



**HAL**  
open science

# La spectrométrie de masse comme nouvelle approche pour mesurer les biomarqueurs en écotoxicologie : Etude de cas chez les larves d'alose *Alosa alosa*

Anavai Tengaripa

## ► To cite this version:

Anavai Tengaripa. La spectrométrie de masse comme nouvelle approche pour mesurer les biomarqueurs en écotoxicologie : Etude de cas chez les larves d'alose *Alosa alosa*. Sciences du Vivant [q-bio]. 2023. hal-03971129

**HAL Id: hal-03971129**

**<https://hal.inrae.fr/hal-03971129>**

Submitted on 3 Feb 2023

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

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

## Rapport de stage du Master 2 Analyse et Contrôle de l'UCBL



# La spectrométrie de masse comme nouvelle approche pour mesurer les biomarqueurs en écotoxicologie : Etude de cas chez les larves d'alose *Alosa alosa*

Ce rapport fait état des résultats obtenus au cours du stage intervenant dans le cadre de la thèse de Marion BLAYA qui porte sur l'étude du déclin de populations d'aloses en Gironde. L'*Alosa alosa* ou grande alose est une espèce de la famille des *Clupeidae*, qui se reproduit en eau douce, notamment dans la Garonne et le Dordogne dont une diminution significative de la population est observée depuis quelques années. Une des hypothèses avancées pour expliquer ce déclin est la dégradation de la qualité environnementale, comme des modifications de température, de pH ou des contaminations chimiques dues aux activités humaines.

L'objectif principal de ce stage était de démontrer le potentiel de la spectrométrie de masse en mode Multiple Reaction Monitoring (MRM) pour l'identification et la quantification de protéines biomarqueurs chez une espèce d'intérêt, mais dont le génome n'est pas séquencé.

Une première étape d'état de l'art a permis de définir les stratégies les plus adéquates à cette étude. Une stratégie bottom-up pour l'identification de peptides rapporteurs a été choisie, suivie du développement d'une approche ciblée par MRM pour la quantification absolue des peptides identifiés via une stratégie AQUA. Une approche phylogénétique a permis d'identifier 14 peptides potentiellement rapporteurs de 7 protéines sélectionnées comme potentiels biomarqueurs.



Toutefois, au terme de ce stage, seuls 3 peptides obtenus par phylogénie ont pu être identifiés, détectés et quantifiés. Si la pertinence et la force de la spectrométrie de masse en écotoxicologie a pu être confirmée en écotoxicologie pour la détection et la quantification de peptides dans des matrices biologiques complexes, les limites d'une stratégie phylogénétique restent un vrai verrou chez certaines espèces. Cependant et en raison de leur forte variabilité individuelle, la quantification de ces trois peptides n'a pas permis de conclure sur l'effet du pH et de la température sur le niveau des protéines cibles.

Réalisé par Anavai TENGARIPA

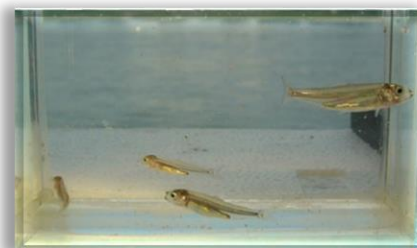
À l'Institut National de la Recherche pour l'Agriculture, l'Alimentation et  
l'Environnement

Encadré par Olivier GEFARD, Aranud SALVADOR, Anabelle ESPEYTE

Du 02 mars au 31 août 2020

Mots clés :

- Alose
- Larve
- MRM (Multiple Reaction Monitoring)
- Phylogénie
- Transition
- Peptides
- Protéines
- Biomarqueur



## **Remerciements**

Je tiens à remercier toutes les personnes qui ont contribué au succès de mon stage et qui m'ont aidée lors de la rédaction de ce mémoire.

Je voudrais dans un premier temps remercier M. Pascal BOISTARD, président du centre Lyon-Grenoble-Auvergne-Rhône-Alpes de INRAE, d'avoir accepté que je puisse effectuer mon stage de fin d'étude au sein des laboratoires INRAE à Villeurbanne.

Je tiens dans un second temps à remercier mon tuteur de stage, M. Olivier GEFFARD pour son accueil, ses recommandations et pour son expertise dont j'ai pu profiter.

De même, je tiens à remercier M<sup>lle</sup>. Anabelle ESPEYTE pour son accueil, pour tout le temps qu'elle a consacré à me former ainsi que pour tous les conseils et les réponses qu'elle m'a apportées.

En outre je souhaite remercier Messieurs Davide DEGLI-ESPOSTI et Arnaud SALVADOR pour leurs réponses aux différentes questions que je me suis posée durant le stage.

Enfin, je tiens à remercier toute l'équipe écotoxicologie pour leur très bon accueil ainsi que pour la bonne ambiance.

## Table des matières

Liste des figures .....	I
Liste des tableaux.....	II
Liste des acronymes et abréviations .....	III
I. Introduction .....	1
II. Présentation de l'entreprise .....	4
1. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement.....	4
2. Equipe Ecotoxicologie .....	4
III. Etat de l'art.....	5
1. Généralités sur l'Alose .....	5
2. Stratégie Top Down pour la spectrométrie de masse .....	6
3. Stratégie Bottom-Up pour la spectrométrie de masse .....	6
4. Approche non ciblée en spectrométrie de masse .....	7
5. Modes de balayage en spectrométrie de masse .....	9
6. Approche ciblée en spectrométrie de masse .....	10
7. Quantification relative .....	11
8. Quantification absolue .....	12
9. Fragmentation des protéines et sélection des transitions pour la spectrométrie de masse MSMS .....	13
10. Choix des protéines d'intérêt en vue d'évaluer l'état de santé des larves d'alose .....	15
11. Stratégie de recherche et d'identification des peptides protéotypiques chez une espèce non-séquencée en vue de mesurer des protéines d'intérêt .....	16
a. Recherche des peptides potentiels.....	17
IV. Matériel et méthode .....	19
1. Recherche et identification des potentiels peptides .....	19
2. Protocole d'extraction des protéines .....	20

3.	Protocole de digestion par la trypsine .....	21
4.	Protocole de concentration et purification sur phase solide (SPE) des peptides 23	
5.	Conditions d'analyse de la méthode chromatographique .....	24
6.	Conditions d'analyse de la méthode MSMS en mode MRM .....	24
a.	Etablissement des transitions MRM .....	25
b.	Optimisation des paramètres de fragmentation .....	25
c.	Optimisation des paramètres de source .....	26
7.	Application de la méthode sur des échantillons de larves à différentes conditions d'exposition (pH ou groupe de géniteur différent) .....	27
8.	Tests et conditions statistiques pour l'analyse de donnée .....	28
a.	Test ANOVA.....	28
b.	Test de Tukey.....	28
V.	Résultats et discussions .....	29
1.	Recherche et identification des potentiels peptides .....	29
2.	Paramètres de source optimisés .....	31
3.	Validation des peptides identifiés.....	32
4.	Application de la méthode pour la quantification des peptides EALVSQLTR, NALAHVQSAR et LLGYQAAAYLDK sur des échantillons de larves exposées à différents pH ou provenant de groupes de géniteurs différents .....	36
a.	Effet du pH sur la quantité de protéines .....	36
b.	Effet du groupe de géniteur sur la quantité de protéine .....	40
VI.	Bilan et perspectives.....	44
	<b>Bibliographie</b> .....	IV
	Annexe 1 : Organigramme INRAE.....	46
	Annexe 2 : Liste complète des peptides potentiels.....	47
	Annexe 3 : Résultats de l'optimisation des paramètres de fragmentation .....	50

## Liste des figures

Figure 1 : Schéma des approches Top Down et Bottom Up pour la protéomique décrites par Gregorich et al, 2014 <sup>15</sup> .....	7
Figure 2 : Schéma des différents modes existant en spectrométrie de masse en tandem.....	9
Figure 3 : Sites de fragmentation et nomenclature générale des fragments issus d'un peptide selon Roepstorff <sup>33</sup> .....	13
Figure 4 : Schéma conceptuel pour la recherche de peptides potentiels .....	16
Figure 5 : Schéma du protocole d'extraction des protéines .....	21
Figure 6 : Schéma du protocole de digestion trypsique .....	22
Figure 7 : Schéma du protocole de concentration et de purification par SPE.....	23
Figure 8 : Chromatogramme du peptide NALAHAVQSAR de la protéine de myosine dans un échantillon de 4 larves .....	29
Figure 9 : Chromatogramme des 14 peptides étalons avant (en rouge) et après (en vert) l'optimisation des paramètres de source .....	31
Figure 10 : Chromatogrammes du peptide NALAHAVQSAR dans un échantillon de 20 larves, léger (en haut) et lourd (en bas).....	33
Figure 11 : Chromatogramme du peptide STLTNPQR dans le cerveau .....	35
Figure 12 : Quantité de peptide en ng/larve (selon le peptide EALVSQLTR), sur l'axe y, en fonction des conditions de pH (7, 8, 9).....	37
Figure 13 : Quantité de peptide en ng/larve (selon le peptide NALAHAVQSAR), sur l'axe y, en fonction des conditions de pH (7, 8, 9) .....	37
Figure 14 : Quantité de peptide en ng/larve (selon le peptide LLGYQAAAYLDK), sur l'axe y, en fonction des conditions de pH (7, 8, 9) .....	38
Figure 15 : Quantité de peptide en ng/larve (selon le peptide EALVSQLTR), sur l'axe y, en fonction du n° série de géniteurs (1, 2) .....	40
Figure 16 : Quantité de peptide en ng/larve (selon le peptide NALAHAVQSAR), sur l'axe y, en fonction du n° série de géniteurs (1, 2).....	41
Figure 17 : Quantité de peptide en ng/larve (selon le peptide LLGYQAAAYLDK), sur l'axe y, en fonction du n° série de géniteurs (1, 2).....	41
Figure 18 : Organigramme généralisé de INRAE site de Lyon-Villeurbanne extrait de la présentation d'accueil des stagiaires .....	46

## Liste des tableaux

Tableau 1 : Tableau indiquant certaines caractéristiques de la Grande Alose .....	5
Tableau 2 : Tableau qui classe les différentes protéines avec leur rôle respectif dans l'organisme .....	15
Tableau 3 : Différentes espèces utilisées pour l'étape d'alignement et chez lesquelles les séquences des protéines d'intérêt sont présentes .....	19
Tableau 4 : Résumé des valeurs de paramètres de source du spectromètre de masse avant et après optimisation .....	27
Tableau 5 : Liste des peptides suspectés expérimentalement chez l'alose dans un pool de 4 larves .....	30
Tableau 6 : Tableau indiquant pour chaque peptides l'intensité avant et après optimisation des paramètres de sources ainsi que le gain de signal obtenu .....	32
Tableau 7 : Tableau résumant les 3 critères de validation pour le peptide NALAHAVQSAR.....	33
Tableau 8 : Tableau indiquant les différents peptides retrouvés dans les différents tissus de l'alose adulte (Muscle, Cerveau, Œuf, Foie) .....	34
Tableau 9 : Tableau indiquant les valeurs absolues des différences de moyenne pour chaque comparaison de groupe pour le peptide NALAHAVQSAR.....	39
Tableau 10 : Liste complète des potentiels peptides .....	47
Tableau 11 : Résultats de l'optimisation des paramètres de fragmentation.....	50

## **Liste des acronymes et abréviations**

aa : acide aminé

ACN : Acétonitrile

AF : Acide Formique

AMBIC : Bicarbonate d'ammonium

AQUA : Absolute Quantification of Protein

DDA : Data Dependant Acquisition

DIA : Data Independant Acquisition

DTT : Dithiothreitol

IAM : Iodoacétamide

INRAE : Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

ISA : Institut des Sciences Analytiques

LOQ : Limite Of Quantification (*fr : Limite de détection*)

MRM : Multiple Reaction Monitoring

PSAQ : Protein Standard Absolute Quantification

QconCat : Quantification Concatmer

SPE : Solid Phase Extraction

SRM : Selected Reaction Monitoring

TIC : Total Ion Current (*fr : Courant Ionique Total*)

TPCK : tosyl-L-phénylalanine chlorométhylcétone

Uniprot : Universal protein ressources

BLAST : Basic Local Alignment



## I. Introduction

Depuis près d'une dizaine d'années, la population d'aloses (*Alosa alosa*), poissons migrateurs de la famille des *Clupeidae*, est en fort déclin dans la Gironde. Afin de comprendre ce phénomène, plusieurs hypothèses ont été étudiées à ce jour<sup>1</sup>. L'augmentation du nombre de barrages le long de la Garonne et de la Dordogne est mise en cause car elle empêcherait la remontée des aloses vers les frayères (zone de ponte)<sup>2</sup>. La présence de poissons prédateurs des aloses, le *Silure*, est également identifiée comme une des causes<sup>1</sup>, avec une prédation trop forte. Cependant, la dégradation de la qualité environnementale des zones de frayères, en lien avec les changements globaux tels que les modifications de températures, de pH ou encore les apports en contaminants chimiques liés aux activités humaines, pourrait également être des causes de ce déclin des populations, mais à ce jour, n'a été que peu abordée chez l'alose.

Les écosystèmes aquatiques sont soumis à de nombreuses pressions humaines, comme leurs contaminations par des composés chimiques tels que les métaux lourds, les hydrocarbures aromatiques polycycliques ou les pesticides. L'écotoxicologie est un domaine scientifique qui a pour objectif d'évaluer l'impact et les effets de ces polluants sur les organismes et les populations à travers, entre autres, l'utilisation de biomarqueurs<sup>3</sup>. Un biomarqueur correspond à tout changement observable et/ou mesurable au niveau moléculaire, biochimique, cellulaire et physiologique qui révèle l'exposition d'un individu à une ou plusieurs substances chimiques polluantes. Ainsi, l'exposition d'un individu sain à certains polluants aura des effets néfastes, apparents ou non, sur sa santé associés à des réponses biochimiques et physiologiques spécifiques<sup>4</sup>. La modulation de toutes protéines nécessaires au développement ou à la survie d'une espèce constitue potentiellement un biomarqueur d'intérêt en écotoxicologie.

Autrefois très utilisés pour l'étude et la quantification de biomarqueurs de santé chez l'homme, les tests enzymatiques (test ELISA par exemple) ont laissé la place à de nouvelles technologies qui ont que très récemment émergées en écotoxicologie pour la mesure de ces biomarqueurs<sup>5</sup>. Parmi elles, il y a particulièrement la spectrométrie de masse en mode Multiple Reaction Monitoring (MRM) reconnu pour sa spécificité, sa sensibilité et sa grande capacité de multiplexage<sup>6</sup> et <sup>7</sup>. Contrairement aux tests enzymatiques qui nécessitent l'utilisation d'un capteur spécifiquement greffé par un

anticorps pour la détection d'une protéine unique, la spectrométrie de masse MRM permet de suivre une multitude de protéines en une seule analyse. La spectrométrie de masse a récemment fait ses preuves pour la quantification multiplexée de biomarqueurs protéiques, en mode protéomique *shotgun* couplé à de la RNA-seq chez le *Gammarus fossarum*<sup>8</sup> et la *Dreissseine polymorpha*<sup>9 et 10</sup> dont la séquence génomique n'était pas connu.

C'est dans ce contexte que s'inscrit mon stage au sein de l'équipe écotoxicologie de l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). Il a pour objectif dans un premier temps d'identifier des peptides d'intérêt chez une espèce non séquencée, l'alose *Alosa alosa*, puis d'optimiser, une méthode de quantification de ces peptides par chromatographie liquide haute performance couplée à la spectrométrie de masse en tandem (HPLC-MSMS) en mode MRM ciblé, afin de pouvoir évaluer à terme leur pertinence comme biomarqueurs. Ce stage s'est réalisé en collaboration avec l'Institut des Sciences Analytiques (ISA) de Lyon et l'UE Ecologie et Ecotoxicologie aquatique de INRAE Bordeaux et plus spécifiquement dans le cadre des travaux de thèse de Marion BLAYA, dont la thèse porte sur l'étude sur le déclin des populations d'aloses en Gironde.

Le premier challenge de ce projet de stage, bien avant le développement de la méthode de quantification, a été de rechercher des peptides protéotypiques pour la mesure et le suivi de protéines d'intérêt. A partir du génome d'un organisme, il est facile de disposer des séquences de protéines et de prédire les peptides potentiellement formés (masse et fragmentation) après une digestion « *in silico* » afin de programmer le spectromètre de masse. Cependant, une des problématiques rencontrées pour de nombreuses espèces d'intérêt environnemental comme le genre *Alosa* est le manque de ressources moléculaires disponibles<sup>11</sup>. Comme le génome d'*Alosa alosa* ainsi que les séquences nucléotidiques ou peptidiques ne sont à ce jour pas encore connus, un des objectifs de mon travail était de montrer si une approche par phylogénie pouvait être une alternative à l'acquisition de données génomiques. La phylogénie correspond à la classification des espèces vivantes en fonction de leurs ressemblances morphologique et/ou moléculaire. Ainsi l'objectif est d'évaluer si les ressemblances moléculaires/génomiques entre espèces séquencées permettent de proposer des peptides d'intérêt chez des espèces non séquencées.

Le deuxième challenge a ensuite été d'optimiser une méthode HPLC-MSMS afin de quantifier les peptides sélectionnés et d'évaluer leur potentiel d'intérêt comme biomarqueurs à travers leur quantification chez les larves venant de géniteurs différents et exposés à des pH variables.

Dans ce mémoire seront présentées dans un premier temps l'entreprise et l'équipe d'accueil. Puis un état de l'art sur les méthodes et techniques existantes en spectrométrie de masse pour une utilisation en protéomique sera fait. Ensuite, la partie expérimentale du travail sera présentée en deux chapitres (matériels et méthodes, résultats et discussion) avant de terminer avec le bilan et les perspectives du stage.

## **II. Présentation de l'entreprise**

### **1. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement**

L'institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) est un organisme de recherche français, dont l'activité est l'étude des relations entre agriculture, environnement et alimentation. Né officiellement le 1<sup>er</sup> janvier dernier de la fusion de deux instituts de recherche, IRSTEA (Institut National en Sciences et Technologies pour l'Environnement et l'Agriculture) et INRA (Institut National de la Recherche Agronomique), ayant des thématiques de recherche complémentaires, INRAE est reconnu comme étant un EPST (Etablissement Public Scientifique et Technologique). Il est présent dans 18 régions françaises. Dans la région Auvergne-Rhône-Alpes, deux centres sont présents, dont le centre Lyon-Grenoble-Auvergne-Rhône-Alpes, réparti en 11 sites et présidé par M. Pascal BOISTARD. Parmi ces 11 sites, il existe le site Lyon-Villeurbanne où s'est déroulé le stage dans l'unité de recherche Riverly au sein de l'équipe Ecotoxicologie dont le responsable est M. Arnaud CHAUMOT. L'organigramme généralisé de INRAE du site Lyon-Villeurbanne est présenté en *Annexe 1*.

