanoESI », est présenté en figure IV-12a. Ce type de source présente une inclinaison d'environ 90° entre le nébuliseur et l'entrée du spectromètre de masse. Cette caractéristique, largement utilisée chez les différents types de constructeurs (Agilent technologies, Waters) permet de dévier massivement les ions neutres (non chargés) de l'entrée du spectromètre de masse, ce qui limite l'encrassement des instruments et permet donc d'espacer leurs nettoyages.

La nouvelle source « CaptiveSpray », illustrée en figure IV-12b, présente une connexion linéaire entre le nébuliseur et l'entrée du spectromètre de masse. Elle est associée à un système d'enrichissement (appelé « NanoBooster », Bruker Daltonique) qui permet d'enrichir le gaz de nébulisation (azote) avec des vapeurs de solvant organique. Le solvant organique recommandé par le constructeur Bruker Daltonique pour l'analyse protéomique est l'acétonitrile. Un gain en termes de sensibilité a été effectivement constaté expérimentalement au sein de notre laboratoire, supposant une meilleure désolvatation des ions en source mais la description de ce mécanisme n'a pas, à notre connaissance, été décrite en détails.

C. Evaluation de la source « CaptiveSpray » équipée d'un « NanoBooster » pour l'analyse protéomique quantitative par rapport à la source « nanoESI ».

Dans cette partie, les deux sources ont été comparées à partir d'échantillons de référence en termes de sensibilité, justesse et linéarité.

C.1. Description des échantillons de référence

Les échantillons modèles utilisés pour réaliser cette évaluation ont été développés au sein de l'infrastructure PROFI dans laquelle est impliquée le LSMBO. Cette infrastructure nationale, qui regroupe trois laboratoires de protéomique en France (EDyP à Grenoble, IPBS à Toulouse et le LSMBO à Strasbourg) a plusieurs objectifs dont le développement de méthodes d'analyses protéomiques quantitatives ciblées ou globales et la formation d'autres laboratoires à l'analyse protéomique. Ainsi, des échantillons de référence ont été développés au sein de cette infrastructure dans le but de pouvoir comparer et étudier les différentes plateformes analytiques utilisées pour l'analyse protéomique. Ces échantillons sont constitués d'un digestat de levure (*Saccharomyces cerevisiae*) dans lequel a été ajouté différentes quantités d'un mélange équimolaire de 48 protéines humaines (UPS1, Sigma Aldrich). Le contenu des échantillons utilisés pour la comparaison des deux sources est présenté en tableau 3. Les différentes quantités d'UPS1 ajoutées au digestat de levure varient entre 250 amol et 25 fmol.

Tableau 3 : Contenu des échantillons ayant servi à la comparaison des deux sources. La totalité des échantillons a été injectée.

	<i>Saccharomyces cerevisiae</i>	Protéines du standard UPS1
<i>Echantillon A (non injecté sur le système couplé à la source « nanoESI »)</i>	1 µg	250 amol
<i>Echantillon B</i>	1 µg	500 amol
<i>Echantillon C</i>	1 µg	1 fmol
<i>Echantillon D</i>	1 µg	5 fmol
<i>Echantillon E</i>	1 µg	10 fmol
<i>Echantillon F</i>	1 µg	25 fmol

L'échantillon A n'a pas été analysé avec le système « NanoESI » classique en raison de la limite de détection de ce système qui n'est pas assez basse. Trois aliquotes étaient disponibles pour chacun des échantillons, ils ont tous été injectés une fois sur le système LC-MS couplé avec chacune des sources.

C.2. Evaluation des sources en terme de sensibilité

Le nombre de protéines d'UPS1 et de protéines de levure (*Saccharomyces cerevisiae*) identifiées à partir des analyses nanoLC-MS/MS acquises en mode DDA pour les échantillons de référence sont présentés en figure IV-13. Le traitement des données pour l'identification des protéines à partir des données spectrales est décrit dans le chapitre II « Stratégie d'analyse protéomique qualitative » en page 37. Les protéomes de l'Homme et de la levure sont connus (même si les banques de données existantes ne sont pas exhaustives en termes de variants de séquence).

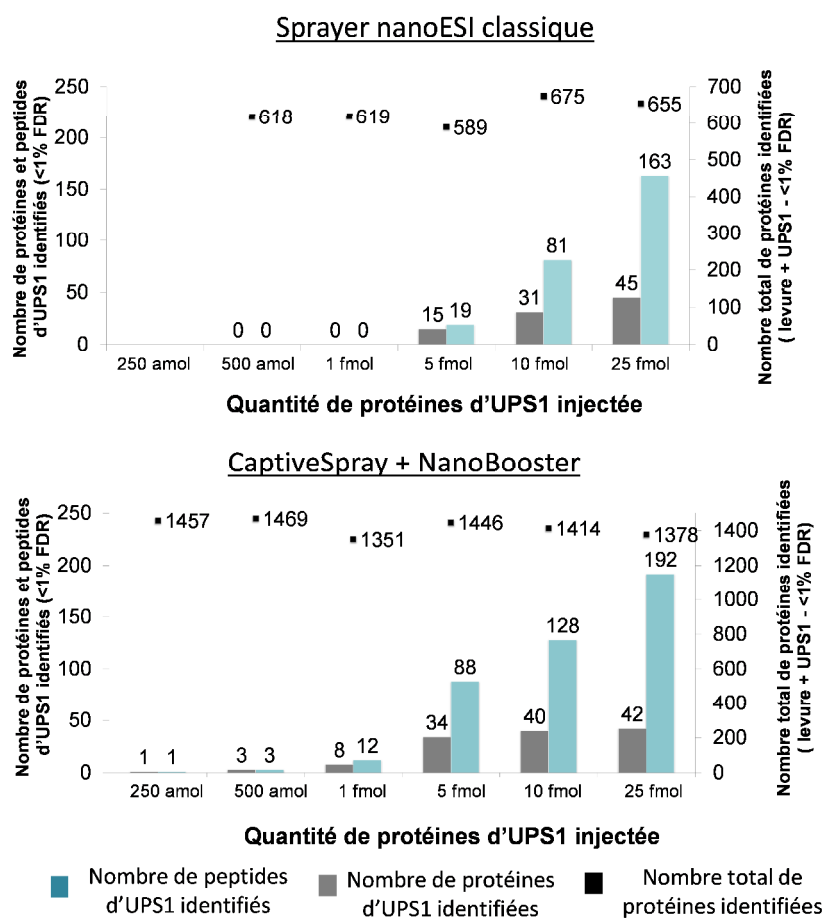


Figure IV-13 Nombre total de protéines identifiées et nombre de protéines et de peptides d'UPS1 identifiés pour la source CaptiveSpray (en haut) et la source « nanoESI » (en bas) (< 1 % FDR)

Le nombre total de protéines identifiées est présenté en rouge. Le nombre total de protéines et peptides d'UPS1 identifiés pour chacun des échantillons sont présentés en gris et bleu respectivement. Le nombre total de protéines identifiées avec la source « CaptiveSpray » (et le « NanoBooster ») pour l'ensemble des échantillons est deux fois plus important qu'avec la source « nanoESI ». Les quantités d'UPS1 les plus faibles détectées pour chacune des sources sont 250 amol et 5 fmol pour les sources « CaptiveSpray » et « nanoESI » respectivement.

On constate d'après les résultats présentés en figure IV-13 que la source « CaptiveSpray » (équipée du « NanoBooster ») permet de doubler le nombre total de protéines identifiées par rapport à la source « nanoESI ». Les quantités d'UPS1 les plus faibles détectées avec chacune des sources sont respectivement 250 amol et 5 fmol pour les sources « CaptiveSpray » et « nanoESI » dans cette

matrice complexe (1 µg de digestat de levure). Le nombre de peptides identifiés pour la « CaptiveSpray » est supérieur à celui de la source « nanoESI » pour chacune des quantités d'UPS1 ajoutées au digestat de levure, en revanche pour l'échantillon F, le nombre de protéines d'UPS1 est supérieur pour la source « nanoSpray ». Ce résultat est un peu inattendu dans la mesure où le nombre de peptide total est plus élevé avec la source « CaptiveSpray ». Plusieurs phénomènes pourraient être à l'origine de ce résultat : les capacités d'ionisation des peptides spécifiques à ces protéines sont différentes en fonction de la source utilisée, la sélection semi-aléatoire des ions pour la fragmentation donne lieu à des effets de sous-échantillonnage (décrits dans la première partie du chapitre) qui varient entre les deux configurations de source ou les critères de filtre des protéines identifiées à moins de 1 % de faux-positifs sont différents pour les sets de données qui sont de tailles très différentes (le nombre de spectres générés avec la « CaptiveSpray » est beaucoup plus élevé que pour la source « nanoESI »). Nous avons extrait les courants d'ions d'un peptide identifié uniquement avec la source « nanoESI » dans les deux sets de données (figure IV-14). Il est possible de constater que le signal du peptide concerné est plus élevé pour la source « CaptiveSpray » que pour la source « nanoESI », ce qui confirme un potentiel artefact d'identification de type sous échantillonnage ou validation statistique des spectres identifiés.

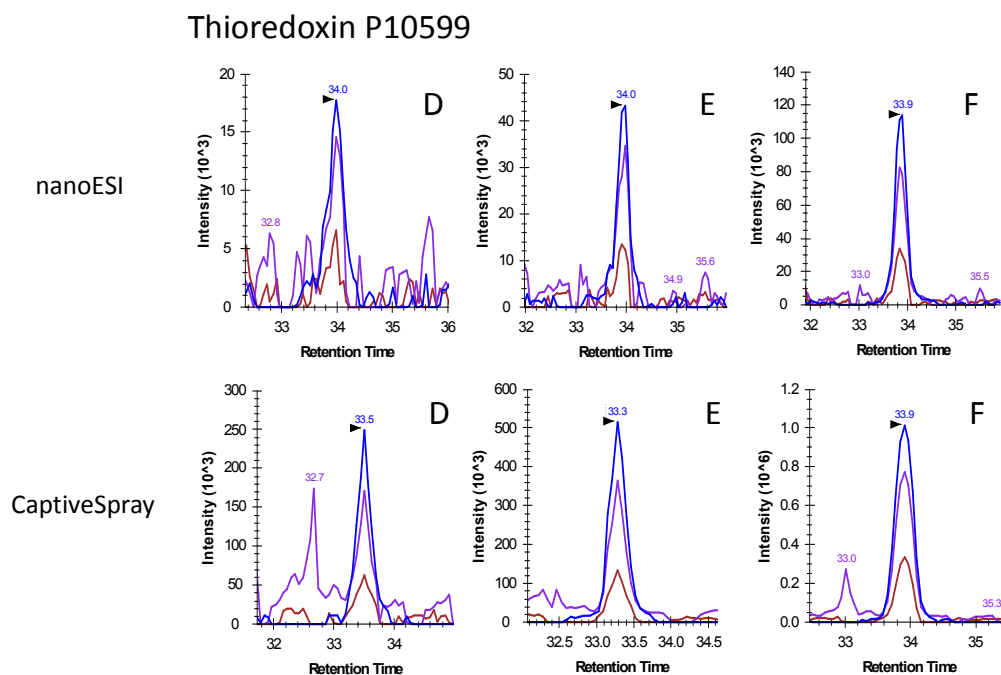


Figure IV-14 Courants d'ions extraits du peptide TAFQEALDAAGDK dans les analyses acquises avec les deux types de source

L'extraction des courants d'ions de ce peptide qui n'a été identifié que pour la source « nanoESI » permet de confirmer une hypothèse probable de sous-échantillonnage ou de différence de validation statistique des spectres entre les deux sets de données concernant la non-identification de ce peptide pour la source « CaptiveSpray ».

Etant donné le nombre total de protéines, ainsi que le nombre de peptides d'UPS1 identifiés pour chacun des échantillons, on peut conclure que la source « CaptiveSpray » équipée du « NanoBooster » permet de gagner en sensibilité et ainsi d'augmenter de façon significative la gamme dynamique de concentration du protéome étudiée par rapport à la source « nanoESI ».

Il est possible de constater que les 48 protéines d'UPS1 n'ont pas toutes été identifiées pour chacune des deux sources. Les protéines qui n'ont pas été identifiées sont des protéines ayant des masses molaires comprises entre 8000 et 20 000 Daltons pour lesquelles le nombre de peptides tryptiques est très réduit, ce qui explique qu'il est plus difficile de les détecter.

C.3. Evaluation des sources en terme de linéarité et de justesse

Bien que cette nouvelle source « CaptiveSpray » ait démontré une meilleure sensibilité que la source « nanoESI » classique en termes de nombre de protéines identifiées à partir de l'analyse d'un mélange complexe, son utilisation éventuelle pour des analyses quantitatives (type XIC ou DIA) ne pourrait être envisagée que si elle permet l'obtention de résultats linéaires (quantité de protéines analysée proportionnelle à l'intensité des signaux mesurés). Dans un premier temps, nous avons donc établi des courbes de linéarité (intensité des signaux détectés par peptide en fonction des quantités d'UPS1) pour évaluer la gamme dynamique d'analyse de la nouvelle source, puis nous avons vérifié que les abondances relatives mesurées pour chaque protéine entre deux échantillons donnés étaient équivalentes en fonction de la source utilisée et correspondaient aux valeurs attendues (justesse).

Cette comparaison a été réalisée d'après l'extraction des courants d'ions de 43 protéines d'UPS1 identifiées avec un taux de faux-positifs inférieur à 1% à partir des analyses nanoLC-MS/MS acquises pour les échantillons (A), B, C, D, E et F pour chacune des configurations de source.

C.3. 1. Evaluation des sources en termes de linéarité

Le traitement des données effectué pour cette évaluation, à l'aide du logiciel Skyline, correspond à la réalisation des étapes suivantes (décrites dans le chapitre « Stratégies d'analyses protéomiques quantitatives » en page 68) :

1. la soumission des spectres issus des analyses nanoLC-MS/MS à l'algorithme de recherche Mascot

2. la création d'une librairie spectrale à partir des spectres dont l'assignation a été validée à moins de 1 % de résultats « faux-positifs »
3. l'extraction des courants d'ions de chaque peptide pour lequel un spectre est présent dans la librairie spectrale, par le logiciel Skyline, dans toutes les analyses à comparer.

Cette évaluation de linéarité a été réalisée d'après les valeurs peptidiques et non protéiques. De plus, les aires des peptides n'ont pas été normalisées et les intégrations des courants d'ions extraits réalisées par le logiciel n'ont été validées manuellement que pour les protéines d'UPS1. Pour cette évaluation, deux états de charge détectés pour un seul peptide seront considérés comme deux entrées peptidiques différentes. La comparaison a porté sur 243 entrées peptidiques différentes.

Deux critères ont été mis en place pour comparer les échantillons analysés avec la source « NanoESI » et les échantillons analysés avec la source « CaptiveSpray » équipée du « NanoBooster »:

- Critère sur les coefficients de variation : les peptides pour lesquels les coefficients de variation, calculés d'après les aires sous la courbe des courants d'ions intégrés entre les trois réplicas analysés par échantillon, étaient supérieurs à 20 % ont été éliminés.
- Critère sur l'« idot product » (produit scalaire isotopique) : L'« idot product » est présenté dans le chapitre « Stratégies d'analyse protéomique quantitatives » en page 68. Les peptides pour lesquels cette valeur était inférieure à 0,8 ont été éliminés.

Le nombre de peptides validés selon ces critères est présenté en tableau 4.

Tableau 4 : Nombre de peptides satisfaisant les critères de coefficients de variation ($\leq 20\%$) et de valeur d'« idot product ». Ces valeurs correspondent aux courants d'ions extraits pour l'ensemble des peptides validant les critères énoncés ci-dessus pour les échantillons A, B, C, D, E et F.

	250 amol	500 amol	1 fmol	5 fmol	10 fmol	25 fmol
« NanoESI »	-	5	1	105	171	196
« CaptiveSpray »	27	18	34	126	177	199

Le nombre de peptides satisfaisant ces critères est nettement plus élevé avec la source « CaptiveSpray » pour tous les échantillons pour lesquels moins de 10 fmol d'UPS1 ont été ajoutés.

Le bénéfice de l'utilisation de cette nouvelle source est donc très net pour la détection de protéines en faibles concentrations.

Les aires sous les courbes (somme de P, P+1, P+2) de tous les peptides qui validaient les critères présentés ci-dessus ont été représentées graphiquement en fonction de la quantité injectée d'UPS1 pour l'ensemble des peptides détectés. Un exemple de courbe de linéarité obtenue pour le peptide CCAAADPHECYAK avec chacune des sources est présenté en figure IV-15.

NanoESI Sprayer

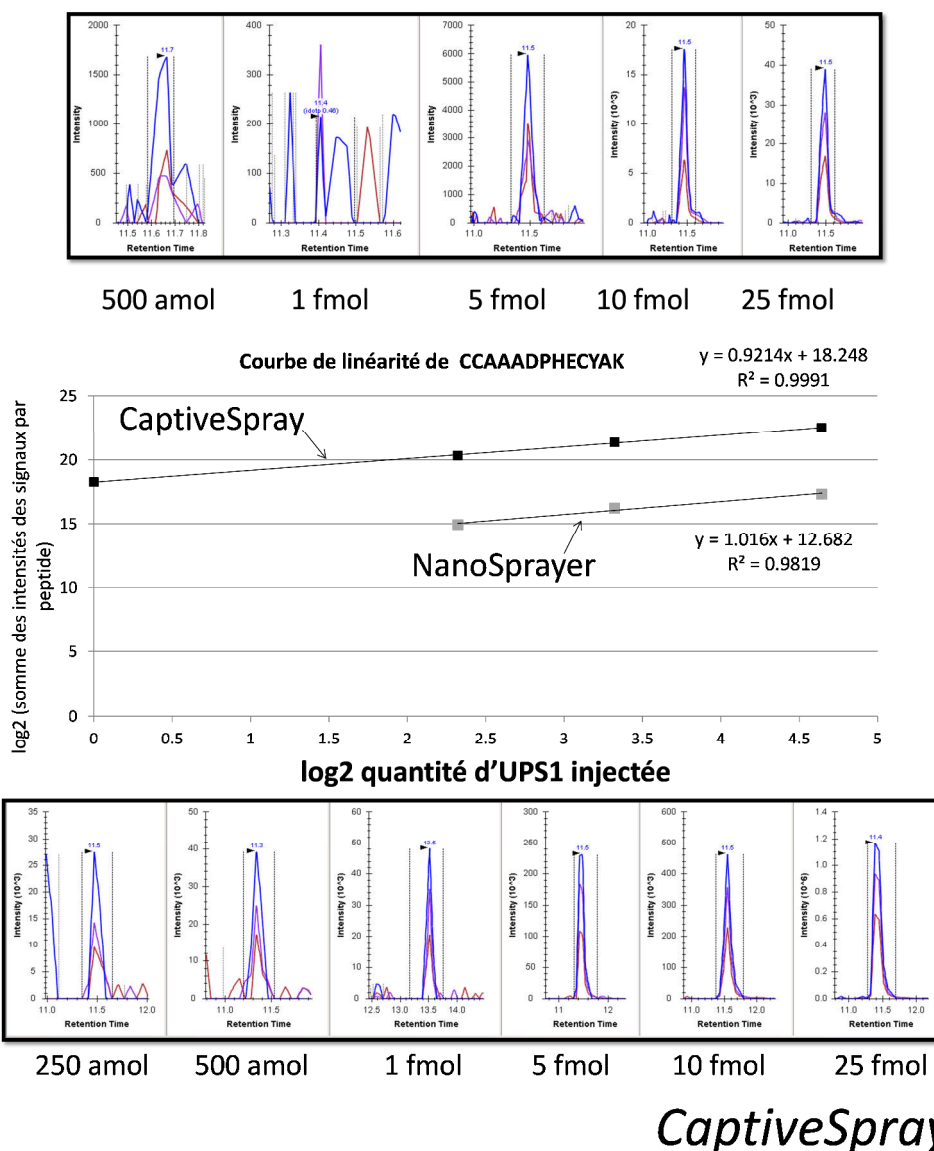


Figure IV-15 Courbes de linéarité et extractions des courants d'ions pour le peptide CCAAADPHECYAK par le logiciel Skyline pour chacune des configurations de source

Selon les critères mis en place pour cette évaluation, la gamme dynamique d'analyse de ce peptide est plus étendue avec la source « CaptiveSpray » qu'avec la source « nanoESI ».

Il est intéressant de constater que les pentes des courbes de linéarité de ce peptide sont très similaires pour chacune des sources. Ce phénomène a été observé pour la plupart des peptides détectés. Ainsi, la source « CaptiveSpray » permet d'obtenir le même type de linéarité de réponse que la source « NanoSpray » mais avec une meilleure sensibilité (étant donné le nombre plus élevé de peptides validant les critères spécifiés et le nombre de protéines identifiées avec chacune des sources). On constate également qu'un très grand nombre de peptides possèdent des coefficients de variations inférieurs à 20 %. Cette valeur seuil de 20 % est validée dans la littérature pour la stratégie ciblée SRM [92]. De ce fait, on peut également choisir ce seuil de validation pour la stratégie XIC qui n'est pas aussi spécifique. Cela signifie que la préparation des échantillons a été parfaitement similaire entre les différents aliquotes et que le couplage LC-MS était stable pendant cette série d'analyse. On peut donc extraire des informations quantitatives fiables à partir de ces données.

Selon les critères de validation établis précédemment, la linéarité de réponse de ce peptide s'étale sur environ un ordre de grandeur (les courants d'ions à 0,250 fmol et 0,500 fmol ne validaient pas les critères définis). Cette gamme est peut-être plus étendue mais ce set de données ne permet pas de le vérifier. Les critères de validation de la linéarité ne sont pas aussi stricts que ceux utilisés pour le projet de développement de méthode d'analyse ciblée pour la quantification de la vitellogénine dans le plasma de tortue Luth, présentée dans la partie résultats en chapitre VI. La stratégie XIC n'est pas aussi précise que la stratégie SRM étant donné la complexité des spectres MS à partir desquels sont extraits les courants d'ions par rapport à celle des signaux enregistrés pour une combinaison de transitions spécifiques analysées en mode SRM. De plus, pour ce type d'analyse quantitative globale, dont les résultats devront être justement validés par une technique orthogonale telle que la SRM, aucun standard n'est utilisé. De ce fait, les linéarités de réponses sur la gamme dynamique analysée des peptides quantifiés ne sont, en général, pas vérifiées pour ce type d'approche.

C.3. 2. Evaluation des sources en termes de justesse

Dans cette partie, les sources seront évaluées en termes de justesse. Tout d'abord, nous allons tester différents modèles de normalisation pour les données générées avec la source « CaptiveSpray ». Ensuite, nous évaluerons les abondances relatives des protéines obtenues pour chacune des sources entre deux échantillons/conditions (E et F pour lesquels respectivement 10 et 25 fmol d'UPS1 ont été ajoutées au microgramme de levure) pour un modèle de normalisation donné.

Pour réaliser cette comparaison, les données ont été traitées avec le logiciel de traitement MSstat (<http://www.msstats.org/>). Cet outil a été développé par le laboratoire d'Olga Vitek [134] pour le

traitement statistique des données de protéomique quantitative générées en mode SRM, XIC ou en mode DIA. L'ensemble de ces stratégies sont décrites dans le chapitre III « Stratégies d'analyse protéomique quantitatives ». Cet outil, disponible au sein du logiciel R (<http://www.r-project.org/>), propose des outils de normalisation (par quantile ou selon le modèle de l'égalité des médianes) et un outil statistique qui permet d'évaluer la significativité des niveaux d'expression d'une ou plusieurs protéines entre deux (ou plusieurs) conditions. Le test réalisé correspond à une analyse de variance basée sur un modèle linéaire à effets mixtes : ce modèle tient notamment compte des différents types de répliques réalisés (techniques ou biologiques). Les différentes données acquises pour une protéine (multiples états de charge, intensités du profil isotopique P, P+1, P+2) sont considérées comme des répliques de mesures de cette même protéine. Classiquement, l'obtention d'une valeur « p » protéique inférieure à un seuil donné (au maximum établi à 0,05), signifie que l'abondance relative d'une protéine entre les conditions comparées est significativement différentielle.

C.3.2.1. Choix du modèle de normalisation à partir des données générées avec la source « CaptiveSpray »

Les abondances relatives des protéines d'UPS1 entre les échantillons/conditions E et F, ainsi que leur significativité, ont été comparées sans normalisation, après que l'ensemble des données ait été normalisé par quantile et selon le modèle de l'égalité des médianes. Ces deux modèles de normalisation sont présentés dans le chapitre III « Stratégies d'analyse protéomique quantitatives » en page 68). La significativité des différences en termes d'abondance relative des protéines d'UPS1 entre les conditions E et F (soit 10 et 25 fmol d'UPS1) a été évaluée par le test statistique proposé par MSstat. Chacune des protéines pour lesquelles des courants d'ions ont été extraits a été testée. Les résultats de la comparaison des protéines d'UPS1 entre les échantillons E et F sont présentés sous la forme de graphique de type « Volcano » (figure IV-16). Ils représentent la valeur $-\log_{10}(p)$ calculée pour chaque protéine en fonction des abondances relatives, exprimées en \log_2 , des protéines entre les conditions E et F. A droite de chacun de ces graphiques sont présentées les zones agrandies pour $p < 0,01$ et une abondance relative supérieure à un facteur 2 ($\log_2(2) = 1$) entre les deux conditions. Les données noires et rouges correspondent respectivement aux protéines de levure et aux protéines d'UPS1.

Le contenu des échantillons comparés est connu : les deux conditions contiennent chacune 1 μg de digestat de levure et 10 ou 25 fmol de protéines d'UPS1 dans les conditions E et F. Le ratio correspondant à l'abondance relative des protéines entre les deux conditions est donc de 1 ($\log_2(1) = 0$) pour les protéines de levure et de 2,5 ($\log_2(2,5) = 1,3$) pour les protéines d'UPS1.

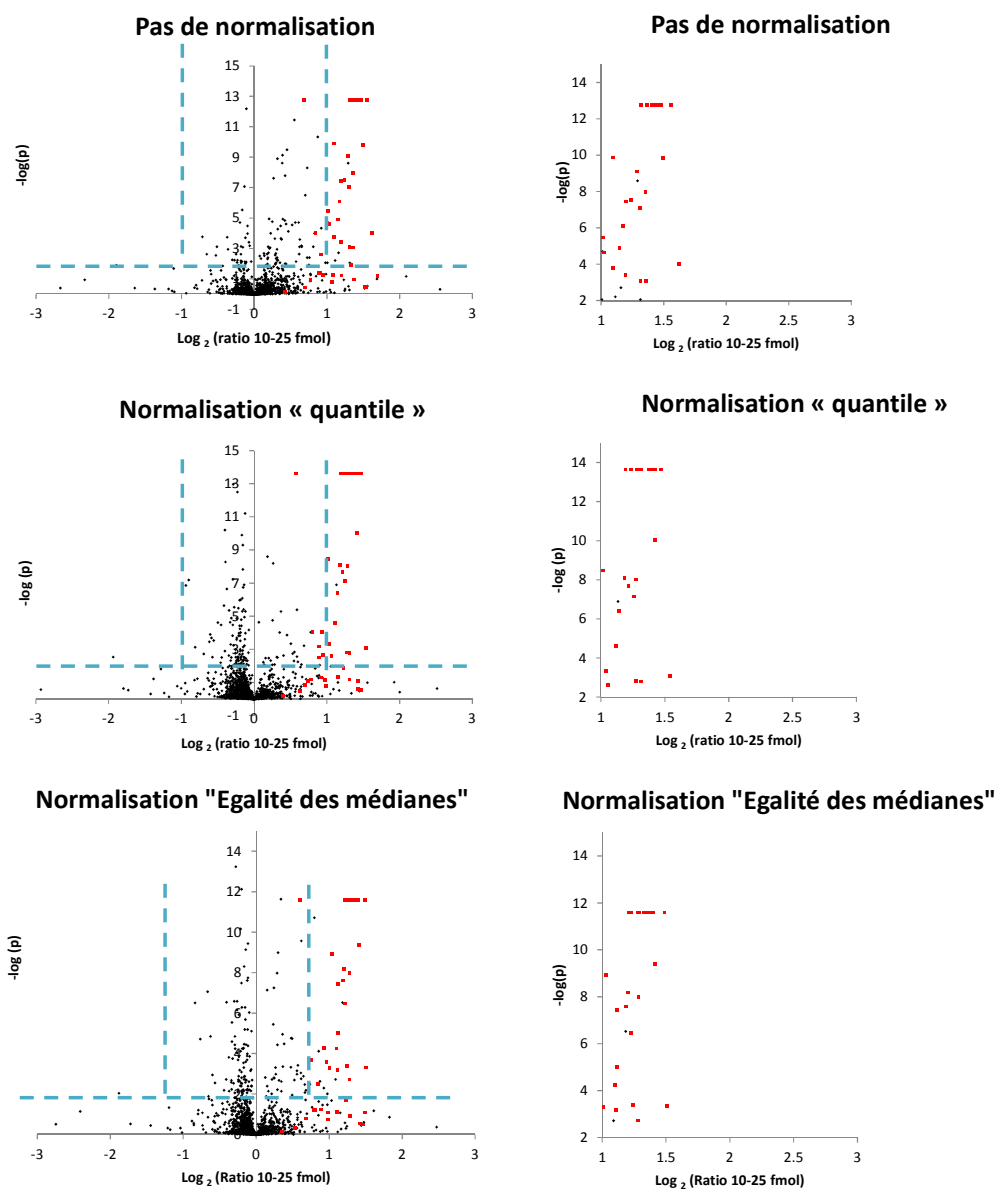


Figure IV-15 Graphiques de type « Volcano » permettant de comparer les abondances relatives des protéines de levure et d'UPS1 entre les échantillons /conditions E et F

Le graphique présente en ordonnée la valeur $-\log(p)$ obtenue après le test statistique basé sur une analyse de variance, proposé par le logiciel MSstat en fonction des abondances relatives des protéines d'UPS1 mesurées entre les échantillons/conditions E et F.

Le seuil de significativité pour cette analyse est établi à $p < 0,01$ (soit $-\log(p) > 2$). Les protéines de levure (en noir) qui possèdent une valeur p inférieure à 0,01 sont donc considérées comme des résultats faux-positifs car la quantité de protéines de levure au sein de chacun des échantillons est identique (soit 1 μg de digestat). Au contraire, les protéines d'UPS1 (en rouge) pour lesquelles la valeur p est supérieure à 0,01, sont des résultats qui peuvent être qualifiés de « faux-négatifs » puisqu'ils ne présentent pas de niveaux d'intensités statistiquement différentiels alors qu'ils le devraient.

L'abondance relative des protéines d'UPS1 entre les deux types d'échantillons (10 et 25 fmol) est en accord avec la valeur attendue ($\log_2(2,5) = 1,3$). En effet, environ 90% des protéines d'UPS1 présentent une abondance relative comprise entre un facteur 1,7 ($\log_2(1,7) = 0,8$) et 3 ($\log_2(3,03) = 1,6$) alors qu'un facteur 2,5 est attendu.

Environ 10 % des protéines de levure possèdent une valeur p inférieure à 0,01, ce qui représente plus de cent protéines alors que nous savons que la quantité de levure dans les échantillons E et F est la même. Il est indispensable, en plus du test statistique et du seuil pour la valeur p, d'établir un ratio de sur/ ou sous-expression minimum au-delà ou en deçà duquel on considère que les protéines présentent effectivement des abondances relatives significativement différentielles. Les stratégies de quantification sans marquage ne présentent pas une reproductibilité très élevée (un seuil de 20 % de coefficient de variation est communément admis pour deux répliques biologiques), ce seuil doit donc être au minimum supérieur à 1,2. Nous avons fixé ce seuil à 2 ($\log_2(2) = 1$) étant donné que le ratio attendu correspondant à l'abondance relative des protéines entre ces deux conditions est de 2,5.

Le nombre de protéines pour lesquelles les niveaux d'expression évalués au sein de cette stratégie sont des « faux-négatifs » (protéines d'UPS1 identifiées comme non différentielles), « faux-positifs » (protéines de levures différentielles par rapport au nombre total de protéines différentielles), ainsi que les « vrais-positifs » (nombre de protéines d'UPS1 détectées comme différentielles) sont présentés en tableau 5.

Tableau 5 : Nombre de résultats « faux négatifs », « faux positifs » et « vrais positifs » pour la comparaison des échantillons E et F pour les trois types de traitement des données générées avec la source « CaptiveSpray » équipée du « NanoBooster » (sans normalisation, normalisation par quantile et selon l'égalité des médianes).

Traitement	Protéines levure différentielles p < 0,01 Ratio > 2 « Faux-positifs »	Protéines UPS1 différentielles p < 0,01 Ratio > 2 « Vrais-positifs »	Taux de faux positif (protéines de levure différentielles/ (protéines de levure différentielles + protéines d'UPS1 différentielles)	Protéines UPS1 non-différentielles p > 0,01 « Faux-négatifs »
Sans normalisation	6	26	19%	14
Normalisation « Egalité des médianes »	3	24	11%	14
Normalisation « quantile »	2	23	8%	14

On constate que les deux types de normalisation donnent des résultats plutôt similaires et que le taux de résultats « faux-positifs » est effectivement réduit grâce à la normalisation des données. Dans la suite du traitement, nous avons choisi d'utiliser la normalisation par quantile car elle présente le taux de « faux-positifs » le plus bas. Pour cette étude, les intégrations des courants d'ions des protéines de levure n'ont pas été validées manuellement, or les protéines de levure qui ont donné lieu aux deux résultats « faux-positifs » (pour l'ensemble de données normalisées par le modèle quantile sélectionné) présentaient des signaux peu intenses ou des intégrations erronées qui auraient été éliminés lors d'une validation manuelle. Par exemple, deux peptides avaient été détectés pour la protéine de levure P13382, qui présente selon le traitement une abondance statistiquement différentielle entre les conditions E et F. Le signal du premier peptide était très proche du signal du bruit de fond. Ce peptide aurait été éliminé manuellement des données puisqu'il ne permettait pas de réaliser une quantification robuste. Une erreur d'intégration des courants d'ions du second peptide a donné lieu au résultat « faux-positif » (illustré en figure IV-16). Cette mauvaise intégration aurait également été corrigée dans le cadre d'une validation manuelle des courants d'ions intégrés. De plus, dans cette étude, l'égalité des variances n'a pas été vérifiée (elle est optionnelle dans le logiciel MSstat), or cette vérification aurait probablement également éliminé le résultat « faux-positif ». Ce résultat « faux-positif » illustre la nécessité de valider, et de corriger si besoin, les intégrations réalisées de façon automatique par les logiciels de traitement. La deuxième protéine de levure qui présentait un résultat « faux positif » était également sujette à des erreurs d'intégrations de courants d'ions ou à des signaux de trop faible intensité. En conclusion, la validation manuelle de ces données aurait permis d'éliminer les résultats « faux-positifs ».

Erreur d'intégration du pic

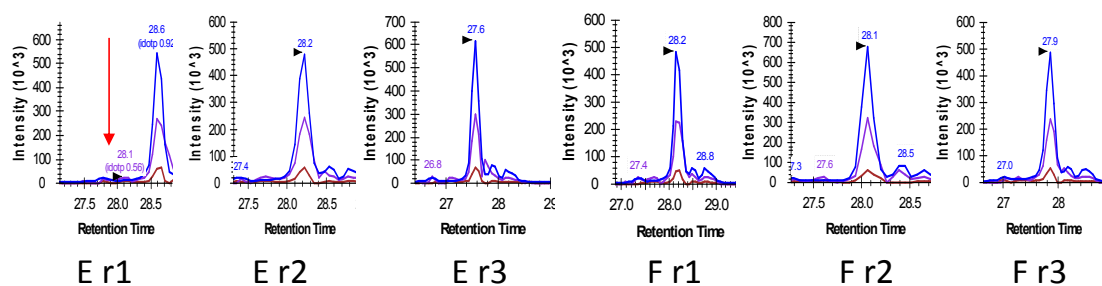


Figure IV-16 Illustration d'une erreur d'intégration d'un courant d'ion pour un peptide donné ayant donné lieu à un résultat faux-positif

L'erreur d'intégration des courants d'ions dans l'analyse E (réplica 1) a conduit à un résultat faux-positif.

Concernant les résultats « faux-négatifs », les protéines d'UPS1 dont l'abondance relative n'est pas ressortie significativement différentielle entre les groupes possédaient une valeur p très proche des

seuils imposés (5 protéines d'UPS1 possédaient une valeur p inférieure à 0,01 mais une abondance relative comprise entre un facteur 1,7 et 2).

Le traitement effectué sur ces échantillons permet donc de constater qu'il est possible de discriminer deux populations de protéines avec une abondance relative au moins supérieure à un facteur deux avec des données générées au travers de ce type de quantification (XIC) sur les analyseurs de type Q-TOF (Bruker Daltonique) équipés d'une source « CaptiveSpray » (et le « NanoBooster ») à condition de valider manuellement les intégrations réalisées par le logiciel de traitement.

Une autre stratégie, permettant de réduire le taux de résultats « faux-positifs », consiste à augmenter le nombre de réplicas analysés par condition. Augmenter le nombre de réplicas permet de minimiser l'impact des valeurs aberrantes (« outliers »). L'étude d'un plus grand nombre de réplicas, et la validation manuelle de l'ensemble des intégrations, auraient donc permis de renforcer nettement la confiance dans les résultats. Dans cette étude, les abondances relatives des peptides quantifiés pour chacune des protéines n'ont pas été évaluées. Cette évaluation aurait également permis d'élever la confiance dans les données quantitatives générées.

On peut donc conclure que cette stratégie permet de discriminer des protéines variant au minimum d'un facteur 2 entre deux conditions. Il faut cependant garder à l'esprit que le pourcentage de « faux-positif » peut rester élevé avec ce type d'approche et qu'il est nécessaire de réaliser un traitement des données rigoureux (tel que la validation manuelle de l'ensemble des intégrations peptidiques). De plus, les abondances relatives des protéines entre les deux conditions sont peu précises. En effet, les stratégies « label free » (sans marquage) ne permettent pas de déterminer précisément les abondances relatives des protéines. Ces erreurs sont notamment dues à la complexité des échantillons et à la présence d'interférences au sein des courants d'ions extraits. Ces interférences correspondent à la présence d'autres peptides qui ont une masse très proches (voire la même) que les peptides quantifiés. Ainsi, plus la résolution de l'instrument utilisé sera élevée, plus le risque d'observer des interférences sera réduit. L'utilisation de spectromètres de masse avec des capacités résolutive plus importantes permet donc d'affiner l'abondance relative mesurée d'une protéine entre deux conditions.

En conclusion, la significativité des résultats obtenus avec ce type d'approche globale qui permet de quantifier plusieurs centaines de protéines simultanément devra être validée à l'aide d'une stratégie quantitative plus précise comme la SRM par exemple ou une autre approche orthogonale.

C.3.2.2. Comparaison des sources

Les résultats obtenus pour la comparaison des échantillons 10 et 25 fmol d'UPS1 acquis en configuration « CaptiveSpray » présentés précédemment ont été comparés aux données obtenues pour les mêmes échantillons avec l'utilisation de la source « nanoESI ». Les graphiques de type « Volcano » des comparaisons de ces deux échantillons sont présentés en figure IV-17 pour chacune des sources. Les données ont été normalisées par le modèle quantile.

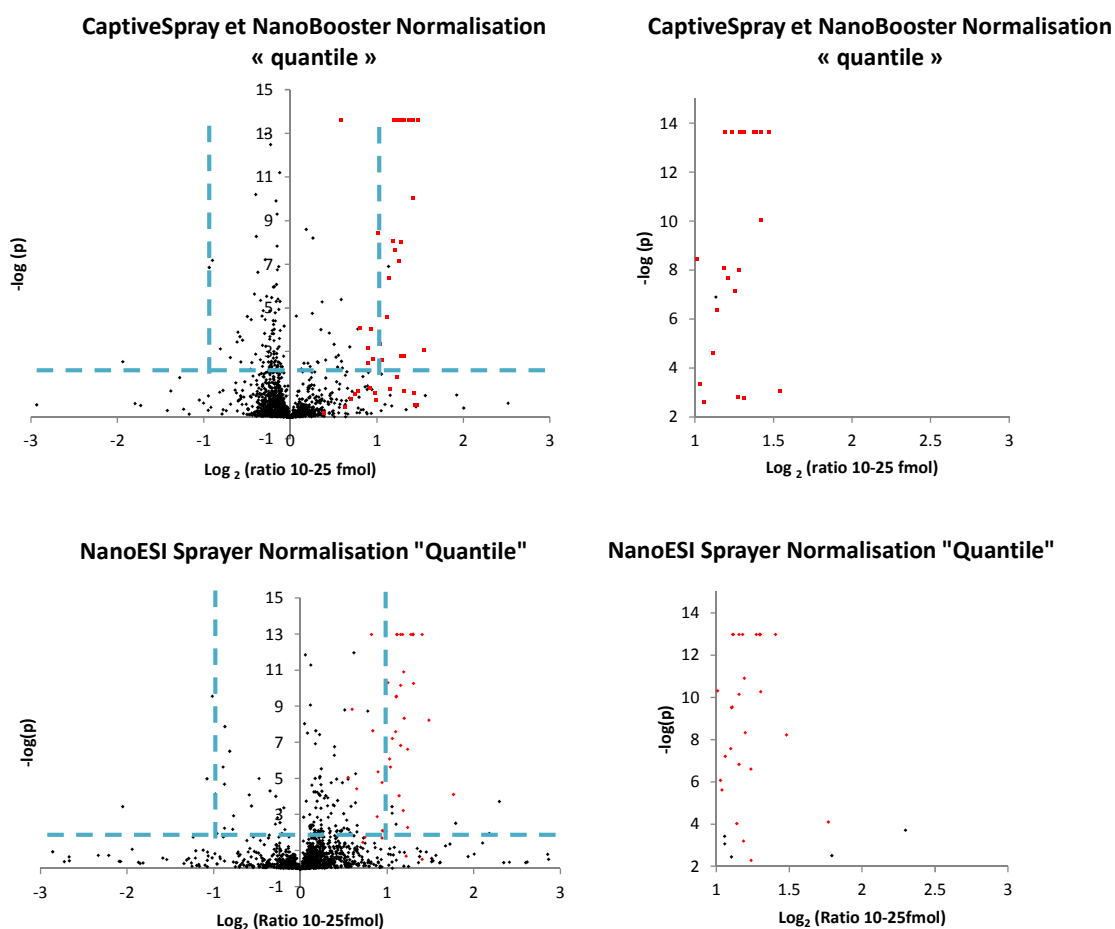


Figure IV-17 Graphiques de type « Volcano » permettant de comparer les abondances relatives des protéines de levure et d'UPS1 entre les échantillons contenant 10 et 25 fmol d'UPS1 pour les deux types de source

Le graphique présente en ordonnée la valeur $-\log(p)$ obtenue après le test statistique basé sur une analyse de variance proposé par le logiciel MSstat pour chaque protéine entre les deux conditions en fonction des abondances relatives des protéines entre les deux conditions/échantillons E et F étudiés. Les données ont été normalisées par quantile.

Les résultats obtenus avec les deux configurations de source semblent comparables. Le nombre de protéines identifiées en fonction des critères de significativité présentés précédemment est présenté en tableau 6. L'abondance relative des protéines d'UPS1 entre les deux groupes (10 et 25 fmol) est en accord avec la valeur attendue ($\log_2(2,5) = 1,3$). Comme pour la source « CaptiveSpray », environ

90% des protéines d'UPS1 présentent une abondance relative comprise entre un facteur 1,7 ($\log_2(1,7) = 0,8$) et 3 ($\log_2(3,03) = 1,6$) alors qu'un facteur égal à 2,5 est attendu.

Les nombres de protéines qui sont des résultats « faux-négatifs » (protéines d'UPS1 identifiées comme non différentielles), « faux-positifs » (protéines de levures différentielles par rapport au nombre total de protéines différentielles), ainsi que les « vrais-positifs » (nombre de protéines d'UPS1 détectées comme différentielles) sont présentés en tableau 6 pour chacune des sources.

Tableau 6 : Nombre de résultats « faux-négatifs », « faux-positifs » et « vrais-positifs » pour la comparaison des échantillons E et F pour les deux sources.

Traitement	Protéines levure différentielles $p < 0,01$ Ratio > 2 Faux-positifs	Protéines UPS1 différentielles $p < 0,01$ Ratio > 2 Vrai-positifs	Taux de faux positif (protéines de levure différentielles)/ (protéines de levure différentielles + protéines d'UPS1 différentielles)	Protéines UPS1 non-différentielles $p > 0,01$ Faux-négatifs
« CaptiveSpray »	2	23	8%	14
« NanoESI »	8	28	22%	5

Il est possible de constater que le nombre de résultats « faux-négatifs » est plus élevé avec la source « CaptiveSpray » qu'avec la source « nanoESI ». La comparaison statistique réalisée dans le logiciel MSstat tient compte du nombre de protéines comparées simultanément selon la correction de Benjamini et Hochberg [134]. Or, pour cette comparaison le nombre de protéines comparées est double pour la source « CaptiveSpray », ainsi les critères de significativité sont plus difficiles à atteindre, ce qui explique ce taux de résultats « faux-négatifs » plus élevé.

En revanche, on constate que le taux de faux-positifs est maintenu plus bas avec la source « CaptiveSpray » qui, en plus, est plus sensible puisqu'elle a permis d'identifier le double de protéines par rapport à la source « nanoESI ». Ces résultats valident donc largement l'utilisation du nouveau type de source « CaptiveSpray » pour la réalisation d'approche quantitative comme la stratégie XIC.

III. Conclusion

Nous avons pu démontrer dans ce chapitre l'importance de l'optimisation fine des paramètres d'acquisition en nanoLC-MS/MS sur le nombre de protéines identifiées au sein d'un même échantillon. Il a également été établi que la nouvelle source « CaptiveSpray » équipée du « NanoBooster » présente un net gain en sensibilité par rapport à l'utilisation du « nanoESI » classique. Le traitement des données XIC effectué sur des échantillons connus a permis de montrer que la nouvelle source « CaptiveSpray », couplée au spectromètre de masse de type Q-TOF, permet la génération de données quantitatives de qualité pour la protéomique.

Chapitre V. Applications et développement de la stratégie quantitative globale XIC à des échantillons biologiques complexes

Chapitre V : Applications et développement de la stratégie quantitative globale XIC à des échantillons biologiques complexes.

I. Implication dans le consortium NOTOX

A. Contexte

La vente des produits cosmétiques dont la toxicité des différents composants utilisés a été testée sur les animaux est interdite en Europe depuis le 11 mars 2013. Il y a donc un besoin urgent de modèles alternatifs à ces tests de toxicité. Le consortium NOTOX, inclus dans l'initiative européenne SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing), a pour but le développement de modèles biologiques pour le développement et la validation d'outils prédictifs de toxicité. Le LSMBO a été impliqué dans le développement des approches protéomiques au sein de ce consortium dans le but d'identifier des signatures protéiques de toxicité à partir d'un modèle hépatique donné et dans des conditions contrôlées. Ces signatures protéiques permettront de modéliser les voies d'hépatotoxicité en réponse à une exposition chronique à une molécule donnée et contribueront à l'établissement d'un modèle de prédiction.

Le modèle sélectionné pour cette étude correspond à une lignée cellulaire hépatique. Le foie, carrefour métabolique de l'organisme, est, en effet, impliqué dans des processus physiologiques majeurs tel que la détoxification des xénobiotiques (molécules étrangères à un organisme mais non issues de l'alimentation) [135]. Le modèle classique de cellules hépatiques en culture, considéré comme le « gold standard », consiste à utiliser des cellules hépatocytes primaires (qui représentent 80 % des cellules hépatiques) car celles-ci expriment les gènes impliqués *in vivo* dans le cas d'hépatotoxicité chronique [136]. Ces cellules proviennent de greffons humains non transplantables. Leurs principales limites résident dans leur disponibilité aléatoire et leur importante variabilité d'un individu à l'autre [137]. De nombreux modèles se rapprochant de ce type d'échantillons ont donc été développés : il existe des lignées cellulaires capables d'être cultivées pendant plusieurs semaines et qui expriment les principales fonctions des cellules hépatocytes [137]. Dans l'idéal, les cellules à l'origine de ces lignées cellulaires proviendraient de tissus sains mais leur disponibilité est limitée, les cellules proviennent donc à l'origine souvent de donneurs malades. Parmi celles-ci, la lignée HepaRG, provient d'une femme infectée par le virus de l'hépatite B et présentant des tumeurs hépatiques. Le prélèvement des cellules hépatiques à l'origine de cette lignée cellulaire a été réalisé en 1999. Cette lignée est composée d'hépatocytes et exprime la plupart des gènes impliqués dans les processus de

toxicité chronique [136]. Ces cellules sont commercialisées et ont la possibilité d'être produite en continu par la société Biopredic. Elles ont donc été sélectionnées pour la réalisation de cette étude par le consortium NOTOX.

Ces cellules peuvent être cultivées selon deux types de culture : bidimensionnelle ou tridimensionnelle. Ces dernières sont plus proches des réelles conditions physiologiques mais sont plus difficiles à mettre en œuvre [138]. Dans un premier temps, les signatures moléculaires d'hépatotoxicité chronique seront recherchées dans le protéome intra- et extracellulaire de ces cellules cultivées dans une configuration bidimensionnelle après leur exposition à une molécule connue pour induire des effets toxiques. Dans un deuxième temps, la même étude sera réalisée pour des cellules cultivées dans une configuration tridimensionnelle.

Dans ce contexte, et au sein du consortium NOTOX, j'ai réalisé une analyse protéomique quantitative sur des cellules hépatiques (HepaRG, culture 2D) après une exposition chronique à l'acide valproïque (VPA), principe actif utilisé contre l'épilepsie et la migraine connu pour provoquer une stéatose hépatique (accumulation de lipides dans le foie qui peut être fatale) chez l'homme. Cette molécule fait office de référence dans le cadre d'études d'hépatotoxicité [139]. Ces données, complétées par d'autres données physiologiques issues d'analyses métabolomiques, transcriptomiques, etc., serviront de base de données pour la modélisation des voies d'hépatotoxicité chronique par ordinateur et le développement des futurs algorithmes de prédiction.

B. Schéma expérimental

Le schéma expérimental complet de l'exposition des cellules HepaRG au VPA est présenté sur la figure V-1. Les analyses protéomiques consistaient à évaluer 13 conditions : 4 groupes « contrôle » aux jours 0, 2, 8 et 14 ainsi que 9 groupes pour lesquels trois quantités différentes d'acide valproïque avait été introduites au milieu de culture (0,05 mM, 0,15 mM et 0,5 mM) aux jours 2, 8 et 14.

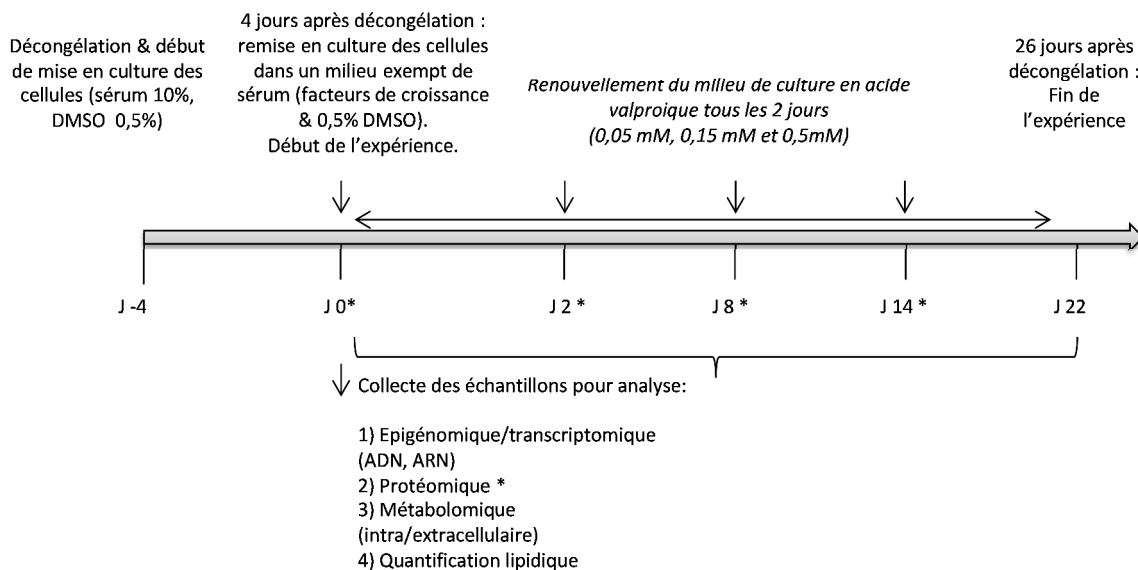


Figure V-1 Schéma expérimental pour l'analyse des cellules HepaRG exposées à l'acide valproïque
Les échantillons ont été collectés aux jours J0, J2, J8, J14. Chaque condition était composée de 3 réplicas de culture. Les échantillons prélevés ont fait l'objet d'analyses épigénomiques, transcriptomiques, protéomiques, métabolomiques et de dosages permettant de caractériser leur état de viabilité. 39 échantillons étaient inclus au total dans l'analyse protéomique.*

II. Mise au point de l'expérience

A. Choix de la stratégie analytique quantitative

Au cours de ma thèse, le laboratoire s'est équipé de nouveaux spectromètres de masse (Q-TOF dernière génération). Ces spectromètres sont, en plus d'être résolutifs (40 000 de résolution pour un m/z de 922 ; Impact-HD, Bruker Daltonique), capables de scanner très rapidement (50 Hz ; Impact-HD, Bruker Daltonique). Ces nouveautés ont permis la mise en place d'un nouveau type de stratégie quantitative au laboratoire : les stratégies sans marquage basées sur l'intégration des aires de courants d'ions extraits (XIC) (décrite dans le chapitre III en page 68 et qui ont également été utilisées pour la comparaison des différentes géométries de source électrospray dans le chapitre IV). Ce type de stratégie permet la quantification de plusieurs centaines de protéines simultanément dans des échantillons complexes, sans limite de conditions à comparer, ce qui convient parfaitement à ce projet qui cumule jusqu'à 13 conditions, et pour lequel un spectromètre de masse précis en masse et résolutif était disponible, comme illustré sur le schéma décisionnel présenté en figure III-12 du chapitre « Stratégies d'analyse protéomique quantitatives » en page 80.

L'approche logicielle qui a été utilisée pour cette étude est basée sur l'extraction et l'intégration des courants d'ions des peptides pour lesquels un spectre MS/MS a été validé. Cette approche a été réalisée au travers de l'utilisation du logiciel Skyline [122]. Ce logiciel, open source, présente

notamment l'intérêt de pouvoir visualiser aisément les intégrations réalisées à partir des courants d'ions. De plus, les deux outils majeurs permettant traiter les données XIC (MaxQuant [123] et Progenesis, Nonlinear Dynamics) n'étaient soit pas encore compatibles avec les données générées par des spectromètres de masse Bruker Daltonique (MaxQuant) soit non disponibles au LSMBO (Progenesis, Nonlinear Dynamics).

B. Choix du protocole de traitement des échantillons pour l'analyse protéomique quantitative

B.1. Points à considérer

Les stratégies quantitatives nécessitent des protocoles de préparations d'échantillons reproductibles. Plus un protocole expérimental de traitement est complexe, plus la probabilité d'introduire un biais est élevée [140]. Les préparations doivent donc rester relativement simples dans le but de mettre au point des méthodes d'analyses robustes et ne pas compliquer massivement l'interprétation des données (avec de multiples fractionnements par exemple). Il est nécessaire de trouver un compromis entre des préparations d'échantillons reproductibles, réalisables sur 78 échantillons simultanément (39 échantillons pour lesquels deux réplicas seront réalisés de façon à évaluer la reproductibilité du protocole de traitement des échantillons), et qui apportent, au travers d'une décomplexification suffisante, une réponse spécifique à la question posée.

