

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE DE PROTÉINES IMPLIQUÉES DANS LA BIOSYNTHÈSE DE BYSSUS
CHEZ LA MOULE PAR SPECTROMÉTRIE DE MASSE EN TANDEM

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN CHIMIE

PAR

MAXIME SANSOUCY

SEPTEMBRE 2021

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

J'aimerais remercier en premier lieu le professeur Pedro Alejandro Segura qui a généreusement accepté cette tâche supplémentaire en tant qu'examineur externe. Aussi, les professeurs Steve Bourgault et Marc Lussier pour avoir accepté mon invitation à se joindre à mon comité d'évaluation dès le début. Leurs conseils et commentaires m'ont permis à chaque étape de mon doctorat de m'améliorer en tant que chercheur. Bien entendu, je ne peux passer sous silence l'incroyable privilège d'avoir étudié sous la direction du professeur Lekha Sleno.

Je profite également de l'occasion pour remercier les technologues et enseignants Isabelle Cloutier, Marie-Claude, Sophie Chen, Sylvie Lemieux et Rami Abdhelaqi avec qui j'ai eu la chance de travailler durant ces cinq dernières années.

J'aimerais aussi remercier les membres du laboratoire Sleno, dont Leanne Ohlund, assistante de recherche, pour son aide indispensable, mais également les membres antérieurs Dr. André LeBlanc et Prof. Makan Golizeh qui m'ont introduit et accueilli au sein de leur équipe à ce moment. Les membres actuels et précédents, Amal Guesmi, Vivaldi Prinville, Dr. Timon Geib, Ons Ousji, Kahina Chabi, Maggy Lépine ainsi que Myriam Mirrault à qui je dois plusieurs moments cocasses et inoubliables.

Finalement, à titre personnel, je remercie sincèrement ma copine Émilie qui a su être à mes côtés tout au long de mes études, mon grand ami Mick, un ami cher et très proche que j'ai la chance de côtoyer depuis la première année du primaire et bien certainement, ma chère mère qui, avec tous les efforts, la détermination et le courage qu'une mère monoparentale doit démontrer, a su transmettre sa passion pour les sciences et développer ma curiosité et ma détermination.

DÉDICACE

[...] from time to time, there arise among
human beings, people who seem to exude love
as naturally as the sun gives us heat.

Alan W. Watts

À ma maman, ma douce Émilie, mon grand
ami Mick et mes amis

TABLE DES MATIÈRES

LISTE DES FIGURES.....	viii
LISTE DES TABLEAUX.....	xvi
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES	xviii
Liste des symboles et des unités.....	xx
RÉSUMÉ.....	xxii
ABSTRACT	xxiii
CHAPITRE I Introduction	24
1.1 Le byssus	25
1.2 Aspects généraux et applications de la protéomique.....	32
1.3 Approches expérimentales.....	34
1.4 Application de la LC-MS/MS à la protéomique.....	39
1.5 Chromatographie 1D <i>versus</i> 2D	49
1.6 Identification de modifications impliquant la DOPA.....	53
1.7 Défis et planification de la recherche	58
1.7.1 Minimiser les effets de matrice	58
1.7.2 Choix de la base de données de protéines.....	60
CHAPITRE II COMPARAISON D'APPROCHES CHROMATOGRAPHIQUES POUR L'ANALYSE DE PEPTIDES PAR LC-MS/MS	63
2.1 Résumé	64
2.2 Introduction.....	64
2.3 Expérimentale.....	67
2.3.1 Matériel	67

2.3.2	Fractionnement par chromatographie d'exclusion stérique	67
2.3.3	Digestion des protéines	67
2.3.4	Fractionnement en phase inverse à pH élevé	68
2.3.5	Fractionnement par échange de cations forts	69
2.3.6	Analyse RP-LC-MS/MS	69
2.3.7	Traitement des données	70
2.4	Résultats et discussion	71
2.4.1	Comparaison entre les approches 1D et 2D	71
2.5	Conclusions.....	80

CHAPITRE III ÉTUDE DE LA BYSSOGÉNÈSE PAR ANALYSES PROTÉOMIQUES SUR LE BYSSUS, LE PIED ET LE MANTEAU CHEZ LES MOULES *MYTILUS* PAR LC-MS/MS

3.1	Résumé	82
3.2	Introduction.....	83
3.3	Expérimentale	86
3.3.1	Matériels.....	86
3.3.2	Extraction et digestion des protéines.....	86
3.3.3	Analyse RP-LC-MS/MS	87
3.3.4	Traitement des données	88
3.4	Résultats et discussion	88
3.5	Conclusions.....	97

CHAPITRE IV ÉTUDE DES GLANDES DU PIED DE LA MOULE *MYTILUS* *EDULIS* PAR DES ANALYSES TRANSCRIPTOMIQUES ET PROTÉOMIQUES 98

4.1	Résumé	99
4.2	Introduction.....	99
4.3	Expérimentale	101
4.3.1	Matériel	101
4.3.2	Collection des glandes du pied de <i>M. edulis</i>	101
4.3.3	Digestion des protéines provenant de chaque glande	102
4.3.4	Analyse RP-LC-MS/MS	103
4.3.5	Analyse transcriptomique des protéines associées au byssus	103
4.3.6	Traitement de données	104
4.4	Résultats et discussion	105

4.5	Conclusions.....	115
-----	------------------	-----

CHAPITRE V CARACTÉRISATION PAR PROTÉOMIQUE ASCENDANTE DE LA PROTÉINE RECOMBINANTE MFP-1 DE LA MOULE *MYTILUS EDULIS* 117

5.1	Résumé	118
5.2	Introduction.....	118
5.3	Expérimentale	120
5.3.1	Matériel	120
5.3.2	Expression et oxydation de la protéine rMFP-1	120
5.3.3	Digestion de la rMFP-1 par la trypsine	121
5.3.4	Hydrolyse complète de la protéine rMFP-1	121
5.3.5	Analyse RP-LC-MS/MS	122
5.4	Résultats et discussion	124
5.4.1	Optimisation de la chromatographie	124
5.4.2	Optimisation des conditions de digestion	125
5.4.3	Hydrolyse complète de la protéine rMFP-1	140
5.4.4	Identification de la protéine mfp-1 du byssus	143

CHAPITRE VI DÉTERMINATION DE RÉACTIONS COVALENTES ENTRE LA DOPA ET AUTRES CATÉCHOLES AVEC LA CYSTÉINE PAR LC-MS/MS 146

6.1	Résumé	147
6.2	Introduction.....	147
6.3	Expérimentale	149
6.3.1	Matériel	149
6.3.2	Oxydation par la tyrosinase.....	150
6.3.3	Dérivation pré-colonne au benzoyle	150
6.3.4	Analyse RP-LC-MS/MS	150
6.3.5	Traitement des données.....	151
6.4	Résultats et discussion	151
6.4.1	Optimisation de l'incubation.....	151
6.4.2	Élucidation des régioisomères du Cys-DOPA	159
6.5	Conclusions.....	162

CHAPITRE VII CONCLUSIONS ET PERSPECTIVES..... 163

7.1	Résumé des découvertes	164
-----	------------------------------	-----

7.2 Perspectives	167
ANNEXE A INVESTIGATING BYSSOGENESIS WITH PROTEOMIC ANALYSIS OF BYSSUS, FOOT AND MANTLE IN <i>MYTILUS</i> MUSSELS BY LC- MS/MS	176
ANNEXE B ÉTUDE DE LA BYSSOGÉNÈSE PAR DES ANALYSES PROTÉOMIQUES SUR LE BYSSUS, LE PIED ET LE MANTEAU CHEZ LES MOULES MYTILUS PAR LC-MS/MS.....	185
ANNEXE C COMPARAISON D'APPROCHES DE PRÉ-FRACTIONNEMENT POUR LES ANALYSES LC-MS/MS EN PROTÉOMIQUE ASCENDANTE	188
ANNEXE D SCHÉMAS DE FRAGMENTATION SUGGÉRÉS POUR LE CYS- DOPAC ET LE CYS-DOPAMINE	207
RÉFÉRENCES.....	209

LISTE DES FIGURES

Figure	Page
1.1 Modélisation 3D du pied de la moule <i>Mytilus edulis</i> . La glande centrale (<i>core gland</i> , section bleue), la glande cuticle (<i>cuticle gland</i> , section rouge) et la glande plaque (<i>plaque gland</i> , section verte) y sont représentées. La figure a été adaptée de (Priemel, Degtyar, Dean, & Harrington, 2017).	27
1.2 Réactions d'oxydation sur la tyrosine donnant la DOPA, responsable de l'adhésion sous-marine de la fibre de byssus. Si la DOPA est oxydée en ortho-quinone, l'efficacité de l'adhésion est diminuée et des réactions croisées entre les chaînes latérales des acides aminés nucléophiles peuvent être observées.	30
1.3 Représentation des quatre principaux domaines des sciences «omiques» ainsi que leur analyte d'intérêt.	33
1.4 Comparaison entre les approches en protéomique impliquant la séparation sur gel et sans gel couplées à la spectrométrie de masse.	35
1.5 Représentation sommaire de la préparation des extraits de protéines dans le but d'une analyse de type ascendante (<i>bottom-up</i> , schéma du haut) ou descendante (<i>top-down</i> , schéma du bas).	36

1.6 Schémas montrant les étapes de réduction des ponts disulfures des protéines par le dithiothréitol (a), suivit de l'alkylation par l'iodoacétamide (b) pour empêcher la formation spontanée de nouveaux ponts disulfures.	38
1.7 Schéma général d'un spectromètre de masse. Selon le type de source utilisée, il est possible que celle-ci soit opérée à pression atmosphérique ou sous-vide.....	40
1.8 Comparaison entre un appareil en mode spatial (gauche) et temporel (droite) employé en spectrométrie de masse en tandem. La figure a été adaptée de <i>Mass Spectrometry, Gross J.H, 2nd edition</i>	42
1.9 Schéma représentant la fragmentation d'un peptide en montrant les liens brisés selon la technique d'activation employée. La structure des ions obtenus par CID en mode positif est également montrée.	46
1.10 Schéma de l'optique d'ion dans le spectromètre de masse hybride quadrupôle-temps de vol TripleTOF™ 5600 de Sciex (disponible en ligne sur sciex.com).	48
1.11 Représentation généralisée de différentes stratégies de combinaison de fractions proposées dans les études en protéomique ascendante. La fenêtre chromatographique est divisée en deux (A) ou en trois portions (B) selon le nombre de fractions à collecter. Les fractions de chaque zone sont ensuite mélangées ensemble en prévision de l'analyse par LC-MS/MS.	52
1.12 Structures de composés électrophiles principalement rencontrés comme xénobiotiques et composés endogènes. La DOPA est un exemple	

particulièrement important puisque cette structure peut être également considérée comme une modification post-traductionnelle.	54
1.13 Représentation simplifiée du métabolisme de la dopamine, donnant lieu à plusieurs intermédiaires réactifs potentiellement impliqués dans les liaisons covalentes avec les cystéines et les lysines des protéines.	57
2.1 Représentation des méthodologies 1D- et 2D-LC couramment employées dans les analyses en protéomique ascendante.....	72
2.2 Chromatogrammes UV pour le fractionnement des peptides par échange de cations forts (haut) et par chromatographie en phase inverse à pH élevé (bas) pour les peptides obtenus par digestion tryptique (signal en bleu) et pepsique (signal en rouge) pour les échantillons de pied de moules.	73
2.3 Séparation des protéines provenant du pied de la moule par chromatographie d'exclusion stérique. Le chromatogramme du haut montre la fenêtre de collection des fractions (de 9 à 29 min) alors que les chromatogrammes ci-dessous montrent le comportement de la séparation lorsque la phase mobile (150 mM NaH ₂ PO ₄) est ajustée à pH 7, pH 3.5 ou contenant 10% d'ACN. Les graphiques dans l'encadré gris illustrent la distribution des masses moléculaires par fraction.	74
2.4 Comparaison entre les spectres TOF-MS obtenus pour l'analyse du peptide QTVAVGIK (<i>m/z</i> 457.78) de la protéine du facteur d'élongation 1-alpha, analysé sans pré-fractionnement, par SCX et par Hp-RP. L'ion précurseur est montré en rouge et l'encadré gris montre un agrandissement de cette zone.	76

2.5 Répartition des protéines identifiées pour les échantillons de pied de <i>Mytilus edulis</i> entre les trois approches.	79
3.1 Distribution relative des protéines provenant de la base de données (a), ainsi que les protéines identifiées à partir de MYTED (b) et MYTCA (c) en combinant les protéines de byssus, de pied et de manteau.	90
3.2 Nombre de protéines identifiées dans les échantillons de pied, de manteau et de byssus à partir de <i>M. edulis</i> (A) et <i>M. californianus</i> (b). Les protéines trouvées dans les échantillons de byssus sont comparées en (C) pour les deux espèces.	91
3.3 Couverture des séquences des protéines provenant des sous-ensembles de pieds et de byssus. La séquence des peptides détectée (à 1% FDR) est présentée en gras pour <i>M. edulis</i> et soulignée pour <i>M. californianus</i>	95
4.1 Distribution des protéines incluses dans la base de données UniProtKb/Swiss-Prot (74 000 séquences) selon leur catégorie (homologie, transcripte, ou séquence prédite).	106
4.2 Regroupement des peptides (à 95% confiance) et de protéines (à 1% FDR) identifiés pour chaque glande du pied et le manteau de la moule <i>M. edulis</i> à partir des séquences obtenues par UniProt (A) et par transcriptomique (B).	109
4.3 Distribution des protéines identifiées pour chaque glande du pied de la moule <i>M. edulis</i> (A) pour les séquences obtenues par UniProt (gauche) et par transcriptomique (droite). La figure présente également une comparaison	

entre les protéines identifiées uniquement dans le pied (B) pour les deux bases de données employées.....	110
5.1 Comparaison de l'allure du pic associé au peptide AKPSYPPTYK lorsque séparé sur des colonnes C12 et biphényle à l'aide d'un gradient de 36 min (à gauche) et de 12 min (droite).....	125
5.2 Chromatogrammes d'ions extraits du peptide tryptique AKPSYPPTYK incubé en présence de différents composés anti-oxydants, pour les versions du peptide sans oxydation (A, m/z 576.31), avec l'addition d'un oxygène (B, m/z 584.30) et deux oxygènes (C, m/z 592.30).....	126
5.3 EIC montrant l'impact de trois concentrations d'acide ascorbique sur la sensibilité du peptide AKPSYPPTYK sans la présence de DOPA (A), avec une DOPA (B) et deux DOPA (C) ajoutée en prévision de la digestion par la trypsine.....	127
5.4 Comparaison des différentes possibilités d'oxydation sur le décapeptide AKPSYPPTYK pour les états de charge +2 (bleu) et +3 (rose). Les résidus tyrosine notés d'un astérisque (Y*) font référence au site d'oxydation associé à la présence de la DOPA.....	128
5.5 Spectre MS/MS obtenus pour les peptides comportant une seule DOPA (A) et deux DOPA (B) issue d'une digestion par la trypsine. Les séries d'ions y (en vert) et b (en bleu) sont présentées sur les deux spectres avec quels fragments internes. La forte intensité des ions y_8 et y_5 est la conséquence de l'effet proline sur la fragmentation du peptide.....	130

5.6 Courant d'ions extraits du peptide AKPSYPPTYK comportant une ou deux DOPA lorsqu'incubé en présence de différents agents anti-oxydants.....	132
5.7 Schéma proposé de la réaction entre l' <i>ortho</i> -quinone (Tyr-9) et la lysine-2 (réaction présentée au centre) formant une liaison croisée intermoléculaire. La réaction avec les groupements thiol (illustré à l'aide du glutathion) et la quinone fut également étudiée.	134
5.8 Schéma proposé pour la fragmentation du peptide AKPSYPPTYK après qu'une liaison intra-moléculaire entre la Lys-2 et la DOPA (Tyr-9) se soit formée. Les rapport m/z en gras peuvent être utilisés ultérieurement comme ions diagnostiques confirmant la présence de la liaison croisée....	135
5.9 Spectre MS/MS du peptide AKPSYPPTYK ayant formé un cycle une fois la lysine-2 attaché à la dopaquinine (Tyr-9). Le séquençage de novo montre l'absence d'ion y et b, potentiellement causé par la liaison croisée, en initiant la fragmentation à partir de la Pro-3. Les ions en gras représentés des fragments pouvant être utilisé comme ions diagnostiques ultérieurement.	136
5.10 Comparaison de l'intensité des signaux obtenus en présence de GSH à différentes concentrations pour le peptide AKPSYPPTYK sans DOPA (gauche), avec une DOPA (centre) et deux DOPA (droite).	137
5.11 Résumé des étapes visant à l'optimisation des incubations de la protéine rMFP-1 DOPA avec le glutathion. Pour chaque scénario, une concentration finale en glutathion et ascorbate de 0.4 mM et 1 mM a été respectivement visée. Une quantité de 2 µg de trypsine a ensuite été ajoutée pour une digestion d'une durée d'une heure.....	138

5.12 Comparaison de l'impact d'une pré-incubation avant ou après l'ajout de GSH sur la sensibilité du peptide tryptique de la rMFP-1 (avec ou sans DOPA). La comparaison montre également l'effet sur la formation de l'adduit GSH avec la version DOPA du peptide AKPSYPPTY*K.....	139
5.13 Spectre MS/MS du peptide AKPSYPPTY*K comportant un adduit GSH. Le séquençage de novo à permis l'identification de plusieurs ions y incluant le GSH dans la structure.	141
6.1 Réaction d'oxydation de la tyrosine en DOPA avec l'aide de la tyrosinase. La fonction dioxygénase de la tyrosinase permet une oxydation successive pour convertir la DOPA en dopaquinone, l'intermédiaire réactif de la DOPA.....	152
6.2 Abondance du Cys-DOPA lorsqu'incubé à différents ratios de cystéine et DOPA (A) et avec une quantité croissante de tyrosinase (B). La réaction conduite avec la tyrosine est également présentée dans les graphiques.	153
6.3 Chromatogramme d'ion extraits présentant la Cys-DOPA (m/z 317.08), la Cys-DA (m/z 273.09) et la Cys-DOPAC (m/z 288.05) incubé avec un ratio catéchol :Cys de 2 :1 dans les conditions optimales.....	154
6.4 Aire sous la courbe moyenne pour Cys-DOPA, Cys-DA et Cys-DOPAC (gauche) ainsi que l'impact de l'excès de catéchol sur la réaction d'addition avec la cystéine (droite).....	155
6.5 Spectre MS ² de la Cys-DOPA, Cys-DA et de la Cys-DOPAC non dérivées. Les ions précurseurs pour chaque composé sont présentés en gras et les	

ions communs sont soulignés. Les valeurs annotées d'un astérix furent identifiés dans deux spectres, mais avec un écart de ± 2 Da.	157
6.6 Schémas proposés de la fragmentation de la Cys-DOPA	158
6.7 Structures et équivalents deutérés de la Cys-DOPA, présentant les isomères 5-S (à gauche) et 2-S (à droite) obtenus à partir de la tyrosine deutérée lorsque celle-ci réagit avec les groupements thiols libres des protéines (ou du glutathion).....	159
6.8 Chromatogramme d'ion extrait pour les isomères 5-S (en bleu) et 2-S (en rose) pour les versions non dérivées, dérivé avec un groupement benzyle, puis dérivé avec deux groupement benzyle. Les valeurs d'aire sous la courbe pour chaque composé sont présentées dans le diagramme à barres sur la droite.	161
7.1 Flux de travail proposé pour une approche quantitative sur les protéines de moule <i>M. edulis</i> utilisant le réactif ICAT afin de marqué les protéines possédant au minimum un résidu cystéine. Les protéines seront digérées par la trypsine dans des conditions similaires que pour les études qualitative. Les peptides n'ayant pas été marqué par ICAT seront collectés et analysés également.....	169
7.2 Flux de travail proposé pour l'identification de la mfp-1 à partir du culot présent dans les échantillons de la glande cuticule. L'utilisation d'agents chaotropiques comme l'urée, la thiourée et le chlorure de guanidinium seront à l'étude ainsi que le temps nécessaire à l'homogénéisation du culot.	170

LISTE DES TABLEAUX

Tableau	Page
1.1 Regroupement des changements bio-mécaniques de la fibre de byssus	31
1.2 Caractéristiques des principaux analyseurs de masses disponibles	44
2.1 Comparaison de chaque approche pour les peptides et les protéines identifiés dans le tissu du pied de la moule <i>Mytilus edulis</i>	77
3.1 Protéines associées à la fabrication de la fibre identifiée dans les échantillons de pied et de byssus pour <i>M. edulis</i> et <i>M. californianus</i>	92
3.2 Résultats des recherches BLAST pour les protéines non caractérisées trouvées dans les échantillons de pied et de byssus.....	96
4.1 Comparaison du nombre de protéines identifiées par la base UniProt et transcriptomique pour <i>Mytilus edulis</i>	107
4.2 Protéines du byssus identifiées à partir de la base UniProt et transcriptomique	113
5.1 Masses monoisotopiques associées à divers états d'oxydation du peptides AKPSYPPTYK pour deux états de charge.....	129

5.2 Probabilités testées sur la proline et la tyrosine afin de permettre l'identification de la protéine mfp-1 provenant de la plaque du byssus de la moule <i>Mytilus edulis</i>	144
7.1 Résumé des travaux futurs pour les projets en cours.	175

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ABC	ammonium bicarbonate (bicarbonate d'ammonium)
ACN	acétonitrile
CID	collision-induced dissociation (dissociation par collisions induites)
Cys-DA	Cystéinyl-dopamine
Cys-DOPA	Cystéinyl-DOPA
Cys-DOPAC	Cystéinul-DOPAC
DAD	diode array detector (détecteur à barette de diode)
DBS	dynamic background subtraction (soustraction dynamique du bruit de fond)
DDM	n-dodecyl B-D-maltoside
DOPA	3,4-dihydroxyphenyl-L-alanine
DTT	dithiothréitol
EIC	extracted ion current (courant d'ions extraits)
ESI	electrospray ionization (électro-pulvérisation)
FDR	false discovery rate (taux de fausses découverte)
FT-ICR	Fourier-Transform Ion Cyclotronic Resonance (résonance cyclotronique d'ion par transformé de Fourier)
GSH	glutathion
HLB	hydrophilic-lipophilic balanced (hydrophilique-lipophile balancé)
Hp-LC	high-pH reverse phase chromatography (chromatographie en phase inverse à pH élevé)
IAM	iodoacétamide
ICAT	isotope-coded affinity tag (étiquette d'affinité codée par isotope)

IT	ion trap analyzer (analyseur à trappe ionique)
LC-MS/MS	liquid chromatography coupled with tandem mass spectrometry (chromatographie liquide couplée à la spectrométrie de masse en tandem)
Lp-RP	low-pH reverse phase chromatography (chromatographie en phase inverse à bas pH)
MALDI	Matrix-Assisted Laser Desorption Ionization (ionisation laser assistée par matrice)
MeOH	méthanol
mfp	mussel foot protein (protéine de pied de moule)
MS	mass spectrometry (spectrométrie de masse)
MS/MS	tandem mass spectrometry (spectrométrie de masse en tandem)
MudPIT	Multidimensional Protein Identification Technology (technologie d'identification de protéine multi-dimensionnelle)
MYTCA	<i>Mytilus californianus</i>
MYTCO	<i>Mytilus coruscus</i>
MYTED	<i>Mytilus edulis</i>
MYTGA	<i>Mytilus galloprovincialis</i>
MYTTR	<i>Mytilus trossulus</i>
PTMP	proximal thread matrix protein (protéine matrice de fibre proximale)
rMFP-1	protéine recombinante mfp-1
RMN	résonance magnétique nucléaire
RP	reverse phase chromatography (chromatographie en phase inverse)
SCX	strong cation exchange (échange ionique par cations forts)
SPE	solid-phase extraction (extraction sur phase solide)
TFA	trifluoroacetic acid (acide trifluoroacétique)
TMP	thread matrix protein (protéine matrice de fibre)
TOF	time of flight analyser (analyseur à temps de vol)
Tris	Tris(hydroxyméthyl)aminométhane

LISTE DES SYMBOLES ET DES UNITÉS

Å	Angstrom
B	secteur magnétique magnetic sector
°C	degré centigrade (Celsius)
cc	centrimètre cube
cps	cycle par seconde
Da	Dalton
ΔE_0	différence de potentiel
E	énergie totale
E_0	énergie potentiel
E_{ex}	énergie en excès
E_{int}	énergie interne
g	gramme, force gravitationnelle
h	heure
η	dureté chimique
k_E	Constante de vitesse en phase gazeuse d'une réaction unimoléculaire
l	litre
M	molaire (molarité, mol/kg ou mol/l)
m	mètre
μ	micron
min	minute
MW	poids moléculaire
m/z	rapport masse-sur-charge
N	nombre d'électron

<i>n</i>	nombre de réplicats analytique
<i>v</i>	état vibrationnel
<i>P</i>	capacité de pic
pI	point isoélectrique
ppm	partie par million
psi	livres par pouces carré
Q	quadripôle
q	quadripôle (mode radio-fréquence seulement)
rpm	tout par minute
RT	temps de retention
<i>s</i>	degré de liberté total
s	seconde
s ⁻¹	par second
S/N	signal-sur-bruit
V	Volt
v/v	concentration volumique (volume/volume)
w/v	concentration massique (masse/volume)
w/w	concentration poids/poids
<i>z</i>	charge

RÉSUMÉ

Plusieurs organismes, comme les araignées, les vers à soie et les moules fabriquent des substances à base de protéines afin de répondre à des demandes particulières. Certains de ces bio-polymères offrent des propriétés anti-bactérienne intéressantes alors que certains possèdent des capacités élastiques, de résistance à la déformation et même d'adhésion particulièrement intéressantes. Plus précisément, dans le cas des moules, celles-ci fabriquent un polymère à base de protéines et de collagène qui surpasse la soie d'araignée. Cette «soie» connue comme le byssus offre une adhésion sous-marine qui a motivé les recherches en bioingénierie depuis les trois dernières décennies.

Les efforts des dernières années se sont essentiellement concentrés sur l'étude et la caractérisation des fibres de byssus afin, d'une part, d'en élucider la composition moléculaire et, d'autre part, de comprendre l'origine de son élasticité et adhésion. En dépit de la diversité des approches analytiques utilisées (chrystalographie des rayons X , spectroscopie par résonance magnétique nucléaire et spectrométrie de masse), l'identité des protéines produites par la moule et celles impliquées dans la fabrication de la fibre de byssus demeure encore limitée à une poignée de candidats. Les récentes avancées en matière de résolution, de sélectivité et de rapidité ont permis à la spectrométrie de masse de devenir la technique par excellence pour les analyses protéomiques. Plus précisément, la chromatographie liquide couplée à la spectrométrie de masse en tandem a largement contribué aux développements de diverses applications de la protéomique.

La présente dissertation a été organisée autour de la préparation des échantillons et du traitement des données afin d'identifier les protéines impliquées dans la biosynthèse du byssus. Dans un premier temps, différentes approches de pré-fractionnement de peptides ont été comparées à une méthode sans pré-fractionnement. Une analyse comparative entre les tissus de pied, de manteau et de byssus a été réalisée dans le but de révéler les protéines de byssus. Ensuite, une approche intégrant transcriptomique et protéomique a été réalisée à partir du transcripte de la moule *M. edulis*. Finalement, une étape plus spécifique sur une protéine recombinante a permis d'étudier la chimie de son groupement catéchol.

Mots clés : protéomique ascendante, LC-MS/MS, moule, byssus, byssogénèse

ABSTRACT

Several organism such as spiders, silkworm, and mussels make a protein-based material to meet specific demands. Some of these biopolymers offer interesting anti-bacterial properties, while some have particularly interesting flexible capacity, load-bearing resistance and even under-water adhesion. More precisely, in the case of mussels, they fabricate a polymer made of a collagen core and proteins that outperforms spider's silk. This "silk" is known as the byssus offers an interesting underwater adhesion that has motivated researches in the field on bioengineering for at least three decades.

The last years have been mostly dedicated on the characterization of the intact material in order to elucidate, on one hand, the molecular and protein composition of byssus. On the other hand, those studies have contributed to the comprehension of the mechanical and adhesion properties. Although many complementary analytical techniques have been employed (i.e. X-ray chrystalography, nuclear magnetic resonance, and mass spectrometry) the identity of the proteins involved in the byssus assembly process is still limited to a handful of candidates. Recent progress in resolution, selectivity and speed have enhance mass spectrometry to be the ideal method for in-deep proteomic analyzes. In addition, liquid chromatography coupled with tandem mass spectrometry has greatly contributed to the development of various application of proteomics.

This dissertation has been organized around sample preparation and data processing to identified proteins involved in byssus biosynthesis. First, sample preparation, including peptide pre-fractionation was compared to maximize proteome coverage. A comparative study between different tissue collected from mussels was used to highlight interesting byssus-related proteins. Then, a transcriptomic approach was integrated, using whole foot and foot gland homogenate, to increase *M. edulis* proteome. Finally, a more specific study on a recombinant version of a byssus proteins helped to understand the chemistry of the catechol group.

Keywords : bottom-up proteomics, LC-MS/MS, mussel, byssogenesis, byssus

CHAPITRE I

INTRODUCTION

Certains processus d'auto-assemblage de bio-polymères reposent sur l'interaction entre plusieurs types de macromolécules, comme les protéines, qui remplissent des fonctions importantes tant sur le point catalytiques que structural. De telle sorte que les caractéristiques observées de ces bio-polymères sont la conséquence directe de modifications ou d'interactions qu'entretiennent certaines protéines entre elles. Par exemple, les propriétés adhésives de la fibre d'araignée reposent essentiellement sur la structure secondaire des protéines sécrétées. Néanmoins, l'identification de ces protéines ainsi que leurs rôles respectifs dans les propriétés mécaniques et/ou adhésives observées sont souvent limitées par les moyens technologiques du moment.

Les dernières avancées technologiques depuis les 30 dernières années ont permis le développement de nouvelles approches analytiques axées sur l'étude de systèmes biologiques. En particulier, la chimie bio-analytique est un bon exemple d'une discipline émergente consacrée aux défis rencontrés lors d'analyses de macromolécules ou bien de composés endogènes. Du point de vue de l'analyse des protéines, l'arrivée de la spectrométrie de masse (MS) a largement contribué à la compréhension de divers systèmes biologiques (Pandey & Mann, 2000), en montrant l'aspect complémentaire que la protéomique apportait à la génomique. Plusieurs approches instrumentales, incluant la cristallographie par diffraction des rayons X et la spectroscopie par résonance magnétique nucléaire (RMN), ont participé à

l'avancement des études en protéomique structurale, en déterminant la structure tertiaire de plusieurs protéines (Sali, Glaeser, Earnest, & Baumeister, 2003). Ces avancées incluent également le développement d'outils bio-informatiques en plus du développement des bases de données essentielles à l'identification des protéines.

La présente dissertation abordera les différents aspects en chimie bio-analytique, allant de la préparation des échantillons biologiques au traitement des données, afin d'apporter une contribution dans la compréhension d'un processus de bio-synthèse relativement bien étudié chez les moules, mais encore aujourd'hui peu compris. Plus précisément, il sera question de la participation de la spectrométrie de masse dans la caractérisation de protéines provenant d'organismes peu étudiés.

1.1 Le byssus

Afin de se développer dans les endroits où les conditions marines sont difficiles (forts courants ou vagues importantes), plusieurs espèces de moules ont la capacité de sécréter une fibre extracellulaire, appelée byssus. Le byssus est un matériau complexe principalement à base de collagène et obtenu à partir d'un organe granulo-musculaire (le pied). Afin de permettre la sécrétion de la fibre, le pied s'attache d'abord au substrat puis, suivant un mouvement de contraction, la fibre de byssus est libérée de la section ventrale (*ventral groove*) selon un processus similaire à un moulage par injection. Au besoin, la moule répète ces étapes afin de s'attacher solidement sur n'importe quel type de surfaces solides. Une moule peut, en moyenne, produire entre 50 à 100 fibres pour assurer son adhésion (Harrington, Masic, Holten-Andersen, Waite, & Fratzl, 2010). La cristallographie de diffraction des rayons X au milieu des années 1990 a révélé le cœur de collagène ainsi que les différentes sections (proximale, distale et non-gradient) de la fibre de byssus (X. Qin & Waite, 1995a). Ces différentes sections sont maintenant reconnues pour attribuer des propriétés mécaniques distinctes à la fibre. La RMN a

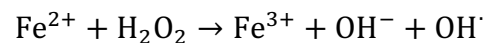
dernièrement offert un portrait plus détaillé de la structure primaire et secondaire des principales protéines du byssus (Arnold et al., 2013).

À la différence de la soie d'araignée, dont la structure semi-cristalline explique partiellement les propriétés élastiques (Simmons, Michal, & Jelinski, 2010), l'élasticité de la fibre de byssus est obtenue par les liaisons de coordinations (liens sacrificiels) entre le centre indole et un atome métallique (le plus souvent de zinc) (Byette, Laventure, Marcotte, & Pellerin, 2016; Schmitt, Politi, Reinecke, & Harrington, 2015). Lorsque la fibre subit une contrainte d'étirement, ce sont ces liens qui se rompent en premier, protégeant ainsi les liaisons covalentes (Harrington & Waite, 2007; Schmitt et al., 2015). Une fois la contrainte relâchée, ces liaisons de coordination se reforment, protégeant ainsi l'intégrité de la fibre. Comme les liaisons nouvellement formées n'impliquent pas forcément les mêmes résidus histidines que ceux précédant la contrainte, la fibre perd de sa flexibilité au fur et à mesure qu'elle subit un stress d'étirement. Par conséquent, la moule doit constamment renouveler les fibres de byssus qu'elle produit afin de rester en place.

La morphologie du pied de la moule révèle plusieurs glandes associées à la fabrication de la fibre de byssus. Plus précisément, le pied est constitué de trois sous-glandes, notamment la glande centrale (*core gland*), la glande cuticule (*cuticle gland*) ainsi que la glande de la plaque (*plaque glande*), tel que représenté par la Figure 1.1. Ces glandes sont responsables de la production des protéines précurseurs du byssus, notamment les pré-collagènes (distale, non-gradient et proximale), les protéines de la matrice enveloppant la fibre (*proximal thread matrix protein*, PTMP; *thread matrix protein*, TMP), ainsi que les protéines de pied (*mussel foot proteins*, mfps) chargées entre autres de l'adhésion, de la bio-mécanique ainsi que de la protection de la fibre.

Une modification post-traductionnelle peu commune, mais particulièrement importante, est observée dans la fibre de byssus. En effet, l'oxydation de certains

résidus tyrosines donne lieu au 3,4-dihydroxyphényl-*L*-alanine (DOPA) qui est associé aux capacités adhésives et mécaniques du byssus selon son emplacement. Cette réaction peut s'obtenir soit par voie enzymatique (tyrosinase, tyrosine hydrolase) ou via la présence d'ions radicalaires hydroxyles obtenus par la réaction de Fenton (voir ci-dessous).



Lorsque cette modification est, d'une part, localisée dans la fibre, la DOPA permet la chélation des ions métalliques, ce qui favorise les propriétés cohésives du byssus. D'autre part, lorsque cette modification est présente dans les protéines qui constituent la plaque du byssus (partie adhésive), la DOPA favorise l'attachement en milieu aqueux.

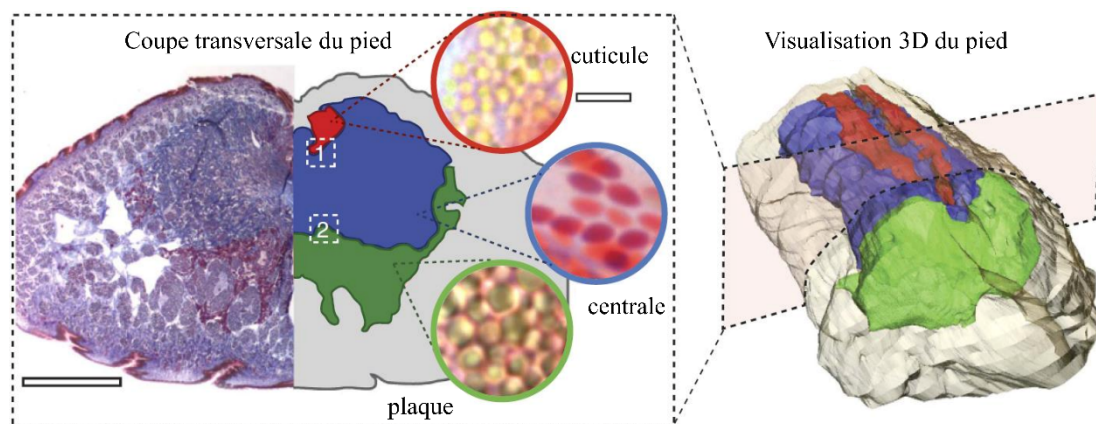


Figure 1.1 Modélisation 3D du pied de la moule *Mytilus edulis*. La glande centrale (*core gland*, section bleue), la glande cuticule (*cuticle gland*, section rouge) et la glande plaque (*plaque gland*, section verte) y sont représentées. La figure a été adaptée de (Priemel, Degtyar, Dean, & Harrington, 2017).

Présentement, seulement trois protéines byssales ont une concentration en DOPA suffisamment importante pour permettre la cohésion et l'adhésion de la fibre. Par exemple, la mfp-1, avec en moyenne de 10 à 15 mol% de DOPA, permet de chélater les ions ferriques et donne lieu au complexe DOPA-Fe (Harrington et al., 2010). Ce complexe favorise la cohésion de la pellicule qui enrobe le byssus et procure une certaine résistance mécanique à la fibre (Harrington et al., 2010). Ce complexe peut également réagir avec l'oxygène moléculaire pour former une semi-quinone, tel que proposé par Wilker (Wilker, 2010) qui d'une part, peut former des liaisons covalentes impliquant d'autres protéines ou, d'autre part, former des liaisons covalentes directement avec le substrat. Toutefois, ce mécanisme reste à démontrer. Néanmoins, l'approche proposée par Wilker pourrait fournir les bases qui expliqueraient les raisons pour lesquelles les propriétés du byssus changent selon l'ion métallique pris en charge par la DOPA (Bouhlef et al., 2017).

Aussi, la présence de la DOPA dans les protéines mfp-3 et 5, localisées dans la plaque, suggère un rôle majeur dans l'adhésion de la moule. À faible pH (sous la valeur du pK_a), la forme réduite de la DOPA est prédominante et favorise les interactions de type dipôle-dipôle. La forme réduite du catéchol permet le déplacement des molécules d'eau qui forment une mince couche à la surface des supports solides (Wei, Yu, Broomell, Israelachvili, & Waite, 2012). Également, les résidus chargés tels que l'arginine et la lysine jouent un rôle similaire en interagissant avec les ions métalliques solvantés (Wei et al., 2012). Par conséquent, la plaque peut initier un contact plus direct avec le substrat et s'y attacher plus solidement.

Le centre catéchol est toutefois sensible à l'auto-oxydation causée par une augmentation du pH et, à long terme (voir Figure 1.2), la forme dopaquinone prend peu à peu plus d'importance et minimise l'adhésion du byssus (Yu et al., 2011). Pour éviter ce genre de problème, la plaque du byssus contient une protéine riche en résidus cystéines. La mfp-6 est présentée comme offrant une protection en contrôlant le pH

localement dans la plaque du byssus. Plus précisément, les résidus cystéines sont les premiers à s'oxyder en formant des ponts disulfures. Toutefois, sous la forme dopaquinone, cette dernière peut former des liens covalents avec les groupements thiols. D'une part, cela minimise le nombre de DOPA disponibles pour l'adhésion et, d'autre part, participe à la liaison des protéines entre elles diminuant l'élasticité et l'adhésion de la fibre. La participation de liaisons covalentes entre les protéines causant l'augmentation de la plasticité (sclérisation) de la fibre reste à l'étude et sera discuté plus en détails à la section 6.

D'un autre côté, l'*ortho*-quinone est considérée comme un accepteur de Michael et par conséquent, elle peut réagir avec les chaînes latérales de certains acides aminés comme la cystéine, l'histidine et la lysine pour donner respectivement les composés Cys-DOPA, His-DOPA et Lys-DOPA (voir Figure 1.2). Bien que pour l'instant seule la modification Cys-DOPA ait été observée dans le byssus de *P. canaliculus* (Zhao & Waite, 2005), peu d'information est actuellement disponible afin de supporter l'impact de ces modifications sur les propriétés mécaniques de la fibre de byssus. Dans le même ordre d'idée, l'hydroxylysine et la desmosine sont des modifications fréquentes dans les tissus à base de collagène ou dans les bio-matériaux aux domaines élastines. Par conséquent, ces modifications peuvent également être présentes dans la fibre de byssus.

Curieusement, la moule a la capacité d'adapter la fabrication de ses fibres afin de répondre aux changements de l'environnement dans lequel elle évolue. Les principaux facteurs connus jusqu'à ce jour sont énumérés dans le tableau 1.1. Bien entendu, il ne s'agit ici que d'une représentation sommaire des facteurs connus qui influencent la byssogénèse. En raison du manque d'informations, les variables associées à la profondeur, la présence de prédateur, de polluants et l'abondance de nourriture ne sont pas incluses dans le tableau. Certains facteurs (e.g. la variabilité de certains métaux

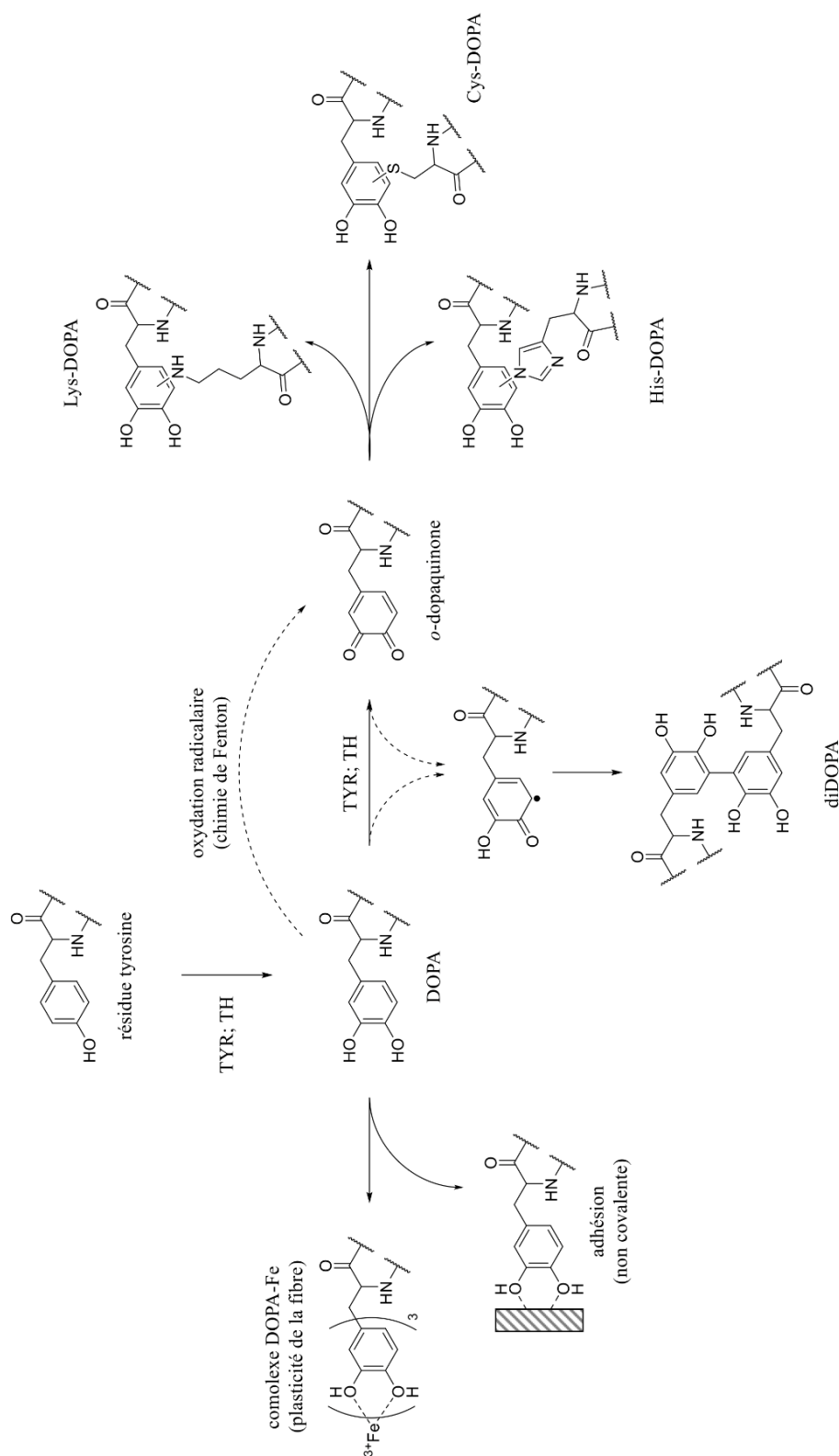


Figure 1.2 Réactions d'oxydation sur la tyrosine donnant la DOPA, responsable de l'adhésion sous-marine de la fibre de byssus. Si la DOPA est oxydée en ortho-quinone, l'efficacité de l'adhésion est diminuée et des réactions croisées entre les chaînes latérales des acides aminées nucléophiles peuvent être observées.

Tableau 1.1 Regroupement des changements bio-mécaniques de la fibre de byssus

Facteur	Conséquences
Ions métalliques	Affectent l'interaction entre les mfp non-adhésives 3 et 5. Altèrent la stabilité du cuticule du byssus. Participe à la formation de liaisons de pontages réversible dans le cuticule de la fibre (Zn et Fe). Certains métaux semblent affecter les propriétés mécaniques plus que certains autres.
Température	Favorise la déamination des arginines des protéines matrices. Augmente la flexibilité de la fibre, cependant diminue la force d'attachement.
pH et force ionique	Favorise la chimisorption de l'oxygène sur la surface et par le fait même la prolifération de bio-filmes (microorganismes). Perte d'adhésion causée par l'oxydation de la DOPA. Décroit l'endurance abrasive de la fibre (pour une diminution d'une unité de pH).
Hydrodynamique	Augmente l'apport en O ₂ ce qui contribue à l'augmentation des liaisons de pontage impliquant la dopaquinone. Produit une fibre plus résistante et moins élastique. Favorise la formation de DOPA et di-DOPA (liaison de pontage).
Géographie	Renforce la fibre (diamètre plus significatif). Offre une meilleure ténacité (adhésion). Altération de la géométrie de la fibre (forme plus hydrodynamique) pour répondre aux courants de la mer. Stimule la sécrétion de fibre en zone plus agitée.
Redox	Promouvoir la plasticité de la fibre. Production d'espèces réactive de l'oxygène via la réaction de Fenton. Cause une augmentation de la concentration du DOPA dans certaine région du byssus (e.g. preCol).
Type de surface	Peut affecter les propriétés élastiques de la fibre.
Microbes	Aide à l'absorption des ions métalliques. Utilise le byssus comme source de nourriture. Dévore le byssus à partir de l'intérieur. Utilise des sidérophores afin de retirer les ions métalliques du cuticule et ainsi entrer dans la fibre.

lourds et la température de l'eau) ont des conséquences similaires sur les propriétés mécaniques de la fibre. De ce fait, il est à prévoir que certains de ces facteurs ont potentiellement un impact additif. Néanmoins, quelques-uns de ces facteurs sont plus difficiles à contrôler afin d'y voir un lien direct avec les modifications des propriétés du byssus. Par exemple, Babarro et Carrington ont corrélié l'influence du site géographique dans lequel les moules se développent et le diamètre de la fibre produite (Babarro & Carrington, 2011). Les travaux de Séguin-Heine et al. ont montré que l'augmentation de la résistance de la fibre était aussi influencée par la concentration en molybdène et en calcium (Seguin-Heine et al., 2014). Par conséquent, l'augmentation de la résistance de la fibre de byssus ne peut être exclusivement associée à la géographie ou la présence de différents métaux lourds.

1.2 Aspects généraux et applications de la protéomique

Au début des années 1990, le *Human Genome Project* fut l'une des premières initiatives scientifiques visant à la compréhension du génome humain¹. Plus de 20 000 gènes furent identifiés et séquencés sur une période d'un peu plus de 10 ans. Ce projet permis, entre autres, plusieurs avancées majeures dans le domaine de la génétique et de la biologie moléculaire. Rapidement, les chercheurs s'accordaient sur le fait que la connaissance du génome à lui seul n'était cependant pas suffisant pour offrir une compréhension complète de certaines maladies et de certains processus biologiques (Fields, 2001).

Du point de vue de la biologie, la génomique se consacre à l'étude des gènes d'un organisme avec les séquences d'ADN comme principale analyte, alors que la transcriptomique s'occupe de la traduction des gènes en question. Au même titre que

¹ <https://www.genome.gov/human-genome-project>

le terme génomique (inventé par le généticien Tom Roderick), le terme protéomique fut initialement introduit en 1975 par O'Farrell, Klose et Scheele (P. R. Graves & Haystead, 2002) et se dédit à l'analyse des protéines. À ce moment, la séparation par électrophorèse sur gel à deux dimensions (2-GE) ne disposait pas des outils adéquats pour la caractérisation des protéines et ne servait qu'essentiellement à la visualisation de ces dernières. L'application du séquençage par la méthode de Sanger, puis ensuite du micro-séquençage par la méthode d'Edman furent considérées comme une avancée majeure dans l'identification des protéines issues d'une séparation sur gel. La figure 1.3 ci-dessus montre la relation qu'entretiennent les quatre grandes sciences omiques.

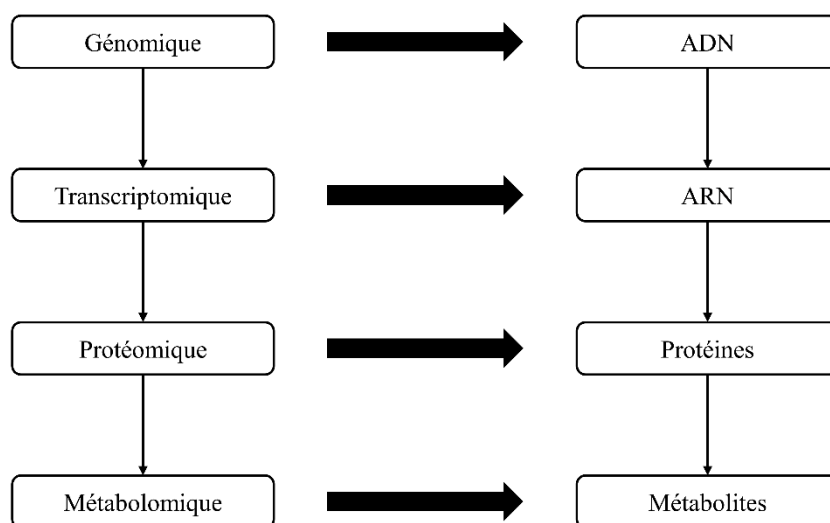


Figure 1.3 Représentation des quatre principaux domaines des sciences «omiques» ainsi que leur analyte d'intérêt.

En 1995, la protéomique fut reconnue plus officiellement comme la science et la technologie se penchant sur l'étude des protéomes. La connaissance préalable du génome rend possible le regroupement de certaines protéines ayant des fonctions similaires ou bien d'associer une fonction commune à plusieurs protéines. Depuis son introduction, la protéomique sert de base pour l'étude de certaines maladies, le

développement de thérapies ciblées, sans mentionner les multiples applications en biologie structurale (P. R. Graves & Haystead, 2002). Toutefois, comme l'ont souligné Hedge et al., l'utilisation d'une seule approche omique ne suffit malheureusement pas à proprement définir un processus biologique dans son ensemble (Hegde 2003). Présentement, les disciplines qui intègrent la transcriptomique à la protéomique demeurent prometteuses, mais se limitent au domaine clinique, avec comme thème central l'étude des différentes formes de cancers et de maladies cardiovasculaires.

De nos jours, la protéomique se voit être une science complémentaire à la génomique et se définit plus précisément comme étant la description du caractère dynamique des protéines d'un système biologique (Liebler, 2002a). La présence de modifications co et post-traductionnelles (réversibles ou irréversibles), dont le nombre est estimé à plus d'un million, contribue à la diversité du protéome (Jensen, 2004). Il est intéressant de noter que la majorité des modifications réversibles jouent un rôle dans la transduction du signal cellulaire, alors que les modifications irréversibles sont particulièrement présentes dans la reconnaissance biologique (D. J. Graves, Martin, & Wang, 1994).

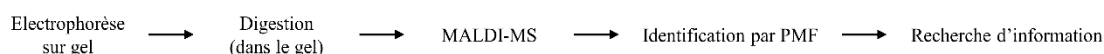
1.3 Approches expérimentales

Les premières approches méthodologiques en protéomique impliquaient une séparation par électrophorèse multi-dimensionnelle sur gel de poly-acrylamide (2D-PAGE), tel que montré par la figure 1.4a. Dans ce type d'approche, la séparation des protéines se fait d'abord par focalisation isoélectrique. Une fois terminé, le gel est retourné à 90 degrés et les protéines sont à nouveau séparées, cette fois, selon leur masse (Klose, 1975; O'Farrell, 1975). Par la suite, les protéines sont directement digérées par la trypsine (dans le gel) pour être ensuite analysées par MALDI-MS. Malgré qu'un grand nombre de protéines ont pu être identifiées par électrophorèse sur gel, le temps nécessaire à la préparation des gels, le manque de reproductibilité ainsi

que le faible domaine de linéarité ont eu raison de cette méthodologie (P. R. Graves & Haystead, 2002).

L'arrivée de la chromatographie liquide (LC) a rapidement remplacé l'électrophorèse sur gel comme méthode principale de séparation. Cette approche dite sans gel (*gel-free*, figure 1.4b) nécessite que les protéines soient d'abord digérées en solutions, puis séparées sur une colonne chromatographique. Plus tard, l'électrophorèse capillaire a aussi été employée comme méthode séparative alternative à la chromatographie liquide. Contrairement à la LC, qui nécessite des quantités de l'ordre du μg -mg, l'électrophorèse capillaire peut séparer des quantités aussi faibles que des ng de peptides ou de protéines. Néanmoins, l'utilisation de solutions tampons pour permettre la migration des composés reste un défi important pour le couplage avec la spectrométrie de masse.

a) approche par séparation sur gel (*gel-based separation*)



b) approche par séparation sans gel (*gel-free separation*)

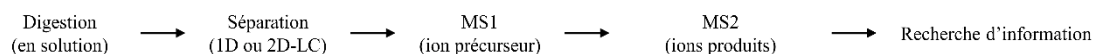


Figure 1.4 Comparaison entre les approches en protéomique impliquant la séparation sur gel et sans gel couplées à la spectrométrie de masse.

Les protéines peuvent être étudiées selon deux axes, avec un exemple de la préparation des échantillons simplifiée pour chaque axe présenté par la figure suivante. Les derniers progrès en matière de sources d'ionisations à pression atmosphérique (API) compatibles à l'étude des macromolécules ont contribué grandement à l'analyse des

peptides et des protéines issues de matrices biologiques. Par conséquent, deux scénarios sont envisageables pour l'étude et la caractérisation des protéines, comme le présente la Figure 1.5 suivante. Dans un premier temps, une méthodologie par protéomique ascendante (*bottom-up* ou *shotgun*) (figure 1.5a) est généralement favorisée pour les études de type profilage de protéines. Les protéines sont d'abord digérées par une enzyme (ou une combinaison d'enzymes) avant de procéder à la détection par LC-MS/MS. La trypsine est l'enzyme généralement préférée pour ce genre d'approche, d'une part pour la sélectivité du clivage du squelette peptidique du côté C-terminal suivant les résidus arginines (R) et lysines (K, sauf si suivit par une proline). D'autre part, une digestion par la trypsine demeure reproductible et la longueur des peptides est adéquate pour les analyses par MS/MS. D'autres enzymes plus sélectives que la trypsine, comme la chymotrypsine ou la protéase V8 (Glu-C), peuvent être utilisées. À l'inverse d'endoprotéases sélectives, la pepsine peut aussi être employée, mais reste moins performante en raison de sa sélectivité plus faible et des peptides plus courts qu'elle engendre.

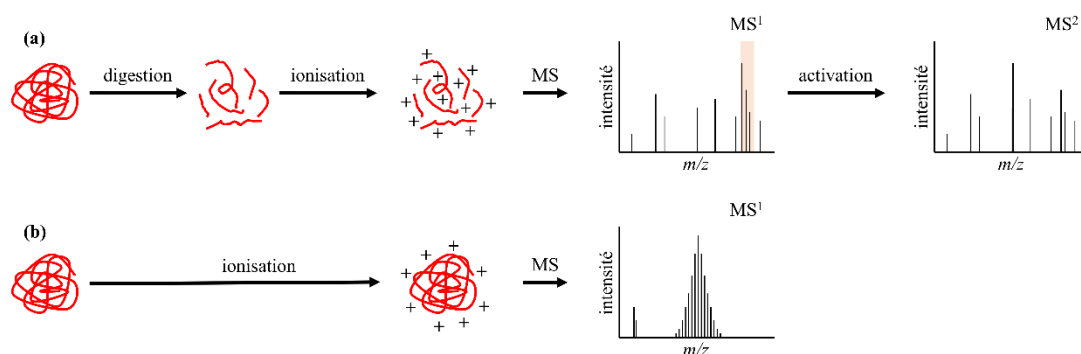
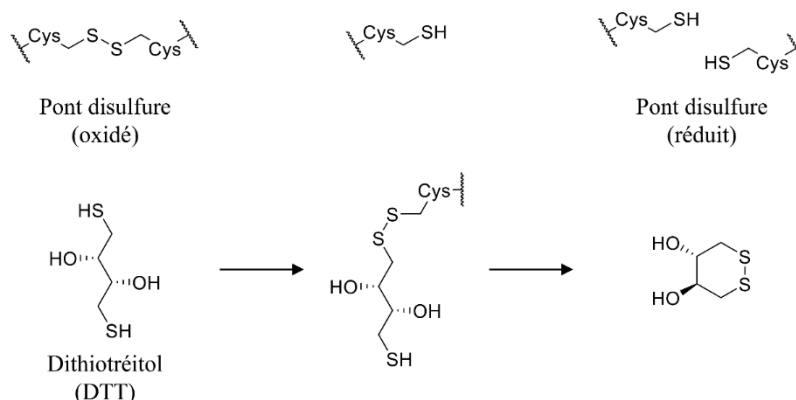


Figure 1.5 Représentation sommaire de la préparation des extraits de protéines dans le but d'une analyse de type ascendante (*bottom-up*, schéma du haut) ou descendante (*top-down*, schéma du bas).

La préparation des échantillons en prévision d'une digestion demande globalement une étape d'alkylation réductrice, avec comme exemple l'utilisation du dithiothreitol (DTT) suivit de l'iodoacétamide (IAM). Comme le montre la Figure 1.6a, cette étape permet d'abord une dénaturation sommaire de la protéine par la réduction des ponts disulfures afin de laisser place au groupement thiols sous sa forme réduite. Ensuite, afin d'éviter la formation spontanée de nouveaux ponts disulfures, l'IAM est ajouté en tant qu'agent alkylant en modifiant les groupements thiols (seconde étape de la Figure 1.6). Ces étapes sont préférablement réalisées à un pH légèrement basique afin de ne pas interférer avec l'alkylation. Comme le pH optimal de la trypsine joue entre des valeurs allant de 8.5 à 8.8, la digestion peut ensuite être effectuée directement. Également, l'utilisation d'enzymes complémentaires moins spécifiques, telle que la pepsine par exemple, permet l'augmentation de la couverture globale du protéome en permettant la détection de peptides non tryptique (Golizeh & Sleno, 2013). Selon les auteurs, une digestion en série n'était pas forcément bénéfique, alors qu'une digestion en parallèle offrait plus d'avantages sur la couverture globale du protéome ainsi que sur le nombre total de peptides séquencés. De plus, une digestion en parallèle demeure particulièrement populaire en offrant de meilleures options pour la quantification. Une fois la digestion terminée, la présence de l'enzyme résiduelle, des sels en excès ainsi que les peptides trop petits pour la MS/MS doivent être retirés du mélange. Plusieurs approches d'extraction, comme l'extraction liquide-liquide ou par précipitation, aident au nettoyage des échantillons. Dans le cas actuel, l'extraction sur phase solide (*Solid-Phase Extraction*, SPE) offre un meilleur taux de récupération des peptides (Namera & Saito, 2013). Puisque les peptides portent aussi des charges, il est possible de les mettre à profit lors de l'étape SPE.

a) réduction des ponts disulfides par le dithiothréitol



b) alkylation des thiols librent par l'iodoacétamide

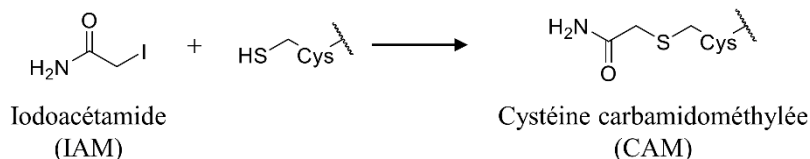


Figure 1.6 Schémas montrant les étapes de réduction des ponts disulfures des protéines par le dithiothréitol (a), suivit de l'alkylation par l'iodoacétamide (b) pour empêcher la formation spontanée de nouveaux ponts disulfures.

L'approche *bottom-up* est généralement favorisée pour étudier les protéines d'un organisme en raison de sa flexibilité. En effet, il est plus simple d'étudier les protéines au niveau des peptides puisqu'il est plus facile de fragmenter des peptides que des protéines pour en extraire l'information. Il est cependant impossible d'étudier les structures tertiaires et quaternaires en protéomique *bottom-up*, puisque les protéines doivent être dénaturées pour que la digestion fonctionne. Dans le cas particulier où une modification post-traductionnelle est le centre d'intérêt de l'étude, la digestion des protéines pose un défi supplémentaire quant à l'identification des peptides modifiés. En effet, ces derniers se trouvent dilués dans l'échantillon et demande par conséquent

une préparation particulière afin de maximiser les chances d'être fragmentés (Liebler, 2002b).

Le deuxième axe de la protéomique consiste à l'analyse de protéines intactes (sans digestion préalable), comme simplifié par la Figure 1.5b. Cette approche descendante (*top-down*) offre une meilleure productivité et une couverture de la séquence primaire nettement supérieure à ce que peut accomplir la protéomique *bottom-up* (Armirotti & Damonte, 2010). De plus, puisque les modifications post-traductionnelles, essentiellement la phosphorylation et la glycosylation, peuvent être observées, les applications *top-down* bénéficient d'une popularité significative pour des applications cliniques (Calligaris, Villard, & Lafitte, 2011).

1.4 Application de la LC-MS/MS à la protéomique

La spectrométrie de masse a initialement été développée par le physicien Anglais Joseph John Thompson en 1897. Les travaux de Thompson se sont concentrés sur l'interaction des particules dans un champs électrique. À ce moment, les ions chargés étaient obtenus à partir de décharges électriques dans un tube sous vide et étaient ensuite séparés selon leur ratio charge élémentaire-sur-masse (e/m). Le spectrographe de masse obtenu sur une plaque photographique se composait de paraboles résultant d'ions ayant percuté la plaque à différentes énergies. Depuis le spectromètre de masse développé par J. J. Thompson, plusieurs appareils ont vu le jour au cours des années qui ont suivi. Cependant, ce n'est qu'au tournant des années 1950 que le premier appareil commercial a été disponible. De nos jours, les spectromètres de masse modernes sont constitués, 1) d'une chambre d'ionisation, 2) d'un ou plusieurs analyseurs de masse et 3) d'un détecteur, tel que présenté par la figure 1.7 ci-dessous.

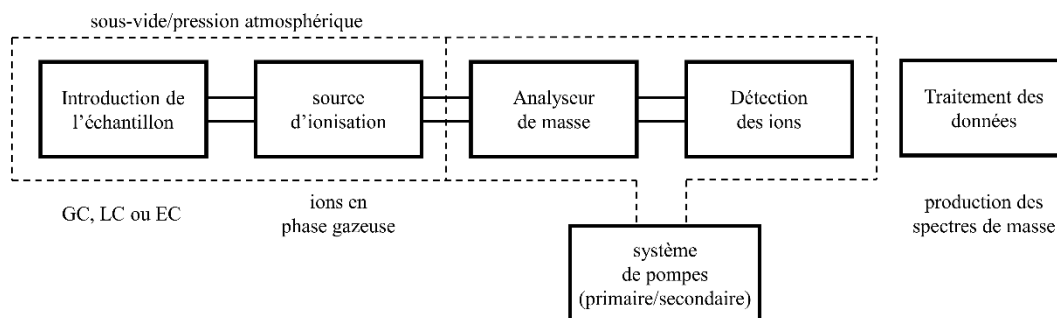


Figure 1.7 Schéma général d'un spectromètre de masse. Selon le type de source utilisée, il est possible que celle-ci soit opérée à pression atmosphérique ou sous-vide.