### **2. Equipe Ecotoxicologie**

L'équipe écotoxicologie de INRAE est pluridisciplinaire (écophysiologie, biochimie, biologie moléculaire, expérimentation et modélisation) et ces activités sont structurées autour de deux objectifs clés :


- La production de connaissances et d'outils pour une meilleure gestion de la qualité des milieux aquatiques
- L'amélioration des réglementations au niveau des contaminations chimiques et des diagnostics des milieux mis en œuvre pour les espèces de l'écosystème

### III. Etat de l'art

#### 1. Généralités sur l'Alose

L'espèce étudiée au cours de ce stage est *Alosa alosa*, aussi appelée grande alose dont les caractéristiques sont présentées dans le *Tableau 1*. Elle fait partie des poissons migrateurs de la famille des *Clupeidae*. Leurs populations peuvent être localisées dans l'Atlantique, en mer Baltique, au Maroc, ou encore en Méditerranée. Elles se reproduisent généralement en mai-juin en eaux douces, notamment dans la Garonne et la Dordogne. Les femelles pondent sur des sites de frai, puis les larves se déplacent en aval du cours d'eau pour se développer en milieu estuarien. Cette période de développement peut durer plusieurs mois. Après avoir atteint un développement suffisant, les jeunes aloses migrent vers la mer où elles vont vivre et se nourrir pendant plusieurs années avant de remonter en eau douce pour se reproduire<sup>12et13</sup>.

Tableau 1 : Tableau indiquant certaines caractéristiques de la Grande Alose. Photo de la grande Alose extraite de Nature Gate<sup>14</sup>

Caractéristique	Grande Alose
Image	
Taille	54 à 80 cm
Poids moyen	1,6 kg
Type d'écailles	Ecailles irrégulières et non alignées
Tâche distinctive	Une seule grosse tâche derrière l'opercule

Au cours du stage, le stade larvaire a été étudié, il correspond au stade de vie le plus sensible de l'espèce et joue donc un rôle important dans le maintien de la population. De ce fait, ce stage ayant pour objectif de caractériser l'alose d'un point de vue protéique (identification de protéines), la technique analytique utilisée est la spectrométrie de masse.

En spectrométrie de masse, méthode analytique utilisée durant le stage, et plus particulièrement en protéomique, deux approches d'identification et de quantification de protéines existent, la stratégie Top Down et la stratégie Bottom Up. Elles ont notamment été explicitées et expérimentées par Gregorich *et al.*<sup>15</sup> dans un travail sur la recherche et le suivi de protéines identifiées chez certaines maladies cardiaques. La *Figure 1* schématise chacune des deux approches.

## **2. Stratégie Top Down pour la spectrométrie de masse**

La stratégie Top Down permet de caractériser et d'identifier les protéines présentes dans l'échantillon étudié. Cette méthode consiste à extraire les protéines d'un échantillon et à l'injecter directement dans le système d'analyse sans processus de digestion.

Même si elle permet de faire de la quantification en spectrométrie de masse haute résolution, cette dernière n'est pas la mieux adaptée pour la quantification. En effet, la quantification d'une protéine entière contenant un grand nombre d'acides aminés entraîne une diminution globale de l'intensité du signal, et donc de la sensibilité, en raison d'une large distribution isotopique et de la répartition du signal sur les différents états de charge. En revanche, cette stratégie présente l'avantage de permettre l'observation des différentes modifications post-traductionnelles (phosphorylations par exemple).

## **3. Stratégie Bottom-Up pour la spectrométrie de masse**

La stratégie Bottom-Up permet d'identifier et de quantifier des peptides rapporteurs d'une protéine. Cette méthode consiste à extraire les protéines d'un échantillon puis de les soumettre à un processus de digestion enzymatique. Ce processus permet de fragmenter une longue chaîne protéique en séquence peptidique grâce à l'action d'une endoprotéase qui va cliver la protéine après des acides aminés spécifiques. Les peptides en solution sont ensuite injectés dans le système d'analyse dont le traitement des données permettra d'identifier et quantifier la protéine.

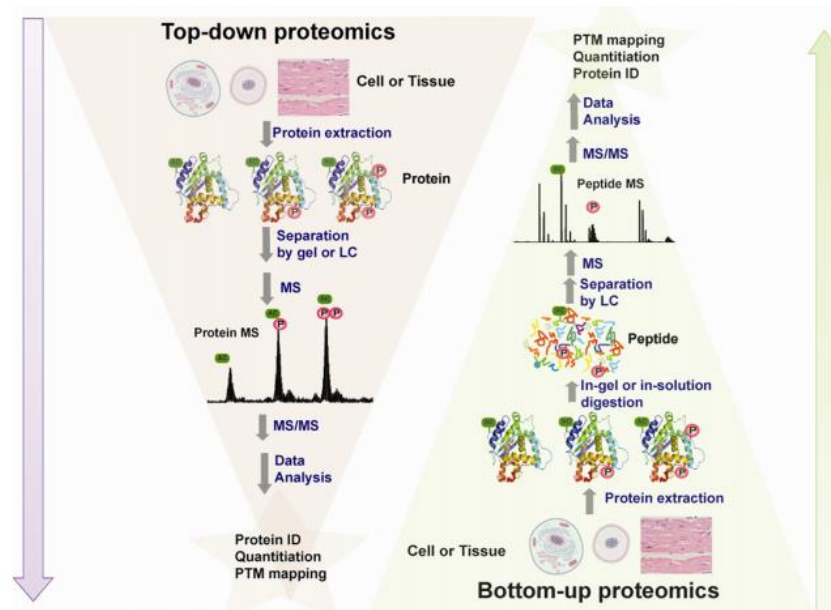


Figure 1 : Schéma des approches Top Down et Bottom Up pour la protéomique décrites par Gregorich et al, 2014<sup>15</sup>

Dans le cadre du stage, l'un des objectifs étant d'identifier et de quantifier des peptides rapporteurs de protéines d'intérêt, l'approche Bottom Up est la plus pertinente. Cette approche est la mieux adaptée pour une quantification en raison d'une LOQ plus faible et donc une meilleure sensibilité. L'enzyme de digestion utilisée est la trypsine, une endoprotéase qui rompt les liaisons peptidiques après les acides aminés basiques que sont la lysine et l'arginine<sup>16</sup>.

Après le choix de la stratégie Bottom Up, il faut définir une approche analytique adaptée. Il existe d'ailleurs deux stratégies possibles en spectrométrie de masse : une approche non ciblée lorsque les protéines à étudier ne sont pas connues et une approche ciblée lorsque les protéines à étudier sont connues.

#### **4. Approche non ciblée en spectrométrie de masse**

Une approche non ciblée, ou globale, consiste à analyser des échantillons biologiques sans à priori sur l'échantillon et d'identifier des biomarqueurs sans information préalable. Deux modes se distinguent :

- Le mode Data Dependant Acquisition (DDA) qui consiste à permuter entre le mode MS et le mode MSMS selon des critères définis (intensité d'un ion par exemple). Ce mode est utilisé notamment pour la création de bases de données de protéines présentes dans un échantillon biologique. Ce mode a été décrit par Bateman en 2013<sup>17</sup> ou encore Hu *et al.* en 2016<sup>18</sup>.

- Le mode Data Independent Acquisition (DIA) qui consiste à fragmenter tous les ions précurseurs selon différentes techniques. La technique Sequential Window Acquisition of all Theoretical mass spectra (SWATH)<sup>18</sup>, par exemple, va sélectionner les ions précurseurs sur des fenêtres définies de  $m/z$ . La technique All Ion Fragmentation (AIF)<sup>19</sup> va fragmenter tous les ions précurseurs et permettre d'obtenir sur le même spectre de masse tous les ions produits. La technique MS<sup>E</sup> <sup>20</sup> permet d'alterner entre des spectres à basse et haute énergie.

Le mode DIA présente l'avantage de ne pas perdre d'information sur l'échantillon analysé. Toutefois, les spectres obtenus sont très complexes et il est alors difficile d'extraire les informations pertinentes. Le mode DDA, de par son mode de fonctionnement, implique une perte d'information, mais présente quant à lui l'avantage de générer des spectres plus faciles à analyser et de pouvoir extraire les spectres de fragmentation des peptides. La combinaison des deux modes permet ainsi d'extraire le maximum d'informations : le mode DDA va permettre de générer des bases de données de spectres de fragmentation spécifiques des peptides qui pourront permettre l'identification des peptides en mode DIA.

En approche non ciblée, il est donc possible d'identifier des protéines présentes dans un échantillon biologique sans information préalable (génomique, biomarqueur connu...). Cependant, en raison de la complexité des échantillons biologiques, ce mode nécessite l'utilisation d'un analyseur haute résolution. C'est d'ailleurs une des raisons qui a guidé notre choix sur une approche ciblée en plus d'avoir besoin d'une quantification plus spécifique qu'en mode *shotgun*.

Toutefois, il existe à ce jour différents types d'analyseurs dont le rôle est de trier les ions en fonction de leurs rapports  $m/z$ . Un analyseur est caractérisé par sa résolution, sa gamme dynamique, sa vitesse de balayage, sa sensibilité et la vitesse à laquelle les ions vont traverser l'analyseur.

Dans le cadre du stage, un triple quadripôle a été utilisé sur lequel 4 modes MSMS sont possibles et seront décrits par la suite, en particulier le mode MRM dont le mode *dynamic* MRM (dMRM), qui nous intéresse particulièrement et qui sera décrit par la suite.



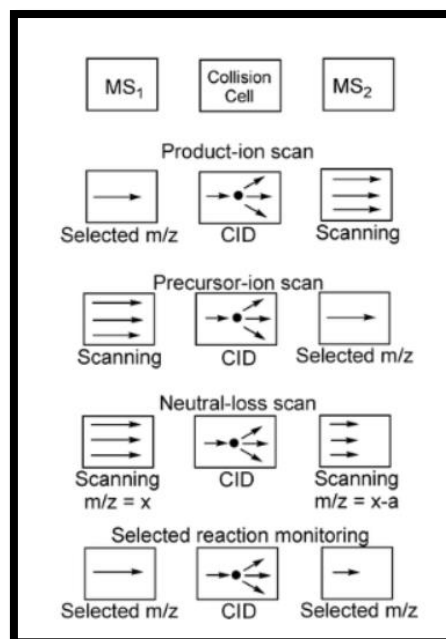
## 5. Modes de balayage en spectrométrie de masse

Il existe différents modes MSMS de balayage schématisés sur la *Figure 2* : tous ces modes donnent une information structurale.

Concernant le mode non ciblé, il existe trois modes de balayage qui lui sont spécifiques :

- Le mode balayage des ions fragments qui permet de sélectionner un ion précurseur et d'effectuer un balayage des ions fragments.
- Le mode balayage des ions précurseurs qui permet de sélectionner un ion fragment et de balayer tous les ions précurseurs et ainsi d'identifier tous les ions précurseurs produisant un même fragment.
- Le mode perte de neutre qui permet d'effectuer un balayage des ions précurseurs et des ions produits avec un différentiel de rapport  $m/z$  entre les deux et ainsi d'identifier tous les ions précurseurs produisant un même fragment neutre (et non détecté).

En plus de l'approche non ciblée en spectrométrie de masse, il existe une autre approche, celle de la spectrométrie de masse MRM en mode ciblé, son principal mode de balayage MSMS étant le mode Multiple Reaction Monitoring (MRM) ou Selected Reaction Monitoring (SRM). Celui-ci, schématisé *Figure 2*, permet de sélectionner un ion précurseur en Q1 et un ion fragment en Q3.



*Figure 2 : Schéma des différents modes existant en spectrométrie de masse en tandem*

Par ailleurs, en protéogénomique la connaissance du génome ou du transcriptome de l'espèce d'intérêt permet d'appliquer une stratégie analytique différente en mode ciblé décrite ci-après.

## **6. Approche ciblée en spectrométrie de masse**

En protéomique, une approche ciblée consiste à évaluer la présence, ou même la quantité, d'une protéine connue dans un échantillon biologique. Plusieurs scientifiques ont d'ailleurs cherché à l'expliquer en expérimentant cette approche, notamment Colangelo *et al.*<sup>21</sup> ou encore Ahn *et al.*<sup>22</sup>. Cette approche ciblée permet donc la quantification d'une protéine spécifique.

Le mode MRM permet donc de détecter en sortie de Q3 un ou plusieurs ions fragments issus de la fragmentation de l'ion précurseur. La détection d'un ion précurseur et de l'un de ces ions fragments correspond à une transition. En général, et c'est le cas dans notre étude, le nombre de transitions minimal approprié pour l'identification d'un peptide s'élève à trois : la première permet de détecter l'ion fragment le plus intense et le moins interféré pour quantifier, tandis que les autres transitions confirment l'identification de la molécule (grâce aux rapports d'aires des transitions qui doivent être constants d'un échantillon à l'autre). La sélection d'une transition plutôt qu'un ion précurseur permet d'améliorer la spécificité de la technique mais aussi de réduire le bruit de fond permettant ainsi l'augmentation du rapport signal sur bruit et donc la diminution de la limite de quantification (LOQ) qui reflète une amélioration de la sensibilité. Par ailleurs, la spectrométrie de masse est généralement couplée à un système de séparation chromatographique. C'est pourquoi l'identification et la quantification d'une molécule sont permises grâce au temps de rétention et à au moins 3 transitions, dont une quantifiante et au moins deux qualifiantes.

Le mode MRM a un temps de cycle de l'ordre de la milliseconde qui permet alors de suivre plusieurs transitions (environ 100 transitions) par molécule associés à une superposition de plusieurs chromatogrammes (3 chromatogrammes pour 3 transitions par exemple). Durant le stage, un triple quadripôle Agilent est utilisé, grâce à ce dernier, il est possible d'acquérir jusqu'à 200 transitions par échantillon. La largeur des pics et la fenêtre de temps étant donc des paramètres limitant dans le suivi d'un nombre important de transitions, le mode dynamic MRM (dMRM) permet de sélectionner des transitions à suivre sur des intervalles de temps définis et plus courts que l'ensemble du chromatogramme permettant ainsi d'augmenter le nombre de

transitions suivies sur l'ensemble du chromatogramme. Ce mode a d'ailleurs permis aux auteurs *J Kao et al*<sup>23</sup> de suivre plus de 1500 transitions et de quantifier 641 peptides en une seule analyse.

Une approche récente a également été développée, permettant de s'affranchir des variations de temps de rétention, notamment lors du transfert d'une méthode d'un appareillage à l'autre, qui peuvent devenir problématiques avec l'approche dMRM : la technique SCOUT-MRM. Cette technique, explicitée par Rougemont *et al*<sup>24</sup> en 2016 ou par Salvador *et al*<sup>25</sup> en 2020, consiste à ajouter à l'échantillon d'intérêt des peptides, appelés SCOUT, qui se répartissent tout au long du chromatogramme. Le spectromètre est ensuite réglé pour détecter les transitions connues du premier peptide SCOUT, entraînant alors, dès leur détection, la recherche des transitions des peptides dont le temps de rétention est situé entre le premier peptide SCOUT et le deuxième peptide SCOUT (groupe 1). Dès la détection du deuxième peptide SCOUT, la détection des transitions des peptides du groupe 1 est stoppée et les transitions des peptides dont le temps de rétention est situé entre le deuxième peptide et le troisième peptide SCOUT (groupe 2) commence. L'analyse se déroule ainsi de suite tout au long du chromatogramme. Cette approche permet une bonne répétabilité des détections de peptides contenus dans des échantillons multiplexés entre différents appareillages en raison de l'indépendance au temps de rétention, contrairement à la technique dMRM.

Au cours de ce stage, la technique dMRM a été utilisée afin de pouvoir suivre un maximum de transitions en une analyse.

Le mode d'acquisition en MRM le plus utilisé pour faire de la quantification est le mode ciblé. Ainsi, après avoir sélectionné ce mode, l'étape suivante est la mise en place d'une stratégie de quantification qui peut être relative ou absolue<sup>26</sup>.

## **7. Quantification relative**

La quantification relative consiste à comparer les intensités des pics de peptides entre l'échantillon d'intérêt et un échantillon témoin.

Deux approches existent :

- Quantification relative avec marquage chimique : la dilution isotopique consiste à marquer l'échantillon témoin avec des isotopes lourds stables, à mélanger l'échantillon témoin avec l'échantillon d'intérêt et à comparer les intensités des pics dans chaque

échantillon. Les réactifs d'isotopes lourds stables doivent avoir une structure permettant d'accéder à différentes fonctions chimiques de la molécule à marquer telles que les amines du côté N-terminal d'une protéine. Pour cette approche, il existe plusieurs réactifs de marquage comme par exemple le réactif mTRAQ qui a été utilisé pour la première fois par Desouza *et al*<sup>27</sup> ou encore le réactif iTRAQ introduit par le groupe Papin<sup>28</sup> en 2004.

- Quantification relative sans marquage qui consiste à analyser séparément l'échantillon témoin et l'échantillon d'intérêt et à comparer les intensités des pics. Afin de s'affranchir des variations instrumentales (rendement d'ionisation...), les spectres sont normalisés par rapport au courant ionique total (TIC). En général, cette approche est la moins coûteuse du fait qu'elle ne nécessite aucun marquage isotopique. Il existe notamment le *Label free* qui consiste à dénaturer les protéines par alkylation suivie d'une digestion enzymatique.

## **8. Quantification absolue**

La quantification absolue est possible par étalonnage externe et interne. Cependant, l'étalonnage externe est difficile à mettre en place du fait de la présence de protéines endogènes ainsi que de certains effets biologiques. Ainsi, durant le stage, la quantification sera faite par étalonnage interne, avec un étalon interne qui respecte certaines conditions. Cet étalon interne doit être différent de l'analyte que l'on cherche à quantifier mais ayant les mêmes propriétés physico-chimiques (temps de rétention, pKa, ...). Les étalons internes les plus adaptés à l'analyse de peptides en LC-MSMS sont donc des peptides marqués avec des isotopes lourds stables (<sup>13</sup>C, <sup>15</sup>N) ajoutés en quantité connue à l'échantillon (dilution isotopique). Différentes stratégies peuvent être employées :

- Stratégie Quantification concatamer<sup>29</sup> (QconCat) qui consiste à produire une ou plusieurs protéines chimériques contenant les peptides d'intérêt marqués par des isotopes lourds stables. Cette stratégie permet de s'affranchir des biais liés à la digestion enzymatique, sans toutefois les éliminer totalement, la protéine chimérique étant différente de la protéine étudiée.

- Stratégie Protein Standard Absolute Quantification<sup>30 ;31</sup>(PSAQ) qui consiste à introduire la protéine d'intérêt totalement marquée par des isotopes lourds stables avant la digestion enzymatique. Cette stratégie permet également de s'affranchir des biais liés à la digestion enzymatique.