Nous avons donc réalisé des tests afin de comparer différents protocoles de traitement des échantillons en évaluant :

- Le nombre de protéines identifiées pour chacun des protocoles testés
- La reproductibilité des informations quantitatives générées

B.2. Comparaison de deux protocoles de traitement des échantillons

B.2.1. Description des protocoles de traitements testés

Deux préparations d'échantillons différentes ont été évaluées sur ces échantillons (la préparation des échantillons est détaillée dans la partie expérimentale en page 215) :

- **Stratégie sans préfractionnement** : cette stratégie consiste à solubiliser l'échantillon, à réduire les ponts disulfures des protéines, à alkyler les cystéines réduites et à digérer enzymatiquement les protéines de l'échantillon en milieu liquide. Après extraction sur phase solide (C18), le digestat peptidique est analysé directement par le couplage LC-MS.
- **Stratégie avec préfractionnement** : cette stratégie est basée sur l'utilisation d'un gel d'électrophorèse SDS-Page. Le niveau de décomplexification de l'échantillon, c'est-à-dire l'étendue de la séparation des protéines dans le gel, est à l'appréciation de l'opérateur. La suite du traitement consiste à exciser les bandes de gel, à réduire les ponts disulfures des protéines, à alkyler les cystéines réduites puis à digérer enzymatiquement les protéines in-gel. La dernière étape consiste à extraire les peptides des bandes de gel pour être analysés sur le couplage LC-MS. Notre choix s'est porté sur 6 bandes (soit une migration des protéines sur environ 1 cm), ce qui représente un compromis entre un temps d'analyse total raisonnable (soit quelques semaines d'analyses au maximum pour les 78 échantillons) et une décomplexification suffisante de l'échantillon.

Le schéma expérimental des tests réalisés pendant la mise au point du protocole de traitement est présenté en figure V-2. Pour réaliser ce test, 3 échantillons exposés à des concentrations différentes en acide valproïque, et pour des durées différentes, ont été mélangés afin d'avoir un échantillon « modèle » (« Pool 0 ») de l'expérience. Cet échantillon modèle a ensuite été divisé en 5 échantillons différents (« Pool 1-Pool 5 »). Trois de ces échantillons ont fait l'objet d'un traitement sans préfractionnement tandis que deux échantillons ont fait l'objet d'un préfractionnement par gel SDS-Page. Une image du gel SDS-Page réalisé est présentée en figure V-3.

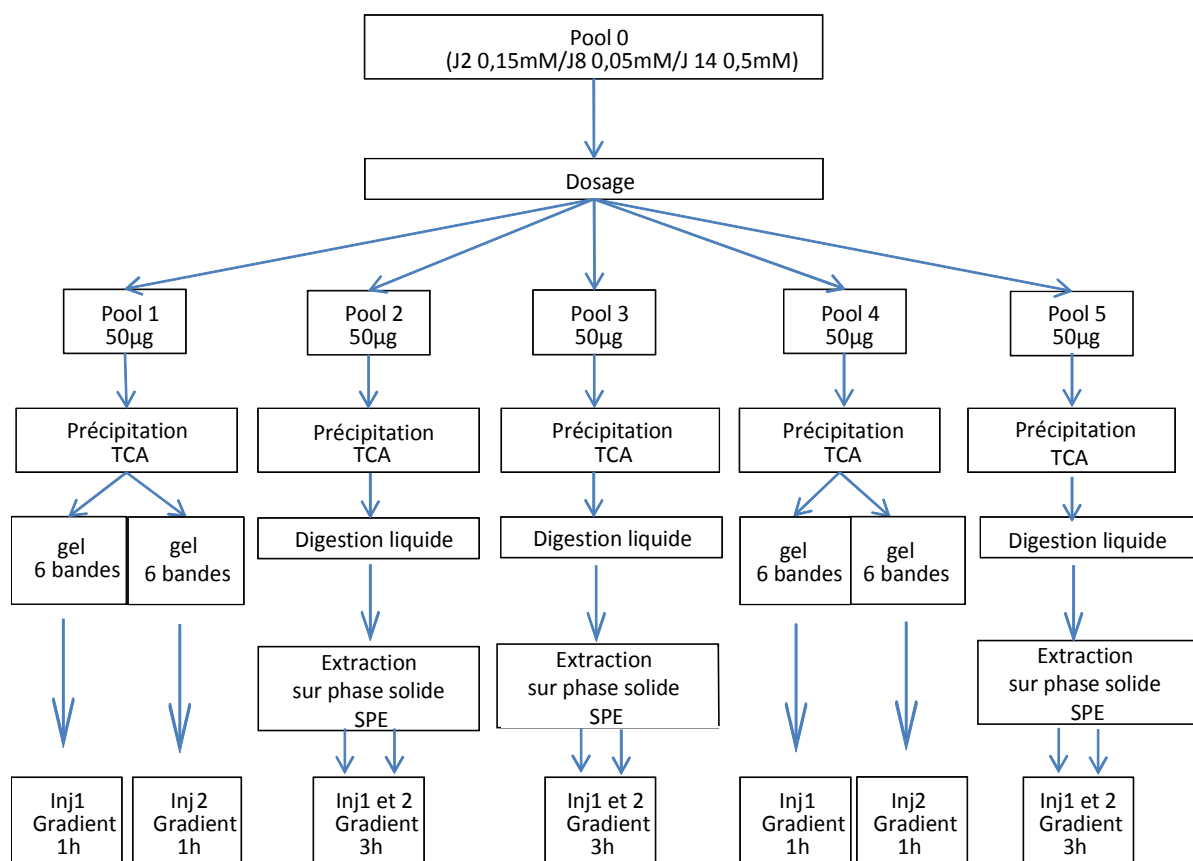


Figure V-2 Schéma expérimental concernant la mise au point du protocole analytique

3 échantillons exposés à des concentrations en acide valproïque différentes, et pendant des durées différentes, ont été mélangés dans le but de créer un échantillon modèle « Pool 0 ». Après dosage, cet échantillon a ensuite été divisé en 5 : « Pools » 1 à 5. Après dosage et précipitation des protéines par une solution de TCA à 10 %, les extraits protéiques de trois de ces échantillons ont été soumis à un traitement sans préfractionnement tandis que deux autres ont été soumis à une décomplexification par gel SDS-Page. L'analyse des peptides issus de la digestion enzymatique in-gel ou off-gel (sans préfractionnement) a été réalisée par un spectromètre de masse de type Q-TOF IMPACT-HD (Bruker Daltonique) couplé à un système UPLC de type NanoAcquity (Waters) équipé d'une colonne C18 Acclaim Pepmap RSLC, Dionex (0,075 x 500 mm, 2 µm). Les échantillons issus du traitement sans préfractionnement par gel ont été analysés deux fois tandis que deux dépôts par gel ont été réalisés (puis analysés par le couplage une fois chacun) pour les échantillons 1 et 4.

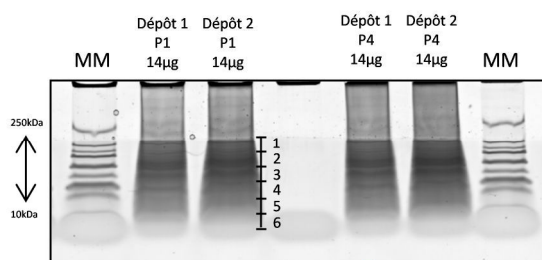


Figure V-3 Image de gel SDS-Page réalisé pendant la mise au point du protocole de traitement des échantillons

Les échantillons « pools 1 et 4 » ont été déposés deux fois sur le gel SDS-Page. Six bandes ont été excisées pour chacune des pistes. Les protéines contenues dans les bandes de gel ont été réduites, alkylées, digérées enzymatiquement in-gel puis les extraits peptidiques ont été analysés avec le couplage LC-MS. La piste MM correspond à un marqueur de masse moléculaire (entre 250 et 10 kDa).

B.2.2. Analyses des échantillons tests sur le couplage LC-MS

L'analyse des peptides issus de la digestion enzymatique in-gel ou off-gel (sans préfractionnement) a été réalisée avec un spectromètre de masse de type Q-TOF IMPACT-HD (Bruker Daltonique) couplé à un système UPLC de type NanoAcquity (Waters) équipé d'une colonne C18 Acclaim Pepmap RSLC, Dionex (0,075 x 500 mm, 2 µm).

La méthode d'acquisition utilisée pour le spectromètre de masse est basée sur les développements présentés au chapitre IV « Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes » en page 84. Un spectre MS est généré toutes les 3 secondes, ce qui est en adéquation avec le système chromatographique utilisé (ainsi que le gradient d'éluion de 180 minutes). Les spectres MS/MS, réalisés en mode DDA, ont été générés à des vitesses comprises entre 2 et 25 Hz. Cette méthode est détaillée dans la partie expérimentale en page 215.

B.2.3. Mise au point des conditions d'analyses chromatographiques

Pour les échantillons ayant subi un traitement avec préfractionnement, seule une heure de gradient chromatographique a été utilisée pour chaque bande de gel de façon à trouver un compromis entre une décomplexification suffisante et un temps d'analyse total raisonnable des 78 échantillons. Dans ces conditions, l'analyse des 6 bandes protéiques pour 78 échantillons était susceptible de nécessiter au moins 4 semaines d'analyse sur le couplage LC-MS.

Différents gradients de chromatographie ont été évalués pour les échantillons faisant issus du traitement sans préfractionnement dans le but de trouver les meilleures conditions d'analyse car leur complexité est supérieure à celles des échantillons qui ont été préfractionnés. Les différentes conditions testées consistaient à allonger les gradients d'éluion de 120 à 180, puis 240 minutes. Les résultats sont présentés en figure V-4.

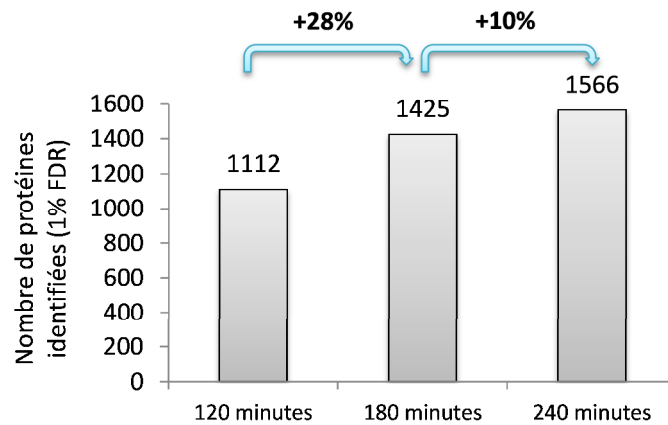


Figure V-4 Nombre de protéines identifiées pour 3 types de gradients d'éluion testés pour les échantillons non préfractionnés

Le nombre de protéines identifiées augmente avec le temps de gradient utilisé.

On peut constater d'après la figure V-4 que le nombre de protéines augmente avec le temps de gradient utilisé, en revanche, le gain entre les gradients de 180 et 240 minutes ne justifie pas l'allongement significatif du temps d'analyse global de tous les échantillons, environ 3-4 jours supplémentaires pour 80 échantillons, en considérant le phénomène de sous échantillonnage concernant l'acquisition des spectres MS/MS décrit ci-dessous.

Phénomène de sous-échantillonnage :

La figure V-5 est représentative d'un phénomène connu dans le domaine de l'analyse d'échantillons complexes par spectrométrie de masse : le sous-échantillonnage. En effet, en mode DDA, seule une partie des peptides seront sélectionnés, isolés puis fragmentés au cours de l'analyse en fonction de l'intensité des précurseurs (décrit au chapitre IV en page 90). Ainsi tous les peptides détectés en MS ne seront pas identifiés dans chacune des analyses. Etant donné le caractère semi-aléatoire de sélection des peptides pour la fragmentation (basée sur leur intensité sur le dernier spectre MS réalisé avant le cycle MS/MS), une manière d'augmenter le nombre de peptides identifiés consiste à analyser plusieurs réplicas pour un même échantillon. Nous avons testé la répétition des analyses pour un même échantillon avec un gradient d'éluion de 180 minutes. Le nombre de protéines identifiées à l'issue de ces multiples analyses sont présentées en figure V-5.

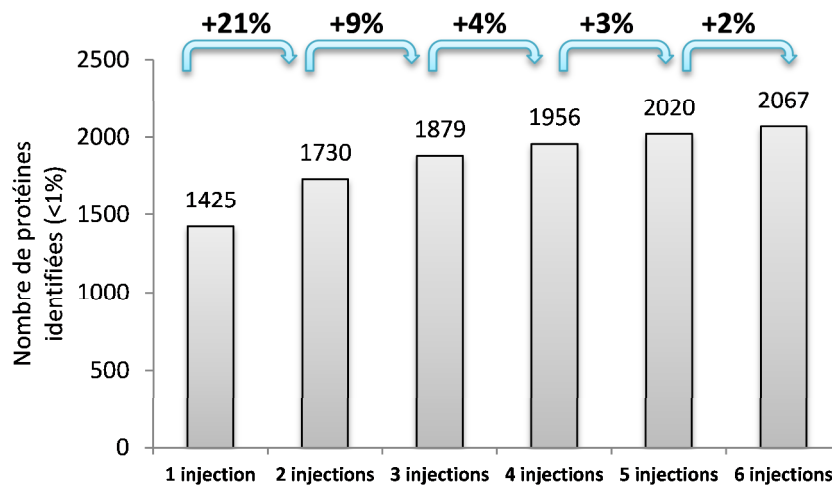


Figure V-5 Illustration du phénomène de sous-échantillonnage

Le nombre de protéines identifiées augmente avec la multiplication des analyses d'un même échantillon signifiant que des peptides différents ont été sélectionnés au cours de chacune d'elles. Le temps de gradient utilisé était de 180 minutes.

On constate d'après la figure V-5 que le nombre de protéines augmente avec le nombre d'analyses effectuées pour un même échantillon, cela signifie que des peptides différents ont été sélectionnés au cours de chacune des analyses. Le phénomène de sous-échantillonnage est non négligeable (au moins 21 % de gain entre la première et la deuxième analyse) malgré les vitesses d'acquisition élevées que peut atteindre l'Impact-HD (jusqu'à 25 Hz pour la méthode utilisée). La stratégie quantitative XIC est une technique quantitative sans marquage nettement moins sensible au sous-échantillonnage que la technique par comptage de spectres par exemple. En effet, les logiciels de traitement de données XIC ont la capacité d'apporter une information quantitative sur un peptide même lorsqu'il n'a pas été identifié dans une analyse donnée à condition qu'il l'ait été dans une autre analyse du même ensemble de données. Ce sous-échantillonnage n'est donc pas problématique pour la stratégie quantitative XIC, à l'inverse, il va permettre de quantifier un plus grand nombre de protéines que si les peptides identifiés avaient été systématiquement les mêmes entre les différentes injections.

On constate que le nombre de protéines supplémentaires identifiées après l'analyse de multiples répliques peut être représenté sous la forme d'une courbe atteignant une asymptote et que le gain en protéines supplémentaires identifiées n'est plus que de 2 % après 6 analyses du même échantillon. Pour un même instrument, plus un échantillon sera complexe, plus l'effet de sous échantillonnage sera grand. De ce fait, si on allonge les temps d'éluion chromatographique, ce qui entrainera une décomplexification plus élevée de cet échantillon, le fait de multiplier les analyses (comme cela sera fait indirectement par l'analyse des 78 échantillons pour l'expérience complète) aura un effet moindre sur le nombre total de protéines identifiées. Le fait d'analyser un très grand nombre

d'échantillons pour ce projet permettra donc de réduire les écarts entre le nombre de protéines identifiées avec les gradients de 180 et 240 minutes respectivement. De ce fait, nous avons choisi d'utiliser un gradient de 180 minutes pour les échantillons qui n'ont pas été décomplexifiés par gel SDS-Page dans le but de réduire le temps d'analyse global pour l'expérience complète.

B.2.4. Comparaison des deux stratégies évaluées en termes de nombre de protéines identifiées

Le nombre de protéines identifiées suite aux traitements avec ou sans fractionnement des échantillons (« Pool 1-Pool 5 ») est présenté en figure V-6.

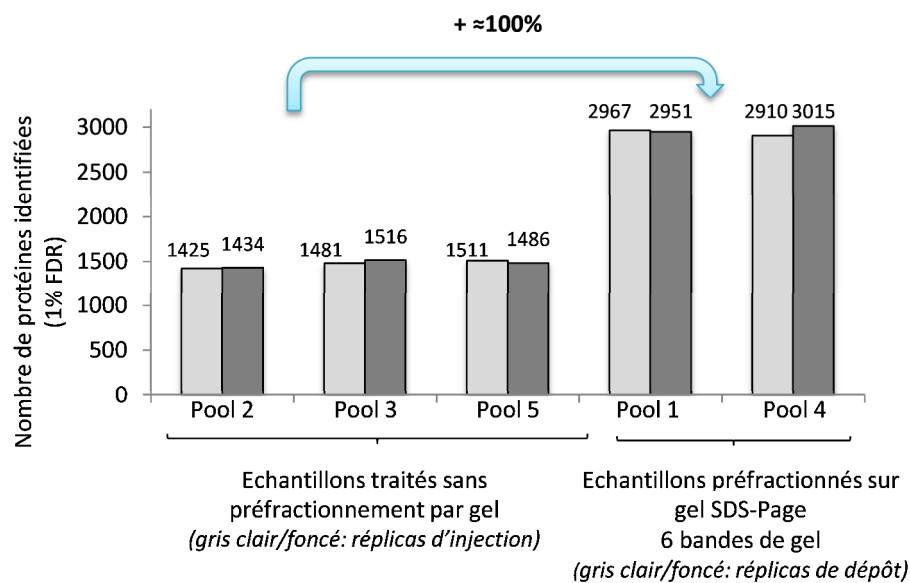


Figure V-6 Nombre de protéines identifiées suite aux deux types de traitement des échantillons testés
Le nombre de protéines identifiées pour les échantillons dont le protocole de traitement a inclus une étape de préfractionnement par gel SDS-Page est deux fois supérieur à celui des échantillons non préfractionnés.

D'après la figure V-6, on constate que la décomplexification par gel SDS-Page permet de doubler, et ce de façon reproductible, le nombre de protéines identifiées par rapport à la stratégie sans préfractionnement. En termes de protéines identifiées, la stratégie par préfractionnement par gel est donc largement supérieure au traitement sans préfractionnement. En complément de ce résultat, une évaluation de la reproductibilité des intensités de courants d'ions générés pour chaque type de traitement a été réalisée.

B.2.5. Comparaison des intensités des courants d'ions extraits pour les deux traitements d'échantillons en termes de reproductibilité

Le traitement des données effectué pour comparer les deux stratégies en termes de reproductibilité a été réalisé grâce au logiciel Skyline et correspond à la réalisation des étapes suivantes (décrites dans le chapitre « Stratégies d'analyses protéomiques quantitatives en page 68 ») :

1. les spectres issus des analyses nanoLC-MS/MS pour chaque type de traitement d'échantillons ont été soumis à l'algorithme de recherche Mascot. Les spectres dont l'assignation a été validée à moins de 1 % de résultats « faux-positifs » ont été regroupés pour former une librairie spectrale par type de traitement réalisé.
2. les courants d'ions de chaque peptide pour lequel un spectre est présent dans la librairie spectrale ont été extraits par le logiciel Skyline à partir des analyses nanoLC-MS/MS correspondantes à chaque type de traitement des échantillons.

La reproductibilité des intensités des courants d'ions a été évaluée pour 3416 protéines pour la stratégie basée sur le préfractionnement des échantillons par gel SDS-Page et 1991 protéines pour la stratégie sans préfractionnement (15199 et 7260 peptides respectivement). Un certain nombre de ces protéines et peptides sont communs aux deux stratégies de traitement comme le présente la figure V-7.

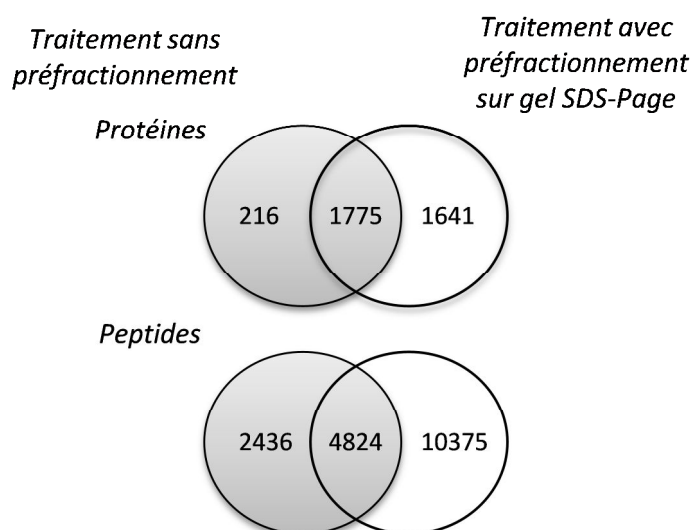


Figure V-7 Nombre de protéines et peptides identifiés pour chacune des stratégies de traitement des échantillons
50 % des protéines identifiées, et 30 % des peptides, sont communs aux deux types de traitement des échantillons.

On peut constater d'après ces résultats qu'en plus d'avoir permis d'identifier le double de protéines, le traitement par préfractionnement a permis d'identifier un nombre beaucoup plus important de

peptides (10375 peptides ont été identifiés spécifiquement pour la stratégie avec préfractionnement contre 2436 pour la stratégie sans préfractionnement).

Afin de comparer les aires sous les courbes des courants d'ions intégrés, les coefficients de variations moyens des peptides ont été calculés pour chacun des traitements. Les données n'ont pas été normalisées pour cette comparaison.

- Avec préfractionnement (gel SDS-Page 6 bandes) : les aires sous la courbe des courants d'ions extraits pour P, P+1 et P+2 par état de charge et par peptide ont été sommées par piste (= 6 bandes). Les coefficients de variations ont été calculés d'après les 4 pistes analysées pour les tests (deux dépôts analysés deux fois)
- Sans préfractionnement : les aires sous la courbe des courants d'ions extraits P, P+1 et +2 ont été sommées par état de charge et par peptide. Les coefficients de variations ont été calculés d'après les 6 analyses réalisées pour les tests (deux analyses de chaque échantillon test préparé).

Les coefficients de variation de l'ensemble des peptides identifiés au sein de chacune des stratégies de préparation sont présentés en figure III-8 sous la forme de boîte à moustaches. Les coefficients de variation de l'ensemble des peptides identifiés au sein de chacune des stratégies à l'exception des peptides contenant des méthionines ont été également représentés sur ce graphique. En effet, ces peptides sont sujets à des modifications post-traductionnelles (oxydation de la fonction thiol des méthionines) dont la proportion peut varier en fonction des étapes à réaliser et du temps de préparation des échantillons.

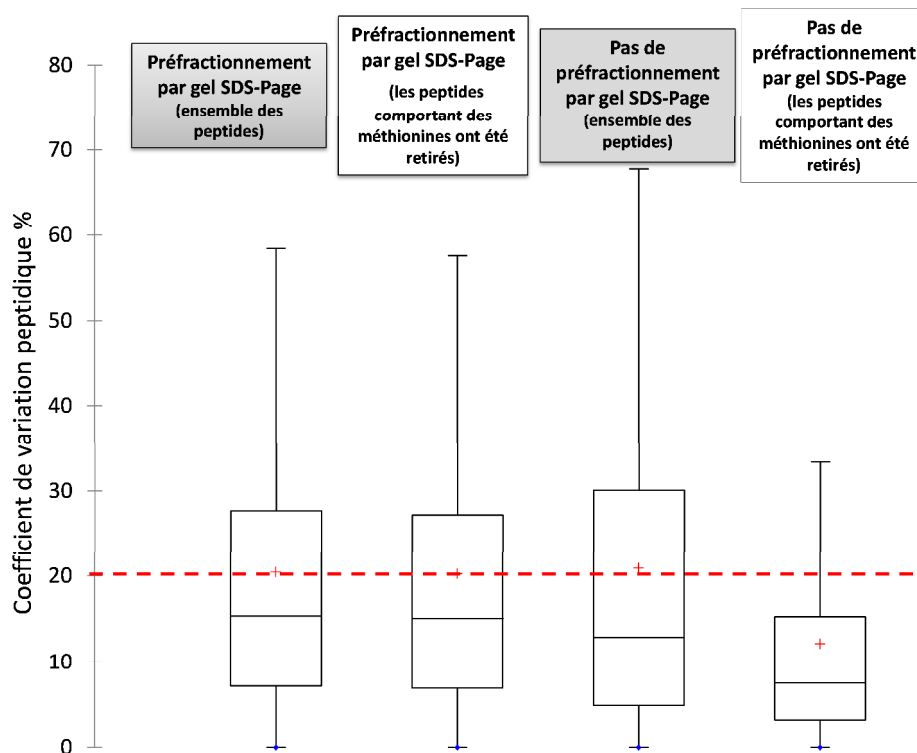


Figure V-8 Boîtes à moustaches représentant les coefficients de variation de l'ensemble des peptides pour chaque type de traitement et ce même ensemble sans les peptides contenant des méthionines

Les coefficients de variation pour l'ensemble des peptides identifiés pour chacune des stratégies semblent équivalents, en revanche, les coefficients de variations sur ces mêmes ensembles de peptides à l'exception des peptides contenant des méthionines sont très inférieurs pour la stratégie sans préfractionnement comparée à celle avec préfractionnement.

La figure V-8 permet de constater que la reproductibilité des courants d'ions intégrés pour l'ensemble des peptides identifiés semble équivalente pour les deux traitements d'échantillons (en gris sur la figure V-8). Les coefficients de variation obtenus pour chaque traitement sont compris entre 10 et 30 % environ. En revanche, une fois les peptides contenant des méthionines éliminés des données (en blanc sur la figure V-8), la stratégie sans préfractionnement apparaît comme plus reproductible par rapport à la stratégie avec préfractionnement par gel SDS-Page. On peut donc conclure de cette figure que les peptides contenant des méthionines sont effectivement sujets à un manque de reproductibilité et qu'ils étaient responsables de l'importante dispersion des coefficients de variation observée pour l'ensemble des peptides à l'issue du traitement sans préfractionnement. Ces résultats suggèrent la présence de multiples formes modifiées pour les peptides contenant des méthionines. Ce phénomène peut s'expliquer par l'étape critique d'extraction et de concentration sur cartouche C18 qui n'est réalisée que pour cette stratégie. En revanche, on peut constater que l'importante dispersion des coefficients de variation obtenus à l'issue du traitement avec préfractionnement n'est pas impactée par l'absence de ces peptides et que le manque de reproductibilité est propre au préfractionnement lui-même.

Nous avons ensuite réalisé la comparaison des peptides communs aux deux stratégies de traitement des échantillons (à l'exception de ceux contenant des méthionines), soit 3198 peptides (figure V-9).

	Avec préfractionnement par gel	Sans préfractionnement
Coefficient de variation moyen (%)	18%	12%
Nb de peptides coefficient de variation $\leq 20\%$	2119/3198 soit 66%	2669/3198 soit 84%

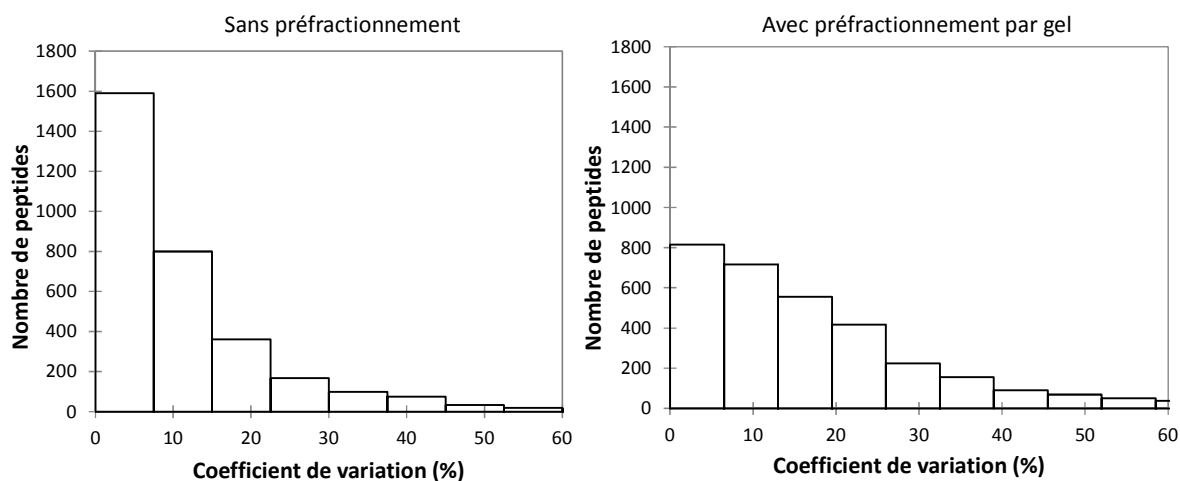


Figure V-9 Histogrammes présentant les coefficients de variation obtenus à partir de l'extraction des courants d'ions de tous les peptides identifiés (à l'exception des peptides contenant des méthionines) pour les deux stratégies

La figure V-9 permet de constater que la stratégie sans préfractionnement présente une reproductibilité plus élevée que la stratégie par gel pour l'ensemble des peptides communs (à l'exception de ceux qui contiennent des méthionines qui n'ont pas été comparés ici). En effet, sur 3198 peptides analysés communs aux deux types de traitement des échantillons, le coefficient de variation moyen des aires des peptides issus des échantillons préfractionnés est d'environ 20 % tandis que ceux issus de la stratégie sans préfractionnement présentent un coefficient moyen de 12 % environ. Près de 90 % des peptides communs possèdent un coefficient de variation inférieur à 20 % pour la stratégie sans préfractionnement, dont plus de la moitié inférieur à 10% comme le montre les histogrammes présentés en figure V-9. Près de 70% des peptides communs possèdent un coefficient de variation inférieur à 20% pour la stratégie avec préfractionnement, et seulement environ 40 % ont un coefficient de variation inférieur à 10 %. De ce fait, sur l'ensemble des peptides communs (à l'exclusion de ceux contenant des méthionines), la stratégie sans préfractionnement présente une reproductibilité largement supérieure à la stratégie avec préfractionnement.

B.2.6. Bilan des tests réalisés

A la suite de ces tests sur les traitements des échantillons, nous avons pu démontrer que la stratégie par préfractionnement par gel SDS-Page permet de doubler le nombre de protéines identifiées mais que ce type de stratégie fournit des données quantitatives moins reproductibles. Ainsi, cette stratégie nous permettrait de réaliser une quantification sur plus de protéines mais de façon moins robuste. **La stratégie sans préfractionnement a donc été choisie pour la réalisation de l'expérience complète.**

III. Etude de la réponse des cellules HepaRG à une exposition à l'acide valproïque

A. Préparation des échantillons

Le contenu protéique des échantillons a tout d'abord été dosé selon la méthode de Bradford afin d'évaluer la quantité de matériel protéique de chaque échantillon. Les résultats sont présentés en figure V-10. Ce dosage permettait également, en parallèle de dosages complémentaires, d'évaluer indirectement la viabilité des cellules au bout de 14 jours d'exposition au VPA.

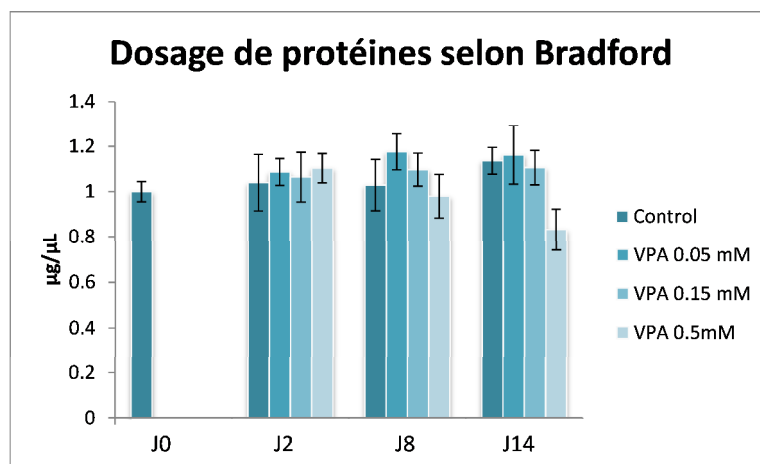


Figure V- 10 Résultat du dosage de protéines selon Bradford

Le contenu protéique de l'ensemble des échantillons a été évalué par dosage de Bradford. Celui-ci semble diminuer de façon significative à l'issue des 14 jours d'exposition.

On peut constater que les cellules ont commencé à se dégrader au jour 14, et ce de façon plus importante pour les échantillons exposés à la concentration d'acide valproïque la plus élevée. La suite du traitement a été réalisée en deux répliques. Le traitement des échantillons est le même que celui « sans préfractionnement » réalisé pendant les tests de traitement des échantillons et est détaillé dans la partie expérimentale. Après digestion enzymatique, les peptides ont été extraits sur

phase solide (cartouche C18) puis une quantité connue de peptides iRT (Biognosis) a été ajoutée à chacun des échantillons. Ces standards possèdent des séquences synthétiques. En général, ils sont utilisés comme repères d'alignement entre différents types de gradients chromatographiques lorsque des méthodes chromatographiques sont amenées à être transposées sur différents couplages. Dans notre cas, ils ont permis de réaliser le suivi du couplage instrumental pendant toute la durée des analyses pendant environ 18 jours consécutifs.

L'analyse des peptides issus de la digestion enzymatique a été réalisée avec un spectromètre de masse de type Q-TOF Impact-HD (Bruker Daltonique) couplé à un système nanoUPLC (NanoAcquity Waters) équipé d'une colonne C18 Zorbax, 0,075 x 250 mm, 1,7 μm (Waters).

B. Identification des protéines et extraction des courants d'ions avec le logiciel Skyline

Les étapes d'identification des protéines et d'extraction des courants d'ions réalisées sont identiques à celles réalisées pour les tests concernant le traitement des échantillons décrits précédemment dans le chapitre. Les points clés concernant le traitement de ce type de données est présenté dans le chapitre « Stratégies d'analyse protéomique quantitatives » en page 68.

2024 protéines (soit 9332 peptides) ont été finalement validées avec un taux de faux-positif inférieur à 1%. Etant donné la faible reproductibilité des intensités des courants d'ions pour les peptides contenant des méthionines (qui représentent 2353 peptides), ceux-ci ont été éliminés des données, de même que les peptides qui n'étaient pas uniques aux protéines identifiées dans cette étude comme il est indiqué dans le chapitre III « Stratégies d'analyse protéomique quantitatives » en page 68.

Les courants d'ions de 6685 peptides (soit 1842 protéines) ont donc été extraits puis intégrés par le logiciel Skyline dans les 78 analyses réalisées. Les extractions de courant d'ions ont été réalisées dans une fenêtre de temps restreinte autour des temps de rétention auxquels les spectres MS/MS de chaque peptide présent dans la librairie spectrale ont été générés (3 minutes). Cette stratégie permet de réduire des possibles erreurs d'intégration des courants d'ions et de raccourcir le temps nécessaire à la réalisation de cette étape. Les intégrations réalisées par le logiciel ont été entièrement vérifiées manuellement dans le but de vérifier la justesse de celles-ci (et éventuellement les corriger).

A la suite de cette validation manuelle des intégrations, 2898 peptides, correspondant à 527 protéines, ont été éliminés des données. Les aires des courants d'ions P, P+1, P+2 ont été sommées pour chacun des peptides (par état de charge) car nous avons constaté que les courants d'ions extraits P+2 sont peu intenses pour la majeure partie des peptides, et sont, de ce fait, sujets aux interférences. Sommer les aires P, P+1, P+2 permet de minimiser leurs effets.

La suite du traitement des données (normalisation, approximation des résultats par protéine puis analyse statistique) a donc été réalisée sur la base de 1315 protéines et 3787 peptides.

C. Suivi du couplage instrumental

Etant donné le nombre d'échantillons à analyser, il est très important de mettre en place un schéma expérimental adapté de façon à ne pas introduire de biais dans l'ordre d'analyse des échantillons et de mettre en place des systèmes de contrôle de la stabilité du couplage LC-MS.

Le schéma expérimental complet comptant 78 échantillons à évaluer (6 réplicas pour chacune des 13 conditions), les analyses ont été réalisées par blocs de 6 x 13 échantillons. Chaque bloc contenait un échantillon de chacune des conditions. Au sein de chaque bloc, les échantillons ont tous été analysés dans un ordre aléatoire comme recommandé dans la littérature lors de l'analyse de plusieurs dizaines d'échantillons consécutivement [141]. Un schéma présentant l'ordre des analyses est présenté en figure V-11. En plus du contrôle réalisé sur les peptides iRT mélangés à tous les échantillons, un second contrôle a été réalisé à partir d'un mélange de tous les échantillons de l'expérience (« échantillon contrôle » nommé « EC »). Un aliquote de cet « échantillon contrôle » a été analysé au début et à la fin de chaque bloc de treize échantillons (soit n = 12). Les peptides iRT ont également été introduits dans ces échantillons « contrôle ».

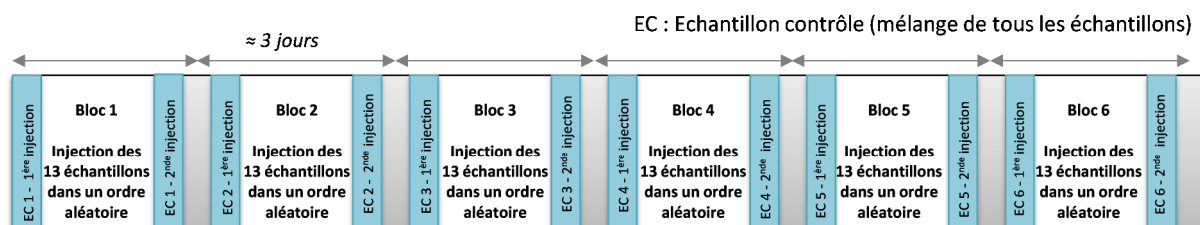


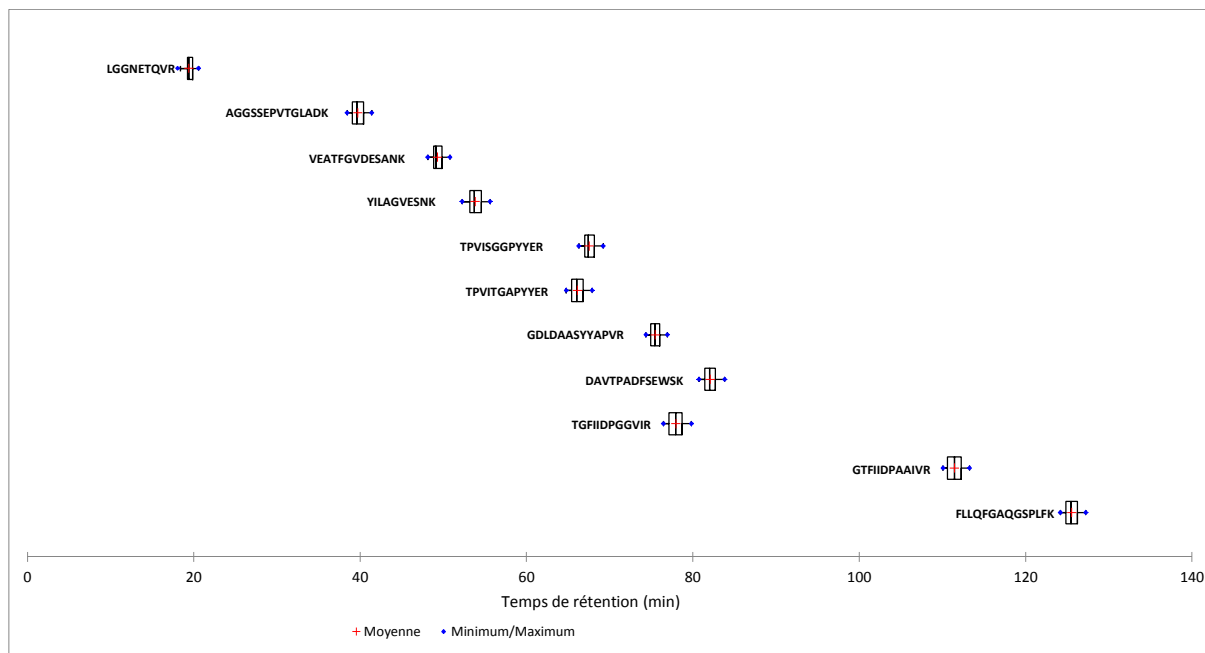
Figure V-11 Schéma représentatif de l'ordre d'analyse des échantillons

C.1. Suivi des « échantillons contrôles »

Les coefficients de variation des peptides standards iRT (à l'exception de deux peptides les plus hydrophobes pour lesquels la rétention chromatographique a été très peu reproductible), leur temps de rétention ainsi que le nombre de protéines identifiés dans chaque « échantillon contrôle » ont été évalués quotidiennement. Les données sur les 12 aliquotes de l'échantillon « contrôle » sont présentées en tableau 1 et figure V-12 et V-13.

Tableau 1 : Moyennes, écarts-types et coefficients de variation issus de l'extraction des courants d'ions des peptides standards iRT pour le suivi de la stabilité du couplage pendant l'expérience

Peptides iRT, Biognosis	Moyenne des aires sous la courbe (somme P, P+1, P+2)	Ecart-type	Coefficient de variation %
FLLQFGAQGSPLFK	1.4E+07	1.7E+06	12.4
GTFIIDPAAIVR	2.4E+07	3.5E+05	1.5
TGFIDPGGVIR	8.1E+07	4.6E+06	5.7
DAVTPADFSEWSK	4.1E+07	3.9E+06	9.4
GDLDAASYAPVR	2.9E+07	2.0E+06	6.7
TPVITSGAPYYER	2.3E+07	1.6E+06	6.9
TPVISGGPYYER	4.6E+07	5.3E+06	11.7
YILAGVESNK	4.9E+07	3.7E+06	7.5
VEATFGVDESANK	3.1E+07	4.0E+06	13.0
AGGSSEPVTLADK	3.1E+07	2.0E+06	6.5
LGGNETQVR	6.3E+06	1.3E+06	20.9



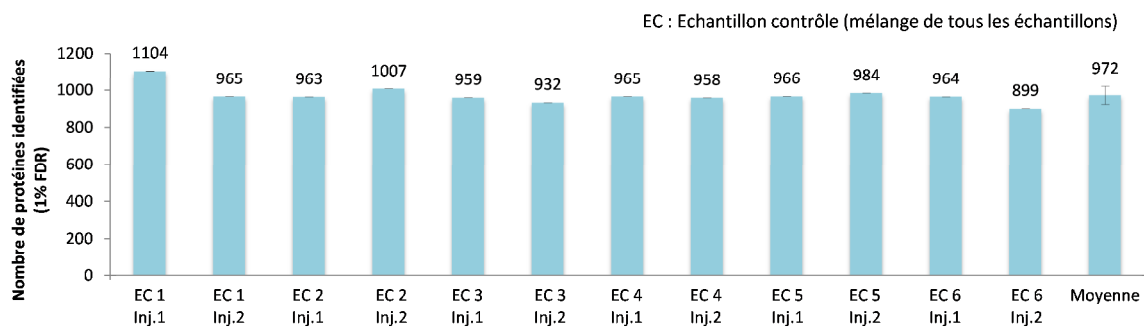


Figure V-13 Nombre total de protéines identifiées dans les 12 aliquotes de l'échantillon « contrôle »
Le nombre de protéines identifiées est resté stable pendant les 18 jours d'analyse à $\pm 10\%$

On peut constater d'après le tableau 1 et la figure V-13 que les coefficients de variation calculés pour les peptides iRT présents dans ces échantillons contrôle sont généralement inférieurs à 20 % (et même souvent inférieurs à 15%). Cette valeur de 20 % est reconnue pour les analyses quantitatives ciblées (SRM) qui permettent un suivi plus spécifique des protéines à quantifier (suivi de multiples transitions par peptide) [92], de ce fait, on peut considérer que la reproductibilité acquise sur ces données est excellente pour cette stratégie qui est plus globale dans la mesure où elle permet de quantifier plusieurs centaines de protéines simultanément. Le nombre de protéines identifiées dans l'échantillon « contrôle » est de 972 protéines ± 49 , ce qui confirme également la stabilité du système LC-MS utilisé pendant les 18 jours d'analyse consécutifs.

De plus, le système chromatographique a montré une excellente reproductibilité avec des coefficients de variation des temps de rétention des peptides iRT inférieurs à 5 % (figure V-12), ce paramètre étant crucial pour ce type de stratégie quantitative basée, en partie, sur la bonne corrélation des temps de rétention des peptides aux temps de rétention enregistrés dans la librairie de spectres servant à l'extraction des courants d'ions.

Ces données attestent de la bonne stabilité du couplage instrumental pendant toute la durée de l'expérience.

C.2. Contrôle sur l'ensemble des échantillons

En complément, une moyenne des coefficients de variation intra-groupe ($n = 6$) de l'ensemble des peptides identifiés et quantifiés dans l'expérience (soit 3787) sont présentés en figure V-14. Ces coefficients de variation ont été calculés avant la normalisation des données mais après la validation manuelle de l'ensemble des courants d'ions.

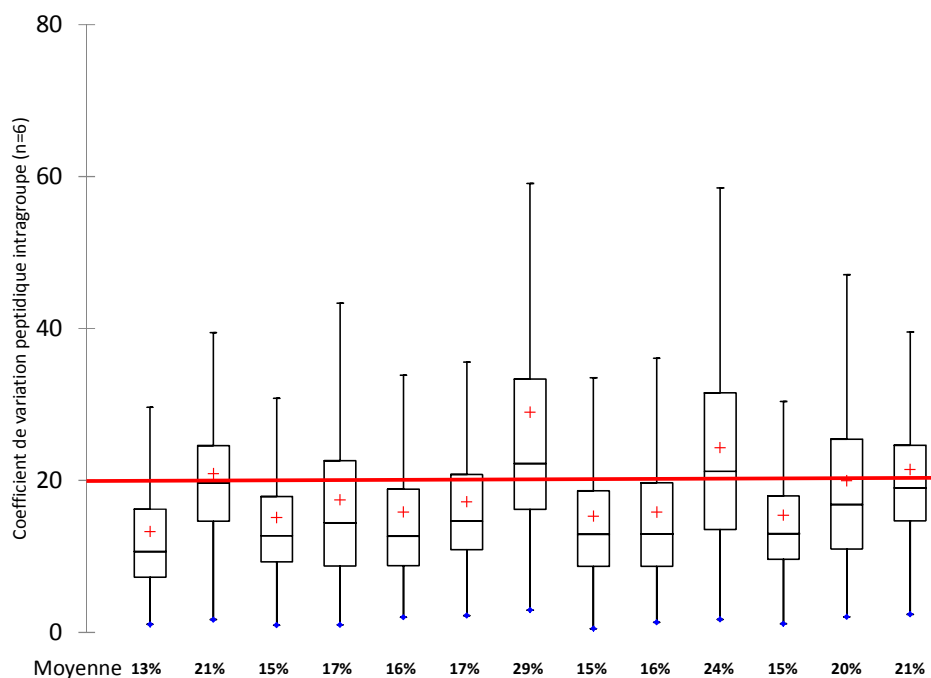


Figure V-14 Coefficients de variation intra-groupe pour les 3787 peptides quantifiés

Les coefficients de variation intragroupes ($n=6$) calculés ont été calculés pour 3787 peptides d'après les aires sous les courbes des courants ions P, P+1 et P+2 sommées pour chaque peptide. Les moyennes de ces coefficients de variation intragroupe sont pour la plupart des 13 conditions inférieures à 20 %, ce qui atteste de la reproductibilité du couplage LC-MS utilisé.

Les moyennes intragroupes des coefficients de variation peptidiques sont largement inférieures à 20% pour la plupart des 13 conditions à comparer. Ces données, non normalisées, valident donc également la bonne robustesse du couplage instrumental utilisé pendant toute la durée de l'expérience.

De plus, la moyenne des coefficients de variation des peptides iRT calculés sur la base des 78 échantillons à analyser (à l'exception d'un peptide hydrophile et de deux hydrophobes qui ne présentaient pas des comportements chromatographiques reproductibles) était de 20 ± 2 %.

C.3. Conclusion sur la stabilité du couplage

L'ensemble des vérifications effectuées sur les 78 échantillons à comparer et sur les échantillons contrôles (« EC ») a permis la validation de la stabilité du couplage instrumental tout au long de l'expérience : les analyses n'ont donc pas été interrompues pour nettoyer le spectromètre de masse. Il est alors possible d'attester de la bonne robustesse du couplage utilisé. Toute variabilité observée sera donc liée au contenu des échantillons biologiques eux-mêmes et ne sera donc pas liée à un biais instrumental.

D. Traitement des données

Il n'existe pas aujourd'hui de traitement de données XIC unique concernant la normalisation ou l'approximation des données peptidiques par protéine. Différents types de traitement des données sont décrits dans la littérature [142]. Ces étapes de traitement sont décrites dans le chapitre « Stratégies d'analyse protéomique quantitatives » en page 68.

Le schéma présenté en figure V-15 résume les différentes étapes que nous avons effectuées sur ces données. Les choix réalisés pour chacune de ces étapes seront justifiés dans la suite du chapitre.

Les différents états de charge détectés par peptide ont été considérés comme des entrées peptidiques différentes.

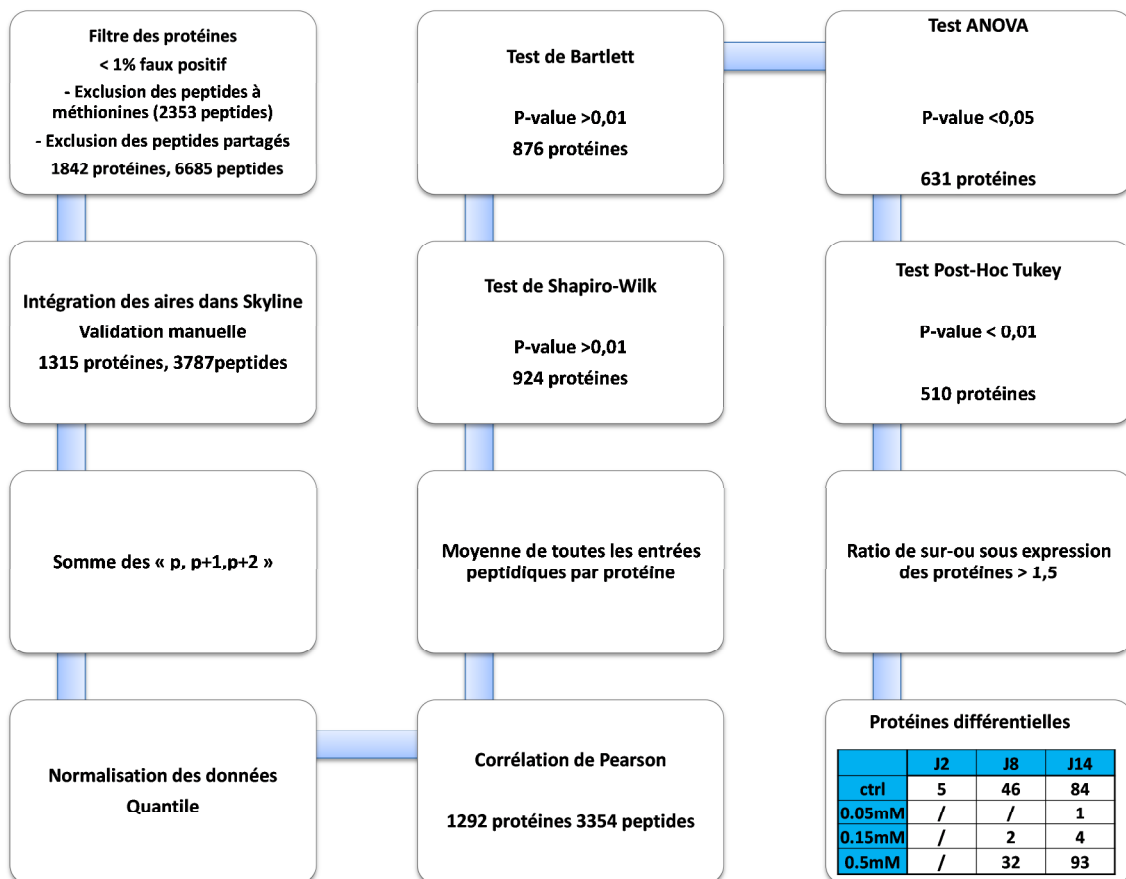


Figure V-15 Résumé du traitement des données réalisé pour le projet NOTOX

D.1. Normalisation des données

Différents modèles de normalisation des données ont été évalués à l'aide du logiciel Normalyzer (version 1.0) [127]. Les corrélations des intensités pour deux échantillons au sein du même groupe sont présentées en Figure V-16 pour les modèles de normalisation les plus utilisés, c'est-à-dire le modèle de normalisation par l'égalité des médiane (MedI-G), par quantile (Quantile) ou sans normalisation (\log_2). Les données ont été transformées en données logarithmiques lors de cette étape.

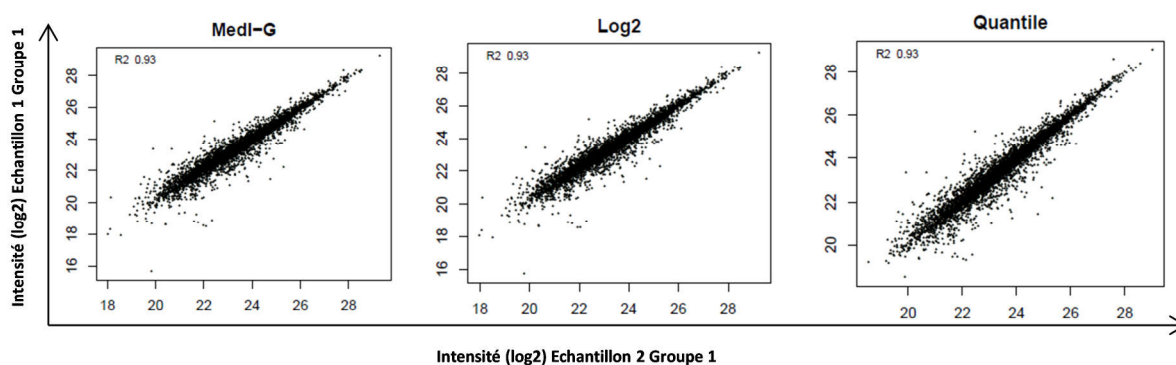


Figure V-16 Corrélation des intensités au sein de deux échantillons du même groupe

Les intensités (\log_2) de chacun des peptides dans une analyse sont représentées en fonction des intensités de ces mêmes peptides dans l'analyse d'un second échantillon faisant partie du même groupe. On peut constater que, pour les trois types de traitement, le nuage de points obtenu sur ces graphiques peut être défini par une équation du type $y = ax + b$ où « a » tend à se rapprocher de la valeur « 1 » et « b » de la valeur « 0 ».

On constate d'après la figure V-16 que les nuages de points peuvent être défini par une droite d'équation $y = ax + b$ avec « a » qui tend vers 1 et « b » qui tend vers 0. De plus, les profils sont très similaires pour les trois types de traitement des données dont un sans étape de normalisation, ce qui conforte la bonne reproductibilité de nos résultats. Il est difficile, à partir de ces représentations de définir quel est le meilleur modèle de normalisation pour ces données.

La figure V-17 présente le coefficient de variation peptidique intra-groupe obtenu à la suite de l'application de chaque modèle de normalisation proposé par le logiciel Normalyzer, le tout rapporté au coefficient de variation intragroupe moyen des données non normalisées. Les modèles auxquels nous nous sommes intéressés dans cette partie sont : \log_2 (données logarithmiques mais non normalisées), MedI (normalisation par l'égalité des médianes) et Quantile. Le détail de l'ensemble des modèles de normalisation proposés par le logiciel Normalyzer (1.0) est présenté en annexe 1.

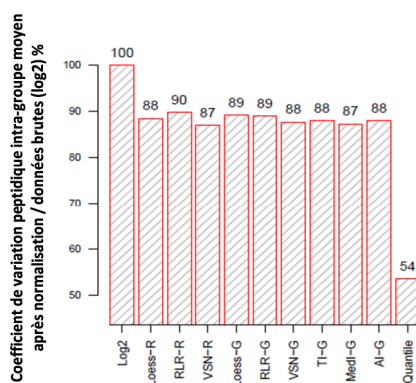


Figure V-17 Coefficient de variation peptidique intragroupe moyen après normalisation (rapporté au coefficient de variation peptidique moyen des données non normalisées) avec différents modèles proposés par l'outil Normalyzer
La plupart des modèles de normalisation présentent des résultats équivalents à l'exception du modèle de normalisation par quantile pour lequel le coefficient de variation intragroupe moyen est fortement diminué.

