

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES
Institut Pluridisciplinaire Hubert Curien, UMR 7178

THÈSE présentée par :

Nina GUILLAUMOT

soutenue le : **25 septembre 2017**

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

Discipline/ Spécialité : Chimie/Chimie Analytique

**Nouvelles applications et opportunités
en protéomique**

THÈSE dirigée par :

Mr. VAN DORSSELAER Alain
Mme SCHAEFFER-REISS Christine

Directeur de recherche, CNRS, Strasbourg
Ingénieur de recherche, CNRS, Université de Strasbourg

RAPPORTEURS :

Mr BULET Philippe
Mr RIVIERE Michel

Directeur de recherche, CNRS, Grenoble
Directeur de recherche, CNRS, Toulouse

A ma mère

A mon père

A Martin

« L'esprit est comme un parachute, il faut qu'il soit ouvert pour qu'il fonctionne » Dalai Lama

Remerciements

En tout premier lieu, je tiens à remercier Sarah Cianférani et Alain Van Dorsselaer de m'avoir ouvert les portes du LSMBO et offert la possibilité de travailler dans un environnement de grande qualité. J'adresse mes remerciements particuliers à Alain pour son encadrement tout au long de cette thèse.

Merci à Christine Schaeffer qui a dirigé ces travaux thèse, merci pour votre patience et votre disponibilité tout au long de ces années et pour m'avoir permis de faire partie de la Team « Bruker » ...Une expérience pleine de rebondissement et très enrichissante !

Je remercie sincèrement Philippe Bulet, et Michel Rivière pour avoir accepté d'évaluer mon travail de thèse.

Un grand merci à la société L'Oréal pour avoir coencadré cette thèse CIFRE. Je tiens à remercier chaleureusement toute l'équipe avec qui j'ai collaboré pendant trois ans : Dominique Bernard, Nükhet Cavusoglu, Stéphane Commo, Emma Debecker-Derezt et Mickael LeBalch. Un très grand merci à Nükhet pour ta patience, ta bonne humeur, ta disponibilité et tes conseils.

Je tiens aussi à remercier vivement l'ensemble des collaborateurs Thierry Rabilloud, Alain. Wagner, Jean-Marc Reichhart avec qui j'ai pu interagir durant ces quelques années. Plus particulièrement merci à Florian Veillard pour tous ces échanges que nous avons eu, merci pour ton temps et ton aide pour la rédaction.

Merci à mes collègues du LMSBO, pour ces années. Merci à Jean Marc, les discussions, les conseils, les anecdotes, les beignets, prend soin de Nelson ! Merci à Danièle, pour tous tes conseils, ta bonne humeur, ton accueil à mon arrivée et surtout merci pour les séances de tortures sportives...Merci à François et Magalie, pour votre amitié pendant ces années. Bonne continuation à vous pour la suite.

Merci aux collègues du R2: Justine, Paola, Lesli, Pauline, Anthony, Oscar, Stéphane, Maxime, Thomas, et bonne continuation dans vos thèses et/ ou projets pro et perso!

Merci, aux bio-informaticiens et à la team qualité! Merci aux anciens du labo avec qui j'ai pu passer de bons moments (Diego, Sarah, Anna, Guillaume, Luc, Seb, Johann, ...). A tous ceux que j'oublie MERCI !

Merci au « Dr Tasher » dit Georg, pour m'avoir appris tant de choses, d'avoir toujours eu du temps pour m'expliquer (et me réexpliquer) les subtilités des Q-TOF, le soir, accoudés à l'Impact ou sur le comptoir d'un bar ! Merci à Macron (Charlotte de son petit nom), pour les « crêpes party », les essayages, le gala, les séances devant le QEx, les péniches, les soirées strasbourgeoises et mon EVJ... Rendez-vous dans le sud tous les deux.

Merci aux girls du bureau Aurélie et Marianne, entre l'Alsace et le Liban, incompréhension et ambiance garanties (On est déjà lundi ?! Non toujours vendredi). Même si parfois la langue française a été plus que malmenée voir piétinée... entre les séances dégustations-goûter et les débats à thème (Nutella Allemand vs Italien) ça a été un réel plaisir de partager les journées au bureau avec vous. Bonne continuation au dernier arrivé dans le bureau David, bon courage pour tes recherches et ta thèse !

Merci à Gauthier pour toutes ces années de « colocation » au laboratoire ... Sans toi je n'aurai jamais connu la ville de Tenteling, le patois, la gastronomie et les coutumes Mosellanes... Tes cool#stories#bro et la façon dont tu emploies le mot « d'ailleurs » vont me manquer. Du Texas, en passant par le Jura (G. Husson, découverte du Mont d'Or, 2016) puis par la Dordogne ...que des bons souvenirs ! Merci à Margaux, mon binôme de toutes les situations, la thèse, la nourriture (service repas assuré 24H/24), le sport (séance en extérieur à -4°C), le covoiturage, les séances réparations-spectro/plomberie, les promenades en tractopelle, les stages intensifs en Photoshop, les apéros, les sorties, les cours d'apprentissage du français à nos collègues Alsa/Libano/Mosellan, mon EVJ inoubliable... GACH, MBD, des réserves alimentaires pour tenir un siège, au mur des « chats » en hommage aux grands moments de réflexions ...sans vous, ça n'aurait pas été pareil, alors merci d'avoir partagé ces trois dernières années avec moi.

Un grand merci à Marine, pour tout, ton amitié, ton accueil, ton soutien/coaching, ton écoute...On vous attend dans le sud pour de nouvelles aventures!

Merci à mes Lingolettes (Alice, Gégé, Justine, Marjo et Virginie), mes années fac ont été « sublimes » grâce à vous, toujours là quand il faut, peu importe la distance, loin des yeux mais près du cœur!

Je tiens à remercier tous les « Favrel-Gauthier-Guigui-Gousset-Van Alderen », ma famille, mes piliers, ma source d'énergie. Merci particulièrement à mes parents, pour m'avoir tout donné, toujours soutenue et pour être là contre vents et marées...

Et enfin merci à Martin, pour me supporter depuis si longtemps, merci pour ce qui a été, ce qui est et tout ce qui va être. Comme promis, le vent nous porte...

Liste des principales abréviations

a.a : acide aminé
Ac : Anticorps
AP-MS : Affinity Purification – Mass spectrometry
CID : Collision Induced Dissociation
Cys (C) : Cystéine
Da : Dalton
DCT : DOPAchrome tautomérase
DDA : Data Dependant Acquisition
DSP : di-thio-bis-succinimidylpropionate
DTT : Dithiothréitol
EDTA : acide éthylène-diamine-tétra acétique
ESI : Electrospray Ionisation
FDR : False Discovery Rate
Gel 1D : gel d'électrophorèse à une dimension
Gram- : bactérie Gram négatif
Gram+ : bactérie Gram positif
HPLC : High Pressure Liquid Chromatography
ICAT : Isotope-Coded Affinity Tags
IMD : Immune Deficiency
IP : ImmunoPrécipitation
iTRAQ® : isobaric Tag for Relative and Absolute
KDa : Kilo Daltons
LC-MS : couplage chromatographie liquide- spectrométrie de masse
LC-MS/MS : couplage chromatographie liquide-spectrométrie de masse en tandem
LPS : LipoPolySaccharides
MALDI : Matrix Assisted Laser Desorption Ionisation
MRM : Multiple Reaction Monitoring
MS : Mass Spectrometry
N-ter : N-terminale
PAMPs : Motifs moléculaires associés aux pathogènes
PM : Poids Moléculaire
PRR : Pattern Recognition Receptors
Psh: Perséphone
PTM: Modification Post Traductionnelle
Q : Quadripôle
QAT : Quaternary Amine Tag
QQQ : Triple quadripôle
RF : Radiofréquence
RP-HPLC: Reversed Phase-HPLC
r-Psh : Perséphone recombinante
SCX : phase échangeuse de cations
SDS : Sodium Dodecyl Sulfate
SILAC : Stable Isotopic Labeling using Amino Acids in Cell Culture
SRM : Selected Reaction Monitoring
TA : Température ambiante
TBP : Tributyl phosphine
TLR : Toll-Like Receptor
TMPP : Tris(2,4,6-triMethoxyPhenyl) Phosphonium
TMPP-Ac-OSu : N-Succinimidylloxycarbonylmethyl) tris(2,4,6-trimethoxyphenyl)phosphonium
TOF : Time Of Flight
Trp (W) : Tryptophane
Tyr (Y) : Tyrosine
TYR : TYRosinase
TYRP1 : TYRosinase related Protein 1
V : Volt

Table de matière générale

Table des matières détaillée	13
Introduction générale	Erreur ! Signet non défini.8
Partie I : Introduction Bibliographie	22
Chapitre I Instrumentation	22
Chapitre II L'analyse Protéomique : spectrométrie de masse en tandem.....	30
Partie II : Résultats.....	42
Chapitre I Etudes de complexes protéiques cibles impliqués dans le mécanisme de la mélanogénèse	42
A. Introduction biologique.....	42
B. Introduction au contexte analytique : caractérisation de complexes multiprotéiques par MS.....	48
C. Résultats	54
D. Conclusion biologique de l'étude	76
E. Conclusion sur les aspects analytiques.....	77
Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP	80
A. Introduction biologique.....	80
B. Introduction au contexte analytique et études de N-terminomique	83
C. Résultats	90
D. Conclusion biologique de l'étude	109
E. Conclusion sur les aspects analytiques.....	111
Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS	113
A. Introduction.....	113
B. Introduction au contexte analytique : objectifs et stratégies analytiques mises en œuvre.....	116
C. Le marquage des cystéines.....	118
D. Le marquage des tryptophanes	130
E. Le marquage des tyrosines.....	143
F. Conclusion biologique de l'étude	147
G. Conclusion sur les aspects analytiques.....	148
Partie III : CONCLUSION GENERALE.....	150
Partie IV : PARTIE EXPERIMENTALE.....	152
REFERENCES BIBLIOGRAPHIQUES	164

Table des matières détaillée

Introduction générale	18
Partie I Introduction Bibliographie	22
Chapitre I Instrumentation	22
A. La spectrométrie de masse, outil indispensable à l'analyse protéomique	22
A.1. Le principe général.....	22
A.2. Les sources d'ionisation	22
A.2.1 La source ESI (<i>Electrospray Ionisation</i>)	22
A.2.2 La source MALDI (<i>Matrix Assisted Laser Desorption Ionisation</i>)	24
A.3. Les analyseurs	25
A.3.1 L'analyseur quadripolaire	25
A.3.2 L'analyseur à temps de vol (TOF)	25
A.3.3 L'analyseur à trappe orbitale (Orbitrap™)	27
A.3.4 Les analyseurs hybrides	28
A.4. Les détecteurs	28
Chapitre II L'analyse Protéomique : spectrométrie de masse en tandem	30
A. La fragmentation peptidique	31
A.1. La nomenclature	31
A.2. Le mode de fragmentation des peptides	31
A.3. Le mode d'acquisition des données	32
B. L'identification et validation des spectres de fragmentation	33
B.1. Les banques de données protéiques	33
B.2. Les algorithmes de comparaison ou moteurs de recherche.....	34
B.3. La validation statistique des données	34
C. Les stratégies de quantification en analyse protéomique	35
C.1. La stratégie de quantification ciblée : SRM (Selected Reaction Monitoring)	36
C.2. Les stratégies de quantification globale du protéome	37
C.2.1 Les approches sans marquage (<i>label free</i>)	38
C.2.2 Les avantages et les limitations des stratégies <i>label free</i>	39
C.2.3 Les approches par marquage isotopique	39
Partie II Résultats	42
Chapitre I Etudes d'un complexe protéique impliqué dans le mécanisme de la mélanogénèse 42	
A. Introduction biologique	42
A.1. La mélanogénèse ou pigmentation	42
A.1.1 Les mélanosomes	42
A.1.2 La synthèse de la mélanine : L'eumélanine et la phéomélanine	43
A.1.3 Les enzymes mélanosomales dans la voie des eumélanines.....	43
A.2. La coopération des enzymes mélanosomales : Formation de complexe	44
A.3. Les objectifs de l'étude	45
B. Introduction au contexte analytique caractérisation de complexes multiprotéiques par MS	48
B.1. Les stratégies basées sur la purification par affinité couplées à la spectrométrie de masse (AP-MS)	48
B.1.1 L'étape d'immunoprécipitation (IP)	49
B.1.2 La stabilisation des interactions protéiques par réticulation chimique	51
B.1.3 Le fractionnement biochimique	51
B.2. Une nouvelle approche par marquage « proximité dépendante » : Le BioID	52
C. Résultats	54
C.1. L'approche de purification de complexe par immunoaffinité et spectrométrie de masse (AP-MS)	54

C.1.1	L'immunoprécipitation : Ac dirigé contre la protéine TYR endogène	54
C.1.1.1	L'identification des protéines présentes par nanoLC-MS/MS	56
C.1.1.2	Conclusion de l'approche d'immunoprécipitation dirigée contre TYR	60
C.1.2	L'immunoprécipitation : Ac dirigé contre la DCT endogène.....	60
C.1.2.1	L'identification des protéines présentes par nanoLC-MS/MS	61
C.1.2.2	Conclusion sur les résultats de l'approche d'IP dirigé contre DCT	64
C.1.3	Conclusion de l'approche par immunoprécipitation dirigé contre une protéine du complexe 64	
C.2.	L'approche par réticulation chimique des protéines (DSP)	65
C.2.1	L'étape de réticulation des protéines par le DSP	65
C.2.1.1	L'identification des protéines identifiées dans les échantillons	67
C.2.1.2	Conclusion sur les résultats de l'approche de réticulation	69
C.2.2	L'immunoprécipitation : Ac dirigé contre TYR recombinante	69
C.2.2.1	L'identification de la TYR par nano-LC-MS/MS	71
C.2.2.2	Conclusion sur les résultats de l'approche d'immunoprécipitation de r-TYR.....	72
C.3.	L'approche par marquage à la biotine « BioID ».....	72
C.3.1	L'identification des protéines présentes dans les lysats	74
C.3.1.1	Conclusion sur les résultats de l'approche BioID	76
D.	Conclusion biologique de l'étude.....	76
E.	Conclusion sur les aspects analytiques	77

Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP 80

A.	Introduction biologique.....	80
A.1.	L'immunité	80
A.1.1	L'immunité innée.....	80
A.1.2	La voie du récepteurs Toll	81
A.1.3	La protéase Perséphone	82
A.2.	Les objectifs de l'étude	83
B.	Introduction au contexte analytique et études de N-terminomique.....	83
B.1.	L'état de l'art en analyse N-terminomique.....	84
B.1.1	La caractérisation des extrémités N-terminales par marquage au TMPP-Ac-OSu	86
B.1.2	L'optimisation du marquage in gel au TMPP et exploitation des données LC-MS/MS	87
C.	Résultats.....	90
C.1.	L'origine et préparation des échantillons de r-Psh incubée	90
C.2.	L'étude de la protéolyse de r-Psh par « Bacillus subtilis » (Gram+).....	93
C.2.1	La détermination des sites de clivage marqués au TMPP	94
C.2.2	La caractérisation fine des produits de dégradations de r-Psh	100
C.2.3	Bilan sur la caractérisation fine des produits protéolytiques de r-Psh par B. subtilis.	103
C.3.	L'étude de la protéolyse de Psh par d'autres pathogènes.....	104
C.3.1	La sélection des pathogènes.....	104
C.3.2	La détermination des sites de clivage des formes maturées de r-Psh	105
C.3.2.1	Les bactéries à Gram+.....	105
C.3.2.2	Les bactéries à Gram-.....	106
D.	Conclusion biologique de l'étude.....	109
D.1.	Les résultats concernant le site d'activation 143-144	109
D.2.	Les résultats inattendus concernant d'autres sites de clivage	110
E.	Conclusion sur les aspects analytiques	111

Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS113

A.	Introduction	113
B.	Introduction au contexte analytique : objectifs et stratégies analytiques mises en œuvre	116
C.	Le marquage des cystéines	118
C.1.	L'état de l'art du marquage chimique des cystéines	118
C.2.	Résultats et discussion	120
C.2.1	Les dérivés APN-TMPP.....	120
C.2.2	Les dérivés ammoniums quaternaires (QAT)	126
C.2.3	Conclusion pour le marquage chimique des cystéines.....	128

D. Le marquage des tryptophanes	130
D.1. L'état de l'art du marquage chimique des tryptophanes	130
D.2. Résultats et discussion	131
D.2.1 Le marquage des tryptophanes de la protéine modèle.	131
D.2.2 Le marquage des tryptophanes d'un échantillon complexe (extrait <i>E. coli</i>).	134
D.2.3 L'enrichissement « en ligne » des peptides NPBS.	139
D.3. Conclusion de la stratégie de marquage chimique des W	142
E. Le marquage des tyrosines	143
E.1. L'état de l'art du marquage chimique des tyrosines	143
E.2. Résultats et discussion	143
E.2.1 Les dérivés de Phényldiazonium.....	143
E.2.2 Phényltriazolédione.....	145
F. Conclusion biologique de l'étude.....	147
G. Conclusion sur les aspects analytiques	148
CONCLUSION GENERALE	150
Partie III PARTIE EXPERIMENTALE.....	152
Chapitre I Etudes de complexes protéiques cibles impliqués dans le mécanisme de la mélanogénèse	152
A. L'immunoprécipitation (IP) dirigée contre TYR et DCT	152
A.1. La préparation des échantillons.....	152
A.2. La préparation du gel 12 % SDS-PAGE	152
A.3. La digestion des protéines « in gel »	152
A.4. L'analyse nanoLC-MS/MS	153
A.4.1 La séparation chromatographique	153
A.4.2 L'acquisition des données MS-MS/MS	153
A.4.3 L'interprétation et validation des données acquises par LC-MS/MS.....	153
B. La stratégie de réticulation des protéines au DSP	154
B.1. La préparation des échantillons.....	154
B.2. La préparation du gel 12 % SDS-PAGE	154
B.3. La digestion des protéines « in gel »	154
B.4. L'analyse nanoLC-MS/MS	154
B.4.1 La séparation chromatographique	154
B.4.2 L'acquisition des données MS-MS/MS	155
B.4.3 L'interprétation des données acquises par LC-MS/MS	155
C. La stratégie de marquage à la Biotine « BiOID ».....	155
C.1. La préparation des échantillons.....	155
C.2. La préparation du gel 12 % SDS-PAGE	155
C.3. La digestion des protéines « in gel »	156
C.4. L'analyse nanoLC-MS/MS	156
C.4.1 La séparation chromatographique	156
C.4.2 L'acquisition des données MS-MS/MS	156
C.4.3 L'interprétation des données acquises par LC-MS/MS	156
Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP	157
A. La préparation des échantillons.....	157
B. Le protocole de marquage au TMPP « in gel ».....	157
C. La digestion des protéines « in gel ».....	157
D. L'analyse nanoLC-MS/MS	158
D.1. La séparation chromatographique.....	158
D.2. L'acquisition des données MS-MS/MS	158
D.3. L'interprétation des données acquises par LC-MS/MS.....	158
Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS	159
A. La préparation des échantillons.....	159

A.1.	Le marquage de la protéine modèle Conalbumine.....	159
A.1.1	Les échantillons contrôles	159
A.1.2	La dérivation des cystéines.....	159
A.1.3	La dérivation des tryptophanes.....	159
A.1.4	La dérivation des tyrosines.....	159
A.2.	Le marquage des protéines « d'E. coli » : échantillon complexe	160
A.2.1	Les échantillons contrôles	160
A.2.2	La dérivation des tryptophanes.....	160
A.3.	La digestion des protéines « in gel ».....	160
B.	L'analyse nanoLC-MS/MS	161
B.1.	La séparation chromatographique.....	161
B.1.1	L'analyse du digestat peptidique de la conalbumine	161
B.1.2	L'analyse du digestat peptidique d'E. coli	161
B.1.2.1	La colonne de piégeage C18	161
B.1.2.2	La colonne de piégeage Phényle	161
B.2.	L'acquisition des données MS-MS/MS	161
B.3.	L'interprétation des données de LC-MS/MS.....	162
B.3.1	La protéine modèle : conalbumine.....	162
B.3.2	L'échantillon biologique : E. coli	162
	REFERENCES BIBLIOGRAPHIQUES	164

Introduction générale

Introduction générale

Le rôle de la spectrométrie de masse (MS), et plus particulièrement de l'analyse protéomique est aujourd'hui largement reconnu en biologie. Elle prend naturellement sa place à côté des méthodes d'analyse génomique et d'analyse transcriptomique qui ont bouleversé nos connaissances des mécanismes moléculaires en biologie.

La protéomique est un domaine scientifique focalisé sur l'identification, la caractérisation et la quantification de toutes les protéines présentes dans un échantillon, à un moment donné et dans des conditions données¹. L'objectif de la protéomique est de fournir des informations sur l'expression de protéines, la localisation cellulaire, les modifications post-traductionnelles (PTM), les interactions entre protéines et le renouvellement ou « turnover » protéique².

La MS des peptides et protéines a pris son essor avec l'invention de deux méthodes d'ionisation de molécules de hauts poids moléculaires, MALDI³ et ESI⁴, dont l'intérêt a été reconnu par un prix Nobel en 2002. Un certain nombre de méthodologies d'analyse protéomique est aujourd'hui bien maîtrisé. Cependant, les demandes en termes d'analyse quantitative et de caractérisations complètes des protéines sont encore souvent au-delà des possibilités des approches classiques (manque de sensibilité, purification insuffisante, difficulté à définir une position N-terminale, robustesse et facilité d'emploi d'une méthode de quantification).

À côté des progrès purement techniques, réalisés principalement par les constructeurs de spectromètres de masse, les progrès qualifiés de « méthodologiques » sont réalisés dans les laboratoires académiques. Ces progrès portent sur la mise sur pied d'un ensemble, souvent très complexe, d'étapes de nature très diverse et qui constituent une stratégie pour établir des données expérimentales fiables, et ainsi faire progresser les connaissances. Ces étapes, qui doivent souvent être optimisées pour chaque projet, sont par exemple la façon d'isoler et de contrôler le matériel biologique qui va être soumis à l'analyse, le paramétrage de la prise de spectres de fragmentation MS/MS dans les expériences de LC-MS/MS, l'exploitation bioinformatique des données MS/MS pour l'identification ou encore la quantification des protéines.

Mon travail de thèse a eu pour objet de développer l'analyse protéomique principalement dans le cadre de trois projets de recherche menés par des biologistes, avec lesquels nous avons collaboré.

Le premier de ces projets a porté sur l'étude de la régulation de la pigmentation de la peau chez l'homme et a été mené en collaboration étroite avec la société L'Oréal. L'objectif de ces travaux était d'étudier la régulation de la mélanogénèse et la formation du complexe protéique impliquant les enzymes mélanosomales. Les résultats de cette étude devraient permettre de concevoir des stratégies pour moduler la production de mélanines. La mélanogénèse est un processus biologique faisant intervenir au moins trois enzymes : TYR, TYRP1 et DCT dont il a été montré qu'elles interagissaient, et formaient probablement un complexe. Pour progresser dans la compréhension du mécanisme de synthèse des mélanines, il était nécessaire de déterminer quelles étaient les autres protéines qui interagissaient avec le complexe TYR/TYRP1/DCT.

L'isolement du complexe sera réalisé à partir de mélanocytes, et sera basé sur l'utilisation de la technique d'immunoaffinité en ciblant tour à tour chacune des trois enzymes. Les contenus protéiques des complexes ainsi isolés seront analysés par gel 1D SDS-PAGE.

Plusieurs possibilités seront alors explorées pour identifier les protéines présentes. La première sera de réaliser directement une analyse nanoLC-MS/MS après immunoprécipitation. Si cette approche ne permet pas l'identification de protéines spécifiques du complexe d'autres alternatives de préparation d'échantillon pourront être tentées, comme par exemple une nouvelle méthode récemment décrite, appelée BioID⁵ ou une méthode basée sur la réticulation chimique (*cross-linking*).

Le second de ces projets a porté sur la détermination d'une étape clé du mécanisme de régulation de la voie du récepteur Toll dans l'immunité innée chez la drosophile. Ce projet a été mené en collaboration étroite avec le Pr. JM. Reichhart (IBMC, Strasbourg). Le problème essentiel concernait les points de clivage pour l'activation de la protéase Perséphone (Psh) qui, à son tour, active indirectement la protéine Spatzle responsable de l'interaction avec le récepteur Toll. Pour cette étude de N-terminomique, nous avons choisi de développer une approche de marquage chimique par le TMPP « *in gel* ». Cette approche devrait permettre de tester l'hypothèse d'une activation de Psh par un clivage spécifique de celle-ci au niveau du motif site « ¹⁴⁴IVGG », par les protéases de pathogènes libérées dans l'hémolymphe lors de l'infection. Par ailleurs, en parallèle de la détermination des nouvelles positions N-terminales ainsi générées, nous utiliserons l'avantage de la stratégie de marquage au TMPP pour couvrir l'ensemble des séquences des produits de dégradation de Psh après clivage. Les études d'incubation de la protéine Perséphone seront d'abord réalisées *in vitro* avec un pathogène à Gram+ (*Bacillus subtilis*) possédant une enzyme spécifique appartenant à la famille des subtilisines. Dans un deuxième temps, des incubations avec des pathogènes à Gram- et Gram+, ne possédant pas d'enzyme du type subtilisine (*Enterococcus faecalis*, *Serratia marcescens* et *Pseudomonas aeruginosa*), seront étudiées. Cette étude devrait permettre de déterminer si les protéases de ces pathogènes activent Perséphone également par clivage au niveau du site « ¹⁴⁴IVGG ».

Le troisième de ces projets avait pour objectif le développement des méthodes de quantification en analyse protéomique par marquage chimique qui seraient particulièrement robustes et faciles à utiliser à grande échelle. Ce projet a été mené en étroite collaboration avec les Dr. T. Rabilloud (CEA Grenoble) et A. Wagner (Université de Strasbourg). Nous explorerons les possibilités de marquage spécifique de trois acides aminés : la cystéine (méthode la plus utilisée pour le marquage chimique des protéines), le tryptophane et la tyrosine. Le réactif de marquage sera choisi selon plusieurs critères. Il devra d'une part être chimiosélectif et augmenter le caractère hydrophobe des peptides marqués. Ceci devra permettre, un enrichissement des peptides de digestion dérivés dans le but de faciliter leur séparation chromatographique et d'éviter le sous-échantillonnage lors de la prise de spectres MS. Le réactif de marquage devra d'autre part conférer aux peptides marqués les propriétés suivantes : fragmentation caractéristique confirmant la présence du réactif et permettant une détection sélective de ces peptides, un facteur de réponse élevé et la possibilité de synthèse facile de formes marquées aux isotopes stables pour la quantification.

Nous présentons ci-après une « Introduction Bibliographique » consacrée uniquement aux aspects analytiques.

Un premier chapitre porte sur l'instrumentation de spectrométrie de masse. Un second chapitre présente les stratégies de quantification en analyse protéomique (0).

Une bibliographie spécifique à chacun des trois projets et portant sur les contextes biologiques sera présentée en début de chaque chapitre consacré à chaque projet dans la partie Résultats (**Erreur ! Source du r envoi introuvable.**).

Partie I Introduction Bibliographie

Partie I Introduction Bibliographique

Chapitre I Instrumentation

A. La spectrométrie de masse, outil indispensable à l'analyse protéomique

A.1. Le principe général

Il aura fallu attendre plusieurs décennies après la première étude publiée par J.J. Thompson (1913)⁶ concernant l'acquisition de spectres de masse de plusieurs composés gazeux, pour que l'analyse de biomolécules soit envisagée par spectrométrie de masse (MS). En effet, jusque dans les années 1990, le séquençage d'Edman était la technique principale de détermination des séquences en acides aminés des protéines⁷. Grâce à la découverte de nouvelles techniques d'ionisation douce par désorption laser assistée par matrice (MALDI)³ et par électrospray (ESI)⁴ dans les années 1990, reconnues par l'obtention d'un prix Nobel en 2002, la spectrométrie de masse a supplanté le séquençage d'Edman. Les performances instrumentales (sensibilité, précision, résolution, etc.) sont, depuis, en constante amélioration et ont fait progressivement de la MS un outil de choix pour l'analyse des séquences, pour la caractérisation des modifications post-traductionnelles et la quantification des protéines en biologie⁸.

Les différents constituants d'un spectromètre de masse ont été représentés dans la Figure 1.

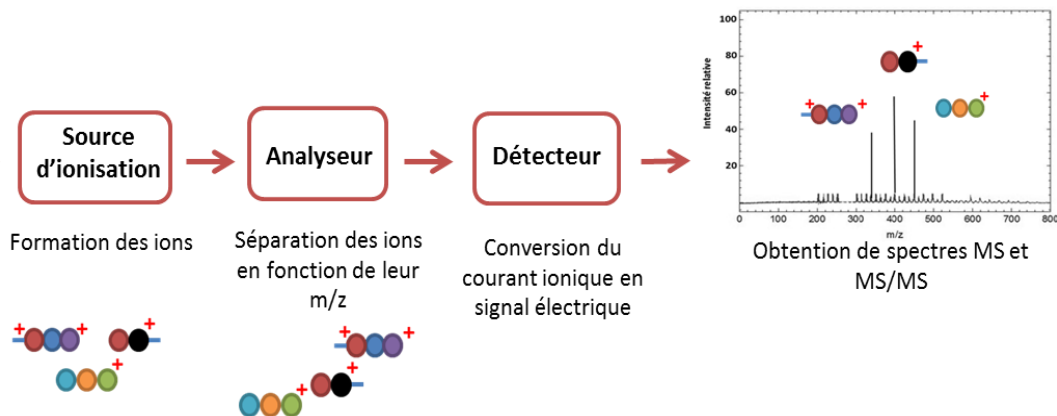


Figure 1: Représentation schématique des différents constituants d'un spectromètre de masse.

Dans les paragraphes qui vont être présentés ci-dessous, seuls les principes théoriques des spectromètres de masse utilisés au cours de mes travaux de thèse seront présentés.

A.2. Les sources d'ionisation

A.2.1 La source ESI (Electrospray Ionisation)

A la fin des années 1960, les expériences menées par M. Dole sur le mécanisme ESI laissent présager que des mesures de molécules de haut poids moléculaires en solution (polymères) pouvaient être réalisées⁹. Il aura fallu attendre les travaux de J. Fenn dans les années 1980, concernant le couplage de la source électrospray

avec un analyseur quadripolaire, pour démontrer que les ions pouvaient être formés sans fragmentation à partir de molécules non volatiles¹⁰.

Les sources ESI mettent en jeu un mécanisme d'ionisation qui assure le passage d'ions présents en solution vers la phase gazeuse, à pression atmosphérique, sous l'effet d'un champ électrique fort. Ce sont des sources d'ionisation dites « douces ».

Production de gouttelettes

C'est l'application d'une différence de potentiel entre le capillaire et l'entrée du spectromètre de masse qui permet à la goutte contenant l'échantillon, de se polariser. L'accumulation de charges, à la pointe du capillaire, déstabilise la gouttelette, qui va alors s'allonger pour prendre la forme d'un cône appelé « cône de Taylor ». Quand les forces à la surface du cône de Taylor deviennent égales à la tension de surface (limite de Rayleigh), des gouttelettes chargées sous forme de spray sont alors émises¹¹.

Diminution de la taille des gouttelettes chargées

Le chauffage de la source va favoriser l'évaporation du solvant. La diminution du rayon des gouttelettes va alors augmenter la densité de charge, atteignant alors la stabilité de la limite de Rayleigh¹². Les forces de répulsion coulombienne vont devenir plus importantes que les forces de tension de surface et plusieurs explosions coulombiennes successives vont générer de multiples gouttelettes dites « filles »¹¹.

Transfert des ions désolvatés en phase gazeuse

Le mécanisme proposé dans le modèle de M. Dole, aussi appelé modèle de la charge résiduelle, suggère que les phénomènes de diminution de la taille des gouttelettes chargées puis d'explosion en gouttelettes filles se répètent. Les ions sont désolvatés progressivement jusqu'à atteindre un seuil « d'un ion par gouttelette »¹³. Un autre mécanisme a été proposé, suggérant la désorption d'ions avant d'atteindre la limite de stabilité de Rayleigh (modèle d'Iribarne et Thomson)¹⁴. Pour les peptides comme pour les protéines, le modèle de M. Dole, apparaît comme le plus susceptible d'intervenir¹⁵.

Le mécanisme d'ionisation par ESI est représenté schématiquement dans la Figure 2.

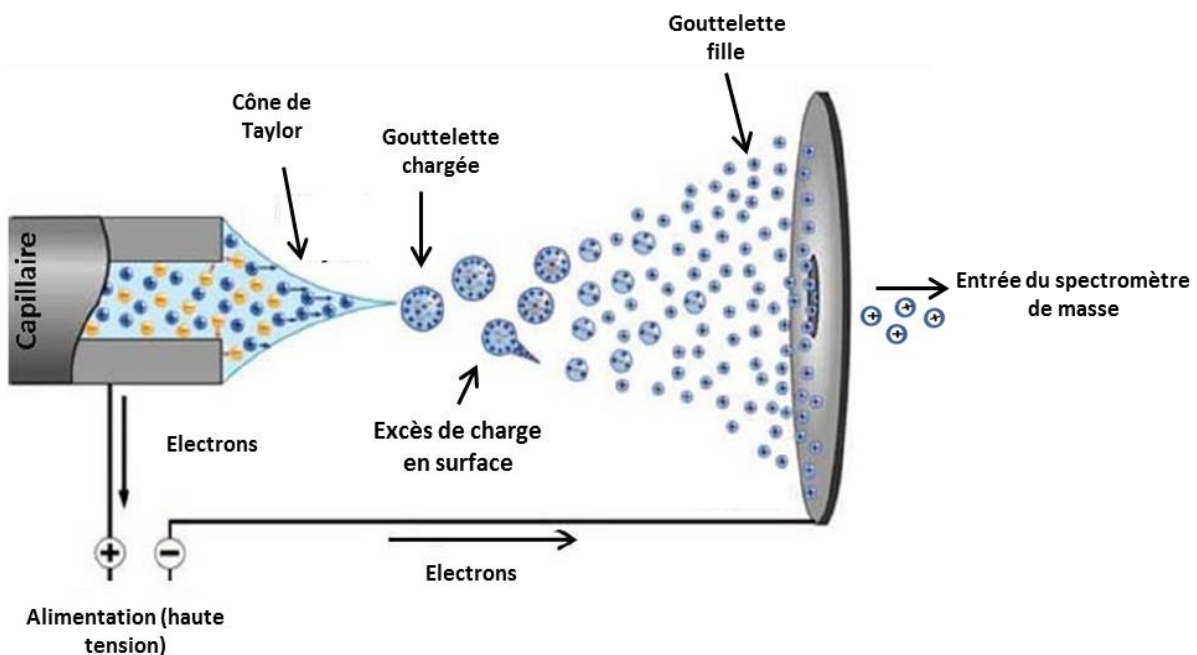


Figure 2: Représentation schématique du modèle suggéré par M. Dole pour la formation d'ions en phase gazeuse avec une source ESI, en mode d'ionisation positif.

En 1996, le développement de la source nano-électrospray (nano-ESI)¹⁶ a apporté de nombreux avantages pour l'analyse nanoLC-MS/MS. Parmi les avantages nous notons la diminution des quantités d'échantillon nécessaires (débit réduit), la réduction des interférences et l'augmentation de l'efficacité d'ionisation des peptides¹⁷.

A.2.2 La source MALDI (*Matrix Assisted Laser Desorption Ionisation*)

L'autre technique majeure d'ionisation douce est la technique MALDI, développée en 1988 par M. Karas et F. Hillenkamp¹⁸. Les analytes d'un échantillon vont être ionisés sous vide après avoir été co-cristallisés avec une matrice organique. Les mécanismes mis en jeu ne sont pas encore clairs pour la communauté scientifique mais le modèle le plus probable est constitué de trois étapes¹⁹⁻²⁰.

Lors d'une première étape, les molécules de la matrice sont irradiées par le rayonnement du laser de longueur d'onde équivalente généralement à celles de la gamme des UV. La matrice, à caractère acide, va donc s'ioniser par échauffement et se désorber, entraînant avec elle les molécules co-cristallisées.

Lors d'une deuxième étape, la matrice va agir comme donneur ou accepteur de protons, ionisant ainsi l'analyte^{18, 21} qui va alors se désorber par diminution des forces de cohésion des cristaux analyte-matrice. Le transfert de protons entre la matrice et l'analyte confère un charge positive aux molécules d'analyte ($[M+H]^+$).

Lors de la troisième étape, les molécules d'analytes sont émises en phase gazeuse²² et les ions générés sont alors détectables selon le mode d'ionisation (positif ou négatif). L'irradiation laser est influencée par la longueur d'onde et la durée d'impulsion²¹. En plus d'être plus tolérante à la présence de sels dans l'échantillon, l'ionisation MALDI conduit en général à la formation d'ions monochargés, ce qui peut faciliter l'interprétation des spectres MS.

Le mécanisme d'ionisation par MALDI est représenté schématiquement dans la Figure 3.

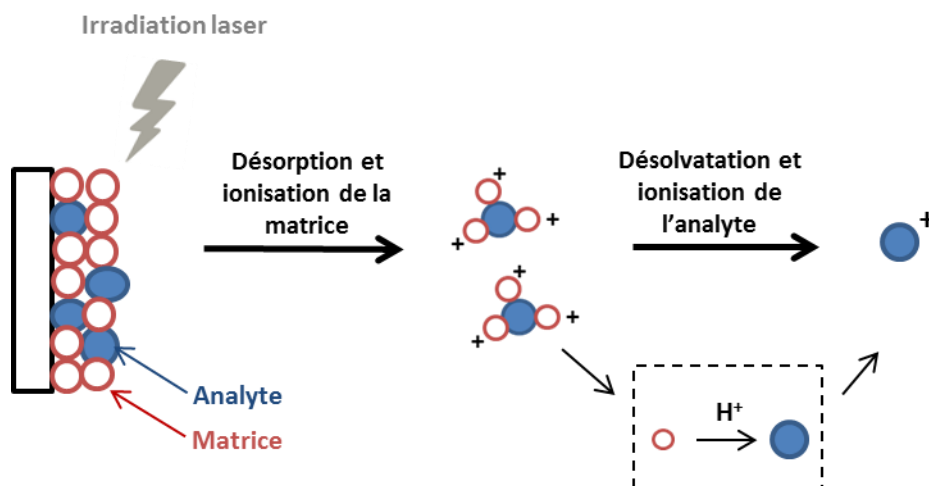


Figure 3: Représentation schématique du modèle le plus probable proposé par E. Hoffmann et V. Stroobant²⁰ pour la formation d'ions en phase gazeuse avec une source MALDI, en mode d'ionisation positif.

A.3. Les analyseurs

Les principaux analyseurs de masse sont les analyseurs quadripolaires, les analyseurs de temps de vol et les trappes orbitale. En fonction de leur géométrie, les analyseurs sont caractérisés par des paramètres clés tels que la gamme de masses analysable, la précision de mesure de masse, la sensibilité et la résolution. Il est possible de combiner l'utilisation de plusieurs types d'analyseurs, pour créer ce que sont communément appelés les analyseurs hybrides, associant les avantages de chacun des composants.

A.3.1 L'analyseur quadripolaire

Il a été développé en 1953 par W. Paul et H. Steinwedel²³, et de par sa capacité à analyser un courant d'ion en continu et son coût, il est l'analyseur basse résolution le plus fréquemment couplé à la source d'ionisation ESI. L'analyseur quadripolaire ou « Q » est composé de quatre électrodes cylindriques parallèles, connectées deux à deux. Le champ électrique, en opposition de phase, appliqué à chaque paire d'électrodes va permettre à l'analyseur de filtrer les ions en fonction de leur rapport masse/charge (m/z). Sous l'application du champ électrique oscillant, la trajectoire des ions accélérés selon l'axe z (vers le détecteur) va être régie par les équations de Mathieu. Le quadripôle va pouvoir séparer les ions en fonction de la stabilité de leur trajectoire entre les électrodes. Pour chaque valeur de potentiel appliqué, les ions avec une trajectoire stable au potentiel donné, sont transmis au détecteur tandis que les autres ions sont déviés²⁰. Le principe de sélection des ions par l'analyseur quadripolaire, à un potentiel fixe donné est schématisé Figure 4.

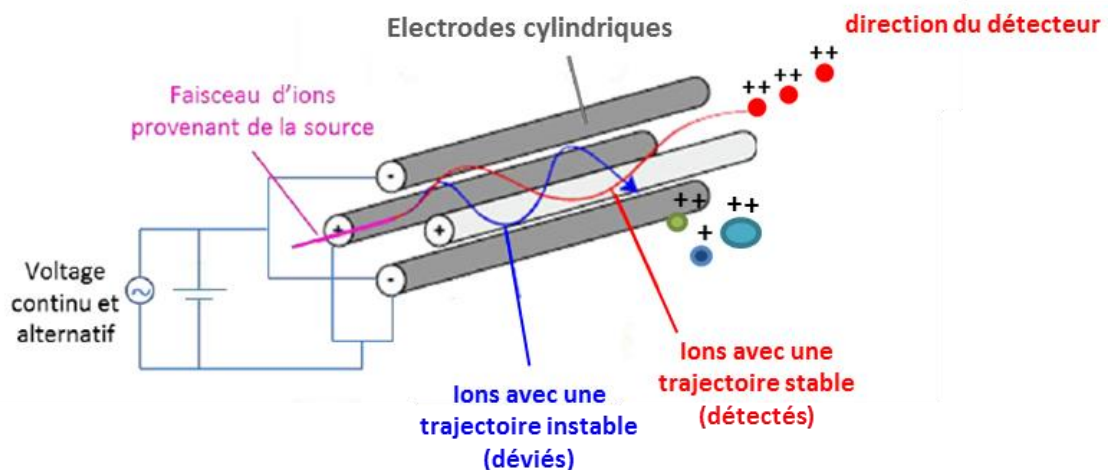


Figure 4: Représentation schématique de la trajectoire stable d'un type ion (en rouge) pour un potentiel fixe et d'une trajectoire instable pour les autres ions (en bleu), dans un analyseur quadripolaire.

A.3.2 L'analyseur à temps de vol (TOF)

Le concept de l'analyseur TOF a été décrit dans les années 1950²⁴. Un vide poussé de 1.10^{-7} mbar va être appliqué dans un tube appelé « tube de vol ». A l'entrée du tube de vol, les ions vont être accélérés grâce à l'application d'un champ électrique. L'accélération électrique des ions entrants va permettre de les séparer selon leur célérité (vitesse de déplacement). Dans un premier temps, une différence de potentiel très élevée

va être appliquée pour conférer une énergie cinétique (E_c) identique aux ions puis dans un deuxième temps, ces ions vont migrer jusqu'au détecteur en fonction de leur célérité. Le temps de vol des ions jusqu'au détecteur permet ensuite de déterminer leur rapport m/z selon les équations suivantes :

En déterminant l'énergie cinétique (1)

$$(1) E_c = 2zeU = \frac{1}{2}mv^2$$

Avec : z = charge de l'ion, e = charge d'un électron, U = différence de potentiel, m = masse de l'ion et v = vitesse de l'ion

La vitesse d'un ion peut alors être déduite avec l'équation suivante (2) :

$$(2) v = \sqrt{\left(\frac{2zeU}{m}\right)}$$

L'équation qui relie alors le temps nécessaire à l'ion pour se déplacer jusqu'au détecteur dans le tube de vol de longueur L est la suivante (3) :

$$(3) t = \frac{L}{v} = \frac{L}{\sqrt{2eU}} * \sqrt{\left(\frac{m}{z}\right)}$$

Peu résolutif à ses débuts, l'analyseur TOF a connu un regain d'intérêt avec le développement de la source MALDI dans les années 1980²⁵.

Le principe de fonctionnement d'un analyseur à temps de vol (TOF) est présenté Figure 5.

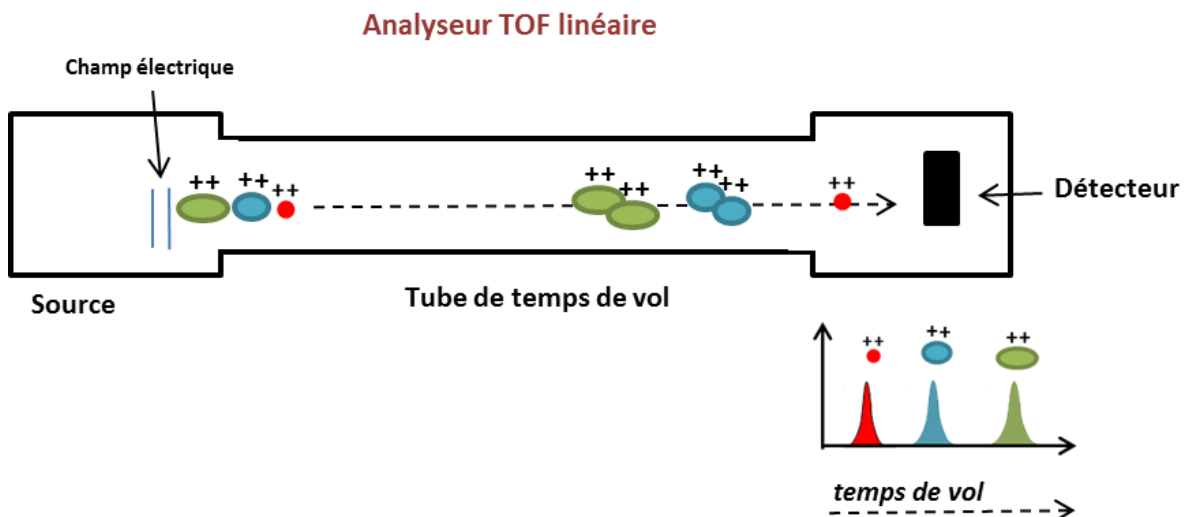


Figure 5: Représentation schématique du fonctionnement d'un analyseur à temps de vol (linéaire).

Les spectres de masses obtenus par un analyseur type TOF correspondent à la somme des spectres générés pour des groupes d'ions propulsés successivement à l'entrée du tube de vol.

En théorie les ions de même rapport m/z acquièrent la même énergie cinétique l'entrée du tube de vol et ont alors le même temps de vol pour arriver au détecteur. En pratique, pour ces ions de même rapport m/z ,

l'énergie cinétique (E_c) peut être légèrement différentes modifiant alors le temps de vol. La résolution de l'analyseur est alors diminuée. Pour contourner ce problème et améliorer la résolution, un réflectron électrostatique (série de lentilles) permettant de focaliser les ions a été ajouté à l'analyseur. Les ions propulsés vont pénétrer dans le réflectron, ceux qui auront une E_c plus élevée, pénétreront plus en profondeur contrairement aux ions avec une E_c plus faible. Ainsi, tous seront ensuite réfléchis en même temps en direction du détecteur²⁶ (amélioration de la résolution).

A.3.3 L'analyseur à trappe orbitale (Orbitrap™)

La trappe orbitale a été développée au début des années 2000 par A. Makarov²⁷. La trappe orbitale aussi appelée Orbitrap™ est un piège à ion, à champ électrostatique traversé par une électrode centrale en forme de fuseau. L'analyseur possède une géométrie particulière avec deux électrodes externes entre lesquelles un champ électrique linéaire est appliqué, ainsi que le long de l'électrode centrale. La forme particulière des électrodes va permettre l'application d'un champ électrostatique (champ dit quadro-logarithmique) dans ce piège à ion de façon à ce que l'oscillation des ions soit harmonique (oscillation avec une fréquence sinusoïdale qui ne dépend que de la masse)²⁸.

Les ions vont entrer dans la trappe avec un mouvement de rotation autour de l'électrode centrale et un mouvement axial le long de cette électrode selon l'axe z (Figure 6). Les ions auront donc une trajectoire identique, en forme de spirale autour de l'électrode centrale. Lorsqu'une rampe de radiofréquence est appliquée entre les deux électrodes externes, l'énergie des ions, qui sont en résonance avec la fréquence, est augmentée. Ainsi ces ions sont éjectés vers les électrodes externes. Les mouvements des ions, selon l'axe z, génèrent un courant induit enregistré par les deux électrodes externes. Les différents m/z des ions dans l'Orbitrap™ vont être déterminés par application d'une transformée de Fourier qui permettra de calculer la fréquence axiale des ions (ω). La vitesse avec laquelle les ions gravitent autour de l'électrode centrale, et donc la fréquence d'oscillation (résonance), est caractéristique du rapport m/z de ces ions (Figure 6).

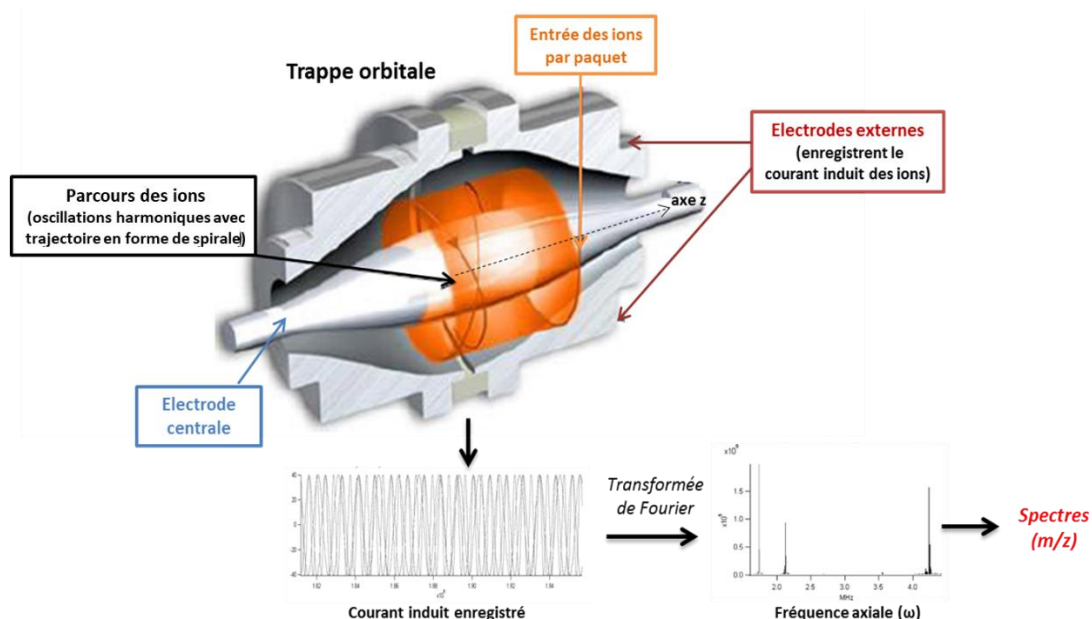


Figure 6: Représentation schématique du fonctionnement d'un analyseur à trappe orbitale. Vue en coupe de la trappe orbitale et la relation entre le courant induit des ions et le spectre de masse.

Un autre type de piège à ion linéaire appelé C-Trap (sous vide poussé), en amont de l'analyseur, va projeter les ions par paquet dans la trappe orbitale, de façon rapide et sans effet « temps de vol » (haute résolution).

A.3.4 Les analyseurs hybrides

L'analyseur dit hybride qui consiste à combiner deux ou trois types d'analyseurs en tandem dans un spectromètre de masse, ce qui permet de fragmenter les peptides. En pratique, l'analyseur quadripolaire est couramment combiné avec l'analyseur Orbitrap™ ou le TOF. L'avantage de la combinaison est d'isoler efficacement les ions par gamme de masse (m/z) avec le Q, de les fragmenter et d'analyser ensuite les fragments dans le TOF ou l'Orbitrap.

Dans certains spectromètres de masse, le quadripôle est également utilisé comme cellule de collision et permet donc de fragmenter les peptides, comme dans les systèmes Q-TOF et QQQ (Triple quadripôles).

En pratique, les QQQ sont peu utilisés pour des analyses exhaustives de protéome (type *discovery proteomic*) mais sont plutôt pour des analyses quantitatives ou de détection ciblée en mode SRM (*Selected Reaction Monitoring*)²⁹⁻³⁰. Les approches de type SRM permettent d'utiliser la fragmentation caractéristique des peptides pour détecter et quantifier de manière spécifique des protéines dans un mélange complexe.

Les spectromètres de masse hybrides type Q-TOF sont bien implantés dans le domaine de l'analyse protéomique haut débit, de par la résolution élevée (~60 000 pour un Impact II de Bruker Daltonics) pour une large gamme de masse, et la précision de masse (<5 ppm) pour la dernière génération d'instruments. Les Q-TOF peuvent acquérir jusqu'à 25 spectres MS/MS par seconde (vitesse de balayage de 25Hz) par exemple, pour l'Impact II (Bruker Daltonics).

Les spectromètres de masse hybrides type Q-Orbitrap™, combine l'avantage apporté par le Q, la C-Trap et la trappe orbitale. Cela permet d'atteindre des résolutions jusqu'à 140 000 (pour un m/z 200) pour un Q Exactive+ (Thermo Scientific™), une précision de masse élevée (<5 ppm) et ils peuvent acquérir jusqu'à 10 spectres MS/MS par seconde (vitesse de balayage de 10 Hz). Ces instruments ont également une sensibilité élevée ce qui est indispensable pour l'analyse protéomique.

A.4. Les détecteurs

Un détecteur a pour rôle de transformer le courant des ions enregistré en courant électrique. Après traitement et amplification, le courant électrique est digitalisé et transformé en une valeur numérique par un système électronique codé (bits). Il a pour rôle de compter les ions arrivants en fonction du temps.

Il existe deux types de détecteur :

- Les ions peuvent être détectés par la production d'un courant d'image induit, comme c'est le cas avec un analyseur à trappe orbitale (cf. Partie I -A.3.3).
- Les ions peuvent être détectés par émission d'électrons avec des détecteurs channeltron³¹ et galette de microcanaux³² (MCP pour *Micro Channel Plate*), Ils sont fréquemment utilisés dans les systèmes instrumentaux type TOF, Q-TOF. Le détecteur MCP en forme de disque est traversé par des multiples microcanaux (diamètre de l'ordre du micron) dont la surface interne est semi-conductrice. Lors du passage des ions, ceux-ci vont venir s'écraser sur cette surface et générer des électrons. Ces électrons

vont alors générer à nouveau d'autres électrons secondaires par impact sur une nouvelle surface semi-conductrice. *In fine*, ce mécanisme d'amplification d'électron va permettre de produire un courant électrique mesurable³³.

Chapitre II L'analyse Protéomique : spectrométrie de masse en tandem

L'analyse protéomique permet l'identification, la caractérisation et la quantification des protéines présentes dans un échantillon. Elle est constituée de cinq grandes étapes qui sont présentées dans la Figure 7. Les protéines extraites des échantillons vont être digérées par une enzymatique après une étape de fractionnement ou d'enrichissement. Le digestat de peptide obtenu va être séparé par chromatographie liquide (LC) et analysé par MS et MS/MS. Les spectres générés vont ensuite être interprétés avec des outils bioinformatiques. Cette approche d'analyse protéomique est aussi appelée approche *bottom up*.

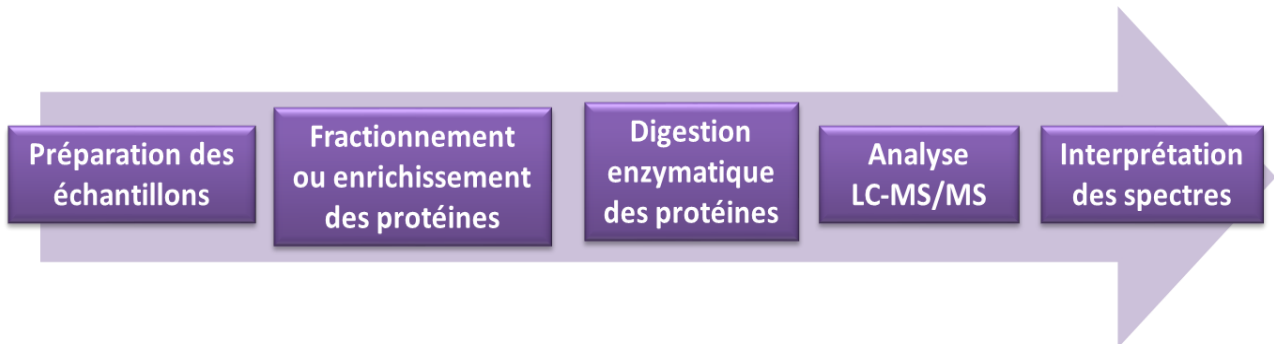


Figure 7: Représentation schématisée des étapes de l'analyse *bottom up* de la préparation des échantillons à l'interprétation des spectres.

Dans les paragraphes suivants, sont détaillées plus en détail les deux dernières étapes de l'analyse protéomique : la fragmentation MS/MS des peptides et les outils bioinformatiques pour l'interprétation des spectres.

Les spectromètres de masse en tandem ou « MS/MS » sont des spectromètres de masses hybrides (Q-TOF, Q-Orbitrap™, ...) et qui permettent d'isoler puis de fragmenter les peptides.

Après séparation chromatographique, les peptides sont ionisés en source. Les ions peptidiques chargés vont être analysés en MS puis vont être sélectionnés en fonction de leur rapport m/z avant d'être fragmentés dans une cellule de collision (MS/MS). Les ions fragments générés vont de nouveau être séparés par un analyseur et être détectés (Figure 8).



Figure 8: Représentation schématisée d'un spectromètre de masse en tandem

A. La fragmentation peptidique

A.1. La nomenclature

Il existe plusieurs modes de fragmentation utilisés pour l'analyse protéomique. Chaque type de fragmentation va générer des ions fragments à partir des précurseurs sélectionnés. En 1990 une nomenclature pour les différents types de fragment générés a été proposée par Biemann³⁴, comme cela est présenté dans la Figure 9. La nomenclature de Biemann classe les fragments de la chaîne peptidique en deux catégories d'ions :

-Le groupe « a, b et c » qui contient la partie N-terminale du peptide précurseur. Les ions correspondants vont porter une charge positive du côté N-Terminal

-Le groupe « x, y et z » qui contient la partie C-terminale du peptide précurseur. Les ions correspondants vont porter une charge positive du côté C-Terminal.

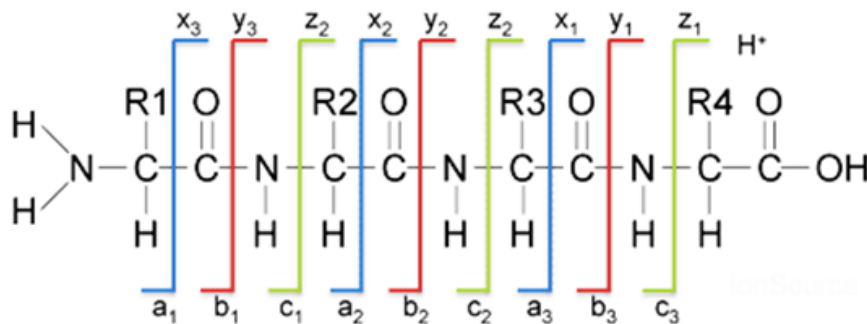


Figure 9 : Représentation schématique de la nomenclature des fragments définis par Biemann en 1990. Les groupements R1, R2, R3 et R4 correspondent aux chaînes latérales des acides aminés.

Les fragments classiquement générés en analyse protéomique sont ceux issus de la rupture de la liaison peptidique (b et y). Les modes de fragmentation les plus fréquemment utilisés en analyse protéomique sont le mode CID (*Collision-Induced Dissociation*) ou le mode HCD (*Higher energy Collisional Dissociation*, terme commercial déposé par la société Thermo Fisher). Le mécanisme de fragmentation CID et HCD sont basés sur le même mode de fonctionnement mais l'énergie est plus élevée dans la cellule de fragmentation HCD. Ils sont présentés dans le paragraphe ci-dessous.

A.2. Le mode de fragmentation des peptides

Dans le mode de fragmentation CID, les ions peptidiques vont être accélérés pour entrer en collision avec des molécules de gaz inerte (Hélium, Azote,...) dans une cellule adaptée, dite « cellule de collision » (cf. Figure 8).

Lors de la collision, l'énergie interne des ions peptidiques va augmenter et entraîner la délocalisation d'un proton mobile. Dans le cas des peptides monochargés, la charge se situe au niveau des fonctions amines des chaînes latérales des résidus arginines (R) et lysines (K). L'énergie nécessaire pour délocaliser cette charge le long du squelette peptidique est élevée et cela explique la difficulté de fragmenter les peptides monochargés.

Les peptides doublement chargés, portent généralement les charges sur les fonctions amine en position N-terminale du peptide et sur des résidus K et R. Le proton situé sur la fonction amine terminale est facilement transférable le long du squelette peptidique. La fragmentation de ces peptides va rompre la liaison amide, chaîne la plus labile selon le modèle du proton mobile³⁵.

Le mode de fragmentation CID apparaît comme la fragmentation la plus adaptée pour les peptides issus d'une digestion trypsique car les peptides sont clivés spécifiquement après les résidus K et R. Cela assure la présence d'un de ces résidus basiques dans le peptide. La fragmentation génère majoritairement des ions fragments de type « b et y ».

Un exemple de spectre du peptide WSVVSNQDVEC(+57)TVVDETK, obtenu par fragmentation en mode CID est présenté Figure 10 pour illustrer le type d'ions fragments générés. L'ajout de 57 Da à la masse de la cystéine correspond à l'alkylation des cystéines (blocage des ponts disulfures par de l'iodoacétamide).

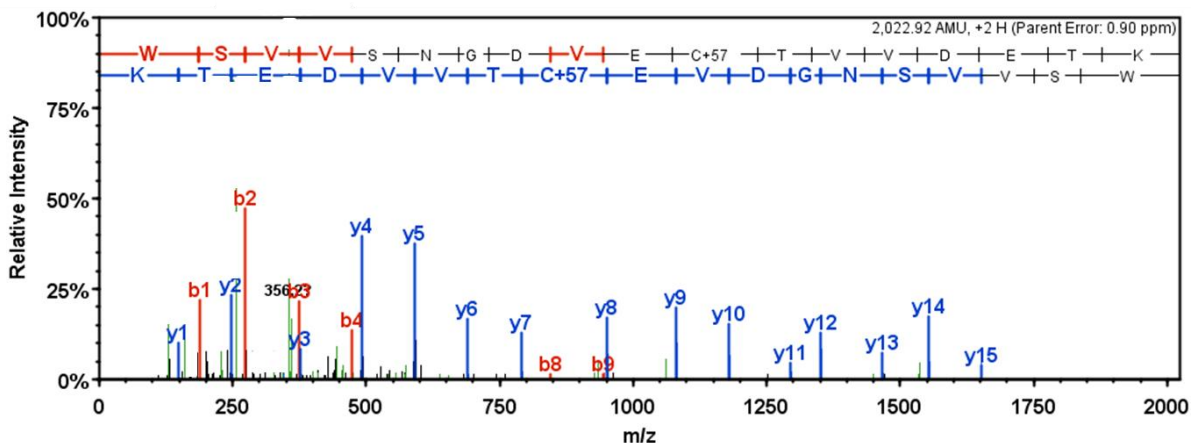


Figure 10: Représentation schématique d'un spectre MS/MS d'un peptide en CID, avec la génération d'ions fragments de type "b et y".

Comme nous le verrons dans le chapitre III, la modification du schéma de fragmentation classique peut être un avantage dans la détection spécifique d'une sous population de peptides.

A.3. Le mode d'acquisition des données

Les données MS/MS générées lors des analyses protéomiques peuvent être obtenues par deux types d'acquisition :

- Le mode d'acquisition dépendant des données (*Data Dependent Acquisition* ou DDA) permet de réaliser de manière automatique des scans MS puis MS/MS en alternance. Après un balayage MS des peptides, les ions les plus intenses sont sélectionnés pour être fragmentés. Bien que largement utilisé, l'inconvénient de ce mode d'acquisition est son incapacité à sélectionner tous les ions pour la MS/MS. Cette sélection stochastique des ions est liée au phénomène de coélution des peptides et à leur efficacité d'ionisation.
- Le mode d'acquisition indépendant des données (*Data Independent Acquisition* ou DIA), développé en 2004 par J.R Yates³⁶, est basé sur la fragmentation de tous les ions peptidiques en sélectionnant séquentiellement une fenêtre de (25 Da) sur une gamme de masse croissante. Ce mode d'acquisition permet de contourner le problème de sélection stochastique du mode DDA. Jusqu'à présent il y a un manque d'outils bioinformatiques efficaces (algorithmes) pour permettre une utilisation facile des données générées par DIA.

La compréhension du fonctionnement de l'interprétation des données nécessite de bien connaître les mécanismes de fragmentation des ions peptidiques.

B. L'identification et validation des spectres de fragmentation

Les stratégies basées sur l'analyse protéomique de l'ensemble des peptides issus d'une digestion enzymatique pour identifier les protéines présentes dans un échantillon, sont désignées par le terme «*bottom-up*».

Une analyse LC-MS/MS d'un digestat protéique peut générer plus de 100 000 spectres de fragmentation. Le traitement et l'interprétation des données obtenues nécessitent l'utilisation d'outils bioinformatiques spécifiques et performants pour remonter à l'identification des protéines à partir des masses des fragments de leurs peptides.

L'interprétation des données MS/MS est possible à l'aide de moteurs de recherche qui vont permettre d'identifier les peptides grâce à leurs fragments par comparaison à des banques de données de référence. Une fois que les données sont interprétées elles vont être validées grâce à des logiciels adaptés.

L'interprétation et la validation des données ont été présentées ci-dessous.

B.1. Les banques de données protéiques

Les banques de données protéiques ont une place centrale dans l'identification des protéines.

Les banques de données protéiques accessibles publiquement ne possèdent pas toutes le même degré de qualité de l'information. L'identification des protéines est entièrement dépendante de la qualité des bases de données. Seules les protéines répertoriées dans la banque peuvent être identifiées dans les échantillons analysés.

Il existe deux principaux types de banques :

- Les banques de dépôt, telles que NCBI (*National Center for Biotechnology Information*) qui répertorient les séquences protéiques issues des banques génomiques. De par leur volume, ces banques exhaustives présentent l'inconvénient de posséder un nombre important de redondances et de ne pas être validées par des experts.
- Les banques corrigées qui répertorient des séquences protéiques décrites dans la littérature et corrigées par des annotations de hautes qualités. La principale ressource disponible est UniProt (*Universal Protein Resource*) dont la base de données est UniProtKB (*UniProt Knowledgebase*). La base de données fournit, entre autres, des informations sur la fonction, la localisation cellulaire, les pathologies associées, les PTM, la structure, etc ... La banque UniprotKB regroupe la banque Swissprot (vérifiée et annotée) et la banque TrEMBL (annotée automatiquement, non validée expérimentalement).

La banque Swissprot a été utilisée pour tous les projets présentés dans la thèse. Le traitement de données a été réalisé avec des banques restreintes à la taxonomie étudiée. Deux raisons à cela, premièrement pour le gain de temps dans la recherche de banques de données (limite l'espace de recherche), deuxièmement, pour limiter la probabilité d'obtenir des faux-positifs (résultats considérés comme vrai alors qu'ils sont faux). Plus l'espace de recherche est grand plus le pourcentage de faux-positifs augmente.

B.2. Les algorithmes de comparaison ou moteurs de recherche

La méthode d'identification des peptides appelée « empreinte de fragments peptidiques » (*Peptide Fragment Fingerprinting*, PFF) a été développée en 2004³⁷. Cette méthode consiste à identifier les protéines présentes dans l'échantillon à l'aide d'une banque de données protéiques. La liste des m/z expérimentaux obtenus à partir des spectres de fragmentation est comparée à la liste des m/z calculés pour les protéines digérées et fragmentées *in silico* dans la banque.

La comparaison des données acquises et celles de la banque de données est réalisée grâce à des algorithmes de moteurs de recherche adaptés. De nombreux moteurs de recherche ont été développés au cours des dernières années, tels que Mascot³⁸⁻³⁹, OMSSA⁴⁰ et Andromeda⁴¹. Néanmoins ce sont les mêmes informations renseignées pour réaliser la comparaison telles que :

- La tolérance de masse acceptée pour les rapports m/z de l'ion peptidique et de l'ion fragment
- Les charges acceptées pour l'ion peptidique et l'ion fragment
- L'enzyme de digestion utilisée et le nombre maximal de clivages manqués tolérés
- La banque de séquences protéiques
- Le mode de fragmentation

Lors de l'interprétation un score (statistique) est calculé pour chaque spectre MS / MS, dépendant de la qualité des spectres. Les scores dépendent de la probabilité que l'assignation faite soit liée au hasard. Plus le score est élevé, plus la probabilité de correspondance est élevée (non liée au hasard) et donc l'identification du peptide est sûre. Un score d'identité est également calculé en fonction de la taille de la banque dans laquelle la recherche a été faite. Ce score permet d'évaluer le score du peptide en fonction de la distribution aléatoire des scores. Plus le nombre d'entrées de protéines est élevé dans une banque, plus le score d'identité sera élevé³⁹.

Dans le travail présenté dans ce manuscrit, les moteurs de recherche utilisés ont été Mascot^{38,39} et Andromeda⁴¹.

B.3. La validation statistique des données

Lors de l'interprétation des spectres il se peut qu'un peptide soit assigné à spectre MS/MS par erreur (résultats faux positifs). Il est donc nécessaire d'estimer le nombre de résultats faux positifs.

Pour cela des stratégies dites de « leurre » appelée stratégie *target-decoy* ont été développées. Le principe de la stratégie repose sur l'implémentation de banques protéiques générées artificiellement avec des séquences protéiques inversées (*decoy*), dans la banque protéique utilisée (*target*). Le nombre de protéines identifiées entre les séquences *target* par comparaison aux séquences *decoy* permet de déterminer un FDR (*False Discovery Rate*) en suivant l'équation :

$$FDR = 2 * \frac{\text{Nombre de decoy identifiés}}{(\text{Nombre de decoy} + \text{Nombre de target})}$$

Classiquement les données obtenues en analyse protéomique sont validées avec un FDR le plus souvent inférieur ou égal à 1% au niveau des spectres assignés à un peptide. Un FDR au niveau protéique peut également être utilisé pour valider les données.

Les logiciels Proline⁴² développé par ProFi (Infrastructure Française de Protéomique, <http://proline.profiroteomics.fr/>), MaxQuant⁴³ avec son propre moteur de recherche (Andromeda) Scaffold⁴⁴ (Proteome software) ont été utilisés dans mes travaux de thèse.

C. Les stratégies de quantification en analyse protéomique

Idéalement, l'analyse protéomique d'un échantillon doit permettre d'identifier les protéines mais également doit fournir des informations quantitatives. Alors que dix ans auparavant, la plupart des expériences protéomiques énuméraient les constituants protéiques d'un système biologique, les mesures quantitatives sont désormais au cœur de pratiquement toutes les études protéomiques d'aujourd'hui.

La complexité élevée des échantillons et la plage dynamique étendue des protéines dans les systèmes vivants représentent des défis dans l'analyse protéomique complète. Les approches par spectrométrie de masse ne cessent d'évoluer pour détecter et quantifier les protéines à faible abondance en présence de protéines à forte abondance⁴⁵. Par conséquent, de nombreuses technologies ont émergé au cours de la dernière décennie, y compris des stratégies analytiques pour effectuer des méthodes d'enrichissement basées sur des méthodes de ciblage. Le marquage chimique est une méthode couramment utilisée dans l'analyse protéomique pour la mesure quantitative⁴⁶.