L'ionisation de substances neutres (médicaments, métabolites, peptides, protéines) s'obtient par le transfert ou le retrait d'espèces chargées (électrons, protons, hydrures ou adduits). Le mécanisme derrière le processus d'ionisation dépend essentiellement du type de source employé. Par exemple, l'énergie des électrons émis lors de l'ionisation par impact électronique (EI, *Electron Ionization*) est le principal facteur influençant l'efficacité de l'ionisation. Dans le cas d'analyse par EI, les molécules doivent être volatiles et stables à haute température, ce qui pose un défi considérable concernant l'analyse des peptides et des protéines. La fin des années 1980 fut marquée par l'arrivée de sources favorables à des applications en protéomique. Parmi ces types de sources, mentionnons la source d'ionisation MALDI (*Matrix-Assisted Laser Desorption Ionisation*) développée en 1988 par Karas et Hillenkamp (Karas & Hillenkamp, 1988) ainsi que la source par électro-nébulisation (*electrospray ionization*, ESI) développée par John B. Fenn (Fenn, Mann, Meng, Wong, & Whitehouse, 1989) qui lui valu le prix Nobel de chimie en 2002. En 1989, John Fenn montra qu'il était désormais possible d'analyser les macromolécules sans risque de décomposition, à partir d'une méthode d'ionisation par électro-nébulisation (Fenn et al., 1989). De manière générale, la distribution des charges à la surface d'une gouttelette se décrit selon le théorème de Gauss. Selon le principe décrit par Fenn et

revisit  par Konermann, la r duction de la gouttelette entra nera une augmentation de la densit   lectronique, causant par la m me occasion l'augmentation de la r pulsion coulombique.   ce moment, la gouttelette atteint la limite de Rayleigh   laquelle le transfert de charge se produit selon diff rents m canismes d pendant de la nature de l'analyte. Cette m thode a l'avantage de cr er des ions multicharg s ($z > 2$) de telle sorte que certains analyseurs de masse ayant un domaine m/z plus restreint peuvent  tre employ s. N anmoins, la production d'ions en phase gazeuse demeure un m canisme complexe qui d pend de la taille des gouttelettes form es et des diff rents degr s de conformation native dans le cas des prot ines, de l' quilibre  lectrostatique ainsi que de l' jection pr matur  des analytes de la gouttelette (Kearle & Tang, 1993; Konermann, Ahadi, Rodriguez, & Vahidi, 2013)..

Une grande vari t  d'analyseur de masse est d sormais disponible (voir tableau 1.2) ce qui a contribu  au d veloppement des applications de la spectrom trie de masse en prot omique. Pour la plupart de ces analyseurs, leur principe de base repose sur la transmission d'ions ayant une trajectoire stable   travers un champs  lectrique et magn tique variable afin d'atteindre la cellule de d tection. Une fois amplifi , le signal est converti en un rapport entre la masse et la charge de l'ion (rapport m/z) pour ainsi composer le spectre de masse.

Dans le cas de l'application en tandem, qui sera abord e plus bas, deux analyseurs de masse sont dispos s l'un   la suite de l'autre (disposition spatiale). La disposition physique de deux analyseurs quadrup le, par exemple, est probablement la meilleure mani re de visualiser un appareil dans la disposition spatiale. Dans ce type d'instrument, les  v nements concernant la s lection d'un ion pr curseur suivi de la d tection des ions fragments se d rouleront dans deux emplacements distincts. Pr cis ment, les ions sont donc transmis du premier analyseur de masse vers une cellule de collision pour y  tre fragment s et les ions sont ensuite achemin s vers le second analyseur de masse qui les guidera vers le d tecteur. Inversement, les instruments tels

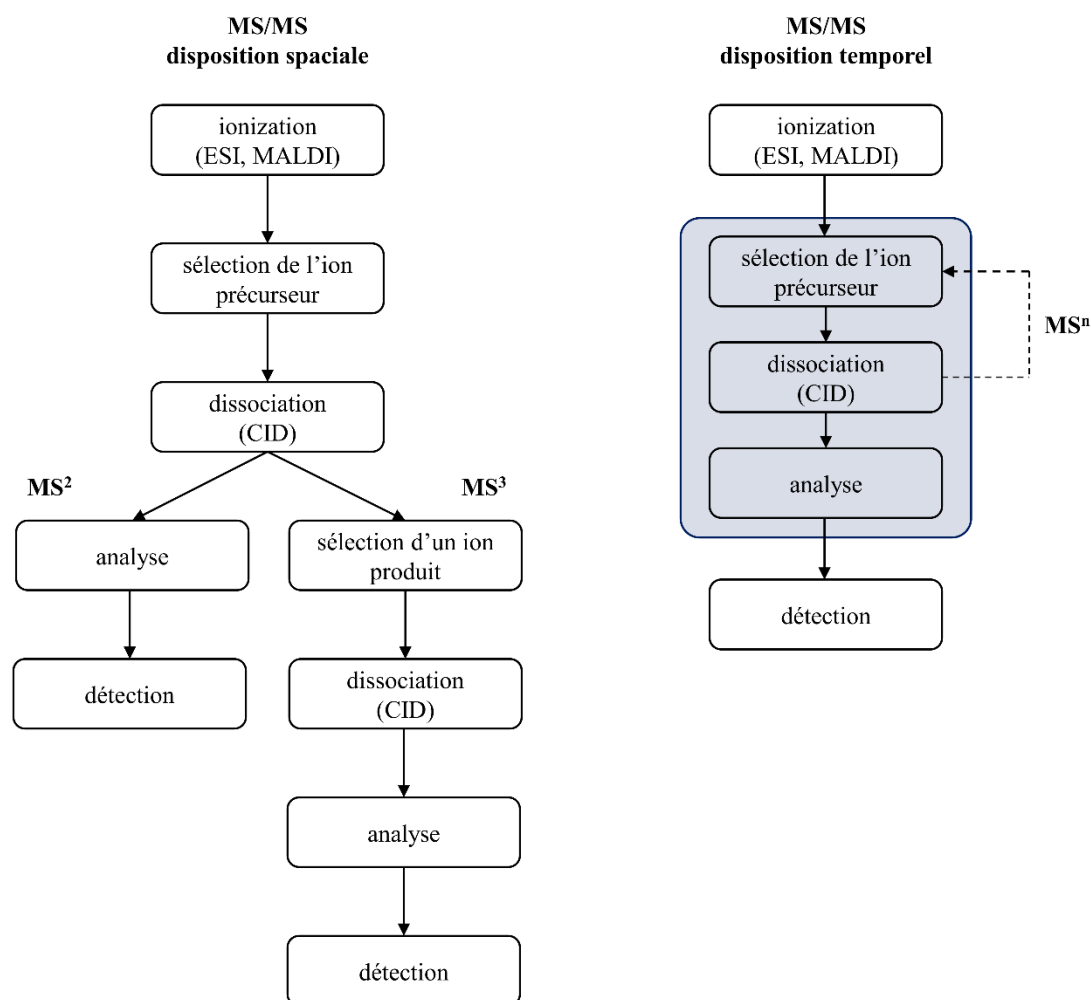


Figure 1.8 Comparaison entre un appareil en mode spatial (gauche) et temporel (droite) employé en spectrométrie de masse en tandem. La figure a été adaptée de *Mass Spectrometry*, Gross J.H, 2nd edition.

que QqIT, FT-ICR et Orbitrap sont considérés comme étant des instruments temporels (McLafferty, 1981), puisque ceux-ci impliquent que l'ion précurseur soit excité sur une période de temps avant d'induire une fragmentation. L'excitation et la fragmentation ont tous deux lieu dans le même emplacement. Par conséquent, la résolution ainsi que la qualité du spectre MS/MS est connue depuis le début des années 1990 comme étant

influencée par le mode (spatiale ou temporel) opéré par l'instrument (J. V. Johnson, Yost, Kelley, & Bradford, 1990). La figure 1.8 illustre la différence entre les dispositions temporelle et spatiale.

L'analyseur quadripôle demeure encore de nos jours le plus populaire, malgré une faible résolution ainsi qu'un domaine de masse relativement limité (voir Tableau 1.2). Cet analyseur remplit plusieurs fonctions, dont celle de filtre de masse et de cellule de collision. Dans le premier cas, la transmission des ions repose sur la stabilité de la trajectoire oscillante (hélicoïdale) de ces derniers, ce qui permet l'élimination de tout ions instables. La combinaison entre l'augmentation d'un courant continu et de radio-fréquences génère des champs alternatifs qui permettent un «balayage» d'une gamme de m/z donnée. Aussi, lorsque l'analyseur quadripôle est opéré en mode radio-fréquence seulement, l'efficacité de la transmission des ions joue un rôle particulièrement significatif en spectrométrie de masse en tandem. Cette fonctionnalité permet de focaliser le rayon d'ions afin de maximiser le transfert de ceux-ci du premier analyseur de masse vers le second. D'autres dispositifs, dont les analyseurs à temps de vol (*time-of-flight*, TOF), les cellules FT-ICR ainsi que les trappes ioniques permettent la transmission de la quasi-totalité des ions produits durant l'ionisation.

Un autre analyseur de masse, l'analyseur TOF, est aussi devenu de plus en plus courant sur les systèmes modernes. Le principe fut introduit par Stephens au milieu des années 1940. Contrairement au quadripôle, l'analyseur TOF opère sans la présence de champs électrique ou magnétique. Le temps de vol t , peut être calculé à partir de l'équation 1 selon une masse m , une énergie cinétique KE ainsi qu'une distance de vol D connus. La séparation des ions à l'intérieur du tube de vol est directement influencée par l'énergie cinétique initiale de l'ion à son entrée dans le tube de vol sous vide.

$$t = \left(\frac{m}{2KE} \right)^{1/2} D \quad \text{Équation 1}$$

Les premières applications des analyseurs à temps de vol ont révélé une faible résolution de masse (entre 300 et 800) en plus d'être dépendant du type de source employée et de la quantité de matériel introduit dans l'appareil (Cotter, 1997). Vers la fin des années 1980, l'introduction d'un réflectron a permis d'atteindre une résolution d'environ 25 fois supérieure au concept original (Grix et al., 1988). Avec une transmission d'ions de près de 90% et sans aucune restriction au terme du domaine de masse, les applications de l'analyseur TOF ont connu une augmentation significative au cours des années qui ont suivi son introduction.

Tableau 1.2 Caractéristiques des principaux analyseurs de masses disponibles

	Q	TI	TOF	TOF (réflectron)	FT-ICR	OT
limite de masse (m/z)	4 000	6 000	>100 000	10 000	30 000	50 000
résolution	2 000	4 000	5 000	20 000	500 000	100 000
précision ^a	100 ppm	100 ppm	200 ppm	10 ppm	< 5 ppm	< 5 ppm
transfert d'ions	continu	pulsé	pulsé	pulsé	pulsé	pulsé
pression (Torr)	10^{-5}	10^{-3}	10^{-6}	10^{-6}	10^{-10}	10^{-10}
disposition possible	QqQ			TOF/TOF (ou PSD)		
mode MS/MS	ions produits ions précurseurs pertes neutres	MS ⁿ ions produits		ions produits	MS ⁿ ions produits	
énergie de collision	faible	faibles		faible ou élevé	faible	

Q : quadripôle; TI : trappe ionique; TOF : temps-de-vol (time-of-flight); FT-ICR : résonance cyclotronique d'ions par transformé de Fourier (Fourier-Transform Ion Cyclotron Resonance); OT : OrbitrapTM

^aprécision mesurée à la mi-hauteur à m/z 1 000

La spectrométrie de masse en tandem est une approche à plusieurs étapes qui permet la collecte d'information structurale. En général, elle implique au moins deux analyseurs de masse (dans le cas d'un appareil spatial) placés l'un à la suite de l'autre. Au moment de l'activation de l'ion précurseur, celui-ci absorbe une certaine quantité d'énergie dans le but d'induire une fragmentation. Les différentes méthodes d'activation possibles impliquent généralement la collision entre des molécules neutres (*collision-induced dissociation*, CID), la capture ou le transfert d'électrons (*electron capture/transfert dissociation*, ECD/ETD) et finalement l'absorption de photons infrarouges (*infrared multiphoton photodissociation*, IRMPD) ou ultra-violet (*ultraviolet photodissociation*, UVPD).

La méthode d'activation par CID demeure, encore de nos jours, l'approche la plus populaire en protéomique ascendante. Suivant la fragmentation, le spectre de masse obtenu se compose d'une série d'ions dont l'intensité du ratio m/z correspond à leur abondance respective. En regroupant les ions formés suivant la direction C-terminale (série y) ou N-terminale (série b) du peptide, il est possible d'élucider la structure primaire (séquençage *de novo*). Au-delà de la série d'ion y/b , d'autres ions peuvent également aider à la reconstruction du peptide, tel que montré par la Figure 1.9, comme les ions a_1 et a_2 par exemple. Néanmoins, il est important de noter que ces ions ne sont pas toujours présents dans les spectres MS/MS et que la série y/b prédomine dans le cas de la fragmentation par CID. Il est possible, quelques fois, d'observer certains acides aminés individuels initialement présents dans le peptide. Cependant, la présence de ces ions dépend fortement de la séquence primaire et ces derniers sont généralement difficilement distinguables du bruit de fond. La figure suivante illustre les principaux ions formés lors de la fragmentation d'un peptide.

Introduit par Cook au début des années 1980, le développement de spectromètre de masses hybrides a permis le développement de plusieurs types d'appareils (QqTOF, QIT/TOF et Orbitrap) au cours des décennies qui ont suivi (Glish

& Burinsky, 2008). La combinaison de différents analyseurs de masse représente vraisemblablement une autre avancée majeure puisque ces systèmes tirent partie des points forts de chacun des analyseurs de masse en cause. Par exemple, dans les

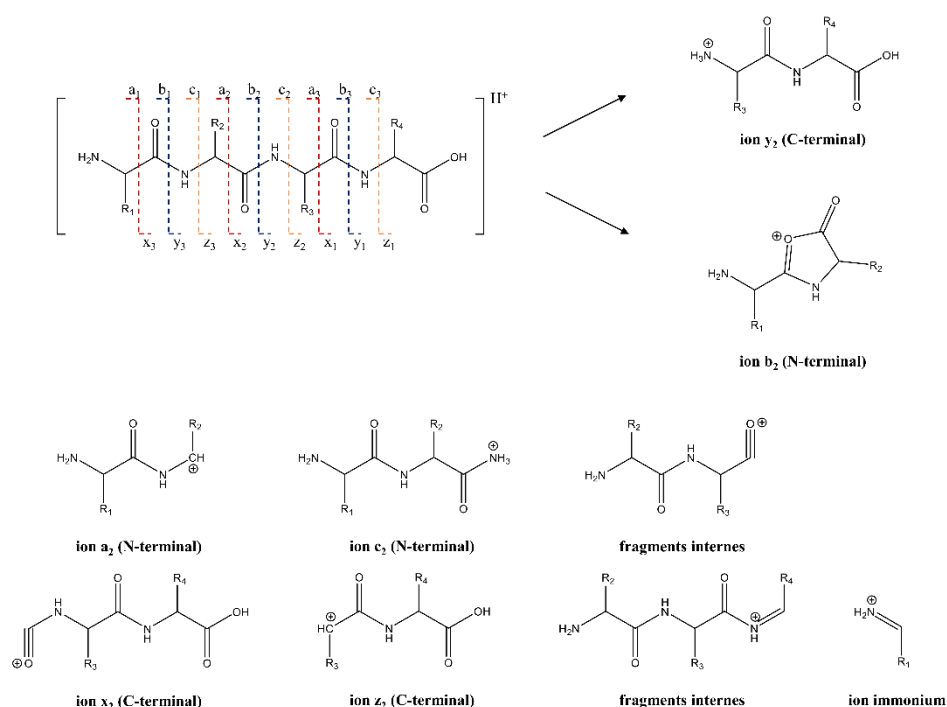


Figure 1.9 Schéma représentant la fragmentation d'un peptide en montrant les liens brisés selon la technique d'activation employée. La structure des ions obtenus par CID en mode positif est également montrée.

appareils de type QTOF, le quadripôle agit comme filtre de masse et permet ainsi une meilleure sélectivité au moment de la sélection de l'ion précurseur. Lorsqu'employé en mode radio-fréquence seul, il performe dans la transmission des ions vers le second analyseur de masse. L'analyseur à temps de vol (*time-of-flight*, TOF), quant à lui, donne accès à la masse exacte, ou mono-isotopique. Celui-ci peut atteindre des résolutions de plus de 100 000 Th (voir Tableau 1.2) selon le mode sous lequel il est employé. Parallèlement, la rapidité à laquelle ces appareils peuvent enregistrer les

spectres de masse ont permis des applications plus poussées en chromatographie (liquide et gazeuse). Tel que souligné par Holland, l'enjeu principal demeure le compromis entre la rapidité et la quantité d'ions détectés afin de ne pas compromettre la précision des ions mesurés (Holland et al., 1983).

À titre d'exemple, le schéma présenté à la Figure 1.10 est celui d'un instrument de type QqTOF (AB Sciex 5600 TripleTOF™). Dans une étude comparative publiée en 2010 dans laquelle différents appareils hybrides ont été considérés, les auteurs ont souligné la sensibilité supérieure obtenue avec un instrument de type Q-TOF par rapport à d'autres appareils hybrides comme les TI-Q ainsi que les trappe ionique linéaires-quadrupôle (Rousu, Herttuainen, & Tolonen, 2010). Dans le même ordre d'idée, toujours selon Rousu, un appareil combinant deux analyseurs complémentaires, jusqu'à un certain niveau, offrait une bien meilleure performance quant à l'identification de divers composés qu'un instrument classique (QqQ). De plus, les auteurs ont montrés que les appareils hybrides offrent une déviation sur la masse de calibration plus faible (stabilité sur la masse), en plus d'une linéarité supérieure, en comparaison aux analyseurs individuels (Eichhorn, Pérez, & Barceló, 2012).

Sur les instruments modernes, l'enregistrement des données provenant de la fragmentation des peptides se fait selon deux modes. Dans le mode d'acquisition dépendante des données (*Data-Dependant Acquisition*, DDA), l'instrument sélectionne en temps réel les ions précurseurs respectant des critères préalablement établis (intensité minimum, exclusion isotopique etc.). Cette approche demeure particulièrement utilisée lors d'études de profilage en protéomique *bottom-up* dans lesquelles la thématique globale reste l'identification de plusieurs protéines. Toutefois, la présence d'artéfacts plus intenses peut venir entraver la prise de décision de l'instrument et ultimement complexifier le spectre MS/MS obtenu. Cependant, l'activation des ions précurseurs est grandement limitée par les critères DDA énumérés plus haut. Par conséquent, seulement 30% des peptides contenus dans un échantillon.

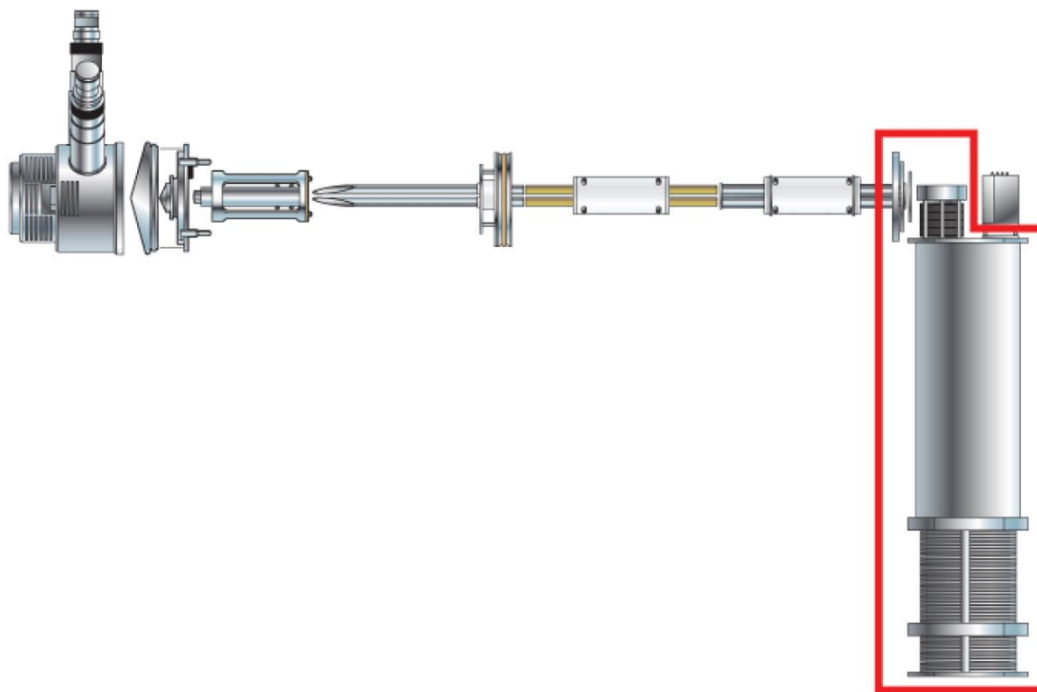


Figure 1.10 Schéma de l'optique d'ion dans le spectromètre de masse hybride quadrupôle-temps de vol TripleTOF™ 5600 de Sciex (disponible en ligne sur sciex.com).

sont activés, puis détectés adéquatement (Michalski, Cox, & Mann, 2011). Kreimer et al. ont montré qu'en adaptant les critères DDA en temps réel (via un algorithme intégré à l'environnement Thermo) il était possible d'accroître d'environ 30% le nombre de peptides détectés provenant d'une mélange complexe (Kreimer et al., 2016). Dans le second mode, l'acquisition indépendante des données (*Data-Independent Acquisition*, DIA), la fenêtre m/z est divisée en fenêtres plus petites, généralement de 1 Da. À chaque temps de cycle (temps nécessaire à balayer la gamme m/z spécifiée), l'instrument balaie chacune des fenêtres et fragmente les ions se trouvant dans ces fenêtres. Plusieurs spectres MS/MS sont alors enregistrés, sans discrimination de l'intensité du précurseur. Ainsi, certains ions précurseurs moins intenses ont une meilleure chance d'être

sélectionnés pour produire un spectre MS/MS. De ce fait, l'influence d'artéfacts est grandement minimisée.

1.5 Chromatographie 1D *versus* 2D

Dans une approche générique en protéomique *bottom-up*, l'emploi de mécanismes de séparation complémentaires, comme l'échange ionique et la phase inverse par exemple, est probablement la façon la plus efficace et la plus populaire. Gidding a publié 1984 ce qui allait devenir les bases de la chromatographie liquide multi-dimensionnelle (2D-LC) (Giddings, 1984). L'auteur a mis l'accent sur la combinaison et la compatibilité des modes de séparation, en plus de la contribution de chaque dimension sur la capacité de pic et des restrictions physico-chimiques imposées par les analytes étudiés (Giddings, 1984). Plus précisément, la capacité de pic d'une approche 2D-LC (P_{2D}) est le produit entre la capacité de pic entre la première et la seconde dimension (P_1 et P_2 , respectivement) telle que présenté par l'équation 2. Dès lors, deux démarches sont possibles pour la séparation des peptides, l'une directe (*on-line*) et l'autre indirecte (*off-line*).

$$P_{2D} = P_1 \times P_2 \quad \text{Équation 2}$$

D'abord, dans une approche directe, les deux phases stationnaires, généralement une résine SCX et C18 poreuse remplissent un capillaire qui joue le rôle de la colonne et fait office d'interface entre le LC et le spectromètre de masse. À titre d'exemple, Washburn et al., en 2001, (Washburn, Wolters, & Yates, 2001) ont démontré l'efficacité de leur méthodologie MudPIT (*Multidimensional Protein Identification Technology*) dans l'étude du protéome de *Saccharomyces cerevisiae* avec une augmentation de plus de 80% des protéines identifiées en comparaison aux études traditionnelles sur gel. Entre autres, les auteurs ont souligné l'étonnante habileté de leur méthode à identifier plusieurs protéines, notamment des protéines membranaires peu

abondantes, normalement ignorées par les approches séparatives sur gel. Différents changements ont depuis été apportés, proposant d'une part une séparation basée sur le point isoélectrique (pI) des peptides ne nécessitant ainsi que peu ou pas de sels lors de l'élution (Dai, Shieh, Sheng, Zhou, & Zeng, 2005) ou d'autre part, l'utilisation de résines anionique et cationique (forte et faible) placées en tandem ou mélangées (Motoyama, Xu, Ruse, Wohlschlegel, & Yates, 2007). Bien que ces améliorations aient apporté de nets avantages sur la détection des peptides par rapport à l'approche MudPIT classique (Motoyama et al., 2007), améliorant l'identification des protéines à plus de 28% (Dai et al., 2005), les méthodes directes par 2D-LC directes sont, en partie, largement critiquées quant à la dépendance et la difficulté d'optimiser individuellement chaque dimension. Également, une instrumentation particulièrement coûteuse et nécessitant un équipement spécifique dans certains cas, a favorisé le développement des approches indirectes.

Dans l'approche indirecte (*off-line*), le mélange de peptides est divisé en plusieurs fractions lors de la première dimension. Les fractions collectées sont ensuite injectées et séparées sur une seconde colonne, généralement une C18, puis analysées par spectrométrie de masse. Le choix des modes de séparation lors de la première dimension est important puisque celui-ci influencera grandement l'orthogonalité de la méthode. Gilar et al. (Gilar, Olivova, Daly, & Gebler, 2005a) ont proposé en 2005 une approche rapide employant le temps de rétention normalisé ($RT_{i(norm)}$) afin d'évaluer la complémentarité de la première dimension. Brièvement, la différence entre le temps de rétention de chaque peptide (RT_i) et du peptide le plus polaire (RT_{min}) est divisée par la différence entre le peptide le plus apolaire (RT_{max}) et le plus polaire (RT_{min}). La résultante de cette équation montre qu'une importante corrélation entre les deux dimensions signifie une faible orthogonalité alors qu'une distribution des peptides plus erratique (faible corrélation) montre une complémentarité particulièrement forte.

$$RT_{i(norm)} = \frac{RT_i - RT_{(min)}}{RT_{(max)} - RT_{(min)}} \quad \text{Équation 3}$$

Dans un autre ordre d'idée, Di Palma et al. ont comparé les méthodes de pré-fractionnement les plus courantes en 2D-LC-MS/MS en termes de leurs aptitudes respectives pour la détection de peptides (Di Palma, Hennrich, Heck, & Mohammed, 2012). Golizeh et al. ont aussi offert une approche similaire pour l'emploi de diverses méthodes de pré-fractionnement par échange ionique (Golizeh, Schneider, Ohlund, & Sleno, 2015). Tel que souligné par les auteurs, l'échange de cations forts est une approche particulièrement efficace, mais qui requière néanmoins une étape supplémentaire afin de retirer les sels non volatiles afin d'éviter que ceux-ci n'interfèrent lors du processus d'ionisation et de détection des peptides.

L'usage de la chromatographie en phase inverse pour le pré-fractionnement des peptides a été introduite par Dai et Mommen et offrait une alternative aux fractionnements par échange ionique (Dai et al., 2005; Mommen, Meiring, Heck, & De Jong, 2013). Le potentiel de cette approche provient du fait que peu ou pas de sels non volatiles sont inclus dans la préparation des phases mobiles. En effet, au contraire de la SCX, les phases mobiles tamponnées utilisées lors d'une séparation en phase inverse contiennent généralement une concentration de 10 mM en sel, minimisant ainsi les risques d'interférence au moment de l'ionisation. Par conséquent, les fractions peuvent être directement analysées avec un minimum de risque par rapport aux effets de matrice lors de l'ionisation.

En phase inverse, la différence de pH entre la première et la seconde dimension est d'une importance cruciale sur l'orthogonalité de la méthode. Par conséquent, la semi-orthogonalité entre l'étape de la collection des fractions et l'analyse MS fut présentée comme un désavantage majeur pour un pré-fractionnement employant une phase inverse (Gilar et al., 2005a; Hao, Ren, Dutta, & Sze, 2013; Song et al., 2009).

D'ailleurs, plusieurs groupes ont adressé cette problématique et suggèrent que certaines fractions non adjacentes (voir Figure 1.11) ou provenant de la même zone d'élution soient mélangées ensemble (Dwivedi et al., 2008; H. Lee et al., 2016; F. Yang, Shen, Camp II, & Smith, 2012). Dans un autre ordre d'idée, Yeung et al. ont proposé une évaluation plus mathématiques de l'orthogonalité de 16 approches 2D-LC en se basant sur leur corrélation respective (Yeung et al., 2020).

Malgré les bénéfices concrets sur l'augmentation de l'orthogonalité et de la couverture globale du protéome étudié, le choix de procéder à une analyse multidimensionnelle ou non demande réflexion. En premier lieu, bien que les analyses 2D-LC permettent la détection et l'identification d'un plus grand nombre de peptides et de protéines respectivement, la quantité de fractions à collecter demeure encore un paramètre arbitraire. En second lieu, les pertes d'échantillons liées au transfert d'une fraction à l'autre sont fortement à prévoir. Par conséquent, cet aspect doit être pris en considération lors d'analyses quantitatives (Di Palma et al., 2012).

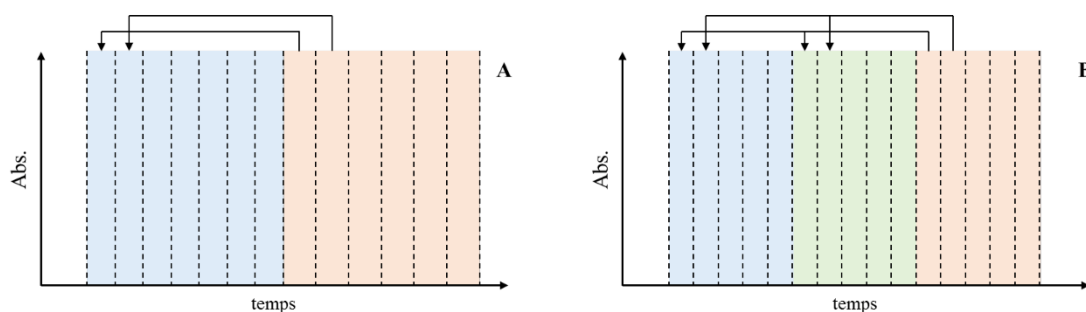


Figure 1.11 Représentation généralisée de différentes stratégies de combinaison de fractions proposées dans les études en protéomique ascendante. La fenêtre chromatographique est divisée en deux (A) ou en trois portions (B) selon le nombre de fractions à collecter. Les fractions de chaque zone sont ensuite mélangées ensemble en prévision de l'analyse par LC-MS/MS.

1.6 Identification de modifications impliquant la DOPA

Les protéines subissent plus de 200 modifications (réversibles et irréversibles) (D. J. Graves et al., 1994), complexifiant l'étude du protéome (Jensen, 2004; Khoury, Baliban, & Floudas, 2011), sans mentionner les modifications causées par les métabolites réactifs de certains xénobiotiques. Plusieurs éléments doivent être pris en compte dans l'étude des modifications covalentes de protéines.

D'abord, la complexité d'un échantillon et les nombreuses étapes de préparation, en plus de la possible suppression ionique causée par des espèces plus abondantes, sont des préoccupations majeures lors de l'identification de protéines modifiées (Liebler, 2002b). Ensuite, la chimie de ces composés influence de manière significative leur réaction avec les groupements nucléophiles de plusieurs macromolécules (Monks & Jones, 2002). Les développements en spectrométrie de masse depuis les dernières années ont contribué à l'avancement de la protéomique en matière de modifications covalentes (Liebler, 2002b). Néanmoins, Geib et al. ont souligné, entre autres, plusieurs obstacles en lien avec la présence d'isomères structurales, le manque d'ions diagnostiques ou bien certaines lacunes lors des mesures DDA (spectres MS/MS de faible qualité ou absent) minant une identification adéquate des sites de modifications (Geib, Thulasingham, Haeggström, & Sleno, 2020).

Parmi les intermédiaires réactifs les plus souvent rencontrés, les structures α,β -insaturées provenant de la peroxydation des lipides (Spickett, 2013) ainsi que les intermédiaires quinone, quinone imine et quinone méthilde font partie des composés connus pour leur importante réactivité avec certaines protéines (Golizeh, Leblanc, & Sleno, 2015; Monks & Jones, 2002; Ousji, Ohlund, & Sleno, 2020). Les quinones offrent un large éventail de métabolites réactifs et bénéficient d'une attention particulière en raison de leur importante présence dans divers produits de consommation (médicament, plastifiants ou agent de conservation). La Figure 1.12 ci-

dessous présente les structures de certains intermédiaires réactifs les plus souvent rencontrés.

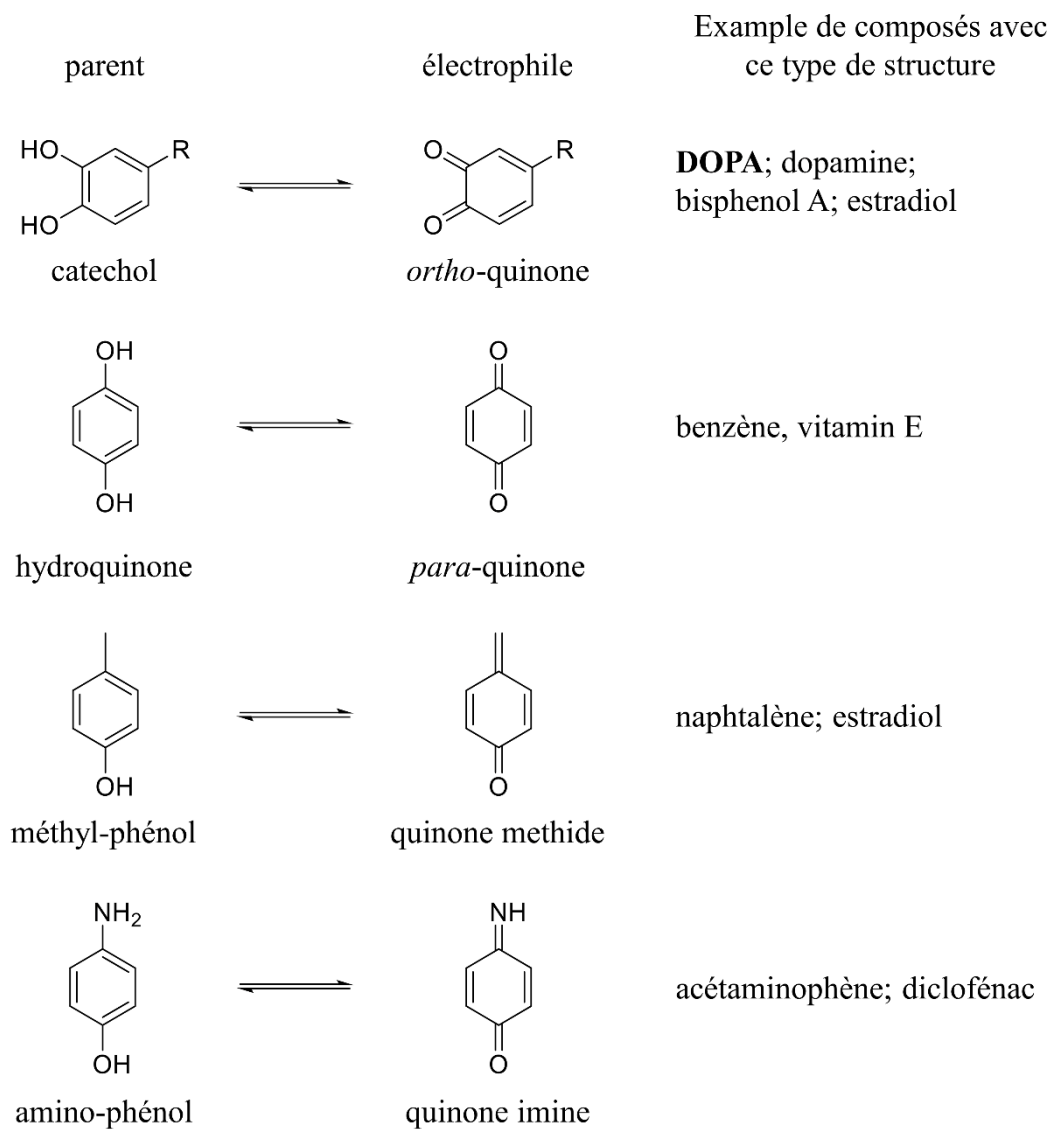
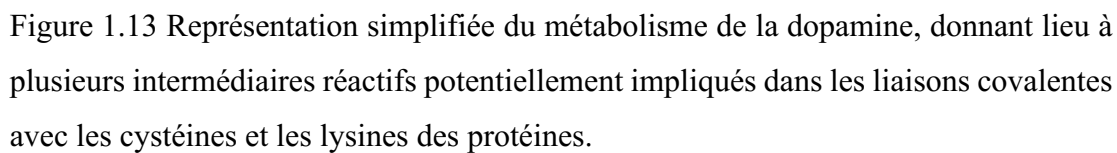


Figure 1.12 Structures de composés électrophiles principalement rencontrés comme xénobiotiques et composés endogènes. La DOPA est un exemple particulièrement important puisque cette structure peut être également considérée comme une modification post-traductionnelle.

Les quinones représentent une classe particulièrement intéressante de composés puisqu'elles peuvent à la fois induire un stress oxydatif et se lier de manière covalente aux groupements nucléophiles des protéines et acides nucléiques (Klopčič & Dolenc, 2019). Par exemple, le stress oxydatif causé par la présence de quinones entraîne la peroxydation des lipides qui, suivant une addition de Michael, est responsable de l'adduction de l'ADN (Rindgen, Seon Hwa Lee, Nakajima, & Blair, 2000) en plus de causer une cytotoxicité suite à la modification de certaines protéines (Galligan et al., 2014; Golizeh, Geib, & Sleno, 2016; Klopčič & Dolenc, 2019). Il est estimé que plus de 40% des composés de la famille des quinoïdes engendreront un intermédiaire réactif durant leur métabolisme (Figure 1.12) (Testa, Pedretti, & Vistoli, 2012). Sur les protéines, par exemple, les groupements thiol (cystéine), indole (histidine) et ϵ -amine (lysine) sont les principaux sites de modification. Néanmoins, la cytotoxicité de la modification dépendra principalement de la stabilité de l'adduit. L'acétaminophène est probablement le meilleur exemple d'un adduit stable causant des dégâts, parfois irréversibles au foie. Durant le métabolisme, l'acétaminophène est oxydé de manière à former une quinone imine réactive (NAPQI) et qui peut réagir avec les thiols des cystéines selon une addition de Michael. Bien que la toxicité de l'acétaminophène soit connue depuis plus de 50 ans environ, le faible pourcentage de la molécule parent convertie en NAPQI (de 5 à 10%) limite la proportion de protéines potentiellement modifiées par ce métabolite réactif. Par conséquent, le peu de protéines modifiées par rapport à la quantité non modifiée pose un défi considérable quant à leur identification (Liebler, 2002b).

Un autre quinoïde important, la DOPA, est un précurseur vital dans le métabolisme de la dopamine, dont les irrégularités dans les régions dopaminergiques du cerveau sont suspectées d'être en lien avec plusieurs maladies cognitives, telles que la schizophrénie (Grabnar, Vovk, Kores Plesnicar, & Boskovic, 2011), l'Alzheimer (Nam et al., 2018) et la maladie de Parkinson (Burbulla et al., 2017). Durant le



ci-dessous. Néanmoins, peu d'études ont adressé la question de la toxicité de ces structures. Par exemple, Meiser et al. (Meiser, Weindl, & Hiller, 2013), ont montré que la dopamine quinone pouvait réagir avec les thiols libres des protéines lorsque les niveaux de glutathion sont rapidement épuisés. Par conséquent, des protéines essentielles dans le maintien de l'équilibre rédox modifiées par la dopamine quinone sont responsables de la cytotoxicité observée (Giroto et al., 2012). Dans le même ordre d'idée, Jinsmaa et al. ont présenté la DOPAL comme étant un métabolite réactif impliqué dans la maladie de Parkinson. La présence de l'aldéhyde peut réagir avec les groupements amine des lysines et former un adduit stable sous la forme d'une base de Schiff. Parallèlement, le noyau catéchol peut aussi réagir avec les thiols libres des protéines. La possibilité de deux sites d'adduction pourrait permettre de potentielles réactions croisées au sein même de la protéine ou bien entre différentes protéines.

D'un autre côté, la DOPA peut également être considérée comme une modification post-traductionnelle importante, comme expliqué à la section 5.3.2. Il n'en demeure pas moins que le noyau catéchol est tout aussi sensible à l'auto-oxydation et peut engendrer des modifications covalentes impliquant d'autres résidus nucléophiles de la protéine. À titre d'exemple, les protéines de moules mfp-1, mfp-3 et mfp-5 sont reconnues pour leur importante concentration en DOPA. Ces protéines sont particulièrement importantes dans l'adhésion et la flexibilité de la fibre de byssus. McDowell et al. ont observé par RMN que des liaisons croisées impliquant la DOPA dans les fibres de byssus produites par *Mytilus edulis* étaient possibles suite à un stress ou durant la maturation de la fibre (McDowell, Burzio, Waite, & Schaefer, 1999). Quelques années plus tard, Zhao et Waite ont postulé que l'abondante présence de cystéines dans la plaque du byssus pouvait étalement donner lieu à des liaisons croisées entre la DOPA de la mfp-1 et les thiols de la mfp-6 (Zhao & Waite, 2005, 2006a). Ceci étant dit, l'impact négatif de la présence des quinones dans les fibres de byssus a été souligné par Yu et al. (Yu et al., 2013) et postulé comme étant responsable de la plasticisation de la fibre. Toutefois, cette avenue demeure encore à ce jour peu explorée.

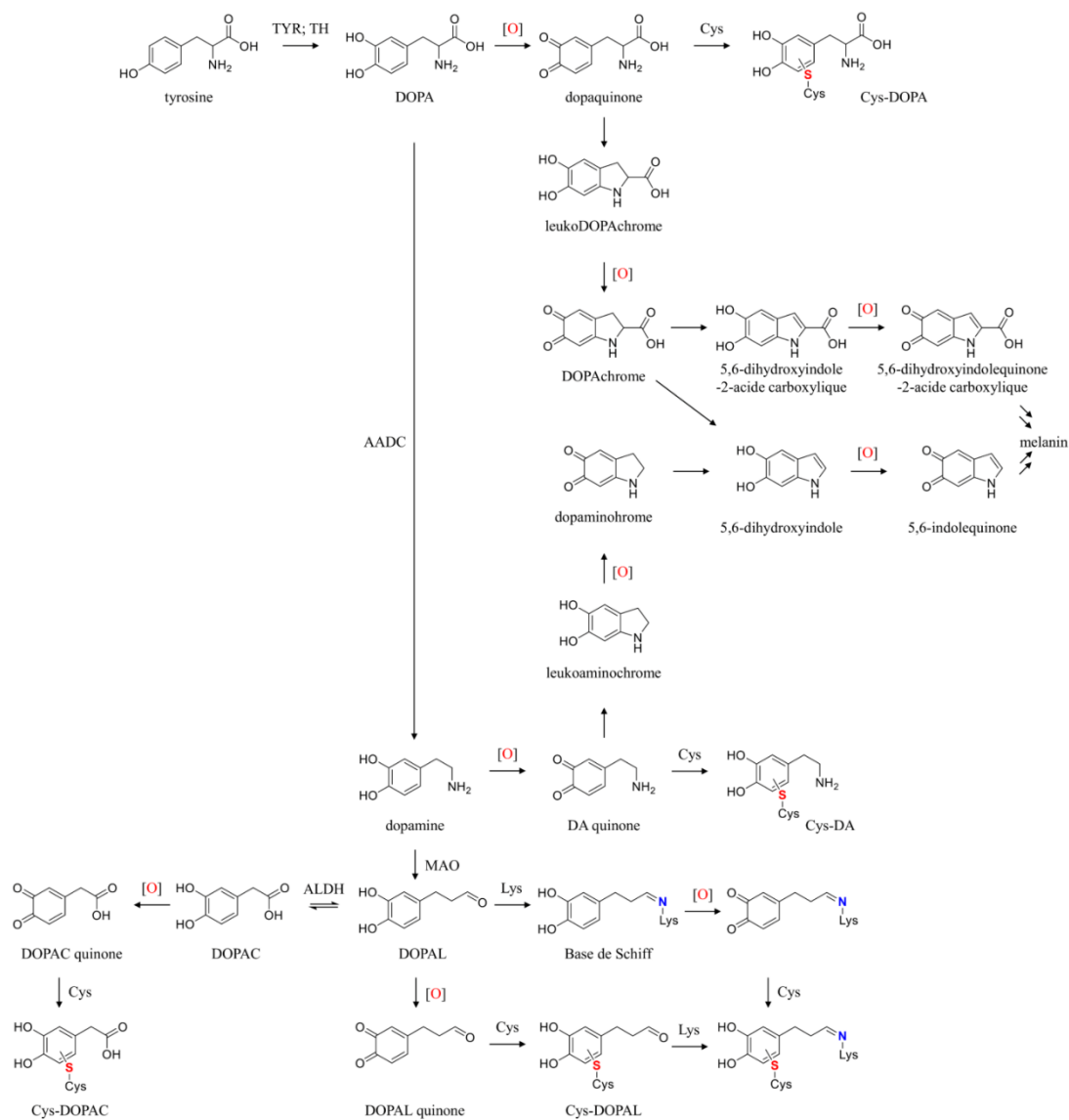


Figure 1.13 Représentation simplifiée du métabolisme de la dopamine, donnant lieu à plusieurs intermédiaires réactifs potentiellement impliqués dans les liaisons covalentes avec les cystéines et les lysines des protéines.

1.7 Défis et planification de la recherche

La solubilisation des protéines une fois extraites pose toutefois plusieurs difficultés analytiques. La présence de domaines trans-membranaires ainsi que les interactions protéines-protéines/nucléotides rendent cette étape particulièrement complexe (Bodzon-Kulakowska et al., 2007). L'utilisation d'agents chaotropes ou de détergents est une manière de dénaturer les protéines pour exposer les résidus chargés au solvant et ainsi favoriser les interactions hydrophiles. Dans le même ordre d'idée, l'ajout de sodium-dodécyl sulfate (SDS) augmente significativement la solubilité des protéines membranaires. Néanmoins, le SDS est reconnu pour interférer lors de la séparation sur gel ainsi que pour causer une importante suppression ionique durant les analyses MS. Des analogues zwitterioniques ou non chargés ont été testés pour l'identification de protéines (Golizeh & Sleno, 2013).

1.7.1 Minimiser les effets de matrice

La suppression ionique, telle qu'introduite par Buhrman et al. en 1996, est une forme d'effet de matrice qui diminue le signal d'un composé au moment de l'ionisation (Buhrman, Price, & Rudewicz, 1996). Par conséquent, la reproductibilité et la sensibilité d'une expérience en sont directement affectées (Volmer & Jessome, 2006). Bien que les mécanismes régissant la suppression ionique restent encore peu compris, malgré qu'ils soient connus depuis plus de 30 ans, Furey et al. ont regroupé les possibles mécanismes derrière la suppression ionique, incluant 1) la compétition entre deux analytes pour atteindre la surface de la gouttelette, 2) la disponibilité des charges selon le type de matrice, 3) la précipitation de l'analyte ou la liaison de ce dernier avec des constituants de la matrice (*e.g* avec des protéines), 4) la neutralisation de l'analyte en phase liquide ou gazeuse et finalement, 5) la présence d'additifs dans les phases mobiles et la technique d'ionisation (Furey, Moriarty, Bane, Kinsella, & Lehane, 2013).

Par ailleurs, la suppression ionique est aussi dépendante du composé analysé comme l'on observé Bonfiglio et al. en 1999. Dans leur étude, les auteurs ont remarqué que la suppression ionique était plus importante pour les composés polaires comme la caféine, que pour les échantillons enrichis de phéancétine, un composé moins polaire (Bonfiglio, King, Olah, & Merkle, 1999). Cette observation s'explique par le fait que les substances peu polaires migrent plus rapidement dans la gouttelette pour se retrouver rapidement à la surface de celle-ci, résultant en un meilleur signal (Konermann et al., 2013). Dans un autre ordre d'idée, l'affinité du proton joue aussi un rôle dans la suppression ionique. Par exemple, dans une étude menée par Ikonoraou et al. et confirmé plus de 30 ans plus tard dans des approches théoriques, les structures azotées ont une affinité plus importante pour le proton en phase gazeuse que les composés riches en oxygène (Chai, Weng, Shen, Sun, & Pan, 2015; Ikonoraou, Blades, & Kebarle, 1990).

Plusieurs options peuvent être considérées afin de 1) identifier la présence d'un effet de matrice et 2) minimiser son impact lors de l'analyse. Une approche simple pour vérifier la présence d'un effet de matrice a été proposé par Choi et al. en 1999, dans laquelle le signal mesuré d'un étalon interne par infusion directe était ensuite comparé à celui d'un échantillon afin d'évaluer l'impact de la suppression ionique (Choi, Gusev, & Hercules, 1999). Gosetti et al. ont aussi montré qu'une bonne manière d'identifier la présence et l'importance d'un effet de matrice était de superposer une courbe de calibration externe avec la courbe obtenue par ajouts dosés (Gosetti, Mazzucco, Gianotti, Polati, & Gennaro, 2007). Selon leur observation, les pentes des courbes d'étalonnage externe et par ajouts dosés devraient se superposer parfaitement en l'absence d'effet de matrice. Parallèlement, une approche plus technique pour diminuer les effets de matrices est de modifier la disposition du capillaire dans la source de l'appareil (Ghosh, Shinde, & Chakraborty, 2012). Initialement, le capillaire était directement orienté face à l'orifice du spectromètre de masse pour ainsi favoriser la transmission des ions. Pour minimiser la suppression ionique, la position du capillaire

a été changée (i.e à 90°) de façon à ce que le capillaire ne soit plus orienté directement face à l'orifice du spectromètre de masse. Par ce changement, moins d'ions entrent dans l'instrument, favorisant sa sensibilité, mais diminuant aussi l'entrée de potentiels interférents.

La préparation des échantillons biologiques nécessite souvent plusieurs étapes d'extraction, de nettoyage et d'enrichissement lors d'analyses de traces, par exemple. Dans certain cas, quelques classes de molécules endogènes sont directement liées à l'origine de l'effet de matrice. Par exemple, les phospholipides représentent une catégorie importante en raison de leur abondance significative (Jain et al., 2014; Jones, Borgmann, Wilkins, & O'Brien, 2006). En effet, les parois cellulaires sont essentiellement composées de phospholipides, qui une fois lysées, peuvent contribuer à la suppression ionique (Bylda, Thiele, Kobold, & Volmer, 2014; Srinivas, 2009). Il est cependant difficile d'éliminer complètement les interférences, ce pourquoi une préparation d'échantillons minutieuse doit être faite. Différentes techniques d'extraction en phase solide ont été testées (voir l'annexe C, section expérimentale) afin de minimiser l'impact de la présence des phospholipides lors de l'élution et l'ionisation des peptides.

1.7.2 Choix de la base de données de protéines

Pendant plus de 30 ans, la composition moléculaire du byssus produit par les moules a été étudié dans le but de comprendre et reproduire les propriétés adhésives et mécaniques (Arnold et al., 2013; DeMartini, Errico, Sjoestroem, Fenster, & Waite, 2017; Gantayet, Rees, & Sone, 2014; S. Li, Xia, Chen, Gao, & Zhan, 2018; C. Qin et al., 2016; X.-X. Qin & Waite, 1998; X. Qin & Waite, 1995b; Suhre, Gertz, Steegborn, & Scheibel, 2014). Malheureusement, l'identification de protéines byssales demeure limitée par l'annotation actuelle des banques de données de protéines. Le répertoire UniProt propose une base de données de moules comportant plus de 74 000 protéines associées à environ 71 espèces de moules, avec moins de 1% (671 protéines)

appartenant à l'espèce *Mytilus edulis*. De plus, la majorité de ces protéines n'a été identifiée qu'à partir d'informations provenant d'études transcriptomiques ou ayant une faible probabilité d'existence. Par conséquent, l'identification de protéines provenant de ces espèces pose un défi considérable.

D'abord, en examinant plus attentivement le statut des protéines contenues dans la base de données UniProt, un peu moins de la moitié des protéines (49%) a été identifiée par homologie avec d'autres espèces produisant elles aussi du byssus. Ceci témoigne du peu de changement observé dans la production des protéines retrouvées dans le byssus, quelle que soit l'espèce de moule. Ensuite, environ 34% de ces protéines sont considérées comme des protéines théoriques. Cette importante proportion peut poser un obstacle significatif lors de l'identification des protéines. En effet, la présence de protéines prédites (ou théoriques) peut être à l'origine de l'augmentation du taux de faux positifs lors de l'identification, comme l'ont suggéré Lubec et Afjehi-Sadat (Lubec & Afjehi-Sadat, 2007). Dans le cas précis des protéines prédites, la séquence primaire offre les bases en prévision de l'identification de la protéine, mais devrait idéalement être accompagnée d'une technique complémentaire afin de supporter les informations obtenues lors du séquençage (Lubec & Afjehi-Sadat, 2007). Les premiers travaux en protéomique employaient fréquemment la dégradation d'Edman dans le but de confirmer la structure primaire d'une protéine. Avec l'arrivée des techniques d'ionisations compatibles à l'analyse des macromolécules et avec les premières ébauches en bio-informatique, Johnson et Walsh ont su développer un algorithme permettant de combiner les informations obtenues par la dégradation d'Edman et par spectrométrie de masse afin de renforcer le séquençage de peptides provenant de mélanges considérés complexes à ce moment (R. S. Johnson & Walsh, 1992).

Ensuite, en 1994, Mann et Wilm ont été les premiers à introduire (Mann & Wilm, 1994) l'utilisation d'une petite séquence de deux ou trois acides aminés pour l'identification de protéines. Cette approche a rapidement gagné en popularité puisqu'il

était désormais possible d'identifier une protéine à partir d'une courte séquence d'acides aminée, même si plusieurs séquences étaient possibles (Mann & Wilm, 1994). En se basant sur le principe proposé par Mann et Wilm, Shilov et al. ont introduit en 2007 un nouvel algorithme de recherche employant de courtes séquences d'acides aminés pour l'identification de protéines à partir des spectres MS/MS de peptides (Shilov et al., 2007). L'algorithme Paragon adapte ses efforts de séquençage en fonction du succès de correspondance de la séquence avec la base de données. Par conséquent, l'algorithme offre un contrôle plus restrictif des faux positifs et permet l'identification de peptides modifiés ou de peptides spécifiques à une catégorie d'enzyme (*e.g.* les peptides tryptiques) (Shilov et al., 2007). Néanmoins, comme l'ont soulevé Wang et Wilson, ce type d'identification par seulement trois acides aminés peut tout de même engendrer des faux positifs et exclure certains peptides importants lors de l'identification de protéines (P. Wang & Wilson, 2013).

Finalement, la flexibilité de certains paramètres de recherche joue un rôle primordial lors de l'identification. Par exemple, dans le cas de l'analyse des protéines provenant des moules, limiter la base de données aux protéines n'appartenant qu'au genre *Mytilus* peut jouer à notre désavantage. Présentement, seulement 671 protéines sont connues comme étant associées à la moule *Mytilus edulis* et 355 pour la moule *Mytilus californianus*. Par conséquent, plusieurs protéines risquent d'être exclues. Il est donc préférable d'inclure plusieurs espèces homologues afin de maximiser l'identification des protéines provenant d'espèces peu étudiées (Cottrell, 2011). Dans le même ordre d'idée, une identification réalisée avec une exactitude trop restreinte sur les ions précurseurs et avec une base de données contenant peu de protéines limitera le nombre de protéines candidates. Par conséquent, comme l'a mentionné Cottrell, la fiabilité des résultats obtenus dans des conditions trop restreintes est aisément discutable (Cottrell, 2011).

CHAPITRE II

COMPARAISON D'APPROCHES CHROMATOGRAPHIQUES POUR L'ANALYSE DE PEPTIDES PAR LC-MS/MS

Pour parvenir à une détection adéquate et reproductible, les approches par chromatographie liquide multi-dimensionnelles sont souvent les approches préconisées pour maximiser la détection des peptides. Cependant, ce type de méthodologies entraînent souvent la perte d'analytes en raison des nombreuses manipulations. Dans ce chapitre, la pertinence de ce genre de préparation sera abordée et critiquée selon le point de vue du nombre de peptides détectés et de protéines identifiées.

Ces expériences ont été réalisées à partir de tissus provenant de moules. Les protéines ont été digérées selon un protocole de digestion enzymatique parallèle, puis analysées selon trois avenues, l'une sans pré-fractionnement, la seconde utilisant un pré-fractionnement par échange de cations forts et la dernière par chromatographie en phase inverse à pH basique. Maxime Sansoucy a procédé à la préparation des échantillons, la rédaction des protocoles analytiques, la réalisation des expériences puis aux traitements des données. La professeure Lekha Sleno a fourni son support et a supervisé les expériences en laboratoire et vérifié l'interprétation des résultats.

2.1 Résumé

La chromatographie multi-dimensionnelle est souvent introduite comme une excellente approche pour l'augmentation de la couverture du protéome d'un organisme. Traditionnellement, les échantillons de protéines sont d'abord digérées par une enzyme relativement sélective avant d'être séparés. Le mélange de peptides est ensuite séparé sur une colonne échangeuse d'ions forts, dans la première dimension, suivit d'une seconde séparation sur une colonne C18 dans la deuxième dimension. De nos jours, l'utilisation de deux mécanismes complémentaires lors de la chromatographie des peptides est considérée comme l'approche classique à adopter pour les études détaillées en protéomique ascendante. Cependant, l'utilisation de phases mobiles difficilement compatibles avec le spectromètre de masse est souvent montrée comme une limitation majeure. Un autre désavantage de ce type de pré-fractionnement est que les fractions collectées doivent être dessalées, ce qui peut engendrer des pertes lors de cette étape supplémentaire. Par conséquent, la détection de peptides peu abondant est souvent difficile dans ce type d'approche analytique.

Les techniques de pré-fractionnement par échange de cations forts et par chromatographie inverse à pH élevé ont été comparées avec une approche directe (sans pré-fractionnement), dans le but d'évaluer la contribution réelle des approches 2D-LC sur la couverture globale du protéome.

2.2 Introduction

Les analyses employant la spectrométrie de masse apportent une aide unique quant à l'identification de protéines issues d'échantillons biologiques. Tel qu'expliqué à la section 1.3, l'approche de type *bottom-up* est la plus souvent choisie pour offrir un meilleur taux de couverture du protéome. Néanmoins, la digestion enzymatique augmente la complexité du mélange étudié. Par conséquent, l'un des principaux enjeux

touche surtout la détection en peu de temps (de l'ordre de quelques secondes) de plus de 100 000 peptides, dans une fenêtre de masse relativement restreinte (typiquement de 1 à 2 Da) (Bonneil, Pfammatter, & Thibault, 2015). De plus, la faible résolution lors de la sélection de l'ion précurseur (typiquement de 1 à 2 Da) nuit considérablement à l'obtention de spectres de masse de bonne qualité (Bonneil et al., 2015).

À cet égard, le pré-fractionnement des échantillons est généralement une approche efficace pour minimiser la densité de certaines régions du spectre de masse. En 2001, Washburn et al. ont démontré l'efficacité d'une séparation par SCX-RP-LC-MS sur l'identification des protéines issues de la bactérie *E. coli* (Washburn et al., 2001). Depuis, plusieurs applications ont été développées ou perfectionnées dans le but précis d'améliorer l'identification des protéines (Law et al., 2015; Manadas, Mendes, English, & Dunn, 2010). Traditionnellement, l'échange de cations fort est une approche courante dans les méthodologies multi-dimensionnelles qui, cependant, comporte son lot de limitations. La présence de sels non-volatils est souvent montrée comme principal point négatif, puisque ceux-ci doivent être préalablement éliminés en prévision de l'analyse MS (Di Palma et al., 2012; Xie, Smith, & Shen, 2012). D'autre part, l'automatisation rend souvent les applications high-throughput compliquée en raison du nombre d'échantillons limité en plus des analyses s'étalonnant sur plusieurs heures.

Parallèlement, au début des années 2000, Gilar et al. ont montré des résultats particulièrement prometteurs pour des méthodes multi-dimensionnelles utilisant un pré-fractionnement en phase inverse (Gilar, Olivova, Daly, & Gebler, 2005b). De plus, le pH joue un rôle considérable dans l'orthogonalité des méthodes RP-RP ainsi que sur la couverture globale du protéome (Di Palma et al., 2012; Gilar et al., 2005a). Dans le même ordre d'idée, la complémentarité de la RP-RP a été comparée à des méthodologies connues pour leur très grande orthogonalité. Le nombre de protéines identifiées par une séparation en phase inverse a été montré supérieure ou comparable

aux approches traditionnelles par SCX et par séparation sur gel (Dowell, Frost, Zhang, & Li, 2008; Y. Wang et al., 2011). Également, le faible taux de ré-échantillonnage (i.e. l'activation du même peptide à plusieurs reprises) et l'importante résolution, mentionné par Dowell, ont grandement contribué à l'augmentation du nombre de peptides identifiés par Hp-RP (Dowell et al., 2008).

Tel que mentionné au chapitre précédent, la faible complémentarité de la RP-RP a été mentionnée à plusieurs occasions comme une limitation majeure. La combinaison de certaines fractions (voir Figure 1.11) a été introduite comme une stratégie judicieuse pour une analyse plus complète du protéome (Song et al., 2010; Yeung et al., 2020). Cette action permet une meilleure utilisation du cycle opératoire de l'instrument et entraîne une augmentation significative des spectres de masse mesurés (Law et al., 2015). Concrètement, la combinaison de fractions se traduit par une plus grande proportion de peptides séquencés, ce qui amène ultimement à un plus grand nombre de protéines identifiées. Toutefois, le nombre de fractions à collecter et la manière dont celles-ci sont combinées demeure encore peu exploré à ce moment. Il est néanmoins accepté que la combinaison de fractions non adjacentes est l'approche à préconiser pour améliorer l'orthogonalité.

Ce chapitre adresse principalement la question de la pertinence d'un pré-fractionnement sur le rendement des protéines identifiées. Certes, les approches multi-dimensionnelles offrent une meilleur reproductibilité, sans mentionner l'amélioration de la qualité des spectres MS¹. Toutefois, peu de preuves expérimentales sont apportées afin de soutenir le choix d'une approche multi-dimensionnelle. Avec une séparation adéquate, il est possible d'équivaloir les approches multi-dimensionnelles quant au nombre de protéines identifiées (Thakur et al., 2011). La majorité des travaux disponibles se basent sur la théorie de la chromatographie afin de justifier l'utilisation de méthodes multi-dimensionnelles. Plus précisément, il sera question ici de fournir

des éléments justificatifs se basant sur la spectrométrie de masse afin de comprendre le réel impact sur le rendement des protéines identifiées.

2.3 Expérimentale

2.3.1 Matériel

Trypsine (bovine), les solvants et autres produits chimiques ont été achetés chez Sigma-Aldrich (St. Louis, MO). L'eau ultrapure a été obtenue à partir d'un système Millipore Synergy UV (Billerica, MA).

2.3.2 Fractionnement par chromatographie d'exclusion stérique

Les échantillons de pieds de moules ont été homogénéisés dans une solution tampon ABC (100 mM, pH 8.5), puis filtrés sur une membrane à poids moléculaire limité à 30 kDa. Les échantillons ont ensuite été dilués avec la phase mobile en préparation pour le fractionnement. Puis, les protéines ont été injectées sur une colonne Agilent Bio-SEC 3000A 300 x 7.8 mm (Agilent Technologies, Palo Alto, CA) en mode d'élution isocratique avec une solution tampon de phosphate de sodim (150 mM, pH). L'absorbance a été mesurée à 220 et 280 nm. 10 fractions éluant entre 1 et 21 min à un débit de 500 µl/min ont été récupérées pour ensuite être séchées en prévision de l'analyse LC-MS/MS.

2.3.3 Digestion des protéines

L'extraction et la digestion de protéines provenant de *Mytilus edulis* et *Mytilus californianus* se sont déroulées selon le protocole détaillé au chapitre 3. Brièvement, les extraits de protéines ont été préparés en prévision d'une alkylation réductrice avec du dithiothréitol et l'iodocaétamine respectivement. Les protéines dénaturées ont ensuite été diluées avec 250 µl du tampon ABC (100 mM, pH 8.5) pour les échantillons digérés par la trypsine ou dilués avec 250 µl d'une solution à 2% TFA dans 20%

méthanol pour les échantillons digérés par la pepsine. Le choix de ces enzymes repose essentiellement sur la complémentarité de leur site de clivage. En effet, la trypsine clive le squelette peptidique généralement du côté C-terminal d'un résidu arginine ou lysine (sauf s'il est suivi d'une proline). La pepsine, quant à elle, est moins sélective que la trypsine et préfère les acides aminés hydrophobiques, comme la phénylalanine ou la leucine par exemple. Une fois la digestion complétée (18h à 37°C), les digestats ont été proprement neutralisés (dans le cas de la pepsine) et dessalés suivant une extraction sur phase solide de type OASIS® HLB SPE (Waters, Milford, MA, 30 mg/1 cc), puis séchés afin d'être resuspendus dans une solution en prévision de la collecte de fractions par Hp-RP ou par SCX.

2.3.4 Fractionnement en phase inverse à pH élevé

Une fois digérés, les échantillons de pied et de manteau de la moule ont été séparés sur un système Agilent 1200 HPLC équipé d'un auto-échantillonneur réfrigéré, d'un dégaseur intégré, d'une pompe binaire, d'un détecteur à barrette de diode ainsi que d'un collecteur de fractions à température contrôlée. Les peptides ont été solubilisés dans la phase mobile A (voir plus bas) et séparés sur une colonne Agilent Pursuit XRs 5µ C18 250 x 4.6 mm (Agilent Technologies, Palo Alto, CA). La collecte des fractions a été opérée à un débit de 600 µl/min avec un gradient d'élution débutant à 5% de la phase mobile B pour les deux premières minutes, suivit d'une croissance linéaire de 5 à 50% B en 12 minutes, pour ensuite atteindre 70% B à 14.5 min et maintenu pendant 6.5 min. Les phases mobiles A et B sont respectivement une solution de 10 mM en acétate d'ammonium (pH 10.0, ajusté avec du NH₄OH) et 90% ACN mélangé avec 10% de la phase mobile A. L'absorbance a été mesurée à 220 et 280 nm. 16 fractions éluant entre 3 et 19 min ont été collectées pour ensuite être combinées (fraction 1+9, 2+10, etc.) en huit fractions finales et séchées en prévision de l'analyse LC-MS/MS.