Notre hypothèse était que le meilleur modèle de normalisation est celui qui présente le coefficient de variation peptidique moyen intragroupe le plus faible. Ce graphique permet d'observer que le modèle de normalisation quantile est le plus puissant. Ce modèle a été choisi dans le but de conforter l'identification des protéines qui ont été évaluées comme significativement différentes entre les différents groupes. De plus, il a été montré dans le chapitre IV « Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes » que ce type de normalisation était en adéquation avec le traitement de données XIC et permettait de réduire le nombre de résultats « faux-positifs » (page 111).

D.2. Rendu des résultats par protéine

Différentes stratégies de traitement des données ont été développées dans le but de passer de multiples valeurs quantitatives peptidiques à une unique valeur par protéine. Les différentes stratégies sont brièvement présentées dans le chapitre III « Stratégies d'analyse protéomique quantitatives ». La stratégie que nous avons retenue consiste à réaliser la moyenne (géométrique) des aires sous les courbes de courants d'ions peptidiques par protéine. Cette approximation est sensible aux valeurs extrêmes, contrairement aux valeurs médianes de ces intensités mais la significativité des différences que nous rapporterons n'en sera que renforcée.

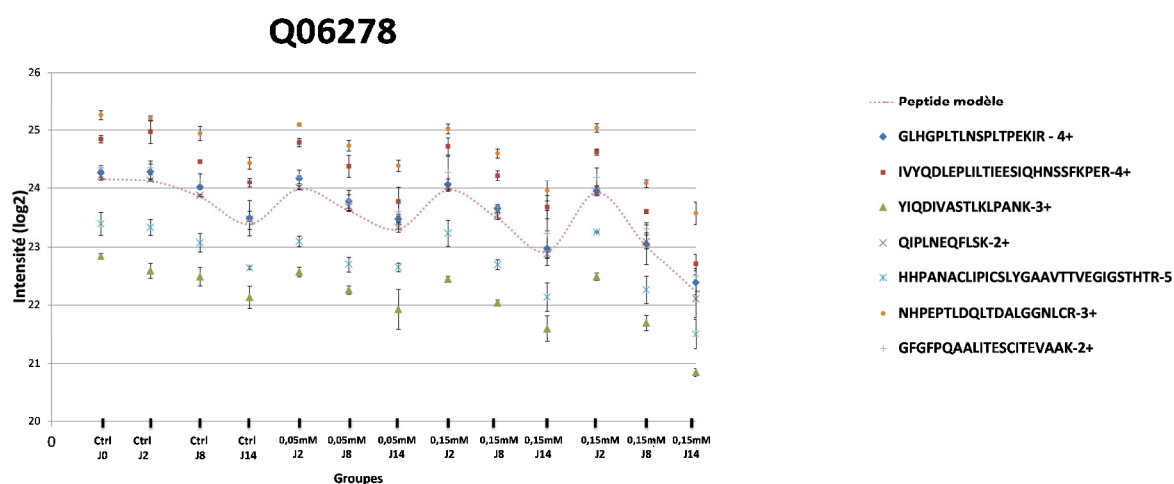
Nous avons pu constater dans nos données que certains peptides ne possèdent pas le même comportement pour une même protéine entre différentes conditions. L'origine de ce phénomène provient probablement de la présence de modifications post-traductionnelles. L'induction de ces modifications par le traitement à l'acide valproïque (ou le temps) présente un intérêt du point de vue

de la réponse physiologique de ces protéines. Cependant, pour donner une réponse correcte, il aurait fallu détecter et quantifier l'ensemble des modifications post-traductionnelles existantes. Or, rechercher l'ensemble des modifications post-traductionnelles aurait conduit à des temps de recherche dans les banques de données protéiques extrêmement longs, de plus, le risque d'augmentation du taux de faux-positif aurait réduit, probablement considérablement, le nombre de protéines identifiées dans ces analyses et de ce fait, n'était donc pas envisageable. Ainsi, avant de valider l'approximation des résultats par protéine (i.e. la moyenne de tous les peptides pour chaque protéine dans chaque condition donnée), nous avons donc voulu vérifier la bonne corrélation des peptides pour chaque protéine quantifiée. Nous avons ainsi voulu répondre à la question suivante : est-ce que tous les peptides répondent de la même façon pour une protéine dans chacune des conditions données ? Par exemple, lorsqu'un peptide donné voit son niveau d'expression multiplié par 2 entre deux conditions, est-ce le cas pour tous les peptides quantifiés pour cette même protéine ? Si oui, l'approximation par protéine sera validée.

Si non, soit :

- nous avons éliminé les peptides mal corrélés sur la base de critères définis dans la suite du chapitre. La valeur protéique a été ensuite recalculée sur la base des peptides qui validaient les critères de corrélation.
- soit nous n'avons pas fourni de valeur quantitative finale pour les protéines concernées.

Nous avons basé notre comparaison sur une corrélation de Pearson [143]. Pour cela, nous avons « créé » un peptide modèle par protéine auquel nous avons comparé tous les autres. L'aire sous la courbe des courants d'ions de ce peptide modèle que nous avons « créé » correspond à la moyenne des aires sous la courbe des courants d'ions extraits pour tous les peptides d'une protéine. Pour cette évaluation, nous avons moyenné les résultats de tous les réplicas n'obtenant plus qu'une seule valeur par groupe. La protéine « Aldehyde oxidase, Q06278 » est présentée en exemple en figure V-18.



Peptides	Coefficient de corrélation/ peptide modèle (en rouge)
QIPLNEQFLSK	1.00
GLHGPLTLNSPLTPEKIR	1.00
NHPEPTLDQLTDALGGNLCR	0.99
GFGFPQAALITESCITEVAAK	0.99
HHPANACLIPICSLYGAAVTTVEGIGSTHTR	0.99
YIQDIVASTLKLPAK	0.99
IVYQDLEPLILTIEESIQHNSSFKPER	0.99

Figure V-18 Evaluation de la corrélation des 7 peptides quantifiés pour la protéine "Aldehyde oxidase"
L'abondance relative du peptide modèle entre les conditions est présentée en rouge.

Nous avons éliminé des données les peptides qui présentaient des coefficients de corrélation trop bas par rapport à l'ensemble des peptides quantifiés par protéine :

- Pour les protéines à un peptide (un seul état de charge détecté) : Pas de corrélation à effectuer, ces protéines ont été conservées mais identifiées comme « protéines à 1 peptide » étant donné la faible robustesse de ces données quantitatives.
- Pour les protéines à deux peptides (ou un seul peptide mais deux états de charge détectés) : Les protéines pour lesquelles les deux peptides avaient un facteur de corrélation inférieur à 0,8 ont été éliminées.
- Pour les protéines avec trois peptides (éventuellement sous la forme de différents états de charge) : si moins de trois peptides pour une protéine avaient un coefficient de corrélation supérieur à 0,8 alors le peptide le moins bien corrélié a été éliminé des données.

- Pour les protéines avec quatre peptides ou plus (peptides ou états de charge) : Si moins de 4 (ou plus) peptides présentaient un facteur de corrélation supérieur à 0,8 alors les peptides les moins bien corrélés (25% des peptides) ont été supprimés des données.

Dans le cas de la protéine Aldehyde oxidase Q06278 citée en exemple, les 7 peptides présentent tous un coefficient de corrélation supérieur à 0,8. Selon nos critères, tous les peptides ont donc été conservés lors du calcul de l'approximation par protéine.

643 peptides ont été éliminés des données à cette étape soit 16 % du nombre total de peptides sur lesquels la quantification a été réalisée. L'élimination de ces peptides a conduit à l'élimination de 23 protéines seulement. Ceci indique que la majorité des peptides présentaient une bonne corrélation pour chacune des protéines et atteste de la qualité de nos résultats.

D.3. Analyse statistique

Une fois la corrélation des peptides au sein de chaque protéine vérifiée, les données protéiques (moyenne de tous les peptides retenus par protéine au sein de chaque groupe) ont ensuite été comparées statistiquement entre les 13 conditions physiologiques (pour rappel, $n = 6$).

La distribution normale des données, ainsi que l'égalité des variances, ont tout d'abord été vérifiées par les tests de Shapiro-Wilk et Bartlett (hypothèse validée pour une valeur $p > 0,01$). Les protéines ne satisfaisant pas ces critères n'ont pas été évaluées dans la suite du traitement. Nous avons pu constater qu'un nombre assez élevé de protéines ne satisfaisait pas les critères de significativité pour le test de Shapiro-Wilk alors que les données ont été transformées en données logarithmiques, ce qui est pourtant connu pour permettre de distribuer les données selon une loi normale. Ce test est donc toujours nécessaire malgré le traitement logarithmique. Les protéines qui présentaient des valeurs p supérieures aux seuils définis ont ensuite été évaluées selon un test ANOVA puis un test Post-hoc de Tukey. Les abondances relatives étaient validées comme significativement différentielles pour des valeurs p respectivement inférieures à 0,05 et 0,01.

Finalement, un ratio de sur-ou sous-expression au-dessus et/ou en deçà duquel nous avons considéré les protéines comme significativement différentielles a également été établi. Nous avons pu constater sur des échantillons dont le contenu était connu au chapitre IV « Optimisations instrumentales sur un analyseur de type Q-TOF pour l'analyse d'échantillons biologiques complexes » que ce seuil était nécessaire pour diminuer le taux de résultats « faux-positif » (soit des protéines

identifiées comme différentielles qui ne devraient pas l'être). Nous avons choisi ce seuil égal à 1,5 en raison des faibles coefficients de variation que nous avons obtenu pour l'ensemble des données et du nombre de répliques par groupe assez élevé ($n = 6$).

Les conditions J2, J8 et J14 (0 mM VPA) ont été comparées à la condition J0 (0 mM VPA). Puis les conditions J2 0,05 mM, 0,15 mM et 0,5 mM ont été comparées à J2 0 mM, de même pour les échantillons prélevés aux jours J8 et J14 qui ont été comparés à leur contrôle respectifs (J8 et J14 0 mM). Le nombre de protéines identifiées dont l'abondance a été identifiée comme statistiquement différentielle est présenté en tableau 2. La liste de ces protéines est présentée en annexe 2.

Tableau 2 : Nombre de protéines dont l'abondance relative entre les groupes est statistiquement différentielle. Chaque condition a été comparée au contrôle prélevé le même jour à l'exception des conditions contrôles J2, J8 et J14 qui ont été comparées à la condition contrôle J0.

	J2	J8	J14
ctrl	5	46	84
0.05mM	/	/	1
0.15mM	/	2	4
0.5mM	/	32	93

On peut constater qu'un très grand nombre de protéines présente une abondance relative significativement différentielle entre les conditions contrôles collectées aux différents jours de prélèvement, ce qui signifie que la simple culture des cellules induit une modification de leur protéome. Les concentrations 0,05 mM et 0,15 mM en VPA ne semblent pas induire d'effets importants sur les cellules étant donné le nombre très faible de protéines identifiées comme différentielles par rapport aux échantillons contrôles.

D.4. Interprétation préliminaire des données

Les niveaux d'expression des protéines ainsi que les valeurs statistiques obtenues pour chacune d'elle ont été soumises au logiciel Ingenuity® Pathway Analysis (IPA®) qui est un logiciel d'aide à l'interprétation des données basé, d'une part, sur l'interprétation des données expérimentales et,

d'autre part, sur des données bibliographiques. Ce logiciel permet d'aider à l'identification des voies métaboliques affectées par les différents traitements auxquelles les cellules ont été exposées.

Tout d'abord, nous avons comparé les échantillons contrôles (comparaison des jours 2, 8 et 14 par rapport au jour 0). Les voies métaboliques impactées par le temps des cellules non exposées au VPA concernent le métabolisme des acides gras. Puis, nous avons comparés tous les échantillons prélevés le 14^{ème} jour entre eux (concentrations 0,05 mM, 0,15 mM et 0,5 mM par rapport au groupe contrôle Jour 14) car le nombre de protéines différentielles détectées y était le plus élevé (tableau 2). Les voies métaboliques impactées par le traitement concernent le métabolisme lipidique, ce qui est attendu puisque l'acide valproïque est connu pour induire une stéatose du foie. L'impact sur ces voies métaboliques a été exacerbé en fonction de la concentration en VPA. Cette interprétation des données très préliminaire permet de conclure que les cellules ont répondu physiologiquement au traitement à l'acide valproïque et que les effets observés sont cohérents avec une stéatose hépatique (une accumulation d'acides gras au niveau cellulaire), validant l'utilisation du modèle de culture cellulaire HepaRG comme modèle d'hépatotoxicité en réponse à une contamination à l'acide valproïque. Les cibles d'intérêt devront être validées par une approche orthogonale comme la SRM par exemple.

IV. Conclusion

Ce chapitre a fait l'objet d'une analyse quantitative globale sur des échantillons issus de lignée cellulaire hépatique. Les difficultés majeures rencontrées consistaient à analyser un grand nombre d'échantillons et à traiter ces données massives pour lesquelles il n'existe pas encore de traitement des données uniques à l'ensemble de la communauté. Pour pallier à ces difficultés, nous avons mis en place un suivi précis du système LC-MS pendant toute la série d'analyses et nous avons mis en place un traitement des données très rigoureux allant jusqu'à la vérification des comportements peptidiques pour chacune des protéines quantifiées, ce qui nous a permis d'augmenter la confiance des données quantitatives générées. L'interprétation préliminaire des données nous a permis d'identifier des voies impactées qui sont tout à fait cohérentes avec les effets connus induits par l'acide valproïque, validant la procédure de traitement de nos données, qui pourront contribuer aux développements des algorithmes de prédiction. Le seul bémol résulte en la stabilité limitée du modèle cellulaire que nous avons pu constater au bout de 14 jours d'expérience.

Chapitre VI : Application de la stratégie quantitative ciblée SRM (Selected Reaction Monitoring) pour la quantification de la vitellogénine dans le plasma de tortue Luth

Chapitre VI : Application de la stratégie quantitative ciblée SRM pour la quantification de la vitellogénine dans le plasma de tortue Luth

Ce projet a été réalisé en collaboration avec le Dr Jean-Yves Georges, DEPE, IPHC, Strasbourg.

I. Contexte de l'étude

La tortue Luth (*Dermochelys Coriacea*, Dc) étant une espèce en voie de disparition, son écologie et sa physiologie font aujourd'hui l'objet de nombreuses études. Les résultats de ces études devraient permettre une meilleure gestion et conservation de cette espèce [144]. Cependant, la période de reproduction des tortues marines pouvant être très longue et les tortues venant pondre sur les plages (par exemple en Guyane pour la tortue Luth) essentiellement durant la nuit et ce, au cours d'une dizaine d'épisodes de ponte espacés de plus d'une semaine, il est très difficile de suivre rigoureusement la totalité des événements de ponte chez un individu pendant la saison. De plus, il n'existe pas de marqueur de statut reproducteur qui permette de connaître facilement le degré d'avancement d'un individu au cours de la saison de ponte. Or, disposer d'une telle information serait crucial pour l'étude des déterminants de la reproduction chez les tortues marines. Par conséquent, le but de ce projet était de développer une mesure quantitative fiable pour connaître le degré d'avancement des tortues Luth pendant la saison de ponte.

Chez les ovipares, la vitellogénine (VTG) est synthétisée par le foie en réponse à une augmentation de la concentration en œstrogène. La VTG est donc aujourd'hui communément utilisée pour mesurer l'exposition aux polluants aquatiques comme les perturbateurs endocriniens [145]. Elle est aussi utilisée pour sexer les animaux puisque le mâle ou les individus immatures sexuellement ne la synthétisent pas naturellement, bien qu'ils en aient la capacité [146]. Chez la femelle, la VTG étant impliquée dans la vitellogénèse, elle est utilisée comme marqueur du statut reproducteur chez *Chelonia mydas* par exemple [147]. Le but de ce projet a donc été de mettre au point une méthode d'analyse quantitative ciblée de cette protéine dans le plasma de tortue Luth, puis de vérifier si son niveau d'expression était effectivement corrélé à l'avancement dans la saison de reproduction.

La principale difficulté du projet est que le protéome de cette tortue n'est pas connu. En amont de ce travail de quantification ciblée, le séquençage de la vitellogénine chez *Dermochelys Coriacea* (Dc) avait été réalisé par une approche de protéomique qualitative combinant de multiples conditions de digestion enzymatique et de multiples modes d'interprétation des données spectrales incluant une

stratégie de séquençage *de novo* massive (travaux de thèse de Thierry Wasselin). Ces stratégies ont été décrites dans le chapitre II « Stratégie d'analyse protéomique qualitative » en page 37. Les séquences de plus de 300 peptides de vitellogénine avaient alors été déterminées et l'existence de deux isoformes différentes de VTG (Dc-VTG1 et Dc-VTG2) avait été montrée, ce qui n'était alors pas connu chez les reptiles.

Au cours de ma thèse, à partir des séquences préalablement déterminées par T. Wasselin, j'ai pris en charge le développement de la méthode de quantification ciblée (SRM) des deux isoformes de VTG chez la tortue Luth. A la suite de la mise au point de la méthode, nous avons pu analyser les échantillons de plasma de deux tortues pour lesquelles des échantillons avaient été collectés au cours de la plupart des épisodes d'une saison de pontes.

Ces travaux ont donné lieu à une publication dans une revue scientifique à comité de lecture [56]. La publication réalisée est présentée à la suite de ce chapitre. Les principales étapes de mises au point de la méthode ainsi que les résultats obtenus sont présentés dans ce chapitre.

II. Choix de la stratégie analytique

Des tests existent pour la quantification ou la détection de la VTG chez les reptiles [145]. Ils sont basés sur l'utilisation d'anticorps tels que les tests ELISA ou Western-Blot. Or, ces tests sont spécifiques (anticorps dirigés contre des séquences précises) et la variabilité inter-espèces ne permet pas de les transposer d'une espèce à l'autre simplement [148]. La stratégie SRM permet de quantifier simultanément plusieurs peptides, ce qui correspondrait à plusieurs analyses Western-Blot (ou tests ELISA) en parallèle. Ainsi, la quantification réalisée est beaucoup plus précise dans la mesure où plusieurs peptides vont permettre la quantification d'une seule protéine. De la même manière, la stratégie SRM permet de quantifier simultanément et spécifiquement plusieurs isoformes dans la mesure où nous sommes capables avec cette technique de différencier, et de quantifier, des peptides uniques à chacune d'elles.

Les standards protéiques de type PSAQTM, présentés dans le chapitre III « Stratégies d'analyses protéomiques quantitatives », ne peuvent être utilisés pour la quantification ciblée de protéines chez des espèces dont le génome n'est pas connu comme *Dermochelys coriacea* car la synthèse de la protéine recombinante nécessite la connaissance de la séquence du gène associé [93]. Il n'est donc pas possible d'utiliser d'autres standards que les peptides pour des organismes non séquencés. La stratégie QconCAT n'a pas été retenue car il est connu que les protéines QconCAT (présentées dans

le chapitre III « stratégies d'analyses protéomiques quantitatives » en page 56) ont un comportement différent de celui des protéines endogènes à quantifier au stade de la digestion enzymatique.

Le choix d'analyser le plasma s'est fait étant donné la facilité de prélèvement de ce type de fluide chez cette espèce. La gamme dynamique de concentration des protéines y est la plus élevée de tout l'organisme (environ 10 ordres de grandeurs) [8], ce qui aurait pu être un frein à une mesure précise de concentration. Seulement, la forte abondance de la protéine (de l'ordre du μM) pendant la saison de ponte a permis de palier à cette difficulté.

III. Les différentes étapes du développement de la quantification par LC-SRM des VTG de tortue Luth

Le principe de la quantification SRM, ainsi que les étapes de développement de ce type de stratégie, a été présenté au sein du chapitre III "Stratégies d'analyses quantitatives en protéomique" en page 50.

Tous les peptides sélectionnés pour le développement de notre méthode présentent une séquence strictement homologue à celles de peptides de VTG d'autres tortues (*Chrysemys picta*, *Mauremys reevesi*). Etant donné la conservation de ces peptides chez *Chrysemys picta* et *Mauremys reevesi*, il est possible que ces séquences soient également conservées chez les autres espèces de tortues pour lesquelles le protéome n'est pas encore séquencé et donc que cette méthode soit directement transposable chez d'autres espèces. Cette stratégie nous a également permis de nous affranchir de toute ambiguïté concernant la validation des séquences déterminées par la stratégie *de novo*. En revanche, le séquençage a permis de sélectionner des peptides protéotypiques se trouvant sur différents domaines ou différentes chaînes de chacune des isoformes, permettant ainsi de quantifier finement chacune d'elles.

Les peptides ont été sélectionnés selon les critères présentés dans le chapitre "Stratégies d'analyses quantitatives en protéomique" en page 53 à l'exception de l'isoforme Dc-VTG1 où un peptide contenant deux méthionines a été sélectionné (MTPVLLPEAVPDIMK). Il s'agissait du seul peptide qui présentait une intensité suffisante pour permettre sa quantification. En général, il est préférable d'éviter la sélection de ce type de peptide qui risque de présenter des formes modifiées (oxydées) qui pourraient induire un biais dans la quantification [81]. Une technique a été reportée dans la littérature permettant de forcer l'oxydation de toutes les méthionines et ainsi d'éviter la présence de

ce type de peptide sous plusieurs formes [149]. En revanche, ces oxydations forcées risquent d'induire des modifications chimiques telle que la formylation de lysine par exemple [150].

Au stade de la mise au point de la méthode, l'ensemble des formes oxydées et non oxydées du peptide a été évalué dans les échantillons de plasma tests. Nous avons pu constater que les niveaux d'intensité des formes oxydées étaient proches des valeurs du bruit de fond. La proportion de peptides portant une ou deux méthionines oxydées a donc été considérée comme négligeable. Les quatre formes endogènes (modifiées ou non) ont été cependant suivies spécifiquement dans l'ensemble des analyses réalisées à partir des échantillons de plasma pour les deux tortues au cours de la saison de ponte afin de contrôler leur abondance.

Concernant l'isoforme Dc-VTG2, quatre peptides ont pu être sélectionnés pour sa quantification (QELTLVEVK, VASPTLESVFK, IQLEIQAGSR et ISSEVTGR).

Le choix des transitions à suivre pour chaque peptide ainsi que l'optimisation des énergies de fragmentation a été réalisé manuellement grâce aux peptides standards, dans la mesure où il n'existe pas d'outil de prédiction ou de base de données SRM pour cette espèce. Nous avons sélectionné trois transitions par peptide. Le temps de scan a été fixé à 160 ms par transition en accord avec le nombre de peptides à quantifier, la largeur des pics chromatographiques et les performances de l'instrument utilisé.

Nous avons également établi des courbes de linéarité pour l'ensemble des peptides sélectionnés à partir des échantillons de plasma tests. Ces courbes de linéarité, réalisées à partir de trois répliques d'analyses, représentent la somme des aires sous les courbes intégrées des trois transitions monitorées par peptide en fonction de la quantité de peptide standard injectée. Les critères que nous avons fixés pour valider la linéarité de réponse des peptides sont les suivants :

- Coefficients de variation entre les répliques d'analyse inférieurs à 20 % calculés d'après les aires sous les courbes des transitions sommées pour chaque peptide
- Tolérance de 20% sur le facteur de réponse entre les abondances mesurées et les quantités injectées
- Tolérance de $\pm 20\%$ sur les quantités mesurées expérimentalement et les quantités théoriques injectées

Nous avons également vérifié, à posteriori, que l'ensemble des concentrations endogènes mesurées étaient comprises dans les gammes de linéarité. Ces courbes nous ont permis de définir une limite de quantification et une limite de détection pour chacun des peptides. Pour cette étape, nous n'avons pas cherché à atteindre les performances maximales de l'instrument mais simplement à couvrir la gamme d'analyse d'intérêt.

Il a ensuite été nécessaire de définir quelle quantité de standard introduire dans les échantillons de façon à avoir l'abondance relative la plus proche possible de 1 entre chaque peptide standard et son homologue endogène. Les concentrations de chaque peptide standard ont donc été ajustées à partir d'échantillons tests pour lesquels les niveaux d'expressions de VTG étaient supposés extrêmes.

Finalement, nous avons mis en place un système de contrôle de façon à évaluer la stabilité du couplage pendant la durée des analyses de plasma des deux tortues prélevées à plusieurs reprises pendant une saison de ponte. Un des échantillons à doser a été analysé tous les deux jours de façon à s'assurer de la stabilité du couplage LC-QQQ utilisé. Nous avons également évalué la reproductibilité des aires sous les courbes des courants d'ions des peptides standards dans l'ensemble des échantillons. Le seuil limite en termes de coefficient de variation était fixé à 20 %.

IV. Résultats et discussion

A. Evaluation de la linéarité de réponse des peptides quantifiés dans des échantillons de plasma

Les courbes de linéarité obtenues pour chacun des peptides sont présentées en figure VI-1. Les concentrations endogènes mesurées à partir des plasmas des deux tortues, calculées à l'issue des analyses, sont représentées par des cadres bleus.

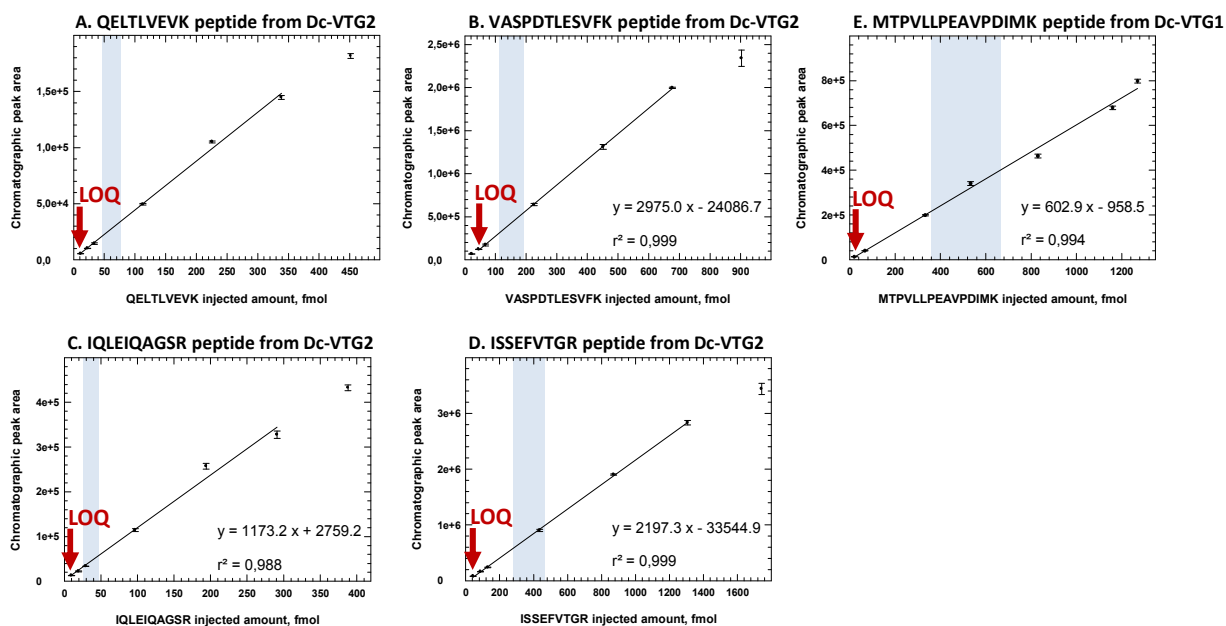


Figure VI -1 Evaluation de la linéarité de réponse des 5 peptides ayant permis la quantification des deux isoformes Dc-VTG1 et Dc-VTG2 dans des échantillons de plasma

Les linéarités de réponse de l'ensemble des peptides s'étalent sur environ 2 ordres de grandeur. Les cadres bleus correspondent aux concentrations endogènes mesurées dans le plasma des deux tortues pour lesquelles des échantillons de plasma ont été collectés pendant toute une saison de ponte. Celles-ci sont en parfaite adéquation avec les linéarités de réponse de chacun des peptides.

Ces courbes démontrent une linéarité de réponse sur une gamme d'environ 2 ordres de grandeur. On constate sur la figure VI-1 que les gammes de linéarité de réponse sont en parfaite adéquation avec les concentrations endogènes mesurées. Ces courbes nous ont permis de définir des limites de détection et de quantification qui sont présentées dans le tableau 1.

Tableau 1 : Limites de détection et de quantification établies pour chacun des peptides quantifiés

Séquence peptidique	Limite de detection (fmol)	Limite de quantification (fmol)
Dc-VTG2		
QELTLVEVK	11,3	11,3
QELTLVEV [Lys(¹³ C6; ¹⁵ N2)]		
VASPDLTLESVFK	22,6	45,1
VASPDLTLESVF [Lys(¹³ C6; ¹⁵ N2)]		
IQLEIQAGSR	9,7	9,7
IQLEIQAGS [Arg(¹³ C6; ¹⁵ N4)]		
ISSEVTGR	43,6	43,6
ISSEVTG [Arg(¹³ C6; ¹⁵ N4)]		
Dc-VTG1		
MTPVLLPEAVPDIMK	20,0	20,0
MTPVLLPEAVPDIM [Lys(¹³ C6; ¹⁵ N2)]		

B. Evaluation de l'abondance des formes oxydées du peptide MTPVLLPEAVPDIMK

L'ensemble des formes modifiées de ce peptide ont également été suivies par le spectromètre de masse dans l'ensemble des échantillons de plasma à doser. Le signal le plus intense acquis pour la forme modifiée la plus abondante (MTPVLLPEAVPDIMK) est comparé au signal le moins intense acquis pour la forme non modifiée de ce même peptide en figure VI-2. Un facteur 100 a été observé entre ces deux formes, ceci a permis de valider la proportion négligeable de ces formes modifiées dans les échantillons de plasma.

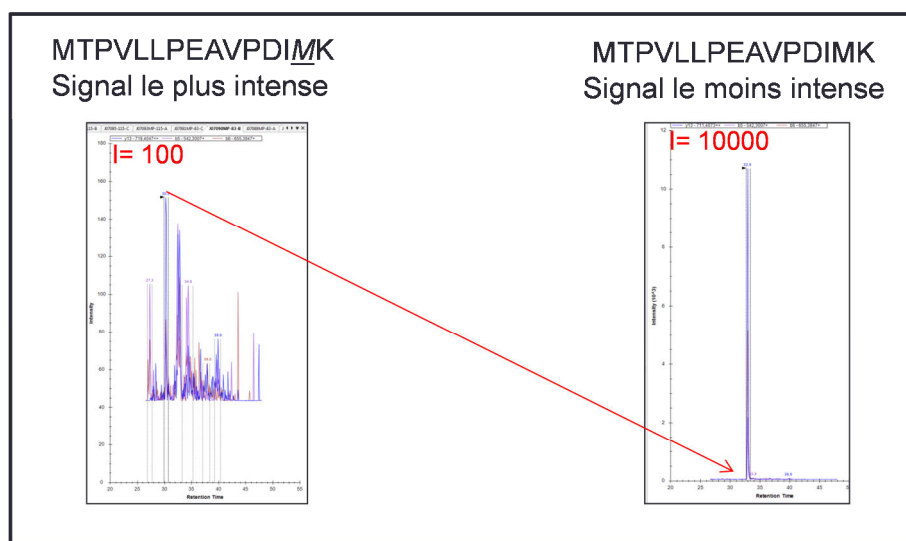


Figure VI-2 Illustration des abondances relatives des formes modifiées et non modifiées du peptide MTPVLLPEAVPDIMK dans les échantillons de plasma

Un facteur 100 a été constaté entre les signaux les plus intenses enregistrés pour la forme modifiée la plus abondante et la forme non modifiée la moins abondante dans les échantillons de plasma des deux tortues suivies au cours de la saison de ponte. La proportion de ces formes oxydées a donc été confirmée comme négligeable.

C. Suivi de la stabilité du système pendant les analyses des échantillons de plasma des deux tortues suivies au cours de la saison de ponte.

La quantification étant basée directement sur les données spectrales, le couplage instrumental a fait l'objet d'un suivi rigoureux tout au long des analyses réalisées. Ce type de contrôle a permis de garantir la stabilité de nos échantillons et la robustesse du couplage instrumental. Pour cela, deux contrôles ont été effectués. Le premier consistait à analyser un même échantillon toutes les 48h. Le coefficient de variation obtenu pour les cinq peptides suivis était inférieur à 20% sur toute la durée de l'expérience. Le second contrôle consistait à vérifier que le coefficient de variation calculé à partir de l'ensemble des aires sous la courbe pour les peptides standards dans tous les échantillons était

bien inférieur à 20%. La stabilité du couplage a ainsi été confirmée ce qui renforce la fiabilité des données quantitatives générées.

D. Concentrations des isoformes Dc-VTG1 et Dc-VTG2 dans les échantillons de plasma de tortue Luth collectés pendant la saison de ponte.

La validation de la linéarité de réponse des peptides ainsi que le suivi rigoureux du système nous a permis d'évaluer les niveaux de concentrations de la VTG dans le plasma de deux tortues au cours d'une saison de ponte. Les niveaux d'expression que nous avons évalués sont représentés en fonction de l'avancement dans la saison de ponte (LRE) pour chacune des tortues en figure VI-3.

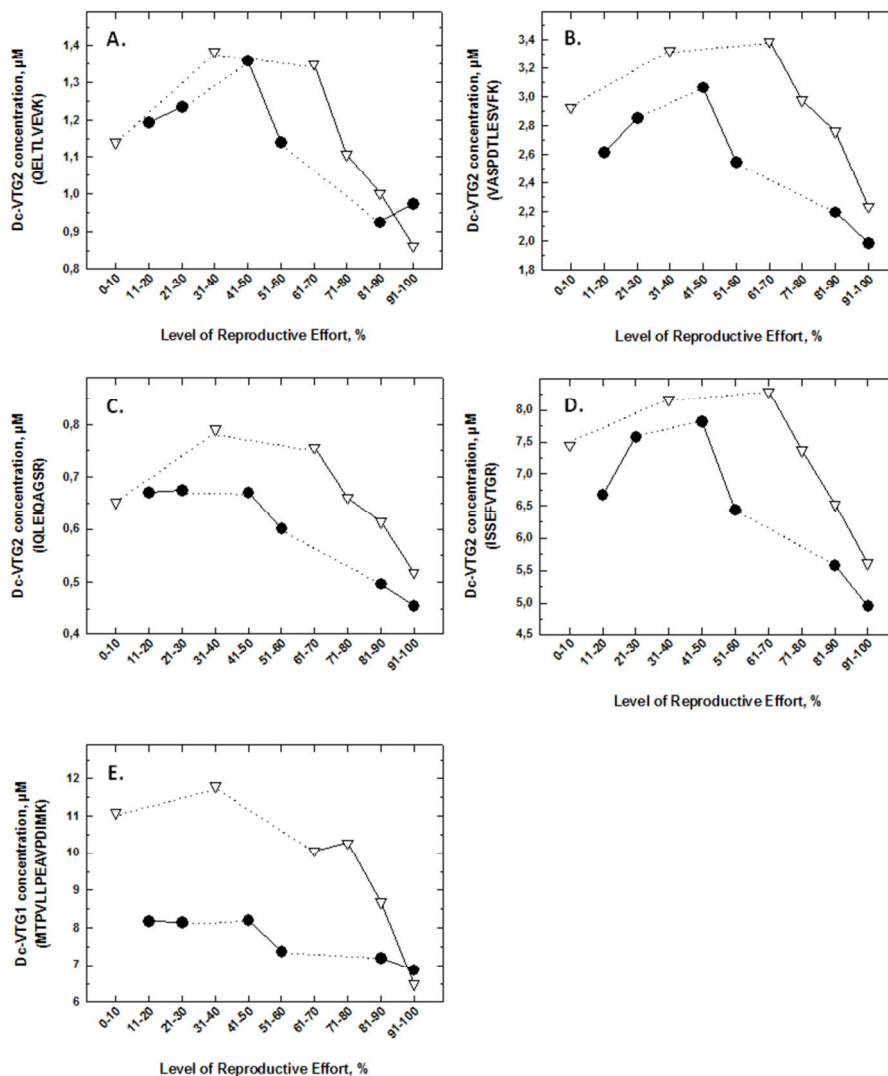


Figure VI-3 Profils de concentration des isoformes Dc-VTG1 et Dc-VTG2 chez deux tortues Luth au cours d'une saison de ponte

Les graphiques représentent les niveaux de concentration mesurés pour chaque peptide en fonction de l'avancement dans la saison de ponte. Les pointillés représentent les pontes pour lesquelles des échantillons n'ont pas pu être collectés. Une tendance à la diminution égale à un facteur 2 est observée pour l'ensemble des peptides quantifiés.

Les concentrations mesurées sont comprises entre 6 et 12 μM pour Dc-VTG1 tandis qu'elles sont comprises entre 0,5 et 9 μM pour Dc-VTG2. Les concentrations mesurées pour les 4 peptides de Dc-VTG2 sont comprises entre 1 et 3 μM pour les peptides QELTLVEVK et VASPDTLSEVFK, entre 0,5 et 1 μM pour IQLEIQAGSR et entre 5 et 9 μM pour ISSEFVTGR.

Les niveaux d'expression des 4 peptides quantifiés pour l'isoforme Dc-VTG2 présentent quelques différences (jusqu'à un facteur 10). La comparaison de la séquence Dc-VTG2 à la séquence VTG2 de l'espèce *Gallus Gallus* avec laquelle elle présente de fortes similarités, nous indique qu'il est possible que les séquences nucléotidiques des gènes codant pour ces protéines soient également similaires. L'étude de la séquence nucléotidique de la poule nous indique que les séquences nucléotidiques respectives des quatre peptides monitorés pour Dc-VTG2 se situent sur des exons différents et qu'il est donc probable que les différences de niveaux d'expressions que nous avons observés résultent de la présence de différents ARNm pour cette protéine. Il est également possible que ces différences de concentrations résultent de la présence de multiples formes processées dans le plasma. L'efficacité de digestion est peut-être également responsable de ces différences.

Cependant, les profils obtenus pour les niveaux d'expression des 5 peptides quantifiés (4 peptides pour Dc-VTG2 et 1 peptide pour Dc-VTG1) sont similaires, à savoir une décroissance d'un facteur 2 environ au cours de la saison de ponte. La précision de la quantification pour l'isoforme 1 est un peu limitée dans la mesure où un seul peptide a été quantifié. En revanche, la tendance à la diminution de la concentration mesurée pour cette isoforme est en adéquation avec celles des peptides de Dc-VTG2 qui diminuent également au cours de la saison de ponte.

V. Conclusion et perspectives

Grâce au développement de cette méthode d'analyse ciblée, nous avons pu constater que les niveaux d'expression de la VTG chez les deux tortues étudiées semblent être directement liés au degré d'avancement dans la saison de ponte (LRE). A la suite de ces travaux, nous avons pu analyser un plus grand nombre de plasmas collectés chez une vingtaine de tortue. Les résultats obtenus sont cohérents avec les niveaux d'expression mesurés pendant la première campagne : ceux-ci diminuent également d'un facteur 2 en fonction de l'avancement dans la saison de ponte. L'exploitation de ces dernières données SRM est en cours par nos collaborateurs. Puisque les niveaux de VTG semblent liés à ceux de l'effort reproducteur et que ces résultats ont été confirmés par l'analyse d'échantillons de plasma chez une vingtaine d'individus différentes, les variations de tout autre paramètre

physiologique mesuré chez les tortues Luth pourront donc maintenant être interprétées avec une meilleure estimation du statut reproducteur des individus concernés.

Chapitre VII : Applications et optimisations de la stratégie quantitative 2D-DIGE (Differential In-Gel Electrophoresis) à des échantillons biologiques complexes

Chapitre VII : Applications et optimisations de la stratégie quantitative 2D-DIGE (Differential In-Gel Electrophoresis) à des échantillons biologiques complexes

Quatre applications de la stratégie quantitative 2D-DIGE sont présentées dans ce chapitre. Les choix méthodologiques effectués pour l'ensemble de ces applications sont discutés conjointement à la fin du chapitre.

I. Etude des compromis évolutifs

Les deux projets présentés dans cette partie ont été réalisés en collaboration avec le Dr. François Criscuolo, DEPE, IPHC, Strasbourg.

A. Etude du coût de la reproduction chez la souris

A.1. Contexte

La biologie évolutive s'intéresse à la diversité des espèces existantes sur Terre, et cherche à identifier les mécanismes ayant conduit, au travers de l'évolution, à une telle diversité [151]. Plus précisément, l'écologie et l'écophysiologie évolutive cherchent à comprendre comment les organismes font face à des compromis (par exemple des compromis énergétiques), c'est à dire comment ils répondent aux variations de leur environnement et quelles sont les conséquences de ces réponses. Le concept central de la « théorie des traits d'histoire de vie » stipule que les processus physiologiques entrent en compétition lorsque la disponibilité des ressources est limitée [152]. Par conséquent, si deux processus partagent les mêmes ressources, alors toute augmentation de l'allocation des dites ressources à l'un de ces processus se fera au détriment de l'autre. C'est pourquoi de tels compromis peuvent entraîner des effets délétères.

Les ressources énergétiques proviennent des réserves endogènes et de l'alimentation. Elles sont de manière basique allouées au maintien des fonctions vitales et à la croissance de l'organisme. Ceci permet d'assurer l'intégrité et la survie de l'organisme. Lorsqu'un animal se reproduit, une partie de l'énergie qui était allouée à la survie et l'intégrité de son organisme est réallouée à la reproduction, d'où de possibles effets délétères ou « coûts » [153, 154]. La figure VII-1 illustre la répartition de l'énergie au sein des différentes fonctions d'un organisme dans un environnement donné.

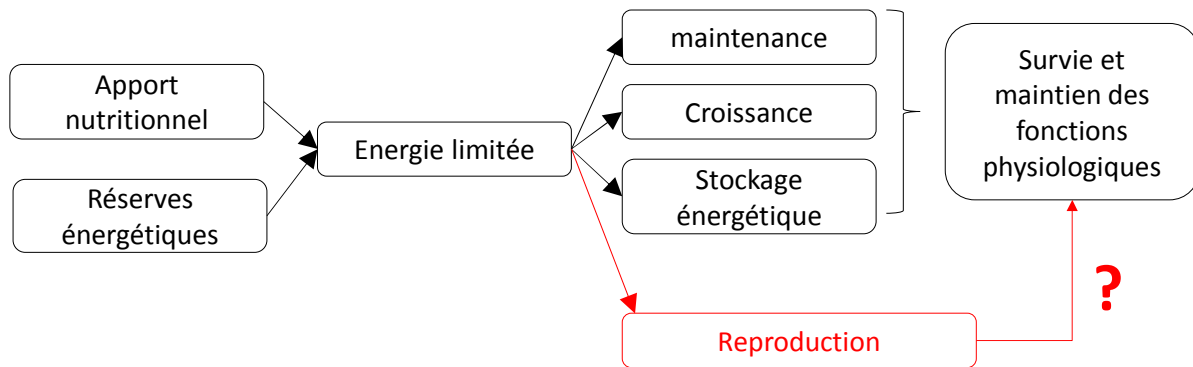


Figure VII-1 Concept de la « théorie des traits d'histoire de vie » appliqué aux compromis énergétiques
L'énergie est répartie entre différentes composantes au sein d'un organisme. Si celui-ci s'implique dans la reproduction, alors une partie de l'énergie disponible n'est plus allouée aux fonctions vitales et il y a un risque pour la survie de l'individu.

Ce projet repose sur l'intérêt grandissant que porte la communauté des écophysiologistes, depuis quelques années seulement, à l'approche protéomique [155]. Il s'agissait en effet de mettre en évidence de nouvelles réponses moléculaires induites dans le contexte des coûts de la reproduction grâce à l'outil protéomique. Les coûts de la reproduction sont de plus en plus étudiés et impliquent entre autres des régulations hormonales, des altérations du métabolisme énergétique, de la fonction immunitaire ou encore l'induction d'un stress oxydatif [154]. Or, la nature des mécanismes induits par exemple au niveau hépatique reste peu connue, d'autant plus que des études récentes semblent invalider l'induction du stress oxydatif hépatique lors d'une manipulation de l'effort de reproduction [156]. Pourtant, puisque le foie est le carrefour clé du métabolisme et qu'il est très sensible à tout changement métabolique et oxydatif [157], l'hypothèse d'un effet de l'effort de reproduction à ce niveau est plausible. Le but de ce projet est donc d'identifier le protéome différentiel hépatique entre différents groupes de souris pour lesquels l'effort reproducteur a été manipulé dans le but d'identifier de nouvelles voies métaboliques qui pourraient être impliquées dans le coût de la reproduction.

Ce projet a fait l'objet d'une publication scientifique dans une revue à comité de lecture présentée dans la suite du chapitre [158].

A.2. Schéma expérimental

L'effort de reproduction a été manipulé chez des souris afin de comparer les effets d'un effort augmenté à ceux d'un effort diminué et d'un effort non manipulé. C'est l'effort de lactation qui a été manipulé étant donné qu'il correspond à la composante la plus demandeuse en énergie chez les mammifères [159]. Une analyse quantitative différentielle du protéome hépatique chez les

différents groupes de souris (contrôle, effort augmenté, et effort diminué ; $n = 4/\text{groupe}$) a ensuite été réalisée par 2D-DIGE (« minimal labeling »).

A.3. Analyse des échantillons

La préparation des échantillons ainsi que le traitement des données sont détaillées dans la publication présentée dans la suite du chapitre.

En résumé, les morceaux de foie ont été broyés à froid et à sec dans de l'azote liquide puis solubilisés dans un tampon dénaturant contenant des détergents et des agents chaotropes. Les protéines ont ensuite été précipitées dans le but d'éliminer la présence d'interférents comme les lipides ou les sels. Après resolubilisation des protéines, celles-ci ont été marquées avec les cyanines Cy3 et Cy5 selon le schéma défini dans la partie expérimentale. Dans le but de minimiser les différences dues aux coefficients d'extinction de chaque CyDyes, nous avons mélangé les marqueurs Cy3 et Cy5 pour chacune des conditions analysées. Les protéines présentes dans un mélange de tous les échantillons, représentant notre standard interne, ont été marquées avec la cyanine Cy2.

Deux échantillons appartenant à deux conditions physiologiques différentes (taille de nichée réduite, augmentée, contrôle) ainsi que le standard interne ont été mélangés puis déposés dans une bande de gel à gradient de pH immobilisé dans le but de séparer les protéines, sous l'effet d'un potentiel électrique, en fonction de leur point isoélectrique. Ces protéines ont ensuite été séparées dans une deuxième dimension selon leur masse moléculaire dans un gel d'acrylamide de type SDS-Page grâce à un système d'électrophorèse vertical. Les gels ont été scannés aux longueurs d'onde d'émission et d'absorption des trois cyanines, générant ainsi trois images par gel. Les images ont ensuite été réalignées par le logiciel d'analyse d'image Samespots (*Nonlinear Dynamics*) qui a permis d'évaluer les abondances relatives des spots entre les images de gel par densitométrie. Le correct alignement des gels réalisé par le logiciel de traitement a été vérifié manuellement pour l'ensemble des spots évalués. En parallèle des gels 2D-DIGE, un autre gel d'électrophorèse pour lequel une plus grande quantité d'extrait protéique avait été chargé a été réalisé. Après alignement de l'image de ce gel coloré au bleu de Coumassie avec celles des gels 2D-DIGE, 384 spots de ce gel ont été excisés dans le but d'identifier les protéines contenues dans ces spots. La validation de l'alignement des deux types de gel a été réalisée d'après l'analyse de protéines contenues dans des spots excisés à partir de chacun des types de gel. Après excision des spots, les ponts disulfures des protéines présentes dans ces spots ont ensuite été réduits puis alkylés et nous avons procédé à une digestion enzymatique in-gel. Après extraction, les peptides issus de ces protéines ont été analysés avec un couplage nanoLC-

ChipCube-IT à l'aide de méthodes d'acquisition de type nanoLC-MS/MS. L'identification des protéines a été réalisée selon la stratégie décrite dans le chapitre II « Stratégie d'analyse qualitative en protéomique » concernant l'analyse du protéome d'espèces séquencées d'après la génération de spectres MS/MS. En plus du maintien du taux de « faux-positif » inférieur à 1% grâce à l'utilisation de la stratégie « target-decoy », deux algorithmes de recherche (Mascot et Omssa) ont été utilisés de façon à augmenter la confiance dans les identifications réalisées. Finalement, l'identification des protéines responsables des variations d'intensité des spots a été réalisée d'après la comparaison du nombre de peptides identifiés pour chacune des protéines dans chaque spot différentiel.

A.4. Résultats

Les analyses par spectrométrie de masse (piège à ions, HCT Plus, Bruker Daltonique) des protéines contenues dans ces spots ont permis la détection et l'identification d'environ 480 protéines uniques avec un taux de faux-positif inférieur à 1%. Plus de 65% des protéines identifiées l'ont été avec plus de 5 peptides. Environ 90% de protéines ont été identifiées avec les deux algorithmes de recherche.

Un exemple d'image de gel 2D-DIGE obtenu est présenté en figure VII-2. L'analyse d'image des gels 2D-DIGE ainsi que l'évaluation statistique des données ont révélées 9 spots protéiques différentiels. La distribution normale des données a été vérifiée par le test de Shapiro-Wilk. L'hypothèse a été validée pour une valeur p supérieure à 0,01. L'égalité des variances a été vérifiée grâce au test de Bartlett. Cette hypothèse a été également validée pour une valeur p supérieure à 0,01. La significativité des abondances relatives des spots a été testée par une analyse ANOVA (le seuil de significativité a été établi pour une valeur p inférieure à 0,05) puis par deux analyses : en composante principale et discriminante. La répartition des spots différentiels sur le gel ainsi que les protéines identifiées dans ces spots sont présentées en annexe 3.

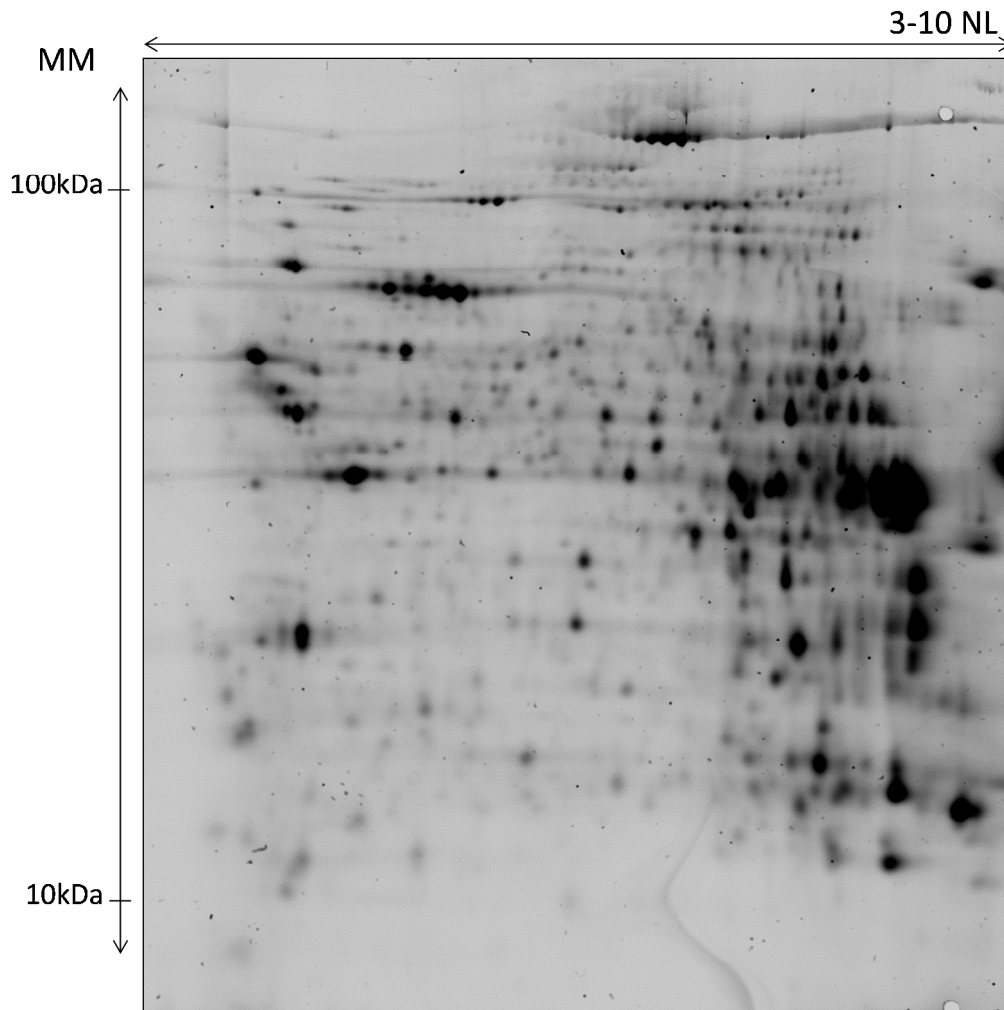


Figure VII-2 Image de gel 2D-DIGE Cy2 obtenue à partir des échantillons de foie
150 µg de protéines ont été déposées sur le gel (50µg pour Cy3, Cy5 et Cy2) La bande de gel à gradient de pH immobilisé utilisée pour la première dimension couvrait la gamme de pH de 3 à 10 (non linéaire). La réticulation du gel en polyacrylamide était de 10%. Les gels ont été scannés à l'aide d'un Ettan DIGE Imager. Cette image a été générée d'après les longueurs d'onde d'émission et d'absorption du fluorochrome Cy2.

L'analyse statistique de ces données a permis l'identification de deux fonctions discriminantes pour les trois groupes comparés.

- Une première fonction discrimine le groupe dont la taille de nichée a été réduite par rapport aux groupes contrôle et celui dont la taille de la nichée a été augmentée. Les protéines impliquées au sein de cette fonction (la regucalcin par exemple) reflètent l'implication des mécanismes des ions Ca^{2+} ainsi que des processus de vieillissement.
- Une seconde fonction discrimine le groupe dont la taille de nichée a été augmentée par rapport au groupe contrôle. Les protéines impliquées au sein de cette fonction (glycine N-methyl transferase et malate dehydrogenase) sont également impliquées dans des processus de vieillissement.

A.5. Conclusion

L'analyse protéomique différentielle réalisée à partir de ces trois groupes de souris n'a pas permis de détecter de réponse moléculaire hépatique spécifique d'un stress oxydatif, ce qui est cohérent avec la littérature [156]. En revanche, notre étude a permis d'identifier de possibles nouveaux mécanismes à l'origine des coûts induits par la reproduction, liés aux processus de vieillissement. Ces résultats donnent donc de nouvelles pistes d'études concernant la compréhension des coûts de la reproduction.

B. Etude du coût de l'activation du système immunitaire chez la souris

B.1. Contexte

Comme le travail présenté précédemment, cette étude est basée sur la compréhension des mécanismes liés aux compromis énergétiques, dans le contexte de la « théorie des traits d'histoire de vie ». Comme la composante reproductrice, le maintien d'un système immunitaire efficace ou l'activation de celui-ci en cas d'infection est demandeur en énergie, ce qui implique des compromis (notamment énergétiques) et des coûts associés [160]. Par ailleurs, les coûts sont ici dépendants de l'âge de l'individu exposé [161-163]. En effet, l'efficacité du système immunitaire diminue avec l'âge, un phénomène connu sous le nom d'immunosénescence [162] [164].

B.2. Schéma expérimental

En collaboration avec le Dr F. Criscuolo, nous avons exposé des souris à un antigène non-viable (le LPS ; lipopolysaccharide) afin d'induire une activation du système immunitaire. Ceci a été réalisé chez des souris plus ou moins âgées (1 mois et 18 mois), tandis que des souris du même âge (1 mois et 18 mois) recevaient une injection contrôle de PBS (Phosphate buffered saline). Une analyse quantitative différentielle (2D-DIGE-MS) du protéome hépatique chez les différents groupes de souris (contrôle « jeunes », contrôle « âgés », LPS « jeunes » et LPS « âgés » ; n = 4/groupe) a ensuite été réalisée.