Chaque méthode de quantification doit être adaptée en fonction de la nature de l'échantillon, le nombre d'échantillon, le nombre de traitements à comparer, le type d'équipement disponible et, bien entendu, le coût et le temps.

Il existe des méthodes de quantification absolue ou relative. La quantification absolue détermine la quantité ou la concentration exacte des protéines dans l'échantillon. La quantification relative différentielle est basée sur la comparaison de deux états biologiques du même échantillon. Les résultats sont alors exprimés sous forme de rapport d'abondance.

Deux grands types de stratégies de quantification peuvent être distingués :

- 1- Quantification globale des protéines dans l'échantillon, réalisée à l'aide de stratégies sans marquage (stratégies dites *label Free*) ou de stratégies avec marquage (portant des étiquettes ou « tag »).
- 2- Quantification ciblée de protéines spécifiques présentant un intérêt pour l'étude, réalisée à l'aide de stratégies SRM ou PRM.

Les différentes stratégies de quantification de protéines sont présentées rapidement dans la Figure 11. Lors de ma thèse nous nous sommes intéressés principalement aux stratégies de protéomique globale.

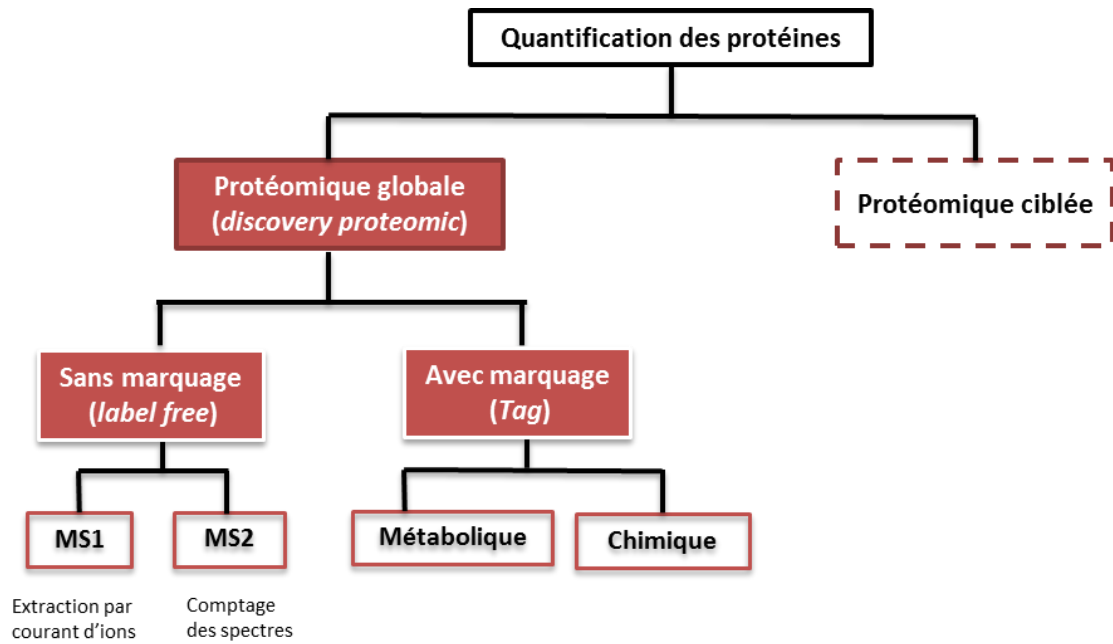


Figure 11: Représentation schématique des principales stratégies protéomiques pour la quantification de protéines

Les stratégies de quantification ciblées de type SRM et PRM sont très brièvement abordées tandis que les stratégies de quantification globales sont présentées de façon plus détaillée dans les paragraphes suivant.

C.1. La stratégie de quantification ciblée : SRM (Selected Reaction Monitoring)

L'approche SRM est une approche quantitative basée sur l'analyse de protéines sélectionnées au préalable. Pour l'analyse quantitative ciblée de ces protéines d'intérêt, une sélection de plusieurs leurs peptides et des fragments de leurs peptides selon différents critères (peptides protéotypiques, pas de méthionine dans la séquence, etc...) va être établie. La sélection d'un peptide avec son fragment est appelée transition. La SRM est principalement réalisée, sur un spectromètre de masse triple quadripôle QQQ, qui permet d'analyser spécifiquement les transitions sélectionnées pour les protéines ciblées.

Les protéines vont être quantifiées par extraction des courants d'ions à partir des fragments (MS2). L'utilisation de standards internes ajoutés en concentrations connues (peptides ou des protéines marqués aux isotopes stables), va permettre de quantifier les protéines avec le rapport des intensités des peptides marqués/peptides endogènes. En fonction du standard utilisé les protéines pourront être quantifiées de manière absolue ou relative.

Une sélection de multiples transitions pour un seul peptide dans une même analyse permet d'augmenter la probabilité de détection et la justesse de la quantification du peptide à quantifier et l'absence d'interférences qui pourraient être lié à la coélution avec d'autres peptides. Ce type d'analyses dites ciblées est beaucoup plus sensible que les stratégies de quantification dites globales.

L'utilisation de standards, permet de s'assurer de la reproductibilité et de corriger les biais techniques.

Il existe deux limitations principales pour cette approche à savoir le nombre limité de peptides qui peuvent être quantifiés dans une analyse LC-SRM (environ 50 peptides par analyse soit 150 transitions) et le temps de développement de la méthode.

Récemment développée, une nouvelle approche appelée PRM⁴⁷ (*Parallel Reaction Monitoring*) permet de suivre simultanément tous les fragments d'un peptide donné avec des instruments de haute résolution (exemple : Q-OrbitrapTM).

C.2. Les stratégies de quantification globale du protéome

Depuis quelques années, des méthodes de quantification à grandes échelles pour l'analyse protéomique par spectrométrie de masse ont été développées⁴⁶. La quantification des protéines permet d'évaluer les variations d'abondance protéique en comparant des quantités relatives de protéines présentes dans différents états biologiques. Pour cela, les intensités des signaux MS ou MS/MS des peptides d'une même protéine sont comparés. Une régulation biologique différentielle pourra, par exemple, être mise en évidence en comparant une condition contrôle versus une condition physiologique particulière.

Les stratégies sans marquage dites *label free* peuvent être plus attrayantes en termes de coûts et de mise en place. Cependant ce type de quantification ne permet pas de détecter des variations fines mais plutôt des différences d'abondance protéiques élevées. Les stratégies avec marquage permettent *a contrario* de déterminer des régulations biologiques plus fines des taux d'expression⁴⁸ protéique. Les différentes stratégies *label free* sont décrites ci-dessous.

Les différentes stratégies de quantification globale sont présentées dans la Figure 12.

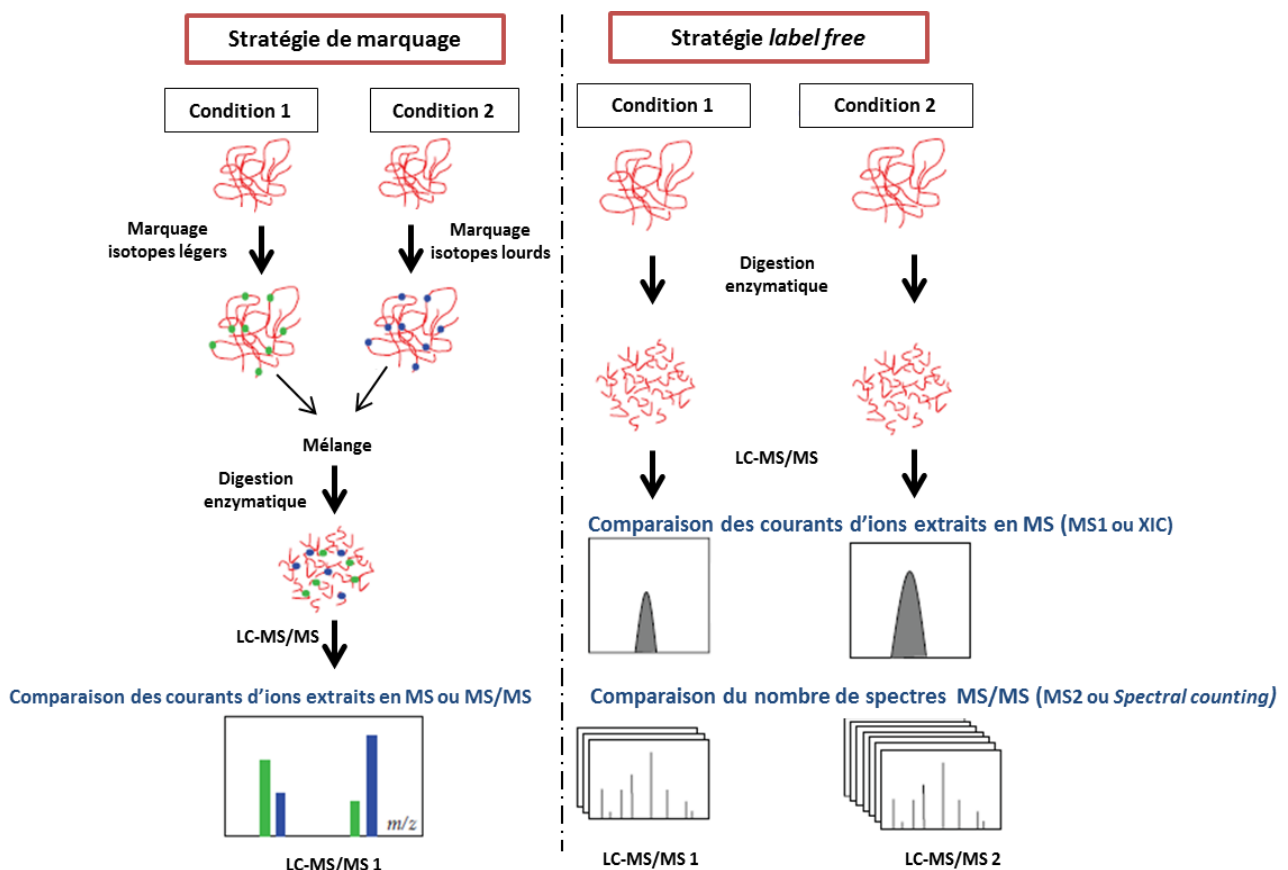


Figure 12 : Représentation schématique des principales stratégies de quantification globale en protéomique. Les stratégies *label free* : qui comparent les aires des pics chromatographiques (MS1, XIC) ou qui comparent le nombre de spectres acquis (MS2 ou *spectral counting*) pour un même peptide dans deux échantillons analysés séparément. Les stratégies par marquage isotopique comparent les aires des pics chromatographiques d'un même peptide dans deux échantillons analysés ensemble (MS et/ou MS/MS). La figure a été adaptée de W. Zhu⁴⁹.

C.2.1 Les approches sans marquage (*label free*)

Le comptage des spectres MS2 (*Spectral counting*).

La méthode de quantification « *spectral counting* » est défini comme le nombre total de spectres de fragmentation générés pour un peptide après analyse LC-MS/MS en mode DDA. Le nombre de spectres est un index d'abondance dans cette méthode, mais il doit bien entendu être ramené à la masse de la protéine.

En théorie, plus une protéine est abondante plus il est probable qu'elle soit sélectionnée en MS et fragmentée en MS/MS⁵⁰. La variation d'abondance d'une protéine d'un échantillon à un autre est mesurée par comptage des spectres entre les analyses MS/MS⁴⁹⁻⁵¹. Un maximum de spectres informatifs doit donc être acquis pour permettre une quantification robuste. L'ajustement du temps d'exclusion dynamique par exemple, fait partie de paramètres instrumentaux à ajuster pour cette stratégie quantitative. En effet, le temps d'exclusion d'un ion va directement impacter sur le nombre de fois où l'ion va être fragmenté et par conséquent sur le nombre de spectres générés. Cela peut affecter le nombre d'identifications. La redondance des protéines majoritaires peut parfois masquer le signal des protéines minoritaires⁴⁸ (effet de « sous échantillonnage »). Le traitement de données de cette approche est relativement simple à mettre en œuvre. Cependant la méthode présente une faible sensibilité par rapport à l'autre méthode de *label free*. Les performances atteintes en termes de vitesse de balayage avec les nouveaux instruments ont relayé le *spectral counting* au second plan au profit de la méthode de quantification par extraction des courants d'ions.

L'extraction des courants d'ions MS1 (XIC).

La méthode de quantification est basée sur l'extraction des courants d'ions MS (XIC, *eXtracted Ion Chromatogram*) en mode DDA. Les informations quantitatives sont obtenues en MS, à partir de l'intégration des aires sous les pics chromatographiques pour un peptide à un temps de rétention et un rapport m/z donné. Le courant d'ion va être extrait dans l'ensemble des analyses LC-MS/MS. Les courants d'ions extraits des peptides d'une protéine pourront ainsi déterminer l'abondance relative de la protéine dans chaque analyse.

L'approche n'est pas limitée par la sélection stochastique du peptide pour la MS/MS. En effet même si un peptide n'a pas été fragmenté dans une analyse LC-MS/MS spécifique (effet de « sous-échantillonnage »), il peut être quantifié de manière différentielle s'il était présent avec une intensité suffisante pendant le balayage MS. Cela rend la quantification par MS1 plus robuste que par MS2 (*Spectral counting*).

Contrairement à la méthode de comptage de spectre, l'extraction des courants d'ions nécessite un traitement bioinformatique adapté. Deux types principaux de logiciels permettent de générer des données XIC:

- Les logiciels qui sont basés sur l'extraction des courants de tous les ions détectés. Ils vont aligner les chromatogrammes et détecter toutes les masses des ions moléculaires (MaxQuant⁴³). Cette approche est plus robuste face aux problèmes de reproductibilité des temps de rétention chromatographique mais nécessite une précision de masse et une haute résolution.
- Les logiciels basés sur l'extraction des courants d'ions des peptides dont l'assignation a été validée. Avec le logiciel Skyline⁵² par exemple, une librairie de spectre MS/MS est créée puis le logiciel extrait les courants d'ions des peptides validés à partir des données brutes.

C.2.2 Les avantages et les limitations des stratégies *label free*

Ces stratégies sans marquage, présentent l'avantage d'être facile : à mettre en œuvre et ne sont pas limitées en termes de nombre d'échantillon à comparer. Peu coûteuses, elles sont théoriquement adaptables à tout type d'échantillon.

Par comparaison aux méthodes de marquage isotopique, dans les méthodes *label free*, les digests peptidiques sont analysés un par un. Il est donc nécessaire de s'assurer de la bonne reproductibilité du système de séparation chromatographique au cours du temps. En effet la robustesse de l'analyse est un point clé pour la qualité de la quantification. L'analyse de réplicas est nécessaire pour permettre de vérifier les biais techniques et biologiques qui peuvent être introduits au cours du temps ou de la préparation d'échantillon. Les paramètres instrumentaux doivent être adaptés en fonction de la complexité des échantillons pour trouver un équilibre entre l'identification et la robustesse des résultats.

Nous notons que l'évolution instrumentale des dernières années et les outils bioinformatiques ont clairement amélioré la fiabilité et la précision des approches *label free*⁴⁹. Cependant le temps d'analyse global pour une expérience et le nombre de données générées (des milliers de spectres par analyse) à traiter peuvent rapidement devenir important en fonction du nombre et de la complexité des échantillons.

C.2.3 Les approches par marquage isotopique

Les approches de quantification par marquage des peptides et des protéines sont basées sur l'utilisation d'isotopes stables. Les peptides marqués aux isotopes lourds et les peptides non marqués ont les mêmes propriétés physico-chimiques seul leur rapport m/z est différent. Les peptides auront donc le même comportement lors de la séparation chromatographique et dans le spectromètre de masse (facteur de réponse, efficacité de l'ionisation, fragmentation). Plusieurs échantillons peuvent ainsi être comparés dans la même analyse LC-MS/MS. Le nombre d'échantillon à comparer est limité en fonction des isotopes utilisés. La quantification relative des protéines est réalisée par comparaison des aires sous les pics chromatographiques en MS ou en MS/MS selon la stratégie. Parmi les principales stratégies de quantification par marquage il y a :

- Le marquage métabolique par incorporation d'acides aminés marqués en culture: SILAC⁵³ (*Stable Isotopic Labeling using Amino Acids in Cell Culture*). SILAC présente l'avantage majeur d'introduire l'étape d'étiquetage au stade le plus précoce possible et permet d'éviter l'introduction de biais techniques qui biaiseraient la quantification.
- Le marquage chimique des protéines et des peptides : Le groupe réactif d'un peptide va être modifié en utilisant des marqueurs isotopiques légers et lourds. Les approches les plus connues sont ICAT⁵⁴ (*Isotope Coded Affinity Tag*, dérivation des résidus spécifiques sur les protéines intactes), iTRAQ⁵⁵ (*isobaric Tags for Relative and Absolute Quantitation*) ou le TMT⁵⁶⁻⁵⁷ (*Tandem Mass Tags*, basée sur la dérivation de résidus spécifiques sur les peptides de digestion). TMT permet de comparer jusqu'à 10 échantillons en MS2 .

Le chapitre III dans la partie II Résultats, est consacré au développement de nouveaux outils de bioconjugaison pour l'analyse protéomique. Dans ce chapitre, les points clés et les limitations des stratégies de quantification par marquage ont été détaillés.

Parti II Résultats

Chapitre I Etudes de complexes protéiques cibles impliqués dans le mécanisme de la mélanogénèse

Partie II Résultats

Chapitre I Etudes d'un complexe protéique impliqué dans le mécanisme de la mélanogénèse

A. Introduction biologique

A.1. La mélanogénèse ou pigmentation

La couleur de la peau et des cheveux repose essentiellement sur la quantité et le type de mélanines qu'ils contiennent. La mélanogénèse, ou pigmentation, est le mécanisme biologique de synthèse des mélanines qui se produit dans les cellules spécialisées, les mélanocytes. Chez l'Homme, la mélanine dans la peau, les cheveux et les yeux, protège le corps contre les agressions environnementales telles que les rayons UV⁵⁸, les radicaux libres et les métaux lourds.

L'hyperpigmentation ou hypopigmentation sont des troubles cutanés importants. Les dysfonctionnements dans la pigmentation et la biogenèse des mélanosomes sont associés à une grande variété de troubles génétiques héréditaires et de maladies pigmentaires (albinisme type oculo-cutané, syndromes de Hermansky-Pudlak...)⁵⁹.

La pigmentation relève d'un processus complexe, qui implique plus de 125 *loci*⁵⁹ et des fonctions cellulaires complémentaires parmi lesquelles la biosynthèse des mélanosomes, l'adressage des protéines mélanosomales, la production de mélanines, le transfert des mélanines au kératinocytes. Ce mécanisme complexe, au cours duquel sont synthétisées les mélanines, se produit dans les mélanocytes au sein des mélanosomes. Les mélanosomes sont des organites de la famille des lysosomes spécialisés dans la synthèse et la distribution de mélanine.

A.1.1 Les mélanosomes

Bien que des observations en microscopie électronique aient permis d'obtenir des informations sur l'ultrastructure des mélanosomes dès le début des années 60, la composition protéique et le mécanisme de biogenèse sont longtemps restés incompris⁶⁰. L'hétérogénéité des mélanosomes observés dans les mélanocytes explique en partie qu'il est été difficile d'appréhender leur formation et leur origine⁶¹. Les mélanosomes ont été classés en quatre stades de développement de I à IV en fonction de leur morphologie et du degré de mélanisation⁶².

Les mélanosomes de stade I et II. Aussi appelés « mélanosomes immatures », ils ne synthétisent pas encore de mélanine. Ils sont composés de fibres protéiques de nature amyloïde, liées à la polymérisation d'une protéine transmembranaire Pmel17⁶³. Une fois que les fibres protéiques sont complètement formées dans les mélanosomes de stade II, la synthèse de mélanine commence⁶⁴. Les protéines impliquées dans la synthèse de la mélanine sont transportées vers les mélanosomes de stade II, et vont permettre progressivement le passage aux stades suivants⁶⁴.

Les mélanosomes de stade III et IV. Ces mélanosomes sont des formes « matures ». Ils contiennent des quantités croissantes de mélanine qui va alors être déposée sur les fibres protéiques⁶³. Les mélanosomes se déplacent de la région périnucléaire vers les extrémités dendritiques des mélanocytes. Les mélanosomes sont ensuite capturés par des filaments d'actine et transférés vers les kératinocytes adjacents⁶⁵.

A.1.2 La synthèse de la mélanine : L'eumélanine et la phéomélanine

La synthèse de la mélanine fait intervenir des enzymes telles que la tyrosinase (TYR, EC 1.14.18.1), enzyme limitante de la mélanogénèse⁶⁶. L'activité de la protéine tyrosinase est nécessaire pour la synthèse de deux types de mélanines : l'eumélanine de couleur marron à noir (la plus photoprotectrice)⁵⁸ et la phéomélanine de couleur jaune à rouge-brun. Ces deux mélanines dérivent du même précurseur, la tyrosine.

L'eumélanine est plus abondante chez les êtres humains, en particulier chez les personnes à peau foncée. La phéomélanine est présente dans les cheveux et la peau humaine mais en quantité plus faible que l'eumélanine. La phéomélanine est retrouvée en quantité importante principalement dans les cheveux roux et dans les éphélides (taches de rousseur)⁵⁸.

La synthèse de l'eumélanine nécessite la liaison de l'hormone alpha MSH au récepteur MC1R qui active la transcription de la tyrosinase. En absence de la signalisation MC1R^{58, 67} la phéomélanogénèse, avec une faible activité de tyrosinase et dépendante de la présence de cystéine⁶⁸, semble être la voie par défaut.

La voie de l'eumélanine est plus largement étudiée et de fait connue. La signalisation par la voie des MC1R active la transcription de la tyrosinase et augmente celle de la protéine *tyrosinase related proteine 1* (TYRP1, EC 1.14.18) et de la protéine DOPAchrome tautomerase (DCT, EC 5.3.2.3)⁵⁸.

A.1.3 Les enzymes mélanosomales dans la voie des eumélanines

La mélanogénèse, mécanisme enzymatique, fait intervenir trois enzymes principales : TYR, TYRP1, et DCT⁶⁸⁻⁶⁹. Ce sont des glycoprotéines membranaires situées préférentiellement dans les organites spécialisés, les mélanosomes. Elles font parties de la famille des métalloprotéines et possèdent par conséquent un groupement prosthétique métallique spécifique tels que le cuivre pour TYR, le zinc pour DCT et enfin le fer pour TYRP1⁷⁰. Ces protéines partagent 40% d'homologie de séquence.

Il a été montré que la TYR oxyde la tyrosine en DOPAquinone⁶⁸. Après cette première étape, la DOPAquinone est spontanément convertie en DOPAchrome (Figure 13). En présence de la DCT, le DOPAchrome est isomérisé en acide 5,6-dihydroxyindole-2-carboxylique (DHICA). En absence de DCT, le DOPAchrome peut alors former spontanément le 5,6-dihydroxyindole (DHI), puis être oxydé en une quinone sous l'action de la tyrosinase. Chez la souris, le DHICA va être oxydée par TYRP1 en quinone instable qui polymérise pour donner l'eumélanine⁶⁸ (Figure 13). Cette activité pour la protéine TYRP1 humaine n'a pas pu être démontrée jusqu'à présent.

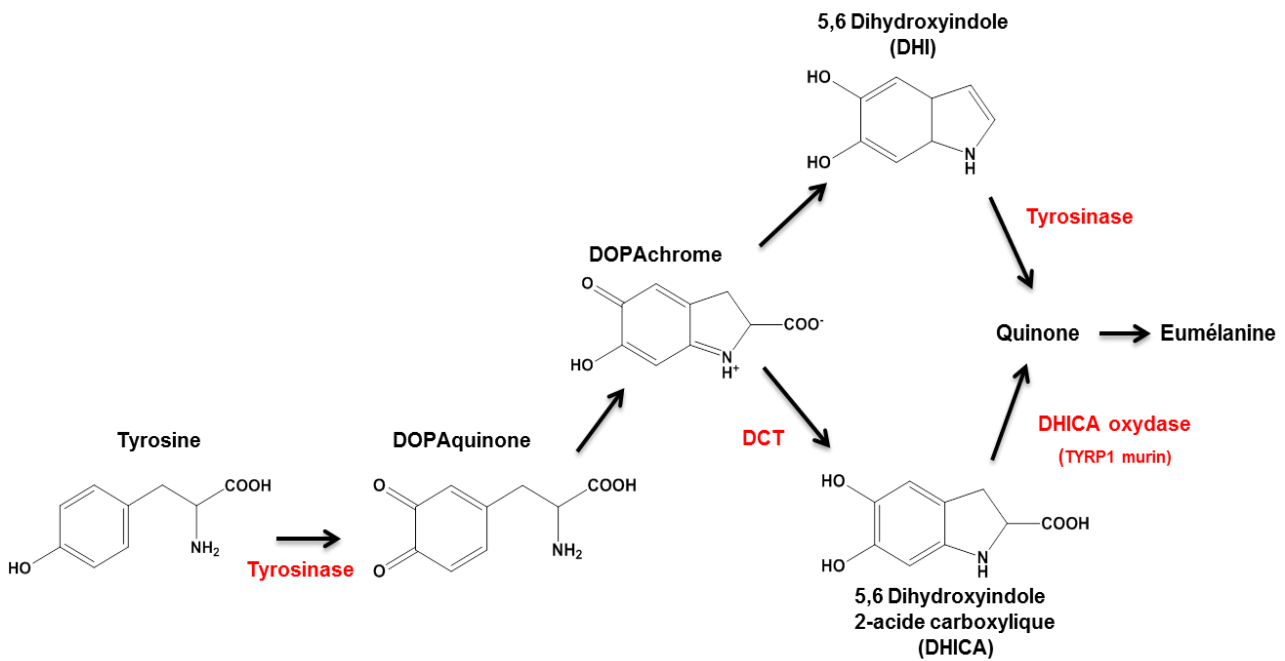


Figure 13: Représentation schématique de la voie de synthèse de l'eumélanine faisant intervenir les trois enzymes mélanosomales : TYR, TYRP1 et DCT.

A.2. La coopération des enzymes mélanosomales : Formation de complexe

Chez l'homme, les mutations du gène de la protéine TYR aboutissent à la forme la plus sévère d'albinisme oculo-cutané (OCA) alors que les mutations du gène de TYRP1 provoquent une forme moins sévère d'OCA⁷¹. Dans la littérature, plusieurs études sur les enzymes de la mélanogénèse ont mis en évidence que les protéines TYRP1 et DCT, en plus d'avoir des fonctions enzymatiques, stabilisent TYR⁷²⁻⁷⁴ et que la coexpression des trois protéines dans les mélanocytes augmente la synthèse de la mélanine⁷⁵⁻⁷⁶.

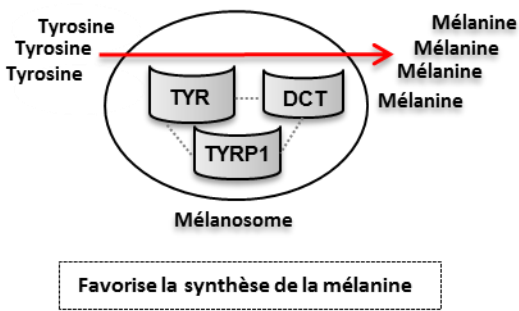
En 1994, une étude a montré la présence des enzymes dans des complexes moléculaires trouvés dans des lysats de mélanocytes par copurification⁷⁷.

La formation de complexe entre TYR et TYRP1 a été montré par des approches de chromatographie sur gel^{66,70}. Le complexe TYR-TYRP1 apparaît favorable à la production de mélanines, l'activité TYR et le taux de TYR étant plus élevés dans des conditions favorables à la formation du complexe avec TYRP1^{66, 70, 77}.

Ces résultats suggèrent que les enzymes forment un complexe multiprotéique à la surface interne de la membrane mélanosomale comme cela est présenté dans la Figure 14.

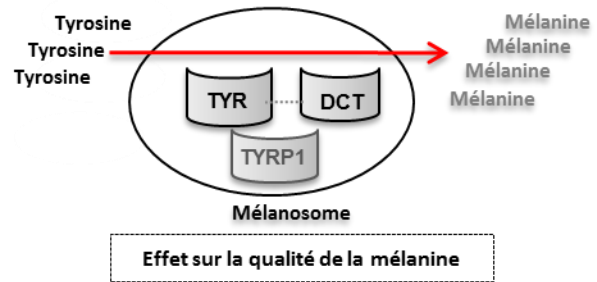
a) Condition physiologique

Co-expression de TYR, TYRP1 et DCT



b) Conditions pathologiques

Mutation de TYRP1



Mutation de TYR

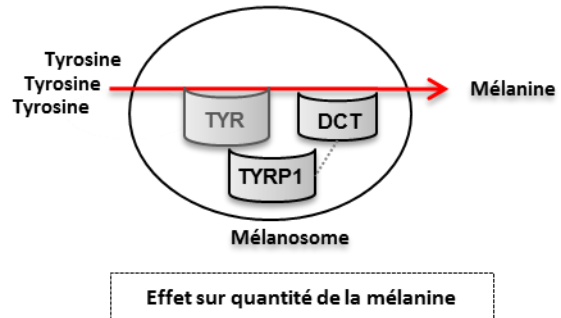


Figure 14: Représentation schématique de l'impact sur la synthèse de la mélanine de l'expression des trois protéines TYR, TYRP1 et DCT. a) En condition physiologique, la coexpression des trois protéines favorise la formation d'un complexe multiprotéique et stabilise l'activité de TYR. b) En conditions pathologiques, l'absence de TYR ou de TYRP1 déstabilise la formation du complexe et altère la synthèse de la mélanine.

A ce jour, aucune analyse par spectrométrie de masse n'a permis de confirmer la présence des trois protéines dans les bandes de hauts poids moléculaires mis en évidence dans les études de coprécipitations publiées. Le mécanisme contrôlant la stabilité de la tyrosinase demeure inconnu jusqu'à présent.

En 2013, après avoir prédit la structure 3D de la tyrosinase humaine W. Wang et son équipe ont proposé des partenaires d'interactions de la tyrosinase⁷⁸. Après simulation de la structure 3D de TYR, l'algorithme PEIMAP (*Protein Experimental Interactome MAP*) a permis de prédire 108 protéines en interaction directe ou indirecte avec TYR, en particulier TYRP1 et DCT. Des acides aminés potentiellement impliqués ont également été déterminés dans ces interactions protéine-protéine.

La connaissance précise de la structure du complexe TYR-TYRP1, des conditions d'une part favorables et d'autre part défavorables à sa stabilité, des mécanismes impliqués dans la stabilisation de l'activité TYR, pourraient permettre de développer des stratégies de stabilisation ou déstabilisation de ce complexe. Ceci pourrait donc aboutir à des innovations et des applications cosmétiques pour moduler la couleur de la peau (ou des cheveux). Il est donc clair qu'une meilleure connaissance des mécanismes de régulation, par exemple les mécanismes enzymatiques sous-jacents de la mélanogénèse, serait utile d'un point de vue biomédical, mais aussi dans le domaine de la cosmétique.

A.3. Les objectifs de l'étude

Pour progresser dans la compréhension du mécanisme de synthèse des mélanines, il est indispensable de déterminer quelles sont les autres protéines qui interagissent avec le complexe TYR/TYRP1/DCT.

L'objectif de notre travail est de développer et d'appliquer des approches d'analyse protéomique pour identifier les protéines partenaires qui interagissent avec le complexe TYR/TYRP1/DCT.

Un prérequis incontournable à l'identification des partenaires. L'identification des partenaires du complexe nécessite un isolement préalable de ce complexe multiprotéique avec une pureté compatible avec les contraintes d'une analyse par spectrométrie de masse. Une grande diversité de méthodes d'isolement de complexes spécifiques a été décrite par les biologistes ces dernières décennies (cf. Partie II -B.). Cet isolement constitue donc une étape clé pour l'identification des protéines partenaires. La purification d'un complexe spécifique à partir d'un extrait protéique nécessite, la plupart du temps, des mises au point pour chaque complexe ciblé. En effet, il s'agit d'isoler à partir d'une matrice biologique, un complexe particulier constitué de protéines associées par des interactions non covalentes spécifiques. En fonction de la nature et de la spécificité des interactions mises en jeu dans un complexe particulier, les conditions expérimentales doivent être optimisées pour conserver ce complexe à l'état natif.

La notion de stringence de la purification lors de l'isolement d'un complexe à partir d'une matrice biologique est un point clé. Cette stringence peut se faire en réalisant une purification plus ou moins poussée (une ou plusieurs étapes consécutives) mais également en jouant sur la stringence de lavage et d'élution qui dépend de la force saline, du pH, de l'hydrophobie, de la nature du tampon, de la température...).

Un milieu à forte stringence provoquera la perte de partenaires ayant des interactions spécifiques fortes avec le complexe. Un milieu à faible stringence ne permettra pas d'éliminer des partenaires non spécifiques. Il faudra donc trouver un juste équilibre entre une stringence qui élimine les interactions non spécifiques et une stringence qui préserve les interactions spécifiques. La frontière entre ces deux conditions peut être très ténue, ce qui explique les difficultés très fréquemment rencontrées pour isoler des complexes.

La spectrométrie de masse : un outil pour valider l'isolement d'un complexe spécifique. Lorsqu'une des nombreuses stratégies développées ces dernières décennies par les biologistes est mise en œuvre pour tenter d'isoler un complexe particulier, il est essentiel de disposer d'un outil, pour déterminer si cette stratégie a effectivement permis d'obtenir le résultat attendu. C'est-à-dire si la stratégie a permis d'identifier un nombre restreint de protéines présentant des interactions spécifiques. C'est le rôle que la spectrométrie de masse peut jouer, comme cela a été décrit ces dernières années dans de nombreuses études d'interactomique. En effet, si une analyse protéomique est réalisée sur l'échantillon sensé contenir un complexe spécifique, la liste de protéines identifiées dans cette étude doit être en cohérence avec ce qui est connu du complexe. Si cette analyse montre que plusieurs dizaines, voire plusieurs centaines de protéines sont présentes, il est à craindre que l'isolement d'un complexe spécifique n'a pas été réalisé.

L'analyse protéomique par spectrométrie de masse est donc un outil performant pour guider les biologistes dans la mise au point d'une méthode d'isolement d'un complexe. Notre laboratoire a développé et appliqué avec succès cette approche dans l'étude de nombreux complexes ces dernières années^{79 80 81}.

Dans le projet décrit ici, comme nous le montrerons, notre contribution a principalement porté sur cet aspect. Il s'agira donc d'adapter et/ou de développer les méthodologies d'analyse protéomique aux contraintes de ce projet, qui sont les faibles quantités ainsi que la complexité du milieu.

L'isolement d'un complexe spécifique n'est pas routinier. De nombreuses stratégies ont été décrites pour l'isolement de complexes multiprotéiques. Nous présenterons les principales d'entre elles dans le chapitre Partie II -B.

La technique la plus utilisée consiste à purifier par affinité et dans des conditions natives (c'est-à-dire en préservant toutes les interactions non covalentes) une protéine d'intérêt, identifiée comme faisant partie du complexe, afin d'enrichir le complexe dans lequel elle serait impliquée. Schématiquement, une protéine cible

ou « appât » est capturée spécifiquement en espérant qu'elle entrainera les protéines qui ont de l'affinité pour elle (Figure 15).

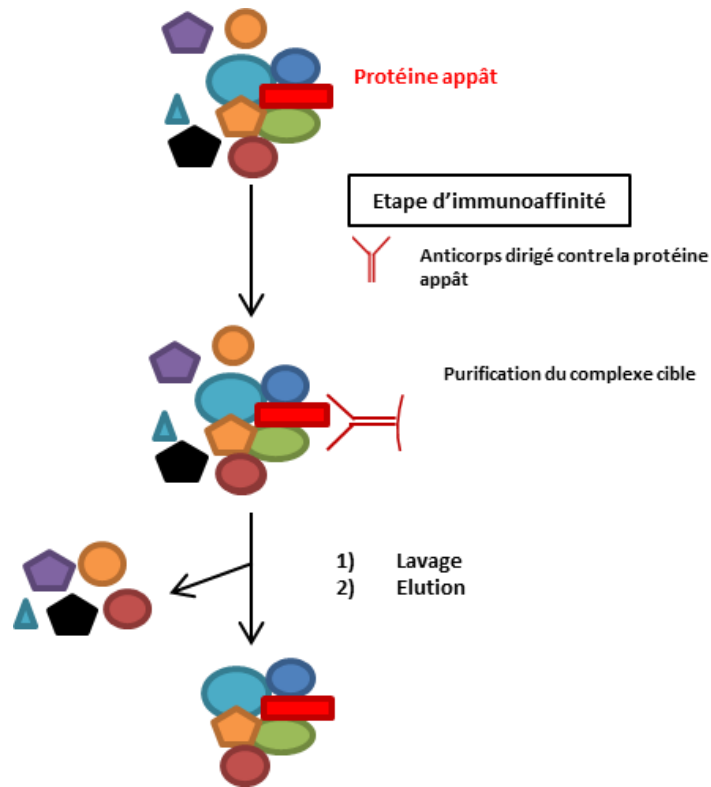


Figure 15: Représentation schématique de la stratégie d'isolement d'un complexe cible par immunoaffinité, basée sur l'utilisation d'anticorps dirigé contre une protéine « appât » du complexe.

L'abondance de littérature sur ce sujet et la diversité des approches décrites, montre que chaque approche nécessite un protocole adapté⁸².

Pour isoler le complexe ternaire TYR/TYRP1/DCT sur lequel seraient fixées les protéines partenaires, la stratégie consiste à trouver les conditions expérimentales qui éliminent les protéines fixées sur le complexe ternaire avec peu d'affinité, donc de façon peu spécifique, et de conserver les protéines partenaires spécifiques. Nous avons utilisé une approche de gel 1D SDS-PAGE et de nano chromatographie liquide de masse en tandem (nanoLC-MS/MS) pour tenter d'optimiser les nombreux protocoles testés pour isoler le complexe ternaire avec ses protéines partenaires spécifiques.

Trois approches ont été testées pour tenter d'isoler le complexe ternaire avec ses partenaires :

- Dans une première approche, nous avons tenté d'obtenir des fractions isolées par immunoaffinité à partir de cultures de mélanocytes (Partie II -C.1).
- Dans une deuxième approche, nous avons adapté une méthode de réticulation chimique (*cross-linking*) sur des cellules ou/et lysat cellulaire pour ponter de façon covalente les lysines impliquées dans les zones d'interaction entre protéines. (Partie II -C.2).
- Dans une troisième approche, nous avons exploré les possibilités d'une nouvelle méthode appelée BioID pour caractériser les protéines partenaires (Partie II -C.3).

L'analyse protéomique de chacune des fractions isolées, grâce aux trois approches mises en œuvre dans ce projet, permettra de mesurer l'efficacité d'isolement de ce complexe ternaire. Si l'analyse protéomique conduit à l'identification de plusieurs centaines de protéines, la méthode d'isolement pourra ainsi être

modifiée. Le rôle de l'analyse protéomique ne se limitera pas uniquement à identifier toutes les protéines présentes mais elle permettra d'orienter le choix de la meilleure stratégie d'enrichissement à utiliser afin d'obtenir des fractions isolées contenant spécifiquement le complexe ternaire et ses interactants potentiels.

B. Introduction au contexte analytique caractérisation de complexes multiprotéiques par MS

Pour assurer leur rôle central dans les mécanismes biologiques dynamiques, les protéines interagissent physiquement les unes avec les autres ainsi qu'avec d'autres molécules (métabolites, lipides...). La formation de complexes protéiques non covalents repose sur des réseaux d'interactions spécifiques. L'étude de la relation structure fonction de ces complexes protéiques est indispensable pour avancer dans la compréhension des processus biologiques. La combinaison des approches de spectrométrie de masse native et des approches d'analyse protéomique, s'est naturellement imposée comme technique d'analyse de choix pour la caractérisation de complexes protéiques.

Les approches d'analyse protéomique dites classiques permettent d'étudier l'agencement primaire de la protéine c'est-à-dire l'ordre des acides aminés (identification, quantification). La MS dite native pourrait permettre, quant à elle, de déterminer la stœchiométrie des partenaires et d'apporter éventuellement des informations sur la structure tertiaire voire quaternaire.

Dans le cadre de la spectrométrie de masse structurale, des informations sur les zones d'interaction pourraient être obtenues. Mais ces études par MS native ne peuvent être envisagées que si nous disposons d'un complexe bien purifié et en quantité suffisante.

L'intérêt croissant pour la détermination de partenaires d'interaction protéiques dans les recherches biologiques a conduit à de nombreuses adaptations et optimisations des stratégies. L'une des principales méthodes d'études des partenaires est l'AP-MS (*Affinity Purification-Mass Spectrometry*), où les partenaires spécifiques d'un complexe vont tenter d'être isolés et identifiés après une ou plusieurs étapes d'immunoaffinité. Cette approche est présentée dans le paragraphe ci-dessous.

B.1. Les stratégies basées sur la purification par affinité couplées à la spectrométrie de masse (AP-MS)

Grâce à la combinaison des techniques de purification par affinité avec la spectrométrie de masse aussi appelée AP-MS (*Affinity Purification-Mass Spectrometry*), il est possible d'isoler des complexes multiprotéiques à partir de lysats cellulaires après une ou plusieurs étapes de purification. Dans la plus part des cas, les stratégies d'AP-MS sont basées sur des méthodes d'immunoprécipitation (IP) avec des anticorps (Ac) et peuvent être combinées avec d'autres approches (fractionnement, réticulation chimique,...). Les protéines présentes dans le complexe sont ensuite identifiées par une analyse protéomique par MS. L'analyse par MS apporte des informations précieuses pour l'identification de partenaires et dans certains cas, pour la détermination de la stœchiométrie du complexe⁸³. Les grandes étapes d'AP-MS et les possibles combinaisons sont présentées Figure 16.

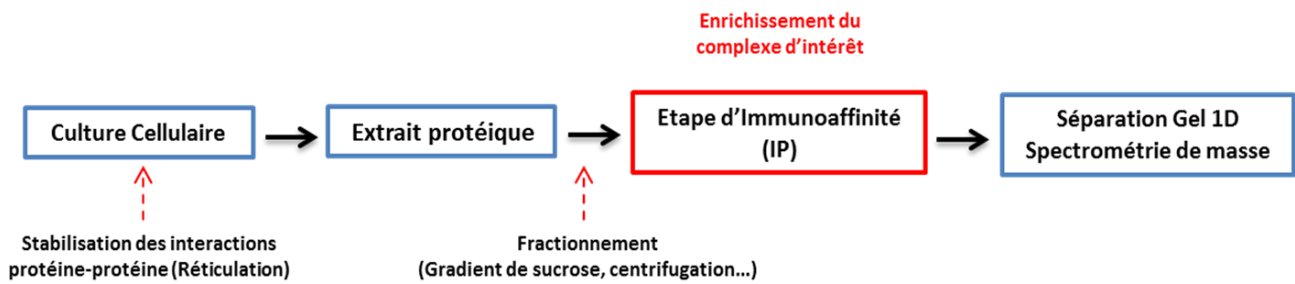


Figure 16: Représentation schématique des étapes des stratégies d'AP-MS, basée sur l'utilisation d'Ac. Les possibilités d'implémenter et de combiner l'AP-MS avec d'autres approches (stabilisation des interactions et fractionnement d'échantillon) sont indiquées par une flèche rouge en pointillé.

Les stratégies d'AP-MS présentent l'avantage de pouvoir être réalisées dans des conditions représentatives du milieu physiologique et ne perturbent pas les PTMs, qui sont souvent cruciales pour l'organisation et/ou l'activité des complexes.

B.1.1 L'étape d'immunoprécipitation (IP)

La protéomique basée sur la spectrométrie de masse combinée à la purification de protéines à base d'étiquette d'affinité est l'une des stratégies les plus efficaces pour isoler et identifier les complexes protéiques.

Ces stratégies de purification par affinité, basée sur une immunoprécipitation (utilisation d'Ac), permettent une purification rapide et efficace de complexes protéiques. L'identification de complexes de protéines, même très peu abondants, avec des spectromètres de masse sensibles modernes est devenue courante.

L'étape d'immunoprécipitation (IP) repose sur l'utilisation d'une protéine « appât » recombinante, impliquée dans un complexe, dont la séquence va être modifiée par l'ajout d'une étiquette spécifique, coté N-ter ou C-ter. L'étiquette reconnue spécifiquement par un anticorps (Ac) va permettre de purifier la protéine recombinante avec ses partenaires d'interaction.

Il existe plusieurs types d'étiquettes (His-TAG, FLAG-TAG...) ⁸⁴⁻⁸⁶ ou combinaisons d'étiquettes (TAP) utilisées avec succès dans les études à haut débit ⁸⁷ (Figure 17). Le choix de l'étiquette fusionnée est déterminant, car elle va influencer la localisation cellulaire, l'activité ou les interactions de la protéine appât...

L'avantage de la stratégie TAP TAG est la pureté de l'échantillon obtenu cependant cette stratégie, lorsqu'elle est utilisée dans les cellules de mammifère et d'insecte, souffre des faibles rendements obtenus ⁸⁶. La culture et la manipulation de ces systèmes de culture exprimant des protéines fusionnées à des étiquettes peuvent nécessiter beaucoup de main-d'œuvre et être coûteuses.

Une autre approche d'immunopurification consiste à utiliser des Ac spécifiques pour une partie de la séquence en acides aminés de la protéine endogène appât impliquée dans le complexe (Figure 17). Dans cette stratégie plusieurs facteurs sont déterminants pour la réussite de la purification le choix de l'Ac, sa spécificité (affinité pour des protéines non spécifiques) et la partie de la protéine contre laquelle l'Ac va être dirigé. Une étude de caractérisation de complexe et de détermination de la stœchiométrie des interactants du complexe, en ciblant six protéines partenaires a d'ailleurs été réalisée avec succès au laboratoire par une stratégie par immunoaffinité couplée à la quantification ciblée (SRM) ⁷⁹.

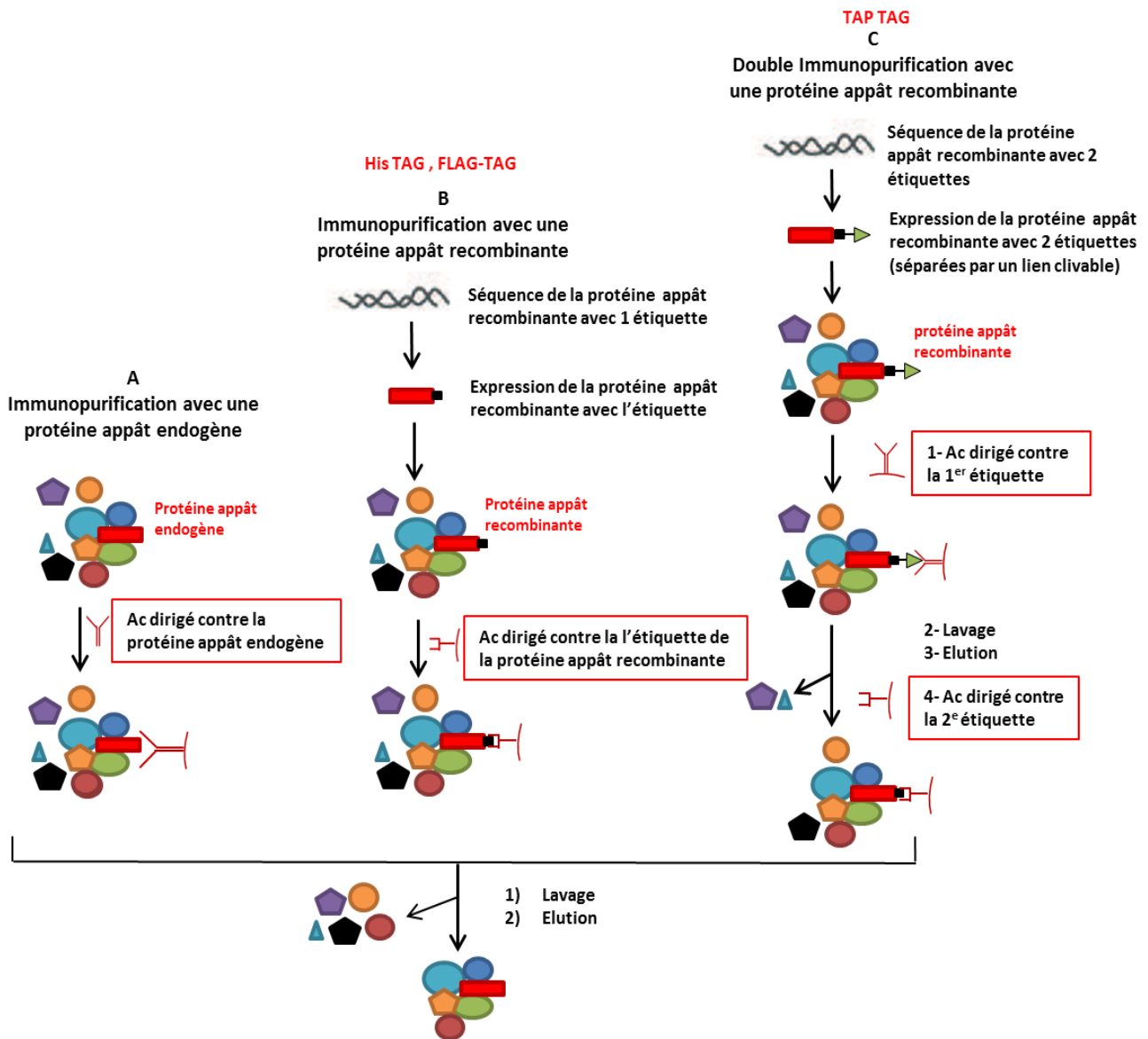


Figure 17: Représentation schématique des stratégies d'immunopurification basées sur l'utilisation d'Ac dirigé contre une protéine appât endogène (A) ou contre une étiquette (B) ou contre plusieurs étiquettes successivement (C) d'une protéine recombinante impliquée dans le complexe cible.

En résumé, la disponibilité, la spécificité d'un Ac pour chaque cible et la mise en place de contrôles appropriés sont les difficultés principales associées à cette approche⁸⁸.

En fonction de la nature et de la spécificité des interactions mises en jeu dans un complexe particulier, des conditions expérimentales précises doivent donc être déterminées afin de conserver les interactions protéine-protéine spécifiques tout en éliminant les interactants non spécifiques. Cependant un milieu à forte stringence provoquera la perte de partenaires ayant des interactions spécifiques fortes avec le complexe⁸⁹. Une des solutions est le maintien des interactions des partenaires d'un complexe par stabilisation par réticulation chimique. La réticulation des protéines permettrait d'extraire de façon plus efficace le complexe et ses partenaires d'interaction.

B.1.2 La stabilisation des interactions protéiques par réticulation chimique

Avant l'étape d'immunoprécipitation il est possible de stabiliser les interactions entre les protéines par réticulation chimique. La réticulation chimique utilise un agent pontant appelé également agent de réticulation, possédant au moins deux groupes réactifs (fonctions amines, sulfhydryles ou carboxyles)⁹⁰ séparés par un bras espaceur (de 5-15 Å).

Ces agents peuvent alors former des liaisons covalentes entre les acides aminés cibles des protéines qui interagissent deux à deux (généralement cystéines ou lysines). Les réticulations vont stabiliser les interactions entre les protéines d'un même complexe lors des différentes étapes de préparation d'échantillon.

Il existe deux principales d'approches par spectrométrie de masse pour étudier les interactions protéiques stabilisées par réticulation chimique :

L'approche *top down* par laquelle les réticulations sont détectées au niveau des protéines intactes d'un complexe, par fragmentation dans l'instrument MS sans digestion enzymatique. L'efficacité d'ionisation et de fragmentation due à la grande taille et à la nature des complexes protéiques restent des limitations. De plus, l'étude est limitée à un nombre restreint de protéines en interaction, et nécessite que les complexes étudiés soient purs. L'obtention de fractions purifiées et les quantités de protéines nécessaires pour l'analyse font parties des limitations de ces méthodes⁹¹.

L'approche *bottom up* est largement utilisée pour la réticulation de protéine en culture par exemple, avec le formaldéhyde et le di-thio-bis-succinimidylpropionate (DSP). Après stabilisation des interactions protéine-protéine, les complexes de haut poids moléculaires sont isolés puis analysés en MS pour identifier de nouveaux partenaires d'interaction⁹². Cette approche sera détaillée dans le paragraphe Partie II -C.2.

B.1.3 Le fractionnement biochimique

Avant l'étape de purification, il est possible de fractionner l'échantillon en fonction de paramètres biochimiques. En effet les propriétés biochimiques particulières d'un complexe cible peuvent être utilisées pour pré-enrichir une fraction d'échantillon avant de réaliser l'étape d'immunoprécipitation. Ce fractionnement permet d'augmenter l'efficacité de l'isolement d'un complexe protéique cible et d'éliminer les interactants non spécifiques⁹³.

Le fractionnement en fonction de la taille (par filtration sur gel ou gradient de densité)⁹⁴, la précipitation sélective et la chromatographie par échange d'ions⁹⁵ ont été largement utilisés pour la séparation et l'enrichissement des complexes protéiques. Par exemple, le complexe impliquant la protéine CDC45 a été étudié après une combinaison d'étapes de chromatographie en série comprenant la filtration sur gel, l'échange d'ions et la chromatographie d'affinité (anti-CDC45)⁹³. Après analyse de la fraction purifiée, dix partenaires d'interaction de CDC45 ont été identifiés par la MS. Un autre exemple d'enrichissement de complexe multiprotéique (1 600 kDa) a été réalisé par chromatographie d'échange d'ions et par filtration sur gel, suivi d'une sédimentation en gradient de glycérol. Après IP, deux protéines partenaires ont été identifiées (TbMP52 et TbMP48)⁹⁵

Les méthodes de fractionnement biochimique combinées à des stratégies d'IP, représentent des approches pertinentes pour les études d'interactomique⁹⁶⁻⁹⁷. Bien que ces approches de fractionnement soient utilisées avec succès, l'adaptation de la stratégie en fonction de la nature du complexe d'intérêt est une limitation du fractionnement⁹⁸.

B.2. Une nouvelle approche par marquage « proximité dépendante » : Le BioID

L'approche « BioID » (marquage à la biotine « proximité-dépendant ») a été décrite pour la première fois par Roux et son équipe en 2013⁵. BioID est une nouvelle technique pour l'étiquetage à la biotine, dépendante de la proximité des protéines dans les cellules. Cette approche est basée sur la fusion de la séquence d'une biotinyllase nommée BirA (*E. Coli*) à la séquence d'une protéine appât, impliquée dans le complexe d'intérêt. Une fois synthétisée, la protéine appât porte donc BirA (35,4 kDa) sur son extrémité N- ou C-terminale.

Après ajout d'ATP et de biotine dans le milieu de culture, BirA qui possède une activité « ligase », catalyse la formation d'un intermédiaire : le 5'-biotinoyl-AMP. Cet intermédiaire réactif, va diffuser et biotinyler les chaînes latérales de lysines primaires accessibles des protéines environnantes⁹⁹. Les protéines partenaires d'interaction et les protéines vicinales, situées donc dans la zone de diffusion du réactif biotinoyl-5'-AMP¹⁰⁰, seront biotinylées (Figure 18). Des études initiales ont suggéré que le rayon de la zone de diffusion est d'environ de 20-30 nm⁵, pour être ensuite restreinte à 10 nm¹⁰¹. Bien que la valeur exacte puisse varier en fonction du complexe protéique à l'étude, ce rayon est suffisant pour atteindre les lysines des protéines partenaires d'interaction directes et vicinales.

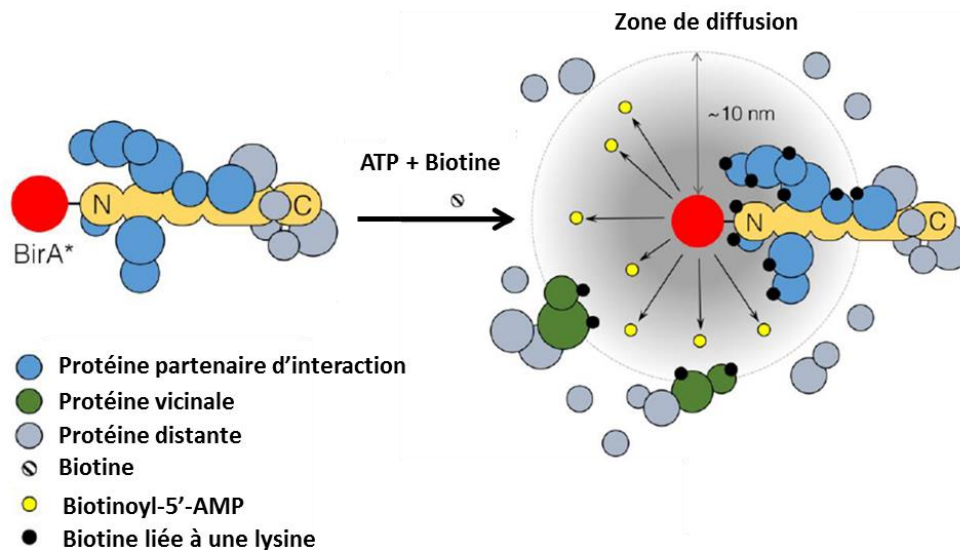


Figure 18: Représentation schématique de la stratégie de biotinylation des protéines dans la stratégie BioID, adapté de Varnaite et al¹⁰⁰.

Après une étape d'extraction protéique, les protéines biotinylées sont enrichies par une étape de purification avec des billes greffées par de la streptavidine. La biotinylation de ces protéines endogènes permet leur isolement sélectif grâce à la forte affinité de la streptavidine pour la biotine. Les protéines biotinylées sont ensuite analysées et identifiées par spectrométrie de masse. Les étapes de la stratégie sont présentées Figure 19. Les protéines identifiées par une approche BioID sont des candidates potentielles pour être des partenaires spécifiques d'interaction de la protéine appât.

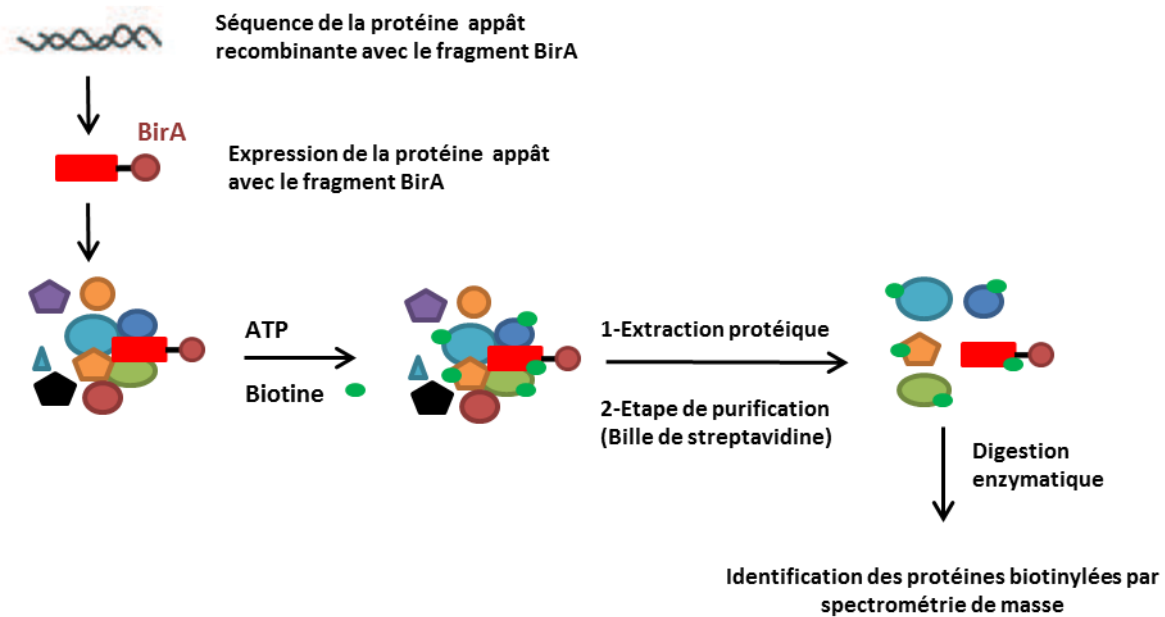


Figure 19: Représentation schématique de la stratégie BioID, pour la détermination de partenaires spécifiques d'un complexe

Dans la littérature, le nombre de publication ne cessent de croître ainsi que les domaines d'application (lamines d'enveloppe nucléaire¹⁰², centrosomes¹⁰³, flagelles associées au cytosquelette¹⁰⁴). Par une approche BioID⁵, K.J. Roux et son équipe ont ainsi identifiées *in vivo* 120 protéines proximales à la lamine A (insoluble) élément structural important de l'enveloppe nucléaire (NE). Autre exemple, O. Mühlemann et son équipe ont identifiés de nouveaux partenaires d'interaction du complexe NMD (impliqué dans la dégradation des ARNm non-sens) potentiellement importants¹⁰⁵. Ces partenaires d'interaction avaient échappé à la détection préalable par stratégie AP-MS avec des Ac, probablement parce qu'ils ne s'associent que faiblement et/ou très transitoirement au complexe NMD.

Ces résultats suggèrent que le BioID offre des avantages significatifs par rapport aux méthodes classiques. D'une part, l'approche permet d'identifier des partenaires d'interactions faibles ou transitoires *in vivo*¹⁰⁶, d'autre part l'applicabilité à des structures subcellulaires insolubles^{100, 106}.

Cette approche permet de fournir une liste de candidats potentiels d'interaction avec le complexe d'intérêt. Les nombreux candidats en tant que partenaires d'interaction ne peuvent pas être tous validés. Des approches complémentaires devront être menées en parallèles pour valider les protéines partenaires sélectionnées.

La modification des résidus lysines, acide aminé cible de la coupure protéolytique trypsique, peut amener la trypsine à manquer des sites de clivage¹⁰⁸. Avec un site de coupure manqué et une biotine ajoutée à la lysine, les peptides générés seront de taille et de masse plus importante. Ces peptides auront donc moins de probabilité d'être détectés par MS. Une autre limitation de la méthode est la taille relativement importante (35,4 kDa) de BirA qui pourrait entraîner une altération de l'adressage mais également de la fonction même des protéines¹⁰⁹.

En conclusion sur l'arsenal analytique actuellement disponible dans la littérature, il ressort que trois voies reposant toutes sur une purification par affinité, pourraient être utilisées pour l'identification des protéines participant au complexe des enzymes mélanosomales TYR/TYRP1/DCT. Ces trois voies ont été explorées pendant mon travail de thèse dans l'ordre suivant : i) la stratégie de piégeage du complexe par capture d'une de ses sous-unités par immunoaffinité, ii) la stratégie de piégeage du complexe grâce à un marquage à la

biotine, iii) la stratégie de stabilisation du complexe par réticulation des sous-unités et son isolement immunoaffinité.

La partie résultats expose successivement nos travaux réalisés avec ces trois stratégies.

C. Résultats

C.1. *L'approche de purification de complexe par immunoaffinité et spectrométrie de masse (AP-MS)*

Dans ce paragraphe nous avons évalué différents protocoles d'AP-MS avec des étapes d'immunoaffinité à l'aide d'anticorps dirigés soit contre TYR soit contre DCT. Pour ces deux approches, outre le choix de la sous-unité qui sera ciblée par l'anticorps (Ac), une étape clé sera évidemment l'optimisation de la stringence du lavage.

L'identification des protéines présentes dans l'isolat a été réalisée par analyse protéomique par spectrométrie de masse et a été utilisée pour guider les biologistes dans la mise au point d'une méthode d'isolement par affinité d'un complexe (AP-MS). Nous avons donc évalué l'enrichissement en matériel mélanosomal de chaque fraction obtenue. Les résultats ont été présentés pour chacune des deux cibles sélectionnées, TYR et DCT.

- Dans un premier temps, une immunoprécipitation (IP) avec un Ac dirigé contre la TYR endogène
- Dans un deuxième temps, une immunoprécipitation (IP) avec un Ac dirigé contre la DCT endogène

Toutes les préparations d'échantillon ont été réalisées par les collaborateurs biologistes (l'Oréal).

C.1.1 L'immunoprécipitation : Ac dirigé contre la protéine TYR endogène

Comme il a été montré que TYR, était une enzyme limitante de la mélanogénèse c'est la première à avoir été sélectionnée comme appât pour l'isolement du complexe d'intérêt. Pour cette étude, une lignée cellulaire de mélanocytes humains MNT1 a été utilisée. Ces cellules sont des lignées tumorales hyperpigmentées, c'est-à-dire naturellement riches en matériel nécessaire à la mélanogénèse.

Les mélanocytes ont été lysés puis centrifugés pour fractionner le matériel protéique en fonction de sa précipitation différentielle pour un tampon donné. Deux étapes successives ont été réalisées pour obtenir un culot contenant le matériel mélanosomal comme cela est présenté Figure 20. Ensuite, une étape d'IP avec un Ac dirigé contre la partie C-terminale de TYR a été réalisée pour tenter d'isoler le complexe d'intérêt (TYR-TYRP1-DCT) et ses partenaires d'interactions spécifiques (protéines mélanosomales).

Afin d'évaluer l'enrichissement de la fraction IP, deux échantillons ont été collectés : l'extrait protéique avant IP (1) et la fraction après IP (2).

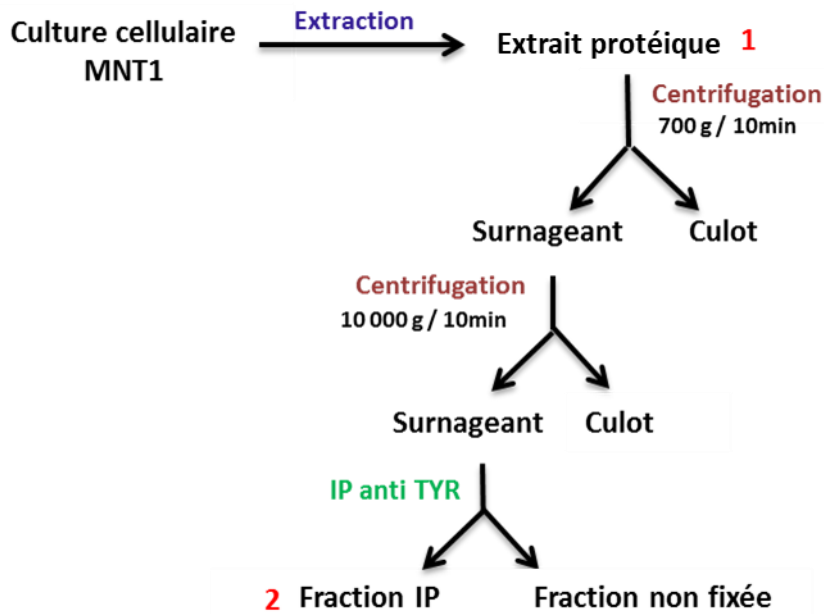


Figure 20: Représentation schématique de la stratégie de préparation d'échantillon avec deux étapes de centrifugation et une étape d'IP (Ac dirigé contre TYR). Les numéros 1 et 2 indiquent les échantillons qui ont été analysés par spectrométrie de masse.

L'extrait protéique (1) et la fraction IP (2) ont été déposés sur gel 1D-PAGE. Les gels obtenus sont présentés Figure 21. Les bandes visibles dans la fraction IP (2) mettent en évidence la présence d'une large population protéique.

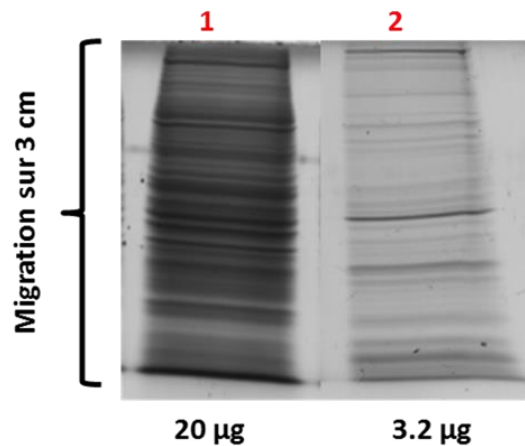


Figure 21: Gel 1D-PAGE du lysat protéique MNT1 avant IP (1), de la fraction obtenue après IP (2) dirigée contre la partie C-terminale de la TYR.

La totalité des pistes de gel 1D-PAGE a été extrudée et a été analysée par nano-LC-MS/MS avec la stratégie présentée Figure 22. Les résultats d'identification obtenus pour chaque bande ont été rassemblés.

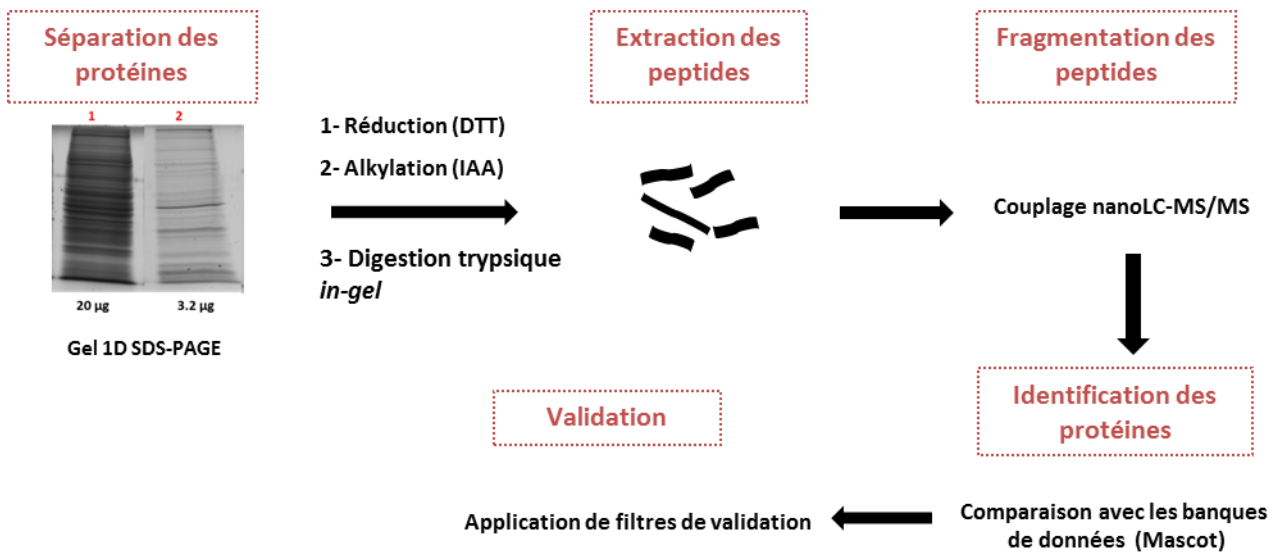


Figure 22: Représentation schématique de la stratégie d'analyse protéomique appliquée pour les échantillons 1 et 2

C.1.1.1 L'identification des protéines présentes par nanoLC-MS/MS

Les résultats d'analyse nanoLC-MS/MS ont permis d'identifier 2506 protéines dans l'extrait protéique MNT1, et plus particulièrement, les trois protéines TYR, TYRP1 et DCT, impliquées dans la mélanogénèse.

