



**HAL**  
open science

# Développements analytiques pour l'analyse multi-omique en chromatographie liquide couplée à la spectrométrie de masse

Julien Faugere

► **To cite this version:**

Julien Faugere. Développements analytiques pour l'analyse multi-omique en chromatographie liquide couplée à la spectrométrie de masse. Chimie analytique. Université de Lyon, 2022. Français. NNT : 2022LYSE1044 . tel-04054983

**HAL Id: tel-04054983**

**<https://theses.hal.science/tel-04054983>**

Submitted on 1 Apr 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.



N° d'ordre NNT :2022LYSE1044

## **THESE de DOCTORAT DE L'UNIVERSITE DE LYON**

opérée au sein de  
**l'Université Claude Bernard Lyon 1**

**Ecole Doctorale N° 206**  
**(Ecole Doctorale de chimie de Lyon)**

**Spécialité de doctorat :**  
**Discipline :** Chimie analytique

Soutenue publiquement le 31/03/2022, par :  
**Julien Faugere**

---

# **Développements analytiques pour l'analyse multi-omique en chromatographie liquide couplée à la spectrométrie de masse**

---

Devant le jury composé de :

Dr. François BECHER, CEA Saclay  
Dr. Frédérique COURANT, Université de Montpellier  
Pr. Nathalie BERNOUD-HUBAC, INSA Lyon  
Pr. Claire DEMESMAY-GUILHIN, UCBL  
Dr. Jean ARMENGAUD, CEA Marcoule  
Dr. Arnaud CHAUMOT, INRAE Villeurbanne

Rapporteur  
Rapporteur  
Examinatrice  
Examinatrice  
Examinateur  
Invité

Pr. Arnaud SALVADOR, UCBL  
Dr. Sophie AYCIRIEX, UCBL

Directeur de thèse  
Co-directrice de thèse





# Université Claude Bernard – LYON 1

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Petru MIRONESCU
Directeur Général des Services	M. Pierre ROLLAND

## COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

## COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

**« Une personne qui n'a jamais commis d'erreurs n'a jamais tenté d'innover. »**

**Albert Einstein**

# Remerciements

Les premières personnes que je tiens à remercier sont mon directeur de thèse, Arnaud Salvador, ainsi que ma co-directrice Sophie Ayciriex. Ils ont su me transmettre l'envie de me perfectionner dans le domaine de la spectrométrie de masse au travers de leurs enseignements et leurs passions. Merci pour vos conseils, votre confiance et pour vos encouragements tout au long de cette thèse. J'ai beaucoup appris à vos côtés et j'espère sincèrement avoir l'occasion de retravailler avec vous.

Je suis également reconnaissant envers mon directeur d'équipe, Jérôme Lemoine, pour avoir cru en moi lors de cet entretien durant l'été 2017 et qui m'a permis d'intégrer cette superbe équipe qu'est Anabio-MS. Ces années passées au sein de l'ISA ont été mémorables. J'ai eu de la chance d'évoluer dans un laboratoire à la pointe de la technologie.

Merci à Sabine Heinisch, responsable de l'équipe chromatographie et techniques couplées de l'ISA, pour m'avoir donné accès au système 2D ainsi qu'à l'IM-QTOF. Ces conseils en chromatographie bidimensionnelle ont été très enrichissants. Je remercie l'ensemble de l'équipe Écotoxicologie de l'INRAe-Villeurbanne pour leurs expertises sur mon espèce sentinelle préférée : ma petite crevette de rivière *Gammarus fossarum*. Je remercie également Jean Armengaud, du laboratoire LI2D du CEA de Marcoule, pour avoir participé à ce projet de thèse à travers son savoir précieux sur les approches en protéogénomique.

Je remercie également mes deux rapporteurs, François Becher et Frédérique Courant ainsi que mes examinateurs Nathalie Bernoud-Hubac, Claire Demesmay-Guilhin, Jean Armengaud, Olivier Berdeaux pour l'évaluation de mes travaux de thèse.

Ces années passées à l'ISA n'auraient pas été les mêmes sans le soutien, la gentillesse, l'humour de toutes les personnes que j'ai pu rencontrer dans cet institut, alors à tous ceux qui ont été là à un moment ou à un autre, merci !

Parmi ces personnes, certaines méritent une attention particulière ...

Merci à Romain, « responsable de plateforme », pour son enseignement du Scout-MRM à l'ancienne, pour sa bonne humeur et nos fous rires ! Après ces quatre ans à le côtoyer j'ai appris que la persévérance est un atout indéniable dans une équipe. Surtout lorsqu'il pense que le PSG peut un jour gagner la LDC (Paris est « presque » magique). Ta tarte au citron meringuée et nos confrontations au badminton vont me manquer (promis un jour je gagnerai un set !).

Merci à tous les doctorants, collègues et amis que j'ai eu la chance de côtoyer pendant ma thèse. Merci, Nour pour tes enseignements, sans toi la thèse ne serait pas une expérience de vie. Merci Nico

pour ces bons moments passés ensemble à discuter du monde geek. Merci à Aurélien et mon Viking préféré JV de m'avoir fait aimer le comique de répétitions et les jeux de mots. Merci à Éva d'avoir partagé avec moi sa passion pour les jeux de société. Merci à Iulia d'avoir été m'a plus grande supportrice et de m'avoir poussé à faire connaître la spéciale Faugere dans le monde entier. Merci à Valérian, Maxime et Delphine de m'avoir accompagné et soutenu dans ma quête d'éradication de l'appellation de la viennoiserie dont il ne faut pas prononcer le nom. Vive la dream team chocolatine ! Merci Houdini de m'en avoir fait voir de toutes les couleurs et m'avoir appris que par moment la chimie c'est de la magie. Un grand merci à mon padawan Thomas sans qui la métabolomique ne serait pas dans ce manuscrit de thèse. Je ne me fais pas de soucis pour le passage de flambeau, car je suis certain que tu vas devenir le plus grand maître Jedi des analyses multi-omiques. Bon courage pour la suite !

Merci à mes ingénieurs préférés pour les moments passés ensemble. Francis, je ne cesserai de te le dire, mais tu as un look d'enfer ! Dire qu'on aurait presque pu faire partie de la même famille. Merci, Maud d'avoir été une si bonne partenaire de tarot, promis un jour j'arrêterai d'être mauvais perdant. Merci Anabelle pour nos moments karaoké au milieu du labo sur des musiques plus que bof, mais tellement bien chantées.

Merci aux personnes côtoyées au quotidien, Julie, Frédéric, Michel, Marion, Fabien, Vincent, Florence, David, Sylvain, Lucile, Laura et l'ensemble de l'équipe Anaquant ainsi que tous ceux que j'oublie.

D'un point de vue plus personnel, il me reste encore plusieurs personnes à remercier...

Un grand merci à mes parents et ma bonne étoile qui m'ont toujours soutenu et accompagné dans les moments de joie, comme dans les moments difficiles. Grâce à vous, j'ai eu la chance de réaliser les études que je souhaitais faire. Merci à ma petite sœur Mélissa, je suis sûr que toi aussi tu vas devenir une crack de la chimie. Comme quoi la chimie c'est de famille. Pour terminer, je remercie Valérie qui m'a encouragé pendant cette thèse, en plus d'avoir la patience de me supporter au quotidien. Elle a toujours été d'un soutien indéfectible et a contribué pour beaucoup dans la rédaction de ce manuscrit de thèse.

## Résumé de la thèse

L'analyse de biomolécules en santé ou en environnement nécessite l'utilisation d'outils analytiques de grande précision ainsi que des méthodes d'acquisition de données adaptées. De nombreux verrous scientifiques peuvent être rencontrés comme le suivi limité d'un grand nombre de molécules en spectrométrie de masse ciblée, les variations d'ionisation ou bien l'analyse simultanée de biomolécules de propriétés physico-chimiques différentes. Les enjeux de cette thèse ont été d'optimiser et d'améliorer l'acquisition des données dans les sciences omiques comme la protéomique, la lipidomique et la métabolomique dans un contexte environnemental. Le premier axe de recherche a été d'améliorer les capacités de multiplexage *via* un nouveau mode d'acquisition en spectrométrie de masse ciblée, appelé Scout-MRM tout en permettant de s'affranchir du temps de rétention. Ce développement innovant a été appliqué pour l'étude de *G. fossarum*, espèce sentinelle en écotoxicologie, *via* le suivi en protéomique ciblée d'un grand multiplex constitué d'un panel de 157 protéines clés. Afin de faciliter le développement de méthodes ciblées en lipidomique et dans une approche de type « plug-and-play », Scout-MRM a été employé sur différentes matrices biologiques. Le deuxième axe de recherche a été de développer un protocole biphasique pour l'extraction simultanée des protéines, des lipides et des métabolites d'un même échantillon. Cette approche multi-omique a permis de mettre en exergue des signatures moléculaires au cours du développement chez les gammarides femelles. Le troisième axe de recherche a été d'explorer les potentialités de la chromatographie bidimensionnelle couplée à la spectrométrie de masse de haute résolution ainsi que différentes configurations d'orthogonalité pour l'analyse des lipides.

**Mots clés** : Spectrométrie de masse ; Multi-omique ; Protéomique ; Lipidomique ; Métabolomique ; Scout-MRM ; Multiplexage ; 2D-LC.

## Thesis abstract

The analysis of biomolecules in health or in environment requires the use of high precision analytical instruments and adapted data acquisition methods. Several scientific challenges can be met such as the limited monitoring of a large number of molecules in targeted mass spectrometry, ionization variations or the simultaneous analysis of biomolecules with different physico-chemical properties. The main objective of this thesis was to optimize and improve data acquisition in the omics field such as proteomics, lipidomics and metabolomics in an environmental context. The first research axis was to improve the multiplexing capabilities *via* a new acquisition mode called Scout-MRM, in targeted mass spectrometry, free from retention time. This innovative development has been applied to the study of *G. fossarum*, a sentinel species widely used in ecotoxicology, *via* the monitoring in targeted proteomics of a large multiplex composed of a panel of 157 key proteins. To facilitate the development of targeted methods in lipidomics and in a "plug-and-play" philosophy, Scout-MRM was used on different biological matrices. The second research axis was to develop a biphasic protocol for the simultaneous extraction of proteins, lipids, and metabolites from the same sample. This multi-omics approach highlighted molecular signatures during the development cycle of female gammarids. The third research axis was to explore the capabilities of two-dimensional chromatography coupled with high resolution mass spectrometry as well as different orthogonality configurations for lipid analysis.

**Keywords :** Mass spectrometry ; Multi-omic ; Proteomic ; Lipidomic ; Metabolomic ; Scout-MRM ; Multiplex approach ; 2D-LC.

# Valorisations des travaux de thèse

## 1. Publications scientifiques

- Faugere J., Gouveia D., Ayciriex S., Chaumot A., Almunia C., François A., Armengaud J., Lemoine J., Geffard O., Degli-Esposti D., Salvador A., 2020. **High-multiplexed monitoring of protein biomarkers in the sentinel *Gammarus fossarum* by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics.** Journal of proteomics 226, 103901.  
DOI: [10.1016/j.jprot.2020.103901](https://doi.org/10.1016/j.jprot.2020.103901)
- Leprêtre M., Palos-Ladeiro M., Faugere J., Almunia C., Lemoine J., Armengaud J., Geffard A., Salvador A., 2020. **From shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential immunomarkers in *Dreissena polymorpha*.** Analytical and Bioanalytical Chemistry 412, 7333–7347.  
DOI: [10.1007/s00216-020-02868-2](https://doi.org/10.1007/s00216-020-02868-2)
- Faugere J., Brunet T., Clément Y., Espeyte A., Geffard O., Lemoine J., Chaumot A., Degli-Esposti D., Ayciriex S., Salvador A. **Development of a multi-omics extraction method for ecotoxicology: in depth investigation of the reproductive cycle of *Gammarus fossarum*.** Soumis dans le Journal Talanta.
- Faugere J., Lemoine J., Ayciriex S., Salvador A. **Toward a targeted method for a rapid lipid profiling of complex lipidome with Scout-MRM.** En préparation.

## 2. Communications orales

- Faugere J., Novembre 2020. **Développement d'approches multi-omiques par LC-MS/MS ciblées chez l'espèce sentinelle *G. fossarum*.** GDR d'écotoxicologie aquatique, Bordeaux.
- Faugere J., Mars 2021. **Développement d'une nouvelle méthode ciblée et hautement multiplexée pour un profilage rapide de lipidomes complexes par Scout-MRM.** Rencontre du club jeune de la Société Française de Spectrométrie de Masse 2021, en ligne.
- Faugere J., Juin 2021. **Développement d'une nouvelle méthode ciblée et hautement multiplexée pour un profilage rapide de lipidomes complexes par Scout-MRM.** Journée Française de spectrométrie de masse 2021, en ligne.

## 3. Communication sous la forme de poster

- Faugere J., Mars 2019. **Développement d'approches multi-omiques par LC-MS/MS ciblée chez l'espèce sentinelle *G. fossarum*.** 13ème Congrès de l'Association Francophone des Sciences Séparatives, Paris.



# Liste des abréviations

## Signification anglaise

<b>1D-LC</b>	:	● One-dimensional liquid chromatography
<b>2D-LC</b>	:	● Two-dimensional liquid chromatography
<b>ACN</b>	:	● Acetonitrile
<b>ACP ou PCA</b>	:	● Principal component analysis
<b>ADN ou DNA</b>	:	● Deoxyribonucleic acid
<b>AFM ou FMA</b>	:	● Multiple Factor Analysis
<b>Ag-LC</b>	:	● Chromatography with silver columns
<b>AMBIC</b>	:	● Amonium bicarbonate
<b>ARN ou RNA</b>	:	● Ribonucleic acid
<b>ARNm ou RNAm</b>	:	● Messenger ribonucleic acid
<b>ATP</b>	:	● Adenosine triphosphate
<b>CE</b>	:	● Collision energy
<b>Cer</b>	:	● Ceramide
<b>CV ou RSD</b>	:	● Relative standard deviation
<b>D1</b>	:	● First dimension
<b>D2</b>	:	● Second dimension
<b>Da</b>	:	● Dalton
<b>DCM</b>	:	● Dichloromethane
<b>DDA</b>	:	● Data Dependant Acquisition
<b>DTT</b>	:	● Dithiothreitol
<b>DG</b>	:	● Diacylglycerol or diglyceride
<b>DI/MS</b>	:	● Direct infusion
<b>DIA</b>	:	● Data Independent Acquisition
<b>DP</b>	:	● Declustering potential
<b>DT</b>	:	● Dwell time
<b>ECN</b>	:	● Equivalent carbon number
<b>ESI(+) ou ESI(-)</b>	:	● Positive or negative electrospray ionisation
<b>EtOH</b>	:	● Ethanol
<b>FA ou AG</b>	:	● Fatty acid

## Signification française

● Chromatographie liquide unidimensionnelle
● Chromatographie liquide bidimensionnelle
● Acétonitrile
● Analyse en composantes principales
● Acide désoxyribonucléique
● Analyse factorielle multiple
● Chromatographie avec des colonnes en argent
● Bicarbonate d'ammonium
● Acide ribonucléique
● Acide ribonucléique messager
● Adénosine triphosphate
● Energie de collision
● Céramide
● Coefficient de variation
● Première dimension
● Deuxième dimension
● Dalton
● Dichlorométhane
● Acquisition dépendante des données
● Dithiothréitol
● Diacylglycérol ou diglycéride
● Infusion directe
● Acquisition indépendante des données
● Potentiel de déclusterisation
● Dwell time
● Nombre équivalent de carbone
● Ionisation électrospray positive ou négative
● Ethanol
● Acide gras

### Signification anglaise

<b>FDA</b>	:	● Food and Drug Administration
<b>FDR</b>	:	● False Discovery Rate
<b>FIA</b>	:	● Flow Injection Analysis
<b>GAPDH</b>	:	● Glyceraldehyde 3-phosphate dehydrogenase
<b>GC</b>	:	● Gas chromatography
<b>GC-FID</b>	:	● Gas chromatography coupled to a flame ionization detector
<b>HILIC</b>	:	● Hydrophilic Interaction Chromatography
<b>HPLC</b>	:	● High performance liquid chromatography
<b>HRMS</b>	:	● High resolution mass spectrometry
<b>IAM</b>	:	● Iodoacetamide
<b>IEC</b>	:	● Ion exchange chromatography
<b>IM</b>	:	● Ion mobility
<b>IPA</b>	:	● Isopropanol
<b>ISTD</b>	:	● Internal standard
<b>JH</b>	:	● Juvenile hormone
<b>LC</b>	:	● Liquid chromatography
<b>LC-LC</b>	:	● Multiple heart cut liquid chromatography
<b>LCxLC</b>	:	● Comprehensive liquid chromatography
<b>LLE</b>	:	● Liquid-liquid extraction
<b>LOD</b>	:	● Limit of detection
<b>LOQ</b>	:	● Limit of quantification
<b>LPA</b>	:	● Lysophosphatidic acid
<b>LPC</b>	:	● Lysophosphocholines
<b>LPE</b>	:	● Lysophosphoethanolamine
<b>LPG</b>	:	● Lysophosphatidylglycerol
<b>LPI</b>	:	● Lysophosphatidylinositol
<b>LPL</b>	:	● Lysophospholipid
<b>m/z</b>	:	● Mass on charge
<b>MAE</b>	:	● Microwave assisted extraction

### Signification française

● Agence des produits alimentaires et médicamenteux
● Taux de fausses découvertes
● Analyse à flux continu
● Glycéraldéhyde 3-phosphate déshydrogénase
● Chromatographie en phase gazeuse
● Chromatographie en phase gaz couplée à un détecteur à ionisation de flamme
● Chromatographie par interaction hydrophile
● Chromatographie liquide haute performance
● Spectrométrie de masse de haute résolution
● Iodoacétamide
● Chromatographie d'échange d'ion
● Mobilité ionique
● Isopropanol
● Standard interne
● Hormone juvénile
● Chromatographie en phase liquide
● Chromatographie liquide de coupe à cœur multiple
● Chromatographie liquide complète
● Extraction liquide-liquide
● Limite de détection
● Limite de quantification
● Acide lysophosphatidique
● Lysophosphocholines
● Lysophosphoéthanolamine
● Lysophosphatidylglycérol
● Lysophosphatidylinositol
● Lysophospholipide
● Masse sur charge
● Extraction assistée par micro-ondes

### Signification anglaise

<b>MALDI</b>	:	● Matrix Assisted Laser Desorption Ionization
<b>ME</b>	:	● Malic enzyme
<b>MeOH</b>	:	● Methanol
<b>MG</b>	:	● Monoacylglycerol or monoglyceride
<b>m-NBA</b>	:	● Meta-nitrobenzyl alcohol
<b>MRM</b>	:	● Multiple reaction monitoring
<b>MRM<sup>3</sup></b>	:	● Multiple reaction monitoring cubed
<b>MRM-HR</b>	:	● High resolution multiple reaction monitoring
<b>MS</b>	:	● Mass spectrometry
<b>MS/MS</b>	:	● Tandem mass spectrometry
<b>MS<sup>1</sup></b>	:	● Precursor ion spectrum
<b>MS<sup>2</sup></b>	:	● Fragment ion spectrum
<b>MTBE</b>	:	● Methyl tert-butyl ether
<b>NPLC</b>	:	● Normal phase liquid chromatography
<b>PA</b>	:	● Phosphatidic acid
<b>PC</b>	:	● Phosphatidylcholine
<b>PCB</b>	:	● Polychlorinated biphenyl
<b>PCs</b>	:	● Principal components
<b>PE</b>	:	● Phosphatidylethanolamine
<b>PG</b>	:	● Phosphatidylglycerol
<b>PI</b>	:	● Phosphatidylinositol
<b>PL</b>	:	● Phospholipid
<b>PLS-DA</b>	:	● Partial least squares discriminant analysis
<b>PRM</b>	:	● Parallel Reaction Monitoring
<b>PS</b>	:	● Phosphatidylserine
<b>PSMs</b>	:	● Peptide-to-spectrum matches
<b>Q</b>	:	● Quadripole
<b>Q/LIT</b>	:	● Hybrid quadrupole / linear ion trap
<b>QQQ ou QqQ</b>	:	● Triple quadrupole
<b>QTrap</b>	:	● Quadripole Ion Trap
<b>R</b>	:	● Resolution

### Signification française

● Ionisation par désorption laser assistée par matrice
● Enzyme malique
● Méthanol
● Monoacylglycérol ou monoglycéride
● Alcool méta-nitrobenzylique
● Multiple reaction monitoring
● Multiple reaction monitoring cube
● Multiple reaction monitoring haute résolution
● Spectrométrie de masse
● Spectrométrie de masse en tandem
● Spectre d'ion précurseur
● Spectres d'ion fragment
● Méthyl tert-butyl éther
● Chromatographie liquide de phase normale
● Acide phosphatidique
● Phosphatidylcholine
● Polychlorobiphényle
● Composantes principales
● Phosphatidyléthanolamine
● Phosphatidylglycérol
● Phosphatidylinositol
● Phospholipide
● Analyse discriminante des moindres carrés partiels
● Parallel Reaction Monitoring
● Phosphatidylsérine
● Correspondances peptide-spectre
● Quadripole
● Quadripôle hybride / trappe d'ion linéaire
● Triple quadrupole
● Piège à ions à quadripolaire
● Résolution

### Signification anglaise

<b>R<sup>2</sup></b>	:	● R-squared
<b>RMN</b>	:	● Nuclear magnetic resonance
<b>RPLC</b>	:	● Reverse phase liquid chromatography
<b>RT</b>	:	● Retention time
<b>S/B ou S/N</b>	:	● Signal to noise
<b>SDS-PAGE</b>	:	● Electrophoresis in polyacrylamide gel containing sodium dodecylsulfate
<b>SEC</b>	:	● Size exclusion chromatography
<b>SIMS</b>	:	● Secondary-ion mass spectrometry
<b>SLE</b>	:	● Solid-liquid extraction
<b>SM</b>	:	● Sphingomyelin
<b>sMRM</b>	:	● Scheduled-MRM
<b>sn</b>	:	● Stereospecific numbering
<b>SPE</b>	:	● Solid-phase extraction
<b>sPRM</b>	:	● Scheduled-Parallel Reaction Monitoring
<b>SRM</b>	:	● Single Reaction Monitoring
<b>SWATH</b>	:	● Sequential Windowed Acquisition of All Theoretical fragment ions
<b>TFA</b>	:	● Trifluoroacetic acid
<b>TG</b>	:	● Triacylglycerol or triglyceride
<b>TOF</b>	:	● Time of flight
<b>UV</b>	:	● Ultraviolet
<b>V<sub>0</sub></b>	:	● Dead volume
<b>VIP</b>	:	● Variable Importance in Projection
<b>VTG</b>	:	● Vitellogenin

### Signification française

● Coefficient de détermination
● Résonance magnétique nucléaire
● Chromatographie liquide de phase inverse
● Temps de rétention
● Signal sur bruit
● Electrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium
● Chromatographie d'exclusion stérique
● Spectrométrie de masse des ions secondaires
● Extraction solide liquide
● Sphingomyéline
● Scheduled-MRM
● Numérotation stéréospécifique
● Extraction en phase solide
● Scheduled-Parallel Reaction Monitoring
● Single Reaction Monitoring
● Acquisition séquentielle fenêtrée de tous les ions fragments théorétiques
● Acide trifluoroacétique
● Triacylglycérol ou triglycéride
● Temps de vol
● Ultraviolet
● Volume mort
● Importance de la variable dans la projection
● Vitellogénine

# Table des matières

<b>Introduction générale</b> .....	<b>1</b>
<b>Chapitre 1 : Étude bibliographique</b> .....	<b>5</b>
1. Les sciences omiques combinées ou approches multi-omiques .....	5
1.1. La génomique .....	5
1.2. Les études phénotypiques .....	6
1.2.1. Transcriptomique .....	6
1.2.2. Protéomique .....	6
1.2.3. La Métabolomique .....	8
1.3. Les approches combinées .....	9
1.4. Approches multi-omiques : émergence du concept et application .....	11
2. La spectrométrie de masse, un outil de choix pour les approches « omiques » .....	13
2.1. Les approches sans <i>a priori</i> .....	14
2.1.1. Le mode d'acquisition DDA .....	14
2.1.2. Le mode d'acquisition DIA .....	15
2.2. Les approches avec <i>a priori</i> .....	17
2.2.1. Le mode d'acquisition <i>Multiple Reaction Monitoring</i> (MRM) .....	17
2.2.2. Le mode d'acquisition <i>Parallele Reaction Monitoring</i> (PRM) .....	18
2.3. Quelles sont les limites des méthodes MRM en termes de capacité de multiplexage et comment les améliorer ? .....	19
2.3.1. Évaluation des capacités de multiplexage en mode MRM .....	19
2.3.2. Capacité de multiplexage du mode MRM versus PRM .....	21
2.3.3. Mise en place de méthodes avec fenêtres de temps (Scheduled-MRM) .....	22
2.3.4. Un nouveau mode d'acquisition : Scout-MRM .....	23
2.3.5. Quels sont les plus grands multiplexes en approche ciblée dans les différentes sciences omiques ? ..	24
2.4. Analyse par infusion directe .....	25
3. Diminution de la complexité d'un échantillon grâce à une dimension de séparation en amont du spectromètre de masse .....	26
3.1. Techniques séparatives en métabolomique polaire .....	26
3.2. Techniques séparatives en lipidomique .....	28
3.3. Techniques séparatives en protéomique .....	29
4. Mise en place d'une méthode chromatographique avancée : la chromatographie liquide bidimensionnelle .....	32
4.1. Les différents modes 2D .....	34
4.1.1. Hors ligne .....	34
4.1.2. En ligne .....	34
4.1.3. Comparaison des 3 modes .....	37
4.2. Problèmes rencontrés en 2D-LC .....	37
4.2.1. Les effets à l'injection .....	37

4.2.2. Couplage avec la spectrométrie de masse et les sources ESI .....	40
4.2.3. Traitement des données .....	41
4.3. La 2D-LC dans le cadre des analyses lipidomiques .....	43
4.3.1. Les approches hors ligne pour l'analyse des lipides .....	45
4.3.2. Les approches en ligne pour l'analyse des lipides .....	45
5. De la préparation omique individualisée à la préparation multi-omique à partir d'un échantillon unique .....	49
5.1. Préparation d'échantillon en protéomique .....	49
5.1.1. Extraction des protéines .....	49
5.1.2. Digestion des protéines lors de l'approche Bottom-up .....	50
5.2. Préparation d'échantillon en lipidomique .....	50
5.3. Préparation d'échantillon en métabolomique .....	51
5.4. Protocole d'extraction multi-omique à partir d'un échantillon unique.....	53
5.4.1. Protocole SIMPLEX et MPLEEx.....	53
5.4.2. Protocole à double extractions.....	56
5.4.3. Protocole « Single-Step extraction » avec séparation chromatographique en HILIC .....	60
5.4.4. Protocole de Vorreiter .....	63
6. Les sciences omiques en écotoxicologie .....	65
6.1. Écotoxicologie et biomarqueur .....	65
6.1.1. Qu'est-ce que l'écotoxicologie ?.....	65
6.1.2. Biomarqueur : outil de prédilection en écotoxicologie .....	66
6.2. Utilisation d'organismes sentinelles en écotoxicologie.....	67
6.2.1. Qu'est-ce qu'une espèce sentinelle ? .....	67
6.2.2. Les espèces sentinelles dans les écosystèmes aquatiques .....	67
6.2.3. L'espèce sentinelle <i>Gammarus fossarum</i> .....	69
6.2.4. Cycle reproducteur et cycle de mue chez l'espèce <i>G. fossarum</i> femelle.....	70
6.3. Objectifs et enjeux des omiques pour l'écotoxicologie.....	72
6.3.1. Apport de la protéomique pour l'écotoxicologie.....	72
6.3.2. La protéogénomique : outil de prédilection pour les organismes non modèle.....	75
6.3.3. Apport de la lipidomique pour l'écotoxicologie.....	77
6.3.4. Apport de la métabolomique pour l'écotoxicologie .....	79
<b>Chapitre 2 : Scout-MRM, un nouveau mode d'acquisition pour augmenter les capacités de multiplexage en écotoxicoprotéomique.....</b>	<b>81</b>
1. Contexte de l'étude.....	81
2. Résumé de la méthodologie pour l'élaboration de la méthode Scout-MRM chez l'espèce sentinelle <i>G. fossarum</i> .....	83
3. Publication n°1 : High-multiplexed monitoring of protein biomarkers in the sentinel <i>G.fossarum</i> by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics .....	85
3.1. Abstract .....	85
3.2. Biological significance .....	86

3.3. Introduction.....	86
3.4. Materials and Methods .....	88
3.4.1. Reagents and chemicals.....	88
3.4.2. Collection and maintenance of <i>G. fassarum</i> organisms .....	88
3.4.3. Sample preparation for shotgun proteomics.....	88
3.4.4. Sample preparation for Scout-MRM assay .....	89
3.4.5. Standards (SIL peptides, scout peptides) preparation .....	89
3.4.6. Nano LC-MS/MS analysis on high resolution mass spectrometry .....	89
3.4.7. Liquid chromatography and targeted mass spectrometry .....	90
3.4.8. MS/MS spectra interpretation from GFOSS database .....	90
3.4.9. Candidate list for Scout-MRM assay development.....	91
3.4.10. Method validation for Scout-MRM assay .....	91
3.4.11. Data analysis .....	92
3.4.12. Application of the assay.....	92
3.5. Results and discussion .....	93
3.5.1. Selection of proteins of interest from proteogenomics and DDA analysis .....	93
3.5.2. Reporter peptide identification and MS optimization in targeted MS approach .....	93
3.5.3. Improvement of multiplexing capacity .....	95
3.5.4. Scout-MRM method optimisation .....	98
3.5.5. Analytical performance evaluation of the assay for quantification .....	100
3.5.6. Application of Scout-MRM assay : comparative proteomic analysis using active biomonitoring in an agriculture watershed.....	103
3.6. Conclusion .....	106
3.7. Supplementary material of the first publication .....	106
3.8. Application de la méthode multiplexée pour l'analyse de cohortes d'échantillons dans le cadre d'études écotoxicologiques.....	107
3.8.1. Transfert de la méthode de dosage protéique appliquée à <i>G. fassarum</i> entre le laboratoire Anabio-MS et l'INRAE.....	107
3.8.2. Mise en évidence d'effets de lots lors d'études comparatives avec des approches protéomiques....	108
3.9. Méthode de transfert rapide entre les approches <i>shotgun</i> et ciblée via l'utilisation du Scout-MRM .....	110
4. Conclusion générale du chapitre.....	112
<b>Chapitre 3 : Profilage rapide de lipidome par Scout-MRM et modélisation des temps de rétention .....</b>	<b>114</b>
1. Contexte de l'étude.....	114
2. Matériels et méthodes.....	115
2.1. Réactifs et produits chimiques .....	115
2.2. Préparation des composés scouts .....	115
2.3. Préparation des échantillons .....	116
2.4. Analyse lipidomique par LC-MS/MS .....	117
2.5. Comparaison des modes Scheduled-MRM versus Scout-MRM .....	118