B.3. Analyse des échantillons

Les étapes de traitement des échantillons, ainsi que les étapes d'analyse des données sont similaires à celles réalisées pour le projet concernant l'étude du coût de la reproduction.

B.4. Résultats

Environ 400 spots de gels ont été excisés à partir du gel coloré au bleu de Coumassie. Les analyses par spectrométrie de masse (piège à ions, HCTPlus, Bruker Daltonique) des protéines contenues dans ces spots ont permis la détection et l'identification d'environ 200 protéines uniques avec un taux de faux-positif inférieur à 1%. Après avoir vérifié la distribution normale des données (test de Shapiro-Wilk, hypothèse validée pour $p > 0,01$) et l'égalité des variances (test de Bartlett, hypothèse validée pour $p > 0,01$), les abondances relatives des spots de gel ont été testées par une analyse de type ANOVA puis Post-hoc de Tukey (les abondances relatives étaient considérées comme différentielles pour une valeur p inférieure à 0,05). L'analyse de ces données a révélé 17 spots protéiques différentiels. Un exemple d'image de gel 2D-DIGE obtenu est présenté en figure VII-3. La répartition des spots différentiels sur le gel ainsi que les protéines identifiées dans ces spots sont présentées en annexe 4.

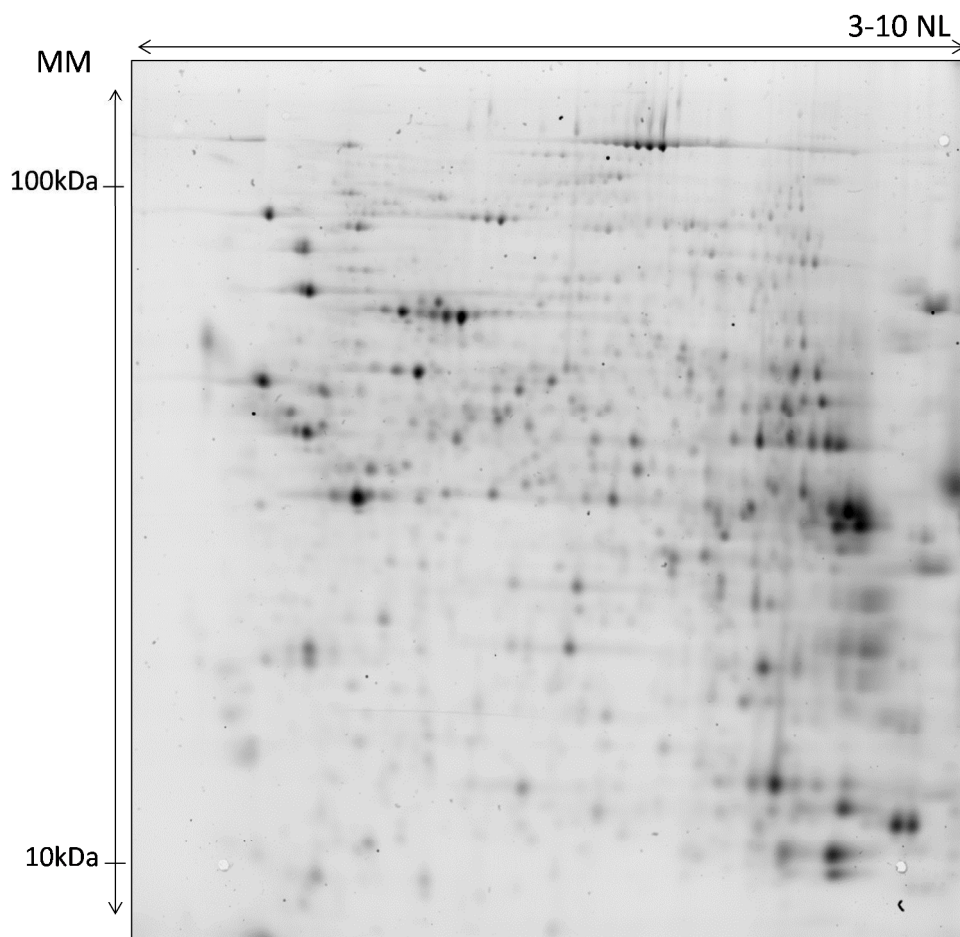


Figure VII – 3 Image de gel 2D-DIGE obtenu pour le standard interne marqué avec Cy2. 150 μg de protéines ont été déposés sur le gel (50 μg pour Cy3, Cy5 et Cy2). La bande de gel à gradient de pH immobilisé utilisée pour la première dimension couvrait la gamme de pH de 3 à 10 (non linéaire). La réticulation du gel en polyacrylamide était de 10%. Les gels ont été scannés à l'aide d'un Ettan DIGE Imager. Cette image a été générée d'après les longueurs d'onde d'émission et d'absorption du fluorochrome Cy2.

En plus des statistiques réalisées en amont, une analyse préliminaire des résultats a été réalisée par notre collaborateur, le Dr François Criscuolo. Une analyse en composante principale a permis de dégager trois facteurs qui permettent de discriminer les 4 groupes de souris (figure VII-4). La catégorisation fonctionnelle des protéines a été obtenue grâce à l'extraction automatique des annotations fonctionnelles des protéines à partir de la base de données « Gene Ontology » à l'aide de la suite logicielle MSDA développée au LSMBO [79].

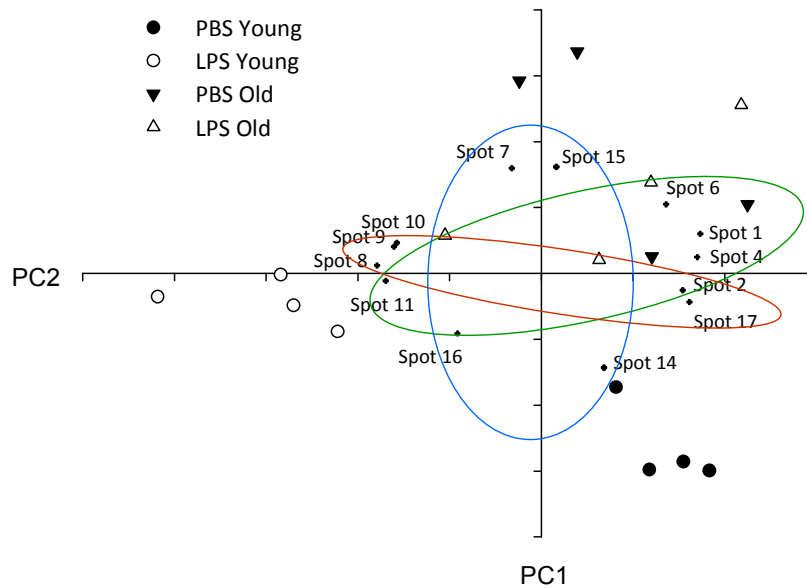


Figure VII-4 Résultats de l'analyse en composante principale
Trois facteurs permettent de discriminer les groupes d'animaux.
Les protéines impliquées dans ces composantes sont impliquées dans
1) (en vert) la réponse au stress,
2) (en rouge) la réponse immunitaire et certaines voies de synthèse
3) (en bleu) les voies cataboliques.

Le premier facteur discrimine les groupes de souris jeunes et âgées exposées au LPS. Chez les souris jeunes, les protéines impliquées sont entre autres liées aux mécanismes de réponse des cellules exposées à un stress oxydatif, alors que cette réponse est peu marquée chez les souris plus âgées.

Le deuxième facteur discrimine les groupes de souris jeunes exposées ou non au LPS. Les protéines surexprimées chez le groupe exposé au LPS sont, comme attendu, liées à l'activation du système immunitaire.

Le troisième facteur discrimine les groupes de souris jeunes et plus âgées. Les protéines impliquées reflètent une protéolyse et un stress oxydatif plus marqués chez les souris âgées.

B.5. Conclusion

Bien que préliminaires, ces analyses révèlent non seulement que l'activation du système immunitaire induit des réponses moléculaires au niveau du foie des souris exposées au LPS, mais aussi que ces réponses diffèrent en fonction de l'âge des souris. En effet, les réponses moléculaires ont surtout été observées chez les souris jeunes, ce qui laisserait supposer, en accord avec le concept d'immunosénescence, que les souris plus âgées sont, soit moins efficaces dans l'activation de leur système immunitaire, soit moins réactives pour se protéger des coûts de cette activation. L'interprétation détaillée de ces données sera réalisée par notre collaborateur.

II. Etude des rythmes circadiens chez la souris

Ce projet a été réalisé en collaboration avec le Dr. Etienne Challet, INCI, Strasbourg.

A. Contexte

Les rythmes biologiques permettent à l'organisme d'anticiper les variations périodiques environnementales afin d'assurer le maintien des fonctions physiologiques et des homéostasies. Ils sont synchronisés par des facteurs environnementaux appelés « Zeitgeber » [165], dont les principaux sont la lumière [166] et la prise alimentaire [167].

Sur une période de 24 heures, les rythmes circadiens contrôlent la plupart des mécanismes biologiques, physiologiques et comportementaux comme par exemple le cycle veille-sommeil, la stabilité de la température corporelle ou encore le métabolisme énergétique. Ils sont produits par plusieurs horloges biologiques, dont une horloge maîtresse située dans l'hypothalamus et des horloges secondaires situées dans les autres structures centrales ou périphériques [168]. Les propriétés qui permettent de les caractériser sont décrites en figure VII-5:

- Le « **MESOR** » (Midline Estimating Statistic Of Rhythm) est le niveau moyen d'une variable rythmique sur la durée d'un cycle complet.
- L'**acrophase** et la **bathyphase** sont les temps écoulés entre une valeur de référence et le temps auquel les variables mesurées sont maximales et minimales respectivement.

- L'**amplitude** est la différence entre le MESOR et l'acrophase.
- La phase correspond au temps écoulé entre un temps de référence (en général minuit) et l'acrophase.
- La période correspond à l'intervalle de temps entre 2 acrophases (cette période est de 24h pour les rythmes circadiens).

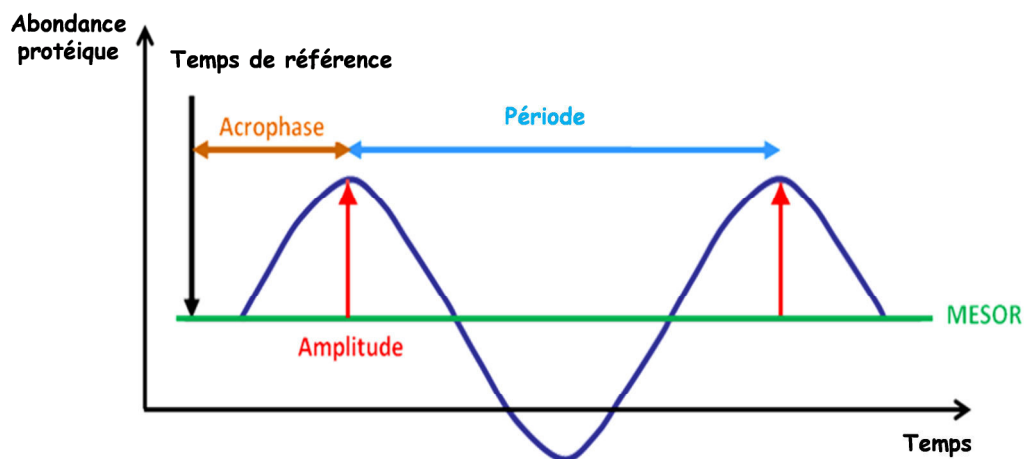


Figure VII-5 Représentation schématique des variables qui caractérisent un rythme circadien

Les niveaux d'expression des protéines qui sont régulées selon les rythmes circadiens peuvent être définis par : le niveau d'expression moyen de cette protéine au cours du temps (le MESOR), l'amplitude de son oscillation et le temps écoulé entre le niveau d'expression maximal de cette protéine et un temps de référence (l'acrophase).

Le fonctionnement des horloges circadiennes est réglé par des gènes spécifiques dits « gènes d'horloge » qui codent des « protéines d'horloge ». L'expression de ces protéines d'horloge est notamment contrôlée par plusieurs boucles de rétroactions majeures [169] présentées en figure VII-6. Les protéines d'horloge BMAL1 (Brain and muscle Arntl-like 1) et CLOCK (Circadian Locomotor Output Cycles Kaput) hétérodimérisées activent la transcription des gènes dont le promoteur possède une séquence E_{BOX} (CACGTG) comme par exemple les gènes d'horloge *Per1-3* (Period homolog 1-3) et *Cry1-2* (Cryptochrome 1-2), *Dec1-2* (Differentiated Embryo Chondrocyte protein 1-2), *Rev-erba/b* (Reverse vira erythroblastis oncogene products) et *Rora/b* (Retinoic acid-related Orphan Receptors). Les protéines DECs et les hétérodimères CRYs-PERs inhibent en retour l'activation de leur propre transcription due à l'action du complexe CLOCK/BMAL1. Les protéines Rev-ERBs et RORs vont plutôt directement inhiber ou activer la transcription de *Bmal1* en se liant à la séquence ROR de son promoteur [169].

Certaines modifications post-traductionnelles participent également à la régulation de ces boucles de rétrocontrôle. Par exemple, les protéines CK1 ϵ et CK1 δ (Casein Kinase 1 epsilon and Casein Kinase 1 delta, lorsqu'elles phosphorylent les protéines PERs, vont contribuer à la dégradation des protéines PERs et la formation des complexes PERs/CRYs [170].

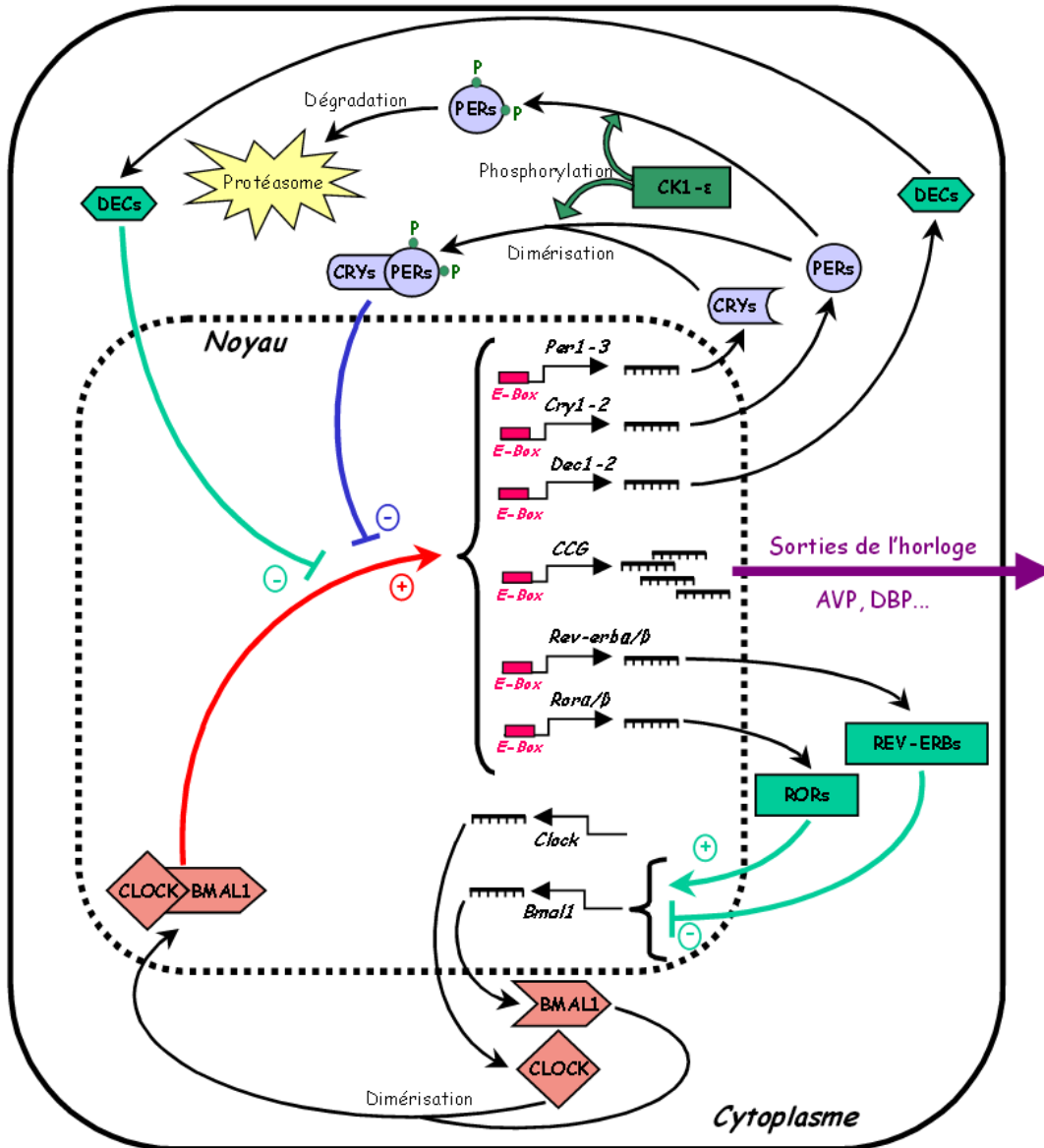


Figure VII-6 Mécanismes moléculaires permettant la genèse d'un rythme circadien au sein du noyau suprachiasmatique de l'hypothalamus (figure extraite de la thèse de Marc Cuesta 2009)

L'expression des protéines d'horloge est régulée par des boucles de rétroaction. Les protéines d'horloge BMAL1 (Brain And Muscle Arntlike 1) et CLOCK (Circadian Locomotor Output Cycles Kaput) hétérodimérisées activent la transcription des gènes dont le promoteur possède une séquence EBOX (CACGTG) comme les gènes d'horloge Per1-3 (Period homolog 1-3) et Cry1-2 (Cryptochrome 1-2), Dec1-2 (differentiated embryo chondrocyte protein1-2), Rev-erba/β (Reverse vira erythroblastis oncogene products) et Rora/β (Retinoic acid-related orphan receptors). Les protéines DEC1 et DEC2 inhibent en retour l'activation de leur propre transcription.

La dérégulation des cycles est connue pour produire des effets délétères chez l'homme [171]. Il semblerait aujourd'hui qu'il existe une horloge circadienne dans la plupart des tissus et nombre reste

à étudier. Une meilleure compréhension de la physiologie circadienne pourrait donner lieu à de nouvelles applications telles que la chrononutrition ou chronothérapie [172]. Dans le but de mieux comprendre cette physiologie circadienne, la rythmicité des protéines présentes dans le noyau suprachiasmatique (l'horloge principale), qui est responsable de la genèse du rythme circadien, a déjà fait l'objet de quelques études [168], tout comme le foie [173] [174]. Cependant, peu de choses sont connues sur les mécanismes d'anticipation des repas. Il a récemment été montré que le cervelet possède une horloge circadienne impliquée dans ces mécanismes d'anticipation [175].

Dans le but d'identifier les protéines régulées selon l'horloge circadienne, nous avons prélevé selon 4 points horaires le cervelet de souris nourries de nuit ad libitum (1 point-horaire toutes les 6 h à partir de midi; n = 5 /point-horaire). Parallèlement, pour étudier l'influence du décalage de l'heure des repas sur l'expression des protéines cérébelleuses et ainsi identifier des protéines possiblement impliquées dans le phénomène d'anticipation des repas, nous avons aussi prélevé selon 2 points-horaires différents (1 point-horaire toutes les 12 h à partir de midi ; n=5 /point-horaire) le cervelet de souris n'ayant eu accès à la nourriture que sur une courte période diurne (de midi à 18h). Ce protocole provoque en effet une activation comportementale diurne (augmentation d'activité locomotrice) des souris dans les heures qui précèdent midi (ce qui n'est pas observé lorsque l'heure des repas n'est pas décalée) et qui a été interprétée comme reflétant une anticipation des repas.

Le schéma expérimental de ce projet est illustré en figure VII-7.

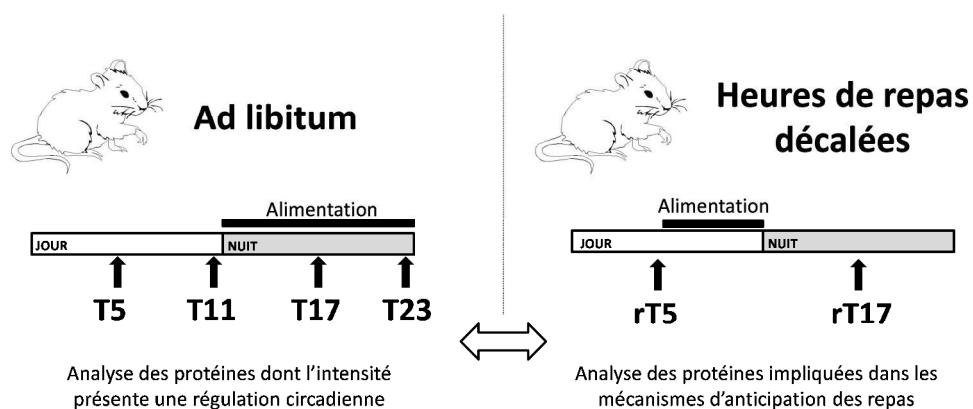


Figure VII-7 Schéma expérimental de l'expérience

Dans un premier temps, le protéome des souris a été comparé entre les différents points horaires chez les souris nourries de nuit ad libitum. La comparaison des 4 points horaires a permis d'identifier les protéines dont l'expression est régulée par l'horloge cérébelleuse. Dans un deuxième temps, le protéome de souris n'ayant eu accès à la nourriture que sur une courte période diurne, a été comparé aux protéomes de souris nourries de nuit ad libitum de façon à identifier des protéines impliquées dans les mécanismes d'anticipation des repas

B. Optimisation du traitement des échantillons

Pour cette expérience, la quantité d'échantillons disponible était suffisamment importante pour nous permettre d'optimiser le protocole d'extraction des protéines en testant plusieurs types de traitement des échantillons. Nous avons comparé l'utilisation du Triton X-100 qui est un détergent non-ionique couramment utilisé en électrophorèse bidimensionnelle [176] puis celle de l'octyl glucoside qui a été décrit dans la littérature comme adapté à la solubilisation, mais également à la dénaturation des protéines du cervelet [177] et compatible avec l'électrophorèse bidimensionnelle [176]. Etant donné le contenu lipidique des échantillons [178], différents types de protocole de traitement des échantillons ont également été comparés, en précipitant les protéines extraites soit avec de l'acide trichloracétique (10%) soit avec de l'acétone. Les extraits protéiques issus de ces différents types de conditions d'extraction ont ensuite été décomplexifiés par gel d'électrophorèse bidimensionnelle (2 gels par condition). Le schéma expérimental concernant les optimisations du protocole de préparation est présenté en figure VII-8. Un exemple d'image des gels 2D obtenue est présenté en figure VII-9.

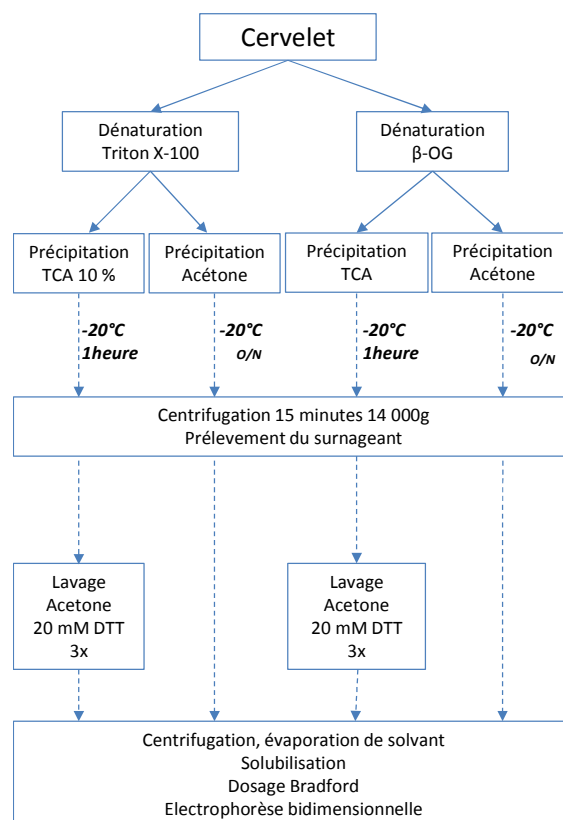


Figure VII-8 Schéma expérimental concernant l'optimisation du protocole de préparation des échantillons

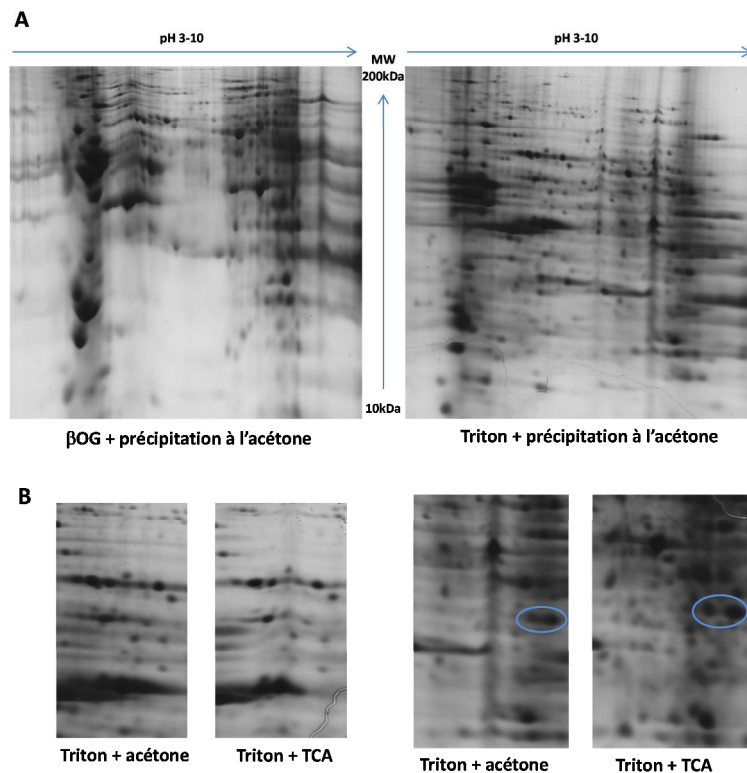


Figure VII-9 Images de gels d'E2D obtenus pendant la phase d'optimisation du traitement des échantillons

En A, est présentée une image des gels pour lesquels les protéines ont été solubilisées soit avec le détergent octyl glucoside (à gauche) soit avec le Triton X-100 (à droite), puis précipitées à l'acétone. En B, deux zones sont présentées (zone de pH acides à gauche et zones de pH basiques à droite) pour lesquels les échantillons ont été traités soit avec de l'acide trichloracétique (10%), soit avec de l'acétone, à la suite de l'utilisation du Triton X-100 comme détergent.

D'après les images de gels présentées en figure VII-9, on constate que la résolution du gel d'électrophorèse est meilleure après l'utilisation du Triton X-100 comme détergent. Concernant le nombre de spots, environ 500 spots ont été détectés après l'utilisation du détergent octyl glucoside alors qu'environ 600 spots ont été détectés après utilisation du Triton X-100. Ainsi, le protocole basé sur l'utilisation du Triton X-100 a été retenu pour la suite de l'expérience. Nous supposons cependant que l'utilisation du détergent octyl glucoside n'est peut-être pas seule responsable du manque de résolution observée sur l'image de gel A (délipidation insuffisante, précipitation des protéines dans la bande de gel à gradient de pH immobilisé, etc.) mais nous n'avons pas identifié précisément l'origine du problème.

Concernant les protocoles de précipitation des protéines, le protocole basé sur l'utilisation de l'acide trichloracétique pour précipiter les protéines a permis de gagner en résolution sur la zone des protéines possédant un point isoélectrique élevé. Dans la littérature, nous avons constaté que la plupart des protéines différentielles identifiées dans une étude 2D-DIGE réalisée à partir d'échantillons de noyau suprachiasmaticque présentaient un point isoélectrique élevé [168], c'est

pourquoi nous avons ensuite décidé de privilégier l'utilisation de l'acide trichloracétique pour la préparation des échantillons à comparer.

C. Analyse des échantillons

La préparation des échantillons est décrite dans la partie expérimentale en page 223. La méthodologie (préparation d'échantillon, réalisation des gels d'électrophorèse, traitement des images avec le logiciel Samespots, identification des protéines par spectrométrie de masse) est la même que pour les projets concernant l'étude des compromis évolutifs. Les analyses par spectrométrie de masse ont été réalisées avec le couplage nanoAcquity (Waters), maXis 4G, (Bruker Daltonique) Ce couplage, sensible et précis en masse, était correctement adapté à l'analyse des spots de gel faiblement complexes.

En revanche, contrairement aux deux applications concernant l'étude des compromis évolutifs, un nouveau système d'électrophorèse a été utilisé pour la réalisation des gels d'électrophorèse bidimensionnelle de ce projet : le système HPE™ Tour FlatTop (Serva). L'intérêt de ce nouveau système (décrit dans la partie « Stratégies d'analyse protéomiques quantitatives » en page 61) est qu'il permet de réaliser des gels de plus grande taille et avec une résolution élevée. De plus, les gels précoulés commercialisés, nécessaires à l'utilisation de ce type de système, sont fixés sur un film plastique et sont donc beaucoup plus faciles à manipuler. Le seul désavantage de ce système est que l'excision des spots redevient manuelle.

Pour ce projet, les identifications par spectrométrie de masse ont été réalisées après avoir rassemblé les spots protéiques issus de plusieurs gels DIGE avant digestion enzymatique. Etant donné le temps d'excision manuel des spots, seuls les spots dont l'abondance a été évaluée comme significativement différentielle ont été excisés et analysés.

D. Résultats

Un exemple d'image de gel 2D obtenue est présenté en figure VII-10. Il est possible de constater le gain en résolution de ce gel par rapport aux gels d'électrophorèse réalisés verticalement pendant la phase d'optimisation (figure IV-9).

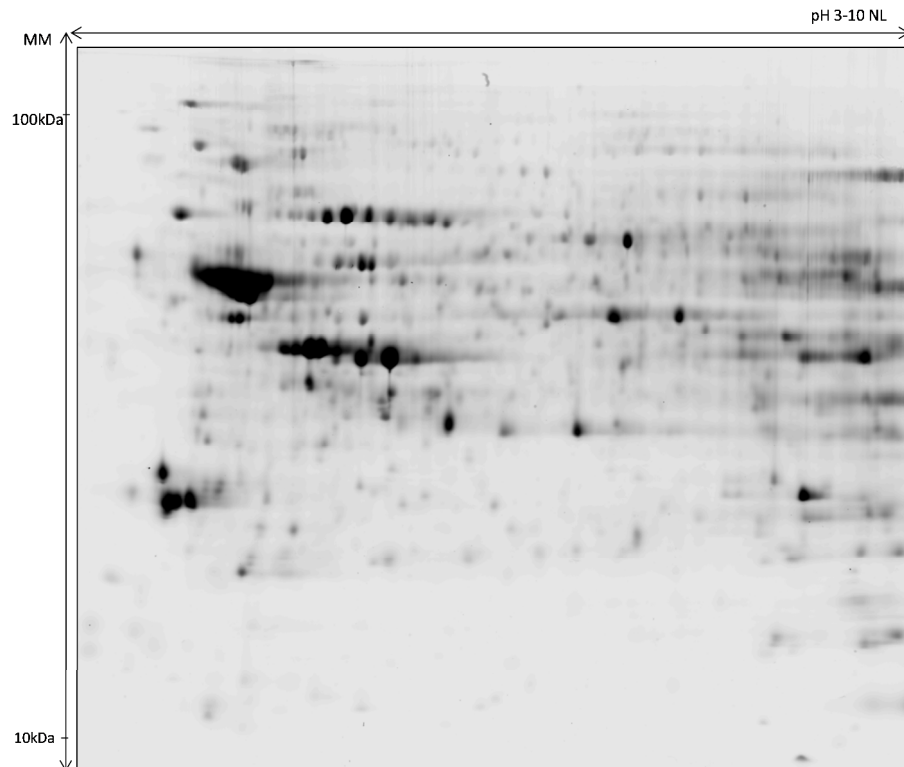


Figure VII-10 Image de gel 2D-DIGE (Cy2) obtenue à partir d'extraits protéiques de cervelet de souris
150 μ g de protéines ont été déposés sur le gel (50 μ g pour Cy3, Cy5 et Cy2) La bande de gel à gradient de pH immobilisé pour la première dimension couvrait la gamme de pH de 3 à 10 (non linéaire) et mesurait 24 cm au lieu de 18 cm pour les approches concernant l'étude des compromis évolutifs. La réticulation du gel en polyacrylamide était de 12,5 %. Les gels ont été scannés à l'aide d'un Ettan DIGE Imager. Cette image a été générée d'après les longueurs d'onde d'émission et d'absorption du fluorochrome Cy2. La résolution des gels réalisés grâce au système d'électrophorèse horizontal est supérieure à celle des gels réalisés grâce au système d'électrophorèse vertical.

D.1. Identification des protéines cérébelleuses régulées par les rythmes circadiens

Dans le but de mettre en évidence les protéines cérébelleuses dont l'expression oscille selon un rythme circadien, les abondances des spots protéiques obtenues chez les 4 groupes de souris nourries pendant la nuit ont été soumises au test du COSINOR qui est classiquement utilisé dans le domaine de la chronobiologie [179]. Nous n'avons conclu à l'oscillation rythmique d'une protéine (à partir de l'intensité des spots protéiques correspondants) que lorsque la significativité ($p < 0,05$) était vérifiée simultanément pour les 2 paramètres que sont l'Amplitude et l'Acrophase. Au final, 25 spots protéiques étaient dans ce cas, la significativité ($p < 0,05$) étant également vérifiée pour le MESOR dans le cas de 15 d'entre eux. La localisation de ces spots sur les gels ainsi que les protéines identifiées au sein de ceux-ci sont présentés en annexe 5. Les protéines identifiées comme différentielles sont impliquées dans les mécanismes de transmission synaptique (Beta synuclein, Synapsin-2, Vesicle fusing ATPase), de biochimie cellulaire (D-dopachrome decarboxylase) ou encore de régulation des métabolismes énergétiques (NAD-dépendant protein deacetylase sirtuin-2).

Deux exemples de profils rythmiques sont présentés en figure VII-11.

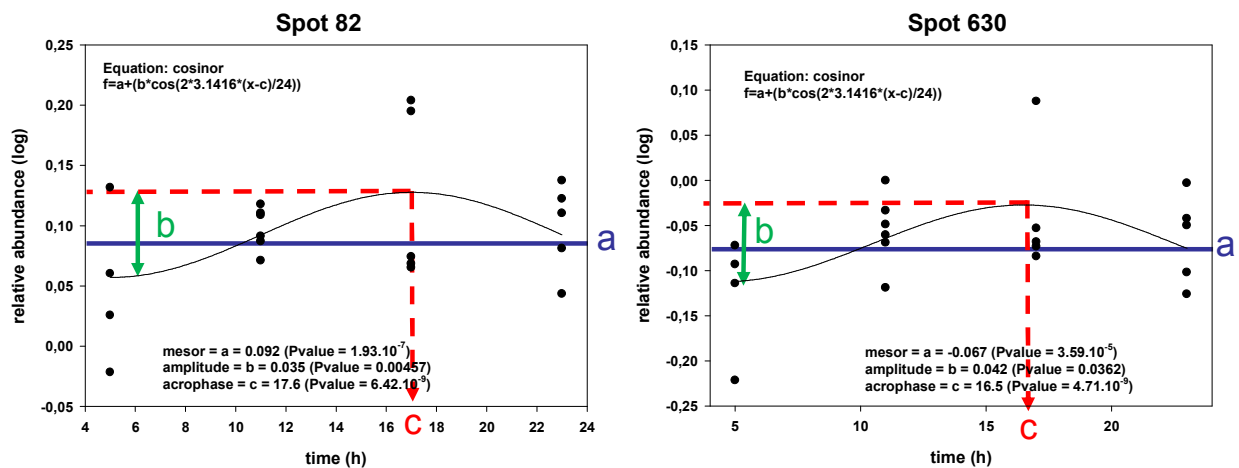


Figure VII-11 Exemples de profils d'expression rythmique de protéines contenues dans 2 spots protéiques différents. Dans les spots 82 et 630, nous avons identifié des protéines qui sont généralement impliquées dans le fonctionnement synaptique. Les abondances relatives observées sont très faibles mais sont en accord avec les valeurs reportées dans la littérature.

La reproductibilité élevée de la stratégie 2D-DIGE nous a permis d'identifier comme différentielles des protéines dont l'abondance relative variait selon une amplitude très faible sur la durée d'un cycle de 24h (1,1-1,2). Cependant, ces faibles valeurs d'amplitude sont en parfait accord avec les variations d'amplitude reportées dans la littérature [168]. Ces résultats seront validés par des approches orthogonales comme des analyses transcriptomiques ou Western-Blot.

D.2. Analyse des protéines impliquées dans les phénomènes d'anticipations des repas

Dans le but de mettre en évidence les protéines cérébelleuses dont l'expression est altérée par le décalage de l'heure des repas et ainsi identifier des candidats susceptibles d'être impliqués dans les mécanismes d'anticipation, les intensités relatives des spots protéiques des groupes de souris témoins (nourries en période nocturne) et des souris « décalées » (nourries en période diurne) ont été comparées par un test ANOVA à deux facteurs (1^{er} facteur : le temps ; 2^{ème} facteur : décalage des repas) suivi d'un test post-hoc de Tukey. L'égalité des variances ainsi que la normalité des valeurs ont été testées au préalable selon les tests de Bartlett et Shapiro-Wilk respectivement ($p > 0,01$). 19 spots protéiques ont ainsi été identifiés comme différentiels ($p < 0,05$ pour ANOVA et Post-hoc de Tukey). Les localisations de ces spots sur les gels ainsi que les protéines identifiées au sein de ces spots sont présentés en annexe 5. Les protéines identifiées comme différentielles appartiennent à la famille des protéines chaperonnes (heat shock protein HSP 90), sont impliquées dans le métabolisme de la glycolyse (glucose-related protein) ou des voies lipidiques (fatty acid binding protein).

Les abondances relatives de deux spots, au sein desquels ont été identifiées des protéines impliquées dans les voies lipidiques ainsi que des protéines chaperonnes impliquées dans les processus de repliement des protéines, sont illustrées en figure VII-12.

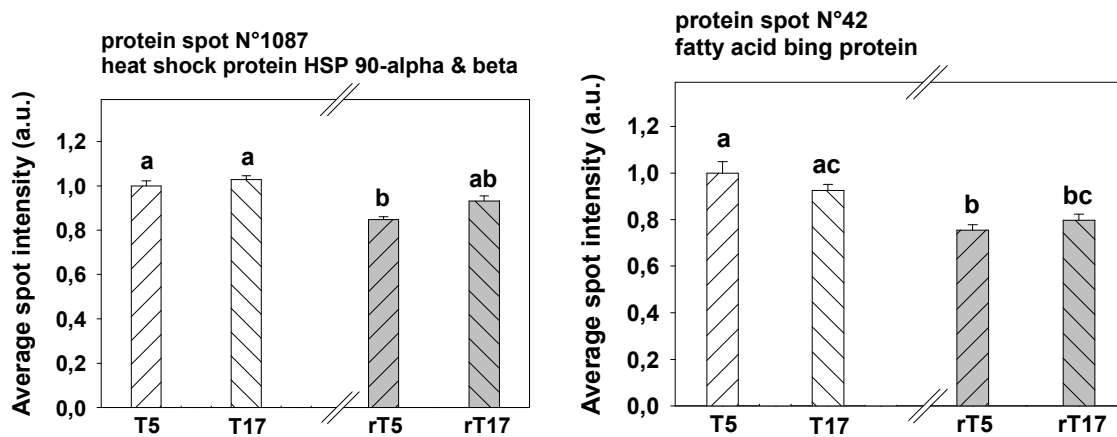


Figure VII-12 Profils d'abondance de deux spots au sein desquels des protéines impliquées dans les voies lipidiques ou chaperonnes ont été identifiées.

Les abondances relatives des spots 1087 et 42 sont significativement différentes entre les temps T5 et rT5.

E. Validation

Bien que le nombre de réplicas par condition soit relativement élevé dans ce projet ($n = 6$), la puissance statistique de notre étude n'est pas suffisante pour conclure de façon définitive quant au profil circadien des protéines identifiées dans cette étude étant donné les très faibles variations rythmiques reportées. Dans le but de compléter ces résultats, des analyses transcriptomiques sont en cours et des mesures en Western-Blot seront réalisées prochainement.

Notre collaborateur va également étudier la séquence des promoteurs des gènes codant les protéines pour lesquelles nos données suggèrent une expression oscillante selon un rythme circadien. Cette analyse permettra de préciser les liens éventuels entre l'expression de ces protéines et sa régulation par les protéines d'horloge.

Notre étude a permis d'identifier des voies potentielles qui pourraient être régulées selon l'horloge circadienne dans le cervelet de souris et/ou qui pourraient être responsables des phénomènes d'anticipation des repas. Les abondances relatives des protéines identifiées sont très faibles mais sont en parfait accord avec la littérature. L'interprétation des résultats, réalisée par notre collaborateur, ainsi que la réalisation de mesures complémentaires est actuellement en cours.

III. Etude des mécanismes d'adaptation de l'ours brun à l'inactivité

Ce projet a été réalisé en collaboration avec le Dr. Stéphane Blanc, DEPE, IPHC, Strasbourg.

A. Contexte

L'atrophie musculaire est préjudiciable pour la mobilité, la production de chaleur et est responsable de désordres métaboliques. Il s'agit d'une des conséquences majeures des processus de vieillissement et d'inactivité physique chez l'Homme [180]. L'atrophie musculaire est également observée dans le cadre d'immobilisation, de maladie et lors de vols spatiaux. Bien que la compréhension des mécanismes impliqués dans les phénomènes d'atrophie musculaire s'améliore [181], il n'y a pas encore de traitement thérapeutique permettant de prévenir ou de lutter contre ses effets. Contrairement à l'Homme, les ours hibernants ont démontré une faculté particulière quant à la préservation de leur masse musculaire en période d'inactivité [182-185]. Plusieurs hypothèses sont explorées pour expliquer le maintien de la masse musculaire de l'ours inactif, qu'il s'agisse d'une meilleure gestion du stress oxydatif induit par l'inactivité, d'un recyclage de l'azote au niveau hépato-intestinal, ou encore de l'existence spécifique d'un facteur antiprotéolytique chez l'ours.

Dans ce contexte, comparer le protéome musculaire de l'ours en période inactive à celui en période active pourrait permettre de compléter ces hypothèses. C'est ce que nous avons réalisé ici chez l'ours brun de Scandinavie. La disponibilité des échantillons est facilitée par une collaboration avec les équipes du « Scandinavian Brown Bear Research Project » (Suède) qui conduit des études sur l'écologie et la conservation de l'ours brun depuis 20 ans.

B. Schéma expérimental

Le schéma expérimental de ce projet consistait à comparer des échantillons de muscle d'ours (n = 6) en période d'activité (collectés au début de l'été) et en période d'inactivité (collectés à la fin de la période d'hibernation).

C. Analyse des échantillons

La préparation d'échantillons ainsi que le traitement des données a été réalisé selon le protocole décrit pour le projet concernant l'étude des protéines cérébelleuses régulées selon les rythmes

circadiens. Les identifications des protéines ont été réalisées d'après les portions de séquences protéiques conservées chez d'autres espèces proches phylogénétiquement.

D. Résultats

Une image d'un des gels 2D-DIGE réalisée est présentée en figure VII-13. L'analyse d'image des gels 2D-DIGE a révélé 27 spots protéiques différentiels selon le test de Student apparié avec $p < 0,05$ et 13 avec $p < 0,01$. La normalité des données ainsi que l'égalité des variances ont été vérifiées grâce aux tests de Shapiro-Wilk et Bartlett respectivement. Ces hypothèses ont été validées si p était supérieure à 0,01. Une image de gel présentant la localisation des spots différentiels ainsi qu'une liste des protéines identifiées dans chacun de ces spots sont présentées en annexe 6.

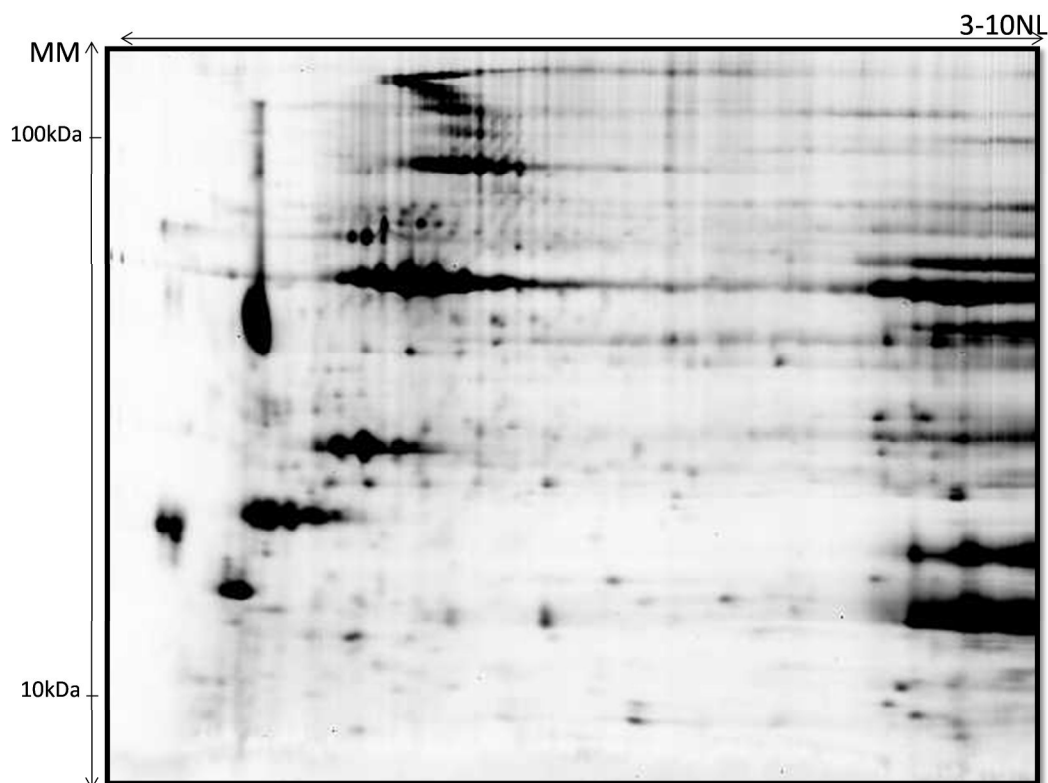


Figure VII-13 Image de gel 2D-DIGE (Cy5) réalisé à partir des extraits protéiques de tissus musculaires d'ours brun
150 µg de protéines ont été déposés sur le gel (50µg pour Cy3, Cy5 et Cy2) La bande de gel à gradient de pH immobilisé pour la première dimension couvrait la gamme de pH de 3 à 10 (non linéaire) et mesurait 24 cm. L'étape de migration des gels selon leur masse moléculaire a été réalisée avec le système d'électrophorèse horizontale. La réticulation du gel en polyacrylamide était de 12,5%. Les gels ont été scannés à l'aide d'un Ettan DIGE Imager. Cette image a été générée d'après les longueurs d'onde d'émission et d'absorption du fluorochrome Cy5.

On peut constater sur cette image de gel qu'un certain nombre de chapelets sont extrêmement abondants (les protéines actine et myosine ont notamment été identifiées dans ces chapelets) au détriment de la détection de nombreux spots protéiques sur ces gels. Dépléter ces protéines aurait permis de détecter plus de spots protéiques, en revanche, il n'existe pas d'anticorps commerciaux à

notre connaissance permettant de les éliminer et la quantité d'échantillon disponible n'était pas suffisante pour pouvoir tester différents protocoles.

E. Conclusion

Les données sont en accord avec une faible réduction de la masse musculaire des ours inactifs. Les régulations moléculaires mises en évidence suggèrent un renforcement des fibres musculaires en période inactive. Par ailleurs, outre les altérations du métabolisme des sucres et des lipides, nombreuses sont les régulations qui suggèrent que les processus de développement des myocytes, de synthèse et de repliement des protéines, mais également de transport des ions Ca^{2+} (essentiels aux contractions) sont non seulement maintenus mais possiblement stimulés en période inactive.

En guise de validation, nous avons comparé les résultats obtenus avec la technique 2D-DIGE à l'analyse par comptage de spectres qui a été initiée après cette expérience sur ces mêmes échantillons par le Dr. Anna Ziemianin.

La stratégie par comptage de spectre a permis d'identifier 27 protéines dont les abondances varient en fonction du statut d'activité de l'ours. Les abondances relatives de deux protéines ont été établies comme statistiquement différentielles avec les deux approches quantitatives utilisées dans cette étude (comptage de spectre et 2D-DIGE). Les niveaux d'expression de ces protéines entre les deux méthodes quantitatives sont présentées en figure VII-14.

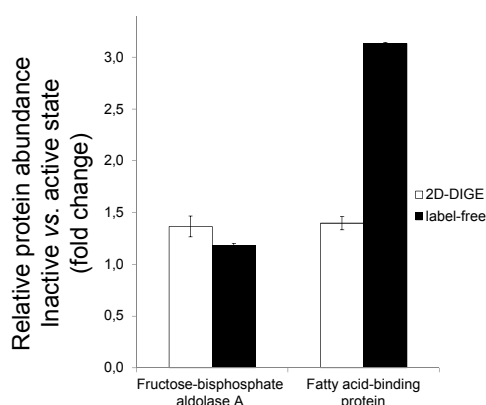


Figure VII-14 Niveaux d'expressions respectifs des deux protéines identifiées comme différentielles communes aux deux approches 2D-DIGE et Label Free

Les deux approches quantitatives conduisent aux mêmes conclusions pour ces deux protéines différentielles.

On constate d'après la figure VII-14 que les sens et amplitudes de variations entre les deux états sont similaires pour les protéines « fructose-biphosphate aldolase A » et « fatty acid-binding protein ». Bien que la précision des amplitudes de variation mesurées ne soit pas très élevée, on peut constater

que les deux stratégies conduisent aux mêmes conclusions, ce qui renforce la confiance dans les résultats que nous avons obtenus. Les deux approches sont donc concordantes et complémentaires puisqu'elles permettent d'identifier 61 protéines musculaires différentielles entre les états actifs et inactifs chez l'Ours brun. Ces données seront vérifiées et complétées par des approches orthogonales.

La combinaison des approches 2D-DIGE-MS et label-free a donc permis de révéler des mécanismes susceptibles d'être liés à la préservation (masse, force et fonctionnalité) paradoxale des muscles de l'ours pourtant resté inactif pendant près de 4 mois lors de sa dormance annuelle.

IV. Discussion méthodologique pour les quatre applications présentées dans ce chapitre

A. Choix de la stratégie 2D-DIGE

Les études présentées dans ce chapitre nécessitaient la comparaison de plus de dix échantillons différents. Lors de la réalisation de ces quatre projets, deux types de couplages étaient alors disponibles au laboratoire pour réaliser des analyses protéomiques globales:

- Un spectromètre de masse de type piège à ions (amaZon/HCTPlus, Bruker Daltonique) couplé à un système nanoLC-ChipCube (Agilent technologies). Ces systèmes « chip » sont des systèmes microfluidiques permettant de réduire les volumes morts, conséquents en analyse nanoLC, mais leur capacité de charge est limitée. Ces couplages étaient relativement rapides (4 Hz) mais peu résolutifs (≈ 2000 pour un m/z égal à 922). Ce système sont donc très pratique pour l'analyse d'échantillons peu complexes comme les spots de gel avec des temps de gradient très courts.
- Deux spectromètres de masse de type Q-TOF (Synapt G1, Waters, maXis 4G, Bruker Daltonique) couplés à des systèmes de chromatographie liquide de type UPLC (décrits en page 16). Ces spectromètres sont résolutifs ($> 10\,000$ pour un m/z égal à 922) et assez rapides (1-2 Hz). Le Maxis 4G possède une capacité résolutive très élevée (environ 50 000 pour le m/z 922). Cette résolution aurait pu permettre la réalisation de stratégies quantitatives type « extraction des courants d'ions (XIC) ». Cependant, au moment de la réalisation de ces projets, les logiciels adaptés au traitement des données sans marquage et permettant l'extraction de courants d'ions à partir des données nanoLC-MS/MS n'étaient soit

pas disponibles (Skyline [122]), soit pas compatibles avec des données générées sur des instruments du constructeur Bruker Daltonique (MaxQuant LFQ [123]).

La stratégie 2D-DIGE ne nécessite pas l'utilisation de spectromètres de masse très rapides ou très résolutifs puisque, d'une part, la quantification est réalisée en amont d'après l'analyse des images de gel grâce à des logiciels de traitement spécifiques, et d'autre part, les extraits protéiques issus de spots de gel 2D sont de faible complexité et donc parfaitement adaptés à la vitesse et à la résolution des couplages dont nous disposons de type nanoLC-ChipCube couplés aux analyseurs de type pièges à ions. La stratégie 2D-DIGE permet, en fonction de l'échantillon, de travailler sur plusieurs centaines de spots simultanément. De plus, le temps d'analyse n'est pas significativement augmenté si on augmente le nombre d'échantillons à analyser puisque plusieurs gels peuvent être réalisés en parallèle. Finalement, la large gamme dynamique de réponse reportée pour la technique 2D-DIGE (4 ordres de grandeur [186]) ainsi que sa reproductibilité élevée [187, 188] en ont fait une technique parfaitement adaptée aux quatre applications présentées dans ce chapitre.

L'approche par comptage de spectres permet éventuellement de valider les résultats obtenus grâce à la stratégie 2D-DIGE comme cela a été réalisé pour l'étude du protéome différentiel de l'ours entre une période active et une période inactive.

L'étude du protéome différentiel de l'ours présentait une difficulté supplémentaire par rapport aux autres applications : le protéome de l'ours n'est pas connu. Or, la résolution élevée de la technologie DIGE et le niveau de décomplexification des échantillons qu'elle produit (jusque quelques protéines par spot seulement), permet d'analyser un nombre élevé de peptides par protéines lors de l'analyse nanoLC-MS/MS, ce qui permet d'augmenter la probabilité d'identifier des peptides par homologie de séquence avec d'autres espèces [75]. A titre d'exemple, on peut comparer le nombre de peptides quantifiés par protéine en moyenne pour la stratégie XIC appliquée dans le chapitre V « Applications et développement de la stratégie quantitative globale XIC à des échantillons biologiques complexes » (en moyenne 6 peptides discriminants quantifiés par protéine, 1 seul peptide quantifié pour 50 % des protéines) au projet 2D-DIGE concernant l'étude du coût de la reproduction pour lequel le contenu d'environ 400 spots de gel révélés au bleu de Coumassie a été analysé (8 peptides discriminants en moyenne par protéine, environ 15 % des protéines identifiées avec un seul peptide). Le nombre de peptides identifiés est supérieur avec la stratégie par gel d'électrophorèse.

B. Validation

Malgré la résolution obtenue avec les gels d'électrophorèse bidimensionnelle, il est possible qu'un spot dont l'abondance relative est différentielle entre différentes conditions, contienne plusieurs protéines. Notre hypothèse était que la protéine identifiée avec le plus grand nombre de peptides était la protéine responsable de la variation d'intensité d'un spot. Cette hypothèse nécessite la validation des cibles potentielles que nous avons évaluées dans nos études par des approches orthogonales.

Les résultats des projets présentés dans ce chapitre ont fait l'objet d'analyses statistiques. Les abondances relatives des protéines identifiées comme différentielles étaient comprises entre 1,1 et 1,5, or le nombre de répliques biologiques utilisés dans la plupart de ces projets était faible ($4 < n < 6$). De plus, un biais a été reporté dans l'utilisation des trois cyanines et notamment dans l'utilisation de la cyanine 2 (qui présente des niveaux de réponse plus faible que les autres fluorochromes) pour le standard interne qui tendrait à induire un nombre plus élevé de résultats faux positifs [189].