2.3.5 Fractionnement par échange de cations forts

Le fractionnement a été réalisé de manière similaire à la procédure mentionnée à la section précédente. Les échantillons digérés de pieds et de manteau de moules ont été solubilisés dans la phase mobile A (voir ci-bas) et séparés sur une colonne Agilent Zorbax 300-SCX 150 x 2.1 mm avec une taille de particules de 5 μm (Agilent Technologies, Palo Alto, CA). La collection des fractions a été opérée sur le même système décrit plus haut, cette fois à un débit de 250 $\mu\text{l/min}$. La proportion de phase mobile B a été augmentée de 0-8% durant les six premières minutes, puis de 8 à 30% B entre 6 et 11 minutes, pour ensuite atteindre 65% B à 20 min, puis 95% B à 25 minutes et maintenue jusqu'à 34 minutes. Les peptides éluant de 1-24 min, pour un total de huit fractions, ont été collectés et sujets à une étape d'extraction en phase solide supplémentaire afin d'en retirer les sels non-volatiles. Une fois dessalées, les fractions ont été séchées en prévision de l'analyse LC-MS/MS.

2.3.6 Analyse RP-LC-MS/MS

Une fois secs, les échantillons ont été solubilisés dans une solution de 10% ACN (100 μl) et 30 μl a été injecté sur une colonne Aeris PEPTIDE XB-C18 100 x 20 mm avec des particules de 1.7 μm , maintenu à 40°C (Phenomenex, Torrance, CA). Un système UHPLC, modèle Nexera (Shimadzu, Columbia, MD) a été opéré à un débit de 300 $\mu\text{l/min}$ avec une proportion de 5% de la phase mobile B maintenu pendant les 2.5 premières minutes. La composition de la phase mobile B a été augmentée de façon linéaire à 30% à 24 minutes, puis à 50% B à 26 minutes et finalement, à 85% B à 26.5 min et maintenu pendant deux minutes.

Les spectres MS et MS/MS ont été obtenus à partir d'un spectromètre de masse hybride quadripôle temps-de-vol TripleTOF™ 5600 (AB Sciex, Concord, ON, Canada). La source d'électro-nébulisation a été opérée en mode positif et maintenue à une température de 500°C, avec un voltage appliqué sur le capillaire de 5 kV et un

potentiel de déclusterisation (DP) de 60 V. La pression des gaz de nébulisation et de séchage a été maintenue à 50 psi. La sonde d'électro-nébulisation a été utilisée pour assimiler le flux provenant du LC alors que la sonde APCI a été employée pour la calibration après quatre injections (en mode MS et MS/MS), à l'aide d'une solution de calibrants préparée dans nos laboratoires. Les mesures TOF-MS ont été réalisées entre m/z 80-1500, avec une soustraction dynamique du bruit de fond (*Dynamic Background Subtraction*, DBS) et des critères de sélection d'information dépendante (*Information-dependent acquisition*, IDA) sur les 15 ions les plus sensibles au-delà de m/z 250. Les spectres MS/MS ont ensuite été enregistrés suivant l'accumulation (50 ms) et la collision ($CE = 30 \pm 10$ V) des ions précurseurs. Le temps de cycle total a été calculé à 1.05 s.

2.3.7 Traitement des données

Les données brutes LC-MS/MS provenant de chaque méthode (sans fractionnement, SCX et Hp-RP) ont été analysées contre la base de données de protéines UniProtKB/Swiss-Prot (téléchargée en décembre 2018) par le logiciel ProteinPilot (version 5.0.2). À partir de la base de donnée Bivalvia (63 461 protéines, 210 espèces), les protéines associées uniquement aux espèces de moules ont été conservées dans la banque de protéines pour l'analyse des données LC-MS/MS. Également, certaines protéines appartenant aux bactéries ainsi qu'aux virus ont été retirées. La base de données finale comportait 7 649 protéines provenant de 81 espèces. Plus de 98% des protéines contenues dans la base de données ne sont pas révisées et la plupart n'a été observée qu'au niveau du transcriptome. Afin d'augmenter le taux de couverture de protéome, les données obtenues pour les digestions par la trypsine et la pepsine ont été combinées lors des recherches. Une fois terminé, les résultats ont été exportés (.mgf, .mzIdentML et le PeptideSummary) puis téléchargés dans le logiciel Scaffold-Q+ (version 4.9.0, Proteome Software). Uniquement les protéines avec un minimum de 2 peptides à 1% FDR ont été considérées.

2.4 Résultats et discussion

2.4.1 Comparaison entre les approches 1D et 2D

Afin de comparer l'impact d'une chromatographie multi-dimensionnelle, des extraits de protéines de moules ont été utilisés. Les protéines provenant du pied et du manteau de la moule ont été extraites puis digérées avec la trypsine et la pepsine pour être finalement analysées par spectrométrie de masse à haute résolution. La Figure 2.1 résume les étapes de la préparation des échantillons pour cette comparaison.

D'abord, les chromatogrammes obtenus lors d'une séparation par échange ionique et par phase inverse sont présentés pour les échantillons de pied de moule (voir

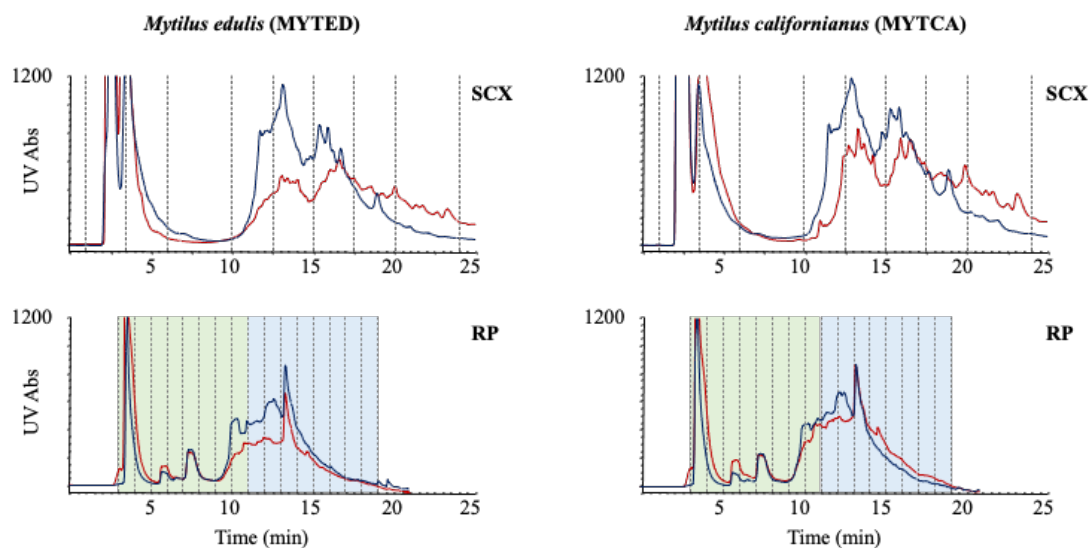


Figure 2.2). Les profils d'élution montrent une importante intensité au niveau du front de solvant, principalement causée par de petits composés polaires et non retenus. La majorité des peptides a été éluée dans une fenêtre de rétention allant de 10 à 20 minutes. Puisque le mécanisme de séparation entre la phase inverse et l'échange ionique est différent, un profil d'élution légèrement plus large a été observé pour les

peptides fractionnés par SCX. Cet élargissement peut s'expliquer par la différence entre les interactions de la phase stationnaire et les peptides lors de l'élution.

L'utilisation d'une phase inverse C18 procure généralement une meilleure résolution en comparaison avec une résine ionique, puisque les peptides sont essentiellement séparés par hydrophobicité (F. Yang, Shen, Camp, & Smith, 2012). L'utilisation d'un pH basique lors du fractionnement est justifiée par sa complémentarité avec la phase mobile acide du LC-MS/MS. En ce sens, la plupart des peptides porteront une charge globale différente sous un pH acide et basique, ce qui aura pour effet de modifier leur hydrophobicité relative.

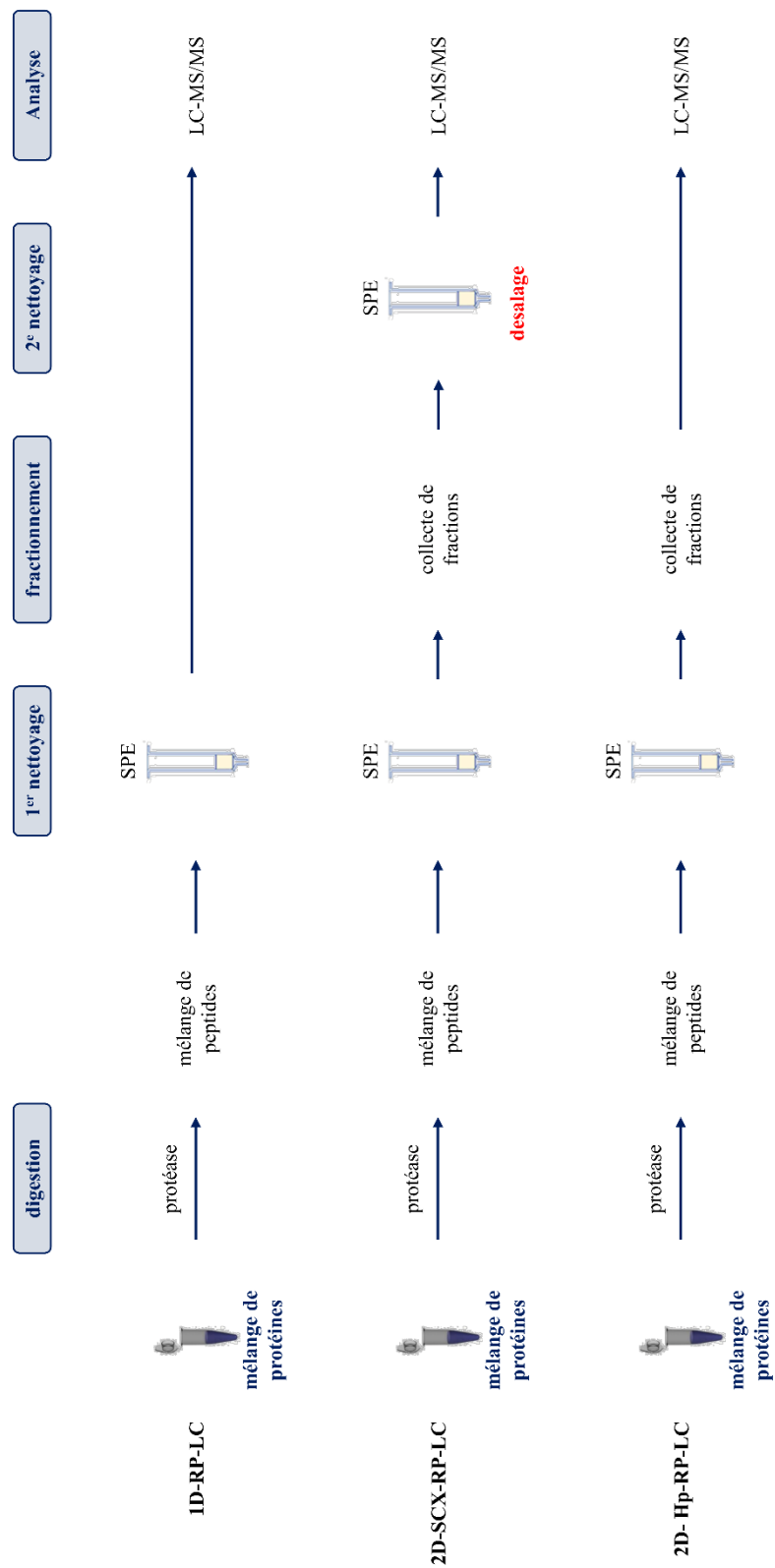


Figure 2.1 Représentation des méthodologies 1D- et 2D-LC couramment employées dans les analyses en protéomique ascendante.

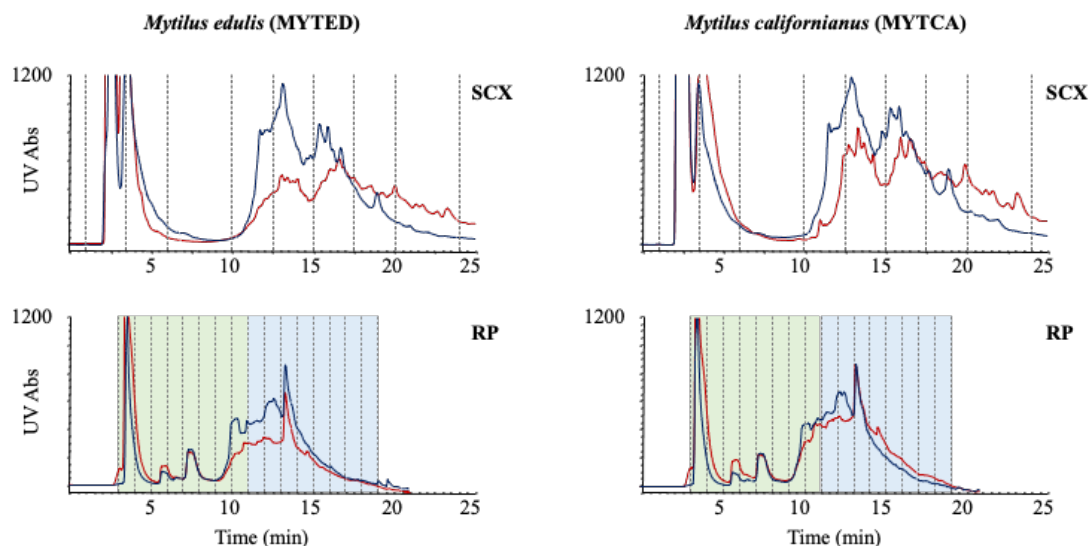


Figure 2.2 Chromatogrammes UV pour le fractionnement des peptides par échange de cations forts (haut) et par chromatographie en phase inverse à pH élevé (bas) pour les peptides obtenus par digestion tryptique (signal en bleu) et pepsique (signal en rouge) pour les échantillons de pied de moules.

Cependant, ce ne sont pas toutes les formes de collecte de fractions qui sont efficaces, même si leur mécanisme de séparation est *a priori* orthogonale. Dans une approche préliminaire, un pré-fractionnement au niveau des protéines par chromatographie d'exclusion stérique a été envisagé. Plusieurs problématiques ont été rencontrées. D'abord, la solubilisation des protéines a été une étape compliquée en raison du choix restreint de solutions pour la solubilisation. En effet, les solutions généralement utilisées pour ce genre de procédure (par exemple le SDS) causent souvent une suppression ionique. Ensuite, les échantillons n'étaient jamais totalement solubilisés, même après plusieurs cycles bain ultrason/vortexe. Donc, ceux-ci ont dû être filtrés avant le pré-fractionnement afin d'éviter de bloquer l'auto-échantillonneur. Par conséquent, une portion significative de l'échantillon est demeurée sur le filtre, ce qui a causé d'importantes pertes. Finalement, la chromatographie par exclusion

stérique est hautement dépendante de la phase mobile, en accord avec les résultats présentés ci-bas (Figure 2.3). Par exemple, les échantillons fractionnés à pH acide semblaient être dénaturés et n'ont donc pas été correctement séparés par la colonne.

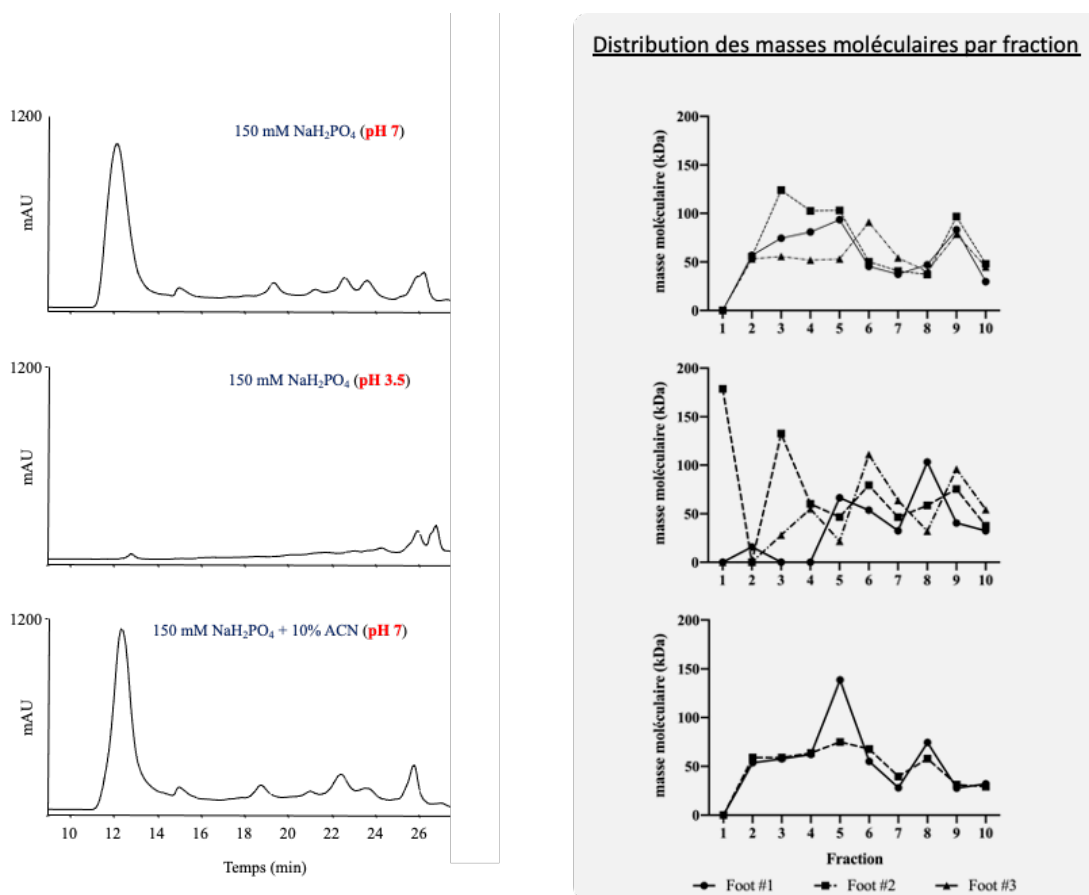


Figure 2.3 Séparation des protéines provenant du pied de la moule par chromatographie d'exclusion stérique. Le chromatogramme du haut montre la fenêtre de collection des fractions (de 9 à 29 min) alors que les chromatogrammes ci-dessous montrent le comportement de la séparation lorsque la phase mobile (150 mM NaH_2PO_4) est ajustée à pH 7, pH 3.5 ou contenant 10% d'ACN. Les graphiques dans l'encadré gris illustrent la distribution des masses moléculaires par fraction.

Lorsque le nombre de protéines par fraction a été examiné, la masse moyenne des protéines collectées pour une fraction donnée variait considérablement. Cette distribution ératique semblait être la conséquence d'interactions de type protéines-protéines, biaisant ainsi la collecte des protéines. Malgré l'ajout d'un agent de réduction dans la solution tampon, peu d'amélioration a été notée et par conséquent, le pré-fractionnement au niveau des protéines n'a pas été considéré pour des analyses subséquentes.

Une méthode LC-MS/MS relativement courte, soit d'une durée de 28 minutes, a été choisie pour cette étude. Parallèlement, différents groupes se sont tournés vers des gradients d'élution nettement plus longs afin d'augmenter la couverture du protéome (Hinzke, Kouris, Hughes, Strous, & Kleiner, 2019; Tsai et al., 2015; Yin et al., 2014). Suivant cette tendance, un gradient d'une heure (58 minutes) a été testé pour améliorer le nombre de peptides séquencés avec assurance par MS/MS. Néanmoins, peu d'amélioration a été notée entre les deux conditions (résultats non présentés). Cette observation s'explique par l'élargissement des pics lorsqu'un gradient plus long est employé et correspond avec la diminution de la sensibilité lors des mesures par spectrométrie de masse. Il est toutefois important d'estimer et d'ajuster la durée du gradient en fonction de la complexité de l'échantillon à analyser.

En divisant les échantillons de protéines digérées, chaque fraction contient un sous-ensemble de peptides qui, une fois élués de la colonne, ont une plus grande chance de produire un spectre MS/MS ayant un meilleur ratio signal-sur-bruit (S/N). La Figure 2.4 montre une comparaison entre les spectres de masse (TOF-MS) obtenus pour un peptide tryptique (QTVAVGVK, m/z 457.78) de la protéine du facteur d'élongation 1- α . En premier lieu, ce peptide a été identifié par les trois approches avec une excellente précision sur la masse (moins de 5 ppm). Bien qu'il se distingue par un ratio S/N relativement important pour l'analyse directe, l'étape de pré-fractionnement a grandement amélioré la qualité du spectre MS et pour ainsi dire, la reproductibilité

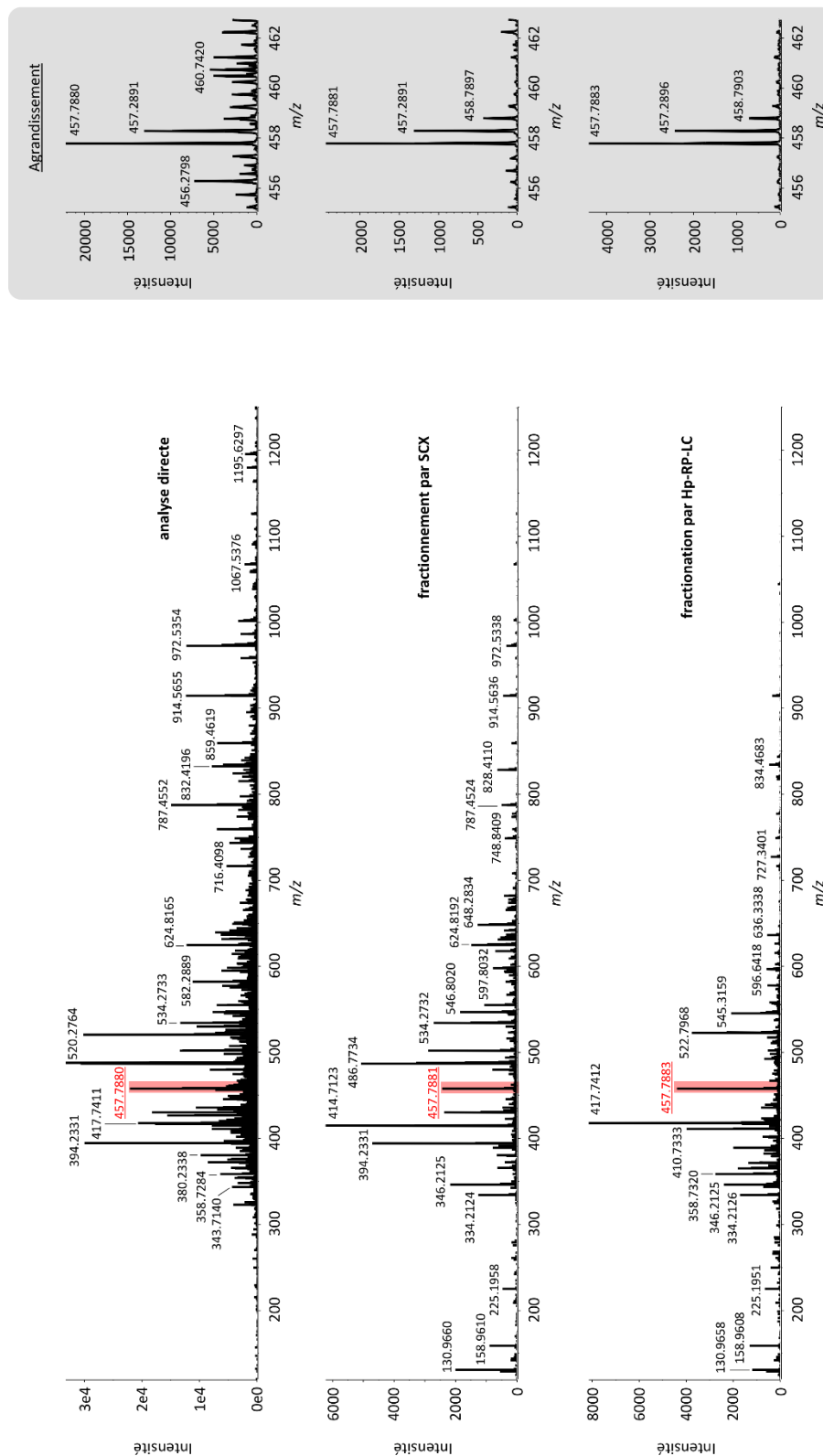


Figure 2.4 Comparaison entre les spectres TOF-MS obtenus pour l'analyse du peptide QTVAVGIK (m/z 457.78) de la protéine du facteur d'élongation 1-alpha, analysé sans pré-fractionnement, par SCX et par Hp-RP. L'ion précurseur est montré en rouge et l'encadré gris montre un agrandissement de cette zone.

lors de l'expérience DDA. Les instruments à haute résolution ont la capacité de montrer la distribution isotopique des peptides et des protéines. Dans le cas des approches SCX et Hp-RP, la signature isotopique du peptide est claire, alors que pour l'analyse directe, le pic associé au carbone 13 se confond dans le bruit de fond de l'appareil. Il s'agit ici d'une importante limitation, puisque l'incapacité à distinguer l'enveloppe isotopique correctement peut entraîner l'exclusion du peptide en question. Mujezinovic et al. ainsi que Slawski et al. (Mujezinovic et al., 2006; Slawski et al., 2012) ont montré qu'avec l'aide d'approximation statistique et de regression linéaire, il était possible d'améliorer l'interprétation des signatures isotopiques dans le cadre de bioanalyses.

L'utilisation d'une étape de pré-fractionnement, particulièrement lorsqu'opérée en mode *off-line*, engendre généralement des pertes d'échantillons. Pour mesurer qualitativement l'impact de la collecte des peptides avant de les analyser par LC-MS, les trois approches ont été comparées entre elles selon le nombre de peptides détectés, le nombre de protéines identifiées ainsi que la couverture moyenne des protéines. Les résultats sont présentés dans le tableau suivant.

Tableau 2.1 Comparaison de chaque approche pour les peptides et les protéines identifiés dans le tissu du pied de la moule *Mytilus edulis*.

	1D	2D (SCX)	2D (RP)
# protéines (2% FDR)	98	114	137
# peptides (1% FDR)	1201	1342	1593
peptide unique/protéine	12.2	11.7	11.6
% couverture de séquence	26.1%	24.7%	24.8%

Le tableau précédant montre bien à quel point les techniques de pré-fractionnement contribuent à l'identification d'un plus grand nombre de protéines. Étonnamment, l'approche par SCX présente des résultats comparables avec les analyses 1D-LC-MS/MS. L'approche la plus performante a été par pré-fractionnement à pH basique, avec une augmentation d'environ 25% de protéines identifiées. Lorsque les trois approches ont été comparées en termes de peptides uniques et de couverture de séquence, il semble juste de mentionner que chacune des approches se sont montrées équivalentes. Il est toutefois pertinent de souligner que les résultats présentés au tableau 2.1 font référence à des expériences à $n = 1$. Le choix de ne pas inclure de réplicats reposait principalement sur 1) le nombre total d'échantillons, 2) le temps alloué pour chaque analyse LC-MS/MS et 3) le temps nécessaire à l'identification des protéines et au traitement des données. Puisque chaque tissu de pied et le manteau devaient être digérés par la trypsine et la pepsine séparément, le nombre d'échantillons à analyser se voyait être doublé. De plus, pour le pré-fractionnement par SCX et Hp-RP, un échantillon était divisé en huit fractions qui allaient ensuite être analysées par LC-MS/MS. Ensuite, l'étape de pré-fractionnement nécessitait plusieurs heures durant lesquelles les échantillons devaient être fractionnés en prévision des analyses par spectrométrie de masse, sans compter la méthode LC-MS/MS qui prenait environ une heure par échantillon. En évaluant chaque aspect technique de notre approche, il a été jugé plus favorable de n'utiliser qu'un seul échantillon par tissu/enzyme dans le but de montrer que les approches multi-dimensionnelles étaient préférables pour l'augmentation de la couverture du protéome, mais que pour des raisons pratiques, une approche par 1D-LC-MS/MS offrirait des résultats comparables. D'un autre côté, les enzymes utilisés pour la digestion des protéines demeurent relativement reproductibles d'une expérience à l'autre. La variabilité entre chaque expérience a de faible chance d'influencer les résultats présentés au tableau 2.1. La Figure 2.5 ci-dessous montre que les approches 2D-LC ont contribué à l'augmentation de la couverture du protéome. Environ de 6 à 18% des protéines ont été seulement identifiées par les approches 2D-LC, avec le fractionnement à pH élevé étant le plus efficace. Dans le même ordre d'idée,

les méthodologies SCX et Hp-RP partagent une sélectivité relativement comparable, puisqu'environ 41% des protéines ont été identifiées à la fois par SCX et Hp-RP.

Les chaînes latérales non polaires de certains peptides octroient une meilleure efficacité lors de l'ionisation par ESI (Cech, Krone, & Enke, 2001). Par conséquent, la contribution des propriétés physico-chimiques détient un rôle majeur dans le succès des approches par fractionnement en phase inverse. De plus, la combinaison de fractions non-adjacentes permet d'obtenir une répartition relativement plus homogène entre les peptides hydrophiles et lipophiles. Par conséquent, le processus d'ionisation est peu ou pas biaisé par la présence de peptides lipophiles et offre ainsi la chance aux peptides à caractère hydrophile la possibilité d'être ionisés. Bien entendu, leur fragmentation dépend aussi du respect des critères de sélection lors de l'analyse DDA.

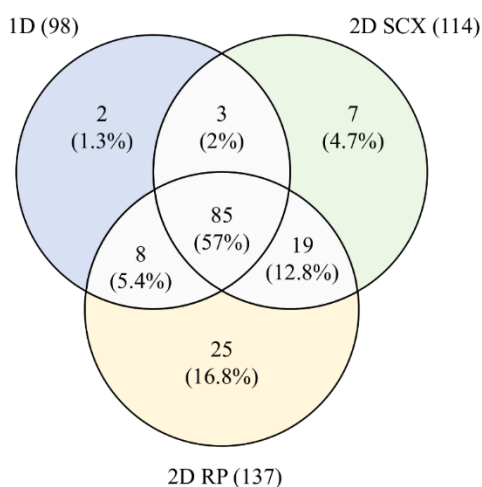


Figure 2.5 Répartition des protéines identifiées pour les échantillons de pied de *Mytilus edulis* entre les trois approches.

Le choix de combiner directement deux fractions non-adjacentes a, en partie, été motivé par le souci de comparer un nombre identique de fractions par rapport à la SCX (huit fractions), mais aussi dans l'objectif d'améliorer immédiatement

l'orthogonalité de cette approche. Dans le but de démontrer que l'utilisation d'une méthodologie par Hp-RP, avec la combinaison de deux fractions non-adjacentes, est réellement la bonne approche, une étude comparative entre la SCX et la RP (à pH basique et acide) a été réalisée sur des protéines de foie de rats. Les détails de ces travaux sont présentés à l'annexe C.

2.5 Conclusions

La comparaison entre une étape de pré-fractionnement (par SCX ou par Hp-RP) et une approche classique par 1D-LC-MS/MS a permis de montrer la contribution du pré-fractionnement sur l'allure générale du spectre MS, la reproductibilité lors de la sélection de l'ion précurseur ainsi que sur la variabilité introduite au moment de la collecte des fractions. De plus, les approches par chromatographie liquide multidimensionnelle sont des méthodes efficaces afin d'améliorer la couverture globale du protéome.

Cependant, le nombre d'échantillons à analyser a beaucoup d'importance dans le choix d'une approche 1D vs. 2D, particulièrement lorsque la collecte de fractions se fait de manière *off-line*. Pour une étude en protéomique *bottom-up* du pied et du manteau de la moule, le pré-fractionnement par SCX et par phase inverse s'est montré efficace dans l'identification de plusieurs protéines. De plus, une complémentarité entre les méthodologies SCX et Hp-RP a été notée. Bien que ces approches se soient montrées quelque peu supérieures à l'analyse directe, peu de bénéfices en a été tirés. Par conséquent, le choix d'une approche 1D-LC a été jugé plus judicieux pour la suite des analyses.

CHAPITRE III

ÉTUDE DE LA BYSSOGÉNÈSE PAR ANALYSES PROTÉOMIQUES SUR LE BYSSUS, LE PIED ET LE MANTEAU CHEZ LES MOULES *MYTILUS* PAR LC-MS/MS

Publié dans Proteomics 2020, 10 septembre 2020, sous le titre :

Investigating Byssogenesis with Proteomics Analysis of Byssus, Foot, and Mantle in
Mytilus Mussels by LC-MS/MS

DOI : 10.1002/pmic.2020000014

Sansoucy M, Tremblay R, Carrington E, Marcotte I, Sleno L

Informations supplémentaires disponibles en ligne :

<https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.2020000014>

Une fois les conditions de pré-fractionnement établies, elles ont été appliquées pour l'étude protéomique de deux espèces de moules, de la famille des *Mytilidae*. Dans ce chapitre, une approche par 1D-LC-MS/MS optimisée a permis d'identifier les protéines ayant potentiellement un lien avec le processus d'auto-assemblage du byssus. De plus, une approche comparative entre le byssus, l'organe producteur du byssus (le pied) ainsi que le manteau de la moule, employé comme tissu contrôle, sera abordée.

Maxime Sansoucy, Réjean Tremblay, Emily Carrington, Isabelle Marcotte et Lekha Sleno sont les co-auteurs de cet article. Maxime Sansoucy a procédé à la recherche bibliographique, la préparation des protocoles, la réalisation des expériences en laboratoire, le traitement des données et la préparation du manuscrit original. La professeure Lekha Sleno a fourni son support et a supervisé le projet, vérifié l'analyse et l'interprétation des données, puis a procédé à la révision finale du manuscrit. Réjean Tremblay et Emily Carrington ont fourni les échantillons de moules et de byssus.

3.1 Résumé

Le byssus de la moule représente une classe de bio-matériaux fascinants ayant une capacité unique pour l'adhésion sur virtuellement tout type de surfaces solides. La fabrication du byssus, ou byssogénèse, est également influencée par plusieurs facteurs environnementaux. Afin de mieux comprendre ce phénomène, les protéines exprimées dans le byssus, dans l'organe producteur du byssus (le pied) ainsi que dans le manteau ont été examinées dans deux espèces de moules par LC-MS/MS. À partir de cette analyse comparative, le manteau a été employé comme tissu contrôle afin de pointer les protéines uniquement exprimées dans le pied. Puisque plusieurs protéines sont susceptibles d'être impliquées dans la fabrication et la régulation des fibres de byssus, les protéines hautement enrichies dans les échantillons de pied ont également été considérées. Bien que limité par les bases de données incomplètes en protéines, ce travail représente un avancement significatif pour l'identification des protéines

biologiquement importantes exprimées dans le pied de la moule, en plus d'accroître nos connaissances actuelles sur le protéome du byssus de ces espèces. Pour une majorité de ces protéines, leur identification ne repose que sur des preuves basées sur la transcriptomique. Plusieurs protéines homologues entre les espèces *M. edulis* et *M. californianus* ont été détectées, en plus d'une importante proportion de protéines associées à d'autres espèces de moules telle que *M. coruscus* et *M. galloprovincialis*. Les données sont disponibles via ProteomeXchange sous l'identifiant 10.6019/PXD021023

Les principales protéines structurales du byssus ont été étudiées précédemment. Il demeure, cependant, un manque quant à la connaissance du processus d'assemblage de ces protéines dans le byssus. Ces travaux apportent un regard détaillé sur les protéines exprimées dans l'organe responsable de la production du byssus et donc, sur leur potentielle implication dans la régulation et la fabrication des fibres. Le byssus ainsi que le manteau de la moule ont aussi été inclus. Considérant le peu d'information disponible sur les espèces étudiées, ce travail contribuera également à une description plus détaillée des protéomes de *Mytilus edulis* et *Mytilus californianus*.

3.2 Introduction

La fibre de byssus représente un matériau biologique offrant des caractéristiques uniques quant à la flexibilité, l'auto-réparation ainsi que pour l'adhésion sous-marine (Harrington & Waite, 2007; X. X. Qin & Waite, 1998; Schmitt et al., 2015). Les fibres peuvent également être adaptées en fonction des différents facteurs environnementaux (Brazee & Carrington, 2006; Seguin-Heine et al., 2014). Dans des eaux turbulentes, par exemple, les moules ont la capacité de modifier la structure secondaire de leurs protéines byssales afin d'augmenter l'endurance abrasive (Arnold et al., 2013). De plus, il a été rapporté qu'une exposition aux métaux lourds interfère avec l'expression de certaines protéines et la byssogénèse elle-même (Raftos

et al., 2016). Bien que ce matériau biologique ait été le sujet de nombreuses études au cours des dernières années, le processus d'assemblage du byssus demeure toutefois peu détaillé.

Ce matériau fibreux est synthétisé et sécrété par la glande du pied de l'animal, un tissu granulo-musculaire. Le byssus est produit à partir de la base (*stern*) du pied et est constitué d'un cœur en collagène pré-pepsinisé et d'une plaque faite principalement de protéines de pied (mfp) dont plusieurs sont responsables de l'attachement sous-marine. Les pré-collagènes distal et proximal (preCol D et P respectivement) offrent des propriétés complémentaires quant au soutien de charge. La région proximale des fibres de byssus est nettement plus flexible que la région distale. Cette flexibilité (ou rigidité) est en partie expliquée par un gradient de distribution le long de la fibre (Pearce & LaBarbera, 2009), en plus de la structure secondaire et la présence de liaisons sacrificielles (Arnold et al., 2013; Harrington & Waite, 2007). Aussi, une modification post-traductionnelle rare, mais cruciale, conduisant à la formation du 3,4-dihydroxyphényl-*L*-alanine (DOPA) renforce la résistance aux cassures de la fibre (Harrington et al., 2010) ainsi que l'adhésion sous l'eau (Wei et al., 2012).

Ces propriétés intrigantes ont motivé plusieurs études adressant la composition protéique du byssus depuis plus de 30 ans. Toutefois, un nombre limité de protéines a été identifié dans les fibres de byssus, incluant les protéines de pieds de moules (*mussel foot proteins*, mfps), les preCols et les protéines de la matrice de la fibres (*matrix thread proteins*) (DeMartini et al., 2017). Néanmoins, plusieurs autres protéines sont impliquées dans la régulation du processus d'assemblage du byssus. La spectrométrie de masse a été un outil inestimable pour la détermination des protéines provenant du byssus en plus de celles identifiées dans d'autres tissus de moules au cours des dernières années (Gantayet, Ohana, & Sone, 2013; Gantayet et al., 2014; S. Li et al., 2018; López, Marina, Vázquez, & Alvarez, 2002). Par exemple, des études utilisant le MALDI-TOF-MS sur *M. edulis* et *M. galloprovincialis* ont montré les protéines les

plus abondantes dans leurs pieds (López et al., 2002). Dernièrement, Gantayet et al. ont détecté par LC-MS/MS environ 204 protéines dans des échantillons de pieds provenant de *Dreissena polymorpha* (Gantayet et al., 2013, 2014). Leur discussion s'est limitée cependant aux protéines les plus abondantes trouvées dans le byssus.

Une limitation importante demeure au niveau des études en protéomiques menées sur ces espèces de moules, en raison d'un manque de protéines dans les bases de données. Plusieurs études ont utilisé la transcriptomique dans le but de caractériser les protéines exprimées dans une espèce de moule donnée cependant, la preuve directe de la présence de ces protéines ne peut pas être faite à partir de ces approches seules. À titre d'exemple, des études par transcriptomique ont été réalisées dernièrement sur *Limnoperna fortunei* (S. Li et al., 2018), *Mytilus coruscus* (C. Qin et al., 2016) et *Mytilus californianus* (DeMartini et al., 2017) dans le but de caractériser les protéines issues d'échantillons de pied, en plus de rapporter de nouvelles enzymes, des protéines à base de collagène ainsi que de nouvelles protéines de pied. Plus important encore, les protéines incluent dans les bases de données sont pour la grande majorité non révisées et sans indication additionnelle de leur existence au niveau de l'expression de protéine.

Dans cette étude, une approche par LC-MS/MS a été employée afin d'identifier les protéines uniquement présentes ou hautement enrichies dans les échantillons de pieds. Par conséquent, ces protéines présenteraient un important potentiel dans la byssogénèse. Le pied et le manteau ont été collectés sur *M. edulis* et *M. californianus* et digérés à l'aide de deux enzymes complémentaires en prévision de l'analyse par LC-MS/MS. Le byssus sécrété par ces deux espèces a également été analysé à des fins comparatives. Cette contribution propose une approche unique quant à l'étude du protéome de deux espèces de moules étroitement similaires, en plus de présenter une liste de protéines spécifiques au pied avec un important potentiel dans la fabrication et la régulation du byssus.

3.3 Expérimentale

3.3.1 Matériels

La trypsine (bovine), la pepsine (porcine), les solvants ainsi que les autres réactifs ont été achetés chez Sigma-Aldrich (St. Louis, MO). L'eau ultrapure a été obtenue à partir d'un système Millipore Synergy UV (Billerica, MA).

3.3.2 Extraction et digestion des protéines

Les moules *Mytilus edulis* (Île-du-Prince-Edward, Canada) ont été expédiées vivantes sur glace. Les tissus de pieds et de manteaux ont été disséqués, congelés instantanément puis lyophilisés. Les échantillons des moules *Mytilus californianus* (Friday Harbor, WA, USA) ont été préparés de la même manière avant d'être expédiés sur glace sèche. Un tampon de bicarbonate d'ammonium (ABC, 1 ml, 100 mM, pH 8.5) a été ajouté aux échantillons secs (environ 20-30 mg pour les pieds et de 15-20 mg pour les manteaux) avant l'homogénéisation mécanique et par sonde ultra sonique. L'homogénat a ensuite été réparti en aliquots de 200 µl. C'est aliquots ont été traités au dithiothréitol, puis par l'iodoacétamide et dilués avec 250 µl du tampon ABC en prévision de la digestion par la trypsine ou dilués avec 250 µl d'une solution de 2% TFA dans 20% méthanol pour les digestions avec la pepsine. Les échantillons ont été digérés sur une période de 18h (avec 20 µg de protéase) à 37°C, puis 50 µl d'une solution de 5% NH₄OH a été ajoutée pour neutraliser la pepsine. Une étape de désalinisation a ensuite été faite sur les digestats en utilisant des cartouches OASIS® HLB SPE (Waters, Millford, MA, 30 mg/1 cc) puis les digestats ont été conservés secs.

Pour les échantillons de byssus (contenant les fibres et les plaques), ceux-ci ont été manuellement réduits à l'état de poudre (1-2 mg) avant l'ajout de 50 µl d'une solution contenant 7 M d'urée et 2 M de thiourée. Les échantillons ont ensuite été soniqués pendant 10 min et chauffés à 50°C pendant 20 min. Une digestion par la

trypsine et la pepsine (25 µg protéase) sur les échantillons de byssus a été faite pendant 18h suivant une réduction alkylatrice ainsi qu'un ajustement de pH. Les digestats ont été sujets à la même extraction SPE que décrite plus haut.

3.3.3 Analyse RP-LC-MS/MS

Les digestats de pieds, de manteaux et de byssus ont été analysés, tel que décrit dans la publication (Sansoucy, Tremblay, Carrington, Marcotte, & Sleno, 2020) (voir annexe A) et à la section 3.3.2 ci-dessus. Brièvement, les peptides ont été injectés (20 µl) sur une colonne Aeris PEPTIDE XB-C18 100 x 2.1 mm maintenue à une température de 40°C (Phenomenex, Torrance, CA). Les phases mobiles A et B ont été préparées à partir d'eau et d'acétonitrile respectivement pour la séparation des peptides et toutes deux acidifiées avec 0.1% d'acide formique.

Les spectres MS et MS/MS ont été obtenus à partir d'un spectromètre de masse hybride quadripôle temps-de-vol TripleTOF™ 5600 (AB Sciex, Concord, ON, Canada). La source d'électro-nébulisation a été opérée en mode positif et maintenue à une température de 500°C, avec un voltage appliqué sur le capillaire de 5 kV et un potentiel de déclusterisation (DP) de 60 V. La pression des gaz de nébulisation et de séchage a été maintenue 50 psi. La sonde d'électronébulisation a été utilisée pour assimiler le flux provenant du LC alors que la sonde APCI a été employée pour la calibration après quatre injections (en mode MS et MS/MS), à l'aide d'une solution de calibrants préparée dans nos laboratoires. Les mesure MS ont été réalisées entre m/z 80-1500, avec une soustraction dynamique du bruit de fond (*Dynamic Background Subtraction*, DBS) et un critère de sélection d'information dépendante (*Information-dependant acquisition*, IDA) sur les 15 ions le plus sensibles au-delà de m/z 250. Les spectres MS/MS ont ensuite été enregistrés suivant l'accumulation (50 ms) et la collision ($CE = 30 \pm 10$ V) des ions précurseurs. Le temps de cycle total a été calculé à 1.05 s.

3.3.4 Traitement des données

Les données LC-MS/MS ont été cherchées à partir d'une base de données provenant du répertoire UniProtKB/Swiss-Prot (disponible en avril 2020) par le logiciel ProteinPilot™ (v5.0.2, Sciex). La base de données choisie contenait 11 946 protéines issues de 81 espèces de moules (Figure 3.1). Les résultats ont ensuite été téléversés dans Scaffold-Q+ (v4.9.0, Proteome Software). Les protéines identifiées ayant au minimum deux peptides à 1 % FDR ont été considérées. L'ensemble des données a été déposé dans le Consortium ProteomeXchange via l'outil de dépôt PRIDE (Deutsch et al., 2017) avec l'identifiant 10.6019/PXD021023.

3.4 Résultats et discussion

Les fibres de byssus ont directement été analysées par LC-MS/MS, en triplicata, puisque le protéome du byssus est considérablement moins complexe que celui du pied ou du manteau. Un gradient de 58 min (au lieu de 36 min) a été employé afin d'augmenter la résolution des peptides des digestats non pré-fractionnés. À l'inverse, les échantillons de pieds et de manteau ont été pré-fractionnés en utilisant deux méthodes complémentaires, dans le but d'augmenter la couverture du protéome. Tel que mentionné au chapitre 2, les peptides ont une meilleure probabilité de produire un spectre MS/MS d'excellente qualité lorsque les échantillons sont pré-fractionnés. Par conséquent, les protéines sont identifiées avec une meilleure confiance. Néanmoins, suite aux observations faites au chapitre précédent, le pré-fractionnement des peptides provenant des digestats de pied et de manteau a été jugé moins avantageux qu'une analyse directe. La même chromatographie qu'employée pour les protéines de byssus a été utilisée pour les échantillons de pieds et de manteau.

Les spectres MS/MS expérimentaux ont été comparés avec une base de données incluant 81 espèces de moules, puisque les bases de protéines actuellement disponibles

pour *Mytilus edulis* et *Mytilus californianus* sont incomplètes. La base de données inclus plusieurs protéines obtenues à partir d'expériences en transcriptomique, avec une infime partie (environ 1%) ayant été préalablement confirmée expérimentalement. La Figure 3.1a regroupe les principales espèces incluses dans la base de données, avec plusieurs séquences de protéines non-révisées provenant du genre *bivavle*. En comparant la distribution des espèces pour les protéines identifiées dans *M. edulis* (MYTED, Figure 3.1b) et *M. californianus* (MYTCA, Figure 3.1c), il est intéressant de remarquer l'importante homologie qui existe entre les espèces étudiées, avec une importante proportion de ces protéines appartenant à *M. coruscus*, *M. galloprovincialis*, et *M. trossulus*. Il est toutefois intéressant de noter que seulement 13 et 8% des protéines identifiées appartiennent à *M. edulis* et *M. californianus* respectivement. Ceci est un exemple concret de l'impact d'une base de données incomplète sur l'identification des protéines. Néanmoins, plusieurs des protéines présentées dans cette étude n'auraient pas été détectées si elles n'avaient pas été au préalable incluent dans la base de données. Présentement, les données LC-MS/MS aident à montrer les régions communes entre plusieurs espèces.

La comparaison entre les protéines identifiées est présentée dans le diagramme de Venn (

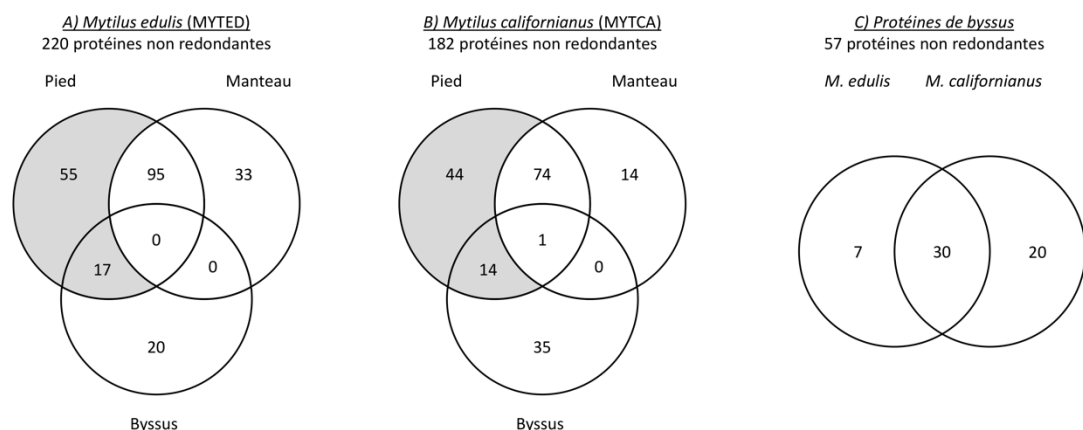


Figure 3.2). Le total de protéines trouvées seulement dans les échantillons de pieds se chiffre à 73 et 58 pour *M. edulis* et *M. californianus* respectivement, en utilisant les protéines trouvées dans le manteau comme échantillon contrôle. Pour ce qui est des analyses sur le byssus, 37 (*M. edulis*) et 50 (*M. californianus*) protéines ont été identifiées selon les critères énoncés plus haut, avec 30 protéines partagées entre *M. edulis* et *M. californianus*. Ces protéines spécifiques au pied de la moule ont été étudiées plus en détails puisque l'un des objectifs de cette étude est de pouvoir relier certaines protéines à la fabrication et la régulation de la fibre de byssus. Une liste complète des protéines identifiées dans les trois types d'échantillons est disponible en ligne (information supplémentaire S1 et S2), ainsi que le nombre de peptides confidentiels à 1 % FDR et le pourcentage de couverture de la séquence. Dans le même ordre d'idée, le tableau S3 (voir information supplémentaire disponible en ligne) compile l'information additionnelle telles que la masse moléculaire, le pI, le pourcentage d'abondance de certains acides aminés, l'indice GRAVY ainsi que des annotations GO pour les protéines uniques au pied.

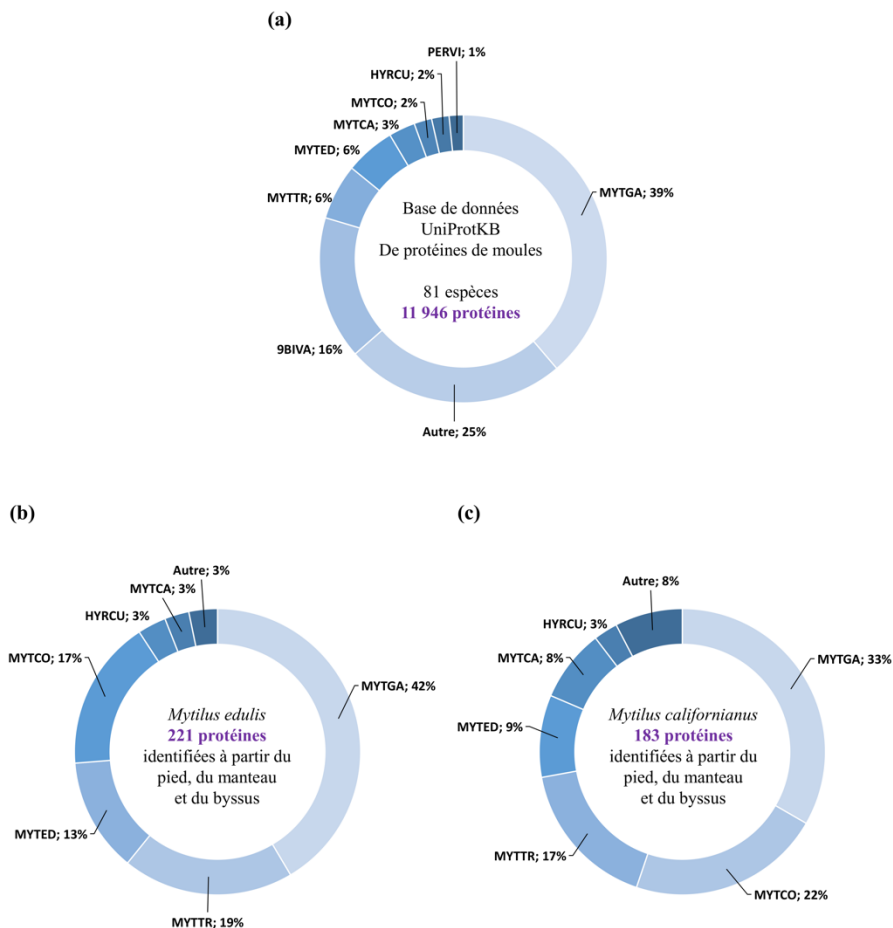


Figure 3.1 Distribution relative des protéines provenant de la base de données (a), ainsi que les protéines identifiées à partir de MYTED (b) et MYTCA (c) en combinant les protéines de byssus, de pied et de manteau.

Pour le byssus, une grande proportion des protéines trouvées n'a pas d'annotation, suivit des protéines impliquées dans des fonctions de liaisons avec métaux. Ces résultats ne sont toutefois pas surprenants, puisqu'il est bien établi que le byssus possède d'importantes capacités de chélation.

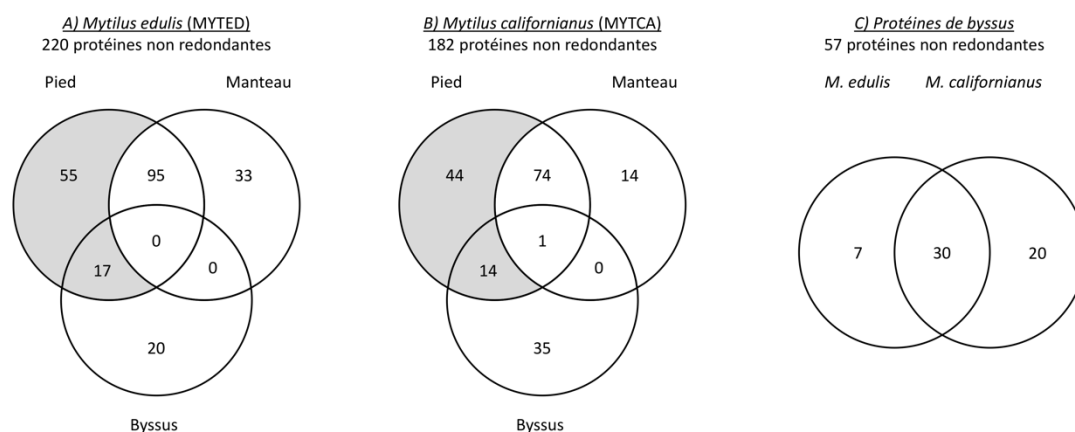


Figure 3.2 Nombre de protéines identifiées dans les échantillons de pied, de manteau et de byssus à partir de *M. edulis* (A) et *M. californianus* (B). Les protéines trouvées dans les échantillons de byssus sont comparées en (C) pour les deux espèces.

Le principal objectif de cette étude visait l'identification de protéines provenant du pied de la moule potentiellement impliquées dans la byssogénèse. Plusieurs protéines identifiées sont connues comme étant des constituants du byssus (J. Herbert Waite, 2017). Le tableau 3.1 liste les protéines trouvées dans les échantillons de pieds et de byssus associées à la fabrication de la fibre. Cette liste inclut aussi des enzymes spécifiques au byssus, telles que la glycosyl-hydrolase, les peroxydases, les tyrosinases et la procollagène-proline di-oxygénase. Finalement, dix protéines distinctes de pieds ont été ajoutées, dont cinq ont été rapportées dans les deux espèces.

Tableau 3.1 Protéines associées à la fabrication de la fibre identifiée dans les échantillons de pied et de byssus pour *M. edulis* et *M. californianus*²

Nom des protéines	
Échantillon de pied	
Byssal calumenin-like 1	Foot protein 6, 10, 12, 18, 19
Byssal glycosyl-hydrolase-like 2	Precollagen (D, P, NG)
Byssal peroxidase-like 2	Precollagen-proline dioxygenase
Byssal protease inhibitor-like 1	Protease inhibitor-like 1
Byssal tyrosinase-like 1, 2	Proximal thread matrix 1, 1a
Collagen-like 5, 6, 7	TSP_1 domain containing 1
Foot protein 2	
Échantillon de byssus	
Byssal glycosyl-hydrolase-like 2	Foot protein 3, 11, 17
Byssal peroxidase-like 1, 2, 4	Mussel byssus collagen-like 3
Byssal protease inhibitor-like 1	Precollagen (D, P, NG)
Byssal protein 1, 2, 3	Protease inhibitor-like 1, 1D, D2
Byssal tyrosinase-like 1, 2	Proximal thread matrix 1, 1a
Collagen chain alpha 1	Thread matrix 2A
Collagen-like 4, 5, 6, 7	TSP_domain containing 1
Foot protein 2, 4, 6, 10, 12	Tyrosinase 5

Les protéines de pied de moules sont bien connues pour leur rôle crucial dans la détermination des propriétés uniques d'adhésion du byssus. Les protéines mfp- 2, 4,

²Les protéines en gras ont été identifiées dans *M. edulis* et *M. californianus*

6, 10, et 12 ont été identifiées dans les échantillons de pied et/ou de byssus pour les deux espèces. La mfp-3, ainsi que trois variants, ont été trouvés dans le byssus de *M. californianus*. Ces protéines de pied ont été identifiées sur la base des séquences disponibles pour MYTED et MYTCA pour la mfp-2 et seulement pour MYTCA pour les protéines mfp-3, 4 et 6 (variants 1 et 2). Le troisième variant pour mfp-6 (provenant de MYTCO) a été identifié dans chacune des espèces, tandis que les variants 1 et 2 n'ont été trouvés que pour *M. californianus*. Six protéines mfps provenant de *M. californianus*, récemment rapportées dans une étude en transcriptomique (DeMartini et al., 2017), ont été détectées ici pour la première fois. Il s'agit des protéines mfp-10, mfp-11, mfp-12, mfp-17, mfp-18 ainsi que la mfp-19, qui ont été exclusivement identifiées dans *M. californianus*. La Figure 3.3 montre la séquence primaire de toutes les protéines mfps détectées, en plus d'indiquer les régions qui ont été identifiées avec un minimum de 95% de confiance dans les échantillons MYTED et MYTCA.

En plus des précollagènes (distal, proximal et non-gradient), deux classes d'enzymes connues pour avoir un impact sur l'adhésion et la plasticité du byssus ont été trouvées. La première classe est responsable de la formation du DOPA et inclus les protéines byssales tyrosinase-homologue 1 et 2 (provenant de MYTCO) ainsi que la tyrosinase 5 (provenant de PERVI). La modification DOPA est proposée comme ayant un rôle dans l'adhésion ainsi que dans les propriétés structurales du byssus (Harrington et al., 2010; Petrone et al., 2015; Wei et al., 2012; Yu et al., 2011). Dans le même ordre d'idées, trois protéines byssales peroxydase-homologues, dernièrement identifiées à partir d'expériences en transcriptomique faites sur *M. coruscus*, ont été identifiées dans notre ensemble de protéines. Le rôle de ces protéines sous-entend qu'elles protègent les fibres de byssus contre les espèces réactives de l'oxygène. Ces radicaux peuvent causer des réactions adverses impliquant la DOPA dans des réactions croisées inter- ou intra-moléculaires entre diverses protéines de byssus. Les liaisons croisées dans le byssus sont soupçonnées d'avoir un impact négatif sur l'adhésion et l'élasticité du

byssus (Zhao & Waite, 2005). Cependant, c'est une piste encore aujourd'hui peu explorée (J. Herbert Waite, 2017; Yu et al., 2013).