2.5.1. Variation du gradient chromatographique .....	118
2.5.2. Modification de la composition de la phase mobile B .....	118
2.5.3. Étude du <i>dwell time</i> et du nombre de points par pic chromatographique en fonction de la répétabilité de mesure .....	119
2.6. Balayage de perte de neutre pour l'analyse des triacylglycérols.....	119
2.7. Validation de la méthode Scout-MRM .....	120
2.8. Acquisition et retraitement des données .....	121
3. Résultats.....	121
3.1. Nomenclature et profils de fragmentation des lipides dans la base de données associée à la méthode Scout-MRM.....	121
3.1.1. Nomenclature des lipides .....	121
3.1.2. Fragmentation des lipides .....	123
3.2. Optimisation du gradient chromatographique.....	126
3.2.1. Choix de la phase stationnaire et des phases mobiles .....	126
3.2.2. Validation du gradient chromatographique .....	127
3.3. Construction de la méthode Scout-MRM .....	131
3.3.1. Critères de sélection d'un scout idéal.....	132
3.3.2. Choix des composés scouts en lipidomique .....	133
3.3.3. Attribution des composés aux différents groupes de scouts.....	134
3.3.4. Quelles sont les limites de la capacité de multiplexage en Scout-MRM ?.....	135
3.4. Règles de sélection des transitions MRM à ajouter dans la méthode Scout-MRM.....	137
3.4.1. Choix des transitions pour les lysophospholipides et les sphingomyélines.....	138
3.4.2. Choix des transitions pour les phospholipides, les céramides et les glycérolipides autres que les triacylglycérols.....	138
3.4.3. Choix des transitions pour les triacylglycérols.....	139
3.5. Construction d'une base de données à l'aide des modèles de rétention.....	142
3.5.1. Modèle du nombre de carbones équivalents (ECN) .....	142
3.5.2. Développement de la base de données.....	143
3.5.3. Incertitude associée à la séparation d'isomères.....	144
3.5.4. Incertitude en dehors des modèles de prédiction.....	145
3.6. Méthodologie pour améliorer la base de données .....	146
3.6.1. 1 <sup>ère</sup> méthodologie : Amélioration de la base de données en temps de rétention à partir de conditions LC imposées .....	147
3.6.2. 2 <sup>ème</sup> méthodologie : Amélioration de la base de données en temps de rétention à partir de conditions LC variables.....	149
3.7. Transférabilité du Scout-MRM .....	151
3.7.1. Indépendance des temps de rétention de la méthodologie Scout-MRM.....	151
3.7.2. Changement de phase stationnaire et potentiels changements de sélectivité en Scout-MRM.....	152
3.8. Évaluation du Scout-MRM en tant que méthode de caractérisation et de quantification simultanée pour l'analyse de lipides.....	156



3.8.1. Répétabilité de la méthode Scout-MRM .....	156
3.8.2. Domaine de linéarité et LOQ de la méthode Scout-MRM .....	157
3.9. Conclusion .....	159
<b>Chapitre 4 : Développement d'une méthode d'extraction multi-omique en écotoxicologie .....</b>	<b>161</b>
1. Contexte de l'étude.....	161
2. Publication n°2 .....	163
2.1. Abstract .....	163
2.2. Introduction.....	164
2.3. Materials and methods.....	166
2.3.1. Reagents and chemicals.....	166
2.3.2. Biological sample collection.....	167
2.3.3. Evaluation of homogenization conditions .....	167
2.3.4. Reference protocol for proteomics sample preparation .....	168
2.3.5. Evaluation of protein fraction after dual protein precipitation procedure .....	168
2.3.6. Reference protocol for metabolomics sample preparation .....	168
2.3.7. Recovery and repeatability evaluation during multi-omics extraction procedure .....	169
2.3.8. Multi-omics extraction procedure for the biological application .....	169
2.3.9. Liquid chromatography - Multiple Reaction Monitoring- mass spectrometry (LC-MRM-MS).....	170
2.3.10. Data analysis .....	171
2.4. Results and discussion .....	172
2.4.1. Comparison of protein extraction methods .....	172
2.4.2. Evaluation of homogenization condition on lipids .....	175
2.4.3. Evaluation of the metabolite fraction.....	176
2.4.4. Comparison of multi-omics extraction method versus single omics .....	177
2.4.5. Repeatability.....	180
2.4.6. Multi-omics method application.....	181
3. Conclusion .....	189
<b>Chapitre 5 : Apport de la chromatographie bidimensionnelle pour la lipidomique .....</b>	<b>190</b>
1. Pourquoi développer des méthodes bidimensionnelles en lipidomique ? .....	190
2. Matériels et méthodes.....	193
2.1. Réactifs et produits chimiques .....	193
2.2. Collecte des échantillons .....	193
2.3. Extraction des lipides.....	194
2.3.1. Broyage et extraction liquide-liquide .....	194
2.3.2. Évaporation et dilution avant injection .....	194
2.4. Étude des effets à l'injection en RPLC-HILIC hors ligne .....	194
2.5. Utilisation du collecteur de fractions pour les analyses HILIC-RPLC en mode hors ligne .....	195
2.6. Paramètres de masse des analyses réalisées sur le système Qtrap 5500 .....	196

2.7. 2D-LC <i>comprehensive</i> et spectrométrie de masse (QTOF) .....	196
2.8. 2D-LC <i>multiple heart cut</i> en ligne et spectrométrie de masse (QTOF) .....	198
2.9. Acquisition et retraitement des données 2D.....	199
3. Résultats et discussion .....	200
3.1. Avantage de la chromatographie HILIC en tant que 1 <sup>ère</sup> dimension lors des analyses 2D-LC pour la lipidomique .....	200
3.2. Choix de la phase stationnaire en 1 <sup>ère</sup> dimension et 2 <sup>ème</sup> dimension .....	201
3.3. Étude des effets à l'injection et mise en place du <i>split</i> .....	203
3.3.1. RPLC en 1 <sup>ère</sup> dimension et HILIC en 2 <sup>ème</sup> dimension.....	204
3.3.2. HILIC en 1 <sup>ère</sup> dimension et RPLC en 2 <sup>ème</sup> dimension.....	206
3.4. Étude du mode <i>comprehensive</i> pour des applications lipidomiques selon la configuration HILICxRPLC-MS/MS .....	207
3.4.1. Présentation du système envisagé .....	207
3.4.2. Mise en évidence d'un manque de résolution en 2 <sup>ème</sup> dimension lors de la configuration HILICxRPLC .....	209
3.5. Problématiques rencontrées lors de l'analyse des lysophospholipides en chromatographie HILIC de 1 <sup>ère</sup> dimension.....	210
3.6. Optimisation de la méthode <i>multiple heart cut</i> HILIC-RPLC-MS/MS.....	212
3.6.1. Présentation du système envisagé .....	212
3.6.2. Mise au point du gradient HILIC en 1 <sup>ère</sup> dimension .....	213
3.6.3. Mise au point du gradient RPLC en 2 <sup>ème</sup> dimension.....	218
3.6.4. Optimisation des paramètres de masse du QTOF .....	220
3.7. Application de la méthode LC-LC en ligne .....	223
3.8. Bilan de l'approche LC-LC en lipidomique comparée à l'approche LCxLC .....	224
3.8.1. Inconvénients du mode LC-LC vis-à-vis du mode LCxLC .....	224
3.8.2. Avantage du mode LC-LC en lipidomique .....	226
4. Conclusion et perspectives.....	228
<b>Conclusion générale.....</b>	<b>230</b>
<b>Annexes .....</b>	<b>234</b>
<b>Références bibliographiques .....</b>	<b>270</b>

## Liste des figures

Figure 1 : Cascade des omiques en fonction de la réponse génotypique et phénotypique [3].....	6
Figure 2 : Approche Top-down <i>versus</i> Bottom-up lors d'une analyse en protéomique (Adapté de la publication de Gregorich <i>et al.</i> [8]). .....	7
Figure 3 : Nombre de publications sur la période de 2005 à 2020 associant les mots clés Spectrométrie de masse à Protéomique ou Lipidomique ou Métabolomique répertoriées sur le site internet « web of science ».....	13
Figure 4 : Principe des analyses en acquisition dépendante des données (DDA) avec un Top 3 comme exemple (Adapté de la publication de Ludwig <i>et al.</i> [29]). .....	15
Figure 5 : Principe du mode DIA. (A) En mode DIA, un seul spectre d'ions précurseurs ( $MS^1$ ) est enregistré, suivi d'une série de spectres d'ions fragments ( $MS^2$ ) avec de larges fenêtres d'isolement en précurseurs ( <i>ex.</i> 25 $m/z$ ) [29]. .....	16
Figure 6 : Principe de la MRM et la MRM <sup>3</sup> . .....	18
Figure 7 : Principe de la PRM.....	19
Figure 8 : Évolution de la capacité de multiplexage en fonction des paramètres de masse (Adapté de la thèse de Blandine Rougemont [39]). .....	21
Figure 9 : Différence entre la MRM et la sMRM. ....	23
Figure 10 : Comparaison de la séparation chromatographique des lipides en mode RPLC et en mode HILIC pour l'analyse lipidique. (A) En HILIC, les lipides d'une même famille se superposent tandis que les familles sont bien séparées entre elles [80]. (B) En mode RPLC, les lipides sont bien séparés au sein d'une même famille. En revanche, il y a une co-élution entre les familles [55]. .....	29
Figure 11 : Séparation d'un ensemble de peptides en HILIC avec 0,5% de m-NBA (p/v) dans la phase mobile (partie inférieure) et RPLC (partie supérieure) [87]. .....	31
Figure 12 : Exemple de mesure de l'espace de séparation pour un système LCxLC [97]. .....	33
Figure 13 : Principe de la chromatographie 2D en mode <i>multiple heart cut</i> [106]. .....	35
Figure 14 : Principe de la chromatographie 2D en mode <i>comprehensive</i> [106]. .....	36

Figure 15 : Évolution de la forme des pics lors de l'augmentation du volume d'injection par rapport au volume mort de la colonne (V0) [114].	39
Figure 16 : Variation de l'aire du pic en fonction de la concentration dans des conditions de total <i>breakthrough</i> [114].	40
Figure 17 : Mise en forme de la collection de chromatogrammes de 2 <sup>ème</sup> dimension en un chromatogramme 2D-LC [121].	42
Figure 18 : Schéma du système d'interface 2D-LC NP-RP basé sur une vanne à deux positions à dix ports avec une interface de séchage en ligne [130].	45
Figure 19 : Schéma d'un système 2D-LC à débit arrêté en 1 <sup>ère</sup> dimension [132].	46
Figure 20 : Schéma du système 2D-LC de type <i>heart cut</i> [128].	47
Figure 21 : Illustration d'un échantillon issu du processus MPLEx [160].	54
Figure 22 : Protocole simplifié des procédures SIMPLEX et MPLEx	54
Figure 23 : Comparaison entre le protocole MPLEx et des protocoles de référence en protéomique [158].	55
Figure 24 : Comparaison entre le protocole SIMPLEX et un protocole de référence en protéomique et métabolomique [159].	56
Figure 25 : Schéma expérimental des méthodes d'extraction classique et double extractions utilisées pour effectuer des analyses métabolomiques et lipidomiques (Adapté de la publication de Villaret-Cazadamont <i>et al.</i> [155]).	57
Figure 26 : Comparaison de la récupération des métabolites polaires avec la méthode classique (barres noires) et la méthode double extractions (barres blanches) [155].	58
Figure 27 : Comparaison de la quantification des classes de lipides entre la méthode classique et la méthode de double extractions [155].	59
Figure 28 : Comparaison de la distribution des espèces de sphingomyélines (SM) et de triglycérides (TG) entre les méthodes d'extraction classique et double extractions [155].	59
Figure 29 : One step extraction versus extraction MTBE : Préparation des échantillons (Adapté de la publication de Medina <i>et al.</i> [166]).	60

Figure 30 : Abondance relative du signal (par classe de lipides) par rapport au signal récupéré avec l'extraction biphasique à l'éther méthyl-tert-butyle (MTBE) [166].	62
Figure 31 : L'abondance relative des classes de métabolites polaires dans les extraits IPA et BUME par rapport à l'extrait plasmatique MeOH [166].	62
Figure 32 : Schéma de la procédure d'extraction permettant d'extraire et de séparer les protéines/métabolites/ADN/ARN à partir d'un extrait cellulaire unique (Adapté de la publication de Vorreiter <i>et al.</i> [167]).	64
Figure 33 : Évaluation du risque écotoxicologique et représentation des réponses induites au sein d'un organisme en réponse à une contamination (Adapté de la thèse de Judith Trapp [170]).	66
Figure 34 : Organisme <i>Gammarus fossarum</i> utilisé comme espèce sentinelle pour le développement d'indicateur de la contamination chimique des milieux aquatiques.	69
Figure 35 : Schéma représentant la procédure utilisée pour déterminer le stade d'inter-mue des organismes <i>G. fossarum</i> [170].	70
Figure 36 : Changements intertégumentales à l'extrémité des périopodes 3 et 4 chez <i>G. fossarum</i> au cours du cycle de mue [195].	71
Figure 37 : Stades de mue corrélés aux stades embryonnaires chez la femelle <i>G. fossarum</i> [170].	72
Figure 38 : Évolution, au cours du temps, des techniques de protéomique utilisées dans les approches écotoxicologiques (Adapté de la thèse de Duarte Domingos Gouveia [204]).	73
Figure 39 : Identification des protéines par protéogénomique. Les données génomiques et transcriptomiques sont utilisées pour générer des bases de données de séquences protéiques permettant l'interprétation de spectres de masse issus d'analyses nano LC-MS/MS avec une approche <i>shotgun</i> .	76
Figure 40 : Localisation du PC 34:1 dans une coupe verticale de <i>G. fossarum</i> par MALDI MSI [198].	78
Figure 41 : Résumé de la méthodologie Scout-MRM dans le cadre de la publication <i>High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics</i> [86].	84
Figure 42 : Identification of endogenous peptides from <i>G. fossarum</i> by MRM and selection of transitions.	94

Figure 43 : Correlation curve plotting the retention time of peptides identified by proteogenomics to the peptide retention time obtained in MRM.. .....	95
Figure 44 : Effects of RT shifts for targeted peptides monitored by Scheduled-MRM® <i>versus</i> Scout-MRM method.....	96
Figure 45 : Scout-MRM concept.....	97
Figure 46 : Correlation study between 3 MRM transitions of the same peptide.. .....	98
Figure 47 : Area ratio between 3 peptides of the same protein.....	99
Figure 48 : Assessment of the correlation study of peptide ratios for the same protein.....	100
Figure 49 : Repeatability of endogenous peptides : Average intensity of the MRM transitions presents in the Scout-MRM method according to their repeatability.....	101
Figure 50 : Variation of the areas of heavy peptides spiked in different gammarid extracts.....	102
Figure 51 : Example of calibration curves for 4 isotopically enriched peptides.....	103
Figure 52 : Comparative proteomic analysis using active biomonitoring in an agriculture watershed. ....	105
Figure 53 : Mise en évidence de l'effet de lots. ....	108
Figure 54 : Stratégie pour convertir une méthode de quantification en protéomique shotgun vers de la protéomique ciblée, en utilisant la méthodologie Scout-MRM [271]. ....	111
Figure 55 : Règles de nomenclature des lipides en accord avec celles utilisées par la base de données Lipid maps [278].. .....	122
Figure 56 : Diversité moléculaire du phospholipide PC 34:1 [279]. ....	123
Figure 57 : Schémas de fragmentation des lipides.....	124
Figure 58 : Différence de structures entre sn-2 et sn-1 pour les LPC. ....	127
Figure 59 : Séparation des LPC en chromatographie de phase inverse avec deux conditions chromatographiques différentes. ....	127
Figure 60 : Problèmes d'interférences isobariques entre deux lipides qui ne diffèrent que d'une insaturation. ....	128

Figure 61 : Recouvrement du massif isotopique du TG 54:3 et du TG 54:4.....	129
Figure 62 : Séparation chromatographique de lipides de <i>G. fossarum</i> selon les conditions chromatographiques décrites dans le tableau n°7. ....	130
Figure 63 : Exemple de séparation chromatographique de TG de <i>G. fossarum</i> selon les conditions chromatographiques décrites dans le tableau n°7. ....	131
Figure 64 : Fonctionnement du déclenchement des groupes scouts en fonction de l'intensité des scouts et de la présence d'interférents.....	132
Figure 65 : Distribution des lipides scouts le long du gradient chromatographique. ....	133
Figure 66 : Remplissage des groupes scouts avec les transitions MRM d'intérêts selon la présence ou l'absence d'une co-élution avec un composé scout.....	134
Figure 67 : Impact de la diminution du <i>dwell time</i> et du nombre de points par pic chromatographique sur la répétabilité de l'intégration des aires d'une transition en Scout-MRM.....	136
Figure 68 : Impact de la diminution du <i>dwell time</i> et du nombre de points par pic chromatographique sur la répétabilité de l'intensité du signal d'une transition en Scout-MRM. ....	137
Figure 69 : Proportion des acides gras les plus souvent retrouvés dans la composition des TG.....	140
Figure 70 : Exemples de courbes de corrélation correspondant au nombre de carbones en fonction du temps de rétention et au nombre de doubles liaisons en fonction du temps de rétention pour des lipides d'une même famille. ....	142
Figure 71 : Impact de la séparation de composés isomères sur la prédiction des temps de rétention. ....	144
Figure 72 : Incertitudes associées à la prédiction des temps de rétention.....	146
Figure 73 : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration. ....	149
Figure 74 : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidique en temps de rétention lorsque les conditions chromatographiques utilisées sont différentes de celles employées pour son élaboration. ....	150

Figure 75 : Impact des décalages en temps de rétention des lipides lors d'une analyse en sMRM et en Scout-MRM.....	152
Figure 76 : Différents cas rencontrés lors de l'évaluation des changements de sélectivité en Scout-MRM lors du passage d'une colonne de phase inverse à une autre. ....	154
Figure 77 : Résultats de l'étude concernant l'évaluation des changements de sélectivité en Scout-MRM pour 4 colonnes en phase inverse par rapport à une colonne de référence.....	155
Figure 78 : Répétabilité des lipides endogènes en fonction de l'aire ou de l'intensité du signal. ....	157
Figure 79 : Workflow of the different proteins, lipids, and metabolites extraction conditions from <i>Gammarus fossarum</i> . ....	173
Figure 80 : Heatmap view of peptides monitored by Scheduled-MRM according to the different protein extraction conditions: (A) methanol (MeOH), lysis buffer including (C) or not a centrifugation step after the extraction (B).....	174
Figure 81 : Comparison of the homogenization conditions on the lipid extraction yield.....	175
Figure 82 : Comparison of protocol "lysis buffer with centrifugation" (C) versus protocol "MeOH" (A) and "lysis buffer without centrifugation" (B) in metabolite analysis.....	177
Figure 83 : Signal variations of peptides by LC-MS/MS when switching from the classical proteomic protocol to the multi-omic protocol. ....	178
Figure 84 : Signal variations of metabolites by LC-MS/MS when switching from the classical metabolomic protocol to the multi-omic protocol. ....	180
Figure 85 : Evaluation of repeatability by the measurement of endogenous peptides (A), lipids (B), and metabolites (C) from <i>G. fossarum</i> . ....	181
Figure 86 : PCA score scatter plots based from the multi-omics data set, namely proteomics data set acquired in ESI(+) (dynamic-MRM) (A), from metabolomics data set acquired in ESI(-) (B), from lipidomics data set acquired in ESI(+) (C), and ESI(-) (D) by LC-ESI-MS/MS (scheduled-MRM). ....	182
Figure 87 : PCA score scatter plot from metabolomics data set acquired in ESI(+) by LC-ESI-MS/MS (scheduled MRM). ....	182
Figure 88 : PLS-DA score scatter plots from the multivariate analysis of the multi-omics data sets: proteomics ESI(+) (A), metabolomics ESI(-) (B), lipidomics ESI(+) (C) and ESI(-) data sets (D).....	183



Figure 89 : Multiple factor analysis (MFA) unveils proteomics, lipidomics and metabolomics variation during the different reproductive stage in female gammarid. ....	188
Figure 90 : Interface pour la chromatographie bidimensionnelle en mode <i>compréhensive</i> (LCxLC) [121]. .....	197
Figure 91 : Interface pour la chromatographie bidimensionnelle en mode <i>multiple heart cut</i> (LC-LC) [121]. ....	198
Figure 92 : Capacité d'isolation des classes lipidiques en fonction de l'ordre des dimensions lors des analyses bidimensionnelles avec une colonne HILIC et RPLC. ....	201
Figure 93 : Compositions à l'élution de composés lipidiques en HILIC en fonction des compositions à l'élution de ces mêmes composés en RPLC. ....	202
Figure 94 : Évaluation du phénomène de <i>breakthrough</i> pour les standards PC et LPC obtenus par RPLC-HILIC hors ligne en faisant varier les volumes d'injection dans la 2 <sup>ème</sup> dimension. ....	204
Figure 95 : Évaluation du phénomène de <i>breakthrough</i> pour les standards PE, PG, PS, DG, TG et Cer obtenus par RPLC-HILIC hors ligne en faisant varier les volumes d'injection dans la 2 <sup>ème</sup> dimension. .....	205
Figure 96 : Évaluation du phénomène de <i>breakthrough</i> pour la famille des LPC obtenu par HILIC-RPLC hors ligne en faisant varier les volumes d'injection dans la 2 <sup>ème</sup> dimension. ....	206
Figure 97 : Système HILICxRPLC-QTOF. ....	208
Figure 98 : Comparaison de la séparation (A) HILIC-MS/MS versus (B) RPLC-MS/MS versus (C) HILICxRPLC-MS/MS lors de l'analyse de 4 lipides (PE 38:6, PE 38:5, PC 34:1, PC 34:2) provenant de deux familles lipidiques différentes. ....	209
Figure 99 : Mise en évidence de larges fenêtres de fractionnement pour les familles de lysophospholipides (ex. LPE et LPC) lors de leur analyse en HILIC-MS/MS. ....	211
Figure 100 : Système HILIC-RPLC/IM-QTOF. ....	212
Figure 101 : Fonctionnement du fractionnement en LC-LC lorsque trois fractions uniques sont séparées par une pause. ....	213
Figure 102 : Séparation des différentes familles lipidiques dans la 1 <sup>ère</sup> dimension en mode HILIC lors de l'analyse d'un extrait de <i>G. fossarum</i> . ....	216

Figure 103 : Séparation des différentes familles lipidiques du gammare dans la première dimension en mode HILIC avec ajout de standards à chaîne d'acides gras courts par classe de lipides. ....	217
Figure 104 : Séparation chromatographique d'origine pour les familles lipidiques (LPL / PL / SM / Cer / DG / TG) avec la méthode RPLC-MS/MS en mode ESI (+).....	219
Figure 105 : Séparation chromatographique optimisée pour les familles lipidiques (PL / SM) avec la méthode RPLC-MS/MS utilisée en deuxième dimension en mode ESI (-). ....	219
Figure 106 : Séparation chromatographique optimisée pour les familles lipidiques (TG / DG / Cer) avec la méthode RPLC-MS/MS utilisée en deuxième dimension en mode ESI (+).....	220
Figure 107 : Optimisation des paramètres de masse (Nébuliseur, débit de <i>sheat gas</i> et de gaz séchant, température du <i>sheat gas</i> et du gaz séchant, tension du <i>nozzel</i> en mode ESI (+) et (-), tension du capillaire en mode ESI (+) et ESI (-) par famille lipidique sur un système IM-QTOF 6560 de chez Agilent® avec un débit de 500 µL/min lors de l'ionisation (n=3).....	221
Figure 108 : Optimisation des énergies de collision en mode ESI (+) et ESI (-) par famille lipidique sur un système IM-QTOF 6560 de chez Agilent® avec un débit de 500 µL/min lors de l'ionisation (n=3). ....	222
Figure 109 : Analyse de la fraction des PC avec l'approche classique RPLC-MS versus <i>multiple heart cut</i> HILIC-RPLC-MS.....	223

# Liste des tableaux

Tableau 1 : Avantages et inconvénients des modes hors ligne, en ligne et à débit stoppé en 2D-LC. [107].	37
Tableau 2 : Avantages et inconvénients des différents systèmes 2D-LC (hors ligne, <i>stop flow</i> , <i>heart cut</i> et <i>comprehensive</i> ) dans le cas des analyses lipidomiques.	44
Tableau 3 : Liste des lipides scouts utilisés dans la méthode.	116
Tableau 4 : Liste et concentrations des standards lipidiques utilisés pour l'étude du nombre de points par pic et du <i>dwell time</i> en fonction de la répétabilité de mesure en Scout-MRM.	119
Tableau 5 : Liste des pertes de neutre lors de l'analyse des triacylglycérols.	120
Tableau 6 : Liste des standards lipidiques utilisés lors de l'étude de répétabilité.	121
Tableau 7 : Paramètres chromatographiques utilisés pour le développement de la méthode Scout-MRM pour l'analyse lipidique.	130
Tableau 8 : Variation du nombre de transitions/groupe scout en fonction du <i>dwell time</i> (DT = 5 ms ou 3 ms ou 1 ms) et du nombre de points/pic chromatographique (10 ou 8 points) dans le cas de pic chromatographique d'une largeur moyenne de 16 secondes.	136
Tableau 9 : Exemple de base de données en temps de rétention pour les DG.	143
Tableau 10 : $\Delta RT_{(isomères)}$ associée à chaque famille de lipides.	145
Tableau 11 : Domaines de linéarité et limites de quantification (LOQ) pour chaque standard lipidique en ESI (+) et ESI (-) obtenues grâce à l'aire ou l'intensité des signaux détectés.	158
Tableau 12 : Evaluation of QRESS metabolites recovery and extraction efficiency.	179
Tableau 13 : Summary of the PLS-DA models performance parameters.	183
Tableau 14 : List of peptides and their corresponding proteins with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted.	184
Tableau 15 : List of metabolites with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted.	185
Tableau 16 : List of lipids with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted.	186

Tableau 17 : List of lipids with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted. .....	187
Tableau 18 : Tensions appliquées pour les analyses en LC-LC pour chaque famille de lipides en fonction de la polarité d'ionisation.....	199
Tableau 19 : Paramètres de masse pour les analyses en LC-LC pour chaque famille de lipide en fonction de la polarité d'ionisation.....	223

## Liste des Annexes

Annexe 1 : Publication "from shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential immunomarkers in <i>Dreissena polymorpha</i> ". .....	234
Annexe 2 : Concentrations des ISTD utilisés pour réaliser la gamme de linéarité en mode ESI (+) et ESI (-). .....	249
Annexe 3 : Structures des lipides. ....	250
Annexe 4 : Masse en Q1, Q3, DP et CE en fonction du mode d'ionisation pour les transitions MRM au sein de la méthode Scout-MRM.....	251
Annexe 5 : Base de données en temps de rétention pour l'élaboration de méthode Scout-MRM en lipidomique.....	253
Annexe 6 : Evaluation of the protein extraction protocols on the peptide's abundance.....	260
Annexe 7 : Comparison of the homogenization conditions on TG 51:1-d5 and PC 11:0/11:0 internal standard (A) and TG species before and after normalization (B).....	261
Annexe 8 : Comparison of the different homogenization conditions on gammarid metabolites. ....	262
Annexe 9: Determination of protein concentration using Bradford assay with classical Proteomics protocol, multi-omics with dual protein precipitation and multi-omics protocol.....	264
Annexe 10 : Comparison of peak area of the peptides with PLS-DA VIP scores (VIP >1) according to the different reproductive stages of female gammarids. ....	265
Annexe 11 : Comparison of peak area of metabolites with PLS-DA VIP scores (VIP >1) according to the different reproductive stages of female gammarids. ....	265
Annexe 12 : Lg factors for MFA. ....	265

Annexe 13 : RV factors for MFA. ....	266
Annexe 14 : Évaluation du phénomène de <i>breakthrough</i> pour les familles lipidiques (SM, Cer, TG, PC, DG et PE) obtenues par HILIC-RPLC hors ligne en faisant varier les volumes d'injection de 5µL à 40µL dans la 2 <sup>ème</sup> dimension, partie 1. ....	267
Annexe 15 : Évaluation du phénomène de <i>breakthrough</i> pour les familles lipidiques (PI, PS et PG) obtenues par HILIC-RPLC hors ligne en faisant varier les volumes d'injection de 5µL à 40µL dans la 2 <sup>ème</sup> dimension, partie 2. ....	268
Annexe 16 : Évaluation de la compatibilité de la spectrométrie de masse en mode ciblé avec les systèmes LCxLC et LC-LC.....	269

## Introduction générale

La biologie des systèmes est un domaine d'étude interdisciplinaire (biologie, chimie analytique, bioinformatique, *etc.*). Elle se concentre sur les interactions complexes au sein d'un réseau d'entités biologiquement pertinent (*ex.* cellule) aussi appelé systèmes biologiques. Les raisonnements et les hypothèses scientifiques dans la biologie des systèmes sont appliqués à l'ensemble du système plutôt que sur les parties qui le composent. Dans le cas où le système étudié correspond au niveau cellulaire d'un organisme, les différentes parties qui le composent sont les biomolécules comme les gènes, les ARN, les protéines, les métabolites et les lipides. Ces composés qui sont en interaction forment le système dont la complexité est difficile à appréhender lorsque la caractérisation de ces compartiments biologiques est réalisée de manière isolée. Il est donc impossible d'avoir une vision globale d'un système sans l'étude de chacune de ces familles de biomolécules communément appelée science omique (génomique, transcriptomique, protéomique, métabolomique, lipidomique). Depuis plusieurs années, les approches multi-omiques se sont développées pour acquérir simultanément différentes données omiques et faciliter la compréhension des mécanismes moléculaires.

La biologie systémique s'articule autour de 4 problématiques qui sont : (i) l'acquisition des données quantitatives pour chaque compartiment constituant le système, (ii) l'élaboration des modèles mathématiques basés sur ces données, (iii) la reconstruction des réactions au niveau de la cellule ou d'un individu et (iv) la mise en place d'hypothèses permettant d'expliquer les mécanismes mis en jeu et la variabilité des données. Les travaux décrits dans ce manuscrit de thèse seront concentrés sur l'optimisation des aspects expérimentaux pour faciliter l'acquisition des données multi-omiques de chaque compartiment constituant le système d'étude. De nombreux verrous, comme l'augmentation des capacités de multiplexage des méthodes d'analyse, la diminution de la complexité de l'échantillon ou la mise en place de protocole multi-omique intégré à partir d'une prise d'essai unique devront être surmontés. Cela permettra d'obtenir une vision analytique précise et juste des différents compartiments biologiques à un instant donné.

La description exhaustive d'un système consiste à élaborer une base de données dans laquelle figurent tous les éléments constitutifs du système biologique étudié. Le premier verrou majeur est donc le manque d'information qui peut altérer le contexte biologique d'un compartiment donné. Pour éviter que cela ne se produise, il est nécessaire de pouvoir analyser le plus grand nombre de molécules et ainsi avoir un multiplexage conséquent. Les approches globales par spectrométrie de masse de haute résolution (HRMS) sont des méthodologies de choix pour l'identification et la caractérisation de milliers de composés en simultanée dans des milieux complexes. Cependant, il n'est plus à démontrer que les méthodes par spectrométrie de masse ciblée, spécialisées dans les méthodes de quantification,

sont plus sensibles, répétables et robustes, mais en contrepartie les capacités de multiplexage sont moins importantes. Dans le cadre de cette thèse, nous avons choisi de privilégier les approches ciblées en mode *Multiple Reaction Monitoring* (MRM) pour disposer des méthodes de quantification les plus performantes. Toutefois, nous avons cherché à augmenter le nombre de molécules analysable simultanément par ce mode d'acquisition. Pour améliorer les capacités de multiplexage en MRM, de nombreux laboratoires utilisent des fenêtres de temps centrées sur la zone d'élution des composés cibles. Toutefois, pour augmenter drastiquement les capacités de multiplexage il faut diminuer les fenêtres de temps ce qui rend cette méthodologie moins robuste puisque la méthode devient sensible aux décalages en temps de rétention. L'un des objectifs de cette thèse est d'augmenter le nombre de composés analysés simultanément par LC-MS/MS ciblée sans sacrifier la robustesse de la méthode. Pour résoudre cette problématique, un nouveau mode d'acquisition basé sur la MRM a été développé dans notre laboratoire. Cette approche, intitulée Scout-MRM, augmente les capacités de multiplexage ainsi que la robustesse puisque son mode de fonctionnement n'est pas basé sur les temps de rétention, mais sur les ordres d'élution des analytes d'intérêt par rapport à des composés présélectionnés appelés scout. Avec ce mode d'acquisition, les groupes de transitions MRM d'intérêt sont déclenchés grâce à la détection de ces scouts. Pour prouver qu'il est possible de construire et d'utiliser en routine des méthodes hautement multiplexées, l'approche Scout-MRM appliquée au domaine de la protéomique a été mise en place dans le 2<sup>ème</sup> chapitre de ce manuscrit. À travers ces travaux, nous avons souhaité mettre en œuvre l'un des plus grands multiplexes pour la quantification de protéines existant à ce jour. Nous avons également voulu démontrer l'intérêt de cette méthodologie lors de la recherche de potentiels biomarqueurs protéiques en écotoxicologie grâce à une application de biosurveillance active dans un contexte de pollution agricole.