Après IP, 1275 protéines ont été identifiées, dont TYR, la protéine appât et TYRP1, partenaire d'interaction décrit dans la littérature⁶⁶. Le chiffre très élevé de 1275 protéines identifiées peut être vu, dans un premier temps, comme un signe d'échec d'isolement spécifique. Néanmoins, il faut garder en tête que les progrès considérables fait en analyse protéomique ces dernières années en terme de sensibilité permettent une dynamique d'au moins mille, voir dix mille entre la protéine identifiée la moins abondante, et la protéine identifiée, la plus abondante.

Nos résultats d'identification doivent donc impérativement être complétés par des résultats semi quantitatifs. Ceci nous permettra de distinguer les protéines spécifiques au complexe (nécessairement abondantes) des protéines non spécifiques et donc constituant le bruit de fond.

Nous présentons ci-après tout d'abord les données d'identification pour la protéine appât TYR. Dans l'optique d'évaluer rapidement l'enrichissement en protéine appât nous avons évalué de manière très grossière l'abondance des protéines en utilisant deux paramètres qui peuvent être considérés comme des indicateurs d'abondance des protéines dans l'échantillon analysé¹¹⁰. Premièrement nous avons classé les protéines en fonction de leur nombre de peptides uniques identifiés, puis en fonction de leur score assigné par Mascot (lié aux nombres de peptides identifiés et à la qualité des spectres assignés)^{38.39}.

Nous avons également exploité nos données en classant toutes les protéines identifiées en fonction de leur localisation cellulaire grâce au logiciel *Gene Ontology* pour déterminer l'enrichissement en protéines supposées être localisées dans le mélanosome et ainsi évaluer la qualité de l'isolement réalisé pour notre complexe d'intérêt.

Identification de la protéine appât TYR.

Après digestion *in silico* à la trypsine, 30 peptides trypsiques devraient être générés et couvrir plus de 70% de la séquence protéique de TYR, dont 14 sont des peptides protéotypiques¹¹¹.

Les peptides protéotypiques sont supposés être identifiés de façon répétée dans une analyse LC-MS/MS d'un mélange complexe¹¹². Les peptides protéotypiques sont avant tout des peptides uniques, ne comportant pas de méthionine, spécifique à une seule protéine, ce qui est primordial important dans un contexte où il y a de l'homologie de séquence entre les protéines.

Au total cinq peptides de TYR ont été détectés dans l'IP. Ces cinq peptides sont uniques à TYR et parmi eux, un seul est protéotypique (versus deux peptides dans l'extrait protéique). Moins de 20% de la séquence protéique de TYR est couverte par l'identification des peptides uniques, comme cela a été reporté dans la Figure 23.

La complexité du mélange (extrait protéique et IP) peut expliquer que tous les peptides protéotypiques n'aient pas été détectés. Cependant et contrairement à ce qui était attendu après une étape d'IP, la couverture de séquence de la protéine appât TYR (ou le nombre de peptides uniques détectés) n'est pas augmentée.

TYR

¹MLLAVLYCLLWSFQTSAGHFPRACVSSKNLMECCPPWSGDRSPCQLSGRGCQNILLSNAPLGPQFPFTGVDDRESWPSVFYNRTCQCSGNF
 MGFNCGNCKFGFWGPNCTERRLLVRRNIFDLSAPEKDKFFAYLTLAKHTISSDYVIPIGTYGQMKNGSTPMFNDINIYDLFVWMHYVVSMDALLGG
 SEIWRDIDFAHEAPAFLPWHRLFLLRWEQEIQKLTGDNFTIPYWDWRDAEKDICTDEYMGQQHPTNPILLSPASFFSSWQIVCSRLEEYNSHQSL
 CNGTPEGPLRRNPGNHDKSRTPRLPSSADVEFCLSLTQYESGSMDKAANFSFRNTLEGFASPLTGIADASQSSMHNALHIYMNMTMSQVQGSAND
 PIFLLHHAFFVDSIFEQWLRHRRLQEVYPEANAPIGHNRESYMVPIPLYRNGDFFISSKDLGYDYSYLQSDPDSFQDYIKSYLEQASRIWSWLLG
 AAMVGAVL TALLAGLVSLLCRHKRQQLPEEKQPLLMEKEDYHSLYQSHL⁵²⁹

Figure 23: Recouvrement de séquence après l'identification des peptides de la protéine (TYR), souligné par un trait bleu pour l'extrait protéique, par un trait rouge pour la fraction IP.

Pour évaluer de manière rapide et simple la qualité de l'enrichissement dans la fraction IP obtenue, nous avons vérifié si la protéine TYR était parmi les protéines les plus abondantes.

Données semi-quantitatives : classement des protéines en fonction du nombre de peptides uniques

Les protéines identifiées dans l'extrait protéique et dans l'IP ont été classées en fonction du nombre de peptides uniques détectés.

Après classement des protéines dans l'ordre décroissant de leurs peptides uniques, la protéine TYR est à la 373^{ème} position. Autrement dit, 30% des protéines dans l'IP ont été identifiées avec plus de peptides uniques que TYR (versus 17% dans l'extrait protéique). Bien entendu, le nombre de peptides d'une protéine est dépendant de la taille de celle-ci, mais TYR ne fait pas partie des protéines dites les plus abondantes dans l'IP. Le nombre de peptides uniques identifiés de TYR a diminué après IP.

La protéine TYRP1, décrite pour former un complexe avec TYR^{66, 77}, a également été identifiée. Comme pour TYR, le nombre de peptides uniques détectés après IP a diminué.

L'histogramme présentant la répartition des protéines en fonction de leurs nombres de peptides uniques dans le lysat et l'IP a mis en évidence la protéine TYR et son partenaire d'interaction TYRP1 n'ont pas été enrichis après IP (Figure 24).

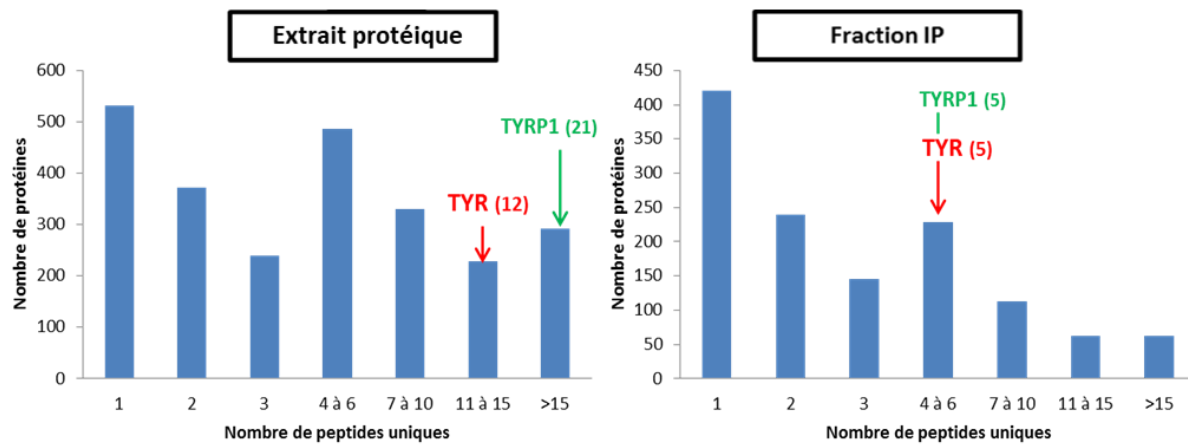


Figure 24: Histogramme de la répartition des protéines en fonction du nombre de peptides uniques identifiés pour chaque protéine. La catégorie dans laquelle se situe la protéine appât TYR est indiquée par une flèche rouge ; TYRP1 est indiquée en vert.

Données semi-quantitatives : classement des protéines en fonction du score protéique

Lors du traitement de données, le moteur de recherche (Mascot) va attribuer un score aux protéines (*protein set score*). Le score protéique dépend d'une part d'un « *ion score* » calculé pour chaque spectre MS/MS (évalue la qualité du spectre) et d'autre part, du nombre de peptides identifiés (uniques et non uniques). Plus le score protéique est élevé, plus l'identification de la protéine est sûre.

Après classement des protéines dans l'ordre décroissant de leurs scores protéiques, TYR est à la 622^{ème} position. Autrement dit 49% des protéines ont été identifiées avec un score plus élevé dans l'IP (versus 24% dans l'extrait protéique). Le score de la protéine TYRP1 est également diminué de manière significative de 753,54 (IP) à 91,28 (extrait protéique).

L'histogramme présentant la répartition des protéines en fonction de leurs nombres de peptides uniques avant et après IP a mis en évidence la protéine TYR et son partenaire d'interaction TYRP1 n'ont pas été enrichis après IP (Figure 25).

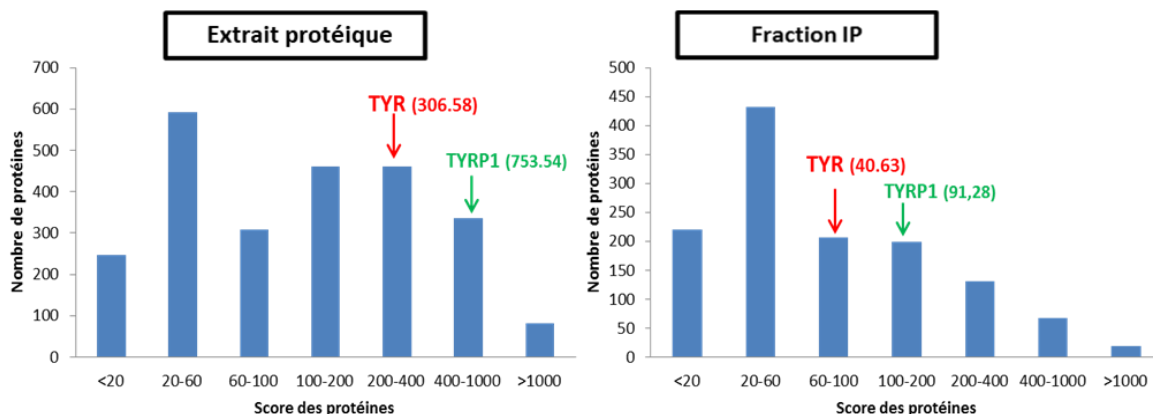


Figure 25: Histogramme de la répartition des protéines en fonction du score protéique assigné par Mascot, identifiées avant et après l'IP. La catégorie dans laquelle se situe la protéine appât TYR est indiquée par une flèche rouge, TYRP1 est indiquée en vert.

Classement des protéines en fonction de leur localisation cellulaire

Dans le but d'évaluer un enrichissement en protéines mélanosomales, nous avons utilisé *Gene Ontology*¹¹⁵ pour classer les données d'identification obtenues par analyses nanoLC-MS/MS. Cet outil permet de classer les gènes des protéines par rapport à la localisation cellulaire des protéines.

Les protéines ont donc été classées en fonction de leur localisation cellulaire à savoir : région extracellulaire, membrane cytoplasmique, cytoplasme, noyaux, organites intracellulaires (appareil de golgi, réticulum endoplasmique, lysosome, ribosome...) et mélanosome. Ce classement nous a permis de mettre en évidence, de manière rapide et claire, le faible pourcentage de protéines localisées dans le mélanosome dans l'IP. Aucun enrichissement en matériel mélanosomal n'a été obtenu après IP (Figure 26).

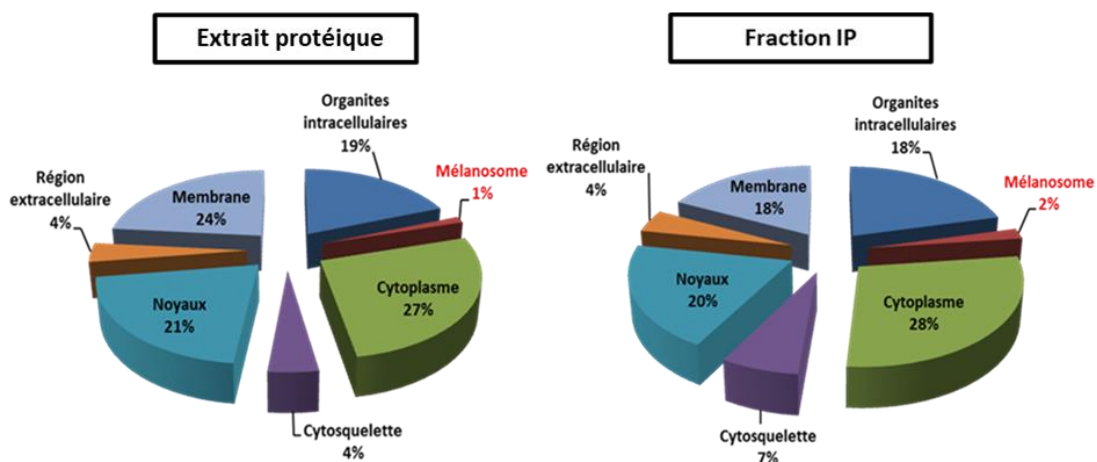


Figure 26: Diagramme de comparaison des protéines identifiées par analyse nanoLC-MS/MS classées en fonction de leur localisation cellulaire, avant et après IP (Ac anti-TYR).

A titre indicatif, une liste de protéines dites abondantes, identifiées avec les scores les plus élevés (30 à 200 peptides uniques), est présentée ci-dessous. La plupart d'entre elles sont des protéines du cytosquelette ou impliquées dans la mobilité cellulaire. Tout au long de leur maturation, les mélanosomes sont acheminés vers la membrane cytoplasmique du mélanocyte pour être transférés dans les kératynocytes. Ce processus dynamique de transport est contrôlé par de nombreux intermédiaires. Il n'est donc pas surprenant d'identifier des protéines du cytosquelette ou impliquées dans la mobilité cellulaire dans la fraction après IP.

Nom de la protéine	Score
Clathrine	2622,4
Filamine	2437,0
Protéine kinase ADN-dépendante	2354,1
Dynéine cytoplasmique	2054,2
Myosine	1947,4
Taline	1777,0
Pyruvate kinase	1688,4
Spectrine	1677,3
Tubuline	1643,6
Actine	1580,3
Protéine mitochondriale "Leucine-rich"	1531,3
Vimentine	1505,4

Tableau 1: Liste des protéines identifiées avec les scores protéiques les plus élevés dans la fraction après IP.

Parmi les protéines prédites pour interagir avec la TYR⁷⁸, seule TYRP1 a été identifiée.

C.1.1.2 Conclusion de l'approche d'immunoprécipitation dirigée contre TYR

L'étude semi-quantitative poussée des protéines identifiées et de leur localisation cellulaire a permis de mettre en évidence qu'aucun enrichissement de la protéine appât TYR n'a été réalisé avec le protocole testé. Le fait que des protéines n'ayant aucune raison de participer au complexe d'intérêt (voir liste ci-dessus) démontre qu'aucune sélectivité n'a pu être mise en œuvre pendant l'expérience. De même le nombre très élevé de protéines présentes, après l'étape de purification, ne rend pas possible la détermination des partenaires d'interactions spécifiques. Ce manque d'efficacité de l'étape d'IP peut être lié soit aux conditions d'éluion ou de lavage trop stringentes (perte de protéine appât et ses partenaires) soit à l'anticorps utilisé (manque d'affinité, peu de spécificité,...).

Il est fort possible que les conditions nécessaires pour éluer certaines protéines fixées de façon non spécifique sur les billes d'affinité soit déjà trop stringentes et dissocient en grande partie le complexe d'intérêt.

Cependant la vérification a posteriori de la qualité de l'anticorps utilisé pour l'immunoprécipitation, s'est avérée faible par WB, ce qui conjugué avec d'autres difficultés, pourrait expliquer l'inefficacité de l'enrichissement. La stratégie d'isolement du complexe a donc été modifiée. DCT, une autre protéine impliquée dans le complexe a été utilisée comme protéine appât pour l'étape d'IP.

C.1.2 L'immunoprécipitation : Ac dirigé contre la DCT endogène

Comme il a été montré DCT (62929 Da), est une enzyme de la mélanogénèse qui participe au complexe d'intérêt (TYR/TYRP1/DCT). Pour ce deuxième essai, DCT a donc été choisie comme protéine appât dans l'étape de purification par immunoaffinité et l'affinité de l'anti-DCT a été validée par WB en amont de l'immunoprécipitation.

Le même protocole de préparation d'échantillon que l'étude précédente a été réalisé sur des cultures cellulaires MNT1 (Figure 20). Afin d'évaluer l'enrichissement de la fraction IP, deux échantillons ont été collectés. Le lysat protéique avant IP (1) et la fraction IP (2).

Les échantillons 1 et 2 ont été déposés sur un gel 1D-PAGE pour séparer les protéines, le gel 1D obtenu est présenté Figure 27. Les protéines ont migré sur trois centimètres pour évaluer rapidement la population protéique et réaliser un fractionnement de l'échantillon en 15 bandes. Les profils des pistes de gel de la fraction IP mettent en évidence à nouveau la présence d'une large population protéique.

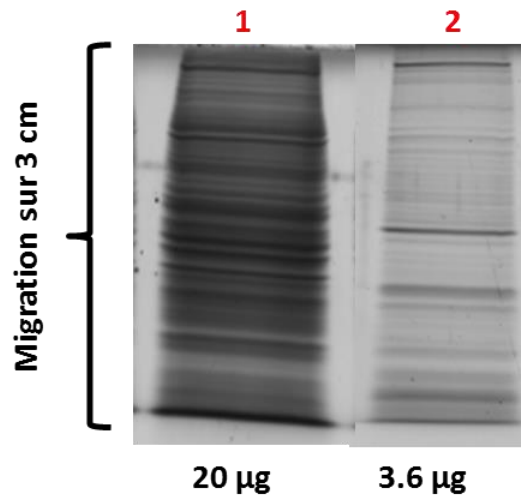


Figure 27: Gel 1D-PAGE de l'extrait protéique MNT1 avant IP (1), de la fraction obtenue après IP (2) dirigée contre la partie C-terminal de la DCT.

La totalité des pistes de gel a été extrudée en plusieurs bandes et a été analysée par nano-LC-MS/MS avec la stratégie présentée Figure 22. Le gradient chromatographique de séparation a été adapté pour l'analyse d'échantillon complexe dans le but d'obtenir une diversité et un nombre de peptides maximum des protéines présentes. Les résultats d'identification obtenus pour chaque bande ont été rassemblés.

C.1.2.1 L'identification des protéines présentes par nanoLC-MS/MS

Les résultats d'analyse nanoLC-MS/MS ont permis d'identifier 2506 protéines dans l'extrait protéique dont les trois protéines TYR, TYRP1 et DCT.

Après IP, 1248 protéines ont été identifiées, dont DCT la protéine appât et TYRP1, son partenaire d'interaction décrit dans la littérature⁷⁷. Ici aussi le chiffre élevé de 1248 protéines identifiées n'est pas pertinent s'il n'est pas complété par des résultats semi quantitatifs.

Nous présentons ci-après tout d'abord les données d'identification pour la protéine appât DCT. Nous présenterons ensuite les données permettant d'évaluer l'abondance des autres protéines en utilisant deux approches : le nombre de peptides uniques identifié pour chaque protéine, et le score assigné par Mascot. Les protéines identifiées ont ensuite été classées en fonction de leur localisation cellulaire grâce au logiciel *Gene Ontology*.

Identification de la protéine appât DCT.

Après digestion *in silico* à la trypsine, les peptides devraient couvrir plus de 60% de la séquence protéique soit 29 peptides tryptiques de DCT (dont 15 sont des peptides protéotypiques¹¹¹).

Au total 16 peptides de DCT ont été détectés dans l'IP. Parmi les 16 peptides uniques qui ont permis d'identifier DCT, six sont protéotypiques (versus cinq peptides dans l'extrait protéique). Plus de 30% de la séquence protéique de DCT est couverte par l'identification des peptides uniques, comme cela a été reporté dans la Figure 28.

La complexité du mélange (extrait protéique et IP) peut expliquer que tous les peptides protéotypiques n'aient pas été détectés. Conformément à ce qui est attendu après une étape d'IP, la couverture de séquence (ou le nombre de peptides uniques détectés) de la protéine appât DCT est augmentée.

```

1MSPLWWGFLLSCLGCKILPGAQGOFPVRCMTVDSLNVKECCPRLGAESANVCGSQOGRGQCTEVRADTRPWSGYPYILRNQDDRELWPRKFFH
RTCKCTGNFAGYNCGDCKFGWGTGPNCCERKKPPVIRQNIHSLSPQEREQFLGALDLAKKRVHPDYVITQHWLGLLGPNGTQPQFANCSYDFFV
WLHYYSVRDRTLGGPRPYRAIDFSHQGPAFVTWHRYHLLCLERDLQRLIGNESFALPYWNFATGRNECDVCTDQLFGAARPDPTLISRNSRFSSW
ETVCDLDDYNHLVTLNCGTYEGLLRNQMGRRNSMKLPTLKDIRDCLSLQKFDNPPFFQNSTFSFRNALEGFDKADGTLDSQVMSLHNLVHSLFN
GTNALPHSAANDPIFVWLHSFTDAIFDEWMKRFNPPADAWPQELAPIGHNRMYNMVPPFPVTNEELFLTSDQLGYSYAIIDLPSVVEETPGWPTT
LLVVMGTLVALVGLFVLLAFLQYRRLRKGTYPLMETHLSSKRYTEEA519

```

Figure 28: Recouvrement de séquence après l'identification des peptides de la protéine (DCT), souligné par un trait bleu pour l'extrait protéique, par un trait rouge pour la fraction IP.

Données semi-quantitatives : classement des protéines en fonction du nombre de peptides uniques

Après classement des protéines dans l'ordre décroissant de leurs peptides uniques, la protéine DCT est à la 155^{ème} position. Autrement dit, 7% des protéines ont été identifiées avec plus de peptides uniques que DCT dans l'IP (versus 45% avant IP). La protéine DCT fait donc partie des protéines abondantes identifiées après l'IP.

De plus, la protéine TYRP1, impliquée dans le complexe est également identifiée. Bien que celle-ci ne soit pas la protéine appât, elle fait partie des protéines les plus abondantes dans l'IP (neuf peptides uniques identifiés). L'histogramme présentant la répartition des protéines en fonction de leur nombre de peptides uniques dans l'extrait protéique et l'IP a mis en évidence que la protéine DCT a été enrichie (Figure 29).

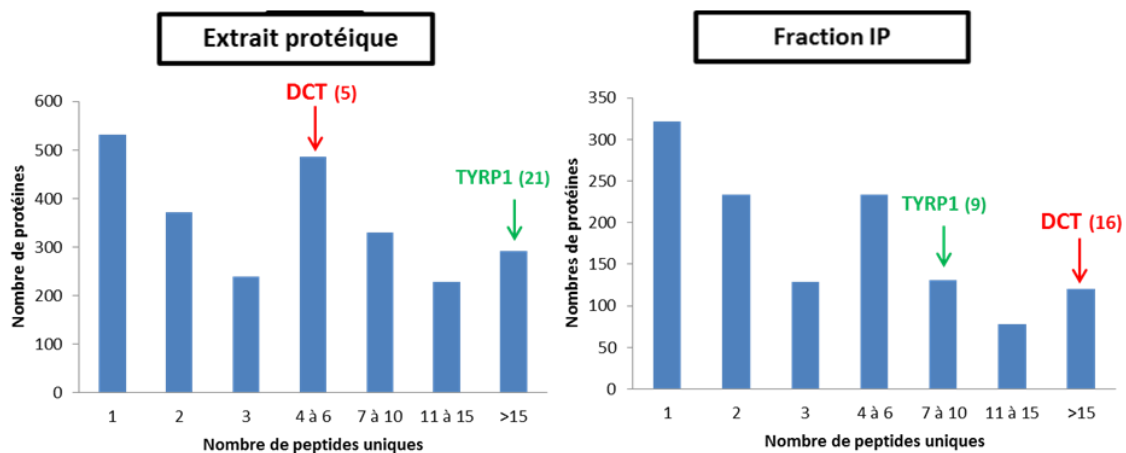


Figure 29: Histogramme de la répartition des protéines en fonction du nombre de peptides uniques identifiés pour chaque protéine. La catégorie dans laquelle se situe la protéine appât DCT est indiquée par une flèche rouge TYRP1 est indiqué en vert.

Données semi-quantitatives : classement des protéines en fonction du score protéique

Après classement des protéines dans l'ordre décroissant de leurs scores protéiques, DCT est à la 147^{ème} position, soit 12% des protéines ont été identifiées avec un score plus élevé après IP (versus 45% avant IP). La protéine DCT fait partie des protéines majoritaires après IP.

De plus, la protéine TYRP1, bien que celle-ci ne soit pas la protéine appât, fait également partie des protéines les plus abondantes dans l'IP. L'histogramme présentant la répartition des protéines en fonction de leurs scores dans l'extrait protéique avant IP et après IP a mis en évidence l'enrichissement en DCT (Figure 30).

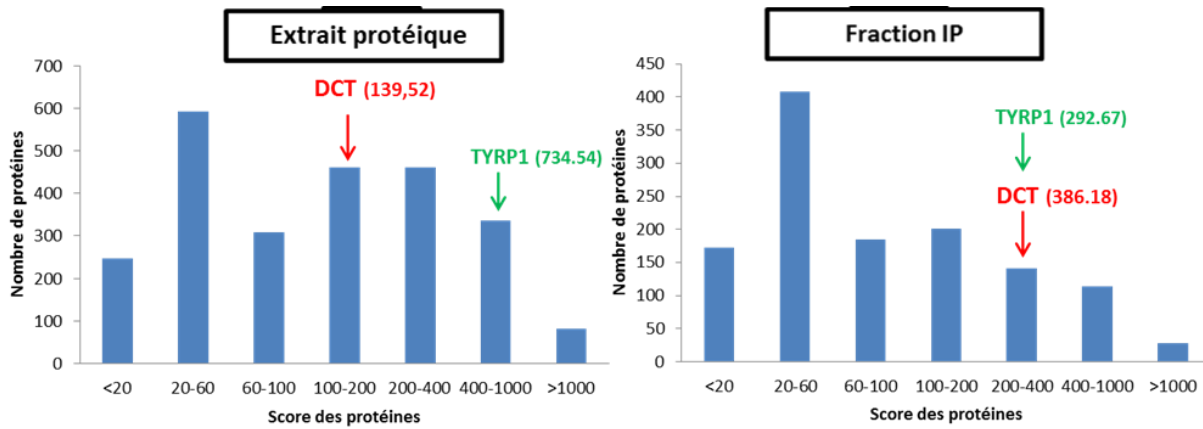


Figure 30: Histogramme de la répartition des protéines en fonction du score protéique assigné par Mascot, dans l'extrait protéique et dans l'IP. La catégorie dans laquelle se situe la protéine appât DCT est indiquée par une flèche rouge, TYRP1 est indiquée en vert.

Classement des protéines en fonction de leur localisation cellulaire

Les protéines ont donc été classées en fonction de leur localisation cellulaire à savoir : région extracellulaire, membrane cytoplasmique, cytoplasme, noyaux, organites intracellulaires (appareil de golgi, réticulum endoplasmique, lysosome, ribosome...) et mélanosome.

Ce classement nous a permis de mettre en évidence, de manière rapide et claire, le faible pourcentage de matériel mélanosomal dans l'IP. Aucun enrichissement en protéines mélanosomales n'a été obtenu après IP (Figure 31).

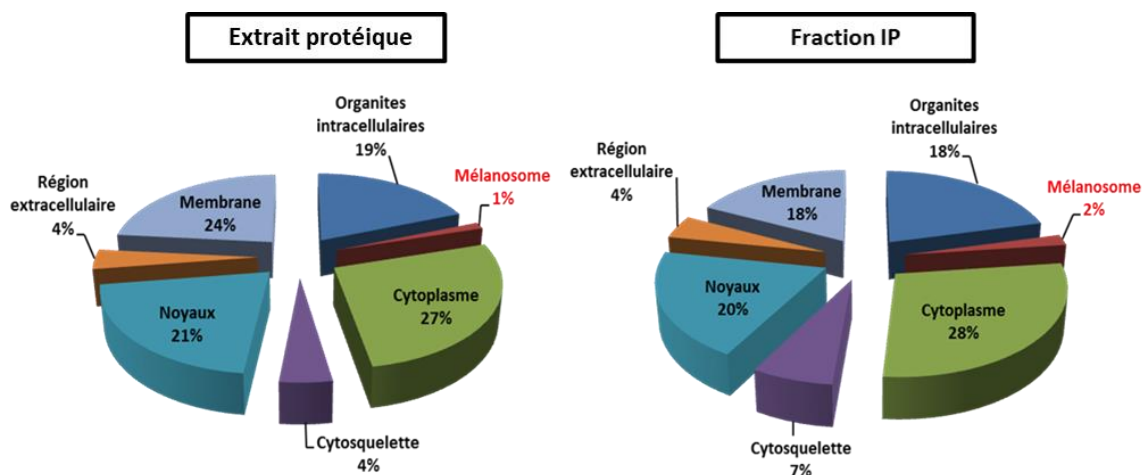


Figure 31: Diagramme de comparaison des protéines identifiées par analyse nanoLC-MS/MS classées en fonction de leur localisation cellulaire, dans l'extrait protéique MNT1 et dans l'IP (DCT).

A titre indicatif, une liste de protéines identifiées avec les scores les plus élevés (entre 30 et 200 peptides uniques) est présentée ci-dessous. La plupart d'entre elles sont des protéines du cytosquelette ou impliquées dans la mobilité cellulaire.

Nom de la protéine	Score
Myosine 9	7444.6
Myosine 10	5002.4
Spectrine	3933.6
Myosine	2976.4
Spectrine	2900.7
Actine	2359.6
Filamine	1479.5
Vimentine	1359.9
Protéine mitochondriale "Stress-70"	1276.8
Phospho-fructo-kinase ATP dépendante	1253.8

Tableau 2: Liste des protéines identifiées avec les scores protéiques les plus élevés dans la fraction IP.

C.1.2.2 Conclusion sur les résultats de l'approche d'IP dirigé contre DCT

L'étude semi-quantitative poussée des protéines identifiées et de leur localisation cellulaire a permis de clairement montrer que la protéine appât DCT a été enrichie après l'IP. Cette étude a également montré que TYRP1, protéine partenaire connue du complexe multiprotéique, était aussi identifiée parmi les protéines les plus abondantes de la fraction.

Malheureusement, le nombre élevé de protéines abondantes identifiées dans l'échantillon (voir liste ci-dessus) invalide en grande partie les résultats de cette expérience.

Les conditions de stringence ne sont donc pas optimum pour isoler le complexe d'intérêt. Bien que le protocole d'immunoaffinité soit le protocole classique, largement utilisé, nos résultats illustrent bien le chemin qu'il reste à parcourir pour optimiser les conditions expérimentales.

C.1.3 Conclusion de l'approche par immunoprécipitation dirigé contre une protéine du complexe

Les résultats d'identification obtenus dans ces expériences ont révélé deux types de problèmes fréquemment rencontrés avec l'isolement de complexe par immunoprécipitation, à savoir l'affinité et /ou la spécificité de l'Ac utilisé et la stringence des conditions de lavage et d'élution.

Bien que ce type de stratégies soit décrit dans la littérature, elles demeurent cependant complètement « complexe-dépendante ». Plusieurs hypothèses peuvent expliquer les difficultés à isoler le complexe d'intérêt :

- 1- L'épitope de la protéine reconnu par l'Ac peut être masqué et empêcher la reconnaissance par l'Ac, ce qui est envisageable dans le cas de complexes multiprotéiques de tailles importantes.
- 2- Il peut y avoir de la réactivité croisée générant des interactions aspécifiques avec certaines protéines.
- 3- Enfin, la quantité de complexe isolé est trop faible pour permettre une identification.

Pour vérifier ou invalider ces hypothèses, un certain nombre de contrôles auraient été nécessaires. Par exemple, l'analyse de la fraction obtenue par IP lorsque le gène de la protéine endogène est éteint (protéine

non exprimée). De même, l'analyse de la fraction IP obtenue avec un Ac non spécifique de la protéine appât peut être réalisée en parallèle pour déterminer les interactants non spécifiques.

En parallèle à l'approche basée sur l'immunoprécipitation, nous avons tenté l'approche de réticulation sur le complexe TYR/TYRP1/DCT. Cette approche est décrite dans le paragraphe ci-dessous.

C.2. *L'approche par réticulation chimique des protéines (DSP)*

L'objectif de cette approche est de stabiliser les interactions protéine-protéine avec un agent pontant, le di-thio-bis-succinimidylpropionate ou DSP), afin de garder les partenaires du complexe lors de son isolement par immunoprécipitation.

Dans un premier temps nous avons donc vérifié l'efficacité du protocole de réticulation chimique sur un extrait protéique total de cellules MNT1. La réticulation doit permettre de stabiliser les interactions de complexe de haut poids moléculaire. L'identification de peptides dérivés au DSP et plus particulièrement ceux de protéines localisées dans le mélanosome a été vérifiée.

Dans un deuxième temps nous avons évalué le protocole d'immunoprécipitation, et plus particulièrement l'affinité de deux Ac dirigés contre la séquence de TYR a été vérifiée sur la protéine recombinante r-TYR, par analyse protéomique.

C.2.1 *L'étape de réticulation des protéines par le DSP*

Le DSP est un agent pontant clivable par réduction d'un pont disulfure et dont le traitement des données, après réticulation, est entièrement compatible avec une stratégie nano-LC-MS/MS. La stratégie de réticulation par le DSP de deux lysines de protéines en interaction est présentée Figure 32.

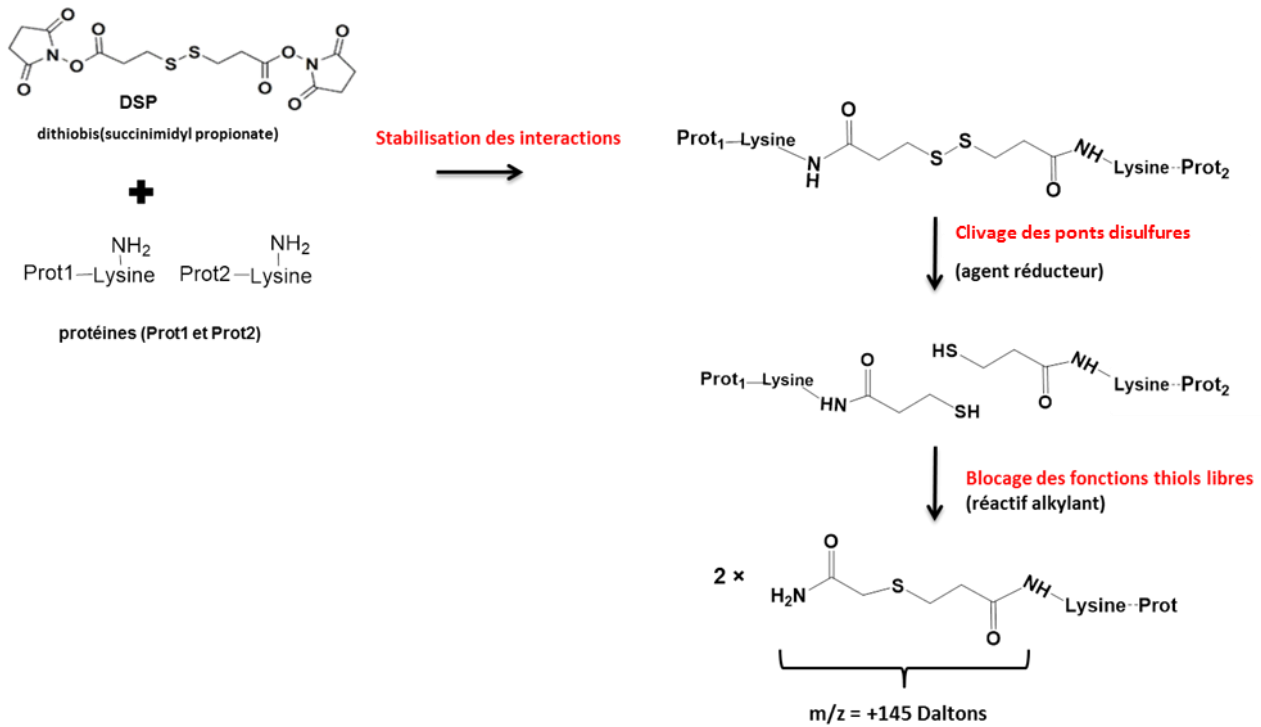


Figure 32: Représentation schématique de la stratégie de réticulation entre les lysines de deux protéines avec le DSP (agent pontant). Après réduction des ponts disulfures du DSP, les peptides portant la modification DSP (+ 145 Da) sont identifiables par Mascot.

Les protéines des culots de culture de cellules MNT1 ont été réticulées avec le DSP. Les différentes étapes de la stratégie sont présentées Figure 33.

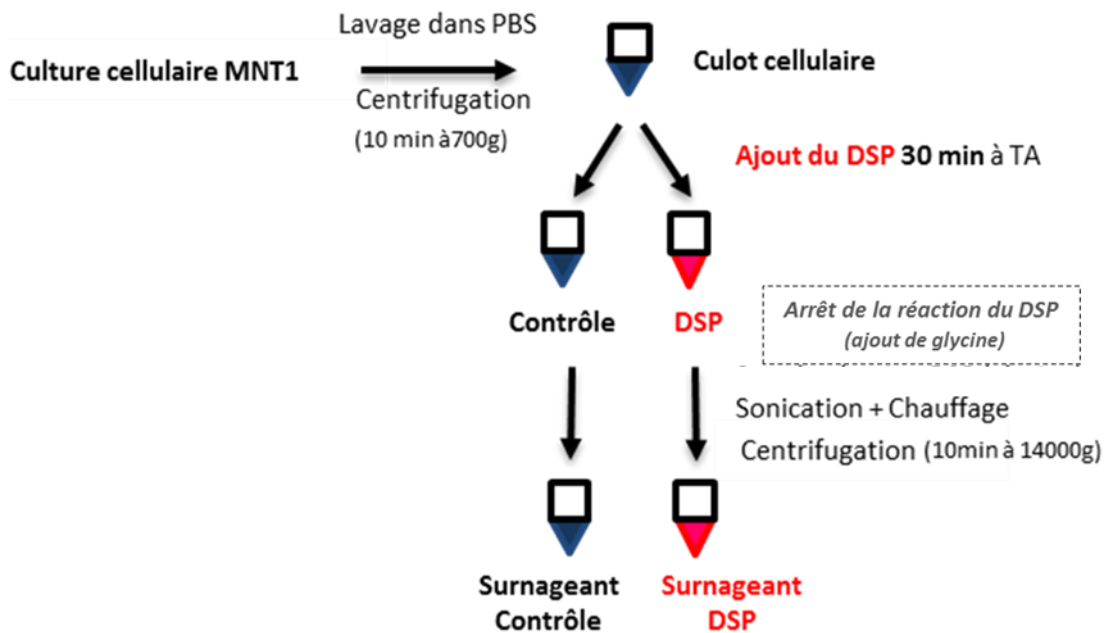


Figure 33: Représentation schématique de la stratégie de réticulation chimique des interactions protéine-protéine par ajout d'un agent pontant (DSP) et le contrôle sans DSP.

Le contrôle et le lysat avec DSP ont été déposés sur gel 1D-PAGE, en conditions non réductrices dans le but de maintenir les réticulations et pouvoir ainsi observer des complexes protéiques de haut poids moléculaires.

Après révélation des protéines au bleu de Coomassie, le gel 1D obtenu est présenté Figure 34. Dans le but de séparer les complexes de haut poids moléculaires qui devraient être stabilisés, les protéines ont été séparées sur 10cm de gel. Des dépôts sont visibles dans les fonds des puits (protéines qui n'ont pas migrées) et dans les zones correspondant aux hauts poids moléculaires. En conditions non dénaturantes, sans chauffage ni réduction des protéines, il est attendu que des agrégats protéiques se forment et soient visibles dans le fond des puits (pas de migration électrophorétique).

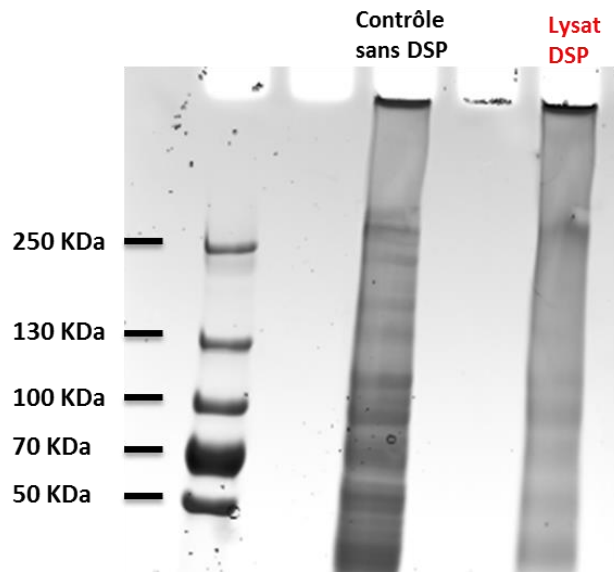


Figure 34: Gel 1D-PAGE du contrôle et de l'échantillon avec DSP en conditions non réductrices.

Une découpe systématique des pistes de gel a été faite. Les bandes de gel 1D ont été extrudées et soumises à la stratégie analytique présentée Figure 22. Le gradient chromatographique utilisé a été adapté pour la séparation d'un mélange complexe peptidique.

Nous avons vérifié si des protéines portant la modification DSP (+145 Da) avaient été identifiées en nous focalisant plus particulièrement aux protéines localisées dans le mélanosomes. Nous avons ensuite vérifié la présence des trois protéines du complexe TYR/TYRP1/DCT dans les bandes analysées.

C.2.1.1 L'identification des protéines identifiées dans les échantillons

Protéines identifiées réticulées au DSP.

Plus de 1500 protéines ont été identifiées dans les bandes de gel analysées. Au total, dans le lysat DSP, 450 protéines ont été identifiées avec au minimum un peptide portant la modification DSP. À l'aide du logiciel *Gene Ontology*, nous avons pu déterminer que 27 de ces protéines réticulées étaient localisées dans le mélanosome (Figure 35).

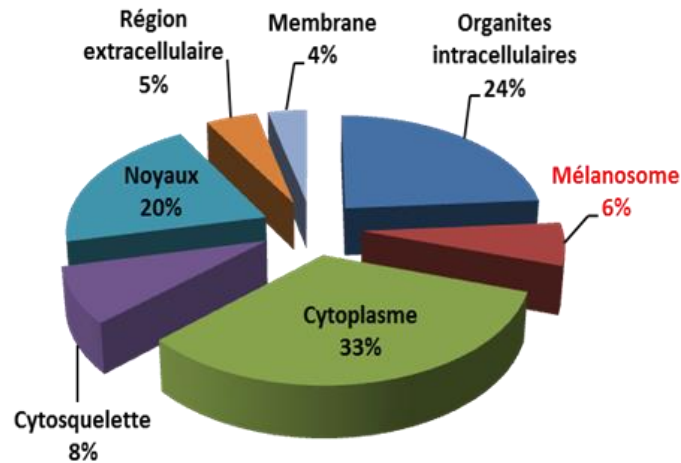


Figure 35: Diagramme de répartition des protéines identifiées avec un peptide DSP, par analyse nanoLC-MS/MS, classées en fonction de leur localisation cellulaire.

Caractérisation de protéines TYR, TYRP1 et DCT.

Les résultats d'analyse nanoLC-MS/MS ont permis d'identifier les protéines TYR, TYRP1 et DCT dans des bandes correspondant aux zones de hauts poids moléculaires dans le contrôle (bandes A1, B1 et C1) et dans le lysat avec DSP (bandes A2, B2 et C2). Nous notons que dans nos conditions de réticulation, la protéine DCT, a été identifiée dans la bande A2 (fond du puit). Les bandes dans lesquelles les trois protéines mélanosomales ont été identifiées sont présentées Figure 36.

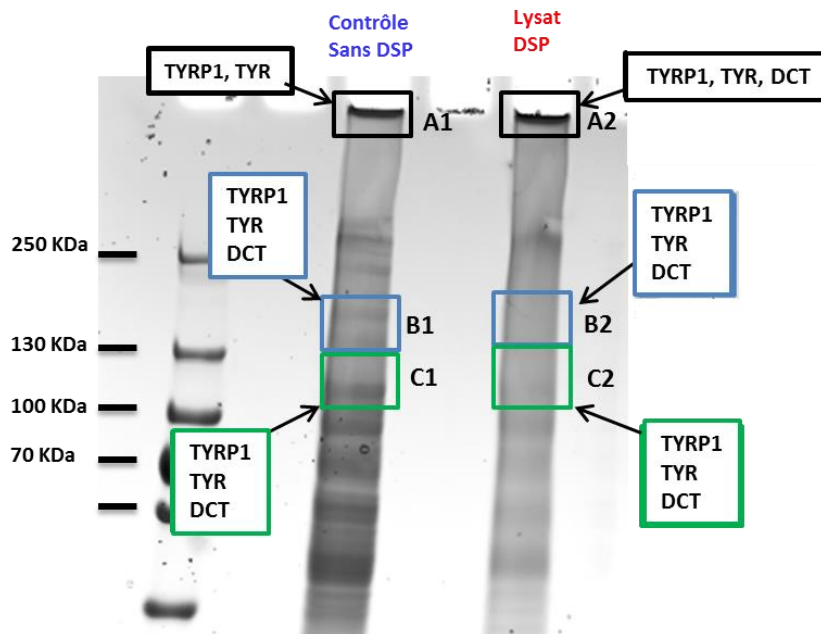


Figure 36: Résultats d'identification des peptides trypsique des protéines TYR, TYRP1 et DCT dans les bandes de haut poids moléculaires dans le contrôle (bandes A1, B1 et C1) et dans le lysat avec DSP (bandes A2, B2 et C2).

Parmi les peptides identifiés de TYRP1 un des peptides est dérivé au DSP dans la bande B2. Cette bande se situe dans une zone correspondant aux hauts poids moléculaires (entre 130 et 250 kDa).

Le spectre de fragmentation MS/MS peptide dérivé DSP 132 NLLDLSKEEK 141 de TYRP1 est présenté Figure 37. Un ajout de masse de 145 Da, correspondant au réactif DSP clivé (réduit) est détecté sur la lysine du peptide.

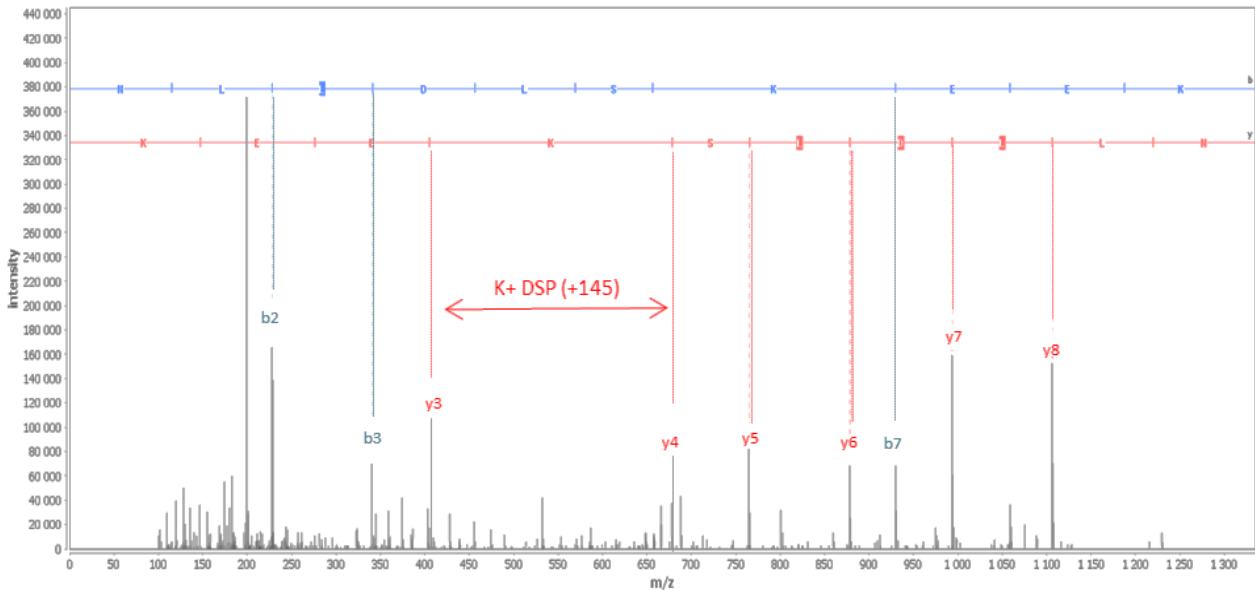


Figure 37: Spectre MS/MS d'un peptide de digestion trypsique $^{132}\text{NLLDLSKEEK}^{141}$ de la protéine TYRP1 dans la bande 2. La lysine du peptide est marquée chimiquement par le DSP (+145 Da) .

C.2.1.2 Conclusion sur les résultats de l'approche de réticulation

L'efficacité du protocole de réticulation des protéines au DSP a donc été vérifiée. Au total, après ajout du DSP plus de 450 protéines ont été d'identifiées avec au minimum un peptide DSP, dont 27 protéines du mélanosome (*Gene Ontology*). La stratégie analytique mise au point a permis d'identifier les trois protéines TYR, TYRP1 et DCT dans des bandes de gel 1D de haut poids moléculaires.

Après stabilisation des interactions protéine-protéine par le DSP, un peptide de TYRP1 a été identifiée portant la modification chimique du DSP réduit (+145 Da).

Les conditions expérimentales de la première étape de notre approche d'isolement de complexe, avant immunoprécipitation, sont validées. Suite aux premiers essais d'immunoprécipitation de la TYR, l'Ac utilisé s'est avéré peu efficace et deux autres anticorps ont été choisis afin de contrôler leur l'affinité pour la TYR.

C.2.2 L'immunoprécipitation : Ac dirigé contre TYR recombinante

Comme il a été montré que TYR était une enzyme limitante de la mélanogénèse, elle a été sélectionnée comme appât pour l'isolement du complexe d'intérêt.

La protéine TYR a été produite par la société GTP technology® (pour l'Oréal) dans un système d'expression de cellules CHO (cellules ovariennes de hamster). La séquence de la protéine a été fusionnée avec une étiquette polyhistidine côté C-terminal pour permettre sa purification du système d'expression.

La TYR recombinante, r-TYR (1-486 a.a), possède une masse moléculaire de 55465.19 Da et contient six sites consensus de N-glycosylation dans sa séquence ; identiques à la protéine humaine. La séquence C-terminale de la protéine humaine diffère de la séquence de la recombinante à partir de l'acide aminé en position 477.

Dans le but de vérifier la spécificité de l'Ac anti-TYR utilisé dans l'expérience précédente nous avons tenté d'immunoprécipiter la r-TYR. L'Ac est un Ac monoclonal nommé A, dirigé contre la séquence entière de TYR. Aucune autre information n'a été fournie par le fabricant. Après une étape d'incubation de r-TYR avec l'Ac, l'échantillon a été élué sur des billes magnétiques et les fractions fixées et non fixées ont été collectées. Après chauffage des billes magnétiques, la fraction des protéines non éluées a également été collectée comme cela est présenté Figure 38.

Nous avons également vérifié la spécificité d'un deuxième Ac monoclonal nommé B dont la partie qui est reconnue de la TYR est communiquée par le fabricant (dirigé contre la partie de la séquence de TYR de l'acide aminé en position 1 à la position 433). L'Ac est lié à une biotine, la forte affinité de la biotine pour la streptavidine devra permettre une meilleure élimination des protéines non spécifiques du complexe dans l'extrait protéique issu des cultures MNT1. Après une étape d'incubation de r-TYR avec l'Ac, l'échantillon a été élué sur des billes greffées avec de la streptavidine et les fractions fixées et non fixées ont été collectées. Après chauffage des billes magnétiques, la fraction des protéines non éluées a également été collectée comme cela est présenté Figure 38.

Nous rappelons que le fournisseur de l'Ac a vérifié sa spécificité pour TYR et que malgré la forte homologie de séquence entre les trois protéines, aucune réaction croisée de l'Ac pour les protéines TYRP1 et DCT n'a été détectée.

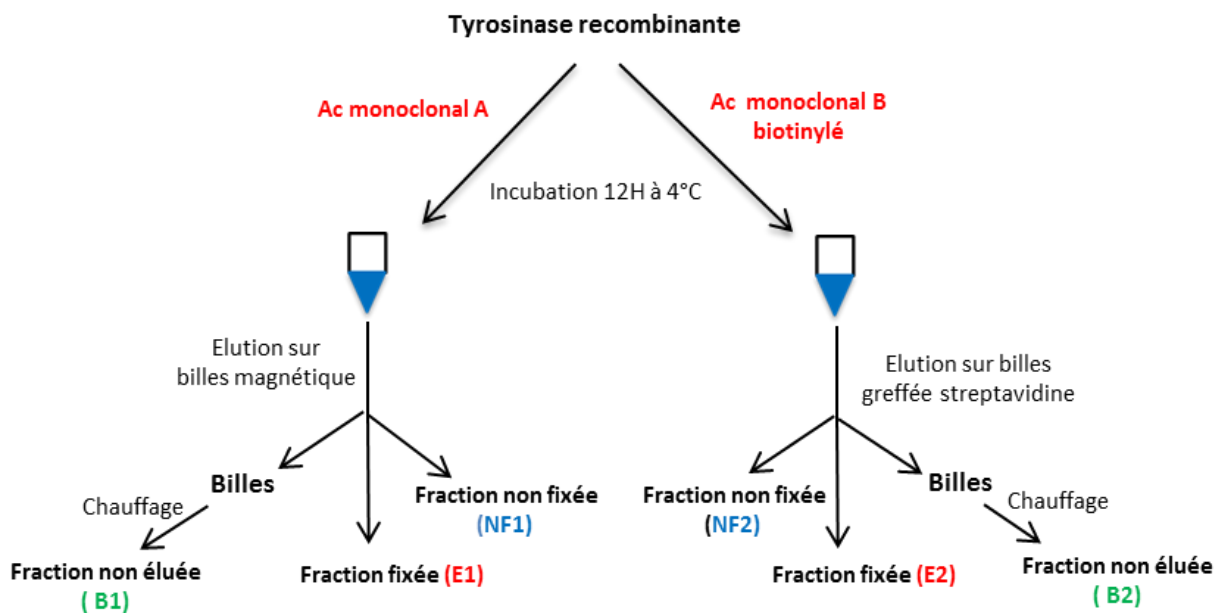


Figure 38: Représentation schématique des étapes d'IP de la protéine r-TYR par deux Ac. A gauche, l'IP avec l'Ac A dirigé contre la séquence entière de r-TYR pour laquelle les fractions E1, NF1 et B1 ont été collectées. A droite, l'IP avec l'Ac B dirigé contre la séquence de r-TYR (position 1^{er}-433^{ème} a.a) pour laquelle les fractions E2, NF2 et B2 ont été collectées.

Toutes les fractions ont été déposées sur gel 1D-PAGE. Après coloration au Bleu de Coomassie, le gel obtenu est présenté Figure 39.

IP avec l'Ac A. Aucune bande n'est visible dans les masses apparentes vers 60 kDa, masse attendue pour la protéine r-TYR. Cependant dans la fraction non éluée (B1) une bande est visible vers 60-70 kDa.

IP avec l'Ac B. Dans la fraction fixée E2. Après l'IP, des bandes de faible intensité, visible vers 70kDa (bande 3), masse attendue pour r-TYR et vers 130 kDa (bande 2), sont visibles. Il est possible que la protéine r-TYR forme un homodimère et ce complexe se situerait donc vers 130 kDa.

Dans la fraction non fixée NF2. Une bande de faible intensité est visible aux alentours de 130 kDa (bande 4).

Dans la fraction non éluée B2. Plusieurs bandes intenses sont visibles entre les masses apparentes de 50 à 70 kDa (bande 5). Après chauffage, il peut s'agir de formes dégradées de r-TYR.

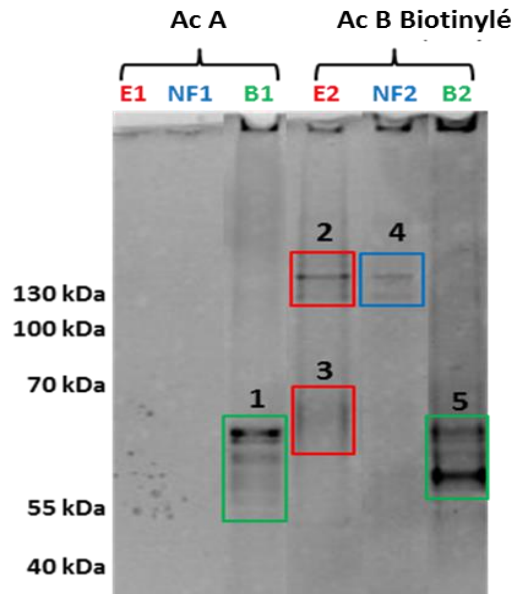


Figure 39: Gel 1D des fractions fixées (E1, E2), des fractions non fixées (NF1, NF2) et des fractions non éluées collectées sur les billes après chauffage (B1, B2).

Les bandes de gel 1D de 1 à 5 ont été extrudées par le collaborateur. Nous avons ensuite soumis ces bandes à une stratégie d'analyse protéomique (cf. Figure 22).

C.2.2.1 L'identification de la TYR par nano-LC-MS/MS

IP avec l'Ac A. Dans la bande 1 (fraction B1) aucun peptide de r-TYR n'a été identifié. Cependant des kératines contaminantes et des fragments d'anticorps ont été identifiés. L'Ac A et les conditions d'incubation et d'élué ne semblent pas permettre d'isoler la protéine r-TYR. L'utilisation de l'Ac A est donc invalidée pour notre approche.

IP avec l'Ac B. Les résultats ont permis d'identifier des peptides de r-TYR dans les bandes 2, 3, 4 et 5. L'Ac biotinylé permet d'immunopurifier r-TYR de manière peu efficace. L'identification de la TYR dans la fraction non fixée et dans la fraction non éluée (reste sur les billes) met en évidence que les conditions d'incubation et d'élué ne sont pas optimum pour isoler efficacement r-TYR.

C.2.2.2 Conclusion sur les résultats de l'approche d'immunoprécipitation de r-TYR

Les résultats de protéomique ont montré que la stratégie avec l'Ac biotinylé dirigé contre r-TYR (1-433^{ième} a.a) disponible commercialement semblait la plus efficace pour « capturer » la protéine appât.

Les conditions d'isolement de la TYR doivent encore être optimisées pour les expériences en culture cellulaire, et il est impératif que des contrôles (collecte des fractions non fixée et non éluée) soient réalisés pour valider la méthode d'IP.

En parallèle de ces expériences d'utilisation d'une méthode de réticulation nous avons exploré la piste d'une technique récente d'enrichissement en partenaires d'intérêt ; le « BioID » décrite ci-dessous.

C.3. L'approche par marquage à la biotine « BioID »

La méthode appelée « BioID » a été décrite pour la première fois en 2013¹⁰⁶ (Partie II -B.2). Le BioID est basé sur un marquage à la biotine dépendant de la proximité entre les partenaires du complexe et la protéine appât. Cette expérience est réalisée dans les cellules en culture. Comme il a été montré que TYR (60 393Da), était une enzyme limitante de la mélanogénèse, elle a donc été sélectionnée comme appât pour tenter l'isolement du complexe d'intérêt.

Comme décrit dans le paragraphe Partie II -B.2, les cellules MNT1 ont été transduites par un plasmide contenant la séquence de la protéine TYR fusionnée à la protéine biotinylyase BirA (*E. Coli*).

Dans une première étape les cellules MNT1 ont été transduites avec un construit génétique Sh-LUC-GFP, (*Green Fluorescent Protein*). La détection et le suivi de la fluorescence émise par la protéine GFP traduite ont permis de valider l'intégration d'un plasmide exogène donc, dans les cellules MNT1.

Dans une deuxième étape, les cellules MNT1 ont alors été transduites avec le gène de la protéine chimère TYR-BirA. L'expression de la protéine TYR-BirA a été vérifiée par une analyse Western Blot après révélation par un Ac anti TYR (Figure 40).

Après révélation, une bande intense est visible sur la membrane de nitrocellulose à 70 kDa, correspondant à la masse attendue de la tyrosinase endogène. Cependant aucune bande n'est visible aux alentours de 100kDa, masse moléculaire attendue de la protéine chimère TYR-BirA.

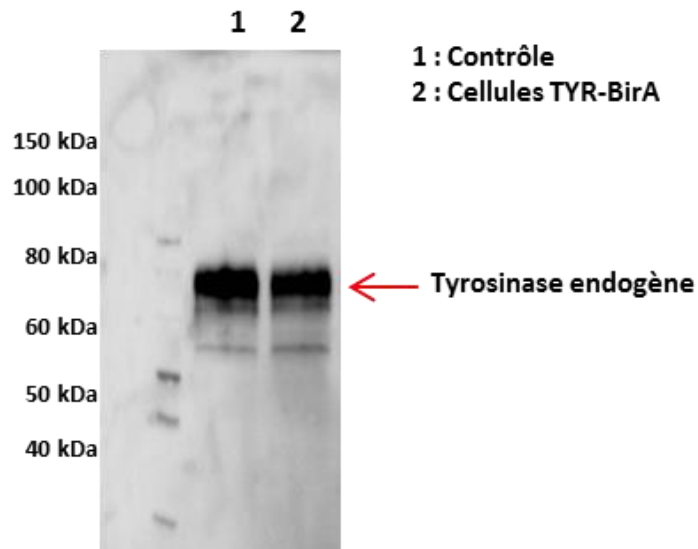


Figure 40 : Western Blot des dépôts de lysats cellulaires MNT1 du contrôle et des cellules transfectées TYR-BirA après révélation par un Ac dirigé contre TYR.

Ce résultat partiellement décevant pourrait être expliqué de la façon suivante. Il est possible que l'étape de traduction ou d'adressage de la protéine chimère ne se soit pas correctement réalisée. Une faible expression de la protéine chimère pourrait également expliquer que la quantité de protéine traduite soit en dessous du seuil de détection par Western Blot.

Un signe d'encouragement vient de la mesure de la quantité d'ARNm de la protéine chimère TYR-BirA par qPCR (*quantitative chain polymerase reaction*) qui a permis de vérifier la présence du transcrit.

Bien que la protéine ne soit pas détectée par Western Blot nous avons donc choisi de poursuivre l'étude avec l'étape de marquage par biotinylation des protéines des cellules transfectées (avec la biotinylyse BirA). Le protocole expérimental et la stratégie d'analyse protéomique sont présentés Figure 41.

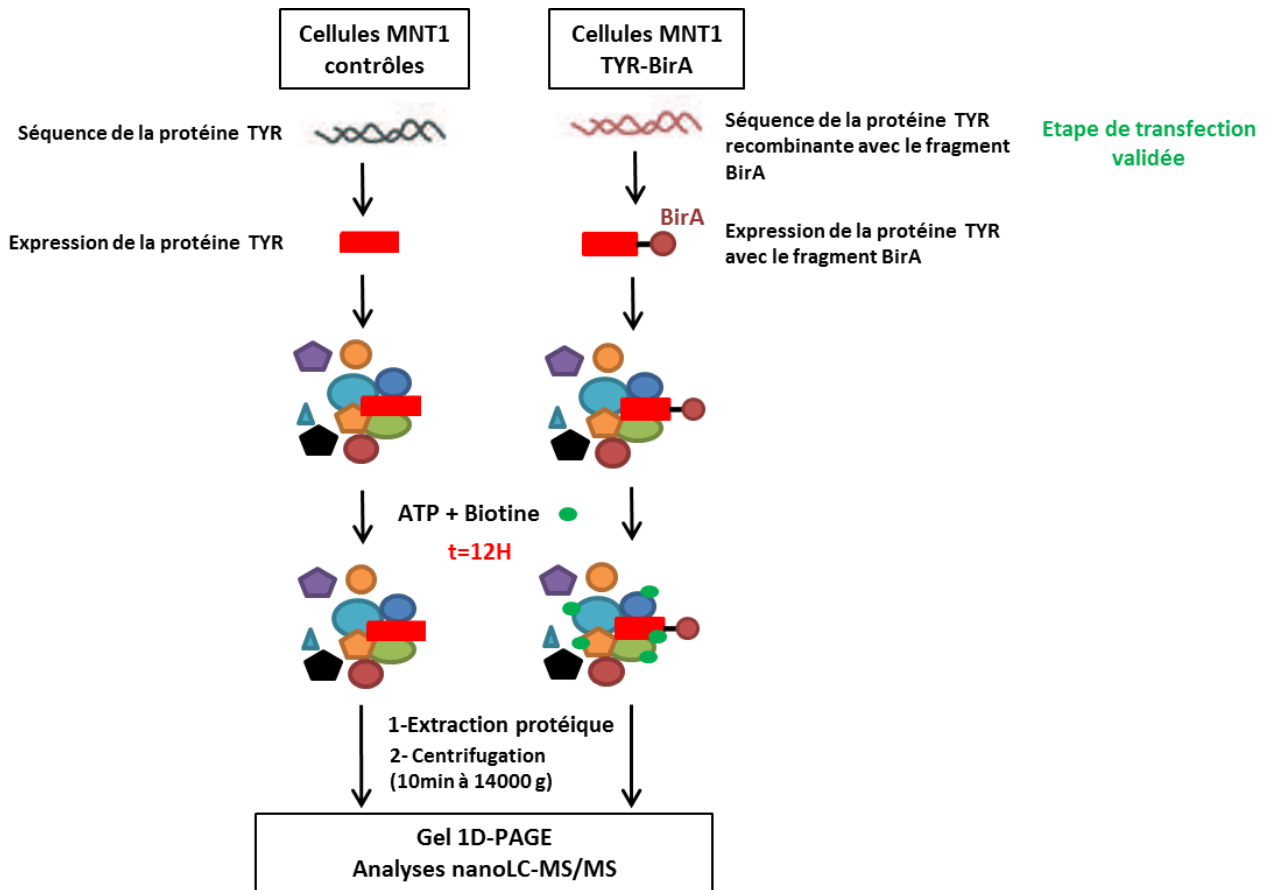


Figure 41: Représentation schématique de la stratégie BioID avec des cellules MNT1, où la séquence de BirA a été fusionnée à la séquence de la protéine TYR. Les deux échantillons, cellules transfectées et le contrôle, ont été analysés par nanoLC-MS/MS pour valider l'étape de biotinylation des protéines avant de mettre au point les conditions expérimentales de purification.

Les analyses protéomiques ont été réalisées de la façon suivante : après avoir extrait les protéines, les lysats ont été centrifugés, le surnageant a été collecté et déposé sur gel 1D-PAGE. Les protéines du lysat ont été concentrées sur gel car nous nous attendions à de faibles quantités de protéines biotinylées dans l'échantillon. La bande de concentration « stacking » a été extrudée. Les protéines ont été réduites et alkylées pour ensuite être digérées par la trypsine. Les digestes peptidiques du contrôle et des cellules transfectées ont été analysés par nanoLC-MS/MS (cf. la stratégie analytique présentée Figure 22).

C.3.1 L'identification des protéines présentes dans les lysats

L'interprétation bioinformatique des données a permis d'identifier les trois protéines TYR, TYRP1 et DCT mélanosomales dans les digestes des cellules transfectées et dans le contrôle.

Bien que le digeste peptidique des cellules transfectées soit complexe (2380 protéines identifiées), les résultats d'analyse ont permis d'identifier 22 protéines avec un peptide biotinylé.

La liaison covalente de la biotine est détectable en MS, par la détection d'un ajout de 226.3 Da à la lysine. Pour exemple, le spectre de fragmentation MS/MS du peptide QEKALQR (Z804A_HUMAN, protéine capable de lier le Zinc) marqué chimiquement par la biotine est présenté Figure 42.

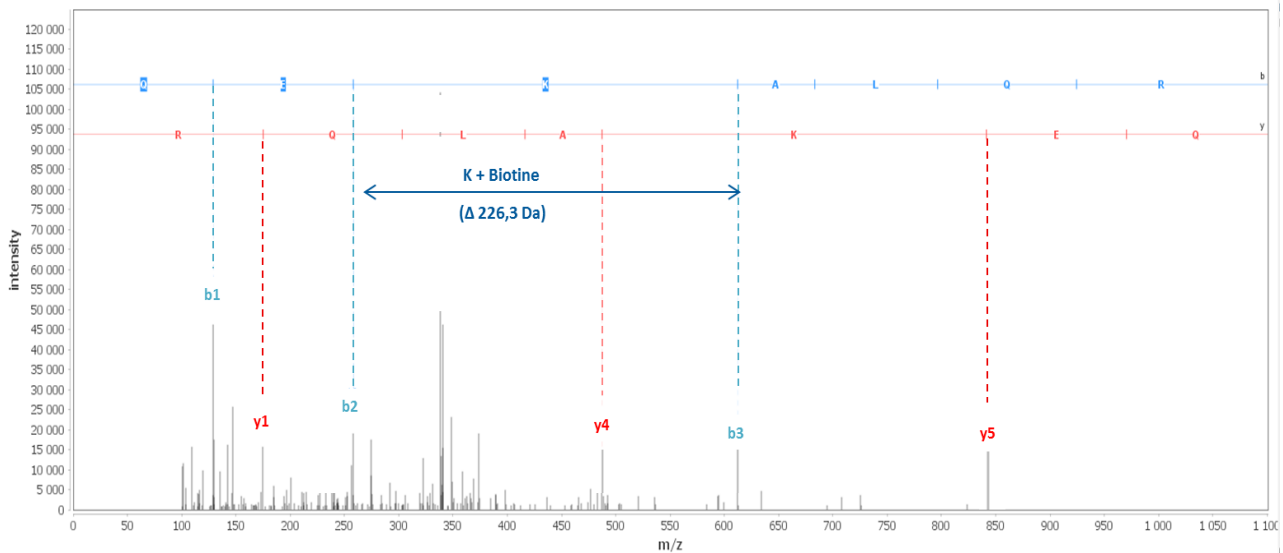


Figure 42: Spectre de fragmentation MS/MS du peptide QEK (+Biotine)ALQR de la protéine Z804A_HUMAN (*Zinc finger protein*, capable de lié le Zinc).

La biotine ne peut être synthétisée dans les cellules humaines, ce qui veut dire que les marquages à la biotine, des peptides identifiés, sont liés à des modifications chimiquement induites par notre protocole.

Malgré la surexpression, en théorie, de la TYR dans les cellules transfectées, le même nombre de peptides uniques de la protéine a été détecté dans le contrôle sans TYR-BirA et dans les cellules transduites. Ceci pourrait s'expliquer par des réactions de biotinylation spontanée dès qu'il y a présence de biotine dans les cellules ou par la présence d'une biotinyllase endogène.