Foot protein 2 (Q1XBT8_MYTED)

MLFSFLLLTCTQLCLGTSRPRYDDDDDDYTTPVYKPSPSRYRPNVNPCLKKPCKYNGVCKPSSGGSYKCVCKGGYGYNCNLKNACKPNQCKN
KSRICIPVGKTFKCECRNGNFGRLCERNVCSNPCKNKGKCAPLGKTGYKCTCSGGYTGPCEVHACKPNPCKNNGRCYPDGKTGYKCKCVGG
YSGPTCQENACKPNPCRNGGKCSADKFGDYTCDCRPGYFGPQCEKYVCAPNPCKNSGICSSDGSGGYRCKCKGGYSGPTCKVNVCKPNPRKN
SGRCVNRKSSSYNCICKGGYSGPTCGVNAACKPNPCKNSGRCVNRKSSSYNCICKGGYSGPKCEHVCKPNPCQNRGRVCVPEGRDGYRCKCVGGY
SGPTCDENVCKPNPCQNKGRCPDNDSDGFKCRCLGGYKGPTECDKPNPCNTKPCNKGKCNNGKTYTCKCAYGYRGRHCTAKAYKPNPCA
SRPCKNRGKCTVKGTGYVCTCAKGYSGRYCALKSPPSYNDDDEY

Foot protein 2 (AOA223HCL0_MYTCA)

MLFSFLLLTCTQLCLGTAPPQYEDDDDDYKPDYAYKSPPKYRPNVNPCLKKPCKYNGQCKPNSGSSYKICIRGGYGYNCNLKNACKPNPCK
NKSCKVPVGKTFKCIICRLGNFGRLCEKNVCSNPCKNRGKCPVGKTGYKCKCGGGYSGPRCEVHVCKPNPCKNKGRCYPDGKTGYKCKCVG
GYSGPTCQENACIPNPCSNNGKCSADKFGDYTCDCRPGYFGPECEKYVCAPNPCKNNGKCSDDGSGGYKCKTGGYSGPTCNVHVCKPNPCE
NRGKCIIPRGNGYKCNCKGGYSGPTCGENVCKPNPCQNRGRCPDNDSDGFKCKCVGGYSGPTCEDKPNPCNSKPCNKGKCIYNGRTYTCKC
AYGYRGRHCTDKAYKPNPCASRPCKNGGRCTVKNGYVCKCTKGYSGKYCALKSPPSYDDEY

Foot protein 3 variant 3 (Q2VPV5_MYTCA)

MNKFSTVLLALVLIGFAVQSDAGYGYDLGYNAPWPYNNGYGYNGYNGYHGRYGWNKGWNGFPWGGSYYGNGKGYLY

Foot protein 3 variant 4 (Q2VPV4_MYTCA)

MNKFSTVLLALVLIGFAVQSDAGYDYKGYNSPWPYNNGYGYNGYNGYHGRYGWNKGWNGFPWGGSYYGNGKGYLY

Foot protein 4 variant 1 (A1XF84_MYTCA)

MKRGLFTILLITAVLVVVAESYGRRYGEPSTGYANIGHRRYERAI SFHRHSHVHGHLHHRHVHRHSLHGHVHMHRVSHRIMHRHRLHGH
VHRHRLHNNHVRHSLHGHVHRHRLHHRVHRHNLHGHVHRHRLHKKHVNHRVHLKHLHKKHQLHGHVHRHQLHKKHVNHRVHLKHLH
KHQVLHGHVHTHRVHLKHHVHKHRLHKKHLHKKHQLHGHVHTHRVHLKHHVHKHRLHKKHLHKKHQLHGHVHM
RVLHKHHVHKHRLHKKHVVHKKHSHRVHLKHHVHKHVEHQHVHKKHVLHHRVHSHHVVHSHVHKHVRVHSHVHKHNVHSHVHRHQI
LHRHVHRHQVVRHVRHRLIAHRHIIHSHQA AVHRHVHTHVFEENFDDGTDVNLRI RHGI IYFGGNTYRLSGRRRFFMTLWQECLESYGSDS
ECFVOLGNQHLFTVVQGHSTSFSDLSNDLHPDNNIEQIANDHVNDIAQSTDGDINDFADTHYNDVAPIADVHVDNIAQTADNHVKNIAQT
AHHHVNDVAQIADHDVNDIGQTAYDVNNIGQTADHDVNDIAQTADHDVNAIAQTADHDVNAIAQTADHDVNDIGDTANSHIVRVQGVAKNH
LYGINKAIGKHIQHLKDVSNRHIEKLNNHATKNLLQSALQHKQQTIEREIQHKRHLSEKEDINLQHENAMKSKVSYDGPVFNEKSVVSNQG
SYNEKVPVLSNGGGYNGKVSALSDQGSYNEGYAY

Foot protein 6 variant 3 (D3JVG2_MYTCO)

MKSIQMILAVFVITALCGIVKSGRGDYDYRGYCSNKGCRPGYFFYDNRGYCKYGTSTRSYKYDCNSYAGCCLPRNPYGTVKYYCTNKYGCNDY
YFYNNKGYYYNKKDYFNCGSYNGCCLRSY

Foot protein 6 variant 1 (Q0H216_MYTCA)

MKSIQMILAVFVITALCGIVESGGGNYRGYCSNKGCRSGYIFYDNRGFCKYGSSTYKYDCGRYAGCCLPRNPYGRVKYYCTKKYSCPDFFY
YNNKGYYYNKKDYFNCGSYNGCCLRSY

Foot protein 6 variant 2 (Q0H215_MYTCA)

MKSIQMILAVFVITALCGIVESGGGNYRGYCSNKGCRSGYIFYDNRGYCKYGSSTYKYDCGRYAGCCLPRNPYGNVYYCTKKNACPKDFYF
YNNKGSYYYKRNAFYDCRSYNGCCLRSY

Foot protein 6 variant 3 (D3JVG2_MYTCO)

MKSIQMILAVFVITALCGIVKSGRGDYDYRGYCSNKGCRPGYFFYDNRGYCKYGTSTRSYKYDCNSYAGCCLPRNPYGTVKYYCTNKYGCNDY
YFYNNKGYYYNKKDYFNCGSYNGCCLRSY

Foot protein 10 (A0A223HCJ8_MYTCA)

MSRLICMLFLVVAAYDYNNSKT**VIVKHSAL**GDEYIGIGIPSVVSYSDFSNNKYHGFKISNGRFSFGGIAYKLKCGQQR**FYNVWSNCWSIYH**
SRSVCFDRLKKQSLFVDYPGGNIIDDGYKAVVVKNAVYGGDDDIGYGSVSY**LDYFNKFHGFKISDNRF**SYGGK**AYTLKCGLQRF**YNVWGGC
 WSNYHSRSVCFDRLMNQNLFAVYPGDYDYGAVYGGD**YDIGVGS**VSY**LDYCNKFHGFKISDNRF**SYGGTAYTLRCGLQRF**YNLWGNCWNSY**
 HSRKVCFDRLRTQNLFAVYPGGYDNDGSVTAVVKNVNSNDLY

Foot protein 11 (A0A223HCP4_MYTCA)

MSYGICVLVMAVAATSGYNTYSGSHHHLRLYRKHIVHSGHYGHHGLAYYKRYHGFKITHGKFAYKGLGYKLTCGKTRFYRLWRSCYGR
 YHSRGRCFGRFRKQHLFAVYKGYSHGSTGHYYKGHGHGYGHGHGYSHGHGHVYIHGHGHGHGHGYSHGHSHVYIHGHGHGHGHGFSGHGHHR
 AAYSYYYKHYGFRHHHRFSYGGKAYKLACHRNRFYSLWNTCYTHYHSRSCYCFGLRQHLFVSYHGGAYGHIHHAAHHGHGTHDLYRHNH
 FLHLKMAKHHVKRIHHAANSHVSKISYLSNKHVQNIQYQAKKHVSRISEANRHINRFRKIVKYHINRLHNHAKHHLSAVENQHKLRLERE
 YQNRHLSNEEHLHNKHEYHISKHSEVGYGYGHGHVVS**YLDYHNTYHGFP**IAEHRFSYAGKAYQLKCESKRFYSLWSDCWSQYHSR**SICFD**
 NLRKQHLFIDYPGDYADEYAPTKIVVKTGGDFDYK

Foot protein 12 (A0A223HCL5_MYTCA)

MACGILVLLIAAIAASTTGYSNDYYPDGKDDYSYKAVNSYSYKPKLVKSTAYSSSSSHVSHHHGHAGY**AGHKLITGHA**HRFKVGRIFKRF
HLRGVELIGRGVR**YNGFYGR**LCSATKLYGALNSCYED**YGDIGDCFGLIKQGLIGRRHIYHKYGHGFGLGVT**FVKRGIVYGR**SFYKFGCRR**
 PKFIRYMTSCYNKFHHASTCVSVLRKRHLIRTPHSYHGHMHLAGYTRFRFRKIRHGRFRYGGKYYKLSCNKRFRFYSHWNRCYGRY**HSVGRCF**
GYLRRKHLFSYNGYSSGVVYNTYGPYSGNITFD**YGYKSHHGFTIRNNKFSYGGIGYQL**MCGRSR**FYRLWGE**CYNQYNSVSSCFGLRLRQ
HLFYRYGGEDYSHVGGDAYGPDYSSGPIFD**HSYSSHHGFR**IRHHR**FSYNGIAYK**LKCSRSR**FYRLWGQ**CYHRY**HSVGTCFGR**LSHHL**FLYR**
 YIGDHHAILNKHLYLHKKIADNHARRIHNVAKSHVAR**ISHVSNKHVRNIQYQANKHVSRIANEARSHMNR**FQQLVDYHLKRLHKHTKYHLSA
 AEHQHKKHLNERQYQNNKHELNVRRHLKKHEYNLKKHGYAYGHGHGHGHGHGHGHGHVVSQDYYSKYHG**FKLDHHRFSYGGKAYQLTCG**
 DSRFYXYWSDCWSHHHSR**SVCFGLRKRHLFDVYSDG**YDEEDDYEPVAKQSAY

Foot protein 17 (A0A223HCQ3_MYTCA)

MLLMYQILCVLLFGVVYVTCGSNKCLSOPDYQLQGYEYSRAAVRGTSQKQLDQLRLGRSSIDVTGMLSRLKKYCYRSNIRSGTARHDDYSSD
 PYTEPQKGKVVVTSLCDSNPTSTTFKGKWDGCLYPEWRNWKCPYVKCNPKLCNGITSPNKYTWFVCQPDSEYIDLLVACPKNKGWFLTT
 KRAWFPCCGCKKYSWQCV

Foot protein 18 (A0A223HCK1_MYTCA)

MKFVIVAVLVALCLAGESAAGRYQSYCDKTLVSVPCAYQNCHNACAYKGCHGHHRCNLNNYKCCCTCTGTHKK

Foot protein 19 (A0A223HCK5_MYTCA)

MFRFTCKILFSIFLTFVLNTEVSGSQCGWGCPTLPPNTVGPCVELCPSAGCDFGYLCCSNGCGHVCKIGIWTCTY

Figure 3.3 Couverture des séquences des protéines provenant des sous-ensembles de pieds et de byssus. La séquence des peptides détectée (à 1% FDR) est présentée en gras pour *M. edulis* et soulignée pour *M. californianus*.

La présente étude inclus également 14 protéines «non caractérisées» obtenues à partir d'information issue d'études en transcriptomiques réalisées sur *M. coruscus*, *M. galloprovincialis* et *M. trossulus* (voir les tableaux supplémentaires S1 et S3 disponibles en ligne). Les séquences de ces protéines ont été soumises à une recherche BLAST (<https://www.uniprot.org/blast/>). Les protéines analogues sont présentées dans le tableau suivant, qui inclus également le pourcentage d'homologie ainsi que le numéro d'accèsion pour ces protéines.

Tableau 3.2 Résultats des recherches BLAST pour les protéines non caractérisées trouvées dans les échantillons de pied et de byssus

# d' ascension de la protéine non caractérisée	protéine homologue (BLAST)	% d' identité	# d' ascension de la protéine similaire
A0A409V718_MYTGA	glycine-rich cell wall structural protein 2-like	47.8	A0A2H3XBY0_PHODC
A0A3L5TUM8_MYTGA	Metalloproteinase inhibitor 1	36.8	A0A2F0BPN3_ESCRO
A0A3R5Q664_MYTGA	C2H2-type zinc finger-containing protein	48.8	Q54GK4_DICDI
A0A3L5TR65_MYTGA	Ventral vesicle protein	37.9	A0A225W0N5_9STRA
A0A3R5TGW4_MYTGA	peroxidase isoform X4	44.4	A0A1S3IU77_LINUN
A0A3L5TR12_MYTGA	Superoxide dismutase [Cu-Zn] 4AP	55.2	A0A369S5E9_9METZ
A0A077H3M5_MYTTR	Thymosin beta 4 X-linked	92.3	A0A3Q3FTL2_9LABR
A0A077GYU5_MYTTR	ATP synthase subunit d, mitochondrial	38.9	A0A433TFP9_ELYCH
A0A3L5TS65_MYTGA	Vacuolar protein sorting-associated protein 13A	56.0	A0A2H3XBY0_PHODC
A0A3L5TRH9_MYTGA	Dysferlin	75.2	A0A2F0BPN3_ESCRO
A0A3G1CJM7_MYTTR	Myosin, essential light chain, adductor muscle	71.6	Q54GK4_DICDI
A0A3L5TRE4_MYTGA	MANSC domain-containing protein	45.7	A0A225W0N5_9STRA
A0A3L5TUS2_MYTGA	Titin	41.1	A0A1S3IU77_LINUN

3.5 Conclusions

Avec cette étude, une liste de protéines spécifiques au pied de la moule potentiellement impliquées dans la byssogénèse a été présentée. Bien entendu, des études plus poussées devront être menées dans le but de clarifier les rôles spécifiques que jouent ces protéines dans l'assemblage de la fibre de byssus. Avec le peu d'information disponible sur le protéome des espèces étudiées, la majorité des protéines incluses dans les bases de données sont purement identifiées sur l'homologie des séquences que partagent certaines espèces entre elles. De plus, il existe trois types de glandes individuelles (centrale, cuticule et plaque) situées dans le pied. Ces glandes sont hautement spécialisées dans la ségrégation des protéines spécifiques au byssus. Une étude portant sur la caractérisation des protéines produites par chaque glande sera un suivi intéressant qui devrait apporter des connaissances supplémentaires sur l'implication de ces protéines dans la byssogénèse. Dès que les bases de données seront plus complètes pour les espèces étudiées, les données de cette étude pourront être interrogées de nouveau. Finalement, il serait intéressant d'étudier les modifications post-traductionnelles sur les différentes protéines du byssus.

CHAPITRE IV

ÉTUDE DES GLANDES DU PIED DE LA MOULE MYTILUS EDULIS PAR DES ANALYSES TRANSCRIPTOMIQUES ET PROTÉOMIQUES

Les travaux présentés au chapitre précédent ont souligné la faiblesse d'une base de données peu détaillée. Quelques protéines du byssus ont précédemment été identifiées dans la moule de Californie, mais également dans la moule *Mytilus edulis* par homologie. Faute d'information complémentaire, une base de données a été construite à partir du transcriptome de *M. edulis*. Dans ce chapitre, les protéines provenant de différentes glandes du pied ont été digérées et analysées par notre méthodologie introduite au chapitre précédent. Les protéines ont été identifiées à l'aide d'une base de données de protéines et d'une banque de données créée à partir du transcriptome de la moule *M. edulis*.

Maxime Sansoucy a procédé à la recherche bibliographique, la préparation des protocoles, la réalisation des expériences en laboratoire, le traitement des données et la préparation du manuscrit original. La professeure Lekha Sleno a fourni son support et a supervisé le projet, vérifié l'analyse et l'interprétation des données. Fransica Jehle a procédé à la collecte des différentes glandes du pied de la moule ainsi que du manteau. Mihai Mesko a procédé à l'expression ainsi qu'à la purification de la protéine recombinante MFP-1. Yonatan Morocz et Daniel John Jackson ont contribué à l'élaboration de la base de données de protéines à partir du transcriptome de la moule *M. edulis*. Les résultats présentés dans ce chapitre feront l'objet d'une publication ultérieure dans le *Journal of Proteome Research*.

4.1 Résumé

Le succès de l'identification des protéines à partir d'une base de donnée protéomique dépend principalement de la complétion de celle-ci. Par exemple, plusieurs protéines identifiées dans notre étude présentée au chapitre 3 ont été identifiées par homologie avec une espèce similaire. Néanmoins, certaines de ces protéines ont été identifiées pour la première fois à partir d'information obtenue expérimentalement dans la moule *Mytilus californianus*.

Cette étude fait suite aux travaux précédents, en impliquant cette fois une base de données obtenue à partir d'information transcriptomique. Cette nouvelle base de données, contenant seulement des protéines retrouvées dans le byssus, a été utilisée à titre de complément à la base protéomique. Bien que les principales protéines structurales du byssus aient été étudiées depuis de nombreuses années, il demeure un manque quant à l'étude du processus d'assemblage de ces protéines dans le byssus. Cette étude apporte un regard détaillé sur les protéines exprimées dans l'organe responsable de la production du byssus, à partir des glandes qui les produisent et indirectement, sur leur potentielle implication dans la régulation et la fabrication des fibres de byssus. Considérant le peu d'information disponible sur les espèces étudiées, ces travaux contribueront également à une description plus détaillée du protéome de la moule *Mytilus edulis*.

4.2 Introduction

Pendant plus de 30 ans, le byssus produit par les moules a été étudié dans le but de comprendre et reproduire les propriétés adhésives et mécaniques (DeMartini et al., 2017; Gantayet et al., 2014; Leblanc et al., 2012; S. Li et al., 2018; C. Qin et al., 2016; X. Qin & Waite, 1995b; X. X. Qin & Waite, 1998; Suhre et al., 2014). Malheureusement, l'identification de ces protéines à partir des bases de données

actuellement disponibles demeure limitée en raison du peu de protéines annotées incluses dans les banques de données.

Une manière de pallier au manque d'annotation est d'utiliser la transcriptomique comme outil complémentaire à la protéomique. La combinaison entre la transcriptomique et la protéomique reste un défi de taille, malgré les récentes avancées technologiques (Nie, Wu, Culley, Scholten, & Zhang, 2007). Néanmoins, plusieurs groupes ont utilisé une approche transcriptomique chez les moules *Limnoperna fortunei* (S. Li et al., 2018) et *Mytilus coruscus* (C. Qin et al., 2016) afin de pallier au manque d'information lors de leurs études protéomiques. Parallèlement, une approche strictement transcriptomique ayant été réalisée sur *Mytilus californianus* a révélé plusieurs protéines de pieds (*mussel foot proteins*, mfps) additionnelles (DeMartini et al., 2017). Cependant, l'existence de ces protéines demeure encore à démontrer d'un point de vue expérimental.

Ces travaux font le suivi de l'étude introduite au chapitre précédent et présentent la contribution de la transcriptomique sur l'identification des protéines provenant de la moule *Mytilus edulis* par LC-MS/MS. Les principales glandes du pied de la moule responsables de la fabrication du byssus ont été collectées et digérées, à l'aide d'enzymes complémentaires, en prévision des analyses par spectrométrie de masse. Les résultats obtenus par la base de données transcriptomique ont été ensuite comparés avec les résultats obtenus à l'aide de la base de données de protéines (UniProtKB/Swiss-Prot).

4.3 Expérimentale

4.3.1 Matériel

La trypsine (bovine), la pepsine (porcine), les solvants ainsi que tout autres produits chimiques ont été acheté chez Sigma-Aldrich (St. Louis, MO). De l'eau ultrapure a été obtenue à partir d'un système Milipore Synergy UV (Billerica, MA).

4.3.2 Collection des glandes du pied de *M. edulis*

Trois glandes du pied de la moule (glande centrale, cuticule et plaque) ont été disséquées de *Mytilus edulis*, tel que détaillé précédemment (Priemel et al., 2017). Brièvement, le pied a été coupé à la base à l'aide d'un scalpel et directement transféré dans un vase de pétri propre. Par la suite, les tissus provenant du pied ont été disséqués en fines sections transversales, sous un microscope, avec l'aide d'un scalpel chirurgical. Un test de coloration réalisé sur le tissu restant a permis de révéler les différentes glandes du pied. La glande de la plaque, qui se situe essentiellement autour de la dépression distale (voir Figure 1.1), a été disséquée puis directement transférée dans des microtubes en polypropylène de 1.5 ml et conservés sur glace. Les glandes cuticule et centrale sont respectivement localisées aux bords supérieurs latéraux et centraux du pied (Figure 1.1). La procédure a ensuite été répétée lors de la collection des échantillons de manteau. L'ensemble des échantillons a été conservé sur glace avant le transport sur glace sèche pour finalement être conservé à -30°C. Pour les fibres de byssus, ces dernières ont été collectées à partir de moules vivantes dans un laboratoire d'aquaculture à l'Institut des Sciences de la Mer de Rimouski (ISMER) à l'UQAR.

4.3.3 Digestion des protéines provenant de chaque glande

Trois échantillons de chaque type (glande centrale, plaque, cuticule et manteau) ont été solubilisés dans une solution de 7 M d'urée et 2 M de thiourée (5 µl par mg de tissu). Les échantillons ont ensuite été homogénéisés à l'aide d'une sonde ultrasonique (6 x 5s), puis conservés sur glace. Les homogénats ont été centrifugés durant 30s (14 000 rpm) afin de précipiter tout matériel insoluble. Un aliquot de 100 µl a été ensuite mélangé à 200 µl d'une solution tampon de bicarbonate d'ammonium (100 mM, pH 8.5). Les échantillons ont été traités avec du dithiothréitol (3 mM), puis par l'iodoacétamide (5 mM) en prévision des digestions enzymatiques. Deux aliquots de 100 µl chacun (l'un pour la trypsine et le second pour la pepsine), préalablement alkylés par l'iodoacétamide, ont été dilués avec 400 µl de la solution tampon de bicarbonate d'ammonium pour les échantillons digérés par la trypsine ou dilués dans 400 µl d'une solution de 2% d'acide formique préparée dans 10% méthanol pour les échantillons digérés par la pepsine. Pour le byssus, 50 µl d'une solution de 7 M d'urée et de 2 M de thiourée ont été ajoutés aux échantillons (entre 1.5-2.0 mg de fibres broyées à l'aide d'un pilon et d'un mortier) pour ensuite être soniqués 15 min et chauffés à 50°C (20 min). Par la suite, 150 µl de la solution tampon de bicarbonate d'ammonium ont été ajoutés précédant l'étape de réduction alkylatrice, tel que décrit plus haut. Ensuite, 400 µl de bicarbonate d'ammonium ou de 400 µl d'une solution 2% d'acide formique préparée dans 10% méthanol ont été ajoutés, pour une digestion avec la trypsine ou la pepsine respectivement. La trypsine et la pepsine (20 µg) ont été ajoutés aux échantillons et incubés durant 18h à 37°C.

Les digestions ont été arrêtées suite à l'ajout de 200 µl d'une solution à 5% d'acide formique pour les échantillons digérés par la trypsine et de 200 µl d'une solution à 5% d'hydroxyde d'ammonium pour les échantillons digérés par la pepsine. Les mélanges de peptides ont été dessalés lors d'une étape d'extraction en phase solide (SPE) employant des cartouches OASIS® HLB SPE (Waters, Milford, MA, 30 mg/1

cc). Les peptides ont été élués avec 100% méthanol, séchés et reconstitués dans 200 µl d'une solution à 10% acétonitrile en prévision des analyses par LC-MS/MS. Les échantillons de byssus ont été traités tel que décrit précédemment (Sansoucy et al., 2020) et digérés de la manière décrite au paragraphe précédent.

4.3.4 Analyse RP-LC-MS/MS

Les échantillons de peptides ont été injectés (20 µl) sur une colonne chromatographique Aeris PEPTIDE XB-C18 100 x 2.1 mm (Phenomenex, Torrance, CA) maintenue à une température de 40°C, avec de l'eau et de l'acétonitrile toutes deux acidifiées avec 0.1% d'acide formique comme phase mobile A et B respectivement. Un système Nexera UHPLC (Shimadzu, Columbia, MD) a été opéré à un débit de 300 µl/min avec le gradient d'élution suivant : 5% B maintenu durant 3 min, augmenté à 30% B à 50 min, à 50% à 55 min, puis finalement maintenu à 90% B de 56 à 58 min. Les spectres MS¹ et MS² ont été enregistrés sur un spectromètre de masse hybride quadripôle temps-de-vol (TripleTOF™ 5600, Sciex, Concord, ON, Canada) opéré en mode positif. La source d'ionisation a été maintenue à une température de 500°C avec un voltage de nébulisation de 5 kV et un potentiel de déclusterisation (DP) de 60 V. La pression associée aux gaz de nébulisation et de séchage a été maintenue à 50 psi, alors que le gaz d'entrée (curtain gas) a été maintenu à une pression de 35 psi. L'enregistrement des spectres MS (m/z 120-1250), suivi de l'enregistrement jusqu'à concurrence de 15 spectres MS/MS (m/z 80-1500), s'est fait en employant la soustraction dynamique du bruit de fond, avec une énergie de collision de 30 ± 10 V en mode de données dépendantes pour un temps de cycle total de 1.05 s.

4.3.5 Analyse transcriptomique des protéines associées au byssus

Le séquençage simple (*single end*) ainsi que le séquençage paillé (*paired end*) ont tous deux été filtrés à plusieurs reprises, sur plusieurs critères, avant l'étape

d'assemblage. Brièvement, la totalité des données brutes a été initialement traitée par une version d'une ligne de commande provenant du logiciel FastQC (version 0.11.9) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Le reste de la séquence adaptatrice a été coupé en utilisant BBDuk à partir de l'ensemble BBTools employé avec le fichier adaptateur par défaut Illumina, ayant une sous-chaîne (k-mer) entre 8 et 25 k-mers. La qualité de la coupe pour chaque lecture a été évaluée à l'aide du logiciel trimmomatic (version 0.36) avant d'appliquer une correction. Par la suite, un k-mer de 50 pb a été utilisé pour les corrections réalisées par Tadpole dans l'environnement BBTools. Les potentiels contaminants provenant des lectures de la ARNr ont été soustraits dans BBDuk, dans l'environnement BBTools. Une librairie de ribo-kmer a été dérivée de la banque de données Silva (Quast et al., 2013) et publiquement fournis par Brian Bushnell avec un k-mer de 31 pb. Finalement, les lectures ont été filtrées selon leur longueur. Chaque lecture filtrée a été analysée par FastQC (version 0.11.9) dans le but d'estimer l'impact global pour l'ensemble des mesures de filtres. Les données ont été assemblées à l'aide du logiciel Trans-Abyss (version 2.0.1) (Robertson et al., 2010). Une analyse par expression différentielle des gènes a été réalisé avec ClC Genomic Workbench (version 20.0.3). La lecture pour le séquençage simple pour chaque tissu et réplicat biologique a été cartographiée à partir du transcriptome groupé en utilisant l'outil *RNA-seq analysis* avec les paramètres par défaut. Une *p*-value corrigée pour le FDR a été employée pour identifier les transcrits exprimés de manières différentes entre les tissus.

4.3.6 Traitement de données

Les données obtenues par LC-MS/MS ont été cherchées à partir les protéines de moules provenant de la base de données UniProtKB/Swiss-Prot (74 377 protéines issues de 71 espèces, juillet 2021) à l'aide du logiciel ProteinPilot (algorithme Paragon) disponible en tant qu'outil de la série 3.1 OneOmics disponible en ligne (oneomics.sciexcloud.com, Sciex, Concord, ON, Canada), avec l'iodoacétamide

spécifié comme agent d'alkylation sur les cystéines, la trypsine ou la pepsine mentionnée pour la digestion ainsi que les modifications biologiques comme effort d'identification. Les recherches ont été réalisées avec une tolérance sur la masse de 0.05 et de 0.1 Da pour les ions précurseurs et produits respectivement. Les spectres MS et MS/MS ont également été cherchés à partir d'une base de données maison construite à partir d'information transcriptomique comportant 629 séquences possibles provenant de la moule *M. edulis*. Finalement, les données ont été cherchées à partir d'une troisième base de données composée de protéines associées uniquement à *Mytilus edulis*, téléchargée à partir du répertoire UniProtKB/Swiss-Prot (671 protéines, juillet 2021). Les protéines identifiées dans au moins deux répliques biologiques et avec un taux de fausse découverte de 1% ont été considérées. Les données ont ensuite été déposées dans le répertoire ProteomeXchange Consortium à l'aide de l'assistant PRIDE (Deutsch et al., 2017) sous l'identifiant de donnée 10.6019/PXD026050.

4.4 Résultats et discussion

En examinant plus attentivement le statut des protéines contenues dans la base de données UniProt, un peu moins de la totalité de ces protéines a été obtenu par homologie avec d'autres espèces, comme l'illustre la figure suivante. Ceci témoigne du peu de changement observé dans les protéines du byssus, puisqu'une grande proportion de ces protéines sont retrouvées dans le byssus. Cependant, l'importante proportion de protéines théoriques (34%) pose un défi supplémentaire.

En effet, la présence de protéines prédites (ou théoriques) peut être à l'origine de l'augmentation du taux de faux positifs lors de l'identification d'une protéine (Lubec & Afjehi-Sadat, 2007). Dans le cas actuel, la séquence primaire offre les bases en prévision de l'identification de la protéine, mais devrait être accompagnée par une technique complémentaire afin de supporter les informations obtenues lors du séquençage *de novo* (Lubec & Afjehi-Sadat, 2007). Par exemple, il était fréquent

d'employer la dégradation d'Edman dans le but de confirmer la séquence primaire. De nos jours, l'utilisation d'approches par étiquetage isotopique a largement surclassé la méthode suggérée en 1992 par Johnson (R. S. Johnson & Walsh, 1992). Dans le cas présent, l'utilisation de l'algorithme Paragon permet de concentrer les efforts d'identification sur certaines régions plus prometteuses, ce qui permet un contrôle plus restrictif des faux positifs (Shilov et al., 2007).

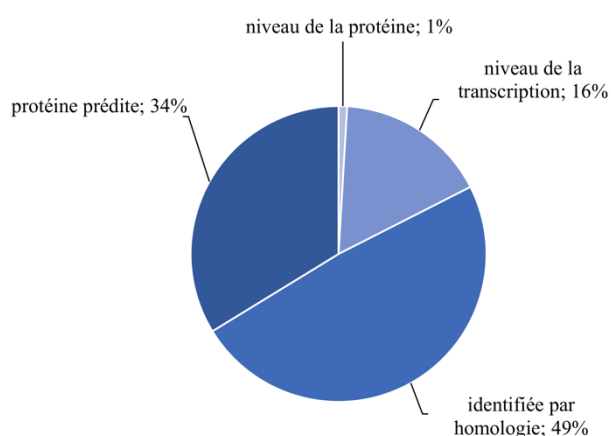


Figure 4.1 Distribution des protéines incluses dans la base de données UniProtKb/Swiss-Prot (74 000 séquences) selon leur catégorie (homologie, transcripte, ou séquence prédite).

Des réplicats biologiques ($n = 3$) pour chaque glande ainsi que pour le manteau de la moule ont été digérés séparément par la trypsine et la pepsine, puis directement analysés par LC-MS/MS, sans étape de pré-fractionnement. Une chromatographie d'environ 60 minutes a été utilisée afin d'améliorer la résolution des peptides. Les données tryptiques et pepsiques ont été combinées lors des recherches pour l'identification des protéines. Chaque glande a été cherchée à partir de la base de données protéomique et transcritomique. Les protéines identifiées (à 1% FDR) avec un minimum de deux peptides à 95% de confiance et présentes dans au moins deux

réplicats ont été conservées. Le Tableau 4.1 ci-dessous et la Figure 4.1 ci-dessus résument les résultats obtenus pour les deux types de base de données.

Tableau 4.1 Comparaison du nombre de protéines identifiées par la base UniProt et transcriptomique pour *Mytilus edulis*

	base de données	centrale	cuticule	plaque	manteau
# peptides (95% conf.)	protéomique	3317	1745	3234	1881
	transcriptomique	439	326	489	6
# protéine (1% FDR)	protéomique	150	102	158	163
	transcriptomique	20	15	29	2
couverture de séquence (moy.)	protéomique	37.3%	34.1%	35.7%	31.6%
	transcriptomique	28.4%	26.7%	28.8%	3.0%

Le tableau 4.1 montre clairement une différence dans le nombre de protéines identifiées par la base de données UniProt versus la base de données préparée selon les informations transcriptomiques. Cet écart s'explique d'une part par la spécificité de la base de données transcriptomiques, qui ne contient que les protéines mfp, PTMP, PTM et pré-collagène trouvées dans le byssus, alors que la base UniProt, quant à elle, inclue plusieurs catégories protéines. D'autre part, les résultats provenant du répertoire UniProt contiennent plusieurs séquences redondantes. Par exemple, les protéines CuZn superoxide dismutase (K4GX76_MYTGA), surperoxide dismutase (K4GY84_MYTGA) et superoxide dismutase [Cu-Zn] (F6KYK2_MYTCH) partagent une très faible homologie lorsque leurs séquences respectives sont alignées. Les informations actuelles ne permettent pas le retrait de l'une ou l'autre de ces protéines et par conséquent, les trois protéines ont été conservées. Bien que le taux moyen de la couverture de la séquence des protéines obtenu à partir de la base transcriptomique demeure inférieur à celui provenant de la base de donnée protéomique, les résultats

obtenus restent intéressants quant à la proportion de protéines identifiées seulement dans le pied.

Une différence marquée entre la base protéomique et transcriptomique a été observée dans le nombre total de protéines identifiées dans le manteau de la moule (tissu contrôle). Puisque la base transcriptomique ne contient que des protéines normalement retrouvées dans le byssus, il n'est toutefois pas surprenant que seulement deux protéines ont été associées au manteau, contrairement à 163 pour la base protéomique. Les six peptides provenant de la base transcriptomique et identifiés dans le manteau sont probablement le fruit d'une contamination lors de la collecte des glandes.

Dans un autre ordre d'idées, les résultats présentés à la Figure 4.2 montrent qu'une excellente précision a été obtenue pour la détection des peptides et pour l'identification des protéines, à l'exception de la glande de la plaque, qui montre une plus grande variabilité. De manière générale, les résultats obtenus à partir de la base de données protéomiques ont montré une bien meilleure précision que pour la base de protéines créée à partir du transcriptome de *M. edulis*. Cet écart s'explique en raison du nombre de protéines (74 377 pour UniProt vs. 629 pour le transcriptome) que contiennent les bases de données respectives. Par conséquent, une plus grande proportion de protéines a été identifiée à partir du répertoire UniProt, ce qui affecte de manière moins significative le coefficient de variation.

Afin de renforcer la confiance dans les protéines identifiées, les réplicats ont été comparés entre eux, pour les deux ensembles provenant de la base Uniprot et de la base transcriptomique, afin de conserver seulement les protéines identifiées dans un minimum de deux réplicats. Une fois les redondances entre espèces retirées pour les échantillons cherchés à partir de la base Uniprot, un total de 146 protéines de pied ont été identifiées, dont 64 de ces protéines ont été attribuées au pied seulement au terme

d'une comparaison avec le manteau (tissu contrôle). Dans le même ordre d'idées, avec l'aide de la base transcriptomique, un total de 44 protéines a été identifié dans le pied de la moule, dont une seule protéine ait aussi été identifiée dans le manteau suite à la comparaison. La Figure 4.3 illustre la répartition des protéines selon chaque glande, pour les bases de données protéomique et transcriptomique employées.

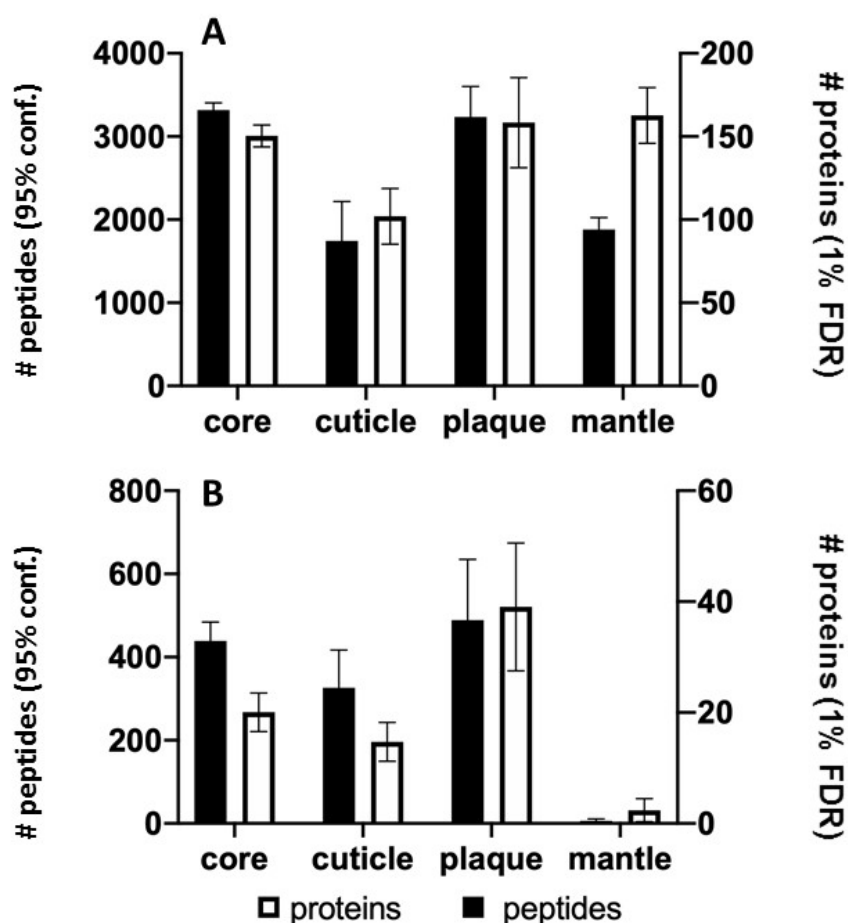


Figure 4.2 Regroupement des peptides (à 95% confiance) et de protéines (à 1% FDR) identifiés pour chaque glande du pied et le manteau de la moule *M. edulis* à partir des séquences obtenues par UniProt (A) et par transcriptomique (B).

Un des avantages de l'utilisation des glandes séparément est l'amélioration du protéome de la moule. Comme l'illustre la Figure 4.3 (A), plusieurs protéines ont été identifiées dans les trois glandes du pied, dont quelques-une associées spécifiquement à la glande centrale, cuticule ou de la plaque. L'extraction des protéines des glandes

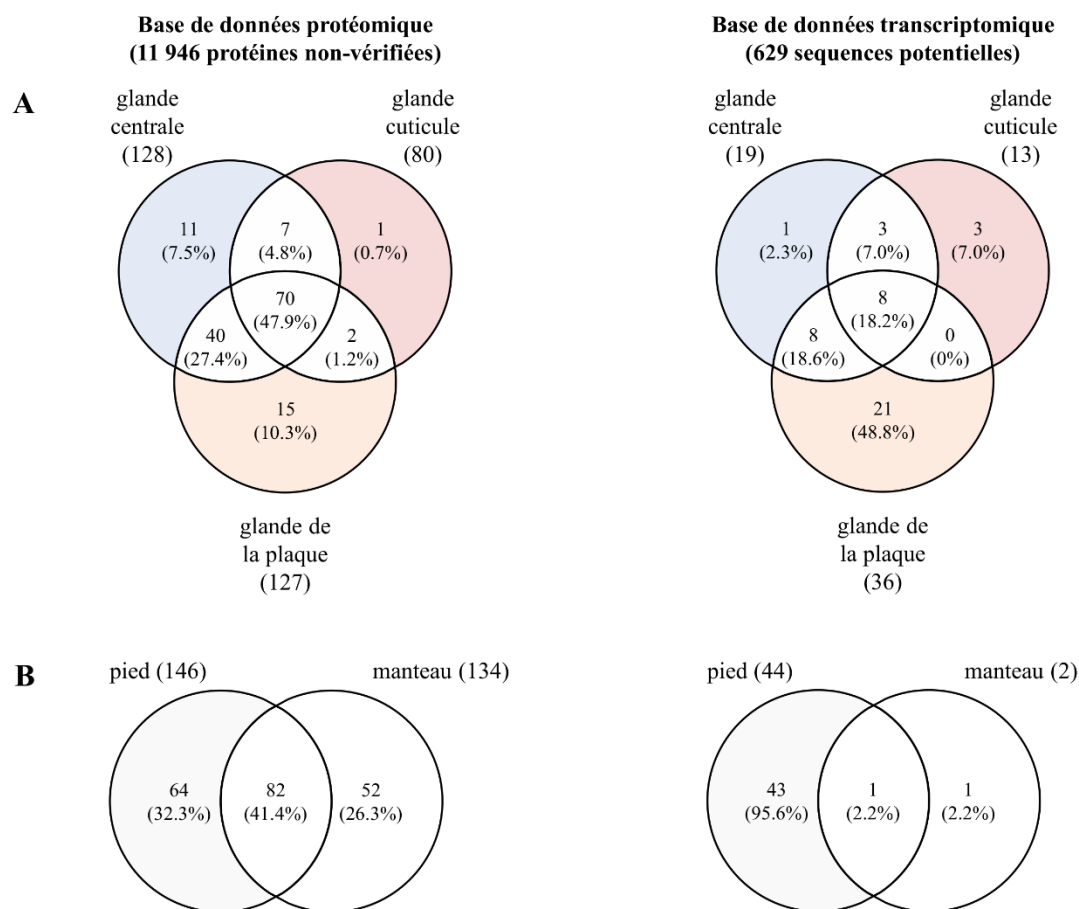


Figure 4.3 Distribution des protéines identifiées pour chaque glande du pied de la moule *M. edulis* (A) pour les séquences obtenues par UniProt (gauche) et par transcriptomique (droite). La figure présente également une comparaison entre les protéines identifiées uniquement dans le pied (B) pour les deux bases de données employées.

centrales et de la plaque semble plus efficace que pour la glande cuticule, dans les conditions expérimentales actuelles. Une explication possible viendrait du fait que la glande centrale possède des vésicules de plus faible diamètre (environ de 0.5 et 1 μm) par rapport aux vésicules des glandes cuticule et de la plaque (Priemel et al., 2017). Les conditions de dénaturation engendrées suite à l'ajout de la solution d'urée et de thiourée n'affecteraient que minimalement la stabilité de ces vésicules et par conséquent, justifierait le nombre inférieur de protéines identifiées.

Dans un autre ordre d'idées, une importante proportion de protéines a été identifiée dans plus d'une glande, avec seulement 11 protéines associées à une seule glande. Il est possible que, lors de la collecte des glandes, les instruments employés aient été contaminés par les protéines les plus abondantes, comme la mfp-2 et les préCols, par exemple, et ce, malgré une procédure minutieuse. D'un autre côté, la littérature mentionne que la sécrétion des protéines impliquées dans l'assemblage du byssus n'emploie qu'une unique glande pour une catégorie de protéines donnée (Priemel et al., 2017; J. H. Waite, 1992). En se basant sur les résultats, il est possible que plus d'une glande soit impliquée dans la production du byssus à différents moments. Par exemple, Priemel et al., ont souligné l'interaction qu'entretiennent les glandes centrales et cuticules durant le processus d'assemblage (Priemel et al., 2017). Toutefois, très peu d'information au sujet de l'implication de plus d'une glande est actuellement disponible.

L'utilisation de la base transcriptomique s'est montrée un atout particulièrement utile pour l'identification de protéines connues comme étant des précurseurs du byssus. Parmi les 19 mfps inclus dans la base Uniprot, seulement les mfp-1 à 3 sont associés à *M. edulis*. Les 16 autres protéines présentes dans la base UniProt proviennent d'études transcriptomiques réalisées sur *M. californianus* et *M. coruscus*. Également, certaines protéines comme les mfp-4 et 6 n'ont pas été identifiées lorsque la base de données protéomiques a été utilisée. À l'aide de la base de données

transcriptomiques, les mfp-4, 6, 10 et 12, ainsi que plusieurs variants potentiels ont été identifiés dans *M. edulis*. Une importante homologie entre les moules *M. edulis* et *M. californianus* peut être déduite pour la protéine mfp-10. En effet, 15 mfp-10 ont été identifiés dans *M. edulis* à partir de leur séquence respective, obtenue d'une analyse transcriptomique. À l'exception de deux protéines, les 13 autres candidates ont toutes été identifiées comme provenant de la glande de la plaque. Également, plusieurs variantes de la protéine PTMP-1 ont été identifiées, essentiellement dans la glande centrale, mais aussi dans les glandes cuticule et de la plaque. Cette classe de protéines est généralement retrouvée dans la glande centrale, aux côtés des pré-collagènes. Le fait que ces PTMP soient identifiées avec un nombre considérable de peptides dans deux glandes soutient l'hypothèse d'une contamination lors de la dissection des glandes, mais également la possibilité d'une interaction entre plusieurs glandes.

L'extraction des protéines influence la décision de considérer, ou non, une protéine comme étant spécifique au pied. Par exemple, lorsque le pied de la moule a directement été homogénéisé, le couple de protéines twitchine et catchine a été identifié à la fois dans le pied et dans le manteau de la moule. Ces deux protéines ont été jugées intéressantes en raison du potentiel rôle qu'elles pourraient jouer dans la byssogénèse. Chez les invertébrés, la protéine catchine est généralement retrouvée dans les tissus musculaires et permet de maintenir une tension tout en minimisant la demande en énergie (Yamada, Yoshio, Oiwa, & Nyitray, 2000). D'un autre côté, la protéine twitchine participerait à la contraction des tissus musculaires des mollusques lorsque les niveaux de calcium ne suffirait plus à maintenir ce tissu sous tension. Malgré le fait que ce couple de protéines ait été retrouvé dans un tissu non-musculaire, et sachant que le pied de la moule doit se contracter afin de permettre la sécrétion du byssus, le manque d'information quantitative pouvant justifier la présence de ces protéines dans le manteau a miné l'implication de ces protéines dans le processus d'assemblage et par conséquent, le couple twitchine et catchine ne peut donc pas être directement associé à la fabrication de la fibre. En utilisant les glandes du pied

Tableau 4.2 Protéines du byssus identifiées à partir de la base UniProt et transcriptomique

Protéomique		# peptide (% couverture de la séquence)		
# d'accession	Nom de la protéine	centrale	cuticule	plaque
A0A193DU95_MYTCO	Byssal calumenin-like protein 1	22 (49.4%)	8 (26.1%)	20 (44.1%)
A0A193DU99_MYTCO	Byssal protease inhibitor-like protein 1	5 (3.5%)	3 (2.5%)	6 (5.3%)
A0A0G2YLH0_MYTCO	Byssal tyrosinase-like protein 1	4 (5.4%)		3 (4.4%)
A0A193DUA5_MYTCO	Byssal tyrosinase-like protein 2	6 (10%)	8 (14.9%)	
A0A0K2D7V5_MYTCO	Collagen-like protein 5	18 (12%)	20 (12.3%)	
A0A0K2D7R0_MYTCO	Collagen-like protein 6	7 (10.9%)		6 (9.1%)
A0A0K2D820_MYTCO	Collagen-like protein 7	4 (11.7%)		6 (18.3%)
Q1XBT8_MYTED	Foot protein 2	9 (22.2%)		14 (29.1%)
A0A223HCJ8_MYTCA	Foot protein 10	2 (5.9%)		6 (10.2%)
A0A223HCL5_MYTCA	Foot protein 12	3 (5.3%)		17 (14.6%)
O76271_MYTED	PreCol-NG	81 (49.8%)	89 (48.9%)	31 (29.7%)
O44367_MYTED	PreCol-D	79 (45.7%)	102 (49.3%)	27 (28.7%)
O16161_MYTED	PreCol-P	16 (26.5%)	22 (27.9%)	
Q8T5C3_MYTED	Proximal thread matrix protein 1b	34 (59.6%)		
Transcriptomique*		# peptide (% couverture de la séquence)		
# d'accession	Nom de la protéine	centrale	cuticule	plaque
MYTCA_R7292100	Foot protein 2	10 (26.3%)		18 (40.4%)
MYTCA_R7292219	Foot protein 4 (variant 1)			23 (28.7%)
MYTCA_R7299012	Foot protein 6 (variant 2)		3 (10.2%)	
MYTCA_R7292290	Foot protein 10	8 (41.1%)		21 (67.5%)
MYTCA_R7293819	Foot protein 12			7 (31.9%)
MYTGA_R7292104	PreCol-D	56 (66.3%)	62 (59.2%)	17 (42.3%)
MYTGA_R7292068	PreCol-P	10 (22.3%)	14 (21.6%)	
MYTGA_R7292044	PreCol-NG	65 (41.4%)	64 (41.2%)	24 (27.1%)
MYTGA_R7304154	Proximal thread matrix protein 1	54 (22.0%)	20 (9.2%)	77 (29.6%)

responsables de la fabrication de la fibre de byssus et en employant la base de données UniProt, plusieurs protéines, dont le couple twitchine et catchine, ont été identifiées sans ambiguïté dans le pied de la moule. Les protéines de la famille twitchine (twitchin, twitchin-like protein 1 et twitchin-like protein 2) ont été identifiées dans les glandes centrale et de la plaque, avec quelques peptides identifiés dans la glande cuticule pour la protéine twitchine. En ce qui concerne la protéine catchine, cette dernière a été identifiée dans les trois glandes.

D'un autre côté, les tyrosinases sont une classe de protéines responsables de l'oxydation des résidus tyrosines en DOPA. Les résultats présentés au chapitre trois faisaient mention de deux tyrosinases byssales qui n'ont été identifiées qu'à partir des échantillons de byssus. Maintenant, ces protéines ont été identifiées dans le pied, avec la tyrosinase-homologue 1 byssale associée aux glandes centrale et de la plaque. La tyrosinase homologue 2 byssale, quant à elle, a été identifiée dans la glande cuticule et également dans la glande centrale. Le fait que ces deux tyrosinases aient été identifiées dans deux régions distinctes du byssus soutient le fait que l'oxydation du résidu tyrosine remplit des fonctions différentes selon la région de la fibre dans laquelle elle se trouve.

Étonnamment, aucune des protéines riches en DOPA, comme les protéines mfp-5, mfp-3 et mfp-1, n'ont été identifiées à partir du répertoire UniProt ou de la base de données transcriptomiques. Pourtant, ces protéines sont incluses dans ces répertoires, avec plusieurs correspondances provenant de différentes espèces dans le cas d'UniProt. Ces protéines bénéficient d'un intérêt particulier en raison de leur importante concentration en DOPA, mais aussi en raison de leur implication dans l'adhésion et la protection de la fibre de byssus. D'autre part, la DOPA a la capacité de former des liaisons croisées avec diverses chaînes latérales nucléophiles (Miserez, Schneberg, Sun, Zok, & Waite, 2008; Zhao & Waite, 2005, 2006b). La formation de ces liaisons croisées est depuis considérée comme étant potentiellement responsable du «vieillessement» du byssus. Également, la possibilité de former des liaisons croisées de

type diDOPA a aussi été une hypothèse envisagée par Burzio et Waite au début des années 2000. Néanmoins, la cinétique menant à la formation des liaisons croisées dépend des résidus situés autour de la DOPA (Jee, Park, & Kim, 2000). L'une des principales limites lors d'analyses de peptides possédant des liaisons croisées provient de la myriade de réactions possibles lors de la fragmentation de ces peptides. Ces nombreuses possibilités complexifient le spectre MS/MS en affectant par la même occasion l'estimation du taux FDR (Luo, Bassett, & Ranish, 2019).

4.5 Conclusions

Les moules ont la capacité de produire une fibre extracellulaire dont la composition des protéines est étudiée depuis plusieurs décennies. Néanmoins, de nombreuses protéines restent encore à découvrir, en plus de clarifier les rôles biologiques de celles actuellement connues. L'utilisation d'une base de données composée de protéines obtenues à partir du transcriptome de la moule *Mytilus edulis* a été introduite comme une excellente alternative afin de palier aux défis apportés par une base de données protéomique incomplète. Avec cette approche, environ 190 protéines de pieds de la moule, provenant des données protéomiques et transcriptomiques combinées ont été identifiées, dont une centaine de protéines associées uniquement à l'organe producteur du byssus. Plusieurs variantes d'une même protéine ont potentiellement été identifiées, mais demeurent à être confirmées.

La dissection des glandes a été aussi un avantage dans l'étude du protéome de *M. edulis*. Alors qu'il est maintenant possible d'associer une glande avec la protéine qu'elle produit, cette approche a été plus avantageuse pour identifier les protéines présentes dans le pied seulement. Par conséquent, quelques protéines intéressantes qui ont été écartées sont maintenant associées au pied de la moule. Cependant, la procédure de dissection reste sensible à la contamination lors de la collecte des glandes et de l'extraction des protéines. Étonnamment, la protéine mfp-1 n'a été identifiée dans

aucune des bases de données employées. Cette protéine est connue pour comporter plusieurs modifications d'hydroxylations sur les résidus tyrosines et prolines. Les approches expérimentales prochaines devront porter une attention particulière lors de l'étape d'extraction dans le but de permettre l'extraction de cette protéine à partir de la glande cuticle, là où sa concentration est présumée la plus abondante.

CHAPITRE V

CARACTÉRISATION PAR PROTÉOMIQUE ASCENDANTE DE LA PROTÉINE RECOMBINANTE MFP-1 DE LA MOULE *MYTILUS EDULIS*

La protéine mfp-1 joue un rôle particulièrement important dans la protection de la fibre de byssus, en recouvrant complètement la plaque et le cuticule de la fibre. La présence relativement importante de DOPA offre à la fibre la résistance mécanique nécessaire via l'implication de la DOPA dans des liaisons de coordination avec différents métaux. Étonnamment, cette protéine n'a été détectée dans aucune des approches méthodologiques précédentes. Ce chapitre abordera les modifications apportées à la méthodologie détaillée aux chapitres 3 et 4 dans le but d'identifier la protéine mfp-1 provenant du byssus de la moule *Mytilus edulis*.

Maxime Sansoucy a réalisé les recherches bibliographiques, la préparation des protocoles expérimentaux, les expériences ainsi que le traitement des données. Mihai Mesko, candidat au doctorat à l'Université McGill, a synthétisé et purifié les versions de la mfp-1 avec et sans DOPA. La professeure Lekha Sleno a soutenu et supervisé le déroulement du projet, la vérification des données et l'interprétation des résultats. Les résultats présentés dans ce chapitre feront partie d'une publication soumise à une date ultérieure.

5.1 Résumé

La protéine mfp-1 demeure l'une des protéines riche en DOPA les plus importantes en raison de son implication dans la protection du byssus. En effet, elle participe, via l'entremise de liaisons de coordination, au maintien de l'intégrité structurale de la fibre. Cependant, la présence de DOPA implique la formation potentielle de liaisons croisées, ce qui complique l'efficacité de l'extraction de la protéine.

Le présent chapitre montre les premières tentatives visant à l'élaboration des conditions expérimentales optimales pour l'analyse de la protéine mfp-1 à partir d'une version recombinante. Puisque la DOPA est une modification susceptible aux conditions de pH, différents agents anti-oxydants ont été testés afin de favoriser la forme réduite de la DOPA. L'identification de la présence de liaisons croisées (intra ou inter-moléculaires) impliquant la forme réactive *ortho*-quinone et certaines chaînes latérales nucléophiles demeure un enjeu particulièrement de taille en spectrométrie de masse. Pour la première fois, l'existence d'un lien croisé au sein du décapeptide composant essentiellement la mfp-1 a été rapportée par LC-MS/MS, avec plusieurs ions diagnostiques.

5.2 Introduction

Le byssus, tel que décrit au chapitre 1, est un matériau biologique fabriqué par les moules afin d'assurer leur attachement sur virtuellement n'importe quel type de support solide. La DOPA, une modification post-traductionnelle rare, mais cruciale pour le byssus, permet de préserver les propriétés adhésives et bio-mécaniques de la fibre. Cependant, des conditions défavorables comme un pH légèrement moins acide ou la présence d'agents oxydants, favorise la conversion de la DOPA en dopaquinone.

Ce changement a été mentionné à quelques reprises comme étant potentiellement néfaste pour l'attachement du byssus (Lin et al., 2007; Yu et al., 2011, 2013; Zhao & Waite, 2006a), mais aussi pour l'intégrité structurale de la fibre. En 2006, Zhao et Waite ont observé que les résidus DOPA des protéines mfp-3 et mfp-5, qui participent différemment au processus d'adhésion, étaient plus susceptibles à l'auto-oxydation (Zhao & Waite, 2006a). Les auteurs ont estimé le nombre de résidus participant aux liaisons croisées avec les cystéines de la mfp-6 à environ de 5 à 10 résidus DOPA/1000. Cependant, peu d'informations est actuellement disponible quant à l'impact négatif de la présence de liaisons covalentes entre plusieurs protéines (B. P. Lee, Messersmith, Israelachvili, & Waite, 2011).

Dans un autre ordre d'idées, l'obtention des mfps directement à partir du byssus ou de l'organe producteur (le pied) demeure une tâche considérable. La plupart des méthodes sont laborieuses et nécessitent une succession de traitement acide, d'agent dénaturant et de purification afin d'obtenir un degré de pureté acceptable (Byette, Pellerin, & Marcotte, 2014; S. Li et al., 2018). Puisque les séquences des mfp sont bien connues (Deming, 1999; B. P. Lee et al., 2011), il est par conséquent possible d'obtenir des versions recombinantes de ces protéines ou d'une section en particulier (Burzio & Waite, 2000; Priemel et al., 2020).

La protéine recombinante mfp-1 (rMFP-1) représente une protéine intéressante pour étudier les potentielles réactions entre les groupements dopaquinones avec les chaînes latérales nucléophiles. Plus précisément, la rMFP-1 contient un motif de 10 acides aminés qui permet une approche facile par protéomique *bottom-up*. En effet, le décapeptide en question (AKPSYPPTYK) est un peptide tryptique et comporte deux tyrosines en position 5 et 9 pouvant être oxydées en DOPA. L'un des premiers enjeux de ce chapitre adressera l'optimisation des conditions expérimentales menant à l'identification de ce peptide. L'utilisation du séquençage *de novo* permettra la localisation de la DOPA sur le peptide.

5.3 Expérimentale

5.3.1 Matériel

La trypsine (bovine), la pronase, les solvants ainsi que tous autres produits chimiques ont été obtenus de Sigma-Aldrich (St. Louis, MO). L'eau ultrapure a été obtenue à partir d'un système Millipore Synergy UV (Billerica, MA).

5.3.2 Expression et oxydation de la protéine rMFP-1

L'expression de la protéine recombinante mfp-1 de la moule *Mytilus edulis* a été réalisée en collaboration avec l'équipe de recherche du Prof. Matt Harrington de l'Université McGill. Brièvement, des cellules DH5 α (100 μ l) propices à l'insertion d'un brin d'ADN extracellulaire ont été décongelées sur glace pendant 30 minutes. Par la suite, une solution de plasmide (10 μ l, 10 mg/ml) a été ajoutée à la solution, pour ensuite être conservée pour une durée supplémentaire de 30 minutes sur glace. Les cellules ont ensuite été mises dans un bain d'eau maintenu à 42°C, pendant 30 secondes, afin de permettre la prise des plasmides, puis immédiatement remises sur la glace pendant 2 minutes. Le milieu liquide Luna broth (1 ml) a été ajouté aux échantillons avant d'être incubé (1h, 37°C, 160 rpm). Un gel LB-Agar dopé avec 50 μ g/ml de sulfate de kanamycine a été inoculé avec un aliquot de 100 μ l de la solution finale. Une fois les plaques séchées, ces dernières ont été placées dans un incubateur, en position inversée (18h, 37°C).

Afin d'obtenir la DOPA sur la protéine recombinante mfp-1 qui n'en contient pas pour le moment, l'oxydation de la mfp-1 a été réalisée *in vitro* à l'aide de la tyrosinase provenant de champignons. La réaction s'est effectuée dans une solution tampon phosphate (100 mM) à laquelle de l'acide citrique (25 mM) et du borate (20 mM) ont été ajoutés afin d'obtenir un pH de 6.8. Une solution de mfp-1 (10 mg/ml) a été mélangée en quantité similaire à la solution de tyrosinase (125 unités/ μ mol

tyrosinase) et incubée durant 24 h. De l'acide acétique glacial a été ajouté au cinquième du volume d'incubation une fois l'incubation terminée. La solution a été centrifugée (12 000 x g, 10 min) en prévision d'une séparation par chromatographie. Le surnageant a ensuite été séparé par chromatographie rapide de protéine en phase liquide (*Fast Protein Liquid Chromatography*, FPLC, BioRad), puis purifié par filtration sur gel (Hiload 16/600 Superdex 200 pg, GE Lige Sciences) afin de retirer la tyrosinase du mélange. La séparation par FPLC s'est effectuée en mode isocratique, en employant une phase mobile contenant 150 mM NaCl acidifiée avec 1% d'acide acétique. La filtration a été suivie par absorbance et les fractions correspondantes à la protéine rMFP-1 ont été combinées, séparées par SDS-PAGE et analysées par spectrométrie de masse afin de confirmer la présence de la DOPA. Le pourcentage en DOPA (mol%) a été calculé à partir de l'intensité des pics obtenus par MS pour les versions oxydée et non-oxydée de la protéine rMFP-1. Les fractions restantes ont été lyophilisées, puis conservées à -20°C.

5.3.3 Digestion de la rMFP-1 par la trypsine

Une quantité de 10 µg de la protéine rMFP-1 a été mélangée avec une solution tampon de bicarbonate d'ammonium (100 mM, pH 8.5). De l'acide ascorbique a ensuite été ajouté au mélange (25 µl, 10 mM) en prévision de la digestion tryptique. Puisque l'échantillon ne comportait qu'une seule protéine, l'étape de réduction alkylatrice a été volontairement omise et la digestion s'est déroulée durant 1h à 37°C avec de la trypsine (2 µg). Les échantillons ont ensuite été transférés dans des vials en préparation pour l'analyse LC-MS/MS.

5.3.4 Hydrolyse complète de la protéine rMFP-1

La digestion complète de la mfp-1 a été réalisée à l'aide de la pronase tel que décrit préalablement par Mitrea et al. (Mitrea, Leblanc, St-Onge, & Sleno, 2010). Brièvement, environ 10 µg de protéine recombinante mfp-1, dont l'un des résidus

tyrosine a été préalablement fonctionnalisé en DOPA, ont été solubilisées dans 200 µl d'une solution tampon Tris (100 mM, pH 9) contenant du glutathion. Les échantillons ont été incubés pour une durée de 2h à 37°C, avant d'être acidifiés avec une solution d'ascorbate (1.3 mM finale). La protéine a été diluée à un volume de 225 µl avec la même solution de Tris mentionnée plus haut, avant l'ajout de la pronase (25 µl, 1 mg/mL). La digestion s'est déroulée sur une durée allant de 0-24h à 50°C, puis elle a été directement injectée sur une colonne de type PFP (voir section 5.4.1) pour les analyses par spectrométrie de masse.

À des fins de comparaison, le même échantillon (aliquot de 100 µl) a été sélectionné pour une digestion complète utilisant une solution 6 M HCl. Sommairement, la même quantité de la protéine mfp-1 (10 µg) a été pré-incubée de la même manière que détaillé au paragraphe précédent. Une fois la pré-incubation terminée, les échantillons ont été dilués dans une solution de 6 M HCl (250 µl) et digérés durant 24h à 120°C. Les tubes ont été ouverts pour permettre l'évaporation de l'acide durant la dernière heure puis, une fois la digestion complétée, 100 µl d'une solution 5% d'hydroxide d'ammonium ont été ajoutés pour neutraliser l'acide. Les échantillons ont ensuite été évaporés à sec (sous N₂) avant d'être resolubilisés dans la solution de 100 mM Tris (contenant 50 mM en chlorure de calcium). Les échantillons ont ensuite été injectés sur la même colonne, en employant la même méthode LC-MS/MS que celle décrite à la section 6.3.4.

5.3.5 Analyse RP-LC-MS/MS

Les échantillons ont été injectés (10 µl) séparément sur une colonne Aeris PEPTIDE XB-C18 (100 x 2.1 mm), une colonne Biphenyl (100 x 2.1 mm, 2.6 µm, 100 Å, Phenomenex) ainsi que sur une colonne Scherzo SM-C18 (150 x 3 mm, 3 µm). Les colonnes ont été installées sur un système Nexera UHPLC (Shimadzu, Columbia, MD, USA), opéré à un débit de 300 µl/min (25, 40 et 50°C) et comme phases mobiles de

l'eau (A) et de l'ACN (B), toutes deux acidifiées avec de l'acide formique (0.1%). Pour le gradient de 36 minutes, celui-ci débute à 5%B durant les 2.5 premières minutes, augmente ensuite à 30% à 24 minutes, à 50% à 26 minutes, ensuite à 85% à 26.5 minutes et est maintenu durant 2 minutes. Pour le gradient de 12 minutes, ce dernier débute à une concentration de phase mobile B à 5% durant les 2.5 premières minutes, pour atteindre 50% à 10 min, ensuite rapidement augmenter à 85% en 0.5 minute pour finalement être maintenu à 85% durant 0.5 minute.

Les spectres MS ont été enregistrés sur un système hybride quadripôle temps-de-vol à haute résolution TripleTOF 5600 (Sciex, Concord, ON, Canada) équipé d'une source d'ionisation DuoSpray (opéré en mode positif). Les réglages de la source sont demeurés les mêmes pour les méthodes de 36 et 12 minutes et ont été fixés à un voltage de 5.5 kV et une température de 450°C, avec un débit de gaz de nébulisation et de séchage (GS1/GS2) de 50 psi, avec finalement un potentiel de déclusterisation (DP) de 60 V. Pour la chromatographie de 36 minutes, un balayage a été réalisé dans une fenêtre de m/z 140 - 1250 (250 ms de temps d'accumulation) pour les spectres MS, suivit d'un balage allant de m/z 80 – 1250, pour l'obtention de spectres MS/MS. Pour ce qui est de la méthode de 12 minutes, le balayage pour l'obtention des spectres MS s'est fait sur une intervalle de m/z 120 – 1250 (250 ms de temps d'accumulation) et allant de m/z 80 – 1500 pour l'obtention des spectres d'ions produits MS/MS. Les critères de sélection IDA sont demeurés les mêmes pour les deux méthodes, soit la sélection des cinq ions les plus intenses avec une restriction d'exclusion de l'ion cible, appliqué durant 20 s après 2 occurrences. De plus, une soustraction dynamique du bruit de fond pour l'ensemble des expériences DDA a également été appliquée. Chaque spectre MS/MS a été obtenu avec un temps d'accumulation de 50 et 150 ms (pour les méthodes de 36 et 12 minutes respectivement) et avec une énergie de collision de 30 ± 10 V. Le temps de cycle total était de 1.1 s. La précision sur la masse a été suivie à l'aide d'une solution de calibration injectée à chaque quatre échantillons.

5.4 Résultats et discussion

La mfp-1 est une protéine qui se caractérise par la répétition du décapeptide AKPSYPPTYK. Plusieurs modifications sur les prolines (hydroxyproline et dihydroxyproline) et les tyrosines (DOPA) ont été identifiées sur le peptide (Taylor, Waite, Ross, Shabanowitz, & Hunt, 1994). Cette protéine bénéficie d'une attention particulière en raison de son rôle dans la protection de la fibre de byssus.

5.4.1 Optimisation de la chromatographie

La présence de plusieurs prolines et de tyrosines rend le peptide relativement polaire. Aussi, les trois prolines présentes dans la séquence altèrent la séparation en causant un dédoublement du pic chromatographique. Cette tendance est la conséquence de l'isomérisation cis/trans du lien peptidique lorsqu'un résidu proline est présent dans la séquence primaire. Cette interconversion a également été observée dans d'autres peptides issus de la digestion de protéines riches en proline, comme l'élastine (Pepe et al., 2013), ou lors d'analyse de certains peptides aux propriétés thérapeutiques (Guttman et al., 2020). Dans l'optique d'améliorer l'apparence du pic, différentes colonnes et conditions chromatographique ont été testées (Figure 5.1).

Bien que ce peptide comporte 10 acides aminées, plusieurs de ceux-ci portent un groupement hydroxyle qui le rend relativement polaire. Par conséquent, le peptide a tendance à éluer dans les 10 premières minutes (d'un gradient de 60 minutes) avec une base s'étalant sur plus de deux minutes (Figure 5.1, colonne Aeris). Lorsque les échantillons ont été de nouveau injectés sur la colonne biphényle, l'élargissement du pic a considérablement été diminué par rapport à la colonne Aeris. L'utilisation de la colonne biphényle favorise les interactions π - π entre les chaînes latérales de la tyrosine et la phase stationnaire, permettant une meilleure rétention. De plus, une chromatographie plus rapide (12 minutes) a été mise en place. Aussi, l'augmentation

de la température a légèrement permis d'améliorer l'apparence du pic lorsque la colonne était maintenue à 50°C.

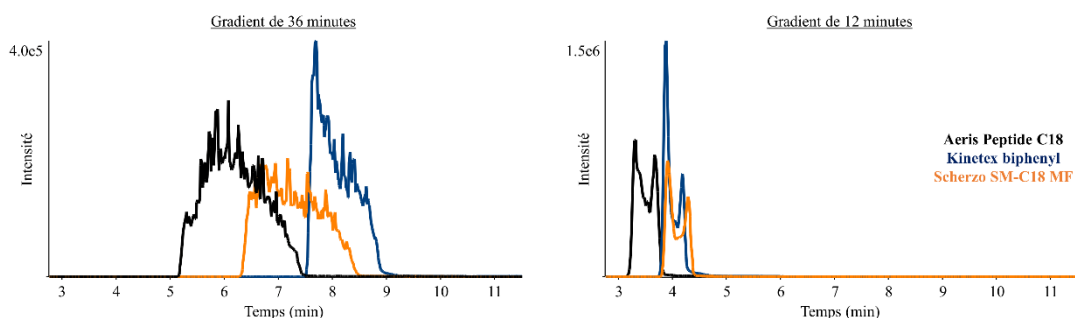


Figure 5.1 Comparaison de l'allure du pic associé au peptide AKPSYPPTYK lorsque séparé sur des colonnes C12 et biphenyle à l'aide d'un gradient de 36 min (à gauche) et de 12 min (droite).