Nous avons conscience de la faible puissance statistique de nos résultats [187, 190] et nous sommes donc conscients des limites de nos études. Cependant, ces limites ne remettent pas en cause le caractère exploratoire de ces recherches et l'objectif annoncé de ne pas « rater » une possible voie physiologique d'intérêt uniquement sur la base d'une puissance statistique trop faible. Les hypothèses soulevées par nos travaux feront l'objet de nouvelles études et/ou de validation avec des méthodes basées sur d'autres stratégies d'analyse protéomique (la SRM par exemple) ou d'immunodétection par exemple.

V. Conclusion

L'utilisation de la technique 2D-DIGE a permis pour ces quatre projets de réaliser des analyses quantitatives ciblant plusieurs centaines de protéines et en analysant 10 à 40 échantillons simultanément à une période où le parc instrumental de notre laboratoire, et la disponibilité des logiciels de traitement des données, ne permettaient pas la réalisation d'analyses quantitatives XIC qui permettent de quantifier plusieurs centaines de protéines, voire des milliers, par analyse nanoLC-MS/MS. Ces projets exploratoires visaient à identifier de nouvelles voies physiologiques d'intérêt. Aujourd'hui, étant donné l'évolution des analyseurs de type Orbitrap ou Q-TOF, ces analyses pourraient être reconduites avec des spectromètres de masse plus rapides, ce qui permettrait, d'une part, de valider les voies identifiées comme différentielles dans ces études et d'autres part, de

détecter des protéines différentielles d'intérêt moins abondantes que celles quantifiées par 2D-DIGE. Par contre, la décomplexification des extraits protéiques avec les gels d'électrophorèse bidimensionnelle est un véritable atout pour l'étude de protéome différentiel de protéomes d'espèces exotiques dont les séquences protéiques ne sont pas connues. La faible spécificité de la technique 2D-DIGE a été reportée dans la littérature au travers de la génération de listes de protéines identifiées comme différentielles de manière récurrente dans les approches 2D-DIGE et qui sont ainsi sujettes à l'appellation « Déjà-vu » [191, 192]. Au sein de nos thématiques, nous ne sommes pas dans un contexte de recherche de biomarqueurs, et nous ne sommes donc pas impactés par un éventuel manque de spécificité. Toute réponse moléculaire pertinente est à nos yeux possiblement d'intérêt pour pouvoir mieux expliquer une situation physiologique donnée.

Travaux supplémentaires

Travaux supplémentaires

D'autres projets ont été initiés pendant cette thèse mais n'ont pas été développés dans ce manuscrit. Voici un bref résumé pour deux d'entre eux:

I. Identification et caractérisation de nouveaux peptides antimicrobiens chez des oiseaux antarctiques et subantarctiques

Ce projet a été réalisé en collaboration avec le Dr Y. Le Maho, DEPE, IPHC, Strasbourg.

Aujourd'hui, le phénomène de résistance aux antibiotiques a explosé, ce qui rend le développement de nouveaux antibiotiques particulièrement difficile. Les peptides antimicrobiens produits par la Nature pourraient donc représenter de nouvelles solutions à explorer.

Alors qu'ils incubent leurs œufs pendant plusieurs semaines, les mâles de manchot royal (*Aptenodytes patagonicus*) peuvent préserver la nourriture qu'ils ont dans l'estomac sans la digérer et sans qu'elle se dégrade [193]. Ceci permet la survie des poussins à l'éclosion. La présence d'activités antimicrobiennes a été démontrée dans les contenus stomacaux préservés [194]. Deux peptides antimicrobiens ont ainsi été isolés et caractérisés, la sphéniscine 1 et la sphéniscine 2, et la sphéniscine 2 s'est révélée très active en milieu salin. L'identification de ce peptide revêt donc une importance particulière, voire une première étape vers un nouveau traitement des infections survenant dans les fluides biologiques riches en sels. Ces infections, qui affectent des millions de personnes chaque année, regroupent essentiellement les infections chroniques respiratoires, les infections oculaires et les infections urinaires.

L'objectif de ce projet était donc de rechercher de nouvelles molécules antimicrobiennes dans divers tissus du Manchot Royal (bourse de Fabritius, os, langue et cuticule d'œuf) dans le but de compléter l'inventaire des peptides antimicrobiens naturels actifs en milieu salé.

Du point de vue analytique, ce projet consistait à décomplexifier par gel SDS-Page des extraits protéiques issus des échantillons biologiques, suivi d'analyses nanoLC-MS/MS à haut débit. Comme dans le projet concernant les protéines de tortue Luth (en page 152), le protéome du manchot royal n'est pas connu. Les identifications ont donc été réalisées grâce aux portions de séquences protéiques conservées chez des organismes séquencés et proches phylogénétiquement [9] dans des

banques de données spécifiques de peptides déjà connus pour leurs propriétés antifongiques ou antimicrobiennes (Antimicrobial Peptide Database, APD). Plusieurs nouveaux peptides antimicrobiens de manchot royal, homologues de peptides antimicrobiens connus chez d'autres espèces, ont été identifiés. Les résultats préliminaires suggèrent la présence ubiquiste de la sphéniscine 2 et de molécules apparentées à la dermcidine (*Homo Sapiens*), la gallinacine (*Gallus Gallus*), la moricine (*Bombyx mori*), l'ostricacine (*Struthio camelus*), la cathelicidine (*Gallus Gallus*), la brévinine (*Rana Guentherli*), la palustrine (*Rana palustris*), la caerine (*Litoria splendida*). Ces analyses seront complétées très prochainement par celle de tissus de manchot Adélie. La poursuite du projet consistera à analyser les nouvelles molécules identifiées par chromatographie liquide (méthode à mettre au point), puis à tester leur activité antimicrobienne avec différentes souches de pathogènes, notamment en présence de sels.

II. Etude des partenaires de la protéine GIP1 chez *Arabidopsis thaliana*

Ce projet a été réalisé en collaboration avec le Dr M.E. Chabouté et Morgane Batzenschlager, IBMP, Strasbourg.

L'équipe d'AC Schmit à l'institut biomoléculaire des plantes (Marie-Edith Chabouté, Etienne Herzog et Morgane Batzenschlager) étudie les complexes de nucléation microtubulaires dans les cellules d'*Arabidopsis thaliana* et plus précisément les protéines impliquées dans l'ancrage et la régulation de ces complexes au sein des cellules.

La protéine GIP (GCP3 Interacting Protein) a été identifiée comme étant impliquée de manière cruciale au sein des complexes de nucléation. Afin de mieux connaître tous les acteurs impliqués dans ces complexes, ce projet consistait à identifier des possibles partenaires d'interaction de GIP *in vivo*. Une immunoprécipitation ciblant GIP a été réalisée, à partir d'échantillons de plantules, dans des conditions non dénaturantes de manière à préserver ces complexes. Nous avons ensuite déposé ces extraits sur gel SDS-Page, puis analysé par spectrométrie de masse le contenu des bandes de gel dans le but de confirmer l'interaction de GIP avec des partenaires déjà connus de nos collaborateurs et d'identifier de nouveaux candidats d'intérêt. La liste de protéines partenaires est actuellement exploitée par nos collaborateurs.

Bilan de la thèse et perspectives

Bilan de la thèse

Les travaux présentés dans cette thèse montrent que l'analyse protéomique, pourvu que les méthodes et stratégies soient optimisées à tous les niveaux, sont parfaitement adaptées pour produire des résultats originaux dans différents domaines de la biologie. Plus particulièrement, dans les domaines de la physiologie et l'écologie évolutive, nous avons ainsi caractérisé divers protéomes en termes d'abondance relative de protéines en réponse à des défis énergétiques (réponses à l'inactivité physique ou à des compromis énergétiques). Dans le domaine de l'écophysiologie, nous avons mis au point une méthode de suivi du statut reproducteur de la tortue Luth, espèce menacée d'extinction dont le protéome n'était pas séquencé. Enfin, dans le domaine de la toxicologie, nous avons pu préciser la réponse d'une lignée cellulaire hépatique à une exposition chronique à l'acide valproïque.

La diversité des problématiques biologiques sur lesquelles j'ai pu travailler m'a permis d'appréhender trois stratégies quantitatives majeures dans le domaine de la protéomique : la stratégie SRM, qui représente la technique de choix en analyse quantitative ciblée, la stratégie 2D-DIGE, une des premières stratégies quantitatives développées en analyse protéomique mais qui permet la quantification simultanée de plusieurs centaines de protéines et qui, associée au marquage des protéines avec des espèces fluorochromes, permet de travailler sur un grand nombre d'échantillons biologiques simultanément et finalement, la stratégie sans marquage basée sur l'extraction des courants d'ions en MS (XIC) qui connaît un nouvel essor actuellement grâce au développement des Orbitrap et Q-TOF de dernière génération.

Dans le domaine de la protéomique quantitative, il n'existe pas de méthode standard, applicable à tout type de problématique biologique. Chaque nouvelle problématique et le traitement de chaque nouveau type d'échantillon a donc requis la mise en place d'un protocole spécifique et la réalisation de nombreuses optimisations et de nombreux développements méthodologiques. En ayant été confrontée à des difficultés multiples, j'ai pu apporter des solutions à tous les stades de l'analyse protéomique : celui de la préparation d'échantillons, de l'acquisition des données ou encore de leur traitement. Par exemple, une faible quantité de matériel biologique disponible a nécessité l'optimisation des méthodes d'acquisition des spectromètres de masse pour pouvoir obtenir la meilleure sensibilité possible, la non-disponibilité des séquences protéiques pour certaines espèces m'a amenée à travailler à partir de séquences conservées chez d'autres espèces proches phylogénétiquement, la gestion de nombreux échantillons a nécessité la mise en place de stratégies

quantitatives adaptées (SRM, XIC, 2D-DIGE) et le traitement des données massives générées en spectrométrie de masse a nécessité le développement de protocoles rigoureux. La figure VIII-1 résume les étapes de développement critiques des analyses protéomiques quantitatives, les optimisations ou développements réalisés et l'apport de chacun d'eux.

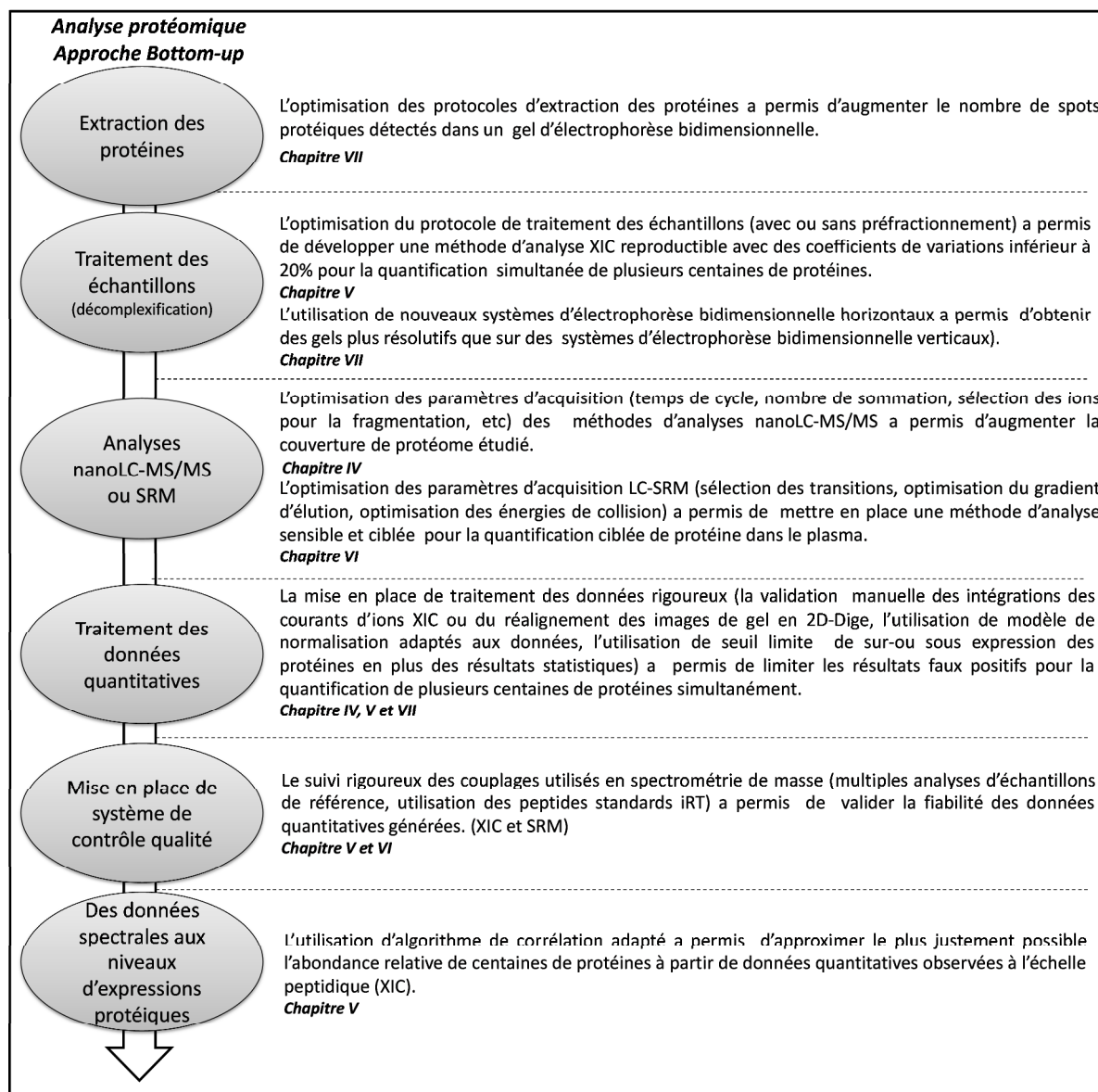


Figure VIII-1 Résumé des différentes étapes de développement d'une méthode de quantification en protéomique.

Perspectives

Comme cela a été montré par ce travail de thèse, la protéomique permet aujourd'hui d'apporter de nombreuses informations biologiques. Il est possible d'identifier dans un échantillon biologique complexe plusieurs centaines, voire milliers de protéines simultanément en une seule analyse et en moins d'une heure. Générer des informations quantitatives concernant les abondances relatives de ces milliers de protéines entre différentes conditions (patho) physiologiques est également devenu possible. Dans le cadre d'études ciblées, la protéomique permet également d'aller jusqu'à la détermination complète des séquences d'acides aminés et permet l'étude de quelques modifications post-traductionnelles (nature, localisation).

Cependant, la protéomique souffre encore de certaines limites :

- L'identification, la caractérisation et la quantification des protéines présentes à de très faibles concentrations est encore limitée. Les protéines les plus abondantes des échantillons masquent en effet celles qui le sont moins. Les méthodes de fractionnement protéiques et peptidiques doivent donc continuer à être développées (par ex. préparations d'organites). Même les analyses ciblées utilisant la stratégie SRM, bien que très sensibles, doivent encore être adaptées pour réaliser des quantifications ciblant des protéines présentes à des quantités subfemtomolaires [195].
- Les données générées par les spectromètres de masse sont de plus en plus massives, ce qui requiert toujours plus d'évolution dans le domaine bio-informatique et toujours plus de puissance de calcul [196]. Les développements réalisés au LSMBO, avec notamment la suite logicielle MSDA dont certains modules utilisent la puissance d'une grille de calcul [79] vont dans ce sens. Cependant, il subsiste des difficultés pour manipuler des fichiers lourds de plusieurs giga-octets ou composer avec des volumes de données tels qu'un tableur comme Excel devient trop petit !
- Le traitement des données issues des diverses approches de protéomique quantitative n'est pas standardisé au sein de la communauté. Plusieurs types de normalisation des données existent, la « traduction » des abondances mesurées au niveau peptidique en abondance protéique et la gestion des peptides « outliers » (dont les niveaux d'expressions diffèrent de ceux d'autres peptides quantifiés pour une même protéine) varient d'une étude à l'autre

(utilisation d'un seul peptide, des plus intenses, de tous ceux détectés, des plus corrélés entre eux, etc.).

- La caractérisation des modifications post-traductionnelles reste encore limitée bien que leur importance biologique soit indéniable. Ceci est principalement dû au fait que leur analyse requiert en général des préparations d'échantillons spécifiques et que leur recherche au sein des données spectrales dépasse souvent les puissances de calculs disponibles.
- L'annotation des génomes séquencés reste encore limitée pour les espèces exotiques, ce qui freine bien sûr les analyses chez les espèces concernées, même si des laboratoires comme le LSMBO n'ont jamais cessé de développer les approches de séquençage *de novo* pour palier à ces difficultés.

Finalement, on peut constater que la protéomique est en constant progrès et que des optimisations et des développements méthodologiques comme ceux décrits dans cette thèse contribuent à son incursion réussie dans la plupart des domaines de la Biologie. Pour aller encore plus loin et faire s'amenuiser les limites listées précédemment, la pluridisciplinarité devra être de mise afin de conjuguer les efforts et compétences des chimistes, analyticiens, biologistes, (bio-) informaticiens, et statisticiens.

Bibliographie

Bibliographie

1. Choi, J. and P. Tontonoz, *Metabolism: Drug discovery goes for a swim*. Nat Chem Biol, 2013. **9**(2): p. 68-9.
2. Altelaar, A.F., J. Munoz, and A.J. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics*. Nat Rev Genet, 2013. **14**(1): p. 35-48.
3. Cox, J. and M. Mann, *Is proteomics the new genomics?* Cell, 2007. **130**(3): p. 395-8.
4. Nesvizhskii, A.I., *Proteogenomics: concepts, applications and computational strategies*. Nat Methods, 2014. **11**(11): p. 1114-25.
5. Nilsson, T., et al., *Mass spectrometry in high-throughput proteomics: ready for the big time*. Nat Methods, 2010. **7**(9): p. 681-5.
6. Richards, A.L., A.E. Merrill, and J.J. Coon, *Proteome sequencing goes deep*. Curr Opin Chem Biol, 2014. **24C**: p. 11-17.
7. Meyer, B., D.G. Papatotiriou, and M. Karas, *100% protein sequence coverage: a modern form of surrealism in proteomics*. Amino Acids, 2011. **41**(2): p. 291-310.
8. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects*. Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
9. Ma, B. and R. Johnson, *de novo sequencing and homology searching*. Mol Cell Proteomics, 2012. **11**(2): p. O111 014902.
10. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review*. Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
11. Domon, B. and R. Aebersold, *Options and considerations when selecting a quantitative proteomics strategy*. Nat Biotechnol, 2010. **28**(7): p. 710-21.
12. Heck, A.J. and J. Krijgsveld, *Mass spectrometry-based quantitative proteomics*. Expert Rev Proteomics, 2004. **1**(3): p. 317-26.
13. Rodriguez-Suarez, E. and A.D. Whetton, *The application of quantification techniques in proteomics for biomedical research*. Mass Spectrom Rev, 2013. **32**(1): p. 1-26.
14. Lohuis, T.D., et al., *Hibernating bears conserve muscle strength and maintain fatigue resistance*. Physiol Biochem Zool, 2007. **80**(3): p. 257-69.
15. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. Anal Chem, 1988. **60**(20): p. 2299-301.
16. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
17. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
18. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down proteomics: facts and perspectives*. Biochem Biophys Res Commun, 2014. **445**(4): p. 683-93.
19. Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics*. Chem Rev, 2013. **113**(4): p. 2343-94.
20. Garcia, B.A., *What does the future hold for Top Down mass spectrometry?* J Am Soc Mass Spectrom, 2010. **21**(2): p. 193-202.
21. Laskay, U.A., et al., *Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments*. J Proteome Res, 2013. **12**(12): p. 5558-69.
22. Wu, C., et al., *A protease for 'middle-down' proteomics*. Nat Methods, 2012. **9**(8): p. 822-4.
23. Capriotti, A.L., et al., *Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics*. J Chromatogr A, 2011. **1218**(49): p. 8760-76.
24. Canas, B., et al., *Trends in sample preparation for classical and second generation proteomics*. J Chromatogr A, 2007. **1153**(1-2): p. 235-58.

25. Speers, A.E. and C.C. Wu, *Proteomics of integral membrane proteins--theory and application*. Chem Rev, 2007. **107**(8): p. 3687-714.
26. Mosley, A.L., et al., *A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus*. J Proteomics, 2009. **72**(1): p. 110-20.
27. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. Nature, 1970. **227**(5259): p. 680-5.
28. Neuhoff, V., et al., *Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250*. Electrophoresis, 1988. **9**(6): p. 255-62.
29. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins*. J Biol Chem, 1975. **250**(10): p. 4007-21.
30. Gorg, A., W. Weiss, and M.J. Dunn, *Current two-dimensional electrophoresis technology for proteomics*. Proteomics, 2004. **4**(12): p. 3665-85.
31. Rabilloud, T., et al., *Two-dimensional gel electrophoresis in proteomics: Past, present and future*. J Proteomics, 2010. **73**(11): p. 2064-77.
32. Westermeier, R. and R. Marouga, *Protein detection methods in proteomics research*. Biosci Rep, 2005. **25**(1-2): p. 19-32.
33. Rabilloud, T., *Silver staining of 2-D electrophoresis gels*. Methods Mol Biol, 1999. **112**: p. 297-305.
34. Marouga, R., S. David, and E. Hawkins, *The development of the DIGE system: 2D fluorescence difference gel analysis technology*. Anal Bioanal Chem, 2005. **382**(3): p. 669-78.
35. Sandra, K., et al., *Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography*. J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **866**(1-2): p. 48-63.
36. Sandra, K., et al., *Highly efficient peptide separations in proteomics. Part 2: bi- and multidimensional liquid-based separation techniques*. J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(11-12): p. 1019-39.
37. Hancock, W.S., et al., *High-pressure liquid chromatography of peptides and proteins VI. Rapid analysis of peptides by high-pressure liquid chromatography with hydrophobic ion-pairing of amino groups*. Journal of Chromatography, 1978. **161**: p. 291-298.
38. Shen, Y., et al., *High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics*. Anal Chem, 2002. **74**(16): p. 4235-49.
39. Percy, A.J., et al., *Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins*. Anal Bioanal Chem, 2012. **404**(4): p. 1089-101.
40. MacNair, J.E., K.C. Lewis, and J.W. Jorgenson, *Ultra-high-pressure reversed-phase liquid chromatography in packed capillary columns*. Anal Chem, 1997. **69**(6): p. 983-9.
41. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals*. Mass Spectrom Rev, 2001. **20**(6): p. 362-87.
42. Dole, M., et al., *Molecular beams of macroions*. J Chem Phys, 1968. **49**: p. 2240-2249.
43. Iribarne J.V. and B. Thomson, *On the evaporation of charged ions from small droplets*. J Chem Phys, 1976. **64**: p. 2287-2294.
44. Paul, W. and h. steinwedel, *Ein neues Massenspektrometer ohne Magnetfeld*. Zeitschrift Naturforschung 1953. **Teil A 8**: p. 448-450.
45. De Hoffman, E. and V. Stroobant, *Spectrométrie de masse Dunod*.
46. Wolff, M.M. and W.E. Stephens, *A Pulsed Mass Spectrometer with Time Dispersion*. Review of Scientific Instruments, 1953. **24**(616).
47. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis*. Science, 2006. **312**(5771): p. 212-7.
48. McLafferty, F.W. and T.A. Bryce, *Metastable-ion characteristics: characterization of isomeric molecules*. . Chem Commun (Camb), 1967: p. 1215-1217.

49. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.
50. Zubarev, R.A., et al., *Electron capture dissociation for structural characterization of multiply charged protein cations*. Anal Chem, 2000. **72**(3): p. 563-73.
51. Dongré, A.R., et al., *Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model*. J Am Chem Soc, 1996. **1996**(118): p. 8365-8374.
52. Sadygov, R.G., D. Cociorva, and J.R. Yates, 3rd, *Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book*. Nat Methods, 2004. **1**(3): p. 195-202.
53. Biemann, K., *Appendix 5. Nomenclature for peptide fragment ions (positive ions)*. Methods Enzymol 1990. **193**: p. 886-887.
54. Swaney, D.L., C.D. Wenger, and J.J. Coon, *Value of using multiple proteases for large-scale mass spectrometry-based proteomics*. J Proteome Res, 2010. **9**(3): p. 1323-9.
55. Hildonen, S., T.G. Halvorsen, and L. Reubsaet, *Why less is more when generating tryptic peptides in bottom-up proteomics*. Proteomics, 2014. **14**(17-18): p. 2031-41.
56. Plumel, M.I., et al., *Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (Dermochelys coriacea)*. J Proteome Res, 2013. **12**(9): p. 4122-35.
57. Swaney, D.L., G.C. McAlister, and J.J. Coon, *Decision tree-driven tandem mass spectrometry for shotgun proteomics*. Nat Methods, 2008. **5**(11): p. 959-64.
58. Chapman, J.D., D.R. Goodlett, and C.D. Masselon, *Multiplexed and data-independent tandem mass spectrometry for global proteome profiling*. Mass Spectrom Rev, 2014. **33**(6): p. 452-70.
59. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
60. Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting*. Curr Biol, 1993. **3**(6): p. 327-32.
61. Thiede, B., et al., *Peptide mass fingerprinting*. Methods, 2005. **35**(3): p. 237-47.
62. Henzel, W.J., C. Watanabe, and J.T. Stults, *Protein identification: the origins of peptide mass fingerprinting*. J Am Soc Mass Spectrom, 2003. **14**(9): p. 931-42.
63. Blueggel, M., D. Chamrad, and H.E. Meyer, *Bioinformatics in proteomics*. Curr Pharm Biotechnol, 2004. **5**(1): p. 79-88.
64. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology*. Nat Biotechnol, 2001. **19**(3): p. 242-7.
65. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database*. J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
66. Geer, L.Y., et al., *Open mass spectrometry search algorithm*. J Proteome Res, 2004. **3**(5): p. 958-64.
67. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-67.
68. Craig, R., J.P. Cortens, and R.C. Beavis, *Open source system for analyzing, validating, and storing protein identification data*. J Proteome Res, 2004. **3**(6): p. 1234-42.
69. Searle, B.C., M. Turner, and A.I. Nesvizhskii, *Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies*. J Proteome Res, 2008. **7**(1): p. 245-53.
70. Wheeler, D.L., et al., *Database resources of the National Center for Biotechnology Information: update*. Nucleic Acids Res, 2004. **32**(Database issue): p. D35-40.
71. Boeckmann, B., et al., *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003*. Nucleic Acids Res, 2003. **31**(1): p. 365-70.

72. O'Donovan, C., et al., *High-quality protein knowledge resource: SWISS-PROT and TrEMBL*. Brief Bioinform, 2002. **3**(3): p. 275-84.
73. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem*. Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.
74. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry*. Nat Methods, 2007. **4**(3): p. 207-14.
75. Rogowska-Wrzesinska, A., et al., *2D gels still have a niche in proteomics*. J Proteomics, 2013. **88**: p. 4-13.
76. Seidler, J., et al., *de novo sequencing of peptides by MS/MS*. Proteomics, 2010. **10**(4): p. 634-49.
77. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.
78. Frank, A. and P. Pevzner, *PepNovo: de novo peptide sequencing via probabilistic network modeling*. Anal Chem, 2005. **77**(4): p. 964-73.
79. Carapito, C., et al., *MSDA, a proteomics software suite for in-depth Mass Spectrometry Data Analysis using grid computing*. Proteomics, 2014. **14**(9): p. 1014-9.
80. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present*. Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
81. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions*. Nat Methods, 2012. **9**(6): p. 555-66.
82. Picotti, P., et al., *Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics*. Cell, 2009. **138**(4): p. 795-806.
83. Cima, I., et al., *Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer*. Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3342-7.
84. Iuga, C., et al., *Proteomic identification of potential prognostic biomarkers in resectable pancreatic ductal adenocarcinoma*. Proteomics, 2014. **14**(7-8): p. 945-55.
85. Stahl-Zeng, J., et al., *High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites*. Mol Cell Proteomics, 2007. **6**(10): p. 1809-17.
86. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial*. Mol Syst Biol, 2008. **4**: p. 222.
87. Kusebauch, U., et al., *Using PeptideAtlas, SRMAtlas, and PASSEL: Comprehensive Resources for Discovery and Targeted Proteomics*. Curr Protoc Bioinformatics, 2014. **46**: p. 13 25 1-13 25 28.
88. Farrah, T., et al., *PASSEL: the PeptideAtlas SRMexperiment library*. Proteomics, 2012. **12**(8): p. 1170-5.
89. Vizcaino, J.A., et al., *A guide to the Proteomics Identifications Database proteomics data repository*. Proteomics, 2009. **9**(18): p. 4276-83.
90. Desiere, F., et al., *The PeptideAtlas project*. Nucleic Acids Res, 2006. **34**(Database issue): p. D655-8.
91. Mohammed, Y., et al., *PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments*. J Proteomics, 2014. **106**: p. 151-61.
92. Carr, S.A., et al., *Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach*. Mol Cell Proteomics, 2014. **13**(3): p. 907-17.
93. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
94. Brun, V., et al., *Isotope-labeled protein standards: toward absolute quantitative proteomics*. Mol Cell Proteomics, 2007. **6**(12): p. 2139-49.
95. Pratt, J.M., et al., *Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes*. Nat Protoc, 2006. **1**(2): p. 1029-43.

96. Gallien, S., et al., *Technical considerations for large-scale parallel reaction monitoring analysis*. J Proteomics, 2014. **100**: p. 147-59.
97. Ronsein, G.E., et al., *Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics*. J Proteomics, 2015. **113**: p. 388-99.
98. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
99. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative*. Nat Chem Biol, 2005. **1**(5): p. 252-62.
100. Schwanhausser, B., et al., *Global analysis of cellular protein translation by pulsed SILAC*. Proteomics, 2009. **9**(1): p. 205-9.
101. Kruger, M., et al., *SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function*. Cell, 2008. **134**(2): p. 353-64.
102. Tall, M.L., et al., *[Injectable preparation of labeled leucine with the carbon 13 for a clinical research program on the Alzheimer disease: Pharmaceutical control of raw materials and the finished product and stability study]*. Ann Pharm Fr, 2015. **73**(1): p. 43-59.
103. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue*. Nat Methods, 2010. **7**(5): p. 383-5.
104. Miyagi, M. and K.C. Rao, *Proteolytic 18O-labeling strategies for quantitative proteomics*. Mass Spectrom Rev, 2007. **26**(1): p. 121-36.
105. Unlu, M., M.E. Morgan, and J.S. Minden, *Difference gel electrophoresis: a single gel method for detecting changes in protein extracts*. Electrophoresis, 1997. **18**(11): p. 2071-7.
106. Timms, J.F. and R. Cramer, *Difference gel electrophoresis*. Proteomics, 2008. **8**(23-24): p. 4886-97.
107. Miller, I., J. Crawford, and E. Gianazza, *Protein stains for proteomic applications: which, when, why?* Proteomics, 2006. **6**(20): p. 5385-408.
108. Campostrini, N., et al., *Spot overlapping in two-dimensional maps: a serious problem ignored for much too long*. Proteomics, 2005. **5**(9): p. 2385-95.
109. Havlis, J. and A. Shevchenko, *Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry*. Anal Chem, 2004. **76**(11): p. 3029-36.
110. Broedel, O., et al., *In-Gel 18O labeling for improved identification of proteins from 2-DE Gel spots in comparative proteomic experiments*. J Proteome Res, 2009. **8**(7): p. 3771-7.
111. Szabo, Z., et al., *Mass spectrometry-based label free quantification of gel separated proteins*. J Proteomics, 2012. **75**(18): p. 5544-53.
112. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags*. Nat Biotechnol, 1999. **17**(10): p. 994-9.
113. Hansen, K.C., et al., *Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography*. Mol Cell Proteomics, 2003. **2**(5): p. 299-314.
114. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
115. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS*. Anal Chem, 2003. **75**(8): p. 1895-904.
116. Merrill, A.E., et al., *NeuCode labels for relative protein quantification*. Mol Cell Proteomics, 2014. **13**(9): p. 2503-12.
117. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics*. Anal Chem, 2003. **75**(24): p. 6843-52.
118. Kovanich, D., et al., *Applications of stable isotope dimethyl labeling in quantitative proteomics*. Anal Bioanal Chem, 2012. **404**(4): p. 991-1009.
119. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics*. Anal Chem, 2004. **76**(14): p. 4193-201.

120. Zhu, W., J.W. Smith, and C.M. Huang, *Mass spectrometry-based label-free quantitative proteomics*. J Biomed Biotechnol, 2010. **2010**: p. 840518.
121. Higgs, R.E., et al., *Quantitative Proteomics via High Resolution MS Quantification: Capabilities and Limitations*. Int J Proteomics, 2013. **2013**: p. 674282.
122. Schilling, B., et al., *Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation*. Mol Cell Proteomics, 2012. **11**(5): p. 202-14.
123. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ*. Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
124. Sandin, M., et al., *Data processing methods and quality control strategies for label-free LC-MS protein quantification*. Biochim Biophys Acta, 2014. **1844**(1 Pt A): p. 29-41.
125. Cappadona, S., et al., *Current challenges in software solutions for mass spectrometry-based quantitative proteomics*. Amino Acids, 2012. **43**(3): p. 1087-108.
126. Callister, S.J., et al., *Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics*. J Proteome Res, 2006. **5**(2): p. 277-86.
127. Chawade, A., E. Alexandersson, and F. Levander, *Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets*. J Proteome Res, 2014. **13**(6): p. 3114-20.
128. Matzke, M.M., et al., *A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments*. Proteomics, 2013. **13**(3-4): p. 493-503.
129. Carrillo, B., et al., *Methods for combining peptide intensities to estimate relative protein abundance*. Bioinformatics, 2010. **26**(1): p. 98-103.
130. Karpievitch, Y., et al., *A statistical framework for protein quantitation in bottom-up MS-based proteomics*. Bioinformatics, 2009. **25**(16): p. 2028-34.
131. Venable, J.D., et al., *Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra*. Nat Methods, 2004. **1**(1): p. 39-45.
132. Purvine, S., et al., *Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer*. Proteomics, 2003. **3**(6): p. 847-50.
133. Moran, D., et al., *Data-independent acquisition (MSE) with ion mobility provides a systematic method for analysis of a bacteriophage structural proteome*. J Virol Methods, 2014. **195**: p. 9-17.
134. Clough, T., et al., *Protein quantification in label-free LC-MS experiments*. J Proteome Res, 2009. **8**(11): p. 5275-84.
135. Seviour, D.K., O. Pelkonen, and J.T. Ahokas, *Hepatocytes: the powerhouse of biotransformation*. Int J Biochem Cell Biol, 2012. **44**(2): p. 257-61.
136. Jennen, D.G., et al., *Biotransformation pathway maps in WikiPathways enable direct visualization of drug metabolism related expression changes*. Drug Discov Today, 2010. **15**(19-20): p. 851-8.
137. Castell, J.V., et al., *Hepatocyte cell lines: their use, scope and limitations in drug metabolism studies*. Expert Opin Drug Metab Toxicol, 2006. **2**(2): p. 183-212.
138. Kikkawa, R., et al., *Investigation of a hepatotoxicity screening system in primary cell cultures - "what biomarkers would need to be addressed to estimate toxicity in conventional and new approaches?"*. J Toxicol Sci, 2005. **30**(1): p. 61-72.
139. Jennings, P., et al., *SEURAT-1 liver gold reference compounds: a mechanism-based review*. Arch Toxicol, 2014. **88**(12): p. 2099-133.
140. Megger, D.A., et al., *Label-free quantification in clinical proteomics*. Biochim Biophys Acta, 2013. **1834**(8): p. 1581-90.
141. Cairns, D.A., *Statistical issues in quality control of proteomic analyses: good experimental design and planning*. Proteomics, 2011. **11**(6): p. 1037-48.
142. Nahnsen, S., et al., *Tools for label-free peptide quantification*. Mol Cell Proteomics, 2013. **12**(3): p. 549-56.

143. Forshed, J., et al., *Enhanced information output from shotgun proteomics data by protein quantification and peptide quality control (PQPQ)*. Mol Cell Proteomics, 2011. **10**(10): p. M111 010264.
144. Wallace, B.P., et al., *Global conservation priorities for marine turtles*. PLoS One, 2011. **6**(9): p. e24510.
145. Jones, P.D., et al., *Vitellogenin as a biomarker for environmental estrogens*. Australasian Journal of Ecotoxicology, 2000. **6**: p. 45-58.
146. Palmer, B.D. and S.K. Palmer, *Vitellogenin induction by xenobiotic estrogens in the red-eared turtle and African clawed frog*. Environ Health Perspect, 1995. **103 Suppl 4**: p. 19-25.
147. Herbst, L.H., et al., *Induction of vitellogenesis by estradiol-17beta and development of enzyme-linked immunosorbent assays to quantify plasma vitellogenin levels in green turtles (Chelonia mydas)*. Comp Biochem Physiol B Biochem Mol Biol, 2003. **135**(3): p. 551-63.
148. Simon, R., et al., *Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum)*. J Chromatogr A, 2010. **1217**(31): p. 5109-15.
149. Simon, R., et al., *Total ApoE and ApoE4 isoform assays in an Alzheimer's disease case-control study by targeted mass spectrometry (n=669): a pilot assay for methionine-containing proteotypic peptides*. Mol Cell Proteomics, 2012. **11**(11): p. 1389-403.
150. Dai, J., et al., *Identification of degradation products formed during performic oxidation of peptides and proteins by high-performance liquid chromatography with matrix-assisted laser desorption/ionization and tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2005. **19**(9): p. 1130-8.
151. Stearns, S.C., *Life history evolution: successes, limitations, and prospects*. Naturwissenschaften, 2000. **87**(11): p. 476-86.
152. Roff, D., *Life History, Evolution of*. Encyclopedia of Biodiversity, 2001. **3**: p. 715-728.
153. Metcalfe, N.B. and P. Monaghan, *Does reproduction cause oxidative stress? An open question*. Trends Ecol Evol, 2013. **28**(6): p. 347-50.
154. Harshman, L.G. and A.J. Zera, *The cost of reproduction: the devil in the details*. Trends Ecol Evol, 2007. **22**(2): p. 80-6.
155. Vandenkoornhuysse, P., et al., *Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics*. Ecol Lett, 2010. **13**(6): p. 776-91.
156. Garratt, M., et al., *Physiological adaptations to reproduction. I. Experimentally increasing litter size enhances aspects of antioxidant defence but does not cause oxidative damage in mice*. J Exp Biol, 2013. **216**(Pt 15): p. 2879-88.
157. Wasselin, T., et al., *Exacerbated oxidative stress in the fasting liver according to fuel partitioning*. Proteomics, 2014. **14**(16): p. 1905-21.
158. Plumel, M.I., et al., *Litter size manipulation in laboratory mice: an example of how proteomic analysis can uncover new mechanisms underlying the cost of reproduction*. Front Zool, 2014. **11**: p. 41.
159. Stier, A., et al., *Constraint and cost of oxidative stress on reproduction: correlative evidence in laboratory mice and review of the literature*. Front Zool, 2012. **9**(1): p. 37.
160. Sheldon, B.C. and S. Verhulst, *Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology*. Trends Ecol Evol, 1996. **11**(8): p. 317-21.
161. Franceschi, C., et al., *Inflamm-aging. An evolutionary perspective on immunosenescence*. Ann N Y Acad Sci, 2000. **908**: p. 244-54.
162. Cao, Z., et al., *Proteomics reveals age-related differences in the host immune response to sepsis*. J Proteome Res, 2014. **13**(2): p. 422-32.
163. O'Connor, J.E., et al., *Systems Biology and immune aging*. Immunol Lett, 2014. **162**(1 Pt B): p. 334-45.
164. Bokov, A., A. Chaudhuri, and A. Richardson, *The role of oxidative damage and stress in aging*. Mech Ageing Dev, 2004. **125**(10-11): p. 811-26.

165. Aschoff, J., *Exogenous and endogenous components in circadian rhythms*. Cold Spring Harb Symp Quant Biol, 1960. **25**: p. 11-28.
166. Golombek, D.A. and R.E. Rosenstein, *Physiology of circadian entrainment*. Physiol Rev, 2010. **90**(3): p. 1063-102.
167. Mendoza, J., *Circadian clocks: setting time by food*. J Neuroendocrinol, 2007. **19**(2): p. 127-37.
168. Deery, M.J., et al., *Proteomic analysis reveals the role of synaptic vesicle cycling in sustaining the suprachiasmatic circadian clock*. Curr Biol, 2009. **19**(23): p. 2031-6.
169. Ko, C.H. and J.S. Takahashi, *Molecular components of the mammalian circadian clock*. Hum Mol Genet, 2006. **15 Spec No 2**: p. R271-7.
170. Eide, E.J., et al., *Control of mammalian circadian rhythm by CKIepsilon-regulated proteasome-mediated PER2 degradation*. Mol Cell Biol, 2005. **25**(7): p. 2795-807.
171. Foster, R.G. and K. Wulff, *The rhythm of rest and excess*. Nat Rev Neurosci, 2005. **6**(5): p. 407-14.
172. Reddy, A.B., et al., *Circadian orchestration of the hepatic proteome*. Curr Biol, 2006. **16**(11): p. 1107-15.
173. Robles, M.S., J. Cox, and M. Mann, *In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism*. PLoS Genet, 2014. **10**(1): p. e1004047.
174. Mauvoisin, D., et al., *Proteomics and circadian rhythms: It's all about signaling!* Proteomics, 2014.
175. Mendoza, J., et al., *The cerebellum harbors a circadian oscillator involved in food anticipation*. J Neurosci, 2010. **30**(5): p. 1894-904.
176. Rabilloud, T., et al., *Detergents and chaotropes for protein solubilization before two-dimensional electrophoresis*. Methods Mol Biol, 2007. **355**: p. 111-9.
177. Shevchenko, G., et al., *Comparison of extraction methods for the comprehensive analysis of mouse brain proteome using shotgun-based mass spectrometry*. J Proteome Res, 2012. **11**(4): p. 2441-51.
178. Cermenati, G., et al., *Lipids in the nervous system: from biochemistry and molecular biology to patho-physiology*. Biochim Biophys Acta, 2015. **1851**(1): p. 51-60.
179. Cornelissen, G., *Cosinor-based rhythmometry*. Theor Biol Med Model, 2014. **11**: p. 16.
180. Sayer, A.A., et al., *New horizons in the pathogenesis, diagnosis and management of sarcopenia*. Age Ageing, 2013. **42**(2): p. 145-50.
181. Bonaldo, P. and M. Sandri, *Cellular and molecular mechanisms of muscle atrophy*. Dis Model Mech, 2013. **6**(1): p. 25-39.
182. Bodine, S.C., *Hibernation: The search for treatments to prevent disuse-induced skeletal muscle atrophy*. Experimental Neurology, 2013. **248**: p. 129-135.
183. Harlow, H.J., et al., *Body mass and lipid changes by hibernating reproductive and nonreproductive black bears (Ursus americanus)*. Journal of Mammalogy, 2002. **83**(4): p. 1020-1025.
184. Lohuis, T.D., H.J. Harlow, and T.D. Beck, *Hibernating black bears (Ursus americanus) experience skeletal muscle protein balance during winter anorexia*. Comp Biochem Physiol B Biochem Mol Biol, 2007. **147**(1): p. 20-8.
185. Narici, M.V. and M.D. de Boer, *Disuse of the musculo-skeletal system in space and on earth*. Eur J Appl Physiol, 2011. **111**(3): p. 403-20.
186. Tonge, R., et al., *Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology*. Proteomics, 2001. **1**(3): p. 377-96.
187. Karp, N.A. and K.S. Lilley, *Maximising sensitivity for detecting changes in protein expression: experimental design using minimal CyDyes*. Proteomics, 2005. **5**(12): p. 3105-15.
188. Viswanathan, S., M. Unlu, and J.S. Minden, *Two-dimensional difference gel electrophoresis*. Nat Protoc, 2006. **1**(3): p. 1351-8.

189. Karp, N.A., et al., *Experimental and statistical considerations to avoid false conclusions in proteomics studies using differential in-gel electrophoresis*. Mol Cell Proteomics, 2007. **6**(8): p. 1354-64.
190. Arentz, G., et al., *State of the art of 2D DIGE*. Proteomics Clin Appl, 2014.
191. Wang, P., F.G. Bouwman, and E.C. Mariman, *Generally detected proteins in comparative proteomics--a matter of cellular stress response?* Proteomics, 2009. **9**(11): p. 2955-66.
192. Petrak, J., et al., *Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins*. Proteomics, 2008. **8**(9): p. 1744-9.
193. Gauthier-Clerc, M., et al., *Penguin fathers preserve food for their chicks*. Nature, 2000. **408**(6815): p. 928-9.
194. Landon, C., et al., *Solution structure of spheniscin, a beta-defensin from the penguin stomach*. J Biol Chem, 2004. **279**(29): p. 30433-9.
195. Sano, S., et al., *Absolute quantitation of low abundance plasma APL1beta peptides at sub-fmol/mL Level by SRM/MRM without immunoaffinity enrichment*. J Proteome Res, 2014. **13**(2): p. 1012-20.
196. Aebersold, R., *A stress test for mass spectrometry-based proteomics*. Nat Methods, 2009. **6**(6): p. 411-2.

Partie expérimentale

Partie expérimentale

Cette partie contient les parties expérimentales des projets qui n'ont pas encore fait l'objet de publication à comité de lecture.

Partie expérimentale : Chapitre V. Applications et développement de la stratégie quantitative globale XIC à des échantillons biologiques complexes

13 groupes d'individus ont été comparés :

- **Jour 0** non contaminé
- **Jour 2** : 0 mM ; 0,05mM VA ; 0,15 mM VA et 0,5 mM VA
- **Jour 8** : 0 mM ; 0,05mM VA ; 0,15 mM VA et 0,5 mM VA
- **Jour 14** : 0 mM ; 0,05mM VA ; 0,15 mM VA et 0,5 mM VA

A. Préparation des échantillons

A.1. Extraction des protéines et dosage Bradford

La culture des cellules a été réalisée par Sebastian Klein (Saarland University). La quantité de cellules estimée dans chaque vial étaient de 2.4 millions de cellules. Les cellules ont été lysées directement dans le vial dans un tampon à base de sarkosyl (0,1%). Après centrifugation, les surnageants ont été collectés puis directement congelés à -80°C.

Ces échantillons ont tout d'abord été dosés selon un dosage de Bradford afin d'évaluer la quantité de matériel protéique extraite de chaque échantillon. Dans la suite du protocole, chaque échantillon a été préparé en dupliquât. 50 µg de chacun des échantillons a été précipité par une solution d'acide trichloracétique finale à 10% (w/v). Les culots ont ensuite été lavés avec du tétrahydrofurane (3x 400 µL) afin d'éliminer le sarkosyl précipité.

A.2. Traitement sans préfractionnement

Le protocole utilisé pour les tests ou le traitement des échantillons finaux par digestion liquide est identique.

50 µg de protéines ont été resolubilisées dans 200 µL d'un tampon de lyse (urée 8 M, NH_4HCO_3 0,1M). Les ponts disulfures ont été réduits (solution de dithiotréthol 700mM, 37°C, 30 minutes), puis alkylés (solution de Iodoacétamide 700mM, température ambiante, obscurité). 1500 µL de tampon NH_4HCO_3 0,1 mM a ensuite été ajouté en vue de la digestion. 0,5 µg de trypsine (Promega) ont été ajoutés pour 50 µg de protéines par échantillon (37°, 12h) Après digestion, les échantillons ont été purifiés sur cartouche C18 (sep-pack C18, conditionnement : 4mL méthanol, 4mL acétonitrile ACN, 6mL HCOOH 0,1%, lavage : 6mL HCOOH 0,1%, élution 600µL 50% ACN, 0,1% HCOOH) puis une quantité connue de peptides iRT (Biognosis) a été mélangée à chacun des échantillons afin d'assurer le suivi du système instrumental. Les peptides iRT ont été repris dans 50 µL de solution de reprise (Biognosis) + 450 µL 0,1% HCOOH . 5 µL de cette solution a été incorporé à chaque échantillon avant l'étape d'évaporation des solvants. Enfin, l'acétonitrile présent dans tous les échantillons a été évaporé (Speed-vac). Les échantillons ont été repris dans 100µL de H_2O (0,1% HCOOH).

A.3. Gels d'électrophorèse SDS-Page (tests de mise au point) et digestion in-gel

Après précipitation, les extraits protéiques ont été resolubilisés dans un tampon type Laemmli (10 mM Tris pH 8, EDTA 1 mM, β ME 5%, SDS 5%, glycérol 10% et bleu de bromophénol 0,10%). Le gel SDS-Page était composé d'un gel de « concentration » et d'un gel de « séparation » tous deux à base de polyacrylamide réticulés à 5% et 12% respectivement (0,37 M Tris pH 8,8, SDS 0,1%, Temed 0,1% et persulfate NH_4 0,025% pour le gel de séparation et 0,15 M Tris pH 6,8, SDS 0,1%, Temed 0,025% et persulfate NH_4 0,1% pour le gel de concentration).

Les conditions électrophorétiques ont été fixées à 50 V pendant 30 minutes. Le gel a été maintenu sous agitation dans une solution de fixation pendant 1 heure (3% H_3PO_4 , 50% $\text{CH}_3\text{CH}_2\text{OH}$) puis maintenu sous agitation pendant deux heures dans une solution de coloration à base de bleu de Coumassie (0,1% Coumassie R250, 10% CH_3COOH , 40% CH_3OH).

Les bandes de gel ont été excisés manuellement à l'aide d'un emporte-pièce (2mm par bande environ).

Les bandes de gel ont été décolorées et les protéines contenues dans ces bandes réduites puis alkylées à l'aide d'une station « Massprep Station » (Waters). Ce traitement combine une étape de décoloration (3 cycles de lavage avec 50 µL de NH_4HCO_3 25 mM et 50 µL d'acétonitrile, 10 minutes), une étape de déshydratation (50 µL d'acétonitrile, 60°C, 5 minutes), une étape de réduction des ponts disulfures (50 µL de 10mM DTT, 25mM NH_4HCO_3 , 60°C, 30 minutes), une étape d'alkylation des

protéines (50 μL de 55mM Iodoacétamide NH_4HCO_3 25mM, 30 minutes), une nouvelle étape de lavage (1 cycle 50 μL de NH_4HCO_3 à 25 mM/ 50 μL d'acétonitrile, 10 minutes) et une étape de déshydratation du gel (50 μL d'acétonitrile, 15 minutes).

Les protéines ont ensuite été digérées enzymatiquement par la trypsine (Promega). Les bandes déshydratées ont été réhydratées avec 6 μL d'une solution de trypsine (12,5 ng/L, NH_4HCO_3 25 mM) + 10 μL de tampon NH_4HCO_3 25 mM. La digestion enzymatique a duré 12h à 37°C. Les peptides ont été extraits des morceaux de gels par ajout de 30 μL d'une solution à 60 % acétonitrile, 0,1% HCOOH. Les surnageants ont été prélevés et l'acétonitrile a été évaporé sous vide (Speed vac, Thermo Scientific).

C. Analyse des protéines en nanoLC-MS/MS (*Impact-HD, Bruker Daltonique*)

Après purification des digestats, les peptides tryptiques ont été analysés sur un système UPLC (nanoAcquity, Waters) couplé à un spectromètre de masse combinant une source « CaptiveSpray » couplée au « NanoBooster » (Bruker Daltonique) et un analyseur de type temps de vol (Impact-HD; Bruker Daltonique). Le couplage nanoHPLC-MS/MS était contrôlé respectivement par les logiciels Bruker compass Hystar (v 3.2) et OtofControl (Rev 3.4, Bruker Daltonik GmbH).

Le système UPLC utilisé combinait une colonne d'enrichissement C18, 180 μm x 20 mm, 5 μm , une colonne de séparation zorbax C18, 250mm x 75 μm , 1.7 μm . Le volume injecté était de 2 μL , le débit de chargement sur la précolonne de 5 $\mu\text{L}/\text{min}$ (0,1% HCOOH, H_2O). Les peptides ont été élués selon un gradient de 180 minutes (respectivement 60 minutes) de 2 à 35% B avec A : H_2O , 0,1% HCOOH et B: Acétonitrile, 0,1% HCOOH (450 nL/min) pour les échantillons traités respectivement en digestion liquide et déposés sur gel SDS-Page.

Les paramètres de sélection des peptides afin de les isoler puis de les fragmenter (mode CID – Collision induced dissociation) sont présentés dans le tableau 1.

Tableau 1: Paramétrage des séquences d'acquisition en spectrométrie de masse

MS	Gamme de masse	100-2200 m/z
	Temps de cycle	3 secondes
Fragmentation CID	Etats de charge	2x-5x
MS/MS	Cycle d'exclusion	1 spectres puis 1 min d'exclusion
	Gamme de masse	100-2200 m/z
	Vitesse de scan (seuil d'intensité)	2-25 Hz

D. Recherche dans les banques de données

L'algorithme de recherche Mascot v2.3.02 (Matrix Science, London, UK) a été utilisé pour l'identification des protéines à partir des données MS et MS/MS. La recherche a été réalisée dans une banque de donnée restreinte à la taxonomie « Homo Sapiens » (Taxon ID 9606) extraite à partir de la base de donnée UniprotKB/Swissprot (téléchargement en février 2014 ; 20372 entrées).

La tolérance accordée sur les erreurs de masse en mode MS était de 15 ppm et 0,02 Daltons pour les MS/MS pour le moteur de recherche Mascot. La tolérance sur le nombre de site de clivage manqué par l'enzyme protéolytique était de 1. La carbamidomethylation des cystéines et l'oxydation des méthionines des protéines ont été spécifiées comme modifications variables.

Afin de ne considérer que les spectres de qualité, des critères de filtres stricts sur les scores d'identifications des ions ont été considérés à l'aide du logiciel Scaffold (v_3_00_07, Proteome software Inc.). L'identification des protéines était validée à partir d'un seul peptide unique à condition de remplir les critères suivants :

- Ion score > 25 pour les résultats provenant du moteur de recherche Mascot
- Pourcentage de faux positif < 1%

Les contaminants usuels, comme la kératine et la trypsine, ont été éliminés des données.

E. Statistiques

Les analyses statistiques ont été réalisées à l'aide du logiciel « R » (v2.8.1 ; <http://www.R-project.org>). Après avoir vérifié la normalité des données (test de Shapiro-Wilk, p-value > 0,01) ainsi que l'égalité des variances (test de Bartlett, p-value > 0,01), les données ont été évaluées selon un test ANOVA (p-value < 0,05) puis un test Post-hoc de Tukey (p-value < 0,01). Un facteur d'expression minimum a également été établi entre deux conditions en plus du test statistique afin de définir la significativité des données.

Partie expérimentale : Chapitre VII. Applications et optimisations de la stratégie quantitative DIGE (Differential In-Gel Electrophoresis) à des échantillons biologiques complexes

A. Etude du coût de l'activation du système immunitaire chez la souris

4 groupes d'individus ont ainsi été comparés :

- PBS young (groupe jeune – pas de challenge immunitaire)
- LPS young (groupe jeune – challenge immunitaire avec injection de LPS)
- PBS old (groupe vieux – pas de challenge immunitaire)
- LPS old (groupe vieux - challenge immunitaire avec injection de LPS)

A.1. Réalisation des gels

Huit gels 2D-DIGE ont été réalisés afin de procéder à une analyse d'image différentielle.

Les 16 échantillons ont été répartis sur les gels 2D-DIGE suivant une répartition aléatoire entre les différents fluorochromes (tableau 2). Les marqueurs « Cy3 » et « Cy5 » (fluorophores) ont servi à marquer les 16 échantillons (50 µg de protéines totales / échantillon, marquées soit par la Cy3, soit par la Cy5). Le marqueur « Cy2 » a servi à marquer un mélange en équiquantité des 16 échantillons (pour dépôt sur chaque gel de 50µg de protéines totales marquées) qui sert de standard interne pour l'analyse différentielle. Les gels DIGE ont été scannés par un Ettan DIGE Imager (GE Healthcare). La résolution a été fixée à 100 µm.