- Stratégie Absolute Quantification of protein<sup>32</sup> (AQUA) qui consiste à ajouter une quantité connue de peptides marqués après l'étape de digestion enzymatique. Cette stratégie ne permet toutefois pas de s'affranchir des biais liés à la digestion enzymatique mais présente l'avantage de ne pas nécessiter la production de protéines marquées qui peut être coûteuse.

La stratégie choisie pour le stage est la stratégie AQUA en quantification absolue.

## **9. Fragmentation des protéines et sélection des transitions pour la spectrométrie de masse MSMS**

L'un des avantages de la spectrométrie de masse MSMS dans l'étude d'échantillons biologiques est la connaissance déjà bien établie des règles de fragmentation. Ceci permet alors de prévoir les fragments qui se formeront. Cette étape de fragmentation est alors capable d'apporter des informations structurales permettant de déduire la séquence peptidique.

Les différents types de fragments produits dépendent notamment de la technique de fragmentation utilisée mais les ruptures dans la liaison peptidique, situées au niveau des liaisons des acides aminés, peuvent intervenir à trois endroits distincts et sont spécifiques à chaque technique. Ainsi, en connaissant la séquence peptidique et la technique de fragmentation, il est possible de prédire les fragments formés.

La technique de fragmentation par collision est la stratégie employée lors de ce stage mais plusieurs stratégies existent formant des ions différents dont la nomenclature proposée par Roepstorff<sup>33</sup> en 1984 est présentée Figure 3 :

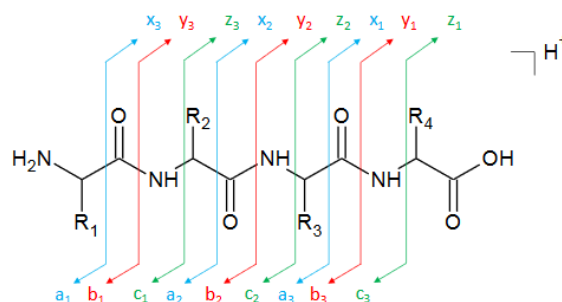


Figure 3 : Sites de fragmentation et nomenclature générale des fragments issus d'un peptide selon Roepstorff<sup>33</sup>

- Fragmentation par collision (CID) à basse énergie : les peptides se fragmentent par collision à basse énergie avec des molécules de gaz inerte formant principalement des ions b et y.
- Fragmentation par collision (CID) à haute énergie : Fragmentation par collision (CID) à haute énergie avec des molécules de gaz inerte formant tous les types d'ions a et x, b et y, c et z.
- Dissociation par transfert d'électrons : les peptides se fragmentent par transfert d'électrons d'un donneur anionique formant préférentiellement des ions c et z.

Toutefois, la prédiction des fragments nécessite la connaissance de la séquence peptidique et donc du génome. C'est pourquoi la méconnaissance du génome de l'aloose a nécessité une stratégie de phylogénie afin d'identifier des protéines d'intérêt dont les séquences peptidiques sont connues et qui sera présentée par la suite.

La connaissance des fragments formés permet ensuite de déterminer les transitions entre l'ion précurseur et l'ion fragment selon plusieurs critères de sélection<sup>34</sup> :

- L'intensité des pics : afin d'optimiser la sensibilité de la technique, les transitions impliquant les pics les plus intenses (ions fragments les plus abondants) sont privilégiées
- La masse des ions précurseur et fragment (ou nombre d'acides aminés) : afin d'optimiser la spécificité, en particulier dans une matrice complexe comme les échantillons biologiques, les fragments présentant un nombre important d'acides aminés sont privilégiés. En effet, une longue séquence peptidique de l'ion fragment diminue la probabilité de détecter ce même fragment issu d'un peptide similaire et proche mais sans intérêt. En général, les fragments présentant de 8 à 20 acides aminés permettent d'obtenir une bonne spécificité et sont appelés « best flyers ».
- Eviter les peptides ayant des acides aminés susceptibles de subir des modifications chimiques ou réarrangements tels que l'oxydation (méthionine), désamidation (Glutamine)...
- L'absence d'interférence sur la transition quantifiante si possible

Dans le cadre du stage, il a été choisi de sélectionner les ions précurseurs ayant entre 6 et 25 acides aminés impliquant des ions fragments ayant un m/z supérieur à 250 Th.

## 10. Choix des protéines d'intérêt en vue d'évaluer l'état de santé des larves d'alose

L'évaluation de l'état de santé d'une espèce nécessite que le choix des protéines d'intérêt soit fait sur la base des connaissances physiologiques. Ainsi, les biomarqueurs protéiques préférentiellement choisis sont ceux associés à un bon développement embryonnaire, comme la myosine et la myogénine pour la contraction et le développement musculaire, mais aussi la vitellogénine associée aux réserves énergétiques utilisées par les embryons durant leur développement. De plus, l'utilisation de certains biomarqueurs connus pour refléter une modulation de leur système de détoxification comme l'acétylcholinestérase, la métallothionéine et le cytochrome P450 1A1<sup>35</sup> et <sup>36</sup> peut également être envisagée. Enfin et pour un contrôle purement biologique l'utilisation de l'expression de la protéine phocéine peut être utilisée comme contrôle de charge protéique total de chaque échantillon (Hu *et al.*, 2016<sup>37</sup>).

Le *Tableau 2* présente les protéines d'intérêt choisies *a priori* comme biomarqueur de l'état de santé d'une espèce au stade larvaire, dans notre cas des larves d'*Alosa alosa* et leur rôle biologique respectif.

*Tableau 2 : Tableau qui classe les différentes protéines avec leur rôle respectif dans l'organisme*

<b>Nom de la Protéine</b>	<b>Rôle biologique/fonctionnel</b>
Myosine	Contraction musculaire
Myogénine	Développement musculaire
Vitellogénine	Alimentation de l'embryon jusqu'à son autonomie
Acétylcholinestérase	Marqueur d'une exposition aux insecticides
Métallothionéine	Gestion des métaux dans l'organisme, son induction marque une exposition à une contamination métallique
Cytochrome P450 A1	Enzyme de détoxification
Phocéine	Protéine de contrôle

## 11. Stratégie de recherche et d'identification des peptides protéotypiques chez une espèce non-séquencée en vue de mesurer des protéines d'intérêt

Une des stratégies utilisées pour la recherche de peptides protéotypiques chez une espèce dont le génome n'est pas disponible, est l'approche phylogénétique. Cette stratégie ayant déjà fait ses preuves dans l'étude sur la vitellogénine comme biomarqueur chez le gammare effectué par Jubeaux *et al*<sup>88</sup>, elle sera utilisée dans le cadre de ce stage. Par ailleurs, il s'avère qu'aucune séquence protéique n'est disponible pour *Alosa Alosa*. Ainsi, l'approche phylogénétique a consisté premièrement à choisir des espèces proches d'un point de vue phylogénétique pour laquelle les séquences génomiques/protéiques sont disponibles pour les protéines d'intérêt. Dans le cas du genre *Alosa*, les bases de données protéiques donnent accès à certaines séquences protéiques chez le hareng ou le poisson zèbre, espèces les plus proches d'*Alosa alosa*.

Cette stratégie est illustrée par la *Figure 4* et sera détaillée par la suite.

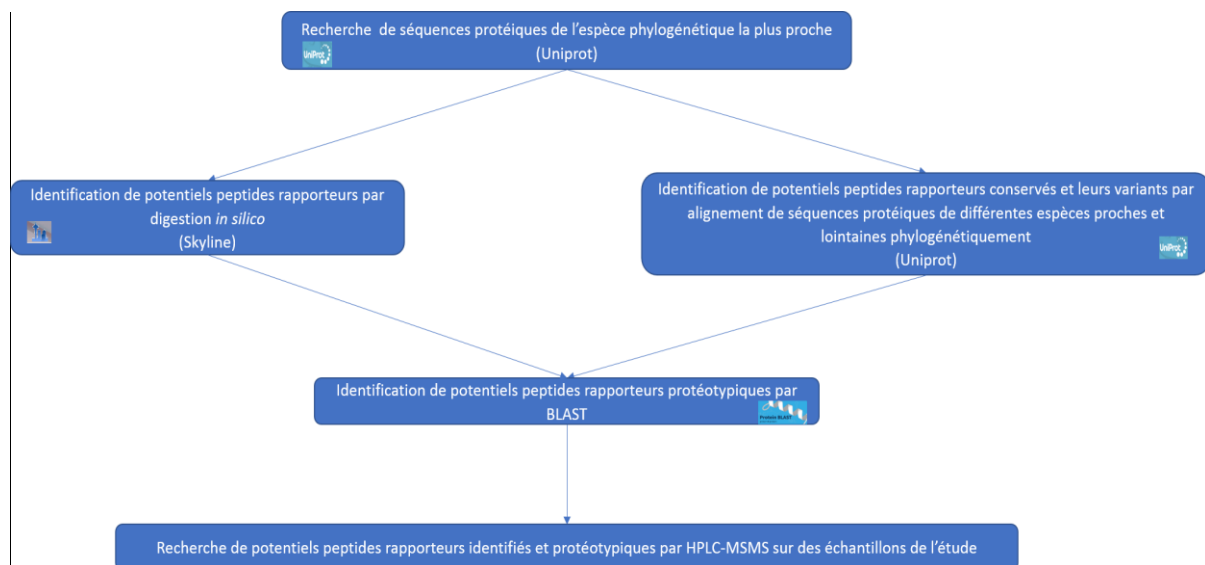


Figure 4 : Schéma conceptuel pour la recherche de peptides potentiels



### **a. Recherche des peptides potentiels**

Dès lors que les différentes séquences protéiques sont déterminées via la base de données *Universal Protein Resources*<sup>39</sup> (*Uniprot*), la seconde étape consiste à rechercher des peptides potentiels en combinant deux étapes : d'une part, une étape de digestion *in silico* des protéines d'intérêt pour obtenir une première liste de peptides potentiels, et d'autre part, une étape d'alignement des séquences protéiques permettant d'identifier certains peptides conservés n'ayant pas été identifiés *in silico*.

#### **i. Identification de peptides et recherche de peptides par digestion *in silico***

Les séquences des protéines sélectionnées subissent une étape de digestion *in silico* afin de sélectionner des séquences peptidiques candidates pour être recherchées en HPLC-MSMS. Cette digestion *in silico* est effectuée avec le logiciel *Skyline* (version *Skyline (64-bit) 4.2.0 19072*). Elle est possible en raison des mécanismes connus de digestion enzymatique par la trypsine : les liaisons peptidiques sont « coupées » au niveau des arginines et des lysines.

Le logiciel *Skyline* permet ainsi de simuler une digestion tryptique et de sélectionner des peptides potentiels. La sélection de ces peptides se faisant alors selon plusieurs critères :

- Le peptide doit être constitué d'au moins 6 acides aminés et d'au plus 25 acides aminés pour optimiser la spécificité
- Les 15 premiers acides aminés côté N-terminal sont exclus car ils sont plus susceptibles de subir des modifications post-traductionnelles
- Les ions précurseurs doivent être doublement ou triplement chargés positivement
- Les ions fragments doivent être mono-chargés positivement
- Sélection des fragments de type y car le mode de fragmentation est un mode CID de basse énergie
- Les ions précurseurs et fragments doivent être compris entre 250 et 1500 Thomson.
- Aucun clivage ne doit être manqué 2
- Les peptides précédés ou succédés par KR, RK, KK ou RR ne sont pas sélectionnés afin d'éviter une digestion aléatoire de la structure

## ii. Recherche de peptides conservés par alignement de séquences

En parallèle de la digestion *in silico* via *Skyline*, une étape d'alignement via le site *Uniprot*<sup>39</sup> en faisant un *BLAST*, des séquences protéiques est faite sur d'autres espèces de poissons. Cette étape permet d'une part d'évaluer la diversité/variabilité des peptides, d'identifier des peptides trypsiques conservés ou d'éventuels variants chez différentes espèces. D'autre part, elle permet aussi de vérifier le niveau de conservation des peptides proposés et sélectionnés via *Skyline*. Ainsi, pour effectuer cette étape, il est d'abord nécessaire de sélectionner, en plus des espèces les plus proches phylogénétiquement de l'espèce d'étude, d'autres espèces de poissons chez lesquelles les protéines d'intérêt sont disponibles. Ainsi, si par alignement, des séquences peptidiques sont communes à toutes ces espèces (peptides conservés), ou se différenciant par 3 acides aminés maximum (variants), elles sont considérées comme susceptibles d'être présentes dans la séquence protéique de l'espèce étudiée. Une fois qu'une liste de peptides potentiels est établie, les séquences peptidiques identifiées pour chaque protéine sont soumises à une recherche d'homologie via le logiciel *Basic Local Alignment Search Tool* (BLAST). Cette étape bio-informatique permet de vérifier si le peptide trypsique identifié *in silico* ou par alignement est spécifique (protéotypique) de la protéine d'intérêt. Si la séquence peptidique présente une homologie avec d'autres séquences protéiques ayant la même fonction biologique chez d'autres espèces, le peptide identifié est considéré comme protéotypique de la fonction biologique d'intérêt. En revanche, si le peptide présente une homologie avec des protéines ayant des fonctions différentes par rapport à celle recherchée, sa spécificité est moindre et il est donc écarté comme potentiel candidat analytique.

Les peptides obtenus par digestion *in silico* via *Skyline*, les peptides conservés ainsi que les variants ayant été identifiés comme protéotypiques (79/107) sont donc retenus comme potentiels biomarqueurs et sont finalement recherchés expérimentalement sur des échantillons de l'espèce étudiée. La liste de 107 peptides retenus est présentée en *Annexe 2*.

## IV. Matériel et méthode

### 1. Recherche et identification des potentiels peptides

Dans le cadre du stage, l'aloise étant une espèce non-séquencée, l'identification des protéines et des peptides d'intérêt a été effectuée par approche phylogénétique avec le hareng (*Clupea Harengus*) comme espèce la plus proche phylogénétiquement. Par ailleurs, le poisson zèbre (*zébra fish*) a été sélectionné comme seconde espèce proche phylogénétiquement de l'aloise pour les protéines de la métallothionéine, l'acétylcholinestérase et la phocéine dont les séquences protéiques ne sont pas répertoriées sur *Uniprot* chez le hareng.

Les différentes espèces utilisées pour l'étape d'alignement sont répertoriées dans le *Tableau 3*. Durant le stage, une variation allant jusqu'à 3 acides aminés différents est acceptée.

*Tableau 3 : Différentes espèces utilisées pour l'étape d'alignement et chez lesquelles les séquences des protéines d'intérêt sont présentes. Un carré vert indique qu'une espèce est utilisée pour l'alignement des séquences de la protéine associée.*

Espèces Protéines	<i>Anguilla anguilla</i> (ANGAN)	<i>Acipenser transmontanus</i> (ACITR)	<i>Anguilla japonica</i> (ANGJA)	<i>Astyanax mexicanus</i> (ASTMX)	<i>Clupea harengus</i> (CLUHA)	<i>Cyprinus carpio</i> (CYPCA)	<i>Danio rerio</i> (DANRE)	<i>Electrophorus electricus</i> (ELEEL)	<i>Esox lucius</i> (ESOLU)	<i>Ictalurus punctatus</i> (ICTPU)	<i>Oncorhynchus mykiss</i> (ONCMY)	<i>Orizias latipes</i> (ORYLA)	<i>Oryzias melastigma</i> (ORYME)	<i>Salmo salar</i> (SALSA)	<i>Silurus lanzhouensis</i> (9_TELE)	<i>Squalius cephalus</i> (SQUCE)	<i>Takifugu rubripes</i> (TAKRU)
Myosine																	
Myogénine																	
Cytochrome																	
Métallothionéine																	
Acétylcholinestérase																	
Vitellogénine																	
Phocéine																	

⇒ Seuls les peptides identifiés comme protéotypiques ont été recherchés expérimentalement.

## **2. Protocole d'extraction des protéines**

Le protocole d'extraction des protéines a été inspiré de la thèse de Gouveia<sup>8,40</sup> et adapté aux Aloses. La *Figure 5* présente le protocole d'extraction des protéines chez les larves d'alose.

Les échantillons initialement stockés à - 80°C ont été placés dans un bain de glace à + 4°C pour éviter toutes réactions de protéases et les larves poolées (entre 5 à 20 larves par pool) ont été placées dans un tube pour broyeur à billes avec une bille de 4mm d'acier chrome. Puis, 600 µL de tampon TRIS avec inhibiteurs à protéases ont été ajoutés, ce qui permet de contrôler le pH de la solution de 7,8 à 8 où l'effet de la trypsine est optimal et où les protéines sont stables.

Après 2 cycles de 25 secondes de broyage à 4 m/s qui permet de lyser mécaniquement les cellules et de libérer les protéines, les tubes (sans les billes) ont été centrifugés à 10000 g pendant 15 min à +4°C. L'étape de centrifugation a pour but de séparer les protéines présentes dans le surnageant des débris de cellules présents dans le culot.

Un volume de 250 µL de surnageant a ensuite été mélangé à 750 µL de solution de délipidation (1 : 1 ; éthanol absolu : diéthylether), dans un tube Eppendorf de 2 mL, avant d'être vortexé puis mis à reposer pendant 10 min dans la glace. Une nouvelle étape de centrifugation a été effectuée pour une durée de 10 min à +4°C à une vitesse de 10000 g. L'ajout du solvant de délipidation permet d'une part de solubiliser les lipides dans le surnageant obtenu et d'autre part de précipiter les protéines au contact de ce solvant organique.

Le surnageant a ainsi été éliminé et le culot protéique re-suspendu avec 250 µL de tampon TRIS sans inhibiteurs et sans triton X. Après évaporation des dernières gouttes de solvant organique, la solution a de nouveau été homogénéisée à l'aide du broyeur à billes pendant 30 secondes à 4 m/s.

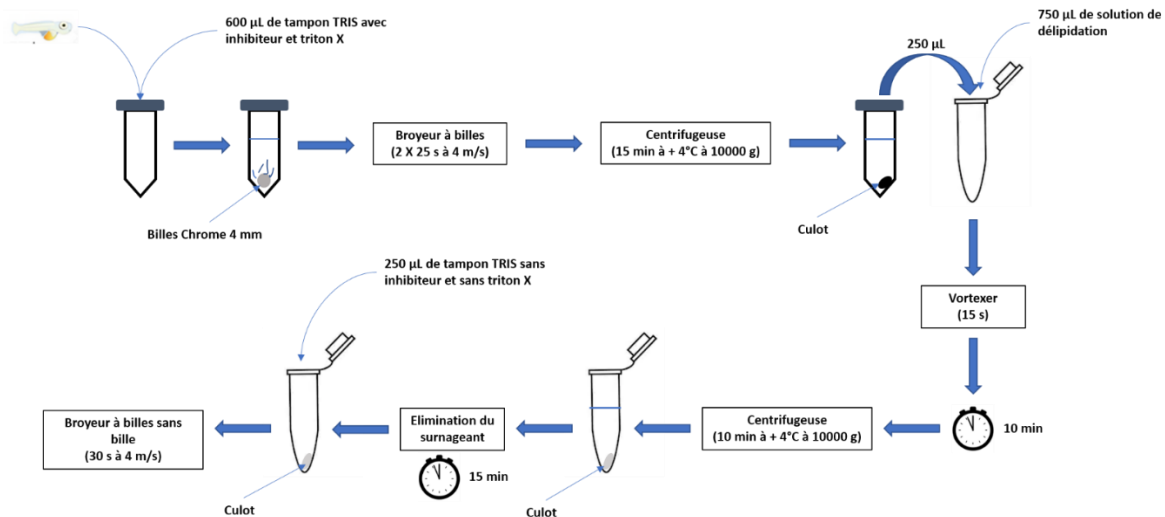


Figure 5 : Schéma du protocole d'extraction des protéines

### 3. Protocole de digestion par la trypsine

Le protocole de digestion par la trypsine va permettre de couper les liaisons peptidiques des protéines au niveau des arginines ou des lysines pour obtenir les peptides. La *Figure 6* présente le protocole de digestion enzymatique par la trypsine.