Dans les deux digestes, la TYR a été identifiée avec six peptides uniques et avec le même score protéique. Trois de ces peptides possèdent des lysines, qui ne sont pas biotinylées. Aucun des peptides identifiés de TYRP1 (19 peptides) et DCT (5 peptides), protéines partenaires impliquées dans le complexe d'intérêt, n'a été biotinylé.

En conclusion, la protéine chimère a peu de chance d'être présente dans les cellules, en tout cas pas à un niveau détectable de l'activité biotinyllase, en accord avec les résultats de WB.

La liste des protéines identifiées biotinylées est présenté Tableau 3, aucune n'a été décrite pour interagir avec le complexe ni même pour être impliquées dans la pigmentation.

sp Q00325 MPCP_HUMAN	Phosphate carrier protein, mitochondrial
sp Q99623 PHB2_HUMAN	Prohibitin-2
sp P19367 HXK1_HUMAN	Hexokinase-1
sp Q9UHD1 CHRD1_HUMAN	Cysteine and histidine-rich domain-containing protein 1
sp Q9Y4G6 TLN2_HUMAN	Talin-2
sp Q7Z6Z7 HUWE1_HUMAN	E3 ubiquitin-protein ligase HUWE1
sp Q02750 MP2K1_HUMAN	Dual specificity mitogen-activated protein kinase kinase 1
sp Q9BW92 SYTM_HUMAN	Threonine--tRNA ligase, mitochondrial
sp Q5JXB2 UE2NL_HUMAN	Putative ubiquitin-conjugating enzyme E2 N-like
sp Q9P265 DIP2B_HUMAN	Disco-interacting protein 2 homolog B
sp Q9BUT1 BDH2_HUMAN	3-hydroxybutyrate dehydrogenase type 2
sp Q96PU8 QKI_HUMAN	Protein quaking
sp P53778 MK12_HUMAN	Mitogen-activated protein kinase 12
sp Q9BZD4 NUF2_HUMAN	Kinetochore protein Nuf2
sp Q9UIF9 BAZ2A_HUMAN	Bromodomain adjacent to zinc finger domain protein 2A
sp Q9H4L7 SMRCD_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1
sp Q460N5 PAR14_HUMAN	Poly [ADP-ribose] polymerase 14
sp O60749 SNX2_HUMAN	Sorting nexin-2
sp Q8NHU6 TDRD7_HUMAN	Tudor domain-containing protein 7
sp Q9NQ55 SSF1_HUMAN	Suppressor of SWI4 1 homologue
sp Q8NGY6 OR6N2_HUMAN	nd

Tableau 3: Liste des protéines identifiées avec un peptide biotinylé dans les cellules MNT1 par l'approche BioID

C.3.1.1 Conclusion sur les résultats de l'approche BioID

Les protéines identifiées biotinyllées ne sont pas décrites pour être impliquées dans la pigmentation. Malgré que les premières étapes de l'approche aient été validées (transfection d'un plasmide exogène, présence de l'ARNm de la protéine chimère), aucun résultat n'a mis en évidence l'expression de la TYR-BirA.

La protéine TYR-BirA, pourrait ne pas être exprimée ou mal adressée. Une des explications au problème rencontré, pourrait être la taille relativement importante de la protéine BirA (35,4 kDa) fusionné à TYR. En effet, la fusion de BirA avec la protéine TYR peut entraîner une altération de la localisation et de la fonction des protéines¹⁰⁰. Une des solutions pour pallier ce problème serait l'utilisation de la protéine chez *A. aeolicus*, qui est la plus petite protéine BirA connue, significativement plus courte que celle chez *E. coli*¹⁰⁹.*

D. Conclusion biologique de l'étude

Malgré l'utilisation des trois approches mise au point et testées pour tenter d'isoler le complexe cible TYR/TYRP1/DCT, aucun partenaire du complexe n'a pu être clairement identifié comme spécifique.

Néanmoins il ressort que :

- L'approche d'IP dirigée contre TYR et DCT a montré que TYRP1 était toujours identifié après IP, mais avec d'autres protéines qui ne devraient pas être impliquées dans la mélanogénèse. Ces résultats

ont été utiles, puisqu'ils ont permis de réaliser que l'anticorps dirigé contre TYR n'était pas suffisamment spécifique. Cette information a été très utile pour l'approche DSP.

- L'approche DSP a démontré la présence des protéines TYR/TYRP1 et DCT dans les complexes de haut poids moléculaires, mais l'étape d'isolement par IP nécessaire pour démontrer la spécificité n'a pu encore être réalisée. Comme pour l'approche précédente, d'autres conditions de stringence devraient donc être testées pour tenter d'isoler le complexe de haut poids moléculaires réticulés qui puisse être capture par une approche IP.
- L'approche BioID a permis d'identifier des protéines marquées à la biotine. Cependant, aucun peptide de digestion de la protéine TYR n'a pu être identifié sous forme biotinylé. De même aucun peptide de digestion des partenaires TYRP1 et DCT n'a pu être identifié sous forme biotinylé. Ce résultat négatif nous empêche de considérer que les autres protéines identifiées avec une biotinylation participent au complexe. D'autres conditions expérimentales pourraient être testées comme par exemple utiliser une autre protéine appât (telle que TYRP1 ou DCT) ou utiliser une culture cellulaire ayant un autre système d'adressage (culture mélanocytaire MHN)

E. Conclusion sur les aspects analytiques

Les résultats obtenus n'ont pas permis de d'identifier de façon sûre de nouveaux partenaires du complexe TYR/TYRP1/DCT. Néanmoins différentes mises au point ont été réalisées pour l'identification des protéines présentes en très petites quantités dans des échantillons très complexes. Ce savoir-faire ouvre la voie à la réalisation de séries d'expériences de mise au point de conditions de stringences optimales. Ces conditions devront permettre d'une part de conserver les interactions spécifiques entre l'anticorps et la protéine appât mais aussi entre les partenaires principaux du complexe, et d'autre part de décrocher des billes d'affinité toutes les autres protéines.

Le rôle crucial de l'analyse protéomique comme guide pour l'identification de partenaires de complexes.

Les différentes méthodes mises en place pour tenter d'isoler le complexe ont mis en évidence les difficultés que représentent les stratégies d'isolement de complexe. L'analyse protéomique par spectrométrie de masse est apparue comme un outil précieux et pertinent pour évaluer la qualité de l'isolement d'un complexe spécifique, et pour guider les biologistes dans les choix de la stratégie à adopter. Il est apparu que l'utilisation de l'analyse protéomique demandait pour chaque approche méthodologique et pour chaque complexe étudié une mise au point très fine.

Importance de données semi-quantitatives. Parmi les mises au point les plus nécessaires dans l'exploitation des données d'expérience de LC-MS/MS, nous avons montré qu'une simple liste de protéines identifiées n'apportait pas vraiment d'informations intéressantes. En effet la sensibilité des spectromètres de masses et l'efficacité des logiciels de bioinformatique d'exploitation des données, permettent d'atteindre une dynamique proche de 10^4 entre les protéines les plus abondantes et les moins abondantes. Il en résulte que, sans données de quantification des listes de protéines identifiées ne présentent que peu d'intérêt dans le type d'étude menée ici.

Il existe des approches dites de quantification *label free* performantes, sans marquage chimique (voir chapitre III), mais elles demandent un long processus d'exploitation des données totalement incompatible avec le type d'étude réalisée dans mon travail.

Nous avons donc utilisé une approche semi-quantitative, plus rapide, qui s'est avérée suffisante pour évaluer la qualité de l'enrichissement en complexe d'intérêt. Cette approche nécessite de déterminer une combinaison pertinente de critères (nombre de peptides uniques identifiés, recouvrement de séquence, score protéique, localisation cellulaire) pour extraire une information de type semi quantitatif.

Parti II Résultats

Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP

Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP

Ce projet est le fruit d'une collaboration étroite avec l'équipe du Pr JM. Reichhart (Institut de Biologie Moléculaire et Cellulaire, UPR 9022 (Réponse Immunitaire et Développement chez les Insectes), Strasbourg).

A. Introduction biologique

A.1. L'immunité

Toutes les espèces animales sont exposées quotidiennement à un grand nombre de micro-organismes potentiellement dangereux. Contre cela, les animaux et les plantes possèdent des mécanismes biologiques de défense qui leur permettent de réagir lors d'infections par des pathogènes. Au cours de l'évolution, la pression de sélection exercée par les agents pathogènes a conduit à une diversité et une complémentarité des réponses immunitaires. Il existe plusieurs réponses de défense adaptative et innée, qui diffèrent par leur rapidité et leur spécificité dans la reconnaissance et le traitement d'une infection pathogène.

La réponse adaptative est spécifique au pathogène envahissant et aux molécules toxiques qu'il sécrète. La capacité de distinguer ce qui est étranger de ce qui ne l'est pas, est une caractéristique fondamentale du système immunitaire inné. La réponse immunitaire innée peut distinguer différentes classes d'agents pathogènes et recruter la forme la plus efficace de réponse adaptative pour les éliminer.

A.1.1 L'immunité innée

L'immunité dite innée, ou encore appelée non spécifique, représente donc la première ligne de protection en réponse à une infection. Ce système de réponse est très conservé au cours de l'évolution.

Les cellules du système immunitaire inné des mammifères et des insectes détectent les pathogènes par l'intermédiaire de récepteurs (appelés PRR pour Pattern Recognition Receptors) reconnaissant des motifs moléculaires spécifiquement exprimés par les microorganismes (appelés PAMPs pour Pathogens Associated Molecular Pattern)¹¹⁶.

Les PAMPs sont des produits spécifiques du métabolisme microbien, qui représentent des signatures du pathogène, comme par exemple les lipopolysaccharides (LPS) chez les bactéries à Gram négatif et les peptidoglycanes chez les bactéries à Gram positif.

La drosophile, *Drosophila melanogaster*, ou mouche du vinaigre est un diptère supérieur, souvent utilisée comme organisme modèle en laboratoire. Son faible coût d'entretien et les nombreuses approches génétiques qu'elle permet ont fait un organisme de premier choix dans l'étude d'un grand nombre de mécanismes moléculaires et biologiques tels que le développement. Dépourvue de réponse immunitaire adaptative, elle s'est également révélée précieuse dans l'élucidation des mécanismes de réponse immunitaire innée.

La réponse immunitaire innée des insectes repose en partie sur la transcription, par les cellules du corps gras (équivalent du foie chez les mammifères), de gènes codant pour des peptides antimicrobiens (ou AMPs pour *Anti-Microbial Peptides*) puis à leur sécrétion dans l'hémolymphe¹¹⁷⁻¹¹⁸. Grâce à aux études génétiques menées chez *D. melanogaster*, il a été établi que l'expression de ces AMPs¹¹⁷, en réponse à une infection, dépendait de deux voies de signalisation distinctes.

La première voie, connue sous le nom de la voie du récepteur Toll, est principalement activée en réponse à une infection pathogène fongique et bactérienne (Gram+)¹¹⁹⁻¹²⁰. Sa stimulation aboutit à la synthèse d'un

premier groupe d'AMPs tel que la drosomyicine¹²¹⁻¹²² en passant par l'activation d'une protéase nommée Spatzle.

La seconde voie immunitaire, appelée «*immune deficiency*» ou IMD, est activée lors d'infection par des bactéries à Gram- par l'intermédiaire du récepteur PGRP-LC et conduit à la synthèse d'un deuxième groupe d'AMPs tels que la diptéricine¹²³⁻¹²⁴ (Figure 43).

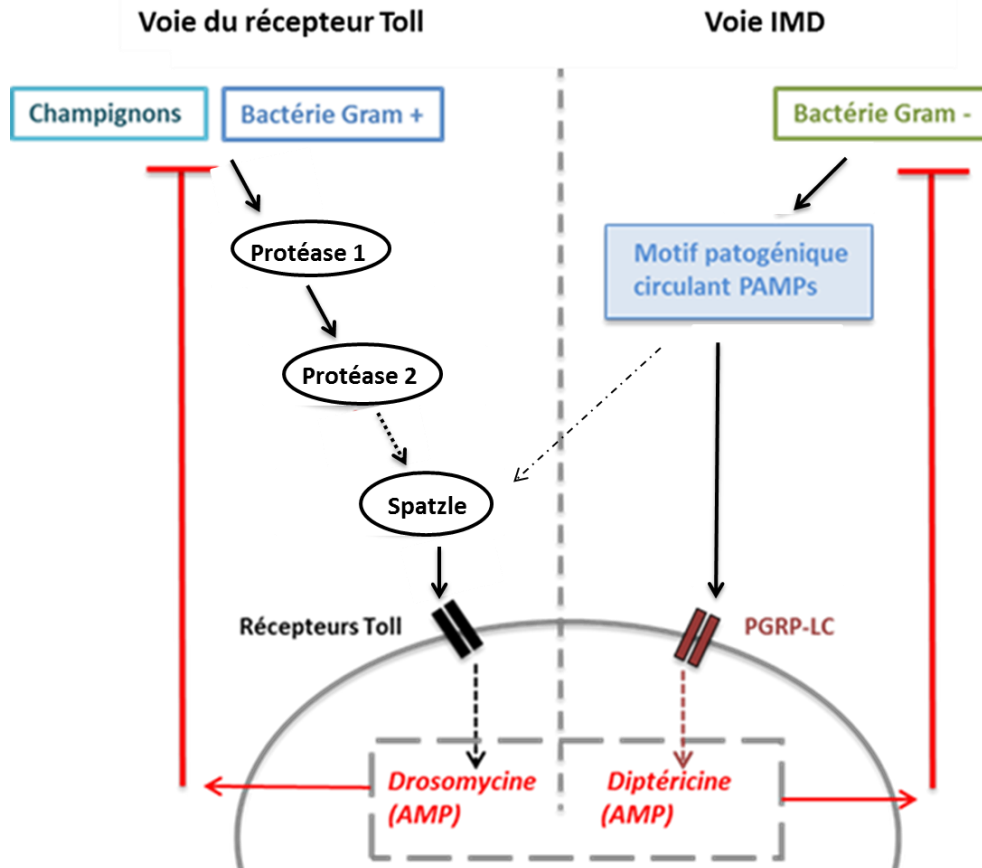


Figure 43: Représentation schématique des deux voies, celle du récepteur Toll et la voie IMD, mises en jeu dans l'immunité innée suite à une infection de pathogène. Les PAMPs interviennent dans la voie IMD mais aussi dans la voie du récepteur Toll.

A.1.2 La voie du récepteurs Toll

Au point de vue évolutif, les récepteurs *Toll-like* (TLR) semblent être une des plus anciennes voies mises en jeu dans le système immunitaire du fait de son omniprésence chez un grand nombre d'espèces. Les récepteurs TLR appartiennent à la famille des récepteurs de reconnaissance des motifs moléculaires, conservés chez de nombreux pathogènes, et qui se situent au niveau des corps gras chez les insectes.

Chez *D. melanogaster*, le récepteur Toll a été initialement découvert pour son rôle dans la mise en place de la polarité dorso ventrale chez l'embryon. L'équipe du Pr Hoffmann a ensuite démontré qu'il jouait également un rôle essentiel dans la réponse immunitaire de la drosophile¹²⁵. Contrairement aux TLRs, le récepteur Toll de la drosophile n'est pas activé directement par des PAMPs mais par la forme clivée et activée de la cytokine Spatzle. L'activation de Spatzle repose sur deux cascades protéolytiques.

La première, appelée voie des PAMPs, est activée par des PRRs circulants reconnaissant les peptidoglycanes bactériens ou fongiques. Cette reconnaissance entraîne l'auto-activation de la protéase modSP (pour modular Serine Protease) puis des protéases Grass et Spatzle Processing Enzyme (ou SPE).

La deuxième, appelée voie des "signaux de danger", n'est pas activée par des PAMPs mais par l'activité protéolytique libérée par les pathogènes. Cette voie est dépendante des protéases Perséphone et SPE, mais son mode d'activation est encore mal compris. Cette deuxième voie des "signaux de danger" est illustrée Figure 44.

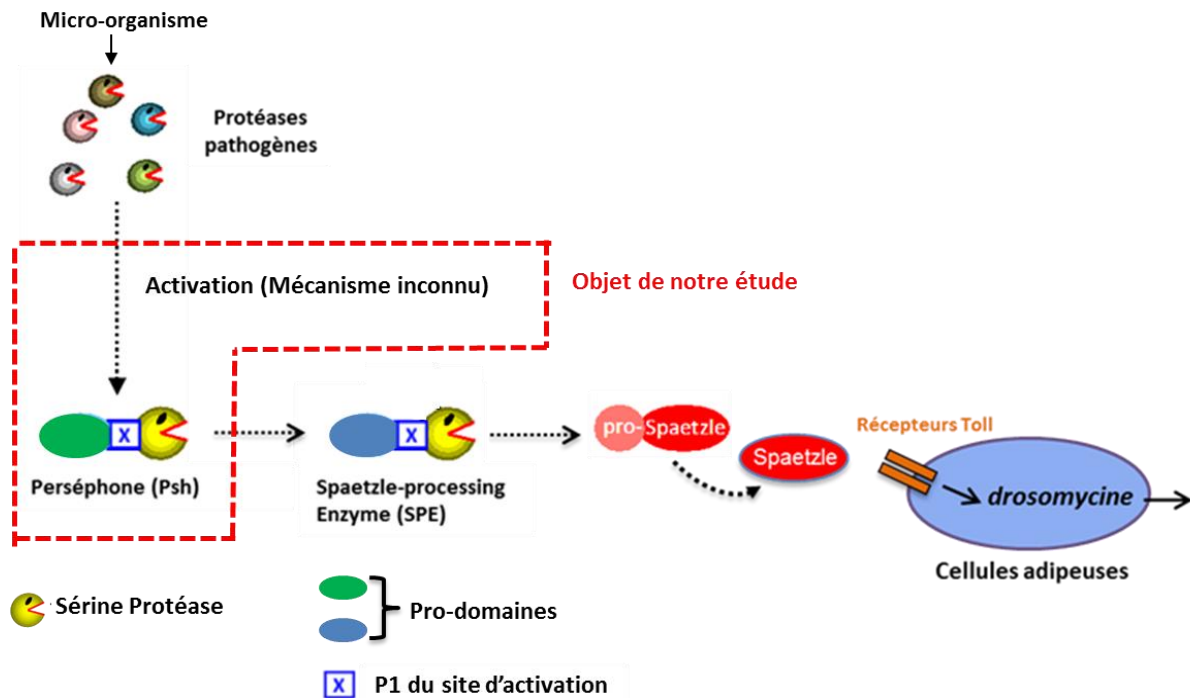


Figure 44: Représentation schématique de la cascade protéolytique mise en jeu dans la voie du récepteur Toll lors d'infection pathogènes (d'après une communication personnelle de Florian Veillard)

A.1.3 La protéase Perséphone

La voie présentée ci-dessus est dépendante de la protéase Perséphone (Psh), protéase à sérine de la famille de la chymotrypsine sécrétée dans l'hémolymphe sous forme de zymogène inactif¹²⁶. Son pro-domaine N-terminal est de type « Clip », un pro-domaine spécifique de certaines protéases à sérine d'insectes. Comme pour toutes les protéases à sérine, ce pro-domaine n'inhibe pas directement le site actif de l'enzyme. Le zymogène est inactif car le domaine catalytique n'est pas dans la bonne conformation structurale. Son activation dépend de la libération de l'extrémité N-terminale du domaine catalytique par un clivage protéolytique précis. Ce clivage intervient en amont d'une séquence hautement conservée chez les protéases à sérine de type « IVGG ». Une fois libérés, ces résidus vont participer à l'organisation structurale du site actif et permettre d'obtenir la conformation active de l'enzyme¹²⁷. Contrairement à la majorité des protéases à sérine qui possèdent une arginine ou une lysine en position P1 de leur site d'activation, Psh possède une histidine à cette position.

Parce que le zymogène de Psh est sensible *in vitro* à l'hydrolyse par la protéase Pr1A (*subtilisin-like serine protease*) du champignon *B. bassiana*, il a été proposé que l'activation de la voie du récepteur Toll par l'activité protéolytique de pathogènes serait due à une activation directe de Psh par les protéases d'origine exogène. Cependant, cette hypothèse soulève de nombreuses interrogations:

1- Bien que le zymogène de Psh soit effectivement hydrolysé par la protéase de *B. bassiana*, les produits d'hydrolyse obtenus n'ont pas été déterminés, et rien ne permet de dire si la forme active de Psh est générée.

2- La majorité des protéases ne clive pas lorsqu'il y a un résidu histidine en position P1 du site d'interaction de substrat

3- De nombreux pathogènes sécrètent des protéases comme facteurs de virulence. Cependant, la nature, la structure et la spécificité de ces protéases ne sont pas conservées d'un pathogène à l'autre.

Ces trois observations soulèvent une question : comment le zymogène de Psh peut être activé par une grande variété de protéases alors que son site d'activation est très spécifique et précis ?

Nous nous sommes donc focalisés sur la détermination précise des sites de coupure de Psh en présence de protéases spécifiques de certains micro-organismes. Comme nous le verrons, nous nous limiterons à une étude *in vitro* des clivages protéolytiques d'une Psh recombinante, avec un suivi cinétique de sa ou de ses formes maturées.

A.2. Les objectifs de l'étude

L'objectif de notre étude porte sur deux problématiques biologiques:

1- La séquence de Psh (**Erreur ! Source du renvoi introuvable.**) contient un motif «¹⁴⁴IVGG» décrit comme étant la position libérée après activation, pour participer à l'organisation structurale du site actif de la protéine. Notre objectif a été de déterminer précisément le site de clivage de Psh dans le cas d'infection par *Bacillus subtilis* (Gram+), qui comme *B. Bassania* sécrète une protéase exogène *subtilisine-like*. Nous avons donc vérifié que le site de clivage se situe bien au niveau de l'histidine en position ¹⁴³ avant le motif «¹⁴⁴IVGG ». Ceci permettrait de confirmer le rôle clé des protéases appartenant à la famille des subtilisines, dans l'activation de la voie Toll par l'intermédiaire de Psh.

2- Le zymogène de Psh peut être activé par une grande variété de protéases alors que son site d'activation est très spécifique et précis. L'hypothèse formulée par l'équipe du Pr Reichhart, et sur laquelle nous avons travaillé, est que Psh doit posséder un mécanisme capable de détecter indirectement la présence de micro-organismes par leur activité enzymatique dans l'hémolymphe¹²⁸.

Les sites de coupures de Psh par des pathogènes ne possédant pas de protéase de type *subtilisine-like* mais activant la voie du récepteur Toll par l'intermédiaire de Psh, ont été déterminés. Nous avons donc vérifié si les sites de coupure étaient au niveau de l'histidine en position 143 avant le motif «¹⁴⁴IVGG ».

B. Introduction au contexte analytique et études de N-terminomique

Les extrémités N-terminales des enzymes ou des pro-enzymes peuvent participer à leur activité, ou à leur stabilité; elles peuvent orchestrer la maturation des protéines via différentes modifications post-traductionnelles. La protéolyse est une PTM majeure et a la particularité d'être irréversible. La maturation protéolytique génère *in fine* une nouvelle extrémité N-terminale aussi appelée néo N-ter des protéines (Figure 45).

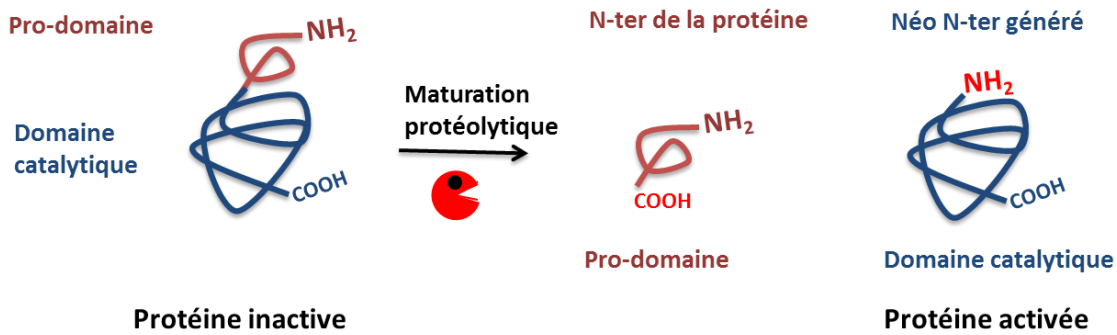


Figure 45: Représentation schématique de l'activation d'une enzyme inactive par une maturation protéolytique générant un néo N-ter de la forme activée de la protéine

L'étude et la caractérisation des produits de dégradations représentent par conséquent un réel enjeu pour comprendre et apporter des nouvelles informations sur les mécanismes biologiques¹²⁹.

Lors de la course au séquençage des génomes commencée dans les années 1970, des études de génomique et de protéomique ont été combinées pour apporter des nouvelles informations sur les séquences d'ADN. La protéogénomique a permis la correction des annotations des génomes grâce aux études de protéomique, et plus particulièrement celles des extrémités N-terminales des protéines aussi nommées «N-terminomiques». Elles sont souvent réalisées en comparant des échantillons provenant d'états physiologiques différents pour déterminer les activités enzymatiques induisant des clivages de protéines et les sites concernés.

Historiquement, la dégradation d'Edman¹³⁰ a été largement utilisée depuis les années 50, mais elle a été remplacée dans les années 90 par l'analyse protéomique bien plus sensible.

L'analyse protéomique, qui passe par l'analyse des peptides de digestion (stratégie appelée *bottom up*), ne permet pas systématiquement de caractériser les peptides N-terminaux. Déterminer l'ensemble des sites de clivage des protéines d'un extrait biologique complexe est donc impossible par analyse protéomique classique. Plusieurs stratégies ont été développées pour atteindre ce but. Elles reposent en général sur un enrichissement des peptides N-terminaux générés par la digestion enzymatique par des méthodes de chromatographie, dans un protocole d'analyse protéomique. Plusieurs approches d'enrichissement des peptides de digestion N-terminaux¹³¹⁻¹³² ont été décrites.

Dans la plupart de ces méthodes d'enrichissement, les ϵ -amines des lysines sont bloquées et les α -amines N-terminales sont marquées par un réactif spécifique sélectionné pour permettre l'enrichissement en ligne par chromatographie. Lors de l'analyse chromatographique les peptides N-terminaux des protéines présentes peuvent être détectés grâce au marquage spécifique et leur séquence déterminée par MS/MS, ce qui permet d'identifier la position N-terminale. Dans le travail présenté ici nous verrons que nous avons travaillé sur la protéine Perséphone isolée, de façon comparative, avant et après incubation avec des pathogènes.

B.1. L'état de l'art en analyse N-terminomique

Avant de présenter l'approche que nous avons utilisé pour notre travail, basée sur le marquage au TMPP (tris(2,4,6-triméthoxyphényl) phosphonium), nous présentons ci-dessous les deux méthodes d'analyse N-terminomique les plus décrites. Ces méthodes ne sont pas compatibles avec l'analyse d'échantillons isolés à partir de bandes de gels 1D-PAGE et nécessitent des quantités de matériel assez importantes. Il s'agit de la méthode TAILS (*Terminal Amine Isotope Labeling of Substrates*) et la méthode COFRADIC (*COmbined FRActional Diagonal Chromatography*).

La méthode TAILS consiste à diméthyliser les groupements α -amines et ϵ -amines primaires des protéines avec du formaldéhyde, dérivation réalisée avant la digestion par la trypsine. Les α -amines libres issues de la digestion enzymatique vont réagir avec un polymère réactif et être piégées en LC, les peptides N-terminaux diméthylés seront alors collectés et analysés par LC-MS/MS. L'utilisation du formaldéhyde marqué aux isotopes lourds et légers puis l'étape d'enrichissement permettent d'identifier et quantifier de façon relative les abondances des peptides N-terminaux par calcul des ratios lourd/léger des signaux MS des peptides N-ter¹³³. Si ce ratio est environ égal à 1, il s'agit d'une extrémité N-terminale de la protéine (avant protéolyse). Si le peptide N-ter n'est pas identifié dans le contrôle (observation d'un singulet), l'extrémité N-ter a alors été générée après protéolyse (Figure 46).

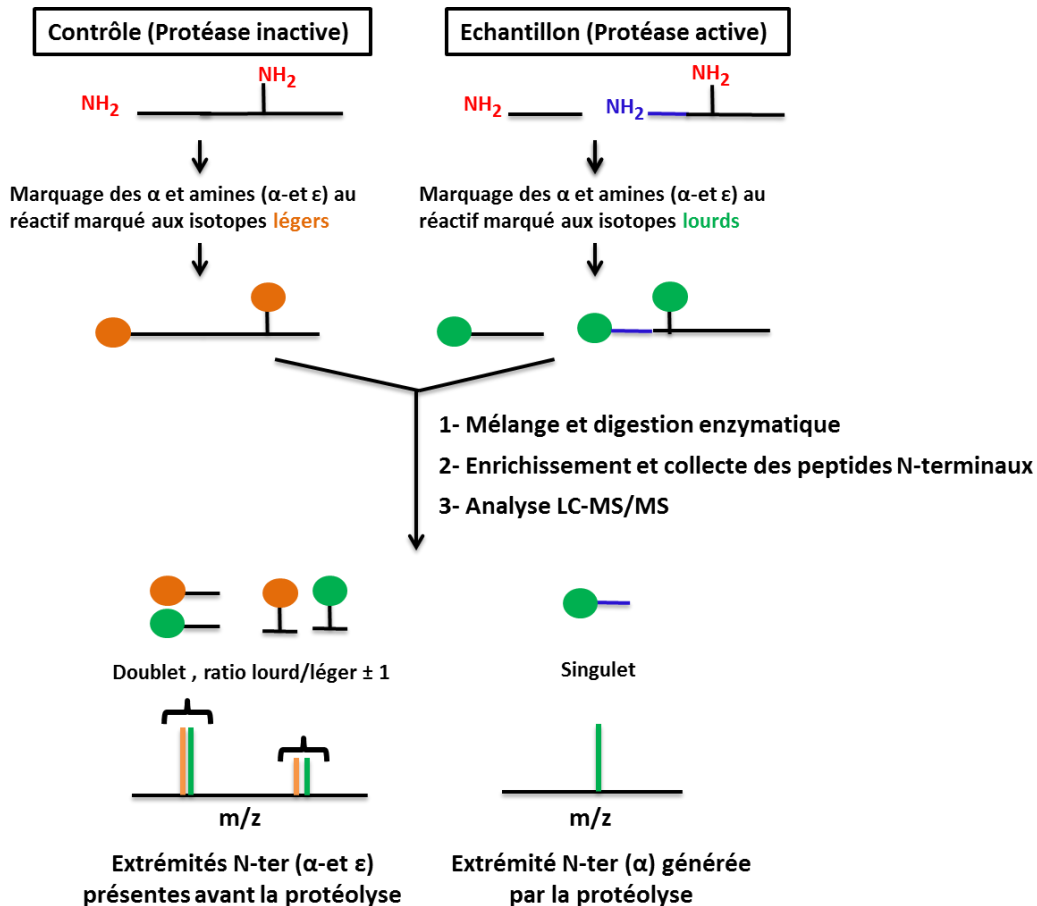


Figure 46: Représentation schématique des étapes de l'approche TAILS. Un échantillon contenant la protéase inactive et un échantillon contenant la protéase active sont comparés par marquage différentiel avec le réactif non marqué et marqué isotopiquement.

Une autre approche de marquage différentiel des N-terminaux avec un réactif isobarique iTRAQ combiné à TAILS est également décrit pour la comparaison relative des N-ter protéiques¹³². Cette approche iTRAQ-TAILS permet de comparer jusqu'à huit échantillons et de quantifier les N-ter en mode MS/MS.

La méthode COFRADIC¹³⁴ est basée sur l'acétylation des groupes amines (α - et ϵ -) avant digestion des protéines. Le mélange peptidique est ensuite séparé par RP-HPLC et chaque fraction peptidique collectée contient donc deux types de peptides: les internes avec un NH₂ libre et les N-terminaux avec une terminaison NH₂ bloquée. Les fonctions amines libres des peptides internes sont ensuite dérivées par l'acide 2,4,6-Trinitrobenzènesulfonique (TNBS). L'ajout de ce groupement hydrophobe va retarder l'élution des peptides internes par RP-HPLC. Cette séparation permet d'obtenir une fraction enrichie en peptides N-terminaux. Ce fractionnement poussé peut cependant conduire à une perte d'échantillon et reste chronophage en termes

de temps d'analyse. L'introduction de réactifs marqués isotopiquement (lourd/léger) permet, également, de quantifier de manière relative les protéines et de déterminer l'origine des extrémités N-ter¹³¹.

La méthode COFRADIC permet également la quantification relative des terminaisons N-ter (comme la trideutéro-acétylation d'amines libres primaires dans la première étape pour déterminer le degré d' α -N-acétylation naturelle) ou par introduction de métabolites marqués par l'approche SILAC¹³¹. Une des variantes rapportées est la méthode chaFRADIC (*charge-based fractional diagonal chromatography*) où les N-ter protéiques sont diméthylés, puis les protéines sont digérées et fractionnées en utilisant une chromatographie d'échange d'ions (SCX)¹³⁵. Les peptides N-ter diméthylés auront un état de charge supérieur aux peptides internes trideutéro-acétylés et pourront alors être séparés par chromatographie SCX-HPLC. Une autre variante a été également décrite, combinant l'avantage du marquage spécifique des N-ter des protéines TMPP (après acétylation des amines des lysines) avec l'efficacité d'enrichissement COFRADIC¹³⁶.

La mise en place d'une stratégie multi étapes comme celle de COFRADIC s'avère fastidieuse et coûteuse en termes de temps et en quantité de matériel biologique.

Les deux stratégies TAILS et COFRADIC, énoncées plus haut, ne nous ont pas semblé adaptées pour l'étude à mener. En effet, dans l'étude des clivages protéolytiques d'une protéine recombinante, nous voulions suivre la cinétique de la formation des formes maturées de la protéine. Or notre étude cinétique de la protéolyse de Psh n'a pu être réalisée qu'à partir de bandes de gel 1D PAGE, principalement pour des raisons de quantité de matériel disponible.

Pour répondre aux exigences de notre étude, nous avons développé une stratégie analytique basée sur l'adaptation de la méthode de marquage TMPP complètement compatible *in gel*¹³⁷. Avec cette méthode de marquage TMPP aucune étape de blocage des ϵ -amines et ni d'enrichissement n'est nécessaire. Cette méthode présente l'énorme avantage d'être compatible avec une stratégie classique de protéomique. Notre objectif analytique a été de mettre en œuvre une méthode permettant de suivre *in vitro* la cinétique de la protéolyse de r-Psh, et de caractériser finement les sites de clivage protéolytique après maturation.

B.1.1 La caractérisation des extrémités N-terminales par marquage au TMPP-Ac-OSu

En 2009, Gallien et *al.* ont publié une méthodologie baptisée N-TOP pour N-Terminal Oriented Proteomics, développée au laboratoire, basée sur une dérivation chimique en solution à l'aide d'un réactif le TMPP-Ac-OSu (pour N-succinimidylloxycarbonylmethyl) tris(2,4,6-triméthoxyphényl)phosphonium, appelé TMPP par la suite). Il est nécessaire de contrôler la régiosélectivité afin que le TMPP-Ac-OSu réagisse préférentiellement avec les α -amines libres primaires et non avec les ϵ -amines de la chaîne latérale des lysines (Figure 47). Pour favoriser la dérivation des α -amines aux extrémités N-ter ($pK_a \sim 7,8$), le pH de la réaction doit être fixé à 8,2¹³⁸. Les amines des chaînes latérales des lysines seront alors protonées et non réactives ($pK_a \sim 11$).

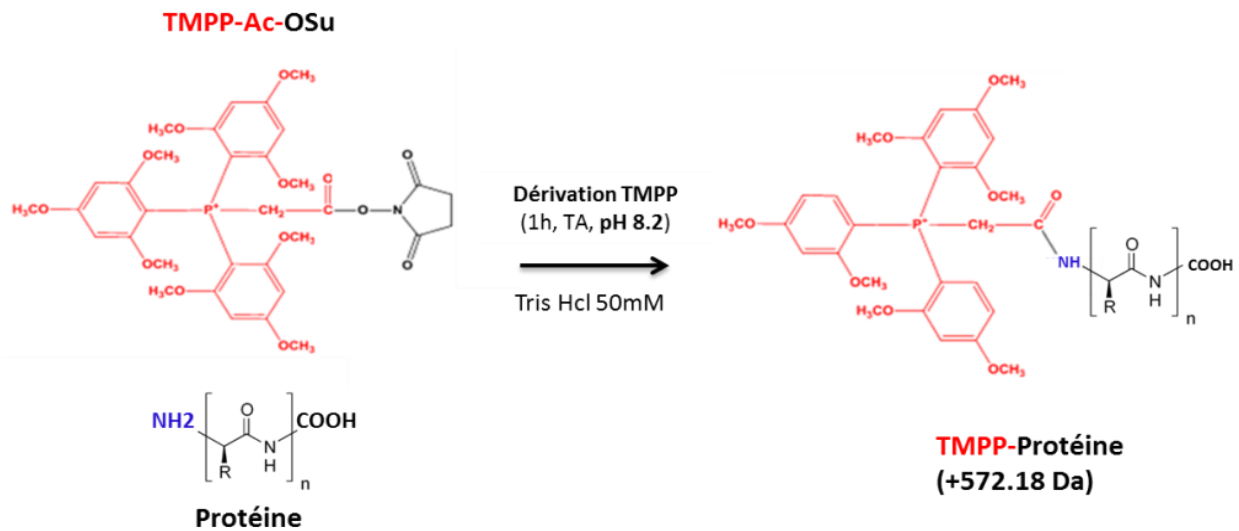


Figure 47: Représentation schématique de la réaction chimique sélective entre le TMPP-Ac-OSu et la fonction amine N-terminale d'une protéine à pH 8,2.

Les peptides sont séparés par chromatographie sur une colonne de phase inverse (RP-HPLC). L'étiquette TMPP des peptides dérivés va permettre de les séparer chromatographiquement du reste des peptides. En fin de chromatogramme une fraction enrichie en peptides N-terminaux dérivés, est ainsi observée permettant d'augmenter leur probabilité d'identification en analyse MS et MS/MS. Cette stratégie permet l'identification des peptides N-terminaux dérivés tout en conservant la détection des peptides internes.

En général, pour les peptides tryptiques internes la fragmentation favorisée en mode CID est la génération de fragments type b et y¹³⁹. De par la présence de la charge permanente (portée par le phosphonium), la fragmentation des peptides dérivés est affectée. La formation de fragments type a et b est alors favorisée. Cette fragmentation atypique nécessite l'assignation manuellement des spectres MS/MS. En 2013, D. Bertaccini a développé au laboratoire, une approche permettant une validation sûre et automatisée pour l'identification des N-ter des protéines, dérivés au TMPP légers et lourds dans un même échantillon.

Le marquage « *in gel* » des protéines au TMPP a été décrit pour la première fois par Ayoub et al.¹⁴⁰, adapté de l'approche N-TOP, pour étudier les processus protéolytiques et caractériser les micro-hétérogénéités des extrémités N-ter des anticorps monoclonaux¹³⁷.

B.1.2 L'optimisation du marquage *in gel* au TMPP et exploitation des données LC-MS/MS

Nous avons dû adapter les conditions de marquage développées par D. Ayoub et al. pour suivre le processus de protéolyse par gel 1D PAGE afin de répondre aux exigences suivantes: rapidité de la méthode et facilité de mise en place du protocole.

Optimisation du marquage *in gel* sur une protéine modèle. Notre projet d'étude a été de suivre la cinétique d'une seule protéine, nous voulions simplifier le protocole et éliminer l'urée (indésirable pour la MS), le détergent utilisé dans la méthode d'origine. Nous voulions également éviter au maximum les espèces observées après la formation des produits des réactions parasites du TMPP-Ac-OSu.

Eluées tardivement dans la même zone d'éluion de peptides hydrophobes portant un groupement TMPP, la présence de ces produits pourrait gêner l'identification des N-ter dérivés. Nous avons donc optimisé une méthode de dérivation *in gel* avec 250 équivalents de TMPP-Ac-OSu dans un tampon Tris HCl. Après avoir testé 200 et 400 équivalents, cela a été le meilleur compromis entre efficacité de marquage et la formation des produits de dégradation du TMPP.

Notre approche a été testée sur une protéine modèle l'hémoglobine (HbA0), constituée de deux chaînes de globines α et β donc deux extrémités N-terminales. Pour être dans le même ordre de grandeur que les échantillons biologiques du projet, l'approche a été testée sur 5 μ g d'hémoglobine.

Après migration sur gel d'électrophorèse 1D-PAGE, la bande de gel extrudée contenant l'hémoglobine est soumise à la dérivation TMPP-Ac-OSu avec 250 équivalents. Un tampon Tris-HCl dont le pH a été ajusté à 8,2 est ensuite ajouté. Au bout d'une heure de l'hydroxylamine est ajoutée pour neutraliser l'excès des molécules de TMPP-Ac-OSu et éliminer les potentiels marquages aspécifiques des groupements hydroxyles des tyrosines. Les produits de dégradation liés à l'hydrolyse du TMPP dans l'eau ou du TMPP avec le Tris HCl, sont éliminés par quatre cycles de lavages de la bande de gel. La bande est alors soumise au protocole classique de digestion à la trypsine pour une analyse protéomique.

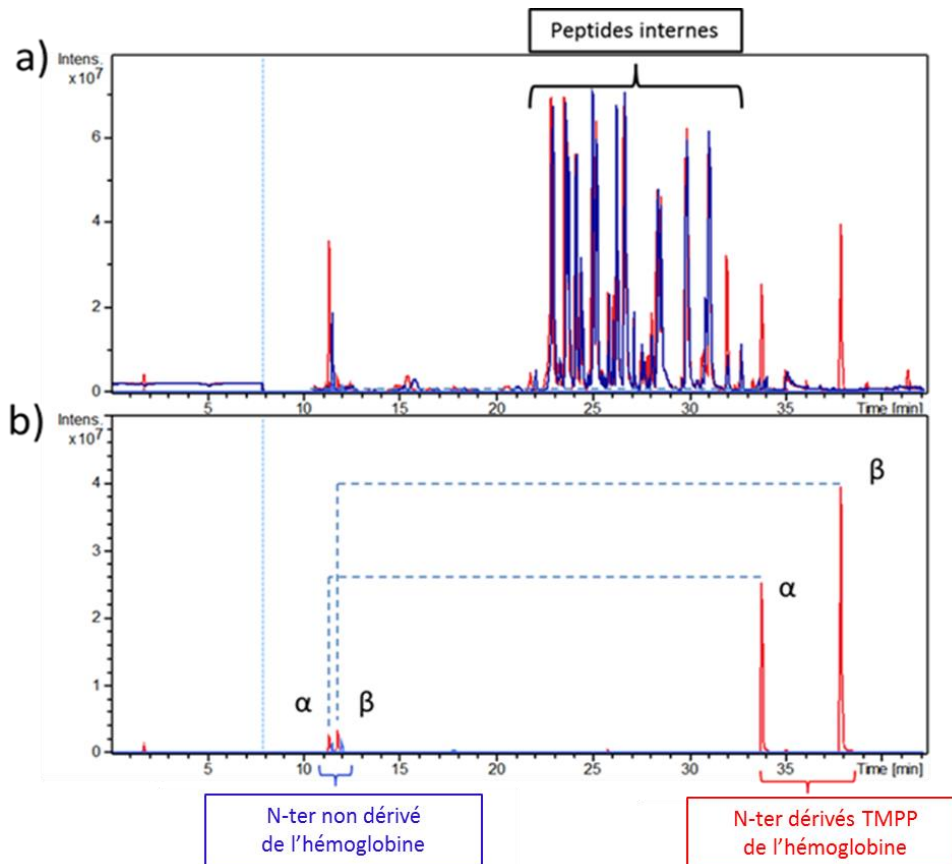


Figure 48: Superposition des BPC (base peak chromatogram) obtenus en injectant un digestat trypsique d'une protéine modèle (l'hémoglobine) dérivée en rouge et non dérivée en bleu (a). Superposition des EIC des peptides N-terminaux des chaînes α et β non dérivés à gauche, dérivés à droite (b).

La Figure 48 présente la séparation des peptides de digestion obtenue lors de l'analyse en LC-MS/MS. Ce chromatogramme montre que les peptides marqués TMPP sont élués plus tardivement que leurs homologues non marqués et que les peptides internes de l'hémoglobine. La charge positive permanente apportée par le TMPP permet une meilleure ionisation des peptides dérivés comme le montrent les courants d'ions extraits pour les peptides N-terminaux qui sont plus intenses que dans le contrôle (Figure 48). La stratégie de marquage au TMPP appliquée à l'hémoglobine permet donc d'identifier distinctement deux extrémités N-ter α et β dérivées au TMPP.

Cette stratégie de marquage *in gel* développée a été appliquée stratégie pour l'étude des produits de dégradation de la protéine recombinante Psh (r-Psh).

Stratégie et exploitation bioinformatique des données LC-MS/MS. Les différentes étapes de la stratégie de dérivation *in gel* développées sont présentées Figure 49. Pour l'étude de la cinétique de Psh, contrairement à un protocole classique utilisant le DTT, le tributylphosphine (TBP) est utilisé pour réduire les ponts disulfures de la protéine avant marquage. Le TBP ne possédant pas de fonctions réactives thiols, il permet d'éviter les réactions parasites avec le TMPP.

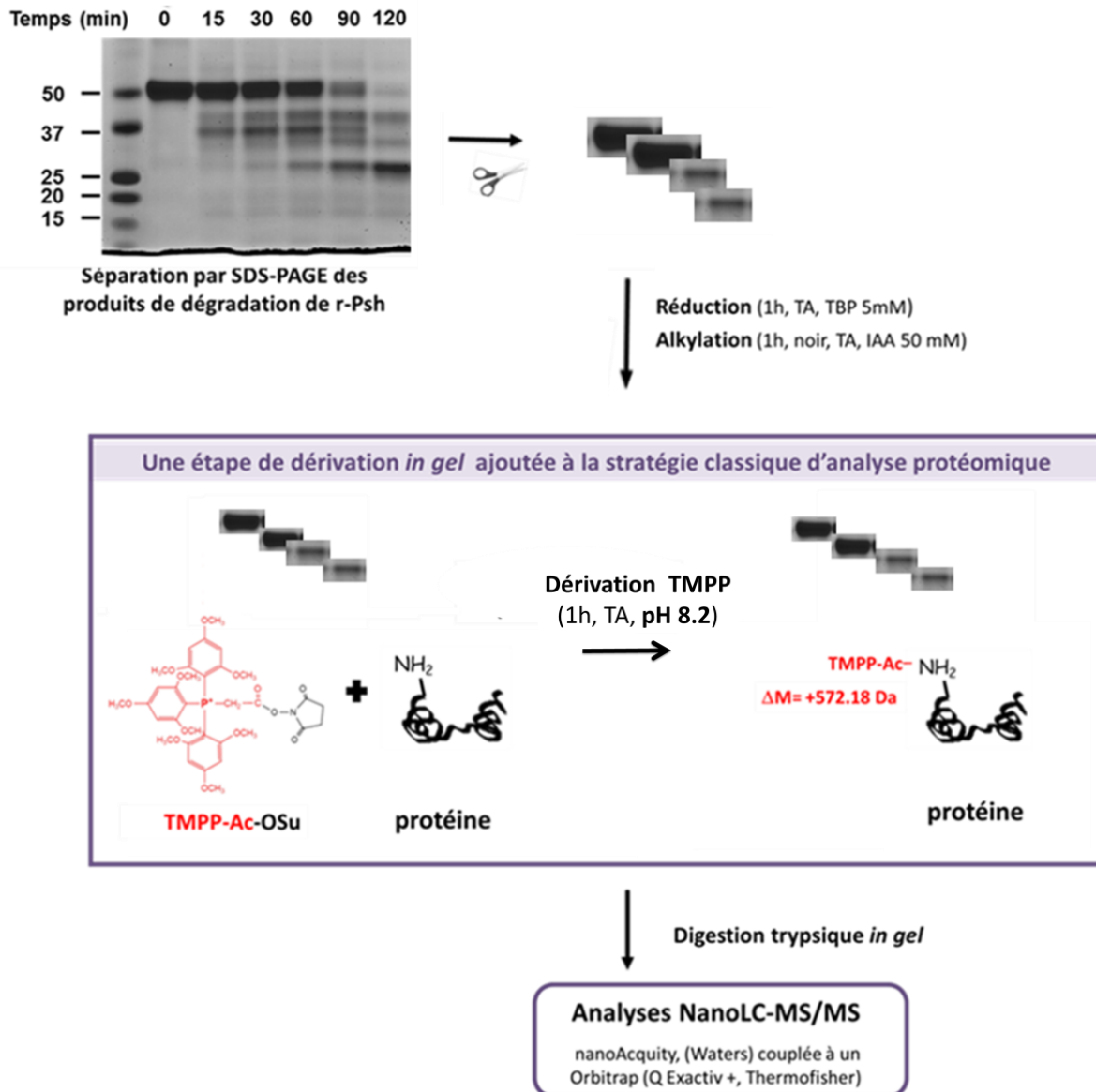


Figure 49: Stratégie de marquage *in gel* au TMPP-Ac-OSu des extrémités N-terminales des produits de dégradation protéolytique de r-Psh.

L'exploitation des données des LC-MS/MS pour l'identification des peptides de digestion nécessite bien sûr l'ajout de la séquence d'acides aminés de notre protéine recombinante dans la banque Swissprot (Uniprot). La modification TMPP (+572 Da) a été ajoutée dans les paramètres du moteur de recherche Mascot^{38,39} afin d'identifier les peptides dérivés au TMPP. La recherche des peptides internes a permis l'identification de la séquence de r-Psh.

Les paramètres de recherche Mascot doivent prendre en compte la fragmentation particulière des peptides portant la modification TMPP. Les fragments «a» caractéristiques de la présence du TMPP sont implémentés dans les critères de recherche MS/MS. Les scores d'ion Mascot des peptides dérivés sont généralement plus faibles que ceux des peptides non dérivés, malgré leur fragmentation de haute qualité. Peu de fragments sont situés dans la zone des masses inférieures des spectres MS/MS et cela crée ainsi un biais dans la notation

des scores attribués par Mascot. Les spectres des peptides marqués au TMPP sont donc vérifiés manuellement et ceci à l'aide du logiciel Scaffold qui permet de visualiser le spectre de fragmentation attribué aux peptides identifiés.

C. Résultats

C.1. L'origine et préparation des échantillons de r-Psh incubée

Stratégie de production et caractérisation de la r-Psh. La séquence Psh de la drosophile est donnée ci-dessous (Figure 50). Nous notons les trois acides aminés impliqués dans la triade catalytique dans le site actif de l'enzyme: l'histidine en position 187, l'aspartate en position 234 et la sérine en position 339, responsable de l'activité protéolytique de la protéase à sérine. Nous remarquons la présence de trois sites consensus de N-glycosylation dans cette séquence.

```
MPLKWSLLLGTFLVLSICSSVEAAVTVGRACKVDTMPGICRTSSDCEPLIDGYIKSGVLTLDV
PSCGLGAWGEIFCCPTKPCCDNSTSVSTSTTKAPMTSGRVDVPTFGSGDRPAVAACKKI
RERKQQRSGNQLVHIVGGYPVDPGVYPHMAAIGYITFGTDFRCGGSLIASRFVLTAA187HCV
NTDANTPAFVRLGAVNIENPDHSYQDIVIRSVKIHPQYVGNKYN234DAILELERDVVETDNIR
PACLHTDATDPPSNSKFFVAGWGVINVTTRASKILLRAGLELVPLDQCNISYAEQPGSIRLLK
QGVIDSLLCAIDQKLIADACKGD339SGGPLIHELNVEDGMYTIMGVISSGFGCATVTPGLYTRV
SSYLDFIGIVWPDNRV
```

Figure 50: Séquence de Psh chez la drosophile. Les sites consensus de N-glycosylation sont encadrés en rouge. L'histidine en position 187, l'asparagine en position 234 et la sérine en position 339, responsables de l'activité de la protéolytique de Psh sont en vert.

La protéine r-Psh a été exprimée en système de lignées cellulaires S2 de drosophile. Afin de permettre une purification de la r-Psh, deux étiquettes ont été ajoutées en position C-terminale de la protéine recombinante: une séquence spécifique (V5) ainsi qu'une séquence polyhistidine. La séquence de r-Psh est donc composée de quatre parties (Figure 51):

- Le peptide de sécrétion (1^{er}-22^{ème} a.a), qui est éliminé lors de la sécrétion de Psh
- Le pro-domaine du (23^{ème}-143^{ème} a.a)
- Le domaine catalytique du (144^{ème}-395^{ème} a.a)
- Les tags nommés V5 (18 acides aminés) et HisTag de six histidines pour la purification (396^{ème}-425^{ème} a.a).

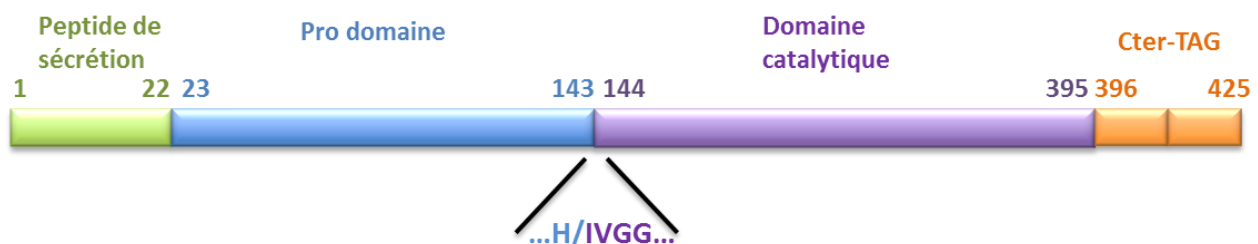


Figure 51 : Représentation schématique de la séquence de la Psh recombinante utilisée dans les expériences *in vitro*. Le site de coupure menant à l'activation du domaine catalytique est mentionné «¹⁴³H/¹⁴⁴IVGG».

Le peptide signal (position 1 à 22) est éliminé par le système de sécrétion dans le milieu extracellulaire. La purification est faite à partir du surnageant cellulaire. Le tag V5 est présent dans le plasmide d'expression utilisé pour des raisons techniques, mais il n'est pas utilisé pour l'étape de purification. Le HisTag (polyhistidine) a été utilisé pour purifier la protéine r-Psh par chromatographie d'affinité sur colonne de nickel-sépharose.

Vérification de l'identité et évaluation de la pureté de r-Psh. Avant de soumettre l'échantillon de r-Psh à des expériences d'incubation avec des pathogènes, il était important de caractériser et d'évaluer la pureté de la protéine obtenue après ce processus de purification. L'échantillon de r-Psh a donc été analysé par spectrométrie de masse en mode LC-MS (voir Figure 52).

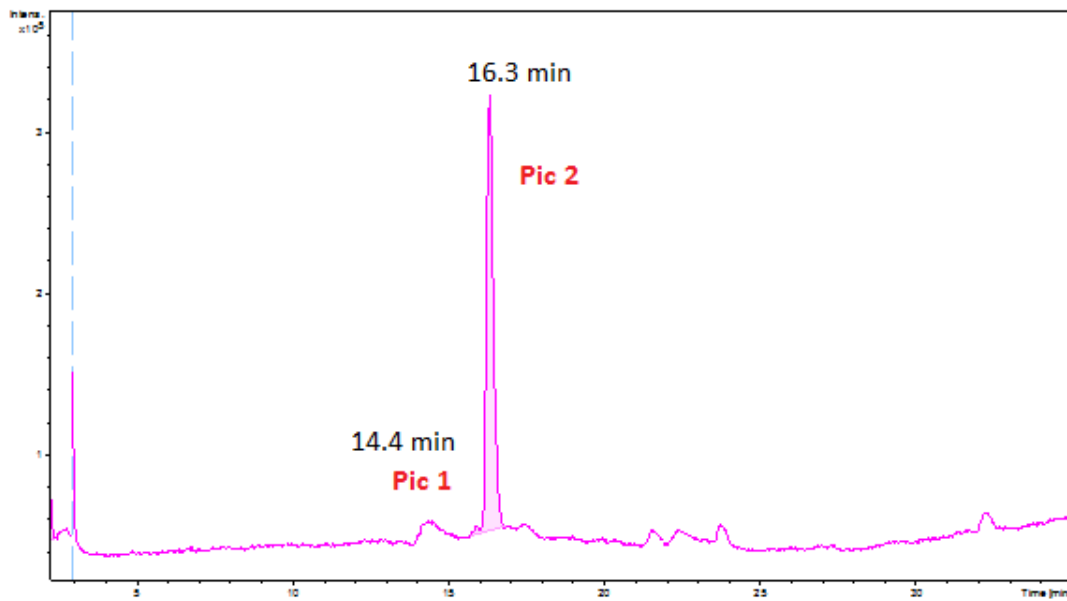


Figure 52 : LC-MS de r-Psh dans le tampon Tris HCl (TIC : Total Ion Current). Pour le pic 2, le spectre ESI-MS a montré une grande hétérogénéité; néanmoins une masse de 49740 Da a pu être mesurée sur le composé qui semble majeur. Pour tous les autres pics observés, aucune masse n'a pu être mesurée, à cause d'une grande hétérogénéité. Ces hétérogénéités sont très probablement causées par des glycosylations très hétérogènes. La colonne de chromatographie est une phase inverse C4 avec élution H₂O/ACN, 0.1% d'acide formique en 40 min. Le spectromètre de masse est un MicroQTOF (Bruker).

Le spectre de masse enregistré pendant l'analyse LC-MS est présenté ci-dessous. L'axe des abscisses correspond au rapport m/z observé pour les ions multichargés observés (Figure 53). Le spectre observé n'est visiblement pas celui d'une protéine homogène, comme attendu à cause des N-glycosylations.

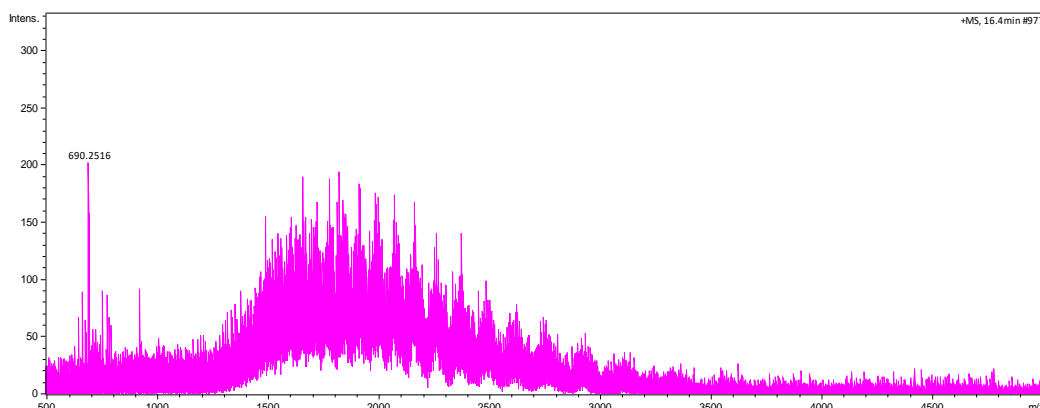


Figure 53 : Spectre de masse du pic chromatographique n° 2 (r-Psh) de r-Psh dans le tampon Tris HCl. Une forme majeure a pu être mesurée à 49 740 +/- 2Da

La masse théorique calculée à partir de la séquence en acides aminés de la r-Psh attendue est de 43 195.19 Da. De nombreuses protéoformes sont présentes. Une forme majeure a néanmoins pu être mesurée avec une masse de 49 740 +/- 2 Da (Figure 54). Ce spectre de masse « brut » obtenu sur r-Psh est ensuite « déconvolué », c'est-à-dire transformé en spectre de masse où l'axe des X correspond à des masses moléculaires exprimées en Da.

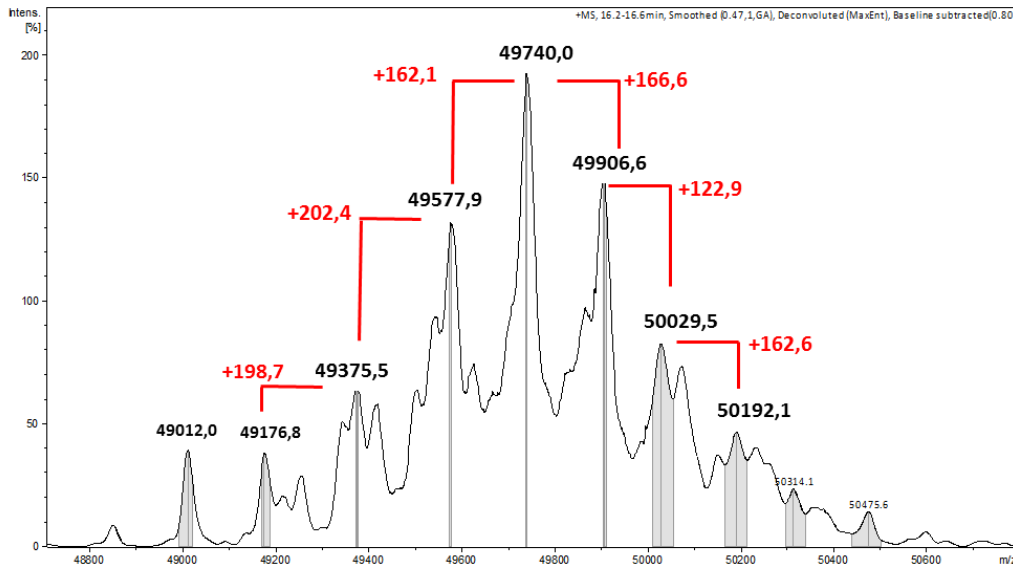


Figure 54: Spectre de masse lissé et déconvolué du pic n°2 (r-Psh) de r-Psh incubée uniquement dans le tampon Tris HCl. Les écarts de masse entre les pics correspondent relativement bien à des masses de sucres, ce qui confirme que l'hétérogénéité observée est bien due à des glycosylations.

Afin de vérifier que toute la séquence de la r-Psh était correcte, nous avons réalisé deux digestions enzymatiques (trypsine et pepsine). Les digestats ont ensuite été analysés par LC-MS/MS. Le recouvrement de la séquence d'acides aminés est proche de 65% comme le montre le Tableau 4 pour la digestion trypsique.

1	MPLKWSLLLG	TFVLISCSSV	EA ²³ AVTVGRAC	KVTDTMPGIC	RTSSDCEPLI
51	DGYIKSGVLT	LNDVPSCGLG	AWGEIFCCPT	KPCCD ^N STIT	SVSTSSTTST
101	KAPMTSGRVD	VPTFGSGDRP	AVAACKKIRE	RKQQRSGNQL	VIHIVGGYPV
151	DPGVYPHMAA	IGYITFGTDF	RCGGSLIASR	FVLTAAHCVN	TDANTPAFVR
201	LGAVNIENPD	HSYQDIVIRS	VKIHPQYVGN	KYNDIAILEL	ERDVVETDNI
251	RPACLHTDAT	DPPSNKFFV	AGWGVLN ^V TT	RARSKILLRA	GLELVPLDQC
301	^N ISYAEQPGS	IRLLKQGVID	SLLC ^A IDQKL	IADACKGDSG	GPLIHELNVE
351	DGM ^Y TIMGVI	SSGFGCATVT	PGLYTRVSSY	LD ^F IEGIVWP	DNRVLES ^R GP
401	FEGKPIPNPL	LGLDSTRTGH	HHHHH		

Tableau 4: Recouvrement de la séquence protéique de r-Psh après digestion par la trypsine (65%)

Nous avons noté que pour chacune des digestions, les peptides contenant les trois sites consensus de N-glycosylation n'ont pas été identifiés. La longueur des peptides trypsiques comportant des sites de N-glycosylation (>13 a.a), ainsi que la masse ajoutée pour les glycosylations attendues, expliquent que ces peptides n'aient pas été identifiés sous leur forme non glycosylée dans les analyses LC-MS/MS.

Cet échantillon de r-Psh a été soumis à deux types d'incubation :

-Une incubation avec le pathogène *Bacillus subtilis* (Gram+), décrite dans le paragraphe Partie II -C.2

-Une incubation avec d'autres pathogènes, *Staphylococcus aureus* (Gram+), *Serratia marcescens* (Gram-) et *Pseudomonas aeruginosa* (Gram-) décrite au paragraphe Partie IIC.3.

C.2. L'étude de la protéolyse de r-Psh par « *Bacillus subtilis* » (Gram+)

Digestion et suivi cinétique par 1D-PAGE.

20 µl de la forme recombinante et purifiée de Psh (0.4 µg/µl) ont été incubée à 29 °C avec 20 µl de surnageant de culture de *B. subtilis* dans du tampon Tris-HCl 0.1 M pH 8. A différents temps d'incubation (5, 15, 30, 60, 120 et 240 min), 5 µg de Psh ont été prélevés et la réaction est stoppée à 100 °C pendant 5 min en présence de tampon échantillon réducteur (NuPAGE LDS Sample buffer, Invitrogen). Deux contrôles ont été réalisés : r-Psh seule ou incubée pendant 240 min dans les mêmes conditions en présence de surnageant préalablement inactivé par la chaleur (10 min à 100 °C). Les échantillons ont ensuite été séparés par électrophorèse SDS-PAGE pendant 2 heures à 100 Volt en utilisant de NuPAGE 4-12% Bis-Tris Gel (Invitrogen) et colorés avec du bleu colloïdal (SimplyBlue Safestain, Invitrogen). Le gel obtenu est présenté Figure 55.

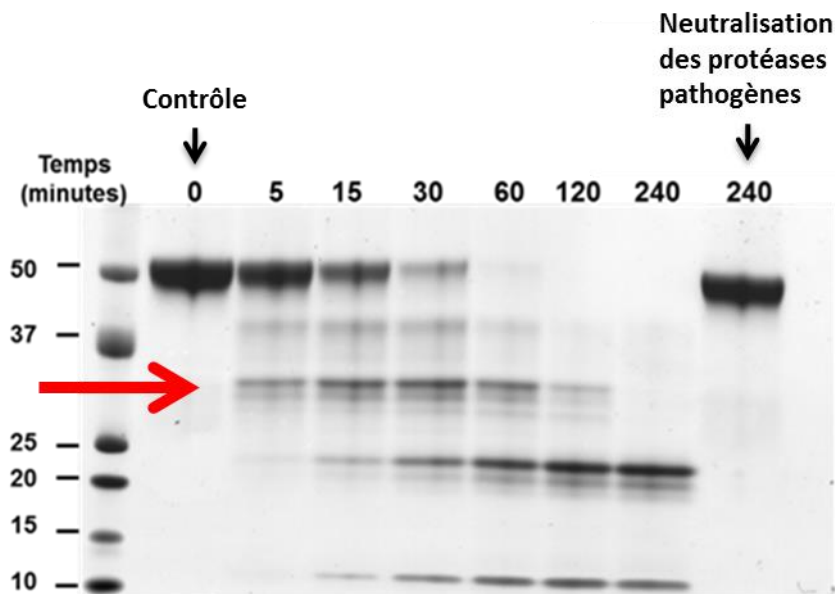


Figure 55: Gel d'électrophorèse SDS 1D-PAGE de la protéolyse de r-Psh par *B. subtilis* en fonction du temps. La flèche rouge indique la bande qui correspondrait au produit d'un clivage en position 143-144.

Ce gel montre la bande 4 à environ 30 kDa qui apparaît en premier et sa masse apparente est celle qui correspond le mieux à une forme de r-Psh maturée au niveau du site d'activation après la position ¹⁴³H (Figure 56). Nous notons qu'au cours de l'incubation, d'autres bandes apparaissent à environ 40, 25 et 10 kDa. Pour déterminer les positions N-terminales générées par l'action de protéases du pathogène, les différentes bandes numérotées qui apparaissent avec le temps, ont été extrudées (Figure 56) et ont été soumises à la dérivation *in gel* au TMPP puis analysées par LC-MS/MS.

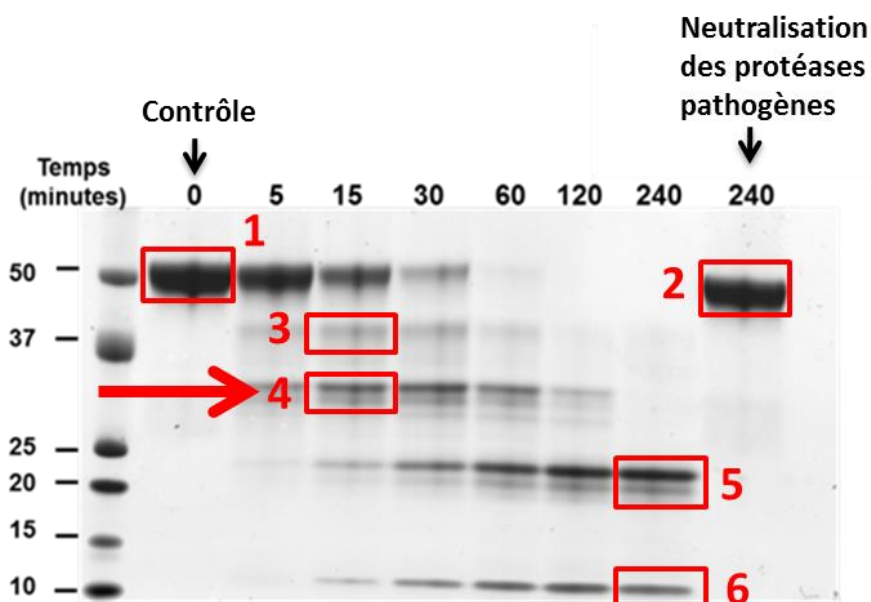


Figure 56: Bandes de gel 1D-PAGE sélectionnées pour être soumises à la dérivation *in gel* puis à l'analyse LC-MS/MS (entourées en rouge). La bande contenant probablement une forme de r-Psh maturée au niveau du site d'activation après la position ¹⁴³H est fléchée en rouge.

Dans un souci de clarté, nous avons choisi de présenter les résultats des analyses protéomiques de la façon suivante:

- Dans un premier temps, nous présentons l'identification des acides aminés N-terminaux dérivés au TMPP pour les six bandes. En effet, ce résultat permet de répondre à la question biologique posée à savoir l'action protéolytique des enzymes du pathogène au niveau du site présumé de clivage de Psh (143-144).
- Dans un deuxième temps, nous présentons une exploitation plus poussée des données de LC-MS/MS de façon à illustrer les capacités de l'analyse protéomique et l'intérêt de développer encore d'autres stratégies. Les données d'identification des différentes bandes et leurs cartes de recouvrement protéique pour les formes de r-Psh des cinq bandes (1, 3, 4, 5 et 6) sont présentées de façon détaillée.

C.2.1 La détermination des sites de clivage marqués au TMPP

Ce paragraphe est focalisé exclusivement sur la recherche des sites dérivés au TMPP dans les six bandes extrudées. Le tableau ci-dessous résume les résultats obtenus.

Numéro de bande	Acide aminé marqué au TMPP	Séquence du peptide N-terminal
1	²³ A	²³ AVTVGR ²⁸
2	²³ A	²³ AVTVGR ²⁸
3	²³ A	²³ AVTVGR ²⁸
4	¹⁴⁴ I	¹⁴⁴ IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF ¹⁷¹
5	¹⁴⁴ I	¹⁴⁴ IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF ¹⁷¹
6	n.d.	n.d

Tableau 5: Résultats des identifications des peptides-N-terminaux dérivés au TMPP pour les six bandes.