L'ensemble des applications MRM est exclusivement utilisé pour le suivi de biomolécules présélectionnées à partir d'analyse par spectrométrie de masse de haute résolution. Toutefois, les récents progrès techniques et méthodologiques apportés par le Scout-MRM ont permis de repousser les limites des capacités de multiplexage en mode ciblé. Désormais, plusieurs centaines de composés peuvent être analysés simultanément et rapidement grâce à l'utilisation de grandes listes de transitions MRM issues de bases de données *in silico*. Ainsi, la méthodologie Scout-MRM permet d'envisager la réalisation d'études de « découverte ciblée » sans analyse HRMS au préalable. Bien que les transitions MRM puissent être déterminées à partir des bases de données, il est nécessaire de connaître l'ordre d'élution des composés d'intérêt par rapport aux scouts pour pouvoir construire les groupes de détection lors de l'élaboration de la méthode. Il est donc important d'avoir des outils de prédiction pour déterminer l'ordre d'élution des composés le long du gradient chromatographique. Le 3<sup>ème</sup> chapitre de ce manuscrit traitera de l'association de la lipidomique et de la méthodologie Scout-MRM puisque les lipides se prêtent particulièrement bien à la réalisation d'étude de découverte ciblée.

En effet, il existe plusieurs approches pour estimer les ordres d'élution des lipides permettant ainsi de construire les méthodes Scout-MRM en lipidomique. De plus, le développement d'une approche hautement multiplexée en lipidomique serait une véritable plus-value, car il n'existe à ce jour que peu de méthodes MRM capables d'analyser et de quantifier précisément de nombreux lipides.

Le deuxième verrou rencontré lors des analyses multi-omiques est l'intégration des analyses du métabolome incluant le lipidome à celui du protéome afin de capitaliser sur leurs complémentarités pour une meilleure corrélation des résultats entre les trois familles omiques. À ce jour, bien qu'utilisant le même outil, les différentes approches omiques par spectrométrie de masse ciblée se font séparément avec des traitements d'échantillon et des prises d'essais multiples (réplicats d'exposition). Cela pose non seulement des problèmes de coûts, de temps et de quantité d'échantillons consommée, mais cela va aussi induire plus de variations inter-individus. Le 2<sup>ème</sup> objectif de cette thèse est donc de développer de nouvelles approches analytiques permettant d'analyser les protéines, les lipides et les métabolites à partir d'une prise d'essai unique. Ainsi une méthode d'extraction multi-omique intégrée sera développée et décrite dans le 4<sup>ème</sup> chapitre de ce manuscrit. Il s'agira d'une extraction liquide/liquide utilisant un mélange de solvants composé d'eau, de méthanol et de MTBE qui induira une séparation biphasique couramment utilisée pour extraire les lipides. Contrairement à une approche lipidomique classique, la fraction aqueuse contenant les protéines et les métabolites ne sera pas éliminée. On conservera ainsi les trois familles omiques. Cette extraction sera adaptée pour l'analyse de matrices complexes solides comme *G. fassarum* en optimisant les différentes conditions d'homogénéisation de l'échantillon. Les performances de cette méthode d'extraction multi-omique seront évaluées et comparées aux méthodes omiques existantes.

Le 3<sup>ème</sup> verrou est que les analyses multi-omiques imposent de prendre en considération l'ensemble des molécules présentes dans le système étudié ce qui implique l'analyse d'échantillon complexe. Cette complexité peut induire des effets matrices impactant la qualité du signal (diminution de l'intensité, apparition d'interférences, etc.). La préparation d'échantillon multi-omique peut réduire la complexité d'un échantillon, mais elle implique de faire des compromis en adoptant une procédure compatible avec l'ensemble des omiques. De ce fait, la simplification de l'échantillon n'est pas aussi importante qu'elle ne le serait si une seule omique avait été pris en considération. Pour contrebalancer ce manque de décomplexification, une dimension de séparation supplémentaire est apportée par les techniques chromatographiques, à savoir la chromatographie liquide dans notre cas. Toutefois, ces méthodes monodimensionnelles (1D-LC) peuvent présenter des limites en termes de capacité de séparation. De nos jours, de nouveaux outils de séparation plus avancés sont disponibles comme la chromatographie liquide bidimensionnelle (2D-LC) qui dispose de deux dimensions séparatoires orthogonales. Le dernier objectif de cette thèse est donc d'évaluer que la 2D-LC est capable de



diminuer la complexité d'un échantillon en réduisant les effets matrices par rapport aux approches 1D-LC. L'état de l'art bibliographique sur l'analyse des lipides a également montré des besoins d'optimisation de séparation. Ainsi le développement d'une approche chromatographique bidimensionnelle couplée à la spectrométrie de masse a été initié en lipidomique. Ces travaux seront décrits dans le 5<sup>ème</sup> et dernier chapitre de cette thèse. Le mode de séparation 2D sélectionné est le mode *multiple heart cut* (HILIC-RPLC-MS/MS). Les avantages et inconvénients de cette méthode chromatographique seront comparés aux approches *comprehensive* (RPLCxHILIC-MS/MS et HILICxRPLC-MS/MS) déjà décrites dans la littérature. Pour finir, une étude de compatibilité de la 2D-LC avec la spectrométrie de masse sera réalisée.

L'ensemble de ces développements sera effectué sur l'espèce sentinelle *G. fossarum*, couramment utilisée pour surveiller l'état des cours d'eau. Cela a permis d'implémenter de nouvelles technologies dans le domaine de l'écotoxicologie en pleine expansion.

Avant de présenter ces résultats, une étude bibliographique portant sur la spectrométrie de masse, les extractions multi-omiques et la 2D-LC est présentée dans le 1<sup>er</sup> chapitre de ce manuscrit.

# Chapitre 1 : Étude bibliographique

## 1. Les sciences omiques combinées ou approches multi-omiques

Au cours des années, les progrès dans le développement des technologies à haut débit d'analyse ont permis d'identifier et de mesurer de nombreuses biomolécules au sein d'une cellule ou d'un tissu biologique [1]. Les données moléculaires ainsi que les domaines scientifiques issus de ces technologies sont qualifiés de omiques. Les sciences omiques regroupent et associent des techniques de chimie analytique, de biochimie et de biologie moléculaire ainsi que les traitements des informations obtenues. Ces données générées permettent de comprendre le fonctionnement des systèmes biologiques, en caractérisant d'un point de vue moléculaire le génotype et le phénotype d'un individu. Il existe une multitude de domaines de recherche omiques dont la génomique, la transcriptomique, la protéomique et la métabolomique.

### 1.1. La génomique

La plus ancienne de ces sciences omiques est la génomique dont les travaux de Watson et Crick sur la structure moléculaire de l'ADN furent déterminants en 1953 [2]. Contrairement à la génétique qui ne s'intéresse qu'à l'étude de gènes spécifiques et leurs hérédités, la génomique désigne l'étude de l'ensemble des gènes ou génome d'un organisme. Il existe deux catégories de génomique [3]:

- La première catégorie est appelée génomique structurale. On peut également la retrouver sous l'appellation génétique moléculaire. Son rôle consiste à étudier la structure des gènes en localisant et annotant les gènes, en déterminant le nombre d'exons, d'introns *etc.*
- La deuxième catégorie est appelée génomique fonctionnelle et consiste en l'étude de la fonction des gènes responsables de la production des protéines.

L'ensemble des informations génétiques qui définit les caractères héréditaires d'un individu est regroupé dans le génome. Cependant, ce génome ne permet pas d'expliquer la totalité des réponses observées chez un organisme ayant subi une perturbation extérieure, puisqu'il ne s'agit que d'informations éventuellement exprimées [3]. À l'inverse du génotype qui représente le patrimoine héréditaire dépendant des gènes, le phénotype regroupe l'ensemble des caractères apparents d'un individu. Le phénotype découle du génotype, mais il peut être modulé par des modifications environnementales [3]. De ce fait, les sciences omiques découlant de la génomique ont été développées pour étudier les réponses subsidiaires de l'ADN (ARN, protéines, métabolites et lipides). Cela a permis de se rapprocher des réponses phénotypiques, comme illustré sur la figure n°1 avec la

cascade des omiques. Plus une science omique est éloignée de la génomique plus la réponse biologique se rapprochera du phénotype plutôt que du génotype.

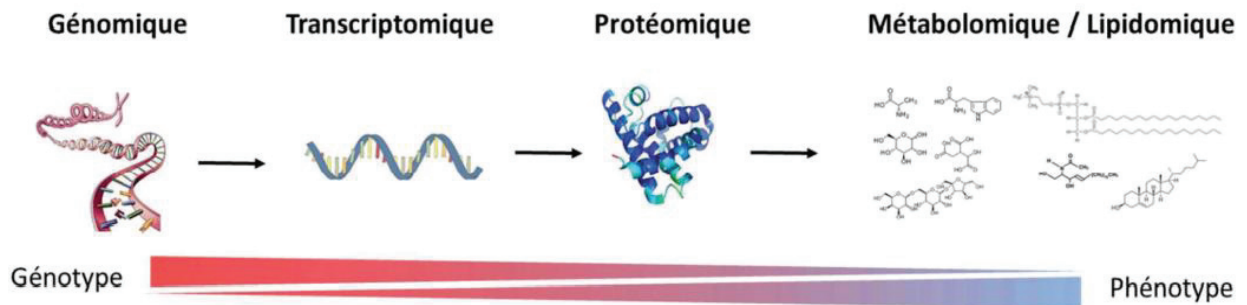


Figure 1 : Cascade des omiques en fonction de la réponse génotypique et phénotypique [3].

## 1.2. Les études phénotypiques

### 1.2.1. Transcriptomique

L'étude des éléments issus de la traduction de l'ADN en ARN est appelée transcriptomique. Le terme de transcriptome est apparu à la fin des années 90 avec les travaux de Velculescu *et al.* [4]. Il représente l'ensemble des transcrits au sein d'une cellule à un instant donné. Ces transcrits peuvent être des ARN messagers (ARNm) qui codent les protéines, mais aussi des ARN de transfert ou ribosomaux acteurs de la synthèse des protéines. À l'inverse du génome qui est constant, le transcriptome est dynamique et peut être modulé par des facteurs extérieurs. La transcriptomique est donc le premier lien qui permet de relier le génome d'un individu à son phénotype à un instant donné [3].

Comme pour les ARN qui sont issus de la transcription des gènes contenus dans l'ADN, il existe un niveau de régulation supplémentaire qui est issu du transcriptome, à savoir la protéomique.

### 1.2.2. Protéomique

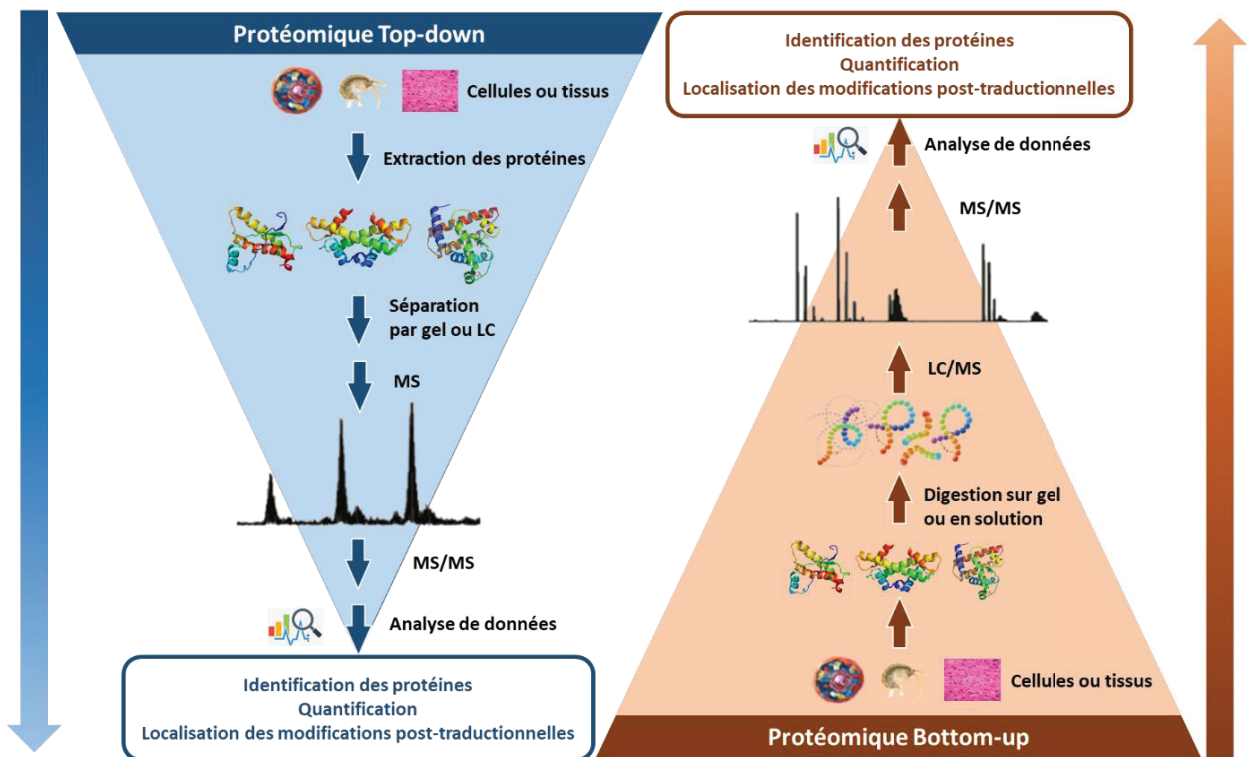
#### 1.2.2.A. Qu'est-ce que la protéomique ?

La protéomique s'est développée à la même période que la transcriptomique avec les travaux de Wilkins *et al.* en 1995 [5]. Elle étudie les produits issus de la traduction des ARNm appelés protéines. De même que le transcriptome, le protéome constitue un ensemble de protéines variant de façon dynamique au cours du temps. Les analyses protéomiques ont pour but d'identifier, de quantifier et de comprendre le fonctionnement des produits de la traduction des ARNm. Elles ont également pour objectif l'étude des modifications post-traductionnelles. Ces modifications des protéines, telles que la phosphorylation, la glycosylation ou la méthylation peuvent moduler le fonctionnement des protéines en faisant varier leur action catalytique ou leur rôle dans la signalisation cellulaire [6,7].

Les informations fournies par le génome peuvent ne pas corrélérer avec le contenu protéique puisqu'en plus des délais et des différences dans les taux de dégradation, l'ARNm n'est pas toujours traduit en protéine [6]. De ce fait, l'étude du protéome permet de mieux comprendre les réponses phénotypiques au sein d'un organisme.

### 1.2.2.B. Les approches en protéomique : Top-down versus Bottom-up

Lorsque l'on souhaite caractériser des protéines par spectrométrie de masse, deux approches existent : Top-down et Bottom-up. Comme illustrée par la figure n°2, la première approche dite descendante consiste à extraire et à analyser les protéines intactes d'un échantillon biologique par LC-MS alors que la deuxième méthodologie dite ascendante analyse des fragments de protéines aussi appelés peptides qui sont rapporteurs de la protéine [8–10].



**Figure 2 : Approche Top-down versus Bottom-up lors d'une analyse en protéomique.** En Top-down (à gauche), les protéines sont extraites des cellules ou tissus avant d'être séparées par gel ou LC, et d'être directement analysées par MS pour obtenir des données protéiques. Par la suite, une protéine spécifique peut être isolée et fragmentée par MS/MS pour obtenir des informations sur sa séquence et ainsi identifier cette protéine via une recherche dans des bases de données. En Bottom-up (à droite), les protéines extraites de cellules ou de tissus sont soumises à une digestion (souvent à l'aide de trypsine) et les peptides résultants sont séparés par LC et analysés par MS. Par la suite, les peptides sont fragmentés et les informations sur la séquence des peptides sont utilisées pour identifier les protéines présentes dans l'échantillon. (Adapté de la publication de Gregorich *et al.* [8]).

Chacune de ces stratégies présente ses propres avantages et inconvénients. Les analyses en Top-down sont particulièrement utilisées lors de la caractérisation structurale et l'identification de modifications post-traductionnelles des échantillons biologiques, avec par exemple les méthylations ou les phosphorylations [11,12]. En revanche, la stratégie Bottom-up permet l'identification de plus de

protéines grâce à une meilleure sensibilité en spectrométrie de masse [13]. Le groupe de Kelleher *et al.* [13] a réussi à identifier 7.4 fois plus de protéines avec l'approche descendante comparée à l'approche ascendante. La première raison est que les protéines sont des biomolécules de hauts poids moléculaires (10 kDa à 1000 kDa) ce qui rend difficile leur analyse en spectrométrie de masse à cause de la limite des gammes de masse. Cela n'est pas le cas des peptides qui sont de plus petits poids moléculaires (< 1500 Da). La seconde raison est que la sensibilité est impactée lors de l'analyse de protéines entières puisque ces molécules peuvent avoir de nombreux états de charges diluant ainsi le signal lors de leur ionisation. Le fractionnement des protéines en peptides de plus petites tailles permet d'avoir des molécules ayant moins d'états de charge (+1 à +3). Le désavantage de la stratégie Bottom-up est que la digestion enzymatique des protéines permettant d'obtenir les peptides tryptiques complexifie la matrice à analyser. Cela augmente drastiquement le nombre de composés dans le milieu nécessitant l'emploi de techniques de séparation en amont du détecteur. Dans le cadre de cette thèse, l'approche Bottom-up sera préférée.

Comme illustré sur la figure n°1, les protéines ne représentent pas l'ultime niveau de la cascade des omiques puisqu'il s'agit des métabolites. La métabolomique est donc le reflet le plus proche entre une réponse biologique et le phénotype d'une cellule ou d'un organisme.

### 1.2.3. La Métabolomique

La métabolomique est le domaine scientifique qui caractérise les biomolécules de faibles poids moléculaires (< 1 KDa ou < 1.5 KDa selon les définitions) dans un organisme biologique. Ces composés peuvent avoir trois origines différentes :

- Les métabolites endogènes qui sont produits naturellement par le métabolisme de l'organisme étudié.
- Les xénométabolites sont des biomolécules produites par un métabolisme indépendant de celui de l'organisme étudié. Par exemple, des métabolites d'origine bactérienne peuvent se retrouver dans le métabolome de l'hôte [14].
- Les xénobiotiques sont des métabolites exogènes qui sont retrouvés dans l'organisme étudié. Ils ne sont pas produits par l'organisme ni par son alimentation naturelle. Il peut s'agir par exemple, de pesticides ou de médicaments ainsi que les produits de leur dégradation.

En résumé, les métabolites sont les composés organiques qui ne sont pas issus de l'expression des gènes (acides aminés, sucres, acides organiques, acides gras, nucléotides, vitamines, stéroïdes, lipides, glycanes, *etc.*) [15,16]. Le métabolome peut être subdivisé en deux parties avec d'une part les métabolites polaires et d'une autre part les métabolites apolaires comme les lipides, par exemple. Ces molécules, propres à la lipidomique, sont des molécules organiques insolubles dans l'eau qui sont

classées dans 8 classes en fonction de leurs structures chimiques : les acides gras, les glycérophospholipides, les glycérolipides, les sphingolipides, les stérols, les prényls, les saccharolipides et les polyketides [17].

Comme indiqué sur la figure n°1, la métabolomique est la dernière étape de la cascade des omiques. Il s'agit donc de la réponse permettant de se rapprocher au plus près des caractères phénotypiques observés chez un individu ayant subi une perturbation (modification génétique, maladie, exposition à une substance toxique, *etc.*). En effet, ces petites molécules découlent en partie des activités des protéines et sont essentielles aux processus de régulation cellulaire [18,19]. Comparée aux autres omiques, la métabolomique s'applique à des molécules souvent présentes dans de nombreux organismes taxonomiquement éloignés. En effet, les métabolismes primaires directement impliqués dans les processus physiologiques principaux (croissance, respiration, développement, reproduction d'un organisme ou d'une cellule, *etc.*) sont relativement bien conservés. Cela permet d'adapter rapidement les méthodes d'analyses globales en métabolomique entre plusieurs espèces très variées. Contrairement à la génomique, la transcriptomique et la protéomique qui sont toutes issues de la séquence de gènes variables, il n'est pas nécessaire d'avoir à disposition un génome séquencé lors d'un changement d'organisme d'étude en métabolomique [18].

À ce jour, la métabolomique et la protéomique permettent d'obtenir les réponses les plus précises d'un organisme vis-à-vis d'une perturbation. Toutefois, les approches combinant simultanément les différentes omiques présentent plus d'avantages lorsque l'on tend vers des approches systémiques.

### 1.3. Les approches combinées

Depuis 30 ans, la biologie moléculaire n'a cessé d'évoluer pour tendre vers un contexte plus global en intégrant l'ensemble des compartiments biologiques ainsi que leurs interactions. Traditionnellement, les chercheurs avaient la volonté d'étudier le fonctionnement de chaque gène de manière individuelle. Cependant, la tendance à évoluer vers la compréhension de l'ensemble des gènes ainsi que des sous-produits qui en découlent (ARN, protéines, métabolites polaires et lipides). Pour étudier ces nouveaux réseaux, un concept moderne intitulé « biologie des systèmes » a émergé. Il se fonde sur le fait que toute entité biologique, que ce soit un génome, un organe, une cellule, *etc.*, représente un système dans lequel les constituants sont en interactions constantes et sont à l'origine des principaux processus biologiques [20]. Dans la biologie, aucun mécanisme ne se produit individuellement sans interaction avec son environnement. L'ensemble des éléments qui constitue le système sont en lien, en opposition ou en synergie. Par exemple, on peut rencontrer dans un système : la communication cellulaire, la circulation des métabolites dans l'organisme, la stimulation ou l'inhibition des produits de l'expression des gènes, *etc.* [21]. La biologie des systèmes repose sur le fait que les processus biologiques

proviennent des interactions de cause à effet entre constituants et donc que leurs origines sont définies comme un comportement collectif au sein des systèmes considérés. Les comportements issus de ces systèmes présentent une grande complexité à cause du nombre important de constituants qui les composent ainsi que leur hétérogénéité. Ces comportements peuvent provenir d'une organisation horizontale, c'est-à-dire à partir d'interactions en réseau, ou d'une organisation verticale avec des interactions multi-échelles entre le niveau moléculaire, cellulaire, *etc.*, jusqu'à l'individu, sa population ou son environnement. Ainsi la biologie des systèmes se définit comme l'étude des interactions complexes dans les organismes biologiques à travers la compréhension de la structure de ces compartiments, de leur dynamique et des processus de régulation les contrôlant.

La biologie des systèmes se concentre sur 4 grandes questions :

- Comment acquérir des données quantitatives pour chaque compartiment constituant le système ?
- Comment élaborer des modèles mathématiques basés sur ces données ?
- Comment reconstruire les réactions au niveau de la cellule ou d'un individu ?
- Quelles sont les hypothèses permettant d'expliquer les mécanismes mis en jeu et la variabilité des données ?

Pour répondre à ces questions et mettre en place des approches répondant au concept de biologie des systèmes, de nombreuses innovations ont dû être développées. Des techniques à haut débit d'analyse ont été élaborées pour permettre l'acquisition d'un large panel d'informations [21]. Les outils informatiques ont dû être améliorés pour intégrer la gestion d'un nombre conséquent de données générées par les hauts débits d'analyses [22].

Les modèles de système biologique bénéficient de l'inclusion de toutes les parties pertinentes du système considéré. Au niveau cellulaire, cela signifie qu'au minimum les "acteurs" les plus importants, c'est-à-dire les biomolécules doivent être étudiés ainsi que leurs interactions. Ces biomolécules, appartenant à différentes omiques, comprennent les gènes, les ARN, les protéines, les métabolites polaires et les lipides. Chacune de ces sciences omiques représente une pièce du puzzle qu'il faut associer pour obtenir une vision plus complète et globale d'un organisme en accord avec la réalité.

Les approches combinées de la biologie des systèmes permettent de tendre vers la biologie intégrative et prédictive. C'est-à-dire que l'analyse des différentes omiques et de leurs interactions dans un environnement donné permet d'appréhender les paramètres influençant un phénomène biologique. Cela a pour but d'élucider les mécanismes fondamentaux d'une pathologie pour proposer de nouvelles cibles thérapeutiques, d'identifier des traitements et d'élaborer de nouvelles méthodes de diagnostic [23].

## 1.4. Approches multi-omiques : émergence du concept et application

Comme discuté dans les paragraphes précédents, les interactions entre les classes biologiques sont omniprésentes au sein d'un individu. Il est donc impossible d'obtenir une vision holistique du comportement d'un organisme lors d'un stress externe dans le cadre d'une étude omique unique [24]. Au cours des dix dernières années, les initiatives visant à améliorer la compréhension des perturbations moléculaires dans un système donné se sont orientées vers les analyses multi-omiques. Cette approche étend les connaissances obtenues en mesurant et corrélant entre elles plusieurs classes biomoléculaires afin de mieux comprendre le phénotype exprimé. De ce fait, la multi-omique permet d'avoir une vision globale de ce qui pourrait se produire (génomique et transcriptomique) et de la façon dont cela se produit (protéomique, métabolomique, lipidomique) [24]. Les hypothèses fondées sur des analyses multi-omiques permettent d'obtenir des résultats plus fiables en se fiant à des informations provenant de sources orthogonales. Les données obtenues en protéomique, lipidomique ou métabolomique peuvent permettre de combler un manque d'informations dans l'une de ces omiques et ainsi valider ou non une hypothèse [25]. Par exemple, des analyses protéomiques réalisées sur l'organisme *Hyallolela Azteca*, un amphipode d'eau douce, contaminé par un herbicide appelé atrazine, semblent indiquer une perturbation de la glycolyse [18]. Cette hypothèse s'appuie sur l'observation d'une baisse des taux de glycéraldéhyde 3-phosphate déshydrogénase (GAPDH), de l'énolase et de l'adénosine triphosphate (ATP) synthase qui sont des protéines associées à la production d'énergie. Cependant, ces données sont en contradiction avec les mesures phénotypiques puisque ces organismes contaminés étaient physiquement plus grands que les organismes témoins. Pour vérifier cette hypothèse, l'analyse des métabolites polaires et non polaires a été réalisée. Cette étude a mis en évidence une régulation à la hausse des carnitines, des acides gras, des glycérols et de l'urée par rapport aux organismes témoins ce qui indique une oxydation des acides gras. En effet, le rôle physiologique des carnitines est de faciliter le transfert des acides gras à travers la membrane mitochondriale [26], où est réalisée la  $\beta$ -oxydation des acides gras qui constitue le principal processus par lequel les acides gras sont oxydés pour générer de l'énergie (ATP) [27]. L'utilisation conjointe des données de protéomique et métabolomique a permis d'affiner les résultats obtenus lors de la contamination de l'organisme *H. Azteca* par de l'atrazine. Il semble que ce pesticide réduit effectivement la glycolyse chez ces amphipodes exposés. Cependant, puisque les invertébrés sont composés principalement de lipides, les organismes contaminés peuvent compter sur la  $\beta$ -oxydation des acides gras pour répondre à leurs besoins énergétiques. Cette étude démontre les avantages de la multi-omique pour comprendre l'impact des contaminants environnementaux sur les organismes exposés [18].



Les analyses multi-omiques présentent de nombreux avantages, mais il existe certaines limitations. La multiplication du nombre d'extractions et du nombre d'analyses augmente nettement le temps global de l'expérimentation ainsi que la complexité des données en termes de retraitement. Il est plus long et difficile de corrélérer les différents résultats issus de l'ensemble des omiques [24]. De plus, le nombre de composés issus de l'association moléculaire présente dans les données multi-omiques rend l'annotation manuelle impossible. Des méthodes informatiques puissantes sont donc nécessaires pour élucider les relations dans un système biologique et tirer le meilleur parti des analyses multi-omiques [24]. Depuis une décennie, de nombreux progrès réalisés dans chacune des omiques ont contribué de façon importante à la faisabilité des expériences multi-omiques en permettant une couverture de molécules analysables plus large aussi appelée multiplexage, une automatisation accrue et une augmentation du débit d'analyse. Malgré ces innovations, aucune technique n'est capable d'analyser et d'identifier l'ensemble des molécules présentes dans chacune des omiques. Ces limitations font que seule une partie des analytes peut être observée, identifiée, annotée et quantifiée de manière convenable. Ces lacunes qui en résultent peuvent altérer considérablement le contexte biologique dans un système donné à cause de cette perte d'informations [24].

En conclusion, la multiplication des méthodes d'extraction, les capacités de multiplexage des méthodes analytiques, l'annotation fiable des analytes et le retraitement des données constituent des obstacles importants à la mise en place de méthodes multi-omiques pour une application pratique et de routine.

Afin de répondre au 1<sup>er</sup> axe de recherche de la biologie des systèmes, à savoir « Comment acquérir des données pour chaque omique qui constitue le système étudié ? », les aspects expérimentaux facilitant l'acquisition des données multi-omiques ont dû être optimisés. De ce fait, une attention particulière sera allouée à la préparation d'échantillon, la méthode de séparation chromatographique et les modes d'acquisitions en spectrométrie de masse dans le reste de ce manuscrit.

C'est dans ce contexte de la biologie des systèmes que la chimie analytique trouve sa place, notamment avec les analyses par spectrométrie de masse (MS). Son intérêt se dévoile à travers sa capacité à détecter et à quantifier des centaines de protéines, de lipides et de métabolites à partir d'un tissu, d'un fluide corporel ou d'une culture cellulaire. La versatilité de la spectrométrie de masse est aussi un avantage puisqu'elle permet de réaliser des études de deux manières différentes : sans *a priori* (méthode non ciblée) et avec *a priori* (méthode ciblée)

## 2. La spectrométrie de masse, un outil de choix pour les approches « omiques »

Depuis plus d'une dizaine d'années, la spectrométrie de masse tient une place centrale dans le cadre d'études omiques comme le montre le nombre de publications croissant associant le mot clé spectrométrie de masse à la protéomique, la lipidomique ou bien la métabolomique (cf. figure n°3). Cette technique analytique est particulièrement adaptée pour l'étude des sciences omiques, car elle permet d'identifier et de quantifier avec précision la présence d'analytes dans un mélange complexe avec une grande sensibilité. En effet, des molécules peuvent avoir le même rapport masse/charge ( $m/z$ ) malgré des structures chimiques très différentes. Cependant grâce aux profils isotopiques et à l'analyse de fragments caractéristiques, la composition élémentaire des molécules peut être identifiée aidant ainsi à l'annotation des composés.