Ce protocole consiste premièrement à transférer dans un tube Eppendorf de 5 mL les 250 µL d'échantillon extrait puis d'ajouter 3 mL de bicarbonate d'ammonium (AMBIC) à 50 mM. Ensuite, un volume de 362 µL de dithiothreitol (DTT) à 15 mM a été ajouté pour rompre les ponts disulfures avant d'être incubé à 60°C pendant 40 min. Cette étape permet de faciliter l'accès aux sites de clivage par l'enzyme de protéolyse lors de la digestion par la trypsine. Dépassé ce délai et après un retour de l'échantillon à température ambiante, un volume de 402 µL d'iodoacétamide (IAM) à 15 mM a été ajouté, puis l'échantillon a été placé à l'abri de la lumière à température ambiante pendant 40 min. L'ajout de l'IAM empêche l'oxydation des cystéines libres et donc la reformation des ponts disulfure.

Après ce temps d'attente, un volume de 150 µL d'une solution de trypsine TPCK, dont la concentration doit être représentative de la quantité de protéines de l'échantillon, a été ajouté avant la mise à incubation à 37°C pendant 1h de l'échantillon. Par ailleurs, cette enzyme généralement traitée avec un inhibiteur, le tosyl-L-phénylalanine chlorométhylcétone (TPCK), va inhiber l'activité de la chymotrypsine synthétisée en même temps que la trypsine au niveau du pancréas, ce dernier pouvant induire des ruptures de liaisons supplémentaires pendant la digestion trypsique sans modifier l'action de la trypsine.

Afin de bloquer la réaction de digestion enzymatique, 40  $\mu\text{L}$  d'acide formique (AF) ont été ajoutés après le temps d'incubation. Puis, 10  $\mu\text{L}$  de solvant d'injection  $\text{H}_2\text{O}/\text{ACN}$  (90/10) acidifié 0.5% AF a été ajouté. Afin de détecter et/ou quantifier le peptide d'intérêt, l'échantillon a été spiké avec 10  $\mu\text{L}$  d'une solution de peptides lourds à 10  $\mu\text{g}/\text{mL}$ . Enfin, une dernière étape de centrifugation à 8000 rpm pendant 5 min a été effectuée afin d'éliminer les résidus susceptibles d'être présents dans le culot et de récupérer le surnageant contenant les peptides.

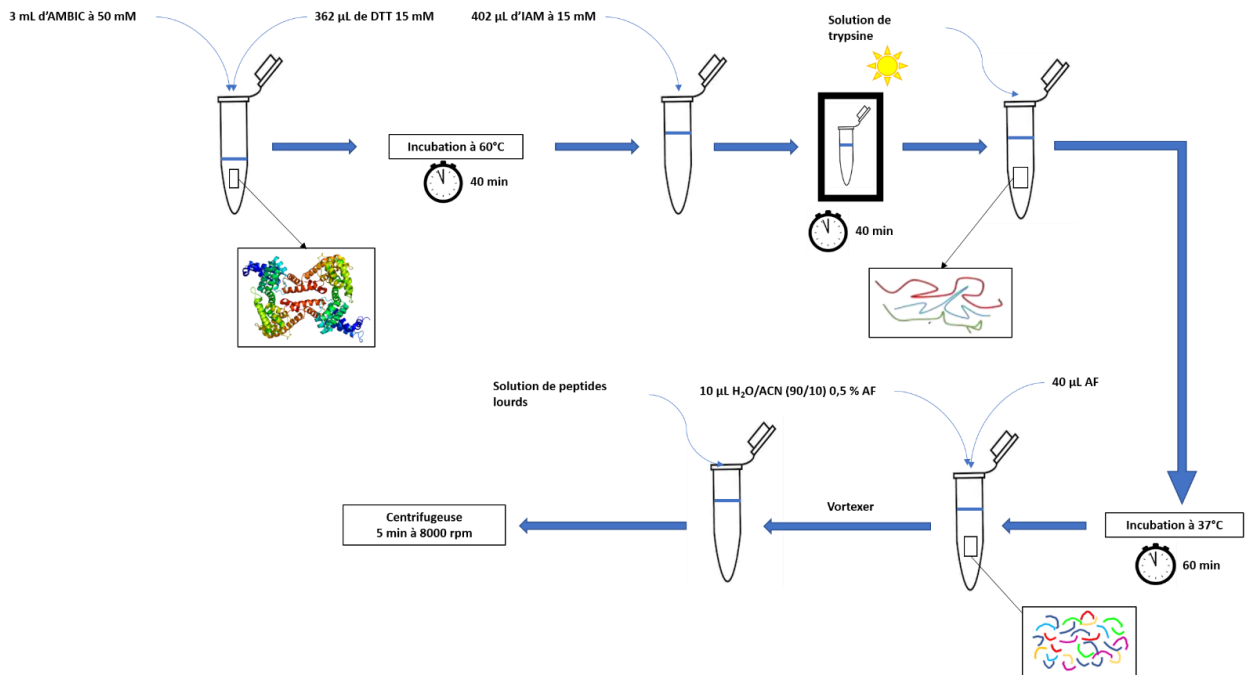


Figure 6 : Schéma du protocole de digestion trypsique

#### **4. Protocole de concentration et purification sur phase solide (SPE) des peptides**

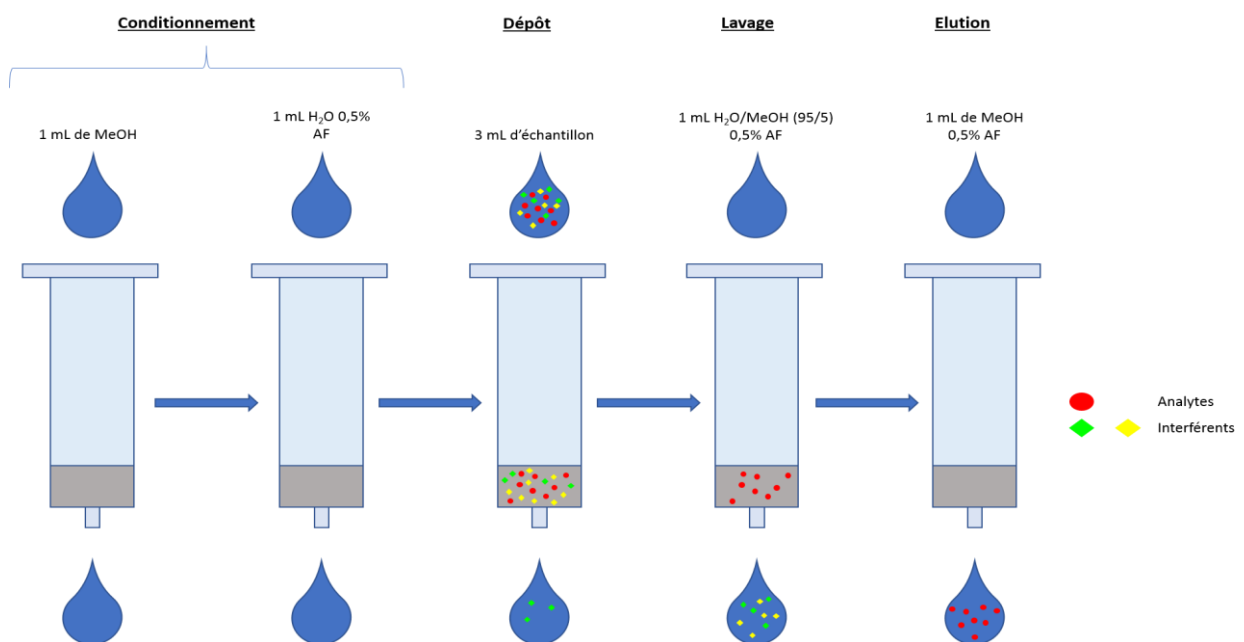
L'extraction sur phase solide permet de concentrer et de purifier les peptides qui ont été extraits par adsorption sur une surface solide. La *Figure 7* présente le protocole d'extraction par SPE.

La SPE ici est réalisée à l'aide de cartouche Oasis HLB 3 cc (60 mg, phase inverse). Ce type de cartouches phase inverse repose sur la rétention des composés par des interactions hydrophobes et l'élution par un solvant organique apolaire.

La phase stationnaire a été conditionnée dans un premier temps avec 1 mL de méthanol afin de déployer les greffons, puis avec 1 mL d'eau acidifiée (0,5% AF) pour équilibrer la cartouche aux mêmes conditions de pH que l'échantillon. Un volume de 3 mL d'échantillon a ensuite été déposé sur la cartouche.

Celle-ci a ensuite été lavée avec 1 mL d'un mélange eau/MeOH (95/5) acidifié (0,5% AF) afin d'éluer les composés non désirés.

Enfin, les peptides ont été élués avec 1 mL de méthanol acidifié (0.5% AF).



*Figure 7 : Schéma du protocole de concentration et de purification par SPE*

Après élution des peptides, l'échantillon a été évaporé à sec sous flux d'azote afin d'éliminer le solvant d'élution. Un volume de 100 µL de méthanol contenant 10% de glycérol a toutefois été ajouté au préalable afin de retenir les peptides, le glycérol ne s'évaporant pas.

L'échantillon séché a ensuite été dissous dans 90 µL de solvant d'injection H<sub>2</sub>O/ACN (90/10) acidifié à 0,5% AF et vortexé. Une étape de centrifugation de 5 min à 15000 rpm est nécessaire avant les analyses afin d'éliminer d'éventuels résidus et impuretés en suspension.

## **5. Conditions d'analyse de la méthode chromatographique**

L'analyse a été réalisée sur un système chromatographique HPLC Agilent 1290 Infinity II. Un volume de 20 µL d'échantillon a été injecté et séparé sur une colonne *Waters Xbridge C18* de dimensions 2,1\*10 mm, avec des particules de 3,5 µm de diamètre. L'élution a été réalisée à un débit de 0,3 mL/min avec un premier pallier de 2 min à 95/5 (H<sub>2</sub>O/ACN) suivi d'un gradient de 36 min de 5 à 35% ACN (pente de 0,83 %/min). Un lavage colonne avec 100 % ACN a ensuite été réalisé pendant 5 min suivi d'un rééquilibrage colonne à 95/5 (H<sub>2</sub>O/ACN) d'une durée de 5 min.

Cette méthode chromatographique est une méthode ayant été optimisée pour la détection et la quantification de peptides chez le gammare et s'est avérée adaptée dans le cas de l'aloose pour l'étape d'identification des protéines.

## **6. Conditions d'analyse de la méthode MSMS en mode MRM**

En spectrométrie de masse en tandem, il est important d'établir et d'optimiser plusieurs conditions comme par exemple les transitions MRM, les paramètres de fragmentation, les paramètres de source ou encore la méthode MSMS selon le mode choisi, ici en mode MRM. Par ailleurs, l'ensemble des spectres de fragmentation sont acquis sur une gamme dynamique allant de 250 à 1500 m/z. De plus, comme une méthode MRM est limitée à 200 transitions, le temps de cycle est limité à 1500 ms pour avoir un nombre de point par pic suffisant, au moins 10 points par pic.



### **a. Etablissement des transitions MRM**

Au vu du nombre important de peptides potentiellement générés par l'étape de protéolyse, il a été important de connaître les transitions MRM qui caractérisent les peptides d'intérêt. Pour cela, le logiciel Skyline a été utilisé, permettant, lorsque les critères de sélection lui sont indiqués, de générer un ensemble de transitions potentielles du peptide après une protéolyse par la trypsine. Dans le cadre du stage et pour éviter de détecter des re-fragmentations, seuls les ions fragments mono chargés ont été retenus et devaient provenir d'ions précurseurs di ou tri chargés. Aucune liste de transitions ne sera donnée en raison de sa longueur trop importante.

### **b. Optimisation des paramètres de fragmentation**

L'optimisation des paramètres de fragmentation doit précéder l'optimisation des paramètres de source. Elle consiste à déterminer les transitions les plus intenses ainsi que les énergies de collision associées et optimales. L'énergie de collision représente l'accélération que subit l'ion précurseur lorsqu'il traverse la cellule de collision et rentre en collision avec les molécules de gaz.

Cette étape a été effectuée sans colonne (en FIA, Flow Injection Analysis) via le logiciel Agilent *Optimizer* sur des solutions standards de peptides individuels. Ainsi, à partir de la masse de la molécule, le logiciel va déterminer le ou les ions précurseurs, le ou les ions fragments et les énergies de collision optimales associés à chaque transition.

Le logiciel *Optimizer* va procéder de la manière suivante :

- Sélection de l'ion précurseur en mode SIM (*Single Ion Monitoring*)
- Détermination des ions fragments issus de cet ion précurseur en mode *Product Ion Scan*
- En mode MRM, le logiciel va déterminer pour chaque transition formée l'énergie de collision qui permet d'obtenir le signal le plus intense

L'ensemble des ions précurseurs, des ions fragments et des énergies de collision optimaux est résumé en *Annexe 3*.

### **c. Optimisation des paramètres de source**

Dès lors que les paramètres de fragmentation ont été fixés, il a été possible d'optimiser les paramètres de source parmi lesquels :

- Les pressions en haute et basse fréquence appliquées à l'*ion funnel* qui est un dispositif permettant de focaliser un faisceau d'ions à la sortie de Q1
- La tension appliquée au niveau du capillaire qui permet de conduire les ions de la source vers le spectromètre de masse en passant d'une pression atmosphérique au vide
- La pression et la tension appliquées au niveau du nébuliseur dont le rôle est de mettre l'échantillon initialement à l'état liquide sous la forme d'un nuage de gouttelettes pour faciliter l'ionisation.
- La température et le débit du gaz séchant et ionisant "*sheath gas*" qui va aider à l'ionisation en réduisant la dispersion des ions du spray.
- La température et le débit d'azote qui permet la désolvatation des molécules de la phase mobile au niveau du capillaire et prévient la formation d'agrégats. Cela permet entre autres de créer un nuage de gouttelettes.

Ces paramètres ont été optimisés à l'aide du logiciel Agilent *Source Optimizer*. Il permet d'optimiser ces paramètres dans une méthode d'acquisition en fonction des composés d'intérêts. La méthode d'acquisition doit être paramétrée de manière à avoir un temps d'analyse le plus court possible (inférieur à 10 min) afin que l'optimisation puisse être réalisée en un temps relativement court. En effet, pour chaque paramètre, le logiciel *Source Optimizer* lance plusieurs runs pouvant aller jusqu'à 56, pour des intervalles et des pas définis. La méthode HPLC utilisée pour cette optimisation a consisté en un gradient de 6 min de 5 à 35% d'ACN (pente de 5%/min), suivi d'un lavage de la colonne avec 100% d'ACN d'une durée de 0,5 min et s'est terminée par un rééquilibrage de la colonne à 95/5 (H<sub>2</sub>O/ACN) d'une durée de 0,5 min, le tout réalisé à un débit de 0,3 mL/min.

Les valeurs des paramètres de source avant et après optimisation sont présentées dans le *Tableau 4*.

*Tableau 4 : Résumé des valeurs de paramètres de source du spectromètre de masse avant et après optimisation*

	Avant optimisation	Après optimisation
Pression en haute fréquence (V)	145	130
Pression en basse fréquence (V)	115	80
Tension au capillaire (V)	4500	4500
Pression au nébuliseur (psi)	40	40
Tension au nébuliseur (V)	0	500
Température du gaz séchant (°C)	250	300
Débit du gaz séchant (L/min)	12	10
Température de l'azote (°C)	210	210
Débit d'azote (L/min)	17	11

## **7. Application de la méthode sur des échantillons de larves à différentes conditions d'exposition (pH ou groupe de géniteur différent)**

La stratégie de quantification utilisée a été la méthode AQUA qui consiste à ajouter en concentration connue une solution de peptides marqués après l'étape de digestion enzymatique. Ainsi, les tests ont été réalisés sur des pools de 5 larves en 3 réplicas par condition d'exposition à différents pH (7, 8 et 9), mais aussi sur deux séries de larves provenant de deux groupes de géniteurs différents.

Il est à noter que lors des expositions il n'y a pas eu de grande différence de pH entre les conditions de pH8 et de pH9.

Pour revenir à la concentration en protéines dans nos échantillons par étalonnage interne en un point, la formule utilisée est :

$$C_{\text{analyte}} = \frac{\text{Aire sous le pic de l'analyte}}{\text{Aire sous le pic de l'éalon interne}} \times C_{\text{étalon interne}}$$

Il a été choisi d'exprimer la quantité de protéines en nanogramme par larve, ainsi en tenant compte des volumes, il suffit de multiplier la concentration par un facteur 16,896/5, 16,896 étant le facteur de dilution appliqué à l'échantillon et 5 le nombre de larves dans le pool.

## **8. Tests et conditions statistiques pour l'analyse de donnée**

### **a. Test ANOVA**

L'ensemble des séries de données ont premièrement été soumis à un test statistique ANOVA pour l'étude des variance, dans le but de savoir s'il y a une différence significative entre les groupes.

Les hypothèses nulle ( $H_0$ ) et relative ( $H_1$ ) ainsi que le coefficient de significativité émis pour les tests ANOVA sont :

- $H_0$  : la quantité de protéine est équivalente entre les différentes conditions de pH et/ou groupe de géniteurs comparés
- $H_1$  : la quantité de protéine est différente entre les différentes conditions de pH et/ou groupe de géniteurs comparés
- $\alpha = 0,05$  : si p-value  $> \alpha$  alors  $H_1$  est rejeté en faveur de  $H_0$  ; si p-value  $< \alpha$  alors  $H_0$  est rejeté en faveur de  $H_1$

### **b. Test de Tukey**

Un test de *Tukey* a été fait, pour l'ensemble des séries de données, dans le but de déterminer quels sont les groupes de mesure qui diffèrent les uns des autres. Ce test n'a été réalisé seulement si le test ANOVA montrait qu'il y avait une différence significative entre les groupes. En effet, ce test permet de faire une comparaison multiple des moyennes en une seule étape, dans le but de savoir si une ou plusieurs moyennes sont significativement différentes l'une de l'autre.