Ces résultats d'identification des acides aminés marqués au TMPP des analyses nanoLC-MS/MS sont décrits en détail ci-dessous:

- Analyse de la bande 1 (contrôle) qui correspond à la forme de Psh intacte (50 kDa)
- Analyse de la bande 2 (contrôle avec inactivation des protéases du pathogène, 50 kDa)
- Analyse de la bande 4 qui correspond à la masse attendue de 30 kDa pour un clivage en position 143-144.
- Analyse des bandes 3, 5 et 6 qui semblent correspondre à des produits inattendus.

Analyse de la bande 1 (contrôle).

Dans le contrôle (t=0 min), à 50 kDa, l'analyse nanoLC MS/MS de la bande de gel a permis d'identifier le peptide $^{23}\text{AVTVGR}^{28}$ dérivé au TMPP en position N-terminale. Le rapport du signal sur bruit mesuré pour ce peptide est intense ($1.2 \cdot 10^9$) comme cela est présenté Figure 57. Le temps de rétention de ce peptide ($t_r=26.1$ minutes) correspond bien à celui d'un peptide portant un groupement hydrophobe TMPP. La position ^{23}A correspond au premier acide aminé de la séquence du pro domaine de Psh, après le peptide de sécrétion. La forme intacte avec le N-ter en position 23 est donc bien celle attendue. Les peptides internes permettent de couvrir plus de 80% de la séquence protéique.

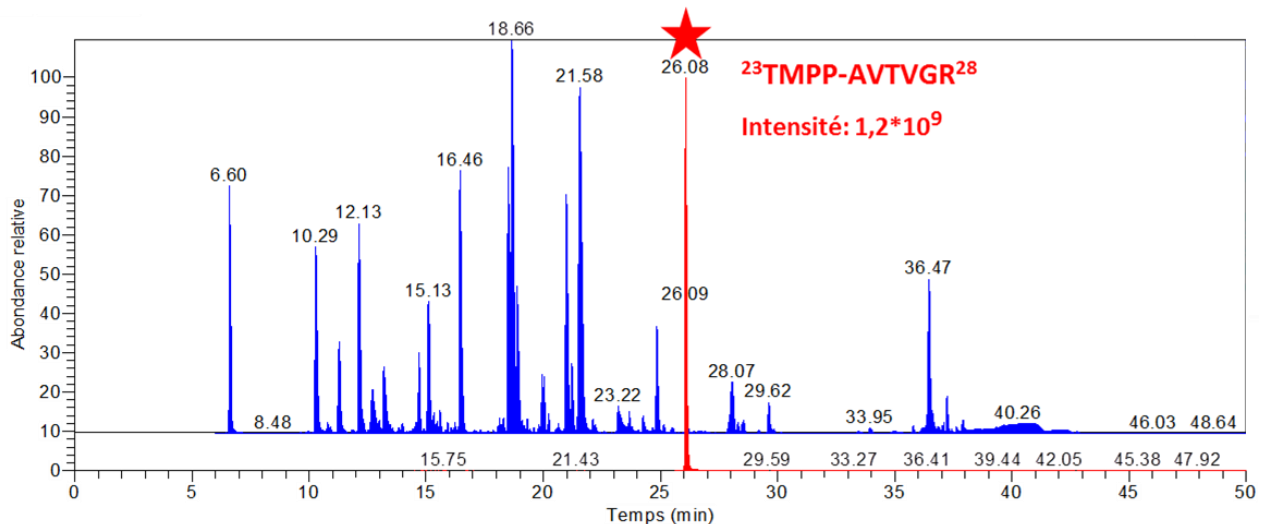


Figure 57: Chromatogramme des peptides tryptiques de r-Psh (BPC, en bleu) de la bande 1 qui est l'échantillon «contrôle». Extraction du courant d'ion m/z 587.77(en rouge) correspondant au peptide N-terminal $^{23}\text{AVTVGR}^{28}$ dérivé au TMPP.

Le spectre attribué au peptide $^{23}\text{AVTVGR}^{28}$, par le moteur de recherche, fournit les informations nécessaires pour valider manuellement son identification. La totalité des fragments «b» générés lors de la fragmentation du peptide dérivé sont présents (fragments «b1 à b5») et permet d'assigner le spectre au peptide $^{23}\text{AVTVGR}^{28}$. Le fragment «b1» permet de vérifier la position du TMPP en N-terminale. Les fragments «a» présents sur le spectre sont une empreinte caractéristique de la présence du réactif (Figure 58).

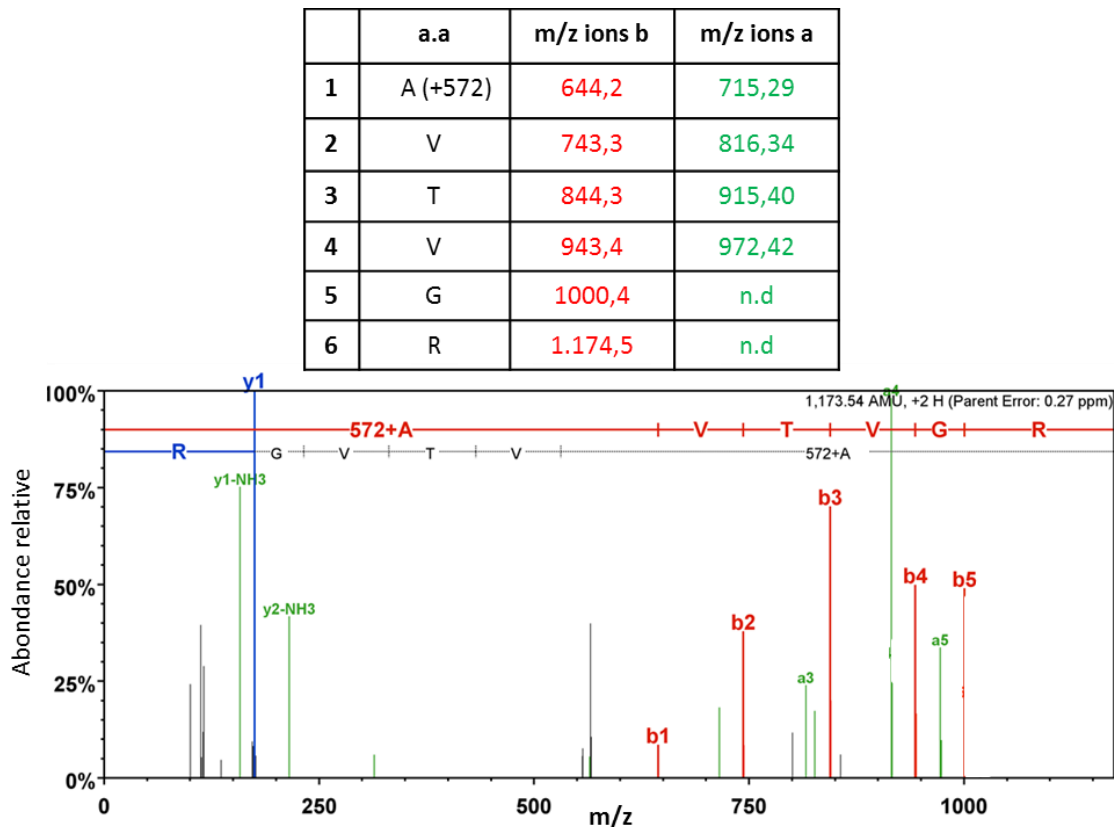


Figure 58: Tableau de fragments générés et spectre de fragmentation MS/MS du peptide TMPP-²³AVTVGR²⁸ identifié dans la bande 1 (50 kDa).

Intérêt de la dérivation au TMPP. Il est à noter que lorsque la dérivation TMPP n'est pas réalisée sur cet échantillon, il ne nous a pas été possible d'identifier le peptide tryptique N-terminal ²³AVTVGR²⁸ de six acides aminés dans l'expérience de LC-MS/MS. Ce peptide correspondant à l'ion moléculaire chargé deux fois (m/z de 301.68) est situé à la limite basse de la gamme de balayage du spectromètre de masse, il est possible qu'il ne soit pas sélectionné en MS. L'un des avantages du marquage au TMPP est donc ainsi confirmé. Ce peptide de seulement six acides aminés a probablement un temps de rétention très faible voire, avoir été élué dans le volume mort. Le peptide N-ter dérivé a été détecté et identifié grâce à la séparation chromatographique liée à la présence du TMPP.

Détection d'autres sites marqués au TMPP. Quatre peptides dérivés au TMPP, et correspondant à des clivages après la position 173, très minoritaires, ont été identifiés, mais avec des intensités de l'ion moléculaire fragmenté 1000 fois moins intenses (intensité 10^5) que pour l'ion moléculaire du peptide TMPP-²³AVTVGR²⁸. Ceci n'est pas surprenant puis l'analyse du produit de départ en LC-MS a mis en évidence la présence de contaminants mineurs (voir Figure 52). Ceci illustre également la capacité d'une dynamique élevée de cette méthodologie. Parallèlement, cette capacité à détecter et identifier des traces de peptides marqués au TMPP, rappelle l'importance qu'il y a à toujours associer une information quantitative, même grossière, pour chaque identification de peptide marqué.

Analyse de la bande 2 (contrôle avec inactivation des protéases du pathogène).

Lorsque les protéases du pathogène sont désactivées, aucun produit de dégradation n'est observé sur le gel 1D-PAGE. La forme Psh intacte, semble donc stable dans le temps, dans les conditions d'incubation.

L'analyse nanoLC-MS/MS de la bande 2 à 50 kDa ($t=240$ minutes) a permis d'identifier le peptide attendu ²³AVTVGR²⁸ dérivé au TMPP en position N-ter. La forme Psh intacte n'est pas dégradée et l'extrémité N-ter protéique est la même que celle identifiée dans le contrôle à $t=0$ minute. Cette étude permet de vérifier que

les produits de dégradation visibles lors de l'incubation de r-Psh avec le pathogène sont liés à l'activité des protéases du pathogène.

Analyse de la bande 4.

L'analyse nanoLC-MS/MS de la bande 4 (30 kDa) a permis d'identifier clairement le peptide ¹⁴⁴IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF¹⁷¹ dérivé au TMPP en N-ter.

Le rapport du signal sur bruit mesuré pour ce peptide est intense ($2.4 \cdot 10^6$). Le temps de rétention élevé de ce peptide ($t_r=25.8$ minutes) correspond bien à celui d'un peptide portant un groupement hydrophobe TMPP (Figure 59).

Le massif isotopique du peptide est celui d'un ion multichargé (4+) caractéristique de la présence du groupement TMPP (charge permanente). La différence de masse mesurée par rapport à la masse prévue à partir de la séquence en acides aminés (16 Da) correspond à l'oxydation de la méthionine.

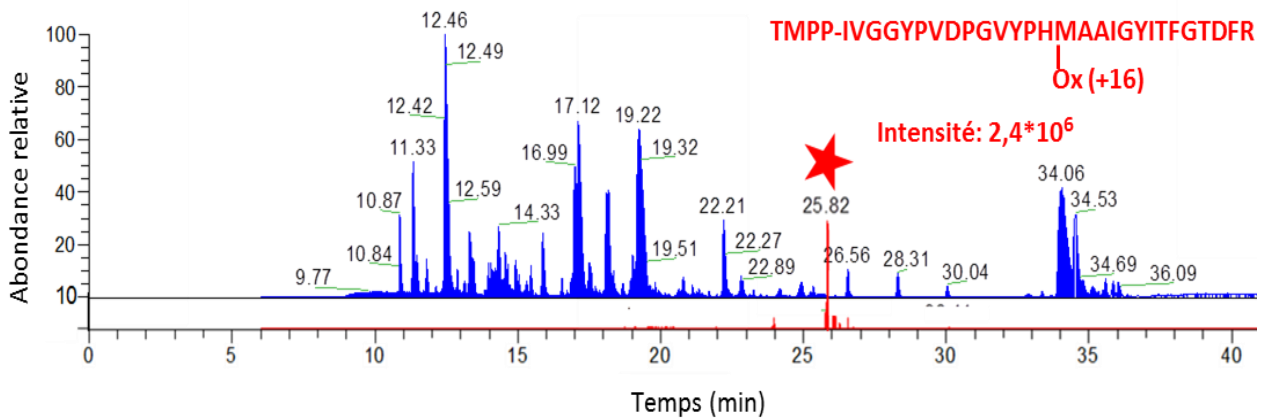


Figure 59: Chromatogramme (en bleu) des peptides tryptiques de r-Psh de la bande 4 (30 kDa). Extraction du courant d'ion m/z 901.17 (en rouge) correspondant au peptide ¹⁴⁴IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF¹⁷¹ dérivé au TMPP ($t_r=25.82$ minutes).

La Figure 60 présente le spectre de fragmentation de l'ion 4+ (m/z=901.17 pour le pic P) du peptide ¹⁴⁴IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF¹⁷¹. Ce spectre est informatif puisque, grâce au fragment « b1 », il permet de confirmer la position N-terminale TMPP, bien que les fragments b1 à b3 soient très faibles, mais significatifs (Figure 60). L'extrémité N-ter ¹⁴⁴I est générée lors de la maturation de r-Psh par les protéases bactériennes.

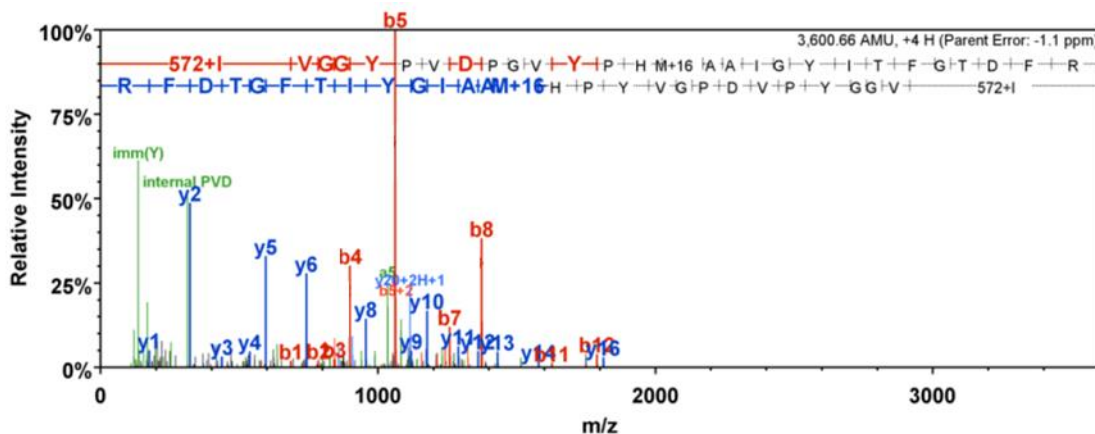


Figure 60: Sélection du spectre MS/MS assigné au peptide ¹⁴⁴TMPP-IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF¹⁷¹ dans la bande 4.

La masse moyenne calculée de la protéine avec une extrémité N-terminale en position 144 est de 30628 Da. L'hétérogénéité de l'échantillon de r-Psh liée aux glycosylations peut expliquer la présence de la forme active de r-Psh à une masse supérieure de celle attendue.

Analyse des bandes 3, 5 et 6.

Bande 3: L'analyse nanoLC-MS/MS de la bande 3 (40 kDa) a permis d'identifier un peptide $\text{TMPP}^{23}\text{AVTVGR}^{28}$, ce qui montre que la réduction de masse moléculaire est nécessairement due à un clivage proche du C-terminal de r-Psh, et à une distance de celui-ci d'environ une centaine d'acides aminés. Cette bande est peu intense et commence à disparaître au bout de 30 minutes.

Bande 5: Cette bande apparaît après la bande 4, et a une cinétique d'apparition très proche de celle de la bande 6. L'analyse nanoLC-MS/MS de la bande 5 (25 kDa) a permis d'identifier le peptide $^{144}\text{IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF}^{171}$ à partir d'un ion moléculaire très intense (intensité= $1 \cdot 10^7$, $t_r=25.4$ minutes). Ceci indique un clivage de la r-Psh en position 143-144. La protéine contenue dans la bande 5 possède donc la même extrémité N-terminale que celle de la bande 4 (30 kDa). Ce peptide a également été identifié dérivé en N-ter, sans oxydation du résidu de méthionine $^{144}\text{IVGGYPVDPGVYPHMAAIGYITFGTDF}^{171}$. Le rapport du signal sur bruit mesuré pour ce peptide non oxydé est également intense ($5.6 \cdot 10^6$; $t_r=26.3$ minutes) (Figure 61).

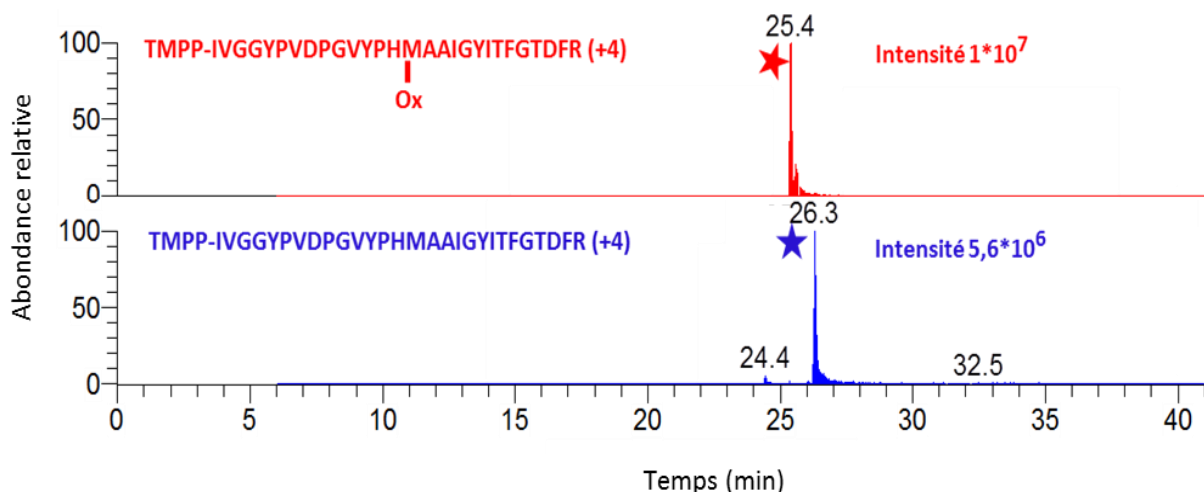


Figure 61 EIC des deux ions dérivés au TMPP de la bande 5 (25 kDa). EIC du m/z 901.17 correspondant au peptide $^{144}\text{IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF}^{171}$ ($t_r=25.4$ min) en rouge. EIC d'ion m/z 897.17 correspondant au peptide $^{144}\text{IVGGYPVDPGVYPHMAAIGYITFGTDF}^{171}$ ($t_r=26.3$ min) en bleu.

Dans une première analyse, les spectres MS/MS présentaient des ions b1 à b3 très faibles. Pour obtenir des spectres MS/MS présentant des ions de b1 à b5 aussi visibles que possible nous avons répété l'expérience LC-MS/MS en utilisant une liste d'inclusion. Cette liste comprenait le m/z des deux ions moléculaires $^{144}\text{IVGGYPVDPGVYPHM(Ox)AAIGYITFGTDF}^{171}$ (m/z 901.17, 4H+) et $^{144}\text{IVGGYPVDPGVYPHMAAIGYITFGTDF}^{171}$ (m/z 897.17, 4H+) avec un temps d'exclusion dynamique de 10 secondes. Nous avons ainsi pu obtenir des spectres MS/MS prouvant de façon non ambiguë la présence du TMPP en position N-terminale. La Figure 62 ci-dessous présente les spectres obtenus.

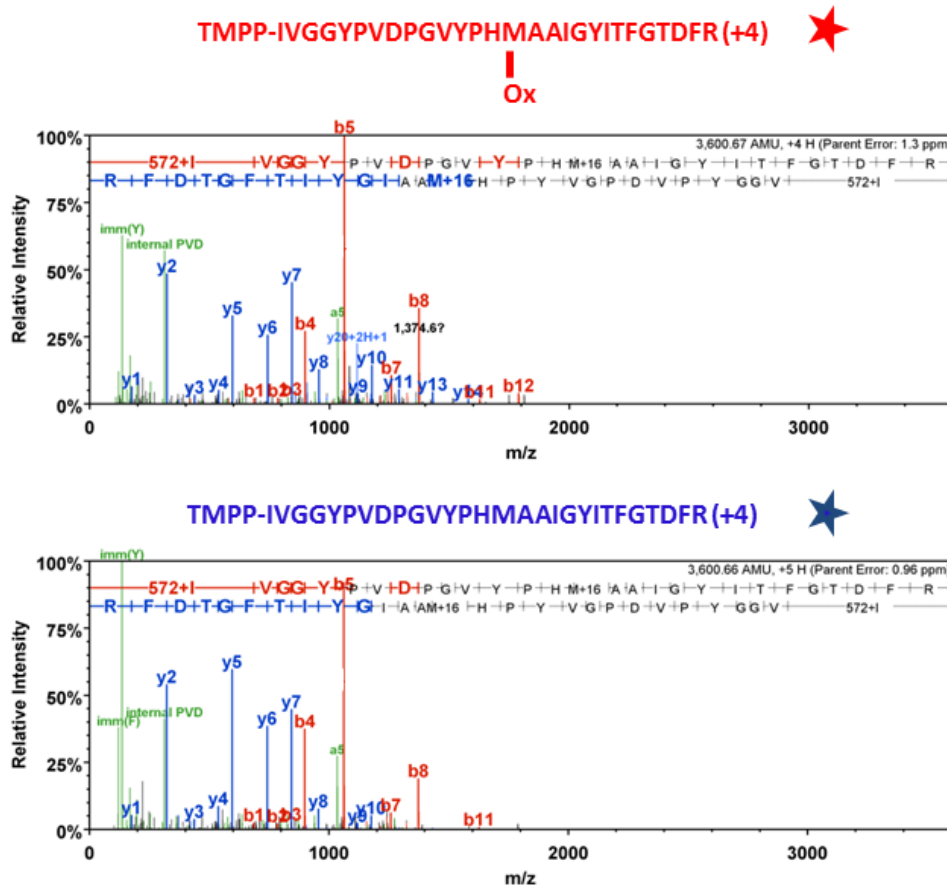


Figure 62: Sélection des spectres MS/MS des peptides dérivés $^{144}\text{IVGGYPVDPGVYPH(Ox)AAIGYITFGTDF}^{171}$ (en rouge) et $^{144}\text{IVGGYPVDPGVYPHMAAIGYITFGTDF}^{171}$ (en bleu) obtenus à l'aide d'une liste d'inclusion implémentée dans la méthode MS, de la bande 5. Les fragments b1 à b3 sont aussi visibles que sur la Figure 60.

Bande 6 : L'analyse nanoLC-MS/MS de la bande 6 (10 kDa) n'a pas permis d'identifier une extrémité N-ter dérivée au TMPP.

Conclusion des sites de clivage marqués au TMPP

Le résultat principal de cette étude est la mise en évidence, grâce à la dérivation TMPP *in gel*, d'une protéine d'environ 30 kDa correspondant à la r-Psh clivée en position 144 dans la bande 4.

Ceci confirme l'hypothèse d'une activation directe de Psh par la protéase *subtilisine-like* sécrétée par *Bacillus subtilis*. Cette étude *in vitro* a permis de mettre en évidence pour la première fois le site d'activation de Psh.

La bande 5, contient également une protéine dont le N-terminal correspond à un clivage de la r-Psh en position 143-144, mais dont la taille apparente sur le gel 1D-PAGE est inférieure d'environ 100 acides aminés par rapport à la bande 4. Ceci indique un clivage du côté C-terminal de cette protéine.

Pour la bande 6, nous n'avons pas pu mettre en évidence de peptide marqué au TMPP. Plusieurs raisons peuvent être invoquées pour expliquer cela, dont par exemple un peptide tryptique N-terminal de très petite taille.

Pour compléter ces résultats et leur conclusion, nous avons réalisé une exploitation en profondeur de toutes les données de LC-MS/MS obtenues, en particulier pour tenter de déterminer les points de clivages du côté C-terminal.

C.2.2 La caractérisation fine des produits de dégradations de r-Psh

Après avoir déterminé les extrémités N-terminales dérivées au TMPP dans les six bandes, nous avons tenté d'obtenir pour chacune des bandes une carte de recouvrement des peptides de digestion aussi complète que possible.

Les séquences protéiques caractérisées pour les bandes 1, 3, 4, 5, 6 sont présentées dans la Figure 63. La bande 2, correspondant à la même masse apparente et identifiée avec le même N-ter que la bande 1, n'a pas été étudiée ici.

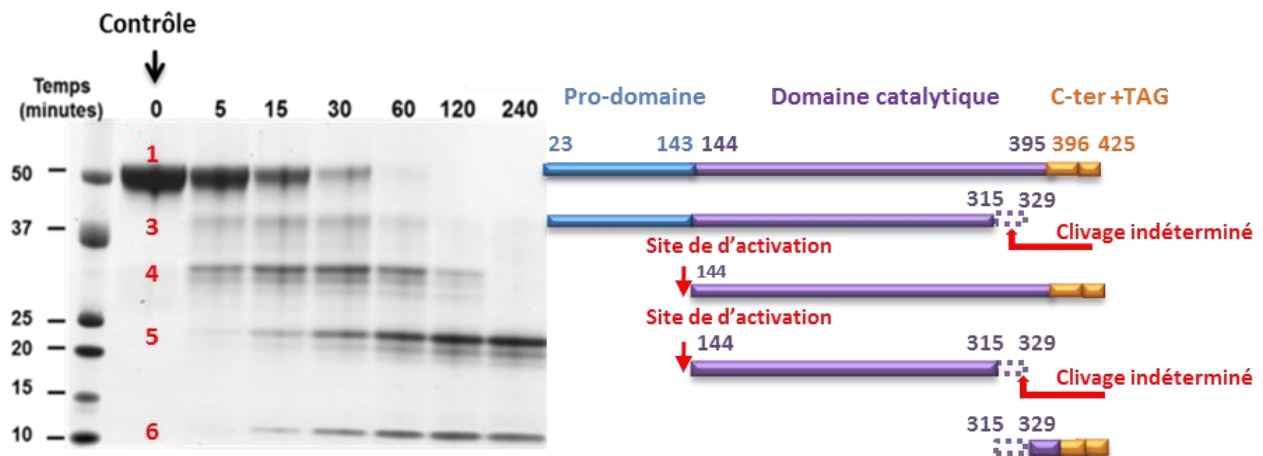


Figure 63: Recouvrement de la séquence protéique déterminée après comparaison des aires sous le pic des peptides identifiés dans les bandes de gel après digestion trypsique

Nous avons comparé les cartes de recouvrement peptidique des différentes bandes à celle de la bande de référence (bande 1) correspondant à la forme intacte de r-Psh dans l'ordre suivant :

- Bande 3 avec la bande 1
- Bande 4 avec la bande 1
- Bande 5 avec la bande 1
- Bande 6 avec la bande 1

Pour déterminer les positions éventuelles de clivage, nous avons tenté de mettre en évidence l'absence de certains peptides de digestion trypsique dans les bandes 3, 4, 5 et 6 par rapport à la bande 1 (la forme intacte de r-Psh dans le contrôle). Comme la comparaison des cartes de recouvrement de séquence protéique ne permet souvent pas de détecter des différences entre échantillons, nous avons apporté une dimension quantitative. Pour cela, nous avons utilisé des diagrammes de comparaison visuelle des intensités peptidiques normalisées pour chaque bande.

Cette approche est bien sûr rudimentaire par rapport aux méthodes de quantification du type *label free* mais demandent un investissement en temps beaucoup plus réduit et réaliste tout en permettant, comme nous le verrons, de faire ressortir des différences significatives.

Les diagrammes que nous avons utilisés reportent les aires sous les pics pour chaque ion moléculaire d'un peptide trypsique identifié (XIC). Chaque aire peptidique a été normalisée par l'aire de deux peptides indépendamment (double vérification), sélectionnés volontairement après le site de clivage ^{143}H (position 180^{ième} et 201^{ième} acide aminé).

Comparaison bande 3 / bande. 1 Les peptides tryptiques des bandes 1 (50 kDa) et 3 (40 kDa) ont été étudiés. Les peptides sont classés de gauche à droite selon l'axe des abscisses en fonction de la position dans la séquence protéique (Figure 64). Dans les bandes 1 et 3 le peptide N-terminal de la protéine dérivé au TMPP a été identifié.

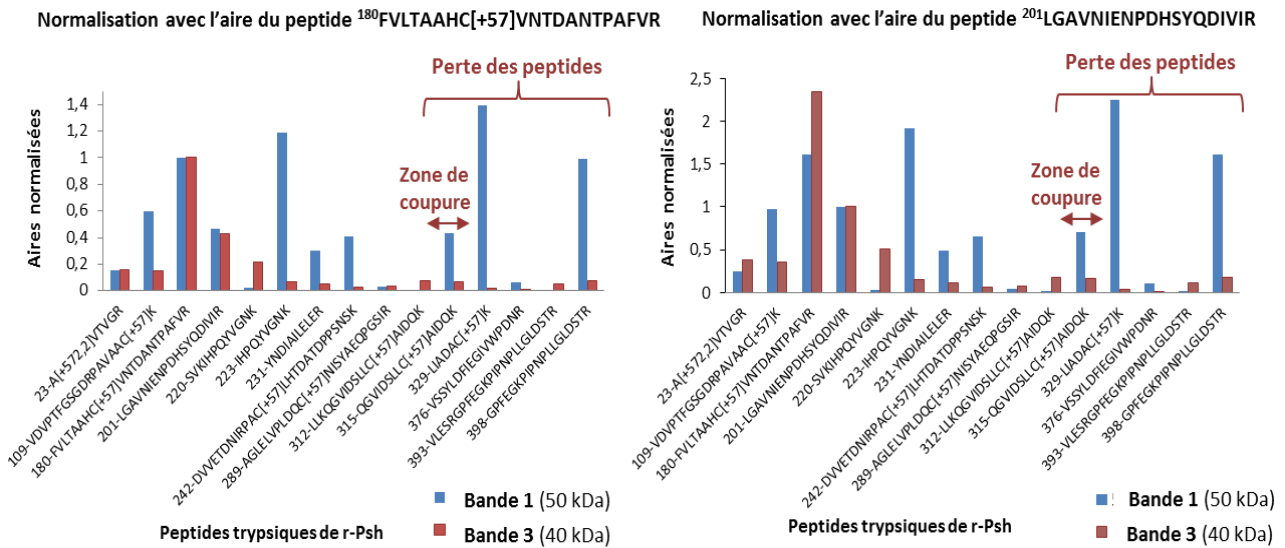


Figure 64: Histogramme de comparaison des aires normalisées des peptides tryptiques de r-Psh dans les bandes 1 (50 kDa) et 3 (40 kDa) lors de l'incubation avec *B. subtilis*.

Après normalisation il apparait que la couverture de séquence de la forme à 50 kDa (bande 1) est complète de la position en N-ter jusqu'à la position 416 (C-ter). Les aires normalisées des peptides tryptiques sont du même ordre de grandeur entre les bandes 1 et 3 de la position ²³A à la position ³¹⁵Q. La perte de la séquence entre les positions 315 à 425 correspond à la perte d'environ 10 kDa. Pour tenter d'apporter plus de précisions, nous avons recherché dans les données MS/MS tous les peptides semi-tryptiques identifiés entre les positions 315 à 329. Cependant, le site de coupure précis n'a pu être déterminé. La bande 3 correspond à la forme de r-Psh avec un clivage indéterminé du côté C-terminal.

Comparaison bande 4 / bande. 1 L'analyse nanoLC-MS/MS a permis d'identifier l'extrémité N-terminale de la forme de r-Psh en position ¹⁴⁴I dérivé au TMPP. L'histogramme présenté Figure 65 met en évidence l'absence de la partie N-ter de la séquence (diminution de l'aire) avant le site d'activation. Après le site d'activation les aires peptidiques normalisées sont du même ordre de grandeur entre les bandes 1 et 4. La bande 4 correspond bien à la forme r-Psh clivée après la position ¹⁴³H.

Normalisation avec l'aire du peptide ¹⁸⁰FVLTAAHCl+57]VNTDANTPAFVR

Normalisation avec l'aire du peptide ²⁰¹LGAVNIENPDHYSQDIVIR

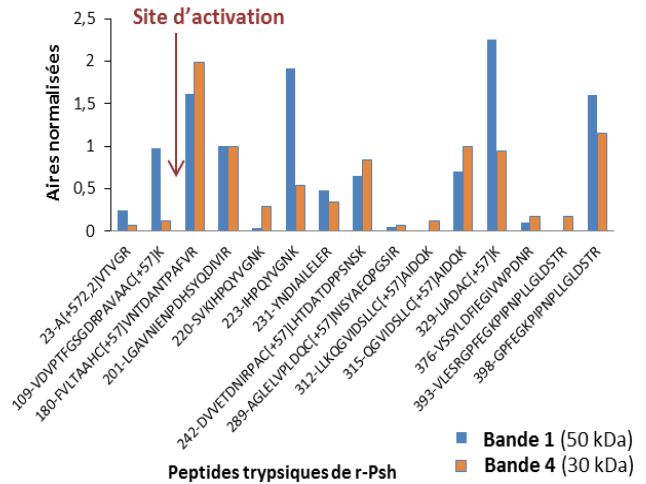
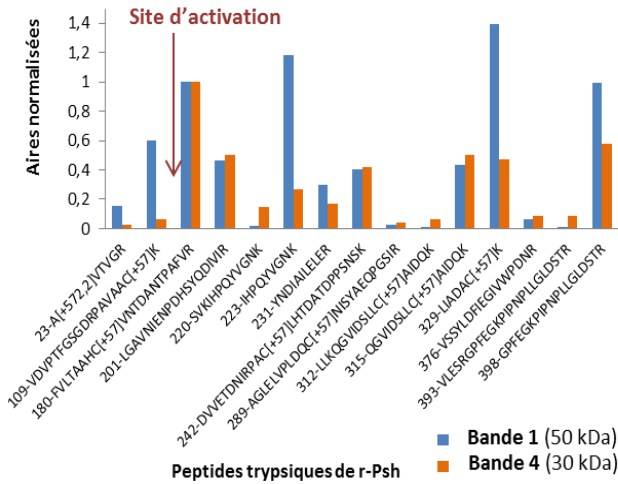


Figure 65: Histogramme de comparaison des aires normalisées des peptides tryptiques de r-Psh dans les bandes 1 (50 kDa) et 4 (30 kDa) lors de l'incubation avec *B. subtilis*.

Comparaison bande 5 / bande 1. L'analyse nanoLC-MS/MS a permis d'identifier l'extrémité N-terminale de la forme de r-Psh en position ¹⁴⁴I dérivé au TMPP. L'histogramme présenté Figure 66 met en évidence la baisse de l'intensité des peptides coté N-ter avant le site d'activation. De plus l'absence de la séquence entre les positions 315 à 425 est observée comme dans la bande 3. La bande 5 correspond à la forme de r-Psh clivée coté N-terminal au niveau du site d'activation 143-144 et clivée du côté C-terminal.

Normalisation avec l'aire du peptide ¹⁸⁰FVLTAAHCl+57]VNTDANTPAFVR

Normalisation avec l'aire du peptide ²⁰¹LGAVNIENPDHYSQDIVIR

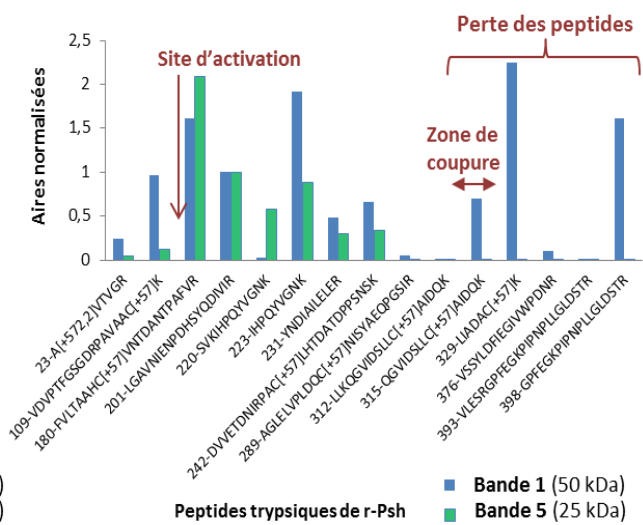
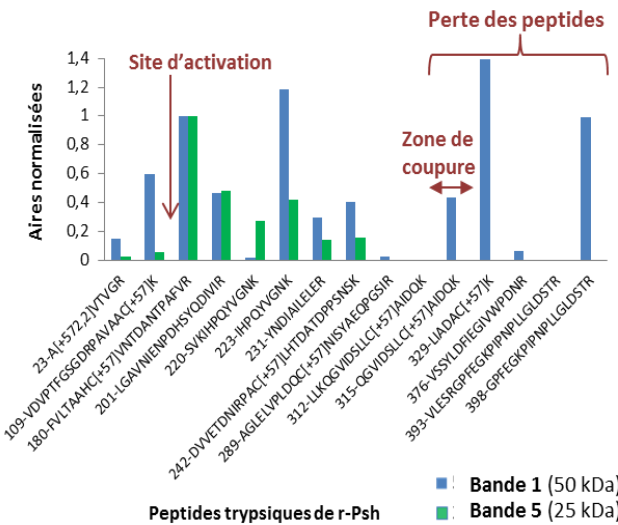


Figure 66: Histogramme de comparaison des aires normalisées des peptides tryptiques de r-Psh dans les bandes 1 (50 kDa) et 5 (25 kDa) lors de l'incubation avec *B. subtilis*.

Comparaison bande 6 / bande 1. L'intensité de la bande 6 augmente proportionnellement à celle de la bande 5. Aucun peptide N-ter n'a été identifié dans cette bande. L'hypothèse la plus logique est qu'il s'agit de la séquence peptidique après la coupure indéterminée coté C-ter dans la séquence de la protéine. Il est nécessaire d'adapter les peptides choisis pour normaliser les aires. Ainsi les peptides en positions 393 et 398 ont été choisis pour la normalisation des aires. L'histogramme présenté Figure 67 met en évidence l'absence de la séquence protéique jusqu'à la position 329 de la partie C-terminale. Il s'agit bien des peptides absents dans la bande 5.

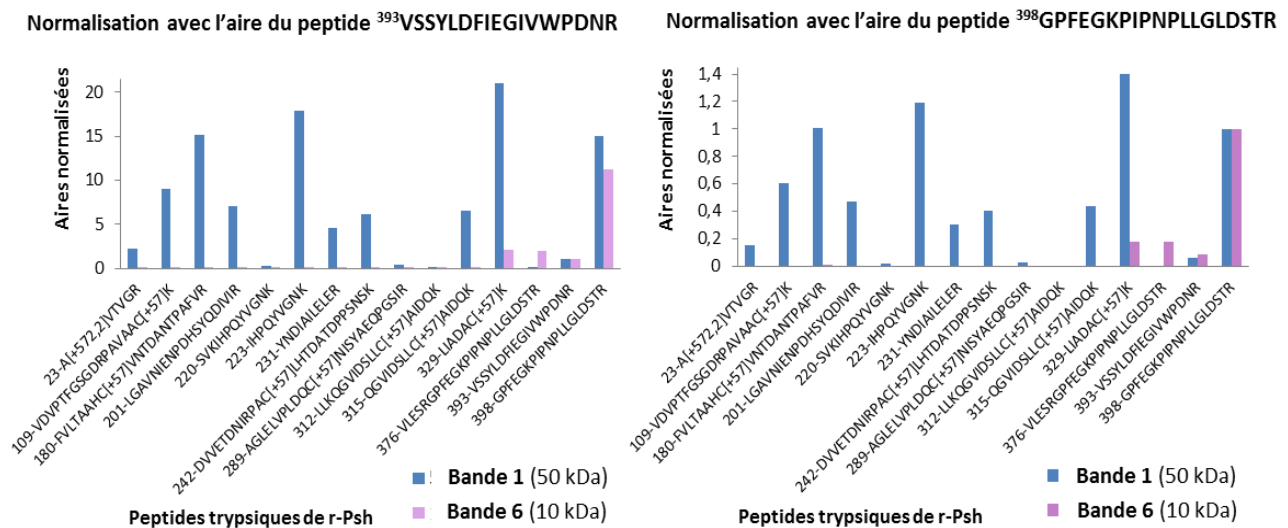


Figure 67: Histogramme de comparaison des aires normalisées des peptides tryptiques de r-Psh dans les bandes 1 (50 kDa) et 6 (10 kDa) lors de l'incubation avec *B. subtilis*.

Remarque sur la diffusion des bandes 1D-PAGE. En théorie lors de la migration d'échantillon dans un gel 1D-PAGE les protéines migrent en fonction de leurs poids moléculaires au niveau de la masse apparente correspondante. Cependant, de faibles fractions protéiques, sous forme de traces, ne migrent pas de manière homogène au reste de la fraction. Ces traces sont détectées grâce à la sensibilité des instruments de spectrométrie de masse de dernière génération. Avant l'apparition de ces nouvelles performances instrumentales, ce phénomène n'était que peu voire pas décrit. Pour exemple, le peptide N-ter de la protéine ²³AVTVGR²⁸ a été identifié dans les bandes 1,2 et 3 de manière intense ainsi que sous forme de traces dans les bandes 4, 5 et 6.

C.2.3 Bilan sur la caractérisation fine des produits protéolytiques de r-Psh par *B. subtilis*.

Après avoir identifié les extrémités N-terminales dérivées au TMPP dans les six bandes, nous avons déterminé les positions éventuelles de clivage en établissant la carte de recouvrement des peptides de digestion pour les bandes 1, 3, 4, 5 et 6 et en introduisant un aspect «quantitatif». La comparaison des aires sous la courbe pour les ions moléculaires dans chaque analyse LC-MS/MS a permis de confirmer :

- Bande 1: séquence complète de la forme intacte de r-Psh
- Bande 3: clivage au niveau C-terminale de la séquence de r-Psh entre les acides aminés 315 et 329
- Bande 4: clivage au niveau du site d'activation 143-144
- Bande 5: deux clivages, l'un au niveau du site d'activation, l'autre du côté C-terminale de r-Psh entre les acides aminés 315 et 329
- Bande 6: séquence C-terminale de r-Psh avec un N-ter situé entre les positions 315 et 329

L'ensemble des résultats obtenus est résumé dans la Figure 68.

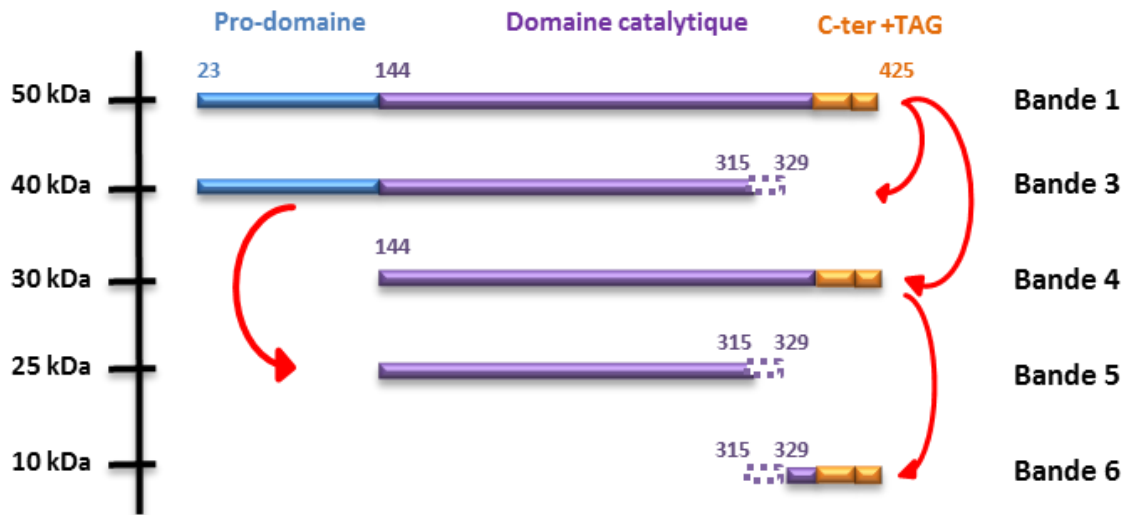


Figure 68: Représentation schématique des produits de dégradation de r-Psh détectés lors de l'analyse 1D-PAGE (Figure 55). Seule la forme à 30 kDa contient les trois acides aminés de la triade du site actif de la protéase à sérine Psh.

L'étude de la cinétique d'apparition a montré que la coupure en position ¹⁴⁴I intervient en premier donc, il n'est pas sûr que les mécanismes de dégradation observée (site de coupure coté C-ter) aient le temps d'intervenir *in vivo*.

C.3. L'étude de la protéolyse de Psh par d'autres pathogènes

Dans cette deuxième partie, les mécanismes d'activation de Psh avec d'autres pathogènes que *B. subtilis* ont été étudiés. Dans un premier temps, nos collaborateurs ont sélectionné des pathogènes capables de stimuler la voie du récepteur Toll par l'intermédiaire de Psh *in vivo*. Dans un deuxième temps, notre objectif analytique a été de déterminer les sites de clivage de r-Psh.

C.3.1 La sélection des pathogènes

L'activation de la voie du récepteur Toll chez la drosophile, par l'intermédiaire de Psh, a été testée avec un panel de pathogènes dont les protéases n'appartiennent pas à la famille des subtilisines. L'équipe du Pr Reichhart a mesuré, au bout de 16 heures le taux de transcription d'un AMP, la drosomycine, induite suite à la stimulation de la voie du récepteur Toll. A ce temps précoce, après une infection pathogénique, seule la cascade protéolytique dépendante de Psh est activée. Dans ces conditions, seuls les pathogènes testés sécrétant des protéases ont entraîné une production de drosomycine en réponse à l'infection chez les drosophiles sauvages. En revanche, cette induction n'a pas été observée chez les drosophiles mutantes nulles pour Psh. Cette approche a permis de sélectionner spécifiquement trois pathogènes activant la voie par l'intermédiaire de Psh pour tester leur capacité à activer pro-Psh (Tableau 6).

Famille du pathogène	Nom du pathogène
Bactérie à Gram+	<i>S. aureus</i>
Bactérie à Gram-	<i>P. aeruginosae</i> <i>S. marcescens</i>

Tableau 6: Présentation des pathogènes capables d'activer la voie du récepteur Toll par l'intermédiaire de Psh en fonction de la famille à laquelle ils appartiennent.

L'activation de la voie du récepteur Toll par l'intermédiaire de Psh n'est donc pas limitée aux bactéries Gram+ et aux champignons contrairement à ce qui était généralement décrit¹⁴¹. Les résultats de l'étude menée par nos collaborateurs ont permis de sélectionner trois pathogènes. Ces pathogènes sont connus pour utiliser leurs protéases comme facteurs de virulence. Les protéolyses de r-Psh avec chacun de ces pathogènes (Gram+ et Gram-) ont été étudiées pour répondre à l'objectif fixé sur des dupliquas biologiques.

C.3.2 La détermination des sites de clivage des formes maturées de r-Psh

C.3.2.1 Les bactéries à Gram+

La protéolyse de r-Psh par *Staphylococcus aureus*

La protéolyse de r-Psh par la bactérie *S. aureus* a été suivie par gel 1D-PAGE. Le gel obtenu est présenté Figure 69 après révélation au bleu colloïdale. Aucune bande à 30 kDa ne semble correspondre à un clivage en position 143-144. Dans les contrôles, les bandes correspondant aux masses apparentes de 50 et 40 kDa ont déjà été étudiées avec *B. subtilis*, elles correspondent respectivement à la forme intacte de r-Psh et à la forme dégradée coté C-ter. Une dégradation aux alentours de 37-35 kDa, dans la bande 1 est visible dans le contrôle. L'intensité des bandes de gel à 37 kDa (bande 2) et 35 kDa (bande 3) a augmenté au cours du temps d'incubation (Figure 69). Seuls les produits de dégradation déjà observés dans le contrôle à t=0 (bande 1) sont observés dans le contrôle à 240 minutes. Les bandes 1, 2 et 3 sélectionnées (Figure 69) ont été extrudées et soumises à la stratégie de dérivation développée.

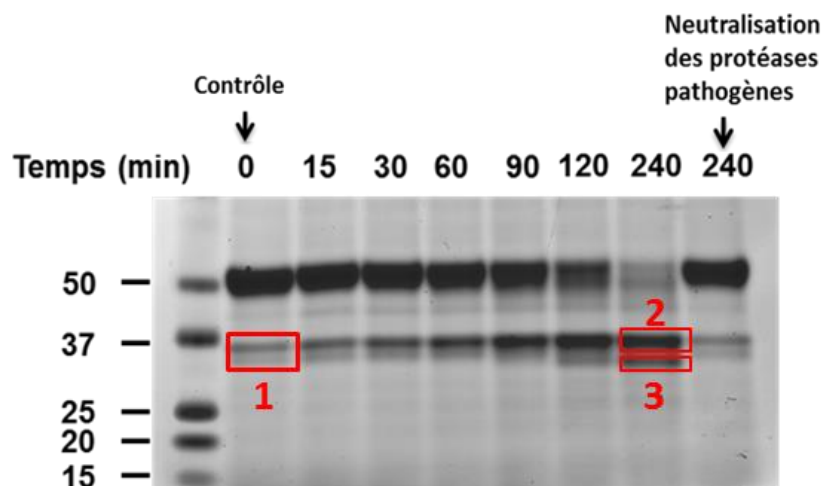


Figure 69: Gel d'électrophorèse SDS PAGE 1D de la protéolyse de r-Psh par *S. aureus* en fonction du temps. Les bandes de gel sélectionnées pour être soumises à la dérivation TMPP sont entourées en rouge

Ce paragraphe est focalisé exclusivement sur la recherche des sites marqués au TMPP dans les trois bandes extrudées. Le tableau ci-dessous résume les résultats d'analyses nanoLC-MS/MS obtenus.

Numéro de bande	Acide aminé marqué au TMPP	Séquence de peptide N-terminal
1	²³ A	²³ AVTVGR ²⁸
2 et 3	¹¹⁶ S ¹¹⁸ D	¹¹⁶ SGDRPAVAACK ¹²⁶ ¹¹⁸ DRPAVAACK ¹²⁶

Tableau 7: Résultats des identifications des peptides N-terminaux dérivés au TMPP pour les trois bandes

Ces résultats d'identification des acides aminés marqués au TMPP des analyses MS sont décrits en détail ci-dessous:

- Analyse de la bande 1 qui correspond à une dégradation de Psh dans le contrôle
- Analyse des bandes 2 et 3 qui correspondent aux masses apparentes de 37 à 35 kDa, non attendues pour un clivage en position 143-144

Analyse de la bande 1. L'analyse nanoLC MS/MS de la bande 1 (50 kDa) a permis d'identifier le peptide N-ter de la forme intacte ²³AVTVGR²⁸ dérivés au TMPP (intensité $1,4 \cdot 10^7$, tr=26.2 minutes). Aucune nouvelle extrémité n'a été identifiée. La perte de 10-15 kDa est probablement liée à des dégradations de la séquence protéique de r-Psh du côté C-ter. L'objectif étant de déterminer les sites de clivage potentiels aucune couvertures de séquence n'a été établie dans cette expérience.

Analyse des bandes 2 et 3. L'analyse nanoLC-MS/MS de la bande 2 (37 kDa) a permis d'identifier deux nouvelles extrémités dérivées en N-ter en position 116 et 118. Le rapport du signal sur bruit mesuré pour les deux peptides est de $3,4 \cdot 10^7$ (tr=21.3 minutes) pour ¹¹⁶SGDRPAVAACK¹²⁶ et de $2,5 \cdot 10^8$ (tr=21.3 minutes) pour ¹¹⁸DRPAVAACK¹²⁶. Les peptides ont été validés à l'aide de leurs spectres de fragmentation MS/MS.

Les masses moyennes théoriques de la protéine r-Psh avec des extrémités N-terminales en position ¹¹⁶S et ¹¹⁸D sont de 33756 et 33612 Da respectivement.

L'analyse nanoLC MS/MS de la bande 3 (35 kDa) a permis d'identifier les deux extrémités dérivées en N-ter en positions 116 et 118. Le rapport du signal sur bruit mesuré pour les deux peptides est de $2,6 \cdot 10^8$ (tr=21.3 minutes) pour ¹¹⁶SGDRPAVAACK¹²⁶ et de $8,3 \cdot 10^7$ (tr=21.3 minutes) pour ¹¹⁸DRPAVAACK¹²⁶.

Conclusion de la protéolyse de r-Psh par *S. aureus*

Le résultat principal de l'action des protéases de la bactérie *S. aureus*, est la mise en évidence de deux formes protéiques aux environs de 37-35 kDa correspondant à des formes de r-Psh clivées spécifiquement aux positions ¹¹⁶S et ¹¹⁸D, amont du site d'activation (143-144).

Bien que ce pathogène active la synthèse de la drosomycine par l'intermédiaire de Psh, aucune forme de r-Psh activée au niveau du site d'activation n'a pu être détectée. Les protéases de la bactérie *S. aureus* ne seraient donc pas capables de cliver pro-Psh *in vitro* au niveau du site d'activation en position ¹⁴³H.

C.3.2.2 Les bactéries à Gram-

La protéolyse de r-Psh par *Serratia marcescens*

La protéolyse de r-Psh par la bactérie *S. marcescens* a été suivie par gel 1D-PAGE. Le gel obtenu est présenté Figure 70 après révélation au bleu colloïdale. Aucune bande à 30 kDa ne semble correspondre à un clivage en position 143-144. Dans le contrôle à t = 0, la bande correspondant à la masse apparente à 50 kDa et 40 kDa ont déjà été étudiée avec *B. subtilis*. Il s'agit de la forme intacte et d'une forme dégradée coté C-terminale de la séquence de r-Psh. Deux produits de protéolyse sont visibles à 37-35 et 25 kDa et correspondent

respectivement aux bandes 1 et 2 sélectionnées (Figure 70). Ces deux bandes ont été extrudées et soumises à la stratégie de dérivation TMPP.

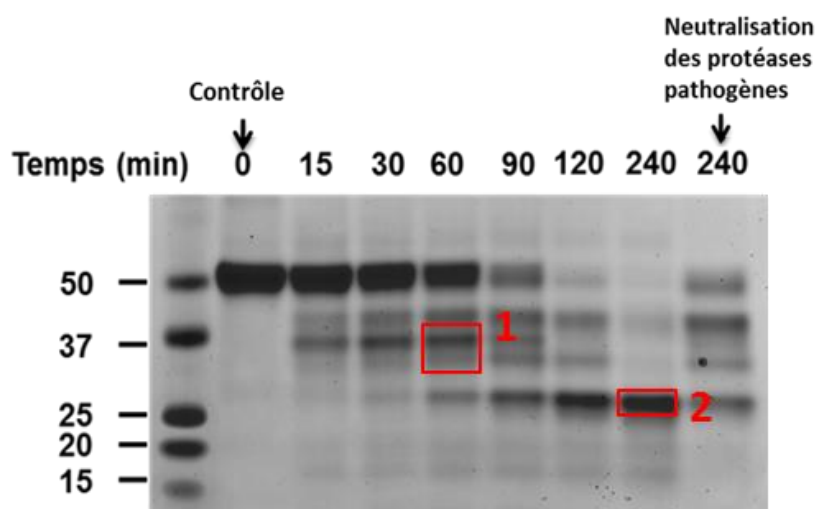


Figure 70: Gel d'électrophorèse SDS PAGE 1D de la protéolyse de r-Psh par *S. marcescens* en fonction du temps. Les bandes de gel sélectionnées pour être soumises à la dérivation TMPP sont entourées en rouge.

Ce paragraphe est focalisé exclusivement sur la recherche des sites marqués au TMPP dans les deux bandes extrudées. Le tableau ci-dessous résume les résultats d'analyses obtenus.

Numéro de la bande	Acide aminé marqué au TMPP	Séquence du peptide N-terminal
1 et 2	¹¹¹ V	¹¹¹ VPTFGSGDRPAVAAC(+57)K ¹²⁶

Tableau 8: Résultats des identifications des peptides N-terminaux dérivés au TMPP pour les deux bandes

Ces résultats d'identification des acides aminés marqués au TMPP sont décrits en détail ci-dessous :

Analyse de la bande 1 qui correspond aux masses de 37-35 kDa

Analyse de la bande 2 qui correspond à la masse de 25 kDa

Analyse de la bande 1. L'analyse nanoLC MS/MS de la bande 1 (37-35 kDa) a permis d'identifier le peptide ¹¹¹VPTFGSGDRPAVAAC(+57)K¹²⁶ dérivé en N-ter au TMPP. Le rapport du signal sur bruit mesuré est de $1.2 \cdot 10^7$ (tr=25.4 minutes). Le peptide a été validé à l'aide de son spectre de fragmentation MS/MS. La masse moyenne théorique de la forme processée de r-Psh avec une extrémité N-terminale en position ¹¹¹V est de 34258 Da.

Analyse de la bande 2. L'analyse nanoLC-MS/MS de la bande 2 (25 kDa) a permis d'identifier le peptide ¹¹¹VPTFGSGDRPAVAAC(+57)K¹²⁶ dérivé en N-ter au TMPP (intensité= $2.8 \cdot 10^8$, tr=25.6 minutes). La dégradation côté C-ter de la protéine explique la présence de cette forme de r-Psh (N-ter en position 111) à la masse apparente de 25 kDa.

Conclusion de la protéolyse de r-Psh par *S. marcescens*

Le résultat principal de l'action des protéases de la bactérie *S. marcescens*, est la mise en évidence d'une forme protéique aux environs de 37-35 kDa correspondant à une forme clivée spécifiquement en position ¹¹¹V, en amont du site d'activation (143-144).

Aucune forme de r-Psh clivée au niveau du site d'activation n'a pas été détectée. Les protéases de la bactérie *S. marcescens* ne seraient donc pas capables d'activer Perséphone *in vitro* au niveau du site prédit d'activation après l'histidine en position ¹⁴³H.

La protéolyse de r-Psh par Pseudomonas aeruginosa

La protéolyse de r-Psh par la bactérie *P. aeruginosa* a été suivie par gel 1D-PAGE. Le gel obtenu est présenté Figure 71 après révélation au bleu de colloïdale. Aucune bande à 30 kDa ne semble correspondre à un clivage en position 143-144. La masse apparente à 50 kDa a déjà été étudiée avec *B. subtilis*, il s'agit de la forme intacte de r-Psh. Des produits de dégradation sont visibles à 37,35 et 25 kDa et correspondent respectivement aux bandes 1, 2 et 3. Ces trois bandes ont été extrudées et soumises à la stratégie de dérivation TMPP.

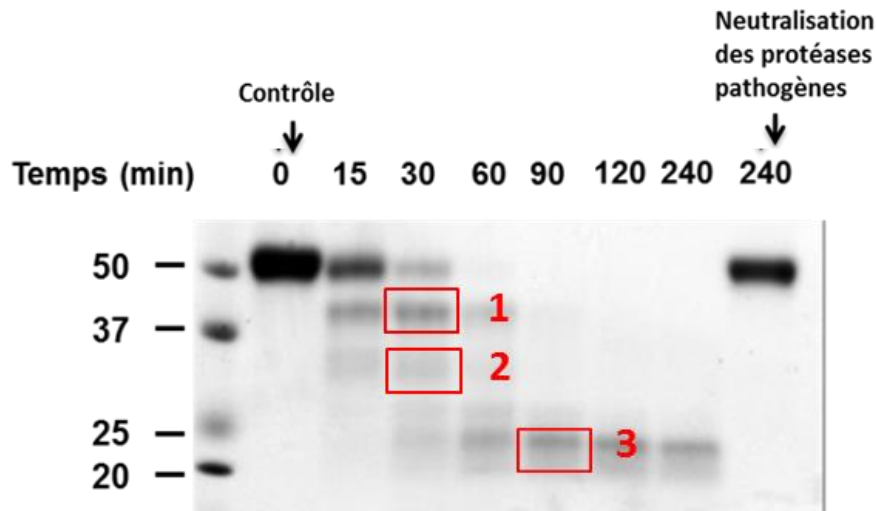


Figure 71: Gel d'électrophorèse SDS PAGE 1D de la protéolyse de r-Psh par *P. aeruginosa* en fonction du temps. Les bandes de gel sélectionnées pour être soumises à la dérivation TMPP sont entourées en rouge.

Ce paragraphe est focalisé exclusivement sur la recherche des sites marqués au TMPP dans les trois bandes extrudées. Le tableau ci-dessous résume les résultats d'analyses nanoLC-MS/MS obtenus.

Numéro de la bande	Acide aminé marqué au TMPP	Séquence du peptide N-terminal
1	¹⁰⁴ M ¹¹⁴ F	¹⁰⁴ MTSGR ¹⁰⁸ ¹¹⁴ FGSGDRPAVAAC[+57]KK ¹²⁶
2	n.d	n.d
3	¹⁰⁴ M ¹¹⁴ F	¹⁰⁴ MTSGR ¹⁰⁸ ¹¹⁴ FGSGDRPAVAAC[+57]KK ¹²⁶

Tableau 9 : Résultats des identifications des peptides N-terminaux dérivés au TMPP pour les trois bandes

Ces résultats d'identification des acides aminés marqués au TMPP sont décrits en détail ci-dessous:

- Analyse de la bande 1 qui correspond à la masse de 37 kDa, non attendue pour un clivage en position 143-144
- Analyse de la bande 2 qui correspond à la masse de 35 kDa, non attendue pour un clivage en position 143-144
- Analyse de la bande 3 qui correspond à la masse de 25 kDa non attendue pour un clivage en position 143-144

Analyse de la bande 1. L'analyse nanoLC MS/MS a permis d'identifier deux peptides ¹⁰⁴MTSGR¹⁰⁸ et ¹¹⁴FGSGDRPAVAAC[+57]KK¹²⁶ dérivés en N-ter au TMPP. Les rapports du signal sur bruit mesurés pour les peptides sont de $1.8 \cdot 10^6$ pour ¹⁰⁴MTSGR¹⁰⁸ (tr=23.5 minutes) et $1.1 \cdot 10^7$ pour ¹¹⁴FGSGDRPAVAAC[+57]KK¹²⁶ (tr=24.2 minutes). Les masses moyennes théoriques de la protéine r-Psh avec des extrémités N-terminales en position ¹⁰⁴M et ¹¹⁴F sont de 35004 et 33960 Da respectivement.

Analyse de la bande 2. L'analyse nanoLC MS/MS n'a permis d'identifier d'extrémité dérivée en N-ter au TMPP. La faible intensité de la bande est corrélée avec la faible quantité de matériel protéique et nous n'avons donc pas pu déterminer une extrémité N-ter dérivée.

Analyse de la bande 3. L'analyse nanoLC MS/MS a permis d'identifier les deux peptides $^{104}\text{MTSGR}^{108}$ (Intensité= $8.9 \cdot 10^5$, tr=23.5 minutes) et $^{114}\text{FGSGDRPAVAAC}^{+57}\text{KK}^{126}$ (Intensité= $4.4 \cdot 10^7$, tr=24.1 minutes) dérivés. Une dégradation côté C-ter de la protéine (décrite **Erreur ! Source du renvoi introuvable.**) explique la présence de ces formes de r-Psh à 25 kDa.

Conclusion de la protéolyse de r-Psh par *P. aeruginosa*

Le résultat principal, de l'action des protéases de la bactérie *P. aeruginosa*, est la mise en évidence d'une forme protéique aux environs de 37 kDa correspondant à des formes clivées spécifiquement aux positions ^{104}V et ^{114}F , en amont du site d'activation (143-144).

Aucune forme de r-Psh clivée au niveau du site d'activation n'a été détectée. Les protéases de la bactérie *P. aeruginosa* ne seraient donc pas capables d'activer Perséphone *in vitro* au niveau du site prédit d'activation après l'histidine en position ^{143}H .

D. Conclusion biologique de l'étude

D.1. Les résultats concernant le site d'activation 143-144

La voie du récepteur Toll est activée par une protéase pathogène type *subtilisine-like*, par clivage protéolytique précis de Psh $_{142}$ qui libère l'extrémité N-terminale du domaine catalytique. Une fois libérés, les résidus de la séquence hautement conservée « $^{144}\text{IVGG}$ », participent à l'organisation structurale du site actif et permettent d'obtenir la conformation active de l'enzyme 127 .

Activation de r-Psh par *B. subtilis*. Le résultat principal de l'étude est la mise en évidence d'un clivage au niveau du site d'activation après la position ^{143}H d'une forme de r-Psh aux alentours de 30 kDa comme attendu. Les résultats de protéomique semblent confirmer l'hypothèse d'une activation directe de la Psh par les protéases de type subtilisines de *Bacillus subtilis* (Gram+). C'est la première fois que le site d'activation de Psh est mis en évidence *in vitro*.

Ces résultats d'identification ont été confortés par une étude plus fine de caractérisation des produits de dégradation de r-Psh qui a consisté à établir des cartes de recouvrement des bandes de gel 1D ont été établies. En effet les diagrammes de recouvrement de séquence mettent en évidence le clivage coté N-terminal niveau du site d'activation de la forme à 30 kDa. De façon intéressante, cette étude a permis de déterminer un point de clivage non attendu coté C-terminal de la protéine.

Activation de r-Psh par d'autres pathogènes. Le résultat principal de cette étude est que la forme de r-Psh clivée après la position 143 et correspondant à la masse calculée de 30 kDa n'est pas visible sur le gel 1D. Cette forme n'a pas été identifiée après dérivation au TMPP des produits de dégradation de r-Psh dans les trois protéolyses suivies. Les protéases sécrétées par les pathogènes *S. aureus*, *S. marcescens*, *P. aeruginosae* contrairement à *B. subtilis* ne semblent pas activer Perséphone *in vitro* au niveau du site prédit d'activation après l'histidine en position ^{143}H . En conclusion, la sécrétion de protéases de type subtilisines semblent être requise pour activer directement Psh au niveau du site d'activation.

D.2. Les résultats inattendus concernant d'autres sites de clivage

Les expériences menées par l'équipe du Pr Reichhart ont mis en évidence que les pathogènes sélectionnés du type bactérie à Gram+ et Gram- activent la voie du récepteur Toll par l'intermédiaire de Psh. Les résultats de spectrométrie de masse ont permis d'identifier des sites de clivage situés dans une zone spécifique entre les positions 104 à 118, commune aux protéolyses réalisées par les trois pathogènes autres que *B. subtilis* (Figure 72).

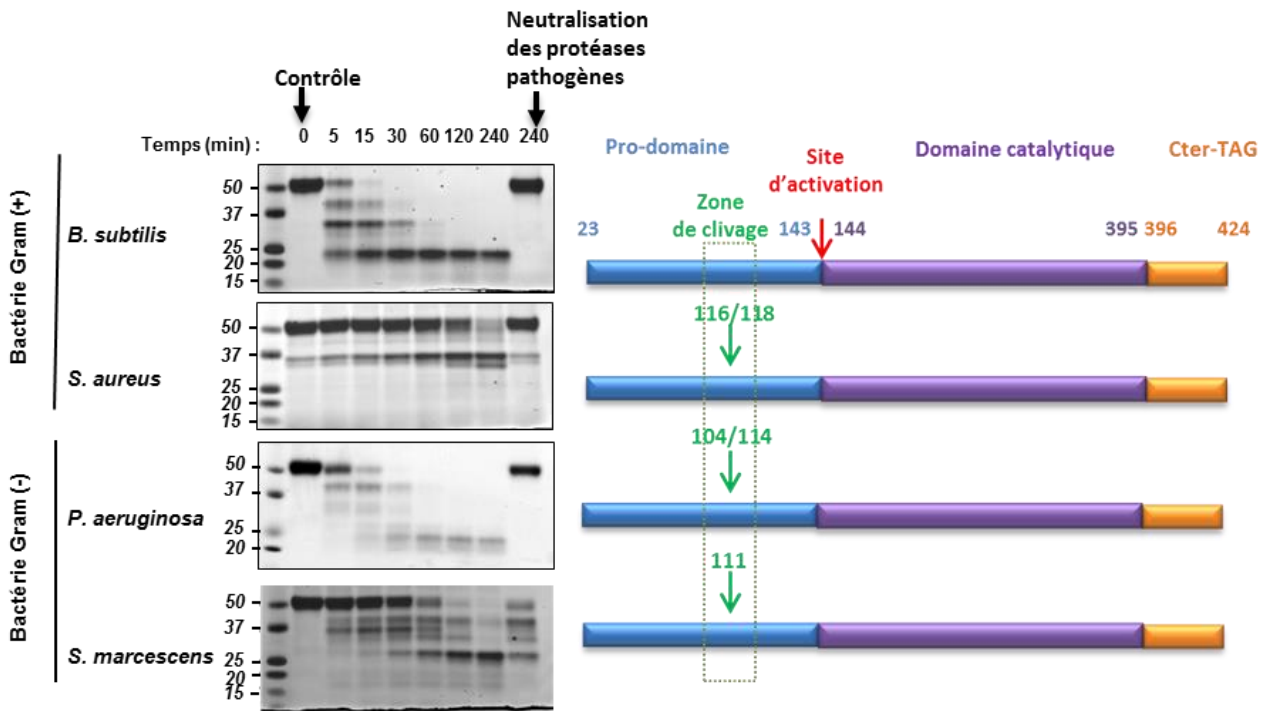


Figure 72: Schéma représentatif des gels 1D-PAGE du suivi cinétique des protéolyses de r-Psh par les quatre pathogènes étudiés (à gauche). Les extrémités N-terminales, identifiées dérivées au TMPP sont indiquées par des flèches rouges (à droite) au niveau du site d'activation, vertes pour les autres.

A partir de ces résultats, l'existence d'un double mécanisme d'activation a pu être proposée par nos collaborateurs biologistes.

Le mécanisme d'activation de Psh par clivage au niveau du site d'activation serait un mécanisme séquentiel. Ce dernier mécanisme interviendrait en deux étapes. Dans un premier temps, Psh serait clivée de manière spécifique dans la région 104-118 hautement sensible à la protéolyse en amont du site du site d'activation «¹⁴⁴IVGG». Dans un deuxième temps, une enzyme spécifique de la drosophile cliverait au niveau du site «¹⁴⁴IVGG». L'étude de construits génétiques de drosophiles adaptés a permis à l'équipe du Pr Reichhart de mettre en évidence l'intervention d'une protéase à cystéine dans la seconde étape d'activation séquentielle. Le site d'activation serait donc «protégé» de l'action de la protéase à cystéine endogène de la drosophile tant que les pathogènes n'auraient pas dégradé une grande partie du pro-domaine de Psh.

Grâce à la stratégie de marquage *in gel* développée, nos résultats ont donc permis aux biologistes d'élucider un mécanisme de régulation de la voie Toll particulièrement fin. Psh apparaît alors comme le senseur de l'activité protéolytique pathogène, indépendamment de la nature, structure ou spécificité des protéases. Bien sûr, ces études menées *in vitro* demanderaient aussi à être menées *in vivo* sur des échantillons d'hémolymphe drosophilienne suite à une infection de pathogène.