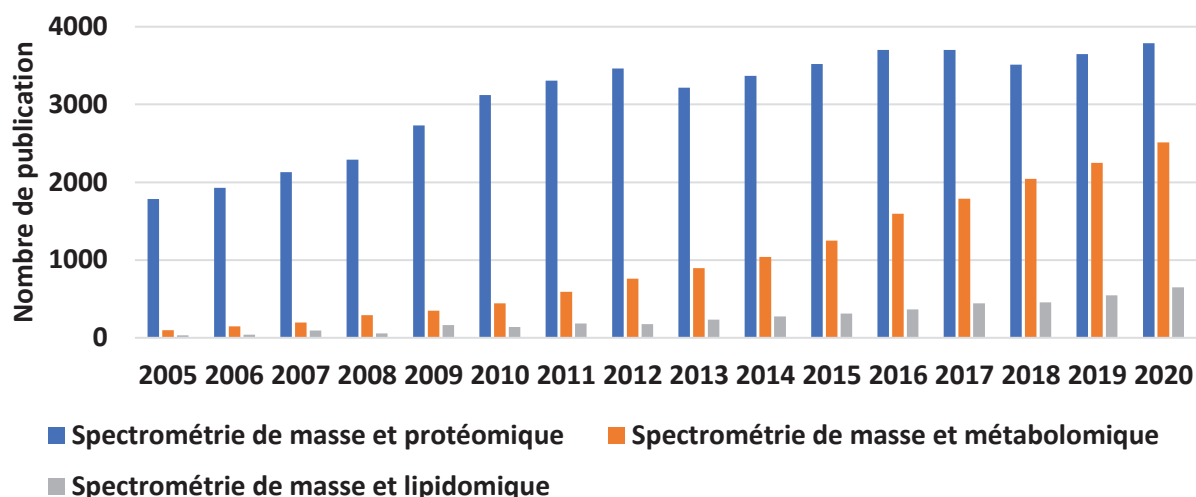


Figure 3 : Nombre de publications sur la période de 2005 à 2020 associant les mots clés Spectrométrie de masse à Protéomique ou Lipidomique ou Métabolomique répertoriées sur le site internet « web of science ».

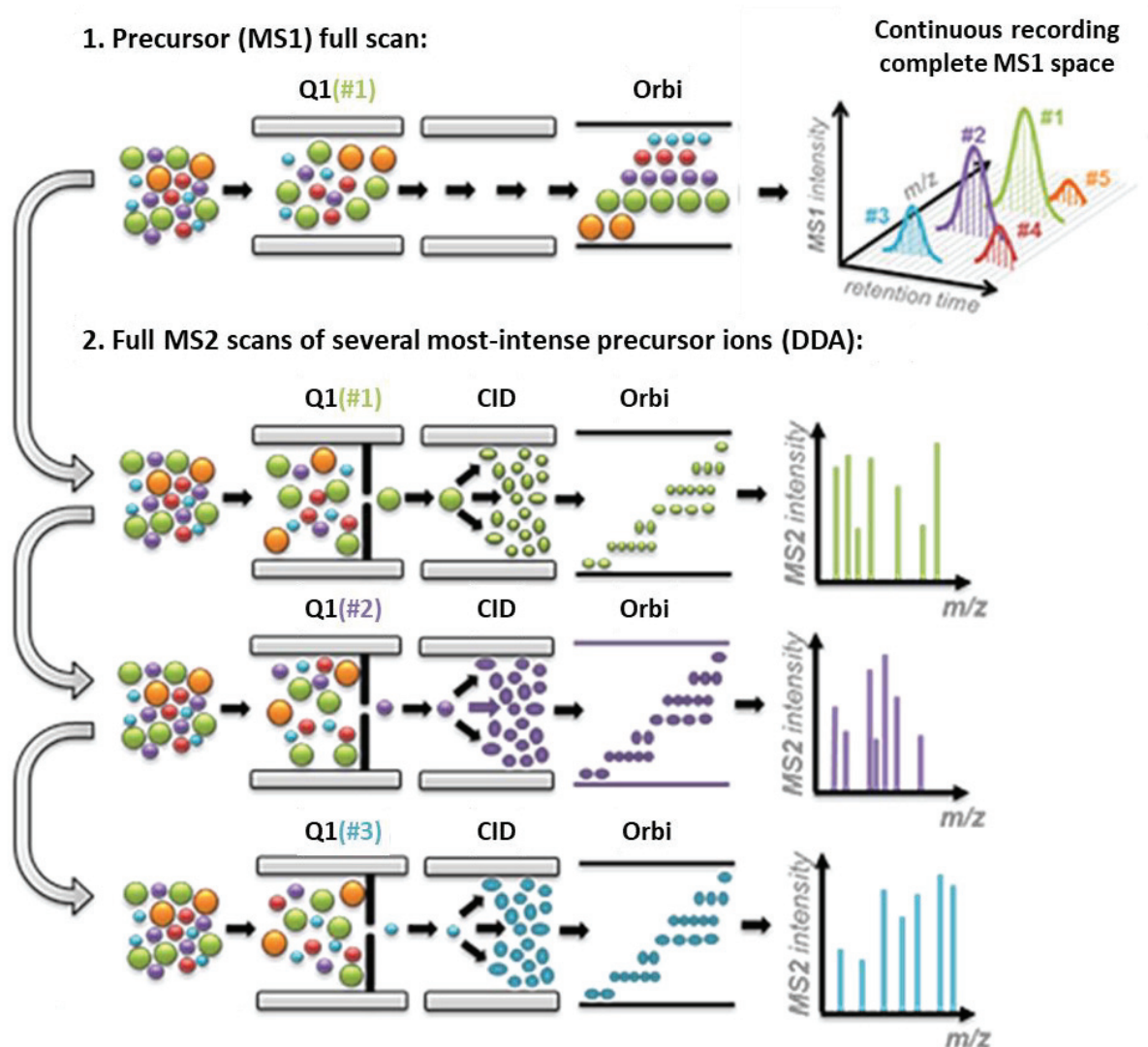
Cette technique est utilisée dans de nombreux domaines à des fins d'analyse structurale, de détermination précise des masses moléculaires et de quantification de biomolécules allant jusqu'à l'attomole par litre ( $10^{-18}$  mol/L) [28]. La spectrométrie de masse permet l'analyse d'un grand nombre de familles de molécules (protéines, lipides, métabolites, sucres, etc.) présentes dans des matrices très différentes (biofluides, objets du patrimoine, etc.). Elle présente également des capacités de multiplexage importantes permettant le suivi de nombreuses molécules en une seule analyse permettant d'avoir des couvertures de protéome et de métabolome conséquentes [28]. Il existe deux types d'approches en spectrométrie de masse à la fois différentes et complémentaires. Les approches sans *a priori* aussi appelées analyses globales et les approches avec *a priori* aussi appelées analyses ciblées.

## 2.1. Les approches sans *a priori*

Dans le cadre des analyses globales, les échantillons sont analysés sans critère de sélection et dans le but de détecter le plus grand nombre de molécules lors de l'analyse. Cette approche s'appuie sur des analyseurs haute résolution comme les temps de vol (TOF) ou la technologie Orbitrap pour déterminer avec une grande précision les masses des composés au sein d'un mélange complexe. Deux modes d'acquisition sont couramment employés : le mode d'acquisition dépendant des données (DDA, *data dependent acquisition*) et le mode d'acquisition indépendant des données (DIA, *data independent acquisition*).

### 2.1.1. Le mode d'acquisition DDA

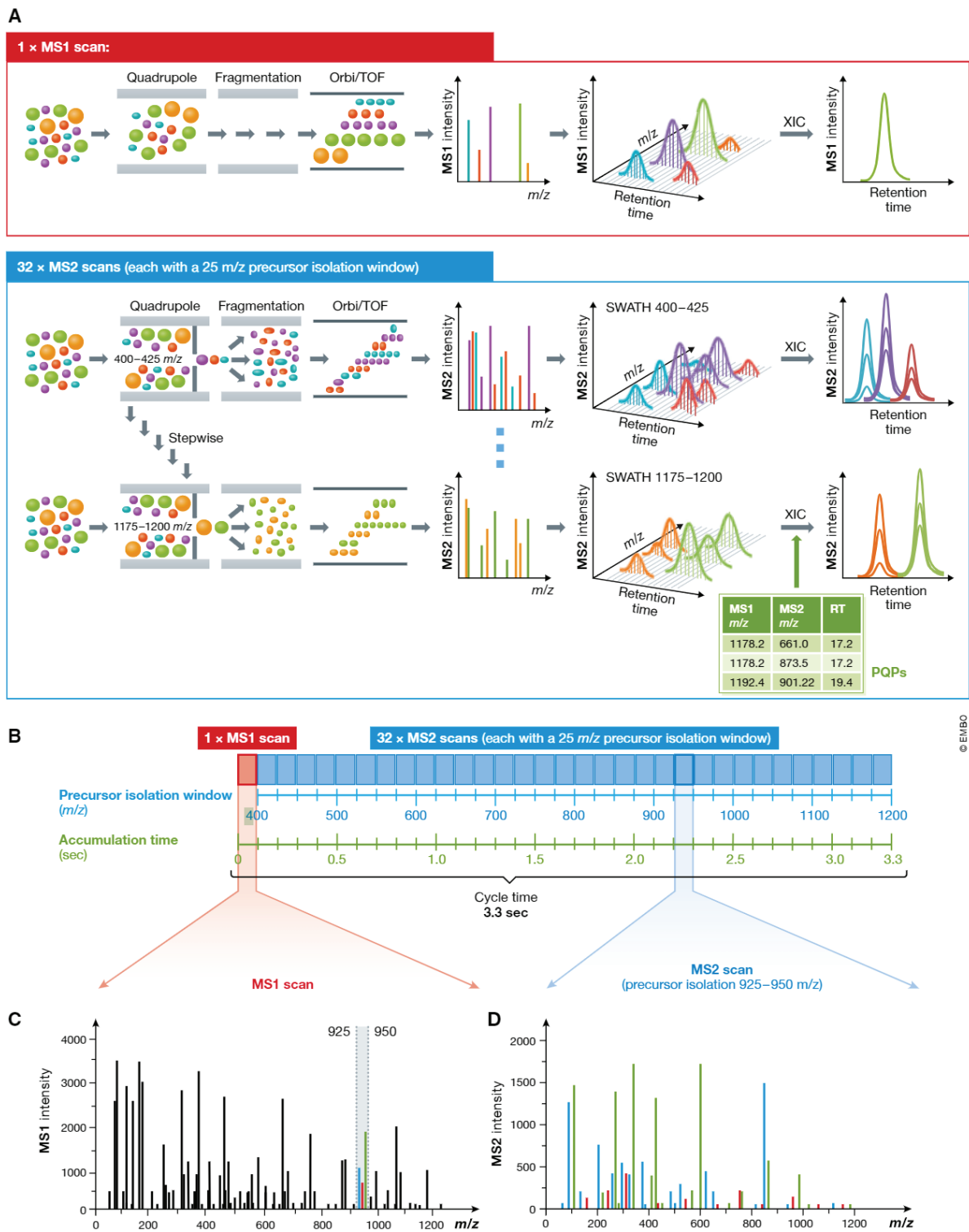
Le mode DDA est réalisé en deux étapes comme indiqué sur la figure n°4 [29]. La première étape consiste à réaliser un scan de surveillance en balayant les rapports  $m/z$  des ions précurseurs. Il est ainsi possible d'interroger les données MS *a posteriori*. Selon les critères définis par l'opérateur, les précurseurs les plus intenses sont ensuite fragmentés successivement dans la cellule de collision avant de réaliser un nouveau balayage de surveillance. On parle alors de Top « n » avec « n » le nombre de précurseurs fragmentés après le premier scan. Cette méthodologie implique le plus souvent 10 à 20 précurseurs (Top10 à 20). Afin d'augmenter le nombre de composés à fragmenter, le principe d'exclusion dynamique peut être associé à la méthode. Le biais le plus important que l'on peut rencontrer en analyse DDA est le sous-échantillonnage dans le cas où la gamme dynamique des molécules analysées représente plusieurs ordres de grandeur. Cela engendre une redondance de sélection des ions prépondérants ainsi qu'un phénomène de valeurs manquantes lorsque des ions moins abondants ne sont jamais sélectionnés ou de façon aléatoire [30]. Cette variabilité induit un manque de reproductibilité [31]. De plus, le manque de sensibilité lors de l'analyse en mode MS simple est également une limitation en mode DDA.



**Figure 4 : Principe des analyses en acquisition dépendante des données (DDA) avec un Top 3 comme exemple.** La première étape de ce mode d'acquisition est de réaliser un balayage de rapports  $m/z$  des ions précurseurs. Les 3 précurseurs les plus intenses sont ensuite fragmentés dans la cellule de collision ( $m/z$  pic vert, pic violet et pic bleu). Une fois les 3 précurseurs fragmentés un nouveau scan de surveillance est réalisé pour détecter de nouveau les 3 précurseurs les plus intenses (Adapté de la publication de Ludwig *et al.* [29]).

### 2.1.2. Le mode d'acquisition DIA

Le second mode utilisé lors des approches sans *a priori* est le mode DIA dont le principe est représenté sur la figure n°5 [29]. Comme pour le mode précédent, le spectromètre de masse réalise à instant donné un balayage de spectre de l'ensemble des précurseurs (*cf.* figure n°5 A). Lors de la seconde étape en DIA, le spectre de masse de l'ensemble des précurseurs est découpé en fenêtres dont le pas de rapport  $m/z$  est fixé par l'expérimentateur (fenêtres de 25  $m/z$  dans la figure n°5 B). Les précurseurs issus de ces fenêtres de masses (*cf.* figure n°5 C) sont envoyés successivement dans la cellule de collision afin d'obtenir des spectres de co-fragmentations (*cf.* figure n°5 D). L'utilisation de bases de données permet de réattribuer les ions fragments issus de la fragmentation des ions précurseurs.



**Figure 5 : Principe du mode DIA.** (A) En mode DIA, un seul spectre d'ions précurseurs ( $MS^1$ ) est enregistré, suivi d'une série de spectres d'ions fragments ( $MS^2$ ) avec de larges fenêtres d'isolement en précurseurs (ex. 25  $m/z$ ). Grâce à des cycles répétés de fenêtres d'isolement de précurseurs consécutives sur une plage de masse définie, un ensemble complet de données est enregistré. Cet ensemble comprend des informations continues sur tous les ions fragments et précurseurs détectables. Ainsi, des chromatogrammes d'ions extraits peuvent être générés au niveau  $MS^2$  et au niveau  $MS^1$ . (B) Le Q-TOF utilise 32 balayages  $MS^2$  avec des incréments définis de 25  $m/z$ , commençant de 400  $m/z$  à 1200  $m/z$ . Un balayage  $MS^1$  complet est enregistré au préalable. En appliquant un temps d'acquisition de 100 ms par balayage, on obtient un temps de cycle total de 3,3 s. (C) Le balayage complet  $MS^1$  détecte tous les précurseurs éluant à un moment donné. Par exemple, dans la gamme de masse de 925 à 950  $m/z$ , trois espèces co-éluées sont détectées (vert, rouge et bleu). (D) Le balayage  $MS^2$  correspondant à une fenêtre d'isolement des précurseurs de 925 à 950  $m/z$  représente un spectre  $MS^2$  mixte avec des fragments des trois espèces [29].

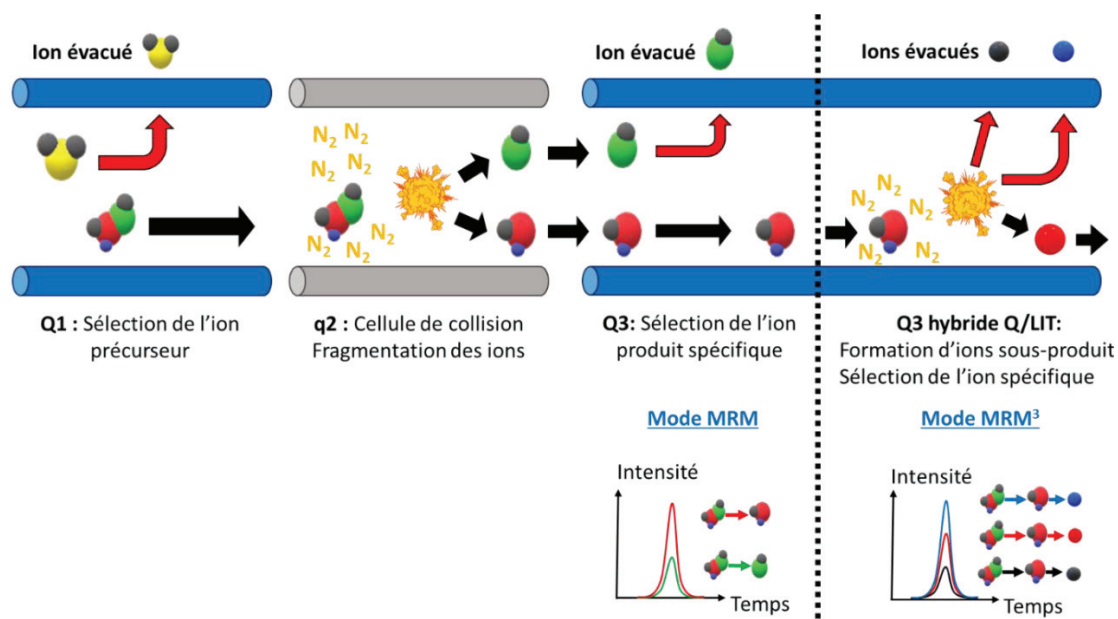
Contrairement au mode DDA, le mode DIA fragmente la totalité des composés ionisés puisqu'il n'y a plus de sélection des ions en fonction de leur intensité [29]. De plus, il est également possible d'interroger à la fois les données MS et MS/MS *a posteriori*. Cependant, en DIA les spectres de masse sont plus complexes à interpréter à cause des co-fragmentations impliquant des algorithmes et des bases de données conséquentes pour le retraitement des données [31,32]. On peut également retrouver le mode DIA sous l'appellation SWATH pour *Sequential Windowed Acquisition of All Theoretical fragment ions* chez Sciex®.

Les approches globales sont donc des méthodologies de choix pour l'identification et la caractérisation de composés dans des milieux complexes dans le cadre d'approches différentielles. En revanche, les appareils permettant de faire de la spectrométrie de masse de haute résolution possèdent des gammes dynamiques allant jusqu'à 4 décades ce qui est plus faible par rapport aux approches ciblées utilisant le mode *Multiple Reaction Monitoring* (MRM) qui eux atteignent 5 décades [29].

## 2.2. Les approches avec *a priori*

### 2.2.1. Le mode d'acquisition *Multiple Reaction Monitoring* (MRM)

À l'inverse des approches globales, seuls les composés d'intérêts sont ciblés dans les approches avec *a priori*. Les appareils de référence utilisés en spectrométrie de masse ciblée sont principalement de type triple quadripôles (QQQ) en utilisant le mode d'acquisition *Single Reaction Monitoring* (SRM) aussi appelé *Multiple Reaction Monitoring* (MRM) dans le cas où plusieurs transitions sont suivies. Comme indiqué sur la figure n°6, le mode MRM consiste à (i) sélectionner un ion précurseur dans le premier quadripôle Q1, (ii) fragmenter ce précurseur dans la cellule de collision q2 et (iii) sélectionner dans le 3<sup>ème</sup> quadripôle Q3, les ions produits du précurseur sélectionné en Q1 avant la détection par l'analyseur en masse. Une transition MRM représente le suivi du couple ion précurseur / ion produit illustré par un pic chromatographique. Bien que le mode MRM emploie des analyseurs basse résolution, la spécification de ce type de méthode est assurée par la double filtration en masse qui permet de simplifier l'échantillon avant le détecteur. Cependant, ce degré de spécificité peut ne pas être suffisant avec la persistance d'interférences isobariques dans des échantillons très complexes.



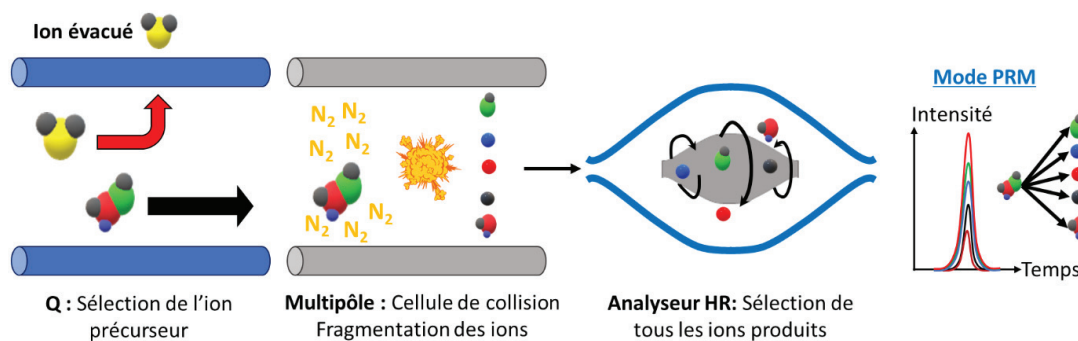
**Figure 6 : Principe de la MRM et la MRM<sup>3</sup>.** Dans le mode MRM, les ions sont filtrés dans le 1<sup>er</sup> quadripôle en fonction de leur  $m/z$  avant d'être envoyés dans le deuxième quadripôle qui est utilisé comme cellule de collision. Pour terminer, les ions fragments issus de la fragmentation d'un composé d'intérêt dans la cellule de collision sont filtrés dans un troisième quadripôle avant d'être détectés. Le mode MRM<sup>3</sup> fonctionne de la même manière que le mode MRM, mais les ions filtrés dans le 3<sup>ème</sup> quadripôle hybride sont isolés, refragmentés puis filtrés avant détection.

Des développements instrumentaux tels que la *Multiple Reaction Monitoring cubed* (MRM<sup>3</sup>) développée par l'équipe Anabio-MS en 2009 permettent de gagner un degré de spécificité supplémentaire [33]. À la différence de la MRM, Le 3<sup>ème</sup> quadripôle hybride permet de piéger les ions dans la trappe d'ion linéaire à injection axiale (Q/LIT). Comme indiqué dans la figure n°6, en MRM<sup>3</sup> les ions fragments spécifiques de l'ion précurseur sont refragmentés dans le Q3 avant de subir un nouveau balayage des ions sous-produits. Cette triple sélectivité conduit à l'augmentation du rapport signal/bruit permettant de gagner en sensibilité tout en limitant le risque d'interférence. Cependant, ce mode d'acquisition n'est pas approprié pour le développement de méthodes hautement multiplexées à cause du temps d'accumulation et d'excitation dans la trappe qui limite la fréquence d'acquisition. En effet, le temps d'acquisition en mode MRM classique pour une transition est d'environ 10 ms contre 350 ms en MRM<sup>3</sup> [33].

## 2.2.2. Le mode d'acquisition *Parallel Reaction Monitoring* (PRM)

Depuis une dizaine d'années, les spectromètres de masse de haute résolution sont très utilisés dans le cadre de suivi de molécules *via* le mode d'acquisition *Parallel Reaction Monitoring* (PRM) de chez Thermo Scientific®. On peut également le retrouver sous l'appellation MRM haute résolution ou MRM-HR chez Sciex®. Ce mode d'acquisition développé par l'équipe de B. Domon [34,35] diffère du mode MRM après l'étape de sélection du composé d'intérêt et de sa fragmentation dans la cellule de collision. En PRM l'ensemble des fragments issus du précurseur sont analysés simultanément comme représenté sur la figure n°7.





**Figure 7 : Principe de la PRM.** Dans le mode PRM, les ions sont filtrés dans le 1<sup>er</sup> quadripôle en fonction de leur  $m/z$  avant d'être envoyés dans le multipôle qui est utilisé comme cellule de collision. Pour terminer, l'ensemble des ions fragments issus de la fragmentation d'un ion d'intérêt sont analysés dans un analyseur de haute résolution.

L'utilisation de la haute résolution permet de différencier les différents rapports de  $m/z$  et de s'assurer qu'ils proviennent du bon composé à quantifier. Il n'y a donc pas besoin de réaliser une deuxième filtration en masse avant sa détection. Le suivi de toutes les transitions par composé permet d'augmenter la sélectivité de la méthode lors de l'analyse d'échantillon complexe en limitant les erreurs dues à la présence de potentiels interférents. Les modes MRM et PRM ont été souvent comparés dans la littérature notamment en protéomique [36–38] en démontrant des performances comparables en termes de sensibilité, de linéarité et de justesse. Comme le mode PRM permet de suivre la totalité des transitions pour un même composé, une grande spécificité est attendue [37]. Il est possible en MRM de suivre la totalité des fragments comme en PRM, mais cela prendra trop de temps puisque chaque transition ne peut être détectée qu'individuellement. La sensibilité supérieure en MRM a conduit la communauté scientifique à la préférer vis-à-vis de la PRM [38] bien que les gains soient souvent faibles.

### 2.3. Quelles sont les limites des méthodes MRM en termes de capacité de multiplexage et comment les améliorer ?

La sensibilité et la robustesse des méthodes ciblées utilisant le mode MRM en font une méthodologie de choix dans la mise en place de méthodes de quantification. Que ce soit en lipidomique, métabolomique ou en protéomique, il est sans cesse nécessaire de développer des méthodes permettant de doser de plus en plus de molécules en une seule analyse. C'est pour cela que dans cette partie nous nous sommes intéressés aux limites de multiplexage des méthodes MRM et aux différentes stratégies déployées pour augmenter la capacité de multiplexage.

#### 2.3.1. Évaluation des capacités de multiplexage en mode MRM

Le développement de méthodes MRM hautement multiplexées nécessite d'optimiser des paramètres clés pour augmenter la vitesse d'acquisition des appareils de type quadripolaire sans diminuer la qualité du résultat analytique (nombre de points par pic et rapport signal/bruit). Le premier paramètre



est le *dwell time* qui correspond au temps de suivi d'une seule transition alloué par le spectromètre de masse. Le deuxième paramètre est le temps de cycle qui correspond à la durée nécessaire pour observer la totalité des transitions présentes dans la méthode. Le dernier paramètre est le temps de repos qui traduit la durée nécessaire pour vider le deuxième quadripôle entre deux transitions MRM. Le temps de repos correspond au *dwell time* minimal permis par le spectromètre de masse (5 ms pour un QTrap 5500 de chez Sciex®).

Dans le cadre d'analyse multiplexée, plus le nombre des transitions MRM par méthode est élevé, plus le temps de cycle augmente comme le montre la formule suivante, qui est applicable pour un *dwell time* identique pour chaque transition :

- **Temps de cycle = (Dwell time + Temps de repos) x Nombre de transitions**

Cependant, le *dwell time* est limité par les capacités de balayage des appareils. Ceci implique qu'il existe un *dwell time* minimal, qui varie de 1 à 5 ms. Il est important de noter que plus la valeur de *dwell time* sera petite, plus le rapport signal/bruit (S/B) sera faible et plus la sensibilité sera diminuée.

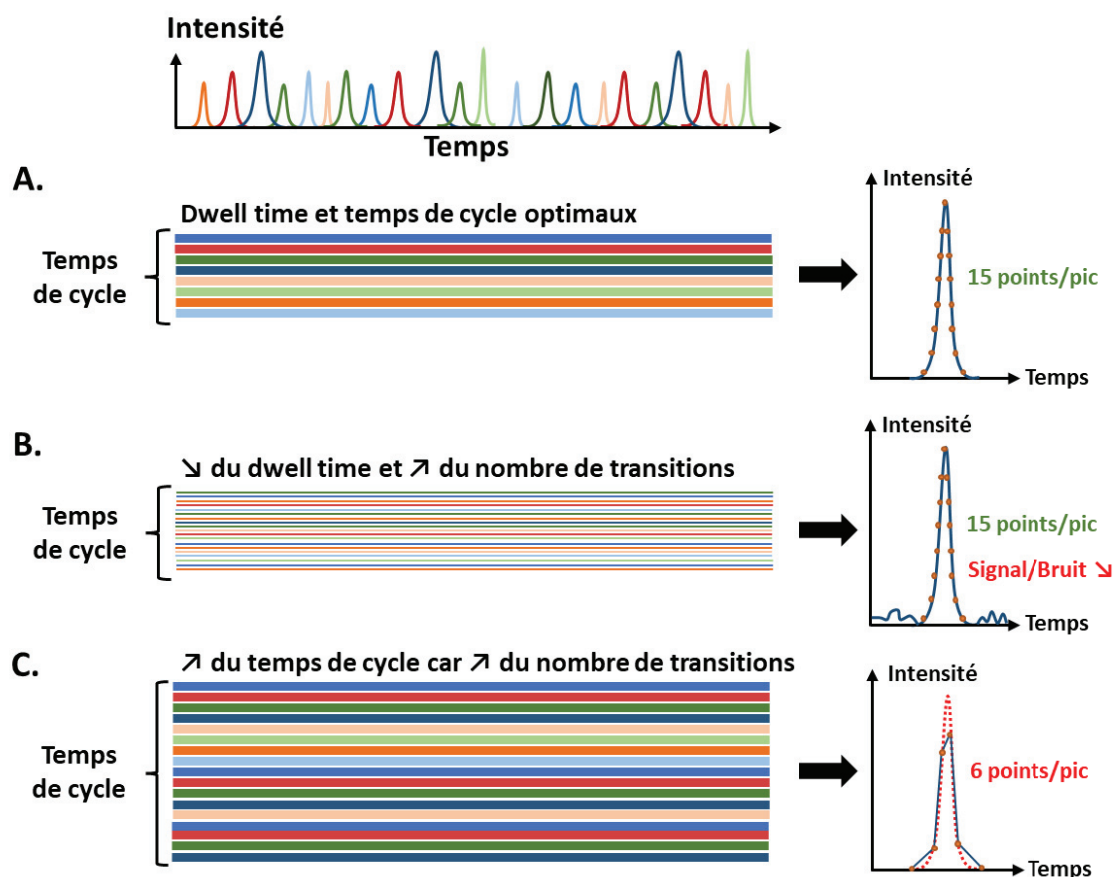
Afin de déterminer les capacités de multiplexage maximales, on peut calculer le nombre maximal de transitions observables par méthode dans des conditions données. Par exemple, pour une largeur de pic moyenne de 15 s et un *dwell time* minimum de 5 ms, on souhaite avoir 12 points par pic. Le nombre de points minimum par pic chromatographique doit être compris entre 10 et 15 afin d'assurer une bonne quantification. Cela nous donne un temps de cycle de 1.25 seconde selon la relation :

- **Nombre de points = Largeur de pics / Temps de cycle**

On divise le temps de cycle (1.25 s) par le *dwell time* (5 ms) et le temps de repos (5 ms) pour obtenir le nombre de transitions observables dans la méthode. Ainsi la méthode peut contenir, dans ces conditions, un maximum de 125 transitions. Les conditions généralement adoptées permettent d'obtenir entre 100 et 150 transitions MRM par méthode.

Comme le montre la figure n°8, l'optimisation d'une méthode MRM couplée à la chromatographie liquide consiste à trouver le bon équilibre entre *dwell time*, temps de cycle et nombre de transitions MRM. Dans cette figure, le premier cas (cf. figure n°8 A) met en évidence la situation optimale dans laquelle le nombre de transitions et le *dwell time* induisent un temps de cycle permettant d'avoir suffisamment de points pour définir le pic chromatographique. Ces conditions induisent également des *dwell time* suffisamment grands pour avoir un signal/bruit acceptable. La figure n°8 B indique que si l'on souhaite augmenter le nombre de transitions MRM sans augmenter le temps de cycle pour conserver le même nombre de points par pic chromatographique, il est nécessaire de diminuer le *dwell time*. Cela se traduira par une diminution du rapport S/B. Le troisième cas (cf. figure n°8 C) indique que

si l'on souhaite augmenter le nombre de transitions sans perdre en sensibilité en gardant un *dwell time* identique, cela se traduira par l'augmentation du temps cycle et la diminution du nombre de points par pic chromatographique. Cette diminution du nombre de points peut entraîner des biais dans la quantification si la totalité de l'aire réelle du pic chromatographique n'est pas intégrée.