Ce test consiste premièrement à calculer une valeur critique, noté  $Q_c$ , et de la comparer avec la valeur absolue des différences de moyenne de chaque comparaison faite. En conséquence, si  $Q_c$  est supérieur à la différence de moyenne, en valeur absolue, alors les groupes comparés ne sont pas différents. Dans le cas contraire, si  $Q_c$  est inférieur à la différence de moyenne, en valeur absolue, alors les groupes comparés présentent une différence significative.

$$Q_c = Q_{\text{value}} \times \sqrt{\frac{\text{moyenne des carrées à l'intérieur d'un groupe}}{n}}$$

Avec  $Q_{\text{value}} = 4,339$ , déterminé pour un risque de 5%, avec un degré de liberté de 6 et un nombre de groupes de 3

$n$  = nombre de données par groupe

## V. Résultats et discussions

### 1. Recherche et identification des potentiels peptides

La liste de peptides identifiés par digestion *in silico* et des peptides étant conservés par alignement, ainsi que leurs variants sont présentés en *Annexe 2*. Par ailleurs, les peptides identifiés comme protéotypiques sont mis en évidence en jaune dans le tableau de l'*Annexe 2*.

Une première recherche expérimentale des peptides protéotypiques sur un pool de 4 larves a permis de mettre en évidence la potentielle présence de 14 peptides chez l'alose. Ces 14 peptides ont été suspectés après la détection d'un signal présentant au moins 3 transitions ou plus superposées du peptide dans l'échantillon de 4 larves. La *Figure 8* est un exemple correspondant au chromatogramme partiel du peptide NALAHAVQSAR dans un échantillon de 4 larves qui montre un signal présentant plus de 3 transitions superposées à 8,2 min.

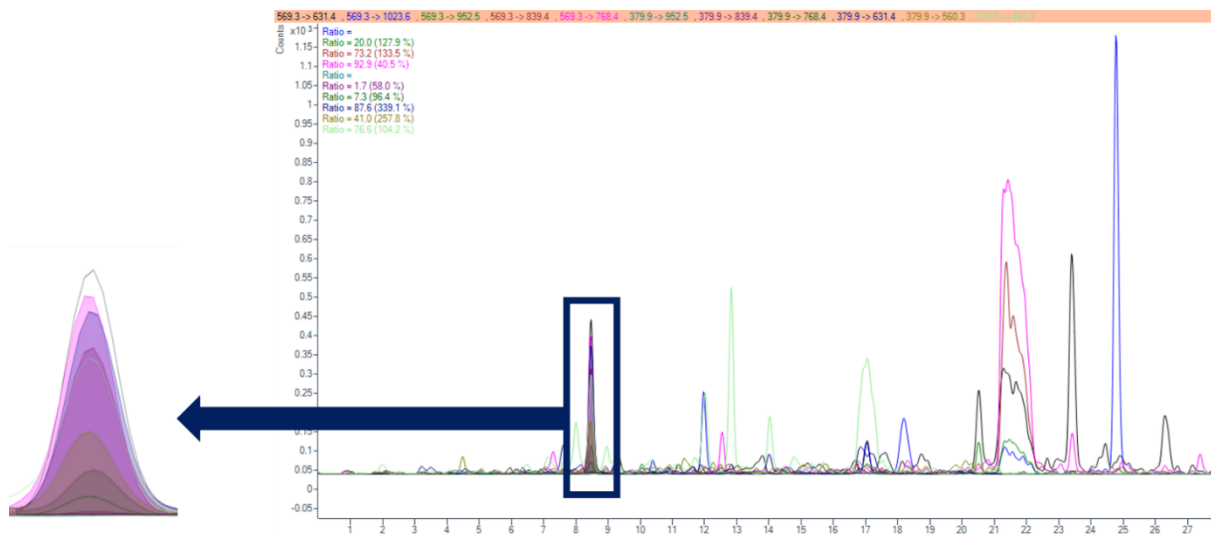


Figure 8 : Chromatogramme du peptide NALAHAVQSAR de la protéine de myosine dans un échantillon de 4 larves. Le zoom du pic d'intérêt à 8,2 min est présenté à gauche du chromatogramme

Ces 14 peptides sont présentés dans le *Tableau 6*. Parmi ces 14 peptides, 10 sont issus de la digestion *in silico* via Skyline tandis que 4 sont issus de la sélection par alignement. Sur les 4 peptides issus de l'alignement, 3 présentent un acide aminé différent de la séquence protéique de l'espèce de référence et sont mis en évidence dans le *Tableau 5*.

Tableau 5 : Liste des peptides suspectés expérimentalement chez l'aloise dans un pool de 4 larves. En rouge, les peptides trouvés par alignement et en gras noir, l'acide aminé différent de la séquence protéique de référence.

Nom de la protéine	Peptides
<b>Myosine</b>	MEIDDLSSNMEAVAK
	NALAHAVQSAR
	EALVSQLTR
<b>Myogénine</b>	MPGGFDQGGYQER
	<b>HQEQPHCPGQCLPWACK</b>
	<b>STLTNPNQR</b>
<b>Vitellogénine</b>	LLGYQAAAYLDK
	SQPLGSIYVK
<b>Acétylcholinestérase</b>	YWANFAR
	SQFDHYSK
<b>Métallothionéine</b>	<b>SCCTCCPSGCSK</b>
<b>Cytochrome</b>	IVSEHYASYDK
	<b>SLAFSTDQVGWVR</b>
<b>Phocéine</b>	QFCLELNLAVK

Afin de confirmer la présence de ces peptides suspectés expérimentalement, des peptides marqués  $^{13}\text{C}$  ont été utilisés. En effet, l'ajout d'un peptide marqué à l'échantillon permet de confirmer l'identification du peptide léger si le peptide marqué et le peptide à identifier présentent :

- Un pic au même temps de rétention
- Au moins 3 transitions superposées communes
- Des rapports de transitions communs d'un échantillon à l'autre

Cependant, il est toutefois nécessaire de procéder à une optimisation des paramètres de sources pour avoir un signal optimal.

## 2. Paramètres de source optimisés

La Figure 9 présente le chromatogramme de 14 peptides étalons lourds, avant et après l'optimisation des paramètres de source.

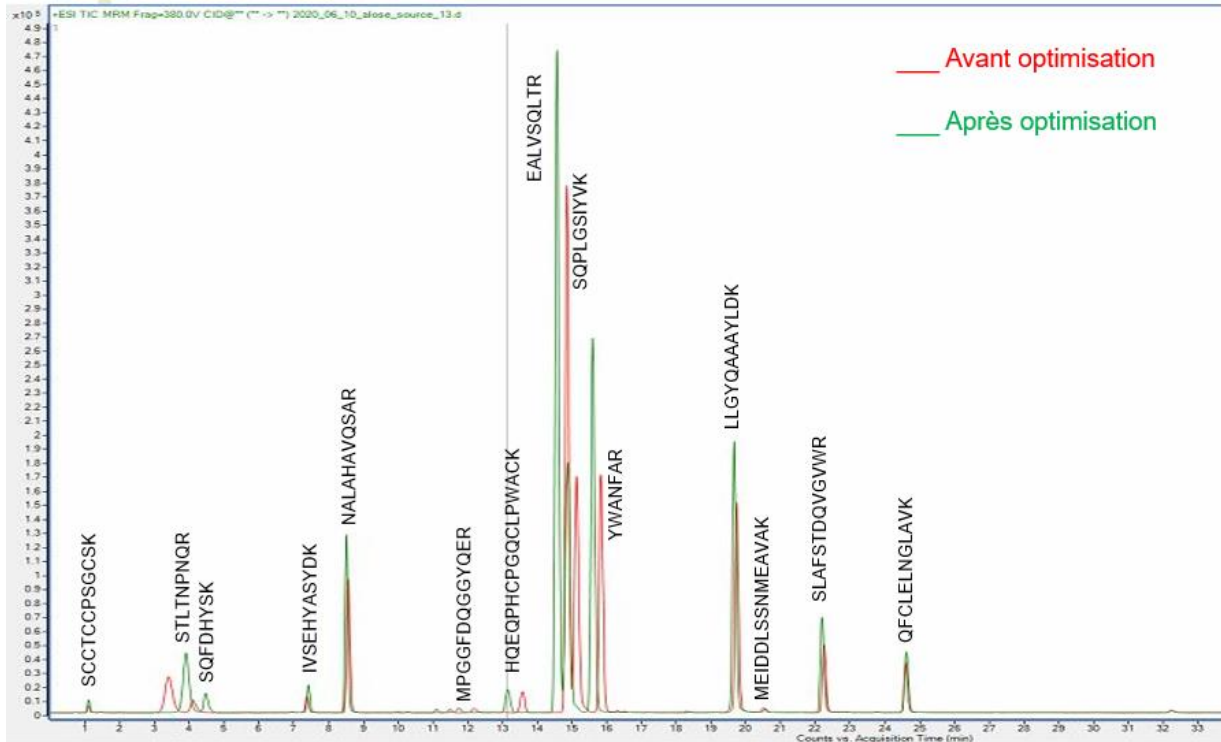


Figure 9 : Chromatogramme des 14 peptides étalons avant (en rouge) et après (en vert) l'optimisation des paramètres de source

L'optimisation des paramètres de source a permis d'augmenter l'intensité du signal d'un facteur 1,4 en moyenne. L'ensemble de ces facteurs sont résumés dans le Tableau 6.

Tableau 6 : Tableau indiquant pour chaque peptides l'intensité avant et après optimisation des paramètres de sources ainsi que le gain de signal obtenu

Protéine	Peptide	Intensité avant optimisation	Intensité après optimisation	Gain de signal
Myosine	NALAHAVQSAR	2,6.10 <sup>4</sup>	3,4.10 <sup>4</sup>	X 1,3
	EALVSQLTR	2,5.10 <sup>5</sup>	3,2.10 <sup>5</sup>	X 1,3
	MEIDDLSSNMEAVAK	1,05.10 <sup>3</sup>	1,4.10 <sup>3</sup>	X 1,3
Myogénine	HQEQPHCPGQCLPWACK	9,5.10 <sup>3</sup>	1,05.10 <sup>4</sup>	X 1,1
	MPGGFDQGGYQER	1,05.10 <sup>3</sup>	1,25.10 <sup>3</sup>	X 1,2
	STLTNPNQR	1,1.10 <sup>4</sup>	1,8.10 <sup>4</sup>	X 1,6
Vitellogénine	LLGYQAAAYLDK	6,25.10 <sup>4</sup>	8,5.10 <sup>4</sup>	X 1,4
	SQPLGSIYVK	1,1.10 <sup>5</sup>	1,15.10 <sup>5</sup>	X 1,05
Acétylcholinestérase	SQFDHYSK	4,6.10 <sup>3</sup>	6,5.10 <sup>3</sup>	X 1,4
	YWANFAR	1,25.10 <sup>5</sup>	1,95.10 <sup>5</sup>	X 1,6
Cytochrome 1A1	IVSEHYASYDK	4.10 <sup>3</sup>	7.10 <sup>3</sup>	X 1,75
	SLAFSTDQGVWR	1,5.10 <sup>4</sup>	2.10 <sup>4</sup>	X 1,3
Métallothionéine	SCCTCCPSGCSK	1,8.10 <sup>3</sup>	4.10 <sup>3</sup>	X 2,2
Phocéine	QFCLELNGLAVK	9,5.10 <sup>3</sup>	1,04.10 <sup>4</sup>	X 1,1

Concernant le bruit de fond, il est le même avant et après optimisation. Le rapport signal sur bruit a donc été amélioré par un facteur 1,4 en moyenne.

Par ailleurs, un décalage en temps de rétention a été observé entre les pics des chromatogrammes avant et après optimisation des paramètres de source. Ce décalage peut s'expliquer soit par un vieillissement de la colonne, soit par une mauvaise stabilisation de la colonne avant injection.

### 3. Validation des peptides identifiés

Après optimisation des paramètres de sources, des analyses sur des pools de tailles différentes (5 larves, 10 larves et 20 larves) ont été effectuées et seuls 3 des 14 peptides (NALAHAVQSAR, EALSQQLTR et LLGYQAAAYLDK) ont réellement été confirmés/validés expérimentalement. La *Figure 10* présente l'exemple du peptide NALAHAVQSAR, dans un échantillon de 20 larves, validé par un peptide marqué avec un temps de rétention de 7,9 min pour le peptide léger et un temps de rétention de 7,9 min pour le lourd.



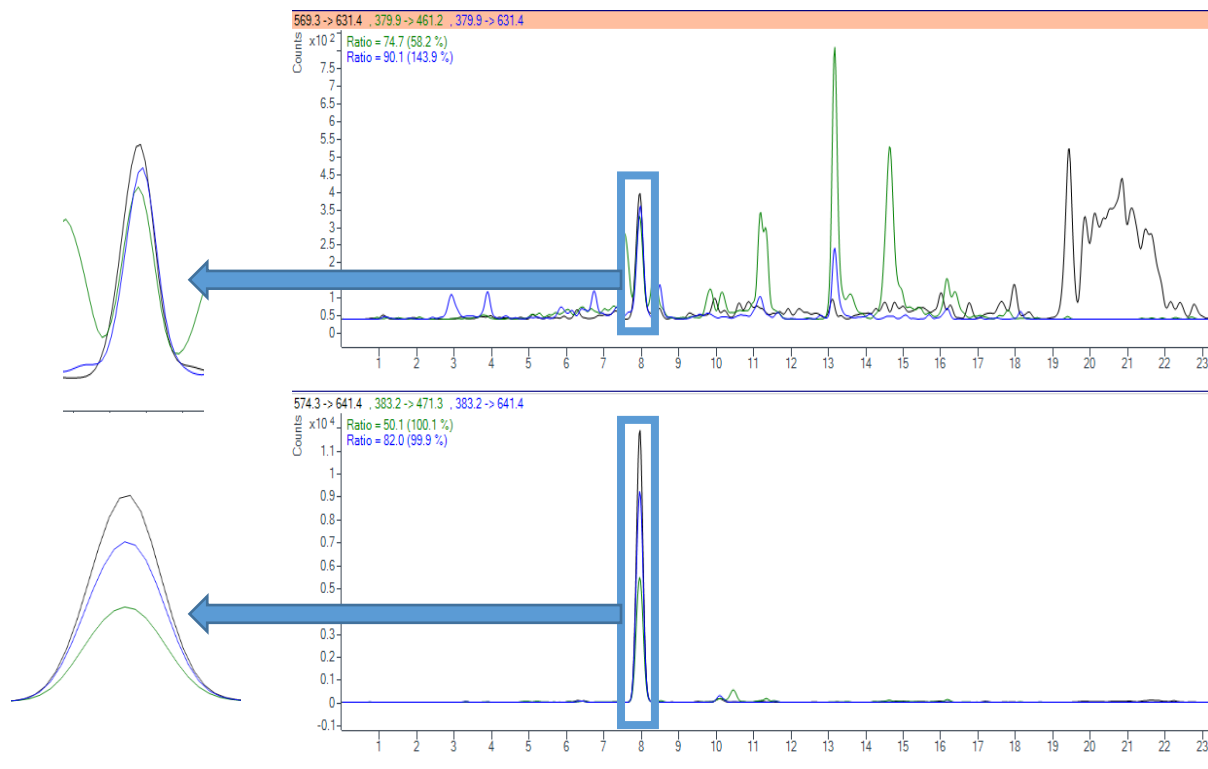


Figure 10 : Chromatogrammes du peptide NALAHAVQSAR dans un échantillon de 20 larves, léger (en haut) et lourd (en bas). Les zooms des pics d'intérêt à 7,9 min sont présentés à gauche des chromatogrammes

Les résultats permettant de répondre aux 3 critères de validation pour le peptide NALAHAVQSAR sont présentés dans le *Tableau 7*.

Tableau 7 : Tableau résumant les 3 critères de validation pour le peptide NALAHAVQSAR

	tr (min)	Ion précurseur	Ion fragment	Ratio
<b>Peptide léger</b>	7,9	569,3	631,4	Quantifiante
		379,9	631,4	79,5
			461,2	55,2
<b>Peptide lourd</b>	7,9	574,3	641,4	Quantifiante
		383,2	641,4	82
			471,3	50,1

Ici, il est possible d'observer que les rapports de transitions entre les transitions quantifiante et les transitions qualifiantes sont globalement constants à  $\pm 20\%$  entre le lourd et le léger. De plus, ils sortent au même temps de rétention ce qui permet alors de confirmer l'identification de ce peptide.

Au vu de ces résultats et du faible nombre de peptides validés, deux hypothèses peuvent être proposées :

- La méthode n'est pas assez sensible pour détecter les peptides aux concentrations de l'échantillon
- La séquence protéique de l'aloise est différente de celle des espèces de référence (hareng et poisson zèbre) utilisées lors de la recherche des peptides potentiels

Après avoir émis ces deux hypothèses, une recherche des 14 peptides a été faite sur des tissus d'aloise adultes (muscle, cerveau, œuf et foie) pour tenter d'écartier l'hypothèse d'un manque de sensibilité. En effet, l'utilisation de tissus d'adultes permet de cibler les tissus riches en protéines d'intérêt, contrairement à l'utilisation de larves entières, dont la concentration en protéines est de 0,2 mg/larve en moyenne. Ainsi, si les protéines dans les tissus sont à des niveaux de concentration détectables et quantifiables par spectrométrie de masse, alors l'absence de signal confirmerait qu'il y a une mauvaise conservation de la séquence protéique.

L'ensemble des données et observations récoltées a permis d'identifier dans des tissus spécifiques (muscle, cerveau, œuf, foie), en plus des 3 peptides déjà confirmés chez la larve, 5 peptides supplémentaires identifiés par alignement (MEIDDLSSNMEAVAK, STLTNPQR, SQPLGSIYVK, SCCTCCPSGCSK, QFCLELNGLAVK) indiqués dans le *Tableau 8*.

*Tableau 8 : Tableau indiquant les différents peptides retrouvés dans les différents tissus de l'aloise adulte (Muscle, Cerveau, Œuf, Foie). Les cases grisées indiquent que le peptide n'est pas présent (absence de signal ou signal ne répondant pas aux critères de validations).*

Protéines	Peptides	Muscle	Cerveau	Œuf	Foie
Myosine	MEIDDLSSNMEAVAK	Validé			
	NALAHAVQSAR	Validé			
	EALVSQLTR	Validé			
Myogénine	MPGGFDQGGYQER				
	HQEQLPHCPGQCLPWACK				
	STLTNPQR		Validé	Validé	
Vitellogénine	LLGYQAAAYLDK		Validé	Validé	Validé
	SQPLGSIYVK		Validé	Validé	Validé
Acétylcholinestérase	YWANFAR				
	SQFDHYSK				
Métallothionéine	SCCTCCPSGCSK	Validé	Validé	Validé	Validé
Cytochrome	IVSEHYASYDK				
	SLAFSTDQVGWVR				
Phocéine	QFCLELNGLAVK		Validé		

Par exemple, la *Figure 11* présente le chromatogramme du peptide STLTNPQR dans le cerveau. Ce dernier montre bien la présence de ce peptide chez l'aloise car les trois conditions (temps de rétention identique, transitions et ratios identiques) ont été remplies.