Tableau 2 : Répartition des échantillons et des marqueurs sur les gels 2D-DIGE

N° GEL	Cy2	Cy3	Cy5
1	SI	140	136
2	SI	143	84
3	SI	137	121
4	SI	134	47
5	SI	86	130
6	SI	85	45
7	SI	49	122
8	SI	46	83

Un gel coloré au bleu de Coumassie contenant plus de matériel protéique que les gels 2D-DIGE (mélange de 62,5 µg de chaque échantillon soit 1000 µg chargés au total) a également été réalisé afin de procéder à l'analyse par spectrométrie de masse des protéines contenues dans les spots de gel (conditions électrophorétiques 2D-DIGE/ Gel bleu identiques).

La similarité des profils de migration des protéines a été vérifiée entre les gels 2D-DIGE qui ont servi à l'analyse d'image et le gel coloré au bleu de Coumassie qui a servi pour l'analyse qualitative des protéines contenues dans les spots protéiques avec la vérification en spectrométrie de masse d'une dizaine de spots DIGE.

A.2. Analyse d'image

L'analyse d'image a été réalisée grâce au logiciel « Progenesis Samespots » (v4.5, Nonlinear Dynamics). Les images de gel ont été alignées les unes par rapport aux autres par le logiciel de façon automatique. Un contrôle visuel a été réalisé sur cet alignement et des corrections éventuelles ont été apportées manuellement. Les « volumes densitométriques » des spots contenant des protéines marquées par les Cy 3 ou Cy5 ont été normalisés par rapport à l'ensemble des spots contenus sur chaque gel et par rapport aux volumes des spots contenant les protéines marquées par le standard interne (Cy2).

A.3. Statistiques

Les données présentées sont les volumes densitométriques (normalisés par rapport aux volumes des spots protéiques marqués par les Cy2), ainsi que les moyennes \pm ESM de ces volumes, des spots protéiques pour chaque groupe. Ces données ont été normalisées par rapport à la moyenne du groupe PBS Young. Les analyses statistiques ont été réalisées à l'aide du logiciel « R » (v2.8.1 ; <http://www.R-project.org>). La distribution normale des variables mesurées a été vérifiée par le test de Shapiro-Wilk et l'égalité des variances par le test de Bartlett. Les variations entre les groupes ont été testées par ANOVA suivie du test post-hoc de Tukey. Le niveau de significativité a été établi à $P > 0,01$ pour les tests de Shapiro-Wilk et de Bartlett et à $P < 0,05$ pour les tests ANOVA et Tukey.

A.4. Analyse des protéines contenues dans les spots de gels en nanoLC-MS/MS

Après digestion in-gel des protéines, les peptides tryptiques ont été analysés sur un système nanoHPLC Chip-Cube (Agilent) couplé à un spectromètre de masse combinant une source

électrospray et un analyseur de type trappe ionique (HCT Plus ; Bruker Daltonique). Le couplage nanoHPLC-MS/MS était contrôlé respectivement par les logiciels ChemStation (v 5.3, Agilent technologies) et EsquireControl (Rev B.01.03, Bruker Daltonique)

Le système nanoLC/chip (Agilent) utilisé combinait une colonne d'enrichissement ZORBAX 300SB-C18, 4 mm, 40 nL, 5 µm, une colonne de séparation ZORBAX 300SB-C18, 43mm x 75 µm, 5 µm et une pointe de spray. Le volume injecté était de 3 µL, le débit de chargement sur la précolonne de 3.75 µL/min (2%AcN, 0,1% HCOOH, H₂O). Les peptides ont été élués selon le gradient : t=0 8% B, t= 7 min 40% B, t=8 min, 70%B, t= 10 min 70% B avec A : H₂O, 2% AcN, 0,1% HCOOH et B: Acétonitrile, 2% H₂O, 0,1% HCOOH (300 nL/min).

Les paramètres de sélection des peptides afin de les isoler puis de les fragmenter (mode CID – Collision induced dissociation) sont présentés dans le tableau 3.

Tableau 3: Paramétrage des séquences d'acquisition en spectrométrie de masse

MS	Accumulation dans la trappe	100 000 ions ou 20 ms
	Vitesse de balayage	8100 m/z par seconde
	Gamme de masse	250-2000 m/z
Fragmentation CID	Nombre d'ions sélectionnés	3 ions dichargés les plus intenses
	Etats de charge	2x
	Cycle d'exclusion	1 minute ou 2 spectres
	Accumulation dans la trappe	300000 ions ou 20 ms
MS/MS	Vitesse de balayage	26000m/z par seconde
	Gamme de masse	50- 2800 m/z

Deux algorithmes de recherche ont été utilisés pour l'identification des protéines à partir des données MS et MS/MS, Mascot v2.4.1 (Matrix Science, London, UK) et Omssa v2.1.9 (Open Mass Spectrometry Search Algorithm). La recherche a été réalisée dans une banque de donnée restreinte à la taxonomie « mus musculus » (souris) extraite à partir de la base de donnée NCBIInr (téléchargement en mai 2011 ; 288050 entrées). La tolérance accordée sur les erreurs de masse en mode MS et MS/MS était de 0,25 Daltons. La tolérance sur le nombre de site de clivage manqué par l'enzyme protéolytique était de 1. La carbamidomethylation des cystéines, l'oxydation des méthionines et l'acétylation N-terminale des protéines ont été spécifiées comme modifications

variables. Afin de ne considérer que les spectres de qualité, des critères de filtres stricts sur les scores d'identifications des ions ont été considérés à l'aide du logiciel Scaffold (v_3_00_07, Proteome software Inc.). L'identification d'une protéine à partir d'un seul peptide était validée pour des scores d'ions MS/MS supérieurs à 45 (Mascot) et pour des valeurs $-\log E$ supérieures à 5.3 (Omssa). L'identification d'une protéine à partir d'au moins 2 peptides était validée pour des scores d'ions MS/MS supérieurs à 30 (Mascot) et pour des valeurs $-\log E$ supérieure à -0,05 (Omssa). Une stratégie target-decoy a été employée pour estimer le taux de faux positifs parmi les identifications. Les contaminants usuels, comme la kératine et la trypsine, ont été éliminés des données.

B. Etude des rythmes circadiens chez la souris

Informations sur les échantillons : 6 groupes de souris ont été comparés.

Souris nourries ab libitum : Echantillonnage à T5, T11, T17 et T23

Souris nourries en période diurne : Echantillonnage à T5 et T17

B.1. Evaluation du protocole de traitement des échantillons

B.1.2. Extraction des protéines

Deux demi cervelets ont été découpés sous azote et solubilisés dans 200-300 μL de tampon de lyse soit :

- 8M Urée, 2M thio-urée, 4% Chaps, 1% DTT, 0,5% Triton X-100, 0,005% TLCK (T7254 Sigma), cocktail commercial d'inhibiteur de protéase 1/50 (P8340, Sigma)
- 7M Urée, 2M thio-urée, 1% Octyl C-D-glucopyranoside (βOG), 10 mM Tris pH 8,8, 0,005% TLCK, cocktail d'inhibiteur de protéase 1/50 (P 8340, Sigma)

Les échantillons ont été vortexés, agités pendant 1h30 puis soniqués pendant 10 -20 secondes (amplitude 10-20%) dans un bain de glace (Digital Cell Disruptor, Branson). Après centrifugation (5 minutes, 1200 g, 4°C), les surnageants ont été collectés puis divisés en deux pour les tests de précipitation des protéines.

B.1.3. Précipitation des protéines

Deux types de précipitations des protéines ont été évalués : précipitation à l'acétone ou à l'acide trichloracétique (TCA, 10%).

1800µL d'acétone froid ou 200-300 µL de TCA (20%) ont été ajoutés aux échantillons extraits avec le triton ou le βOG. Les échantillons ont été incubés à -20°C (une nuit pour le traitement à l'acétone, 1h pour le traitement au TCA). Les échantillons ont ensuite été centrifugés pendant 15 minutes (4°C, 14000 g) et les surnageants ont été collectés. Les culots résiduels de la précipitation par acétone ont été séchés au speed-vac tandis que les culots résiduels de la précipitation au TCA ont été lavés trois fois avec 200 µL d'acétone, 20 mM DTT (jusqu'à l'élimination complète de toute présence d'acide vérifié par papier pH). Les culots ont ensuite été séchés au speed-vac. Les échantillons ont ensuite été repris dans un tampon de lyse (7M Urée, 2M thio-urée, 4% Chaps, 30 mM Tris pH 8,5) vortexés, agités à température ambiante pendant 30 minutes puis soniqués pendant 10 secondes (amplitude 20%, Digital Cell Disruptor, Branson). Les extraits ont ensuite été dosés par dosage de Bradford.

B.1.4. Dosage des protéines et vérification du dosage par gel SDS-Page

La concentration des protéines dans les échantillons a été mesurée selon la méthode de Bradford (Protein Assay Dye Reagent Concentrate, 500-0006, Bio-Rad). La gamme d'étalonnage a été réalisée à partir d'albumine de sérum bovin (0- 10µg). Le dosage de Bradford réalisé a été validée par réalisation d'un gel SDS-Page (10x10 cm): les protéines de chaque échantillon ont été déposées séparément en équiquantité sur le gel (10 µg de protéines de chaque échantillon repris dans une solution tampon à 10 mM Tris pH 8, EDTA 1mM, βME 5%, SDS 5%, glycérol 10% et bleu de bromophénol 0,10%). Le gel 1D était composé d'un gel de « concentration » (permettant la concentration des échantillons en début de gel) et d'un gel de « séparation » tous deux à base de polyacrylamide réticulés à 5% et 12%, respectivement (0,37 M Tris pH 8.8, SDS 0,1%, Temed 0,1% et Persulfate NH₄ 0,025% pour le gel de séparation et (0,15 M Tris pH 6.8, SDS 0,1%, Temed 0,025% et Persulfate NH₄ 0,1% pour le gel de concentration. Les conditions électrophorétiques ont été fixées à 50 V pendant 90 minutes puis le voltage a été augmenté à 100V pendant environ 1h. Le gel a été agité dans une solution de fixation pendant 30 minutes (3% H₃PO₄, 50% CH₃CH₂OH) puis laissé pendant une nuit dans une solution de coloration à base de bleu de Coumassie (0,1% Coumassie R250, 10% CH₃COOH, 40% CH₃OH). Le dosage était considéré comme valide si l'intensité des bandes entre les échantillons était similaire.

B.1. 5. Electrophorèse bidimensionnelle des gels tests (migration verticale)

300 µg de protéines ont été repris dans un volume de 400 µL d'une solution d'urée 7M, thio-urée 2M, 2% CHAPS, 2% DTT, 2% Ampholytes (servalytes 3-10, Serva) et quelques gouttes de bleu de bromophénol. Le tout a été déposé sur une bande de gel à gradient de pH immobilisé (ReadyStrip IPG, Biorad) de 18 cm pH 3-10 non linéaire (Ready IPG strips, Bio-Rad). La réhydratation passive a duré 2h à température ambiante. Chaque bande de gel a été recouverte de 2mL d'huile végétale. La réhydratation active a ensuite été réalisée par l'application d'un courant de 50V pendant une nuit à température ambiante. Après réhydratation, un gradient de tension a été appliqué (0-200V/1h linéaire ; 200-1000 V/ 4h linéaire ; 1000-5000 V/ 16h linéaire, 5000V/ 7h). Les strips ont ensuite été congelés à -80°C.

Pour la réalisation de la migration électrophorétique, les strips ont décongelés pendant 15 minutes. Ils ont ensuite été incubés pendant 30 minutes dans une solution d'équilibration (6M Urée, 50 mM Tris-HCl pH 8,8, 30% glycérol, 2% SDS avec 1 de DTT puis 30 minutes dans le même tampon avec 2,5% d'iodoacétamide. Les strips ont ensuite été rincés avec de l'eau puis incubés quelques minutes dans du tampon d'électrophorèse (0,3% Tris, 1.5% glycine, 0,1% SDS). Les strips ont ensuite été déposés sur de gels de polyacrylamide réticulé à 12% (0,37M Tris pH 8,8, 0,1% SDS, 0,1% Temed, 0,025% persulfate d'ammonium). Les conditions électrophorétiques ont été fixées à 5 mA par gel pendant 15h (Protean II Cell, Bio-Rad).

Les gels ont ensuite été fixés dans une solution de fixation pendant 30 minutes (3 % H₃PO₄, 50 % CH₃CH₂OH) puis laissé pendant une nuit dans une solution de coloration à base de bleu de Coumassie (0,1% Coumassie R250, 10% CH₃COH, 40% CH₃OH). Les gels ont été scannés après rinçage.

B.2. Analyse 2D-DIGE sur l'ensemble des échantillons

B.2.1. Extraction des protéines

Les protéines ont été extraites selon le protocole décrit ci-dessus à base de Triton X-100.

B.2.2. Précipitation des protéines

Le contenu protéique des échantillons a été précipité selon le protocole basé sur l'utilisation de l'acide trichloracétique décrit ci-dessus.

B.2.3. Dosage de Bradford et vérification par gel 1D

Les extraits protéiques ont été dosés par dosage de Bradford comme décrit ci-dessus. La validité du dosage a été vérifiée par gel SDS-Page comme décrit ci-dessus.

B.2.4. Marquage des protéines pour analyse 2D-DIGE

Echantillons : Les cyanines Cy2, Cy3 et Cy5 (CyDyes, Amersham Biosciences, GE Healthcare), anhydres, ont été reconstituées dans du diméthylformamide anhydre. Le marquage se fait par ajout de 400 picomoles de CyDyes à 50 µg de protéines (reprises dans 20 µL de solution à base d'urée (7M), thio-urée (2M), CHAPS 4%, et 30 mM Tris). Après centrifugation, les échantillons ont été incubés à l'obscurité pendant 30 minutes dans la glace. 1 µL de solution stop de lysine (10mM) a alors été ajouté à chaque échantillon, puis les échantillons ont été incubés une seconde fois à l'obscurité pendant 10 minutes dans la glace.

Standard interne : 25 µg de protéines de chaque échantillon ont été mélangés. Le mélange a ensuite été repris dans 120 µL de solution à base d'urée (7M), thio-urée (2M), CHAPS 4%, et 30 mM Tris. Le marquage a ensuite été réalisé avec 2,4 ng de CyDyes selon le procédé énoncé précédemment.

B.2.5. Electrophorèse bidimensionnelle (E2D)

Deux échantillons marqués (soit 2 x 50 µg de protéines) ainsi que 50 µg de protéines du standard interne marqué également ont été mélangés avant la première étape d'isoélectrofocalisation selon la répartition du tableau 1 présenté ci-dessous.

Tableau 4 : Répartition des marqueurs Cy2, Cy3 et Cy5 en fonction des échantillons et des gels (SI= Standard interne)

N° gel	Cy2	Cy3	Cy5	N° gel	Cy2	Cy3	Cy5
1	IS	T5-1	T11-1	10	IS	T17-3	T23-3
2	IS	T5-2	T17-1	11	IS	T17-4	rT5-3
3	IS	T23-1	T5-3	12	IS	rT17-3	T17-5
4	IS	rT5-1	T17-2	13	IS	rT5-4	T23-4
5	IS	T5-5	rT17-1	14	IS	rT17-4	T23-5
6	IS	T11-2	T5-4	15	IS	rT17-5	rT5-5
7	IS	T23-2	T11-3	16	IS	T11-6	T17-6
8	IS	rT5-2	T11-4	17	IS	rT5-6	T5-6
9	IS	T11-5	rT17-2	18	IS	T23-6	rT17-6

SI : standard interne

Le mélange de 150 µg de protéines a été repris dans un volume de 500 µL d'une solution d'urée 7M, thio-urée 2M, 2% CHAPS, 2% DTT, 2% Ampholytes (servalytes 3-10, Serva) et quelques gouttes de bleu de bromophénol. Le tout a été déposé sur une bande de gel à gradient de pH immobilisé (ReadyStrip IPG, Biorad) de 24 cm pH 3-10 non linéaire (Ready IPG strips, Bio-Rad). La réhydratation passive a duré 1h. Chaque bande de gel a été recouverte de 2mL d'huile végétale. La réhydratation active a ensuite été réalisée par l'application d'un courant de 50 V pendant une nuit. Après réhydratation, un gradient de tension a été appliqué (0-200V/1h linéaire ; 200-1000 V/ 4h linéaire ; 1000-5000 V/ 16h linéaire, 5000V/ 7h). La dernière étape a été rallongée de façon à atteindre 95 500Vh. L'isoélectrofocalisation a duré 30 heures (Protean ief cell, Biorad). Les strips ont ensuite été congelés à -80°C.

La migration des gels 2D a été réalisée à l'horizontal selon le protocole développé pour la tour HPE (HPE, FlatTop Tower, Serva). Les gels utilisés étaient des gels de polyacrylamide précoulés réticulés à 12,5%. Après décongélation, les strips ont été incubés à l'obscurité pendant 15 minutes dans le tampon d'équilibration du kit (Serva) avec 1% DTT puis 15 minutes dans le tampon d'équilibration (Serva) avec 2,5% d'iodoacétamide. Les bandes de gel ont ensuite été déposées sur les gels précoulés. Les conditions électrophorétiques suivies sont présentées en tableau 5.

Tableau 5 : Conditions électrophorétiques suivies pour la migration 2D

Étape	Max V	Max mA	Max W	Temps
1	100	7	1	30 min
2	200	13	3	30 min
3	300	20	5	10 min
4	1000	40	30	3h50
5	1500	45	40	40 min

Les scans des gels ont été réalisés à partir d'un Ettan DIGE Imager (**Cy2** longueur d'onde d'absorption 491 ± 3 nm, longueur d'onde d'émission 506 ± 5 nm, 1 minute d'exposition ; **Cy3** longueur d'onde d'absorption 553 ± 3 nm, longueur d'onde d'émission 572 ± 5 nm, 0,2 minutes d'exposition ; **Cy5** longueur d'onde d'absorption 648 ± 3 nm, longueur d'onde d'émission 669 ± 5 nm, 0,5 minutes d'exposition).

B.2.6. Analyse d'image des gels 2D-DIGE (Nonlinear Dynamics, Samespots)

L'analyse d'image à partir des gels DIGE a été réalisée grâce au logiciel « Progenesis Samespots » (v4.5, Nonlinear Dynamics). Les images de gel ont été alignées les unes par rapport aux autres par le logiciel de façon automatique. Un contrôle visuel a été réalisé sur cet alignement et des corrections éventuelles ont été apportées manuellement. Les « volumes densitométriques » des spots contenant des protéines marquées par les fluorochromes Cy2, Cy3 et Cy5 ont été normalisés par rapport au « volume » de l'ensemble des spots contenus sur chaque gel, puis les valeurs Cy3 ou Cy5 ont été normalisée par rapport aux valeurs de standard interne.

B.3. Statistiques

Les analyses statistiques ont été réalisées à l'aide du logiciel « R » (v2.8.1 ; <http://www.R-project.org>). La normalité des valeurs ainsi que l'égalité des variances ont été vérifiées par les tests de Shapiro-Wilk et Bartlett. Les niveaux de significativité retenues étaient $p > 0,01$.

Les données ont été testées selon le test du COSINOR qui permet d'évaluer la rythmicité des données selon l'équation $y = A + (B \times \cos (2\pi (x-C)/24))$ avec A le MESOR, B l'amplitude et C l'acrophase et selon une ANOVA à deux facteurs suivi d'un test post-hoc de Tukey. Les niveaux de significativité retenus étaient $p\text{-value} < 0,05$. L'oscillation rythmique d'une protéine a été validée que lorsque la

significativité ($p < 0,05$) était vérifiée simultanément pour les 2 paramètres que sont l'Amplitude et l'Acrophase.

B.4. Analyse des protéines contenues dans les spots de gels DIGE

B.4.1. Réduction-Alkylation-digestion

Les spots protéiques différentiels ont été excisés manuellement à partir des gels DIGE (pool de 4 spots). Les spots de gel ont été lavés et les protéines contenues dans ces spots réduites puis alkylées à l'aide d'une station « Massprep Station » (Waters). Ce traitement combine une étape de lavage (3 cycles de lavage avec 50 μL de NH_4HCO_3 25 mM et 50 μL d'acétonitrile, 10 minutes), une étape de déshydratation (50 μL d'acétonitrile, 60°C, 5 minutes), une étape de réduction des ponts disulfures (50 μL de 10mM DTT, 25mM NH_4HCO_3 , 60°C, 30 minutes), une étape d'alkylation des protéines (50 μL de 55mM Iodoacétamide NH_4HCO_3 25mM, 30 minutes), une nouvelle étape de lavage (1 cycle 50 μL de NH_4HCO_3 à 25 mM/50 μL d'acétonitrile, 10 minutes) et une étape de déshydratation du gel (50 μL d'acétonitrile, 15 minutes).

Les protéines ont ensuite été digérées enzymatiquement par la trypsine (Promega). Les spots de gel déshydratés ont été réhydratés avec 8 μL et 20 μL d'une solution de trypsine (6,7 ng/L, NH_4HCO_3 25mM) pour les spots de gels DIGE et coloré au bleu de Coumassie respectivement. La digestion enzymatique a duré 12h à 37°C. Les peptides ont été extraits des morceaux de gels par ajout de 30 μL d'une solution à 60% acétonitrile, 0,1% HCOOH. Les surnageants ont été prélevés et l'acétonitrile a été évaporé sous vide (Speed vac, Thermo Scientific).

B.4.2. Analyses nanoLC-MS/MS (maXis)

Après digestion in-gel des protéines contenues sur les gels DIGE, les peptides tryptiques ont été analysés sur un système UPLC (nanoAcquity, Waters) couplé à un spectromètre de masse combinant une source électrospray et un analyseur de type temps de vol (maXis ; Bruker Daltonique). Le couplage nanoHPLC-MS/MS était contrôlé respectivement par les logiciels Bruker compass Hystar (v 3.2) et OtofControl (Rev 3.4, Bruker Daltonik GmbH).

Le système UPLC utilisé combinait une colonne d'enrichissement C18, 180 μm x 20mm, 5 μm , une colonne de séparation C18, 250mm x 75 μm , 1,7 μm . Le volume injecté était de 4 μL , le débit de chargement sur la précolonne de 5 $\mu\text{L}/\text{min}$ (0,1% HCOOH, H_2O). Les peptides ont été élués selon le

gradient : t=0 min 1% B, t= 9 min 35% B, t=10 min, 90%B, t=11 min 1% B avec A : H₂O, 0,1% HCOOH et B: Acétonitrile, 0,1% HCOOH (450 nL/min).

Les paramètres de sélection des peptides afin de les isoler puis de les fragmenter (mode CID – Collision Induced Dissociation) sont présentés dans le tableau 6.

Tableau 6 : Paramétrage des séquences d’acquisition en spectrométrie de masse

MS	Gamme de masse	100-2500 m/z
	Nombre d’ions sélectionnés	5 ions
Fragmentation CID	Etats de charge	2x-5x
MS/MS	Cycle d’exclusion	2 spectres puis 1 min d’exclusion
	Gamme de masse	100-2500 m/z

B.4.3. Recherche dans les banques de données protéiques

L’algorithme de recherche Mascot v 2.4.1 (Matrix Science, London, UK) a été utilisé pour l’identification des protéines à partir des données MS et MS/MS des spots découpés sur les gels DIGE à l’issue des tests. Ces identifications ont été également validées par un second moteur de recherche : Omssa v2.1.9, Open Mass Spectrometry Search Algorithm. La recherche a été réalisée dans une banque de donnée restreinte à la taxonomie « mus musculus » extraite à partir de la base de donnée UniprotKB/Swissprot (téléchargement en juin 2013 ; 16731 entrées).

La tolérance accordée sur les erreurs de masse en mode MS était de 15 ppm et 0,02 Daltons pour les MS/MS pour le moteur de recherche Mascot et Omssa. La tolérance sur le nombre de site de clivage manqué par l’enzyme protéolytique était de 1. La carbamidomethylation des cystéines, l’oxydation des méthionines et l’acétylation N-terminale des protéines ont été spécifiées comme modifications variables.

Afin de ne considérer que les spectres de qualité, des critères de filtres stricts sur les scores d’identifications des ions ont été considérés à l’aide du logiciel Scaffold (V 3.6.5), Proteome software Inc.). L’identification des protéines était validée à partir d’un seul peptide unique à condition de remplir les critères suivants :

- Ion score – identity score > 12 ; ion score > 25 ; taux de faux positifs < 1% pour les résultats provenant du moteur de recherche Mascot
- $-\log(\text{E-value}) > 8,1$; taux de faux positifs < 1% pour les résultats provenant du moteur de recherche Omssa

Une stratégie target-decoy a été employée pour estimer le taux de faux positifs parmi les identifications. Les contaminants usuels, comme la kératine et la trypsine, ont été éliminés des données. Les kératines peuvent également provenir d'une contamination de l'échantillon lors du prélèvement de celui-ci, les kératines dont les points isoélectriques et poids moléculaires étaient cohérents avec les positions sur les gels d'électrophorèse où les kératines ont été identifiées n'ont pas été éliminées des données car elles peuvent interférer dans les résultats (=être responsable de la variation d'intensité du spot).

C. Etude des mécanismes d'adaptation de l'ours brun à l'inactivité

C.1. Informations sur les échantillons

Muscle Vastus Lateralis = VL / avant (juin-été) et après (février-hiver) hibernation.

Remarque 1 : Chaque individu est son propre contrôle (ex : VL1h = même individu que VL1e)

Echantillons : (n = 6 / group):

Group VL avant hibernation : W1104 E, W1209E, W1110E, W1204E, W1105E, W1207E et W1202E

Group VL après hibernation : W1207F, W1110F, W1204F, W1209F, W1105F, W1202F, W1104F

Les échantillons ont été conservés à -80°C.

C.2. Traitement des échantillons

C.2.1. Extraction des protéines

Les échantillons de muscle congelés ont été broyés à l'aide d'un broyeur à froid et à sec de type Mikro-Dismembrator (Sartorius ; une minute à 3000 tours/minutes). Environ dix milligrammes de chaque échantillon réduit en poudre ont été repris dans 300-600 µL d'une solution à base d'urée (7M), thio-urée (2M), CHAPS 4%, dithiothréitol (DTT) 1%, TLCK (inhibiteur de trypsine) 0,005% et inhibiteurs de protéases (cocktail commercial dilué 1 :50, P8340, Sigma). Les échantillons ont été homogénéisés par agitation (2 heures) et sonication (10 secondes, 135 Watts).

Les protéines ont été précipitées par ajout de 5 volumes d'acétone froid. Après une nuit à -20°C, les échantillons ont été centrifugés (14 minutes, +4 °C, 14000xg). Les surnageants ont été éliminés et les culots de protéines séchés par évaporation sous vide (Speed-vac, Thermo Scientific). Chaque culot a été repris dans une solution à base d'urée (7M), thio-urée (2M), Tris pH 8.5 (30mM) et CHAPS 4%. Le pH de chaque échantillon a été ajusté à pH 8,5 (tris 1,5 M pH 8,8). Après homogénéisation par agitation (vortex) et sonication (10 secondes, 135 watts) tout en maintenant les tubes dans la glace, les échantillons ont été centrifugés afin d'éliminer les éventuels débris cellulaires.

C.2.2. Dosage de Bradford et vérification par gel 1D

La concentration des protéines dans les échantillons a été mesurée selon la méthode de Bradford (Protein Assay Dye Reagent Concentrate, ref 500-0006, Bio-Rad). La gamme d'étalonnage a été réalisée à partir d'albumine de sérum bovin (0- 10 μ g). Le dosage de Bradford réalisé a été validée par réalisation d'un gel SDS-Page (10x10 cm): les protéines de chaque échantillon ont été déposées séparément en équiquantité sur le gel (10 μ g de protéines de chaque échantillon repris dans une solution tampon à 10 mM Tris pH 8, EDTA 1mM, β ME 5%, SDS 5%, glycérol 10% et bleu de bromophénol 0,10%). Le gel 1D était composé d'un gel de « concentration » (permettant la concentration des échantillons en début de gel) et d'un gel de « séparation » tous deux à base de polyacrylamide réticulés à 5% et 12%, respectivement (0,37 M Tris pH 8,8, SDS 0,1%, Temed 0,1% et Persulfate NH₄ 0,025% pour le gel de séparation et (0,15 M Tris pH 6,8, SDS 0,1%, Temed 0,025% et Persulfate NH₄ 0,1% pour le gel de concentration. Les conditions électrophorétiques ont été fixées à 50 V pendant 90 minutes puis le voltage a été augmenté à 100V pendant environ 1h. Le gel a été agité dans une solution de fixation pendant 30 minutes (3% H₃PO₄, 50% CH₃CH₂OH) puis laissé pendant une nuit dans une solution de coloration à base de bleu de Coumassie (0,1% Coumassie R250, 10% CH₃COOH, 40% CH₃OH). Le dosage était considéré comme valide si l'intensité des bandes entre les échantillons était similaire.

C.2.3. Marquage des protéines pour analyse 2D-DIGE

Echantillons : Les cyanines Cy2, Cy3 et Cy5 (CyDyes, Amersham Biosciences, GE Healthcare), anhydres, ont été reconstituées dans du diméthylformamide anhydre. Le marquage se fait par ajout de 400 picomoles de CyDyes à 50 μ g de protéines (reprises dans 20 μ L de solution à base d'urée (7M), thio-urée (2M), CHAPS 4%, et 30mM Tris). Après centrifugation, les échantillons ont été incubés à l'obscurité pendant 30 minutes dans la glace. 1 μ L de solution stop de lysine (10mM) a alors été ajouté à chaque échantillon, puis les échantillons ont été incubés une seconde fois à l'obscurité pendant 10 minutes dans la glace.

Standard interne : 25 μ g de protéines de chaque échantillon ont été mélangés. Le mélange a ensuite été repris dans 120 μ L de solution à base d'urée (7M), thio-urée (2M), CHAPS 4%, et 30mM Tris. Le marquage a ensuite été réalisé avec 2,4 ng de CyDyes selon le procédé énoncé précédemment.

C.2.4. Electrophorèse bidimensionnelle (E2D)

Deux échantillons marqués (soit 2 x 50 µg de protéines) ainsi que 50 µg de protéines du standard interne marqué également ont été mélangés avant la première étape d'isoélectrofocalisation selon la répartition du tableau 7 présenté ci-dessous.

Tableau 7 : Répartition des marqueurs Cy2, Cy3 et Cy5 en fonction des échantillons et des gels

N° GEL	Cy2	Cy3	Cy5	N° GEL	Cy2	Cy3	Cy5
1	SI	1204E	1204F	4	SI	1202F	1202E
2	SI	1209F	1209E	5	SI	1207E	1207F
3	SI	1110E	1110F	6	SI	1105F	1105 ^E
7	SI	1104E	1104F	-	-	-	-

SI : standard interne

Le mélange de 150 µg de protéines a été repris dans un volume de 530 µL d'une solution d'urée 7M, thio-urée 2M, 2% CHAPS, 2% DTT, 2% Ampholytes (servalytes 3-10) et quelques gouttes de bleu de bromophénol. Le tout a été déposé sur une bande de gel à gradient de pH immobilisé (ReadyStrip IPG, Biorad) de 24 cm pH 3-10 non linéaire (Ready IPG strips, Bio-Rad). La réhydratation passive a duré 1h. Chaque bande de gel a été recouverte de 2mL d'huile végétale. La réhydratation active a ensuite été réalisée par l'application d'un courant de 50 V pendant une nuit. Après réhydratation, un gradient de tension a été appliqué (0-200V/1h linéaire ; 200-1000 V/ 4h linéaire ; 1000-5000 V/ 16h linéaire, 5000V/ 7h). La dernière étape a été rallongée de façon à atteindre 90 000 VH. L'isoélectrofocalisation a duré 18 heures (Protean ief cell, Biorad). Les strips ont ensuite été congelés à -80°C.

La migration des gels 2D a été réalisée à l'horizontal selon le protocole développé par le constructeur de la tour HPE (HPE, FlatTop Tower, Serva). Les gels utilisés étaient des gels de polyacrylamide réticulés à 12,5%. Après décongélation, les strips ont été incubés à l'obscurité pendant 15 minutes dans le tampon d'équilibration du kit (Serva) avec 1% DTT puis 15 minutes dans le tampon d'équilibration (Serva) avec 2,5% d'iodoacétamide. Les strips ont ensuite été déposés sur les gels précoulés. Les conditions électrophorétiques suivies sont présentées en tableau 8.

Tableau 8 : Conditions électrophorétiques suivies pour la migration 2D

Etapes	Max V	Max mA	Max W	Temps
1	100	7	1	30 min
2	200	13	3	30 min
3	300	20	5	10 min
4	1000	40	30	3h50
5	1500	45	40	40 min

Les gels ont ensuite été fixés au moins une heure. Les scans des gels ont été réalisés à partir d'un Ettan DIGE Imager (**Cy2** longueur d'onde d'absorption 491 ± 3 nm, longueur d'onde d'émission 506 ± 5 nm; **Cy3** longueur d'onde d'absorption 553 ± 3 nm, longueur d'onde d'émission 572 ± 5 nm; **Cy5** longueur d'onde d'absorption 648 ± 3 nm, longueur d'onde d'émission 669 ± 5 nm).

C.2.5. Analyse d'image des gels 2D-DIGE (Nonlinear Dynamics, Samespots)

L'analyse d'image à partir des gels DIGE a été réalisée grâce au logiciel « Progenesis Samespots » (v4.5, Nonlinear Dynamics). Les images de gel ont été alignées les unes par rapport aux autres par le logiciel de façon automatique. Un contrôle visuel a été réalisé sur cet alignement et des corrections éventuelles ont été apportées manuellement. Les « volumes densitométriques » des spots contenant des protéines marquées par les fluorochromes Cy2, Cy3 et Cy5 ont été normalisés par rapport au « volume » de l'ensemble des spots contenus sur chaque gel, puis les valeurs Cy3 ou Cy5 ont été normalisée par rapport aux valeurs de standard interne.

D.2. Statistiques

Les analyses statistiques ont été réalisées à l'aide du logiciel « R » (v2.8.1; <http://www.R-project.org>).

Les échantillons analysés étaient appariés (type avant-après hibernation) et nous avons $n=7$ /groupe. Les différences obtenues entre les groupes ont donc été testées par le test de Student apparié (two-sided paired t-test). La normalité des valeurs ainsi que l'égalité des variances ont été vérifiées au préalable par les tests de Shapiro-Wilk et Bartlett.

Le niveau de significativité que nous avons retenu était $p < 0,05$ pour le test de Student apparié et $p > 0,01$ pour les tests de Bartlett et Shapiro-Wilk.

E.2. Analyse des protéines contenues dans les spots de gels DIGE

E.2.1. Réduction-Alkylation-digestion

Les spots protéiques ont été excisés manuellement puis rassemblés à partir des gels 1, 2, 3 et 4.

Les spots de gel ont été lavés et les protéines contenues dans ces spots réduites puis alkylées à l'aide d'une station « Massprep Station » (Waters). Ce traitement combine une étape de lavage (3 cycles de lavage avec 50 μL de NH_4HCO_3 25 mM et 50 μL d'acétonitrile, 10 minutes), une étape de déshydratation (50 μL d'acétonitrile, 60°C, 5 minutes), une étape de réduction des ponts disulfures (50 μL de 10 mM DTT, 25 mM NH_4HCO_3 , 60°C, 30 minutes), une étape d'alkylation des protéines (50 μL de 55 mM Iodoacétamide NH_4HCO_3 25 mM, 30 minutes), une nouvelle étape de lavage (1 cycle 50 μL de NH_4HCO_3 à 25 mM/50 μL d'acétonitrile, 10 minutes) et une étape de déshydratation du gel (50 μL d'acétonitrile, 15 minutes).

Les protéines ont ensuite été digérées enzymatiquement par la trypsine (Promega). Les 4 spots de gel rassemblés déshydratés ont été réhydratés avec 5 μL d'une solution de trypsine (12,5ng/L, NH_4HCO_3 25 mM) + 20 μL de tampon NH_4HCO_3 25mM. La digestion enzymatique a duré 12h à 37°C. Les peptides ont été extraits des morceaux de gels par ajout de 30 μL d'une solution à 60% acétonitrile, 0,1% HCOOH . Les surnageants ont été prélevés et l'acétonitrile a été évaporé sous vide (Speed vac, Thermo Scientific).

E.2.2. Analyses nanoLC-MS/MS (maXis)

Après digestion in-gel des protéines contenues sur les gels DIGE, les peptides tryptiques ont été analysés sur un système UPLC (nanoAcquity, Waters) couplé à un spectromètre de masse combinant une source électrospray et un analyseur de type temps de vol (maXis ; Bruker Daltonique). Le couplage nanoHPLC-MS/MS était contrôlé respectivement par les logiciels Bruker compass Hystar (v 3.2) et OtofControl (Rev 3.4, Bruker Daltonik GmbH).

Le système UPLC utilisé combinait une colonne d'enrichissement C18, 180 μm x 20mm, 5 μm , une colonne de séparation C18, 250mm x 75 μm , 1,7 μm . Le volume injecté était de 4 μL , le débit de

chargement sur la précolonne de 5 $\mu\text{L}/\text{min}$ (0,1% HCOOH , H_2O). Les peptides ont été élués selon le gradient : $t=0$ min 1% B, $t=9$ min 35% B, $t=10$ min, 90%B, $t=11$ min 1% B avec A : H_2O , 0,1% HCOOH et B: Acétonitrile, 0,1% HCOOH (450 nL/min).

Les paramètres de sélection des peptides afin de les isoler puis de les fragmenter (mode CID – Collision induced dissociation) sont présentés dans le tableau 9.

Tableau 9: Paramétrage des séquences d’acquisition en spectrométrie de masse

MS	Gamme de masse	100-2200 m/z
	Nombre d’ions sélectionnés	6 ions
Fragmentation CID	Etats de charge	2x-5x
MS/MS	Cycle d’exclusion	1 spectres puis 1 min d’exclusion
	Gamme de masse	100-2200 m/z

E.2.3. Recherche dans les banques de données protéiques

L’algorithme de recherche Mascot v2.3.02 (Matrix Science, London, UK) a été utilisé pour l’identification des protéines à partir des données MS et MS/MS des spots découpés sur les gels DIGE à l’issue des tests. Ces identifications ont été également validées par un second moteur de recherche : Omssa v2.1.7, Open Mass Spectrometry Search Algorithm. La recherche a été réalisée dans une banque de donnée restreinte à la taxonomie « mammalia » (mammifères, Taxon ID 40674) extraite à partir de la base de donnée UniprotKB/Swissprot (téléchargement en décembre 2012 ; 66243 entrées).

La tolérance accordée sur les erreurs de masse en mode MS était de 15 ppm et 0,02 Daltons pour les MS/MS pour le moteur de recherche Mascot et Omssa. La tolérance sur le nombre de site de clivage manqué par l’enzyme protéolytique était de 1. La carbamidomethylation des cystéines, l’oxydation des méthionines et l’acétylation N-terminale des protéines ont été spécifiées comme modifications variables.

Afin de ne considérer que les spectres de qualité, des critères de filtres stricts sur les scores d’identifications des ions ont été considérés à l’aide du logiciel Scaffold (v_3_00_07, Proteome

software Inc.). L'identification des protéines était validée à partir d'un seul peptide unique à condition de remplir les critères suivants :

- Ion score – identity score > 7 pour les résultats provenant du moteur de recherche Mascot
- $-\log(\text{E-value}) > 7$ pour les résultats provenant du moteur de recherche Omssa

Une stratégie target-decoy a été employée pour estimer le taux de faux positifs parmi les identifications. Les contaminants usuels, comme la kératine et la trypsine, ont été éliminés des données. Les kératines peuvent également provenir d'une contamination de l'échantillon lors du prélèvement de celui-ci, les kératines dont les points isoélectriques et poids moléculaires étaient cohérents avec les positions sur les gels d'électrophorèse où les kératines ont été identifiées n'ont pas été éliminées des données car elles peuvent interférer dans les résultats (=être responsable de la variation d'intensité du spot).

Annexe 1 : Modèle de normalisation proposés par le logiciel Normalyzer (1.0)

[127]

Global : In global methods, normalization is done across all samples without considering any replicate grouping. The basic assumption for any global normalization is that most of the measured variables (genes/peptides/proteins) are not differentially expressed across conditions. This can be confirmed with the DE plots of Log2 transformed data at the end of the Normalyzer report. This hypothesis is generally true for most high throughput datasets. In case the assumption doesn't stand, local normalization can be optionally done.

Local : Local normalization is only done within the replicate groups without any overall/global normalization. This could be useful in cases where global normalization is not possible and when too few variables are measured.

Log 2 : Data is log2 transformed.

LOESS : Log2 transformed data is normalized by LOESS method using the function “normalizeCyclicLoess » For the global normalization, each sample is normalized to the average of all samples, whereas in the local method, each sample within the replicate group is normalized to the average of that replicate group.

Robust Linear Regression (RLR) : Log2 transformed data is normalized by robust linear regression (“rlm” function). For the method, each sample is normalized to the median of all samples, while in the local method, each sample in a replicate group is normalized to the median of samples in the replicate group.

Variance Stabilization Normalization (VSN) : Log2 transformed data is normalized using the function “justvsn” from the vsn package. Global and local normalization are done as described in the RLR section above.

Total Intensity (TIG) : Intensity of each variable in a given sample is divided by the sum of intensities of all variables in the sample and then multiplied by the median of ‘sum of intensities of all variables in all samples’. The normalized data is then transformed to Log2 Median Intensity (MedIG)

Intensity of each variable in a given sample is divided by the median of intensities of all variables in the sample and then multiplied by the mean of 'median of sum of intensities of all variables in all samples'. The normalized data is then transformed to Log2.

Average Intensity (AIG) : Intensity of each variable in a given sample is divided by the mean of sum of intensities of all variables in the sample and then multiplied by the mean of 'mean of sum of intensities of all variables in all samples'. The normalized data is then transformed to Log2.

QuantileG : Quantile normalization is done by the function "normalize.quantiles" from the package preprocessCore.

NormFinder(NFG) : NormFinder normalization is done using the R script obtained from the developers of NormFinder algorithm. The R script was modified for incorporation in the Normalyzer pipeline. Normalization by NormFinder is a global normalization method which can be used even when few variables are measured as long as control variables are included in the assay. In this method, two variables with least local and global variance are selected by NormFinder. The intensity of each variable in a sample is divided by the mean of intensities of the two NormFinder identified control variables from the corresponding samples and multiplied by the mean of all intensities of the control variables. The data is then log2 transformed.

Annexe 2 : Liste des protéines identifiées comme différentielles concernant l'étude des cellules HepaRG en réponse à une exposition à l'acide valproïque

(tests de Shapiro-Wilk ($p > 0,01$) ; Bartlett ($p > 0,01$), ANOVA ($p < 0,05$) et Post Hoc de Tukey ($p < 0,01$))

La couleur rose indique que l'abondance relative des protéines entre les deux conditions comparées était supérieure ou égale à 1,5.

Nom des protéines	Numéro d'accession Uniprot	D2_ctrl-D0_ctrl	D8_ctrl-D0_ctrl	D8_ctrl-D2_ctrl	D14_ctrl-D0_ctrl	D2_0.05mM-D2_ctrl	D8_0.05mM-D8_ctrl	D14_0.05mM-D14_ctrl	D2_0.15mM-D2_ctrl	D8_0.15mM-D8_ctrl	D14_0.15mM-D14_ctrl	D2_0.5mM-D2_ctrl	D8_0.5mM-D8_ctrl	D14_0.5mM-D14_ctrl
Fucose mutarotase	A2VDF0	3.56E-01	2.64E-02	9.78E-01	2.70E-03	9.88E-01	1.00E+00	1.00E+00	1.99E-01	1.00E+00	8.88E-01	2.39E-01	2.21E-01	9.26E-01
Deoxyribonuclease-2-alpha	O00115	9.99E-01	2.72E-01	7.90E-01	3.63E-02	1.00E+00	7.47E-01	4.08E-01	1.00E+00	8.02E-01	8.91E-01	1.00E+00	3.11E-01	0.00E+00
Membrane-associated progesterone receptor component 1	O00264	1.00E+00	9.72E-01	6.75E-01	4.53E-01	1.00E+00	8.19E-01	1.00E+00	1.00E+00	6.13E-01	1.00E+00	9.98E-01	3.32E-01	5.00E-04
Chloride intracellular channel protein 1	O00299	1.37E-01	0.00E+00	1.20E-03	0.00E+00	9.88E-01	4.71E-01	1.00E+00	1.00E+00	1.36E-01	9.98E-01	9.05E-01	1.00E+00	1.44E-01
Procollagen-lysine,2-oxoglutarate 5-hydroxylase	O00469	8.55E-01	6.00E-04	4.46E-02	1.00E-04	1.00E+00	9.61E-01	1.00E+00	1.00E+00	4.77E-01	7.54E-01	9.96E-01	1.44E-01	1.00E+00
Myosin regulatory light chain 12B	O14950	8.57E-01	9.00E-03	3.46E-01	3.00E-04	1.00E+00	1.00E+00	9.99E-01	1.00E+00	1.00E+00	5.44E-01	1.00E+00	1.00E+00	7.92E-01
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	P49748	0.00E+00	0.00E+00	5.26E-01	0.00E+00	9.99E-01	0.00E+00	0.00E+00	1.00E+00	0.00E+00	0.00E+00	6.89E-01	0.00E+00	0.00E+00
UDP-glucose 6-dehydrogenase	O60701	5.71E-01	1.00E-04	1.83E-02	0.00E+00	1.00E+00	9.76E-01	5.70E-01	1.00E+00	2.61E-01	1.55E-01	9.91E-01	1.30E-03	0.00E+00
Peroxisomal 2,4-dienoyl-CoA reductase	Q9NUI1	0.00E+00	0.00E+00	1.00E+00	0.00E+00	5.77E-01	9.92E-01	9.97E-01	1.70E-03	1.00E+00	2.95E-01	1.00E-04	1.30E-03	5.00E-04
SH3 domain-binding glutamic acid-rich-like protein CREG1	O75368	1.00E+00	8.99E-01	6.01E-01	9.75E-02	1.00E+00	8.11E-01	9.03E-01	1.00E+00	9.92E-01	2.92E-01	1.00E+00	7.30E-02	5.00E-04
Protein CREG1	O75629	9.93E-01	1.00E+00	8.63E-01	1.00E+00	1.00E+00	1.00E+00	9.01E-01	1.00E+00	6.68E-01	8.61E-01	1.00E+00	1.97E-01	1.30E-03
Cytosolic 10-formyltetrahydrofolate synthetase subunit g, mitochondrial	O75891	1.00E+00	0.00E+00	0.00E+00	0.00E+00	9.98E-01	3.17E-01	9.59E-01	8.19E-01	8.90E-03	2.00E-04	2.51E-02	0.00E+00	0.00E+00
ATP synthase subunit g, mitochondrial	O75964	7.59E-01	1.00E+00	9.83E-01	9.99E-01	1.00E+00	1.00E+00	9.98E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.27E-01	7.00E-04
Aflatoxin B1 aldehyde reductase member 3	O95154	1.00E+00	1.00E-04	1.00E-04	0.00E+00	4.27E-01	1.00E+00	9.91E-01	1.00E+00	1.00E+00	4.04E-02	1.00E+00	8.80E-03	0.00E+00
Adenylyl-sulfate kinase	O95340	9.43E-02	1.63E-02	1.00E+00	2.70E-03	9.21E-01	9.82E-01	9.91E-01	3.78E-01	2.67E-01	1.58E-01	6.60E-03	7.00E-04	0.00E+00
6-phosphogluconolactonase	O95479	8.41E-01	1.19E-01	1.90E-03	6.00E-02	1.00E+00	8.57E-01	1.00E+00	1.00E+00	7.30E-02	1.00E-04	2.88E-02	9.00E-04	0.00E+00
Fructose-bisphosphate aldolase B	P05062	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00	6.35E-01	9.26E-01	2.91E-01	0.00E+00	0.00E+00	1.00E-04	0.00E+00	0.00E+00
Apoptosis-inducing factor 1, mitochondrial	O95831	6.73E-01	1.00E-03	1.35E-01	7.00E-04	1.00E+00	1.00E+00	1.42E-01	9.80E-01	6.05E-01	0.00E+00	1.53E-01	4.00E-04	0.00E+00
N(G),N(G)-dimethylarginine dimethylaminohydrolase	O95865	9.57E-01	2.55E-02	4.20E-01	4.10E-03	1.00E+00	1.00E+00	9.50E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Ceruloplasmin	P00450	1.00E+00	2.30E-03	4.20E-03	3.90E-03	9.91E-01	1.00E+00	1.00E+00	1.00E+00	9.29E-01	1.32E-01	9.94E-01	5.54E-01	1.32E-02
Ornithine carbamoyltransferase, mitochondrial	P00480	9.99E-01	1.76E-01	2.39E-02	9.00E-04	1.00E+00	1.00E+00	1.14E-01	1.00E+00	9.99E-01	9.32E-01	8.95E-01	1.00E+00	1.22E-01
Complement factor B	P00751	9.94E-01	7.26E-01	1.46E-01	1.00E+00	5.35E-01	9.83E-01	9.99E-01	9.20E-01	1.00E+00	9.23E-01	1.34E-01	4.72E-01	0.00E+00
Alpha-1-antitrypsin	P01009	1.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00	5.84E-01	8.26E-01	8.88E-02	1.10E-03	6.60E-02	2.99E-02	0.00E+00	0.00E+00
Alpha-1-antichymotrypsin	P01011	9.52E-01	1.20E-03	4.21E-02	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	7.00E-01	8.97E-01	1.00E+00	1.90E-03
Complement C3	P01024	5.50E-01	9.96E-01	9.37E-02	8.66E-01	4.45E-01	3.99E-01	9.84E-01	7.97E-01	2.63E-01	6.35E-02	1.68E-01	1.40E-03	0.00E+00
Fibrinogen alpha chain	P02671	1.00E+00	0.00E+00	0.00E+00	0.00E+00	9.35E-01	7.50E-01	1.00E+00	1.00E+00	4.34E-01	1.50E-01	3.50E-01	3.00E-04	0.00E+00
Fibrinogen gamma chain	P02679	8.37E-01	1.00E-03	7.00E-02	1.32E-01	9.99E-01	9.91E-01	1.00E+00	1.00E+00	1.00E+00	4.63E-01	1.00E+00	2.12E-01	1.00E-04
Serum amyloid P-component	P02743	1.71E-01	2.60E-03	8.19E-01	2.30E-03	1.00E+00	1.00E+00	1.00E+00	1.00E+00	7.76E-01	8.62E-01	1.00E+00	4.15E-01	1.00E+00
Alpha-1-acid glycoprotein 1	P02763	1.00E+00	1.00E-04	5.00E-04	6.52E-02	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.18E-01	9.86E-01	2.08E-02	0.00E+00
Vitamin D-binding protein	P02774	1.00E+00	0.00E+00	1.00E-04	1.20E-03	5.82E-01	8.01E-01	1.00E+00	8.72E-01	6.94E-01	3.64E-02	5.97E-02	4.40E-03	2.00E-04
Serotransferrin	P02787	9.72E-01	9.99E-01	1.00E+00	4.00E-04	8.93E-01	7.87E-01	8.58E-01	1.00E+00	9.98E-01	2.73E-02	8.35E-01	2.67E-02	0.00E+00
Vitronectin	P04004	1.00E+00	2.21E-01	3.40E-01	3.08E-02	9.72E-01	1.00E+00	1.00E+00	1.00E+00	9.88E-01	6.75E-02	1.00E+00	1.46E-01	0.00E+00
Lupus La protein	P05455	9.99E-01	1.79E-01	6.42E-01	6.00E-04	9.54E-01	9.82E-01	3.80E-01	1.00E+00	1.00E+00	2.16E-01	9.17E-01	8.35E-01	1.60E-03
Integrin beta-1	P05556	9.92E-02	0.00E+00	2.50E-03	0.00E+00	9.98E-01	1.00E+00	4.99E-02	1.00E+00	1.00E+00	5.16E-02	4.06E-01	5.00E-04	0.00E+00
Glucose-6-phosphate isomerase	P06744	3.81E-01	1.00E+00	3.04E-01	1.00E+00	1.00E+00	9.92E-01	9.73E-01	4.00E-01	1.00E+00	1.20E-03	1.00E+00	9.98E-01	0.00E+00
Annexin A2	P07355	6.00E-01	0.00E+00	1.10E-02	0.00E+00	9.99E-01	9.50E-01	9.54E-01	1.00E+00	1.00E+00	1.69E-01	1.00E+00	9.88E-01	1.00E+00
Tubulin beta chain	P07437	7.77E-01	0.00E+00	4.50E-03	1.19E-02	9.92E-01	4.40E-02	1.00E+00	4.28E-01	4.37E-01	1.00E+00	5.16E-01	6.70E-03	1.00E+00
Cathepsin L1	P07711	5.72E-01	6.00E-04	1.22E-01	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.69E-01	1.00E+00	7.68E-01	9.99E-01	1.00E+00
Galactose-1-phosphate uridylyltransferase	P07902	1.00E+00	8.10E-02	3.22E-01	1.01E-02	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	5.01E-01	1.00E+00	6.91E-01	9.50E-03
Methylcrotonyl-CoA carboxylase subunit alpha, mitochondrial	Q96RQ3	4.00E-04	0.00E+00	1.00E+00	0.00E+00	9.66E-01	9.98E-01	9.50E-01	7.90E-01	9.97E-01	9.87E-01	9.91E-01	2.60E-01	1.00E+00
Alcohol dehydrogenase 4	P08319	1.00E+00	1.00E+00	1.00E+00	7.71E-01	1.00E+00	1.00E+00	3.51E-01	9.56E-01	6.70E-01	1.20E-01	1.00E+00	2.80E-03	6.18E-02
Integrin alpha-5	P08648	7.51E-01	3.20E-03	2.45E-01	2.80E-03	1.00E+00	1.00E+00	2.22E-01	9.06E-01	9.89E-01	1.00E+00	1.00E+00	9.99E-01	1.12E-02
Cytoskeleton-associated protein 4	Q07065	8.00E-04	8.20E-01	6.73E-02	9.96E-01	3.40E-01	9.34E-01	5.50E-02	6.07E-02	4.43E-01	1.00E-04	2.60E-03	0.00E+00	0.00E+00
Poly [ADP-ribose] polymerase 1	P09874	3.06E-01	0.00E+00	0.00E+00	0.00E+00	1.00E+00	7.76E-01	4.30E-03	1.00E+00	8.44E-02	1.56E-02	8.98E-01	1.80E-03	4.40E-03
Cytochrome c oxidase subunit SB, mitochondrial	P10606	5.93E-01	6.46E-01	1.00E+00	4.80E-03	1.00E+00	9.97E-01	9.99E-01	9.98E-01	1.00E+00	1.00E+00	4.34E-01	2.76E-01	0.00E+00
Microtubule-associated protein tau	P10636	1.00E+00	1.00E+00	9.60E-01	6.14E-02	1.00E+00	1.00E+00	2.54E-02	1.00E+00	9.94E-01	1.03E-02	9.99E-01	7.31E-02	0.00E+00
Clusterin	P10909	9.98E-01	9.63E-02	5.14E-01	3.34E-01	9.88E-01	1.00E+00	8.99E-01	1.00E+00	1.00E+00	2.34E-01	9.51E-01	9.99E-01	0.00E+00
Lysosome-associated membrane glycoprotein 1	P11279	1.00E+00	9.20E-01	9.99E-01	8.25E-01	1.00E+00	9.94E-01	1.00E+00	8.02E-01	9.83E-01	6.29E-01	3.12E-01	4.00E-03	4.00E-04
Glucose-6-phosphate 1-dehydrogenase	P11413	1.00E+00	0.00E+00	1.00E-04	0.00E+00	1.00E+00	9.93E-01	4.61E-01	1.00E+00	2.03E-01	3.33E-01	9.88E-01	1.00E-04	5.00E-04