**Figure 8 :** Évolution de la capacité de multiplexage en fonction des paramètres de masse. Les lignes horizontales représentent les fenêtres de détection de chaque transition MRM dans la méthode. La largeur de ces lignes représente le *dwell time* appliqué pour chaque transition MRM. Plus la ligne est épaisse plus le *dwell time* est important. La somme des largeurs de ces lignes représente le temps de cycle de la méthode. (A) Dans la partie supérieure, on se trouve dans les conditions optimales avec un nombre de points par pic suffisant et un bon rapport S/B. (B) La partie centrale représente une méthode où le nombre de transitions augmente avec une diminution du *dwell time* afin de conserver le même temps de cycle que les conditions optimales. On a donc un nombre suffisant de points par pic, mais une diminution du rapport S/B. (C) La partie inférieure représente une méthode où le nombre de transitions augmente sans diminuer le *dwell time* ce qui implique une augmentation du temps de cycle. On a donc moins de points par pic chromatographique (adapté de la thèse de Blandine Rougemont [39]).

### 2.3.2. Capacité de multiplexage du mode MRM versus PRM

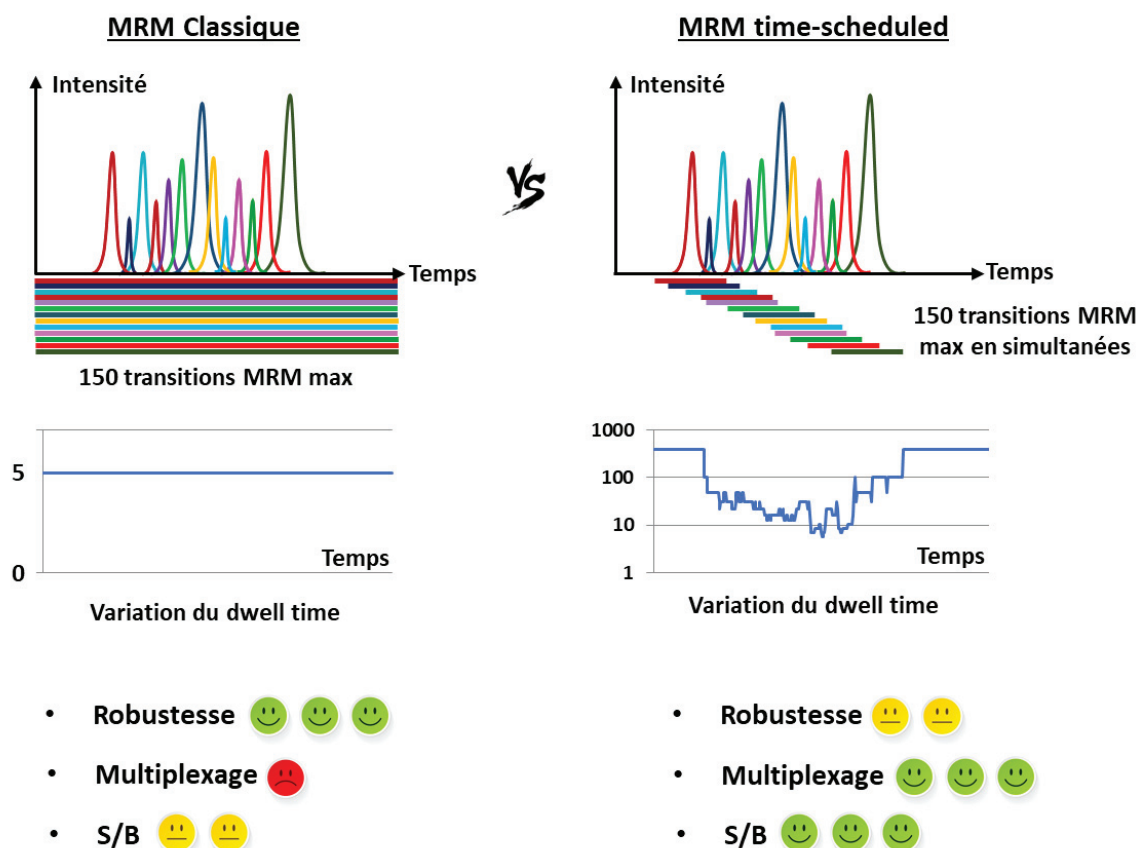
L'une des faiblesses du mode PRM est le faible taux de balayage qui limite la capacité des expériences MS/MS lorsqu'elles sont réalisées à grande échelle [40]. Les récents progrès technologiques ont permis d'augmenter ce taux de balayage pour les appareils de type Q-TOF afin de permettre le multiplexage d'un grand nombre de précurseurs [41]. Par exemple, le spectromètre de masse TripleTOF® 5600 peut fournir jusqu'à 100 expériences MS/MS par seconde [41]. Cependant, par rapport à la MRM qui peut réaliser plusieurs centaines d'expériences MS/MS en accord avec le nombre de points par pic

chromatographique (*dwell time* de 1 à 5 ms), la capacité de multiplexage en PRM est encore inférieure dans les mêmes conditions [42]. Cela est seulement vrai dans le cas où l'on accorde plus d'importance au nombre de précurseurs par rapport au nombre de fragments par composé (par exemple lorsque 3 transitions suffisent dont une quantifiante et deux qualifiantes). Dans le cas où il est nécessaire d'avoir la totalité des fragments par peptide, la PRM sera plus appropriée. En effet, l'analyseur TOF compte le nombre d'ions de n'importe quel rapport  $m/z$  dans chaque mesure alors que le quadripôle ne compte que les ions d'un seul rapport  $m/z$  dans chaque mesure. Ainsi, dans le cas d'une analyse hautement multiplexée d'un échantillon complexe il faudra plus de temps au QQQ pour balayer la totalité des rapports  $m/z$  que pour un TOF.

### 2.3.3. Mise en place de méthodes avec fenêtres de temps (Scheduled-MRM)

Le mode MRM permet de suivre entre 100 et 150 transitions MRM par méthode en fonction des conditions chromatographiques et des différents paramètres cités précédemment. Dans le cas où l'on utilise 3 transitions par composé pour plus de spécificité cela représente le dosage simultané de 30 à 50 composés. Pour augmenter le nombre de composés analysés par méthode MRM, il est nécessaire de mettre en place des méthodes avec des fenêtres de détection pour l'acquisition des rapports  $m/z$  au sein des zones d'intérêt du chromatogramme. Le mode permettant de mettre en place des fenêtres de détection avec le mode MRM possède plusieurs noms commerciaux : Scheduled-MRM chez Sciex<sup>®</sup>, Dynamic-MRM chez Agilent<sup>®</sup> et Time-MRM chez Thermo Scientific<sup>®</sup>. Pour plus de simplicité, ce mode d'acquisition sera appelé méthode *scheduled* ou sMRM. Comme son nom l'indique, la sMRM utilise des fenêtres temporelles qui enregistrent les signaux des transitions seulement dans la zone d'élution des composés d'intérêt (*cf.* figure n°9). Ce fenêtrage autour des temps de rétention permet d'augmenter les capacités de multiplexage de la méthode en augmentant le nombre de transitions maximales. Puisque le spectromètre de masse n'est plus obligé de suivre l'ensemble des transitions pendant la totalité de l'analyse, de nouvelles transitions MRM sont suivies. En mode *scheduled*, le temps de cycle est fixe pendant toute l'analyse. De ce fait, le *dwell time* est ajusté automatiquement à chaque fenêtre de temps de rétention de manière à avoir le *dwell time* le plus grand tout en respectant le temps de cycle de la méthode. Cela permet de gagner en sensibilité en améliorant le rapport signal/bruit par rapport à une méthode MRM. Les méthodes *scheduled* nécessitent de connaître précisément les temps de rétention des composés d'intérêts dans des conditions séparatives données. Cette limitation implique d'avoir une méthode robuste pour ne pas perdre d'informations dues au décalage des temps de rétention. Les moindres variations analytiques comme l'utilisation de différentes configurations de chaînes chromatographiques, les différences de débits appliqués, la longueur et la largeur de la colonne impliquent des réajustements au niveau de ces fenêtres. Cette problématique des décalages en temps de rétention sera rediscutée plus précisément dans le chapitre

2 du manuscrit. La méthodologie utilisée en sMRM peut également être appliquée au mode PRM [40,43,44].



**Figure 9 : Différence entre la MRM et la sMRM.** En MRM, les transitions sont suivies pendant tout le chromatogramme ce qui réduit la capacité de multiplexage et limite le rapport signal/bruit. En effet, le *dwell time* est fixe dans le temps avec des valeurs faibles dans le cas d’une méthode hautement multiplexée. La robustesse est importante, car ce mode est indépendant des temps de rétention. En sMRM, les transitions sont réparties dans des fenêtres de temps ce qui augmente la capacité de multiplexage ainsi que le rapport signal/bruit, car le *dwell time* est variable dans le temps en fonction du nombre de transitions superposées. La robustesse est plus faible, car ce mode est dépendant des temps de rétention qui peuvent varier.

### 2.3.4. Un nouveau mode d’acquisition : Scout-MRM

Un nouveau mode d’acquisition a été récemment développé permettant d’augmenter drastiquement les capacités de multiplexage des méthodes MRM appelé Scout-MRM [45–47]. Ce nouveau mode d’acquisition développé en 2016 est le fruit d’une collaboration entre l’équipe Anabio-MS et la société Sciex®. La particularité de ce mode d’acquisition consiste à s’affranchir des décalages en temps de rétention *via* la détection de composés exogènes ou endogènes appelés scout. Lorsque l’intensité de la transition MRM d’un scout dépasse un certain seuil, un groupe de transitions non ordonnées de composés qui lui est attribué est déclenché. L’acquisition de ces transitions se poursuit jusqu’à ce qu’un autre scout soit détecté et ainsi de suite. La méthode Scout-MRM est donc une évolution de la sMRM vers un mode dynamique qui s’affranchit des décalages en temps de rétention des composés. Le principe du mode Scout-MRM et son utilisation seront approfondis dans le chapitre 2 du manuscrit.

À ce jour, la méthodologie Scout-MRM a été développée sur les appareils de type QTrap de la société Sciex (QTRAP 6500, QTRAP 6500+ et QTRAP 7500).

### 2.3.5. Quels sont les plus grands multiplexes en approche ciblée dans les différentes sciences omiques ?

Dans le domaine de la santé, de nombreuses méthodes à haut degré de multiplexage ont été mises en place pour le dosage ciblé de protéines par spectrométrie de masse. Dans le cas de la scheduled-PRM (sPRM), Schilling *et al.* [48] ont développé une méthode permettant l'analyse simultanée de 466 peptides et 2694 transitions chez la bactérie *E. coli*. En sMRM Burgess *et al.* [49] sont capables de suivre 800 peptides avec un total de 2 400 transitions dans le plasma humain. Pour finir, Rougemont *et al.* [47] ont mis en place le plus grand multiplexe à ce jour en Scout-MRM avec le suivi en simultané de 782 peptides et un total de 2346 transitions MRM réparties dans 19 groupes. Pour atteindre de telles capacités de multiplexage, il est nécessaire d'avoir de longs gradients chromatographiques pour répartir les composés le long du chromatogramme (> 1h). Cela limite au maximum la superposition des fenêtres de détection et permet d'avoir le plus de groupes scouts possible. L'utilisation de la spectrométrie de masse comme outil de mesure en protéomique avec des approches ciblées est très récente en écotoxicoprotéomique. En 2018, le plus grand multiplexe dans le cas d'analyses protéomiques d'espèces sentinelles a été développé par Charlot *et al.* avec le suivi de 71 peptides et 213 transitions en sMRM [50]. Nous avons tiré profit du mode d'acquisition Scout-MRM pour augmenter les capacités de multiplexage des méthodes utilisées en écotoxicologie (*cf.* chapitre 2).

Dans le cas des analyses de lipides, Zhou *et al.* ont élaboré une méthode à partir d'un appareil Q-TOF pour suivre 222 espèces moléculaires à travers 22 classes de lipides dans du sérum humain grâce au mode PRM [51]. Il s'agit de l'un des plus grands multiplexes de lipides en mode PRM. L'un des plus importants multiplexes en mode MRM a été mis en place par Weir *et al.* [52] avec l'analyse de 312 lipides répartis dans 23 familles lipidiques. L'ensemble des transitions de leur méthode est réparties en 2 méthodes sMRM avec d'une part les glycérolipides et d'une autre le reste des familles de lipides. Pour atteindre un tel degré de multiplexage, les auteurs ont utilisé des fenêtres de temps de  $\pm 30$  secondes pour encadrer l'élution des lipides. Cela laisse une faible marge de manœuvre dans le cas de potentiel décalage en temps de rétention. Dans le cas des analyses en lipidomique, il existe peu d'articles scientifiques illustrant de très grands multiplexes avec le mode MRM. En général, les études s'appliquent uniquement à quelques classes de lipides spécifiques ou même parfois une seule par injection [53,54]. Par ailleurs, les analyses reposent majoritairement sur l'utilisation d'appareils de spectrométrie de masse haute résolution. La principale limite des approches MRM est que le suivi des transitions spécifiques exclut l'enregistrement et le retraitement des données post-acquisition des lipides non identifiés [55].

À ce jour, le plus grand multiplexe dans le cas des analyses ciblées en métabolomique a été réalisé par Gonzalez-Dominguez *et al.* [56] dans une étude réalisée en 2020. L'objectif de cette étude est de réaliser, pour la première fois, le développement d'une méthode en métabolomique à haut débit pour mesurer l'exposome humain de façon complète et quantitative. Pour cela, les auteurs ont pris en compte les métabolites endogènes impliqués dans les voies métaboliques essentielles comme la glycolyse et le cycle de Krebs, les métabolites liés à l'alimentation, les médicaments couramment consommés, *etc.* La méthode permet de quantifier 1019 métabolites présents dans le plasma et l'urine chez l'Homme. Ce multiplexe est d'autant plus complet que, pour plus de 98% des composés, 2 transitions MRM sont utilisées alors que la majorité des multiplexes en métabolomique n'utilise qu'une seule transition MRM. Au total, 458 composés sont analysés avec une source d'ionisation électrospray en mode positif (ESI (+)) soit 916 transitions MRM et 561 composés en mode négatif (ESI (-)) soit 1098 transitions MRM. Ce degré de multiplexage est atteint grâce à la technologie sMRM avec des fenêtres de détection comprises entre 50 et 100 secondes. Ces fenêtres de temps rendent cette méthode particulièrement sensible aux décalages en temps de rétention.

Une fois le mode d'acquisition sélectionné, il est nécessaire de se poser des questions sur le système d'introduction de l'échantillon le plus adapté pour l'analyse d'échantillon complexe dans le spectromètre de masse. L'approche la plus simple est de se reposer sur les capacités de discrimination des spectromètres de masse en introduisant l'échantillon directement dans l'appareil.

### 2.4. Analyse par infusion directe

L'approche la plus simple pour commencer la caractérisation d'un échantillon inconnu consiste à effectuer une analyse globale par spectrométrie de masse de haute résolution (HRMS) [53,57–59]. L'une des méthodes consiste à infuser l'échantillon de façon continue directement au niveau de la source du spectromètre de masse sans séparation chromatographique en amont. Cette technique dénommée infusion directe (DI/MS) s'appuie sur la capacité du spectromètre de masse haute résolution à séparer les composés en fonction de leurs rapports  $m/z$ . Elle s'emploie couramment avec des spectromètres de masse tels que l'Orbitrap ou les TOF qui possèdent une très grande résolution permettant de différencier des composés ayant des rapports  $m/z$  proches. Cependant, des phénomènes de suppression d'ions peuvent se produire *via* l'ionisation simultanée de nombreux composés et risquent d'engendrer une perte de sensibilité. Cela est dû essentiellement au changement de propriété des gouttelettes par les interférents lors du passage de la phase liquide à gaz dans la source ESI [60]. En effet, des espèces moins volatiles que le composé d'intérêt peuvent modifier l'efficacité de formation des gouttelettes et perturber l'évaporation de l'analyte [61]. Les espèces moins volatiles peuvent ainsi entraîner la coprécipitation de l'analyte diminuant le nombre d'ions transmis jusqu'au détecteur [62]. L'autre facteur limitant de ce mode d'analyse est son incapacité à

séparer des composés isobares. Par exemple, les lipides qui possèdent de nombreuses formes isomères de même masse, comme les lysophospholipides dont la position de la chaîne d'acides gras peut se trouver en sn-1 ou sn-2, ne peuvent pas être différenciés (*cf.* chapitre 3 partie 3.2) [63]. Ce mode DI/MS est très utilisé en analyse lipidomique, car ce type de quantification a l'avantage, par rapport à la LC-MS, d'avoir un environnement stable d'ionisation qui n'est pas influencé par les variations de phase mobile ou de matrice [58]. De plus, cette approche est simple à mettre en place et nécessite peu d'optimisation.

Pour éviter ces contraintes et ainsi améliorer les limites de détection, il faut utiliser un système chromatographique tel que la chromatographie liquide (LC) ou la chromatographie gazeuse (GC) couplée au système MS afin de décomplexifier l'échantillon en amont de la détection [55]. D'autant plus que l'approche DI/MS pour les échantillons complexes est difficilement compatible avec les spectromètres de masse de faible résolution.

### **3. Diminution de la complexité d'un échantillon grâce à une dimension de séparation en amont du spectromètre de masse**

#### **3.1. Techniques séparatives en métabolomique polaire**

La chromatographie en phase liquide est le mode de séparation le plus utilisé pour l'analyse du métabolome polaire, car il est directement compatible avec l'analyse d'échantillons aqueux [64] et qu'il ne nécessite pas d'étapes de dérivation pour rendre volatils les composés qui ne le sont pas. Cette condition est indispensable en chromatographie en phase gazeuse (GC) [18]. L'utilisation de la chromatographie à polarité de phase inversée (RPLC) est limitée lors de l'analyse d'espèces polaires et/ou ioniques en raison de la faible rétention de ces molécules. D'autres mécanismes de séparation sont nécessaires pour assurer une rétention suffisante et donc la bonne résolution des analytes polaires. Une meilleure rétention permet également de diminuer les effets matrices. Ainsi, la chromatographie d'interaction hydrophile (HILIC) ou la chromatographie liquide en phase normale (NPLC) peuvent être utilisées. Cependant, la HILIC reste le mode chromatographique le plus utilisé que ce soit en écotoxicologie [65,66] ou lors des études cliniques [67–69]. Cela est dû au couplage difficile entre la phase normale et les sources électrospray des spectromètres de masse. Les phases mobiles courantes en NPLC contiennent de grandes quantités de solvants de faible polarité tels que l'hexane ou l'heptane. Ces solvants ne permettent pas de former des sprays stables pour favoriser la formation d'analytes chargés en phase gaz lors de l'étape d'ionisation en raison de leur faible conductivité, de leur tension superficielle et de leur constante diélectrique [70].



La HILIC est souvent considérée comme une variante de la NPLC puisqu'elle utilise les mêmes phases stationnaires. En revanche, les phases mobiles sont très différentes de celles employées en NPLC et se rapprochent plus de celles utilisées en RPLC. Dans la littérature, les séparations chromatographiques en mode HILIC sont souvent réalisées avec des phases stationnaires à base de silices vierges ou greffées avec des groupements polaires (amine, amide, diol, cyano, *etc.*) [71]. Les phases mobiles sont composées d'un mélange de solvant organique (généralement de l'acétonitrile) et d'eau [71]. Le mécanisme de rétention principal est encore sujet à discussion, mais il est admis que les molécules d'eau de la phase mobile agissent avec les groupements polaires de la phase stationnaire et forment après un certain temps d'équilibre une couche aqueuse à la surface de cette même phase. Ainsi, un analyte polaire dissout dans la phase mobile est partagé entre la couche aqueuse semi-immobilisée et la phase mobile contenant une composition aqueuse. Plus l'analyte est hydrophile, plus l'équilibre de partage est déplacé vers la couche d'eau de la phase stationnaire, et plus l'analyte est retenu [72]. Inversement, lors d'un gradient chromatographique, plus la composition de la phase mobile aqueuse augmente, plus les composés hydrophiles sont transmis de la couche d'eau semi-immobilisée à la phase mobile pour être élués. Toutefois, il existe d'autres mécanismes de rétention théorisés qui peuvent rentrer en jeu, notamment les liaisons hydrogène entre les espèces polaires neutres, ainsi que les mécanismes électrostatiques et les interactions de Van der Waals de type dipôle-dipôle. Cela indique que le mécanisme de la HILIC se distingue significativement de la NPLC [72] puisque dorénavant la polarité de l'analyte n'est pas le seul facteur à prendre en compte pour expliquer la rétention comme c'est majoritairement le cas en NPLC et RPLC.

Il est important de souligner que la HILIC possède tout de même des inconvénients comme le temps d'équilibrage pour former la couche d'eau à la surface de la phase stationnaire. Ce temps d'équilibre est généralement 2 à 4 fois plus long qu'en RPLC [73]. D'après Buszewski *et al.* la chromatographie HILIC possède également une mauvaise répétabilité des temps de rétention comparée à la RPLC [72]. De plus, il est souvent nécessaire d'ajouter un tampon dans les phases mobiles (*ex.* formiate d'ammonium) pour affiner les pics chromatographiques [74,75]. Ces sels doivent être volatils pour être compatibles avec la spectrométrie de masse. Toutefois, l'utilisation de tampons salins peut entraîner une diminution du signal en ESI qui augmente avec la concentration en sels (gamme de concentration couramment utilisée en HILIC : 5-50 mM). Il a été démontré que 5 mM de formiate d'ammonium peut provoquer une plus grande suppression du signal pour les produits pharmaceutiques acides et basiques par rapport à l'utilisation de simples solutions d'acide formique à 0,1% [74].

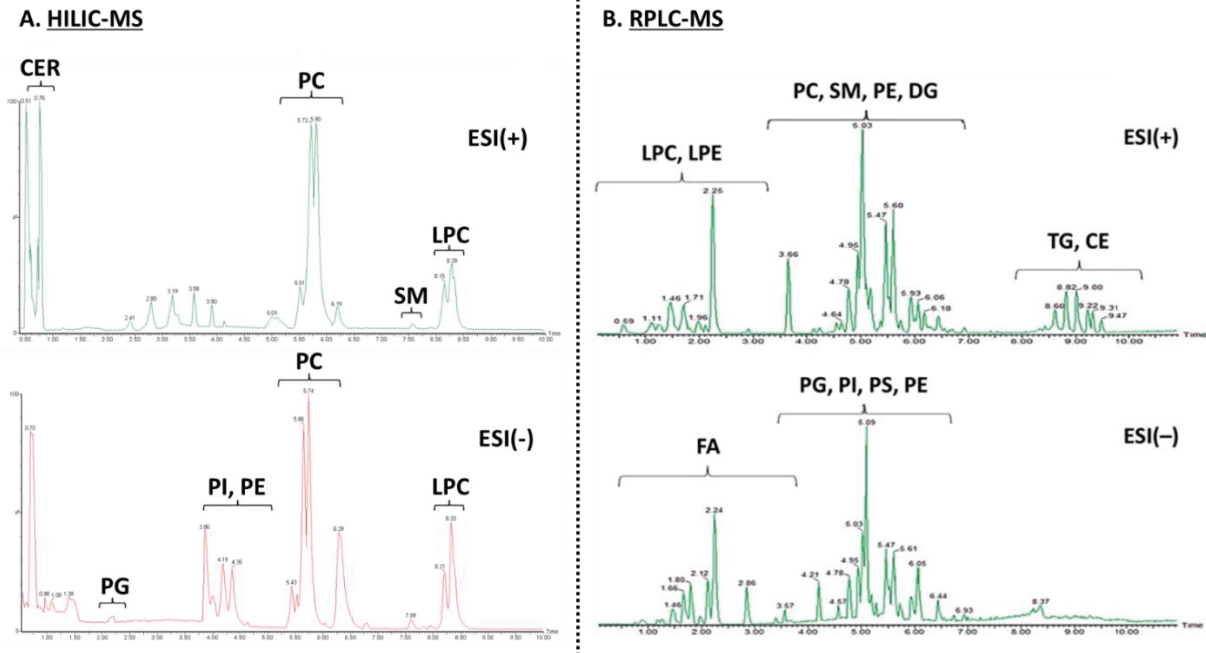
Malgré cela, la HILIC présente des avantages certains lors de la mise en place d'une étude en métabolomique focalisée sur les métabolites grâce à sa capacité de séparation des composés polaires et sa compatibilité avec la spectrométrie de masse.



### 3.2. Techniques séparatives en lipidomique

Bien que la chromatographie en phase gaz soit encore utilisée en lipidomique, notamment pour l'analyse des chaînes d'acides gras libres ou extraites à partir de lipides intacts [76,77], la chromatographie liquide lui est souvent préférée lorsque c'est possible. En effet, La GC est une technique limitante, car les composés doivent être thermiquement stables avec une pression de vapeur suffisamment élevée pour se volatiliser pendant l'injection. Les méthodes de dérivation chimique (méthylation, silylation, *etc.*) ou l'estérification par catalyse acide peuvent résoudre ce problème de volatilité [77–79]. Toutefois, cette réaction chimique peut modifier la reproductibilité de l'analyse en raison d'une efficacité de dérivation différente pour chaque lipide. La dérivation implique également de multiples étapes de réaction augmentant le temps d'analyse global et le risque de perte d'échantillon lors de la procédure [78].

Les trois principaux modes de séparation utilisés en LC-MS appliqués à la lipidomique sont la chromatographie de phase normale (NPLC), la chromatographie d'interaction hydrophile (HILIC) et la chromatographie liquide à polarité de phases inversée (RPLC) [53,55]. Les deux premières reposent sur la séparation des lipides en fonction de leur tête polaire. Ainsi les composés les plus polaires sont les plus retenus sur la colonne. Par exemple, comme on peut l'observer sur la figure n°10, les lipides ne possédant pas de tête polaire comme les triacylglycérols (TG), les diacylglycérols (DG) et les céramides (Cer) ne sont pas ou très peu retenus en comparaison des phospholipides qui possèdent par exemple une tête choline, éthanolamine, *etc.* Les lysophospholipides qui possèdent une tête polaire en plus de n'avoir qu'une seule chaîne d'acide gras au lieu de deux vis-à-vis des phospholipides sont les composés les plus retenus. Il est important de noter que les têtes polaires ne varient pas au sein d'une même famille lipidique et que la HILIC ainsi que la NPLC induisent peu de séparation selon la longueur des chaînes d'acide gras [58]. C'est pour cela que dans ces modes de séparation l'ensemble des lipides issus d'une même classe sont élués dans la même zone du chromatogramme comme indiqué sur la figure n°10 A. La chromatographie HILIC et de phase normale permettent donc d'obtenir une bonne séparation des familles au détriment de la séparation au sein d'une même classe de lipides [58]. Pour les mêmes raisons qui sont énoncées dans la partie précédente (*cf.* partie 3.1), la chromatographie HILIC est souvent préférée à la NPLC pour l'analyse des lipides.



**Figure 10 : Comparaison de la séparation chromatographique des lipides en mode RPLC et en mode HILIC pour l'analyse lipidique.** (A) En HILIC, les lipides d'une même famille se superposent tandis que les familles sont bien séparées entre elles [80]. (B) En mode RPLC, les lipides sont bien séparés au sein d'une même famille. En revanche, il y a une co-élution entre les familles [55].

Contrairement à la NPLC et la HILIC, la RPLC est réalisée sur une phase stationnaire apolaire et repose sur la séparation des lipides en fonction de leurs caractères hydrophobes portés par les chaînes d'acides gras [55]. Les composés les plus retenus sont donc les composés les plus apolaires (ex. lysophospholipides < phospholipides et sphingolipides < glycérolipides). Les différences entre chaînes de carbones pour une même classe lipidique assurent la bonne résolution au sein de celle-ci. Toutefois, ces chaînes d'acides gras sont présentes pour l'ensemble des lipides et il ne s'agit pas de structures différenciantes entre classes lipidiques. Cela peut conduire à une co-élution entre familles de lipides comme illustré dans la figure n°10 B avec la superposition des sphingomyélines (SM), des diacylglycérols (DG) et des phospholipides (PL) dont font partie les phosphatidylcholines (PC), les phosphatidyléthanolamines (PE), les phosphatidylinositols (PI), les phosphatidylsérines (PS) ainsi que les phosphatidylglycérols (PG). Pour les composés les plus apolaires, ce sont les TG qui peuvent être co-élus avec les esters de cholestérols. La RPLC permet donc d'obtenir une bonne séparation au sein des familles lipidiques au détriment de la séparation entre classes de lipides [55].

### 3.3. Techniques séparatives en protéomique

Les peptides sont des composés présentant une diversité physico-chimique très importante (hydrophobie, charge, point isoélectrique, taille, etc.). De ce fait, il est possible de les séparer par presque tous les modes de chromatographie liquide comme la chromatographie d'interaction hydrophile (HILIC), la chromatographie d'échange d'ions (IEC), la chromatographie d'exclusion de taille

(SEC), *etc.* [81]. La chromatographie de phase inverse (RPLC) est l'approche la plus courante en protéomique, car elle est reproductible, compatible avec la spectrométrie de masse et permet la rétention de la plupart des peptides [81–83].

L'acétonitrile (ACN) est le solvant organique le plus utilisé en RPLC auquel sont souvent ajoutés des additifs acides. Ces additifs agissent comme agent d'appariement d'ions pour former des paires d'ions neutres avec les acides aminés basiques des peptides (arginine, lysine et histidine). Cet appariement renforce le caractère hydrophobe de l'ion ce qui conduit à une meilleure rétention. Ces additifs acides diminuent également les interactions secondaires en protonant les sites chargés des groupements silanols qui peuvent interagir avec les groupements acides des peptides (asparagine, glutamine) limitant l'élargissement des pics chromatographiques [84,85]. L'acide trifluoroacétique (TFA) est l'additif de référence pour améliorer les performances chromatographiques des peptides. En effet, il s'agit d'un acide fort ( $pK_a = 0.3$ ) qui forme des paires d'ions très efficaces pour l'obtention de pic chromatographique fin et symétrique. Malheureusement, le TFA agit comme un fort suppresseur de signal lorsqu'il est couplé avec des spectromètres de masse. L'interaction entre le TFA et le peptide est tellement forte qu'il est compliqué de les dissocier. Cela empêche l'ionisation des molécules en mode ESI (+) à cause de la formation de paires d'ions neutres. De plus, le TFA rend le spray d'ionisation instable à cause de sa conductivité et de sa tension superficielle élevée nuisant ainsi à la détection par spectrométrie de masse [84,85]. Le TFA est souvent substitué par l'acide formique ( $pK_a = 3.75$ ), car bien qu'il présente des performances chromatographiques moins importantes (appariement d'ion plus faible), les propriétés de détection en spectrométrie de masse sont améliorées. Il s'agit du meilleur compromis entre performance chromatographique et sensibilité en spectrométrie de masse [84,85].