5.4.2 Optimisation des conditions de digestion

Sachant que la modification DOPA est relativement sensible à l'oxydation, différentes conditions ont été testées afin de préserver la forme réduite. Dans un premier temps, ces tests ont été réalisés avec la forme non-modifiée de la rMFP-1 afin de mesurer l'impact de chacun de ces composés sur l'activité de la trypsine. Des essais antérieurs ont montré qu'une concentration finale de 0.4 mM en BHT, borate et ascorbate offrait des résultats favorables. Par conséquent, la comparaison de chaque composé anti-oxydant a été réalisée à cette concentration. Finalement, les échantillons ont été digérés dans des conditions similaires (2 µg trypsine, 1h à 37°C) en prévision de l'analyse LC-MS/MS.

Bien que la rMFP-1 non-modifiée ne possède pas de DOPA pouvant être protégée, le chromatogramme d'ions extrait y étant associé (Figure 5.2A) montre une amélioration suite à l'ajout d'acide ascorbique. L'ajout de borate et de BHT, quant à eux, ont généré un signal comparable. Des changements plus marqués ont été observés

lorsque le peptide possède un ou deux DOPA sur la séquence, tel que montré par les Figure 5.2B et C respectivement. Un signal maximal est obtenu suivant l'ajout d'acide ascorbique, alors que les signaux les plus faibles ont été mesurés lorsque la protéine a été incubée sans agent anti-oxydant ou avec du BHT. De plus, pour le peptide ne possédant qu'une DOPA, l'ajout du borate montre un effet intéressant (Figure 5.2B, EIC vert). En effet, le borate a la capacité de former un complexe avec des diols vicinaux et ainsi stabiliser la forme réduite de la DOPA. Néanmoins, la présence de borate augmente légèrement le pH de la solution à des valeurs entre 8.5-9.5, ce qui favorise la conversion de la DOPA en dopaquinone.

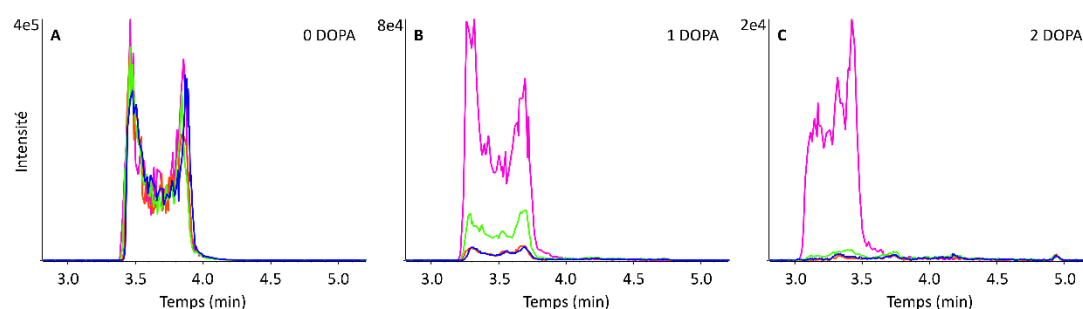


Figure 5.2 Chromatogrammes d'ions extraits du peptide tryptique AKPSYPPTYK incubé en présence de différents composés anti-oxydants, pour les versions du peptide sans oxydation (A, m/z 576.31), avec l'addition d'un oxygène (B, m/z 584.30) et deux oxygènes (C, m/z 592.30).

Des essais supplémentaires ont été réalisés afin d'établir la concentration finale d'ascorbate optimale. Dans un premier temps, l'augmentation de la concentration finale d'ascorbate ne semble pas affecter l'activité de la trypsine. En effet, comme le montre la Figure 5.3, pour les digestions sans ascorbate, l'ascorbate a une concentration de 0.4 mM, une concentration de 1 mM et une concentration en ascorbate de 2 mM. La présence ou non d'acide ascorbique ne semble pas affecter la sensibilité du peptide lorsque celui-ci ne porte aucune modification. Cependant, lorsque la DOPA est

présente (Figure 5.3B et C), l'ajout de l'acide ascorbique est clairement bénéfique afin d'obtenir une sensibilité supérieure. Par conséquent, les incubations futures ont été réalisées avec une concentration finale de 400 μ M d'acide ascorbique.

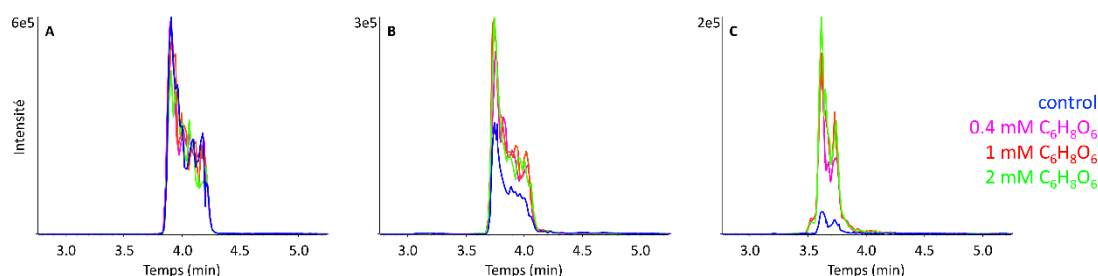


Figure 5.3 EIC montrant l'impact de trois concentrations d'acide ascorbique sur la sensibilité du peptide AKPSYPPTYK sans la présence de DOPA (A), avec une DOPA (B) et deux DOPA (C) ajoutée en prévision de la digestion par la trypsine.

Parallèlement, plusieurs résidus incluant les deux tyrosines peuvent être oxydés. Par conséquent, les XIC obtenus à partir des masses exactes des peptides correspondant aux différentes possibilités d'oxydation ont été superposés afin d'identifier l'état d'oxydation prédominant (Figure 5.4). Aussi, l'exactitude des masses monoisotopiques mesurées pour l'ensemble des possibilités d'oxydation sont présentées à partir des informations recueillies dans le Tableau 5.1 ci-dessous. Par ailleurs, il a été remarqué que le décapeptide issu de la digestion de la protéine mfp-1 purifiée à partir du pied de la moule *M. edulis* peut inclure plusieurs résidus oxydés, dont les tyrosines et les prolines, avec la possibilité d'une di-hydroxyproline (Taylor et al., 1994). Ces états d'oxydation ont également été observés par Burzio et Waite en 2000 lorsque le décapeptide a été oxydé en présence de tyrosinase et de périodate (Burzio & Waite, 2000).

Comme le montre la figure ci-dessous, la forme prédominante d'oxydation du peptide implique la présence d'une seule DOPA, avec une légère proportion associée à une DOPA supplémentaire. En accord avec l'erreur sur la masse (Tableau 5.1), aucune autre oxydation (hydroxyproline, di-hydroxyproline et TOPA) n'a été identifiée.

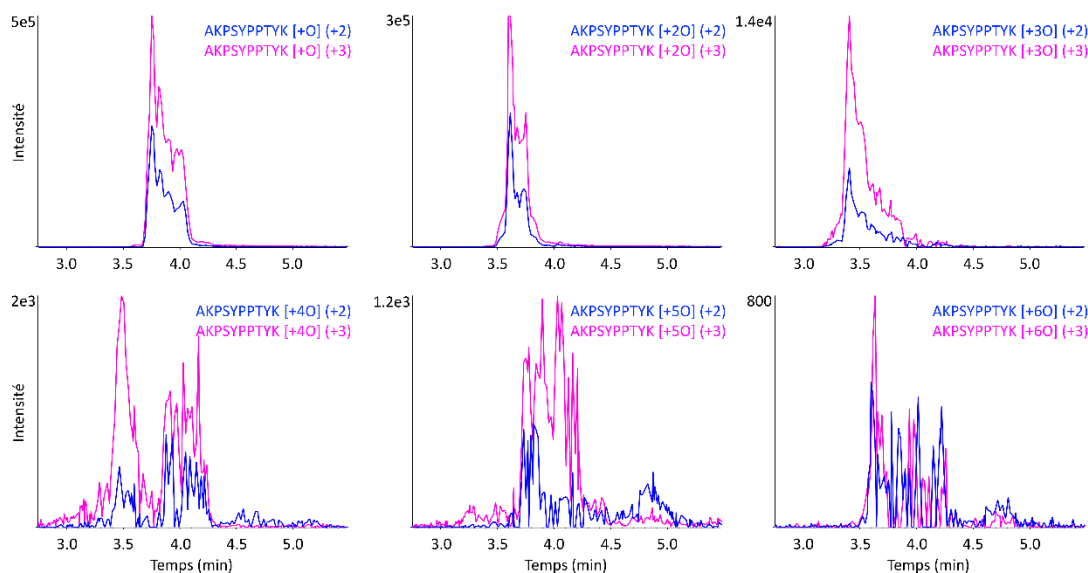


Figure 5.4 Comparaison des différentes possibilités d'oxydation sur le décapeptide AKPSYPPTYK pour les états de charge +2 (bleu) et +3 (rose). Les résidus tyrosine notés d'un astérisque (Y*) font référence au site d'oxydation associé à la présence de la DOPA.

En comparaison, la présence de deux DOPA sur le peptide résultait en un signal moins sensible que lorsqu'une seule DOPA était présente. Cette observation porte à croire que la présence d'une DOPA supplémentaire est la conséquence d'une oxydation excessive par la tyrosinase ou d'un stress oxydatif. Par conséquent, seules les versions non-modifiées, [+O] et [+2O] ont été considérées pour les expériences subséquentes.

Tableau 5.1 Masses monoisotopiques associées à divers états d'oxydation du peptides AKPSYPPTYK pour deux états de charge.

modification	séquence ^a	Formule	z	m/z observé	ppm
aucune	AKPSYPPTYK	C ₅₅ H ₈₂ N ₁₂ O ₁₅	2	576.3087	0.6
			3	384.5417	0.8
[+O]	AKPSY <u>P</u> PTY*K	C ₅₅ H ₈₂ N ₁₂ O ₁₆	2	584.3063	0.7
			3	389.8733	0.7
[+2O]	AKPSY*PPTY*K	C ₅₅ H ₈₂ N ₁₂ O ₁₇	2	592.3038	0.8
			3	395.2052	1.4
[+3O]	AKPSY* <u>P</u> PTY**K	C ₅₅ H ₈₂ N ₁₂ O ₁₈	2	600.3014	2.7
			3	400.5370	1.8
[+4O]	AKPSY**PPTY**K	C ₅₅ H ₈₂ N ₁₂ O ₁₉	2	608.3277	48.4
			3	405.8672	-1.8
[+5O]	AKPSY** <u>P</u> *PTY**K	C ₅₅ H ₈₂ N ₁₂ O ₂₀	2	616.2641	-50.0
			3	411.1928	-16.5
[+6O]	AKPSY**P** <u>P</u> TY**K	C ₅₅ H ₈₂ N ₁₂ O ₂₁	2	624.2800	-21.0
			3	421.8238	-92.6

Y : tyrosine; Y* : DOPA; Y** : TOPA; P : proline; P* : hydroxyproline; P** : dihydroxyproline

^aLes résidus surlignés représentent les différentes combinaisons d'oxydation.

L'une des interrogations se concentrait sur la localisation de la modification DOPA. Dans leurs travaux, Taylor et al. ont proposé que les deux tyrosines aient été converties en DOPA (Taylor et al., 1994). Puisque les deux tyrosines sont susceptibles d'être modifiées, la fragmentation du peptide a été induite de façon théorique à l'aide d'un outil disponible en ligne (<http://prospector.ucsf.edu/prospector/mshome.htm>) afin de connaître l'emplacement de la modification. À partir de spectres MS/MS, trois ions ont été identifiés avec la tyrosine du côté C-terminal (Tyr-9) modifiée (noté en rouge dans le spectre de la Figure 5.5A). Parmi ces ions, deux d'entre eux, à *m/z* 360.1541 et *m/z* 475.2169, ont confirmé la présence de la DOPA du côté C-terminal puisqu'ils

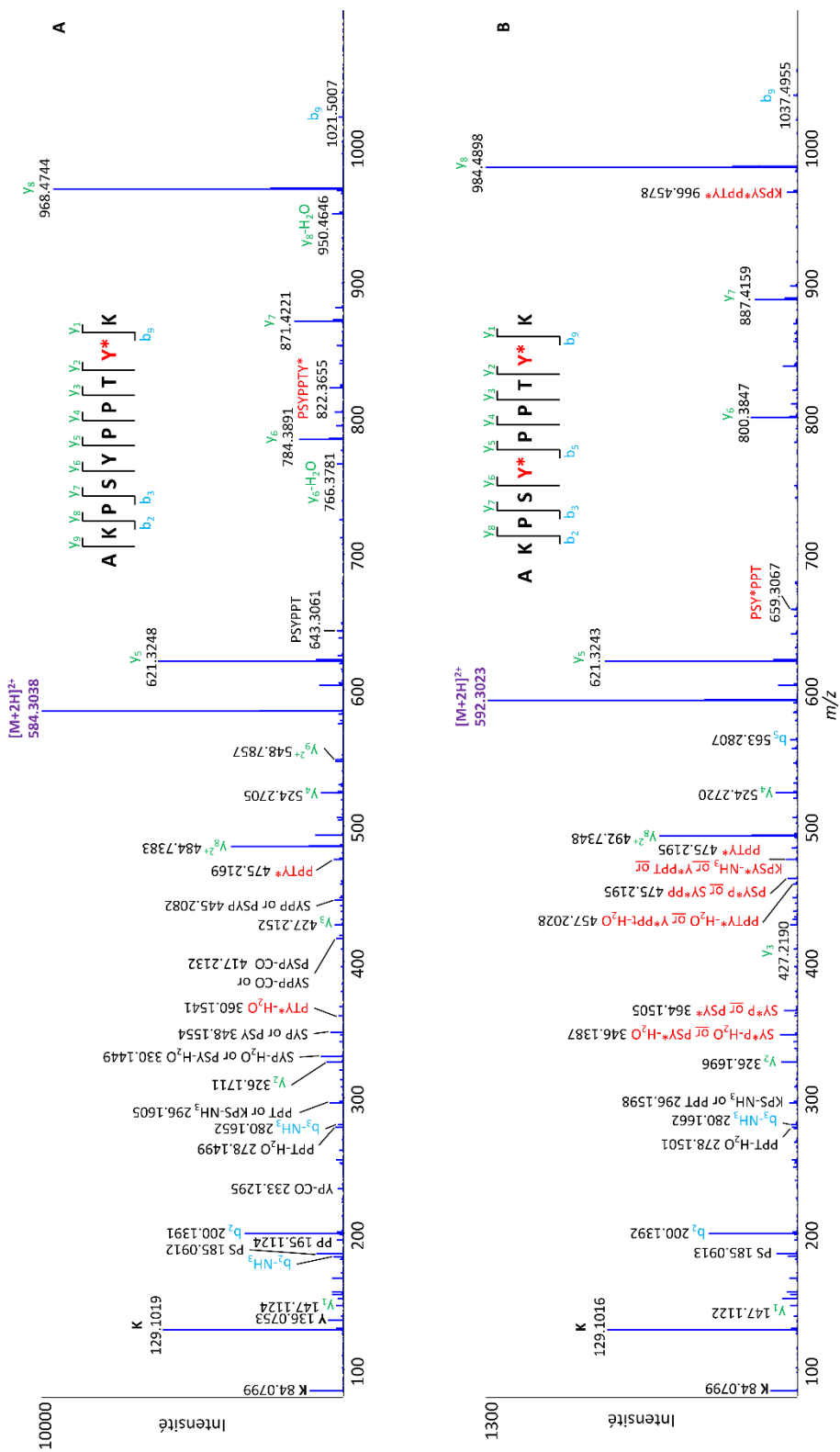


Figure 5.5 Spectre MS/MS obtenus pour les peptides comportant une seule DOPA (A) et deux DOPA (B) issue d'une digestion par la trypsine. Les séries d'ions y (en vert) et b (en bleu) sont présentées sur les deux spectres avec quels fragments internes. La forte intensité des ions y_8 et y_5 est la conséquence de l'effet proline sur la fragmentation du peptide.

n'étaient pas présents dans la liste de fragments théoriques lorsque la tyrosine du côté N-terminal était modifiée.

Du point de vue de la spectrométrie de masse, la présence de plusieurs prolines, et même de prolines vicinales, influence considérablement la fragmentation du peptide en question. Comme l'illustre la Figure 5.5 les spectres de masses du peptide AKPSYPPTYK sont largement prédominés par deux ions acylium (y_5 et y_8) qui correspondent à la fragmentation du lien entre la lysine et la tyrosine respectivement. Ce genre de signature est couramment rencontrée dans les peptides possédant plusieurs prolines. Maux et al. ont expliqué qu'en raison d'une importante basicité et d'une contrainte stérique empêchant la réorganisation du N-terminale, l'ion y correspondant était par conséquent favorisé (Maux, Enjalbal, Martinez, & Aubagnac, 2002). D'un autre côté, la présence de résidu favorisant la séquestration du proton mobile, comme les lysines et les arginines, ou causant un encombrement stérique, favorisera le clivage en C-terminal du peptide (Breci, Tabb, Yates, & Wysocki, 2003).

Une particularité a été observée lors de la comparaison des différentes conditions d'incubation pour le peptide comportant une DOPA (voir Figure 5.6). Pour trois des conditions testées (control, borate et BHT), un pic supplémentaire aux environs de 4.5 minutes a été détecté, avec un m/z de 582.29 (en vert dans la Figure 5.6), qui correspond à la perte de 2 Da (deux protons) par rapport au m/z 584.30. Burzio et Waite ont aussi observé un écart de 2 Da et ont soulevé la possibilité que l'une des lysines du peptide ait réagit avec la dopaquinone. Cependant, aucune information supplémentaire n'a été apportée de la part des auteurs à ce propos. Primel et al. ont proposé un mécanisme semblable et ont postulé que la formation d'une base de Schiff entre la lysine et la DOPA serait à l'origine d'un lien covalent au sein d'une protéine semblable à la rMFP-1 DOPA (Priemel et al., 2020). Cependant, la cinétique entre les amines primaires et les *o*-quinone est relativement lente. Yang et al. ont remarqué entre autres qu'une réaction de Michael était nettement favorisée par rapport à la formation

d'une base de Schiff (J. Yang, Saggiomo, Velders, Stuart, & Kamperman, 2016). Aussi, la base de Schiff obtenue demeure une structure particulièrement sensible qui peut être rapidement dissociée. Par conséquent, il est peu probable qu'une telle structure se soit formée dans les conditions expérimentales actuelles.

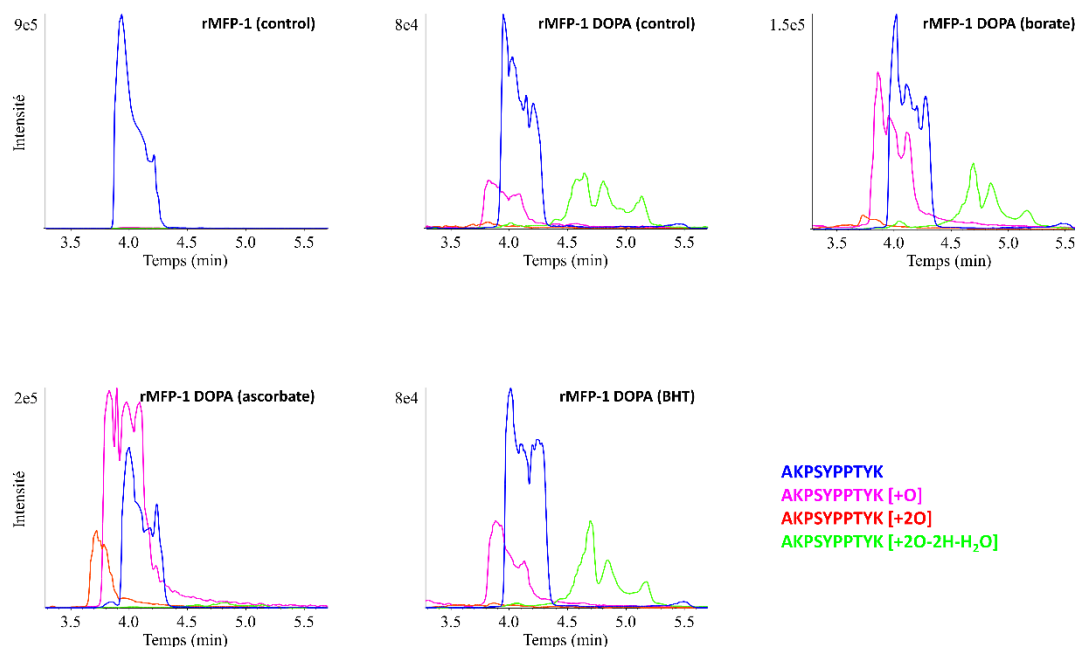


Figure 5.6 Courant d'ions extraits du peptide AKPSYPPTYK comportant une ou deux DOPA lorsqu'incubé en présence de différents agents anti-oxydants.

Lorsque le spectre MS/MS enregistré pour l'ion à m/z 582.29 (voir Figure 5.9) a été comparé à ceux présentés à la Figure 5.5, aucun ion de la série y n'a été détecté à l'exception de l'ion b_2 . L'absence des ions y_5 et y_8 est encore plus intéressant compte tenu qu'ils proviennent de la fragmentation des liens peptidiques K/P et Y/P. Ceci porte à croire qu'un réarrangement empêche le peptide de se fragmenter afin de permettre le séquençage du peptide. Tel que mentionné, la réaction entre une amine primaire et une quinone demeure particulière. En effet, la cinétique de l'amine ϵ de la lysine demeure faible, en comparaison avec d'autres chaînes latérales, en raison de la valeur du pK_a

relativement élevé (Bolton, Turnipseed, & Thompson, 1997). Cependant, il n'est pas impossible de voir ce genre de modification du point de vue de la protéine. En effet, le groupe Lau a observé qu'un motif dans la séquence primaire de la cytochrome c oxydase et d'autres protéines riches en lysine accorderait à certaines lysines la nucléophilicité nécessaire pour réagir avec les quinones (Fisher et al., 2007; Labenski, Fisher, Lo, Monks, & Lau, 2009). Autrement dit, la présence de plusieurs lysines côte-à-côte diminue la valeur du pK_a , permettant ainsi la réaction avec la quinone. Dans la séquence de la rMFP-1, deux lysines sont séparées par une alanine et, afin de minimiser la répulsion coulombique, l'une des deux lysines ne porte pas de charge. Par conséquent, la chaîne latérale est désormais disponible pour réagir avec la quinone. La Figure 5.7 suivante montre la réaction proposée entre la lysine du côté N-terminal et la tyrosine (DOPA) du côté C-terminal.

Une fois la protéine digérée par la trypsine, la résultante de la formation d'un nouveau lien covalent entre la lysine et la DOPA cyclisera le peptide en question, ce qui le stabilisera. La différence de 2 Da mentionnée plus haut est causée par l'oxydation de la DOPA en quinone lors du processus d'ionisation. En effet, Zhang et Bartels ont observé que la forme oxydée de certaines quinones, une fois modifiées par des amines primaires, prédominait le spectre MS (F. Zhang & Bartels, 2004). L'un des premiers résidus à être fragmenté est la proline du côté N-terminale. Ceci concorde, d'une part, avec l'effet proline décrit plus haut, et d'autre part avec le fait que cet ion aurait généralement conduit à l'obtention de l'ion y_8 . Des pertes neutres caractéristiques de la sérine, de la tyrosine et de la proline ont été observées et sont à l'origine des principaux pics (en gras dans la figure) présentés dans le spectre MS/MS. De plus, comme le montre la Figure 5.8 certains des ions rapportés contiennent la liaison croisée entre la lysine et la DOPA. Éventuellement, ces ions pourront servir à confirmer la présence de cette liaison intra-moléculaire à partir de la mfp-1 identifiée dans les échantillons de byssus.

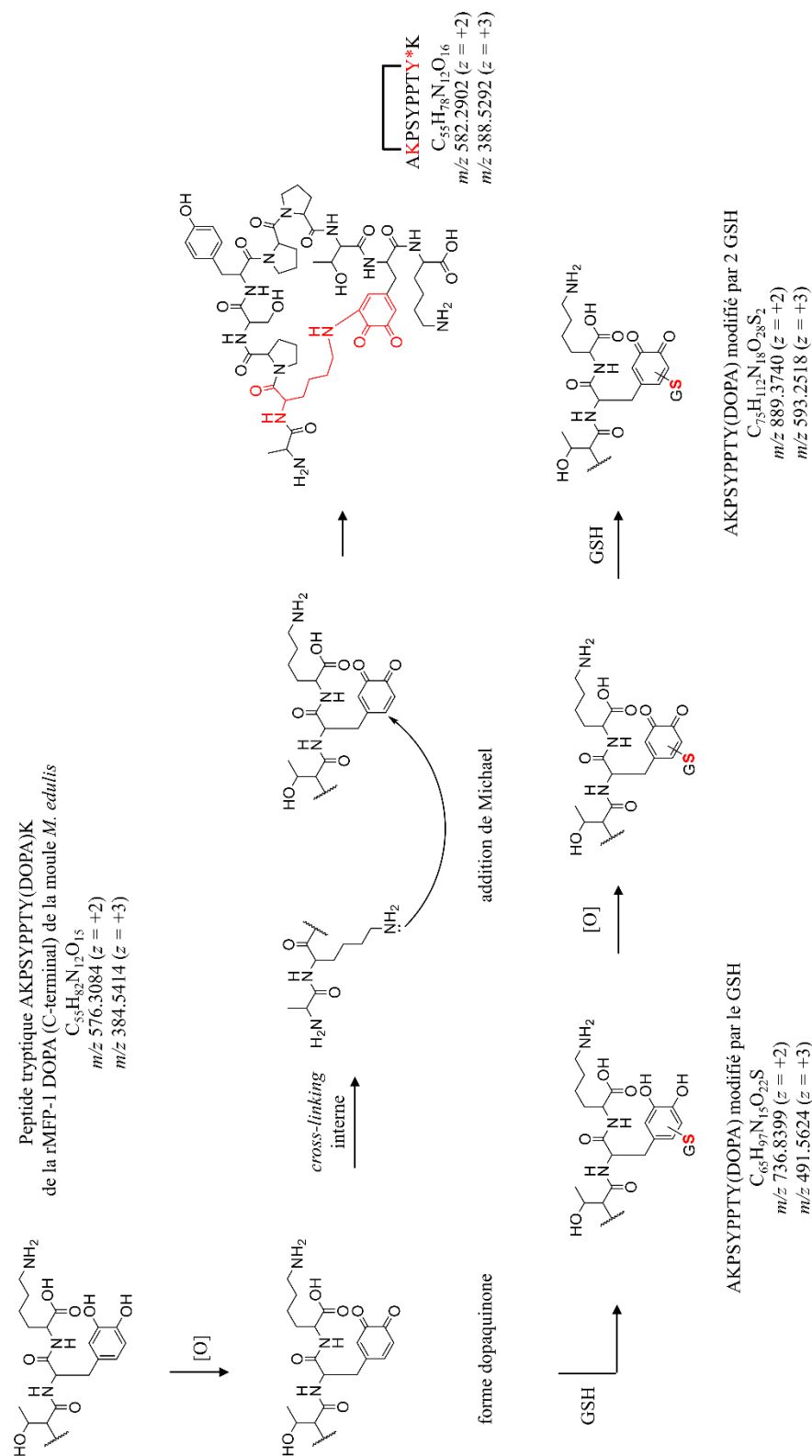


Figure 5.7 Schéma proposé de la réaction entre l'*ortho*-quinone (Tyr-9) et la lysine-2 (réaction présentée au centre) formant une liaison croisée intermoléculaire. La réaction avec les groupements thiol (illustré à l'aide du glutathion) et la quinone fut également étudiée.

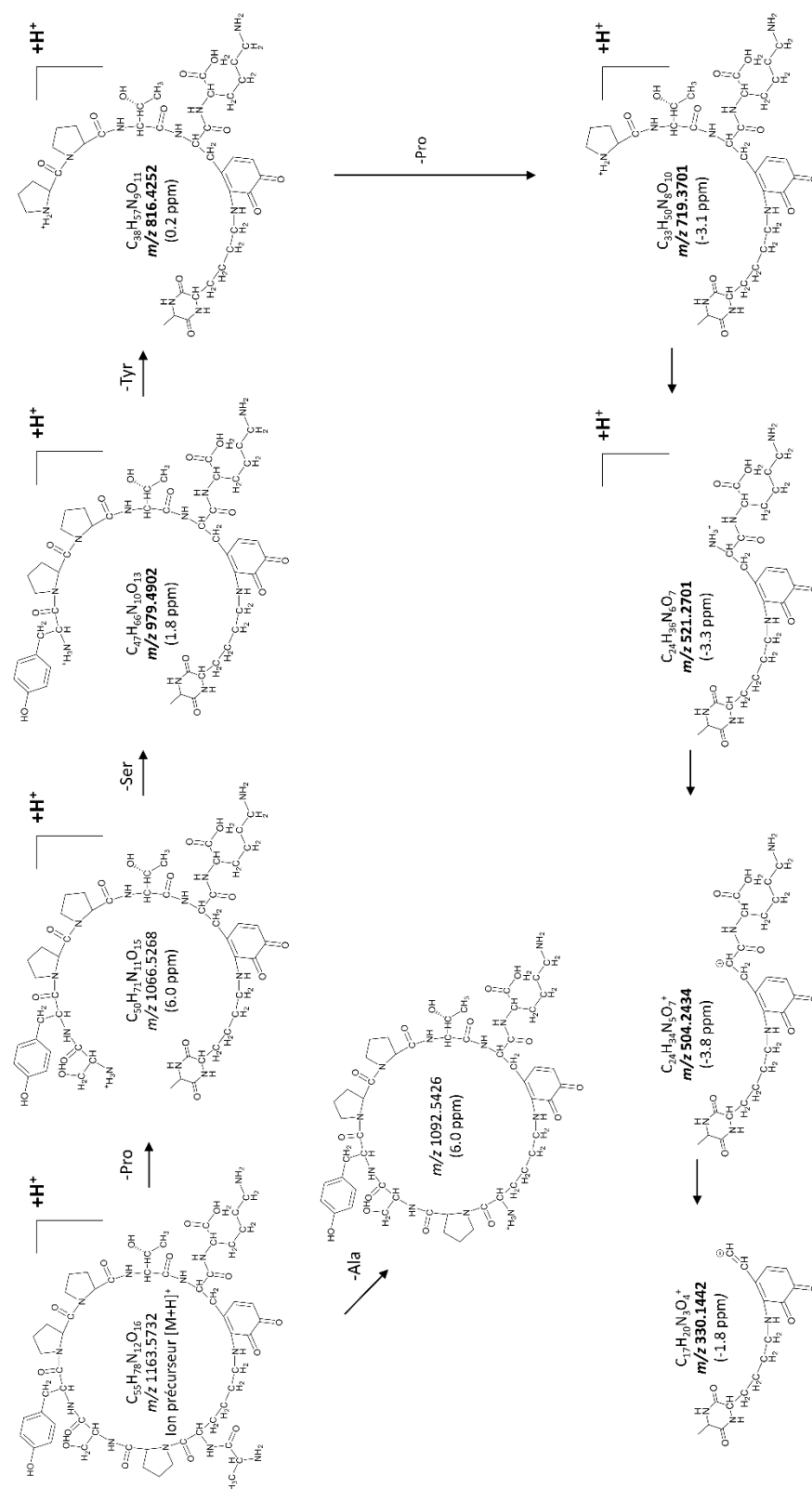


Figure 5.8 Schéma proposé pour la fragmentation du peptide AKPSYPPTYK après qu'une liaison intra-moléculaire entre la Lys-2 et la DOPA (Tyr-9) se soit formée. Les rapport m/z en gras peuvent être utilisés ultérieurement comme ions diagnostiques confirmant la présence de la liaison croisée.

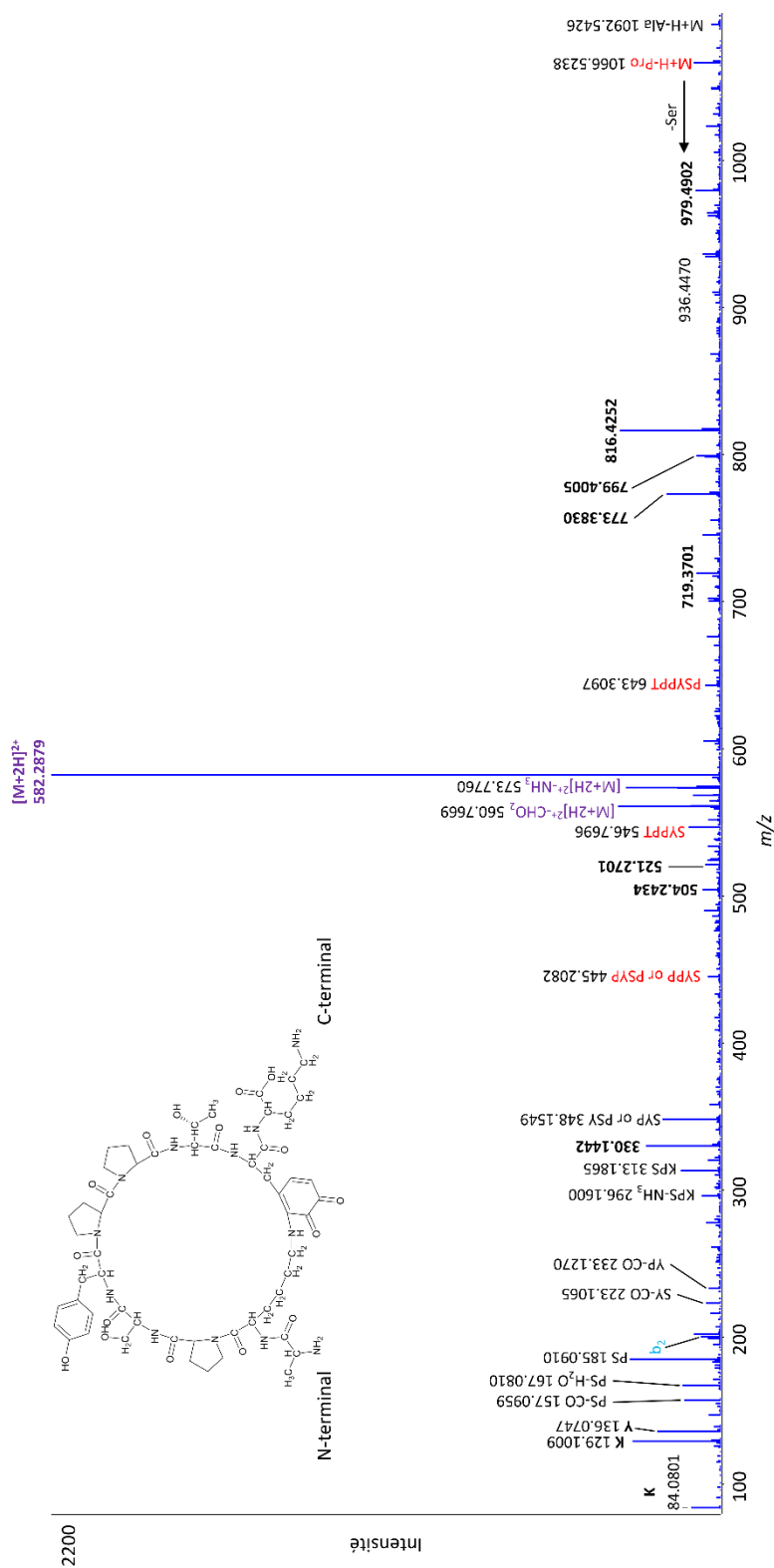


Figure 5.9 Spectre MS/MS du peptide AKPSYPPTYK ayant formé un cycle une fois la lysine-2 attaché à la dopaquinine (Tyr-9). Le séquençage de novo montre l'absence d'ion y et b, potentiellement causé par la liaison croisée, en initiant la fragmentation à partir de la Pro-3. Les ions en gras représentent des fragments pouvant être utilisés comme ions diagnostiques ultérieurement.

Puisque la forme dopaquinone de la protéine recombinante mfp-1 peut aussi réagir avec les groupements thiols (Figure 5.7), cette dernière a été incubée (1h à 37°C) en présence de glutathion (GSH) à des concentrations allant de 0.2 à 2 mM. L'augmentation de la concentration de GSH n'interfère pas avec le signal du peptide non-modifié néanmoins, pour la version du peptide comportant une DOPA, l'intensité maximale est obtenue avec la concentration la plus faible (0.2 mM) en GSH, comme l'illustre la figure ci-dessous, pour ensuite diminuer. Le signal du peptide comportant deux DOPA présente une tendance relativement similaire, à l'exception que le signal demeure stable peu importe la concentration de GSH ajoutée.

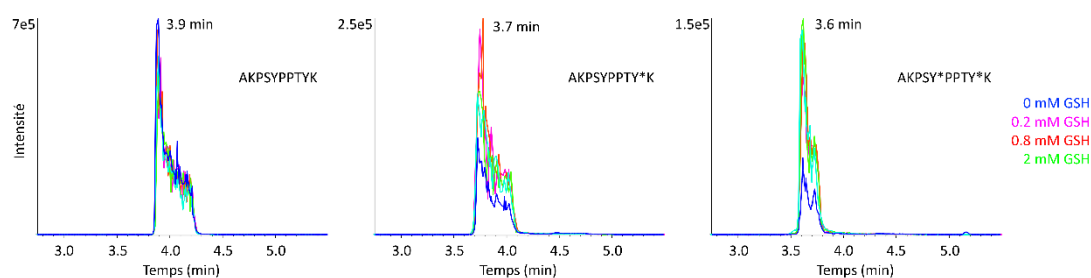


Figure 5.10 Comparaison de l'intensité des signaux obtenus en présence de GSH à différentes concentrations pour le peptide AKPSYPPTYK sans DOPA (gauche), avec une DOPA (centre) et deux DOPA (droite).

Le glutathion est un agent anti-oxydant naturellement présent dans les cellules et offre une protection contre les agents oxydants comme les métaux lourds, et permet aussi la capture de substances réactives. Par conséquent, il n'est pas surprenant qu'il empêche l'oxydation de la DOPA avant de réagir avec la dopaquinone. Sachant que le groupement thiol peut réagir avec la dopaquinone, tel qu'il sera présenté au chapitre 6, une série d'essais visant à connaître le moment opportun pour l'ajout du GSH a été réalisée. La Figure 5.11 ci-dessous résume les étapes des trois scénarios envisagés.

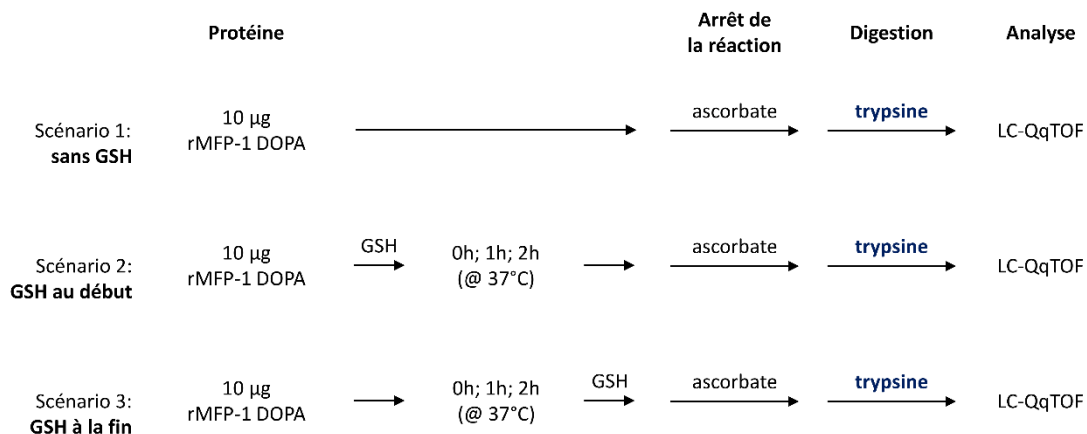


Figure 5.11 Résumé des étapes visant à l'optimisation des incubations de la protéine rMFP-1 DOPA avec le glutathion. Pour chaque scénario, une concentration finale en glutathion et ascorbate de 0.4 mM et 1 mM a été respectivement visée. Une quantité de 2 µg de trypsin a ensuite été ajoutée pour une digestion d'une durée d'une heure.

Au terme de ces essais, comme le montrent les chromatogrammes suivants (Figure 5.12), le signal de AKPSYPPTY*K diminue rapidement lorsque le glutathion n'est pas ajouté ou ajouté après la période d'incubation. Par la suite, dans le second scénario, lorsque le GSH est ajouté dès le début, le signal de AKPSYPPTY*K demeure relativement au terme des deux pré-incubations, supportant l'effet de protection du GSH par rapport à l'oxydation de la DOPA en dopaquinone. Néanmoins, un signal associé à la formation d'un adduit GSH est détecté avec une excellente exactitude (<1 ppm) et atteint rapidement un sommet après 1h de pré-incubation. Finalement, le dernier scénario implique l'ajout du GSH au terme de la période de pré-incubation. Le chromatogramme d'ions extraits obtenu montre clairement l'effet négatif sur la formation de l'adduit GSH, avec un signal maximum mesuré immédiatement suivant l'ajout de l'acide ascorbique ($t = 0$).

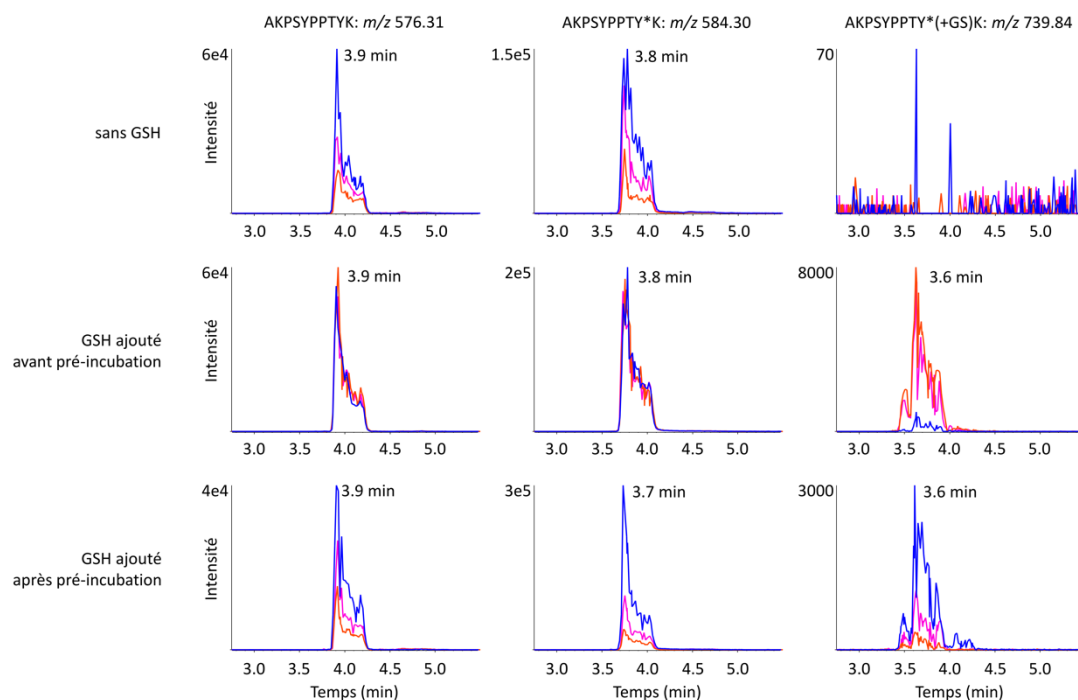


Figure 5.12 Comparaison de l'impact d'une pré-incubation avant ou après l'ajout de GSH sur la sensibilité du peptide tryptique de la rMFP-1 (avec ou sans DOPA). La comparaison montre également l'effet sur la formation de l'adduit GSH avec la version DOPA du peptide AKPSYPPTY*K.

L'analyse du spectre MS/MS de l'adduit GSH sur le peptide comportant une DOPA a permis la réalisation du séquençage du peptide (Figure 5.13). Dans un premier temps, plusieurs ions de la série *y* et quelques-uns de la série *b* ont été attribués à la plupart des pics détectés. Comme observé à la Figure 5.13, les ions y_5 et y_8 prédominent le spectre MS/MS en raison de l'effet proline. Certains de ces ions, par exemple, correspondent à une perte d'eau (-18 Da) ou à une fragmentation du GSH au niveau de l'acide γ -glutamique (-129 Da). Il s'agit de l'une des pertes neutres caractéristique du glutathion qui permet de confirmer la présence du GSH sur le peptide.

Une fois l'addition de Michael entre la dopaquinone et le groupement thiol complétée, il n'est pas rare de voir une seconde addition sur le groupement catéchol (Xu, Huang, Kramer, & Hawley, 1996). Par conséquent, une multitude de réactions est possible comme l'illustre la

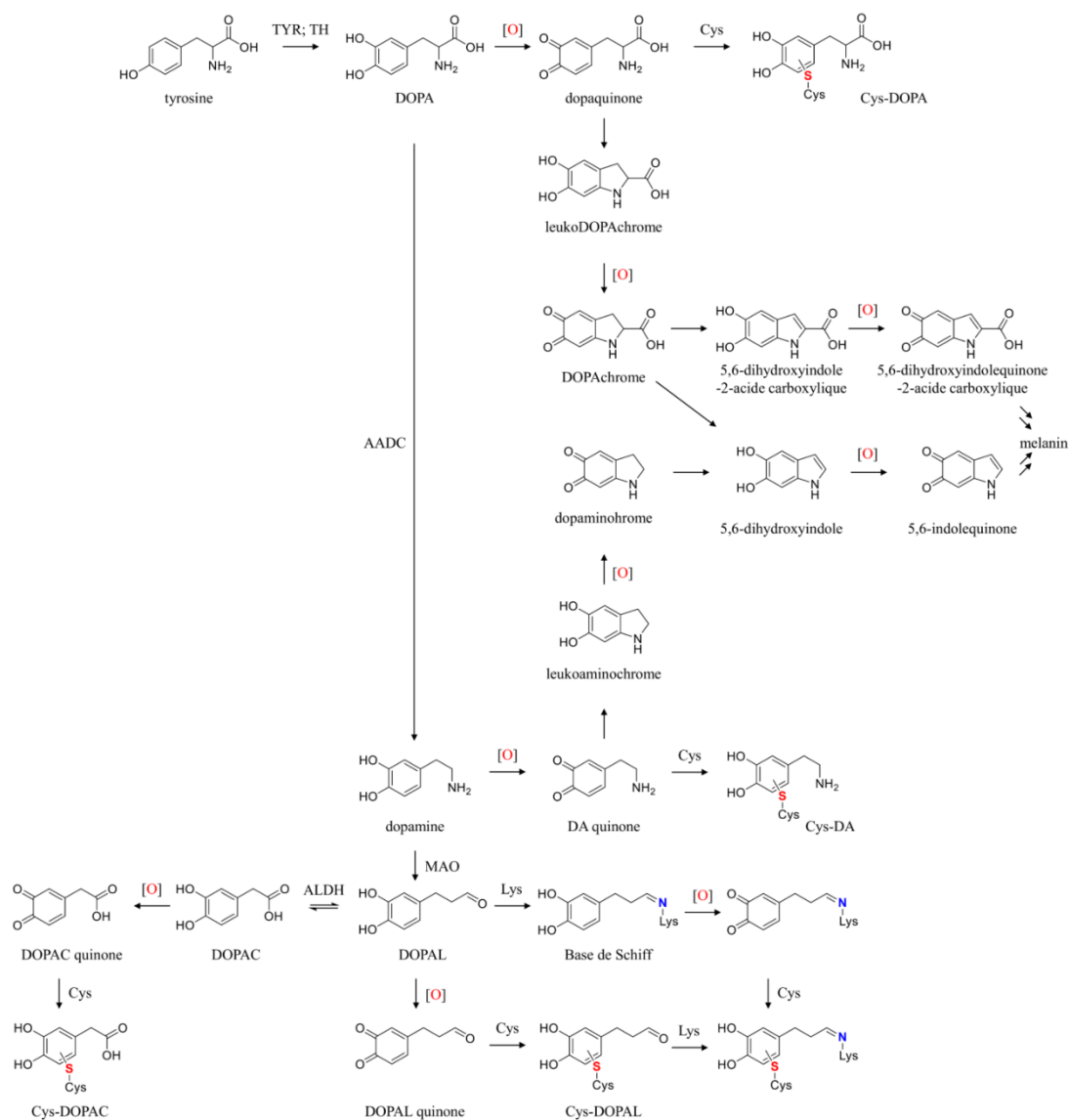


Figure 1.13 Représentation simplifiée du métabolisme de la dopamine, donnant lieu à plusieurs intermédiaires réactifs potentiellement impliqués dans les liaisons covalentes avec les cystéines et les lysines des protéines.

. Cependant, aucune de ces possibilités n'a été identifiée avec la méthode actuelle et ce, même lorsque les incubations ont été faites entre la DOPA et la cystéine (voir chapitre 6). Il est donc peu probable que ce type de réactions successives se déroule au niveau de la protéine rMFP-1 lorsqu'elle est oxydée en raison de l'encombrement stérique. Aussi, la réaction avec la chaîne latérale de la lysine joue un rôle considérable dans la stabilité de l'adduit. Par conséquent, les conditions d'ionisation et de fragmentation à eux seuls peuvent être néfastes pour une détection adéquate du peptide comportant plus d'un adduit.

5.4.3 Hydrolyse complète de la protéine rMFP-1

Une hydrolyse acide à haute température est généralement la méthode la plus courante pour l'étude de la composition en acides aminés d'une protéine. Cette approche implique l'utilisation d'une solution d'acide chlorhydrique à 6 M, chauffée entre 110-120°C durant 24h. Cette méthode a d'ailleurs été employée par Zhao et Waite afin de montrer l'existence du Cys-DOPA dans les fibres de byssus de la moule verte (Zhao & Waite, 2005). Toutefois, l'emploi d'une concentration relativement élevée d'acide chlorhydrique maintenu à haute température peut s'avérer problématique pour la détection de certains acides aminés, comme la cystéine (Leblanc et al., 2012). Par ailleurs, il a été observé que certains acides aminés subissaient des modifications lors de l'hydrolyse (Fountoulakis & Lahm, 1998).

Afin d'identifier la Cys-DOPA à partir de protéines liées, une hydrolyse complète a été réalisée à l'aide d'une solution 6 M HCl (24h, 110°C) et, en parallèle, avec de la pronase. Comme mentionné précédemment, les hydrolyses complètes

employant du HCl sont probablement les approches les plus courantes. Cependant, certains acides aminés peuvent être dégradés ou subir des modifications (déamidation,

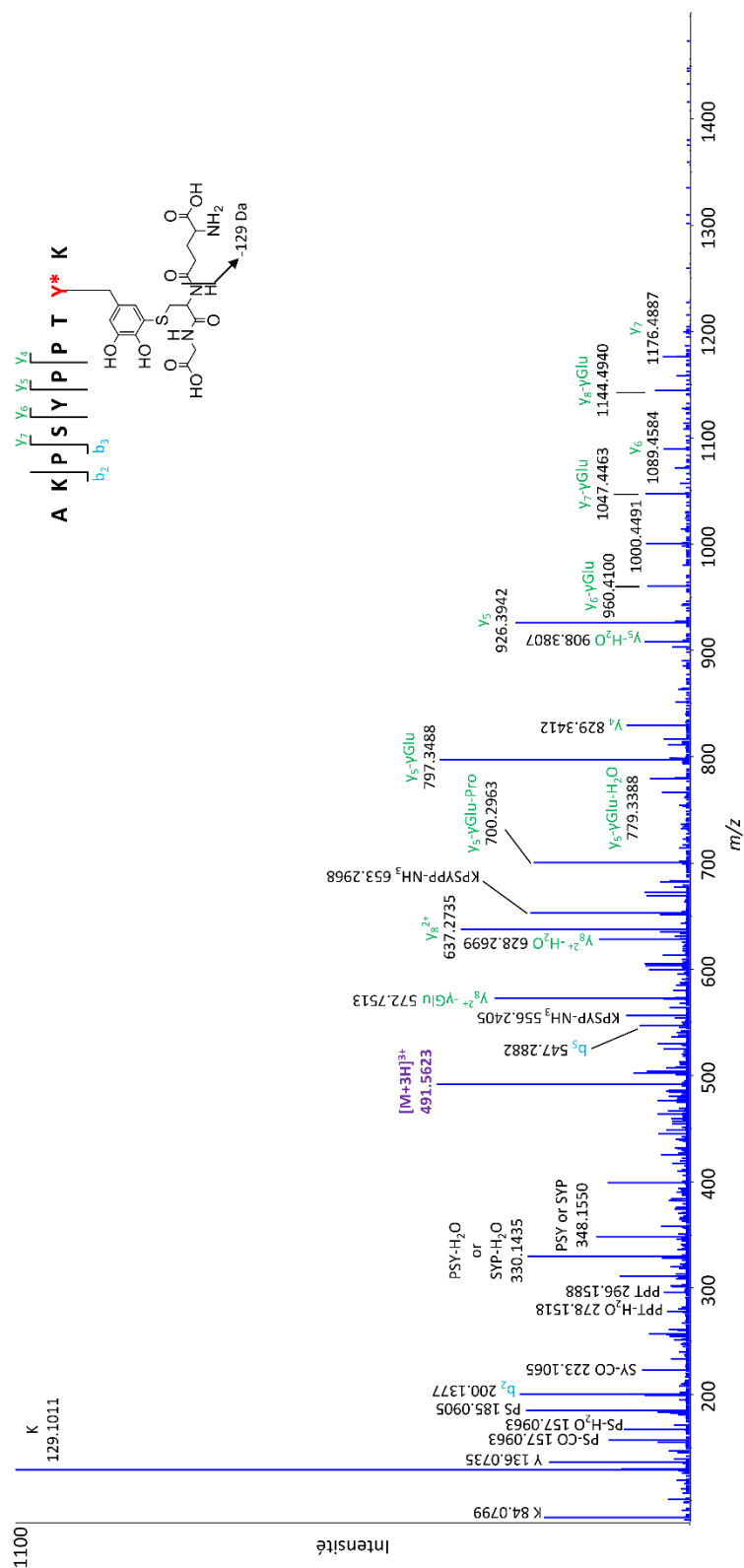


Figure 5.13 Spectre MS/MS du peptide AKPSYPPTY*K comportant un adduit GSH. Le séquençage de novo à permis l'identification de plusieurs ions y incluant le GSH dans la structure.

halogénéation et oxydation) durant la digestion (Fountoulakis & Lahm, 1998; Leblanc et al., 2012). Malgré tout, la cystéine demeure stable, une fois attachée à la dopaquinone, pour survivre aux conditions extrêmes de la digestion (Zhao & Waite, 2006a). Parallèlement, une approche alternative employant une combinaison de protéases a été introduite (Mitrea et al., 2010). Cette approche a l'avantage d'être sélective à la dissociation des liens peptidiques en plus d'employer des températures beaucoup moins extrêmes (40-50°C). Malheureusement, aucune des digestions réalisées par HCl ou par la pronase n'a produit des résultats concluants.

Une quantité de 10 µg de protéine recombinante mfp-1 a été mélangée avec une solution tampon Tris. (pH 8) et incubée durant une période de 0; 1; 2; 4; 6; 18 et 24h en présence de glutathion. De l'acide ascorbique a été ajouté au terme du temps d'incubation afin d'interrompre la réaction entre la DOPA et le GSH. Environ 250 µg de pronase, préalablement dissoute dans une solution tampon Tris contenant 50 mM en CaCl₂, a été ajouté au mélange de rMFP-1. Parallèlement, pour l'hydrolyse acide, une quantité similaire de rMFP-1 a été incubée avec du GSH et hydrolysée à l'aide de 6 M HCl durant 6 et 24h. Les échantillons ont ensuite été analysés par LC-MS/MS avec un gradient de 12 min et une colonne biphényle. Une partie de l'échantillon a été réservée pour une dérivation pré-colonne utilisant le composé benzoyle afin d'augmenter la rétention sur la colonne.

La progression de la digestion par la pronase a été suivie à l'aide du signal de la tyrosine. Cependant, le signal associé à la tyrosine augmente rapidement au terme d'une digestion d'une heure et demeure constante par la suite. La dérivation chimique a montré que la tyrosine était bel et bien détectée avec un signal important après la première heure. Les résultats provenant de l'hydrolyse acide n'ayant pas été concluant, des tests additionnels ont été réalisés afin de s'assurer que la pronase était vraiment désactivée. De l'acétonitrile (1 :1 v/v) a été ajouté à cet effet. Également, un groupe d'échantillon a été digéré parallèlement à une température de 25 et 37°C à des fins

comparatives. Autrement, les conditions expérimentales ont été les mêmes que celles énoncées au paragraphe précédent. Les échantillons ont ensuite été centrifugés pour en retirer le surnageant et évaporés sous azote afin d'être solubilisés à nouveau dans une solution 10% ACN. Une fois de plus, une partie du digestat a été réservée pour la dérivation chimique, puis dérivé avec le composé benzoyle au terme de l'hydrolyse. Les tests n'ont présenté aucun résultat concluant.

5.4.4 Identification de la protéine mfp-1 du byssus

L'un des constats soulevé dans les chapitres 3 et 4 a été l'absence de la protéine mfp-1 dans les échantillons de pied et de byssus. Sachant que cette protéine est incluse dans la base de données UniProt, et avec les informations recueillies à partir des tests menés précédemment sur la protéine recombinante MFP-1, des essais supplémentaires ont été réalisés sur des échantillons de byssus de la moule *M. edulis*.

Dans un premier temps, les échantillons digérés de pied de moules utilisés lors de l'étude présentée au chapitre 3 ont été cherchés de nouveau dans la base UniProt à l'aide du logiciel ProteinPilot. Ce logiciel offre la possibilité de modifier les probabilités associées à une modification post-traductionnelle pour un résidu donné. Dans le cas actuel, sachant que l'une des tyrosines est potentiellement oxydée en DOPA et qu'il est supposé que l'hydroxyproline y est présent, les paramètres initiaux en lien avec la présence de DOPA et d'hydroxyproline ont été augmentés de manière à favoriser la détection du peptide. Le tableau suivant présente les conditions initiales et finales de probabilités associées à ces modifications pour la tyrosine et la proline.

L'augmentation des probabilités n'a pas dépassé les 10% pour l'oxydation [+O] de la proline et la tyrosine. Aussi, la probabilité associée à la dihydroxyproline [+2O] a été maintenue à une valeur de 0.1%. Au-delà de ces valeurs, la pertinence des rapports m/z correspondant à la présence de ces modifications est largement mise en doute.

Tableau 5.2 Probabilités testées sur la proline et la tyrosine afin de permettre l'identification de la protéine mfp-1 provenant de la plaque du byssus de la moule *Mytilus edulis*.

résidue	modification	probabilité initiale	probabilité testée
tyrosine (Y)	oxidation [+O]	1.3%	10%
proline (P)	oxidation [+O]	1.3%	10%
	oxidation [+2O]	0.1%	0.1%

D'une part, après un certain pourcentage, l'algorithme «force» à associer des pics n'étant pas des peptides à correspondre aux valeurs de m/z recherchées et d'autre part, l'abondance de ces modifications demeure relativement faible. Par conséquent, un travail supplémentaire de vérification doit être mené afin de montrer que les pics identifiés sont bel et bien des peptides provenant de la protéine étudiée. Malheureusement, la protéine mfp-1 n'a pas été identifiée à partir des digestions de pied de la moule *M. edulis*.

Ensuite, une tentative d'identifier la mfp-1 à partir de la plaque du byssus a été réalisée. La mfp-1 est une protéine qui recouvre la totalité de la fibre ainsi que la plaque du byssus. Dans cette expérience, la plaque de la fibre de byssus a d'abord été prélevée et broyée, puis mélangée à une solution de 7 M d'urée et 2 M de thiourée. Les échantillons furent soniqués et chauffés à 50°C (Sansoucy et al., 2020) pour ensuite être incubés avec ou sans acide ascorbique. Finalement, une digestion enzymatique à l'aide de la trypsine a été réalisée pour une durée de 1-4h avant que les échantillons soient analysés par LC-MS/MS. L'évaluation des XIC pour les masses exactes présentées au tableau 6.1, jusqu'à concurrence de l'ajout de 6 oxygènes, a été considérée. Toutefois, aucun des échantillons ne semblait contenir le peptide

AKPSYPPTYK recherché, d'une part en raison de l'allure générale des XIC qui montraient essentiellement la présence d'artéfacts et d'autre part, en raison de l'importante erreur sur les masses exactes mesurées qui se situaient nettement en-deçà de la tolérance acceptée par l'instrument (> 30 ppm pour certains cas) pour les pics potentiellement intéressants.

Finalement, une dernière tentative a été réalisée à partir de la glande cuticule du pied de *M. edulis*. Cette glande est riche en mfp-1 et offre ainsi la meilleure chance de détecter cette protéine dans les conditions de préparation et d'analyse actuelles. La glande a été homogénéisée dans la même solution d'urée et de thiourée avant d'être mélangée à de l'acide ascorbique afin de favoriser la forme catéchol du peptide. La digestion tryptique des échantillons s'est déroulée dans les mêmes conditions (2 μ g, 37°C, 1h) avant d'être directement analysée par LC-MS/MS. Malheureusement, les résultats de cette approche ont, une fois de plus, été non concluants dans les conditions expérimentales proposées. À ce stade, il n'est pas totalement clair si l'absence de la protéine est causée par la présence des liaisons croisées impliquant les résidus DOPA ou en raison de notre procédure d'extraction. Plus de détails seront fournis aux chapitres suivants concernant les expériences futures à ce sujet. Aussi, il demeure à déterminer si les liaisons croisées se forment avant ou après la sécrétion du byssus. Ces liaisons peuvent être la résultante d'une oxydation à long terme, mais peuvent aussi être présentes dès la sécrétion de la fibre afin de la protéger immédiatement de l'environnement extérieur.

CHAPITRE VI

DÉTERMINATION DE RÉACTIONS COVALENTES ENTRE LA DOPA ET AUTRES CATÉCHOLES AVEC LA CYSTÉINE PAR LC-MS/MS

L'hydroxylation des résidus tyrosines en un acide aminé peu commun, le 3,4-dihydroxylphényl-*L*-aniline (DOPA), représente une modification particulièrement intéressante, d'une part corrélée avec diverses maladies neurodégénératives et désordres cognitifs et, d'autre part, les propriétés adhésives et bio-mécaniques observées chez les moules sont en partie la conséquence de cette modification. Parallèlement, la DOPA a aussi été utilisée comme biomarqueur afin d'identifier les protéines exposées à un stress oxydatif. Néanmoins, l'identification de cette modification demeure un défi considérable compte tenu que peu de protéines, autres que les protéines de moules, sont connues comme ayant la DOPA comme modification.

D'une part, ce chapitre présentera les conditions dans lesquelles la DOPA peut réagir avec certains acides aminés nucléophiles. Il sera question dans cette première partie des conditions optimales menant à l'identification de modification de protéines en considérant l'intermédiaire réactif de la DOPA, la dopaquinone, comme composé modifiant les protéines. Ensuite, il sera question de la DOPA en tant que modification présente sur les protéines. De plus les approches requises afin d'identifier des liaisons croisées au sein d'une protéine recombinante de la moule *M. edulis* dans laquelle la DOPA pourrait réagir avec un ou plusieurs sites nucléophiles seront présentées. Ce chapitre fera office d'une publication ultérieure.

6.1 Résumé

L'oxydation de la tyrosine en 3,4-dihydroxyphenyl-*L*-alanine (DOPA) occupe une place particulièrement importante dans le métabolisme de la dopamine. Plusieurs maladies neurodégénératives ainsi que certains désordres cognitifs ont comme point commun un dérèglement dans les niveaux de dopamine.

D'un autre côté, la DOPA peut également se retrouver comme modification post-traductionnelle dans certaines protéines. Bien que cette modification demeure relativement rare, elle est à l'origine des propriétés adhésives et bio-mécaniques observées dans le byssus de la moule. Dernièrement, certains groupes de recherche œuvrant dans le domaine des biomatériaux ont mentionné la possibilité de réactions entre la DOPA et certaines chaînes latérales nucléophiles afin de former des liaisons inter ou intra-protéines.

Puisque ces hypothèses n'ont pas été explorées en profondeur, l'objectif premier de ce chapitre sera d'établir les conditions optimales pour la réaction entre la cystéine et la DOPA. La dopaquinone sera obtenue à partir d'incubation employant la tyrosinase à différents ratios DOPA : cystéine.

6.2 Introduction

L'obtention de la DOPA est une étape déterminante dans le métabolisme de la dopamine et implique généralement l'une des deux enzymes suivantes; la tyrosine hydrolase ou la tyrosinase. Certains désordres cognitifs tels que la schizophrénie (Grabnar et al., 2011) ou des maladies neurodégénératives telles que la maladie d'Alzheimer (Nam et al., 2018) ou de Parkinson (Burbulla et al., 2017) partagent comme point commun une déficience en dopamine dans les régions dopaminergiques du cerveau. L'une des explications de ce déséquilibre repose sur le fait que le

groupement catéchol de la dopamine et de la DOPA est particulièrement sensible à l'oxydation et donne naissance à un intermédiaire réactif, la dopamine quinone et la dopaquinone. Ces intermédiaires sont connus pour réagir avec les nucléophiles doux, comme les thiols des cystéines, ce qui conduit à la formation de liaisons covalentes (Y. Li, Jongberg, Andersen, Davies, & Lund, 2016). Par exemple, des travaux publiés en 2012 ont montré l'aptitude de la dopamine quinone à se lier avec la protéine de signal redox, Redox Sensor DJ-1, via les cystéines 53 et 106, connectant ainsi les deux protéines entre elles (Giroto et al., 2012). Suite à cette modification, la protéine n'est plus en mesure d'offrir une protection adéquate face à un éventuel stress oxydatif.