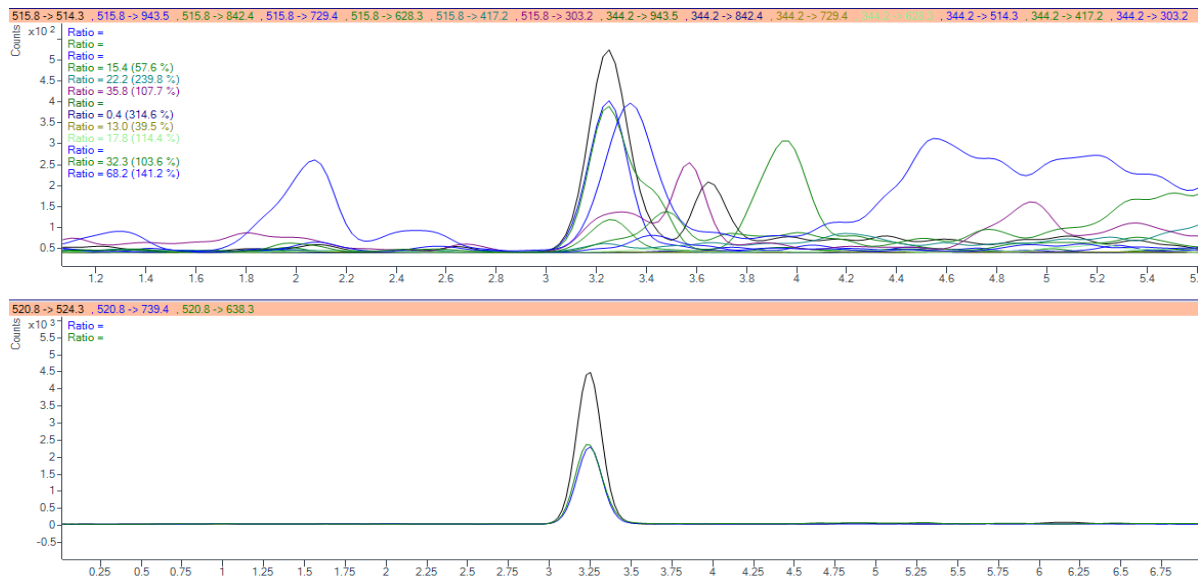


Figure 11 : Chromatogramme du peptide STLTPNQR dans le cerveau

Plusieurs peptides (6 peptides) non détectés chez les larves n'ont toujours pas été détectés chez l'aloise adulte. En effet, pour les peptides non validés, soit les temps de rétention des peptides lourds ne correspondaient pas aux signaux observés pour les peptides légers, soit aucun signal n'a été détecté. Malheureusement, au vu des résultats, aucune des hypothèses citées précédemment ne peut réellement être écartée. Effectivement, pour 6/14 peptides non validés, l'hypothèse d'un manque de sensibilité chez la larve n'est pas totalement vraie. De même, l'hypothèse d'une mauvaise conservation des séquences protéiques entre les espèces plus ou moins proches phylogénétiquement ne peut être écartée car 8 peptides ont pu être validés. Ainsi, la connaissance du génome ou du transcriptome semble être importante puisqu'elle permettrait de déduire les séquences protéiques et ainsi s'affranchir de ces doutes concernant la sensibilité et la conservation des séquences protéiques.

#### **4. Application de la méthode pour la quantification des peptides EALVSQLTR, NALAHVQSAR et LLGYQAAAYLDK sur des échantillons de larves exposées à différents pH ou provenant de groupes de géniteurs différents**

Au vu du faible nombre de peptides validés, la phase d'application n'a été fait que sur trois peptides validés qui sont EALVSQLTR et NALAHAVQSAR spécifiques de la myosine ainsi que LLGYQAAAYLDK spécifique de la vitellogénine. La quantification de ces peptides va permettre de comparer les quantités de protéines (myosine et vitellogénine) chez des larves d'aloise ayant été exposées à différentes conditions de pH (7, 8 et 9) ou provenant de deux groupes de géniteurs différents (série 1 et série 2). Comme, il a été précisé dans la partie matériel et méthode, les conditions d'exposition à pH8 et pH9 sont similaires avec respectivement les pH expérimentaux de 8,28 et 8,25.

##### **a. Effet du pH sur la quantité de protéines**

Cette partie fait état des études, tests et interprétations des résultats obtenus lors de la quantification des protéines myosine et vitellogénine pour l'étude de l'effet du pH sur le développement de la larve et donc sur la quantité de chaque protéine.

Les *Figures 12, 13 et 14* représentent respectivement les quantités de myosine selon EALVSQLTR et NALAHAVQSAR et les quantités de vitellogénine selon LLGYQAAAYLDK en fonction des conditions de pH auxquelles les larves ont été exposées.

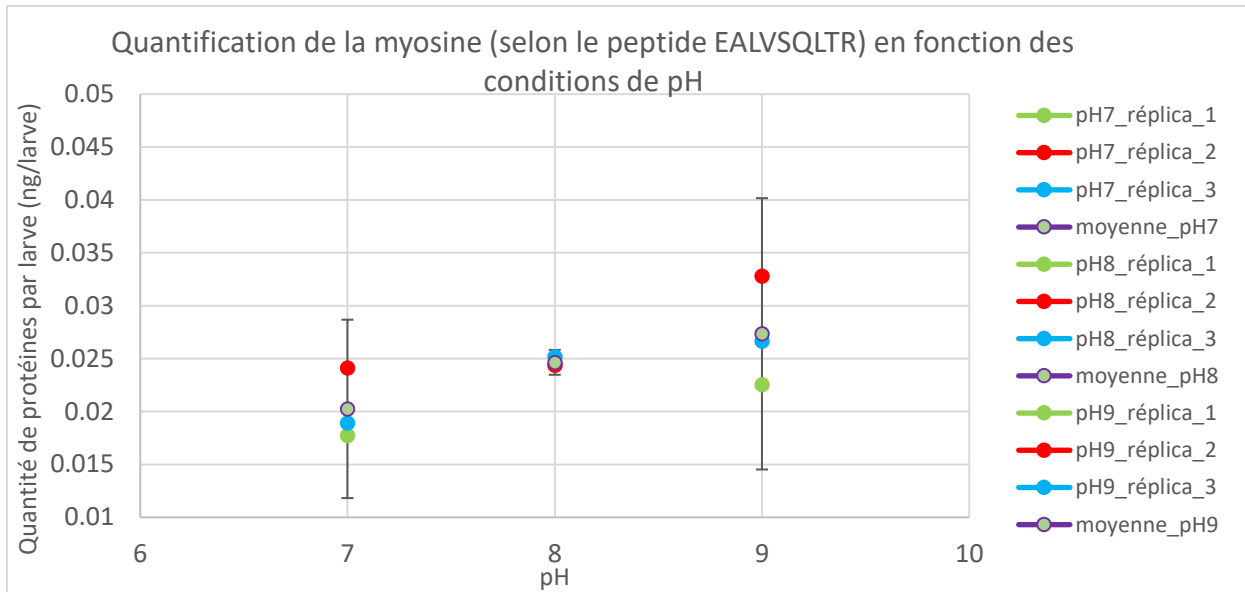


Figure 12 : Quantité de peptide en ng/larve (selon le peptide EALVSQLTR), sur l'axe y, en fonction des conditions de pH (7, 8, 9). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encerclé en violet : moyenne de la quantité de protéines par larve en ng pour chaque condition de pH ; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.

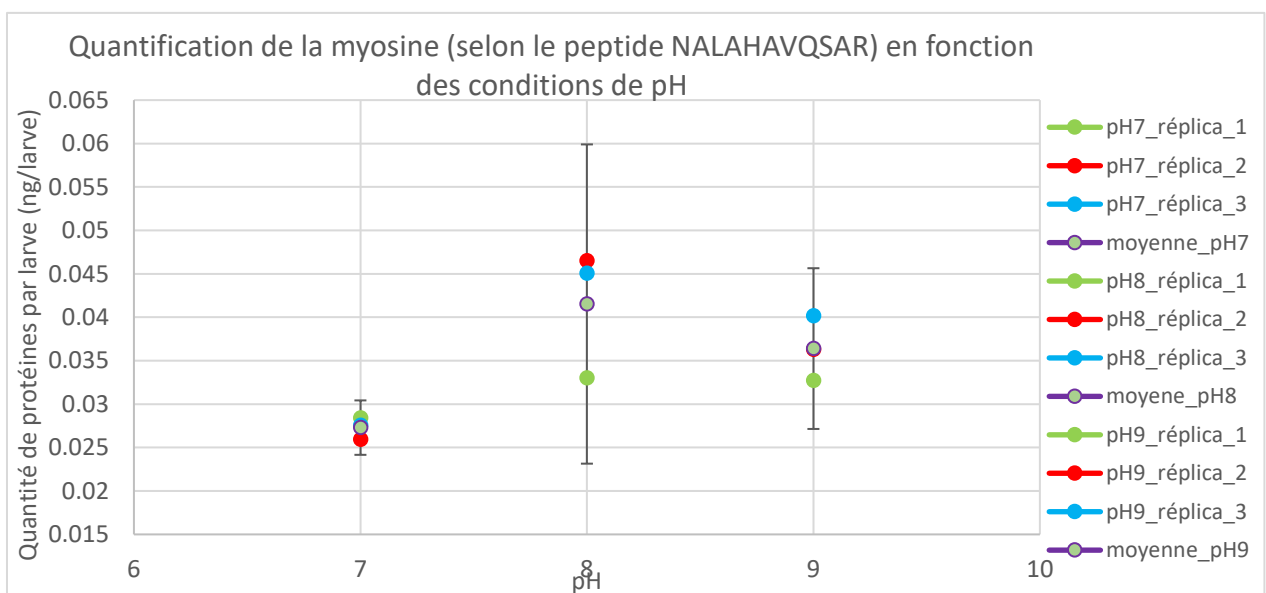


Figure 13 : Quantité de peptide en ng/larve (selon le peptide NALAHAVQSAR), sur l'axe y, en fonction des conditions de pH (7, 8, 9). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encerclé en violet : moyenne de la quantité de protéines par larve en ng pour chaque condition de pH ; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.

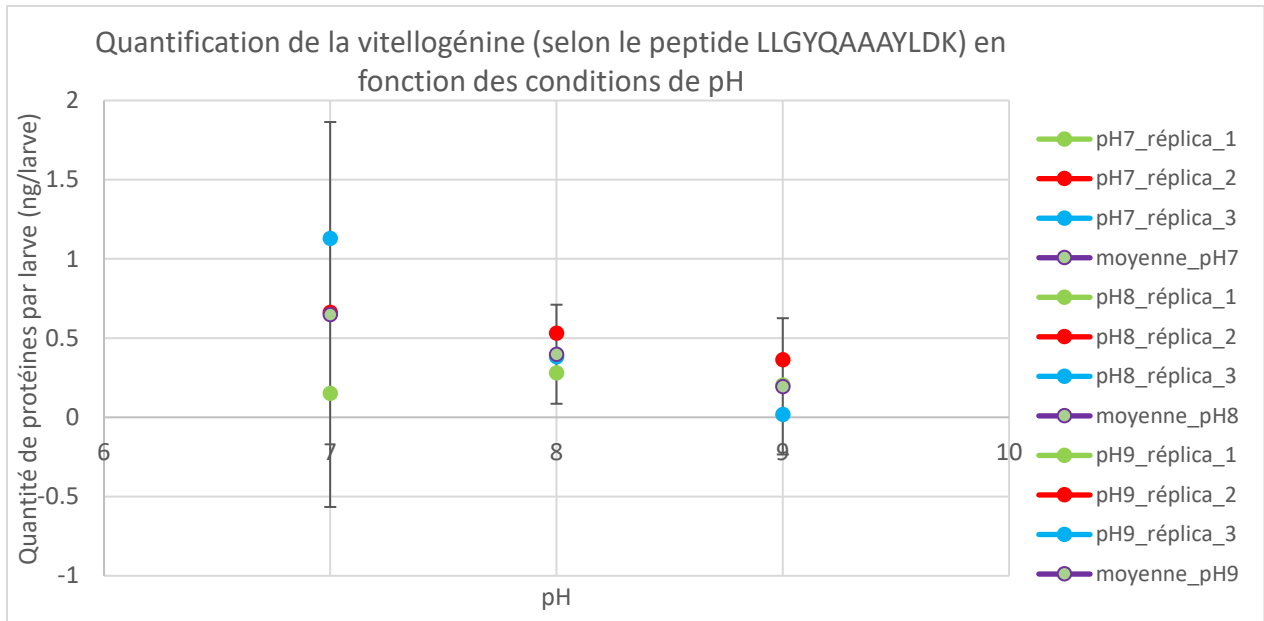


Figure 14 : Quantité de peptide en ng/larve (selon le peptide LLGYQAAAYLDK), sur l'axe y, en fonction des conditions de pH (7, 8, 9). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encadré en violet : moyenne de la quantité de protéines par larve en ng pour chaque condition de pH ; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.

Les Figure 12 et 13 semblent montrer respectivement que pour les peptides EALVSQLTR et NALAHAVQSAR une quantité de protéine détecté plus faible à pH7. En revanche, la Figure 14 quant à elle montre qu'en moyenne, la quantité de protéine détectée est plus élevée à pH7.

Par ailleurs, les différents tests ANOVA montrent que :

- Pour le peptide EALVSQLTR, il n'y a pas de différence de quantité de protéine selon le pH d'exposition. En effet, avec une p-value de 0,1245 supérieure à 0,05 le coefficient de significativité,  $H_1$  est rejeté en faveur de  $H_0$ .
- Pour le peptide NALAHAVQSAR, il existe une différence entre les groupes de larves exposés à différents pH (7, 8, 9). Effectivement, avec une p-value de 0,0013 inférieure à 0,05 le coefficient de significativité,  $H_0$  est rejeté en faveur de  $H_1$ .
- Pour le peptide LLGYQAAAYLDK, il n'existe pas de différence de quantité de protéine entre les groupes de larves à différent pH. Il est vrai qu'avec une p-value de 0,2721 supérieure à 0,05 le coefficient de significativité,  $H_1$  est rejeté en faveur de  $H_0$ .

Maintenant qu'il a été montré que pour le peptide NALAHAVQSAR, les groupes diffèrent les uns des autres en termes de quantité de protéines, un test de *Tukey* permet de connaître quels sont ces groupes.

La valeur de  $Q_c$  pour NALAHAVQSAR est de 0,0063. Ainsi, le *Tableau 9* présente les résultats de ce test sous forme matriciel (3,3) pour ce peptide. Ici, les valeurs présentées correspondent aux valeurs absolues des différences de moyenne pour chaque comparaison.

*Tableau 9 : Tableau indiquant les valeurs absolues des différences de moyenne pour chaque comparaison de groupe pour le peptide NALAHAVQSAR. En rouge, les valeurs qui sont supérieures à  $Q_c$  et indique donc que les deux groupes sont différents l'un de l'autre.*

NALAHAVQSAR	pH7	pH8	pH9
pH7	0,0000		
pH8	<b>0,0142</b>	0,0000	
pH9	<b>0,0091</b>	0,0051	0,0000

Le test de *Tukey* montre pour NALAHAVQSAR qu'il y a une différence significative entre les larves exposées à pH8 et pH7, ainsi qu'entre celles exposées à pH9 et pH7. En ce qui concerne les peptides EALVSQLTR et LLGYQAAAYLDK, les tests de *Tukey* réalisés ont montré qu'il n'y a pas de différence significative entre les groupes de larves exposés aux différents pH. En effet, chaque valeur absolue des différences de moyenne est inférieure aux  $Q_c$  respectivement de 0,0089 et 0,7720.

Ici, il est possible d'observer une légère différence entre les profils des peptides EALVSQLTR et NALAHAVQSAR, notamment au niveau des larves exposées à pH7. Effectivement, comme il a été montré lors de l'interprétation des tests ANOVA et *Tukey*, pour le peptide EALVSQLTR la quantité de myosine à pH7 est équivalente aux conditions de pH8 et pH9 contrairement au peptide NALAHAVQSAR où la quantité de protéine à pH7 est différente de pH8 et pH9.

Cette différence peut être expliquée par un problème de sensibilité du spectromètre de masse vis-à-vis de l'un des peptides. Ce manque de sensibilité peut être dû à la faible quantité de matrice biologique utilisée. En outre, cette étude pourrait permettre de montrer qu'une exposition à des pH différents n'influence pas le développement de la larve dans le cas où le manque de sensibilité de la méthode ne joue pas sur le peptide EALVSQLTR mais sur NALAHAVQSAR. A l'inverse, si le manque de sensibilité joue sur le peptide EALVSQLTR, et compte tenu des observations faites sur les *Figures 12 et 13*, il semblerait qu'une exposition à un pH7 pourrait influencer le développement de la larve. Ainsi, comme les peptides EALVSQLTR et NALAHAVQSAR sont des peptides de la myosine, une protéine du muscle, si la quantité de protéine est plus élevée cela pourrait traduire un développement musculaire plus avancé. Dans le cas présent, le fait que la quantité de protéines soit plus faible à

pH7 qu'aux pH8 et pH9 montre alors un développement moins avancé lorsque la larve est exposée à un pH7.

### **b. Effet du groupe de géniteur sur la quantité de protéine**

Cette partie fait état des études, des tests et des interprétations des résultats obtenus lors de la quantification des protéines myosine et vitellogénine pour l'étude de l'effet du groupe de géniteurs sur le développement de la larve et donc sur la quantité de protéine.

Les *Figures 15, 16 et 17* représentent respectivement les quantités de myosine selon EALVSQLTR et NALAHVQSAR et les quantités de vitellogénine selon LLGYQAAAYLDK en fonction du groupe de géniteurs.