Pour l'analyse des peptides fortement polaires et hydrophiles, faiblement retenus en RPLC, il faut privilégier la séparation sur colonne HILIC d'autant plus que ce changement de phase stationnaire peut être accompagné par une augmentation de sensibilité. Lors des analyses RPLC utilisant l'ACN comme solvant d'élution il faut au maximum 25 à 30 % d'ACN pour éluer la totalité des peptides [86] alors que dans l'idéal il en faudrait 80% pour une désolvatation optimale. En effet, l'acétonitrile présente une volatilité plus grande et une tension de surface plus faible que l'eau ce qui favorise la désolvatation des gouttelettes chargées lors de l'ionisation en ESI [73]. Dans le cas des peptides, il faut environ 70 à 80% de phase organique pour réaliser leurs éluions en HILIC. De ce fait, le signal est plus intense en HILIC qu'en RPLC. Le gain de signal obtenu est en moyenne d'un facteur 2 à 4, mais il pourrait être plus important, car les peptides possèdent plusieurs états de charge qui diluent le signal comparé aux petites molécules monochargées [87]. Les peptides tryptiques ne contiennent qu'une seule lysine ou arginine localisée sur le C-terminal (*cf.* partie 5.1.2). Comme ces acides aminés sont les plus basiques, les protons resteront préférentiellement sur ces acides aminés lors de la fragmentation. Dans le cas

des ions polychargés, une des charges sera localisée en C-terminal et les autres seront placées sur les liaisons peptidiques afin de permettre la fragmentation selon le mécanisme du proton mobile [88–90]. Dans le cas des peptides monochargés, le proton sera emprisonné en C-terminal ce qui demandera plus d'énergie pour réaliser sa fragmentation induisant un rendement d'ionisation plus faible que pour un polychargé. C'est pour cela que les ions dichargés sont préférés aux monochargés pour la mise en place de méthodes de quantification. En ESI, l'état de charge moyen dépend de la tension de surface de la goutte lors de l'ionisation. Plus celle-ci est importante plus le composé sera chargé. Comme la tension de surface de l'eau est plus élevée que celle de l'acétonitrile, l'état de charge des peptides est plus faible en HILIC qu'en RPLC. De ce fait, la proportion de peptides dichargés est plus faible en HILIC qu'en RPLC contribuant davantage au phénomène de dilution du signal au profit de l'ion monochargé. Pour contrecarrer cet effet en déplaçant l'état de charge vers des niveaux supérieurs (monochargé vers dichargé) il faut ajouter dans la phase mobile un adduit avec une tension de surface supérieure à l'ACN comme l'alcool méta-nitrobenzylique (m-NBA). Cela permet d'obtenir un gain de sensibilité proche de ceux obtenus pour les petites molécules ne possédant qu'un seul état de charge [87]. Simon *et al.* [87] ont illustré cette augmentation de signal à travers la figure n°11.

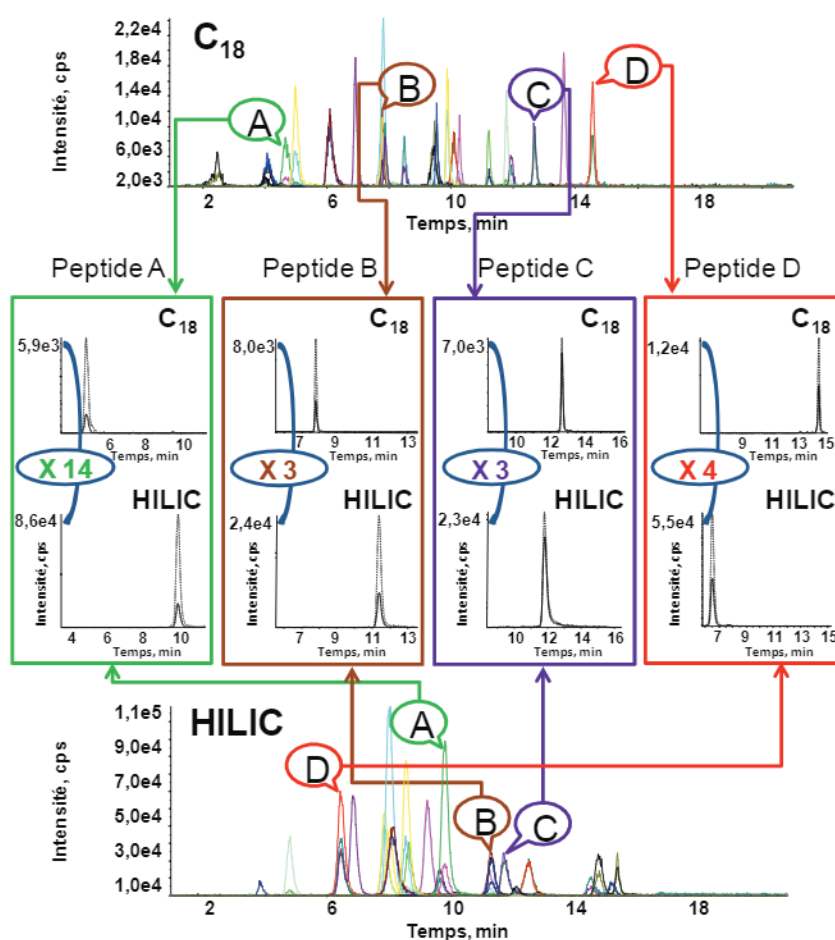


Figure 11 : Séparation d'un ensemble de peptides en HILIC avec 0,5% de m-NBA (p/v) dans la phase mobile (partie inférieure) et RPLC (partie supérieure) [87].

Le gain de signal en HILIC avec utilisation du m-NBA par rapport à la RPLC est compris entre un facteur 3 et 14 lorsque l'on observe l'augmentation de signal pour les 4 peptides mis en évidence. La HILIC est donc une bonne alternative à la RPLC, car en plus d'améliorer la sensibilité elle offre la possibilité d'obtenir une séparation différente de la RPLC grâce à son orthogonalité permettant ainsi de lever de possibles co-élutions. Cependant, son principal défaut est sa capacité de charge limitée qui est plus faible que la RPLC. Lorsque les quantités injectées sont équivalentes, l'HILIC permet d'abaisser les limites de quantification, mais lorsqu'une grande quantité d'échantillons est disponible ce gain de sensibilité peut être contrebalancé en RPLC en augmentant les quantités injectées [62].

La chromatographie liquide unidimensionnelle est largement appliquée dans le domaine des sciences omiques. Cependant, de nouveaux besoins analytiques ont émergé depuis les dix dernières années pour élargir le champ d'application de la LC aux échantillons complexes [91]. De nos jours, les analystes ont tendance à ne plus vouloir quantifier quelques composés dans un milieu complexe, mais plutôt à quantifier le plus grand nombre de ces inconnus parfois même à l'état de traces [91]. Ces méthodes de séparation classiques n'ont pas toujours un pouvoir de résolution suffisant pour séparer l'ensemble des composés cibles dans un échantillon complexe. Les chercheurs en métabolomique et en protéomique sont confrontés à cette problématique, car les besoins en termes de nombre de composés à identifier et à mesurer pour ces biomolécules sont en augmentation afin de tendre vers la biologie des systèmes [91,92]. Cependant, toute augmentation du pouvoir de séparation chromatographique qui ne résulte pas d'un changement d'instrumentation ou de la conséquence du progrès dans la science des matériaux nécessitera souvent une augmentation de temps [91,92]. Il n'est pas possible de poursuivre dans cette voie puisque le temps d'analyse global est l'un des paramètres clés dans le développement analytique tant d'un point de vue scientifique qu'économique. Un concept différent comme la chromatographie bidimensionnelle est nécessaire [91].

#### **4. Mise en place d'une méthode chromatographique avancée : la chromatographie liquide bidimensionnelle**

La réponse au manque de séparation de la chromatographie monodimensionnelle est le développement de systèmes chromatographiques multidimensionnels. D'après Giddings *et al.* [93], le terme séparation multidimensionnelle qui comprend également la chromatographie bidimensionnelle (2D-LC) fait référence à une technique dans laquelle : (i) les composés subissent deux ou plusieurs séparations indépendantes et (ii) que les composés correctement résolus dans une dimension restent résolus tout au long du processus de séparation. Les configurations dans lesquelles la phase mobile sortant d'une colonne est envoyée directement dans une autre (colonnes directement branchées en série) ne peuvent donc pas être assimilées à de la 2D-LC [93,94].

En chromatographie, la qualité de séparation est souvent évaluée par le calcul de la résolution entre deux composés en déterminant la différence de temps de rétention associée aux largeurs des pics chromatographiques. Cependant, cette grandeur est limitée, car elle ne prend en compte que des séparations deux à deux et non pas de l'ensemble des composés au sein du chromatogramme. C'est pour cela qu'une autre grandeur appelée capacité de pics a été proposée par Giddings [95] puis améliorée par Horvath [96] pour l'adapter aux analyses en mode gradient. Cette capacité de pics mesure le nombre de pics qui peut être contenu dans une fenêtre de temps avec une résolution fixée (en général  $R=1$ ). Dans des conditions normales en termes de temps d'analyse et de pression, les analyses unidimensionnelles ont généralement des capacités de pics inférieures à 500. Pour augmenter cette capacité de pics, il est nécessaire d'optimiser la méthode chromatographique au maximum [97]. On peut alors atteindre une capacité de pics proche de 1000 [98]. Dans les meilleures conditions et avec un temps d'analyse de 1 journée, la valeur maximale atteignable est de 1400-1600 en 1D-LC [99]. Dans le cas où l'on souhaite augmenter davantage le nombre de composés à séparer dans un échantillon complexe, il sera nécessaire de mettre en place des méthodes bidimensionnelles. En effet, Karger [100], Giddings [101] et Guiochon [102] ont démontré que la capacité de pics en 2D-LC correspond au produit des capacités de pics des deux colonnes qui composent le système à deux dimensions. L'orthogonalité entre les deux colonnes est le premier critère pour vérifier le degré de pertinence d'un système 2D. Ce degré d'orthogonalité correspond aux différences de sélectivité entre les deux dimensions et il peut être évalué comme étant l'espace bidimensionnel qui est occupé par les pics sur le diagramme de rétention. Plus il sera important, plus le système aura de chance de séparer des composés co-élus dans la 1<sup>ère</sup> dimension comme le montre la figure n°12.

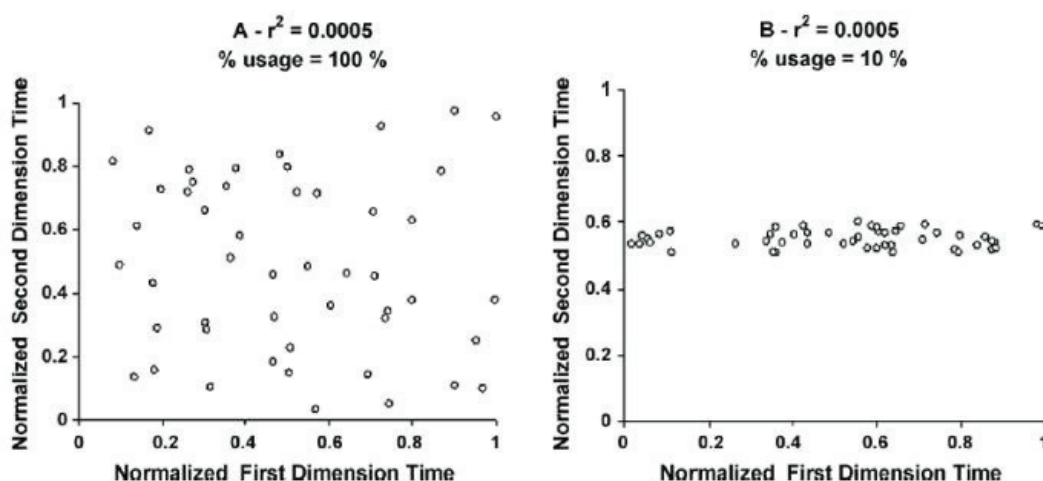


Figure 12 : Exemple de mesure de l'espace de séparation pour un système LCxLC. Plus le pourcentage d'usage est important, plus le système bidimensionnel est orthogonal [97].

En effet, l'espace de séparation à droite de la figure (10%) est moins important qu'à gauche (100%) puisque les temps de rétention sont plus resserrés et donc moins bien séparés. Ce phénomène est appelé diagonalisation des temps de rétention et il traduit le manque d'orthogonalité.

Le nom chromatographie liquide bidimensionnelle regroupe différentes techniques qui peuvent s'effectuer en ligne ou hors ligne. Il existe également un mode utilisant des débits stoppés. Ces différents modes sont utilisés en fonction des problématiques, des échantillons et des appareils disponibles, car ils présentent chacun des avantages et des inconvénients.

### 4.1. Les différents modes 2D

#### 4.1.1. Hors ligne

Les systèmes chromatographiques 2D en mode hors ligne consistent à séparer l'échantillon grâce à une première séparation appelée 1<sup>ère</sup> dimension (D1). Les différentes fractions issues de la D1 sont ensuite collectées avant d'être évaporées puis redissoutes dans un solvant approprié pour être injecté dans la 2<sup>ème</sup> dimension séparative (D2). Les avantages de ce type de fonctionnement sont que les deux dimensions sont indépendantes l'une de l'autre. En effet, il n'y a aucune limite de temps de séparation dans la 2<sup>ème</sup> dimension, car les deux séparations se font de manière indépendante. De plus, cette séparation bidimensionnelle peut être réalisée sur un seul système chromatographique. En revanche, le temps d'analyse est généralement long, car il correspond au temps d'analyse de la D1 auquel est ajouté le temps d'analyse de la D2 multiplié par le nombre de fractions récoltées. Par exemple, certaines séparations 2D peuvent durer plusieurs heures (13 heures pour 16 fractions [103]) et même jusqu'à plusieurs jours (plus de 80 heures pour 27 fractions [104]). Les fractions sont manipulées par l'expérimentateur entre les deux dimensions ce qui augmente le risque de perte, de contamination et le temps d'analyse.

#### 4.1.2. En ligne

Les modes en ligne utilisent deux systèmes chromatographiques simultanément pour analyser en continu les fractions issues de la 1<sup>ère</sup> dimension dans la 2<sup>ème</sup> dimension. Il existe deux configurations nommées coupe à cœur ou *heart-cutting* et la configuration totale ou *comprehensive*. L'ensemble des modes en ligne ont l'avantage d'être entièrement automatisables.

##### 4.1.2.A. Mode *multiple heart cut* (LC-LC)

Le mode *multiple heart cut* consiste à sélectionner les fractions d'intérêts avant de les injecter dans la 2<sup>ème</sup> dimension comme le montre la figure n°13. Pour réaliser cette configuration qui est abrégée LC-LC [105,106], il faut deux systèmes chromatographiques séparés par une vanne permettant de stocker



les fractions collectées issues de la 1<sup>ère</sup> dimension avant leurs analyses dans la seconde dimension. En général, ce mode est utilisé lorsqu'il y a peu de composés d'intérêts qui doivent être isolés, car le nombre de fractions pouvant être collectées est limité par le nombre de boucles de stockage disponibles dans le système. Dix fractions peuvent être collectées dans le système vendu par Agilent® représenté sur la figure n°13 puisque chaque deck possède cinq boucles de stockage plus une boucle de passage. Comme pour le mode hors ligne le temps d'analyse dans la D2 est indépendant de la D1. Cependant, la composition à l'élution à l'issue de la 1<sup>ère</sup> dimension peut impacter la deuxième séparation, car les deux dimensions sont directement connectées. En LC-LC, le temps d'analyse total correspond au temps d'analyse de la D1 auquel est ajouté le temps d'analyse de la D2 multiplié par le nombre de fractions.

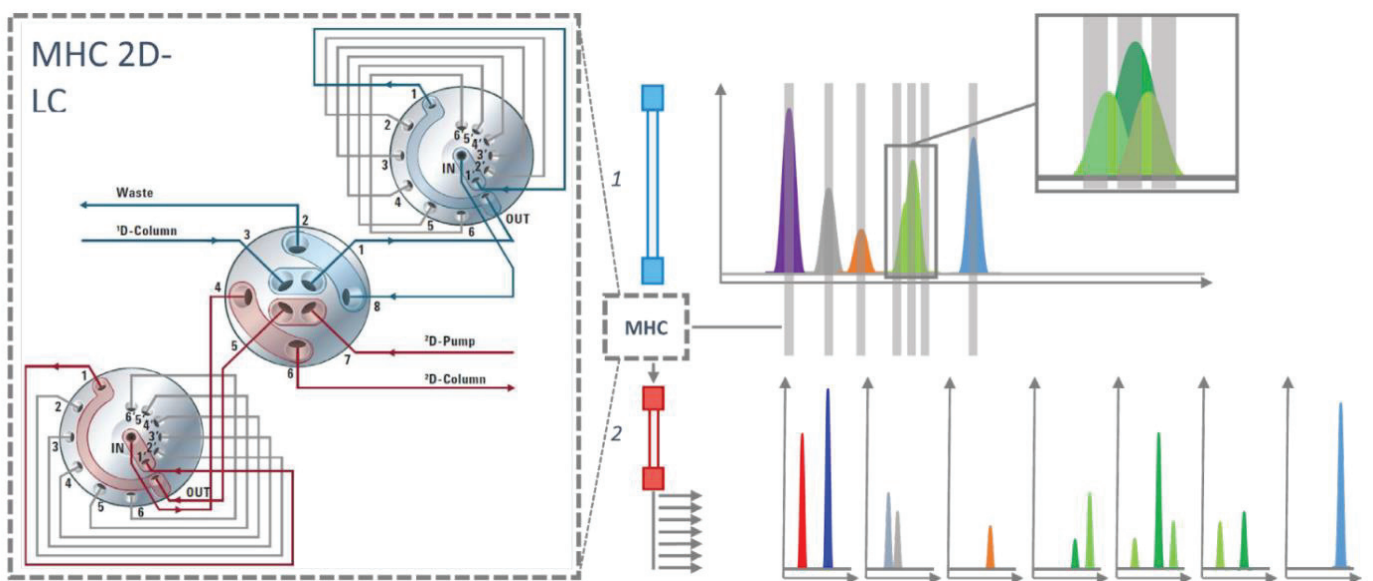


Figure 13 : Principe de la chromatographie 2D en mode *multiple heart cut* [106].

#### 4.1.2.B. Mode *comprehensive* (LCxLC)

Le mode *comprehensive* consiste à récupérer l'ensemble des fractions issues de la 1<sup>ère</sup> dimension et à les analyser tour à tour dans la seconde. Cette configuration qui est abrégée LCxLC [105,106] nécessite d'utiliser deux systèmes chromatographiques séparés par une vanne servant d'interface entre les deux dimensions comme illustrée sur la figure n°14.



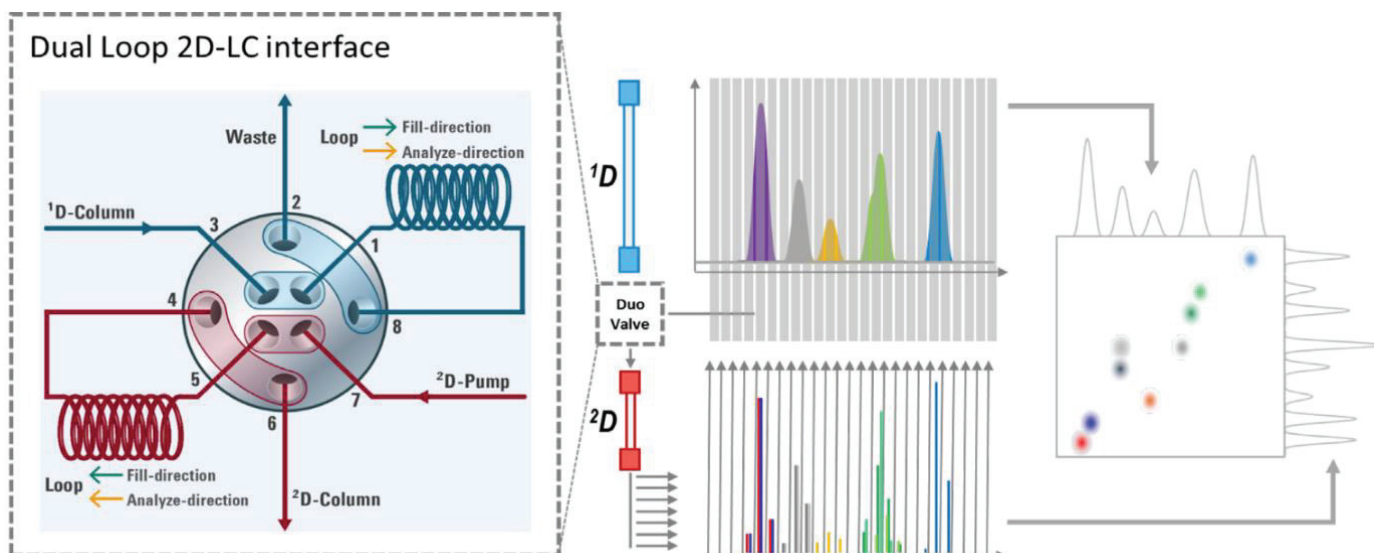


Figure 14 : Principe de la chromatographie 2D en mode *comprehensive* [106].

Ce mode implique que le temps d'analyse de la 2<sup>ème</sup> dimension soit très rapide (inférieur à la minute), car il doit être inférieur au temps de collecte de la fraction suivante dans la 1<sup>ère</sup> dimension (Temps de séparation en D2 < [Volume d'une boucle / Débit en D1]) [107]. En LCxLC, le temps d'analyse correspond au temps total de la 1<sup>ère</sup> dimension. Comme pour le mode *multiple heart cut* la composition à l'élution de la fraction éluee à l'issue de la D1 peut impacter la deuxième séparation, car les deux dimensions sont directement connectées.

#### 4.1.2.C. Mode à débit stoppé

Le mode chromatographique 2D à débit stoppé, aussi appelé *stop and go* ou *stopflow*, consiste à arrêter le débit de la 1<sup>ère</sup> dimension lorsqu'une fraction est transférée vers la 2<sup>ème</sup> dimension. Une fois la séparation dans la 2<sup>ème</sup> dimension effectuée, le débit de la 1<sup>ère</sup> dimension est remis en route jusqu'à ce qu'une autre fraction soit transférée dans la deuxième colonne et ainsi de suite. En mode débit stoppé, le temps d'analyse est généralement long, car il correspond au temps d'analyse de la D1 auquel est ajouté le temps d'analyse de la D2 multiplié par le nombre de fractions. Dans le cas d'une analyse d'un échantillon protéique après digestion tryptique, Sommella *et al.* ont un temps d'analyse de 17h [108].

Les modes hors ligne, en ligne (LC-LC et LCxLC) et à débit stoppé présentent chacun des avantages et inconvénients qui leurs sont propres. Il est donc important de les comparer pour savoir lequel sera le plus en adéquation avec les objectifs de la méthode 2D qui sera développée.

### 4.1.3. Comparaison des 3 modes

Dans le cadre des analyses bidimensionnelles, le mode hors ligne permet d'avoir des capacités de pics plus importantes que le mode en ligne au détriment du temps d'analyse. En effet Beelders *et al.* [109] ont comparé le mode hors ligne et LCxLC (HILIC en D1 et RPLC en D2) pour mettre en évidence que le mode hors ligne possède une capacité de pics 4 fois plus importante que le mode en ligne, mais avec un temps d'analyse 16 fois plus grand. Le mode débit stoppé est un compromis aux deux autres types de systèmes 2D. Il permet d'obtenir des capacités de pics intermédiaires avec des temps d'analyse supérieurs au mode en ligne. Toutefois, l'approche *stop and go* est peu utilisée à cause de la dispersion générée dans la 1<sup>ère</sup> dimension lorsque le débit est arrêté. Les approches en ligne sont les plus complexes à développer d'un point de vue technologique, car il faut avoir à disposition deux chaînes chromatographiques en plus de l'interface de modulation placée entre les deux dimensions. Les avantages et inconvénients des trois modes sont résumés dans le tableau n°1. Comme nous pouvons le voir, le mode en ligne présente le plus d'avantages en proposant un bon compromis entre capacité de pics et temps d'analyse. D'autant plus que l'ensemble du processus est entièrement automatisable. C'est pourquoi ce mode sera préféré dans les travaux présentés dans ce manuscrit.

	Hors ligne	En ligne	Débit stoppé
Capacité de pics	++	+	++
Temps d'analyse	-	++	-
Automatisation	-	++	+
Problème de pollutions	-	+	+
Perte d'échantillons	-	+	+
Appareillage nécessitant une interface complexe (Vanne automatique, boucle de stockage, etc.)	++	-	+
Compatibilité entre D1 et D2	+	+/-	+/-

**Tableau 1 : Avantages et inconvénients des modes hors ligne, en ligne et à débit stoppé en 2D-LC.** ++ très positif, + positif, - négatif, +/- négatif dans certains cas [107].

## 4.2. Problèmes rencontrés en 2D-LC

### 4.2.1. Les effets à l'injection

L'un des problèmes majeurs rencontrés lors des analyses bidimensionnelles en ligne est la compatibilité des solvants entre les deux dimensions, surtout si l'on désire avoir les systèmes les plus orthogonaux possibles. Le premier critère à vérifier est la miscibilité entre les phases mobiles des deux dimensions. Le second est la force d'éluion entre les phases mobiles. Par exemple lors du couplage

HILICxRPLC, l'ACN est la phase mobile qui est considérée comme le solvant le moins éluant en HILIC alors qu'il est le plus éluant en phase inverse. Le couplage de ces deux dimensions engendre d'importants effets à l'injection dans la 2<sup>ème</sup> dimension, comme des déformations de pics chromatographiques et des phénomènes de percées aussi appelés *breakthrough* [110–114]. Ce phénomène, représenté sur la figure n°15, correspond à la séparation d'un pic chromatographique en deux pics chromatographiques distincts. Le premier pic est élué au temps mort et le deuxième au temps de rétention normal du composé avec plus ou moins de déformation à l'avant du pic. Ce dédoublement est dû à la division des composés après leur introduction dans la colonne chromatographique (injection en 1<sup>ère</sup> dimension et/ou introduction en 2<sup>ème</sup> dimension) dans le cas où le solvant d'injection possède un pouvoir éluant trop important. Une partie des analytes est directement entraînée par la phase mobile (élution au temps mort) alors que l'autre est retenue sur la phase stationnaire (élution au temps de rétention normal). Comme indiqué sur la figure n°15, plus la proportion de solvant fort est importante plus les effets à l'injection seront prononcés ce qui induira une perte de sensibilité. La stratégie à mettre en place pour éviter ce problème récurrent en 2D-LC en ligne consiste à limiter la quantité injectée dans la 2<sup>ème</sup> dimension. Pour cela, il est possible de diviser le débit en entrée de la 2<sup>ème</sup> dimension en ajoutant une jonction en « T » ou *split* diminuant ainsi la proportion de solvant fort injecté. Les fractions issues de la 1<sup>ère</sup> dimension peuvent également être diluées entre les deux dimensions en introduisant un co-solvant (solvant de *make-up*) avec un pouvoir éluant plus faible [114]. Une solution récente et innovante proposée par Chapel *et al.* [114] consiste à augmenter le volume d'injection jusqu'à l'obtention du phénomène de total *breakthrough*. Ils ont montré que cette condition particulière permet d'avoir un pic parfaitement défini malgré la persistance du pic de *breakthrough* lorsque le volume d'injection est suffisamment important (*cf.* figure n°15 F). La figure n°16 indique que la condition de total *breakthrough* permet d'obtenir des réponses linéaires lorsque la concentration augmente permettant ainsi de faire des études quantitatives sur ce type de pics chromatographiques [114]. L'utilisation de la division du débit, de la dilution avec le co-solvant et du total *breakthrough* entraîne toujours une perte de sensibilité par rapport à la condition sans effet à l'injection [115–117].

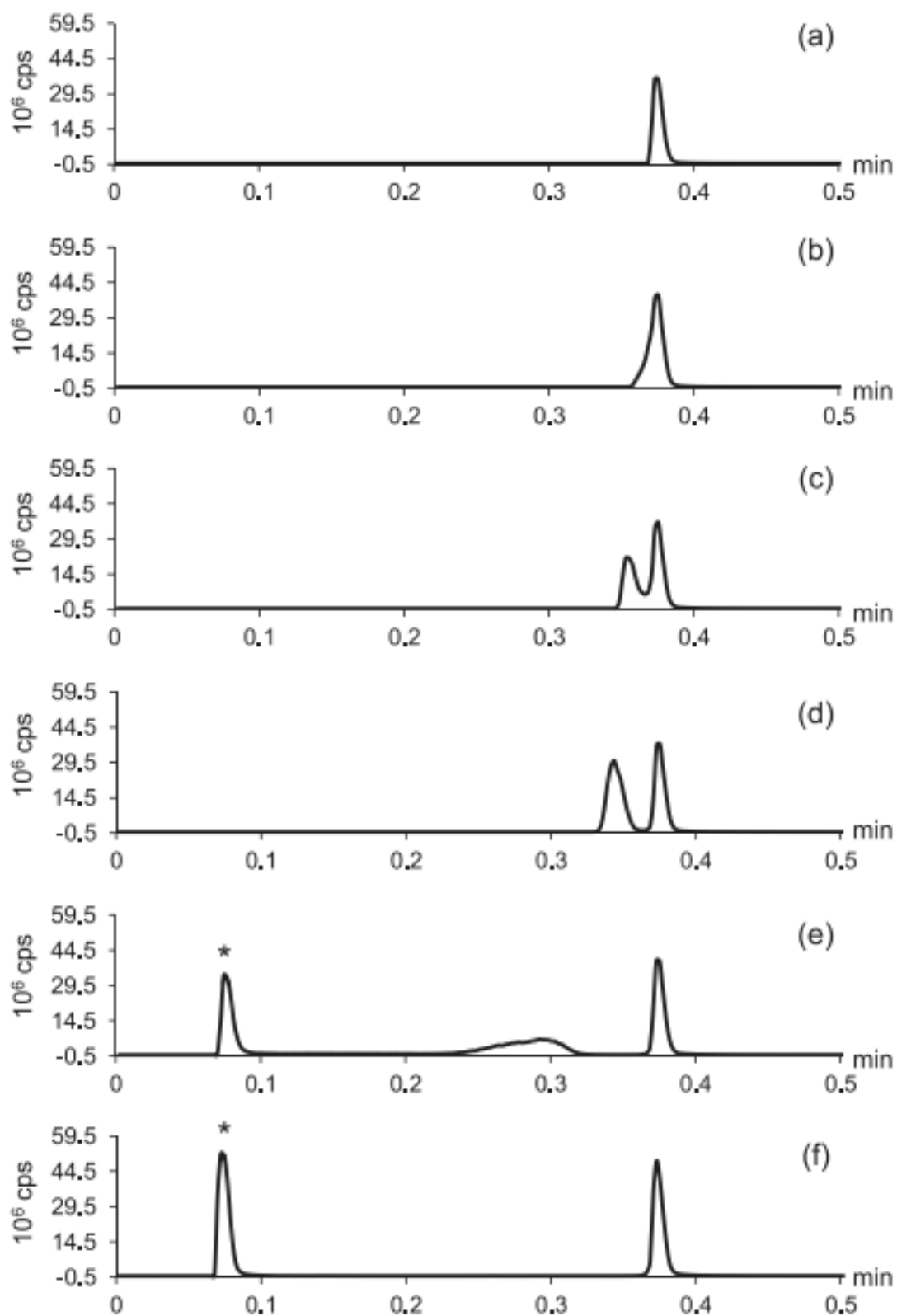
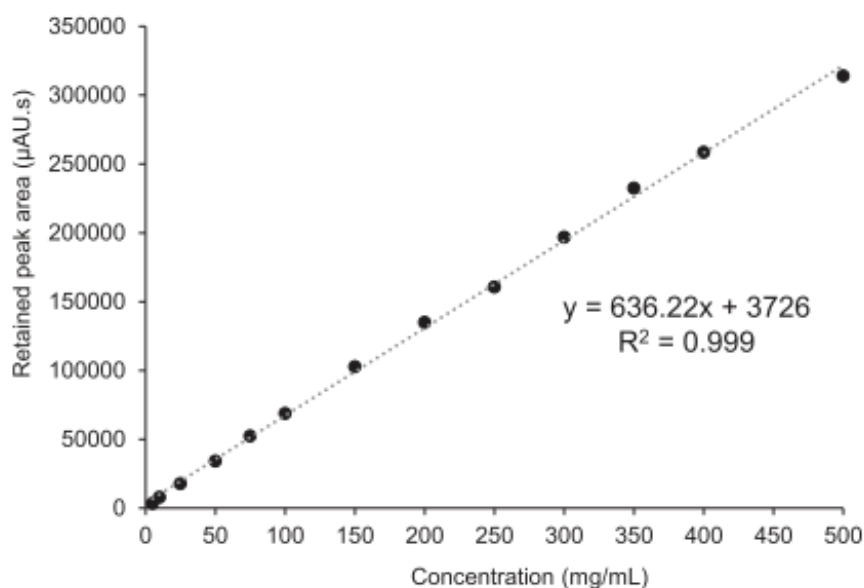


Figure 15 : Évolution de la forme des pics lors de l'augmentation du volume d'injection par rapport au volume mort de la colonne ( $V_0$ ). (a) 1  $\mu\text{L}$  injecté (1.5%  $V_0$ ), (b) 1.5  $\mu\text{L}$  injecté (2.1%  $V_0$ ), (c) 1.8  $\mu\text{L}$  injecté (2.5%  $V_0$ ), (d) 2.2  $\mu\text{L}$  injectés (3%  $V_0$ ), (e) 3.5  $\mu\text{L}$  injectés (4.8%  $V_0$ ), (f) 4.5  $\mu\text{L}$  injectés (6.2%  $V_0$ ). L'astérisque indique le pic de *breakthrough* [114].