La DOPA peut aussi être considérée comme une modification post-traductionnelle importante. En 2010, Zang et al. ont utilisé cette modification comme biomarqueur potentiel afin d'évaluer l'exposition à un stress oxydatif (X. Zhang et al., 2010). Plus précisément, ils ont observé que les protéines directement impliquées dans la chaîne de respiration cellulaire au sein de la mitochondrie étaient plus susceptibles de subir un stress oxydatif. Les protéines comme la déshydrogénase malate 2 étaient donc plus susceptibles de contenir une DOPA. La présence de la DOPA est relativement rare en tant que PTM et sa présence est peut-être la conséquence d'une exposition accrue ou répétée à des espèces réactives de l'oxygène.

Chez les moules, certaines protéines sont connues comme étant des protéines riches en DOPA. Les protéines mfp-1 et mfp-5 du byssus de la moule, contenant environ de 15 à 30 mol% en DOPA, sont un bon exemple (B. P. Lee et al., 2011). La présence du DOPA dans le byssus permet à la moule de s'attacher solidement sur n'importe quelle surface solide tout en permettant le renforcement de l'intégrité structurel de la fibre de byssus. Toutefois, cette dernière perd de son élasticité (plastification) avec le temps, la rendant plus sensible aux éventuels stress mécaniques et donc plus fragile. L'une des hypothèses qui a été émise suggère qu'en raison d'un stress oxydatif important, la densité électronique du DOPA est réorganisée afin de

former l'intermédiaire réactif dopaquinone pour ensuite réagir avec les groupements thiols des protéines riches en cystéines présentes dans le byssus (Yu et al., 2011; Zhao & Waite, 2006a).

Actuellement, seul Zhao et Waite ont présenté des preuves convenables quant à la présence de liaisons covalentes entre la DOPA et la cystéine dans le byssus (Zhao & Waite, 2006a). Une analyse plus avancée a permis d'obtenir des ions caractéristiques qui ont servi à une identification rapide de la présence de la Cys-DOPA (Zhao & Waite, 2005). D'autres chaînes latérales, comme l'histidine et la lysine, peuvent aussi réagir avec la DOPA, mais avec une cinétique nettement plus lente en comparaison avec les cystéines (Y. Li et al., 2016). La modification His-DOPA, moins commune, a été observée dans des échantillons de becs de calamars (Miserez et al., 2008) ainsi que dans la carapace de certains insectes (Kerwin et al., 1999).

Les travaux présentés dans ce chapitre porteront sur l'optimisation des conditions expérimentales menant à l'obtention de la Cys-DOPA. Différents acides aminés tels que l'histidine et la lysine seront également testés afin de montrer leur aptitude respective à former un adduit avec la dopaquinone.

6.3 Expérimentale

6.3.1 Matériel

La cystéine, la 3,4-dihydroxyphenyl-L-alanine (DOPA), la dopamine, l'acide 3,4-dihydroxyphenylacetic (DOPAC), la tyrosinase de champignon, la phénylalanine ainsi que l'acide formique (grade MS) ont tous été obtenus de Sigma-Aldrich (St. Louis, MO, USA). La tyrosine deutérée (*L*-4-Hydroxyphenyl-2,6-*d*₂-alanine) a été achetée chez CDN Isotopes (Pointe-Claire, QC, Canada).

6.3.2 Oxydation par la tyrosinase

Une solution fraîche de tyrosinase (5 µg) a été mélangée avec une solution de cystéine et de DOPA dans les ratios molaires allant de 10:1; 5:1, 2:1; 1:1; 1:2; 1:5; 1:10 Cys:DOPA préparée en triplicatas (pour un volume final de 200 µl) et incubée pendant 20 min à 25°C dans un tampon phosphate (50 mM, pH 6-7). Pour arrêter la réaction avec la tyrosinase, le pH de la solution a été amené à une valeur de 5. La cinétique de la réaction a également été étudiée entre la cystéine et la DOPA. Des échantillons ont été incubés pour une durée maximale de 60 min, avec une quantité de tyrosinase allant de 1 à 10 µg. Les incubations ont été réalisées en triplicatas avec des ratios Cys :DOPA de 10 :1 et 2 :1. Des incubations supplémentaires avec le DOPAC, la dopamine, la tyrosine et la phénylalanine ont été réalisées à des ratios de 10 :1 et 5 :1 selon les conditions décrites plus haut avec 5 µg de tyrosinase.

6.3.3 Dérivation pré-colonne au benzoyle

Afin d'améliorer la rétention sur la colonne C18, les échantillons ont été dérivés selon une procédure préalablement décrite (Wagner et al., 2015). Sommairement, une fois l'incubation avec la tyrosinase complétée, les échantillons ont été dilués par un facteur de quatre dans un tampon borate (50 mM, pH 9). L'agent dérivatif, le *N*-benzoyloxysuccinimide (2.5 mM), est ensuite directement ajouté au mélange en prévision de l'analyse par spectrométrie de masse.

6.3.4 Analyse RP-LC-MS/MS

Les échantillons ont été injectés (5 µl) sur une colonne Luna PFP (150 x 2 mm, avec des particules de 3 µm, avec des pores de dimension 100 Å (Phenomenex, Torrance, CA, USA). Un système Nexera UHPLC (Shimadzu, Columbia, MD, USA), opéré à un débit de 300 µl/min (25°C), a été utilisé avec de l'eau (A) et de l'ACN (B), tous deux acidifiés avec de l'acide formique (0.1%). Le gradient débute à 3% de la phase mobile B pour les 1.8 min, il est ensuite augmenté à 55% à 10 min, puis à 85%

en une minute et maintenu durant 3 min. Les spectres MS ont été enregistrés sur un système hybride quadripôle temps-de-vol à haute résolution TripleTOF 5600 (Sciex, Concord, ON, Canada) équipé d'une source d'ionisation DuoSpray (opéré en mode positif). Les réglages de la source ont été fixés à un voltage de 5.5 kV et une température de 450°C, avec un débit de gaz nébulisation GS1 et de séchage GS2 de 50 psi, avec finalement un potentiel de déclusterisation de 60 V. Un balayage a été réalisé dans une fenêtre de m/z 80 – 925 (300 ms de temps d'accumulation) pour les spectres MS, suivit d'un balayage allant de m/z 50 – 700 pour l'obtention des spectres d'ions produits MS/MS sur les cinq ions les plus intenses. Une restriction d'exclusion de l'ion cible a été appliquée durant 20 s après 2 occurrences avec une soustraction dynamique du bruit de fond pour l'ensemble des expériences DDA. Chaque spectre MS/MS ont été obtenus avec un temps d'accumulation de 150 ms et une énergie de collision de 30 ± 10 V. Le temps de cycle total était de 1.1 s. L'exactitude sur la masse a été suivie à l'aide d'une solution de calibration injectée à chaque quatre échantillons.

6.3.5 Traitement des données

Le traitement des données MS et MS/MS s'est effectué à l'aide du logiciel PeakView (version 2.2) pour la vérification des temps de rétention et de l'exactitude sur la masse. Ensuite, pour une approche semi-quantitative, les données ont été importées dans le logiciel MultiQuant (version 2.3) pour une comparaison des aires sous les pics.

6.4 Résultats et discussion

6.4.1 Optimisation de l'incubation

Comme mentionné plus haut, l'ajout de la tyrosinase est l'une des manières d'obtenir la DOPA et ultimement, la dopaquinone (Figure 6.1). Puisque la dopaquinone est un intermédiaire réactif, elle ne peut donc pas être analysée directement, par

conséquent, la cystéine a été ajoutée afin de piéger la dopaquinone. Dans le but de déterminer les conditions expérimentales optimales, la DOPA et la cystéine ont été incubées avec des quantités variables de tyrosinase. Également, le temps de réaction entre la DOPA et la cystéine a été évalué. De plus, dans une seconde expérience, la DOPA a été remplacée par la tyrosine pour réagir avec la cystéine. Les échantillons ont été analysés par LC-MS/MS et les résultats sont présentés dans la Figure 6.2.

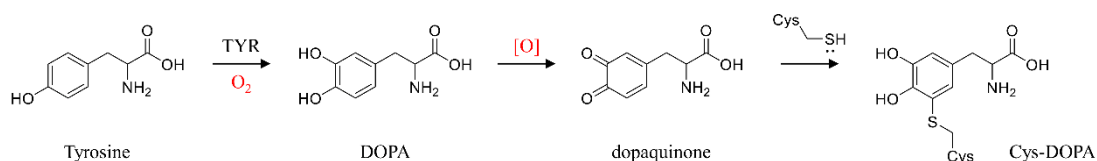


Figure 6.1 Réaction d'oxydation de la tyrosine en DOPA avec l'aide de la tyrosinase. La fonction dioxygénase de la tyrosinase permet une oxydation successive pour convertir la DOPA en dopaquinone, l'intermédiaire réactif de la DOPA.

Sur la base de la spectrométrie de masse à haute résolution, le Cys-DOPA a été identifié avec une importante exactitude de masse (sous les 1 ppm). La quantité de DOPA ajoutée par rapport à la quantité de cystéine est un facteur important dans le rendement de la réaction. En effet, la quantité de Cys-DOPA obtenue décline lorsque la quantité de cystéine dans le mélange diminue. Tel que présenté dans la Figure 6.2A, le rendement pour les ratios de 10 :1, 5 :1 et 2 :1 Cys :DOPA est comparable, avec une excellente reproductibilité. Pour les incubations réalisées avec la tyrosine, la quantité de Cys-DOPA produite diminue rapidement lorsque la réaction se déroule sous un ratio équimolaire, puis atteint son plus bas niveau lorsque la quantité de DOPA surpasse la quantité de cystéine dans le mélange. Ceci s'explique par un changement de conformation dans la structure tertiaire de la tyrosinase lorsque la DOPA (ou la tyrosine) est ajoutée en excès (inactivation suicide). La forme *oxy*-tyrosinase est altérée après le transfert d'un proton au groupement catéchol. Plus précisément, ce transfert implique le relargage de l'un des atomes de cuivre du site actif, ce qui entraîne l'inactivation de

l'enzyme (Muñoz-Muñoz et al., 2010; Ramsden & Riley, 2014). Par ailleurs, l'excès de DOPA peut aussi générer des radicaux *o*-semiquinones via une réaction de dismutation du groupement catéchol. La densité électronique du catéchol se réorganise pour donner la forme *o*-quinone qui polymérise rapidement. Cette réaction est d'ailleurs visible puisque les agrégats se formant produisent une coloration rouge-brune.

Dans une expérience parallèle, la réaction a été réalisée avec une quantité variable de tyrosinase, avec différents temps de réaction. À partir des tests menés dont les résultats sont présentés dans la Figure 6.2B, la quantité de tyrosinase ajoutée lors de la réaction semble être le facteur affectant significativement le rendement de la réaction, plus que le temps d'incubation. Les incubations faites avec 1 et 2 μg ont présenté peu de variations entre un temps d'incubation de 10 et 60 minutes. Inversement, une augmentation majeure a été notée pour les échantillons incubés avec 5 et 10 μg de tyrosinase lorsque le temps d'incubation passe de 10 à 20 minutes.

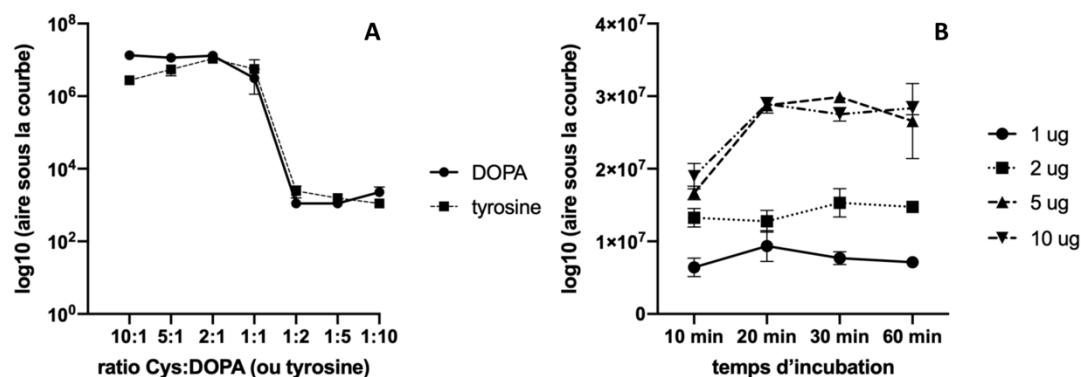


Figure 6.2 Abondance du Cys-DOPA lorsqu'incubé à différents ratios de cystéine et DOPA (A) et avec une quantité croissante de tyrosinase (B). La réaction conduite avec la tyrosine est également présentée dans les graphiques.

Par conséquent, une incubation de plus de 20 minutes n'offre pas un rendement supérieur. Des niveaux comparables de Cys-DOPA ont été obtenus avec 5 et 10 μg de tyrosinase, ce qui sous-entend que la réaction semble atteindre un plateau au-delà de ces quantités.

En plus de la DOPA, la dopamine et le DOPAC, qui sont issus du métabolisme de la DOPA, peuvent également former des liaisons covalentes avec les cystéines libres. Afin de comparer la réactivité à laquelle la cystéine est modifiée, la dopamine et le DOPAC ont été incubés dans les mêmes conditions expérimentales (i.e. 5 μg de tyrosinase durant 20 minutes). Des incubations supplémentaires avec la tyrosine et la phénylalanine ont été réalisées en parallèle dans le but d'évaluer leur aptitude respective à conduire à la DOPA. Tous les adduits ont été identifiés avec une excellente exactitude sur la masse. Une première observation peut être faite à partir de la figure ci-dessous : le Cys-DOPA n'est pas formé lorsque la phénylalanine est employée

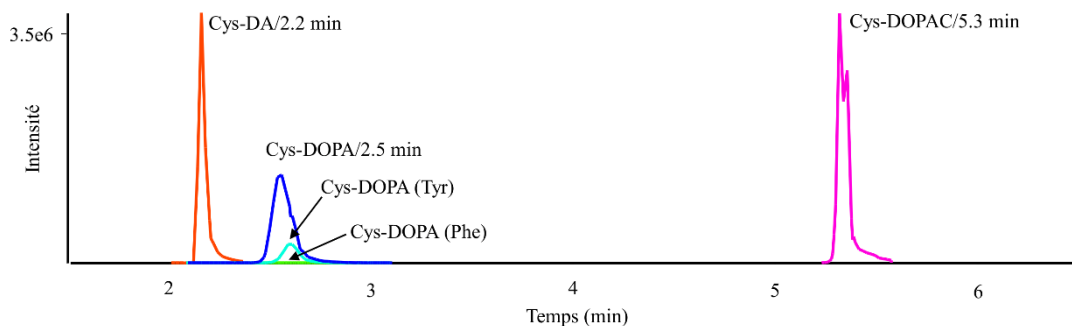


Figure 6.3 Chromatogramme d'ion extraits présentant la Cys-DOPA (m/z 317.08), la Cys-DA (m/z 273.09) et la Cys-DOPAC (m/z 288.05) incubé avec un ratio catéchol :Cys de 2 :1 dans les conditions optimales.

comme substrat. Contrairement aux incubations faites avec la phénylalanine, le Cys-DOPA a été détecté à partir des incubations où la tyrosine était employée comme

substrat cependant, en plus faible intensité que pour les incubations faites avec la DOPA. Il est toutefois intéressant de noter que la présence du groupement carboxylique ne semble pas affecter l'efficacité de l'ionisation, d'autant plus que le signal est comparable à celui mesuré pour la Cys-DA.

Lorsque les valeurs d'aire sous la courbe sont comparées pour les trois composés

(

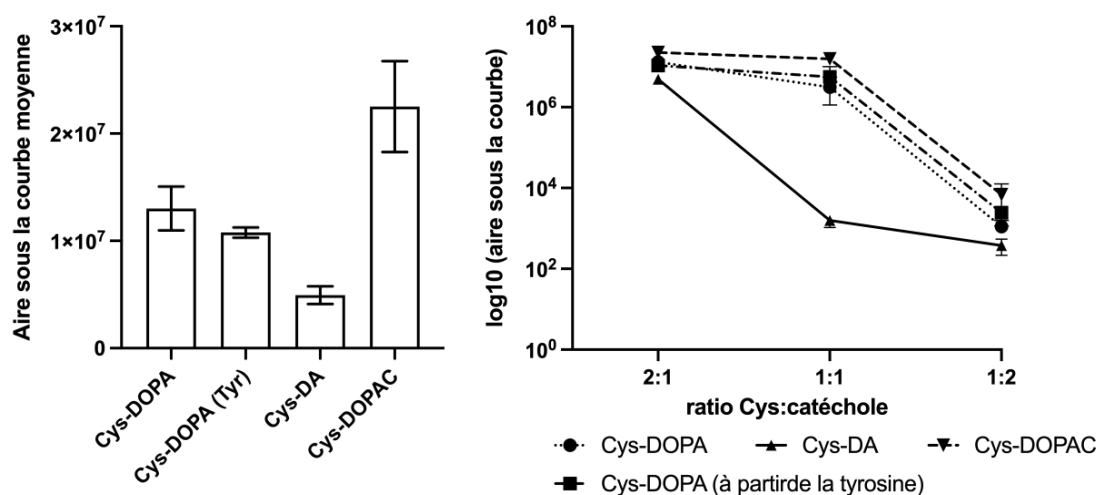


Figure 6.4, à gauche), la Cys-DOPAC est vraiment plus élevée que pour la Cys-DOPA et la Cys-DA. Ensuite, le choix du substrat pour la tyrosinase a aussi un impact sur la quantité de Cys-DOPA formée. En effet, la tyrosine s'est montrée un substrat comparable à la DOPA. Toutefois, l'oxydation de la phénylalanine ne s'est pas produite avec la tyrosinase, supportant le fait que le substrat doit contenir au préalable un groupement hydroxyle afin d'interagir correctement avec le centre actif de l'enzyme. Dans le même ordre d'idées, l'effet suicide sur l'enzyme par la présence en excès d'un

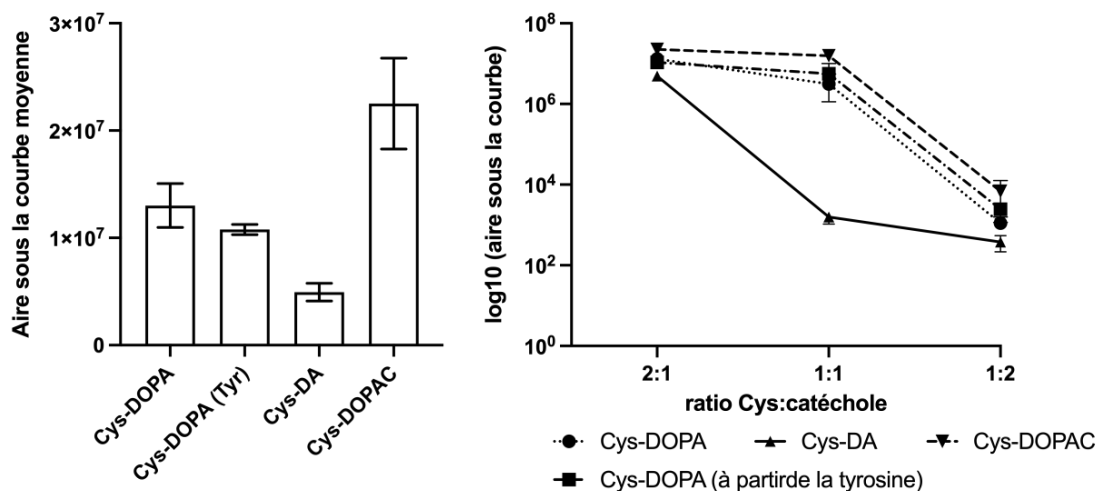


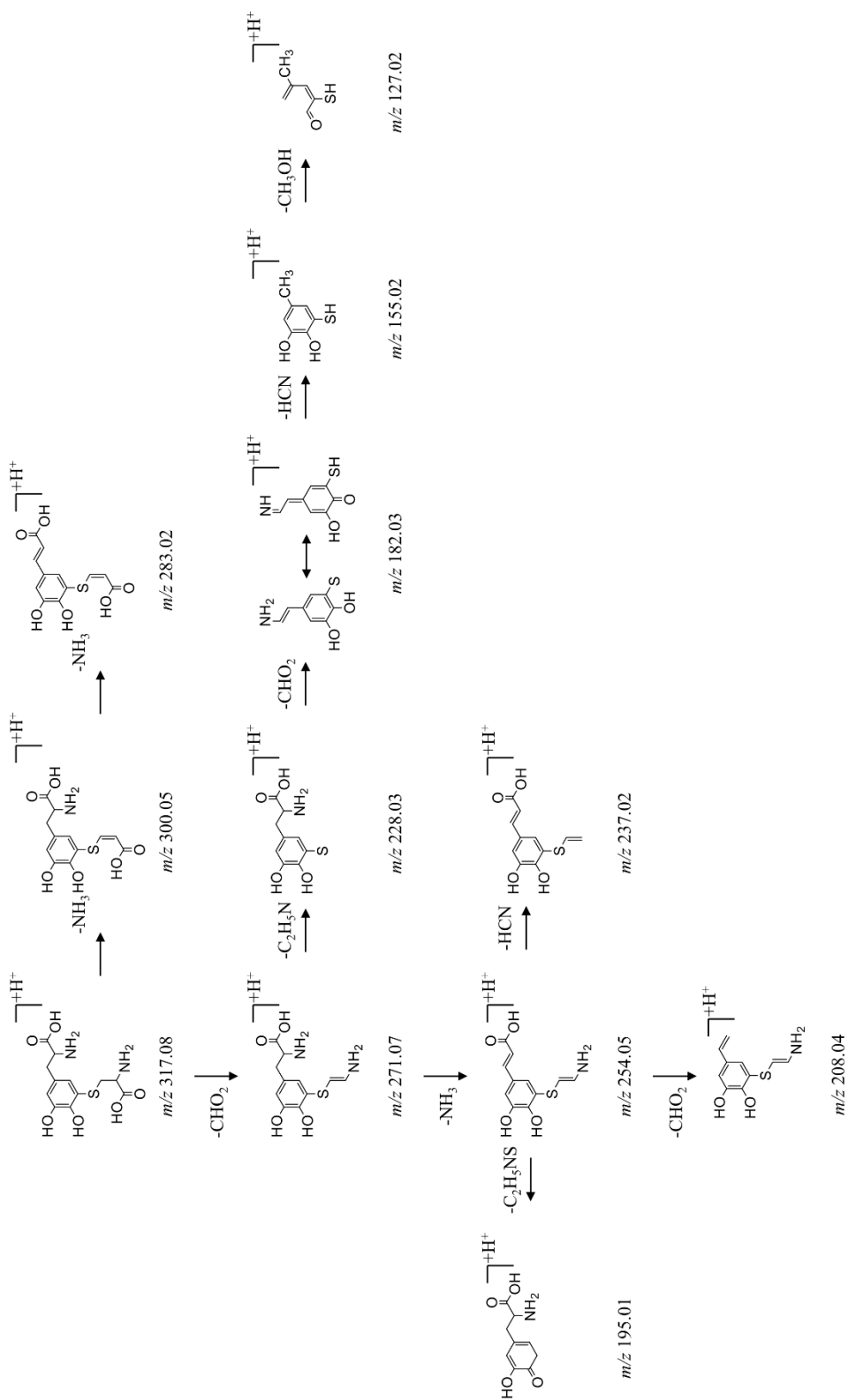
Figure 6.4 Aire sous la courbe moyenne pour Cys-DOPA, Cys-DA et Cys-DOPAC (gauche) ainsi que l'impact de l'excès de catéchol sur la réaction d'addition avec la cystéine (droite).

groupement catéchol n'est pas spécifique à la DOPA, tel que montré ci-dessus. L'absence d'un groupement carboxylique semble accélérer l'effet d'inhibition, ce qui expliquerait pourquoi seule la Cys-DA décroît plus rapidement à un ratio 1 : 1 que pour les incubations réalisées avec la DOPA, la tyrosine et la DOPAC.

D'autres incubations ont été faites avec l'histidine, la tyrosine et la lysine comme nucléophile pour remplacer la cystéine. Cependant, les tests ont révélé que seul le groupement thiol avait la nucléophilicité nécessaire pour réagir avec la quinone selon une addition de Michael. De plus, même si les réactions ont été conduites à des pH afin de favoriser la disponibilité de la paire d'électrons libres pour la lysine et l'histidine, aucune addition impliquant le noyau indole ou du groupe amine-ε n'a été observée.

Dans un autre ordre d'idées, lorsque les spectres MS/MS du Cys-DOPA, Cys-DA et Cys-DOPAC sont comparés entre eux (Figure 6.5), seuls quelques ions produits sont communs aux trois composés (surligné dans la figure). Un profil de fragmentation

pour la Cys-DOPA est suggéré à la Figure 6.6



L'assignation des pics pour le Cys-DOPA a été facilitée par les travaux de Zhao et Waite (Zhao & Waite, 2005). Il a été remarqué que pour le spectre MS/MS de la Cys-DA, certains des ions diagnostiques de la Cys-DOPA sont également présents, mais mesurés avec un écart de 2 Da. L'origine de l'écart de 2 Da entre certains ions produits s'explique par la différence de masse entre l'ion précurseur de la Cys-DA (m/z 273.09) et l'ion obtenu suite à une perte neutre du groupement acide carboxylique (-46 Da) provenant de la Cys-DOPA (m/z 271.07). La DOPA et la dopamine sont des composés semblables, sauf au niveau de la présence d'un groupement carboxylique. Ainsi, lorsque l'ion précurseur du Cys-DA est activé, la fragmentation suit des routes similaires à celles du Cys-DOPA, ce qui explique la similitude entre leur spectre MS/MS. Ces voies de fragmentation demeurent différentes entre les deux molécules. D'un autre côté, le Cys-DOPAC offre un spectre MS/MS bien différent de celui obtenu pour la dopamine et la DOPA. Son spectre MS/MS est prédominé par un pic de base à m/z 199.01 qui correspond à la perte neutre de la

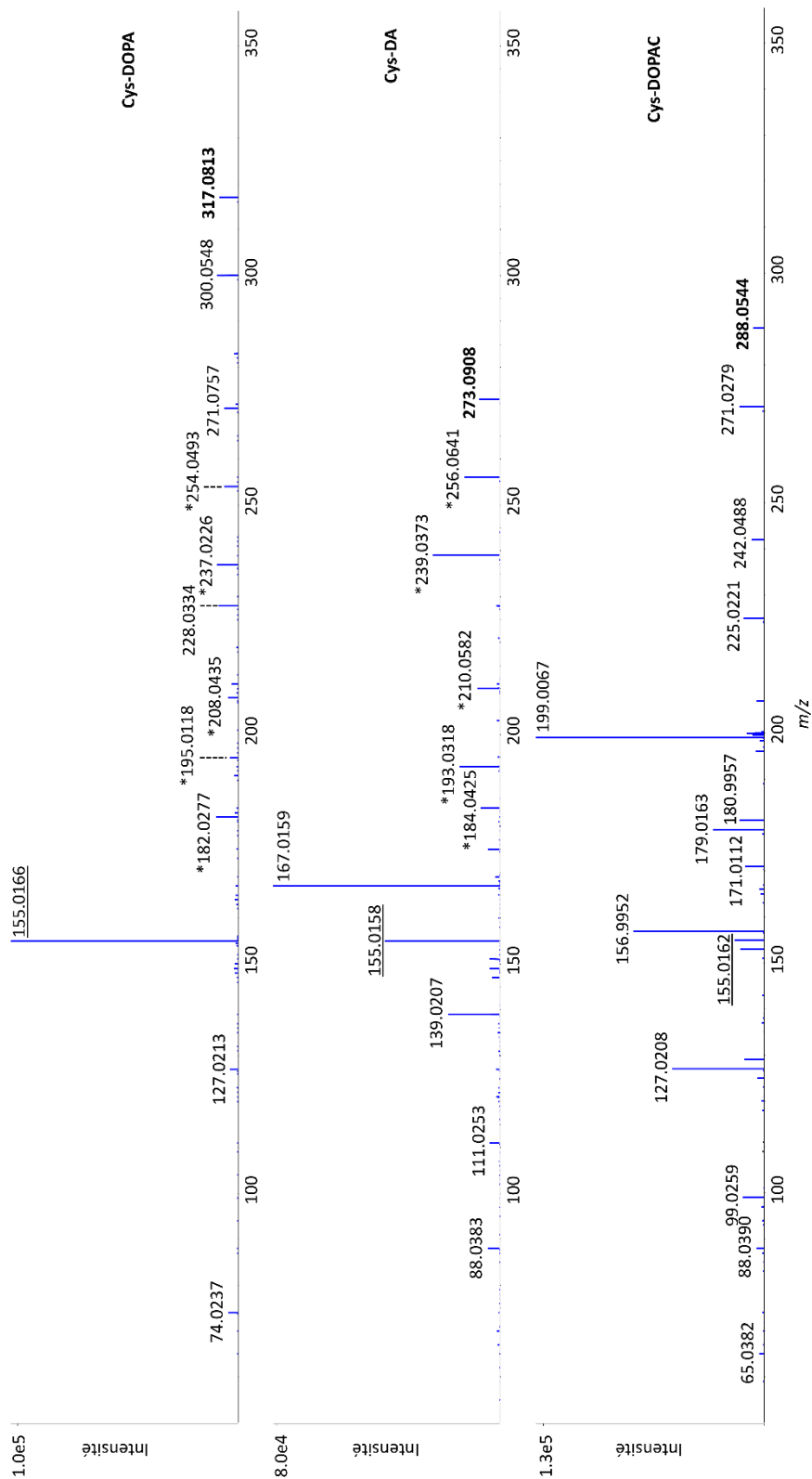


Figure 6.5 Spectre MS² de la Cys-DOPA, Cys-DA et de la Cys-DOPAC non dérivées. Les ions précurseurs pour chaque composé sont présentés en gras et les ions communs sont soulignés. Les valeurs annotées d'un astérisque furent identifiés dans deux spectres, mais avec un écart de ± 2 Da.

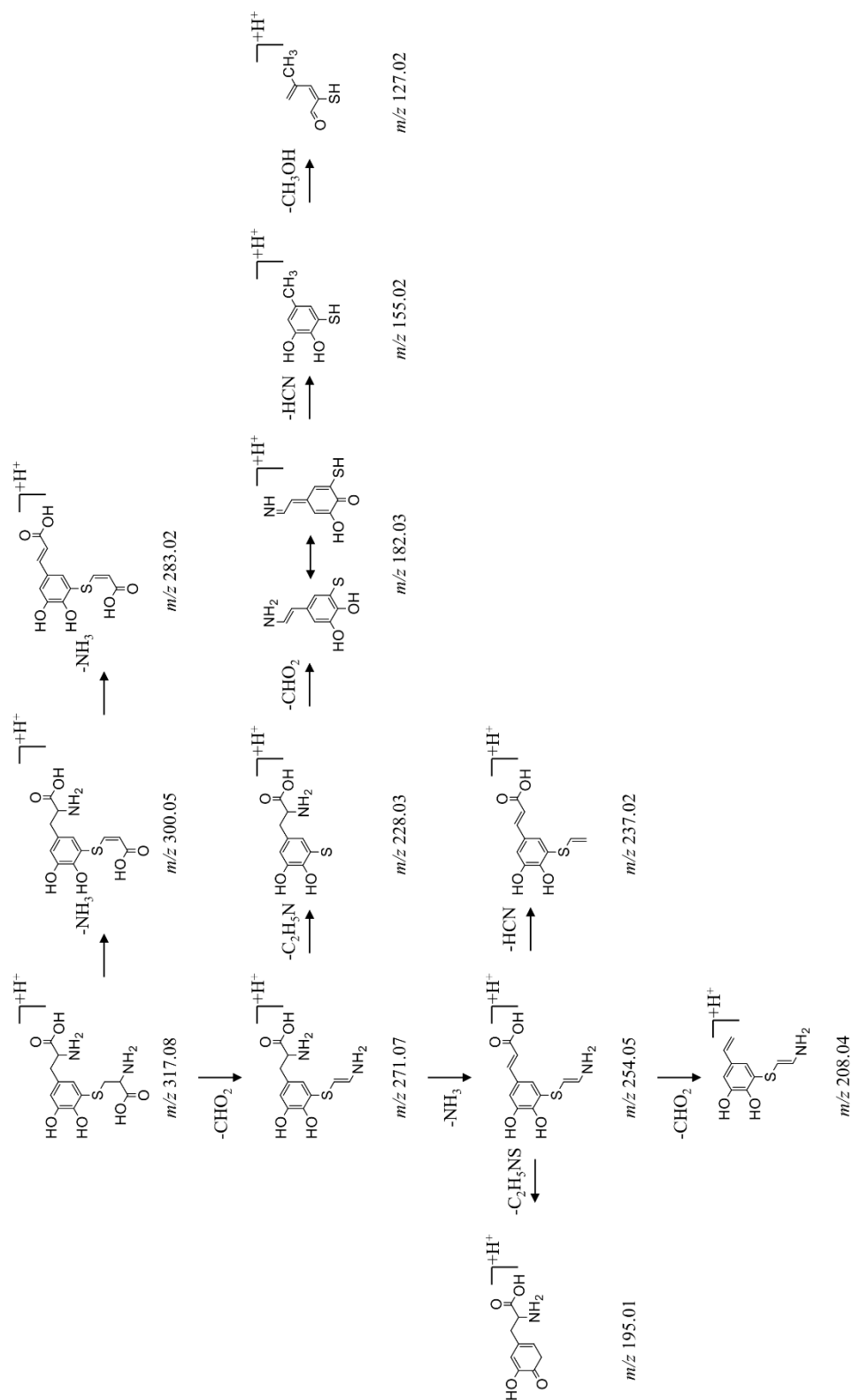


Figure 6.6 Schémas proposés de la fragmentation de la Cys-DOPA

cystéine (-89.0477 Da, C₃H₇NO₂), avec quelques pics en moindre intensité. Des schémas de fragmentation pour la Cys-DA et la Cys-DOPAC sont proposés en annexe.

6.4.2 Élucidation des régioisomères du Cys-DOPA

Durant l'addition de Michael entre la cystéine et l'ortho-quinone, deux isomères sont possibles, l'un à la position 5-*S* et le second à la position 2-*S* (Figure 6.7).

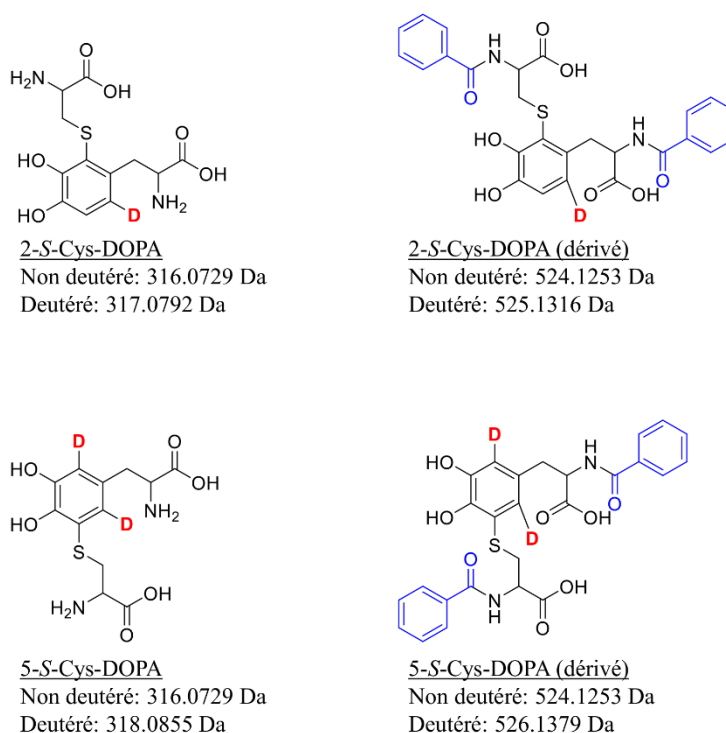


Figure 6.7 Structures et équivalents deutérés de la Cys-DOPA, présentant les isomères 5-*S* (à gauche) et 2-*S* (à droite) obtenus à partir de la tyrosine deutérée lorsque celle-ci réagit avec les groupements thiols libres des protéines (ou du glutathion).

La littérature disponible souligne toutefois une préférence pour l'isomère 5-*S* et le présente comme étant le composé prédominant. (Huang, Xu, Hawley, Hopkins, & Kramer, 1998; Jameson, Zhang, Jameson, & Linert, 2004; Shen, Xia, Wrona, &

Dryhurst, 1996). Néanmoins, ces affirmations se fondent sur l'utilisation de techniques analytiques peu spécifiques en lien avec la distinction d'isomères. En fait, il s'agit de composés isobariques (de même masse), ce qui rendra difficile de les discerner par les approches génériques en chromatographie et spectroscopie classiques.

Par conséquent, une approche simple pour permettre leur séparation (au point de vue de la spectrométrie de masse) est l'utilisation de composés isotopiques stables. L'incorporation de ces isotopes (^{13}C , ^2H ou ^{14}N) aura un effet négligeable sur la chromatographie, mais apportera une augmentation de la masse caractéristique du composé étudié. Ainsi, à l'aide d'instruments à haute résolution, il sera désormais possible de faire la distinction entre l'isomère 5-S et l'isomère 2-S.

Un premier essai utilisant un étiquetage isotopique stable (*stable-isotope labeling*) a permis de différencier ces deux isomères, selon deux scénarios. Dans le premier, l'isomère 5-S a été obtenu lorsque les deux atomes de deutériums restent en place au moment de l'oxydation par la tyrosinase (Figure 6.7, coin supérieur gauche). Dans ce cas, un écart de +2 Da est observé sur le m/z mesuré. Dans le second scénario, si l'un des deux atomes de deutérium est retiré au moment où la cystéine réagit avec la quinone, un écart de masse de +1 Da sera mesuré, ce qui correspond à l'isomère 2-S.

Bien qu'il soit possible de pouvoir les différencier à partir de leur m/z respectif, ces isomères ont les mêmes propriétés physico-chimiques et ne peuvent donc pas être séparés sur une colonne chromatographique. À cet égard, un deuxième essai a été réalisé en employant une méthode de dérivation chimique développée par Wagner et al. (Wagner et al., 2015). Cette méthode permet sélectivement d'ajouter un groupement succinimide sur les amines primaires et ainsi d'augmenter la rétention sur une colonne classique C18. Puisque les isomères 2- et 5-S n'ont pas la même conformation finale, ils n'interagiront pas de la même manière avec la colonne et pourront ainsi être séparés. Sur la base des spectres de masse à haute résolution, le chromatogramme d'ions extrait

des deux espèces a été obtenu à partir de leur ratio masse-sur-charge respectif
(

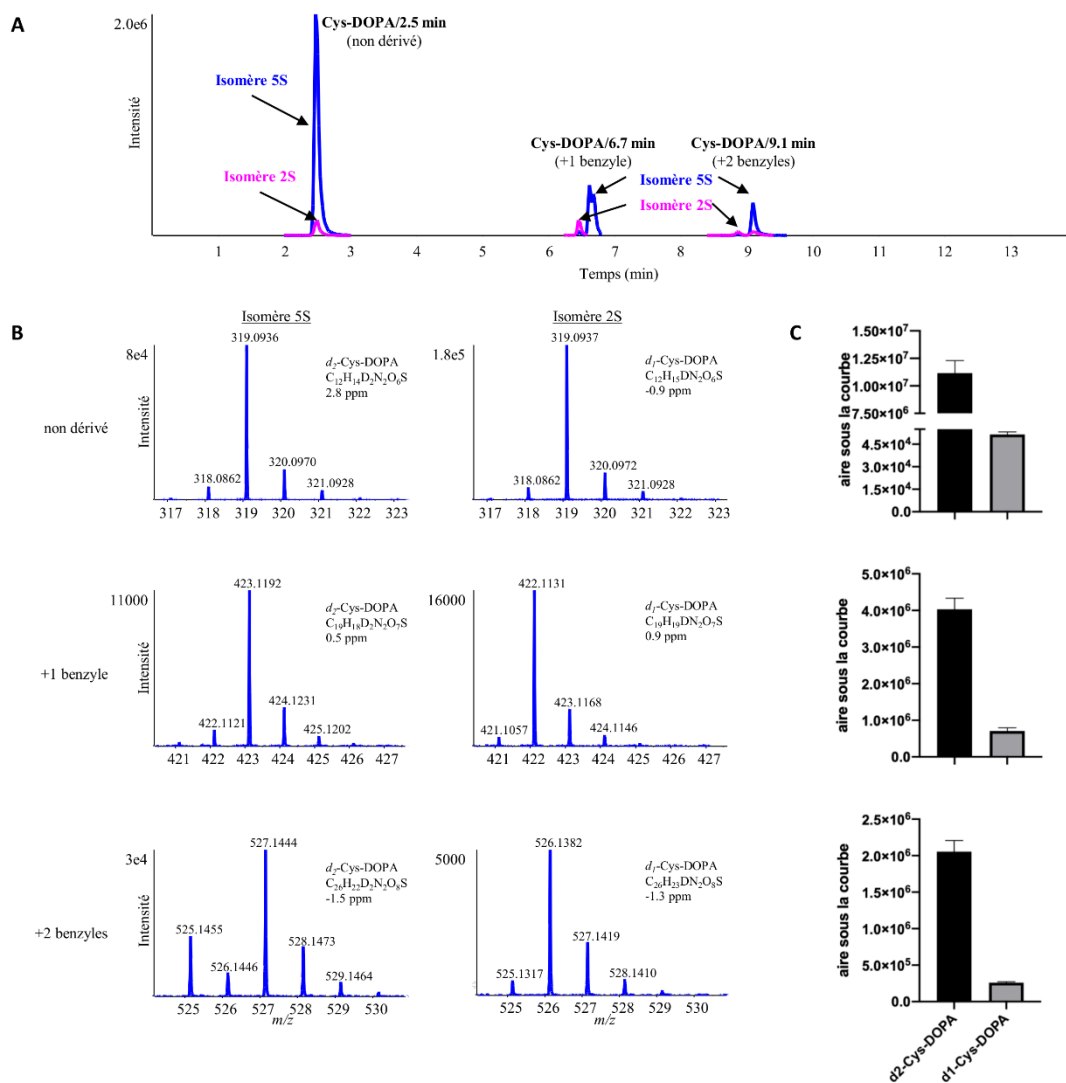


Figure 6.8). Tel qu'illustré par la

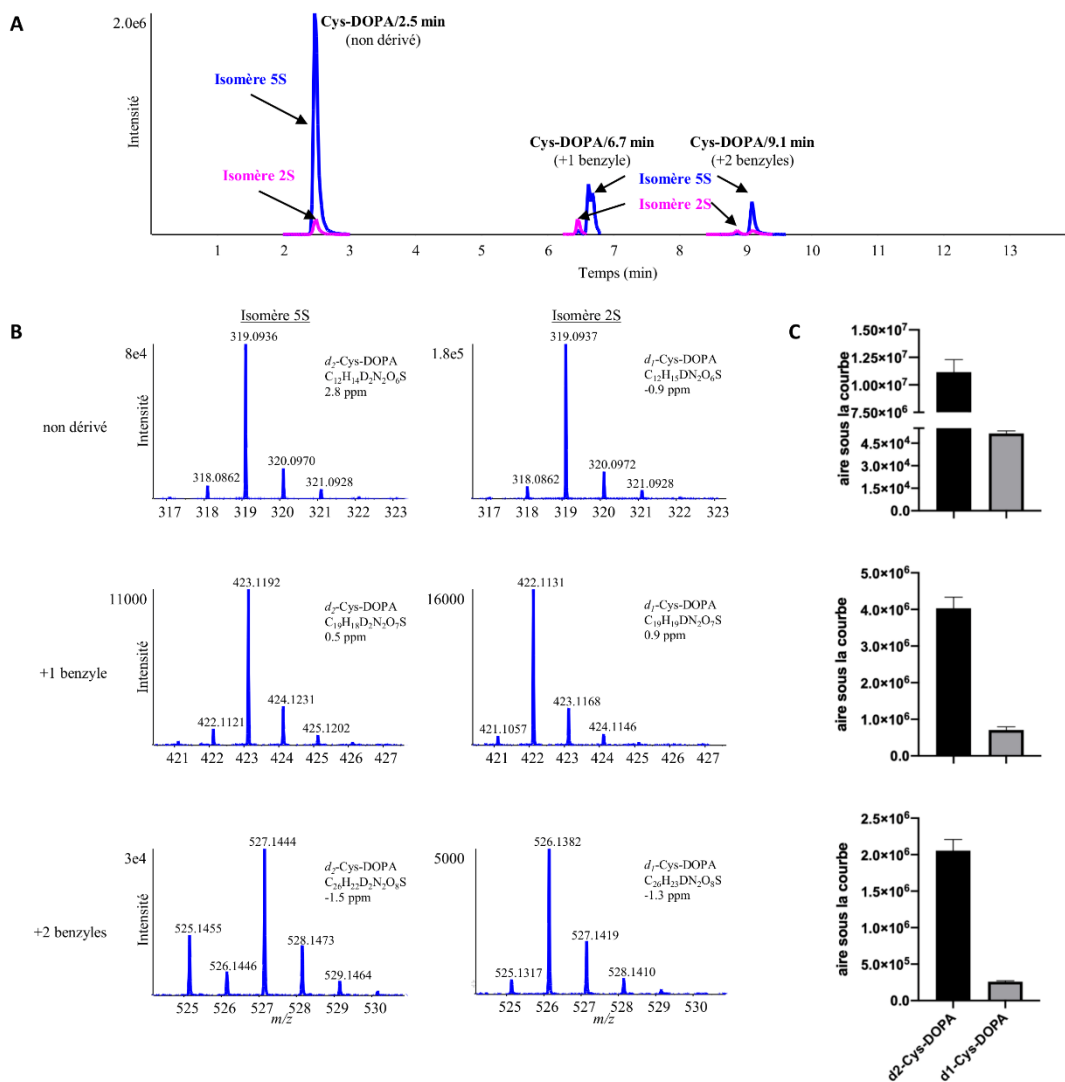


Figure 6.8, l'abondance relative de l'isomère 2-S par rapport

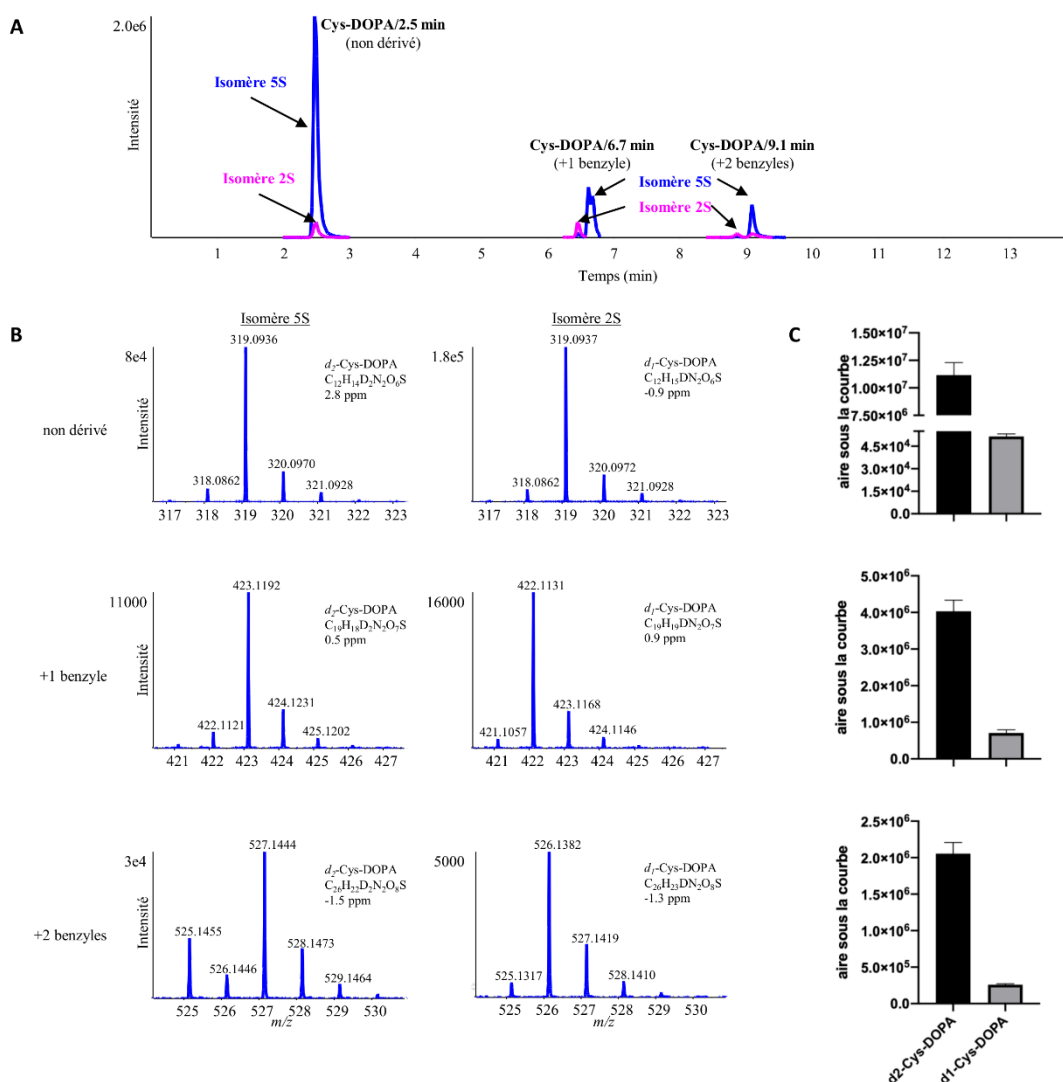


Figure 6.8 Chromatogramme d'ion extrait pour les isomères 5-S (en bleu) et 2-S (en rose) pour les versions non dérivées, dérivé avec un groupement benzyle, puis dérivé avec deux groupement benzyle. Les valeurs d'aire sous la courbe pour chaque composé sont présentées dans le diagramme à barres sur la droite.

à l'isomère 5-S est clairement négligeable. Cette tendance est aussi observée lorsqu'une dérivation chimique pré-colonne est employée.

6.5 Conclusions

La réaction entre une o-quinone et la cystéine demeure un point essentiel pour l'étude de liaisons croisées au sein des protéines riches en DOPA et cystéine présentes dans la fibre de byssus. Le ratio entre la cystéine et la DOPA a été un élément essentiel dans l'optimisation des paramètres d'incubation. Entre autres, l'excès de groupement thiol prévient l'inhibition suicide causée par un surplus de DOPA. Parallèlement, d'autres composés comportant un centre catéchol, issu du métabolisme de la DOPA et de la dopamine, ont montré eux aussi une excellente réactivité en présence de cystéine. Néanmoins, les résultats n'ont pas été concluants quant à la réaction entre des lysines et histidines libres avec la dopaquinone. Pour réagir avec la dopaquinone, l'environnement local dans lequel se trouveront les résidus lysine et histidine est le principal facteur qui influencera la réactivité des chaînes latérales.

L'utilisation de la tyrosine deutérée a singulièrement permis l'identification de deux isomères structuraux possibles (2-*S* et 5-*S*) avec l'isomère 5-*S* identifié comme étant l'isomère majeur. Cet ajout est un avantage majeur puisque jusqu'à ce jour, l'identification des deux isomères a souvent été biaisée ou faite avec des approches méthodologiques qui permettaient une identification rudimentaire. Néanmoins, l'identification de liaisons croisées à partir d'échantillons de byssus devra être abordée lors de travaux futurs.

CHAPITRE VII

CONCLUSIONS ET PERSPECTIVES

L'un des aspects déterminant et décisif en chimie bio-analytique demeure la préparation des échantillons. Tout au long de la présente dissertation, notre travail a démontré la pertinence de chacune des étapes lors de la préparation des échantillons. En ce sens, l'utilisation de méthodologies par chromatographie multi-dimensionnelle est généralement efficace pour l'augmentation de la couverture du protéome. Néanmoins, des pertes peuvent être encourues lors de la manipulation des fractions collectées et par conséquent, cette méthodologie s'avère peu favorable lors d'essais quantitatifs.

Le profilage de protéines provenant d'organismes peu étudiés, comme les moules, pose un défi considérable. Aussi, l'utilisation de bases de données protéomiques peu ou pas annotées complique l'identification des protéines. Par conséquent, la vérification des protéines identifiées demande une attention spécifique, principalement en ce qui a trait à 1) au contrôle du pourcentage FDR adéquat appliqué aux peptides ainsi qu'aux protéines, 2) le nombre de peptides utilisés pour l'identification de la protéine en question et 3) l'unicité des peptides détectés.

La présente dissertation a mis de l'avant, d'un point de vue analytique, les considérations à prendre pour maximiser l'étude du protéome, tout organisme confondu. Les sections suivantes résument les principales avancées réalisées et proposent des avenues intéressantes à explorer lors d'expériences futures.

7.1 Résumé des découvertes

Dans un premier temps, le chapitre 2 a montré l'impact d'un pré-fractionnement par SCX et RP (à pH 10) sur la couverture globale du protéome en comparaison avec une approche directe. Traditionnellement, la SCX est utilisée lors d'étape de collecte de peptides en raison de son importante complémentarité avec la séparation en phase inverse directement couplée à la spectrométrie de masse (Washburn et al., 2001). Le spectre de masse de l'ion précurseur est clairement de meilleure qualité à l'issue d'approches par pré-fractionnement (Figure 2.4). Il est donc favorable, pour la sélection de l'ion précurseur, de procéder à une étape de pré-fractionnement. Malgré une augmentation minime de la couverture du protéome, l'approche sans pré-fractionnement offrait des résultats comparables, bien qu'inférieurs, aux approches par chromatographie multi-dimensionnelle. L'un des facteurs décisifs autre que le nombre total de protéines a été le nombre de peptides par protéine, qui demeure sensiblement le même pour les trois approches. Aussi, le pourcentage de couverture des protéines est même légèrement inférieur pour les approches SCX et RP. Le nombre d'échantillons influence aussi la décision d'inclure ou non une étape supplémentaire de fractionnement. Hinzke et al. ont obtenu le même constat dans leurs travaux, c'est-à-dire que malgré un plus grand nombre de protéines identifiées par 2D-LC, le choix d'utiliser une approche 1D-LC avec une colonne et une chromatographie plus longue amélioreraient considérablement la productivité des analyses (Hinzke et al., 2019).

Notre plus récente publication (Sansoucy et al., 2020) a montré l'efficacité d'une approche par 1D-LC-MS/MS dans l'identification de protéines potentiellement impliquées dans la biosynthèse du byssus chez les moules *Mytilus edulis* et *M. californianus*. Une base de données composée de protéines associées essentiellement au genre *Mytilus* a été employé. Néanmoins, l'identification de ces protéines a posé un défi considérable compte tenu du fait que les bases de données actuellement disponibles renferment peu d'annotation sur les protéines qu'elles contiennent. Par conséquent,

l'utilisation de critères plus restrictifs (peptides à 1% FDR, protéines identifiées avec un minimum de 2 peptides) a permis de produire une liste de protéines plus susceptibles d'être présentes dans les échantillons de moules et ultimement, d'être associées à la byssogénèse. Notre approche a aussi fourni des informations physico-chimiques supplémentaires, telles que le pourcentage en résidus tyrosine, cystéine et acide afin de relier certaines protéines à des processus importants comme l'adhésion (protéines riches en DOPA) ou à des fonctions anti-oxydantes (protéines riches en cystéines). Finalement, l'identité de 13 protéines initialement considérées comme étant non caractérisées a été présentée pour la première fois. Certaines protéines importantes, comme les mfp riches en DOPA, n'ont malheureusement pas été identifiées par cette approche. D'autres expériences visant à la détection de ces protéines sont actuellement en cours.

Afin de palier au manque d'annotation de la base de données protéomique, une base de données a été créée à partir du transcriptome de la moule *Mytilus edulis* et utilisée à des fins de comparaison. Les résultats présentés au chapitre 4 ont montré une bonne corrélation avec les résultats présentés au chapitre 3, avec plusieurs protéines nouvellement identifiées en tant que possibles variantes. Aussi, les analyses en triplicata pour chaque glande ont montré une bonne répétabilité, avec de faibles coefficients de variation. À l'exception de deux protéines qui ont été identifiées dans le manteau de la moule (tissu contrôle), 43 ont été considérées comme étant uniques à l'organe producteur de byssus, selon les résultats obtenus à partir de la banque de protéines déduites du transcriptome. D'autre part, les analyses individuelles des trois glandes du pied de la moule ont grandement contribué à l'amélioration de la couverture du protéome en identifiant les protéines uniques à chaque glande. Cette étude se montre complémentaire aux travaux précédemment présentés et publiés. Cette complémentarité est montrée au tableau 4.1 avec un exemple de quelques protéines identifiées seulement à partir de la banque de donnée protéomique ou transcriptomique.

Malheureusement, aucune des protéines riches en DOPA (mfp-1, 3 et 5) n'a été identifiée à l'aide de la base transcriptomique à ce moment.

Puisqu'aucune des approches précédentes n'a permis l'identification des protéines riches en DOPA, une méthodologie employant une version recombinante de la mfp-1 a été développée en parallèle. Cette protéine demeure relativement complexe à extraire, probablement à cause de la présence de liaisons croisées entre les chaînes latérales nucléophiles et la DOPA. Par conséquent, l'un des objectifs visait à connaître les conditions expérimentales favorisant la forme réduite de la DOPA, afin de minimiser les réactions croisées. Une concentration de quelques millimolaires en acide ascorbique a été jugée suffisante pour maintenir la DOPA sous sa forme réduite, sans risque pour la digestion enzymatique. Un autre avantage de notre méthode est qu'en raison de l'absence de résidus cystéine, la rMFP-1 a été digérée immédiatement par la trypsine sans devoir procéder à l'étape de réduction alkylatrice. Grâce au séquençage *de novo* du spectre MS/MS du décapeptide, nos résultats ont montré que la tyrosine du côté C-terminal (Tyr-9) était majoritairement oxydée en DOPA. La présence d'une seconde DOPA, sur la tyrosine N-terminal (Tyr-5) a également été rapportée, mais en moins grande proportion. Finalement, les mesures réalisées sur notre appareil à haute résolution ont montré, pour la première fois, l'existence d'une liaison croisée impliquant la lysine (Lys-2) et la DOPA pour former un lien intra-moléculaire. Dans le cas des lysines, la réaction entre l'amine primaire et la quinone nécessite des conditions particulières favorisant la nucléophilicité de la chaîne latérale (Labenski et al., 2009). Néanmoins, la chimie derrière ce type d'addition nucléophile demeure encore à clarifier (J. Yang, Cohen Stuart, & Kamperman, 2014). D'un autre côté, la stabilité de l'adduit demeure au centre de la détection de ce dernier. Par conséquent, la formation d'une base de Schiff, comme l'ont proposé Priemel et al. (Priemel et al., 2020) est donc peu probable, sachant que celle-ci est rapidement fragmentée lors de l'étape d'activation de l'ion précurseur. Le séquençage *de novo* performé sur ce peptide a montré la série d'événements durant la fragmentation et a produit plusieurs ions

diagnostiques qui seront nécessaires pour une identification à partir d'échantillons de byssus.

Les travaux présentés au chapitre 6 ont démontré la contribution de la spectrométrie de masse à haute résolution dans l'élucidation de liaisons covalentes entre les groupements catéchols et thiols. Aussi, ce chapitre a apporté une meilleure compréhension de la chimie de la DOPA. L'approche analytique développée visait essentiellement à l'établissement des conditions expérimentales favorisant la réaction entre la cystéine et la DOPA. En cours de route, l'utilisation d'un équivalent deutéré de la tyrosine a permis d'étudier un peu plus en détails la présence de deux isomères de position de la Cys-DOPA qui, jusqu'à présent, avait été largement négligé. Nos travaux ont montré que l'isomère 5-S de la Cys-DOPA était bel et bien l'isomère prédominant et que dans les conditions expérimentales actuelles, une quantité négligeable de l'isomère 2-S était obtenue. Dans un cadre appliqué aux protéines, la présence de la Cys-DOPA demeure un atout majeur puisque cette dernière est mentionnée par Waite (J. Herbert Waite, 2017) et Yu (Yu et al., 2013) comme altérant les propriétés bio-mécaniques et adhésives de la fibre de byssus.

7.2 Perspectives

Une avenue intéressante concernant les résultats présentés aux chapitre 3 et 4 serait de développer l'aspect quantitatif de l'étude du protéome de la moule *M. edulis*, à partir des extraits de pieds et de manteaux. En effet, ce manque a essentiellement été responsable du retrait de plusieurs protéines intéressantes, comme les protéines twitchine, catchine et LIM, partagées par le pied et le manteau des moules à l'étude. N'ayant pas d'information disponible quant à l'abondance de ces protéines dans leur tissu respectifs, il était difficile de relier ces protéines au processus de byssogénèse.

Bien qu'intéressantes, les méthodes quantitatives en protéomique *bottom-up* restent largement dépendantes de la complexité des échantillons étudiés (Domon & Aebersold, 2010). Une approche par étiquetage d'affinité codée par isotopes (ICAT, *Isotope-Coded Affinity Tag*) propose une alternative intéressante pour la quantification des protéines en général, mais également pour la quantification des protéines de faible abondance (Gygi, Rist, Griffin, Eng, & Aebersold, 2002). Également, avec des conditions de pH ainsi qu'une chimie similaire avec l'iodoacétamide employée lors de l'étape de réduction alkylatrice, l'incorporation du composé ICAT dans la procédure actuelle ne devrait pas poser de défi technique supplémentaire lors de la préparation des échantillons (Figure 7.1). Une fois modifiées, les protéines seraient digérées par la trypsine selon la même procédure utilisée tout au long de cette dissertation. La figure suivante présente le flux de travail proposé pour une application en protéomique quantitative chez *Mytilus edulis*. Cependant, le couple de peptides ICAT léger/lourd peut ne pas être adéquatement superposé durant la chromatographie lorsque le composé deutéré est utilisé, compliquant parfois l'identification des peptides intéressants. Hansen et al. ont d'ailleurs proposé une solution avantageuse en employant une version du réactif ICAT marqué au carbone 13 au lieu de deutérium (Hansen et al., 2003).

Cette avenue constituerait, au meilleur de nos connaissances, la première application quantitative en protéomique *bottom-up* chez les moules. Par conséquent, l'implantation d'une méthodologie quantitative apporterait 1) une meilleure connaissance de l'abondance des protéines identifiées dans le pied et le manteau de la moule, à partir d'une sélection de protéines intéressantes pour la byssogénèse, 2) l'approfondissement du protéome de la moule *M. edulis* en incluant à l'étude des protéines peu abondantes initialement ignorées par les approches précédentes et finalement, 3) la transposition de cette approche aux protéines de byssus afin de mesurer les différences entre les moules *M. edulis* et *M. californianus*.