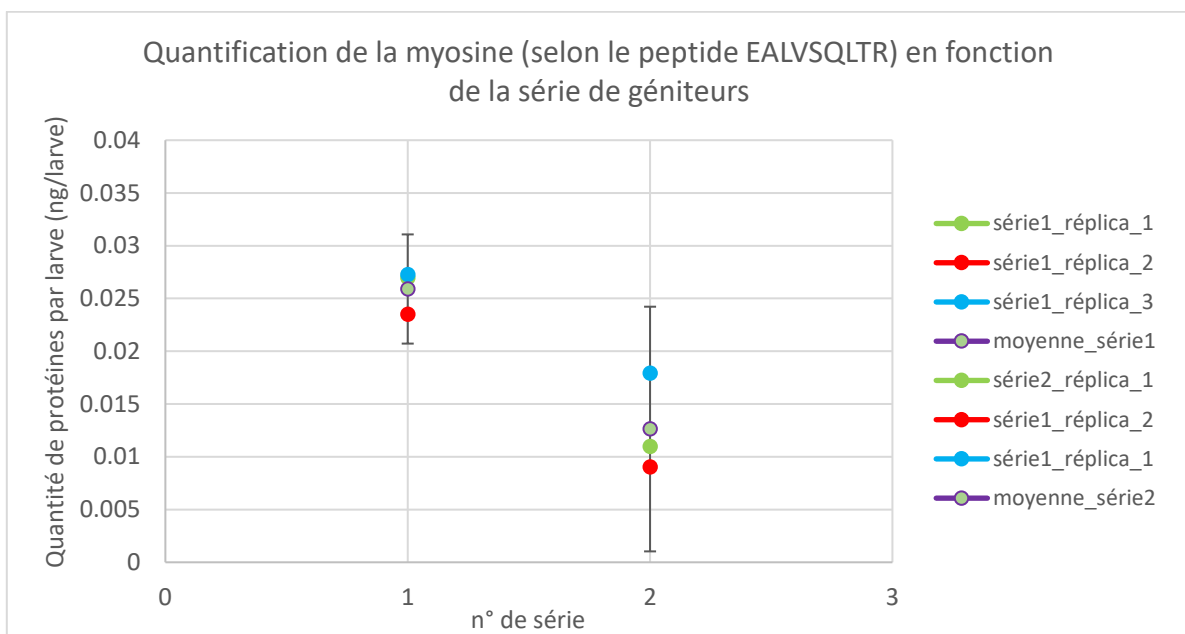


Figure 15 : Quantité de peptide en ng/larve (selon le peptide EALVSQLTR), sur l'axe y, en fonction du n° série de géniteurs (1, 2). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encerclé en violet : moyenne de la quantité de protéines par larve en ng pour chaque série de géniteurs; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.



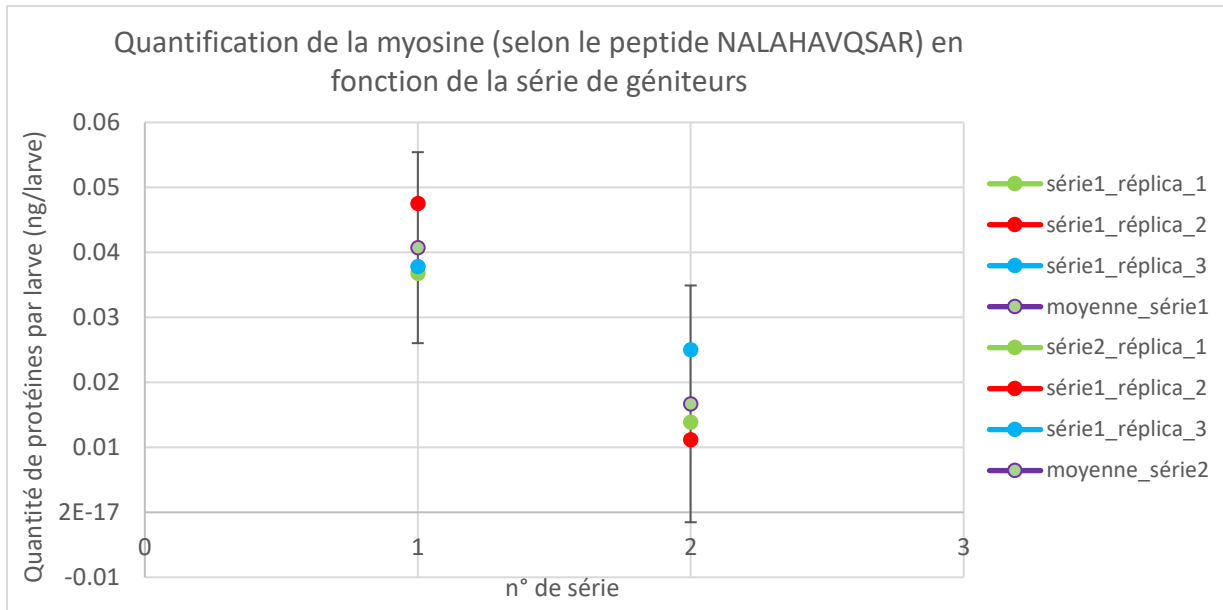


Figure 16 : Quantité de peptide en ng/larve (selon le peptide NALAHAVQSAR), sur l'axe y, en fonction du n° série de géniteurs (1, 2). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encadré en violet : moyenne de la quantité de protéines par larve en ng pour chaque série de géniteurs; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.

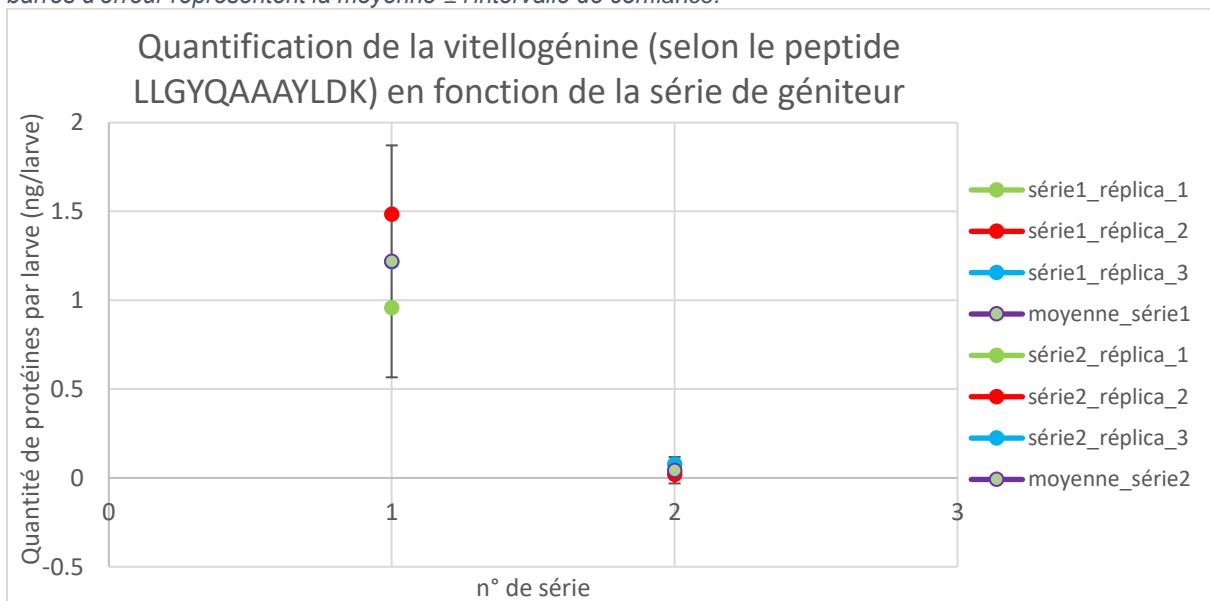


Figure 17 : Quantité de peptide en ng/larve (selon le peptide LLGYQAAAYLDK), sur l'axe y, en fonction du n° série de géniteurs (1, 2). Point vert : quantité de protéines par larve en ng pour le pool 1 ; Point rouge : quantité de protéines par larve en ng pour le pool 2 ; Point bleu : quantité de protéines par larve en ng pour le pool 3 ; Point bleu encadré en violet : moyenne de la quantité de protéines par larve en ng pour chaque série de géniteurs; Les barres d'erreur représentent la moyenne  $\pm$  l'intervalle de confiance.

L'ensemble des figures (15, 16, 17) montrent une quantité de protéines détectée plus élevée dans le groupe de géniteur de la série 1.

Comme pour l'étude de l'effet du pH sur la quantité de protéines chez les larves, des tests statistiques ANOVA et Tukey ont été réalisés dans le but de vérifier s'il existe une différence significative de quantité de protéines entre les deux séries de géniteurs. La

valeur  $Q_{\text{value}}$  utilisée pour le calcul de  $Q_c$  est égale à 5,91 et a été recherchée dans les tables pour un risque de 0,05, un degré de liberté de 3 et un  $n$  (nombre de groupes) de 3. Le nombre de mesures par groupe utilisé pour le calcul de  $Q_c$  est comme précédemment de 3.

Les différents tests ANOVA montrent que pour chaque peptide, il existe une différence significative entre les deux séries de géniteurs. En effet, avec respectivement des  $p$ -value de 0,0108, 0,0115 et 0,0015 pour EALVSQLTR, NALAHAVQSAR et LLGYQAAAYLDK toutes inférieures à 0,05 le coefficient de significativité, les résultats montrent qu'il existe bien une différence de quantité de protéines entre les deux séries quel que soit le peptide quantifié.

Compte tenu du faible nombre de mesures effectuées pour chaque série de géniteurs, le test ANOVA n'est pas suffisant pour conclure sur la différence significative entre les groupes. Ainsi, les tests de *Tukey* montrent :

- Pour le peptide EALVSQLTR avec une différence de moyenne, en valeur absolue, entre les 2 séries de 0,0133 supérieure à  $Q_c$  égal à 0,0123 permet de conclure sur la différence significative avec un risque de 5% entre les deux séries de géniteurs.
- Pour le peptide NALAHAVQSAR, la différence de moyenne entre les séries, en valeur absolue, est de 0,0240 et est supérieur à  $Q_c$  qui vaut 0,0227. Cela permet alors de conclure avec un risque de 5% que la quantité de protéine est significativement différente entre les deux séries de géniteurs.
- Pour le peptide LLGYQAAAYLDK, cette différence de moyenne est de 1,1458 contre un  $Q_c$  de 0,6381 indiquant de ce fait la différence significative avec un risque de 5% entre les deux séries de géniteurs.

L'ensemble des tests réalisés dans le cadre de cette étude a montré que le groupe de géniteurs peut également influencer le développement de la larve. De plus, comme les deux lots de géniteurs ont été soumis aux mêmes conditions (température, pH, ...), mais présentent toutefois des teneurs en protéines différentes, il est important de prendre en compte ce paramètre dans la suite des études.

Par ailleurs, l'observation faite sur la quantité de protéine détectée plus élevée pour le groupe de géniteurs de la série 1 montre, pour les peptides de la myosine, un développement plus avancé des larves. Concernant le peptide LLGYQAAAYLDK de la protéine de la vitellogénine, si la quantité de protéine est faible alors un

développement plus avancé de la larve pourrait en être la cause. En effet, la vitellogénine étant une protéine de l'œuf, présente dans le sac vitellin, qui permet à l'embryon de se nourrir avant d'atteindre un développement suffisant pour le faire seul. Ainsi, plus le sac vitellin est petit, plus la quantité de vitellogénine est faible et donc plus la larve est développée. De ce fait, il semblerait que les larves issues de la série 1 ont un développement moins avancé que celles de la série 2. Cette dernière remarque contredit totalement celle faite pour la myosine qui montre un développement plus avancé de la série 1. Il est alors possible de conclure sur la difficulté d'interprétation des résultats d'un peptide à l'autre.

## **VI. Bilan et perspectives**

Ce stage intervient dans la thèse de Marion BLAYA sur l'étude du déclin des populations d'aloses dans la Gironde. L'objectif principal de ce stage était de mettre en avant la spectrométrie de masse pour la mesure de protéines biomarqueurs et évaluer la pertinence d'une approche alternative pour l'identification de peptides chez les espèces non séquencées mais ayant un intérêt environnemental.

Le premier challenge a été d'identifier des peptides protéotypiques en l'absence de génome chez l'alose, pour sept biomarqueurs protéiques sélectionnés. Ces sept protéines (myosine, myogénine, vitellogénine, métallothionéine, acétylcholinestérase, cytochrome et phocéine), déjà reconnues comme étant des biomarqueurs pertinents ont été recherchées et identifiées par phylogénie, stratégie innovante mise en œuvre par Jubeaux *et Al.*<sup>38</sup> dans le cas de la vitellogénine chez le gammare. Les peptides spécifiques de ces protéines ont ainsi été recherchés par spectrométrie de masse MSMS en mode M RM après digestion enzymatique par la trypsine. Cependant, seuls trois peptides spécifiques de deux protéines, myosine et vitellogénine, sur 14 peptides recherchés pour l'ensemble des protéines d'intérêt ont pu être détectés, validés et quantifiés.

L'optimisation d'une méthode HPLC-MSMS a représenté quant à elle le deuxième challenge. Durant ce stage, seuls les paramètres de fragmentation et les paramètres de source ont été optimisés. Ces optimisations ont permis d'améliorer la sensibilité de l'instrument vis-à-vis de chaque peptide et de détecter jusqu'à 3 peptides de façon spécifique. Les quantifications ont pu mettre en évidence des différences de développement de larves suite à l'exposition à des pH différents ou aux groupes de géniteurs dont elles sont issues. Toutefois, une caractérisation de chaque individu en termes de taille et de poids doit être envisagé pour la suite du projet et constitue un challenge de travailler avec des échantillons de petite taille. En effet, il est actuellement difficile de conclure correctement sur l'effet du pH ou du groupe de géniteur sur le développement de la larve d'alose. De plus, cette dernière condition nécessite une capacité à faire les mesures sur une larve ou un groupe de larves ayant quasiment les mêmes caractéristiques.

Ainsi, l'ensemble des résultats acquis durant ce stage a démontré le potentiel de la spectrométrie de masse pour la détection et l'identification de peptides dans des matrices biologiques complexes. Toutefois, la stratégie phylogénétique a montré de

fortes limites en termes d'identification de peptides potentiels pour une espèce de poissons dont le génome n'est pas séquencé. De ce fait, un séquençage du génome ou du transcriptome semble nécessaire pour avoir la séquence exacte des protéines et éviter tous les questionnements et les doutes qu'il y a eu durant le stage, sur la sensibilité et la conservation des séquences protéiques.

Pour finir, ce stage a été marqué par la crise sanitaire du Covid-19 obligeant la suspension des activités sur site pendant deux mois. Cette période a nécessité une adaptation du stage afin de poursuivre les activités à distance qui se sont résumées à la formation sur les logiciels de retraitement de données et à des recherches bibliographique figurant en première partie de ce rapport. Le confinement général du pays a également conduit à la perte d'une partie des échantillons qui auraient dû être disponibles pour les travaux du stage, limitant ainsi l'avancé dans le développement de la méthode de quantification.

## Bibliographie

- (1) Alose : déjà douze ans d'interdiction de pêche !  
<https://www.ladepeche.fr/2020/01/20/alose-deja-douze-ans-dinterdiction-de-peche,8673011.php> (accessed Apr 20, 2020).
- (2) Disparition des poissons migrateurs dans l'estuaire de la Gironde - Sénat  
<https://www.senat.fr/questions/base/2012/qSEQ120701208.html> (accessed Apr 20, 2020).
- (3) Flammarion, P.; Devaux, A.; Garric, J. MARQUEURS BIOCHIMIQUES DE POLLUTION DANS LES ÉCOSYSTÈMES AQUATIQUES CONTINENTAUX. EXEMPLES D'UTILISATION ET PERSPECTIVES POUR LE GESTIONNAIRE. *Bull. Fr. Pêche Piscic.* **2001**, No. 357–360, 209–226.  
<https://doi.org/10.1051/kmae/2001045>.
- (4) Depledge, M. H.; Fossi, M. C. The Role of Biomarkers in Environmental Assessment (2). Invertebrates. *Ecotoxicology* **1994**, 3 (3), 161–172.  
<https://doi.org/10.1007/BF00117081>.
- (5) 2017\_spectrométrie Comme Nouvelle Approche.Pdf.
- (6) Picotti, P.; Aebersold, R. Selected Reaction Monitoring–Based Proteomics: Workflows, Potential, Pitfalls and Future Directions. *Nat. Methods* **2012**, 9 (6), 555–566.  
<https://doi.org/10.1038/nmeth.2015>.
- (7) Vidova, V.; Spacil, Z. A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition. *Anal. Chim. Acta* **2017**, 964, 7–23.  
<https://doi.org/10.1016/j.aca.2017.01.059>.
- (8) Gouveia, D. D. Molecular approaches for the discovery, development and application of toxicity biomarkers in gammarids. 284.
- (9) Leprêtre, M.; Almunia, C.; Armengaud, J.; Salvador, A.; Geffard, A.; Palos-Ladeiro, M. The Immune System of the Freshwater Zebra Mussel, *Dreissena Polymorpha*, Decrypted by Proteogenomics of Hemocytes and Plasma Compartments. *J. Proteomics* **2019**, 202, 103366. <https://doi.org/10.1016/j.jprot.2019.04.016>.
- (10) Leprêtre, M.; Almunia, C.; Armengaud, J.; Le Guernic, A.; Salvador, A.; Geffard, A.; Palos-Ladeiro, M. Identification of Immune-Related Proteins of *Dreissena Polymorpha* Hemocytes and Plasma Involved in Host-Microbe Interactions by Differential Proteomics. *Sci. Rep.* **2020**, 10 (1), 6226. <https://doi.org/10.1038/s41598-020-63321-z>.
- (11) Velotta, J. P.; Wegrzyn, J. L.; Ginzburg, S.; Kang, L.; Czesny, S.; O'Neill, R. J.; McCormick, S. D.; Michalak, P.; Schultz, E. T. Transcriptomic Imprints of Adaptation to Fresh Water: Parallel Evolution of Osmoregulatory Gene Expression in the Alewife. *Mol. Ecol.* **2017**, 26 (3), 831–848. <https://doi.org/10.1111/mec.13983>.
- (12) Rougier, T.; Lambert, P.; Drouineau, H.; Girardin, M.; Castelnaud, G.; Carry, L.; Aprahamian, M.; Rivot, E.; Rochard, E. Collapse of Allis Shad, *Alosa Alosa*, in the Gironde System (Southwest France): Environmental Change, Fishing Mortality, or Allee Effect? *ICES J. Mar. Sci.* **2012**, 69 (10), 1802–1811.  
<https://doi.org/10.1093/icesjms/fss149>.
- (13) Jolly, M. T.; Aprahamian, M. W.; Hawkins, S. J.; Henderson, P. A.; Hillman, R.; O'Maoiléidigh, N.; Maitland, P. S.; Piper, R.; Genner, M. J. Population Genetic Structure of Protected Allis Shad (*Alosa Alosa*) and Twaité Shad (*Alosa Fallax*). *Mar. Biol.* **2012**, 159 (3), 675–687. <https://doi.org/10.1007/s00227-011-1845-x>.
- (14) Grande alose, *Alosa alosa* - Poissons - NatureGate  
<http://www.luontoportti.com/suomi/fr/kalat/grande-alose> (accessed Aug 18, 2020).
- (15) Gregorich, Z. R.; Chang, Y.-H.; Ge, Y. Proteomics in Heart Failure: Top-down or Bottom-Up? *Pflüg. Arch. - Eur. J. Physiol.* **2014**, 466 (6), 1199–1209.  
<https://doi.org/10.1007/s00424-014-1471-9>.