E. Conclusion sur les aspects analytiques

Grâce à la stratégie de dérivation des extrémités N-terminales au TMPP, les sites de clivages protéolytiques de la protéase Perséphone *in vitro*, induits par les protéases de plusieurs types de pathogènes, ont pu être déterminés précisément. Ceci a nécessité l'optimisation de différentes étapes de notre méthodologie. Nous ne nous sommes pas limités à la détermination de position de clivage, mais nous avons exploité en profondeur les données de LC-MS/MS. Cette étude détaillée nous a permis de distinguer plusieurs protéoformes possédant le même site de clivage en position N-terminale mais avec des masses moléculaires nettement différentes observées sur gel 1D-PAGE. Nous avons déterminé exactement la séquence de ces protéoformes.

Il est ainsi apparu que certaines protéoformes possédaient les trois acides aminés nécessaires à une activité de sérine protéase, alors que d'autres avaient perdu l'un d'entre eux (³³⁹Ser). Ce type d'information, de façon général, est capital pour les aspects fonctionnels des protéases.

Ceci illustre d'une part les possibilités de la méthode de marquage par le TMPP, lorsqu'elle est optimisée, pour l'étude des clivages protéolytiques. D'autre part, il y a toujours un intérêt à mener une caractérisation aussi complète que possible de l'ensemble des formes protéiques générées dans ce type d'étude, et en particulier en innovant dans la façon d'exploiter les données de LC-MS/MS.

Si les résultats obtenus dans cette étude *in vitro* devaient être confirmés par une étude *in vivo*, la méthodologie que nous avons utilisée dans ce travail devrait être adaptée, principalement à cause de la complexité d'un échantillon d'hémolymphe. Pour augmenter la sensibilité de l'approche dans une étude *in vivo* et donc dans un échantillon plus complexe, des méthodes pour cibler plus particulièrement les ions moléculaires désormais connus grâce aux résultats pourront être mises en place. Une méthode d'identification ciblée telle que la PRM (*Parallel Reaction Monitoring*) permet la détection de plusieurs signaux d'un même ion peptidique de manière sensible avec une grande spécificité. Elle est une approche de choix pour abaisser les limites de détection.

Parti II Résultats

Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS

Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS

Ce projet est le fruit de collaborations étroites avec les Dr Thierry Rabilloud (UMR 5249, CEA de Grenoble) et Alain Wagner (UMR7199, Université de Strasbourg).

A. Introduction

De nombreuses méthodologies de quantification en analyse protéomique ont émergées ces dix dernières années⁴⁶. Un grand nombre d'entre elles passent par un marquage chimique. Le marquage chimique consiste à modifier chimiquement (dérivation chimique) des protéines ou des peptides à l'aide de réactifs qui réagissent spécifiquement avec les groupes fonctionnels des chaînes latérales de certains acides aminés. Ces réactifs peuvent être éventuellement eux-mêmes marqués aux isotopes stables, ce qui ouvre la porte à une diversité de stratégies de quantification.

Les réactifs de marquage sont souvent conçus pour permettre un enrichissement des protéines ou des peptides de digestion ce qui a pour but de faciliter leur séparation chromatographique, leur détection en MS et leur enrichissement. La conception du réactif de marquage et la méthodologie d'analyse constituent donc une réelle stratégie analytique. Ces stratégies sont souvent baptisées par un acronyme, par commodité. Parmi les stratégies les plus connues (qui désignent des méthodologies analytiques et des réactifs de marquage spécifiques), nous pouvons citer :

- ICAT⁵⁴ (*Isotope Coded Affinity Tag* ; dérivation des résidus spécifiques sur les protéines intactes),
- iTRAQ⁵⁵ (*isobaric Tags for Relative and Absolute Quantitation*) ou le TMT⁵⁶⁻⁵⁷ (*Tandem Mass Tags*) qui utilisent des réactifs isobariques (dérivation de résidus spécifiques sur les peptides de digestion)
- SILAC⁵³ (*Stable Isotopic Labeling using Amino Acids in Cell Culture* ; marquage métabolique par incorporation d'acides aminés marqués).

La Figure 73 présente les principales stratégies de quantification fréquemment utilisées.

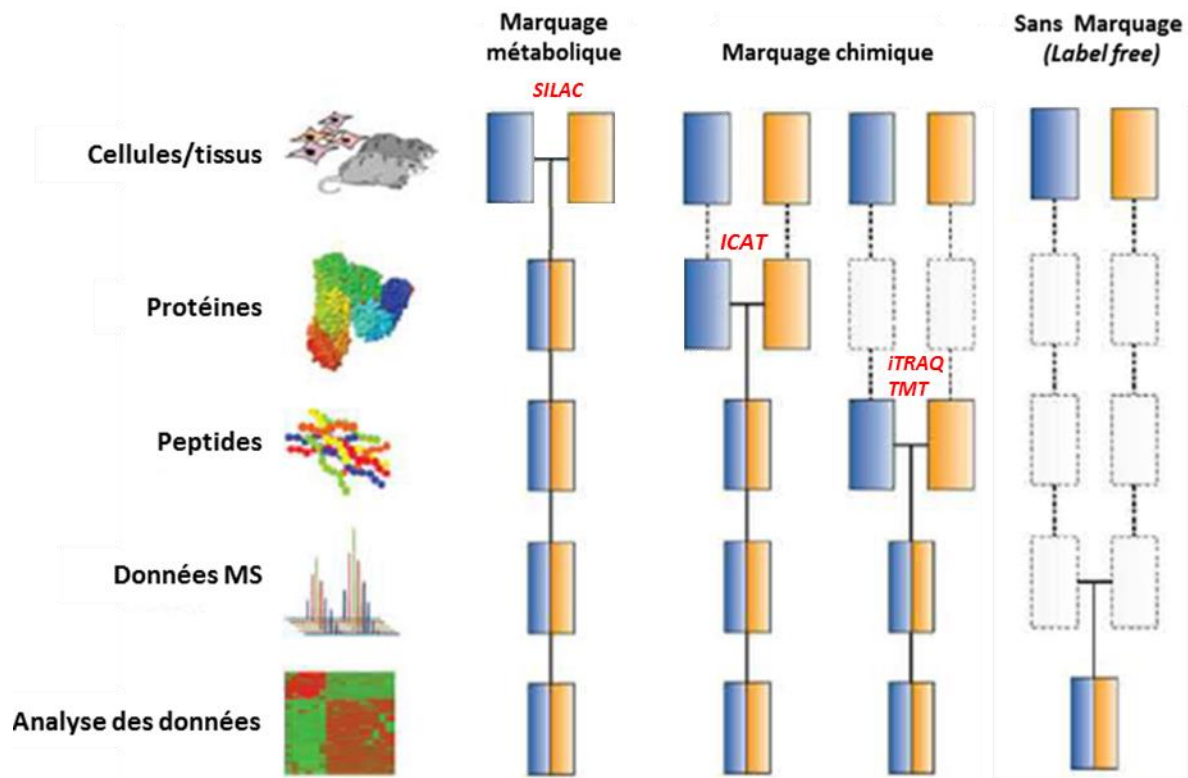


Figure 73: Représentation schématique des stratégies de quantification par spectrométrie de masse en protéomique, adaptée de Bantsheff⁴⁶. Les rectangles bleus et jaunes représentent deux conditions expérimentales à comparer. Horizontalement, sont présentées les grandes étapes de la stratégie d'analyse en protéomique. Les lignes verticales indiquent lorsque les échantillons sont mélangés. Les lignes en pointillés indiquent les points potentiels où des biais expérimentaux peuvent être induits.

La clé d'une bonne méthode: la réduction de la complexité de l'échantillon. Comme présenté ci-dessus, les stratégies de marquage sélectif de certains acides aminés ont donc pour objectif de réduire la complexité de l'échantillon constitué par les peptides de digestion de l'échantillon biologique à analyser afin de réaliser la quantification des protéines présentes. Cette réduction de complexité se fait en «isolant» les peptides de digestion marqués des peptides non marqués. Le mélange de peptides marqués sera donc moins complexe que l'échantillon de départ, ce qui rendra son analyse quantitative plus facile. Malheureusement, le fait d'avoir perdu les peptides non marqués va diminuer le nombre de protéines identifiables ou la qualité de leur identification (recouvrement de séquence). Il y aura donc eu perte d'information, puisque tous les peptides de digestion issus d'une même protéine n'auront pas été nécessairement marqués.

En fait, c'est la façon d'isoler les peptides de digestion marqués des peptides non marqués qui détermine s'il y aura, ou non, perte d'information.

Si l'isolement des peptides marqués consiste à éliminer de l'échantillon tous les peptides non marqués, (piégeage, affinité,...) seules les protéines ayant générées un ou plusieurs peptides de digestion marqués seront détectées et identifiées.

Si l'isolement est réalisé grâce à une chromatographie qui élue, par exemple, d'abord les peptides non dérivés, et ensuite les peptides dérivés, alors l'analyse apportera les données nécessaires pour la quantification, sans que les informations d'identification des protéines soient perdues.

Protéomique quantitative et protéomique exploratoire : un choix à faire. Les stratégies basées sur une réduction de complexité du mélange de peptides de digestion marqués, peuvent donc induire une perte d'information, celle portée par les peptides non marqués, et qui permet l'identification des protéines

présentes. Ces stratégies ne conviennent pas à des problématiques de protéomique exploratoire (*discovery proteomics*)⁴⁵ qui ont pour objectif de caractériser de façon aussi poussée que possible toutes les protéines présentes. Les stratégies de marquage sont donc plutôt conçues pour intervenir dans des problématiques pour lesquelles les protéomes des échantillons seraient déjà bien caractérisés. Ces stratégies sont donc plutôt adaptées à l'étude des changements importants des profils protéomiques, et ceci sur un grand nombre d'échantillons (analyses haut débit).

Choix de l'acide aminé cible. La fréquence et l'universalité du, ou des, acides aminés ciblés par le marquage chimique jouent donc un rôle essentiel dans l'établissement de la stratégie analytique.

Plus l'acide aminé portant le groupe fonctionnel ciblé par le marquage sera fréquent, plus l'analyste disposera de peptides spécifiquement détectés pour doser chaque protéine. Mais dans ce cas, le mélange de peptides marqués restera complexe, et donc l'analyse quantitative demandera plus de mise au point à cause des interférences possibles et pourra être moins précise.

En revanche si la distribution de cet acide aminé ciblé est universelle, mais peu fréquente (idéalement, toute protéine ne contiendrait qu'un seul acide aminé cible présent dans un peptide de digestion protéotypique) alors la situation serait optimale. En effet, la fraction des peptides de digestion comporterait alors un nombre réduit de peptides, donc une sous-population de peptides facile à analyser pour la quantification.

De ce point de vue, la cystéine est apparue comme le meilleur candidat et est de loin la méthode la plus utilisée pour le marquage chimique des protéines¹⁴³, en raison de sa grande réactivité, de sa faible abondance (1.37%)¹⁴⁴ et de sa distribution universelle dans les protéomes.

Un succès limité pour l'analyse quantitative par marquage chimique

Il faut bien constater que la plupart de ces méthodes de marquage chimique n'ont pas été utilisées en routine par la communauté des protéomistes, malgré un certain nombre d'études réalisées⁴⁸. En effet, la plupart des méthodes décrites dans la littérature, et parfois proposées commercialement, n'ont jamais été testées de façon poussées ni validées sur de « vrais » échantillons biologiques. Certaines en sont restées au stade de preuve de concept et ne sont pas considérées comme robustes et fiables. Une connaissance insuffisante de l'étape clef de marquage chimique est sans doute à incriminer à des réactions chimiques incomplètes, des réactions secondaires ou encore à un manque de spécificité vis-à-vis de l'acide aminé ciblé.

Dans ce contexte, nous nous sommes proposés de revisiter les stratégies de quantification par marquage chimique en analyse protéomique. Nous sommes revenus sur plusieurs aspects de ces stratégies afin de lever certains obstacles qui empêchent leur utilisation en routine.

Nous nous sommes focalisés sur les points suivants :

- Premièrement, le marquage chimique repose sur l'utilisation de réactions chimiques (chimie de maleimide, ester de N-hydroxysuccinimide ou NHS) qui ont été développées pour la bioconjugaison de protéines¹⁴³. Cependant et de façon générale, ni la sélectivité chimique, ni les réactions secondaires, ni le rendement de la réaction, ni la stabilité des produits formés¹⁴⁵ n'ont pu être vérifiés ou même évalués dans le cas d'expériences réelles de validation sur des protéomes connus.
- Deuxièmement, la conservation de l'intégrité de l'échantillon, l'hydrolyse des liaisons peptidiques à pH basique, la suppression de signal en ionisation électrospray, la conservation des fragmentations MS/MS informatives, sont autant de points critiques qui n'ont pas été optimisés.

- Troisièmement, la plupart des méthodes impliquent une dérivation au niveau peptidique (comme par exemple les approches iTRAQ ou TMT) et non au niveau protéique (donc avant l'étape de digestion enzymatique). Certes, la dérivation au niveau protéique est plus difficile à mettre en œuvre que celle au niveau peptidique mais elle a l'avantage de permettre de réaliser les mélanges d'extraits protéiques dont on veut déterminer les variations avant l'étape de digestion. Or cette étape de digestion est souvent source d'erreurs (biais expérimentaux).

Les trois points cités ci-dessus permettent de comprendre pourquoi peu de méthodes de marquage chimique sont aujourd'hui reconnues comme des méthodes de protéomiques quantitatives alternatives et couramment utilisées.

Nous pouvons donc penser que, une fois que les différents obstacles mentionnés ci-dessus auront été franchis, les stratégies basées sur le marquage chimique pourraient devenir des stratégies largement utilisées avec succès pour les études protéomiques à grande échelle. Ces stratégies devraient permettre la réalisation d'analyses protéomiques quantitatives robustes et relativement peu coûteuses. Cependant, ces stratégies seront réservées à des études quantitatives de protéomes ayant déjà fait l'objet d'une étude de protéomique exploration.

B. Introduction au contexte analytique : objectifs et stratégies analytiques mises en œuvre

Notre projet consiste donc à re-visiter les stratégies de marquage chimique (ou *tagging*) qui ciblent spécifiquement certains acides aminés, peu fréquents dans les séquences protéiques (cystéine, tryptophane et tyrosine). Ces stratégies doivent bien évidemment être compatibles avec les outils de l'analyse protéomique classique. Ceci devrait permettre de développer des approches robustes et quantitatives qui sont indispensables pour de nombreuses études biologiques, comme, par exemple, les études de toxicoprotéomique.

Dans ce contexte, nous proposons une stratégie en **deux étapes clés** pour réduire la complexité d'un échantillon dans une analyse protéomique *bottom-up* (Figure 74) et ainsi diminuer le sous-échantillonnage durant l'acquisition des spectres MS/MS et le caractère stochastique de l'analyse MS/MS en mode DDA :

- **Etape 1** : Cibler spécifiquement un acide aminé par marquage chimique, c'est-à-dire une sous-population de peptides de digestion, ce qui permet d'identifier des protéines sans l'analyse redondante de l'ensemble complet de peptides protéolytiques. Nous pensons qu'il s'agit là d'un bon compromis entre un seul peptide marqué et facilement détectable versus des dizaines de peptides non marqués mais dans un mélange très complexe. Cette fraction de peptides contenant les acides aminés ciblés va néanmoins permettre de donner une bonne représentation quantitative du protéome étudié.
- **Etape 2** : "Décomplexifier" le mélange peptidique. Notre approche est basée sur une séparation chromatographique qui sépare les deux familles de peptides de digestion, les peptides marqués et les peptides non marqués.

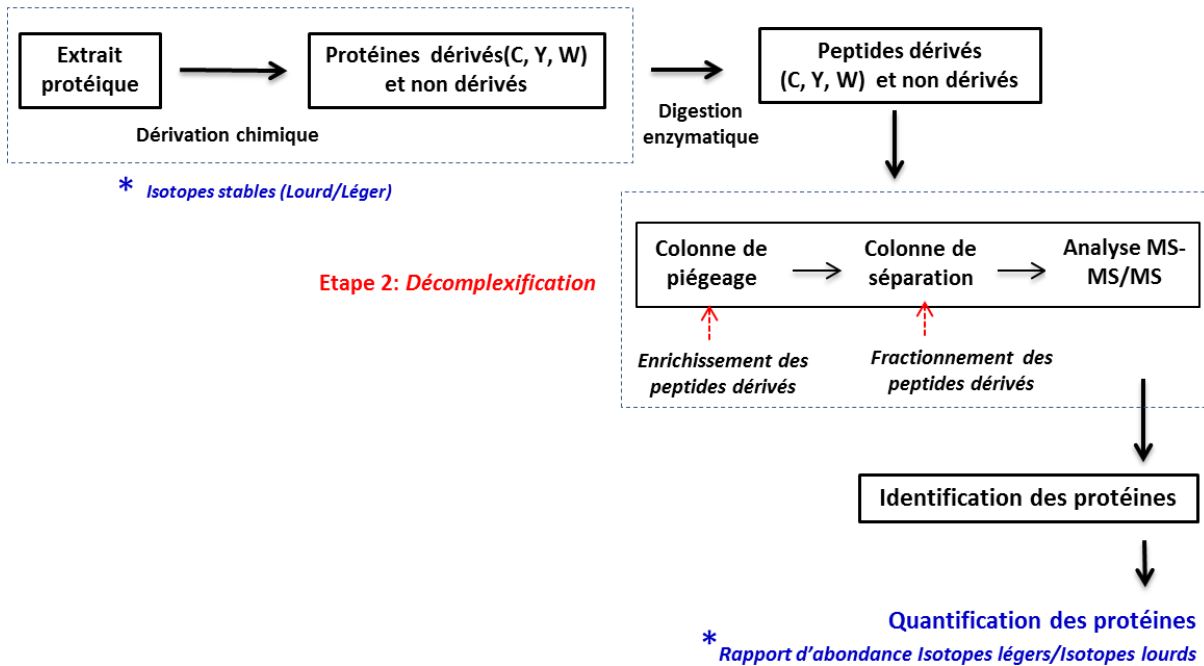
Etape 1: Marquage chimique

Figure 74: Représentation schématique de la stratégie expérimentale de l'approche de marquage chimique d'acides aminés ciblés (Etape 1). Egalement présenté dans la stratégie, une étape de réduction de la complexité du digestat peptidique (Etape 2) en fractionnement par séparation chromatographique les peptides dérivés (colonne de séparation) ou/et en enrichissant « en ligne » les peptides dérivés (colonne de piégeage). L'abondance relative des protéines de deux échantillons peut être déterminée avec l'utilisation de réactifs marqués aux isotopes stables (lourd/léger).

L'idée majeure qui sous-tend notre projet est donc de permettre, par un choix judicieux du réactif de dérivation, de différencier chromatographiquement une certaine catégorie de peptides (les peptides marqués) qui fourniront ainsi plus facilement des données analytiques (spectres MS/MS de qualité).

Cette simplification du mélange peptidique est envisagée à deux niveaux.

Premièrement, en utilisant un réactif chimique qui va augmenter le caractère hydrophobe des peptides dérivés et ainsi augmenter leur temps de rétention en chromatographie liquide afin de les élué dans une zone moins « riche » du chromatogramme (après l'ensemble des peptides non dérivés). Ceci correspond à la philosophie de l'approche de marquage au TMPP que le laboratoire a déjà développé avec succès pour la caractérisation des extrémités N-terminales¹³⁷. Ce marquage chimique permettra de simplifier le mélange peptidique et donc de limiter une éventuelle compétition entre les peptides marqués et les non marqués. Le fait de différencier chromatographiquement les peptides marqués des peptides non marqués aura plusieurs effets bénéfiques pour l'acquisition des données MS/MS dans les expériences de LC-MS/MS. Au niveau de la séparation chromatographique, puisque les peptides marqués seront élués sans interférence avec le reste du digestat peptidique non dérivé et les probabilités d'obtenir un spectre MS/MS de qualité seront ainsi augmentées. Au niveau de l'ionisation des peptides co-élués, une diminution de la compétition pour l'ionisation électrospray (ESI) sera également escomptée.

Deuxièmement, nous nous proposons d'explorer une seconde approche pour simplifier le mélange peptidique et ceci en rajoutant une étape d'enrichissement en amont de la séparation chromatographique. Cette étape d'enrichissement devra pouvoir être réalisée « en ligne » afin de limiter les étapes en amont de l'analyse LC-MS/MS. Cette étape supplémentaire devrait donc également permettre de diminuer la compétition pour l'ionisation entre les peptides dans la source du spectromètre masse mais aussi lors de la chromatographie ce qui devrait augmenter la qualité et la reproductibilité des mesures.

Nous nous sommes focalisés sur le marquage chimique de trois acides aminés: la cystéine, la tyrosine et le tryptophane et avons exploré plusieurs natures de réactifs chimiques.

Pour chacun des réactifs chimiques testés, les conditions expérimentales ont été mises au point sur une protéine modèle, en gardant à l'esprit que l'objectif principal est de développer une méthode de marquage chimique qui soit totalement compatible avec un protocole classique d'analyse protéomique pour l'étude d'échantillons biologiques.

La protéine modèle utilisée pour ces études est l'ovotransferrine, de poulet (*Gallus gallus*), aussi appelée « conalbumine », choisie en raison de son poids moléculaire relativement élevé de 75 kDa et de sa composition variée en acides aminés. La séquence de la conalbumine est composée de 30 cystéines, 21 tyrosines et de 10 tryptophanes (Figure 75).

WCTISSPEEKKCNNLRDLTQQERISLTCVQKATYLDCKIAIANNEADAISLDGGQAFEAGLAPY
 KCLKPIAAEVYEHTEGSTTSYYAVAVVKKGTEFTVNDLQGKTSCHTGLGRSAGWNIPIGTLHHRG
 AIEWEGIESGSVEQAVAKFFSASCVPGATIEQKLCRQCKGDPKTKCARNAPYSGYSGAFHCLK
 DGKGDVAFVKHTTVNENAPDQKDEYELLCLDGSRQPVDNYKTCNWARVAHAHVVARDDN
 KVEDIWSFLSKAQSDFGVDTKSDFHFGPPGKKDPVLKDLLFKDSAIMLKRVPSLMDSQLYLG
 FEYYSAIQSMRKDQLTSPRENRIQWCAVKGDEKSKCDRWSVVSNGDVECTVVDETKDCIIKI
 MKGEADAVALDGGGLVYTAGVCGLVPVMAERYDDESEQCSKTDERPASYFAVAVARKDSNVN
 WNNLKGKKSCHTAVGRTAGWVIPMGLIHNRTGT CNFDEYFSEGCAPGSPNSRLCQLCQGS
 GGIPPEKCVASSHEKYFGYTGALRCLVEKGDVAFIQHSTVEENTGGKKNKADWAKNLQMDDF
 ELLCTDGRANVMDYRECNLAEVPTHAVVVRPEKANKIRDLLERQEKRFVNGSEKSKFMM
 FESQNKDLLFKDLTKCLFKVREGTTYKEFLGDKFYTVISSLKT CNPSDILQMCSFLEGK

Figure 75: Séquence en acide aminé de la conalbumine de poulet. Les 30 cystéines sont indiquées en rouge, les 21 tyrosines en bleu et les 10 tryptophanes en vert.

Dans une première partie, nous présenterons les études menées sur la cystéine, principalement avec une série de réactifs chimiques qui avaient déjà été appliqués avec succès au laboratoire pour dériver les cystéines au niveau peptidique¹⁴⁶.

Une seconde partie sera consacrée à une étude extensive d'un réactif de marquage du tryptophane (NPBS) décrit dans la littérature par l'équipe de K. Tanaka¹⁴⁷. Les conditions expérimentales ont été mises au point sur la protéine modèle puis ont ensuite été validées sur un protéome plus complexe (*E. coli*), représentatif d'un protéome classique, afin d'évaluer la robustesse et la reproductibilité de la méthode développée.

Une troisième partie présentera très succinctement les résultats préliminaires obtenus sur deux nouveaux réactifs de marquage de la tyrosine.

C. Le marquage des cystéines

C.1. L'état de l'art du marquage chimique des cystéines

Plus de 90% des protéines humaines contiennent des cystéines et la digestion *in silico* du protéome humain montre que 15% de tous les peptides tryptiques contenant au moins une cystéine dans leur séquence seraient détectables par LC-MS/MS¹⁴⁸. Cette observation combinée avec la présence de groupement thiol sur leur chaîne latérale, font de la cystéine une cible attractive pour le marquage comme le dénote la littérature^{143, 149}. La cystéine est le seul acide aminé qui porte un groupement thiol (-SH), qui

dépasse largement, particulièrement sa forme déprotonée thiolate (-S⁻), la réactivité de toute autre fonction nucléophile présente dans les protéines. Par conséquent, la dérivation chimiosélective des cystéines est de loin la méthode la plus utilisée pour le marquage chimique des protéines.

Parmi les nombreuses méthodes de modification chimique des cystéines décrites dans la littérature, les réactifs tels que les maléimides N-substitués, les 4-vinylpyridines et les iodoacetamides (IAA) sont parmi les plus utilisés¹⁴³. Tous ces réactifs présentent des inconvénients qui sont essentiellement des réactions secondaires (dans le cas de l'IAA et des maléimides) et l'instabilité des produits formés dans des "environnements" biologiques dû aux réactions réversibles des fonctions thiols et autres réactions secondaires¹⁴⁵. De ce fait, le marquage thiol-spécifique des cystéines est une cible très valorisante et un challenge en synthèse organique.

Historiquement la première stratégie de marquage chimique des protéines développée a été l'approche ICAT (Isotope Affinity Tag)⁵⁴. ICAT est un réactif composé d'iodoacétamide, lié à une biotine à travers un bras espaceur, qui va réagir sélectivement avec les fonctions thiols des cystéines. Le marquage au niveau protéique, avant la digestion enzymatique, permet de réduire la variabilité entre échantillon en évitant l'introduction de biais expérimentaux. Une étape d'enrichissement des peptides contenant des cystéines dérivées est réalisée grâce à la biotine, par une chromatographie d'affinité (Streptavidine). La fraction enrichie en peptides dérivés est ensuite analysée par LC-MS/MS.

Initialement, l'utilisation du réactif ICAT marqué isotopiquement (¹H/²H) a été développée pour déterminer l'abondance relative des protéines entre deux échantillons en comparant le rapport d'abondance des aires des peptides ICAT (lourd/léger). Les peptides marqués isotopiquement possèdent ainsi la même efficacité d'ionisation. Cependant les peptides ¹H et les peptides ²H ne sont pas coélus en phase inverse. Cela a donc conduit au développement d'un marquage isotopique à l'aide des isotopes ¹³C et ¹⁵N qui n'induisent pas de séparation chromatographique entre les peptides marqués aux isotopes lourds/légers. Le nombre de peptides à cystéine limite donc le nombre de peptides identifiables et quantifiables, c'est une des principales limitations de cette technique. C'est pourquoi le marquage des lysines a été développé avec les approches iTRAQ ou TMT⁵⁶⁻⁵⁷. Elle est basée sur la dérivation spécifique par le N-hydroxy-succinimide (NHS) des fonctions amines libres des lysines, cinq fois plus abondantes que les cystéines. L'utilisation de réactifs marqués aux isobares (même masse) permet un multiplexage d'échantillons (jusqu'à 10 comparaisons possibles tout en permettant une coélution). Les lysines sont dérivées au niveau peptidique puis les digestats des échantillons sont mélangés avant l'analyse LC-MS/MS. Du fait de la coélution des peptides isobares pour chaque échantillon, la complexité peptidique n'est pas augmentée lors de la séparation par chromatographie. Lors de la fragmentation les peptides isobares génèrent un ion rapporteur de bas poids moléculaire caractéristique de chaque réactif isobarique. L'abondance relative des protéines entre les échantillons peut alors être déterminée en comparant le rapport d'abondance des aires peptidiques. Le coût de ces études en multiplexage et la dérivation chimique après la digestion enzymatique (introduction de biais) sont les limites de ce type d'approche¹⁵⁰.

En 2014, l'iodoTMT¹⁵¹, nommé ainsi pour sa fonction iodoacétamide utilisée pour dériver spécifiquement les sites nitrosylés des cystéines au niveau protéique a été décrit (Figure 76). Le marquage chimique par l'iodoTMT suivi d'une étape d'enrichissement par immunopurification (Ac anti-TMT) permet d'identifier les sites et de quantifier les protéines S-nitrosylées dans six échantillons simultanément.

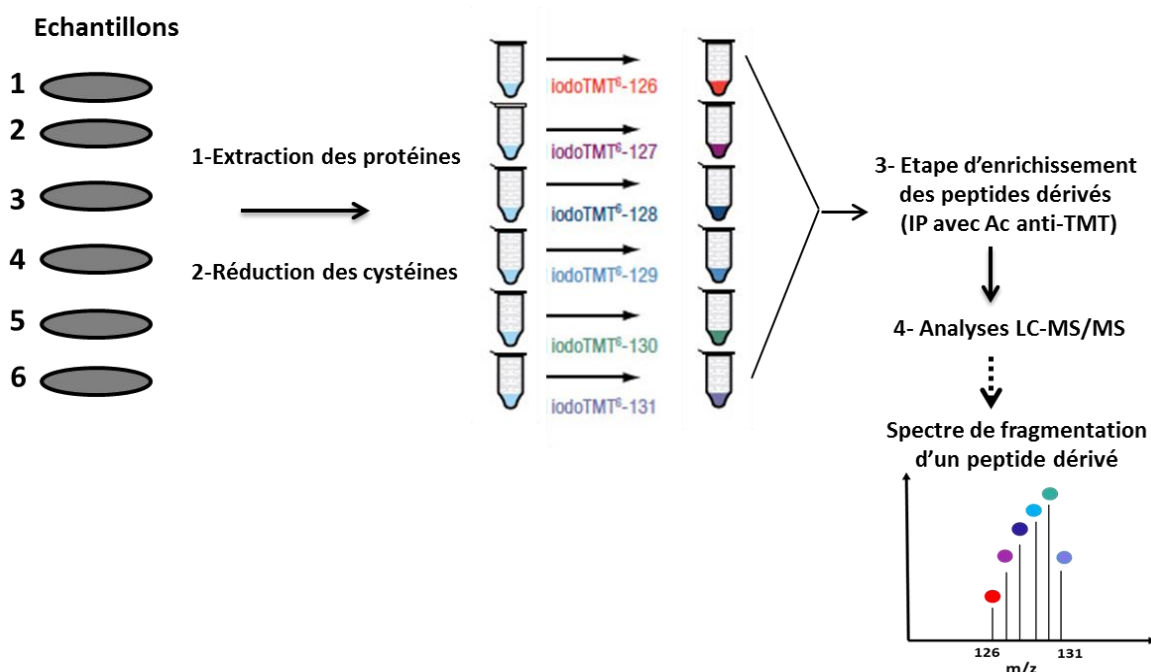


Figure 76: Représentation schématique de la stratégie de marquage par le réactif iodoTMT marqués aux isotopes stables, adaptée de Qu et al.¹⁵¹

Bien qu'il existe un certain nombre de stratégie de dérivation chimique de la cystéine dans la littérature, elles ne sont que peu utilisées en routine dans les laboratoires. Nous avons voulu lever les verrous qui empêchent leur utilisation et nous nous sommes intéressés à de nouvelles dérivations chimiques de la cystéine avec de nouveaux outils de bioconjugaison.

C.2. Résultats et discussion

C.2.1 Les dérivés APN-TMPP

Dans un premier temps, nous avons testé une série de réactifs de marquage qui porte un groupement de liaison (3-arylpropionitrile, APN) capable de réagir spécifiquement avec les résidus cystéines, lié à un groupement fonctionnel portant une charge permanente (tris(2,4,6-triméthoxyphényl) phosphonium, TMPP) qui permet d'améliorer la détection en MS des peptides à cystéines.

Ces APN-TMPP avaient déjà été appliqués avec succès au laboratoire pour dériver les cystéines au niveau peptidique¹⁴⁶. Cette réaction d'addition d'alcyne déficient en électron¹⁵² avec les fonctions thiols des cystéines est présentée Figure 77.

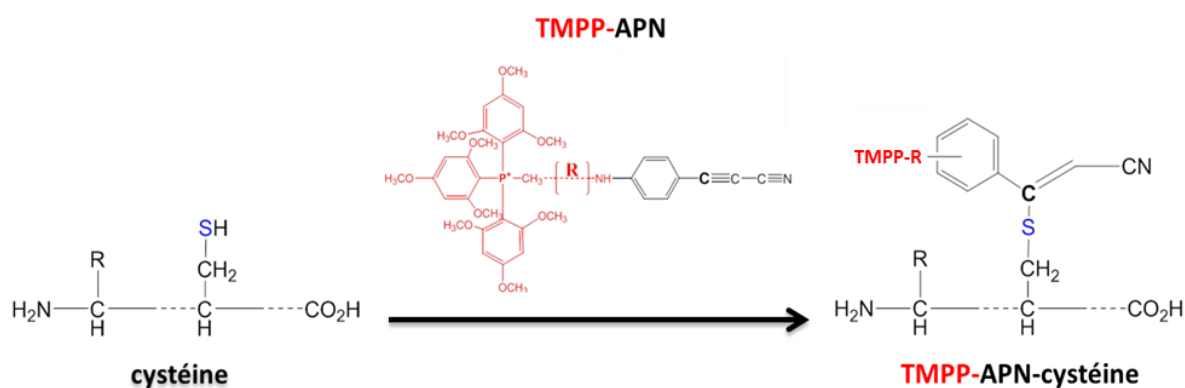


Figure 77: Représentation schématique de la réaction de marquage chimique entre la fonction thiol de la cystéine et la triple liaison carbone du réactif de marquage APN-TMPP.

Plusieurs dérivés APN-TMPP ont été synthétisés par les chimistes en modifiant la structure du groupement R dans le but d'améliorer la solubilité des réactifs en milieu aqueux.

Ces dérivés APN-TMPP présentent l'avantage d'être hydrophobes et doivent donc permettre d'augmenter les temps de rétention en chromatographie des peptides dérivés afin de les éluer dans une zone moins «riche» du chromatogramme (après l'ensemble des peptides non dérivés). Ils pourront également être synthétisés sous forme lourde (marqué aux ¹³C) pour le développement d'une stratégie de quantification (marquage de la partie TMPP).

Nous avons donc développé une approche de marquage chimique des cystéines au niveau protéique et une stratégie analytique, présentées (Figure 78). Premièrement, les cystéines réduites de la protéine ont été dérivées par l'APN-TMPP. Deuxièmement, les cystéines qui n'ont pas réagi avec l'APN-TMPP ont été bloquées chimiquement par alkylation (avec de l'iodoacétamide). Un contrôle, sans l'étape de dérivation chimique par l'APN-TMPP, a été réalisé en parallèle.

Les résultats présentés ci-dessous sont ceux obtenus avec les conditions expérimentales optimisées. La conalbumine a été solubilisée dans un tampon urée à pH 8. Après ajout d'une solution d'APN-TMPP à 6 μ M (60°C pendant 30 minutes), les échantillons ont été soumis à la stratégie analytique présentée Figure 78.

Dans le but d'évaluer la stratégie de marquage, les paramètres tels que la chimiospécificité, la fragmentation, l'efficacité de marquage, le facteur d'ionisation et le comportement chromatographique des peptides dérivés ont été étudiés. Nous présenterons ci-dessous les résultats obtenus avec le réactif APN-TMPP dont la structure est présentée Figure 79.

Résultats des études menées sur le réactif APN-TMPP

La chimiospécificité. Les peptides tryptiques identifiés dans le digestat de la conalbumine ont permis de déterminer un recouvrement de séquence de plus de 80% dans le marquage APN-TMPP et dans le contrôle. Au total 25 cystéines ont été identifiées dérivées par l'APN-TMPP dans le marquage, et 25 dérivées par alkylation dans le contrôle. Aucune réaction secondaire n'a été observée sur les chaînes latérales des lysines (amine primaire) ou les tyrosines (phénol). Le marquage des cystéines par l'APN-TMPP permet d'identifier autant de peptides à cystéines que dans le contrôle.

La fragmentation MS/MS. La charge permanente apportée par le TMPP conduit à une fragmentation MS/MS atypique, tel que décrit dans nos travaux sur le marquage des extrémités N-terminales, les algorithmes de recherche utilisés en analyse protéomique leur confèrent indûment des scores souvent faibles, mais qui ne sont pas directement corrélés à une mauvaise « qualité » de spectres MS/MS. Il était donc nécessaire de vérifier manuellement l'interprétation de ces spectres au cours de l'étape de validation.

La vérification manuelle des spectres de MS/MS des peptides APN-TMPP a mis en évidence que les scores assignés n'étaient pas impactés par la fragmentation caractéristique (fragments « b et a » générés).

A titre d'exemple le spectre de fragmentation du peptide tryptique CVPGATIEQK, Figure 80, montre une absence du fragment b1 et/ou a1 peut être expliquée par l'ajout d'un delta de masse importante apportée par le réactif (+756 Da) coté N-terminale du peptide.

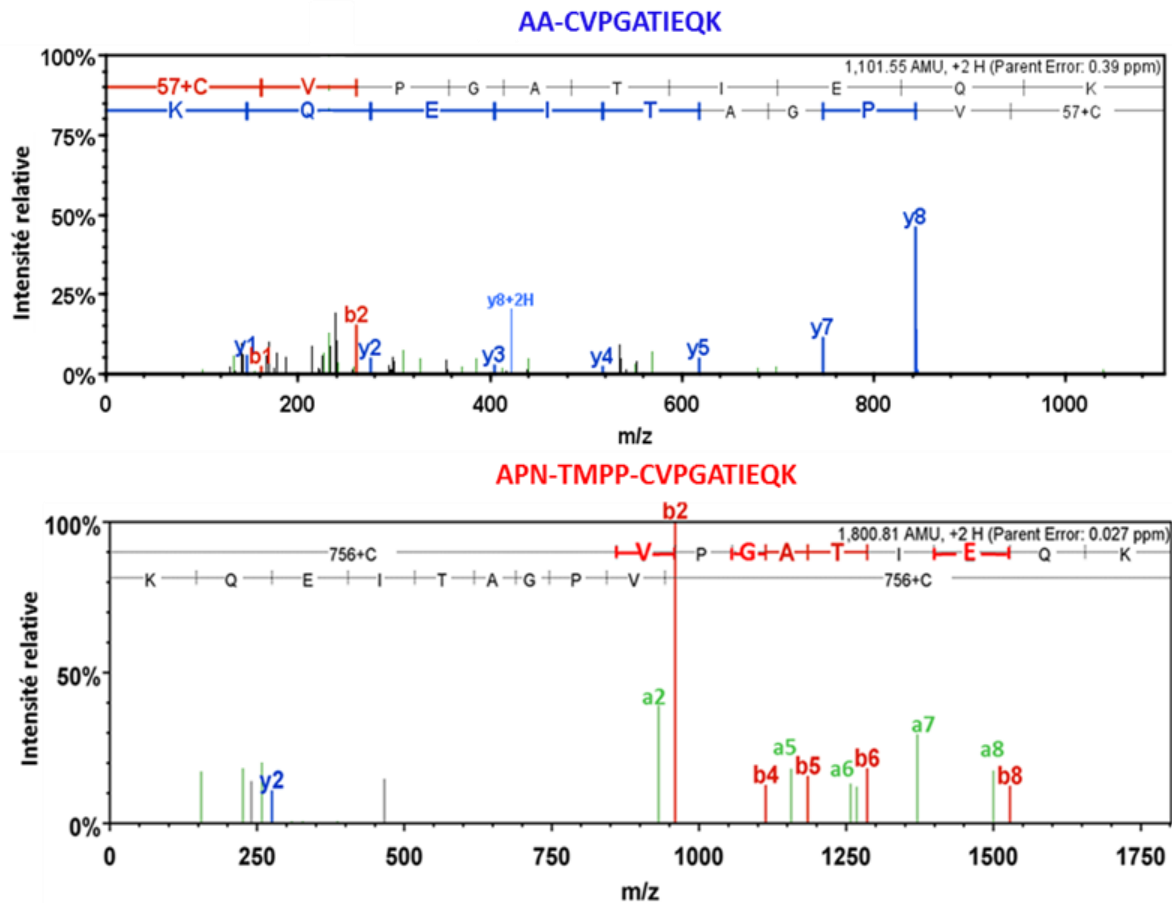


Figure 80: Comparaison des spectres de fragmentation MS/MS du peptide tryptique de la conalbumine AA CVPGATIEQK dans le contrôle et du peptide APN-TMPP-CVPGATIEQK dans l'analyse de marquage.

L'efficacité du marquage. Les résultats obtenus ont montré que la réaction de dérivation a été réalisée avec succès sur la protéine modèle. L'extraction des courants d'ion des peptides non dérivés avant et après marquage par l'APN-TMPP a permis d'observer que la réaction est presque totale avec un rendement proche de 90%.

L'efficacité de marquage a été mise en évidence Figure 81, par comparaison des courants d'ion des peptides AA-CVPGATIEQK avant (intensité $3 \cdot 10^8$) et après marquage (intensité $2 \cdot 10^7$).

Le facteur de réponse. Les résultats ont également montré une augmentation du facteur de réponse des peptides après marquage à l'APN-TMPP. L'extraction des courants d'ion des peptides après marquage a mis en évidence un facteur de réponse cinq fois plus élevé par rapport aux peptides avant marquage.

L'augmentation du facteur de réponse a été mise en évidence Figure 81, par comparaison des courants d'ion des peptides dérivés APN-TMPP-CVPGATIEQK (intensité $1.6 \cdot 10^9$) et AA-CVPGATIEQK dans le contrôle (intensité $3 \cdot 10^8$).

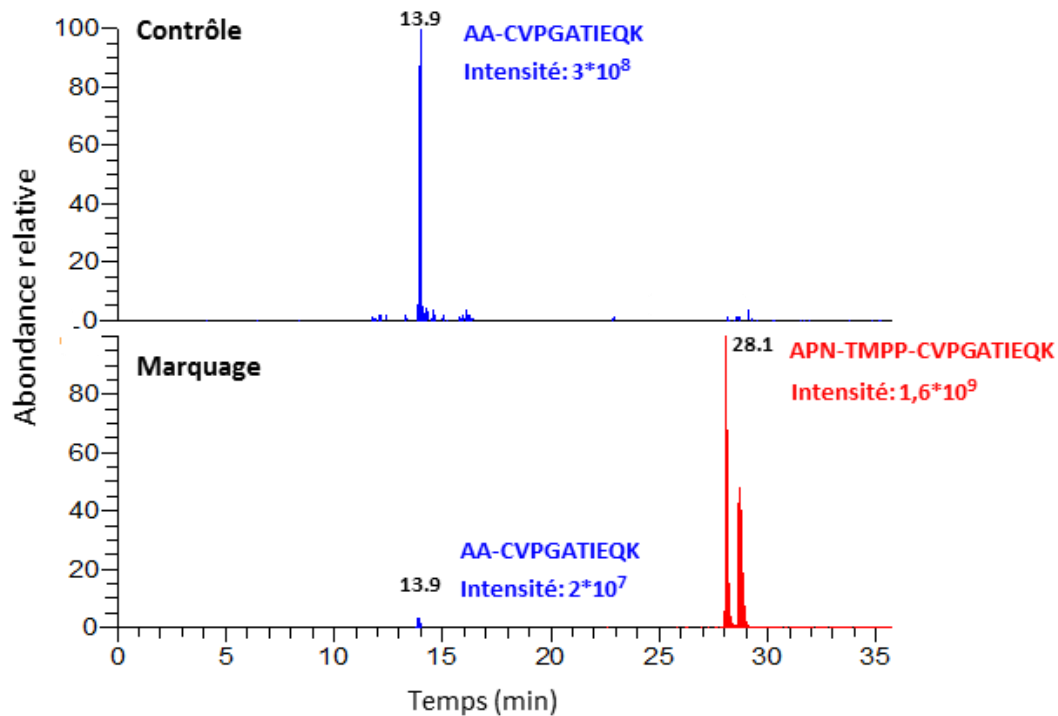


Figure 81: Comparaison des courants d'ions (XIC ; *eXtracted Ion Chromatogram*) du peptide tryptique AA-CVPGATIEQK de la conalbumine, dans le contrôle et le marquage (en bleu). XIC du peptide tryptique APN-TMPP-CVPGATIEQK dans le marquage (en rouge).

Cependant, l'extraction du courant d'ion des peptides dérivés a fait apparaître deux pics distincts sur le chromatogramme correspondant à des diastéréoisomères Z et E. La formation des deux diastéréoisomères a été observée pour l'ensemble des peptides à cystéine dérivés par l'APN-TMPP, avec un ratio d'environ 60/40. A titre d'exemple, Figure 82, l'extraction du courant d'ion du peptide APN-TMPP-CVPGATIEQK met en évidence deux pics aux temps de rétention 28.1 et 28.7 minutes (ratio 65/35).

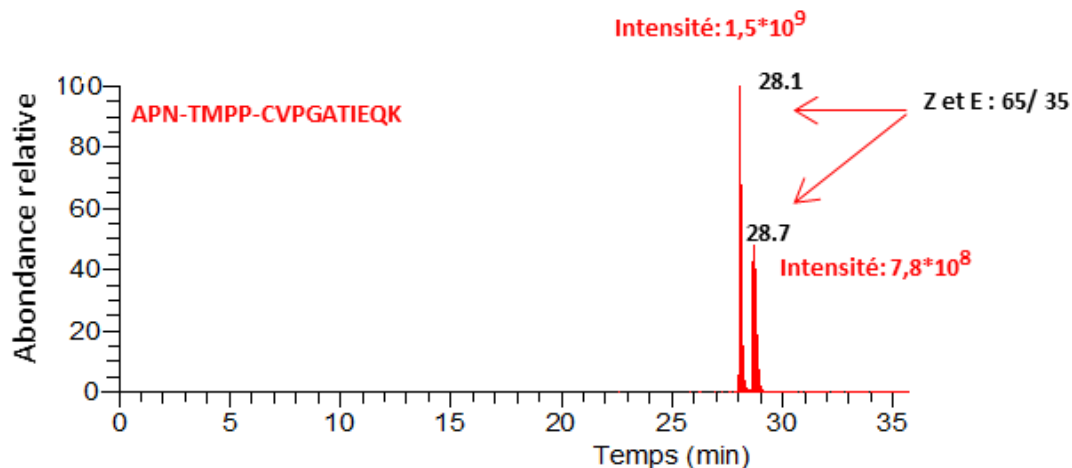


Figure 82: Extraction du courant d'ion (XIC ; *eXtracted Ion Chromatogram*) du peptide tryptique APN-TMPP-CVPGATIEQK dans l'analyse de marquage du digestat de la conalbumine.

La formation de ces diastéréoisomères n'a pas été observée lors de l'étude du marquage au niveau peptidique réalisé sur un digestat moins complexe (lysozyme)¹⁴⁶. Les conditions chromatographiques utilisées pour ce digestat n'ont pas été optimisées pour un mélange complexe, donc les diastéréoisomères n'étaient pas séparés. De plus, la réaction de marquage a été faite à température ambiante, contrairement

à nos conditions opératoires à 60°C. Or, la température peut influencer la chimiosélectivité d'une réaction. Nous avons réalisé des essais à température ambiante, qui n'ont pas permis de limiter la formation des diastéréoisomères (ratio 60/40).

L'un des objectifs des méthodes de marquage est de réduire la complexité de l'échantillon constitué par les peptides de digestion de l'échantillon biologique à analyser. L'impact de l'ajout d'APN-TMPP, hydrophobe, sur le comportement chromatographique (phase inverse) des peptides dérivés a donc été évalué ci-dessous.

Le fractionnement par séparation chromatographique. Les résultats obtenus ont montré l'enrichissement d'une zone du chromatogramme correspondant à la zone d'éluion des peptides hydrophobes.

La comparaison des chromatogrammes (BPC) des digestats peptidiques avant et après marquage a mis en évidence un décalage des temps de rétention des peptides marqués APN-TMPP, et la séparation de ceux-ci avec le reste des peptides tryptiques (Figure 83).

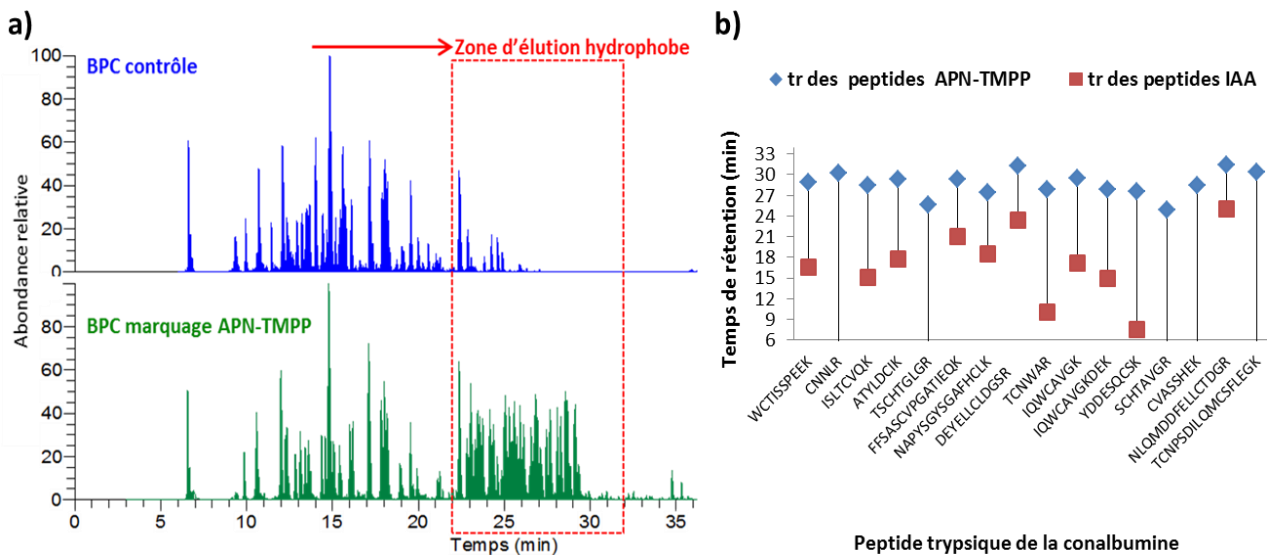


Figure 83: a) Comparaison des chromatogrammes (BPC ; *Base Peak Chromatogram*) des digestats peptidiques de la conalbumine contrôle et marqué. Le BPC du contrôle est présenté en haut (bleu), celui du marquage par l'APN-TMPP en bas (vert). b) Graphique de comparaison des temps de rétention (tr), en rouge des peptides tryptiques IAA de la conalbumine (contrôle), en bleu des peptides APN-TMPP.

Ces résultats mettent en évidence l'éluion plus tardive des peptides dérivés APN-TMPP, due à une meilleure rétention sur la colonne en phase inverse. Cet isolement des peptides de digestion marqués des peptides non marqués permet d'augmenter les probabilités d'obtenir un spectre MS/MS de qualité et de diminuer la compétition pour l'ionisation électrospray (ESI).

C.2.2 Les dérivés ammoniums quaternaires (QAT)

En 2004 Ren et al ont décrit la dérivation chimiospécifique des cystéines avec un dérivé QAT—acrylamidopropyle, sur une protéine modèle. Les QAT possèdent une charge permanente (ammonium) dans leurs structures qui leur confèrent donc les avantages décrits pour le TMPP (augmentation de l'ionisation, fragmentation caractéristique en CID). La stratégie développée par Ren et al, permet d'enrichir les peptides

QAT par chromatographie SCX¹⁵³. Cependant, cette stratégie multi-étapes n'a pas été appliquée, à notre connaissance, à l'étude d'échantillons biologiques complexes.

Un réactif type QAT-phényle, synthétisé, couplé à un groupement iodoacétate a été testé pour dériver la conalbumine. La réaction de marquage des cystéines est présentée Figure 84.

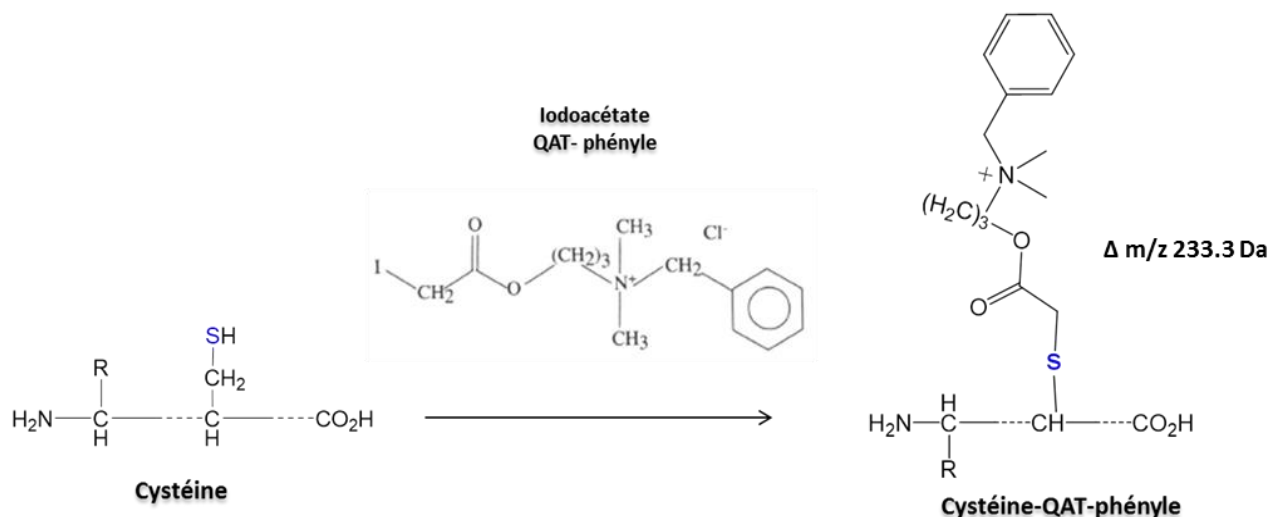


Figure 84: Représentation schématique de la réaction de marquage chimique des cystéines par le dérivé QAT-phényle. La masse ajoutée au peptide dérivé, est de 233.3 Da.

La conalbumine a été solubilisée dans un tampon urée à pH 8. Après ajout d'une solution de QAT-phényle à 1mM dans du DHIC (30°C pendant 30 minutes), l'échantillon a été analysé par gel 1D avant d'être digéré et soumis au protocole analytique présentée Figure 78.

Dans cet essai préliminaire, nous nous sommes essentiellement focalisés sur l'étude de la fragmentation des peptides dérivés par le QAT.

La fragmentation MS/MS. La vérification manuelle des spectres MS/MS des peptides dérivés par le QAT-phényle a mis en évidence une fragmentation caractéristique, équivalente à celle obtenue pour les dérivés phosphonium. Les fragments « b et a » sont générés préférentiellement aux fragments de type « y ». A titre d'exemple, des spectres de fragmentation MS/MS pour les peptides tryptiques WC(QAT)TISSPEEK et WC(IAA)TISSPEEK, sont présentés Figure 85.

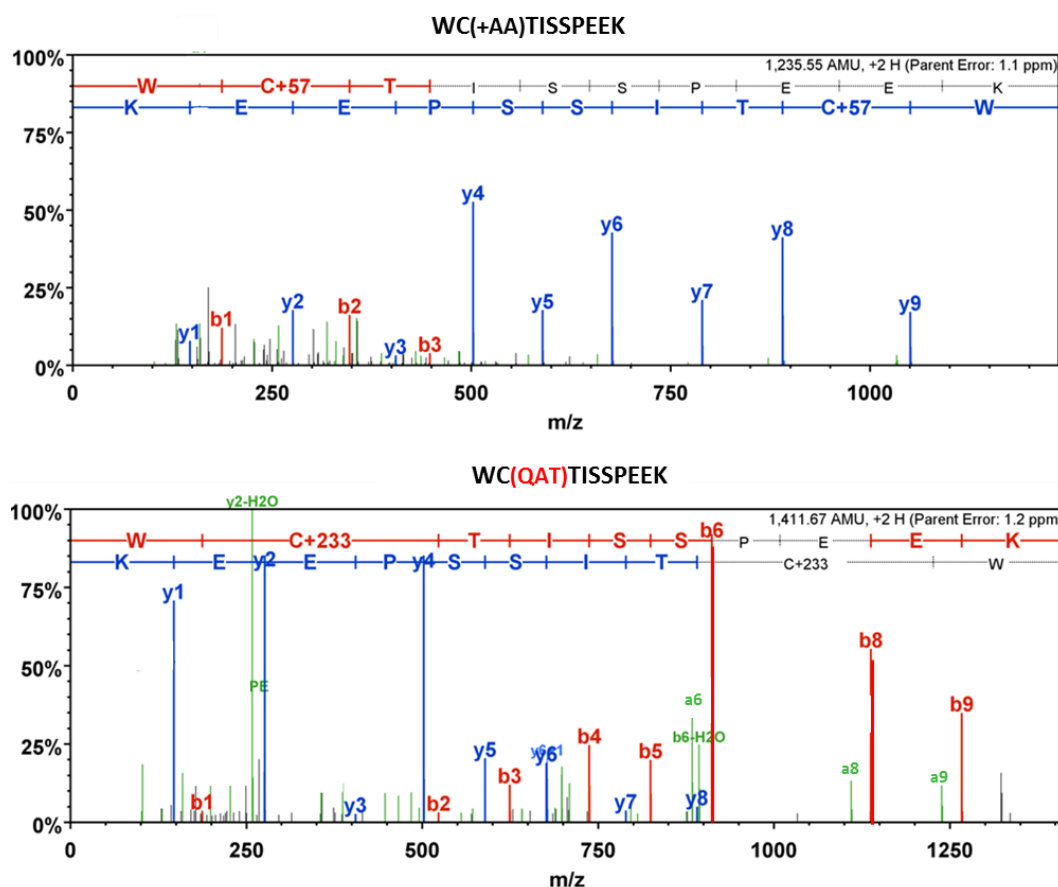


Figure 85: Comparaison des spectres de fragmentation MS/MS des peptides WC(+AA)TISSPEEK en bleu (en haut) et WC(QAT)TISSPEEK en rouge (en bas) en CID.

Les essais préliminaires ont mis en évidence la chimiospécificité de la réaction des QAT-phényle pour les cystéines de la conalbumine. Aucune réaction secondaire n'a été observée avec les chaînes latérales des lysines ou des tyrosines. Contrairement aux dérivés APN-TMPP, la solubilité des réactifs QAT permet de travailler à des concentrations molaires de réactif plus importantes.

Les temps de rétention de peptides marqués QAT-phényle et de leurs homologues dérivés IAA dans le contrôle ont été comparés. L'ajout du QAT ne permet pas de séparer les peptides marqués du reste du digestat peptidique. Cependant, une modification de la structure chimique du réactif pourrait permettre de gagner en hydrophobie et donc d'augmenter le temps de rétention de ces peptides dérivés QAT.

C.2.3 Conclusion pour le marquage chimique des cystéines

Les dérivés APN-TMPP.

La dérivation à l'APN-TMPP a répondu à la plupart des objectifs fixés pour un réactif de marquage chimique, une forte réactivité, une charge permanente et une structure hydrophobe. Donc toutes les caractéristiques sont réunies pour faire de l'APN-TMPP un bon candidat pour une stratégie de marquage.

Cependant, l'étude a également montré la formation inattendue de deux diastéréoisomères ce qui n'est pas compatible avec une méthode de quantification. Premièrement, parce que l'intensité du signal du peptide dérivé est divisée en deux. L'avantage du groupement TMPP (augmentation de l'ionisation et donc du facteur de réponse) qui permet d'augmenter la probabilité de détection des peptides dérivés est donc minimisé. Deuxièmement, parce que la formation de deux diastéréoisomères par peptide dérivé (Z et E) double la

complexité des peptides dérivés du digestat peptidique. Or l'objectif fixé des stratégies de marquage est la réduction de complexité du digestat pour rendre l'analyse quantitative plus facile sans avoir d'information redondante.

Ces réactifs restent toutefois très prometteurs pour la caractérisation de protéines recombinantes, notamment pour le greffage des molécules sur les anticorps dans le cadre d'ADC (antibody-drug conjugate), une classe importante de médicaments biopharmaceutiques puissants. Récemment un nouveau réactif de couplage avec l'APN a été mis au point, pour la bioconjugaison des fonctions thiol dans la préparation d'ADC décrit pour être stable dans le plasma¹⁵⁴.

Pour contourner le problème de formation de diastéréoisomères, il a donc été envisagé de remplacer la partie réactive (APN) par un groupement, moins réactif, de type iodoacétate¹⁵⁵. La synthèse d'un composé iodoacétate couplé au TMPP est en cours par les chimistes (laboratoire du Dr Alain Wagner) et sera évalué sur la conalbumine par la suite au laboratoire.

En conclusion sur notre travail de dérivation des cystéines, il ressort que le fractionnement chromatographique mis au point, qui élue d'abord les peptides non dérivés, et ensuite les peptides dérivés, constitue un très gros avantage pour l'étude de protéomes complexes. Il apporterait les données nécessaires pour une quantification, sans que les informations d'identification des protéines soient perdues. Ce travail sera poursuivi avec l'optimisation nécessaire de la structure du réactif pour éviter la formation de diastéréoisomères.

En parallèle à notre étude des dérivés APN-TMPP, la piste d'une autre série de réactifs, possédant également une charge permanente, apportée par des ammoniums quaternaires, a été explorée. Les résultats préliminaires obtenus avec ces derniers sont présentés succinctement ci-dessous

Les dérivés ammoniums quaternaires

Cette étude préliminaire a montré que les ammoniums quaternaires semblent être des candidats prometteurs pour une stratégie de marquage chimique. La dérivation avec le réactif QAT-phényl a en effet montré une chimiospécificité pour les cystéines, une fragmentation caractéristique, telle une signature des peptides à cystéines dérivés. Cependant, les conditions expérimentales doivent être encore optimisées pour augmenter le rendement de réaction (tampon de solubilisation et pH).

Dans le but de réduire la complexité du digestat peptidique dans une étude de protéome complexe, une des perspectives serait de modifier la structure du QAT pour augmenter l'hydrophobie des peptides dérivés et donc modifier leur rétention chromatographique. A titre d'exemple, la structure d'un dérivé QAT-naphtalène plus volumineux et plus hydrophobe est présentée Figure 86.

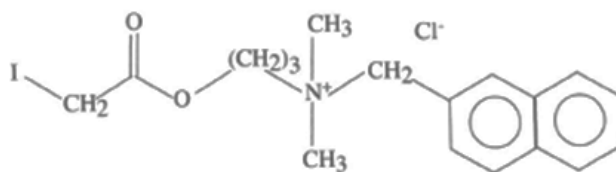


Figure 86: Structure chimique d'un dérivé QAT-naphtalène. La masse ajoutée au peptide dérivé est de 284,2 Da.

D. Le marquage des tryptophanes

D.1. L'état de l'art du marquage chimique des tryptophanes

Le tryptophane (Trp ou W), est le deuxième acide aminé le moins abondant (après la cystéine) avec une fréquence de l'ordre 1% (en fonction des organismes)¹⁴⁴. Cependant, environ 90% des protéines contiennent au moins un résidu tryptophane dans leur séquence¹⁵⁶. Le cycle aromatique (indole) des tryptophanes est un groupement réactif dont la dérivation chimique est peu décrite.

En 2004, l'une des premières approches sélectives de dérivation par des composés diazovinyl-substitués a été décrite par Antos¹⁵⁷.

En 2007, une dérivation des peptides à W avec un malondialdéhyde couplé à une étape d'enrichissement¹⁵⁸ ou plus récemment, une dérivation par des réactifs iodés hypervalents¹⁵⁹ ont été reportées. Cependant, ces méthodes n'ont pas été appliquées dans des stratégies de protéomique quantitative.

Une des approches, décrite pour la première fois en 2003 par K. Tanaka et son équipe¹⁴⁷, a plus particulièrement retenu notre attention. Le réactif de marquage utilisé est le NPBSCI (pour 2-nitrobenzenesulfonyl chloride). Pour éviter la confusion avec le NBS (N-bromosuccinimide), par la suite nous l'appellerons NPBSCI, c'est le réactif qui donne ensuite le dérivé NPBS. Le NPBSCI réagit avec le noyau indole des tryptophanes comme cela est présenté Figure 87.

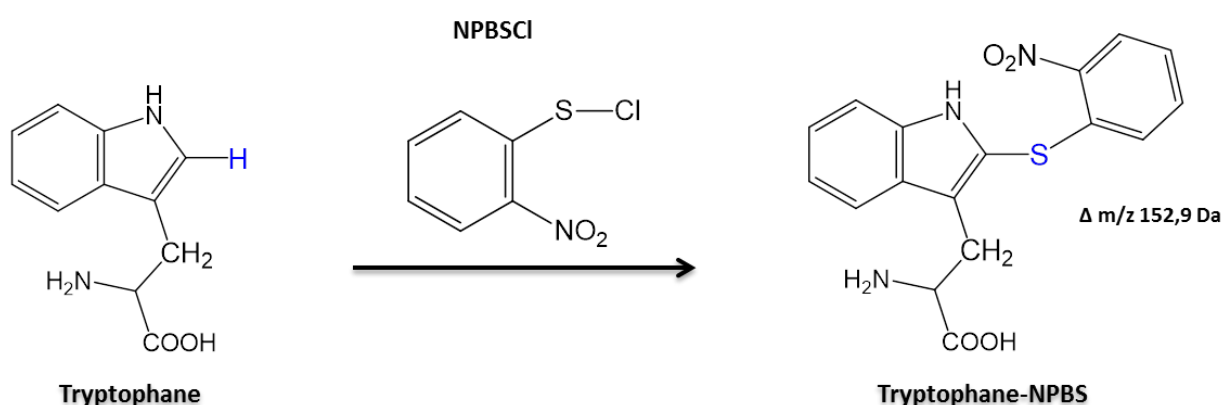


Figure 87 Représentation schématique de la réaction de marquage chimique entre le tryptophane et le NPBSCI.

La stratégie analytique repose sur une étape d'enrichissement réalisée par chromatographie d'affinité entre les peptides dérivés au NPBS et une colonne Sephadex¹⁶⁰ ou phényl sépharose¹⁶¹. Ce type d'interaction entre des cycles aromatiques de la phase stationnaire et ceux des peptides portant une étiquette NPBS est appelé « pi-stacking ».

Les peptides dérivés enrichis sont ensuite séparés par chromatographie et les fractions collectées sont analysées par MALDI-TOF¹⁶²⁻¹⁶³. Malgré la « décomplexification » des échantillons en ciblant une sous-population, un grand nombre de peptides sont encore présents dans les fractions enrichies.

L'abondance relative des protéines peut être déterminée entre deux échantillons par marquage aux isotopes stables (NPBS ¹²C/¹³C). La stratégie a été appliquée avec succès pour l'identification de nouveaux biomarqueurs du stress oxydatif dans les hépatocytes humains¹⁶⁰. Cependant, les nombreuses étapes de la stratégie sont chronophages et augmentent la possibilité d'introduire des biais expérimentaux. A ce jour, aucune étude LC-MS/MS d'échantillons complexes n'a été faite, nous avons donc voulu explorer cette stratégie de marquage et l'adapter pour des analyses protéomiques basées sur de la nanoLC-MS/MS.

D.2. Résultats et discussion

D.2.1 Le marquage des tryptophanes de la protéine modèle.

Une première étude des conditions expérimentales du marquage des tryptophanes par l'approche NPBS a été réalisée sur une protéine modèle (pH, température, tampon). Les différents protocoles de marquage testés sur la protéine modèle sont présentés Figure 88. Je détaillerai les principaux protocoles expérimentaux mis au point ci-dessous.

Pour mettre au point les conditions expérimentales, deux tampons de d'extraction couramment utilisés pour l'étape d'extraction protéique (urée et SDS) et deux solutions de solubilisation pour le NPBS, une solution à 25 % l'acide acétique (AcOH) et une solution de tétraméthyle urée à pH 2.4 (TMU), ont été testés. En milieu acide, la réactivité sélective du NPBS avec le tryptophane prévaut sur celle avec la cystéine¹⁶⁴.

Après dépôt sur gel 1D-PAGE, la protéine a été digérée *in gel* puis le digestat peptidique a été analysé par nanoLC-MS/MS (colonne de piégeage classique C18).

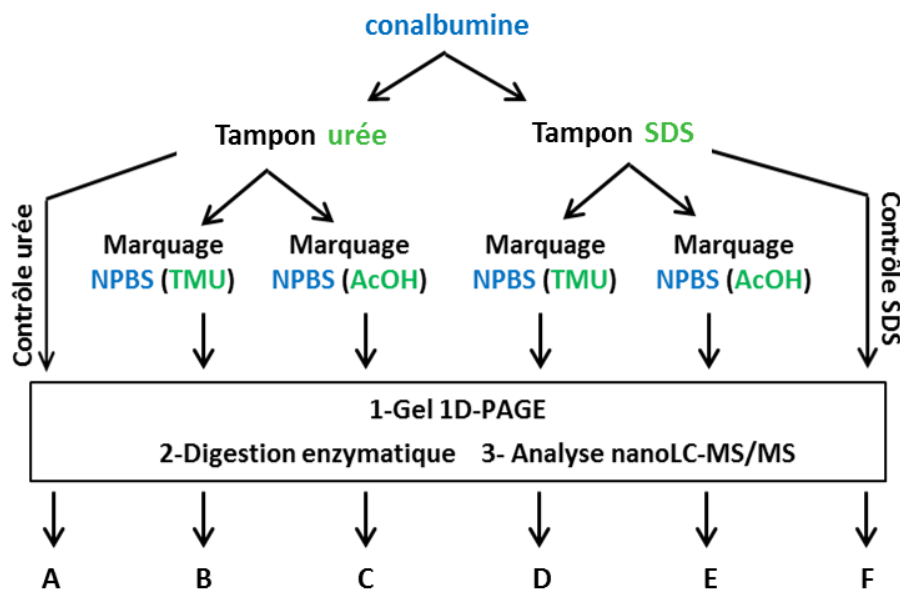


Figure 88: Représentation schématique des différents protocoles de marquage chimique des tryptophanes par le NPBS testés. Deux tampons (urée et SDS) et deux solutions (AcOH, TMU) ont été optimisés (condition B, C, D et E). Deux contrôles, urée et SDS ont été faits (A et F).

Les résultats du marquage des tryptophanes de la conalbumine

Quatre conditions expérimentales B, C, D, E de marquage et leurs contrôles A et F, présentées dans la Figure 88, ont été testées sur la conalbumine. Les digestats tryptiques générés pour les échantillons ont été analysés par nanoLC-MS/MS.

Dans le but d'évaluer la stratégie de marquage, les paramètres tels que la chimiospécificité, la fragmentation, l'efficacité de marquage et le comportement chromatographique des peptides dérivés ont été étudiés. La reproductibilité du marquage a été vérifiée sur des répliques de marquage (A1, A2, A3, B1,...).

La chimiospécificité. Les peptides identifiés dans le digestat de la conalbumine ont permis de déterminer un recouvrement de séquence de la protéine de plus de 80 % quelle que soit la condition expérimentale. Mais,

seule la condition E (SDS et AcOH) a permis d'identifier les dix résidus W présents dans la séquence de la conalbumine (Figure 89).