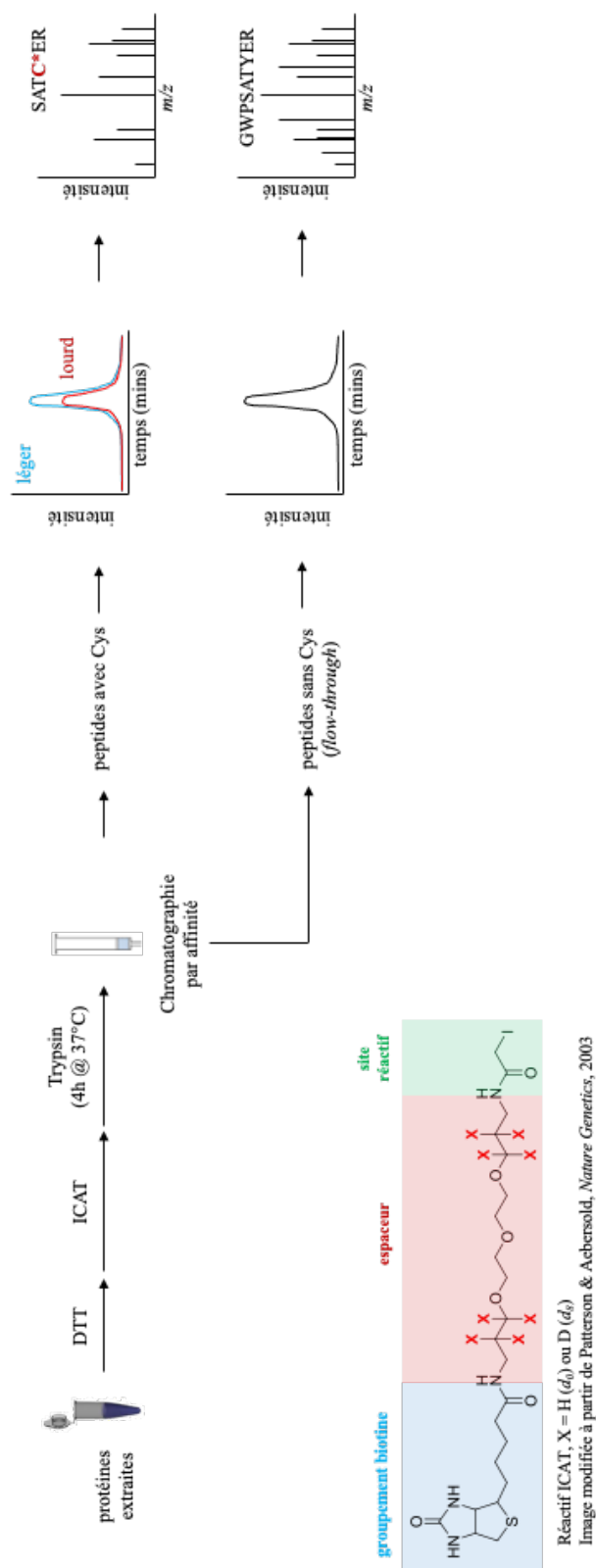


Figure 7.1 Flux de travail proposé pour une approche quantitative sur les protéines de moule *M. edulis* utilisant le réactif ICAT afin de marquer les protéines possédant au minimum un résidu cystéine. Les protéines seront digérées par la trypsine dans des conditions similaires que pour les études qualitative. Les peptides n'ayant pas été marqué par ICAT seront collectés et analysés également

En ce qui concerne les travaux présentés en lien avec la protéine rMFP-1, il est clair que la présence de liaisons croisées nuit considérablement à la détection de cette protéine provenant de la fibre de byssus ainsi que de la glande cuticule qui la produit. Lors de la préparation des échantillons de glandes, un culot a été observé après le traitement avec la solution d'urée et de thiourée. La formation de liaisons croisées au sein de la protéine mfp-1 et ce, même en faible quantité, peut causer l'agrégation de la protéine et expliquer le culot observé. Une modification intéressante à apporter au protocole actuel serait de réaliser une extraction en milieu acide, comme l'ont proposé Waite et Qin en 2001 (J. H. Waite & Qin, 2001) et Danner et al. en 2012 (Danner, Kan, Hammer, Israelachvili, & Waite, 2012) pour l'identification de la mfp-5, la protéine la plus riche en DOPA. À quelques différences près, les auteurs ont utilisé des conditions acides (5% acide acétique) et dénaturantes (8 M urée, 6 M Guanidine-HCl) afin de favoriser l'extraction. L'utilisation de plusieurs agents chaotropiques combinés peut s'avérer problématique en prévision de la digestion par la trypsine.

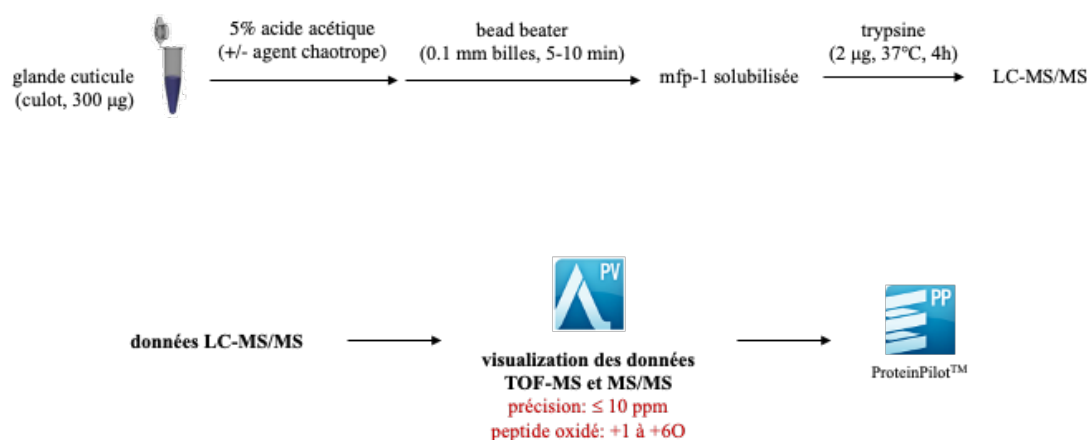


Figure 7.2 Flux de travail proposé pour l'identification de la mfp-1 à partir du culot présent dans les échantillons de la glande cuticule. L'utilisation d'agents chaotropiques comme l'urée, la thiourée et le chlorure de guanidinium seront à l'étude ainsi que le temps nécessaire à l'homogénéisation du culot.

Comme le protocole actuel ne prévoit pas d'étapes supplémentaires visant leur retrait, comme une dialyse, il sera important de s'assurer que leur utilisation n'entrave pas les étapes subséquentes. Dans l'optique de solubiliser le culot contenant potentiellement la mfp-1, un appareil de type *bead beater* sera utilisé pour la mise en solution du culot. Les échantillons seront mélangés avec une solution d'acide acétique contenant ou non une solution d'urée) et de billes afin de favoriser la solubilisation du culot. Plusieurs diamètres (0.1, 0.5 et 1.4 mm) et types (verre et céramique) de billes sont disponibles selon le type d'échantillons à homogénéiser. La Figure 7.2 ci-dessus propose un flux de travail pour de futures analyses.

Dans un autre ordre d'idées, le pré-fractionnement d'un mélange de peptides est généralement perçu comme un avantage. Une chromatographie par SCX est traditionnellement implantée dans les approches 2D-LC en raison de son importante orthogonalité avec la seconde chromatographie. Malgré certains désavantages techniques, comme l'importante concentration en sels non-volatiles, la SCX a été considérée depuis les 20 dernières années comme la méthode par excellence pour la collecte de peptides. Les travaux présentés à l'annexe C montrent l'ébauche d'une comparaison complète entre différents modes de pré-fractionnement.

La comparaison entre les différentes approches de pré-fractionnement reste actuellement à optimiser. Comme le montrent les chromatogrammes UV (Annexe C, figure 2) ainsi que les cartes de densité (Annexe C, figure 3), pour la collecte de 16 fractions en phase inverse et par SCX, plusieurs fractions semblent ne pas contenir de peptides. Ce manque d'homogénéité dans la collecte de fractions est aussi observé lorsque plusieurs sont mélangées dans l'objectif d'améliorer l'orthogonalité. Dans le cas d'un pré-fractionnement par Lp-RP, par exemple, il ne serait pas juste de croire que l'augmentation de l'orthogonalité soit réellement causée par la combinaison de fractions. Selon la méthode proposée à la section expérimentale de l'annexe C, le mélange de la première fraction (sans peptide) et la neuvième (avec peptides) biaise le

véritable impact de la combinaison de fractions. Par conséquent, quelques modifications impliquant soit 1) de commencer la collecte de fractions au moment où les peptides éluent de la colonne (à environ 10 minutes), 2) de débiter la chromatographie à une concentration en phase organique plus importante ou finalement, 3) d'adapter le gradient linéaire composé actuellement utilisé de manière à éluer les peptides plus rapidement.

L'adaptation de la fenêtre de chromatographie semble l'option la plus simple en ce moment, puisque le comportement de l'élution des peptides dans les conditions de gradient et de débit actuel est connu. Cet ajustement demanderait de revoir le temps alloué par fraction (actuellement environ 1 min/fraction) afin de couvrir correctement la zone d'élution. Dans le même ordre d'idées, la durée totale du gradient devra aussi être ajustée de manière à ce que le volume collecté par fraction soit acceptable. Parallèlement, Boichenko et al. ont également observé que la majorité des peptides détectés par Hp-RP proviennent des fractions collectées lorsque la composition en phase organique devenait plus importante, avec peu de peptides récoltés dans les premières fractions (Boichenko, Govorukhina, Van Der Zee, & Bischoff, 2013). Débiter la chromatographie à des pourcentages en phase organique légèrement inférieurs à ceux nécessaires pour l'élution des peptides (par exemple 30-35%) augmenterait grandement l'utilisation de l'espace chromatographique. Ainsi, le nombre de fractions possédant des peptides est plus susceptible de représenter le nombre de fractions collectées. La collection de peptides pourra donc débiter dès les premiers instants de la séparation, ce qui diminuera le temps total d'analyse. Également, le retour aux conditions initiales se fera plus rapidement. Finalement, la méthode actuellement en place utilise un gradient linéaire composé, avec une composition de phase mobile débutant à 5% B pour les deux premières minutes. L'accroissement rapide (d'au moins 5%B/min) occasionnerait aussi une meilleure utilisation de la fenêtre chromatographique lors du pré-fractionnement en phase inverse

et offrirait relativement les mêmes avantages qu'énoncer avec un gradient débutant à une composition en phase mobile B supérieure à 5%.

Une fois la chromatographie optimisée, la prochaine étape visera à identifier le nombre de fractions à collecter en prévision de les combiner ensuite. Cet aspect demeure un point essentiel à étudier, puisque plusieurs groupes de recherche ont collecté un nombre important de fractions avant de les mélanger afin d'améliorer la couverture du protéome (Batth, Francavilla, & Olsen, 2014; Y. Wang et al., 2011; F. Yang, Shen, Camp II, et al., 2012; Yeung et al., 2020). Dans les résultats présentés à l'annexe C de cette dissertation, la combinaison de deux fractions non-adjacentes a été montrée comme une avenue intéressante pour les étapes de fractionnement en phase inverse. Plus précisément, dans le cas d'une collecte de fraction à pH 3, la combinaison d'au moins trois fractions semblait une voie intéressante à explorer dans l'augmentation de la couverture du protéome. Pour l'instant, le nombre de fractions collectées s'est limité à 24 et des expériences supplémentaires, tel qu'illustré ci-bas, permettraient de connaître le nombre maximal de fractions à collecter et combiner afin d'obtenir un compromis dans l'amélioration de la couverture du protéome.

Le choix de l'algorithme de recherche serait aussi un élément à considérer lors d'expérience futures. La qualité des spectres MS/MS ainsi que la présence de modifications post-traductionnelles sont généralement les facteurs affectant l'identification des protéines. Une autre barrière significative lors de l'identification des protéines est de pouvoir corréler la masse du peptide précurseur avec une liste de peptides obtenus *in silico*. Cette liste peut ou non inclure des modifications post-traductionnelles ainsi que les possibles mauvais clivages. L'algorithme de recherche Paragon, développé par Sciex, a été le principal outil aidant à l'identification des protéines au cours des différents projets de recherches présentés dans cette dissertation. Cet algorithme a la capacité d'adapter l'effort de recherche sur certaines régions du peptide lors du séquençage *de novo*, ce qui est un point particulièrement bénéfique pour

l'identification des peptides portant des modifications post-traductionnelles (Shilov et al., 2007). De nos jours, l'utilisation d'un seul algorithme de recherche est probablement peu approprié pour des études protéomiques plus avancées. Idéalement, comme pour les approches chromatographiques, l'utilisation d'un minimum de deux algorithmes complémentaires est hautement recommandé (Yuan, Lin, Molden, & Garcia, 2014). Par exemple, la possibilité d'un pré-traitement est envisageable, comme l'ont proposé Jagtab et al en 2012 (Jagtap et al., 2012). Plus précisément, les auteurs ont montré que l'implantation d'une étape de pré-traitement par l'algorithme Andromeda (MaxQuant) maximisait la qualité des spectres obtenus. Ces spectres ont ensuite été soumis au second algorithme (Paragon) dans le but de maximiser l'identification de peptides portant des modifications post-traductionnelles. Puisque MaxQuant emploie un algorithme qui fonctionne sur des bases différentes que Paragon, la combinaison de ces deux algorithmes complémentaires augmenterait significativement le nombre total de protéines identifiées (Yuan et al., 2014).

Une ébauche d'une telle comparaison avait été réalisée sur les protéines de la fraction S9 provenant de foie de rats. Elle impliquait les algorithmes de recherche Mascot, MaxQuant, X!Tandem (via Scaffold) et Paragon (via ProteinPilot), qui sont généralement utilisés dans les approches *bottom-up*. À l'exception de Paragon, l'une des difficultés provient de la sélection des modifications post-transductionnelles que le logiciel considère lors du séquençage *de novo*. Les logiciels restreignent à cinq le nombre de modifications à choisir en prévision de l'identification des protéines. La qualité des spectres MS/MS ainsi que la présence de modifications post-traductionnelles sont des exemples de facteurs affectant l'identification des protéines à partir d'une base de données (Lubec & Afjehi-Sadat, 2007). Par conséquent, le biais dans la sélection de ces modifications influence la réussite de l'identification d'un peptide. La comparaison montrait que chaque algorithme de recherche apportait son lot de protéines uniques.

Les aspects expérimentaux abordés dans cette dissertation ont mis les bases quant à l'application de nos méthodes analytiques dans le cadre d'une étude protéomique sur diverses espèces de moules. Néanmoins, l'identification de protéines provenant d'organismes peu étudiés posait un défi considérable. Afin de palier à ce manque, une banque de protéines associées à plusieurs espèces de moules a été employée dans le but de maximiser le nombre de candidats. Les résultats de notre étude dernièrement publiée rapportaient plusieurs protéines nouvellement identifiées expérimentalement. Notre méthodologie montre le potentiel d'identifier certaines protéines déduites de données transcriptomiques provenant d'espèces peu caractérisées. En terminant, le tableau suivant regroupe les éléments à considérer pour les prochaines analyses pour les protéines de moules et de foie de rats, ainsi que pour l'étude des protéines riches en DOPA.

Tableau 7.1 Résumé des travaux futurs pour les projets en cours.

Protéomique chez les moules
Développer l'aspect quantitatif à l'aide du marqueur isotopique ICAT pour la détection des protéines de faibles abondances. Permettre aussi de mesurer l'abondance relative de certaines protéines trouvées à la fois dans le pied et le manteau.
Approches multi-dimensionnelles par SCX et RP
Ajuster la fenêtre de collecte de fractions selon le gradient actuel Ajuster le gradient de manière à distribuer les peptides dans les 16 fractions Évaluer la combinaison de 2, 3 et 4 fractions sur l'orthogonalité
Identification de la modification DOPA chez les protéines du byssus
Procéder à l'extraction de la protéine mfp-1 à l'aide d'une solution acide Identifier les liaisons croisées impliquant les résidus cystéines ou lysines à partir des ions diagnostiques observés.

ANNEXE A

INVESTIGATING BYSSOGENESIS WITH PROTEOMIC ANALYSIS OF BYSSUS, FOOT AND MANTLE IN *MYTILUS* MUSSELS BY LC-MS/MS

Publié dans Proteomics 2020, 10 septembre 2020, sous le titre :

DOI : 10.1002/pmic.2020000014

Sansoucy M, Tremblay R, Carrington E, Marcotte I, Sleno L

Informations supplémentaires disponibles en ligne :

<https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.2020000014>

Abstract

Mussel byssus represents a fascinating class of biological materials with a unique capacity to adhere onto virtually any solid surface. Proteins expressed in byssus, the byssal-producing organ (foot) as well as mantle tissue from *M. edulis* or *M. californianus* were analysed by LC-MS/MS. The mantle was used as a control tissue to pinpoint unique proteins from the foot samples potentially involved in byssogenesis. This work represents an important step towards identifying biologically important proteins expressed in foot, as well as extending our knowledge on the byssus proteome. Considering the minimal proteomics data of the studied species, this data also contributes to a more complete description of *Mytilus edulis* and *Mytilus californianus* proteomes. All data are available via ProteomeXchange with identifier 10.6019/PXD021023.

Introduction

Byssus threads represent a biological material with unique flexibility, self-healing, and underwater adhesion properties.(Harrington & Waite, 2007; X.-X. Qin & Waite, 1998; Schmitt et al., 2015) They can also adapt to different environmental factors.(Brazee & Carrington, 2006; Seguin-Heine et al., 2014) In turbulent waters, for example, mussels modify the secondary structure of their byssal proteins to increase abrasion endurance.(Arnold et al., 2013) This fibrous material is synthesized and secreted from the animal's foot gland, a granulo-muscular tissue. It consists of a stern, a pre-pepsinized collagen core and a plaque made primarily of foot proteins, many responsible for underwater attachment. Though byssus has been the subject of many studies, byssal assembly remains a poorly described process.

These intriguing properties have motivated structural investigations addressing the protein composition of byssus for over 30 years. A limited number of proteins have

been identified in the byssus, including mussel foot proteins (mfps), precollagens and matrix thread proteins,(DeMartini et al., 2017) however, many more proteins are involved in the biofabrication process and its regulation. Mass spectrometry has been invaluable for the determination of proteins in byssus and other mussel tissues over the years.(Gantayet et al., 2013, 2014; S. Li et al., 2018; López et al., 2002) An important limitation, however, exists for these species, due to the lack of complete protein databases. Several transcriptomic studies have been performed to describe putative proteins in mussel, however, with only indirect proof of the proteins' importance.

In this study, LC-MS/MS was applied to identify proteins in the foot organ potentially important during byssogenesis. Mussel foot and mantle tissues from *M. edulis* and *M. californianus* were digested with two complementary enzymes prior to LC-MS/MS analysis. Byssus samples from both species were also analysed. This contribution is a unique look at the mussel proteome from two closely-related species, while presenting a list of foot-specific proteins potentially involved in byssus fabrication and regulation.

Experimental

Mytilus edulis mussels (Prince Edward Island, Canada) were shipped live on ice, tissues dissected, flash frozen and lyophilized. *Mytilus californianus* (Friday Harbor, WA, USA) were prepared similarly prior to being shipped for analysis. For each sample type, three biological replicates were digested with trypsin and three biological replicates were also digested with pepsin. Ammonium bicarbonate buffer (100 mM, pH 8) was added to dried tissues prior to homogenisation, and resulting extracts divided into 200 μ L aliquots (3-5 mg tissue each). Reductive alkylation was performed prior to trypsin (pH 8) or pepsin (pH 2) digestions. Samples were digested overnight (20 μ g protease) at 37°C, de-salted using OASIS© HLB cartridges (Waters, Milford, MA, 30 mg/1 cc) and dried. For byssus samples (containing threads and

plaques), 50 μ L 7 M urea and 2 M thiourea was added to ground samples (1-2 mg), sonicated and heated at 50°C (20 min). Trypsin and pepsin (25 μ g protease) digestions were performed as above. Resuspended extracts (100 μ L 10% ACN) were injected (30 μ L) onto an Aeris PEPTIDE XB-C18 100 x 2.1 mm column at 40°C (Phenomenex, Torrance, CA), with water and acetonitrile, both containing 0.1% formic acid. A Nexera UHPLC system (Shimadzu, Columbia, MD) was operated at 300 μ L/min. After a 3 min hold at 5% ACN, the organic composition was increased to 30% at 50 min, 50% at 55 min, then to 90% within one minute. TOF-MS and MS/MS spectra were acquired on a hybrid quadrupole time-of-flight TripleTOF™ 5600 mass spectrometer (AB Sciex, Concord, ON, Canada) in positive electrospray mode at 500°C, with ionspray voltage of 5000 V, declustering potential (DP) 60 V, nebulizer and drying gases 50 psi. TOF-MS (m/z 120-1250), followed by MS/MS acquisition (m/z 80-1500), used dynamic background subtraction and information-dependant acquisition on the 15 most intense ions, with collision energy of 30 ± 10 V and a total cycle time of 1.05 s. LC-MS/MS data for trypsin and pepsin for each biological replicate were searched separately against a protein database from UniprotKB/Swiss-Prot (released on April 2020), using ProteinPilot™ software (version 5.0.2), from 81 species (figure 1a). Search results were then combined and processed in Scaffold-Q+ (v4.9.0, Proteome Software). Proteins with a minimum of two peptides at 1% FDR, and seen in a minimum of two biological replicates, were considered. Data have been deposited to the ProteomeXchange Consortium via the PRIDE(Deutsch et al., 2017) partner repository with the dataset identifier 10.6019/PXD021023.

Results and discussion

The results presented here were compiled from triplicate tryptic and peptic digests of byssus, foot and mantle tissues from *M. edulis* and *M. californianus*. All listed proteins were identified with a minimum of two peptides at 1% FDR and found in at least two biological replicates.

The data was searched with a database including 81 mussel species, since complete protein databases are not yet available for *M. edulis* and *M. californianus*. This database consisted mainly of putative proteins from transcriptomic analysis, with a very small proportion (about 1%) having previous experimental evidence. Figure 1a shows the proportion of proteins by species in the search database. By comparing the distribution of species for identified proteins in *M. edulis* (MYTED, Fig 1b) and *M. californianus* (MYTCA, Fig 1c), a good sequence homology from *M. edulis* and *M. californianus* is seen with *M. coruscus*, *M. galloprovincialis*, and *M. trossulus* proteins. It is striking that, of all proteins detected, only 13% (MYTED) and 8% (MYTCA) were attributed to their respective species. This is, of course, a consequence of incomplete databases for such species. Without searching protein sequences from other species, most proteins would not have been detected. This LC-MS/MS data thus serves to confirm overlapping regions within proteins from different species.

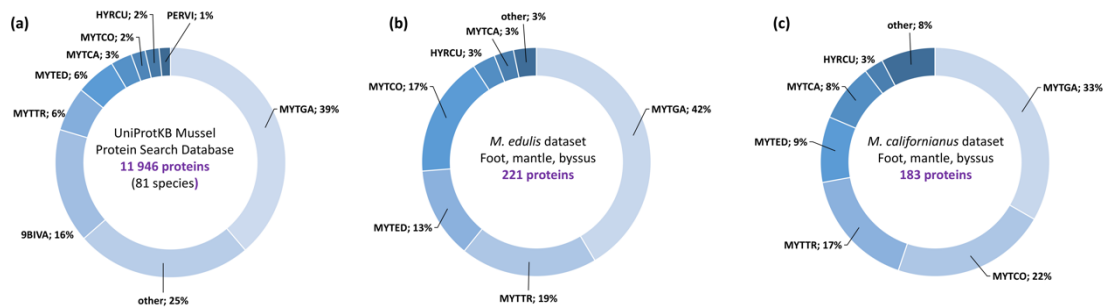


Figure 1. Relative species distribution of proteins, from search database (a), as well as from MYTED (b) and MYTCA (c) datasets combining foot, mantle and byssus analyses

The comparison of proteins identified is shown in Venn diagrams (figure 2). There were 73 and 58 “foot-specific” proteins found in the MYTED and MYTCA datasets, using the mantle as a control tissue. From byssus analyses, 37 and 50 proteins were confidently identified from MYTED and MYTCA, respectively, with 30 proteins

in common. The foot-specific and byssus proteins were investigated further, since the aim of this work was to unveil proteins potentially involved in byssus fabrication and regulation. The details of each complete protein list from the three tissues can be found in the Supplemental Information (Tables S1 and S2), along with the number of confident peptides and sequence coverage for each. Table S3 compiles the 62 foot-specific proteins compiled using the mantle from both species as a control tissue, with additional information such as molecular weight, pI, % abundances for specific amino acids, GRAVY scores and GO annotations. For the byssus and foot-specific proteins, the largest proportion did not have any associated GO terms, followed by proteins with involving binding and metal-ion binding. This is expected, since it is well known that byssus has very strong adhesive properties and metal binding capacities.

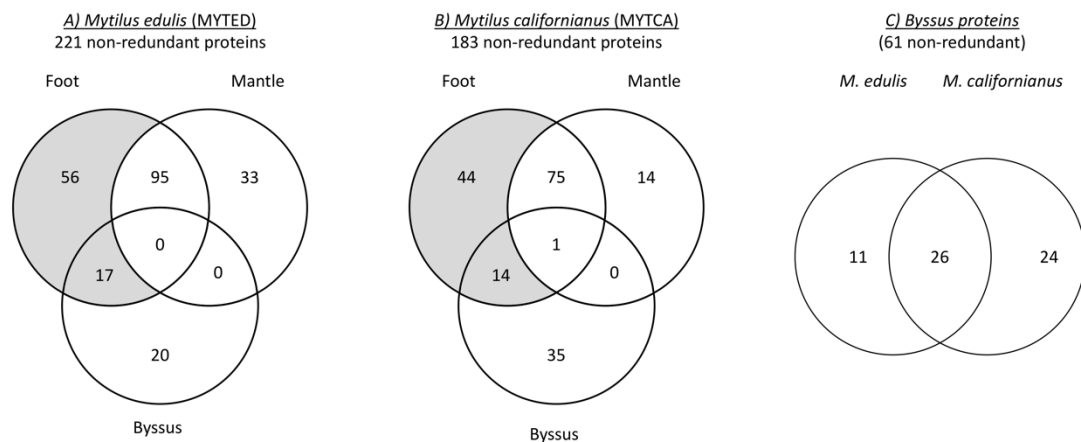


Figure 2. Number of confidently identified proteins detected in foot, mantle, and byssus from *Mytilus edulis* (A) and *Mytilus californianus* (B). Proteins found in byssus were compared between the two species (C).

The main goal of this study was to identify proteins from the mussel's foot potentially implicated in byssogenesis. Many proteins known to be main constituents of byssus itself were detected, as well as several previously implicated in the formation

of byssus.(J. Herbert Waite, 2017) Table 1 lists thread-related proteins found in foot and byssus, including specific enzymes known to be byssal-related, such as glycosyl-hydrolase, peroxidase, tyrosinase, procollagen-proline dioxygenase. The three procollagens (D, NG, P), as well as thread matrix and protease inhibitor proteins are also listed. Ten distinct foot proteins are included, five of which were detected in both species.

Table 1. Thread-related proteins found in foot and byssus samples from *M. edulis* and *M. californianus**

Foot		Byssus	
Byssal calumenin-like 1	Procollagen (D, P, NG)	Byssal glycosyl-hydrolase-like 2	Foot protein 3, 11, 17
Byssal glycosyl-hydrolase-like 2	Procollagen-proline dioxygenase	Byssal peroxidase-like 1, 2 , 4	Mussel byssus collagen-like 3
Byssal peroxidase-like 2	Protease inhibitor-like 1	Byssal protease inhibitor-like 1	Procollagen (D, P, NG)
Byssal protease inhibitor-like 1	Proximal thread matrix 1, 1a	Byssal protein 1, 2, 3	Protease inhibitor-like 1, D1, D2
Byssal tyrosinase-like 1, 2	TSP_1 domain containing 1	Byssal tyrosinase-like 1, 2	Proximal thread matrix 1, 1a
Collagen-like 5, 6, 7		Collagen chain alpha 1	Thread matrix 2A
Foot protein 2		Collagen-like 4, 5, 6, 7	TSP_1 domain containing 1
Foot protein 6, 10, 12, 18, 19		Foot protein 2, 4, 6, 10, 12	Tyrosinase 5

*proteins in bold were detected in both species

Mussel foot proteins are well known for their crucial roles in determining the unique adhesive properties of byssus. Foot proteins 2, 4, 6, 10 and 12 were detected in foot and/or byssus samples of both species. Mfp-3, in three of its variant forms, was detected in *M. californianus* byssus samples. These proteins were identified based on sequences from MYTED and MYTCA for mfp-2, and MYTCA for mfp 3, 4 and 6 (variants 1/2). Variant 3 of mfp-6 (from MYTCO) was found in both species, whereas variants 1 and 2 were detected only in *M. californianus*. A supplemental figure (figure S1) shows the amino acid sequences of all detected mussel foot proteins, and indicates which regions were confidently identified in MYTED and MYTCA samples.

Six *M. californianus* foot proteins recently reported in a transcriptomic study,(DeMartini et al., 2017) were detected here for the first time. These proteins were mfp-10, mfp-11, mfp-12, mfp-17, mfp-18, mfp-19. Mfp-10 and -12 were detected in both species, whereas mfp-11, 17, 18 and 19 were exclusively detected in *M. californianus*.

In addition to the byssal precollagens (distal, proximal, and non-gradient), two enzyme classes known to impact byssus adhesion and plasticity were found. The first enzyme class is responsible for the formation of DOPA and includes byssal tyrosinase-like proteins 1 and 2 (from MYTCO) and tyrosinase 5 (from PERVI). This post-translational modification is argued to have a determinant role in adhesion and structural properties.(Harrington et al., 2010; Petrone et al., 2015; Wei et al., 2012; Yu et al., 2011) Also, three byssal peroxidase-like proteins, previously identified from transcriptomic data in *M. coruscus*, were found in our dataset. These proteins are believed to protect the threads against reactive oxygen species, which may initiate adverse reactions involving DOPA related intra- or inter-molecular crosslinking between byssus proteins.(Zhao & Waite, 2005) Crosslinking in byssus is believed to negatively affect thread adhesion and elasticity, however, this still needs of further study.(J. Herbert Waite, 2017; Yu et al., 2013)

The present study also includes 14 “uncharacterized” proteins obtained from transcriptomic evidence in *M. coruscus*, *M. galloprovincialis*, and *M. trossulus* (see Table S1 and S3). The sequences of these proteins were submitted to a BLAST search (<https://www.uniprot.org/blast/>), and analogous proteins are presented in Supplemental Table S4, with the % sequence homology for each.

From this work, we have presented a list of foot-specific proteins potentially involved in byssogenesis. Of course, more detailed studies must be performed to clarify the specific roles of these proteins in byssus assembly. Hampered by a lack of complete

proteomic databases for the studied species, many proteins were solely identified on the basis of overlapping regions from related species. Once more complete protein databases become available for these species, this data could be further interrogated.

Acknowledgements

The authors would like to acknowledge the *Fonds de Recherche du Quebec - Nature et Technologies* (FQRNT) for the financial support of this research.

ANNEXE B

ÉTUDE DE LA BYSSOGÉNÈSE PAR DES ANALYSES PROTÉOMIQUES SUR LE BYSSUS, LE PIED ET LE MANTEAU CHEZ LES MOULES MYTILUS PAR LC-MS/MS

INFORMATIONS SUPPLÉMENTAIRES AU CHAPITRE 3

Publié dans Proteomics 2020, 10 septembre 2020, sous le titre :

Investigating Byssogenesis with Proteomic Analysis of Byssus, Foot, and Mantle in
Mytilus Mussels by LC-MS/MS

DOI: 10.1002/pmic.202000014

Sansoucy M, Tremblay R, Carrington E, Marcotte I, Sleno L

Informations supplémentaires:

<https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.202000014>

Accession number	Protein name	M. edulis # unique peptides (% sequence coverage)	M. californianus # unique peptides (% sequence coverage)	Ambiguous accession number
A0A3L5TR56_MYTGA	40s s28 ribosomal protein	2 (24.6%)	2 (24.6%)	
A0A077GZL7_MYTTR	60S ribosomal protein L13	4 (19.8%)		
A0A3L5TRV2_MYTGA	Adp-ribosylation 1 factor-like protein	4 (21.9%)		
A0A3L5TSG1_MYTGA	Bifunctional 1-like 3- phosphoadenosine 5- phosphosulfate synthase		6 (17.3%)	
A0A193DU95_MYTCO	Byssal calumenin-like protein 1	33 (70.6%)	26 (63.4%)	
A0A193DU98_MYTCO	Byssal glycosyl-hydrolase-like protein 2		15 (44.1%)	
A0A193DUA2_MYTCO	Byssal peroxidase-like protein 2		7 (10.4%)	
A0A193DU99_MYTCO	Byssal protease inhibitor-like protein 1	7 (5.8%)		
A0A193DUA5_MYTCO	Byssal tyrosinase-like protein 2	9 (14%)	15 (24.4%)	
A0A0G2YLH0_MYTCO	Byssal tyrosinase-like protein-1	3 (6.9%)		
A0A0K2D7V5_MYTCO	Collagen-like protein 5	27 (19.2%)	12 (16.1%)	
A0A0K2D7R0_MYTCO	Collagen-like protein 6	4 (5.7%)		
A0A0K2D820_MYTCO	Collagen-like protein 7	6 (17.9%)		
A0A3L5TR45_MYTGA	Cystathionine gamma-lyase	8 (25.1%)	6 (21.2%)	M1FBU6_MYTTR*
M1FHV2_MYTED*	D-octopine dehydrogenase	27 (66.8%)	5 (32.1%)	
A0A140D2S1_MYTGA	eIF4G	9 (8.7%)		
A0A3L5TSE2_MYTGA	Eukaryotic 1 translation initiation factor 3 subunit	5 (11.9%)		A0A223HCL0_MYTCA*
Q1XBT8_MYTED*	Foot protein 2	9 (18.5%)	4 (13.7%)	
Q0H216_MYTCA	Foot protein-6 variant1		31 (62.8%)	
Q0H215_MYTCA	Foot protein-6 variant2		9 (47.9%)	
A0A223HCJ8_MYTCA	Foot protein 10	9 (17.6%)		
A0A223HCL5_MYTCA	Foot protein 12	15 (13.5%)		
A0A223HCK1_MYTCA	Foot protein 18		13 (37.5%)	
A0A223HCK5_MYTCA	Foot protein 19		10 (79.7%)	
A0A0K0YAY3_MYTCO	Heparan sulfate proteoglycan- like protein-1	60 (15.6%)		
A0A0S1LIC6_MYTCO	HSP90		4 (23.3%)	
A9YLA7_MYTED	Mannose-6-phosphate isomerase	3 (9.6%)		
A0A3L5TS17_MYTGA	Methyltransfer_dom domain- containing protein	14 (50.4%)		
V9P9M1_MYTGA	Mitogen-activated protein kinase	3 (10.8%)		
A0A3L5TU74_MYTGA	Nop domain-containing protein	5 (13.5%)		
E1B300_CRIPL	Peroxiredoxin	8 (28.1%)	6 (26.0%)	
Q6PTI4_MYTCA	Phosphofructokinase	2 (18.3%)	4 (43.8%)	

A0A3L5TTA2_MYTGA	Poly endoribonuclease-c-like-specific	5 (25.8%)		A9QLN2_MYTCA*
O44367_MYTED*	Precollagen D	31 (42.5%)	55 (43.2%)	A9QLN3_MYTCA*
O76271_MYTED*	Precollagen NG (Nongradient byssal)	38 (51.3%)	58 (50.9%)	A9QLN4_MYTCA*
O16161_MYTED*	Precollagen P	28 (33.6%)	18 (37.3%)	
I1YP73_MYTGA	Procollagen-proline dioxygenase	19 (43.9%)	13 (34.6%)	
A0A0K2D7Q5_MYTCO	Protease inhibitor-like protein 1	50 (31.6%)	44 (26.7%)	
Q8T5C2_MYTGA	Proximal thread matrix protein 1		7 (52.3%)	
Q8T6U5_MYTED	Proximal thread matrix protein 1 variant a	81 (88.2%)	31 (56.7%)	
Q0ZPQ2_MYTGA	Ras		4 (30.2%)	
A0A3R5TRR7_MYTGA	Ras-related rab-43-like protein	3 (5.7%)	3 (5.7%)	
A0A077GYU8_MYTTR	Ribosomal protein L19	3 (12.9%)	4 (17.3%)	
A0A077GYT9_MYTTR	Ribosomal protein L22	4 (28.7%)		
Q6PTI5_MYTCA	S-adenosylmethionine synthase	6 (22.3%)	2 (10.3%)	
Q8MUX9_DREPO	Serine/threonine-protein phosphatase		4 (24.0%)	
A0A0K0YB57_MYTGA	SRAV-rich protein 1	31 (72.8%)		
A0A409V6J6_MYTGA	Thioredoxin domain-containing protein	3 (27.3%)		
A0A077H0P3_MYTTR	Transcription factor BTF3	3 (26.7%)	5 (36.4%)	
A0A0K0YB28_MYTGA	Transgelin-like protein 1	21 (90.9%)		
A0A0K0YB26_MYTGA	Transgelin-like protein 4	4 (90.9%)		
A0A3R5Q2X9_MYTGA	Transmembrane 9 emp24 domain-containing protein	5 (44.8%)		
G0YGG2_MYTED	Triosephosphate isomerase	2 (57.7%)		
TPM_PERVI	Tropomyosin	7 (89.4%)		
A0A3L5TTI3_MYTGA	Tryptophan-trna cytoplasmic		4 (18.4%)	
A0A0K2D7M8_MYTCO	TSP_1 domain containing protein 1	33 (29.6%)	12 (19.8%)	
A0A3L5TS65_MYTGA	Uncharacterized protein	4 (20.8%)	4 (18.0%)	
A0A3L5TRH9_MYTGA	Uncharacterized protein		3 (16.0%)	
A0A3L5TRE4_MYTGA	Uncharacterized protein	6 (32.6%)		
A0A3L5TR65_MYTGA	Uncharacterized protein	11 (37.8%)		
A0A5Q0TNM5_MYTCO	Whirlin	7 (41.5%)		
A0A0K2D7N2_MYTCO	YGH-rich protein 3	3 (5.4%)		

La protéine en gras représente une séquence révisée dans la base de donnée UniProt

*Les peptides identifiés dans MYTED sont pour *M. edulis* et les peptides dans MYTCA sont pour *M. californianus*

ANNEXE C

COMPARAISON D'APPROCHES DE PRÉ-FRACTIONNEMENT POUR LES ANALYSES LC-MS/MS EN PROTÉOMIQUE ASCENDANTE

L'utilisation d'une approche par pré-fractionnement permet l'identification de plusieurs peptides et protéines, tel que présenté au chapitre précédent. Par contre, le nombre d'échantillons doit être pris en considération dans le cas où une étape de pré-fractionnement des peptides est considérée plus sérieusement. Une approche par SCX est généralement l'approche par excellence, malgré plusieurs désavantages associés à la préparation des échantillons en prévision de l'analyse LC-MS/MS.

Dans ce chapitre, il sera question, d'une part, de comparer les résultats obtenus par SCX et par RP employant des phases mobiles basique et acide. D'autre part, la combinaison de différentes fractions a souvent été introduite comme méthode efficace afin d'améliorer l'orthogonalité. La pertinence de combiner plusieurs fractions afin d'améliorer l'orthogonalité sera aussi abordée dans le présent chapitre.

Maxime Sansoucy et Lekha Sleno sont les co-auteurs de cet article. Maxime Sansoucy a procédé à la recherche bibliographique, la préparation des protocoles, la réalisation des expériences en laboratoire, le traitement des données et la préparation du manuscrit original. La professeure Lekha Sleno a fourni son support et a supervisé le projet. Ce chapitre fera l'objet d'une publication soumise au journal *Rapid Communication in Mass Spectrometry* ultérieurement.

Résumé

Rationnel. Les approches par chromatographie liquide multi-dimensionnelle sont des approches utiles pour l'augmentation de la couverture du protéome. La chromatographie par échange de cations forts (SCX) est la méthode par excellence pour le pré-fractionnement des peptides. Une comparaison systémique entre une collection de fractions par chromatographie liquide en phase inverse, à pH basique et acide, ainsi que par SCX a été réalisée afin d'évaluer leur orthogonalité respective. De plus, l'impact de la combinaison de plusieurs fractions a été abordé pour chacune des approches.

Méthode. Les protéines provenant de la fraction S9 de foie de rat ont été digérées à l'aide de deux enzymes complémentaires (la trypsine et la pepsine). Les échantillons ont été extraits sur phase solide, puis différentes conditions de lavage acides et basiques ont été testées sur trois résines polymorphiques. Pour les analyses 2D-LC-MS/MS, les peptides ont été fractionnés par échange de cations et en phase inverse. La collecte de fractions en phase inverse s'est effectuée à l'aide de phase mobile basique et acide (pH 10 et pH 3 respectivement). Les analyses ont été réalisées sur un instrument hybride quadripôle-temps de vol opéré en mode *data-dependant*.

Résultats. Pour les extractions en phase solide, une meilleure récupération des peptides s'est faite à l'aide d'une résine de type HLB, avec un lavage à l'eau nanopure. Les échantillons pré-fractionnés selon une approche par phase inverse à pH élevé ont généré de meilleurs résultats en terme du nombre de protéines identifiées, du nombre de peptides uniques ainsi que par le recouvrement de la séquence. Une quantité comparable de protéines identifiées par les approches par chromatographie en phase inverse à pH acide et par SCX a été rapportée. Toutefois, les résultats obtenus par Lp-RP sont améliorés lorsque deux ou trois fractions ont été combinées.

Conclusions. Une comparaison détaillée entre les stratégies de pré-fractionnement a montré que la Hp-RP a été hautement complémentaire aux conditions acides des analyses LC-MS/MS. Aussi, la combinaison de fractions permet d'améliorer l'orthogonalité ainsi que le recouvrement du protéome. L'utilisation de la chromatographie en phase inverse est par conséquent une alternative intéressante puisqu'elle ne nécessite aucune étape supplémentaire de désallage, contrairement à la SCX. Finalement, les peptides collectés à pH élevé (~ 10) ont été identifiés avec une confiance supérieure, avec une meilleure chance pour les peptides moins abondants d'être séquencés.

Introduction

Traditionnellement, les approches par chromatographie d'échange ionique (off-line et on-line) servent de première dimension lors du pré-fractionnement des peptides dans les analyses 2D-LC-MS/MS. Malgré que cette approche offre une importante orthogonalité avec la phase inverse utilisée lors des analyses LC-MS/MS, les phases mobiles utilisées en chromatographie échangeuse d'ion contiennent de grandes quantités de sels non-volatiles qui peuvent rapidement détériorer le processus d'ionisation (Beaudry & Vachon, 2006). Des approches alternatives employant des concentrations moins importantes durant la séparation des peptides ont été par conséquent introduites. Par exemple, Dai et al. ont employé un gradient de pH, avec une concentration dans les millimolaires avec lequel une meilleure efficacité à la séparation a été obtenue. Dans leurs études, les auteurs ont rapporté des peptides moins abondants qui, normalement, co-éluaient dans les conditions standard d'échange ionique (Dai et al., 2005). Pareillement, Mommen et al. ont adapté la technologie MudPIT en éliminant la présence de sels non-volatiles dans la composition de la phase mobile, ce qui a permis l'identification d'une plus grande proportion de protéines (Mommen et al., 2013).

La chromatographie en phase inverse a initialement été considérée comme une approche semi-orthogonale, en raison d'une sélectivité similaire entre les premières et secondes dimensions (Dowell et al., 2008). Lors d'études comparatives, la chromatographie en phase inverse a été comparée à des techniques véritablement orthogonales, incluant la SCX, la séparation sur gel ainsi que la procédure en phase inverse à taux élevé de récupération de protéines. Il a été montré, à partir de ces études, que le pré-fractionnement des peptides, à pH basique ou acide, offrait des résultats comparables, sinon meilleurs, que ceux obtenus par SCX par exemple (Dowell et al., 2008; Y. Wang et al., 2011). Également souligné par Dowell, la RP a l'avantage d'un faible débit de ré-échantillonnage des peptides lors de l'activation, en plus d'un pouvoir de résolution nettement supérieur, ce qui contribue largement à l'amélioration du nombre de peptides identifiés par Hp-RP (Dowell et al., 2008).

Un autre avantage de la chromatographie en phase inverse comme première dimension est que l'analyse LC-MS/MS peut être directement faite sur les fractions collectées, sans risque de causer une suppression ionique. Néanmoins, la nature semi-orthogonale du fractionnement par RP lors de la séparation des peptides est souvent montrée comme étant un désavantage majeur (Gilar et al., 2005a; Hao et al., 2013; Song et al., 2009). Plusieurs groupes se sont penchés sur la question et ont proposé la combinaison de fractions non-adjacentes, collectées à différents moments de la chromatographie (Dwivedi et al., 2008; H. Lee et al., 2016; F. Yang, Shen, Camp, et al., 2012). Le fractionnement par chromatographie en phase inverse a été largement amélioré et son utilisation dans les approches en protéomique *bottom-up* est devenu de plus en plus répandue depuis les dernières années (Callipo et al., 2011; Gokce, Andrews, Dean, & Muddiman, 2011; Tomlinson & Chicz, 2003; Yeung et al., 2020). Malgré tout, l'emploi de la SCX comme étape de pré-fractionnement demeure l'approche préférée. Le nombre de fractions par RP et la manière dont celles-ci sont combinées reste, à ce jour, peu exploré et devrait être clarifié à l'aide de conditions expérimentales.

Les approches classiques en protéomique *bottom-up* demandent une étape de désalage après la digestion enzymatique. L'extraction en phase solide (SPE) est une étape de nettoyage particulièrement utile et simple en protéomique *bottom-up*. Les résines commercialement disponibles permettant le désalage de l'échantillon incluent différentes options, comme la silice, une résine échangeuse d'ions ainsi que des groupements électrostatiques imprimés sur polymères (Tubaon, Haddad, & Quirino, 2017). Des montages *on-line* et *in-line* ont été conçus de manière à 1) maximiser la récupération des peptides lorsqu'une quantité limitée de l'échantillon est disponible, comme lors d'études de protéines biomarqueurs de cancer (Hustoft et al., 2013) et 2) améliorer la productivité de la méthode lors d'applications en bioanalyses (Falkenby et al., 2014). Règle générale, la SPE est principalement employée afin de réduire les effets de matrices causés par l'excès de sels, retirer les protéines non digérées et pré-concentrer l'échantillon en prévision de l'analyse par LC-MS/MS.

Cette étude comparative évaluera l'analyse globale du protéome d'un échantillon biologique complexe, provenant de la fraction S9 de foie de rats. Trois stratégies de pré-fractionnement ont été employées pour la collection de peptides, essentiellement la SCX, la RP à pH basique ainsi que la RP à pH acide. Aussi, l'impact de la combinaison de fractions a été évalué sur la couverture globale du protéome.

Expérimental

Matériel

Les extraits de protéines de la fraction S9 de foie (20 mg/ml) de rat mâle (Sprague-Dawley) proviennent de Corning Life Sciences (Tewksbury, MA). La trypsine (bovine), la pepsine (porcine), le dithiotréitol (DTT), l'iodoacétamide (IAM), l'urée, la thiourée, l'acide formique de grade HPLC, l'acétate d'ammonium, le bicarbonate d'ammonium (ABC), l'acétonitrile et le méthanol ont tous été achetés chez

Sigma-Aldrich (St. Louis, MO). L'eau ultrapure a été obtenue à partir du système Millipore Synergy UV (Billerica, MA).

Digestion des protéines

Les protéines de la fraction S9 de foie de rats (environ 1 mg, 50 µl) ont été solubilisées dans une solution contenant 7 M d'urée et 2 M de thiourée (50 µl) et diluée 1 :1 avec une solution de 100 mM du tampon ABC (pH 8.5). Une alkylation réductrice a été réalisée (10 mM DTT durant 20 min suivit de 25 mM IAM pour 15 min, 37°C, dans le noir), suivie d'un second ajout de DTT (10 mM pour réagir avec l'excès d'IAM). Les échantillons ont ensuite été dilués avec 200 µl de la même solution d'ABC utilisée plus haut pour une digestion par la trypsine ou avec la même quantité d'une solution de 2% acide formique préparée dans 20% méthanol en prévision d'une digestion par la pepsine. Les échantillons ont ensuite été incubés selon un ratio enzyme : protéine de 1 : 50 durant 4h à 37°C pour les deux protéases. Les digestions ont été arrêtées suivant l'ajout de 500 µl d'une solution 1% d'acide formique pour la trypsine ou de 50 µl d'une solution à 5% NH₄OH pour la pepsine avec un ajout supplémentaire de 400 µl d'eau en prévision de l'extraction en phase solide sur des cartouches 1 cc (30 mg) OASIS® HLB (Waters, Milford, MA). Les peptides ont été élués de la cartouche à l'aide de méthanol (1 ml), puis évaporés à sec sous azote et conservés à -30°C en attendant la collection de fractions.

Chromatographie en phase inverse

Les peptides séchés ont été mis de nouveau en solution dans la phase mobile A (120 µl, voir détails ci-dessous) et centrifugés durant 5 min à 14 462 x g avant l'injection (100 µl) sur une colonne Agilent Zorbax Extend C18, 250 x 4.6 mm, 5 micron (80 Å) (Agilent Technologies, Palo Alto, CA), employant un Agilent 1200 Series HPLC équipé d'un auto-échantillonneur à température contrôlée, d'un dégazeur

intégré, d'un système de pompe binaire, d'un détecteur à barrette de diode (*diode array detector*, DAD) ainsi qu'un collecteur de fractions réfrigéré. La collecte de fractions en phase inverse a été réalisée à un débit de 600 µl/min avec un gradient d'élution débutant à 5% B pendant 1.5 min, suivie d'une augmentation à 30% à 15 min, à 50% à 17 min, puis à 80% en 1 min, maintenu durant 2.6 min. Les phases mobiles A et B sont respectivement 10 mM de NH₄OAc (pH 10) et d'un mélange de 10% A dans 90% ACN pour un pré-fractionnement à pH basique. Les phases mobiles A et B ont été remplacées par de l'eau et ACN, toutes deux acidifiées avec 0.1% d'acide formique lorsque les peptides sont séparés à pH acide. L'absorbance UV a été réglée à 220 et 280 nm. La fenêtre de fractionnement a été ajustée de manière à collecter 16 fractions (1.1 min par fraction) à partir de 3 à 20.6 min ou 24 fractions (0.73 min par fraction) entre 3 et 20.5 min. Les fractions ont ensuite été séchées sous un débit d'azote et conservées à -30°C en prévision de l'analyse LC-MS/MS.

Chromatographie par échange de cations forts

Les peptides ont été remis en solution dans la phase mobile A (120 µl, voir détails ci-dessous) et centrifugés (5 min, 14 462 x g) avant l'injection de 100 µl sur une colonne Agilent Zorbax-300, 150 x 2.1 mm, 5 micron (Agilent Technologies, Palo Alto, CA) et employant le même système HPLC décrit pour le fractionnement en phase inverse. L'appareil a été opéré à un débit de 250 µl/min avec un gradient d'élution allant de 0 à 8% B pour les six premières minutes, puis augmenté à 30% à 11 min, à 65% à 20 min, pour ensuite atteindre 95% en 5 min et maintenu durant 10 min. Les phases mobiles A et B sont respectivement des solutions tampons préparées avec 10 mM K₂HPO₄ dissout dans 25% ACN (pH 2.75) et 1 M KCl dissout dans la phase mobile A. La collecte des fractions a été répartie sur 24 min selon l'intervalle suivant : 1-4, 4-6, 6-10, 10-12.5, 12.5-15, 15-17.5, 17.5-20, 20-24 min. De plus, les intervalles ont été ajustés de manière à collecter 16 fractions sur 24 minutes. Les fractions ont ensuite été

évaporées à sec sous un débit d'azote avant d'être conservées à -30°C pour les analyses LC-MS/MS.

Analyse RP-LC-MS/MS

Les fractions séchées préparées dans une solution 10% ACN (100 µl) et 30 µl ont été injectées sur une colonne Aeris PEPTIDE XB-C18 100 x 200 mm avec une taille de particule de 1.7 µm et maintenues à une température de 40°C (Phenomenex, Torrance, CA). Un système UHPLC, modèle Nexera (Shimadzu, Columbia, MD) a été opéré à un débit de 300 µl/min avec une proportion de 5% de la phase mobile B et maintenu pendant les 2.5 premières minutes. La composition de la phase mobile B a été augmentée de façon linéaire à 30% à 24 minutes, puis à 50% B à 26 minutes et finalement, à 85% B à 26.5 min et maintenu pendant deux minutes.

Les spectres TOF-MS et MS/MS ont été obtenus à partir d'un spectromètre de masse hybride quadripôle temps-de-vol TripleTOF™ 5600 (AB Sciex, Concord, ON, Canada). La source d'électronébulisation a été opérée en mode positif et maintenue à une température de 500°C, avec un voltage appliqué sur le capillaire de 5kV et un potentiel de déclusterisation de 60 V. La pression des gaz de nébulisation et de séchage a été maintenue à 50 psi. La sonde d'électronébulisation a été utilisée pour assimiler le flux provenant du LC alors que la sonde APCI a été employée pour la calibration après quatre injections (en mode MS et MS/MS), à l'aide d'une solution de calibrants préparée dans nos laboratoires. Les mesures TOF-MS ont été réalisées entre m/z 80-1500, avec une soustraction dynamique du bruit de fond (*Dynamic Background Subtraction*, DBS) et un critère de sélection d'information dépendante (Information-dependant acquisition, IDA) sur les 15 ions les plus sensibles au-delà de m/z 250. Les spectres MS/MS ont été enregistrés suite à l'accumulation (50 ms) et la collision (CE = 30 ± 10 V) des ions précurseurs. Le temps de cycle total calculé est de 1.05 s.

Traitement des données

Les données MS/MS ont été cherchées à partir de la base de protéines provenant du répertoire UniprotKB/Swiss-Prot pour l'espèce *Rattus norvegicus* (8 054 protéines, disponible en date du 7 mars 2019) par le logiciel ProteinPilot™ (Sciex, version 5.0.2). Pour chaque approche, les fractions provenant des échantillons digérés par la trypsine et la pepsine ont été traitées individuellement. Les formats .mgf, .mzIdent et le PeptideSummary ont ensuite été importés, puis combinés dans le logiciel Scaffold Q+ (Proteome Software, version 4.10.0). À partir de l'interface Scaffold, le seuil minimal à 1% FDR a été fixé pour les peptides et les protéines. Également, seulement les protéines identifiées avec un minimum de deux peptides ont été considérées.

Résultats et discussion

Les effets de matrice sont une préoccupation majeure en spectrométrie de masse car l'efficacité de l'ionisation est directement affectée par la présence d'interférents en concentration impotante. Par conséquent, l'élimination des contaminants tels que l'ADN et les lipides est un prérequis lors d'analyses par spectrométrie de masse. Une façon d'amoindrir l'impact d'un effet de matrice est de réaliser une séparation multi-dimensionnelle du mélange de peptides. Différentes approches lors de l'extraction par SPE ont d'abord été testées.

Comparaison de la préparation des échantillons

Les fractions S9 de foie de rats ont été choisies comme échantillons complexes représentatifs pour cette étude puisque la fraction S9 est composée à la fois de protéines cytosoliques et microsomales. Elle est disponible à partir d'une source commerciale. La reproductibilité de ce type d'échantillon a été mise de l'avant afin de comparer différents paramètres affectant la couverture du protéome. Ce type d'échantillon inclus aussi plusieurs sources d'interférents potentiels qui sont susceptibles de causer des

effets de matrices (Bylda et al., 2014; Srinivas, 2009). Pour maximiser le nombre de peptides détectés, trois cartouches d'extraction sur phase solide ont été testées avec différentes étapes de lavage. Chaque condition a été testée en triplicat ($n = 3$) pour une analyse comparative.

Les protéines de la fraction S9 ont d'abord été digérées à l'aide de deux enzymes complémentaires (la trypsine et la pepsine), puis transférées sur une résine polymérique hydrophilique-lipophile balancée (HLB). Un lavage à l'eau en prévision de l'élution des peptides avec du méthanol a été utilisé comme méthode générique. Parallèlement, les digestats ont été lavés avec une solution d'acide formique ou d'hydroxide d'ammonium à quatre concentrations différentes (0.1; 0.5; 1 et 5%). Le raisonnement derrière ces tests vise à réduire les composés interférants, tels que les phospholipides avant l'analyse par LC-MS/MS. Les résultats de cette comparaison sont présentés à la figure suivante (figure A. 1). Afin d'évaluer l'importance statistique, un test de Student (t -test) utilisant la correction de Welch's a été appliqué aux échantillons. De légères variations ont été observées dans les différentes conditions de lavage pour la cartouche HLB, cependant, aucun des lavages à l'acide formique et d'hydroxide d'ammonium ne présentait une importance statistique ($p < 0.05$). En terme de protéines identifiées et de couverture globale du protéome, (voir tableau 1), chaque condition a été jugée comparable à la méthode générique. Par conséquent, un simple lavage à l'eau ultrapure a été préféré pour la cartouche HLB.

Dans une comparaison parallèle, le digestat de protéines a été transféré sur deux résines échangeuses d'ion distinctes, notamment les résines MAX et MCX. À des fins de comparaison, le pH des échantillons a été ajusté à 3, 7 et 10 avant le transfert sur la résine. Les peptides ont ensuite été élués avec une solution d'acide formique ou d'hydroxide d'ammonium (à 2 et 5%) préparée dans du méthanol. Le nombre de peptide détectés a ensuite été comparé à la méthode générique par HLB (figure A. 1a). Lorsque l'élution est réalisée selon les recommandations du fabricant (pH ajusté),

aucune différence statistique n'a été observée, en comparaison avec un lavage à l'eau ultrapure sur une cartouche HLB (méthode générique, figure A. 1b). Dans le même ordre d'idées, l'élution des peptides faite à l'aide des solutions à 2 et 5% d'acide formique dans le méthanol a généré un nombre comparable de peptides pour la cartouche MAX.

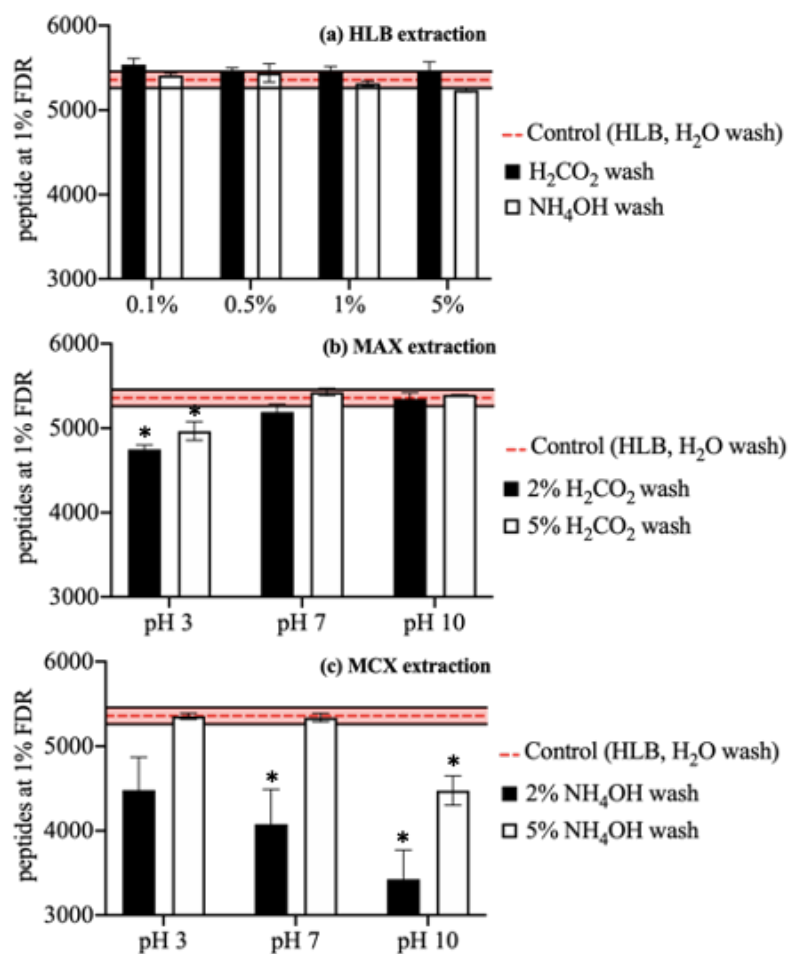


Figure 1 Récupération des peptides obtenus avec des lavages à l'acide formique et à l'hydroxyde d'ammonium pour une extraction HLB (a) et comparé avec un lavage à l'eau (méthode générique) (encadré rouge).

Les conditions d'élution furent aussi testées et le tableau A. 1 résume les résultats obtenus pour l'ensemble des lavages et cartouches testés. Cependant, dans le cas de la cartouche MCX, la collecte des peptides avec une solution à 5% d'hydroxide d'ammonium a donné de meilleurs résultats que pour l'élution réalisée avec une solution à 2%, pour chaque pH testé. Dans un autre ordre d'idées, le pré-fractionnement d'un mélange de peptide est généralement perçu comme un avantage. Une chromatographie par SCX est traditionnellement implantée dans les approches 2D-LC en raison de son importante orthogonalité avec la seconde chromatographie. Malgré certains désavantages techniques, comme l'importante concentration en sels non-volatiles, la SCX a été considérée depuis les 20 dernières années comme la méthode par excellence pour la collecte de peptides. Les travaux présentés dans cette annexe montrent l'ébauche d'une comparaison complète entre différents modes de pré-fractionnement pour une extraction utilisant des cartouches MAX (b) et MCX (c) et comparé avec la méthode générique. Les résultats montrant un changement significatif ($p < 0.05$) sont marqués d'un astérisque.

Pour le pré-fractionnement des peptides, la SCX est généralement le choix par excellence en protéomique *bottom-up*. Cependant, l'utilisation de sels dans la préparation des phases mobiles peut causer une suppression du signal, ce qui impose l'ajout supplémentaire d'une étape de désalage avant l'analyse LC-MS/MS. Alternativement, le fractionnement en phase inverse a l'avantage de ne pas avoir recours à l'addition de sel et par conséquent, ne nécessite pas une étape de désalage, ce qui limite les risques de perte d'échantillon durant la préparation. Les digestions par la trypsine et la pepsine ont été collectées en groupe de 8, 16 et 24 fractions et la fraction non-adjacente a été mélangée. La combinaison de fractions a entre autres permis la comparaison de la couverture du protéome obtenue pour un fractionnement réalisé par chromatographie en phase inverse à pH basique et pH acide ainsi que par SCX. La figure 2 suivante illustre un chromatogramme UV représentatif pour chaque méthode de pré-fractionnement.

Tableau 1 Résultats de la couverture du protéome pour différents protocoles SPE

OASIS HLB SPE			
wash	sample	# proteins (SD)	% sequence coverage (avg)
H ₂ O	control	596 (11.0)	28.2%
H ₂ CO ₂	0.1%	586 (4.9)	29.6%
	0.5%	588 (11.5)	29.2%
	1%	586 (5.0)	29.2%
	5%	584 (5.9)	29.2%
NH ₄ OH	0.1%	598 (2.0)	28.4%
	0.5%	601 (13.0)	28.5%
	1%	581 (5.9)	28.6%
	5%	588 (4.2)	28.0%
OASIS MAX SPE			
elution	sample pH	# proteins (SD)	% sequence coverage (avg)
2% H ₂ CO ₂ (MeOH)	3	551 (4.0)	26.7%
	7	574 (3.6)	28.0%
	10	579 (12.7)	28.4%
5% H ₂ CO ₂ (MeOH)	3	567 (5.9)	26.8%
	7	589 (6.7)	28.5%
	10	586 (5.9)	28.6%
OASIS MCX SPE			
elution	sample pH	proteins avg (SD)	% sequence coverage (avg)
2% NH ₄ OH (MeOH)	3	502 (41.8)	27.9%
	7	468 (36.4)	27.1%
	10	407 (27.3)	26.2%
5% NH ₄ OH (MeOH)	3	577 (2.1)	29.1%
	7	569 (11.9)	29.2%
	10	489 (13.9)	28.4%

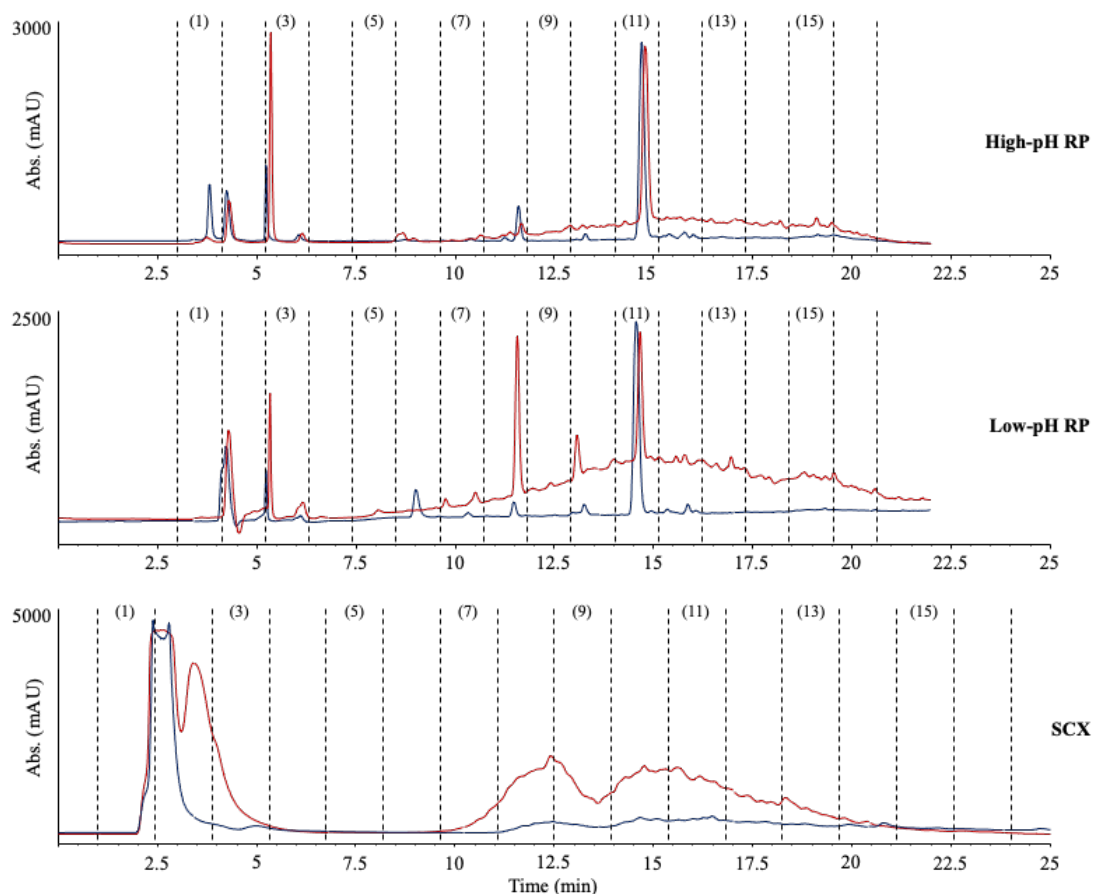


Figure 2 Chromatogramme LC-UV pour les peptides fractionnés en phase inverse à pH basique et acide ainsi que par échange de cations forts, pour une digestion trypsique (en bleu) et une digestion pepsique (en rouge) de protéines provenant de la fraction S9 de foi de rats.