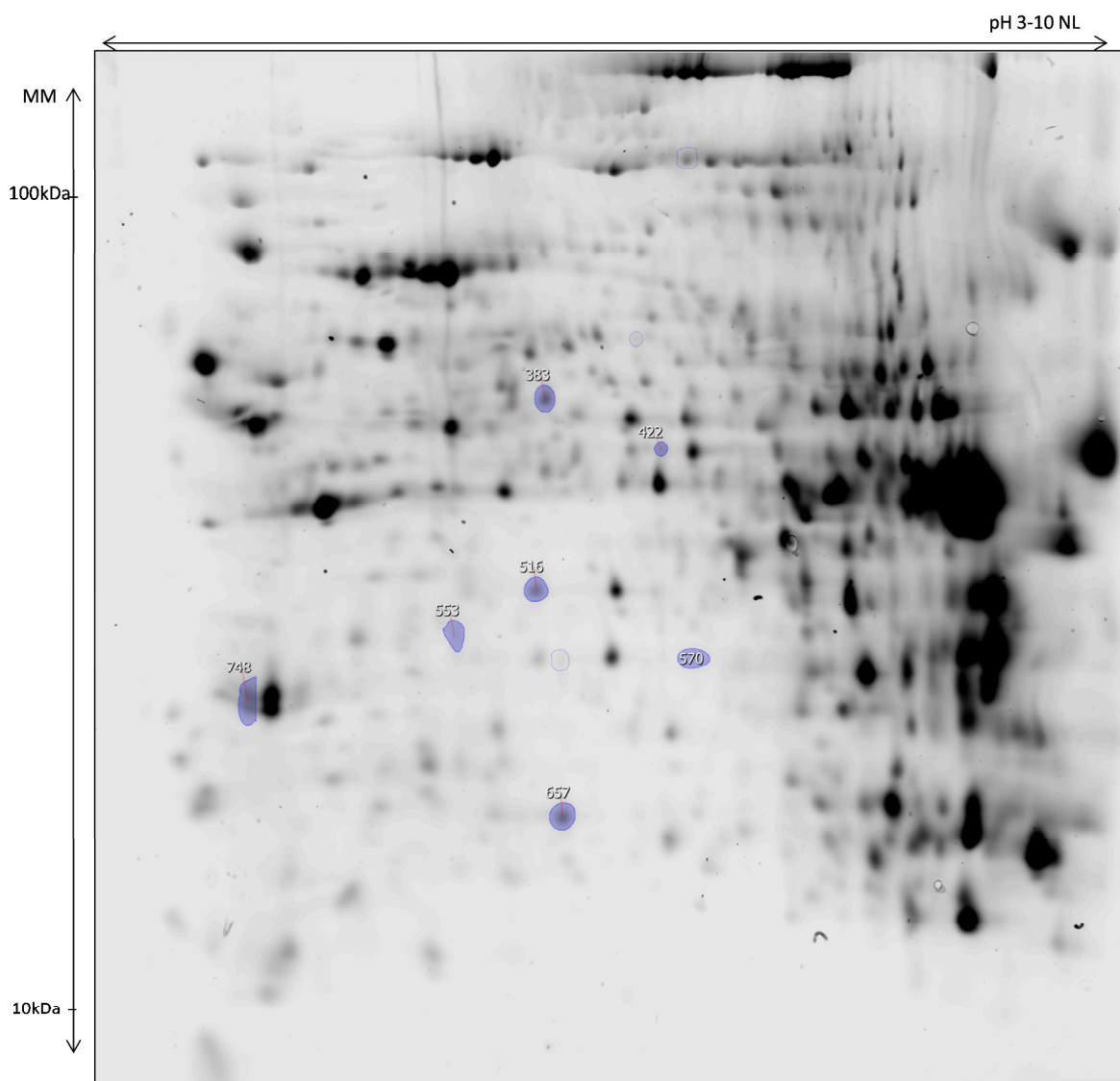
Nom des protéines	Numéro d'accension Uniprot	D2_ctrl-D0_ctrl	D8_ctrl-D0_ctrl	D8_ctrl-D2_ctrl	D14_ctrl-D0_ctrl	D2_0.05mM-D2_ctrl	D8_0.05mM-D8_ctrl	D14_0.05mM-D14_ctrl	D2_0.15mM-D2_ctrl	D8_0.15mM-D8_ctrl	D14_0.15mM-D14_ctrl	D2_0.5mM-D2_ctrl	D8_0.5mM-D8_ctrl	D14_0.5mM-D14_ctrl
Phosphate carrier protein, mitochondrial	Q00325	9.00E-04	6.91E-01	1.20E-01	1.00E+00	2.85E-02	7.17E-01	6.53E-02	1.47E-01	7.90E-01	3.19E-01	5.60E-03	2.00E-04	0.00E+00
Annexin A3	P12429	7.70E-01	1.41E-02	5.60E-01	0.00E+00	1.00E+00	9.46E-01	1.00E+00	1.00E+00	9.99E-01	1.00E+00	1.00E+00	1.00E+00	6.16E-01
Alpha-actinin-1	P12814	1.00E+00	2.96E-01	2.95E-01	1.20E-03	9.95E-01	1.00E+00	9.67E-01	1.00E+00	4.08E-01	7.30E-03	1.00E+00	2.30E-03	0.00E+00
Xaa-Pro dipeptidase	P12955	7.61E-01	4.50E-03	2.99E-01	1.22E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.37E-01	9.82E-01	9.98E-01	5.79E-01	1.00E+00
X-ray repair cross-complementing protein 5	P13010	9.57E-01	1.65E-02	3.19E-01	1.30E-03	1.00E+00	1.00E+00	9.96E-01	9.98E-01	1.00E+00	1.00E+00	1.00E+00	9.98E-01	1.00E+00
Lysosome-associated membrane glycoprotein 2	P13473	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.50E-01	5.62E-01	9.81E-01	1.00E+00	6.59E-01	8.57E-01	5.33E-01	3.60E-03	4.08E-02
CD59 glycoprotein	P13987	6.04E-02	0.00E+00	2.04E-01	0.00E+00	9.49E-01	9.95E-01	2.81E-02	8.79E-01	9.92E-01	9.10E-03	5.32E-02	0.00E+00	0.00E+00
Farnesyl pyrophosphate synthase	P14324	9.97E-01	4.00E-04	4.70E-03	0.00E+00	1.00E+00	8.70E-01	9.98E-01	9.99E-01	1.27E-01	8.00E-04	5.66E-01	0.00E+00	0.00E+00
Glutamine synthetase	P15104	3.15E-01	1.00E+00	7.78E-01	6.00E-04	9.98E-01	1.00E+00	2.94E-02	5.75E-01	8.29E-01	5.00E-03	5.73E-01	8.90E-02	0.00E+00
Aminopeptidase N	P15144	6.36E-01	3.01E-02	8.71E-01	3.20E-03	1.58E-01	1.62E-02	1.88E-02	1.73E-02	1.00E-04	2.00E-04	6.00E-04	0.00E+00	0.00E+00
Arylsulfatase A	P15289	9.74E-01	9.76E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.70E-01	9.84E-01	9.98E-01	2.17E-01	1.00E+00	7.80E-01	2.00E-04
Unconventional myosin-Ic	O00159	1.40E-03	4.93E-01	0.00E+00	1.00E+00	1.00E+00	9.96E-01	2.91E-01	5.72E-01	1.00E+00	3.66E-01	3.90E-01	9.34E-01	1.00E-04
N-acetylglucosamine-6-sulfatase	P15586	9.99E-01	1.05E-01	1.34E-02	2.60E-03	9.99E-01	6.96E-01	1.24E-01	9.98E-01	1.20E-02	3.00E-04	1.00E+00	0.00E+00	0.00E+00
Short-chain specific acyl-CoA dehydrogenase, mitochondrial	P16219	6.04E-01	9.99E-01	9.82E-01	2.70E-03	9.99E-01	3.26E-01	9.54E-01	1.00E+00	9.21E-01	1.00E+00	9.92E-01	9.97E-01	9.93E-01
NADPH-cytochrome P450 reductase	P16435	1.00E+00	8.00E-04	2.00E-04	0.00E+00	7.84E-01	9.99E-01	9.82E-01	4.63E-01	6.97E-01	9.45E-01	1.11E-01	1.00E+00	5.85E-01
Fumarylacetoacetase	P16930	6.92E-01	9.11E-01	1.00E+00	5.61E-02	9.91E-01	1.00E+00	1.00E+00	4.33E-01	9.71E-01	9.87E-01	5.75E-01	1.00E+00	1.00E-03
Stathmin	P16949	1.49E-01	0.00E+00	3.00E-03	0.00E+00	9.99E-01	9.94E-01	5.45E-01	1.00E+00	1.00E+00	9.76E-01	1.39E-01	9.41E-01	0.00E+00
Ganglioside GM2 activator	P17900	8.65E-01	1.00E+00	8.76E-01	7.68E-01	1.00E+00	9.43E-01	1.00E+00	9.69E-01	1.73E-02	1.00E+00	1.00E+00	4.90E-03	5.30E-03
Nucleolin	P19338	1.00E+00	1.11E-01	4.55E-01	7.80E-03	1.00E+00	1.00E+00	9.83E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.95E-01	1.00E+00
Spermidine synthase	P19623	9.94E-01	6.52E-01	9.98E-01	1.00E+00	9.95E-01	1.00E+00	9.53E-01	3.21E-01	3.51E-01	7.94E-01	9.17E-01	5.80E-03	6.50E-03
Thymidine phosphorylase	P19971	8.27E-01	3.00E-03	1.81E-01	2.60E-01	1.00E+00	9.09E-01	1.00E+00	9.45E-01	9.62E-02	6.93E-02	9.11E-01	4.00E-04	2.00E-04
Serine-pyruvate aminotransferase	P21549	1.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00	5.35E-01	6.97E-01	9.98E-01	1.79E-01	4.90E-03	5.70E-03	0.00E+00	0.00E+00
Elongation factor 1-alpha 2	Q05639	1.60E-03	0.00E+00	1.78E-02	0.00E+00	1.00E+00	9.99E-01	9.99E-01	1.00E+00	8.08E-01	9.55E-01	9.76E-01	1.00E+00	6.46E-01
Voltage-dependent anion-selective channel protein 1	P21796	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.97E-01	8.64E-01	9.63E-01	9.83E-01	3.84E-01	9.99E-01	1.62E-01	1.00E-04
rRNA 2'-O-methyltransferase fibrillar	P22087	1.00E+00	9.96E-01	1.00E+00	2.27E-01	9.63E-01	9.97E-01	9.96E-01	9.41E-01	1.00E+00	5.80E-01	1.00E+00	9.76E-01	2.50E-03
UDP-glucuronosyltransferase 1-1	P22309	9.99E-01	5.00E-03	4.17E-02	1.44E-01	1.00E+00	6.03E-01	1.00E+00	1.00E+00	7.72E-02	9.49E-01	9.94E-01	3.70E-03	9.93E-01
Carnitine O-palmitoyltransferase 2, mitochondrial	P23786	9.97E-01	3.31E-01	9.03E-01	4.00E-04	9.98E-01	9.53E-01	5.99E-01	1.00E+00	6.43E-02	1.73E-02	9.59E-01	4.30E-03	6.50E-03
Heterogeneous nuclear ribonucleoprotein D0	Q14103	1.60E-03	0.00E+00	8.99E-01	0.00E+00	1.00E+00	9.99E-01	1.00E+00	5.28E-01	9.65E-01	6.62E-01	2.88E-01	6.90E-03	2.59E-01
Amine oxidase [flavin-containing] B	P27338	5.96E-01	4.10E-03	4.25E-01	4.00E-04	1.00E+00	9.89E-01	1.91E-01	9.99E-01	5.87E-01	2.40E-03	1.00E+00	2.00E-04	0.00E+00
14-3-3 protein theta	P27348	9.48E-01	1.00E-04	4.80E-03	0.00E+00	9.88E-01	1.00E+00	9.99E-01	1.00E+00	9.70E-01	1.00E+00	1.00E+00	1.00E+00	7.85E-01
Dipeptidyl peptidase 4	P27487	9.99E-01	9.67E-01	5.40E-01	7.32E-01	1.00E+00	2.40E-01	1.00E+00	1.00E+00	3.63E-01	7.76E-01	1.00E+00	7.20E-03	1.10E-03
Transketolase	P29401	9.90E-01	3.44E-02	3.52E-01	2.00E-04	1.00E+00	1.00E+00	9.68E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.94E-01	9.02E-01
Phenazine biosynthesis-like domain-containing protein	P30039	1.00E+00	2.04E-01	5.72E-01	2.00E-03	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	3.50E-01	4.89E-01	2.24E-01	1.00E-04
Enoyl-CoA hydratase, mitochondrial	P30084	8.81E-01	0.00E+00	4.00E-04	0.00E+00	1.00E+00	1.00E+00	9.86E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.89E-01	5.95E-01
UMP-CMP kinase	P30085	1.00E+00	5.92E-01	2.51E-01	1.00E+00	1.00E+00	9.99E-01	9.70E-01	1.00E+00	9.88E-01	5.80E-01	9.99E-01	1.00E+00	3.70E-03
Leukocyte elastase inhibitor	P30740	9.01E-01	0.00E+00	7.00E-04	0.00E+00	9.99E-01	1.00E+00	1.00E+00	9.93E-01	1.00E+00	3.50E-03	1.00E+00	9.60E-01	0.00E+00
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	P31040	1.04E-01	3.30E-03	9.48E-01	6.53E-01	8.54E-01	3.32E-01	4.20E-02	1.00E+00	1.75E-02	0.00E+00	1.01E-01	0.00E+00	0.00E+00
S-adenosylmethionine synthase isoform type-2	P31153	1.00E+00	3.68E-02	5.01E-02	7.89E-02	1.00E+00	8.69E-02	9.97E-01	1.00E+00	5.70E-02	4.00E-04	1.00E+00	0.00E+00	0.00E+00
Calcium-binding mitochondrial carrier protein ScaMC-1	Q6NUK1	2.60E-03	1.00E+00	1.40E-03	7.27E-01	9.99E-01	9.25E-01	1.57E-01	1.00E+00	3.53E-01	1.02E-02	9.96E-01	5.80E-03	0.00E+00
Cytidine deaminase	P32320	8.95E-01	2.88E-01	9.96E-01	0.00E+00	1.00E+00	9.99E-01	9.57E-01	9.42E-01	7.33E-01	1.00E+00	6.52E-01	8.39E-01	8.59E-01
Lipid-phosphate phosphatase	P34913	9.98E-01	3.60E-01	9.06E-01	1.51E-01	1.00E+00	9.90E-01	8.72E-01	9.96E-01	1.00E+00	7.93E-01	6.47E-01	1.18E-01	1.00E-04
Glypican-1	P35052	9.66E-01	1.16E-02	2.30E-01	1.00E-03	1.00E+00	1.00E+00	9.09E-01	1.00E+00	9.94E-01	1.00E+00	1.00E+00	1.00E+00	5.16E-01
Amylo-alpha-1,6-glucosidase	P35573	7.49E-01	0.00E+00	2.20E-03	0.00E+00	1.00E+00	9.99E-01	1.00E+00	9.89E-01	1.00E+00	4.50E-02	9.76E-01	7.74E-02	0.00E+00
Cation-independent mannose-6-phosphate receptor	P11717	2.80E-03	7.90E-03	1.00E+00	0.00E+00	9.41E-01	9.40E-01	4.90E-02	7.99E-01	9.70E-01	3.99E-01	3.60E-01	3.34E-01	8.31E-02
ATP synthase subunit gamma, mitochondrial	P36542	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.99E-01	1.00E+00	9.12E-01	1.00E+00	1.00E+00	8.88E-01	9.87E-01	1.00E+00	6.60E-03
Transgelin-2	P37802	9.95E-01	7.00E-04	1.11E-02	3.00E-03	9.98E-01	1.00E+00	9.74E-01	9.99E-01	1.00E+00	4.93E-02	1.00E+00	1.00E+00	9.60E-03

Nom des protéines	Numéro d'accension Uniprot	D2_ctrl-D0_ctrl	D8_ctrl-D0_ctrl	D8_ctrl-D2_ctrl	D14_ctrl-D0_ctrl	D2_0.05mM-D2_ctrl	D8_0.05mM-D8_ctrl	D14_0.05mM-D14_ctrl	D2_0.15mM-D2_ctrl	D8_0.15mM-D8_ctrl	D14_0.15mM-D14_ctrl	D2_0.5mM-D2_ctrl	D8_0.5mM-D8_ctrl	D14_0.5mM-D14_ctrl
Nicotinamide N-methyltransferase	P40261	4.13E-01	4.12E-01	1.90E-03	1.35E-01	7.99E-01	1.00E+00	1.00E+00	7.96E-01	1.00E+00	8.71E-02	9.90E-01	1.00E+00	1.00E-04
Long chain 3-hydroxyacyl-CoA dehydrogenase	P40939	1.00E+00	4.44E-02	3.90E-02	0.00E+00	1.02E-01	4.12E-01	3.19E-02	9.05E-02	6.70E-03	0.00E+00	3.00E-04	0.00E+00	0.00E+00
Myosin-9	P35579	2.90E-03	1.44E-02	1.00E+00	4.10E-03	1.00E+00	1.00E+00	6.72E-01	3.77E-01	2.07E-02	0.00E+00	3.40E-03	7.10E-03	0.00E+00
Aldo-keto reductase family 1 member C3	P42330	9.71E-01	2.80E-03	6.95E-02	0.00E+00	7.36E-01	9.16E-01	3.21E-01	9.94E-01	9.98E-01	6.46E-01	1.00E+00	4.04E-01	5.51E-01
3-ketoacyl-CoA thiolase, mitochondrial	P42765	1.00E+00	3.00E-04	3.00E-04	0.00E+00	1.00E+00	9.97E-01	9.47E-01	8.95E-01	3.11E-01	4.00E-04	9.81E-01	1.72E-02	6.00E-04
Lysosomal Pro-X carboxypeptidase	P42785	8.22E-01	1.00E+00	5.24E-01	8.24E-01	1.00E+00	7.66E-01	9.60E-01	1.00E+00	2.55E-02	7.66E-01	1.00E+00	2.00E-04	4.44E-02
Yorkie homolog	P46937	1.00E+00	2.94E-02	6.70E-03	1.93E-01	8.98E-01	8.66E-01	9.35E-01	1.00E+00	9.48E-01	1.00E+00	9.87E-01	9.43E-01	8.03E-01
Isocitrate dehydrogenase [NADP], mitochondrial	P48735	9.95E-01	4.00E-04	5.50E-03	0.00E+00	9.98E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	7.02E-01	1.00E+00	1.00E+00
Signal recognition particle 9 kDa protein	P49458	1.00E+00	2.12E-02	4.59E-02	1.90E-03	1.00E+00	1.00E+00	9.99E-01	1.00E+00	9.99E-01	1.00E+00	1.00E+00	1.00E+00	5.18E-01
Ketohexokinase	P50053	1.00E+00	1.30E-03	2.20E-03	7.90E-03	1.00E+00	9.99E-01	1.00E+00	1.00E+00	7.26E-01	7.12E-01	1.00E+00	2.56E-02	2.60E-03
Serpin H1	P50454	7.53E-01	2.60E-03	0.00E+00	9.96E-01	6.71E-01	1.00E+00	9.99E-01	3.58E-02	1.00E+00	4.92E-01	1.30E-01	1.00E+00	2.13E-01
Palmitoyl-protein thioesterase 1	P50897	1.00E+00	7.96E-01	8.93E-01	1.19E-02	1.00E+00	1.00E+00	2.30E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.74E-01	0.00E+00
D-3-phosphoglycerate dehydrogenase	O43175	6.40E-03	0.00E+00	1.00E-04	0.00E+00	6.75E-01	3.56E-02	0.00E+00	6.90E-02	0.00E+00	0.00E+00	4.34E-02	0.00E+00	0.00E+00
Ribonuclease UK114	P52758	7.58E-01	4.47E-01	1.00E+00	3.06E-01	7.73E-01	9.94E-01	9.96E-01	1.76E-01	9.95E-01	2.09E-01	8.31E-02	8.00E-04	9.37E-02
Dipeptidyl peptidase 1	P53634	1.00E+00	1.00E+00	9.84E-01	6.24E-01	1.00E+00	8.95E-01	4.16E-02	1.00E+00	6.89E-01	1.60E-03	9.98E-01	3.90E-03	0.00E+00
UDP-glucuronosyltransferase 2B15	P54855	1.00E+00	1.44E-01	1.04E-01	7.00E-04	1.00E+00	1.00E+00	9.27E-01	9.96E-01	9.99E-01	9.29E-01	2.59E-01	6.30E-02	0.00E+00
Heterogeneous nuclear ribonucleoproteins C1/C2	P07910	7.30E-03	9.96E-01	8.64E-02	3.46E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	6.24E-01	9.99E-01	9.84E-01	5.40E-03
3-ketoacyl-CoA thiolase	P55084	7.40E-03	2.00E-04	9.32E-01	0.00E+00	9.88E-01	6.20E-03	7.00E-03	4.90E-01	0.00E+00	0.00E+00	1.00E-04	0.00E+00	0.00E+00
Microsomal triglyceride transfer protein large subunit	P55157	1.00E+00	1.42E-01	2.13E-01	7.00E-04	1.00E+00	1.00E+00	9.80E-01	9.98E-01	1.00E+00	9.96E-01	1.00E+00	9.78E-01	1.00E-04
Double-stranded RNA-specific adenosine deaminase	P55265	1.00E+00	1.19E-01	3.63E-01	7.20E-03	9.85E-01	1.00E+00	4.00E-04	9.94E-01	6.19E-01	2.43E-02	9.81E-01	3.03E-01	3.10E-03
Eukaryotic translation initiation factor 6	P56537	9.87E-01	8.23E-01	1.67E-01	1.00E+00	2.35E-01	1.00E+00	1.00E+00	1.42E-01	1.00E+00	1.76E-01	4.33E-02	2.67E-01	1.00E-04
Proteasome activator complex subunit 3	P61289	8.49E-01	1.56E-02	4.90E-01	7.30E-03	1.00E+00	1.00E+00	9.00E-01	1.00E+00	1.00E+00	9.97E-01	1.00E+00	8.87E-01	9.97E-01
Heterogeneous nuclear ribonucleoprotein K	P61978	5.07E-01	0.00E+00	0.00E+00	0.00E+00	9.99E-01	1.00E+00	1.36E-01	1.00E+00	1.00E+00	2.28E-01	1.00E+00	9.44E-01	7.10E-03
SUMO-conjugating enzyme UBC9	P63279	9.97E-01	4.13E-02	3.23E-01	4.00E-04	9.96E-01	9.50E-01	1.00E+00	7.70E-01	9.94E-01	9.56E-01	1.00E+00	1.00E+00	6.33E-01
Casein kinase II subunit beta	P67870	9.50E-01	9.51E-02	8.03E-01	6.20E-03	1.00E+00	1.00E+00	2.34E-02	9.91E-01	1.00E+00	9.87E-01	9.95E-01	7.11E-01	6.94E-01
Leukotriene-B(4) omega-hydroxylase 1	P78329	9.56E-01	9.77E-02	3.40E-03	5.00E-04	1.00E+00	1.00E+00	1.00E+00	9.84E-01	1.00E+00	8.56E-01	1.00E+00	6.06E-01	2.10E-03
4-aminobutyrate aminotransferase, mitochondrial	P80404	5.58E-01	0.00E+00	3.70E-03	0.00E+00	9.66E-01	9.13E-01	1.00E+00	9.62E-01	1.00E+00	1.00E+00	8.54E-01	9.99E-01	9.32E-02
Serine/arginine-rich splicing factor 3	P84103	9.53E-01	1.50E-01	8.99E-01	1.20E-03	1.00E+00	8.50E-01	7.91E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.75E-01
Histone H3.3	P84243	2.96E-01	5.30E-03	8.01E-01	0.00E+00	1.00E+00	1.00E+00	7.74E-01	1.00E+00	1.29E-01	1.60E-03	9.42E-01	0.00E+00	0.00E+00
15-hydroxyprostaglandin dehydrogenase [NAD(+)]	P15428	1.06E-02	0.00E+00	0.00E+00	0.00E+00	9.96E-01	4.69E-02	1.00E+00	9.71E-01	3.50E-03	2.60E-03	9.97E-01	0.00E+00	7.24E-02
Sorbitol dehydrogenase	Q00796	9.35E-01	2.60E-03	9.45E-02	0.00E+00	1.00E+00	9.89E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.91E-01	9.06E-01	6.65E-02
Hydroxymethylglutaryl-CoA synthase, cytoplasmic	Q01581	1.00E+00	1.80E-03	1.10E-03	0.00E+00	8.38E-01	4.61E-01	9.58E-01	3.50E-01	1.26E-02	4.10E-03	2.26E-01	5.00E-03	0.00E+00
Dihydroorotate dehydrogenase (quinone), mitochondrial	Q02127	9.90E-01	7.77E-01	1.00E+00	9.09E-02	9.98E-01	1.00E+00	4.21E-01	1.00E+00	1.00E+00	8.88E-01	1.00E+00	9.65E-01	2.10E-03
Gamma-glutamyl phosphate reductase	P54886	1.09E-02	7.30E-03	1.00E+00	0.00E+00	1.00E+00	1.00E+00	9.10E-01	1.00E+00	3.78E-01	9.99E-01	1.00E+00	4.34E-01	9.96E-01
Cytochrome P450 4A11	Q02928	8.58E-01	4.08E-02	7.36E-01	2.00E-04	1.00E+00	3.12E-01	1.18E-01	1.00E+00	7.00E-03	4.09E-01	1.00E+00	3.83E-02	9.89E-01
Aminoacylase-1	Q03154	5.27E-01	1.40E-02	7.96E-01	8.00E-04	9.92E-01	8.23E-01	5.23E-01	9.73E-01	9.96E-01	7.50E-03	5.68E-01	9.71E-02	0.00E+00
Mevalonate kinase	Q03426	1.00E+00	1.00E+00	1.00E+00	5.53E-01	9.99E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	6.80E-03
Lon protease homolog, mitochondrial	P36776	1.13E-02	8.57E-02	9.99E-01	0.00E+00	9.98E-01	1.00E+00	1.00E+00	9.98E-01	3.01E-01	6.18E-01	1.00E+00	1.15E-02	6.70E-01
60S ribosomal protein L18	Q07020	1.00E+00	5.00E-04	1.00E-04	4.00E-03	9.97E-01	9.16E-01	4.43E-01	1.00E+00	9.09E-01	1.00E+00	2.81E-01	1.04E-01	7.31E-01
Prolow-density lipoprotein receptor-related protein 1	Q07954	1.00E+00	7.51E-02	1.45E-01	8.00E-04	1.00E+00	9.93E-01	4.72E-01	1.00E+00	9.99E-01	8.91E-01	1.00E+00	9.48E-01	2.31E-01
ATP-dependent RNA helicase A	Q08211	1.00E+00	2.14E-01	2.49E-01	2.00E-03	1.00E+00	1.00E+00	6.74E-01	1.00E+00	1.00E+00	9.97E-01	1.00E+00	1.00E+00	9.77E-01
Galectin-3-binding protein	Q08380	9.83E-01	1.00E+00	7.72E-01	9.37E-01	1.00E+00	9.43E-01	9.07E-02	1.00E+00	4.20E-02	3.45E-02	1.00E+00	4.00E-04	0.00E+00
Acyl-coenzyme A synthetase ACSM2A, mitochondrial	Q08AH3	1.14E-01	0.00E+00	0.00E+00	0.00E+00	1.00E+00	6.36E-01	9.79E-01	1.00E+00	8.22E-01	4.66E-02	8.30E-01	0.00E+00	0.00E+00
Neuroblast differentiation-associated protein AHNAK	Q09666	9.96E-01	2.00E-04	2.20E-03	0.00E+00	6.07E-01	1.00E+00	1.00E+00	3.77E-01	9.63E-01	1.00E+00	2.37E-02	9.04E-01	1.00E+00

Nom des protéines	Numéro d'accession Uniprot	D2_ctrl-D0_ctrl	D8_ctrl-D0_ctrl	D8_ctrl-D2_ctrl	D14_ctrl-D0_ctrl	D2_0.05mM-D2_ctrl	D8_0.05mM-D8_ctrl	D14_0.05mM-D14_ctrl	D2_0.15mM-D2_ctrl	D8_0.15mM-D8_ctrl	D14_0.15mM-D14_ctrl	D2_0.5mM-D2_ctrl	D8_0.5mM-D8_ctrl	D14_0.5mM-D14_ctrl
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	Q13011	3.06E-01	2.00E-04	1.04E-01	0.00E+00	1.00E+00	9.13E-01	1.39E-02	9.81E-01	4.76E-02	2.70E-03	1.00E+00	4.50E-03	0.00E+00
Chromobox protein homolog 3	Q13185	3.28E-01	4.60E-03	7.32E-01	0.00E+00	1.00E+00	1.00E+00	2.59E-02	8.86E-01	1.00E+00	4.96E-01	4.32E-01	8.36E-01	4.55E-02
Selenium-binding protein 1	Q13228	8.73E-01	2.06E-01	4.60E-03	1.23E-02	9.92E-01	9.99E-01	2.54E-01	1.00E+00	2.64E-02	5.00E-04	9.84E-01	0.00E+00	0.00E+00
Transcription intermediary factor 1-beta	Q13263	1.00E+00	1.00E-03	3.20E-03	1.42E-02	1.00E+00	9.96E-01	1.24E-02	1.00E+00	5.02E-01	1.00E+00	9.97E-01	2.96E-01	4.40E-01
Ras GTPase-activating protein-binding protein 1	Q13283	8.46E-01	6.81E-02	8.69E-01	1.10E-03	9.99E-01	9.97E-01	9.62E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.97E-01	9.97E-01
Acid ceramidase	Q13510	1.00E+00	1.00E+00	9.96E-01	1.00E+00	1.00E+00	8.55E-01	9.28E-01	1.00E+00	8.74E-01	3.95E-01	1.00E+00	2.30E-03	1.69E-02
Sorting nexin-1	Q13596	9.09E-01	6.90E-03	2.35E-01	2.50E-03	1.00E+00	1.00E+00	2.29E-01	1.00E+00	1.00E+00	1.00E+00	9.95E-01	9.64E-01	1.00E+00
BAG family molecular chaperone regulator 3	O95817	1.30E-02	1.31E-02	1.00E+00	3.00E-04	1.00E+00	1.00E+00	7.61E-01	9.27E-01	1.00E+00	3.87E-01	3.98E-01	7.35E-01	2.00E-04
Bile acid-CoA:amino acid N-acyltransferase	Q14032	7.51E-02	0.00E+00	1.10E-03	0.00E+00	9.92E-01	1.00E+00	1.00E+00	8.47E-01	7.40E-01	1.29E-01	3.61E-01	0.00E+00	0.00E+00
Glycine N-methyltransferase	Q14749	1.00E+00	1.00E+00	1.00E+00	7.10E-02	1.00E+00	4.28E-02	9.72E-01	9.55E-01	5.00E-04	2.23E-02	6.62E-01	0.00E+00	0.00E+00
Prostaglandin reductase 1	Q14914	9.99E-01	9.85E-01	6.46E-01	1.96E-01	6.77E-01	9.95E-01	1.96E-01	7.79E-01	1.00E+00	2.12E-02	9.97E-01	3.30E-03	0.00E+00
Plectin	Q15149	1.00E+00	2.00E-04	2.00E-04	4.52E-02	1.00E+00	9.61E-01	8.24E-01	1.00E+00	1.00E+00	4.15E-01	9.23E-01	1.00E+00	8.00E-04
Splicing factor 3B subunit 3	Q15393	8.91E-01	4.13E-01	1.00E+00	5.30E-03	1.00E+00	9.98E-01	2.51E-01	9.96E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.11E-01
Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial	Q16134	1.00E+00	1.00E+00	1.00E+00	2.05E-02	8.18E-01	8.48E-02	4.49E-02	8.41E-01	5.00E-04	2.00E-04	4.46E-01	0.00E+00	0.00E+00
Septin-7	Q16181	1.00E+00	9.00E-04	3.00E-03	0.00E+00	1.00E+00	8.81E-01	9.78E-01	9.99E-01	9.26E-01	1.00E+00	1.00E+00	7.48E-01	1.00E+00
Cleavage and polyadenylation specificity factor subunit 6	Q16630	9.70E-01	7.47E-02	2.90E-03	1.12E-01	9.46E-01	7.38E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.72E-01	9.09E-01	1.00E+00
Kynurenine--oxoglutarate transaminase 1	Q16773	7.99E-02	0.00E+00	4.86E-02	0.00E+00	9.99E-01	9.19E-01	1.00E+00	9.96E-01	1.18E-01	1.60E-03	8.38E-01	1.70E-03	0.00E+00
Reticon-1	Q16799	9.82E-01	2.75E-01	9.46E-01	1.00E-04	1.00E+00	9.88E-01	9.99E-01	1.00E+00	9.92E-01	1.00E+00	1.00E+00	1.00E+00	8.31E-01
Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	Q16836	8.38E-01	9.10E-03	3.70E-01	0.00E+00	1.00E+00	8.94E-01	9.82E-01	1.00E+00	9.16E-01	1.00E+00	1.00E+00	9.99E-01	9.88E-01
UTP--glucose-1-phosphate uridylyltransferase	Q16851	2.09E-01	0.00E+00	1.70E-03	0.00E+00	1.00E+00	5.96E-01	9.91E-01	8.23E-01	2.20E-03	0.00E+00	9.93E-02	0.00E+00	0.00E+00
FAD-AMP lyase (cyclizing)	Q3LXA3	8.54E-01	0.00E+00	0.00E+00	1.00E-04	1.00E+00	9.76E-01	1.00E+00	4.82E-01	2.00E-03	3.00E-04	2.25E-02	0.00E+00	0.00E+00
Quinone oxidoreductase PIG3	Q53FA7	9.14E-01	1.80E-03	7.92E-02	0.00E+00	1.00E+00	1.00E+00	9.97E-01	9.84E-01	9.00E-01	5.00E-04	8.94E-01	4.22E-02	4.00E-04
Torsin-1A-interacting protein 1	Q5JTV8	1.00E+00	9.80E-01	1.00E+00	8.10E-03	1.00E+00	1.00E+00	1.14E-01	1.00E+00	1.00E+00	8.02E-01	9.96E-01	1.00E+00	1.30E-03
Long-chain fatty acid transport protein 3	Q5K4L6	3.25E-01	1.14E-02	9.11E-01	0.00E+00	1.00E+00	1.00E+00	8.87E-01	1.00E+00	1.00E+00	9.53E-01	1.00E+00	1.00E+00	8.82E-01
Acyl-coenzyme A synthetase ACSM2B, mitochondrial	Q68CK6	9.24E-01	0.00E+00	1.70E-03	0.00E+00	1.00E+00	9.71E-01	9.64E-01	9.63E-01	1.93E-01	1.00E-04	6.44E-01	0.00E+00	0.00E+00
Twinfilin-2	Q6IBS0	3.89E-01	1.24E-01	1.00E+00	4.00E-04	1.00E+00	1.00E+00	4.23E-01	9.93E-01	1.00E+00	9.90E-01	9.00E-01	9.99E-01	3.78E-01
Cystathionine beta-synthase	P35520	1.55E-02	0.00E+00	0.00E+00	0.00E+00	1.00E+00	4.80E-01	9.68E-01	1.00E+00	5.14E-01	9.99E-01	8.00E-01	1.08E-02	6.30E-01
Acyl-coenzyme A synthetase ACSM5, mitochondrial	Q6NUN0	8.03E-01	5.40E-03	2.99E-01	1.00E-04	1.00E+00	9.62E-01	1.36E-01	9.11E-01	4.53E-01	3.83E-01	1.00E+00	4.39E-02	6.04E-01
Polymerase I and transcript release factor	Q6NZI2	1.00E+00	9.86E-01	9.62E-01	6.70E-03	1.00E+00	9.94E-01	1.00E+00	1.00E+00	7.67E-01	1.00E+00	1.00E+00	9.25E-01	1.48E-01
Acylpyruvate FAHD1, mitochondrial	Q6P587	9.23E-01	1.95E-01	9.66E-01	3.60E-03	1.00E+00	1.00E+00	4.46E-01	1.00E+00	6.32E-01	9.30E-01	1.00E+00	1.89E-01	0.00E+00
Nicotinate phosphoribosyltransferase	Q6XQN6	9.91E-01	3.80E-01	9.68E-01	3.73E-01	1.00E+00	9.96E-01	1.00E+00	9.87E-01	9.76E-01	8.69E-01	8.26E-01	1.80E-01	5.60E-03
Hydroxysteroid dehydrogenase-like protein 2	Q6YN16	6.05E-01	1.35E-01	9.98E-01	2.70E-03	1.00E+00	5.14E-02	4.60E-03	1.00E+00	2.20E-03	0.00E+00	1.00E+00	0.00E+00	0.00E+00
Acyl-CoA dehydrogenase family member 11	Q709F0	2.41E-01	2.10E-03	6.51E-01	2.50E-03	1.00E+00	1.00E+00	8.78E-01	1.00E+00	1.00E+00	1.00E+00	9.86E-01	1.00E+00	1.00E+00
Myosin-14	Q7Z406	9.90E-01	0.00E+00	0.00E+00	0.00E+00	9.99E-01	9.49E-01	8.62E-01	3.97E-01	9.99E-01	4.30E-03	8.90E-01	2.84E-02	0.00E+00
L-xylulose reductase	Q7Z4W1	1.00E+00	1.00E+00	1.00E+00	2.65E-01	9.98E-01	2.25E-02	9.76E-01	1.00E+00	1.70E-03	8.55E-01	4.86E-01	0.00E+00	4.90E-03
Probable D-lactate dehydrogenase, mitochondrial	Q86WU2	1.81E-01	7.80E-03	9.60E-01	2.00E-04	7.12E-01	1.00E+00	8.78E-01	9.37E-01	1.00E+00	8.80E-01	9.05E-01	1.00E+00	1.00E+00
Enoyl-CoA hydratase domain-containing protein 2, mitochondrial	Q86YB7	1.00E+00	9.88E-01	9.94E-01	9.94E-01	9.84E-01	1.55E-01	1.51E-01	9.44E-02	1.15E-02	0.00E+00	2.24E-01	1.90E-03	3.00E-04
Phospholipase D3	Q8IV08	9.77E-01	1.00E-04	2.40E-03	0.00E+00	1.00E+00	9.51E-01	3.74E-02	8.82E-01	7.48E-01	3.19E-01	6.18E-01	3.60E-03	0.00E+00
Proline-, glutamic acid- and leucine-rich protein 1	Q8IZL8	9.94E-01	2.22E-02	1.40E-03	2.70E-01	9.87E-01	1.00E+00	2.29E-01	9.61E-01	9.43E-01	1.00E+00	9.95E-01	2.28E-01	1.00E+00
ER membrane protein complex subunit 1	Q8N766	1.00E+00	9.89E-01	1.00E+00	9.92E-01	1.00E+00	1.00E+00	3.44E-01	1.00E+00	9.07E-01	9.85E-01	1.00E+00	4.92E-01	2.30E-03
Putative phospholipase B-like 2	Q8NHP8	1.00E+00	9.74E-01	7.49E-01	1.00E+00	9.70E-01	1.00E+00	1.00E+00	9.89E-01	9.99E-01	1.16E-01	8.24E-01	6.03E-01	0.00E+00
CD99 antigen-like protein 2	Q8TCZ2	9.99E-01	7.90E-03	7.29E-02	1.00E-04	1.00E+00	8.29E-01	1.04E-01	1.00E+00	9.80E-01	9.92E-01	1.00E+00	2.14E-01	3.00E-04
Palladin	Q8WX93	8.16E-01	3.00E-04	2.92E-02	0.00E+00	1.00E+00	9.22E-01	2.42E-01	1.00E+00	6.74E-01	9.84E-01	1.00E+00	9.99E-01	3.69E-02
Estradiol 17-beta-dehydrogenase 8	Q92506	9.61E-01	4.00E-04	1.31E-02	1.00E-04	9.51E-01	1.00E+00	9.97E-01	1.00E+00	1.00E+00	8.45E-01	1.00E+00	1.00E+00	2.33E-01
Probable ATP-dependent RNA helicase DDX17	Q92841	7.37E-01	4.39E-02	8.72E-01	1.00E-04	9.79E-01	1.00E+00	7.36E-01	1.00E+00	9.99E-01	1.00E+00	9.98E-01	9.99E-01	1.24E-01

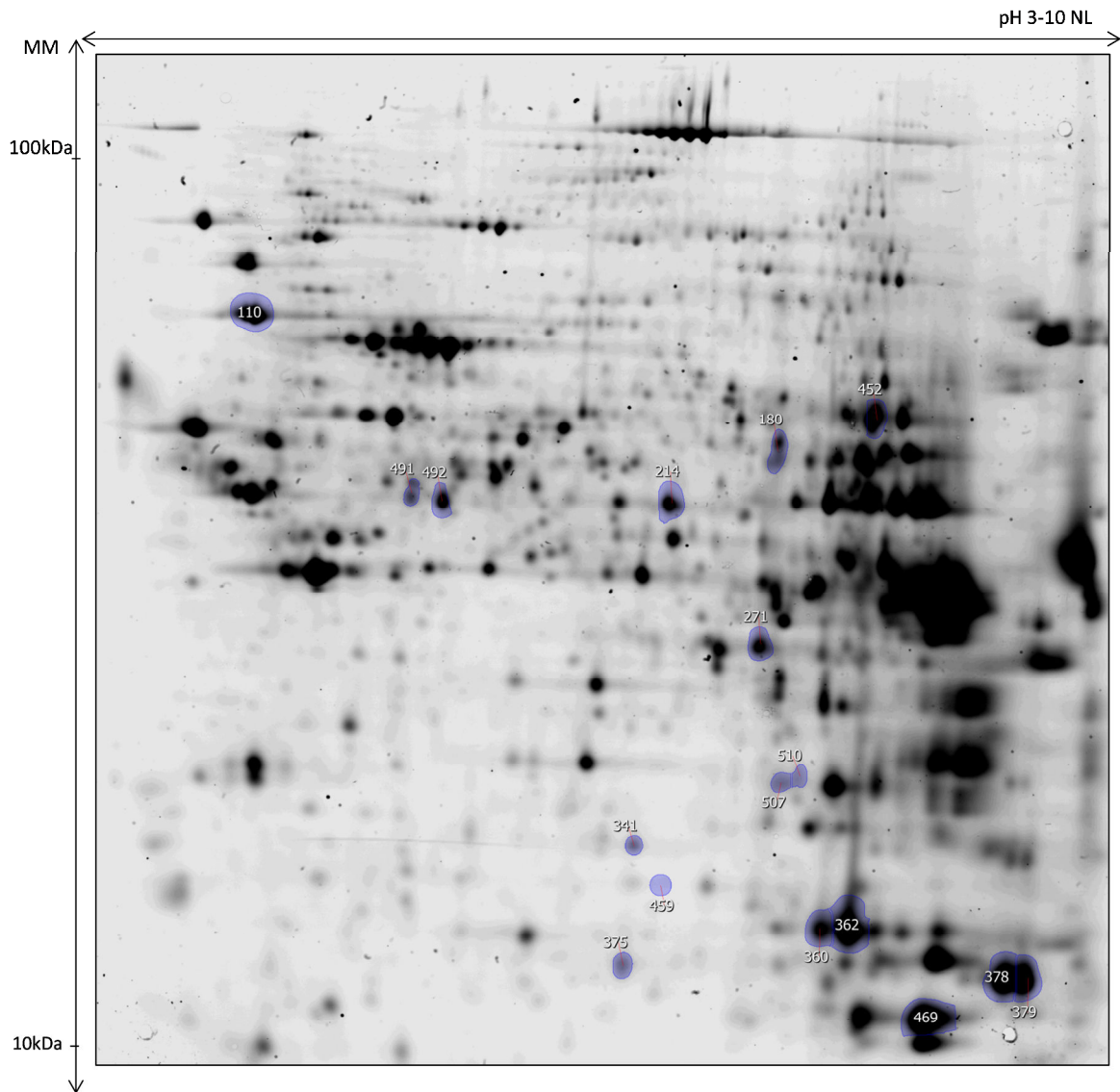
Nom des protéines	Numéro d'accension Uniprot	D2_ctrl-D0_ctrl	D8_ctrl-D0_ctrl	D8_ctrl-D2_ctrl	D14_ctrl-D0_ctrl	D2_0.05mM-D2_ctrl	D8_0.05mM-D8_ctrl	D14_0.05mM-D14_ctrl	D2_0.15mM-D2_ctrl	D8_0.15mM-D8_ctrl	D14_0.15mM-D14_ctrl	D2_0.5mM-D2_ctrl	D8_0.5mM-D8_ctrl	D14_0.5mM-D14_ctrl
Golgi apparatus protein 1	Q92896	5.55E-01	4.24E-02	9.58E-01	1.10E-03	1.00E+00	1.00E+00	7.94E-02	9.23E-01	1.00E+00	3.51E-01	9.63E-01	9.90E-01	3.00E-04
Far upstream element-binding protein 2	Q92945	9.42E-01	1.84E-02	3.77E-01	0.00E+00	1.00E+00	1.00E+00	4.54E-01	1.00E+00	1.00E+00	1.00E+00	9.96E-01	1.00E+00	6.12E-02
Far upstream element-binding protein 1	Q96AE4	9.89E-01	3.40E-03	5.88E-02	1.30E-03	1.00E+00	9.16E-01	2.87E-02	1.00E+00	9.90E-01	3.45E-01	8.60E-01	1.23E-02	1.14E-01
Isochorismatase domain-containing protein 1	Q96CN7	9.92E-01	9.14E-02	6.40E-03	2.28E-01	8.71E-01	9.33E-01	1.00E+00	9.86E-01	7.90E-01	1.00E+00	1.00E+00	9.46E-01	1.00E+00
Receptor expression-enhancing protein 6	Q96HR9	9.12E-01	7.17E-01	1.00E+00	2.70E-03	1.00E+00	9.41E-01	1.00E+00	9.76E-01	8.71E-01	9.93E-01	9.76E-01	1.00E+00	2.08E-01
Mannose-1-phosphate guanylyltransferase alpha	Q96UJ6	9.92E-01	4.54E-02	4.05E-01	8.03E-02	3.45E-01	7.96E-01	1.00E+00	1.00E+00	9.44E-01	9.99E-01	9.93E-01	8.80E-03	1.00E+00
Up-regulated during skeletal muscle growth protein 5	Q96IX5	1.00E+00	9.88E-01	1.00E+00	1.00E+00	1.00E+00	4.04E-01	9.82E-01	1.00E+00	4.85E-01	9.99E-01	9.92E-01	6.05E-01	5.00E-03
Adipocyte plasma membrane-associated protein	Q9HDC9	1.59E-02	0.00E+00	0.00E+00	0.00E+00	5.28E-01	5.75E-01	4.81E-01	7.65E-01	1.05E-01	5.62E-01	1.00E+00	3.00E-04	4.72E-02
Niban-like protein 1	Q96TA1	8.90E-01	4.09E-02	6.91E-01	1.00E-04	1.00E+00	4.99E-01	9.99E-01	1.00E+00	9.99E-01	9.99E-01	1.00E+00	9.99E-01	6.87E-01
Microsomal glutathione S-transferase 2	Q99735	9.21E-01	9.97E-01	1.00E+00	1.00E+00	9.90E-01	9.40E-01	9.99E-01	1.00E+00	7.92E-01	1.00E+00	1.00E+00	1.04E-01	4.10E-03
Vitamin K epoxide reductase complex subunit 1	Q9BQB6	1.00E+00	1.00E+00	1.00E+00	9.79E-01	1.00E+00	9.80E-01	2.01E-01	1.00E+00	7.30E-02	8.74E-02	4.62E-01	1.00E-04	0.00E+00
45 kDa calcium-binding protein	Q9BRK5	9.41E-01	1.50E-03	5.46E-02	2.55E-01	9.99E-01	9.70E-01	9.63E-01	1.00E+00	8.51E-01	1.00E+00	1.00E+00	6.58E-01	1.00E+00
Pyruvate kinase PKLR	P30613	1.68E-02	0.00E+00	0.00E+00	0.00E+00	9.87E-01	1.00E+00	1.00E+00	6.94E-01	4.00E-04	7.40E-03	1.75E-02	0.00E+00	0.00E+00
Cancer-related nucleoside-triphosphatase	Q9BSD7	1.00E+00	9.99E-01	1.00E+00	4.38E-01	1.00E+00	1.00E+00	1.62E-01	1.00E+00	1.00E+00	8.77E-01	8.08E-01	6.31E-01	2.40E-03
Peroxisomal trans-2-enoyl-CoA reductase	Q9BY49	3.20E-01	1.19E-01	1.00E+00	8.24E-01	6.60E-01	7.84E-01	1.00E+00	7.90E-03	1.50E-01	7.75E-01	2.37E-01	3.10E-03	2.00E-04
Ester hydrolase C11orf54	Q9H0W9	9.99E-01	4.00E-04	4.00E-03	4.00E-04	9.35E-01	1.00E+00	1.00E+00	1.00E+00	9.98E-01	8.07E-01	1.00E+00	1.51E-02	5.60E-03
Actin-like protein 8	Q9H568	1.00E+00	4.00E-03	1.60E-02	2.82E-01	9.98E-01	1.00E+00	1.00E+00	7.72E-01	9.94E-01	7.99E-01	9.92E-01	8.37E-01	9.98E-01
Sideroflexin-1	Q9H9B4	3.19E-01	1.80E-03	5.12E-01	4.00E-04	9.61E-01	9.92E-01	4.67E-01	1.00E+00	4.93E-01	1.90E-03	1.00E+00	2.70E-03	0.00E+00
Glyoxalase domain-containing protein 4	Q9HC38	9.99E-01	2.30E-03	3.00E-04	5.20E-03	9.79E-01	6.42E-01	1.00E+00	8.77E-01	1.22E-01	7.81E-01	9.94E-01	1.08E-02	4.60E-01
Regulation of nuclear pre-mRNA domain-containing protein 1B	Q9NQG5	4.76E-01	1.21E-01	4.00E-04	9.99E-01	9.25E-01	4.75E-01	1.00E+00	8.89E-01	3.45E-01	9.68E-01	9.75E-01	4.44E-01	9.96E-01
Sialic acid synthase	Q9NR45	6.52E-01	9.07E-01	1.00E+00	3.90E-03	1.00E+00	3.04E-01	1.00E+00	9.95E-01	9.97E-01	1.00E+00	9.57E-01	7.40E-01	1.00E+00
EH domain-containing protein 2	Q9NZN4	1.00E+00	9.81E-01	9.58E-01	4.30E-03	1.00E+00	9.95E-01	9.90E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Peroxisomal sarcosine oxidase	Q9P0Z9	1.00E+00	3.32E-02	5.12E-02	7.40E-03	1.00E+00	9.97E-01	9.98E-01	1.00E+00	9.33E-01	1.09E-01	1.01E-02	0.00E+00	0.00E+00
7-dehydrocholesterol reductase	Q9UBM7	9.77E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.83E-01	1.00E+00	9.99E-01	1.00E+00	4.28E-01	9.96E-01	1.00E+00	3.60E-03
Histone deacetylase 6	Q9UBN7	9.67E-02	0.00E+00	1.32E-02	0.00E+00	1.00E+00	9.25E-01	1.00E+00	3.68E-01	9.99E-01	9.97E-01	8.90E-01	2.35E-01	0.00E+00
Saccharopine dehydrogenase	Q9UDR5	2.56E-01	6.96E-02	1.00E+00	7.00E-04	1.00E+00	7.37E-01	2.82E-01	9.93E-01	9.95E-01	7.92E-01	9.53E-01	2.50E-01	0.00E+00
2-hydroxyacyl-CoA lyase 1	Q9UJ83	9.92E-01	3.00E-01	2.92E-02	1.43E-01	9.77E-01	7.69E-01	1.00E+00	1.00E+00	6.97E-01	6.80E-01	1.00E+00	1.00E-04	0.00E+00
Hydroxyacid oxidase 1	Q9UJM8	9.99E-01	0.00E+00	2.00E-04	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.78E-01	5.29E-01	1.74E-01	1.30E-01
F-box only protein 2	Q9UK22	1.00E+00	5.00E-04	1.40E-03	5.00E-04	1.00E+00	1.00E+00	8.61E-01	9.92E-01	8.75E-01	9.96E-01	6.19E-01	9.84E-01	1.00E+00
Sarcosine dehydrogenase, mitochondrial	Q9UL12	9.98E-01	1.00E-04	1.50E-03	1.10E-03	4.58E-01	1.12E-01	9.91E-01	2.84E-01	1.56E-02	8.93E-01	6.20E-01	2.00E-04	2.60E-02
Long-chain-fatty-acid-CoA ligase 5	Q9ULC5	8.14E-01	1.00E-03	8.00E-02	0.00E+00	1.00E+00	9.56E-01	1.00E+00	1.00E+00	3.39E-01	1.97E-02	5.75E-01	0.00E+00	0.00E+00
HIG1 domain family member 1A, mitochondrial	Q9Y241	9.49E-01	8.63E-01	1.00E+00	1.70E-03	5.87E-01	1.19E-01	9.91E-01	1.00E+00	5.07E-02	7.75E-01	1.00E+00	1.58E-02	1.00E-03
PCTP-like protein	Q9Y365	1.00E+00	0.00E+00	0.00E+00	0.00E+00	9.78E-01	9.96E-01	9.28E-01	9.65E-01	7.86E-01	1.00E+00	9.88E-01	5.40E-03	4.30E-02
FACT complex subunit SPT16	Q9Y5B9	3.43E-01	1.00E-04	8.26E-02	0.00E+00	1.00E+00	1.00E+00	1.61E-01	1.00E+00	1.00E+00	9.17E-01	9.98E-01	1.00E+00	6.16E-02
Chloride intracellular channel protein 4	Q9Y696	1.00E+00	9.81E-01	9.99E-01	1.11E-02	8.75E-01	9.92E-01	6.73E-01	9.88E-01	9.33E-01	1.00E+00	9.99E-01	2.29E-01	6.40E-03
Aldehyde dehydrogenase X, mitochondrial	P30837	3.38E-02	0.00E+00	0.00E+00	0.00E+00	1.00E+00	1.00E+00	9.95E-01	1.00E+00	6.99E-01	9.95E-01	1.00E+00	1.22E-01	1.70E-03
Alcohol dehydrogenase 1B	P00325	9.96E-01	3.00E-04	4.00E-03	0.00E+00	9.96E-01	9.97E-01	9.99E-01	1.00E+00	7.89E-01	7.40E-03	7.32E-01	0.00E+00	0.00E+00
Alcohol dehydrogenase 1C	P00326	7.91E-01	0.00E+00	6.00E-04	0.00E+00	1.00E+00	1.00E+00	2.96E-01	1.00E+00	1.89E-01	5.00E-04	2.64E-01	0.00E+00	0.00E+00
WD repeat-containing protein 1	O75083	3.42E-02	0.00E+00	2.79E-02	0.00E+00	1.00E+00	9.96E-01	5.89E-01	1.00E+00	9.79E-01	1.00E+00	9.99E-01	1.00E+00	9.99E-01
Long-chain-fatty-acid-CoA ligase 1	P33121	1.81E-01	0.00E+00	4.51E-02	0.00E+00	7.50E-01	9.00E-04	0.00E+00	9.64E-02	0.00E+00	0.00E+00	5.00E-04	0.00E+00	0.00E+00
X-ray repair cross-complementing protein 6	P12956	4.53E-02	0.00E+00	1.40E-03	0.00E+00	1.00E+00	9.86E-01	2.86E-01	9.03E-01	9.88E-01	1.00E+00	3.79E-01	3.07E-01	5.53E-01
Talin-1	Q9Y490	6.57E-01	0.00E+00	2.00E-04	0.00E+00	1.00E+00	9.99E-01	9.99E-01	1.00E+00	9.37E-01	9.95E-01	1.00E+00	4.63E-01	8.69E-01

Annexe 3 : Image de gel présentant les spots identifiés comme différentiels dans le projet concernant l'étude du cout de la reproduction et liste des protéines identifiées dans ces spots de gel