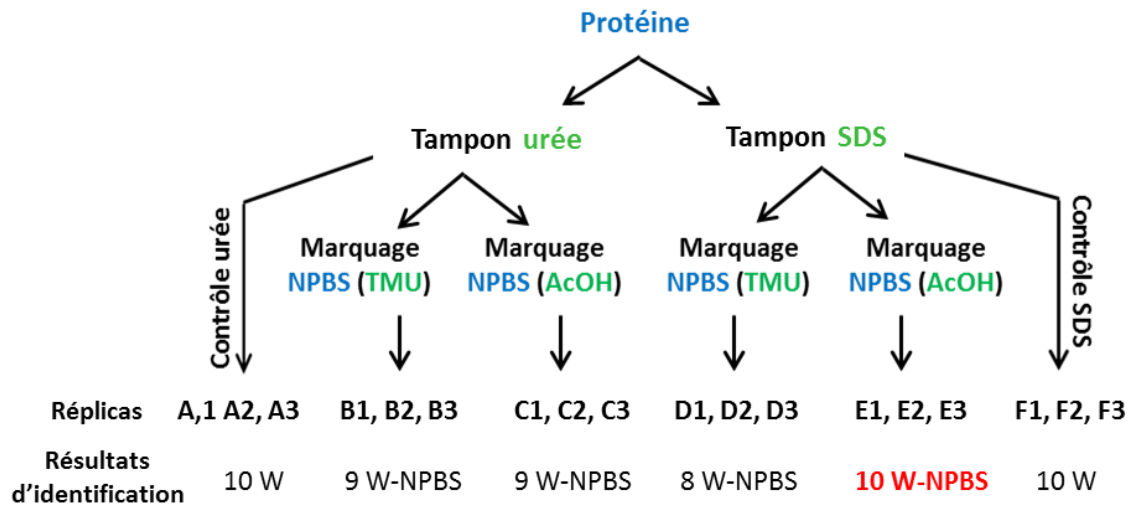


Figure 89: Résultats d'identification des peptides tryptiques à tryptophanes (W) de la conalbumine, dérivés par le NPBS dans les quatre conditions expérimentales de marquage (B à E) et non dérivés dans les contrôles (A et F). La condition de marquage E, qui a permis d'identifier tous les W de la conalbumine, est en rouge

Les résultats obtenus avec la condition E sont reproductibles car les mêmes peptides ont été identifiés dans chaque réplique de marquage.

L'efficacité de marquage. La comparaison de l'intensité des peptides dérivés a mis en évidence une efficacité de marquage plus élevée avec la condition expérimentale E (SDS, AcOH).

A titre d'exemple, la comparaison des intensités du peptide WC(+AA)TISSPEEK avant et après marquage est présentée Figure 90. La comparaison du peptide WC(+AA)TISSPEEK et du peptide NPBS-WC(+AA)TISSPEEK illustre l'efficacité de marquage plus élevée obtenue avec la condition E.

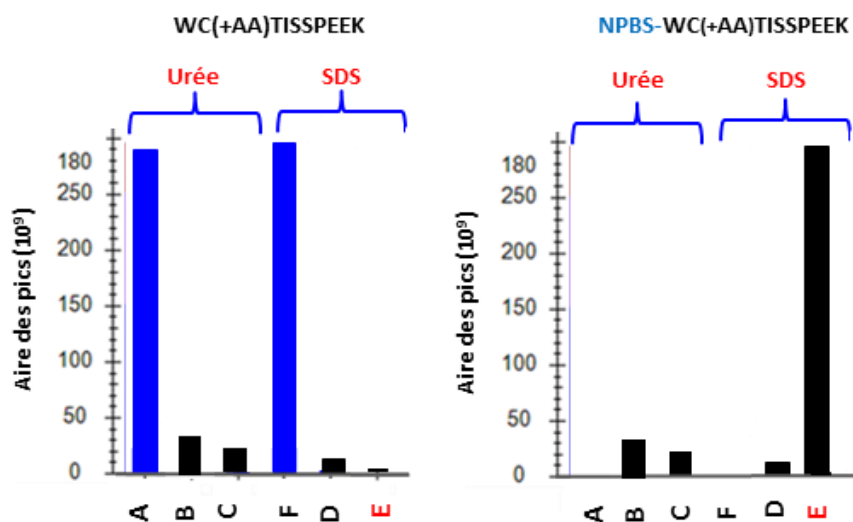


Figure 90: Extraction des aires sous la courbe pour les peptides WC(+AA)TISSPEEK et NPBS-WC(+AA)TISSPEEK dans les conditions de marquages (B,C,F et E) et leurs contrôles (A et D). Le meilleur rendement de marquage est obtenu dans la condition E.

Les résultats obtenus ont montré un décalage des temps de rétention des peptides dérivés NPBS d'environ cinq minutes avec notre gradient d'élution de 40 minutes (Figure 91).

A titre d'exemple, le chromatogramme du peptide VEDIWSFLSK avant et après marquage, est présenté Figure 91. Le peptide dérivé est séparé chromatographiquement de son homologue non marqué de quatre minutes.

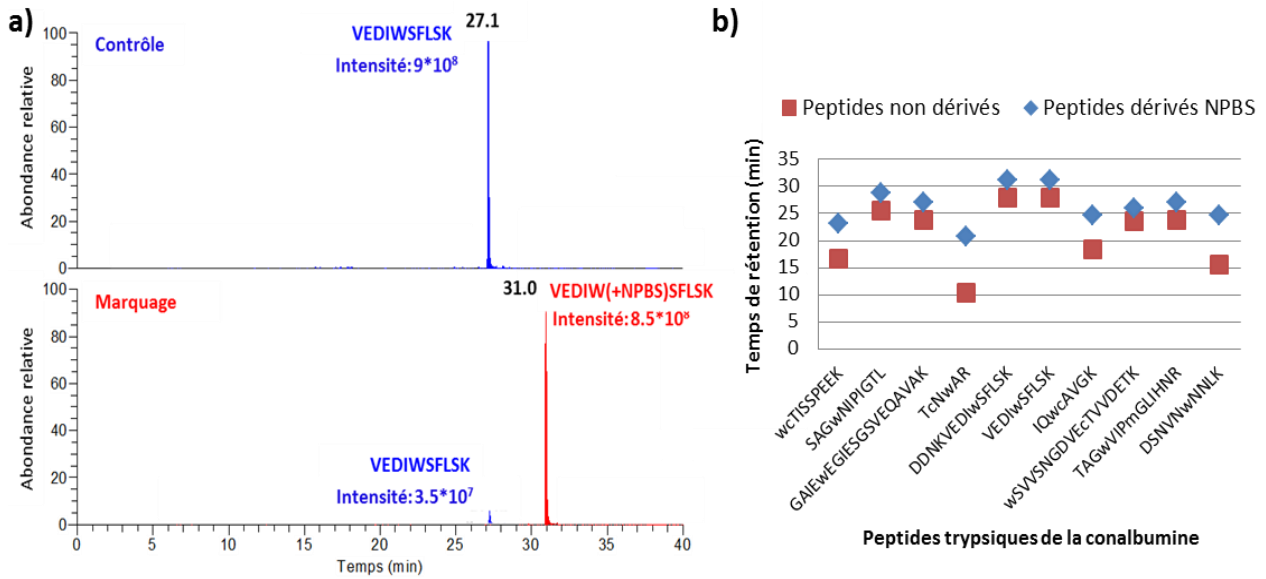


Figure 91: a) Comparaison des extractions des courants d'ions (XIC) du peptide trypsique VEDIWSFLK non dérivé de la conalbumine, dans le contrôle et le marquage (en bleu). EIC du peptide trypsique VEDIW(+NPBS)SFLK dans le marquage (en rouge) obtenu avec la condition E. b) Graphique de comparaison des temps de rétention, en rouge des peptides trypsiniques de la conalbumine non dérivés et en bleu des peptides dérivés.

Fragmentation. L'étude de la fragmentation des peptides dérivés au NPBS a montré que le marquage des peptides par ce réactif ne modifie pas le schéma de fragmentation classique en CID. Le marquage est détecté par spectrométrie de masse par un ajout de 153 Da. Cela n'affecte pas l'efficacité de fragmentation des peptides. La comparaison des spectres MS/MS d'un peptide avec et sans marquage est présentée Figure 92. Avec ou sans marquage, la fragmentation du peptide en CID génère des ions fragments de type « b et y » comme attendu.

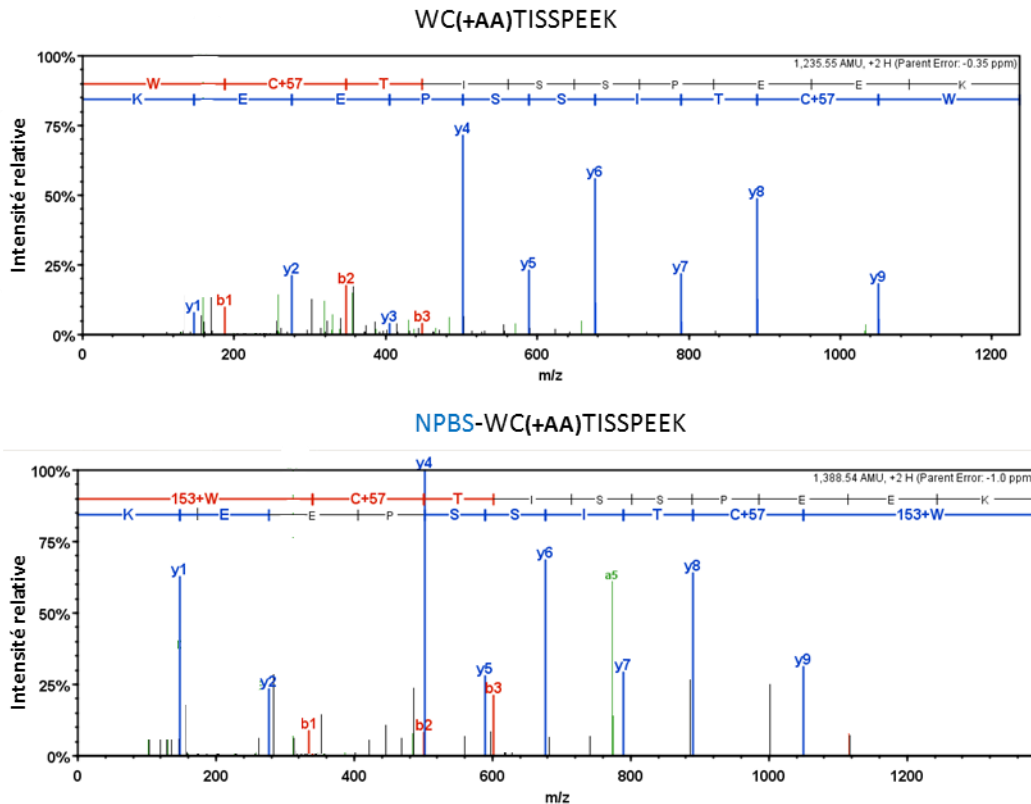


Figure 92: Comparaison des spectres de fragmentation MS/MS des peptides WC(+AA)TISSPEEK en bleu (en haut) et NPBS-WC(+AA)TISSPEEK (en bas) en CID.

Conclusion sur l'étude d'une protéine modèle.

Les conditions expérimentales avec un tampon SDS et de l'acide acétique, ont permis d'identifier les 10 W de la conalbumine dérivés au NPBS, de façon reproductible. Le rendement de marquage pour cette condition a été estimé proche de 50%. Les rendements de marquage atteignant rarement les 100%, et notre projet étant de développer d'une stratégie robuste et reproductible, un rendement proche de 50% est tout à fait acceptable.

Les études de quantification menées sur des échantillons biologiques complexes se sont souvent révélées biaisées à cause de variabilités techniques. La reproductibilité de la stratégie de dérivation des tryptophanes développée a donc été évaluée sur protéome plus complexe (*E. coli*), représentatif d'un protéome classique. En réduisant le nombre d'étapes, par rapport à la stratégie initialement décrite, la variabilité technique devrait être minimisée au profit de la reproductibilité.

D.2.2 Le marquage des tryptophanes d'un échantillon complexe (extrait *E. coli*).

Les extraits de cultures cellulaires nécessitent l'introduction d'étapes supplémentaires adaptées (nature du tampon, précipitation des protéines) pour l'analyse protéomique. Ainsi une étape de précipitation des protéines (éthanol) a été introduite dans la condition expérimentale optimisée (E) sur la protéine modèle. Cette étape permettra également d'éliminer l'excès de NPBS.

En parallèle, un autre tampon de solubilisation a également été testé, le CTAC¹⁶⁵ (cétyltriméthylammonium chlorure, ammonium quaternaire cationique). Il est utilisé dans les désinfectants, entant qu'agent de lyse bactérienne.

Contrairement au SDS chargé négativement, le CTAC est chargé positivement et va permettre de précipiter tous les intermédiaires métaboliques chargés négativement (acides nucléiques, lipides,...). Le CTAC permet d'éliminer certains métabolites provenant du milieu de culture (par exemple le glutathion) qui pourraient interférer avec le réactif de marquage.

L'utilisation du tampon CTAC nécessite impérativement de passer par une étape de précipitation, car il n'est pas compatible avec une migration électrophorétique. En effet contrairement au SDS, le CTAC ne se lie pas de manière homogène aux protéines perturbent la migration électrophorétique des protéines.

Les différentes conditions expérimentales mises en œuvre pour le marquage d'extrait d'*E. coli* sont présentées Figure 93.

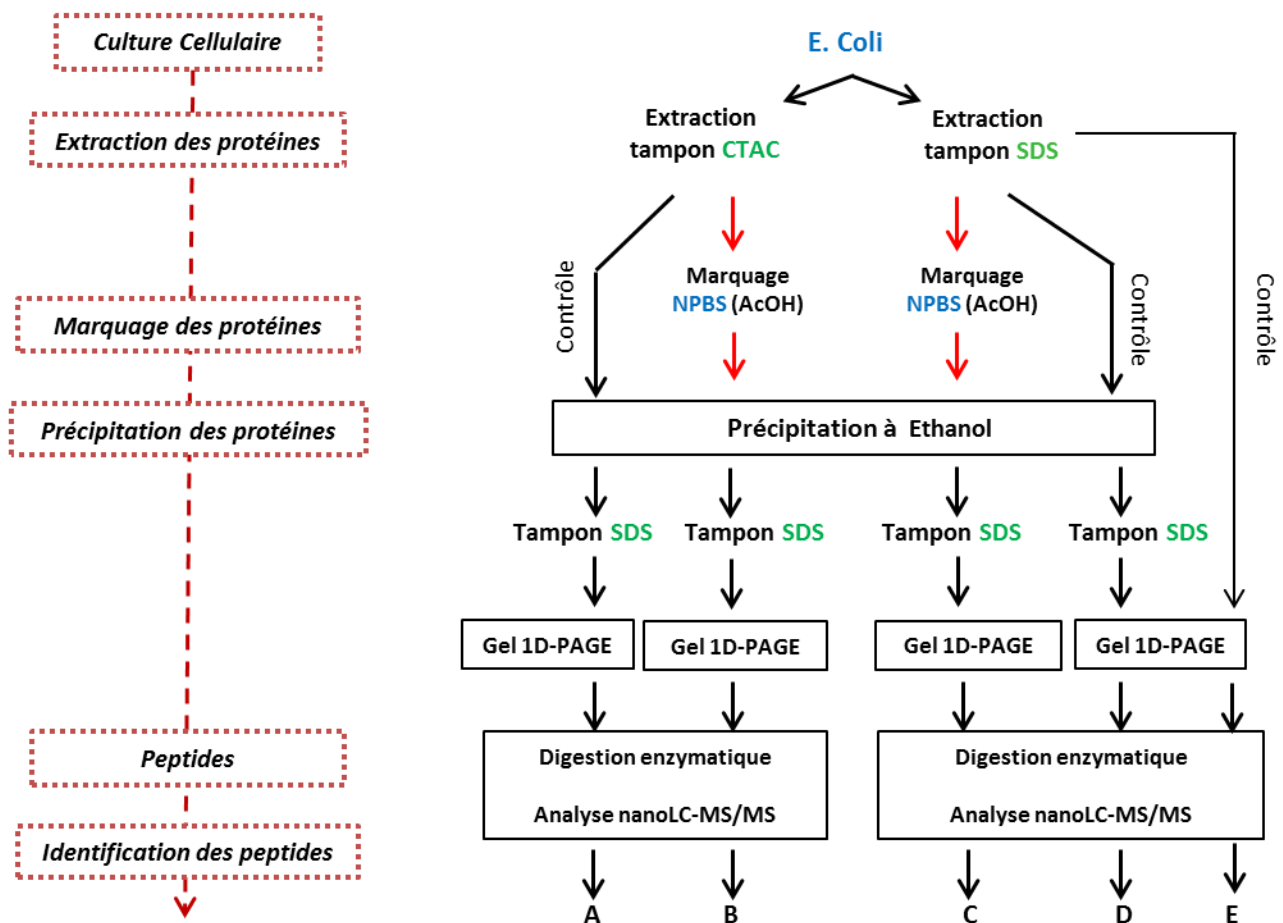


Figure 93: Représentation schématique de la stratégie d'optimisation des conditions expérimentales de marquage d'un échantillon complexe (*E. coli*). Deux conditions expérimentales de marquage (B et C) et deux contrôles avec précipitation des protéines (A, D) et un sans précipitation (E), ont été réalisés.

Dans le but d'évaluer la reproductibilité de ce protocole de marquage chimique des tryptophanes sur le protéome complexe d'*E. coli*, trois répliques de marquage ont été réalisées pour chaque condition expérimentale testée.

Les résultats de marquage des tryptophanes de l'extrait d'*E. coli*.

La précipitation des protéines est une étape qui permet d'éliminer les molécules incompatibles avec l'analyse MS. Le contrôle, réalisé sans étape de précipitation (Condition E), nous a permis de vérifier que la précipitation n'influe pas de manière drastique sur le nombre de peptides et de protéines identifiées.

Je détaillerai donc les résultats obtenus avec une étape de précipitation, qui correspondent aux conditions B (CTAC) et C (SDS). Pour valider les peptides dérivés, nous n'avons pris en compte que ceux qui ont été identifiés dans au moins deux répliques sur trois.

Le nombre de peptides dérivés au NPBS. La Figure 94 montre que le nombre de protéines totales identifiées est supérieur avec la condition B (CTAC). Cette condition a permis d'identifier 200 peptides dérivés en plus par rapport à la condition C (SDS). Les résultats ont montré que les peptides dérivés par l'étiquette NPBS représentaient environ 11% des peptides totaux identifiés dans le digestat d'*E. coli*, dans les deux conditions.

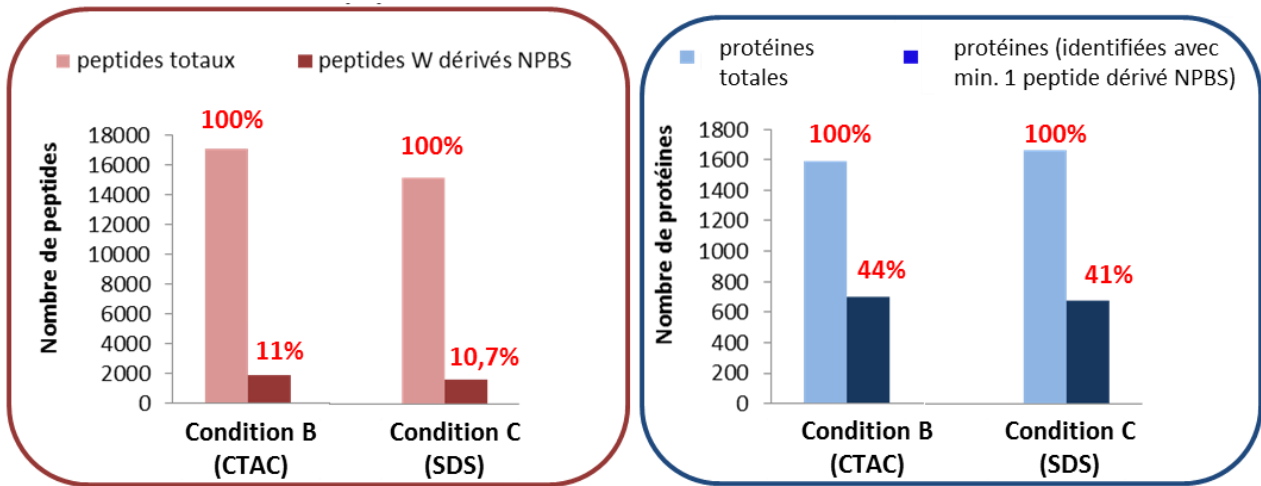


Figure 94: A gauche, l'histogramme de comparaison du nombre de peptides totaux et dérivés NPBS identifiés pour chaque condition (peptides identifiés dans au moins deux répliques sur trois). A droite, l'histogramme de comparaison du nombre de protéines totales identifiées (identifiées avec au minimum un peptide dérivé), pour chaque condition.

La reproductibilité des identifications. Une meilleure reproductibilité des identifications des peptides dérivés NPBS a été obtenue avec la condition B (CTAC). Les diagrammes de Venn montre que, parmi les peptides dérivés identifiés dans deux répliques sur trois, 74% sont communs aux trois répliques avec la condition B et 69% avec la condition C (Figure 95).

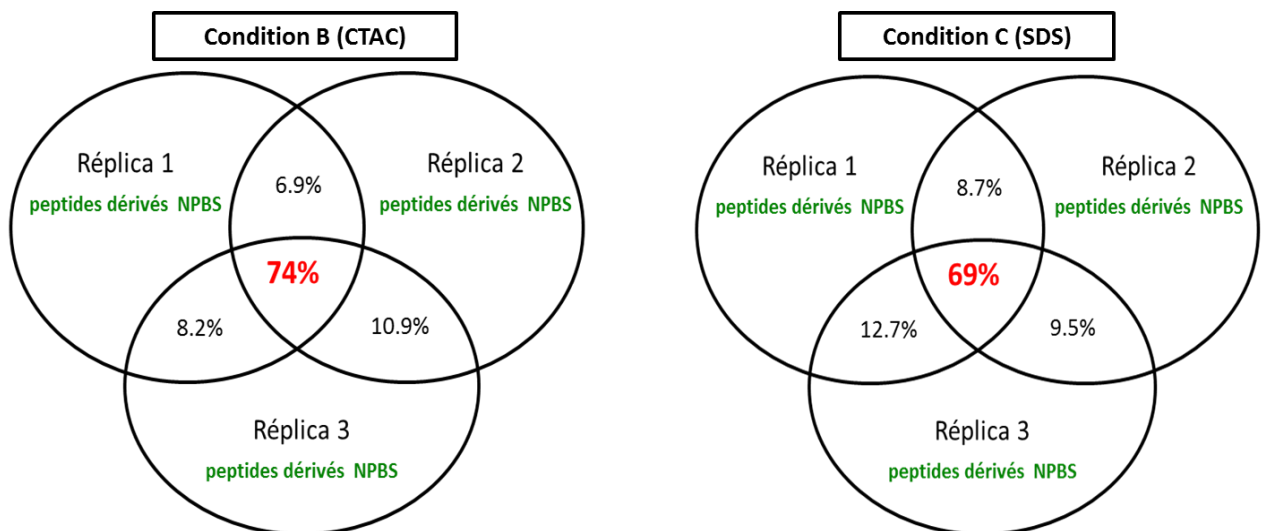


Figure 95: Diagramme de Venn représentant le nombre de peptides dérivés NPBS identifiés entre les répliques de marquage pour les conditions B (CTAC) et C (SDS). A gauche, la condition B : 74 % des peptides dérivés NPBS sont communs aux trois répliques. A droite, la condition C : 69% des peptides NPBS sont communs aux trois répliques.

La reproductibilité des peptides identifiés est élevée dans les deux conditions expérimentales, et plus particulièrement dans la condition B (CTAC).

Efficacité du marquage chimique. Pour évaluer l'efficacité du marquage entre les conditions B et C nous avons classé en deux catégories les peptides dérivés au NPBS dans chaque expérience.

- les peptides qui ont été identifiés uniquement dérivés NPBS
- les peptides qui sont identifiés sous les deux formes (dérivés NPBS et non dérivés).

Cette représentation (Figure 96) montre que plus de 79% des peptides dérivés dans B et 92% dans C, ont été identifiés uniquement portant l'étiquette NPBS, ce qui dénote d'une l'efficacité de marquage élevée.

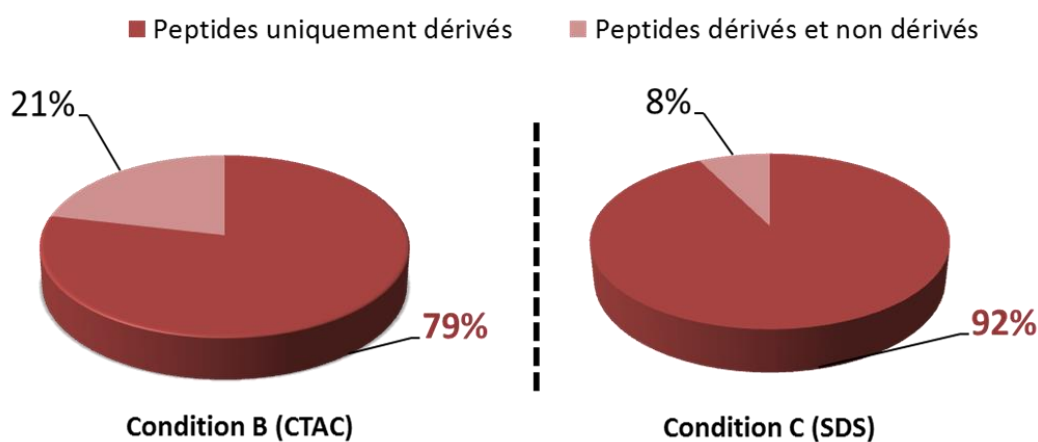


Figure 96: Histogramme de comparaison des peptides identifiés uniquement dérivés au NPBS et des peptides identifiés dérivés au NPBS et non dérivés, pour les conditions expérimentales B (CTAC) et C (SDS).

Dans la littérature, l'efficacité de marquage chimique des stratégies de dérivation, n'est en général, pas évaluée sur un protéome complexe. Notre étude détaillée a permis de montrer l'efficacité de marquage élevée sur une protéine modèle puis sur un protéome plus complexe.

Distribution des coefficients de variation entre le contrôle et le marquage.

Nous avons voulu évaluer l'introduction de biais techniques (entraînant une variabilité des mesures des courants d'ions extraits) avec l'étape de marquage par rapport au contrôle, sur les courants d'ions extraits des peptides en MS.

Un coefficient de variation (CV) a été calculé pour les valeurs extraites pour chaque peptide entre les réplicas de chaque condition. Le CV est une mesure quantitative de précision exprimant une variation par rapport à la valeur observée moyenne¹⁶⁶.

$$CV \% = 100 * \frac{[\text{Ecart type}]}{[\text{Moyenne}]}$$

Dans la stratégie de quantification d'extraction des courants d'ions MS1 (XIC), les informations quantitatives sont obtenues en MS, à partir de l'intégration des aire des pics chromatographiques pour chaque peptide (Partie I -C.). Le logiciel MaxQuant⁴³ a été utilisé pour valider les identifications et pour extraire les données XIC.

Résultats. L'introduction de l'étape de marquage n'a pas augmenté la variabilité des courants d'ions extraits par rapport au contrôle. La représentation de la distribution des CV a permis de mettre en évidence que le marquage augmente la valeur médiane des CV (27.48 vs 20.48) mais cette valeur reste tout à fait acceptable pour une stratégie de quantification MS1 (XIC) (Figure 97).

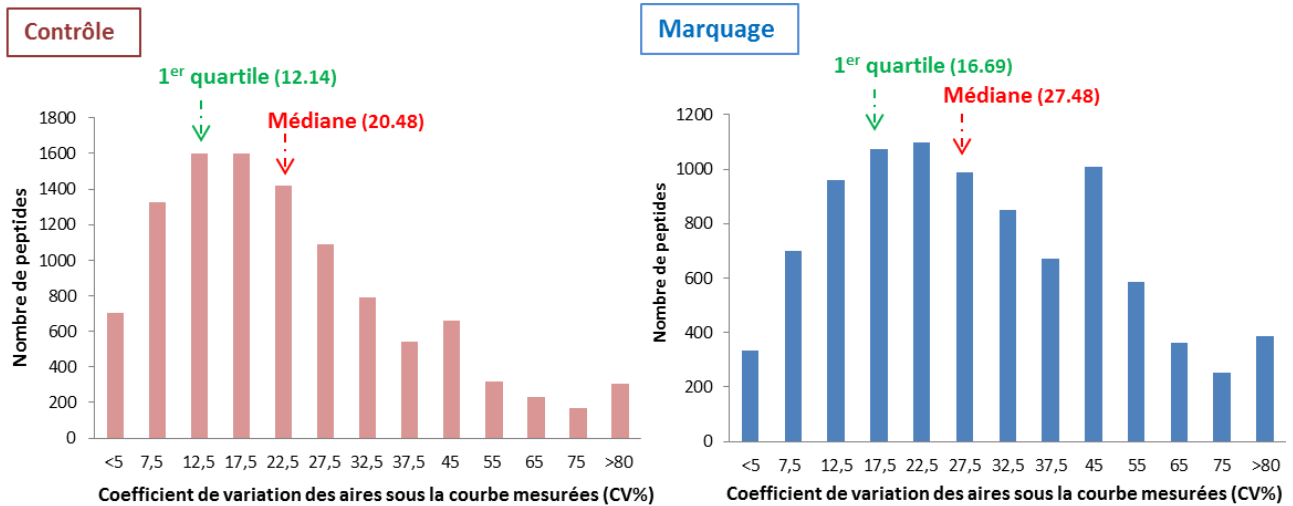


Figure 97: Représentation schématique de la distribution des CV des courants d'ions extraits (MS1) pour chaque peptide identifié dans les répliques des conditions « contrôle » et « marquage ».

Séparation chromatographique. Les chromatogrammes obtenus avec les conditions expérimentales B et C ont montré une élution plus tardive des peptides dérivés par rapport à leurs homologues non dérivés, d'environ dix minutes avec un gradient d'élution de 120 minutes. La Figure 98, montre que les peptides portant l'étiquette NPBS dans les conditions B (CTAC) sont élués dans une zone plus hydrophobe du chromatogramme que leurs homologues non dérivés dans le contrôle.

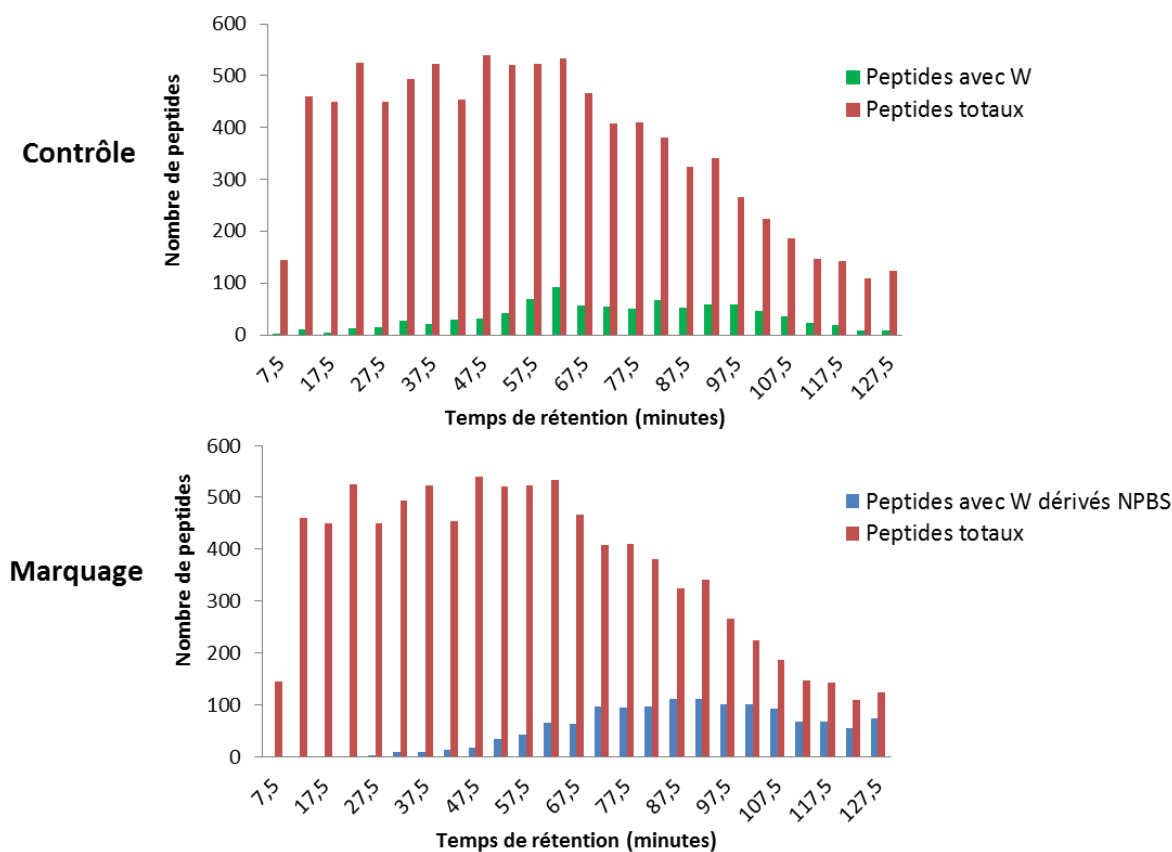


Figure 98 : Histogrammes des temps de rétention des peptides tryptiques du digestat d'*E. coli* identifiés dans le contrôle (totaux et avec W) et le marquage (totaux et dérivés) avec la condition B (CTAC)

Conclusion sur le marquage des tryptophanes d'un échantillon complexe (*E. coli*).

Les différentes conditions expérimentales testées sur des répliques ont montré que les meilleurs résultats ont été obtenus avec le tampon d'extraction CTAC.

Après avoir développé les conditions de marquage pour cibler spécifiquement un acide aminé (Etape 1 de la Figure 74, Partie II -B. , nous avons ensuite exploré la seconde étape pour décomplexifier le mélange peptidique. Nous avons donc testé un enrichissement « en ligne » des peptides dérivés pour les séparer plus efficacement des peptides non dérivés en rajoutant une étape d'enrichissement avant la colonne analytique de séparation.

D.2.3 L'enrichissement « en ligne » des peptides NPBS.

Nous avons testé une étape d'enrichissement facile à mettre en place, en remplaçant la pré-colonne de piégeage (ou colonne de chargement) C18, utilisée dans un système nanoLC-MS/MS classique, par une colonne avec une phase plus sélective pour les composés aromatiques, de type phényle.

Lors de l'étape de chargement du digestat peptidique, les peptides portant un NPBS devraient donc avoir une affinité spécifique pour les groupements phényles, interaction « *pi stacking* » et liaisons hydrophobes, et ainsi être mieux piégés que les peptides non dérivés. Par conséquent, le digestat élué vers la colonne analytique devrait être enrichi en peptides dérivés et être plus « simple ».

Cette étape doit permettre de diminuer la compétition pour l'ionisation entre les peptides dans la source du spectromètre masse ce qui devrait augmenter la qualité et la reproductibilité des mesures.

Dans un premier temps, nous avons évalué l'étape d'enrichissement « en ligne » (pré-colonne phényle), entre la condition de marquage et son contrôle.

Dans un deuxième temps, nous avons comparé la reproductibilité des peptides dérivés identifiés avec une étape d'enrichissement « en ligne » (pré-colonne phényle) par comparaison à une stratégie classique (pré-colonne C18).

Dans un troisième temps, nous avons évalué la variabilité des courants d'ions extraits mesurés des peptides avec l'étape d'enrichissement « en ligne » par rapport à la stratégie classique.

Des triplicatas pour les conditions de marquage et le contrôle d'extrait d'*E. coli* ont été réalisés et analysés.

Comparaison Marquage vs Contrôle (pré-colonne phényle)

L'identification de 1458 peptides à tryptophane (soit 15% des peptides totaux vs 11% dans le contrôle (Figure 99) a montré que ces peptides ont une meilleure affinité que les non marqués avec la phase stationnaire phényle. Dans l'expérience de marquage parmi les 1458 peptides, 1308 sont identifiés portant l'étiquette NPBS (soit 90 % des peptides à W).

Les peptides à tryptophanes sont mieux retenus grâce à l'affinité entre les phényles de la phase stationnaire et l'étiquette NPBS, et donc permet d'enrichir le digestat peptidique élué en peptides à W. (Figure 99).

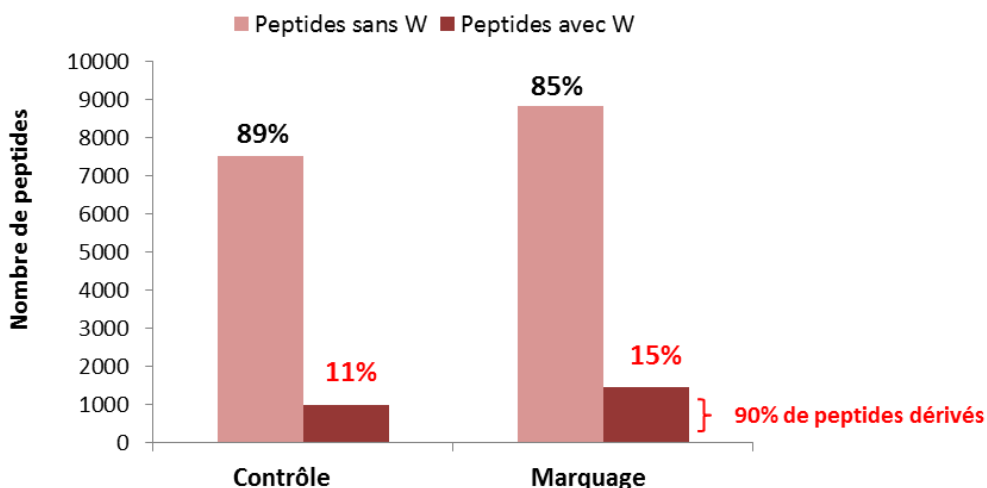


Figure 99: Histogramme de comparaison du nombre de peptides tryptiques identifiés sans W et ceux avec un W (dont W dérivé NPBS), dans le marquage et le contrôle (*E. coli*) après enrichissement sur pré-colonne de piégeage phényle.

Pour évaluer la robustesse de cette méthode d'enrichissement avec pré-colonne phényle par rapport à une C18, nous avons ensuite évalué la reproductibilité des courants d'ions extraits pour les peptides dérivés, utilisés dans la stratégie de quantification XIC ((Partie I -C.).

Comparaison des pré-colonnes phényle et C18

Un nombre moins élevé de peptides totaux a été identifié avec la pré-colonne phényle par rapport à une stratégie classique (C18). Comme attendu, les peptides présentant peu d'affinité pour la pré-colonne phényle sont moins retenus (Figure 100).

Parmi, 13% des peptides détectés sont dérivés au NPBS avec la phase stationnaire phényle contre 14% dans la stratégie classique (C18), la différence n'est pas significative (Figure 100). Le nombre de peptides NPBS

détectés est identique dans les deux cas. Ces peptides NPBS permettent d'identifier un nombre équivalent de protéines (43% vs 44% entre les deux phases stationnaires).

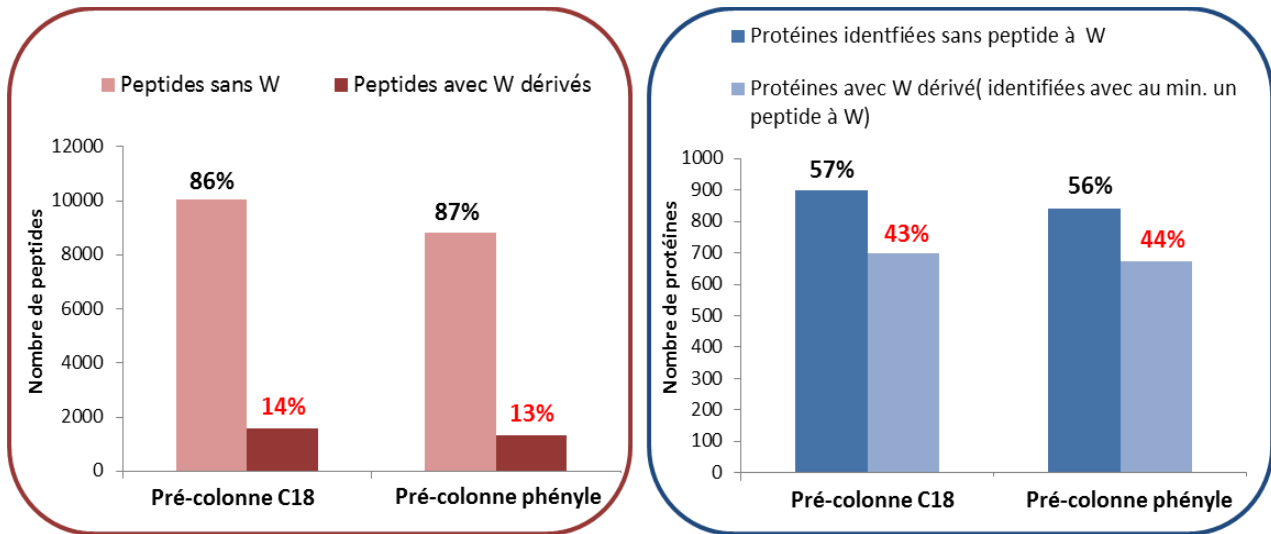


Figure 100 : Histogrammes de comparaison des peptides identifiés sans W et avec W dérivés (à gauche). Histogrammes de comparaison des protéines identifiées avec et sans W dérivés, avec la pré-colonne phényle et avec la pré-colonne C18.

Après avoir étudié l'impact de la phase stationnaire phényle sur le nombre d'identifications, nous avons évalué la reproductibilité des courants d'ions extraits, qui est le paramètre essentiel dans les stratégies de quantification.

Comparaison de la reproductibilité des signaux XIC (MS1) : pré-colonne phényle vs C18

Généralement dans les méthodes de quantification, un CV < à 20% est acceptable pour la variation des signaux XIC entre les répliques. Nous avons donc appliqué ce critère aux données.

Les résultats ont mis en évidence la reproductibilité comparable des courants d'ions extraits des peptides dérivés avec la phase stationnaire phényle par rapport à la stratégie classique. La Figure 101 montre que la reproductibilité des mesures d'intensité des peptides dérivés est du même ordre de grandeur avec les deux types de pré-colonnes.

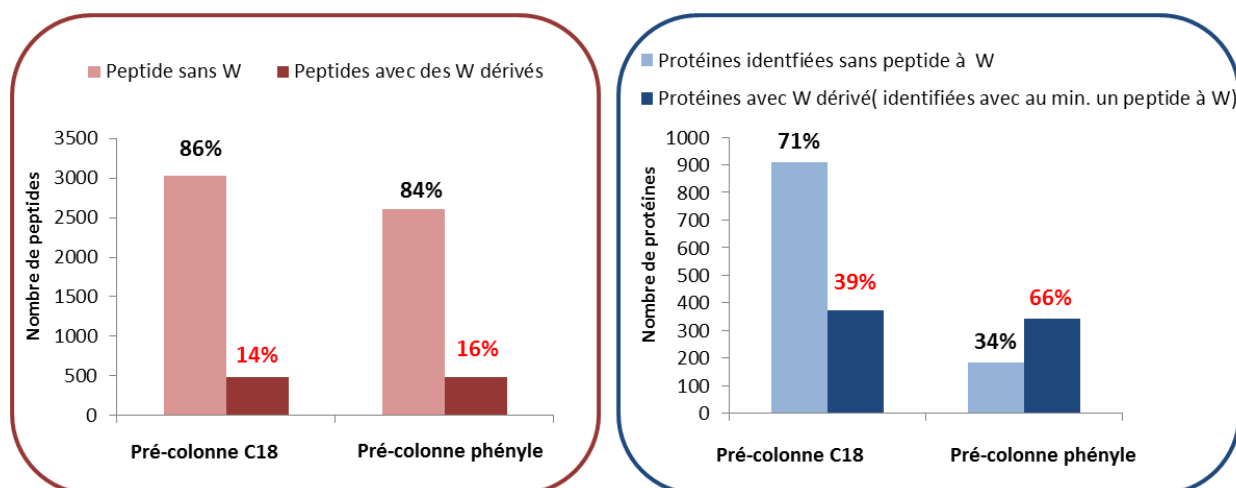


Figure 101 : Histogrammes de comparaison des peptides identifiés sans W et avec W dérivés et des protéines identifiées avec ces peptides, avec une étape d'enrichissement (pré-colonne phényle) et avec la stratégie classique (C18).

L'utilisation d'une pré-colonne phényle a montré une baisse du nombre de peptides totaux détectés (3028 avec la pré-colonne C18 vs 2602 avec la pré-colonne phényle) tout en piégeant efficacement les peptides portant l'étiquette NPBS. Cet « appauvrissement » en peptides non dérivés était un des objectifs fixés pour notre stratégie. L'appauvrissement en peptides non dérivés doit être encore amélioré si nous voulons décomplexifier le digestat et augmenter la probabilité de détection des peptides dérivés.

D'autres catégories de colonnes de piégeage peuvent être testées en modifiant l'affinité de la phase stationnaire, par exemple avec d'autres colonnes d'interactions hydrophobes, tout en restant compatible avec un système nanoLC-MS/MS classique.

D.3. Conclusion de la stratégie de marquage chimique des W

L'objectif de cette étude était de revisiter la stratégie de marquage basée sur le marquage à l'aide du NPBSCL décrite initialement par K. Tanaka puis optimisée par E. Matsuo et son équipe¹⁶⁰. Grâce à l'étude détaillée de plusieurs paramètres (chimiosélectivité, efficacité de marquage, ...) nous avons mis au point une stratégie de marquage des tryptophanes au NPBS pour l'étude d'une protéine modèle puis d'un protéome complexe pour des analyses nanoLC-MS/MS.

L'exploitation des données obtenues avec les conditions expérimentales sélectionnées sur un protéome complexe a montré l'efficacité de marquage avec ce réactif et la reproductibilité en termes de peptides identifiés dérivés au NPBS.

Les essais préliminaires avec une pré-colonne de piégeage phényle pour séparer les peptides dérivés du reste du digestat, ont mis en évidence la reproductibilité des mesures XIC, par rapport à la pré-colonne de piégeage C18. La décomplexification, peut encore être améliorée car de nombreux peptides non dérivés sont encore identifiés. Cette étape pourra être optimisée en utilisant d'autres types de greffage pour moduler l'affinité de la colonne pour les peptides dérivés NPBS. Par exemple, des colonnes de silice greffée avec des groupements de type naphtyl (avantage de la présence de deux cycles aromatiques) pourraient être plus efficace pour piéger plus efficacement les peptides dérivés par le NPBS.

Dans la littérature la stratégie de marquage était décrite avec une étape de digestion liquide¹⁶⁰ pour contourner les fluctuations liées aux stratégies de gel d'électrophorèse. Cependant, la préparation d'échantillons par des approches en solution ou en gel, a une forte influence sur la précision et la robustesse

des méthodes de protéomique quantitative. Malgré l'aspect chronophage, le gel 1D-PAGE présente l'avantage d'être compatible avec les détergents (tels que la SDS) pour la solubilisation des protéines. Récemment, la stratégie des tubes-gel¹⁶⁷, qui permet de piéger les protéines dans un gel de polyacrylamide sans électrophorèse, a été développée au laboratoire. Rapide et robuste, cette technique pourra être testée pour la préparation d'échantillons.

La prochaine étape consiste à utiliser le marquage isotopique (NPBS lourd/léger, disponible commercialement) pour quantifier de manière relative les protéines présentes dans deux échantillons.

E. Le marquage des tyrosines

E.1. L'état de l'art du marquage chimique des tyrosines

Contrairement aux méthodes de marquage des cystéines, les approches de dérivation des tyrosines (Tyr ou Y) ciblant leur chaîne latérale aromatique sont très peu décrites dans la littérature¹⁶⁸⁻¹⁶⁹. Avec une fréquence d'abondance de plus de 3% (en fonction des organismes) et de par leur propriété « redox », les tyrosines et leurs PTM¹⁷⁰⁻¹⁷¹, sont des cofacteurs essentiels impliqués dans divers processus biologiques. Pour étudier ces processus, plusieurs méthodes de marquages chimiques ont été décrites¹⁷². Parmi celles-ci, des dérivations du groupement phénol des tyrosines par des imines, des diazocarboxylates ont été reportées^{173,174}.

La réactivité du noyau aromatique (nucléophile) est utilisée depuis longtemps avec les sels de diazonium pour produire des colorants diazoïques. A l'exception de quelques rares études décrites dans la littérature, comme celle du protéome mitochondrial *in vivo*,¹⁷⁵ aucune approche de protéomique quantitative basée sur la dérivation chimique des tyrosines n'est utilisée en routine par les protéomistes.

Les dérivés phényldiazoniums sont des réactifs bien connus dans la chimie des colorants. Le dérivé appelé Fast Red (4-Chloro-2-methylbenzenediazonium) par exemple, est commercialisé en tant que réactif pour les colorations histochimiques. Nous avons donc exploré une stratégie de marquage au niveau protéique avec le FastRed.

Les dérivés PTAD (4-phenyl1,2,4-triazoline-3,5-dione) sont une autre catégorie de réactifs, dont la réactivité sélective pour les tyrosine au niveau peptidique, a été décrite par Ban^{169, 176}. Il a été montré qu'en milieu aqueux la réaction était chimiospécifique et l'adduit formé était stable. Cependant la dérivation du PTAD au niveau protéique n'a pas été étudié. En parallèle du Fast Red nous avons donc testé la réactivité du PTAD au niveau du marquage chimique d'une protéine modèle.

Les paramètres tels que la chimiospécificité, le profil isotopique, l'efficacité de marquage et la séparation chromatographique des peptides dérivés ont été étudiés.

E.2. Résultats et discussion

E.2.1 Les dérivés de Phényldiazonium

Les dérivés phényldiazoniums tel que le Fast Red réagissent par substitution électrophile avec les phénols des chaînes latérales des tyrosines (Tyr). Pour des raisons de mésomérie (délocalisation des électrons), le couplage a lieu en position *ortho* du groupement OH lorsque la position *para* est substituée comme cela est présenté Figure 102.

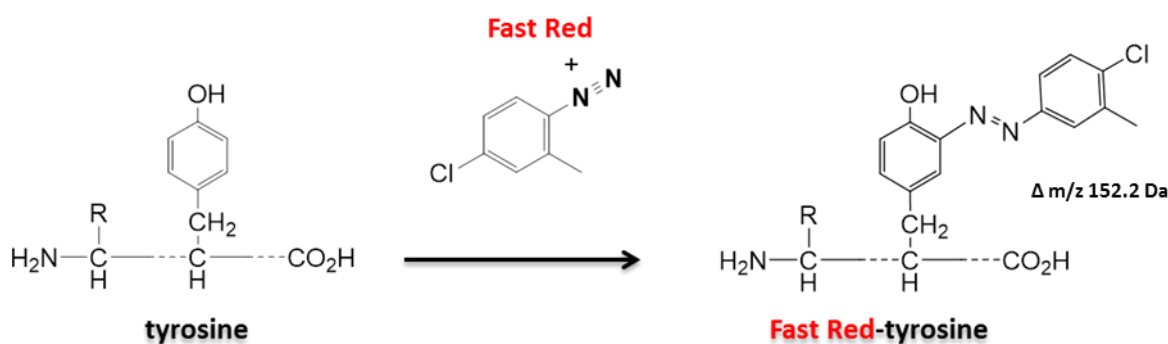


Figure 102: Représentation schématique de la réaction chimique entre le phénol de la tyrosine et réactif de marquage appelé Fast Red (4-Chloro-2-methylbenzenediazonium).

Nous avons choisi ce réactif pour la réactivité du diazonium et pour la présence d'un chlore dans sa structure. Le chlore est présent naturellement sous forme de deux isotopes. L'abondance relative des isotopes du chlore est approximativement 75% pour ^{35}Cl et 25% pour ^{37}Cl . L'abondance des deux isotopes permet de générer un profil isotopique (P, P+1, P+2) en MS avec des abondances relatives caractéristiques de la présence de chlore. La présence d'un doublet représenterait donc un moyen de sélectionner les peptides dérivés par ce réactif halogéné.

Nous avons donc développé une approche de marquage chimique des tyrosines par le réactif Fast Red au niveau protéique. Pour mettre au point les conditions de marquage, le pH (7.5, 8.9) et le tampon (Urée, éthanolamine ou SDS) ont été testés.

Les résultats présentés ci-dessous sont ceux obtenus avec les conditions expérimentales qui ont permises d'obtenir les meilleurs résultats de marquage. La conalbumine a été solubilisée dans un tampon à pH 9 (urée 8M, éthanolamine, HCl). Après ajout d'une solution de Fast Red à 20mM (0°C pendant 30 minutes), la protéine a été déposée sur gel 1D-PAGE et analysée par nanoLC-MS/MS.

Dans ces essais préliminaires, nous nous sommes essentiellement focalisés sur la chimiosélectivité et le profil isotopique des peptides dérivés par le Fast Red.

Résultats d'analyses nanoLC-MS/MS.

La chimiospécificité. Les résultats ont montré que toutes les tyrosines identifiées n'étaient pas dérivées par le Fast Red. En effet, seulement 11 peptides à tyrosines ont été identifiés dérivés sur les 17 détectées dans les contrôles.

Nous avons estimé que la dérivation des tyrosines n'est que de l'ordre de 10 % (comparaison des intensités des peptides non dérivés dans le contrôle et dans l'analyse de marquage) donc une efficacité de marquage faible.

Le profil isotopique. Les abondances relatives des ions P, P+1, P+2 des peptides dérivés au Fast Red sont caractéristiques de la présence d'un chlore. La comparaison du profil isotopique d'un peptide ANVMDYR et de son homologue dérivé a mis en évidence cette caractéristique comme cela est présenté Figure 103.

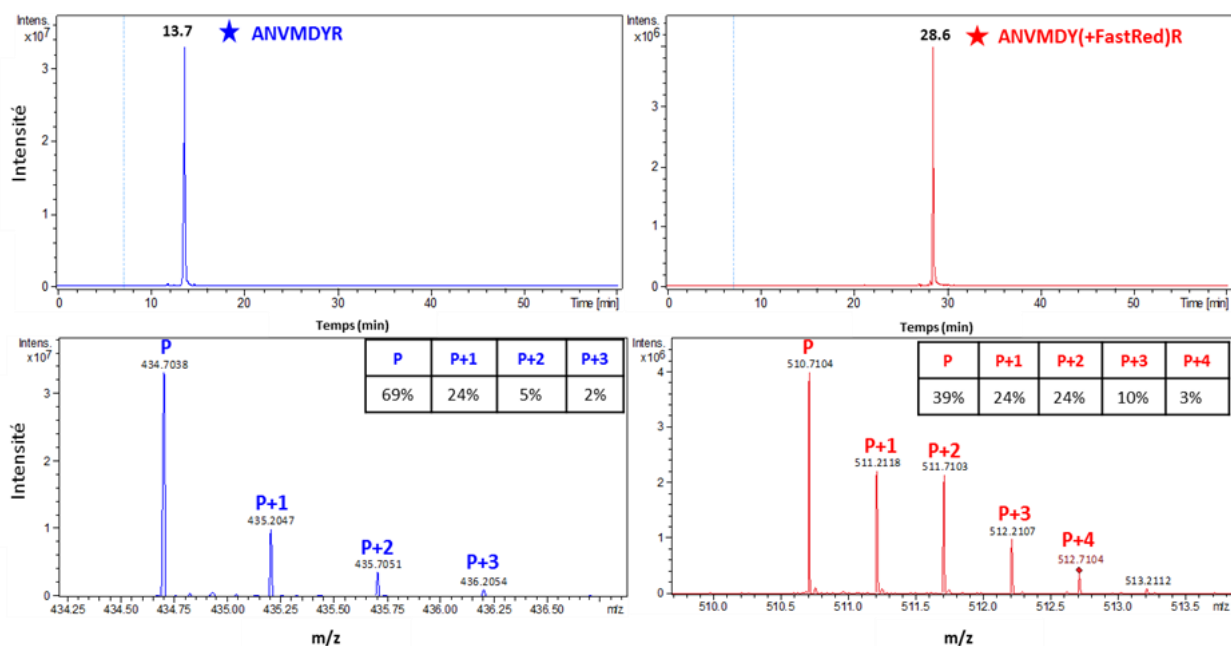


Figure 103: Comparaison des courants d'ions (XIC) des peptides ANVMDYR en bleu et ANVMDY(+FastRed)R en rouge. Le profil isotopique des deux ions peptidiques correspondant est présenté en dessous avec la proportion relative pour chaque pic P, P+1, P+2...

Ce type de profil isotopique peut permettre de détecter spécifiquement les peptides dérivés, ce qui est un avantage pour des stratégies de quantification.

La séparation chromatographique des peptides dérivés « Fast Red ». Les résultats obtenus ont montré l'enrichissement d'une zone du chromatogramme correspondant à la zone d'éluion des peptides hydrophobes. L'extraction des courants d'ions des peptides contenant des tyrosines a mis en évidence un décalage des temps de rétention des peptides dérivés Fast Red, de dix minutes en moyenne avec un gradient d'éluion court (50 minutes). A titre d'exemple, la Figure 103 montre une différence d'environ 15 minutes entre l'éluion du peptide ANVMDYR avec et sans étiquette.

Conclusion sur le marquage des tyrosines par le Fast Red.

La dérivation des tyrosines par le Fast Red a répondu à la plupart des objectifs fixés pour un réactif de marquage chimique à savoir une dérivation sélective des Tyr, un profil isotopique caractéristique et une séparation chromatographique des peptides dérivés. Dans les conditions testées, l'efficacité de marquage est encore faible mais elle pourra être améliorée par l'utilisation de différentes conditions expérimentales. Le Fast Red reste un bon candidat pour une stratégie de marquage, qui jusque-là n'avait été testé qu'au niveau peptidique.

En parallèle de ces composés, nous avons exploré la piste d'une série de dérivés : les phényltriazolédione dont la réactivité sélective pour les tyrosines a déjà été décrite. Notre attention s'est particulièrement portée sur un dérivé substitué en *para* par du Brome.

E.2.2 Phényltriazolédione

La réaction de dérivation des tyrosines par les dérivés de type PTAD est présentée Figure 104.

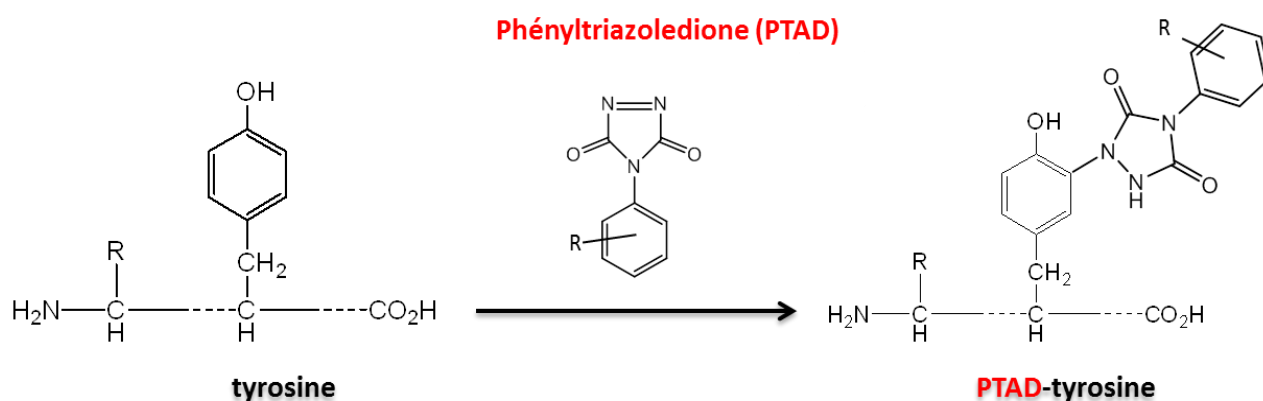


Figure 104: Représentation schématique du marquage chimique entre la fonction phénol de la tyrosine et le réactif de marquage de type PTAD.

Initialement le choix s'était orienté sur un dérivé PTAD bromé (R=Br), en position *para* du phényle, non décrit dans la littérature mais qui permettait, grâce à la présence du Brome (Br), d'obtenir un profil isotopique caractéristique, donc une signature des peptides dérivés. Cependant les premiers résultats lors de la mise au point des conditions, ont mis en évidence un manque de réactivité du PTAD-Br. Le Br en position *para*, désactiverait la réactivité nucléophile du PTAD. Nous avons ensuite voulu vérifier la réactivité de marquage du réactif PTAD non substitué (pas de groupement R).

Pour mettre au point les conditions de marquage, les mêmes protocoles expérimentaux (tampons de solubilisation et pH) que ceux utilisés pour le marquage chimique avec le Fast Red ont été testés. La protéine a été déposée sur gel 1D-PAGE et analysée par nano-LC-MS/MS.

Les résultats les plus probants ont été obtenus avec un tampon SDS à pH 8.

Les critères évalués sont la chimiospécificité des réactifs pour les tyrosines et l'efficacité de marquage.

Résultats du marquage des tyrosines avec le PTAD

La chimiosélectivité. Le marquage avec le PTAD a permis d'identifier 16 tyrosines dérivées sur les 21 identifiées dans le contrôle, ce qui montre une meilleure réactivité du PTAD par rapport au FastRed.

L'interprétation des données nanoLC-MS/MS du digestat de la conalbumine a permis de mettre en évidence que le marquage chimique ne fait pas perdre d'information car plus de 80 % de recouvrement de séquence ont été obtenu après marquage. La digestion trypsique de la conalbumine génère des peptides contenant jusqu'à trois tyrosines qui seront très difficiles à identifier (la masse ajoutée par trois marquages sur un même peptide ou pour des raisons d'encombrement stérique au niveau de la réaction chimique).

L'efficacité de marquage. Les résultats obtenus ont montré une réaction de marquage incomplète. La comparaison des courants d'ions extraits, des peptides non dérivés avant et après le marquage, a mis en évidence une efficacité de marquage faible. A titre d'exemple, l'efficacité de marquage au PTAD du peptide DEYELL(+AA)LDGSR est présentée Figure 105.

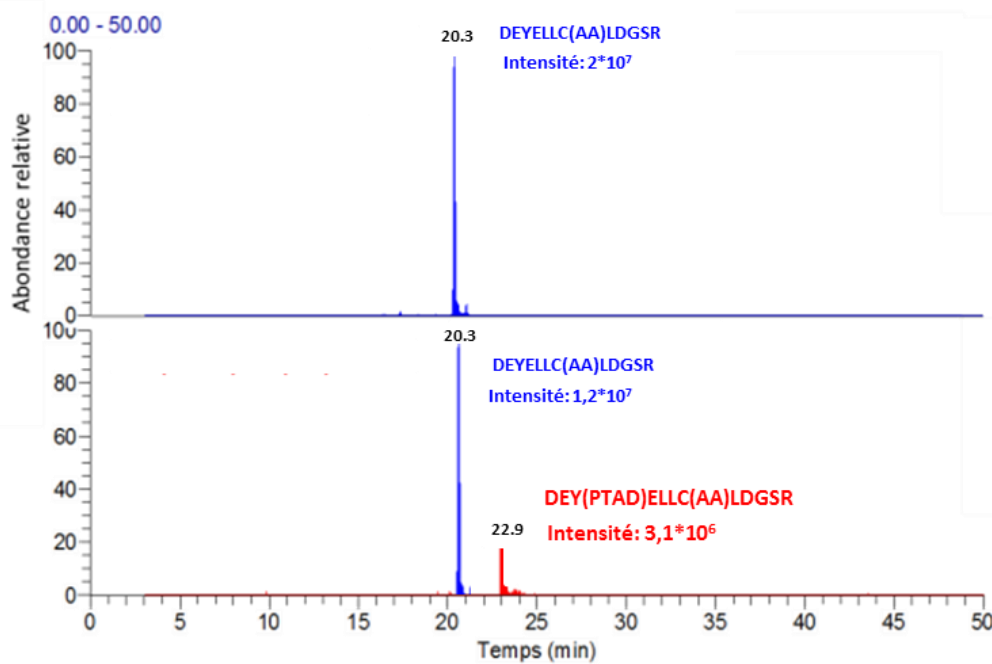


Figure 105: Comparaison des courants d'ions extraits (XIC) des peptides tryptiques de la conalbumine DEYELLC(+AA)LDGSR en bleu et DEY(+PTAD)ELLC(AA)LDGSR en rouge.

Les résultats obtenus ont également mis en évidence un décalage des temps de rétention des peptides dérivés NPBS d'environ trois minutes avec notre gradient d'élution (50 minutes). Cette différence des temps de rétention ne permet pas de séparer efficacement les peptides dérivés du reste du digestat.

Conclusion sur le marquage des tyrosines par le PTAD.

La dérivation chimique des tyrosines par le PTAD a mis en évidence une meilleure chimiospécificité par rapport au Fast Red avec les Tyr au niveau protéique. Ce réactif est un bon candidat pour une stratégie de marquage chimique et par sa chimiosélectivité pour les tyrosines. Cette réactivité pourrait être combinée avec un réactif tel que le TMPP ou un ammonium quaternaire (QAT) dans le but d'exploiter d'une part le ciblage spécifique des tyrosines et d'autre part l'avantage conféré par une charge permanente.

F. Conclusion biologique de l'étude

Parmi les expériences qui ont été présentées dans ce chapitre, la stratégie de marquage chimique des tryptophanes, revisitée, apparaît clairement comme la stratégie la plus aboutie dans des essais de quantification de toutes les protéines identifiées dans le protéome d'un échantillon biologique complexe tel qu'un extrait total d'*E. coli*.

Les avantages mis en évidence lors de l'étude poussée des paramètres tels que la chimiospécificité, l'efficacité de rendement et la robustesse en font une méthode prometteuse pour une stratégie de quantification d'une sous-population de peptides. Peu coûteuse en matériel et en temps, elle pourra être intégrée dans une stratégie d'analyse protéomique classique et donc être appliquée en routine par les protéomistes pour l'étude d'un grand nombre d'échantillons.

L'étape d'enrichissement reste encore à être optimisée pour gagner en robustesse, mais la stratégie est d'ores et déjà tout à fait compatible pour la quantification des protéines avec l'utilisation d'isotopes lourds.

Cette étape pourra être appliquée à l'étude de « vrais » échantillons biologiques, en toxicoprotéomique par exemple.

G. Conclusion sur les aspects analytiques

Notre objectif a donc été de développer une méthode capable d'évaluer rapidement la dynamique des protéomes, offrant ainsi une vue d'ensemble de la variation des protéines en réponse à des stimuli biologiques. Bien qu'elles ne soient pas nouvelles, ces approches de dérivations spécifiques d'acides aminés sont souvent parasitées, avec des obstacles techniques (absence de spécificité chimique, la faible reproductibilité) et plus particulièrement lorsque le marquage chimique est effectué sur des mélanges complexes. Nous avons donc pu apporter dans ce travail d'importantes améliorations dans ce domaine.

Marquages des cystéines. L'étude des marquages avec les dérivés APN-TMPP et QAT a mis en évidence la dérivaison chimiosélective des cystéines avec une bonne efficacité de marquage. Malgré la formation de diastéréoisomères, les dérivés APN-TMPP ont permis de séparer de manière efficace les peptides dérivés du reste du digestat augmentant ainsi leur probabilité de détection. L'utilisation des dérivés APN-TMPP n'est pas envisageable pour développer une stratégie de quantification des protéines. Cependant ces dérivés représentent des candidats prometteurs pour des stratégies de « *proteom discovery* » ou pour la caractérisation de protéines recombinantes, notamment pour le greffage des molécules sur les anticorps dans le cadre d'ADC (antibody-drug conjugate).

Marquages des tryptophanes. La stratégie re-visitée de marquage des tryptophanes est la méthode la plus aboutie de notre projet. Etape par étape, les paramètres LC, MS et MS/MS ont été étudiés et évalués pour l'étude d'un protéome complexe, ce qui n'est généralement pas le cas dans le développement de stratégie quantitative. Les résultats ont mis en évidence la robustesse de la méthode qui permettra d'étudier les grands changements du niveau d'expression des protéines en utilisant le marquage isotopique.

Marquages des tyrosines. Après des nombreux essais réalisés les résultats obtenus ont bien montré toute la difficulté d'appliquer les outils de bioconjugaison de la littérature pour la dérivaison chimique des protéines avec les contraintes liées à l'analyse protéomique. Le réactif testé PTAD, apparait comme un candidat prometteur pour la dérivaison chimiospécifique des tyrosines. Décrits pour sa forte réactivité, il pourra être combiné avec d'autres réactifs de marquages (TMPP, ammoniums quaternaires) pour atteindre les objectifs attendus d'un réactif de marquage chimique

Ce travail nous a permis de dégager de nombreuses voies originales pour le développement de nouvelles méthodes de quantification en protéomique. La plupart des voies explorées demandent encore à être optimisées, mais certaines d'entre elles, telle que le marquage des tryptophanes peuvent déjà être considérées comme opérationnelles.

Nous avons montré qu'une étude détaillée des aspects chimiques ouvre la porte à une grande variété d'outils de quantification protéomique qui pourront être utilisés en fonction du type d'études à mener. Chacune d'entre elles sera sélectionnée en fonction des objectifs et de la nature particulière des échantillons étudiés, ainsi que des attentes en termes de robustesse, de dynamique, de précision, et de débit.

CONCLUSION GENERALE

CONCLUSION GENERALE

Le fil conducteur de mon travail a été l'identification, mais aussi la quantification de protéines d'un protéome, et ceci avec un maximum de finesse structurale.

Nous avons montré dans les deux premiers chapitres qu'il est souvent peu utile de produire de longues listes de protéines identifiées sans apporter des données semi-quantitatives. Ceci est vrai même lorsque le sujet de l'étude ne porte pas a priori sur la quantification de protéines. La raison de cette nécessité de donner une information quantitative pour toute protéine identifiée dans un protéome vient des progrès considérables en sensibilité grâce aux performances des spectromètres de masse, mais aussi des logiciels de bioinformatique qui exploitent les données de LC-MS/MS. Ces progrès permettent d'avoir une dynamique de près de 10 000 entre les protéines les plus abondantes identifiées dans un protéome et les protéines identifiées les moins abondantes.