- (16) Landolt, M. L. *Myxosoma Cerebralis* : Isolation and Concentration from Fish Skeletal Elements — Trypsin Digestion Method. *J. Fish. Res. Board Can.* **1973**, 30 (11), 1713–1716. <https://doi.org/10.1139/f73-271>.
- (17) Bateman, N. W.; Goulding, S. P.; Shulman, N. J.; Gadok, A. K.; Szumlinski, K. K.; Wu, C. C. Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA)\*□S. 10.
- (18) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000Research* **2016**, 5, 419. <https://doi.org/10.12688/f1000research.7042.1>.
- (19) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation\*□S. 10.
- (20) Zhao, Y.-Y.; Lin, R.-C. UPLC–MSE Application in Disease Biomarker Discovery: The Discoveries in Proteomics to Metabolomics. *Chem. Biol. Interact.* **2014**, 215, 7–16. <https://doi.org/10.1016/j.cbi.2014.02.014>.
- (21) Colangelo, C. M.; Chung, L.; Bruce, C.; Cheung, K.-H. Review of Software Tools for Design and Analysis of Large Scale MRM Proteomic Datasets. *Methods* **2013**, 61 (3), 287–298. <https://doi.org/10.1016/j.ymeth.2013.05.004>.
- (22) Ahn, Y. H.; Ji, E. S.; Oh, N. R.; Kim, Y.-S.; Ko, J. H.; Yoo, J. S. Differential Proteomic Approach for Identification and Verification of Aberrantly Glycosylated Proteins in Adenocarcinoma Lung Cancer (ADLC) Plasmas by Lectin-Capturing and Targeted Mass Spectrometry. *J. Proteomics* **2014**, 106, 221–229. <https://doi.org/10.1016/j.jprot.2014.04.031>.
- (23) You, J.; Kao, A.; Dillon, R.; Croner, L. J.; Benz, R.; Blume, J. E.; Wilcox, B. A Large-Scale and Robust Dynamic MRM Study of Colorectal Cancer Biomarkers. *J. Proteomics* **2018**, 187, 80–92. <https://doi.org/10.1016/j.jprot.2018.06.013>.
- (24) Rougemont, B.; Bontemps Gallo, S.; Ayciriex, S.; Carrière, R.; Hondermarck, H.; Lacroix, J. M.; Le Blanc, J. C. Y.; Lemoine, J. Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by *Dickeya Dadantii* Proteomic Analysis during Plant Infection. *Anal. Chem.* **2017**, 89 (3), 1421–1426. <https://doi.org/10.1021/acs.analchem.6b03201>.
- (25) Salvador, A.; Carrière, R.; Ayciriex, S.; Margoum, C.; Leblanc, Y.; Lemoine, J. Scout-Multiple Reaction Monitoring: A Liquid Chromatography Tandem Mass Spectrometry Approach for Multi-Residue Pesticide Analysis without Time Scheduling. *J. Chromatogr. A* **2020**, 461046. <https://doi.org/10.1016/j.chroma.2020.461046>.
- (26) Ankney, J. A.; Muneer, A.; Chen, X. Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annu. Rev. Anal. Chem.* **2018**, 11 (1), 49–77. <https://doi.org/10.1146/annurev-anchem-061516-045357>.
- (27) Desouza, L.; Taylor, A.; Li, W.; Minkoff, M.; Romaschin, A.; Colgan, T.; Siu, K. Multiple Reaction Monitoring of MTRAQ-Labeled Peptides Enables Absolute Quantification of Endogenous Levels of a Potential Cancer Marker in Cancerous and Normal Endometrial Tissues. *J. Proteome Res.* **2008**, 7, 3525–3534. <https://doi.org/10.1021/pr800312m>.
- (28) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**, 3 (12), 1154–1169. <https://doi.org/10.1074/mcp.M400129-MCP200>.
- (29) Harwood, M. D.; Achour, B.; Russell, M. R.; Carlson, G. L.; Warhurst, G.; Rostami-Hodjegan, A. Application of an LC–MS/MS Method for the Simultaneous

- Quantification of Human Intestinal Transporter Proteins Absolute Abundance Using a QconCAT Technique. *J. Pharm. Biomed. Anal.* **2015**, *110*, 27–33. <https://doi.org/10.1016/j.jpba.2015.02.043>.
- (30) Jaquinod, M.; Trauchessec, M.; Huillet, C.; Louwagie, M.; Lebert, D.; Picard, G.; Adrait, A.; Dupuis, A.; Garin, J.; Brun, V.; Bruley, C. Mass Spectrometry-Based Absolute Protein Quantification: PSAQ<sup>TM</sup> Strategy Makes Use of “Noncanonical” Proteotypic Peptides. *PROTEOMICS* **2012**, *12* (8), 1217–1221. <https://doi.org/10.1002/pmic.201100538>.
- (31) Adrait, A.; Lebert, D.; Trauchessec, M.; Dupuis, A.; Louwagie, M.; Masselon, C.; Jaquinod, M.; Chevalier, B.; Vandenesch, F.; Garin, J.; Bruley, C.; Brun, V. Development of a Protein Standard Absolute Quantification (PSAQ<sup>TM</sup>) Assay for the Quantification of Staphylococcus Aureus Enterotoxin A in Serum. *J. Proteomics* **2012**, *75* (10), 3041–3049. <https://doi.org/10.1016/j.jprot.2011.11.031>.
- (32) Rauh, M. LC–MS/MS for Protein and Peptide Quantification in Clinical Chemistry. *J. Chromatogr. B* **2012**, *883–884*, 59–67. <https://doi.org/10.1016/j.jchromb.2011.09.030>.
- (33) Roepstorff, P.; Fohlman, J. Letter to the Editors. *Biol. Mass Spectrom.* **1984**, *11* (11), 601–601. <https://doi.org/10.1002/bms.1200111109>.
- (34) James, A.; Jorgensen, C. Basic Design of MRM Assays for Peptide Quantification. In *LC-MS/MS in Proteomics*; Cutillas, P. R., Timms, J. F., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; Vol. 658, pp 167–185. [https://doi.org/10.1007/978-1-60761-780-8\\_10](https://doi.org/10.1007/978-1-60761-780-8_10).
- (35) Gagnaire, B.; Geffard, O.; Noury, P.; Garric, J. In Vivo Indirect Measurement of Cytochrome P450-Associated Activities in Freshwater Gastropod Molluscs. *Environ. Toxicol.* **2010**, *25* (6), 545–553. <https://doi.org/10.1002/tox.20515>.
- (36) Noury, P.; Geffard, O.; Tutundjian, R.; Garric, J. Non Destructive in Vivo Measurement of Ethoxyresorufin Biotransformation by Zebrafish Prolarva: Development and Application. *Environ. Toxicol.* **2006**, *21* (4), 324–331. <https://doi.org/10.1002/tox.20184>.
- (37) Hu, Y.; Xie, S.; Yao, J. Identification of Novel Reference Genes Suitable for QRT-PCR Normalization with Respect to the Zebrafish Developmental Stage. *PLOS ONE* **2016**, *11* (2), e0149277. <https://doi.org/10.1371/journal.pone.0149277>.
- (38) Jubeaux, G.; Audouard-Combe, F.; Simon, R.; Tutundjian, R.; Salvador, A.; Geffard, O.; Chaumot, A. Vitellogenin-like Proteins among Invertebrate Species Diversity: Potential of Proteomic Mass Spectrometry for Biomarker Development. *Environ. Sci. Technol.* **2012**, *46* (11), 6315–6323. <https://doi.org/10.1021/es300550h>.
- (39) UniProt <https://www.uniprot.org/> (accessed May 4, 2020).
- (40) Multiplexed assay for protein quantitation in the invertebrate *Gammarus fossarum* by liquid chromatography coupled to tandem mass spectrometry - PubMed <https://pubmed.ncbi.nlm.nih.gov/28417180/> (accessed Aug 25, 2020).



### Annexe 1 : Organigramme INRAE

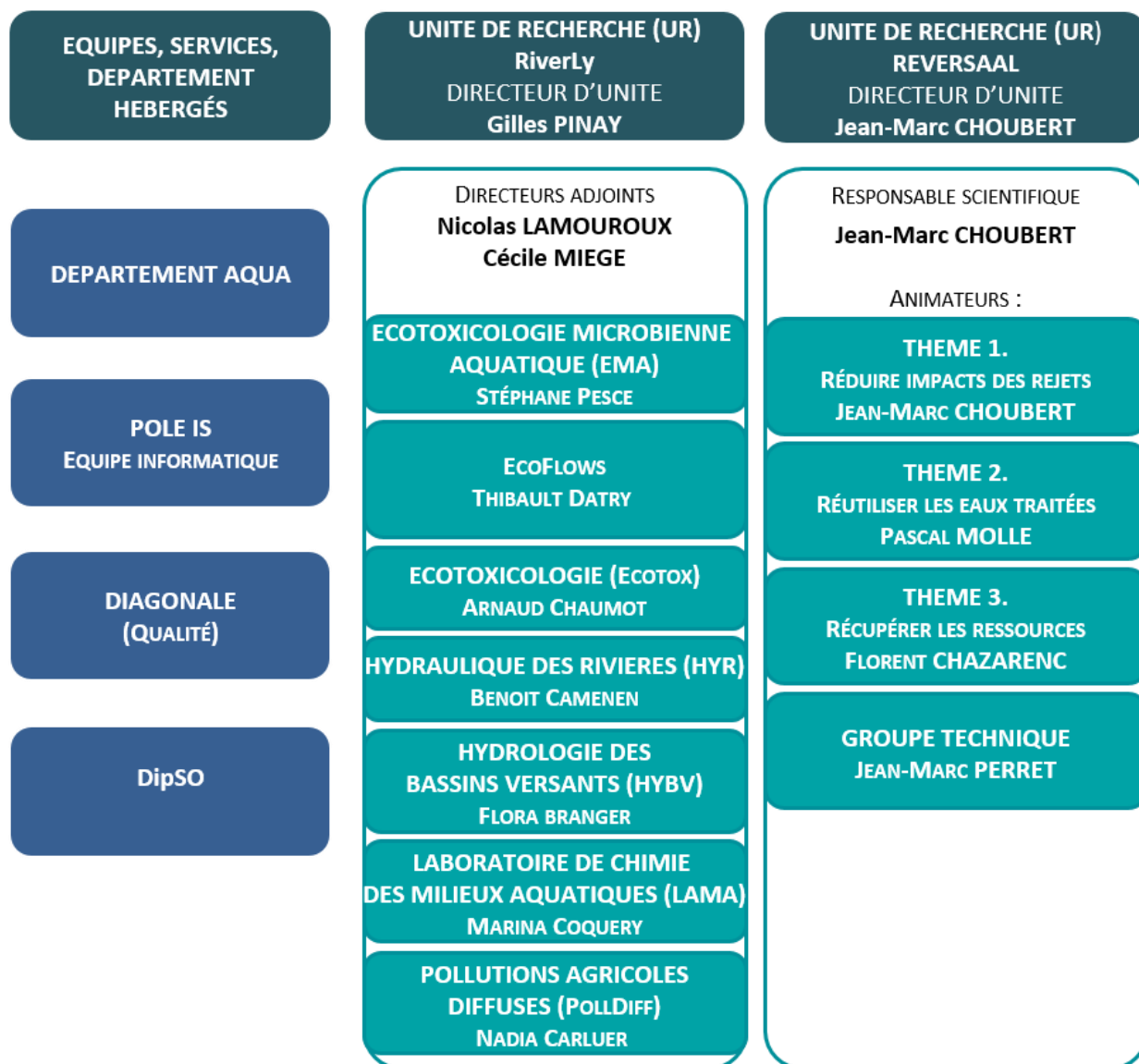


Figure 18 : Organigramme généralisé de INRAE site de Lyon-Villeurbanne extrait de la présentation d'accueil des stagiaires

## Annexe 2 : Liste complète des peptides potentiels

Tableau 10 : Liste complète des potentiels peptides. Peptides skyline = peptides retrouvés par digestion *in silico* via Skyline. Peptides conservés ajoutés manuellement sur Skyline = peptides identifiés comme très conservés ainsi que leurs potentiels variants lors de l'étape d'alignement

Nom de la protéine	Peptides skyline	Peptides ajoutés manuellement
Myosine	MEIDDLSSNMEAVAK	QAFTQQIEDLK
	LQTENGEFSR	
	TLEDQLSEIK	
	NALAHAVQSAR	
	QLNDMNAQR	
	EALVSQLTR	
	SDENSR	
Myogénine	MPGGFDQGGYQER	HQEQPHCPGQCLPWACK
	GSMMLGLCGDGR	HPEP_HCPGQCLPWACK
	LLSGMVGGLEDK	HQEQQHCPGQCLPWACK
	SAIQYIER	HSE-PHCPGQCLPWACK
	LQALVSSLNQQDHDQTGGLHYR	NQEQEHCPGQCLPWACK
	AAQPPR	STLMNPNQR
	VSSSSDQSGSTK	STLTNPNQR
		SAIQYIEK
		SSMVGLCGDSR
	SSMMGLCGDGR	
Vitellogénine	AAEAVANADIDEITLLVK	AYLAGAATDVEFGLR
	AEAGLLGASPAAR	SYFAGAAADVLEFGVR
	ALAVLPAK	GILNILQLNFK
	ALQDLEDDR	
	AYLAGAAADVLEVGVR	
	CLPGCLPVK	
	CYSVEPVLIR	
	DPFVPATK	
	EINIIDELSPEGR	
	EELTVAIPPER	
	FEYTNQVVGK	
	FFGQEIAFANIDK	
	FIQLIQVLR	
	FLGNAVAPAVVVILR	
	GILNILQLNLK	
	GNAPDLAPGNVAIK	
	ILPGFGSAAASLPIR	
IPAIVTTYATR		
LALPLPIALPIGADAVIR		

	<b>LLGYQAAAYLDK</b>	
	LTSALAAQFLPIK	
	NIEDLAAER	
	<b>QEYSTPSGYQTK</b>	
	QMIVDGQESK	
	<b>QNQNVYELQEAGAQGVCK</b>	
	<b>SQPLGSIYVK</b>	
	SSGSSASSFEAIYK	
	<b>TEGIQEALLK</b>	
	TFYVISEDVK	
	<b>VASLENIEAIWSQFK</b>	
	VETLAFAR	
	<b>VLISALAENIYLLK</b>	
	VQADAILALR	
	<b>VQTIISALAETDNWK</b>	
Acétylcholinestérase	LPVPDR	YLAHSEK
	KPWNNVFEAK	<b>NQFDHYSK</b>
	YLAYTEK	
	VVVVSMNYR	
	<b>VGAFGFLALNGSSDAPGNVGLYDQR</b>	
	<b>LALQWVQENIHFFGGNPK</b>	
	<b>LVGCTWGNDELIDCLR</b>	
	<b>HPQELIDQEWQVLPWSSFR</b>	
	<b>DNESLISR</b>	
	EDFLESVK	
	<b>DAMDDIVGDQNVICPLQHFAK</b>	
	<b>YWANFAR</b>	
	<b>HVGLNTEPMK</b>	
	TQFCALWNR	
	<b>LLNITDNIDVER</b>	
	<b>WSSYMMHWK</b>	
	<b>SQFDHYSK</b>	
Métallothionéine	<b>CASGCVCK</b>	<b>SCCSCCPGCSK</b>
	<b>GNSCGSSCCQ</b>	<b>SCCPCCPSGCSK</b>
		<b>SCCSCCPSTCSK</b>
		<b>SCCTCCPSGCSK</b>
		<b>SCCSCCPTGCSK</b>
		<b>SCCPCCPSDCSK</b>
		<b>SCCACCPGCTK</b>
		TGTCNCGATCK
Cytochrome	FINDGK	<b>SLAFSTDQGVVWR</b>

	SLAFSTDQAGVWR	IVSEHYESYDK
	SFSTLEGENPAYSCVLEEHISK	IVREHYDTYDK
	EGLYLVEQLXTVMDDSGSFDPR	LVMEHYDTFDK
	FNTFVQR	IVSEHYATYNK
	<b>IVSEHYASYDK</b>	
	SVGMMGGGLR	
Phocéine	SDCSNIEK	IPEPPEGQDEGVWK
	IMEPPEGQDEGVWK	
	<b>QFCLEINGLAVK</b>	
	ECPAIDYTR	
	HTLDGAACLLNSNK	
	ESSVAK	
	<b>IFSHAYFHHR</b>	
	YENETFLCHR	
	YNLMSK	
	<b>DNLIVPILEEEVQSATAGESDA</b>	

## Annexe 3 : Résultats de l'optimisation des paramètres de fragmentation

Tableau 11 : Résultats de l'optimisation des paramètres de fragmentation. Ici il y a pour chaque peptide d'intérêt, les transitions (ion précurseur + ion fragment) les plus intenses ainsi que l'énergie de collision la plus optimale associée à chaque transition. En rouge, l'ion fragment qui forme avec l'ion précurseur associé la transition quantifiante.

Protéines	Peptides	Ions précurseurs	Ion fragments	Energie de collision optimale (eV)	
Myosine	MEIDDLSSNMEAVAK	830,88	1287,59	25	
			<b>944,45</b>	25	
			857,52	25	
	NALAHAVQSAR	574,31	383,21	<b>641,36</b>	25
				641,36	17
		343,19	13		
	EALVSQLTR	513,79		713,41	18
				<b>614,34</b>	14
				527,31	18
Myogénine	MPGGFDQGGYQER	484,56	719,33	13	
			<b>605,29</b>	11	
			442,22	13	
	HQEQPHCPGQCLPWACK	535,98		<b>669,32</b>	13
				386,19	21
				315,15	25
	STLTNPQR	520,77		739,37	13
				638,32	17
				<b>524,28</b>	13
Vitellogénine	LLGYQAAAYLDK	667,36	<b>1107,56</b>	20	
			887,47	20	
			270,15	28	
	SQPLGSIYVK	550,31		<b>884,53</b>	13
				674,39	21
				417,25	25
Acétylcholinestérase	YWANFAR	469,23	774,39	14	
			<b>588,31</b>	14	
			517,27	14	
	SQFDHYSK	510,23		<b>804,37</b>	17
				657,3	17
				542,28	21
Métallothionéine	SCCTCCPSGCSK	736,25	1224,43	25	
			803,32	25	
			<b>643,29</b>	25	
Cytochrome	IVSEHYASYDK	450,55	754,35	17	
			591,28	21	
			<b>270,15</b>	21	
	SLAFSTDQVGWVR	738,37		<b>1057,53</b>	23
				527,29	27
				371,2	23
Phcéine	QFCLEINGLAVK		964,59	27	
			<b>722,46</b>	23	
			609,38	23	