# Samespots	ANOVA	Moyenne "Control"	Ecart-type	Moyenne "Enlarged"	Ecart-type	Moyenne "Reduced"	Ecart-type	Protein name	Protein accession numbers	Nombre de peptide identifiés
657	0,01084950	1,0	0,1	0,8	0,1	1,2	0,2	Indolethylamine N-methyltransferase	gi 731019	11
516	0,01175589	1,0	0,1	0,9	0,1	1,1	0,1	Fructose-1,6-bisphosphatase 1	gi 14547989	17
570	0,02097399	1,0	0,1	0,9	0,03	0,9	0,1	Glycine N-methyltransferase	gi 15679953	9
570	0,02097399							malate dehydrogenase 1, NAD (soluble), isoform CRA_c	gi 148675904	6
422	0,03166672	1,0	0,1	0,8	0,04	0,8	0,1	Alpha-enolase	gi 13637776	15
422	0,03166672							Adenosylhomocysteinase	gi 21431841	15
553	0,03761520	1,0	0,2	1,1	0,1	1,2	0,1	N(G),N(G)-dimethylarginine dimethylaminohydrolase 1	gi 45476974	14
383	0,04398258	1,0	0,0	1,0	0,1	1,1	0,0	Selenium-binding protein 1	gi 148840436	21
748	0,04519445	1,0	0,1	0,9	0,1	1,2	0,2	Regucalcin	gi 2498920	11

Annexe 4 : Image de gel présentant les spots identifiés comme différentiels dans le projet concernant l'étude du cout de l'activation du système immunitaire et liste des protéines identifiées dans ces spots de gel



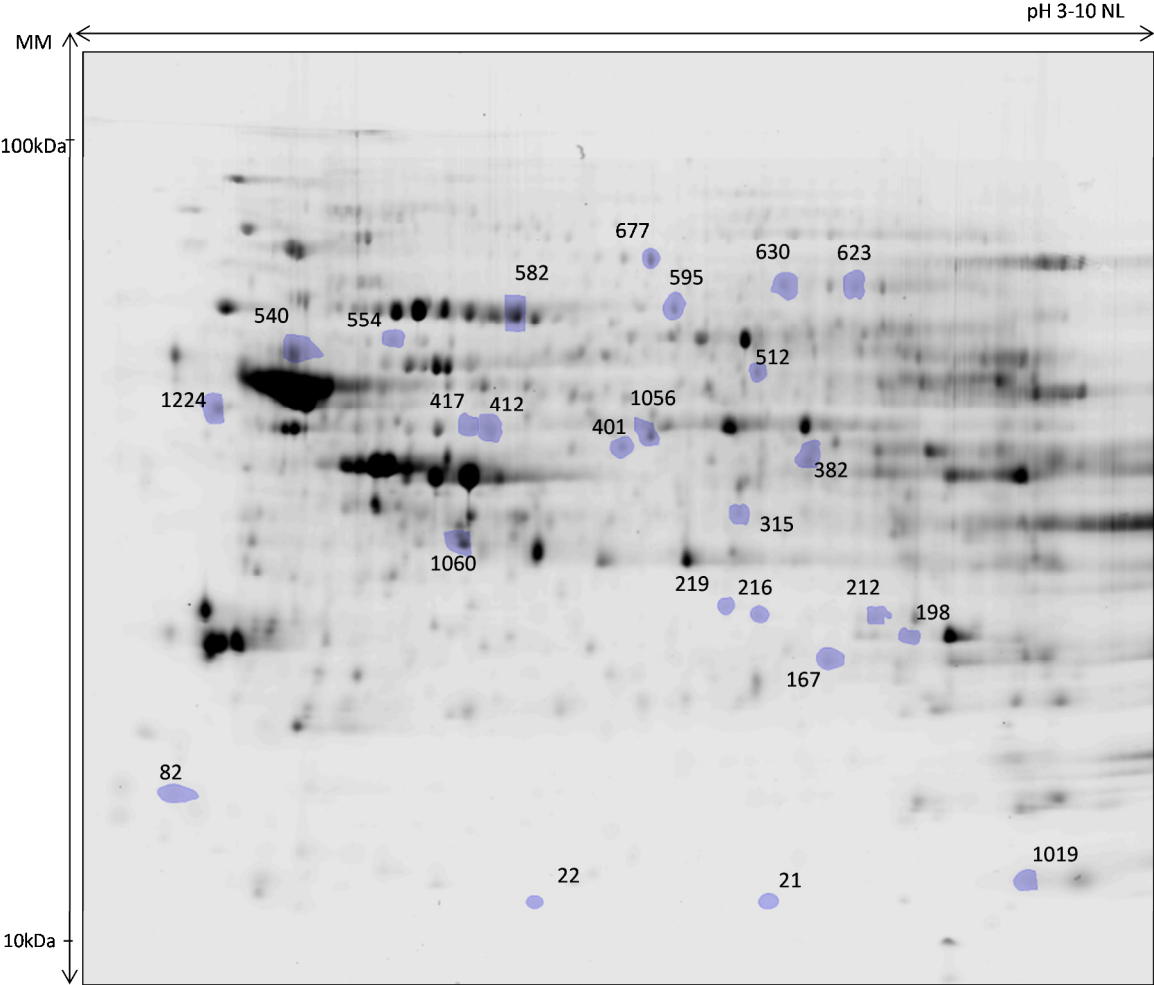
# Samespots	Anova	Test Post- Hoc de Tukey	moyenne PBS Young	ecart-type	moyenne LPS Young	ecart-type	moyenne PBS Old	ecart-type	moyenne LPS Old	ecart-type	Nom des protéines majoritaires identifiées	Numéro d'accession	Nombre de peptides	Algorithme de recherche
110	0.0021121	LPS Young-LPS Old	1	0.1	1.3	0.1	1.0	0.1	1.1	0.1	unnamed protein product	gi 12835845	24	mascot
110	0.0021121	PBS Young-LPS Young	1	0.1	1.3	0.1	1.0	0.1	1.1	0.1				
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	aldehyde dehydrogenase family 6, subfamily A1, isoform CRA_a	gi 148670856	9	mascot
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2		gi 148670856	8	omssa
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	Cytosol aminopeptidase	gi 124028616	13	mascot
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2		gi 124028616	16	omssa
180	0.0436822	PBS Young-PBS Old	1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	Dihydropolyl dehydrogenase, mitochondrial	gi 118572640	16	mascot
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2		gi 118572640	17	omssa
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	Fibrinogen beta chain	gi 67460959	14	mascot
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2		gi 67460959	15	omssa
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	pyruvate kinase isozymes R/L isoform 1	gi 153792131	3	mascot
180	0.0436822		1	0.0	1.2	0.1	1.3	0.2	1.3	0.2	unnamed protein product	gi 74146312	3	omssa
214	0.0023182	PBS Young-LPS Old	1	0.1	1.3	0.1	1.4	0.1	1.2	0.1	4-trimethylaminobutyraldehyde dehydrogenase	gi 115334671	3	mascot
214	0.0023182	PBS Young-PBS Old	1	0.1	1.3	0.1	1.4	0.1	1.2	0.1		gi 115334671	4	omssa
214	0.0023182	PBS Young-LPS Young	1	0.1	1.3	0.1	1.4	0.1	1.2	0.1	Aldehyde dehydrogenase, mitochondrial	gi 1352250	17	mascot
214	0.0023182		1	0.1	1.3	0.1	1.4	0.1	1.2	0.1		gi 1352250	16	omssa
271	0.0227063		1	0.04	0.8	0.0	0.9	0.1	0.9	0.2	Alcohol dehydrogenase class-3	gi 110825702	3	mascot
271	0.0227063		1	0.04	0.8	0.0	0.9	0.1	0.9	0.2		gi 110825702	3	omssa
271	0.0227063	PBS Young-LPS Young	1	0.04	0.8	0.0	0.9	0.1	0.9	0.2	sorbitol dehydrogenase precursor	gi 1009706	13	mascot
271	0.0227063		1	0.04	0.8	0.0	0.9	0.1	0.9	0.2		gi 1009706	15	omssa
341	0.0292196	PBS Young-LPS Young	1	0.1	0.8	0.1	0.9	0.1	1.0	0.1	3-hydroxyanthranilate 3,4-dioxygenase	gi 61211578	14	mascot
341	0.0292196		1	0.1	0.8	0.1	0.9	0.1	1.0	0.1		gi 61211578	16	omssa
360	0.0234152	PBS Old-LPS Young	1	0.2	0.8	0.1	1.2	0.1	1.2	0.3	carbonic anhydrase 3, isoform CRA_b	gi 148673185	9	mascot
360	0.0234152		1	0.2	0.8	0.1	1.2	0.1	1.2	0.3		gi 148673185	11	omssa
360	0.0234152		1	0.2	0.8	0.1	1.2	0.1	1.2	0.3	unnamed protein product	gi 12836667	5	omssa
360	0.0234152		1	0.2	0.8	0.1	1.2	0.1	1.2	0.3	unnamed protein product	gi 12844989	3	omssa
360	0.0234152	LPS Young-LPS Old	1	0.2	0.8	0.1	1.2	0.1	1.2	0.3	unnamed protein product	gi 12836667	4	mascot

# Samespots	Anova	Test Post- Hoc de Tukey	moyenne PBS Young	ecart-type	moyenne LPS Young	ecart-type	moyenne PBS Old	ecart-type	moyenne LPS Old	ecart-type	Nom des protéines majoritaires identifiées	Numéro d'accesion	Nombre de peptides	Algorithme de recherche
362	0.005485	PBS Old-LPS Young	1	0.1	0.8	0.1	1.3	0.0	1.2	0.3	Adenylate kinase 2, mitochondrial	gi 224471907	5	mascot
362	0.005485		1	0.1	0.8	0.1	1.3	0.0	1.2	0.3		gi 224471907	7	omssa
362	0.005485	LPS Young-LPS Old	1	0.1	0.8	0.1	1.3	0.0	1.2	0.3	carbonic anhydrase 3, isoform CRA_b	gi 148673185	12	mascot
362	0.005485		1	0.1	0.8	0.1	1.3	0.0	1.2	0.3		gi 148673185	15	omssa
375	0.0006505	LPS Young-LPS Old	1	0.04	0.8	0.1	1.1	0.1	1.0	0.1	peroxiredoxin-6 [Mus musculus]	gi 6671549	17	mascot
375	0.0006505	PBS Old-LPS Young	1	0.04	0.8	0.1	1.1	0.1	1.0	0.1		gi 3219774	15	omssa
375	0.0006505	PBS Young-LPS Young	1	0.04	0.8	0.1	1.1	0.1	1.0	0.1				
378	0.0120935	PBS Young-LPS Young	1	0.1	0.8	0.1	1.0	0.1	1.0	0.1	Chain A, Mouse C14a Glutathione-S-Transferase Mutant In Complex With S-Hexyl Glutathione	gi 126031446	3	omssa
378	0.0120935		1	0.1	0.8	0.1	1.0	0.1	1.0	0.1	Glutathione S-transferase kappa 1	gi 47116757	3	omssa
378	0.0120935	LPS Young-LPS Old	1	0.1	0.8	0.1	1.0	0.1	1.0	0.1	Glutathione S-transferase P 1	gi 121747	3	mascot
378	0.0120935		1	0.1	0.8	0.1	1.0	0.1	1.0	0.1	unnamed protein product	gi 12832492	6	omssa
378	0.0120935	PBS Old-LPS Young	1	0.1	0.8	0.1	1.0	0.1	1.0	0.1		gi 193703	8	mascot
379	0.0112227		1	0.1	0.8	0.0	1.0	0.1	1.0	0.1	Chain A, Mouse C14a Glutathione-S-Transferase Mutant In Complex With S-Hexyl Glutathione	gi 126031446	3	omssa
379	0.0112227		1	0.1	0.8	0.0	1.0	0.1	1.0	0.1	Glutathione S-transferase kappa 1	gi 47116757	3	omssa
379	0.0112227	LPS Young-LPS Old	1	0.1	0.8	0.0	1.0	0.1	1.0	0.1	Glutathione S-transferase P 1	gi 121747	3	mascot
379	0.0112227		1	0.1	0.8	0.0	1.0	0.1	1.0	0.1	unnamed protein product	gi 12832492	6	omssa
379	0.0112227	PBS Young-LPS Young	1	0.1	0.8	0.0	1.0	0.1	1.0	0.1		gi 193703	8	mascot
452	0.0111751	LPS Young-LPS Old	1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	carbonic anhydrase 3, isoform CRA_b	gi 148673185	4	mascot
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1		gi 148673185	7	omssa
452	0.0111751	PBS Old-LPS Young	1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	fibrinogen, alpha polypeptide, isoform CRA_a	gi 148683476	10	mascot
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1		gi 148683476	9	omssa
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	M2-type pyruvate kinase	gi 1405933	3	omssa
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	malic enzyme, supernatant, isoform CRA_a	gi 148694554	6	omssa
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	mCG126417, isoform CRA_a	gi 148698805	9	mascot
452	0.0111751		1	0.04	1.3	0.1	1.0	0.1	1.0	0.1		gi 148698805	9	omssa
452	0.0111751	PBS Young-LPS Young	1	0.04	1.3	0.1	1.0	0.1	1.0	0.1	NADP-dependent malic enzyme	gi 148694554	5	mascot

# Samespots	Anova	Test Post- Hoc de Tukey	moyenne PBS Young	ecart-type	moyenne LPS Young	ecart-type	moyenne PBS Old	ecart-type	moyenne LPS Old	ecart-type	Nom des protéines majoritaires identifiées	Numéro d'accension	Nombre de peptides	Algorithme de recherche
459	0.0042887		1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	Full=Haloacid dehalogenase-like	gi 81904469	2	omssa
459	0.0042887		1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	hydrolase domain-containing protein 3	gi 81904469	3	mascot
459	0.0042887		1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	nitrilase 1, isoform CRA_a	gi 148707145	7	omssa
459	0.0042887	PBS Old-LPS Young	1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	Omega-amidase NIT2	gi 81881899	2	mascot
459	0.0042887		1	0.1	1.2	0.1	0.8	0.4	0.9	0.1		gi 81881899	2	omssa
459	0.0042887	LPS Young-LPS Old	1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	unnamed protein product	gi 12833077	2	mascot
459	0.0042887		1	0.1	1.2	0.1	0.8	0.4	0.9	0.1	unnamed protein product	gi 12833077	8	omssa
469	0.0181072		1	0.2	0.9	0.1	0.6	0.1	0.6	0.3	Chain A, Mouse C14a Glutathione-S-Transferase Mutant In Complex With S-Hexyl Glutathione	gi 126031446	8	omssa
469	0.0181072	PBS Young-LPS Old	1	0.2	0.9	0.1	0.6	0.1	0.6	0.3	Glutathione S-transferase P 1	gi 121747	7	mascot
469	0.0181072		1	0.2	0.9	0.1	0.6	0.1	0.6	0.3	peroxiredoxin 1	gi 123230137	4	omssa
469	0.0181072	PBS Young-PBS Old	1	0.2	0.9	0.1	0.6	0.1	0.6	0.3		gi 547923	4	mascot
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	Glutathione synthetase	gi 1708057	15	mascot
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2		gi 1708057	14	omssa
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	gi 41688568	8	mascot
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2		gi 41688568	7	omssa
491	0.0403159	PBS Young-LPS Young	1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	pyruvate kinase isozymes R/L isoform 1	gi 153792131	4	mascot
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	S-adenosylmethionine synthase	gi 81902386	8	mascot
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	isoform type-1	gi 81902386	11	omssa
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2	unnamed protein product	gi 74146312	6	omssa
491	0.0403159		1	0.2	1.3	0.1	1.1	0.1	1.1	0.2				
492	0.0401659		1	0.2	1.3	0.1	1.1	0.1	1.2	0.1	S-adenosylmethionine synthase	gi 81902386	10	mascot
492	0.0401659		1	0.2	1.3	0.1	1.1	0.1	1.2	0.1	isoform type-1	gi 81902386	14	omssa
492	0.0401659		1	0.2	1.3	0.1	1.1	0.1	1.2	0.1	Selenide, water dikinase 2	gi 172046691	4	mascot
492	0.0401659		1	0.2	1.3	0.1	1.1	0.1	1.2	0.1		gi 172046691	4	omssa
492	0.0401659	PBS Young-LPS Young	1	0.2	1.3	0.1	1.1	0.1	1.2	0.1	Selenium-binding protein 1	gi 148840436	3	mascot
492	0.0401659		1	0.2	1.3	0.1	1.1	0.1	1.2	0.1		gi 148840436	5	omssa
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1	fructose-bisphosphate aldolase B	gi 15723268	3	mascot
507	0.0253936	PBS Young-PBS Old	1	0.04	0.8	0.1	0.8	0.1	0.8	0.1	Glycerol-3-phosphate dehydrogenase	gi 121557	3	mascot
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1	[NAD+], cytoplasmic	gi 121557	4	omssa
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1	Glycine N-methyltransferase	gi 15679953	6	mascot
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1		gi 15679953	9	omssa
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1	L-lactate dehydrogenase A chain	gi 126048	3	mascot
507	0.0253936		1	0.04	0.8	0.1	0.8	0.1	0.8	0.1		gi 126048	4	omssa
510	0.0138934	PBS Young-LPS Young	1	0.1	0.7	0.1	0.7	0.2	0.8	0.1	Expressed sequence AU021838	gi 10946854	3	mascot
510	0.0138934	PBS Young-PBS Old	1	0.1	0.7	0.1	0.7	0.2	0.8	0.1	Glycerol-3-phosphate dehydrogenase	gi 121557	22	mascot
510	0.0138934		1	0.1	0.7	0.1	0.7	0.2	0.8	0.1	[NAD+], cytoplasmic	gi 121557	22	omssa
510	0.0138934		1	0.1	0.7	0.1	0.7	0.2	0.8	0.1		gi 15679953	3	omssa
510	0.0138934		1	0.1	0.7	0.1	0.7	0.2	0.8	0.1	Glycine N-methyltransferase	gi 15679953	3	mascot

Annexe 5a : Image de gel présentant les spots identifiés comme différentiels dans le projet concernant l'étude des rythmes circadiens et liste des protéines identifiées dans ces spots de gel

(Analyse selon le test du Cosinor)

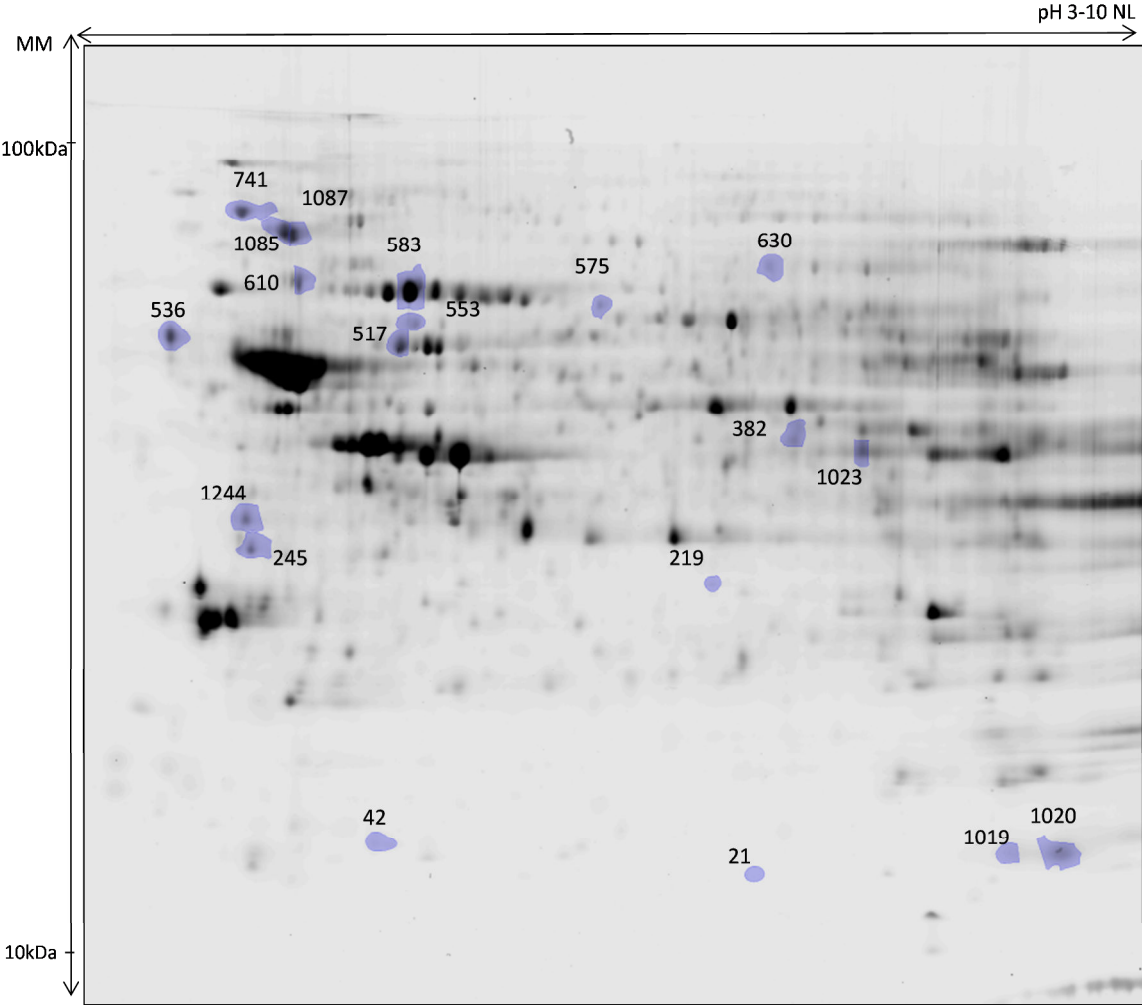


#Samespot	MOY T5	Ecart-type	MOY T11	Ecart-type	MOY T17	Ecart-type	MOY T23	Ecart-type	Algorithme de recherche	Protéines majoritaires détectées dans les spots de gel différentiels	Numéro d'accession	Nombre total de peptides Mascot	Nombre total de spectres Mascot	Nombre total de peptides Omssa	Nombre total de spectre Omssa
21	1	0.2	0.9	0.1	0.8	0.1	1.0	0.2	Mascot+Omssa	D-dopachrome decarboxylase	sp O35215	8	20	8	34
22	1	0.1	0.9	0.1	0.8	0.1	0.8	0.05	Mascot+Omssa	D-dopachrome decarboxylase	sp O35215	6	7	8	9
22	1	0.1	0.9	0.1	0.8	0.1	0.8	0.05	Mascot+Omssa	SH3 domain-binding glutamic acid-rich-like protein 2	sp Q8BG73	3	4	7	9
82	1	0.2	1.1	0.0	1.2	0.2	1.1	0.1	Mascot+Omssa	Beta-synuclein	sp Q91Z73	3	3	2	-
82	1	0.2	1.1	0.0	1.2	0.2	1.1	0.1	Mascot+Omssa	DNA-directed RNA polymerases I, II, and III subunit	sp Q923G2	2	2	2	-
167	1	0.2	1.1	0.2	1.3	0.2	1.1	0.2	Mascot+Omssa	Triosephosphate isomerase	sp P17751	3	3	2	3
167	1	0.2	1.1	0.2	1.3	0.2	1.1	0.2	Mascot+Omssa	Glutathione S-transferase Mu 5	sp P48774	1	1	1	1
167	1	0.2	1.1	0.2	1.3	0.2	1.1	0.2	Mascot+Omssa	Programmed cell death protein 10	sp Q8VE70	1	1	1	1
198	1	0.1	1.0	0.1	1.2	0.1	1.0	0.1	Mascot+Omssa	Carbonic anhydrase 1	sp P13634	6	7	6	7
198	1	0.1	1.0	0.1	1.2	0.1	1.0	0.1	Mascot+Omssa	Phosphoglycerate mutase 1	sp Q9DBJ1	5	5	9	10
212	1	0.1	1.1	0.1	1.3	0.2	1.0	0.2	Mascot+Omssa	Carbonic anhydrase 2	sp P00920	5	5	5	5
212	1	0.1	1.1	0.1	1.3	0.2	1.0	0.2	Mascot+Omssa	Phosphoglycerate mutase 1	sp Q9DBJ1	4	4	3	3
212	1	0.1	1.1	0.1	1.3	0.2	1.0	0.2	Mascot+Omssa	Proteasome subunit alpha type-4	sp Q9R1P0	2	2	2	2
216	1	0.1	1.1	0.1	1.1	0.1	1.0	0.1	Mascot+Omssa	Carbonic anhydrase 2	sp P00920	6	9	6	8
216	1	0.1	1.1	0.1	1.1	0.1	1.0	0.1	Mascot+Omssa	Eukaryotic translation initiation factor 4H	sp Q9WUK2	15	52	10	40
219	1	0.1	0.9	0.1	0.9	0.1	1.0	0.1	Mascot+Omssa	Proteasome subunit alpha type-1	sp Q9R1P4	12	12	9	-
315	1	0.1	1.0	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	Elongation factor 2	sp P58252	6	7	6	-
315	1	0.1	1.0	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	NAD-dependent protein deacetylase sirtuin-2	sp Q8VDQ8	7	7	7	-
315	1	0.1	1.0	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	Transaldolase	sp Q93092	10	11	9	-
382	1	0.02	1.1	0.04	1.1	0.04	1.1	0.03	Mascot+Omssa	Beta-centractin	sp Q8R5C5	1	1	1	1
382	1	0.02	1.1	0.04	1.1	0.04	1.1	0.03	Mascot+Omssa	Alpha-centractin	sp P61164	23	69	19	82
382	1	0.02	1.1	0.04	1.1	0.04	1.1	0.03	Mascot+Omssa	2',3'-cyclic-nucleotide 3'-phosphodiesterase	sp P16330	11	18	10	19
382	1	0.02	1.1	0.04	1.1	0.04	1.1	0.03	Mascot+Omssa	Glutamine synthetase	sp P15105	9	15	12	23
382	1	0.02	1.1	0.04	1.1	0.04	1.1	0.03	Mascot+Omssa	Phosphoglycerate kinase 1	sp P09411	13	21	12	23
401	1	0.1	0.9	0.1	0.9	0.1	0.9	0.01	Mascot+Omssa	Keratin, type II cytoskeletal 1	sp P04264	11	11	9	9
401	1	0.1	0.9	0.1	0.9	0.1	0.9	0.01	Mascot+Omssa	Keratin, type I cytoskeletal 10	sp P13645	9	10	7	8
401	1	0.1	0.9	0.1	0.9	0.1	0.9	0.01	Mascot+Omssa	Ornithine aminotransferase, mitochondrial	sp P29758	8	9	9	10

# Samespot	MOY T5	Ecart-type	MOY T11	Ecart-type	MOY T17	Ecart-type	MOY T23	Ecart-type	Algorithme de recherche	Protéines majoritaires détectées dans les spots de gel différentiels	Numéro d'accession	Nombre total de peptides Mascot	Nombre total de spectres Mascot	Nombre total de peptides Omssa	Nombre total de spectre Omssa
412	1	0.1	1.0	0.03	0.9	0.1	0.9	0.1	Mascot+Omssa	Actin-related protein 3	sp Q99JY9	3	3	9	11
412	1	0.1	1.0	0.03	0.9	0.1	0.9	0.1	Mascot+Omssa	Spliceosome RNA helicase Ddx39b	sp Q9Z1N5	9	10	10	14
412	1	0.1	1.0	0.03	0.9	0.1	0.9	0.1	Mascot+Omssa	Alpha-enolase	sp P17182	4	4	6	7
412	1	0.1	1.0	0.03	0.9	0.1	0.9	0.1	Mascot+Omssa	Cytochrome b-c1 complex subunit 1, mitochondrial	sp Q9CZ13	13	15	20	25
417	1	0.1	1.0	0.1	1.0	0.1	0.9	0.03	Mascot	Alpha-internexin	sp P46660	1	1	-	-
417	1	0.1	1.0	0.1	1.0	0.1	0.9	0.03	Mascot-Omssa	Cytochrome b-c1 complex subunit 1, mitochondrial	sp Q9CZ13	4	2	3	5
417	1	0.1	1.0	0.1	1.0	0.1	0.9	0.03	Mascot-Omssa	Spliceosome RNA helicase Ddx39b	sp Q9Z1N5	2	2	2	2
512	1	0.1	0.9	0.1	0.9	0.1	0.9	0.1	Mascot+Omssa	Glucose-6-phosphate 1-dehydrogenase X	sp Q00612	5	5	6	6
512	1	0.1	0.9	0.1	0.9	0.1	0.9	0.1	Mascot+Omssa	Synapsin-2	sp Q64332	4	4	6	8
540	1	0.1	0.9	0.1	0.9	0.1	0.9	0.05	Mascot+Omssa	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	sp Q76M23	27	92	24	108
540	1	0.1	0.9	0.1	0.9	0.1	0.9	0.05	Mascot+Omssa	ATP synthase subunit beta, mitochondrial	sp P56480	12	18	11	18
540	1	0.1	0.9	0.1	0.9	0.1	0.9	0.05	Mascot+Omssa	Rab GDP dissociation inhibitor alpha	sp P50396	25	85	26	117
554	1	0.1	1.0	0.1	1.1	0.1	1.0	0.05	Mascot+Omssa	Heterogeneous nuclear ribonucleoprotein K	sp P61979	12	12	7	-
582	1	0.1	0.8	0.0	0.8	0.2	0.9	0.1	Mascot-Omssa	Serum albumin	sp P07724	11	11	7	8
595	1	0.1	1.0	0.1	0.9	0.1	0.9	0.0	Omssa	Glycerol-3-phosphate dehydrogenase, mitochondrial	sp Q64521	-	-	1	1
623	1	0.1	1.0	0.1	1.1	0.0	1.1	0.0	Mascot+Omssa	Vesicle-fusing ATPase	sp P46460	8	8	11	12
630	1	0.1	1.2	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	Radixin	sp P26043	9	11	11	16
630	1	0.1	1.2	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	Vesicle-fusing ATPase	sp P46460	29	54	36	84
630	1	0.1	1.2	0.1	1.2	0.2	1.1	0.1	Mascot+Omssa	Dihydropyrimidinase-related protein 1	sp P97427	9	9	10	15
677	1	0.1	1.0	0.1	0.9	0.1	0.9	0.05	Mascot+Omssa	Mitochondrial inner membrane protein	sp Q8CAQ8	42	155	40	167
1019	1	0.3	1.0	0.2	1.3	0.3	1.0	0.1	Mascot+Omssa	Hemoglobin subunit beta-1	sp P02088	13	123	13	172
1056	1	0.1	1.0	0.04	1.0	0.03	0.9	0.04	Mascot+Omssa	Alpha-enolase	sp P17182	15	30	13	33
1056	1	0.1	1.0	0.04	1.0	0.03	0.9	0.04	Mascot+Omssa	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	sp Q91VD5	18	88	18	107
1056	1	0.1	1.0	0.04	1.0	0.03	0.9	0.04	Mascot+Omssa	Tubulin alpha-1B chain	sp P05213	12	28	16	40
1056	1	0.1	1.0	0.04	1.0	0.03	0.9	0.04	Mascot+Omssa	Tubulin alpha-1C chain	sp P68373	12	1	16	1
1056	1	0.1	1.0	0.04	1.0	0.03	0.9	0.04	Mascot+Omssa	Tubulin alpha-4A chain	sp P68368	1	2	2	4
1060	1	0.1	1.0	0.04	0.9	0.1	0.9	0.04	Mascot-Omssa	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	sp P62880	4	5	5	6
1060	1	0.1	1.0	0.04	0.9	0.1	0.9	0.04	Mascot-Omssa	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	sp Q9D051	8	17	8	20
1224	1	0.1	0.9	0.1	0.9	0.1	0.9	0.1	Mascot+Omssa	ATP synthase subunit beta, mitochondrial	sp P56480	12	12	11	-
1224	1	0.1	0.9	0.1	0.9	0.1	0.9	0.1	Mascot+Omssa	Secernin-1	sp Q9CZC8	9	13	7	-

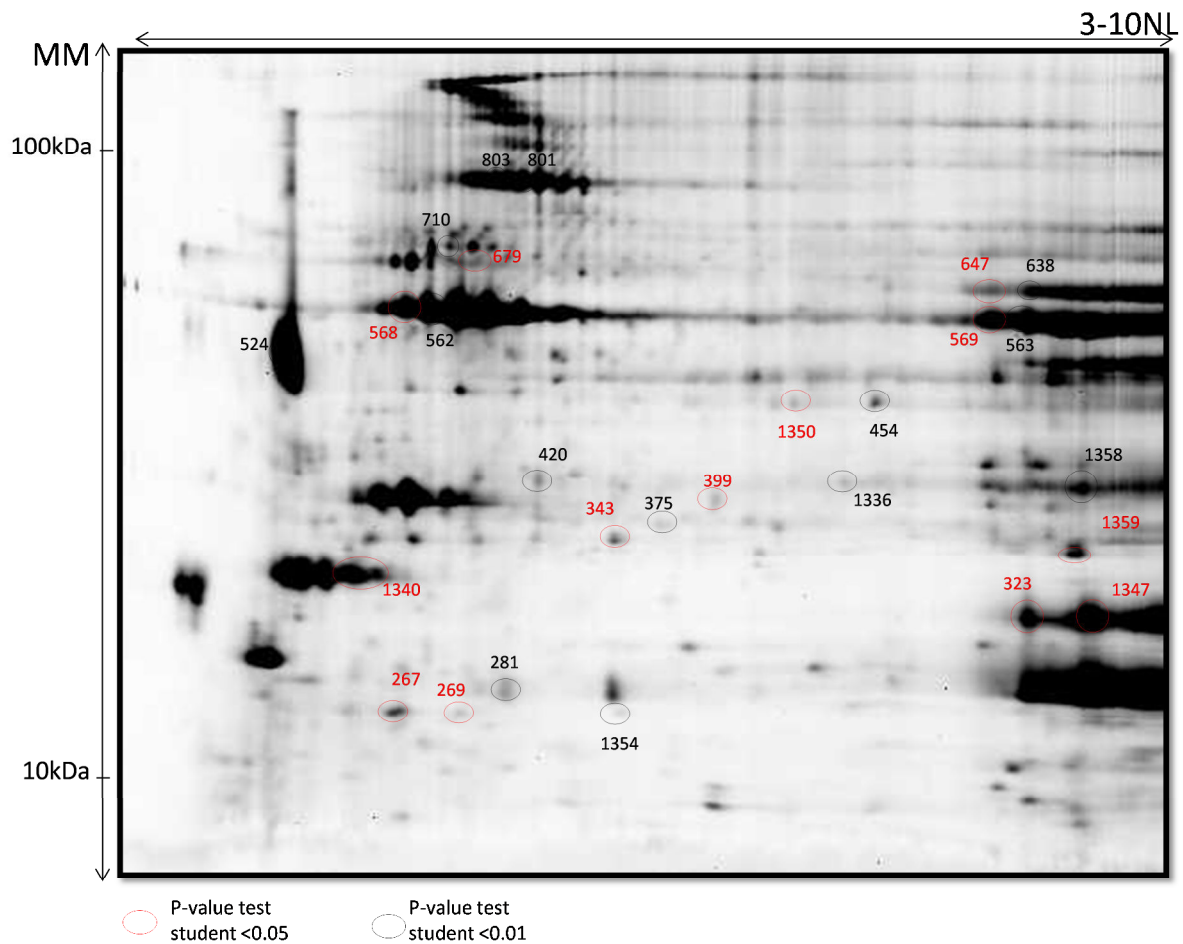
Annexe 5b : Image de gel présentant les spots identifiés comme différentiels dans le projet concernant l'étude des rythmes circadiens et liste des protéines identifiées dans ces spots de gel

(Analyse selon le test ANOVA à deux facteurs suivie du test post hoc de Tukey)



# Samespots	Test Post hoc de Tukey	Moy t5	Ecart-type	Moy t17	Ecart-type	Moy rt5	Ecart-type	Moy rt17	ecart-type	Algorithme de recherche	Protéines majoritaires détectées dans les spots de gel différentiels	Protein accession numbers	Nombre total de peptide Mascot	Nombre total de spectres Mascot	Nombre total de peptide Omssa	Nombre total de spectres Omssa
21	T17-rT17 T17-rT5	0.004	0.096	-0.101	0.044	0.053	0.086	0.032	0.073	Mascot+Omssa	D-dopachrome decarboxylase	sp O35215	8	20	8	34
42	T17-rT17 T5-rT17 T17-rT5 T5-rT5	0.076	0.045	0.043	0.028	-0.045	0.033	-0.022	0.034	Mascot+Omssa	Fatty acid-binding protein, brain	sp P51880	7	25	7	26
219	T17-rT17	0.008	0.026	-0.046	0.054	-0.001	0.025	0.017	0.033	Mascot+Omssa	Proteasome subunit alpha type-1	sp Q9R1P4	12	12	9	-
245	T5-rT5	0.036	0.042	0.021	0.019	-0.036	0.047	-0.015	0.030	Mascot+Omssa	Annexin A5	sp P48036	25	112	21	104
										Mascot+Omssa	Ubiquitin thioesterase OTUB1	sp Q7TQI3	11	30	13	41
382	T5-rT17 T17-rT5 T5-rT5	-0.062	0.009	-0.030	0.016	0.003	0.025	-0.005	0.013	Mascot+Omssa	Beta-centractin	sp Q8R5C5	1	1	1	1
										Mascot+Omssa	Alpha-centractin	sp P61164	23	69	19	82
										Mascot+Omssa	2',3'-cyclic-nucleotide 3'-phosphodiesterase	sp P16330	11	18	10	19
										Mascot+Omssa	Phosphoglycerate kinase 1	sp P09411	13	21	12	23
517	T5-rT17	-0.049	0.046	-0.008	0.030	-0.001	0.020	0.005	0.024	Mascot+Omssa	60 kDa heat shock protein, mitochondrial	sp P63038	28	36	27	-
										Mascot+Omssa	Alpha-internexin	sp P46660	19	21	11	-
536	T17-rT5 T5-rT5	0.010	0.012	-0.001	0.037	-0.061	0.028	-0.028	0.028	Mascot+Omssa	Calreticulin	sp P14211	21	76	20	85
553	rT5-rT17	0.002	0.021	0.009	0.027	0.034	0.012	-0.005	0.011	Mascot+Omssa	Heterogeneous nuclear ribonucleoprotein K	sp P61979	19	124	18	137
575	rT5-rT17	-0.027	0.027	-0.004	0.035	0.016	0.026	-0.036	0.031	Mascot+Omssa	Dihydropyrimidinase-related protein 2	sp O08553	16	20	16	-
583	T17-rT5 T5-rT5	0.046	0.015	0.044	0.013	0.000	0.019	0.024	0.021	Mascot+Omssa	Stress-70 protein, mitochondrial	sp P38647	12	27	13	33
										Mascot+Omssa	Heat shock cognate 71 kDa protein	sp P63017	33	237	33	289
										Mascot+Omssa	V-type proton ATPase catalytic subunit A	sp P50516	26	61	27	71
610	T17-rT5	0.046	0.015	0.044	0.013	0.000	0.019	0.024	0.021	Mascot+Omssa	ATP synthase subunit beta, mitochondrial	sp P56480	12	21	12	22
										Mascot+Omssa	78 kDa glucose-regulated protein	sp P20029	27	61	26	73
										Mascot+Omssa	Spectrin alpha chain, non-erythrocytic 1	sp P16546	15	21	18	25
630	T5-rT17 T5-rT5	-0.125	0.066	-0.038	0.071	-0.024	0.025	-0.027	0.040	Mascot+Omssa	Radixin	sp P26043	9	11	11	16
										Mascot+Omssa	Vesicle-fusing ATPase	sp P46460	29	54	36	84
										Mascot+Omssa	Dihydropyrimidinase-related protein 1	sp P97427	9	9	10	15
741	T17-rT5	0.009	0.028	0.025	0.039	-0.022	0.015	-0.005	0.015	Mascot+Omssa	Endoplasmic reticulum chaperone protein	sp P08113	27	56	26	64
										Mascot+Omssa	Neurofilament medium polypeptide	sp P08553	12	13	6	6
1019	T5-rT5	-0.031	0.153	0.102	0.093	0.176	0.050	0.064	0.033	Mascot+Omssa	Hemoglobin subunit beta-1	sp P02088	13	123	13	172
1020	rT5-rT17 T17-rT5	0.063	0.090	0.053	0.089	0.167	0.029	0.046	0.037	Mascot+Omssa	Hemoglobin subunit beta-1	sp P02088	12	177	13	258
1023	T5-rT5	-0.017	0.022	0.002	0.033	0.030	0.014	0.013	0.021	Mascot+Omssa	Fructose-bisphosphate aldolase C	sp P05063	21	69	25	83
										Mascot+Omssa	Septin-5	sp Q9Z2Q6	18	48	17	65
1085	T17-rT5	-0.009	0.030	0.006	0.027	-0.050	0.032	-0.015	0.017	Mascot+Omssa	Heat shock protein HSP 90-alpha	sp P07901	40	135	42	174
										Mascot+Omssa	Heat shock protein HSP 90-beta	sp P11499	12	22	15	27
1087	rT5-rT17 T17-rT17 T17-rT5 T5-rT5	0.001	0.020	0.013	0.017	-0.071	0.017	-0.031	0.029	Mascot+Omssa	Heat shock protein HSP 90-alpha	sp P07901	39	145	40	158
										Mascot+Omssa	Heat shock protein HSP 90-beta	sp P11499	22	61	23	71
1244	T5-rT17	0.005	0.058	-0.020	0.023	-0.034	0.016	-0.060	0.020	Mascot+Omssa	Carbonic anhydrase-related protein	sp P28651	10	25	10	38

Annexe 6 : Image de gel présentant les spots identifiés comme différentiels dans le projet concernant l'étude de l'ours à l'état actif contrairement à l'état inactif et liste des protéines identifiées dans ces spots de gel



#Same-spot	Moyenne été (actif)	Ecart-type	Moyenne hiver (inactif)	Ecart-type	Algorithme de recherche	Nom des protéines majoritaires identifiées	Numéro d'accension	Nombre total de peptide Omssa	Nombre total de peptide Mascot
267	1.3	0.1	0.9	0.2	Omssa-Mascot	Actin, cytoplasmic 1 OS=Canis familiaris	sp O18840	2	2
267	1.3	0.1	0.9	0.2	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	2	2
267	1.3	0.1	0.9	0.2	Omssa-Mascot	Cytochrome c oxidase subunit 5A, mitochondrial OS=Nycticebus coucang	sp B0VYY2	1	2
269	1.4	0.5	0.6	0.2	Omssa-Mascot	Actin, cytoplasmic 1 OS=Canis familiaris	sp O18840	2	2
269	1.4	0.5	0.6	0.2	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	2	2
269	1.4	0.5	0.6	0.2	Omssa-Mascot	Cytochrome c oxidase subunit 5A, mitochondrial OS=Nycticebus coucang	sp B0VYY2	1	2
323	0.8	0.2	1.2	0.1	Mascot	Alpha-S2-casein OS=Bos taurus	sp P02663	-	1
343	1.4	0.4	0.9	0.2	Mascot	Protein DJ-1 OS=Rattus norvegicus	sp O88767	-	1
343	1.4	0.4	0.9	0.2	Omssa-Mascot	ATP synthase subunit d, mitochondrial OS=Bos taurus	sp P13620	2	1
399	0.8	0.1	1.2	0.3	Omssa	Peroxiredoxin-6 OS=Mus musculus	sp O08709	3	-
399	0.8	0.1	1.2	0.3	Omssa-Mascot	Peroxiredoxin-6 OS=Homo sapiens	sp P30041	3	3
568	0.8	0.1	1.0	0.1	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	21	23
568	0.8	0.1	1.0	0.1	Omssa-Mascot	Actin, cytoplasmic 1 OS=Canis familiaris	sp O18840	13	14
569	0.8	0.2	1.1	0.1	Omssa	Aspartate aminotransferase, cytoplasmic OS=Bos taurus	sp P33097	8	-
569	0.8	0.2	1.1	0.1	Omssa	Aspartate aminotransferase, cytoplasmic OS=Equus caballus	sp P08906	9	-
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	10	7
647	0.8	0.2	1.3	0.2	Omssa	Beta-enolase OS=Bos taurus	sp Q3ZC09	16	-
647	0.8	0.2	1.3	0.2	Omssa-Mascot	Beta-enolase OS=Sus scrofa	sp Q1KYT0	16	15
647	0.8	0.2	1.3	0.2	Omssa-Mascot	Beta-enolase OS=Rattus norvegicus	sp P15429	16	15
679	1.3	0.2	0.9	0.1	Mascot	Desmin OS=Bos taurus	sp O62654	-	21
679	1.3	0.2	0.9	0.1	Mascot	Desmin OS=Mesocricetus auratus	sp P02541	-	20
679	1.3	0.2	0.9	0.1	Omssa	Desmin OS=Canis familiaris	sp Q5XFN2	26	-
679	1.3	0.2	0.9	0.1	Omssa-Mascot	ATP synthase subunit beta, mitochondrial OS=Rattus norvegicus	sp P10719	12	11
1340	1.4	0.2	1.0	0.2	Mascot	Actin, cytoplasmic 1 OS=Canis familiaris	sp O18840	-	3
1340	1.4	0.2	1.0	0.2	Omssa	Troponin T, fast skeletal muscle OS=Bos taurus	sp Q8MKI3	1	-
1340	1.4	0.2	1.0	0.2	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	4	3
1340	1.4	0.2	1.0	0.2	Omssa-Mascot	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform OS=Homo sapiens	sp P10916	2	2
1340	1.4	0.2	1.0	0.2	Omssa-Mascot	Myosin regulatory light chain 2, skeletal muscle isoform type 2 OS=Orctolagus cuniculus	sp P02608	1	1

#Same-spot	Moyenne été (actif)	Ecart-type	Moyenne hiver (inactif)	Ecart-type	Algorithme de recherche	Nom des protéines majoritaires identifiées	Numéro d'accension	Nombre total de peptide Omssa	Nombre total de peptide Mascot
1347	0.9	0.2	1.3	0.1	Mascot	Hemoglobin subunit alpha OS=Ursus maritimus	sp P68235	-	2
1347	0.9	0.2	1.3	0.1	Mascot	Creatine kinase M-type OS=Canis familiaris	sp P05123	-	3
1347	0.9	0.2	1.3	0.1	Mascot	Myoglobin OS=Meles meles	sp P02157	-	5
1347	0.9	0.2	1.3	0.1	Omssa	Myoglobin OS=Lutra lutra	sp P11343	6	-
1347	0.9	0.2	1.3	0.1	Omssa	Glyceraldehyde-3-phosphate dehydrogenase OS=Mustela putorius furo	sp A3FKF7	1	-
1347	0.9	0.2	1.3	0.1	Omssa	Myoglobin OS=Globicephala melas	sp P02174	5	-
1347	0.9	0.2	1.3	0.1	Omssa	Histone-lysine N-methyltransferase 2D OS=Homo sapiens	sp O14686	1	-
1347	0.9	0.2	1.3	0.1	Omssa	Hemoglobin subunit alpha OS=Ursus maritimus	sp P68235	2	-
1347	0.9	0.2	1.3	0.1	Omssa-Mascot	Myoglobin OS=Eschrichtius gibbosus	sp P02177	7	4
1347	0.9	0.2	1.3	0.1	Omssa-Mascot	Hemoglobin subunit beta OS=Ursus maritimus	sp P68011	1	2
1347	0.9	0.2	1.3	0.1	Omssa-Mascot	L-lactate dehydrogenase A chain OS=Mus musculus	sp P06151	1	1
1350	0.9	0.2	1.2	0.2	Mascot	Malate dehydrogenase, cytoplasmic OS=Felis catus	sp Q7YRU4	-	3
1359	0.8	0.2	1.2	0.3	Omssa-Mascot	Alpha-crystallin B chain OS=Bos taurus	sp P02510	5	8
281	0.9	0.2	1.2	0.3	Omssa	Fatty acid-binding protein, heart OS=Sus scrofa	sp O02772	4	-
281	0.9	0.2	1.2	0.3	Omssa-Mascot	Fatty acid-binding protein, heart OS=Spermophilus tridecemlineatus	sp Q99P61	3	4
281	0.9	0.2	1.2	0.3	Omssa-Mascot	Fatty acid-binding protein, heart OS=Homo sapiens	sp P05413	4	3
375	1.0	0.1	1.1	0.1	Mascot	Protein DJ-1 OS=Rattus norvegicus	sp O88767	-	1
420	1.2	0.3	0.9	0.1	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	13	14
454	0.9	0.2	1.2	0.2	Omssa	Troponin T, fast skeletal muscle OS=Bos taurus	sp Q8MKI3	1	-
454	0.9	0.2	1.2	0.2	Omssa-Mascot	Malate dehydrogenase, cytoplasmic OS=Felis catus	sp Q7YRU4	3	5
524	0.9	0.2	1.1	0.1	Omssa-Mascot	Tropomyosin beta chain OS=Mus musculus	sp P58774	15	16
524	0.9	0.2	1.1	0.1	Omssa-Mascot	Tropomyosin alpha-1 chain OS=Rattus norvegicus	sp P04692	10	13
562	0.8	0.2	1.0	0.1	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	26	23
563	0.8	0.2	1.1	0.1	Omssa	Fructose-bisphosphate aldolase A OS=Homo sapiens	sp P04075	11	-
563	0.8	0.2	1.1	0.1	Omssa-Mascot	Creatine kinase M-type OS=Canis familiaris	sp P05123	20	17
563	0.8	0.2	1.1	0.1	Omssa-Mascot	Creatine kinase M-type OS=Rattus norvegicus	sp P00564	18	14
563	0.8	0.2	1.1	0.1	Omssa-Mascot	Fructose-bisphosphate aldolase A OS=Orctolagus cuniculus	sp P00883	11	12
563	0.8	0.2	1.1	0.1	Omssa-Mascot	Creatine kinase M-type OS=Homo sapiens	sp P06732	17	14
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Aspartate aminotransferase, cytoplasmic OS=Sus scrofa	sp P00503	10	8
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Fructose-bisphosphate aldolase A OS=Homo sapiens	sp P04075	11	11
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Fructose-bisphosphate aldolase A OS=Orctolagus cuniculus	sp P00883	11	11
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Creatine kinase M-type OS=Canis familiaris	sp P05123	18	17
569	0.8	0.2	1.1	0.1	Omssa-Mascot	Creatine kinase M-type OS=Rattus norvegicus	sp P00564	16	15
638	0.9	0.2	1.2	0.2	Omssa	Beta-enolase OS=Bos taurus	sp Q3ZC09	18	-
638	0.9	0.2	1.2	0.2	Omssa-Mascot	Beta-enolase OS=Sus scrofa	sp Q1KYT0	19	16
638	0.9	0.2	1.2	0.2	Omssa-Mascot	Beta-enolase OS=Rattus norvegicus	sp P15429	18	16
710	0.8	0.3	1.3	0.3	Omssa	Desmin OS=Canis familiaris	sp Q5XFN2	21	-
710	0.8	0.3	1.3	0.3	Omssa-Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	10	11
710	0.8	0.3	1.3	0.3	Omssa-Mascot	Desmin OS=Bos taurus	sp O62654	21	21

#Same-spot	Moyenne été (actif)	Ecart-type	Moyenne hiver (inactif)	Ecart-type	Algorithme de recherche	Nom des protéines majoritaires identifiées	Numéro d'accension	Nombre total de peptide Omssa	Nombre total de peptide Mascot
801	0.9	0.2	1.2	0.3	Mascot	Myosin-1 OS=Bos taurus	sp Q9BE40	-	13
801	0.9	0.2	1.2	0.3	Mascot	Myosin-2 OS=Bos taurus	sp Q9BE41	-	14
801	0.9	0.2	1.2	0.3	Omssa	Myosin-2 OS=Canis familiaris	sp Q076A7	13	-
801	0.9	0.2	1.2	0.3	Omssa	Myosin-2 OS=Equus caballus	sp Q8MJV1	13	-
801	0.9	0.2	1.2	0.3	Omssa-Mascot	Myosin-7	sp P49824	12	10
801	0.9	0.2	1.2	0.3	Omssa-Mascot	Heat shock cognate 71 kDa protein OS=Equus caballus	sp A2Q0Z1	19	13
801	0.9	0.2	1.2	0.3	Omssa-Mascot	Serum albumin OS=Felis catus	sp P49064	14	12
801	0.9	0.2	1.2	0.3	Omssa-Mascot	Serum albumin OS=Canis familiaris	sp P49822	19	15
803	0.8	0.2	1.2	0.4	Omssa-Mascot	Heat shock cognate 71 kDa protein	sp A2Q0Z1	19	17
803	0.8	0.2	1.2	0.4	Omssa-Mascot	Serum albumin OS=Felis catus	sp P49064	11	13
803	0.8	0.2	1.2	0.4	Omssa-Mascot	Serum albumin OS=Canis familiaris	sp P49822	15	18
1336	0.9	0.2	1.2	0.2	Mascot	Actin, cytoplasmic 1 OS=Canis familiaris	sp O18840	-	1
1336	0.9	0.2	1.2	0.2	Mascot	Actin, alpha skeletal muscle OS=Homo sapiens	sp P68133	-	1
1336	0.9	0.2	1.2	0.2	Mascot	Myosin-2 OS=Bos taurus	sp Q9BE41	-	2
1336	0.9	0.2	1.2	0.2	Mascot	Proteasome subunit beta type-7 OS=Sus scrofa	sp A1XQU1	-	2
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Proteasome subunit alpha type-3 OS=Mus musculus	sp O70435	1	1
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Heat shock protein beta-1 OS=Canis familiaris	sp P42929	2	2
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Triosephosphate isomerase OS=Canis familiaris	sp P54714	4	3
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Myosin-1 OS=Bos taurus	sp Q9BE40	1	2
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Myosin-2 OS=Canis familiaris	sp Q076A7	1	2
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Myosin-2 OS=Equus caballus	sp Q8MJV1	1	2
1336	0.9	0.2	1.2	0.2	Omssa-Mascot	Pyruvate kinase PKM OS=Felis catus	sp P11979	1	1
1354	0.9	0.2	1.2	0.2	Mascot	Tubulin alpha-4A chain OS=Homo sapiens	sp P68366	-	1
1354	0.9	0.2	1.2	0.2	Omssa	Fatty acid-binding protein, heart OS=Sus scrofa	sp O02772	5	-
1354	0.9	0.2	1.2	0.2	Omssa	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5 OS=Bos taurus	sp P23935	1	-
1354	0.9	0.2	1.2	0.2	Omssa-Mascot	Fatty acid-binding protein, heart OS=Homo sapiens	sp P05413	6	5
1354	0.9	0.2	1.2	0.2	Omssa-Mascot	Fatty acid-binding protein, heart OS=Spermophilus tridecemlineatus	sp Q99P61	4	4
1358	0.9	0.2	1.1	0.1	Omssa-Mascot	Triosephosphate isomerase OS=Canis familiaris	sp P54714	14	14



Marine PLUMEL

Optimisations des stratégies analytiques quantitatives en protéomique : application à l'étude des réponses adaptatives du métabolisme chez divers organismes



L'analyse protéomique consiste à caractériser l'ensemble des protéines exprimées dans une cellule, dans un état et à un temps donné (le protéome). L'analyse protéomique tend aujourd'hui à fournir des informations quantitatives, cependant, pour pouvoir générer des données fiables et sensibles, des optimisations méthodologiques doivent être apportées à chaque problématique biologique.

L'objectif de cette thèse a été d'adapter les stratégies protéomiques quantitatives existantes à diverses problématiques biologiques. Ce travail de thèse a ainsi contribué à apporter des résultats originaux concernant l'étude de désordres métaboliques. Il a également permis la mise au point d'une méthode de suivi du statut reproducteur de la tortue Luth. Finalement, il a permis d'évaluer la réponse d'une lignée cellulaire hépatique à une intoxication chronique à l'acide valproïque. Ces données alimenteront des algorithmes de prédiction de toxicité hépatique chronique au travers du consortium NOTOX (EU).

Mots-clés : Analyse protéomique, quantification, spectrométrie de masse

Proteomics consists in the characterization of all the proteins expressed in a physiological state and at a given time (the proteome). Today, proteomics tends to bring quantitative information. However, in order to generate reliable, sensitive and robust data, whatever the biological question, methodological improvements need to be done.

The aim of this PhD thesis was to apply and adapt quantitative proteomics strategies to specific biological issues as there is no universal quantifying method. This PhD thesis contributed to bringing original results concerning the study of metabolic disorders. It has also contributed to setting-up a targeted approach in order to quantify the level of reproductive effort of Leatherback Turtles. Finally, this PhD thesis also contributed, through the consortium NOTOX (EU), to the evaluation of physiological answers of hepatic cell lines exposed to valproic acid. These data will be used in order to generate hepatotoxicity prediction algorithms.

Key-words: Proteomics, quantification, mass spectrometry