Nous avons été confronté à trois problématiques qui nécessitaient toutes d'améliorer les possibilités de l'analyse protéomique dans plusieurs directions, et ceci principalement :

- Dans une étude sur la mélanogénèse, nous avons mis en place des stratégies analytiques et des protocoles de préparation des échantillons qui ont permis d'évaluer la qualité de l'enrichissement de fractions immunopurifiées de complexes protéiques et de déterminer ainsi rapidement quelles étaient les protéines majeures pouvant éventuellement être considérées comme des partenaires du complexe d'intérêt.
- Dans une étude du mécanisme d'activation d'une protéase importante de l'immunité innée, nous avons mis au point une stratégie de marquage spécifique des positions N-terminales des protéines et avons déterminé très précisément les séquences des différents fragments générés lors des clivages. L'action protéolytique des enzymes de plusieurs pathogènes a pu ainsi être très précisément décrite. L'hypothèse du rôle clé de la protéase pathogène subtilisine pour l'activation de la protéine Perséphone dans la voie du récepteur Toll a pu être clairement démontrée.
- Dans une étude portant sur de nouvelles méthodes de quantification protéomique pour la mise en évidence de fluctuations fortes du protéome d'un système biologique nous avons décrit plusieurs nouveaux réactifs de marquage sur le tryptophane, la cystéine et la tyrosine. Ces méthodes ont le potentiel de permettre des études à haut débit tout en étant très robustes et pourraient être appliquées à de nombreuses études pour lesquelles seuls de forts changements métaboliques présentent un intérêt, mais doivent être recherchés sur un très grand nombre d'échantillons.

Notre travail a illustré l'intérêt de développer de nouveaux modes de marquages chimiques. Ceci dans le but d'aller vers des méthodes de quantification mieux adaptées à la diversité des études en protéomique, ce dont la biologie a besoin aujourd'hui. Les réactifs de marquage que nous avons étudiés correspondent à des familles chimiques dont les propriétés peuvent maintenant être modulées de nombreuses façons. Ces propriétés permettront d'adapter au mieux les études de quantifications protéomiques à la nature et aux contraintes des études biologiques à mener.

PARTIE IV PARITE EXPERIMENTALE

Partie III PARTIE EXPERIMENTALE

Chapitre I Etudes de complexes protéiques cibles impliqués dans le mécanisme de la mélanogénèse

A. L'immunoprécipitation (IP) dirigée contre TYR et DCT

A.1. La préparation des échantillons

Les cultures cellulaires MNT1 ont été faites par les collaborateurs biologistes avec le protocole du kit de préparation de Thermo Scientific™ : Pierce™ Direct Magnetic IP/Co-IP.

Très brièvement, les cellules ont été lysées dans un tampon de lyse (pH 7,4, Tris à 0,025 M, NaCl à 0,15M, EDTA à 0,001 M, 1% de NP40 et 5% glycérol). Le lysat protéique a été ensuite centrifugé à 700 g pendant 10 minutes. Le surnageant collecté a été resolubilisé dans 500 µL de tampon de lyse puis centrifugé à 10000 g pendant 10 minutes. Le surnageant protéique a ensuite été incubé avec les billes magnétiques couplées aux Ac (anti TYR ou anti DCT). Après incubation, les billes ont été collectées avec un support magnétique. Elles ont été lavées avec 500 µL de tampon aqueux à pH 2.

La concentration de protéine a ensuite été estimée par un dosage de Bradford (Protein Assay Dye Reagent Concentrate, (réf 500-0006, Bio-Rad). Une gamme d'étalonnage avec de l'albumine (Bovine) a été réalisée (0- 10 µg/µL).

A.2. La préparation du gel 12 % SDS-PAGE

Un gel de « concentration », réticulé à 5%, (0,15 M Tris, pH 6,8, SDS 0,1%, Temed 0,025% et Persulfate NH₄ 0,1%) et un gel de « séparation », réticulé à 12%, (0,37 M, Tris pH 8,8, SDS 0,1%, Temed 0,1% et Persulfate NH₄ 0,025%) ont été préparés.

L'extrait protéique et la fraction IP ont été resolubilisés dans un tampon Laemmli à 10 mM (Tris pH 8, EDTA 1mM, βME 5%, SDS 5%, glycérol 10% et bleu de bromophénol 0,10%). 20 µg de protéines de l'échantillon de l'extrait protéique, 3,2 µg de la fraction IP et 10 µL d'une solution de marqueurs de masses moléculaires Precision Plus Protein™ Dual Color (Bio-Rad) ont été déposés sur le gel.

La migration des échantillons déposés a été effectuée à l'aide d'un MiniProtean III (Bio-Rad, Munich Allemagne) sous voltage constant. Les conditions électrophorétiques ont été fixées à 50 V pendant 60 minutes puis le voltage a été augmenté à 120 V pendant 20 minutes.

Les protéines dans le gel ont été fixées pendant 30 minutes sous agitation (3% H₃PO₄, 50% CH₃CH₂OH) puis colorées pendant 6 heures dans une solution à base de bleu de Coomassie (0,1% Coomassie R250, 10% CH₃COOH, 40% CH₃OH).

A.3. La digestion des protéines « in gel »

Les bandes de gel extrudées ont été décolorées avec une solution tampon (CH₃CN /NH₄HCO₃, 75/25 v/v, 25 mM), puis déshydratées avec 50 µL de CH₃CN. Les ponts dissulfures des protéines ont ensuite été réduit avec du dithiothréitol (10 mM), 30 minutes à 60°C et 30 minutes à température ambiante (TA) puis alkylés avec de l'iodoacétamide (55 mM) pendant 20 minutes à TA dans l'obscurité. Les bandes de gel ont été lavées (3 fois) avec une solution tampon (NH₄HCO₃/ CH₃CN, 50/50 v/v, 25 mM), puis déshydratées avec 50 µL de CH₃CN.

Les protéines ont été digérées par une solution de trypsine (Promega) à 12,5 ng/ μL (NH_4HCO_3 , 25mM), avec un rapport 1/100 (enzyme/protéine), pendant 12 heures à 37° C.

Après ajout de 25 μL d'une solution aqueuse de 80 % (v/v) CH_3CN et 0,1 % (v/v) HCOOH sous agitation pendant 90 minutes à température ambiante, les peptides tryptiques ont été extraits. La solution a été évaporée à sec. Les peptides extraits ont été solubilisés dans 100 μL d'une solution aqueuse à 0,1 % HCOOH (v/v).

A.4. L'analyse nanoLC-MS/MS

A.4.1 La séparation chromatographique

Les extraits peptidiques ont été analysés par nanoLC-MS/MS sur un système chromatographique nanoACQUITY Ultra-Performance-LC (Waters). Les peptides ont été piégés sur une colonne de piégeage de 20 x 0.18 mm, 5 μm Symmetry C18 (Waters Corp.), pendant 3 minutes à 5 $\mu\text{L}/\text{min}$ avec 99% de solvant A (0.1% de HCOOH dans H_2O) et 1% de solvant B (0.1% HCOOH dans le CH_3CN). Les peptides ont ensuite été séparés avec un débit de 450 nL/min, avec un gradient de 1-35% de solvant B pendant 79 min à 60°C sur une colonne chromatographique Acquity UPLC® BEH130 C18 (Waters Corp., 75 μm x 250 mm, 1.7 μm de taille de particule).

A.4.2 L'acquisition des données MS-MS/MS

Le système de chromatographie liquide est couplé à un spectromètre de masse de type Orbitrap avec un Q Exactive + (Thermo Scientific™). Une tension de 1,8 kV est appliquée au capillaire de l'électrospray. La température de la source NSI (NanoSpray Ionisation) est de 250°C. Les données sont acquises en mode DDA avec une bascule automatique entre le mode MS (50 ms/scan sur une gamme de masse de 300 à 1800 m/z et une résolution à 70000) et le mode MS/MS (100 ms/scan sur une gamme de masse de 200 à 2000 m/z et une résolution à 17500). Les dix ions les plus intenses sont sélectionnés pour la MS/MS avec un seuil d'intensité fixé à 2×10^5 . Les ions monochargés sont exclus et ne sont donc pas sélectionnés pour la MS/MS. Une exclusion dynamique des ions est fixée à 60 secondes. L'acquisition des données est réalisée à l'aide du logiciel Xcalibur v.3.0.63 (Thermo Fisher Scientific™).

A.4.3 L'interprétation et validation des données acquises par LC-MS/MS

Le traitement des données de MS/MS est réalisé à l'aide de l'algorithme de recherche Mascot v2.3.02 (Matrix Science, London, UK) dans une banque de données de protéines Swissprot (SP) *Homo sapiens* (taxonomie 9606, 20305 entrées). Les paramètres de recherche utilisés sont : une tolérance sur la masse de l'ion précurseur de 5 ppm et de 0,07 Da sur celle des ions fragments. Le nombre de sites de clivage manqué toléré par la trypsine est fixé à 1. La carbamidométhylation des cystéines (+57 Da) et l'oxydation des méthionines (+16 Da) des protéines ont été définies comme modifications variables possibles.

Les données ont été validées avec le logiciel Proline⁴² (ProlineStudio Release 1.4). Les résultats ont été filtrés avec un FDR inférieur ou égal à 1% au niveau peptidique.

B. La stratégie de réticulation des protéines au DSP

B.1. La préparation des échantillons

Les préparations de cultures cellulaires MNT1 réticulées ont été réalisées par les collaborateurs biologistes avec le protocole du kit de préparation de Thermo Scientific™ : DTSSP KIT (3,3'-dithiobis (sulfosuccinimidylpropionate)).

Très brièvement, les cellules MNT1 ont été récoltées (19×10^6) et remises en suspension dans 20mL de tampon PBS (Phosphate buffered saline) avant d'être centrifugées à 700g pendant 10minutes. Le culot protéique obtenu a été suspendu dans du tampon phosphate (phosphate à 0,1M, NaCl à 0,15M; pH 7,2). Puis le réactif de réticulation DSP a été ajouté pour obtenir une concentration finale de 2 mM, à température ambiante. Au bout de 30 minutes d'incubation, les cellules ont été rincées avec 2mM de d'une solution de glycine (1M, pH7,5). Puis 750 μ L d'un tampon de lyse (Tris-HCl 10 mM, 2,2% SDS, 1% Triton X-100, 10 % Glycérol, pH7) ont été ajoutés. Les protéines ont été soniquées (2 à 3 pulse pendant 3 minutes) puis dénaturées par chauffage (90°C pendant 2 minutes). Le lysat protéique a ensuite été centrifugé à 14000g pendant 10 minutes. Le surnageant protéique a été collecté.

La concentration de protéine a ensuite été estimée un dosage de Bradford (Protein Assay Dye Reagent Concentrate, réf 500-0006, Bio-Rad). Une gamme d'étalonnage avec de l'albumine (Bovine) a été réalisée (0-10 μ g/ μ L).

B.2. La préparation du gel 12 % SDS-PAGE

Le protocole de préparation du gel 1D-PAGE de polyacrylamide est en tout point le même que celui décrit dans la partie ci-dessus (cf Partie III -A.2).

25 μ g de protéines et 10 μ L d'une solution de marqueurs de masse moléculaires Precision Plus Protein™ Dual Color (Bio-Rad) ont été déposés sur le gel.

La migration des échantillons déposés a été effectuée à l'aide d'un MiniProtean III (Bio-Rad, Munich Allemagne) sous voltage constant Les conditions électrophorétiques ont été fixées à 50 V pendant 60 minutes puis le voltage a été augmenté à 120 V pendant 120 minutes.

Les protéines dans le gel ont été fixées pendant 30 minutes sous agitation (3% H₃PO₄, 50% CH₃CH₂OH) puis colorées pendant 6 heures dans une solution de coloration à base de bleu de Coomassie (0,1% Coomassie R250, 10% CH₃COOH, 40% CH₃OH).

B.3. La digestion des protéines « in gel »

Le protocole de digestion des protéines est en tout point identique à celui décrit ci-dessus (cf.Partie III -A.3)

B.4. L'analyse nanoLC-MS/MS

B.4.1 La séparation chromatographique

Les paramètres du système de piégeage et de séparation des peptides sont en tout point les mêmes que ceux décrits dans la Partie III -A.4.1. Après avoir été piégés, les peptides ont été élués sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-35% de solvant B pendant 120 minutes à 60°C.

B.4.2 L'acquisition des données MS-MS/MS

Les paramètres d'acquisition des données MS-MS/MS sont en tout point les mêmes que ceux décrits dans la Partie IIIA.4.2A.4.2.

B.4.3 L'interprétation des données acquises par LC-MS/MS

Le traitement des données de MS/MS est en tout point identique à celui décrit dans la partie Partie III -A.4.3. La carbamidométhylation des cystéines (+57 Da), l'oxydation des méthionines (+16 Da) et la modification chimique du DSP réduit sur les lysines (+145 Da) des protéines ont été définies comme modifications variables possibles.

Les données ont été validées avec le logiciel Proline⁴² (ProlineStudio Release 1.4). Les résultats ont été filtrés avec un FDR inférieur ou égal à 1% au niveau peptidique.

C. La stratégie de marquage à la Biotine « BioID »

C.1. La préparation des échantillons

Les cultures cellulaires MNT1 ont été faites par les collaborateurs biologistes avec le protocole décrit par K. Roux¹⁰⁶ en 2013.

Très brièvement, les cellules MNT1 ont été récoltées ($2,6 \cdot 10^5$) et transfectées avec un plasmide viral exprimant la séquence de la protéine TYR-BirA (MOI de 40, polybrène 2 μ g/ μ L). Ensuite, les cellules ont été exposées à 50 μ M de biotine (concentration finale) pour stimuler la biotinylation pendant 12 heures avant la récolte. Les cellules ont été lavées avec dans 20mL de PBS avant d'être centrifugées à 700g pendant 10min. Le culot protéique obtenu a été suspendue dans 100 μ L d'un tampon de lyse (Tris-HCl 10 mM pH7,2 , 2% SDS, 1% Triton X-100, 10 % Glycérol). Ensuite, le lysat a été soniqué (2 à 3 pulse pendant 3 minutes) puis chauffé (90°C pendant 2 minutes) avant d'être centrifugé à 14000g pendant 10 minutes. Le surnageant protéique a été collecté.

La concentration de protéine a ensuite un dosage de Bradford (Protein Assay Dye Reagent Concentrate, réf 500-0006, Bio-Rad). Une gamme d'étalonnage avec de l'albumine (Bovine) a été réalisée (0- 10 μ g/ μ L).

C.2. La préparation du gel 12 % SDS-PAGE

Le protocole de préparation du gel 1D-PAGE de polyacrylamide est en tout point le même que celui décrit dans la partie ci-dessus (cf Partie III -A.2).

25 μ g de protéines et 10 μ L d'une solution de marqueurs de masse moléculaires Precision Plus Protein™ Dual Color (Bio-Rad) ont été déposés sur le gel.

La migration des échantillons déposés a été effectuée à l'aide d'un MiniProtean III (Bio-Rad, Munich Allemagne) sous voltage constant Les conditions électrophorétiques ont été fixées à 50 V pendant 60 minutes. Les protéines n'ont pas été séparées, elles ont migré en une seule bande (gel « stacking ») dans le but de rendre compatible l'échantillon avec la MS et pour supprimer le détergent et les éventuels sels présents.

Les protéines dans le gel ont été fixées pendant 30 minutes sous agitation (3% H₃PO₄, 50% CH₃CH₂OH) puis colorées pendant 6H dans une solution de coloration à base de bleu de Coomassie (0,1% Coomassie R250, 10% CH₃COOH, 40% CH₃OH).

C.3. La digestion des protéines « in gel »

Le protocole de digestion des protéines est en tout point identique à celui décrit dans la Partie III -A.3.

C.4. L'analyse nanoLC-MS/MS

C.4.1 La séparation chromatographique

Les paramètres du système de piégeage et de séparation des peptides sont en tout point les mêmes que ceux décrits dans la Partie III -A.4.1. Après avoir été piégés, les peptides ont été élués sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-35% de solvant B pendant 120 minutes à 60°C.

C.4.2 L'acquisition des données MS-MS/MS

Les paramètres d'acquisition des données MS-MS/MS sont en tout point les mêmes que ceux décrits dans la Partie III -A.4.2Partie IIIA.4.2.

C.4.3 L'interprétation des données acquises par LC-MS/MS

Le traitement des données de MS/MS est en tout point identique à celui décrit dans la Partie III -A.4.3. La carbamidométhylation des cystéines (+57 Da), l'oxydation des méthionines (+16 Da) et la biotinylation des lysines (+226 Da) des protéines ont été définies comme modifications variables possibles.

Les données ont été validées avec le logiciel Proline⁴² (ProlineStudio Release 1.4). Les résultats ont été filtrés avec un FDR inférieur ou égal à 1% au niveau peptidique.

Chapitre II Etude des clivages protéolytiques d'une enzyme par marquage au TMPP

A. La préparation des échantillons

Les échantillons ont été préparés par le Dr Florian Veillard de l'équipe du Pr JM. Reichhart (Institut de Biologie Moléculaire et Cellulaire, UPR 9022 (Réponse Immunitaire et Développement chez les Insectes, Strasbourg).

Etude de la protéolyse de r-Psh par « Bacillus subtilis » (Gram+)

Les échantillons et les gels ont été préparés par les collaborateurs biologistes. 20 μL de la forme recombinante purifiée de Psh (0,4 $\mu\text{g}/\mu\text{L}$) ont été incubés à 29 °C avec 20 μL de surnageant de culture de *B. subtilis* dans du tampon Tris-HCl 0,1 M pH 8. A différents temps d'incubation (5, 15, 30, 60, 120 et 240 minutes), 5 μg de Psh ont été prélevés et la réaction est stoppée à 100 °C pendant 5 minutes en présence d'un tampon réducteur (NuPAGE LDS Sample buffer, Invitrogen). Deux contrôles ont été réalisés : r-Psh seule ou incubée pendant 240 min dans les mêmes conditions en présence de surnageant préalablement inactivé par la chaleur (10 minutes à 100 °C). Les échantillons ont ensuite été séparés par électrophorèse SDS-PAGE (NuPAGE 4-12%, Invitrogen) pendant 2 heures (100 V) et colorés avec du bleu de Coomassie (SimplyBlue Safestain, Invitrogen).

Pour les études des protéolyses de r-Psh par d'autres pathogènes le même protocole de préparation a été réalisé.

Les échantillons biologiques ont été préparés en duplicatas.

B. Le protocole de marquage au TMPP « in gel »

Les bandes de gel 1D-PAGE extrudées ont été lavées par une solution tampon ($\text{CH}_3\text{CN}/\text{NH}_4\text{HCO}_3$, 75/25 v/v, 25 mM), puis déshydratées avec 50 μL de CH_3CN . Les ponts disulfures des protéines ont été réduits par du tributylphosphine (5 mM) pendant une heure à température ambiante (TA) puis alkylés par l'iodoacétamide (50 mM) pendant 20 minutes à TA. Ensuite la protéine a été dérivée par 250 équivalents TMPP-Ac-Osu. 50 μL de tampon Tris-HCl (100 mM) dont le pH a été ajusté à 8,2, ont ensuite été ajoutés dans les puits. Au bout d'une heure une solution d'hydroxylamine (0,1 M) a été ajoutée. Les bandes de gel ont été lavées par la solution tampon ($\text{CH}_3\text{CN}/\text{NH}_4\text{HCO}_3$, 75/25 v/v, 25 mM), puis déshydratées avec 50 μL de CH_3CN (quatre cycles de lavages).

C. La digestion des protéines « in gel »

Les protéines ont été digérées par une solution de trypsine (Promega) à 12,5 ng/ μL (NH_4HCO_3 , 25mM), avec un rapport 1/100 (enzyme/protéine), pendant 12 heures à 37° C.

Après ajout de 25 μL d'une solution aqueuse de 80 % (v/v) CH_3CN et 0,1 % (v/v) HCOOH sous agitation pendant 90 minutes à température ambiante, les peptides tryptiques ont été extraits. La solution a été évaporée à sec. Les peptides extraits ont été repris dans 10 μL d'une solution aqueuse à 0,1 % HCOOH (v/v).

D. L'analyse nanoLC-MS/MS

D.1. La séparation chromatographique

Les paramètres du système de piégeage et de séparation des peptides sont en tout point les même que ceux décrits dans la Partie III -A.4.1. Après avoir été piégés, les peptides ont été élués sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-50% de solvant B pendant 50 minutes à 60°C.

D.2. L'acquisition des données MS-MS/MS

Le système de chromatographie liquide est couplé à un spectromètre de masse de type Orbitrap avec un Q Exactive + (Thermo Scientific™). Une tension de 1,8 kV est appliquée au capillaire de l'électrospray. La température de la source NSI (*NanoSpray Ionisation*) est de 250°C. Les données sont acquises en mode DDA avec une bascule automatique entre le mode MS (50 ms/scan sur une gamme de masse de 300 à 1800 m/z et une résolution à 70000) et le mode MS/MS (100 ms/scan sur une gamme de masse de 200 à 2000 m/z et une résolution à 17500). Les dix ions les plus intenses sont sélectionnés pour la MS/MS avec un seuil d'intensité fixé à 2×10^5 . Les ions monochargés sont exclus et ne sont donc pas sélectionnés pour la MS/MS. Une exclusion dynamique des ions est fixée à 10 secondes. L'acquisition des données est réalisée à l'aide du logiciel Xcalibur v.3.0.63 (Thermo Fisher Scientific™).

D.3. L'interprétation des données acquises par LC-MS/MS

Le traitement des données de MS/MS est réalisé à l'aide du moteur de recherche Mascot v2.3.02 (Matrix Science, London, UK) dans une banque de données de protéines Swissprot (SP) de *Drosophila melanogaster* (taxonomie 7227, 3227 entrées) dans laquelle nous avons introduit la séquence de r-Psh. Les paramètres de recherche utilisés sont : une tolérance sur la masse de l'ion précurseur de 5 ppm et de 0,07 Dalton sur celle des ions fragments. Le nombre de sites de clivage manqué toléré par la trypsine est fixé à 1. La carbamidométhylation des cystéines (+57 Da), l'oxydation des méthionines (+16 Da) et la modification induite par le TMPP (+572 Da) des protéines ont été définies comme modifications variables possibles.

Les spectres ont été vérifiés et validés manuellement à l'aide du logiciel Scaffold⁴⁴. La probabilité de correspondance entre les données expérimentales et théoriques a été fixée à *Ion Score* ≥ 0 et *Ion-Identity Score* = -40.

Chapitre III De nouveaux outils de bioconjugaison pour l'analyse protéomique par MS

A. La préparation des échantillons

Les cultures cellulaires et les marquages chimiques ont été réalisés par le Dr. Thierry Rabilloud.

A.1. *Le marquage de la protéine modèle Conalbumine*

Les préparations d'échantillon ont été réalisées en triplicatas.

A.1.1 Les échantillons contrôles

100 µg de conalbumine ont été solubilisés dans 100 µl de tampon à pH 8 (urée 8 M, Tris HCl 0,125 M). Les ponts disulfures de la protéine ont ensuite été réduits par le tris(2-carboxyethyl)phosphine ou TCEP à 10 mM pendant une heure, puis alkylée par l'iodoacétamide (IAA) à 50 mM pendant 30 minutes à 30°C. La protéine a été ensuite diluée dans du tampon Laemmli (10 mM Tris pH 8, EDTA 1 mM, β-mercaptoéthanol 5%, SDS 5%, glycérol 10% et bleu de bromophénol 0,10%) et 2 µg ont été déposés sur un gel 1D-PAGE.

A.1.2 La dérivation des cystéines

100 µg de conalbumine ont été solubilisés dans 100 µl de tampon à pH 8 (urée 8M, Tris HCl 0.125M). Les ponts disulfures de la protéine ont ensuite été réduits par le TCEP (10 mM) pendant une heure, puis marquée par soit l'APN-TMPP, solubilisée dans une solution 80 % CH₃CN et 20% H₂O (v/v) (6 mM), soit par l'ammonium quaternaire phényle (1mM), pendant 30 minutes à 30°C. La protéine a été ensuite diluée dans du tampon Laemmli et 2 µg ont été déposés sur un gel 1D-PAGE.

Les réactifs APN-TMPP testés ont été synthétisés par l'équipe du Dr A. Wagner.

Les réactifs ammoniums quaternaires testés ont été synthétisés par le Dr T. Rabilloud.

A.1.3 La dérivation des tryptophanes

100 µg de conalbumine ont été solubilisés dans 100 µl dans un tampon à pH 8 (SDS 2%, Tris HCl 0,125M) ou dans un tampon à pH 8 (urée 8M, Tris HCl 0,125M). La protéine est ensuite réduite par le TCEP (10 mM) pendant une heure, puis marquée par le NPBS (10 mM), solubilisé dans la tetraméthylurée (TMU) ou l'acide acétique (AcOH) pendant 30 minutes à 60°C. La protéine a été ensuite diluée dans du tampon Laemmli et 2 µg ont été déposés sur un gel 1D-PAGE.

A.1.4 La dérivation des tyrosines

FastRed

100 µg de conalbumine ont été solubilisés dans 100 µL de tampon à pH 9 (urée 8M, éthanolamine, HCl). La protéine est ensuite réduite par le TCEP (10 mM) pendant une heure, puis alkylée par l'IAA (50 mM) pendant 30 minutes à 30°C. La protéine est ensuite marquée par le Fast Red (20 mM) pendant 30 minutes à 0°C. La protéine a été ensuite diluée dans du tampon Laemmli et 2 µg ont été déposés sur un gel 1D-PAGE.

PTAD

100 µg de conalbumine ont été solubilisés dans 100 µl de tampon à pH 8 (SDS 2%, Tris HCl 0.125M). La protéine est ensuite réduite par le TCEP (10 mM) pendant une heure, puis alkylée par l'IAA (50 mM) pendant 30 minutes à 30°C. La protéine est ensuite marquée par le PTAD (50 mM) pendant 30 minutes à 25°C. La protéine a été ensuite diluée dans du tampon Laemmli et 2 µg ont été déposés sur un gel 1D-PAGE.

A.2. Le marquage des protéines « d'E. coli » : échantillon complexe

Les préparations d'échantillon ont été réalisées en triplicatas.

A.2.1 Les échantillons contrôles

Les protéines des cellules de cultures d'*E. coli* ont été extraites par un tampon SDS (2% acide pyruvique 100mM w/v, NaOH 50mM, pH 2.4) ou CTAC (2.5% acide pyruvique 100mM w/v, NaOH 50mM, pH 2.4). Les protéines ont ensuite été précipitées à l'éthanol, avant d'être resolubilisées dans un tampon Laemmli et 100 µg de protéines ont été déposés sur un gel 1D-PAGE.

A.2.2 La dérivation des tryptophanes

Les protéines des cellules de cultures d'*E. coli* ont été extraites par un tampon SDS (2% acide pyruvique 100mM w/v, NaOH 50mM, pH 2.4) ou CTAC (2.5% acide pyruvique 100mM w/v, NaOH 50mM, pH 2.4). Les tryptophanes des protéines ont été marqués pendant 30 minutes à 60°C par le NPBS (10 mM), solubilisé ou l'acide acétique (AcOH). Les protéines ont ensuite été précipitées à l'éthanol, avant d'être resolubilisées dans un tampon Laemmli et 100 µg de protéines ont été déposés sur un gel 1D-PAGE.

A.3. La digestion des protéines « in gel »

Les protéines ont migré jusqu'à former une bande de concentration (gel « stacking »). Les protéines dans les gels 1D-PAGE, ont été fixées pendant 30 minutes sous agitation (3% H₃PO₄, 50% CH₃CH₂OH) puis colorées dans une solution de coloration à base de bleu de Coomassie (0,1% Coomassie R250, 10% CH₃COOH, 40% CH₃OH).

Les bandes « stacking » de gel, extrudées ont été décolorées, réduites DTT (10 mM) pendant une heure, puis alkylées à l'IAA (50 mM) pendant 30 minutes à 30°C. Elles ont été lavées avec une solution tampon (NH₄HCO₃/CH₃CN, 50/50 v/v, 25 mM), puis déshydratées avec 50 µL de CH₃CN (quatre fois).

Les protéines ont été digérées par une solution de trypsine (Promega) à 12,5 ng/ µL (NH₄HCO₃, 25m M), avec un rapport 1/100 (enzyme/protéine), pendant 12 heures à 37° C.

Après ajout de 25 µL d'une solution de 80 % (v/v) CH₃CN et 0,1 % (v/v) HCOOH sous agitation pendant 90 min à température ambiante, les peptides de digestion ont été extraits des bandes de gel 1D-PAGE. Puis ils ont été évaporés.

Pour le digestat de la conalbumine, les peptides extraits ont été resolubilisés dans 20 µL d'une solution aqueuse à 0,1 % (v/v) HCOOH.

Pour le digestat d'*E. coli*, les peptides extraits ont été repris dans 200 µL d'une solution aqueuse à 0,1 % (v/v) HCOOH.

B. L'analyse nanoLC-MS/MS

B.1. La séparation chromatographique

B.1.1 L'analyse du digestat peptidique de la conalbumine

Les paramètres du système de piégeage et de séparation des peptides sont en tout point les mêmes que ceux décrits dans la Partie III -A.4.1. Après avoir été piégés, les peptides ont été élués sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-50% de solvant B pendant 50 minutes à 60°C.

B.1.2 L'analyse du digestat peptidique d'*E. coli*

B.1.2.1 La colonne de piégeage C18

Les extraits peptidiques ont été analysés par nanoLC-MS/MS sur un système chromatographique nanoACQUITY Ultra-Performance-LC (Waters). Les peptides ont été piégés sur une pré-colonne de piégeage de 20 x 0.18 mm, 5 µm **Symmetry C18** (Waters Corp.), pendant 3 minutes à 5 µL/min avec 99% de solvant A (0.1% de HCOOH dans H₂O) et 1% de solvant B (0.1% HCOOH dans le CH₃CN). Les peptides ont ensuite été séparés sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-50% de solvant B pendant 120 min à 60°C.

B.1.2.2 La colonne de piégeage Phényle

Les extraits peptidiques ont été analysés par nanoLC-MS/MS sur un système chromatographique nanoACQUITY Ultra-Performance-LC (Waters). Ils ont été piégés sur une pré-colonne de piégeage de 20 x 0.18 mm, 5 µm **XBridges Phényl** (Waters), pendant 3 minutes à 5 µL/min avec 99% de solvant A (0.1% de HCOOH dans H₂O) et 1% de solvant B (0.1% HCOOH dans le CH₃CN). Les peptides ont ensuite été séparés sur une colonne analytique avec un débit de 450 nL/min, avec un gradient de 1-50% de solvant B pendant 120 min à 60°C.

B.2. L'acquisition des données MS-MS/MS

La chromatographie liquide est couplée à un spectromètre de masse de type Orbitrap avec un Q Exactive + (Thermo Scientific™). Une tension de 1,8 kV est appliquée au capillaire de l'électrospray. La température de la source NSI (*NanoSpray Ionisation*) est de 250°C. Les données sont acquises en mode DDA avec une bascule automatique entre le mode MS (50 ms/scan sur une gamme de masse de 300 à 1800 m/z et une résolution à 70000) et le mode MS/MS (100 ms/scan sur une gamme de masse de 200 à 2000 m/z et une résolution à 17500). Les 10 ions les plus intenses sont sélectionnés pour la MS/MS pour chaque spectre MS avec un seuil d'intensité fixé à 2×10^5 .

Les ions monochargés sont exclus et ne sont donc pas sélectionnés pour la MS/MS. Une exclusion dynamique d'ions est fixée à 60 secondes. L'acquisition des données est réalisée à l'aide du logiciel Xcalibur v.3.0.63 (Thermo Fisher Scientific™).

B.3. L'interprétation des données de LC-MS/MS

B.3.1 La protéine modèle : conalbumine

Le traitement des données de MS/MS a été réalisé à l'aide du moteur de recherche Mascot v2.3.02 (Matrix Science, London, UK) dans une banque de données de protéines Swissprot (SP). Les paramètres de recherche utilisés ont une tolérance sur la masse de l'ion précurseur de 5 ppm et de 0,07 Da sur celle des ions fragments. Le nombre de site de clivage possiblement manqué par la trypsine est fixé à 1. La carbamidométhylation des cystéines (+57 Da), l'oxydation des méthionines (+16 Da) et la modification induite par le NPBS (+152 Da) des protéines ont été définies comme modifications variables possibles.

Pour rassembler et valider les identifications peptidiques obtenues, le logiciel Scaffold a été utilisé. La probabilité de correspondance entre les données expérimentales et théoriques a été fixée à *Ion Score* ≥ 0 et *Ion-Identity Score* = -40.

B.3.2 L'échantillon biologique : *E. coli*

Les données ont été traitées en utilisant le logiciel MaxQuant⁴³ (version 1.4.3.17) et son moteur de recherche intégré (Andromeda)⁴¹. Les recherches ont été effectuées dans la banque de données de protéines Swissprot (SP) de *E. Coli* (taxonomie 83333, 6216 entrées). La tolérance de masse des précurseurs a été fixée à 5 ppm et à 0,07 Da pour les ions fragments. Le nombre de site de clivage possiblement manqué par la trypsine est fixé à 1. La carbamidométhylation des cystéines (+57 Da), l'oxydation des méthionines (+16 Da) et la modification induite par le NPBS (+152 Da) ont été définies comme des modifications variables possibles. Le nombre minimal de peptides uniques a été fixé à un. Les résultats ont été filtrés avec un FDR inférieur ou égal à 1% au niveau peptidique.

Les valeurs des aires sous la courbe, extraites pour chaque peptide ont ensuite été normalisées à l'aide du package R (v3.0.3; <http://www.R-project.org>) nommé Normalyzer¹⁷⁷. La méthode de normalisation utilisée est la normalisation par quantile (*Quantile normalisation*).

REFERENCES BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

1. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004**, *5* (9), 699-711.
2. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013**, *113* (4), 2343-94.
3. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1988**, *2* (8), 151-153.
4. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
5. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B., A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196* (6), 801-10.
6. Donnan, F. G., Rays of positive electricity and their application to chemical analyses. By Sir J. J. Thomson, O. M. F.R.S. Second edition. Pp. x + 237. London: Longmans, Green and Co., 1921. Price 16s. *Journal of the Society of Chemical Industry* **1923**, *42* (36), 861-861.
7. Edman, P., A method for the determination of amino acid sequence in peptides. *Archives of biochemistry* **1949**, *22* (3), 475.
8. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198-207.
9. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular Beams of Macroions. *The Journal of Chemical Physics* **1968**, *49* (5), 2240-2249.
10. Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984**, *88* (20), 4451-4459.
11. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **2001**, *20* (6), 362-87.
12. Duft, D.; Achtzehn, T.; Muller, R.; Huber, B. A.; Leisner, T., Coulomb fission: Rayleigh jets from levitated microdroplets. *Nature* **2003**, *421* (6919), 128.
13. Dole, M.; Mack, L.; Hines, R. L.; F.C., M.; Ferguson, L. D.; Alice, M., Molecular beams of macroions. *J Chem Phys* **1968**, *49*, 2240-2249.
14. Iribarne J.V.; Thomson, B., On the evaporation of charged ions from small droplets. *J Chem Phys* **1976**, *64*, 2287-2294.
15. Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **2000**, *35* (7), 804-17.
16. Wilm, M.; Mann, M., Analytical properties of the nano-electrospray ion source. *Anal Chem* **1996**, *68* (1), 1-8.
17. Karas, M.; Bahr, U.; Dülcks, T., Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius' Journal of Analytical Chemistry* **2000**, *366* (6), 669-676.
18. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **1988**, *60* (20), 2299-2301.
19. Zenobi, R.; Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews* **1998**, *17* (5), 337-366.
20. De Hoffmann E, S. V., Mass Spectrometry: Principles and Applications. *Wiley* **2007**.
21. Dreisewerd, K., The desorption process in MALDI. *Chemical reviews* **2003**, *103* (2), 395-426.
22. Knochenmuss, R., A quantitative model of ultraviolet matrix-assisted laser desorption/ionization. *Journal of Mass Spectrometry* **2002**, *37* (8), 867-877.
23. Paul, W.; Steinwedel, h., Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift Naturforschung* **1953**, Teil A *8*, 448-450.
24. Wolff, M. M.; Stephens, W. E., A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments* **1953**, *24* (616).
25. Mamyrin, B. A., Laser assisted reflectron time-of-flight mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **1994**, *131*, 1-19.
26. Fancher, C. A.; Woods, A. S.; Cotter, R. J., Improving the sensitivity of the end-cap reflectron time-of-flight mass spectrometer. *Journal of Mass Spectrometry* **2000**, *35* (2), 157-162.
27. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **2000**, *72* (6), 1156-62.
28. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005**, *40* (4), 430-43.
29. Domon, B.; Aebersold, R., Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212-7.
30. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012**, *9* (6), 555-66.

31. Kurz, E. A., Effects of high-current input pulses upon Channeltrons. *Rev Sci Instrum* **1979**, *50* (11), 1492.
32. Audier, M.; Delmotte, J. C.; Boutot, J. P., Multiplicateur à galette de microcanaux : amélioration des performances de gain et de dynamique de détection. *Revue de Physique Appliquée* **1978**, *13* (4), 188-194.
33. De Hoffman, E.; Stroobant, V., Spectrométrie de masse Dunod.
34. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **1990**, *193*, 886-887.
35. Dongré, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H., Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am Chem Soc* **1996**, *118*, 8365-8374.
36. Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **2004**, *1* (1), 39-45.
37. Blueggel, M.; Chamrad, D.; Meyer, H. E., Bioinformatics in proteomics. *Curr Pharm Biotechnol* **2004**, *5* (1), 79-88.
38. Pappin, D. J.; Hojrup, P.; Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **1993**, *3* (6), 327-32.
39. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-67.
40. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H., Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3* (5), 958-64.
41. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011**, *10* (4), 1794-805.
42. Carapito, C.; Lane, L.; Benama, M.; Opsomer, A.; Mouton-Barbosa, E.; Garrigues, L.; Gonzalez de Peredo, A.; Burel, A.; Bruley, C.; Gateau, A.; Bouyssié, D.; Jaquinod, M.; Cianferani, S.; Burlet-Schiltz, O.; Van Dorsselaer, A.; Garin, J.; Vandembrouck, Y., Computational and Mass-Spectrometry-Based Workflow for the Discovery and Validation of Missing Human Proteins: Application to Chromosomes 2 and 14. *J Proteome Res* **2015**, *14* (9), 3621-34.
43. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26* (12), 1367-72.
44. Searle, B. C., Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10* (6), 1265-9.
45. Vaudel, M.; Sickmann, A.; Martens, L., Introduction to opportunities and pitfalls in functional mass spectrometry based proteomics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014**, *1844* (1, Part A), 12-20.
46. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012**, *404* (4), 939-65.
47. Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B., Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* **2012**, *11* (12), 1709-23.
48. Elliott, M. H.; Smith, D. S.; Parker, C. E.; Borchers, C., Current trends in quantitative proteomics. *J Mass Spectrom* **2009**, *44* (12), 1637-60.
49. Zhu, W.; Smith, J. W.; Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, *840518* (10), 10.
50. Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004**, *76* (14), 4193-201.
51. Lundgren, D. H.; Hwang, S. I.; Wu, L.; Han, D. K., Role of spectral counting in quantitative proteomics. *Expert review of proteomics* **2010**, *7* (1), 39-53.
52. Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; Verdin, E.; Kahn, C. R.; MacCoss, M. J.; Gibson, B. W., Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline: APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION. *Molecular & Cellular Proteomics : MCP* **2012**, *11* (5), 202-214.
53. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, *1* (5), 376-86.
54. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999**, *17* (10), 994-9.
55. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004**, *3* (12), 1154-69.
56. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75* (8), 1895-904.
57. Werner, T.; Sweetman, G.; Savitski, M. F.; Mathieson, T.; Bantscheff, M.; Savitski, M. M., Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. *Analytical Chemistry* **2014**, *86* (7), 3594-3601.

58. Alonso, S.; Izagirre, N.; Smith-Zubiaga, I.; Gardeazabal, J.; Diaz-Ramon, J. L.; Diaz-Perez, J. L.; Zelenika, D.; Boyano, M. D.; Smit, N.; de la Rúa, C., Complex signatures of selection for the melanogenic loci TYR, TYRP1 and DCT in humans. *BMC Evol Biol* **2008**, *8* (74), 1471-2148.
59. Chi, A.; Valencia, J. C.; Hu, Z. Z.; Watabe, H.; Yamaguchi, H.; Mangini, N. J.; Huang, H.; Canfield, V. A.; Cheng, K. C.; Yang, F.; Abe, R.; Yamagishi, S.; Shabanowitz, J.; Hearing, V. J.; Wu, C.; Appella, E.; Hunt, D. F., Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes. *J Proteome Res* **2006**, *5* (11), 3135-44.
60. Seiji, M.; Fitzpatrick, T. B.; Simpson, R. T.; Birbeck, M. S., Chemical composition and terminology of specialized organelles (melanosomes and melanin granules) in mammalian melanocytes. *Nature* **1963**, *197*, 1082-4.
61. Raposo, G.; Tenza, D.; Murphy, D. M.; Berson, J. F.; Marks, M. S., Distinct Protein Sorting and Localization to Premelanosomes, Melanosomes, and Lysosomes in Pigmented Melanocytic Cells(♣). *The Journal of Cell Biology* **2001**, *152* (4), 809-824.
62. Basrur, V.; Yang, F.; Kushimoto, T.; Higashimoto, Y.; Yasumoto, K.; Valencia, J.; Muller, J.; Vieira, W. D.; Watabe, H.; Shabanowitz, J.; Hearing, V. J.; Hunt, D. F.; Appella, E., Proteomic analysis of early melanosomes: identification of novel melanosomal proteins. *J Proteome Res* **2003**, *2* (1), 69-79.
63. Delevoeye, C.; Giordano, F.; van Niel, G.; Raposo, G., La biogenèse des mélanosomes: L'échiquier de la pigmentation. *Medecine sciences : M/S* **2011**, *27* (2), 153-162.
64. Raposo, G.; Marks, M. S., Melanosomes--dark organelles enlighten endosomal membrane transport. *Nat Rev Mol Cell Biol* **2007**, *8* (10), 786-97.
65. Virador, V. M.; Muller, J.; Wu, X.; Abdel-Malek, Z. A.; Yu, Z. X.; Ferrans, V. J.; Kobayashi, N.; Wakamatsu, K.; Ito, S.; Hammer, J. A.; Hearing, V. J., Influence of alpha-melanocyte-stimulating hormone and ultraviolet radiation on the transfer of melanosomes to keratinocytes. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2002**, *16* (1), 105-7.
66. Kobayashi, T.; Hearing, V. J., Direct interaction of tyrosinase with Tyrp1 to form heterodimeric complexes in vivo. *Journal of Cell Science* **2007**, *120* (24), 4261-4268.
67. Ito, S., A Chemist's View of Melanogenesis. *Pigment Cell Research* **2003**, *16* (3), 230-236.
68. Wakamatsu, I. a., <Chemistry of melanin_the pigmentary system 2d edition_2006.pdf>. **2006**.
69. Solano, F., Melanins: Skin Pigments and Much More—Types, Structural Models, Biological Functions, and Formation Routes. *New Journal of Science* **2014**, *Volume 2014*, 28.
70. Jimenez-Cervantes, Molecular Interactions within the Melanogenic Complex: Formation of Heterodimers of Tyrosinase and TRP1 from B16 Mouse Melanoma. **1998**.
71. Boissy, R. E.; Zhao, H.; Oetting, W. S.; Austin, L. M.; Wildenberg, S. C.; Boissy, Y. L.; Zhao, Y.; Sturm, R. A.; Hearing, V. J.; King, R. A.; Nordlund, J. J., Mutation in and lack of expression of tyrosinase-related protein-1 (TRP-1) in melanocytes from an individual with brown oculocutaneous albinism: a new subtype of albinism classified as "OCA3". *American journal of human genetics* **1996**, *58* (6), 1145-56.
72. Kobayashi, T.; Imokawa, G.; Bennett, D. C.; Hearing, V. J., Tyrosinase stabilization by Tyrp1 (the brown locus protein). *J Biol Chem* **1998**, *273* (48), 31801-5.
73. Hearing, V. J., Functional Properties of Cloned Melanogenic Proteins. **1992**.
74. Kobayashi, T.; Urabe, K.; Winder, A.; Jimenez-Cervantes, C.; Imokawa, G.; Brewington, T.; Solano, F.; Garcia-Borrón, J. C.; Hearing, V. J., Tyrosinase related protein 1 (TRP1) functions as a DHICA oxidase in melanin biosynthesis. *Embo J* **1994**, *13* (24), 5818-25.
75. Manga, P.; Sato, K.; Ye, L.; Beermann, F.; Lamoreux, M. L.; Orlow, S. J., Mutational analysis of the modulation of tyrosinase by tyrosinase-related proteins 1 and 2 in vitro. *Pigment Cell Res* **2000**, *13* (5), 364-74.
76. Zhao, H.; Eling, D. J.; Medrano, E. E.; Boissy, R. E., Retroviral infection with human tyrosinase-related protein-1 (TRP-1) cDNA upregulates tyrosinase activity and melanin synthesis in a TRP-1-deficient melanoma cell line. *J Invest Dermatol* **1996**, *106* (4), 744-52.
77. Orlow, S. J.; Zhou, B. K.; Chakraborty, A. K.; Drucker, M.; Pifko-Hirst, S.; Pawelek, J. M., High-molecular-weight forms of tyrosinase and the tyrosinase-related proteins: evidence for a melanogenic complex. *J Invest Dermatol* **1994**, *103* (2), 196-201.
78. Wang, W.; Park, D.; Ji, S.; Yin, S.-J.; Qian, G.-Y.; Chung, H. Y.; Yang, J.-M.; Lee, J.; Park, Y.-D., Computational prediction for the protein interactions of tyrosinase: Protein experimental interactome MAP. *Process Biochemistry* **2013**, *48* (4), 638-648.
79. Sangare, L. O.; Alayi, T. D.; Westermann, B.; Hovasse, A.; Sindikubwabo, F.; Callebaut, I.; Werkmeister, E.; Lafont, F.; Slomianny, C.; Hakimi, M. A.; Van Dorsselaer, A.; Schaeffer-Reiss, C.; Tomavo, S., Unconventional endosome-like compartment and retromer complex in *Toxoplasma gondii* govern parasite integrity and host infection. *Nature communications* **2016**, *7*, 11191.
80. Stojko, J.; Fioulaine, S.; Petiot-Becard, S.; Van Dorsselaer, A.; Meinel, T.; Giglione, C.; Cianferani, S., Ion mobility coupled to native mass spectrometry as a relevant tool to investigate extremely small ligand-induced conformational changes. *Analyst* **2015**, *140* (21), 7234-45.

81. Stojko, J.; Dupin, A.; Chaignepain, S.; Beaurepaire, L.; Vallet-Courbin, A.; Van Dorsselaer, A.; Schmitter, J.-M.; Minvielle-Sébastien, L.; Fribourg, S.; Cianféroni, S., Structural characterization of the yeast CF IA complex through a combination of mass spectrometry approaches. *International Journal of Mass Spectrometry*.
82. Kocher, T.; Superti-Furga, G., Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **2007**, *4* (10), 807-15.
83. Hernandez, H.; Robinson, C. V., Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protocols* **2007**, *2* (3), 715-726.
84. Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology* **2003**, *60* (5), 523-33.
85. Krogan, N. J.; Cagney, G.; Yu, H.; Zhong, G.; Guo, X.; Ignatchenko, A.; Li, J.; Pu, S.; Datta, N.; Tikuisis, A. P.; Punna, T.; Peregrín-Alvarez, J. M.; Shales, M.; Zhang, X.; Davey, M.; Robinson, M. D.; Paccanaro, A.; Bray, J. E.; Sheung, A.; Beattie, B.; Richards, D. P.; Canadien, V.; Lalev, A.; Mena, F.; Wong, P.; Starostine, A.; Canete, M. M.; Vlasblom, J.; Wu, S.; Orsi, C.; Collins, S. R.; Chandran, S.; Haw, R.; Rilstone, J. J.; Gandhi, K.; Thompson, N. J.; Musso, G.; St Onge, P.; Ghanny, S.; Lam, M. H. Y.; Butland, G.; Altaf-Ul, A. M.; Kanaya, S.; Shilatifard, A.; O'Shea, E.; Weissman, J. S.; Ingles, C. J.; Hughes, T. R.; Parkinson, J.; Gerstein, M.; Wodak, S. J.; Emili, A.; Greenblatt, J. F., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **2006**, *440* (7084), 637-643.
86. Li, Y., Commonly used tag combinations for tandem affinity purification. *Biotechnology and Applied Biochemistry* **2010**, *55* (2), 73-83.
87. Gingras, A. C.; Caballero, M.; Zarske, M.; Sanchez, A.; Hazbun, T. R.; Fields, S.; Sonenberg, N.; Hafen, E.; Raught, B.; Aebersold, R., A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol Cell Proteomics* **2005**, *4* (11), 1725-40.
88. Dunham, W. H.; Mullin, M.; Gingras, A. C., Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **2012**, *12* (10), 1576-90.
89. Rashid, K. A.; Hevi, S.; Chen, Y.; Le Caherec, F.; Chuck, S. L., A proteomic approach identifies proteins in hepatocytes that bind nascent apolipoprotein B. *J Biol Chem* **2002**, *277* (24), 22010-7.
90. Rappsilber, J., The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology* **2011**, *173* (3), 530-540.
91. Mallick, P.; Kuster, B., Proteomics: a pragmatic perspective. *Nat Biotech* **2010**, *28* (7), 695-709.
92. Vasilescu, J.; Guo, X.; Kast, J., Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **2004**, *4* (12), 3845-54.
93. Moyer, S. E.; Lewis, P. W.; Botchan, M. R., Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* **2006**, *103* (27), 10236-41.
94. Weiner, O. D.; Rentel, M. C.; Ott, A.; Brown, G. E.; Jedrychowski, M.; Yaffe, M. B.; Gygi, S. P.; Cantley, L. C.; Bourne, H. R.; Kirschner, M. W., Hem-1 Complexes Are Essential for Rac Activation, Actin Polymerization, and Myosin Regulation during Neutrophil Chemotaxis. *PLoS Biology* **2006**, *4* (2), e38.
95. Panigrahi, A. K.; Gygi, S. P.; Ernst, N. L.; Igo, R. P.; Palazzo, S. S.; Schnauffer, A.; Weston, D. S.; Carmean, N.; Salavati, R.; Aebersold, R.; Stuart, K. D., Association of Two Novel Proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA Editing Complex. *Molecular and Cellular Biology* **2001**, *21* (2), 380-389.
96. Ducut Sigala, J. L.; Bottero, V.; Young, D. B.; Shevchenko, A.; Mercurio, F.; Verma, I. M., Activation of transcription factor NF-kappaB requires ELKS, an IkappaB kinase regulatory subunit. *Science* **2004**, *304* (5679), 1963-7.
97. Smits, A. H.; Vermeulen, M., Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends in biotechnology* **2016**, *34* (10), 825-34.
98. Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **2007**, *8* (8), 645-54.
99. Choi-Rhee, E.; Schulman, H.; Cronan, J. E., Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein Science : A Publication of the Protein Society* **2004**, *13* (11), 3043-3050.
100. Varnaite, R.; MacNeill, S. A., Meet the neighbours: mapping local protein interactomes by proximity-dependent labelling with BioID. *Proteomics* **2016**.
101. Kim, D. I.; Kc, B.; Zhu, W.; Motamedchaboki, K.; Doye, V.; Roux, K. J., Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111* (24), E2453-E2461.
102. Mehus, A. A.; Anderson, R. H.; Roux, K. J., Chapter One - BioID Identification of Lamin-Associated Proteins. In *Methods in Enzymology*, Katherine, L. W.; Arnoud, S., Eds. Academic Press: 2016; Vol. Volume 569, pp 3-22.
103. Gupta, G. D.; Coyaud, E.; Goncalves, J.; Mojarad, B. A.; Liu, Y.; Wu, Q.; Gheiratmand, L.; Comartin, D.; Tkach, J. M.; Cheung, S. W.; Bashkurov, M.; Hasegan, M.; Knight, J. D.; Lin, Z. Y.; Schueler, M.; Hildebrandt, F.; Moffat, J.; Gingras, A. C.; Raught, B.; Pelletier, L., A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **2015**, *163* (6), 1484-99.
104. Morriswood, B., Form, Fabric, and Function of a Flagellum-Associated Cytoskeletal Structure. *Cells* **2015**, *4* (4), 726-47.

105. Schweingruber, C.; Soffientini, P.; Ruepp, M. D.; Bachi, A.; Muhlemann, O., Identification of Interactions in the NMD Complex Using Proximity-Dependent Biotinylation (BioID). *PLoS ONE* **2016**, *11* (3).
106. Roux, K. J.; Kim, D. I.; Burke, B., BioID: a screen for protein-protein interactions. *Current protocols in protein science* **2013**, *74*, Unit 19.23.
107. Kim, D. I.; Roux, K. J., Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends in Cell Biology* **26** (11), 804-817.
108. Mendoza, V. L.; Vachet, R. W., Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass spectrometry reviews* **2009**, *28* (5), 785-815.
109. Kim, D. I.; Jensen, S. C.; Noble, K. A.; Kc, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J., An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell* **2016**, *27* (8), 1188-1196.
110. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **2005**, *4* (9), 1265-72.
111. Mohammed, Y.; Borchers, C. H., An extensive library of surrogate peptides for all human proteins. *J Proteomics* **2015**, *129*, 93-7.
112. Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R., Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **2007**, *25* (1), 125-31.
113. Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004**, *3* (6), 608-14.
114. Klapper, M. H., The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochemical and biophysical research communications* **1977**, *78* (3), 1018-24.
115. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **2000**, *25* (1), 25-9.
116. Medzhitov, R.; Janeway, C. A., Jr., Decoding the patterns of self and nonself by the innate immune system. *Science* **2002**, *296* (5566), 298-300.
117. Meister, M.; Lemaître, B.; Hoffmann, J. A., Antimicrobial peptide defense in Drosophila. *BioEssays : news and reviews in molecular, cellular and developmental biology* **1997**, *19* (11), 1019-26.
118. Lemaître, B.; Hoffmann, J., The host defense of Drosophila melanogaster. *Annual review of immunology* **2007**, *25*, 697-743.
119. Ferrandon, D.; Imler, J.-L.; Hetru, C.; Hoffmann, J. A., The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat Rev Immunol* **2007**, *7* (11), 862-874.
120. Lemaître, B.; Reichhart, J. M.; Hoffmann, J. A., Drosophila host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc Natl Acad Sci U S A* **1997**, *94* (26), 14614-9.
121. Fehlbaum, P.; Bulet, P.; Michaut, L.; Lagueux, M.; Broekaert, W. F.; Hetru, C.; Hoffmann, J. A., Insect immunity. Septic injury of Drosophila induces the synthesis of a potent antifungal peptide with sequence homology to plant antifungal peptides. *J Biol Chem* **1994**, *269* (52), 33159-63.
122. Hoffmann, J. A., The immune response of Drosophila. *Nature* **2003**, *426* (6962), 33-38.
123. Lemaître, B.; Kromer-Metzger, E.; Michaut, L.; Nicolas, E.; Meister, M.; Georgel, P.; Reichhart, J. M.; Hoffmann, J. A., A recessive mutation, immune deficiency (imd), defines two distinct control pathways in the Drosophila host defense. *Proc Natl Acad Sci U S A* **1995**, *92* (21), 9465-9.
124. Wicker, C.; Reichhart, J. M.; Hoffmann, D.; Hultmark, D.; Samakovlis, C.; Hoffmann, J. A., Insect immunity. Characterization of a Drosophila cDNA encoding a novel member of the dipterin family of immune peptides. *J Biol Chem* **1990**, *265* (36), 22493-8.
125. Lemaître, B.; Nicolas, E.; Michaut, L.; Reichhart, J.-M.; Hoffmann, J. A., The Dorsoventral Regulatory Gene Cassette spätzle/Toll/cactus Controls the Potent Antifungal Response in Drosophila Adults. *Cell* **1996**, *86* (6), 973-983.
126. Ligoxygakis, P.; Pelte, N.; Hoffmann, J. A.; Reichhart, J. M., Activation of Drosophila Toll during fungal infection by a blood serine protease. *Science* **2002**, *297* (5578), 114-6.
127. Kellenberger, C.; Leone, P.; Coquet, L.; Jouenne, T.; Reichhart, J.-M.; Roussel, A., Structure-Function Analysis of Grass Clip Serine Protease Involved in Drosophila Toll Pathway Activation. *Journal of Biological Chemistry* **2011**, *286* (14), 12300-12307.
128. Chamy, L. E.; Leclerc, V.; Caldelari, I.; Reichhart, J.-M., Danger signal and PAMP sensing define binary signaling pathways upstream of Toll. *Nature immunology* **2008**, *9* (10), 1165-1170.
129. Doucet, A.; Butler, G. S.; Rodriguez, D.; Prudova, A.; Overall, C. M., Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* **2008**, *7* (10), 1925-51.

130. Gooley, A. A.; Ou, K.; Russell, J.; Wilkins, M. R.; Sanchez, J. C.; Hochstrasser, D. F.; Williams, K. L., A role for Edman degradation in proteome studies. *Electrophoresis* **1997**, *18* (7), 1068-72.
131. Van Damme, P.; Van Damme, J.; Demol, H.; Staes, A.; Vandekerckhove, J.; Gevaert, K., A review of COFRADIC techniques targeting protein N-terminal acetylation. *BMC Proceedings* **2009**, *3* (Suppl 6), S6-S6.
132. Prudova, A.; auf dem Keller, U.; Butler, G. S.; Overall, C. M., Multiplex N-terminome Analysis of MMP-2 and MMP-9 Substrate Degradomes by iTRAQ-TAILS Quantitative Proteomics. *Molecular & Cellular Proteomics* **2010**, *9* (5), 894-911.
133. Doucet, A.; Overall, C. M., Protease proteomics: Revealing protease in vivo functions using systems biology approaches. *Molecular Aspects of Medicine* **2008**, *29* (5), 339-358.
134. Staes, A.; Impens, F.; Van Damme, P.; Ruttens, B.; Goethals, M.; Demol, H.; Timmerman, E.; Vandekerckhove, J.; Gevaert, K., Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat. Protocols* **2011**, *6* (8), 1130-1141.
135. Venne, A. S.; Vogtle, F. N.; Meisinger, C.; Sickmann, A.; Zahedi, R. P., Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. *J Proteome Res* **2013**, *12* (9), 3823-30.
136. Bland, C.; Hartmann, E. M.; Christie-Oleza, J. A.; Fernandez, B.; Armengaud, J., N-Terminal-oriented proteogenomics of the marine bacterium *roseobacter denitrificans* Och114 using N-Succinimidylloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) labeling and diagonal chromatography. *Mol Cell Proteomics* **2014**, *13* (5), 1369-81.
137. Ayoub, D.; Bertaccini, D.; Diemer, H.; Wagner-Rousset, E.; Colas, O.; Cianferani, S.; Van Dorsselaer, A.; Beck, A.; Schaeffer-Reiss, C., Characterization of the N-Terminal Heterogeneities of Monoclonal Antibodies Using In-Gel Charge Derivatization of alpha-Amines and LC-MS/MS. *Anal Chem* **2015**, *87* (7), 3784-90.
138. Gallien, S.; Perrodou, E.; Carapito, C.; Deshayes, C.; Reytrat, J. M.; Van Dorsselaer, A.; Poch, O.; Schaeffer, C.; Lecompte, O., Ortho-proteogenomics: multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res* **2009**, *19* (1), 128-35.
139. Paizs, B.; Suhai, S., Towards understanding the tandem mass spectra of protonated oligopeptides. 1: mechanism of amide bond cleavage. *Journal of the American Society for Mass Spectrometry* **2004**, *15* (1), 103-113.
140. Ayoub, D., Vers une étude approfondie des protéomes: caractérisation des extrémités N-terminales des protéines *Thèse de doctorat, Ecole doctorales des sciences chimiques UMR7 178-IPHC: 2012.*
141. Hoffmann, J. A.; Reichhart, J. M., Drosophila innate immunity: an evolutionary perspective. *Nat Immunol* **2002**, *3* (2), 121-6.
142. Gottar, M.; Gobert, V.; Matskevich, A. A.; Reichhart, J. M.; Wang, C.; Butt, T. M.; Belvin, M.; Hoffmann, J. A.; Ferrandon, D., Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell* **2006**, *127* (7), 1425-37.
143. Giron, P.; Dayon, L.; Sanchez, J.-C., Cysteine tagging for MS-based proteomics. *Mass Spectrometry Reviews* **2011**, *30* (3), 366-395.
144. UniProtKB, Swiss-Prot protein knowledgebase release 2016_11-statistics. In <http://web.expasy.org/docs/relnotes/relstat.html>, 2016.
145. Baldwin, A. D.; Kick, K. L., Tunable Degradation of Maleimide-Thiol Adducts in Reducing Environments. *Bioconjugate Chemistry* **2011**, *22* (10), 1946-1953.
146. Koniev, O.; Leriche, G.; Nothisen, M.; Remy, J.-S.; Strub, J.-M.; Schaeffer-Reiss, C.; Van Dorsselaer, A.; Baati, R.; Wagner, A., Selective Irreversible Chemical Tagging of Cysteine with 3-Arylpropionitriles. *Bioconjugate Chemistry* **2014**, *25* (2), 202-206.
147. Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O., An approach to quantitative proteome analysis by labeling tryptophan residues. *Rapid Communications in Mass Spectrometry* **2003**, *17* (14), 1642-1650.
148. Tambor, V.; Hunter, C. L.; Seymour, S. L.; Kacerovsky, M.; Stulik, J.; Lenco, J., CysTRAQ — A combination of iTRAQ and enrichment of cysteinyl peptides for uncovering and quantifying hidden proteomes. *Journal of Proteomics* **2012**, *75* (3), 857-867.
149. Liu, T.; Qian, W. J.; Chen, W. N.; Jacobs, J. M.; Moore, R. J.; Anderson, D. J.; Gritsenko, M. A.; Monroe, M. E.; Thrall, B. D.; Camp, D. G., 2nd; Smith, R. D., Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: the human mammary epithelial cell proteome. *PROTEOMICS* **2005**, *5* (5), 1263-73.
150. Rauniyar, N.; Yates, J. R., Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research* **2014**, *13* (12), 5293-5309.
151. Qu, Z.; Meng, F.; Bomgarden, R. D.; Viner, R. I.; Li, J.; Rogers, J. C.; Cheng, J.; Greenlief, C. M.; Cui, J.; Lubahn, D. B.; Sun, G. Y.; Gu, Z., Proteomic Quantification and Site-Mapping of S-Nitrosylated Proteins Using Isobaric iodoTMT Reagents. *Journal of Proteome Research* **2014**, *13* (7), 3200-3211.
152. Shiu, H. Y.; Chan, T. C.; Ho, C. M.; Liu, Y.; Wong, M. K.; Che, C. M., Electron-deficient alkynes as cleavable reagents for the modification of cysteine-containing peptides in aqueous medium. *Chemistry* **2009**, *15* (15), 3839-50.
153. Ren, D.; Penner, N. A.; Slentz, B. E.; Mirzaei, H.; Regnier, F. E., Enrichment of Cysteine-Containing Peptides from Tryptic Digests Using a Quaternary Amine Tag. *J. Proteome Res.* **2003**, *2*, 321.

154. Kolodych, S.; Koniev, O.; Baatarkhuu, Z.; Bonnefoy, J. Y.; Debaene, F.; Cianferani, S.; Van Dorsselaer, A.; Wagner, A., CBTF: new amine-to-thiol coupling reagent for preparation of antibody conjugates with increased plasma stability. *Bioconjug Chem* **2015**, *26* (2), 197-200.
155. Shimada, T.; Kuyama, H.; Sato, T. A.; Tanaka, K., Development of iodoacetic acid-based cysteine mass tags: detection enhancement for cysteine-containing peptide by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Biochem* **2012**, *421* (2), 785-7.
156. Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J., Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **2003**, *21* (5), 566-9.
157. Antos, J. M.; Francis, M. B., Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution. *Journal of the American Chemical Society* **2004**, *126* (33), 10256-10257.
158. Foettinger, A.; Melmer, M.; Leitner, A.; Lindner, W., Reaction of the indole group with malondialdehyde: application for the derivatization of tryptophan residues in peptides. *Bioconjug Chem* **2007**, *18* (5), 1678-83.
159. Hansen, M. B.; Hubalek, F.; Skrydstrup, T.; Hoeg-Jensen, T., Chemo- and Regioselective Ethynylation of Tryptophan-Containing Peptides and Proteins. *Chemistry* **2015**, *4* (10), 201504462.
160. Matsuo, E.-i.; Watanabe, M.; Kuyama, H.; Nishimura, O., A new strategy for protein biomarker discovery utilizing 2-nitrobenzenesulfonyl (NBS) reagent and its applications to clinical samples. *Journal of Chromatography B* **2009**, *877* (25), 2607-2614.
161. Takami, Y.; Uto, H.; Tamai, T.; Sato, Y.; Ishida, Y.; Morinaga, H.; Sakakibara, Y.; Moriuchi, A.; Oketani, M.; Ido, A.; Nakajima, T.; Okanoue, T.; Tsubouchi, H., Identification of a novel biomarker for oxidative stress induced by hydrogen peroxide in primary human hepatocytes using the 2-nitrobenzenesulfonyl chloride isotope labeling method. *Hepatology research : the official journal of the Japan Society of Hepatology* **2010**, *40* (4), 438-45.
162. Matsuo, E.; Toda, C.; Watanabe, M.; Ojima, N.; Izumi, S.; Tanaka, K.; Tsunasawa, S.; Nishimura, O., Selective detection of 2-nitrobenzenesulfonyl-labeled peptides by matrix-assisted laser desorption/ionization-time of flight mass spectrometry using a novel matrix. *PROTEOMICS* **2006**, *6* (7), 2042-9.
163. Matsuo, E.; Toda, C.; Watanabe, M.; Iida, T.; Masuda, T.; Minohata, T.; Ando, E.; Tsunasawa, S.; Nishimura, O., Improved 2-nitrobenzenesulfonyl method: optimization of the protocol and improved enrichment for labeled peptides. *Rapid Commun Mass Spectrom* **2006**, *20* (1), 31-8.
164. Scoffone, E.; Fontana, A.; Rocchi, R., Sulfonyl halides as modifying reagents for polypeptides and proteins. I. Modification of tryptophan residues. *Biochemistry* **1968**, *7* (3), 971-979.
165. Murray, M. G.; Thompson, W. F., Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **1980**, *8* (19), 4321-4325.
166. Findlay, J. W. A.; Smith, W. C.; Lee, J. W.; Nordblom, G. D.; Das, I.; DeSilva, B. S.; Khan, M. N.; Bowsher, R. R., Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *Journal of Pharmaceutical and Biomedical Analysis* **2000**, *21* (6), 1249-1273.
167. Muller, L.; Fornecker, L.; Van Dorsselaer, A.; Cianferani, S.; Carapito, C., Benchmarking sample preparation/digestion protocols reveals tube-gel being a fast and repeatable method for quantitative proteomics. *Proteomics* **2016**, *16* (23), 2953-2961.
168. McKay, Craig S.; Finn, M. G., Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chemistry & Biology* **2014**, *21* (9), 1075-1101.
169. Ban, H.; Gavriluk, J.; Barbas, C. F., Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *Journal of the American Chemical Society* **2010**, *132* (5), 1523-1525.
170. Zhan, X.; Desiderio, D. M., Nitroproteins from a human pituitary adenoma tissue discovered with a nitrotyrosine affinity column and tandem mass spectrometry. *Anal Biochem* **2006**, *354* (2), 279-89.
171. Rikova, K.; Guo, A.; Zeng, Q.; Possemato, A.; Yu, J.; Haack, H.; Nardone, J.; Lee, K.; Reeves, C.; Li, Y.; Hu, Y.; Tan, Z.; Stokes, M.; Sullivan, L.; Mitchell, J.; Wetzell, R.; Macneill, J.; Ren, J. M.; Yuan, J.; Bakalarski, C. E.; Villen, J.; Kornhauser, J. M.; Smith, B.; Li, D.; Zhou, X.; Gygi, S. P.; Gu, T. L.; Polakiewicz, R. D.; Rush, J.; Comb, M. J., Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **2007**, *131* (6), 1190-203.
172. Jones, L. H.; Narayanan, A.; Hett, E. C., Understanding and applying tyrosine biochemical diversity. *Molecular BioSystems* **2014**, *10* (5), 952-969.
173. Minakawa, M.; Guo, H.-M.; Tanaka, F., Imines that React with Phenols in Water over a Wide pH Range. *The Journal of organic chemistry* **2008**, *73* (21), 8669-8672.
174. Joshi, N. S.; Whitaker, L. R.; Francis, M. B., A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J Am Chem Soc* **2004**, *126* (49), 15942-3.
175. Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y., Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging. *Science (New York, N.Y.)* **2013**, *339* (6125), 1328-1331.
176. Ban, H.; Nagano, M.; Gavriluk, J.; Hakamata, W.; Inokuma, T.; Barbas, C. F., Facile and Stable Linkages through Tyrosine: Bioconjugation Strategies with the Tyrosine-Click Reaction. *Bioconjugate Chemistry* **2013**, *24* (4), 520-532.

177. Chawade, A.; Alexandersson, E.; Levander, F., Normalyzer: A Tool for Rapid Evaluation of Normalization Methods for Omics Data Sets. *Journal of Proteome Research* **2014**, *13* (6), 3114-3120.

Nouvelles applications et opportunités en protéomique

Résumé

Les objectifs de mes travaux de thèse étaient de développer de nouvelles méthodes d'identification, de caractérisation et de quantification de protéines, mieux adaptées à la diversité des études en protéomique, ce dont la biologie a besoin aujourd'hui.

L'analyse protéomique par spectrométrie de masse est apparue comme un outil précieux et pertinent pour évaluer la qualité de l'isolement d'un complexe spécifique, et pour guider les biologistes dans les choix de la stratégie à adopter.

La stratégie de marquage de N-terminomique développée a permis de caractériser un processus de maturation biologique en déterminant précisément les sites d'activation de la protéine Perséphone par marquage spécifique des extrémités N-terminales. Ce travail a permis d'élucider un nouveau mécanisme fin de régulation dans l'immunité innée chez la drosophile

De nouveaux modes de marquages ont été mis au point et les familles chimiques des réactifs de marquage étudiés permettront d'adapter au mieux les études de quantifications protéomiques à la nature et aux contraintes des études biologiques à mener.

Mots clés : Protéomique _ Spectrométrie de masse _ Complexe _ N-terminomique _ Marquage _ Caractérisation _ Quantification

Summary

The aim of this work was to develop new methods for the identification, characterization and quantification of proteins best suited to a large diversity of proteomics studies, which is nowadays essential to biology.

Our work has shown that proteomic analysis based on mass spectrometry can be a valuable and relevant tool to evaluate the isolation strategy efficiency set up for a specific complex and thus guide the biologists in their choice.

The N-terminomic labeling strategy developed allowed us to describe a biological maturation process by determining precisely the Persephone protein activation sites using specific labeling of the successively generated N-terminal extremities. This work allowed elucidating a new regulation mechanism in the Drosophila innate immunity system.

New chemical labeling reagents to target specific amino acids (cysteine, tyrosine and tryptophan) have been set up for fast mass-spectrometry based proteomics.

These labeling strategies combined with proteomic tools will allow developing a robust and quantitative approach essential for biological studies.

Keywords : Proteomics _ Mass spectrometry _ Multiprotein complex _ N-terminomic _ Chemical labeling_ Characterization _ Quantification