**Figure 16 :** Variation de l'aire du pic en fonction de la concentration dans des conditions de total *breakthrough*. Solvant d'injection H<sub>2</sub>O/ACN (30/70 ; v/v), 3,5 μL (4,8% V0) injectés sur une colonne HILIC. Utilisation d'un détecteur UV à 210 nm pour détecter des peptides [114].

#### 4.2.2. Couplage avec la spectrométrie de masse et les sources ESI

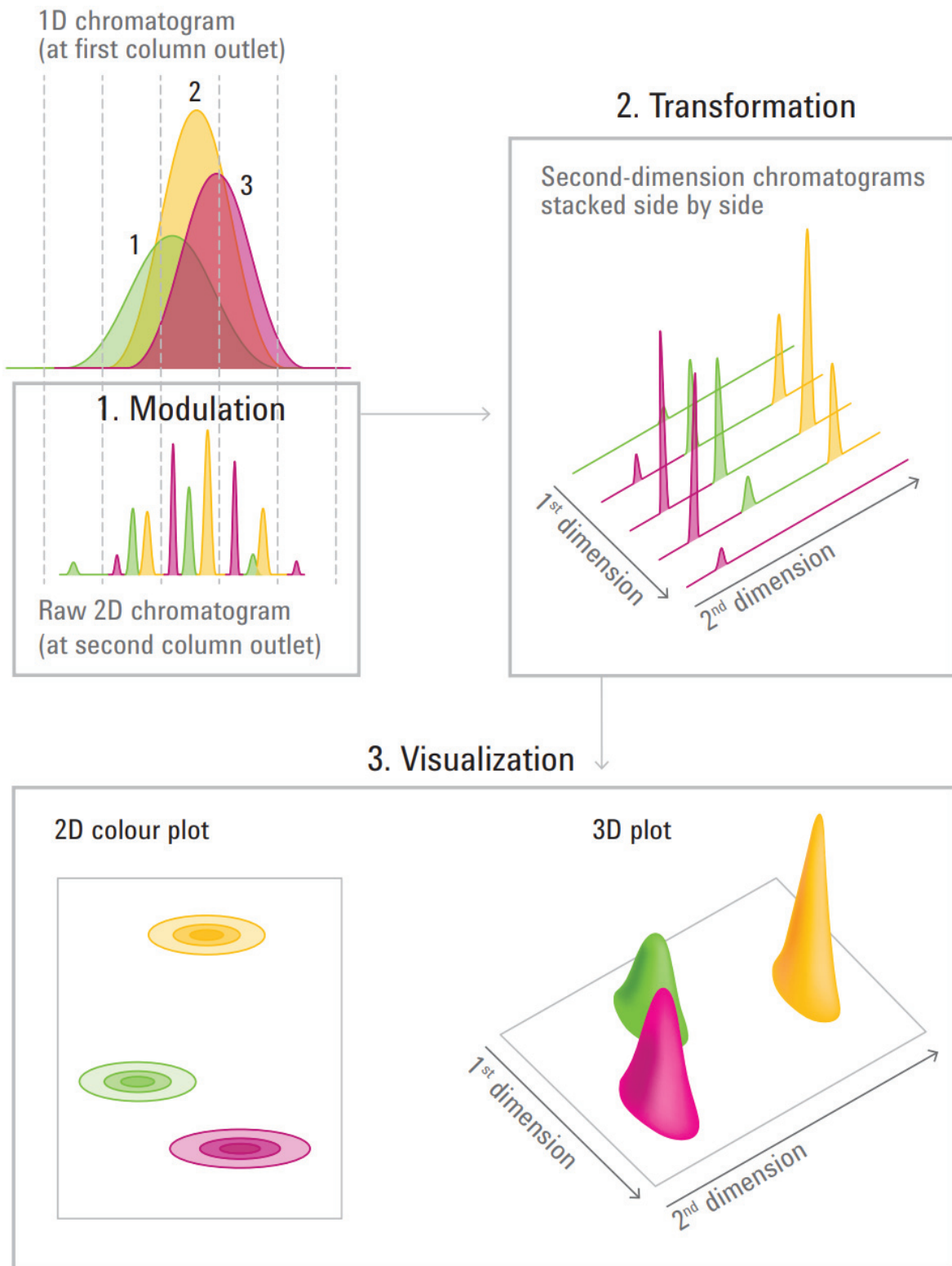
Les spectromètres de masse à impact électronique (EI), à ionisation chimique (CI) ou à thermospray sont des détecteurs sensibles au débit massique introduit dans la source de l'appareil. Contrairement à ces détecteurs, la source électrospray (ESI) possède une sensibilité à la concentration qui est le résultat d'effets contradictoires. L'ESI-MS est une suite d'étapes comprenant la nébulisation, le chargement des gouttelettes, l'évaporation du solvant et la libération des ions qui sont conduits dans le spectromètre de masse. Plus le nombre d'ions introduits dans le spectromètre de masse sera important plus la réponse obtenue le sera également. Cependant, l'augmentation du débit à l'entrée du spectromètre de masse n'induit pas forcément une augmentation de sensibilité à la suite d'un afflux plus important d'ions dans la source. L'évaporation du solvant joue un rôle important dans le mécanisme de l'électrospray. Un débit élevé peut donc réduire l'efficacité de la désolvatation des composés d'intérêts. Par conséquent, l'augmentation du débit diminue la libération d'ions à partir des gouttelettes et limite donc la sensibilité du spectromètre de masse [118]. De nos jours, tous les fabricants proposent des sources électrospray fonctionnant à des débits de l'ordre du millilitre par minute [119]. Toutefois, les débits inférieurs à 500 μL/min permettent une meilleure désolvatation et donc une ionisation optimale [119]. La mise en place d'un système 2D en mode *comprehensive* implique des débits supérieurs à 1 mL/min afin d'avoir une séparation la plus rapide possible en 2<sup>ème</sup> dimension. Pour résoudre ce problème de compatibilité de débit, il faut mettre en place une jonction « T » à l'entrée du spectromètre de masse afin de diviser le débit par un facteur fixé par l'expérimentateur [119]. Cette division de débit ou *split* peut s'accompagner d'une perte de sensibilité, car une partie des composés ne sera pas envoyée dans le détecteur. C'est pour cela qu'il est important

d'évaluer le ratio gain / perte de sensibilité en fonction de la division et du débit en amont du spectromètre de masse. Le *split* diminuera la quantité potentielle d'ions introduits dans le spectromètre de masse, mais en contrepartie il facilitera la désolvation des ions et donc la quantité d'ions libérés à partir des gouttelettes [118]. Lorsque c'est possible, les jonctions « T » sont à éviter pour que le système reste aussi simple et robuste que possible. Cela élimine le risque de blocage d'un capillaire de transfert de faible diamètre [118].

L'autre facteur à prendre en considération lors du couplage de la spectrométrie de masse avec des systèmes 2D-LC est la fréquence d'acquisition. Il est couramment admis que pour définir un pic chromatographique il faut entre 10 et 15 points par pic. Dans le cas d'une configuration classique, les pics chromatographiques possèdent des largeurs de pic de 10 secondes ou plus. Dans le cas de la mise en place de méthode LCxLC, les pics peuvent être réduits à une largeur de 1 seconde ou moins toujours dans le but d'avoir une séparation la plus rapide possible en 2<sup>ème</sup> dimension [120]. Ainsi, il peut être compliqué d'avoir suffisamment de points pour définir de tels pics chromatographiques en LCxLC. Voilà pourquoi les spectromètres de masse de type temps de vol (*Time of flight* ou TOF) sont principalement couplés au système LCxLC, car ils présentent les vitesses d'acquisitions les plus importantes.

### 4.2.3. Traitement des données

La mise en place d'un système chromatographique bidimensionnel implique une complexité dans le traitement des données, car on rajoute une dimension supplémentaire qui permet d'obtenir une quantité d'informations plus importante que les méthodes 1D. De ce fait, il est primordial d'utiliser des outils informatiques pour traiter les chromatogrammes obtenus [92]. Pour cela, il existe plusieurs approches pour traiter ce type de données 2D. Une des possibilités est de traiter les données fraction par fraction ce qui permet d'utiliser les outils classiques en 1D-LC. Cependant, cela peut rapidement devenir délicat lorsque l'on se retrouve avec plus de 3 dimensions à exploiter (par exemple : temps de rétention D1, temps de rétention D2, intensité, spectre de masse) impliquant l'utilisation d'outil chimométrique complexe. L'autre solution est de transformer le fichier de données en image à 2 ou 3 dimensions. La figure n°17 montre des exemples de représentation des données. La première (3D plot) représente les pics chromatographiques en 3 dimensions. La deuxième représentation (2D colour plot) illustre le même chromatogramme vu du dessus. Cette méthode de transformation en fichier image donne des résultats plus simples à interpréter et est actuellement disponible commercialement sous forme de logiciel (ex : LC image).



**Figure 17 : Mise en forme de la collection de chromatogrammes de 2<sup>ème</sup> dimension en un chromatogramme 2D-LC.** Dans la 1<sup>ère</sup> partie, les pics chromatographiques verts, jaunes et roses qui constituent un pic composite non résolu en 1D-LC sont séparés en 2D-LC. Dans la 2<sup>ème</sup> partie, les chromatogrammes de 2<sup>ème</sup> dimension sont transformés afin que la série soit plus visible et que les pics individuels jaunes, verts et roses augmentent et diminuent au fur et à mesure que le pic de 1<sup>ère</sup> dimension est élué. Dans la 3<sup>ème</sup> partie, deux visualisations différentes du chromatogramme 2D sont présentées, la représentation en 2D et en 3D [121].

La chromatographie bidimensionnelle a déjà été appliquée à des analyses en protéomiques [114] et métabolomiques [122]. Dans cette thèse, cette technique séparative a été utilisée pour les analyses en lipidomique. La suite de l'état de l'art se focalisera donc sur les différentes configurations déjà mises en place pour l'analyse de cette famille omique.

### 4.3. La 2D-LC dans le cadre des analyses lipidomiques

Lors des analyses lipidiques en chromatographie bidimensionnelle, les phases inverses ont déjà été associées à des phases polaires. L'avantage de coupler ces deux dimensions est de combiner la séparation des lipides d'une même famille en RPLC avec la séparation des différentes classes lipidiques en HILIC ou NPLC comme expliqué précédemment (*cf.* partie 3.2). Ainsi cela permet de diminuer les effets matrices subits par les lipides minoritaires et d'éviter la présence de composés isobares entre les familles lipidiques [123]. Les différentes configurations 2D qui ont déjà été mises en place dans le cadre d'analyses lipidomiques sont référencées dans le tableau n°2. Les avantages et les inconvénients de chacun de ces systèmes ont été résumés.



Système 2D	Avantages	Inconvénients	Biblio
Hors ligne D1 : RPLC D2 : Ag-LC	<ul style="list-style-type: none"> <li>Séparation entre acide gras cis et trans</li> <li>Deux dimensions totalement indépendantes</li> <li>Capacité de pics totalement optimisée</li> <li>Retraitement facilité et compatible avec tout type de MS</li> <li>Simplicité du montage</li> </ul>	<ul style="list-style-type: none"> <li>Analyse d'une seule famille (TAG)</li> <li>Non automatisable</li> <li>Temps d'analyse long</li> <li>Évaporation entre les deux séparations</li> <li>Risque de perte d'échantillon accrue</li> </ul>	[124]
Hors ligne D1 : HILIC D2 : RPLC	<ul style="list-style-type: none"> <li>Deux dimensions totalement indépendantes</li> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Capacité de pics totalement optimisée</li> <li>Retraitement facilité et compatible avec tout type de MS</li> <li>Simplicité du montage</li> </ul>	<ul style="list-style-type: none"> <li>Non automatisable</li> <li>Temps d'analyse long</li> <li>Évaporation entre les deux séparations</li> <li>Risque de perte d'échantillon accrue</li> </ul>	[125]
LC-LC <i>Heart cut</i> D1 : NPLC D2 : RPLC	<ul style="list-style-type: none"> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Automatisable</li> <li>2<sup>ème</sup> dimension indépendante de la 1<sup>ère</sup></li> <li>Retraitement facilité et compatible avec tout type de MS</li> <li>Évaporation automatique entre les deux dimensions</li> </ul>	<ul style="list-style-type: none"> <li>Risque de perte d'échantillon lors de l'évaporation</li> <li>Complexité du montage</li> <li>1<sup>ère</sup> dimension dépendante de la 2<sup>ème</sup></li> <li>Capacité de pics plus faible que le mode hors ligne</li> </ul>	[126,127]
LC-LC <i>Heart cut</i> D1 : HILIC D2 : RPLC	<ul style="list-style-type: none"> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Automatisable</li> <li>2<sup>ème</sup> dimension indépendante de la 1<sup>ère</sup></li> <li>Retraitement facilité et compatible avec tout type de MS</li> </ul>	<ul style="list-style-type: none"> <li>Pas de <i>multiple heart cut</i> possible avec ces systèmes</li> <li>1 injection nécessaire par fraction</li> <li>1<sup>ère</sup> dimension dépendante de la 2<sup>ème</sup></li> <li>Diminution obligatoire du pouvoir éluant du solvant d'injection en 2D</li> </ul>	[128,129]
<i>Stop flow</i> D1 : NPLC D2 : RPLC	<ul style="list-style-type: none"> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Retraitement facilité et compatible avec tout type de MS</li> <li>Automatisable</li> <li>2<sup>ème</sup> dimension indépendante de la 1<sup>ère</sup></li> </ul>	<ul style="list-style-type: none"> <li>Débit stoppé en 1<sup>ère</sup> dimension</li> <li>Élargissement des pics</li> <li>Évaporation entre les deux dimensions</li> </ul>	[130]
<i>Stop flow</i> D1 : HILIC D2 : RPLC	<ul style="list-style-type: none"> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Retraitement facilité et compatible avec tout type de MS</li> <li>Automatisable</li> <li>2<sup>ème</sup> dimension indépendante de la 1<sup>ère</sup></li> </ul>	<ul style="list-style-type: none"> <li>Débit stoppé en 1<sup>ère</sup> dimension</li> <li>Élargissement des pics</li> <li>Diminution obligatoire du pouvoir éluant du solvant d'injection en 2D</li> <li>1<sup>ère</sup> dimension dépendante de la 2<sup>ème</sup></li> </ul>	[131,132]
LCxLC D1 : RPLC D2 : HILIC	<ul style="list-style-type: none"> <li>Gain de temps par rapport aux autres modes 2D</li> <li>La totalité des fractions est envoyée dans la 2<sup>ème</sup> dimension</li> <li>Automatisable</li> </ul>	<ul style="list-style-type: none"> <li>Les deux dimensions sont dépendantes l'une de l'autre</li> <li>Retraitement complexe des données et compatible avec les TOF uniquement</li> <li>Capacité de pics plus faible que les autres modes 2D</li> <li>Complexité du montage</li> <li>Diminution obligatoire du pouvoir éluant du solvant d'injection en 2D</li> <li>Temps d'équilibrage important en HILIC</li> </ul>	[123,133]
LCxLC D1 : HILIC D2 : RPLC	<ul style="list-style-type: none"> <li>Fractionnement des familles lipidiques puis séparation au sein de ces familles</li> <li>Gain de temps par rapport aux autres modes 2D</li> <li>La totalité des fractions est envoyée dans la 2<sup>ème</sup> dimension</li> <li>Automatisable</li> </ul>	<ul style="list-style-type: none"> <li>Les deux dimensions sont dépendantes l'une de l'autre</li> <li>Retraitement complexe des données et compatible avec les TOF uniquement</li> <li>Capacité de pics plus faible que les autres modes 2D</li> <li>Complexité du montage</li> <li>Diminution obligatoire du pouvoir éluant du solvant d'injection en 2D</li> <li>RPLC pas assez résolutive avec un gradient court.</li> </ul>	[123]

Tableau 2 : Avantages et inconvénients des différents systèmes 2D-LC (hors ligne, *stop flow*, *heart cut* et *comprehensive*) dans le cas des analyses lipidomiques.

### 4.3.1. Les approches hors ligne pour l'analyse des lipides

Comme on peut le voir dans le tableau n°2, les approches en mode hors ligne sont les plus faciles à mettre en place. En effet comme ces deux dimensions sont indépendantes l'une de l'autre cela permet d'avoir les capacités de pics les plus importantes. On peut ainsi coupler tous les systèmes chromatographiques sans se soucier de la compatibilité entre les phases mobiles puisque l'on peut évaporer et resolubiliser l'échantillon entre les deux dimensions. Holcapek *et al.* ont ainsi pu coupler la chromatographie de phase inverse à de la chromatographie ionique utilisant des colonnes en argent (Ag-LC) afin de séparer les isomères cis et trans des triacylglycérols [124]. Cependant, ces approches sont difficilement automatisables, nécessitent des temps d'analyse longs et peuvent engendrer des risques de perte d'échantillons entre les dimensions [124,125].

### 4.3.2. Les approches en ligne pour l'analyse des lipides

#### 4.3.2.A. Choix de la phase stationnaire polaire : HILIC ou NPLC ?

En 2D-LC en ligne, la RPLC est plus souvent associée à la HILIC plutôt qu'à la NPLC. En effet, ces deux modes de chromatographies utilisent des phases mobiles similaires et donc miscibles entre elles. Cela permet d'éviter l'étape de séchage de l'échantillon entre les deux dimensions qui est indispensable pour les analyses 2D utilisant la RPLC couplée à la NPLC [126,127,130]. Cette étape d'évaporation peut être automatisée comme le montre la figure n°18.

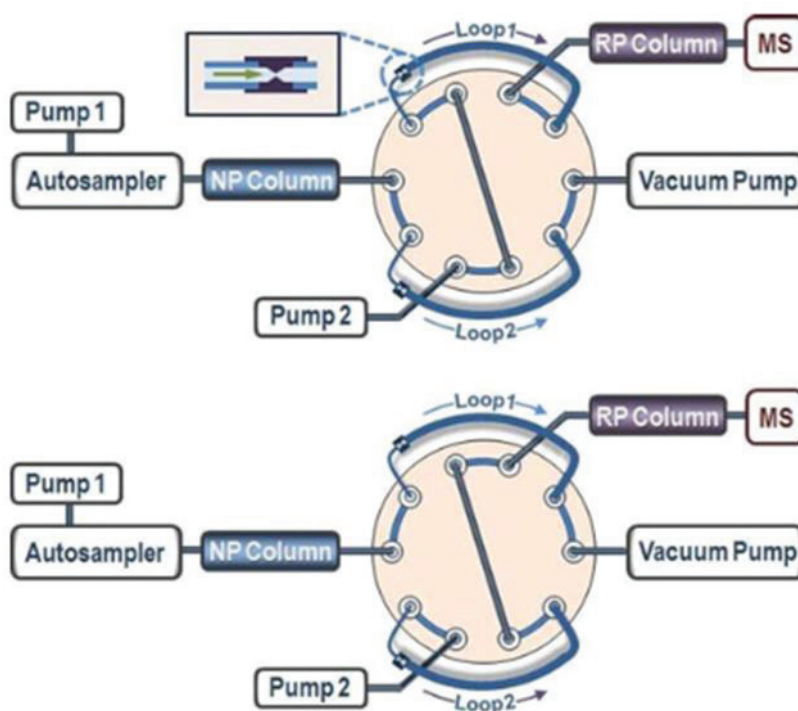
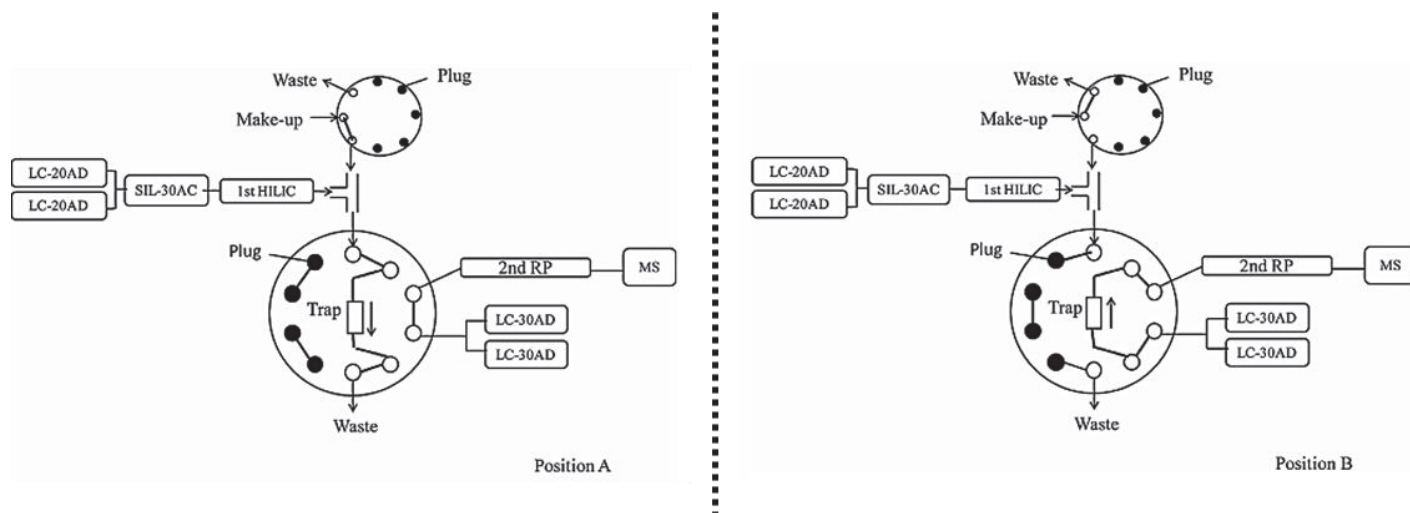


Figure 18 : Schéma du système d'interface 2D-LC NP-RP basé sur une vanne à deux positions à dix ports avec une interface de séchage en ligne [130].

Pendant qu'une fraction est analysée dans la 2<sup>ème</sup> dimension, la sortie de la colonne de 1<sup>ère</sup> dimension est reliée à une pompe à vide qui sèche la phase mobile. Lorsque la vanne permute, la phase mobile de la 2<sup>ème</sup> dimension resolubilise les lipides avant de les envoyer dans la colonne de phase normale. Parallèlement, la fraction suivante est de nouveau évaporée grâce à la pompe à vide et ainsi de suite. Malgré le fait que cette étape d'évaporation puisse être automatisable [126,127,130], cela augmente le risque de perte d'échantillon. De plus, la solubilisation des TG peut être compliquée sans la présence de solvants chloroformés. Pour finir, la HILIC est plus compatible que la NPLC avec les sources électrospray des spectromètres de masse (cf. partie 3.1).

#### 4.3.2.B. Approche à débit stoppé en lipidomique

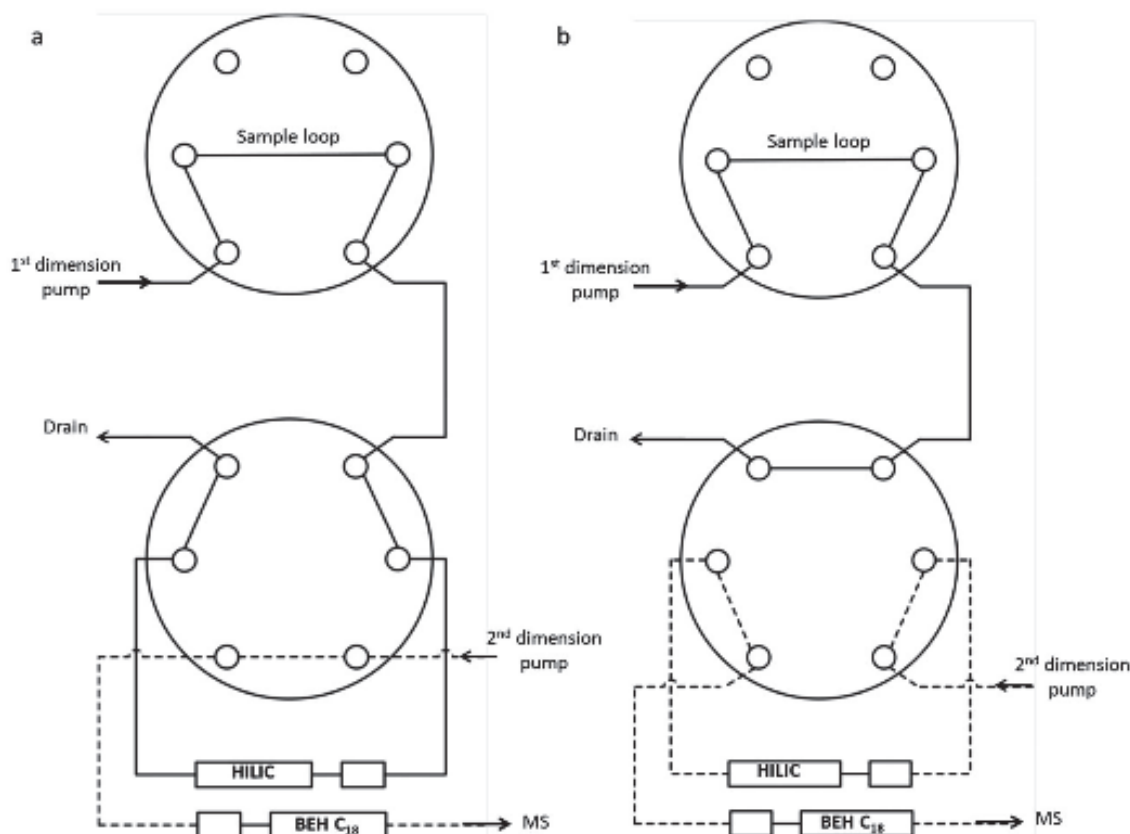
Différents types de 2D-LC en ligne sont décrits dans la littérature lors de l'analyse de lipide comme indiqué dans le tableau n°2. Le premier est la *stopflow* 2D-LC qui combine les avantages des méthodes en ligne ainsi que des méthodes hors lignes en stoppant le débit de la 1<sup>ère</sup> dimension lors de l'analyse d'une fraction en 2<sup>ème</sup> dimension [130–132]. Cependant, le fait de stopper le débit dans la 1<sup>ère</sup> dimension peut induire des élargissements de pics importants. Wang *et al.* ont imaginé un système de cartouche de piégeage accompagné d'un solvant de *make-up*, présenté sur la figure n°19, pour limiter le problème d'incompatibilité de solvant et d'élargissement de pic [132]. En effet, la cartouche entre les deux dimensions permet de capturer les lipides en tête de colonne, limitant ainsi la déformation de pic générée par l'arrêt du débit en D1 tout en diluant et éliminant la phase mobile issue de la 1<sup>ère</sup> dimension résolvant ainsi le problème de *breakthrough*.



**Figure 19 : Schéma d'un système 2D-LC à débit arrêté en 1<sup>ère</sup> dimension.** Les composés issus de la première dimension ont été piégés en position (A). Les composants piégés ont été injectés dans la deuxième colonne en position (B). En position (B), le débit est arrêté dans la 1<sup>ère</sup> dimension [132].

4.3.2.C. Approche *heart cut* en lipidomique

Le deuxième type de 2D-LC en ligne pour les lipides est la chromatographie *heart cut* (LC-LC) permettant de sélectionner une fraction issue de la 1<sup>ère</sup> dimension et de l'envoyer dans la 2<sup>ème</sup>. Comme indiqué dans la figure n°20, les fractions issues de la séparation en 1<sup>ère</sup> dimension (HILIC) sont directement transférées dans la 2<sup>ème</sup> dimension (RPLC) grâce à une rotation de vanne.



**Figure 20 : Schéma du système 2D-LC de type *heart cut*.** Avec la vanne en position 1 (à gauche), les lipides ont été retenus dans la colonne HILIC pendant que la phase mobile traverse la colonne. Avec une programmation en temps de la vanne (0,9 - 2,4 min pour les céramides/C1P et 3,2 - 6,0 min pour les PC/SM), la vanne a été déviée en position 2 (à droite). Les lipides retenus sont envoyés dans la colonne RP (BEH C18) grâce aux éluants de la pompe de deuxième dimension pour une séparation supplémentaire [128].

Sachant que le temps d'analyse en D2 est plus important que la différence de temps entre deux fractions issues de la D1, il n'est pas possible de stocker simultanément deux fractions indépendantes. Cette approche nécessite de devoir injecter l'échantillon autant de fois qu'il y a de fractions à analyser [128,129]. Pour pouvoir analyser successivement plusieurs fractions issues d'une même injection, il est nécessaire d'utiliser un système de vanne comportant des collecteurs de fractions. À ce jour, aucune méthode utilisant les vannes de *multiple heart cut* n'a été développée en lipidomique. Cependant, Li *et al.* [127] ont réussi à mettre au point une alternative aux vannes de collecte en se servant d'un système d'évaporation automatisable placé entre la dimension NPLC et RPLC (*cf.* figure n°18). Pendant qu'une fraction déjà évaporée issue de la 1<sup>ère</sup> dimension est envoyée dans la 2<sup>ème</sup> dimension, la fraction

suivante est récoltée dans une seconde boucle afin d'être à son tour évaporée, le tout sans arrêter le débit de la 1<sup>ère</sup> dimension. Il est important de noter qu'avec ce système il est obligatoire d'analyser la totalité du chromatogramme. Il n'est pas possible de sélectionner seulement certaines fractions à envoyer en 2D pour gagner du temps d'analyse lorsque l'on étudie des familles lipidiques en particulier (le système n'est pas capable d'éliminer des fractions entre les deux dimensions).