La fraction S9 est une source riche en protéines membranaires, qui sont généralement plus accessibles à des protéases non-spécifiques, telle que la pepsine (Golizeh & Sleno, 2013). La pepsine clive préférentiellement près des résidus aromatiques ou hydrophobiques (Dunn, 2002), alors que la trypsine clive du côté C-terminale des lysines et arginines. Par conséquent, la combinaison de ces enzymes complémentaires durant le traitement des données s'avère une judicieuse stratégie pour l'identification

de protéines (Choudhary, Wu, Shieh, & Hancock, 2003; Golizeh & Sleno, 2013; Tsiatsiani & Heck, 2015). En jugeant les cartes de densité (figure A.3) préparées à l'aide des temps de rétention normalisés lors de l'analyse LC-MS/MS (Gilar et al., 2005a), le fractionnement par RP à pH basique est nettement plus performant pour la séparation des peptides que ne l'a été la RP à pH acide et la SCX.

Dans une série d'expériences parallèles, les digestats de protéines ont été divisés en 16 et 24 fractions durant le fractionnement à pH basique et acide, pour ensuite être combinées en huit fractions finales. La combinaison de fractions non-adjacentes est l'une des stratégies permettant de contrer la faible orthogonalité avec la RP-LC-MS/MS (Y. Wang et al., 2011; F. Yang, Shen, Camp, et al., 2012), comme l'illustre la figure A.3 pour les peptides pré-fractionnés à pH acide. Pour ce qui est de Hp-RP, la combinaison de fraction montre une meilleure distribution du TIC tout au long des huit fractions. Il semble que les TIC, provenant de la combinaison de 3 fractions non-adjacentes, soient moins intenses lorsqu'ils sont comparés avec une collecte de 16 fractions. La faible complémentarité d'une collecte de fractions à pH acide est clairement démontrée lorsque 16 fractions individuelles ont été analysées. L'intensité du signal mesuré par le TIC a montré une corrélation directe avec le temps de rétention des peptides collectés. Le mélange de fractions non-adjacentes améliore la complémentarité d'une approche de fractionnement par phase inverse à pH acide (figure A.3, centre droite).

L'approche par pré-fractionnement à pH basique montre une distribution des peptides comparables à celle obtenue par SCX. Cette observation est en lien direct avec des études antérieures qui ont souligné l'importance du pH sur la sélectivité des peptides sur la complémentarité des approches en phase inverse avec les plate formes LC-MS/MS (Di Palma et al., 2012; Gilar et al., 2005a). Étonnamment, lorsque le nombre de protéines identifiées à partir de 16 fractions (pH 10 et pH 3) a été comparé

avec les résultats obtenus par SCX (8 fractions individuelles), l'approche par pré-fractionnement

Tableau 2 Couverture du protéome selon trois approches de pré-fractionnement

Hp-RP			
# fractions	# peptides	# protéines	% couverture de séquence (moy.)
16	22 228	791	40.3%
8 (à partir de 16)	17 971	777	37.2%
24 (à partir de 16)	16 433	772	35.5%
Lp-RP			
# fractions	# peptides	# protéines	% couverture de séquence (moy.)
16	9 447	606	31.0%
8 (à partir de 16)	8 976	570	31.5%
24 (à partir de 16)	10 812	660	32.2%
SCX			
# fractions	# peptides	# protéines	% couverture de séquence (moy.)
8	7 235	507	28.2%
16	8 168	546	28.4%
8 (à partir de 16)	8594	567	27.5%

en phase inverse a généralement mieux performé que l'approche par SCX, avec 32 protéines identifiées seulement à partir des approches RP, comme le montre la figure A.4. Les résultats présentés dans le tableau A.2 ont révélé que le fractionnement des peptides via une chromatographie à pH acide est relativement équivalent à un pré-fractionnement par SCX. À partir de la figure A.4, la collecte de fractions à pH basique

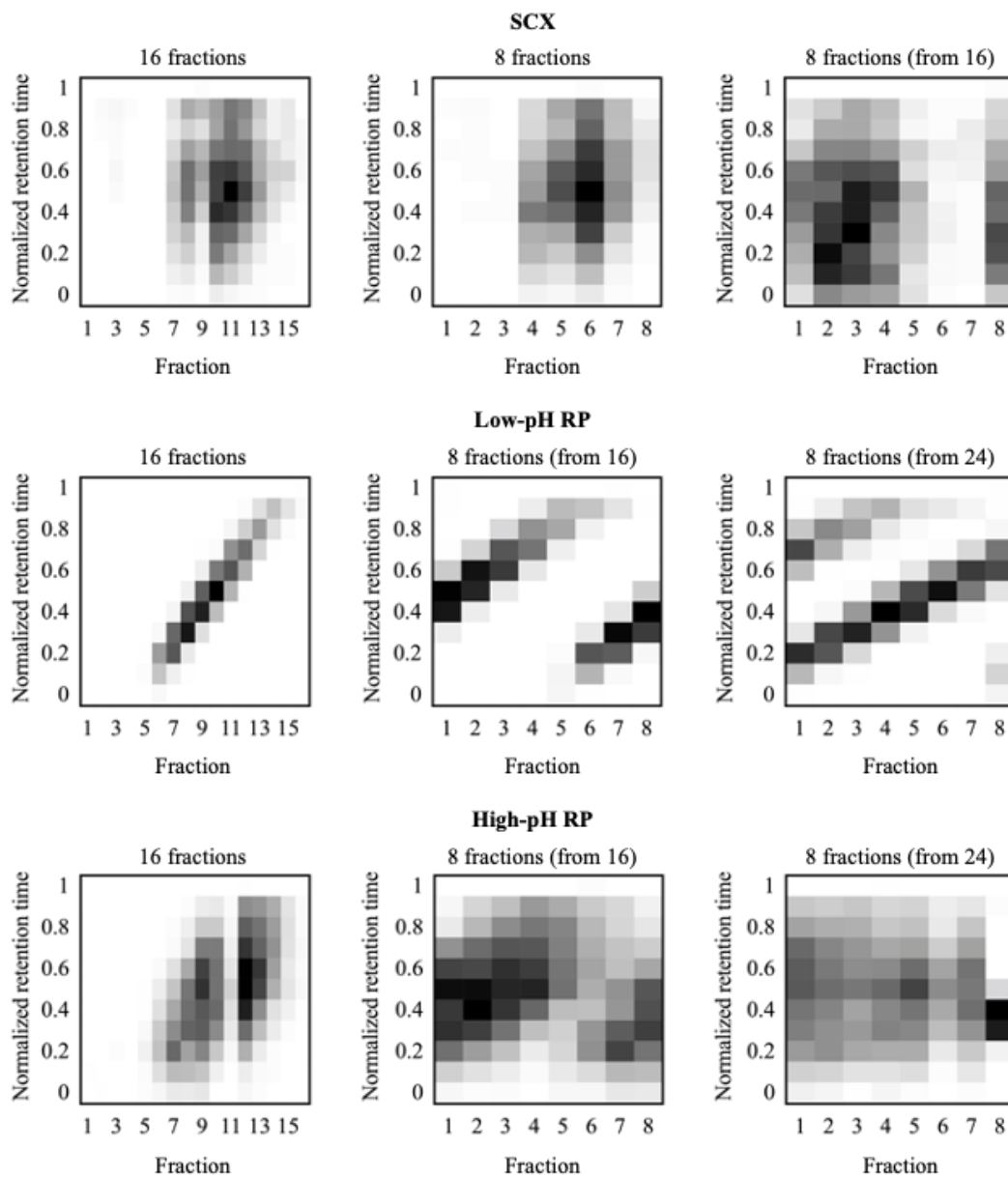


Figure 3 Distribution des peptides par fraction en utilisant le temps de rétention normalisé pour un fractionnement par SCX (au-dessus), par Hp-RP (au centre) et la Lp-RP (au-dessous) pour une digestion tryptique. Noter que la coloration représente l'intensité normalisée du TIC provenant des signaux obtenus par LC-MS/MS.

semble offrir de meilleurs résultats et que la combinaison de 16 à 8 fractions est adéquate afin d'obtenir d'excellents résultats. Il est intéressant de noter que la collection de 24 fractions n'a pas amélioré le nombre de protéines identifiées, probablement en raison de la dilution de l'échantillon causé par la courte fenêtre de fractionnement. De plus, l'analyse de huit fractions est deux fois plus rapide que l'analyse de 16 fractions individuelles, avec un minimum de bénéfice en terme de protéines identifiées et d'augmentation du taux de couverture du protéome. Par conséquent, la collection de 16 fractions combinées en huit fractions finales a été la stratégie préférée pour le pré-fractionnement des peptides par Hp-RP.

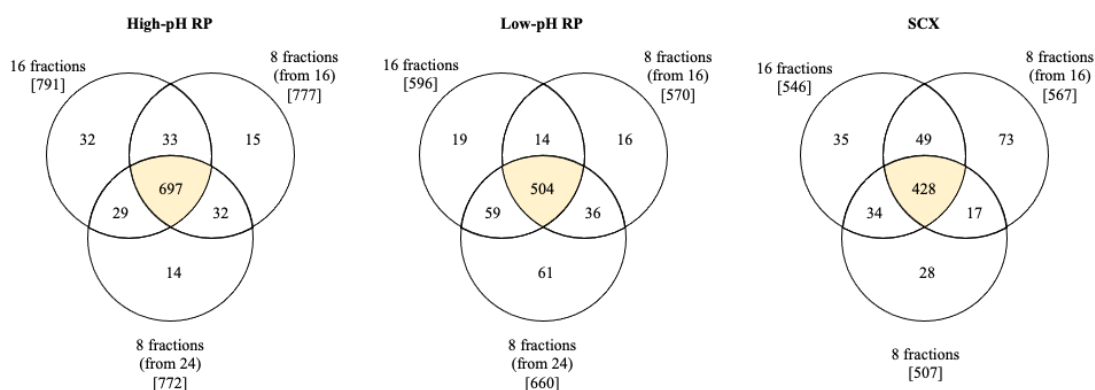


Figure 4 Diagramme de Venn comparant le nombre de protéines (à 1% FDR) identifiées à partir d'un fractionnement par chromatographie en phase inverse à pH basique et acide ainsi que par SCX, avec ou sans combinaison de fractions.

Conclusions

Le pré-fractionnement d'un échantillon de protéines suite à la digestion demeure la méthode à ce jour la plus efficace afin d'améliorer le nombre de protéines identifiées. La composition des phases mobiles utilisées lors de la séparation des peptides influence considérablement la détection des peptides durant les mesures par spectrométrie de masse. Dans cette étude comparative, deux techniques de pré-fractionnement utilisant la chromatographie en phase inverse (à pH basique et acide) ont été comparées avec un pré-fractionnement utilisant une résine échangeuse de cations forts.

À pH basique, un plus grand nombre de peptides et de protéines ont été identifiés lorsqu'une technique de fractionnement par chromatographie en phase inverse a été employée, surpassant les résultats obtenus par SCX et ce, indépendamment de la manière dont les fractions ont été combinées. La combinaison de fractions non-adjacentes donne lieu à des mélanges plus homogènes en ce qui concerne la dimension des peptides et leurs propriétés physico-chimiques. Ceci offre l'opportunité d'une plus grande chance à certains peptides d'être détectés. De ce fait, la combinaison de fractions a été montrée comme étant bénéfique, en particulier pour les échantillons fractionnés à pH acide.

Finalement, même si la combinaison de plusieurs fractions a été présentée comme étant particulièrement bénéfique dans le cas du Lp-RP, par exemple, cette action augmente tout de même les risques de perte d'échantillon et devrait être choisie avec précaution. Le choix de la méthode de fractionnement demeure une décision cruciale avec un impact significatif dans la productivité ainsi que sur les résultats analytiques.

ANNEXE D

SCHÉMAS DE FRAGMENTATION SUGGÉRÉS POUR LE CYS-DOPAC ET LE CYS-DOPAMINE

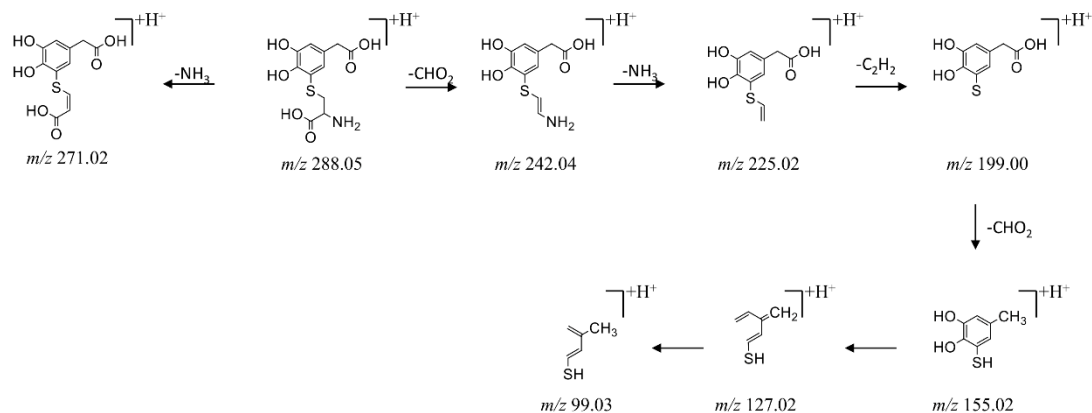


Schéma de fragmentation proposé pour la Cys-DOPAC

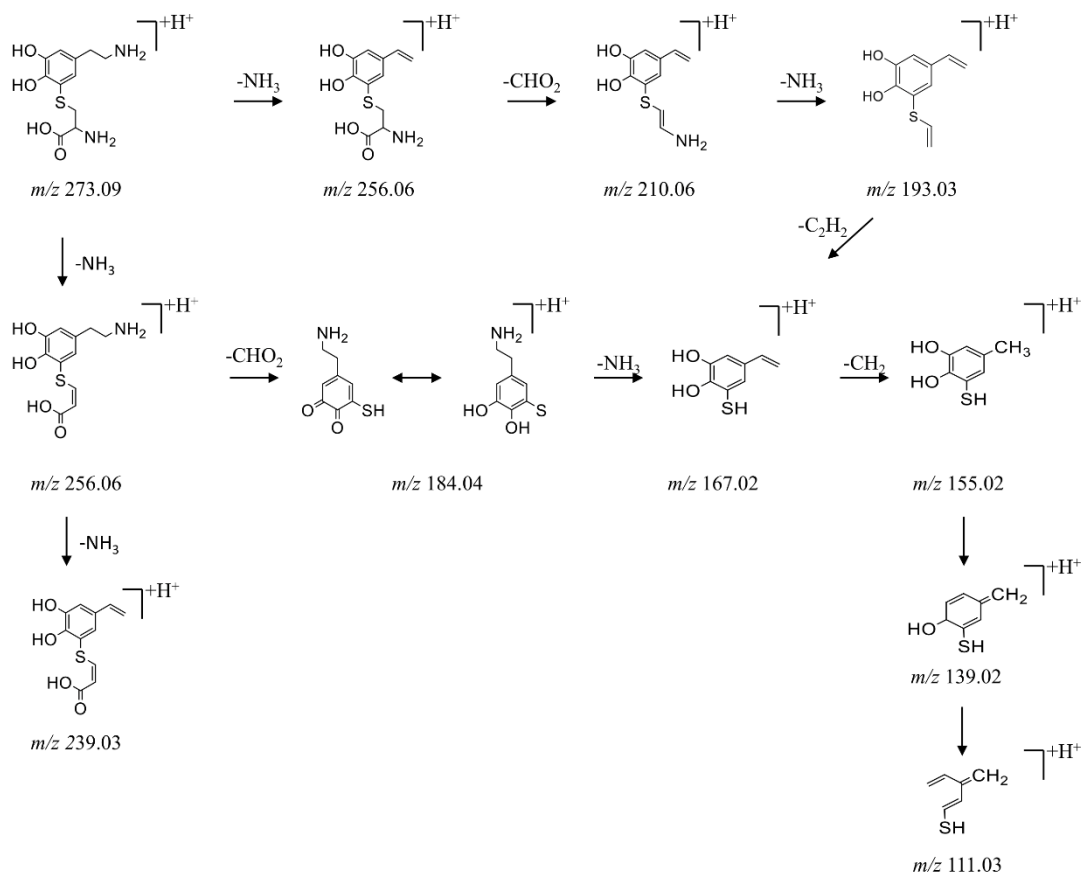


Schéma de fragmentation proposé pour la Cys-dopamine

RÉFÉRENCES

- Armirotti, A., & Damonte, G. (2010). Achievements and perspectives of top-down proteomics. *Proteomics*, 10(20), 3566–3576. <https://doi.org/10.1002/pmic.201000245>
- Arnold, A. A., Byette, F., Séguin-Heine, M. O., Leblanc, A., Sleno, L., Tremblay, R., ... Marcotte, I. (2013). Solid-state NMR structure determination of whole anchoring threads from the blue mussel *mytilus edulis*. *Biomacromolecules*, 14(1), 132–141. <https://doi.org/10.1021/bm301493u>
- Babarro, J. M. F., & Carrington, E. (2011). Byssus secretion of *Mytilus galloprovincialis*: Effect of site at macro- and micro-geographical scales within Ría de Vigo (NW Spain). *Marine Ecology Progress Series*, 435(August), 125–140. <https://doi.org/10.3354/meps09200>
- Batth, T. S., Francavilla, C., & Olsen, J. V. (2014). Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *Journal of Proteome Research*, 13, 6176–6186.
- Beaudry, F., & Vachon, P. (2006). Electrospray ionization suppression, a physical or a chemical phenomenon? *Biomedical Chromatography*, 20(2), 200–205. <https://doi.org/10.1002/bmc.553>
- Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., ... Silberring, J. (2007). Methods for samples preparation in proteomic research. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 849(1–2), 1–31. <https://doi.org/10.1016/j.jchromb.2006.10.040>
- Boichenko, A. P., Govorukhina, N., Van Der Zee, A. G. J., & Bischoff, R. (2013). Multidimensional separation of tryptic peptides from human serum proteins using reversed-phase, strong cation exchange, weak anion exchange, and fused-core fluorinated stationary phases. *Journal of Separation Science*, 36(21–22), 3463–3470. <https://doi.org/10.1002/jssc.201300750>
- Bolton, J. L., Turnipseed, S. B., & Thompson, J. A. (1997). Influence of quinone methide reactivity on the alkylation of thiol and amino groups in proteins: Studies

- utilizing amino acid and peptide models. *Chemico-Biological Interactions*, 107(3), 185–200. [https://doi.org/10.1016/S0009-2797\(97\)00079-3](https://doi.org/10.1016/S0009-2797(97)00079-3)
- Bonfiglio, R., King, R. C., Olah, T. V., & Merkle, K. (1999). The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. *Rapid Communications in Mass Spectrometry*, 13(12), 1175–1185. [https://doi.org/10.1002/\(SICI\)1097-0231\(19990630\)13:12<1175::AID-RCM639>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0231(19990630)13:12<1175::AID-RCM639>3.0.CO;2-0)
- Bonneil, E., Pfammatter, S., & Thibault, P. (2015). Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *Journal of Mass Spectrometry*, 50(11), 1181–1195. <https://doi.org/10.1002/jms.3646>
- Bouhleb, Z., Genard, B., Ibrahim, N., Carrington, E., Babarro, J. M. F., Lok, A., ... Marcotte, I. (2017). Interspecies comparison of the mechanical properties and biochemical composition of byssal threads. *The Journal of Experimental Biology*, 220(6), 984–994. <https://doi.org/10.1242/jeb.141440>
- Brazee, S. L., & Carrington, E. (2006). Interspecific comparison of the mechanical properties of mussel byssus. *Biological Bulletin*, 211(3), 263–274. <https://doi.org/10.2307/4134548>
- Breci, L. A., Tabb, D. L., Yates, J. R., & Wysocki, V. H. (2003). Cleavage N-terminal to proline: Analysis of a database of peptide tandem mass spectra. *Analytical Chemistry*, 75(9), 1963–1971. <https://doi.org/10.1021/ac026359i>
- Buhrman, D. L., Price, P. I., & Rudewicz, P. J. (1996). A Study of Ion Suppression. *Journal of the American Society for Mass Spectrometry*, 7(11), 1099–1105. <https://doi.org/10.1021/jasms.8b00806>
- Burbulla, L. F., Song, P., Mazzulli, J. R., Zampese, E., Wong, Y. C., Jeon, S., ... Krainc, D. (2017). Dopamine oxidation mediates mitochondrial and lysosomal dysfunction in Parkinson's disease. *Science*, 357(6357), 1255–1261. <https://doi.org/10.1126/science.aam9080>
- Burzio, L. A., & Waite, J. H. (2000). Cross-linking in adhesive quinoproteins: Studies with model decapeptides. *Biochemistry*, 39(36), 11147–11153. <https://doi.org/10.1021/bi0002434>
- Byette, F., Laventure, A., Marcotte, I., & Pellerin, C. (2016). Metal-Ligand Interactions and Salt Bridges as Sacrificial Bonds in Mussel Byssus-Derived Materials. *Biomacromolecules*, 17(10), 3277–3286. <https://doi.org/10.1021/acs.biomac.6b01010>

- Byette, F., Pellerin, C., & Marcotte, I. (2014). Self-assembled pH-responsive films prepared from mussel anchoring threads. *Journal of Materials Chemistry B*, 2(37), 6378–6386. <https://doi.org/10.1039/c4tb01021c>
- Bylda, C., Thiele, R., Kobold, U., & Volmer, D. A. (2014). Recent advances in sample preparation techniques to overcome difficulties encountered during quantitative analysis of small molecules from biofluids using LC-MS/MS. *Analyst*, 139(10), 2265–2276. <https://doi.org/10.1039/c4an00094c>
- Calligaris, D., Villard, C., & Lafitte, D. (2011). Advances in top-down proteomics for disease biomarker discovery. *Journal of Proteomics*, 74(7), 920–934. <https://doi.org/10.1016/j.jprot.2011.03.030>
- Callipo, L., Capriotti, A. L., Cavaliere, C., Gubbiotti, R., Samperi, R., & Laganà, A. (2011). Evaluation of different two-dimensional chromatographic techniques for proteomic analysis of mouse cardiac tissue. *Biomedical Chromatography*, 25(5), 594–599. <https://doi.org/10.1002/bmc.1487>
- Cech, N. B., Krone, J. R., & Enke, C. G. (2001). Predicting electrospray response from chromatographic retention time. *Analytical Chemistry*, 73(2), 208–213. <https://doi.org/10.1021/ac0006019>
- Chai, Y., Weng, G., Shen, S., Sun, C., & Pan, Y. (2015). The protonation site of para-dimethylaminobenzoic acid using atmospheric pressure ionization methods. *Journal of the American Society for Mass Spectrometry*, 26(4), 668–676. <https://doi.org/10.1007/s13361-014-1069-7>
- Choi, B. K., Gusev, A. I., & Hercules, D. M. (1999). Postcolumn introduction of an internal standard for quantitative LC-MS analysis. *Analytical Chemistry*, 71(18), 4107–4110. <https://doi.org/10.1021/ac990312o>
- Choudhary, G., Wu, S. L., Shieh, P., & Hancock, W. S. (2003). Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *Journal of Proteome Research*, 2(1), 59–67. <https://doi.org/10.1021/pr025557n>
- Cotter, R. J. (1997). *Time-of-Flight Mass Spectrometry - Instrumentation and Applications in Biological Research*.
- Cottrell, J. S. (2011). Protein identification using MS/MS data. *Journal of Proteomics*, 74, 1842–1851.
- Dai, J., Shieh, C. H., Sheng, Q. H., Zhou, H., & Zeng, R. (2005). Proteomic analysis with integrated multiple dimensional liquid chromatography/mass spectrometry

- based on elution of ion exchange column using pH steps. *Analytical Chemistry*, 77(18), 5793–5799. <https://doi.org/10.1021/ac050251w>
- Danner, E. W., Kan, Y., Hammer, M. U., Israelachvili, J. N., & Waite, J. H. (2012). Adhesion of mussel foot protein Mefp-5 to mica: An underwater superglue. *Biochemistry*, 51(33), 6511–6518. <https://doi.org/10.1021/bi3002538>
- DeMartini, D. G., Errico, J. M., Sjoestroem, S., Fenster, A., & Waite, J. H. (2017). A cohort of new adhesive proteins identified from transcriptomic analysis of mussel foot glands. *Journal of the Royal Society Interface*, 14(131). <https://doi.org/10.1098/rsif.2017.0151>
- Deming, T. J. (1999). Mussel byssus and biomolecular materials. *Current Opinion in Chemical Biology*, 3(1), 100–105. [https://doi.org/10.1016/S1367-5931\(99\)80018-0](https://doi.org/10.1016/S1367-5931(99)80018-0)
- Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., ... Vizcaino, J. A. (2017). The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research*, 45(D1), D1100–D1106. <https://doi.org/10.1093/nar/gkw936>
- Di Palma, S., Hennrich, M. L., Heck, A. J. R., & Mohammed, S. (2012). Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of Proteomics*, 75(13), 3791–3813. <https://doi.org/10.1016/j.jprot.2012.04.033>
- Domon, B., & Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology*, 28(7), 710–721. <https://doi.org/10.1038/nbt.1661>
- Dowell, J. A., Frost, D. C., Zhang, J., & Li, N. (2008). Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Analytical Chemistry*, 80(17), 6715–6723. <https://doi.org/10.1021/ac8007994>
- Dunn, B. M. (2002). Structure and mechanism of the Pepsin-like family of aspartic peptidases. *Chemical Reviews*, 102(12), 4431–4458. <https://doi.org/10.1021/cr010167q>
- Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M., Ens, W., Standing, K. G., ... Krokhn, O. V. (2008). Practical Implementation of 2D HPLC Scheme with Accurate Peptide Retention Prediction in Both Dimensions for High-Throughput Bottom-Up Proteomics. *Analytical Chemistry*, 80(18), 7036–7042. <https://doi.org/10.1021/ac800984n>

- Eichhorn, P., Pérez, S., & Barceló, D. (2012). Time-of-Flight Mass Spectrometry Versus Orbitrap-Based Mass Spectrometry for the Screening and Identification of Drugs and Metabolites: Is There a Winner? In *Comprehensive Analytical Chemistry* (Vol. 58). <https://doi.org/10.1016/B978-0-444-53810-9.00009-2>
- Falkenby, L. G., Such-Sanmartín, G., Larsen, M. R., Vorm, O., Bache, N., & Jensen, O. N. (2014). Integrated solid-phase extraction-capillary liquid chromatography (speLC) interfaced to ESI-MS/MS for fast characterization and quantification of protein and proteomes. *Journal of Proteome Research*, 13(12), 6169–6175. <https://doi.org/10.1021/pr5008575>
- Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 146(4929), 64–71.
- Fields, S. (2001). Proteomics in genomeland. *Science*, 291(5507), 1221–1224. <https://doi.org/10.1126/science.291.5507.1221>
- Fisher, a a, Labenski, M. T., Malladi, S., Gokhale, V., Bowen, M. E., Milleron, R. S., ... Lau, S. S. (2007). Quinone rxn Electrophile Binding Motif Cytochrome c. *Biochemistry*, 46, 11090–11100. Retrieved from http://pubs3.acs.org/acs/journals/doilookup?in_doi=10.1021/bi700613w
- Fountoulakis, M., & Lahm, H.-W. (1998). Hydrolysis and amino acid composition analysis of proteins. *Journal of Chromatography A*, 826(2), 109–134.
- Furey, A., Moriarty, M., Bane, V., Kinsella, B., & Lehane, M. (2013). Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta*, 115, 104–122. <https://doi.org/10.1016/j.talanta.2013.03.048>
- Galligan, J. J., Rose, K. L., Beavers, W. N., Hill, S., Tallman, K. A., Tansey, W. P., & Marnett, L. J. (2014). Stable histone adduction by 4-oxo-2-nonenal: A potential link between oxidative stress and epigenetics. *Journal of the American Chemical Society*, 136(34), 11864–11866. <https://doi.org/10.1021/ja503604t>
- Gantayet, A., Ohana, L., & Sone, E. D. (2013). Byssal proteins of the freshwater zebra mussel, *Dreissena polymorpha*. *Biofouling*, 29(1), 77–85. <https://doi.org/10.1080/08927014.2012.746672>
- Gantayet, A., Rees, D. J., & Sone, E. D. (2014). Novel Proteins Identified in the Insoluble Byssal Matrix of the Freshwater Zebra Mussel. *Marine Biotechnology*, 16(2), 144–155. <https://doi.org/10.1007/s10126-013-9537-9>
- Geib, T., Thulasingham, M., Haeggström, J. Z., & Sleno, L. (2020). Investigation of

- Clozapine and Olanzapine Reactive Metabolite Formation and Protein Binding by Liquid Chromatography-Tandem Mass Spectrometry. *Chemical Research in Toxicology*. <https://doi.org/10.1021/acs.chemrestox.0c00191>
- Ghosh, C., Shinde, C. P., & Chakraborty, B. S. (2012). Influence of ionization source design on matrix effects during LC-ESI-MS/MS analysis. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 893–894, 193–200. <https://doi.org/10.1016/j.jchromb.2012.03.012>
- Giddings, J. C. (1984). Two-dimensional separations: concept and promise. *Analytical Chemistry*, 56(12). <https://doi.org/10.1021/ac00276a003>
- Gilar, M., Olivova, P., Daly, A. E., & Gebler, J. C. (2005a). Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry*, 77(19), 6426–6434. <https://doi.org/10.1021/ac050923i>
- Gilar, M., Olivova, P., Daly, A. E., & Gebler, J. C. (2005b). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *Journal of Separation Science*, 28(14), 1694–1703. <https://doi.org/10.1002/jssc.200500116>
- Giroto, S., Sturlese, M., Bellanda, M., Tessari, I., Cappellini, R., Bisaglia, M., ... Mammi, S. (2012). Dopamine-derived quinones affect the structure of the redox sensor DJ-1 through modifications at Cys-106 and Cys-53. *Journal of Biological Chemistry*, 287(22), 18738–18749. <https://doi.org/10.1074/jbc.M111.311589>
- Glish, G. L., & Burinsky, D. J. (2008). Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, 19(2), 161–172. <https://doi.org/10.1016/j.jasms.2007.11.013>
- Gokce, E., Andrews, G. L., Dean, R. A., & Muddiman, D. C. (2011). Increasing proteome coverage with offline RP HPLC coupled to online RP nanoLC-MS. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 879(9–10), 610–614. <https://doi.org/10.1016/j.jchromb.2011.01.032>
- Golizeh, M., Geib, T., & Sleno, L. (2016). Identification of 4-hydroxynonenal protein targets in rat, mouse and human liver microsomes by two-dimensional liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM*, 30(13), 1488–1494. <https://doi.org/10.1002/rcm.7577>
- Golizeh, M., Leblanc, A., & Sleno, L. (2015). Identification of Acetaminophen Adducts of Rat Liver Microsomal Proteins using 2D-LC-MS/MS. *Chemical Research in Toxicology*, 28(11), 2142–2150.

<https://doi.org/10.1021/acs.chemrestox.5b00317>

- Golizeh, M., Schneider, C., Ohlund, L. B., & Sleno, L. (2015). Multidimensional LC-MS/MS analysis of liver proteins in rat, mouse and human microsomal and S9 fractions. *EuPA Open Proteomics*, 6, 16–27. <https://doi.org/10.1016/j.euprot.2015.01.003>
- Golizeh, M., & Sleno, L. (2013). Optimized proteomic analysis of rat liver microsomes using dual enzyme digestion with 2D-LC-MS/MS. *Journal of Proteomics*, 82, 166–178. <https://doi.org/10.1016/j.jprot.2013.02.001>
- Gosetti, F., Mazzucco, E., Gianotti, V., Polati, S., & Gennaro, M. C. (2007). High performance liquid chromatography/tandem mass spectrometry determination of biogenic amines in typical Piedmont cheeses. *Journal of Chromatography A*, 1149(2), 151–157. <https://doi.org/10.1016/j.chroma.2007.02.097>
- Grabnar, I., Vovk, T., Kores Plesnicar, B., & Boskovic, M. (2011). Oxidative Stress in Schizophrenia. *Current Neuropharmacology*, 9(2), 301–312. <https://doi.org/10.2174/157015911795596595>
- Graves, D. J., Martin, B. L., & Wang, J. H. (1994). *Co- and Post-translational Modification of Proteins* (1st ed.). Oxford University Press.
- Graves, P. R., & Haystead, T. A. J. (2002). Molecular Biologist's Guide to Proteomics. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS Microbiology and Molecular Biology Reviews*, 66(1), 39–63. <https://doi.org/10.1128/MMBR.66.1.39>
- Grix, R., Kutscher, R., Li, G., Grüner, U., Wollnik, H., & Matsuda, H. (1988). A time-of-flight mass analyzer with high resolving power. *Rapid Communications in Mass Spectrometry*, 2(5), 83–85. <https://doi.org/10.1002/rcm.1290020503>
- Gygi, S. P., Rist, B., Griffin, T. J., Eng, J., & Aebersold, R. (2002). Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *Journal of Proteome Research*, 1(1), 47–54. <https://doi.org/10.1021/pr015509n>
- Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A., & Burlingame, A. L. (2003). Mass spectrometric analysis of protein mixtures at low levels using cleavable ¹³C-isotope-coded affinity tag and multidimensional chromatography. *Molecular & Cellular Proteomics: MCP*, 2(5), 299–314. <https://doi.org/10.1074/mcp.M300021-MCP200>
- Hao, P., Ren, Y., Dutta, B., & Sze, S. K. (2013). Comparative evaluation of

- electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and high-pH reversed phase (Hp-RP) chromatography in profiling of rat kidney proteome. *Journal of Proteomics*, 82, 254–262. <https://doi.org/10.1016/j.jprot.2013.02.008>
- Harrington, M. J., & Waite, J. H. (2007). Holdfast heroics: comparing the molecular and mechanical properties of *Mytilus californianus* byssal threads. *Journal of Experimental Biology*, 210(24), 4307–4318. <https://doi.org/10.1242/jeb.009753>
- Harrington, Masic, A., Holten-Andersen, N., Waite, J. H., & Fratzl, P. (2010). Iron-Clad Fibers: A Metal-Based Biological Strategy for Hard Flexible Coating. *Science*, 328, 216–221.
- Hinzke, T., Kouris, A., Hughes, R. A., Strous, M., & Kleiner, M. (2019). More is not always better: Evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. *Frontiers in Microbiology*, 10(FEB), 1–13. <https://doi.org/10.3389/fmicb.2019.00238>
- Holland, J. F., Enke, C. G., Allison, J., Stults, J. T., Pinkston, J. D., Newcome, B., & Watson, J. T. (1983). Mass Spectrometry on the Chromatographic Time Scale: Realistic Expectations. *Analytical Chemistry*, 55(9), 997A-1012A. <https://doi.org/10.1021/ac00260a003>
- Huang, X., Xu, R., Hawley, M. D., Hopkins, T. L., & Kramer, K. J. (1998). Electrochemical oxidation of N-acyldopamines and regioselective reactions of their quinones with N-acetylcysteine and thiourea. *Archives of Biochemistry and Biophysics*, 352(1), 19–30. <https://doi.org/10.1006/abbi.1997.0567>
- Hustoft, H. K., Brandtzaeg, O. K., Røgeberg, M., Misaghian, D., Tørrsetnes, S. B., Greibrokk, T., ... Lundanes, E. (2013). Integrated enzyme reactor and high resolving chromatography in “sub-chip” dimensions for sensitive protein mass spectrometry. *Scientific Reports*, 3, 1–7. <https://doi.org/10.1038/srep03511>
- Ikonoraou, M. G., Blades, A. T., & Kebarle, P. (1990). Investigations of the Electrospray Interface for Liquid Chromatography/Mass Spectrometry. *Analytical Chemistry*, 62(9), 957–967. <https://doi.org/10.1021/ac00208a012>
- Jagtap, P., Bandhakavi, S., Higgins, L., McGowan, T., Sa, R., Stone, M. D., ... Griffin, T. J. (2012). Workflow for analysis of high mass accuracy salivary data set using MaxQuant and ProteinPilot search algorithm. *Proteomics*, 12(11), 1726–1730. <https://doi.org/10.1002/pmic.201100097>
- Jain, M., Ngoy, S., Sheth, S. A., Swanson, R. A., Rhee, E. P., Liao, R., ... Nilsson, R. (2014). A systematic survey of lipids across mouse tissues. *American Journal of*

Physiology - Endocrinology and Metabolism, 306(8), 854–868.
<https://doi.org/10.1152/ajpendo.00371.2013>

- Jameson, G. N. L., Zhang, J., Jameson, R. F., & Linert, W. (2004). Kinetic evidence that cysteine reacts with dopaminoquinone via reversible adduct formation to yield 5-cysteinyl-dopamine: An important precursor of neuromelanin. *Organic and Biomolecular Chemistry*, 2(5), 777–782. <https://doi.org/10.1039/b316294j>
- Jee, J., Park, S., & Kim, H. (2000). *Tyrosinase-induced cross-linking of tyrosine-containing peptides investigated by matrix- assisted laser desorption / ionization time-of-flight mass spectrometry*. 1567, 1563–1567.
- Jensen, O. N. (2004). Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology*, 8(1), 33–41. <https://doi.org/10.1016/j.cbpa.2003.12.009>
- Johnson, R. S., & Walsh, K. A. (1992). Sequence analysis of peptide mixtures by automated integration of Edman and mass spectrometric data. *Protein Science*, 1(9), 1083–1091. <https://doi.org/10.1002/pro.5560010902>
- Johnson, J. V., Yost, R. A., Kelley, P. E., & Bradford, D. C. (1990). Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps. *Analytical Chemistry*, 62(20), 2162–2172. <https://doi.org/10.1021/ac00219a003>
- Jones, J. J., Borgmann, S., Wilkins, C. L., & O'Brien, R. M. (2006). Characterizing the phospholipid profiles in mammalian tissues by MALDI FTMS. *Analytical Chemistry*, 78(9), 3062–3071. <https://doi.org/10.1021/ac0600858>
- Karas, M., & Hillenkamp, F. (1988). Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Analytical Chemistry*, 60(20), 2299–2301. <https://doi.org/10.1021/ac00171a028>
- Kearle, P., & Tang, L. (1993). From Ions in Solution To Ions in the Gas Phase. *Analytical Chemistry*, 65(22), 972A–986A. <https://doi.org/10.1021/ac00070a715>
- Kerwin, J. L., Turecek, F., Xu, R., Kramer, K. J., Hopkins, T. L., Gatlin, C. L., & Yates, J. R. (1999). Mass spectrometric analysis of catechoi-histidine adducts from insect cuticle. *Analytical Biochemistry*, 268(2), 229–237. <https://doi.org/10.1006/abio.1998.3069>
- Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Scientific Reports*, 1, 1–5. <https://doi.org/10.1038/srep00090>

- Klopčič, I., & Dolenc, M. S. (2019). Chemicals and Drugs Forming Reactive Quinone and Quinone Imine Metabolites. *Chemical Research in Toxicology*, 32(1), 1–34. <https://doi.org/10.1021/acs.chemrestox.8b00213>
- Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues - A novel approach to testing for induced point mutations in mammals. *Human Genetics*, 26(3), 231–243. <https://doi.org/10.1007/BF00281458>
- Konermann, L., Ahadi, E., Rodriguez, A. D., & Vahidi, S. (2013). Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry*, 85, 2–9.
- Kreimer, S., Belov, M. E., Danielson, W. F., Levitsky, L. I., Gorshkov, M. V., Karger, B. L., & Ivanov, A. R. (2016). Advanced Precursor Ion Selection Algorithms for Increased Depth of Bottom-Up Proteomic Profiling. *Journal of Proteome Research*, 15(10), 3563–3573. <https://doi.org/10.1021/acs.jproteome.6b00312>
- Labenski, M. T., Fisher, A. A., Lo, H. H., Monks, T. J., & Lau, S. S. (2009). Protein electrophile-binding motifs: Lysine-rich proteins are preferential targets of quinones. *Drug Metabolism and Disposition*, 37(6), 1211–1218. <https://doi.org/10.1124/dmd.108.026211>
- Law, H. C. H., Kong, R. P. W., Szeto, S. S. W., Zhao, Y., Zhang, Z., Wang, Y., ... Chu, I. K. (2015). A versatile reversed phase-strong cation exchange-reversed phase (RP-SCX-RP) multidimensional liquid chromatography platform for qualitative and quantitative shotgun proteomics. *Analyst*, 140(4), 1237–1252. <https://doi.org/10.1039/c4an01893a>
- Leblanc, A., Arnold, A. A., Genard, B., Nadalini, J. B., Heine, M. O. S., Marcotte, I., ... Sleno, L. (2012). Determination of isotopic labeling of proteins by precursor ion scanning liquid chromatography/tandem mass spectrometry of derivatized amino acids applied to nuclear magnetic resonance studies. *Rapid Communications in Mass Spectrometry*, 26(10), 1165–1174. <https://doi.org/10.1002/rcm.6204>
- Lee, B. P., Messersmith, P. B., Israelachvili, J. N., & Waite, J. H. (2011). Mussel-Inspired Adhesives and Coatings. *Annual Review of Materials Research*, 41(1), 99–132. <https://doi.org/10.1146/annurev-matsci-062910-100429>
- Lee, H., Mun, D.-G. G., So, J. E., Bae, J., Kim, H., Masselon, C., & Lee, S.-W. W. (2016). Efficient Exploitation of Separation Space in Two-Dimensional Liquid Chromatography System for Comprehensive and Efficient Proteomic Analyses. *Analytical Chemistry*, 88(23), 11734–11741. <https://doi.org/10.1021/acs.analchem.6b03366>

- Li, S., Xia, Z., Chen, Y., Gao, Y., & Zhan, A. (2018). Byssus structure and protein composition in the highly invasive Fouling Mussel *Limnoperna fortunei*. *Frontiers in Physiology*, 9(APR), 1–14. <https://doi.org/10.3389/fphys.2018.00418>
- Li, Y., Jongberg, S., Andersen, M. L., Davies, M. J., & Lund, M. N. (2016). Quinone-induced protein modifications: Kinetic preference for reaction of 1,2-benzoquinones with thiol groups in proteins. *Free Radical Biology and Medicine*, 97, 148–157. <https://doi.org/10.1016/j.freeradbiomed.2016.05.019>
- Liebler, D. C. (2002a). *Introduction to Proteomics*. Humana Press.
- Liebler, D. C. (2002b). Proteomic approaches to characterize protein modifications: New tools to study the effects of environmental exposures. *Environmental Health Perspectives*, 110(SUPPL. 1), 3–9. <https://doi.org/10.1289/ehp.02110s113>
- Lin, Q., Gourdon, D., Sun, C., Holten-Andersen, N., Anderson, T. H., Waite, J. H., & Israelachvili, J. N. (2007). Adhesion mechanisms of the mussel foot proteins mfp-1 and mfp-3. *Proceedings of the National Academy of Sciences of the United States of America*, 104(10), 3782–3786. <https://doi.org/10.1073/pnas.0607852104>
- López, J. L., Marina, A., Vázquez, J., & Alvarez, G. (2002). A proteomic approach to the study of the marine mussels *Mytilus edulis* and *M. galloprovincialis*. *Marine Biology*, 141(2), 217–223. <https://doi.org/10.1007/s00227-002-0827-4>
- Lubec, G., & Afjehi-Sadat, L. (2007). Limitations and pitfalls in protein identifications by mass spectrometry. *Chemical Reviews*, 107(8), 3568–3584. <https://doi.org/10.1021/cr068213f>
- Luo, J., Bassett, J., & Ranish, J. (2019). Identification of Cross-linked Peptides Using Isotopomeric Cross-linkers. *Journal of the American Society for Mass Spectrometry*, 30(9), 1643–1653. <https://doi.org/10.1007/s13361-019-02253-z>
- Manadas, B., Mendes, V. M., English, J., & Dunn, M. J. (2010). Peptide fractionation in proteomics approaches. *Expert Review of Proteomics*, 7(5), 655–663. <https://doi.org/10.1586/epr.10.46>
- Mann, M., & Wilm, M. (1994). Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Analytical Chemistry*, 66(24), 4390–4399. <https://doi.org/10.1021/ac00096a002>
- Maux, D., Enjalbal, C., Martinez, J., & Aubagnac, J. L. (2002). New example of proline-induced fragmentation in electrospray ionization mass spectrometry of peptides. *Rapid Communications in Mass Spectrometry*, 16(15), 1470–1475.

<https://doi.org/10.1002/rcm.741>

- McDowell, L. M., Burzio, L. A., Waite, J. H., & Schaefer, J. (1999). Rotational echo double resonance detection of cross-links formed in mussel byssus under high-flow stress. *Journal of Biological Chemistry*, 274(29), 20293–20295. <https://doi.org/10.1074/jbc.274.29.20293>
- McLafferty, F. W. (1981). Mass spectrometry Mass spectrometry. *Science*, 214, 280–287.
- Meiser, J., Weindl, D., & Hiller, K. (2013). Complexity of dopamine metabolism. *Cell Communication and Signaling*, 11(1), 1–18. <https://doi.org/10.1186/1478-811X-11-34>
- Michalski, A., Cox, J., & Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research*, 10(4), 1785–1793. <https://doi.org/10.1021/pr101060v>
- Miserez, A., Schneberk, T., Sun, C., Zok, F. W., & Waite, J. H. (2008). The transition from stiff to compliant materials in squid beaks. *Science*, 319(5871), 1816–1819. <https://doi.org/10.1126/science.1154117>
- Mitreá, N., Leblanc, A., St-Onge, M., & Sleno, L. (2010). Assessing covalent binding of reactive drug metabolites by complete protein digestion and LC-MS analysis. *Bioanalysis*, 2(7), 1211–1221. <https://doi.org/10.4155/bio.10.74>
- Mommen, G. P. M., Meiring, H. D., Heck, A. J. R., & De Jong, A. P. J. M. (2013). Mixed-bed ion exchange chromatography employing a salt-free pH gradient for improved sensitivity and compatibility in MudPIT. *Analytical Chemistry*, 85(14), 6608–6616. <https://doi.org/10.1021/ac400995e>
- Monks, T., & Jones, D. (2002). The Metabolism and Toxicity of Quinones, Quinonimines, Quinone Methides, and Quinone-Thioethers. *Current Drug Metabolism*, 3(4), 425–438. <https://doi.org/10.2174/1389200023337388>
- Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., & Yates, J. R. (2007). Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Analytical Chemistry*, 79(10), 3623–3634. <https://doi.org/10.1021/ac062292d>
- Mujezinovic, N., Raidl, G., Hutchins, J. R. A., Peters, J. M., Mechtler, K., & Eisenhaber, F. (2006). Cleaning of raw peptide MS/MS spectra: Improved protein identification following deconvolution of multiply charged peaks, isotope clusters,

- and removal of background noise. *Proteomics*, 6(19), 5117–5131. <https://doi.org/10.1002/pmic.200500928>
- Muñoz-Muñoz, J. L., Acosta-Motos, J. R., Garcia-Molina, F., Varon, R., Garcia-Ruiz, P. A., Tudela, J., ... Rodríguez-López, J. N. (2010). Tyrosinase inactivation in its action on dopa. *Biochimica et Biophysica Acta - Proteins and Proteomics*, Vol. 1804, pp. 1467–1475. <https://doi.org/10.1016/j.bbapap.2010.02.015>
- Nam, E., Derrick, J. S., Lee, S., Kang, J., Han, J., Lee, S. J. C., ... Lim, M. H. (2018). Regulatory Activities of Dopamine and Its Derivatives toward Metal-Free and Metal-Induced Amyloid- β Aggregation, Oxidative Stress, and Inflammation in Alzheimer's Disease. *ACS Chemical Neuroscience*, 9(11), 2655–2666. <https://doi.org/10.1021/acscchemneuro.8b00122>
- Namera, A., & Saito, T. (2013). Recent advances in unique sample preparation techniques for bioanalysis. *Bioanalysis*, 5(8), 915–932. <https://doi.org/10.4155/bio.13.52>
- Nie, L., Wu, G., Culley, D. E., Scholten, J. C. M., & Zhang, W. (2007). Integrative analysis of transcriptomic and proteomic data: Challenges, solutions and applications. *Critical Reviews in Biotechnology*, 27(2), 63–75. <https://doi.org/10.1080/07388550701334212>
- O'Farrell, P. H. (1975). High Resolution of Proteins * Electrophoresis. *The Journal of Biological Chemistry*, 250(10), 4007–4021. <https://doi.org/10.1016/j.bbi.2008.05.010>
- Ousji, O., Ohlund, L., & Sleno, L. (2020). Comprehensive in Vitro Metabolism Study of Bisphenol A Using Liquid Chromatography-High Resolution Tandem Mass Spectrometry. *Chemical Research in Toxicology*, 33(6), 1468–1477. <https://doi.org/10.1021/acs.chemrestox.0c00042>
- Pandey, A., & Mann, M. (2000). Proteomics to study genes and genomes. *Nature*, 405(June), 837–846. <https://doi.org/10.1111/j.1749-6632.1999.tb08531.x>
- Pearce, T., & LaBarbera, M. (2009). A comparative study of the mechanical properties of Mytilid byssal threads. *Journal of Experimental Biology*, 212(10), 1442–1448. <https://doi.org/10.1242/jeb.025544>
- Petrone, L., Kumar, A., Sutanto, C. N., Patil, N. J., Kannan, S., Palaniappan, A., ... Miserez, A. (2015). Mussel adhesion is dictated by time-regulated secretion and molecular conformation of mussel adhesive proteins. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms9737>

- Priemel, T., Degtyar, E., Dean, M. N., & Harrington, M. J. (2017). Rapid self-assembly of complex biomolecular architectures during mussel byssus biofabrication. *Nature Communications*, 8. <https://doi.org/10.1038/ncomms14539>
- Priemel, T., Palia, R., Babych, M., Thibodeaux, C. J., Bourgault, S., & Harrington, M. J. (2020). Compartmentalized processing of catechols during mussel byssus fabrication determines the destiny of DOPA. *Proceedings of the National Academy of Sciences of the United States of America*, 117(14), 7613–7621. <https://doi.org/10.1073/pnas.1919712117>
- Qin, C., Pan, Q., Qi, Q., Fan, M., Sun, J., Li, N., & Liao, Z. (2016). In-depth proteomic analysis of the byssus from marine mussel *Mytilus coruscus*. *Journal of Proteomics*, 144, 87–98. <https://doi.org/10.1016/j.jprot.2016.06.014>
- Qin, X.-X., & Waite, J. H. (1998). A potential mediator of collagenous block copolymer gradients in mussel byssal threads. *Proceedings of the National Academy of Sciences*, 95(18), 10517–10522. <https://doi.org/10.1073/pnas.95.18.10517>
- Qin, X., & Waite, J. H. (1995a). Exotic collagen gradients in the byssus of the mussel *Mytilus edulis*. *The Journal of Experimental Biology*, 198(Pt 3), 633–644. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7714453>
- Qin, X., & Waite, J. H. (1995b). Exotic collagen gradients in the byssus of the mussel *Mytilus edulis*. *The Journal of Experimental Biology*, 198(Pt 3), 633–644. <https://doi.org/10.1242/jeb.198.3.633>
- Qin, X. X., & Waite, J. H. (1998). A potential mediator of collagenous block copolymer gradients in mussel byssal threads. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), 10517–10522. <https://doi.org/10.1073/pnas.95.18.10517>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), 590–596. <https://doi.org/10.1093/nar/gks1219>
- Raftos, D. A., Melwani, A. R., Haynes, P. A., Muralidharan, S., Birch, G. F., Amaral, V., ... Taylor, D. A. (2016). The biology of environmental stress: Molecular biomarkers in Sydney rock oysters (*Saccostrea glomerata*). *Environmental Science: Processes and Impacts*, 18(9), 1129–1139. <https://doi.org/10.1039/c6em00322b>
- Ramsden, C. A., & Riley, P. A. (2014). Tyrosinase: The four oxidation states of the

- active site and their relevance to enzymatic activation, oxidation and inactivation. *Bioorganic and Medicinal Chemistry*, Vol. 22, pp. 2388–2395. <https://doi.org/10.1016/j.bmc.2014.02.048>
- Rindgen, D., Seon Hwa Lee, Nakajima, M., & Blair, I. A. (2000). Formation of a substituted 1,N6-etheno-2'-deoxyadenosine adduct by lipid hydroperoxide-mediated generation of 4-oxo-2-nonenal. *Chemical Research in Toxicology*, 13(9), 846–852. <https://doi.org/10.1021/tx0000771>
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., ... Birol, I. (2010). De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), 909–912. <https://doi.org/10.1038/nmeth.1517>
- Rousu, T., Herttuainen, J., & Tolonen, A. (2010). Comparison of triple quadrupole, hybrid linear ion trap triple quadrupole, time-of-flight and LTQ-orbitrap mass spectrometers in drug discovery phase metabolite screening and identification in vitro - amitriptyline and verapamil as model compounds. *Rapid Communications in Mass Spectrometry*, 24(7), 939–957. <https://doi.org/10.1002/rcm.4465>
- Sali, A., Glaeser, R., Earnest, T., & Baumeister, W. (2003). From words to literature in structural proteomics. *Nature*, 422(6928), 216–225. <https://doi.org/10.1038/nature01513>
- Sansoucy, M., Tremblay, R., Carrington, E., Marcotte, I., & Sleno, L. (2020). Investigating Byssogenesis with Proteomic Analysis of Byssus, Foot and Mantle in Mytilus Mussels by LC-MS/MS. *Proteomics*, 2000014. <https://doi.org/10.1002/pmic.202000014>
- Schmitt, C. N. Z., Politi, Y., Reinecke, A., & Harrington, M. J. (2015). Role of sacrificial protein-metal bond exchange in mussel byssal thread self-healing. *Biomacromolecules*, 16(9), 2852–2861. <https://doi.org/10.1021/acs.biomac.5b00803>
- Seguin-Heine, M. O., Lachance, A. A., Genard, B., Myrand, B., Pellerin, C., Marcotte, I., & Tremblay, R. (2014). Impact of open sea habitat on byssus attachment of suspension-cultured blue mussels (*Mytilus edulis*). *Aquaculture*, 426–427, 189–196. <https://doi.org/10.1016/j.aquaculture.2014.02.006>
- Shen, X. M., Xia, B., Wrona, M. Z., & Dryhurst, G. (1996). Synthesis, redox properties, in vivo formation, and neurobehavioral effects of N-acetylcysteinyl conjugates of dopamine: Possible metabolites of relevance to Parkinson's disease. *Chemical Research in Toxicology*, 9(7), 1117–1126. <https://doi.org/10.1021/tx960052v>
- Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., ...

- Schaeffer, D. A. (2007). The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. *Molecular & Cellular Proteomics*, 6(9), 1638–1655. <https://doi.org/10.1074/mcp.T600050-MCP200>
- Simmons, A. H., Michal, C. A., & Jelinski, L. W. (2010). Molecular Orientation and Two-Component Nature of the Crystalline Fraction of Spider Dragline Silk Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/2890379>. *Advancement Of Science*, 271(5245), 84–87.
- Slawski, M., Hussong, R., Tholey, A., Jakoby, T., Gregorius, B., Hildebrandt, A., & Hein, M. (2012). Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching. *BMC Bioinformatics*, 13(1), 1–18. <https://doi.org/10.1186/1471-2105-13-291>
- Song, C., Yu, Z., Zou, H., Han, G., Jiang, X., Ye, M., ... Chen, R. (2009). Reversed-Phase-Reversed-Phase Liquid Chromatography Approach with High Orthogonality for Multidimensional Separation of Phosphopeptides. *Analytical Chemistry*, 82(1), 53–56. <https://doi.org/10.1021/ac9023044>
- Song, C., Yu, Z., Zou, H., Han, G., Jiang, X., Ye, M., ... Zou, H. (2010). Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Analytical Chemistry*, 82(1), 53–56. <https://doi.org/10.1021/ac9023044>
- Spickett, C. M. (2013). The lipid peroxidation product 4-hydroxy-2-nonenal: Advances in chemistry and analysis. *Redox Biology*, Vol. 1, pp. 145–152. <https://doi.org/10.1016/j.redox.2013.01.007>
- Srinivas, N. R. (2009). Dodging matrix effects in liquid chromatography tandem mass spectrometric assays - Compilation of key learnings and perspectives. *Biomedical Chromatography*, 23(5), 451–454. <https://doi.org/10.1002/bmc.1152>
- Suhre, M. H., Gertz, M., Steegborn, C., & Scheibel, T. (2014). Structural and functional features of a collagen-binding matrix protein from the mussel byssus. *Nature Communications*, 5, 3392. <https://doi.org/10.1038/ncomms4392>
- Taylor, S. W., Waite, J. H., Ross, M. M., Shabanowitz, J., & Hunt, D. F. (1994). trans-2, 3-cis-3, 4-Dihydroxyproline, a New Naturally Occurring Amino Acid, Is the Sixth Residue in the Tandemly Repeated Consensus Decapeptides of an Adhesive Protein from *Mytilus edulis*. *Journal of the American Chemical Society*, 116(23), 10803–10804. <https://doi.org/10.1021/ja00102a063>

- Testa, B., Pedretti, A., & Vistoli, G. (2012). Reactions and enzymes in the metabolism of drugs and other xenobiotics. *Drug Discovery Today*, 17(11–12), 549–560. <https://doi.org/10.1016/j.drudis.2012.01.017>
- Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., & Mann, M. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular and Cellular Proteomics*, 10(8), 1–9. <https://doi.org/10.1074/mcp.M110.003699>
- Tomlinson, A. J., & Chiczy, R. M. (2003). Microcapillary liquid chromatography/tandem mass spectrometry using alkaline pH mobile phases and positive ion detection. *Rapid Communications in Mass Spectrometry*, 17(9), 909–916. <https://doi.org/10.1002/rcm.1001>
- Tsai, T. H., Song, E., Zhu, R., Di Poto, C., Wang, M., Luo, Y., ... Ransom, H. W. (2015). LC-MS/MS-based serum proteomics for identification of candidate biomarkers for hepatocellular carcinoma. *Proteomics*, 15(13), 2369–2381. <https://doi.org/10.1002/pmic.201400364>
- Tsiatsiani, L., & Heck, A. J. R. (2015). Proteomics beyond trypsin. *FEBS Journal*, 282(14), 2612–2626. <https://doi.org/10.1111/febs.13287>
- Tubaon, R. M., Haddad, P. R., & Quirino, J. P. (2017). Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics*, 17(20). <https://doi.org/10.1002/pmic.201700011>
- Volmer, D. A., & Jessome, L. L. (2006). Ion Suppression: A Major Concern in Mass Spectrometry. *LCGL North America*, 24(5), 498–510.
- Wagner, M., Ohlund, L. B., Shiao, T. C., Vézina, A., Annabi, B., Roy, R., & Sleno, L. (2015). Isotope-labeled differential profiling of metabolites using N-benzoyloxysuccinimide derivatization coupled to liquid chromatography/high-resolution tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 29(18), 1632–1640. <https://doi.org/10.1002/rcm.7264>
- Waite, J. H. (1992). The formation of mussel byssus: anatomy of a natural manufacturing process. *Results and Problems in Cell Differentiation*, 19, 27–54. https://doi.org/10.1007/978-3-540-47207-0_2
- Waite, J. H., & Qin, X. (2001). Polyphosphoprotein from the adhesive pads of *Mytilus edulis*. *Biochemistry*, 40(9), 2887–2893. <https://doi.org/10.1021/bi002718x>
- Waite, J. Herbert. (2017). Mussel adhesion – essential footwork. *The Journal of Experimental Biology*, 220(4), 517–530. <https://doi.org/10.1242/jeb.134056>

- Wang, P., & Wilson, S. R. (2013). Mass spectrometry-based protein identification by integrating de novo sequencing with database searching. *BMC Bioinformatics*, *14* Suppl 2(Suppl 2). <https://doi.org/10.1186/1471-2105-14-S2-S24>
- Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., ... Smith, R. D. (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics*, *11*(10), 2019–2026. <https://doi.org/10.1002/pmic.201000722>
- Washburn, M. P., Wolters, D., & Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, *19*(3), 242–247. <https://doi.org/10.1038/85686>
- Wei, W., Yu, J., Broomell, C., Israelachvili, J. N., & Waite, J. H. (2012). Hydrophobic Enhancement of Dopa-Mediated Adhesion in a Mussel Foot Protein. *Journal of the American Chemical Society*, *135*(1), 377–383. <https://doi.org/10.1021/ja309590f>
- Wilker, J. J. (2010). Marine bioinorganic materials: mussels pumping iron. *Current Opinion in Chemical Biology*, *14*(2), 276–283. <https://doi.org/10.1016/j.cbpa.2009.11.009>
- Xie, F., Smith, R. D., & Shen, Y. (2012). Advanced proteomic liquid chromatography. *Journal of Chromatography A*, *1261*, 78–90. <https://doi.org/10.1016/j.chroma.2012.06.098>
- Xu, R., Huang, X., Kramer, K. J., & Hawley, M. D. (1996). Characterization of products from the reactions of dopamine quinone with N-acetylcysteine. *Bioorganic Chemistry*, *24*(1), 110–126. <https://doi.org/10.1006/bioo.1996.0011>
- Yamada, A., Yoshio, M., Oiwa, K., & Nyitray, L. (2000). Catchin, a novel protein in molluscan catch muscles, is produced by alternative splicing from the myosin heavy chain gene. *Journal of Molecular Biology*, *295*(2), 169–178. <https://doi.org/10.1006/jmbi.1999.3349>
- Yang, F., Shen, Y., Camp, D. G., & Smith, R. D. (2012). High-pH reversed-phased chromatography with fraction concatenation for 2D proteomic analysis. *Expert Reviews in Proteomics*, 129–134.
- Yang, F., Shen, Y., Camp II, D. G., & Smith, R. D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Reviews in Proteomics*, *9*, 129–134.
- Yang, J., Cohen Stuart, M. A., & Kamperman, M. (2014). Jack of all trades: Versatile

- catechol crosslinking mechanisms. *Chemical Society Reviews*, 43(24), 8271–8298. <https://doi.org/10.1039/c4cs00185k>
- Yang, J., Saggiomo, V., Velders, A. H., Stuart, M. A. C., & Kamperman, M. (2016). Reaction pathways in catechol/primary amine mixtures: A window on crosslinking chemistry. *PLoS ONE*, 11(12), 1–17. <https://doi.org/10.1371/journal.pone.0166490>
- Yeung, D., Mizero, B., Gussakovsky, D., Klaassen, N., Lao, Y., Spicer, V., & Krokhin, O. V. (2020). Separation Orthogonality in Liquid Chromatography-Mass Spectrometry for Proteomic Applications: Comparison of 16 Different Two-Dimensional Combinations. *Analytical Chemistry*, 92(5), 3904–3912. <https://doi.org/10.1021/acs.analchem.9b05407>
- Yin, X., Liu, X., Zhang, Y., Yan, G., Wang, F., Lu, H., ... Yang, P. (2014). Rapid and sensitive profiling and quantification of the human cell line proteome by LC-MS/MS without prefractionation. *Proteomics*, 14(17–18), 2008–2016. <https://doi.org/10.1002/pmic.201300510>
- Yu, J., Wei, W., Danner, E., Ashley, R. K., Israelachvili, J. N., & Waite, J. H. (2011). Mussel protein adhesion depends on interprotein thiol-mediated redox modulation. *Nature Chemical Biology*, 7(9), 588–590. <https://doi.org/10.1038/nchembio.630>
- Yu, J., Wei, W., Menyo, M. S., Masic, A., Waite, J. H., & Israelachvili, J. N. (2013). Adhesion of mussel foot protein-3 to TiO₂ surfaces: The effect of pH. *Biomacromolecules*, 14(4), 1072–1077. <https://doi.org/10.1021/bm301908y>
- Yuan, Z. F., Lin, S., Molden, R. C., & Garcia, B. A. (2014). Evaluation of proteomic search engines for the analysis of histone modifications. *Journal of Proteome Research*, 13(10), 4470–4478. <https://doi.org/10.1021/pr5008015>
- Zhang, F., & Bartels, M. J. (2004). Structural analysis of naphthoquinone protein adducts with liquid chromatography/tandem mass spectrometry and the scoring algorithm for spectral analysis (SALSA). *Rapid Communications in Mass Spectrometry*, 18(16), 1809–1816. <https://doi.org/10.1002/rcm.1552>
- Zhang, X., Monroe, M. E., Chen, B., Chin, M. H., Heibeck, T. H., Schepmoes, A. A., ... Qian, W.-J. (2010). Endogenous 3,4-Dihydroxyphenylalanine and Dopamine Modifications on Protein Tyrosine. *Molecular & Cellular Proteomics*, 9(6), 1199–1208. <https://doi.org/10.1074/mcp.M900321-MCP200>
- Zhao, H., & Waite, J. H. (2005). Coating proteins: Structure and cross-linking in fp-1 from the green shell mussel *Perna canaliculus*. *Biochemistry*, 44(48), 15915–15923. <https://doi.org/10.1021/bi051530g>

- Zhao, H., & Waite, J. H. (2006a). Linking adhesive and structural proteins in the attachment plaque of *Mytilus californianus*. *Journal of Biological Chemistry*, 281(36), 26150–26158. <https://doi.org/10.1074/jbc.M604357200>
- Zhao, H., & Waite, J. H. (2006b). Linking adhesive and structural proteins in the attachment plaque of *Mytilus californianus*. *Journal of Biological Chemistry*, 281(36), 26150–26158. <https://doi.org/10.1074/jbc.M604357200>