### 4.3.2.D. Approche *comprehensive* en lipidomique

Le dernier type de chromatographie bidimensionnelle utilisé en lipidomique est la 2D-LC *comprehensive* (LCxLC) qui consiste à transférer la totalité des fractions de la 1<sup>ère</sup> dimension vers la 2<sup>ème</sup>. Cette approche est celle qui présente le plus de facilité pour son automatisation et la grande capacité de séparation. Cependant, sa principale limitation est le temps d'analyse très court en 2<sup>ème</sup> dimension qui est imposé par le temps de modulation entre les deux dimensions. Cette limitation du temps d'analyse en D2 diminue le pouvoir de séparation de la 2<sup>ème</sup> dimension [123,133]. Les méthodes LCxLC ont déjà été évaluées pour les analyses lipidomiques à travers deux configurations, la RPLCxHILIC [123,133] et la HILICxRPLC [123]. *Xu et al.* ont démontré que la 2D-LC *comprehensive* permet bien d'obtenir de meilleurs résultats que la 1D-LC en termes de nombres de composés détectés [123]. Ils ont pu détecter 1784 lipides en RPLCxHILIC et 1059 lipides en HILICxRPLC contre 418 en RPLC et 212 en HILIC. Ces résultats montrent également que la configuration RPLCxHILIC permet de détecter le plus de lipides. Cette différence entre les deux configurations LCxLC s'explique par le fait que la RPLC, placée en D2, ne permet pas d'atteindre des capacités de séparations pour les lipides suffisantes en utilisant des temps de séparation de l'ordre de la minute. La HILIC permet de séparer les lipides malgré un temps d'analyse court [133]. Toutefois, placer la HILIC en 2<sup>ème</sup> dimension n'est pas la condition idéale, car les temps d'équilibrage pour ces types de phases chromatographiques sont longs [73] ce qui est un désavantage en 2D lorsque l'on cherche à avoir un gradient le plus rapide possible dans la dernière dimension.

Le développement de méthode hautement multiplexée permettant l'analyse simultanée de plusieurs milliers de composés associée à des techniques de séparation chromatographique performante permet la mise en place d'études multi-omiques. En plus de la méthode d'analyse, il est nécessaire d'optimiser la préparation d'échantillon qui se trouve en amont de la séparation chromatographique afin d'extraire chaque famille omique à partir d'un échantillon unique.

## **5. De la préparation omique individualisée à la préparation multi-omique à partir d'un échantillon unique**

### **5.1. Préparation d'échantillon en protéomique**

#### **5.1.1. Extraction des protéines**

Les deux grands domaines des sciences omiques, à savoir la génomique et la transcriptomique possèdent des méthodes d'extractions universelles avec des tampons standardisés. Cela est rendu possible par le fait que les acides nucléiques sont relativement uniformes vis-à-vis de leurs propriétés physicochimiques. Dans le cas de la protéomique, cela n'est pas envisageable, car les groupes fonctionnels des chaînes latérales ont une diversité importante en termes d'hydrophobicité et de charge. La conséquence est qu'il ne sera probablement jamais envisageable de mettre au point un tampon d'extraction protéique universel [134,135].

La première étape de la préparation d'échantillon protéique est l'extraction des protéines des échantillons biologiques en réalisant une lyse des membranes cellulaires qui peut être accompagnée du broyage de l'échantillon lorsque celui-ci est solide. Lors de l'extraction, il est nécessaire de maintenir le pH entre 7 et 8 pour que l'activité des protéases libérées en solution ne soit pas favorisée. De plus, il faut garder l'échantillon au frais et utiliser des inhibiteurs de protéase (EDTA, leupeptine, aprotinine, etc.) pour limiter la digestion prématurée et non contrôlée des protéines d'intérêts. Il est également possible d'utiliser des détergents comme le triton X-100 pour faciliter la solubilisation des protéines membranaires [135,136]. La présence de sel dans le tampon permet de réaliser un choc osmotique qui lyse les cellules fragiles. En effet, les cellules qui se trouvent dans une solution hypo-osmotique vont vouloir rétablir l'équilibre osmotique en faisant pénétrer de l'eau dans la cellule ce qui à terme provoque la rupture de la membrane plasmique. La deuxième étape consiste à éliminer les contaminants non protéiques comme l'ADN, les sucres, les lipides et les sels. La méthode la plus simple pour réaliser cette étape est la précipitation des protéines grâce à un solvant organique qui sépare les composés d'intérêts du reste des molécules toujours en solution [135]. La dernière phase de l'extraction protéomique est le fractionnement des protéines. Cette étape est cruciale pour la décomplexification des échantillons notamment pour la matrice plasma et sérum où plus de la moitié des protéines sont des albumines. Plusieurs méthodes existent pour éliminer les protéines majoritaires, mais aussi pour préconcentrer des protéines d'intérêts. La technique qui offre le plus de sélectivité est la chromatographie d'affinité [137] et l'immunoprécipitation [138]. Cependant, l'efficacité de ces méthodologies dépend de la disponibilité des supports d'affinité ou d'anticorps ainsi que de la qualité entre la relation protéine cible / anticorps.

### 5.1.2. Digestion des protéines lors de l'approche Bottom-up

Une fois extraites les protéines sont dénaturées grâce à une étape de réduction et d'alkylation en milieu salin afin de rompre les ponts disulfures. Cela a pour effet de détruire les structures secondaire et tertiaire des protéines. L'étape suivante est l'étape clé de l'approche Bottom-up qui consiste à réaliser une protéolyse par digestion enzymatique. Celle-ci est réalisée classiquement par une endoprotéase à sérine, la trypsine qui clive les liaisons peptidiques après les résidus arginine (R) et lysine (K) du côté C terminal [139]. La présence de ces résidus basiques à l'extrémité du C terminal permet de générer des ions « Y » (après fragmentation) de masse élevée facilitant ainsi l'interprétation des spectres de masse [140]. Néanmoins, certaines séquences d'acides aminés peuvent conduire à l'absence ou au mauvais clivage de la liaison peptide [141,142]. Pour minimiser ces phénomènes, plusieurs types de trypsine existent comme la trypsine TPCK qui augmente le rendement de peptides tryptiques lors de la digestion en diminuant l'activité chymotrypsique. La chymotrypsine hydrolyse préférentiellement les protéines après les acides aminés tyrosine, tryptophane, phénylalanine, méthionine ou leucine. Une fois l'échantillon digéré, il peut être purifié grâce à une étape d'extraction sur phase solide (SPE). Cette étape permet de réaliser un dessalage de l'échantillon et/ou de purifier les peptides. Elle permet également de préconcentrer l'échantillon et de le placer dans les conditions optimales pour la séparation chromatographique en changeant le solvant qui servira lors de l'injection en chromatographie liquide.

### 5.2. Préparation d'échantillon en lipidomique

Les applications biologiques en lipidomique nécessitent la mise en place de protocoles d'extraction rapides, répétables et capables d'extraire des composés possédant une forte variété de polarité. Ils doivent être également capables de décomplexifier la matrice en ne récupérant que les composés d'intérêts. Pour cela, il existe plusieurs procédures qui utilisent les extractions liquide/liquide (LLE) ou les extractions sur phase solide (SPE). Introduits dans les années 70, le protocole de Folch [143] puis celui de Bligh and Dyer [144] figurent parmi les plus utilisés. Ces protocoles utilisent un mélange ternaire composé de chloroforme, méthanol et d'eau afin d'induire une séparation de phase. Cette séparation permet de récupérer la phase organique inférieure qui contient les lipides et d'éliminer les métabolites polaires et les protéines qui restent dans la phase aqueuse. Le chloroforme peut être remplacé par le dichlorométhane qui est moins toxique [55]. En 2008 est introduite une nouvelle procédure par Matyash *et al.* [145]. Celle-ci remplace la phase organique par du methyl tert-butyl éther (MTBE) qui a l'avantage d'inverser l'ordre des phases lors de la LLE puisque le MTBE a une densité plus faible que l'eau. La phase organique correspond à la phase supérieure, ce qui facilite sa récupération et limite la pollution de l'échantillon par la phase aqueuse. Depuis son introduction, le protocole de Matyash a tendance à remplacer les protocoles employant le chloroforme [55]. Les méthodes SPE sont



aussi utilisées lors des analyses lipidomiques dans le cas où l'on souhaite fractionner les familles de lipides ou bien les préconcentrer. Cette procédure est recommandée si des fractions lipidiques spécifiques interfèrent lors des analyses, ou lorsque des caractérisations approfondies des classes de lipides sont nécessaires au lieu d'un profilage lipidomique plus complet [55]. Après les étapes de LLE et/ou de SPE, l'échantillon peut être évaporé puis resuspendu dans un solvant approprié avant injection.

### 5.3. Préparation d'échantillon en métabolomique

La grande diversité chimique des métabolites fait de leur extraction une tâche difficile [146]. Pour trouver un équilibre entre une interférence minimale de la matrice et une récupération maximale des métabolites, de nombreux efforts sont faits actuellement pour optimiser les protocoles d'extraction. Ceux-ci peuvent varier et dépendent considérablement des types d'échantillons utilisés ainsi que des gammes de concentration des métabolites à extraire. Typiquement, des métabolites présents en forte quantité dans des tissus seront suffisamment extraits pour pouvoir être détectés, cependant les métabolites minoritaires vont nécessiter une extraction plus complexe et spécifique pour être extraits en quantité suffisante. Les méthodes d'extractions usuelles en métabolomique comprennent l'extraction liquide-liquide (LLE), l'extraction solide-liquide (SLE), l'extraction en phase solide (SPE), l'extraction assistée par micro-ondes (MAE), *etc.* [146]. L'un des points communs de chaque protocole concernant l'extraction des métabolites est le *quenching* des métabolites dans leur matrice. Le *quenching* consiste en la limitation de la dégradation des métabolites due à plusieurs causes : dégradation enzymatique, dégradation thermique, dégradation chimique par métabolisation, *etc.* [147]. Idéalement, les techniques métabolomiques devraient représenter l'état métabolique au moment exact où un échantillon est prélevé et dans les conditions environnementales dans lesquelles il évolue. La congélation de l'échantillon à -20°C ou -80°C est fortement recommandée pour ralentir les cinétiques de dégradation et ainsi conserver la majorité des métabolites, mais cela peut ne pas être suffisant notamment dans le cas d'un stockage sur le long terme. De Koning et Van Dam ont été les premiers auteurs à combiner l'extinction rapide et la séparation des cellules du surnageant en une seule procédure [148]. Pour cela, ils ont utilisé du méthanol froid mélangé à de l'eau pour réaliser l'arrêt instantané du métabolisme en inactivant les enzymes par dénaturation tout en séparant les métabolites des protéines en les faisant précipiter.

Dans la littérature, la SLE est l'une des techniques les plus répandues pour les échantillons complexes solides tels que l'amphipode d'eau douce *G. fossarum* ainsi que les échantillons de plantes [65,66,149,150]. Les échantillons préalablement stockés à -20/-80°C sont homogénéisés sous forme de poudre puis extraits avec un mélange de solvants généralement organiques et polaires (méthanol, éthanol, ACN) associés à une phase aqueuse [151]. Le choix des solvants doit tenir compte de leur



compatibilité avec les instruments d'analyse sauf si le protocole se termine par une évaporation totale du solvant d'extraction. Le méthanol (MeOH) est souvent choisi comme solvant majoritaire pour l'extraction de métabolites, car il permet d'obtenir des rendements d'extractions satisfaisants à la fois pour les métabolites polaires et les lipides apolaires. Comme indiqué précédemment, il présente également un intérêt non négligeable pour le *quenching* de l'échantillon. Il est utilisé en mélange avec l'eau dans la plupart des cas avec une proportion supérieure pour la phase organique (MeOH/H<sub>2</sub>O ; 8/2 ; v/v par exemple) [66,151]. La LLE ou la SLE peut se finaliser de deux façons différentes : soit par évaporation totale du solvant d'extraction suivi de la reprise dans la phase mobile, soit par dilution de l'extrait. La dilution n'est possible que si l'on peut suffisamment diluer pour faire correspondre le solvant d'échantillon aux conditions du début du gradient chromatographique et ainsi éviter les phénomènes de *breakthrough*. (cf. 4.2.1). Il ne semble pas y avoir de différence notable entre les deux procédures au niveau des recouvrements et elles sont toutes deux assez présentes dans la littérature [151].

Dans le cas des analyses de matrices liquides, faibles en protéine, comme les urines, le type d'extraction typique est relativement plus simple. En effet, comme l'urine est composée à 95 % d'eau avec une teneur en protéines plus faible par rapport à des fluides biologiques comme le plasma, la préparation des échantillons est un processus simple dont la complexité diffère en fonction de l'objectif de l'expérience [152]. Dans le cas des analyses non ciblées, la procédure de préparation d'échantillon non sélective la plus simple est la dilution de l'échantillon avec les solvants appropriés suivi de l'injection directe dans le système analytique après une éventuelle étape de centrifugation [153]. Tandis qu'en analyse ciblée, où un groupe limité de métabolites est étudié, des techniques d'extraction analytique sélective, telles que la SPE ou la LLE, peuvent être utilisées à des fins de préconcentration et de nettoyage [152]. Pour ce qui est de l'extraction non ciblée des échantillons liquides, riches en protéines, comme le plasma sanguin, une étape de précipitation de ces macromolécules est nécessaire grâce à un solvant organique (MeOH). Comme pour l'urine, l'extraction ciblée du plasma reste dominée par la SPE et la LLE afin de séparer les métabolites d'intérêts des autres biomolécules [154].

Chacune des procédures citées ci-dessus a été optimisée pour chaque omique. Cependant, l'application de protocoles différents lors d'une étude multi-omique présente de nombreux désavantages en termes de temps d'analyse et de capacité d'interprétation des données.

## 5.4. Protocole d'extraction multi-omique à partir d'un échantillon unique

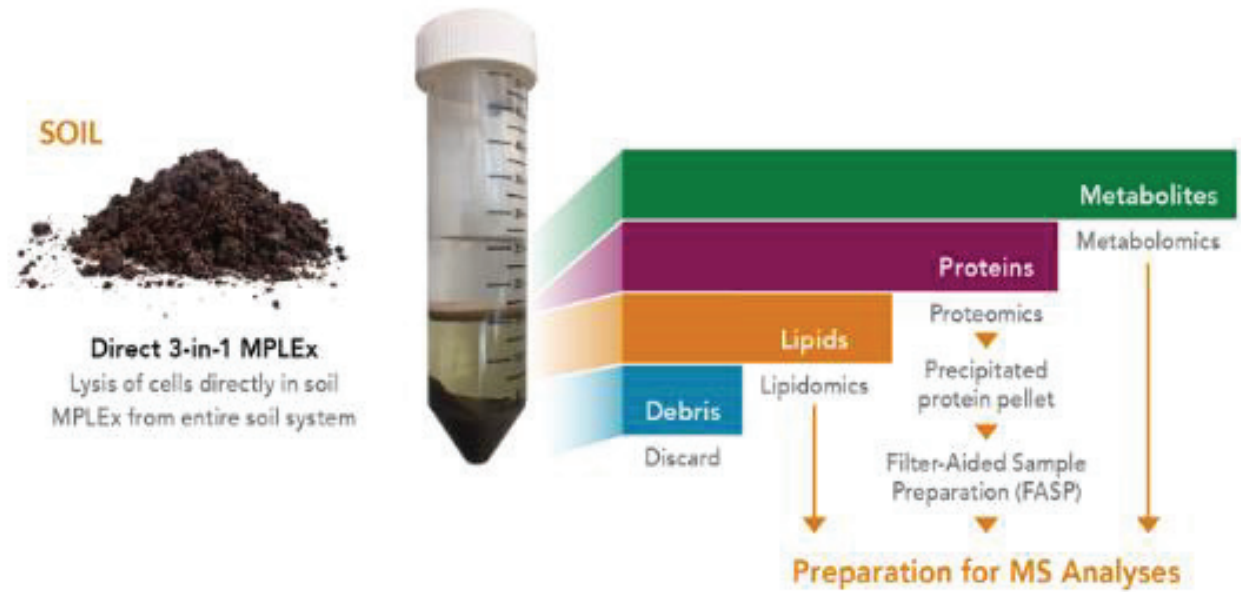
Les méthodes d'extraction multi-omiques doivent permettre l'extraction d'un grand nombre de molécules présentant des propriétés physicochimiques très différentes les unes des autres (acides nucléiques/protéines/métabolites). Dans l'idéal, ces protocoles d'extractions doivent être réalisés à partir d'un échantillon unique afin de limiter au maximum les variations interindividus. Cela permettra d'avoir une meilleure corrélation des résultats en facilitant l'intégration des données statistiques puisque l'ensemble des données omiques proviendront du même individu [155]. Les extractions multi-omiques uniques permettent également d'économiser des d'échantillons précieux en limitant le nombre d'extractions. Il ne sera donc plus nécessaire de posséder plusieurs répliques d'un échantillon afin de réaliser chaque omique. Cette procédure unique permet également un gain de temps, car la récupération des différentes fractions omiques à partir d'extractions distinctes est longue [156–158]. Les protocoles expérimentaux en multi-omique à partir d'un échantillon unique sont dérivés de l'extraction des métabolites et des lipides avec comme première étape l'ajout d'un solvant organique froid (MeOH) [158] afin d'extraire et de stopper le métabolisme. Les solvants d'extraction des lipides sont ajoutés pour séparer les métabolites polaires et apolaires dans des phases distinctes. Pour finir, les composés ayant précipité au contact des solvants organiques, comme les protéines, sont récoltés.

Les deux principaux protocoles pour l'extraction simultanée des lipides, des protéines et des métabolites polaires sont le protocole MPLEx pour "Metabolite, Protein and Lipid Extraction" [158] et le protocole SIMPLEX pour "Simultaneous Metabolite, Protein, Lipid EXtraction" [159].

### 5.4.1. Protocole SIMPLEX et MPLEx

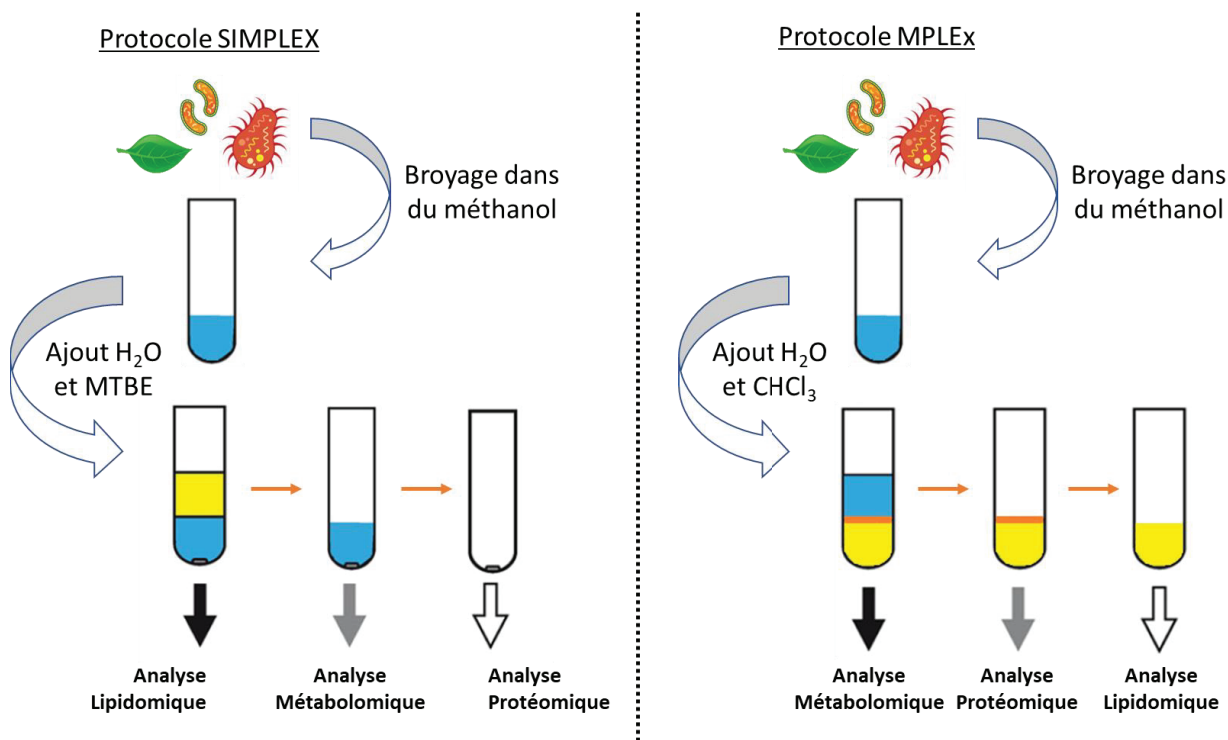
Le MPLEx et le SIMPLEX se basent respectivement sur les extractions de Bligh and Dyer [144] et de Matyash [145] qui sont utilisées en lipidomique et qui ont été adaptées pour l'analyse des autres omiques comme la protéomique et la métabolomique.

La méthode MPLEx utilise un mélange ternaire à base de chloroforme/méthanol/eau permettant d'avoir une séparation de phases comme représentée sur les figures n°21 et 22. La phase supérieure correspond à la fraction aqueuse qui contient les métabolites polaires et la phase inférieure correspond à la fraction organique contenant les lipides. Ces deux phases sont séparées par l'interphase qui contient les protéines [158,160,161].



**Figure 21 : Illustration d'un échantillon issu du processus MPEX.** Cette image présente la séparation de phase dans laquelle les métabolites, les protéines et les lipides sont simultanément extraits du même échantillon de sol pour des analyses par spectrométrie de masse [160].

Le protocole SIMPLEX s'appuie sur les mêmes principes que le MPEX en utilisant un mélange MTBE/méthanol/eau représenté sur la figure n°22.

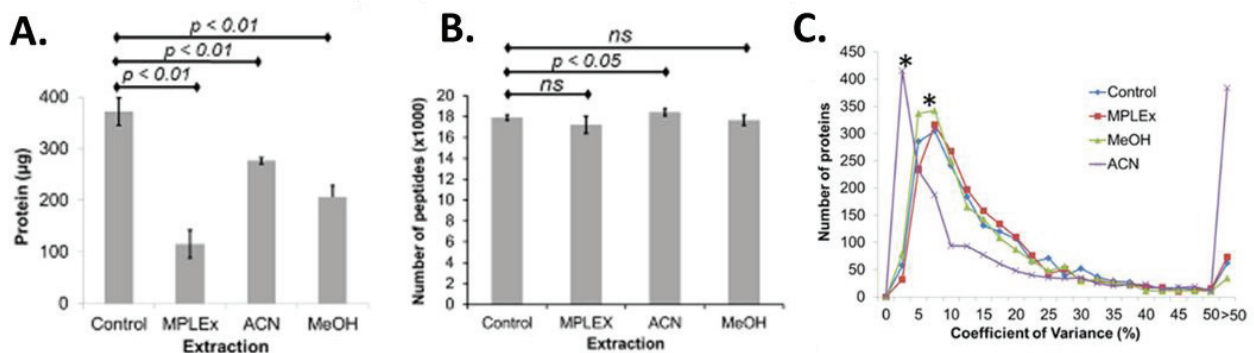


**Figure 22 : Protocole simplifié des procédures SIMPLEX et MPEX.** Ces deux protocoles permettent de réaliser une séparation de phases dans laquelle les métabolites, les protéines et les lipides sont simultanément extraits du même échantillon. Le protocole SIMPLEX utilise un mélange ternaire composé de H<sub>2</sub>O/MeOH/MTBE qui conduit à l'obtention d'une phase organique supérieure. Le protocole MPEX utilise un mélange ternaire composé de H<sub>2</sub>O/MeOH/chloroforme qui conduit à l'obtention d'une phase organique inférieure.

L'utilisation du MTBE à la place du chloroforme permet d'inverser l'ordre des fractions lors de la séparation de phase en utilisant un solvant de densité plus faible que l'eau. La fraction lipidique devient la phase supérieure et la fraction des métabolites polaires devient la phase inférieure. L'autre différence notable entre ces deux protocoles est la position des protéines. En effet, les protéines se trouvent sous forme de culots protéiques ayant précipité au fond du tube dans la phase aqueuse ce qui permet une récupération plus facile que lorsqu'elle se trouve à l'interphase [145,159]. Ainsi lorsque l'on souhaite automatiser le protocole d'extraction multi-omique grâce à des systèmes robotisés de manipulation des liquides il est préférable d'utiliser la procédure SIMPLEX [162].

Les protocoles MPLEx ou SIMPLEX permettent de choisir quelle fraction on souhaite mettre en valeur en choisissant quelle sera la phase supérieure afin de faciliter sa récupération sans risquer la contamination par les autres [53,145]. De plus, le protocole utilisant le MTBE serait, d'après de récentes publications, plus efficace pour l'extraction des glycérophospholipides, des céramides et des acides gras insaturés, tandis que le protocole au chloroforme serait plus efficace pour la récupération des acides gras saturés et des plasmalogènes [53,163].

Nakayasu *et al.* ont comparé le protocole MPLEx et deux protocoles classiquement utilisés en protéomique [158]. À travers la figure n°23, ils ont démontré que même si le protocole multi-omique possède un taux de récupération des protéines moins important que les protocoles de référence (*cf.* figure n°23 A) il permet d'obtenir une couverture de protéome comparable. En effet, la figure n°23 B indique que le nombre de peptides détectés est similaire pour l'ensemble des protocoles.

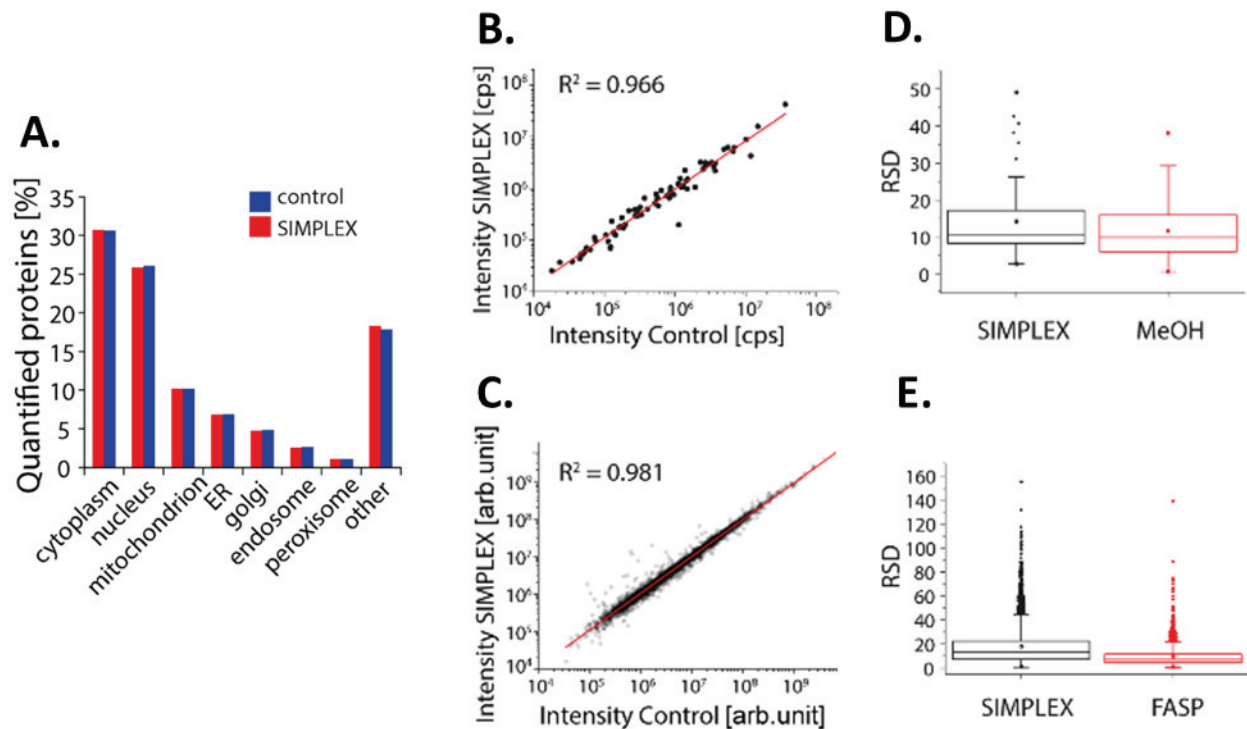


**Figure 23 : Comparaison entre le protocole MPLEx et des protocoles de référence en protéomique.** Extraction des protéines de *S. oneidensis* selon les protocoles MPLEx, acétonitrile (ACN) et méthanol (MeOH). Un échantillon a été digéré avec de la trypsine sans extraction préalable (Control) comme contrôle. (A) Récupération des protéines après extraction. (B) Nombre de peptides identifiés dans les différentes extractions. (ns = non significatif). (C) Distribution des coefficients de variation entre les protéines identifiées dans les échantillons extraits avec différentes méthodes [158].

Aucune différence significative dans les distributions des coefficients de variation (CV) n'a été observée en comparant le MPLEx aux autres protocoles puisque les CV de la grande majorité des protéines sont inférieurs à 25% (*cf.* figure n°23 C). L'extraction à l'ACN présente une distribution bimodale, avec des

CV très faibles et très élevés (*cf.* figure n°23 C), ce qui suggère que certaines protéines ne sont pas précipitées de manière reproductible avec ce solvant. Ce phénomène pourrait être dû au fait que l'ACN ne précipite pas complètement les petites protéines [164].

Coman *et al.* ont réalisé le même type d'étude en comparant cette fois-ci le protocole SIMPLEX à un protocole utilisé classiquement en protéomique et un protocole utilisé en métabolomique. Leurs résultats présentés sur la figure n°24 montrent une couverture du protéome et du métabolome avec une forte corrélation [159]. En effet, le nombre de peptides détectés et quantifiés est similaire entre le protocole SIMPLEX et celui de protéomique de référence (*cf.* figure n°24 A). De plus, que ce soit en protéomique ou en métabolomique les intensités du protocole multi-omique et de référence (*cf.* figures n°24 B et 24 C) présentent une forte corrélation et une excellente répétabilité (*cf.* figures n°24 D et 24 E) avec des CV inférieurs à 20%.

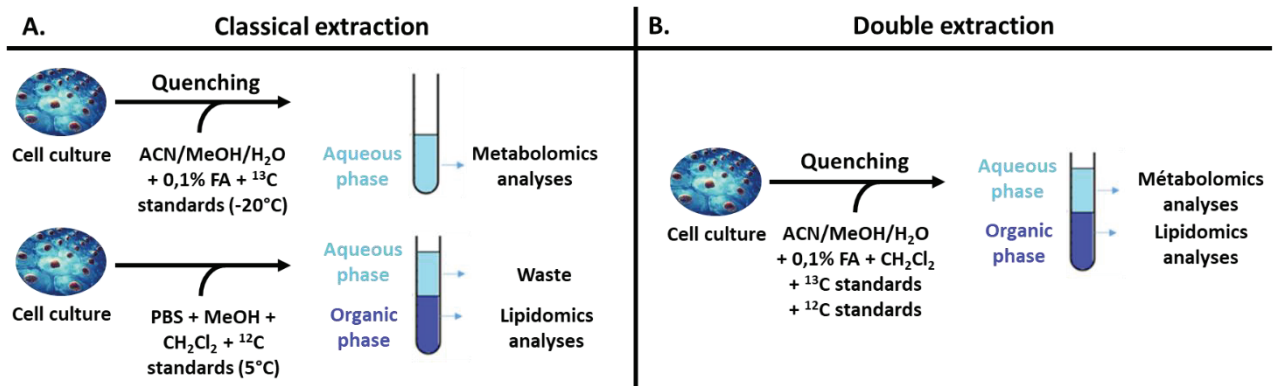


**Figure 24 : Comparaison entre le protocole SIMPLEX et un protocole de référence en protéomique et métabolomique.** (A) Distribution relative des protéines dans différents compartiments cellulaires entre le système SIMPLEX et le protocole de référence en protéomique. Au total, 3327 protéines ont été quantifiées. (B) Corrélation entre l'intensité des métabolites obtenue avec le protocole SIMPLEX et celui de référence. (C) Corrélation entre l'intensité des peptides obtenue avec le protocole SIMPLEX et celui de référence. (D) et (E) Moyenne, médiane de l'écart type des métabolites et des protéines pour le protocole SIMPLEX et la référence avec leurs erreurs associées (n=3 pour toutes les expériences) [159].

#### 5.4.2. Protocole à double extractions

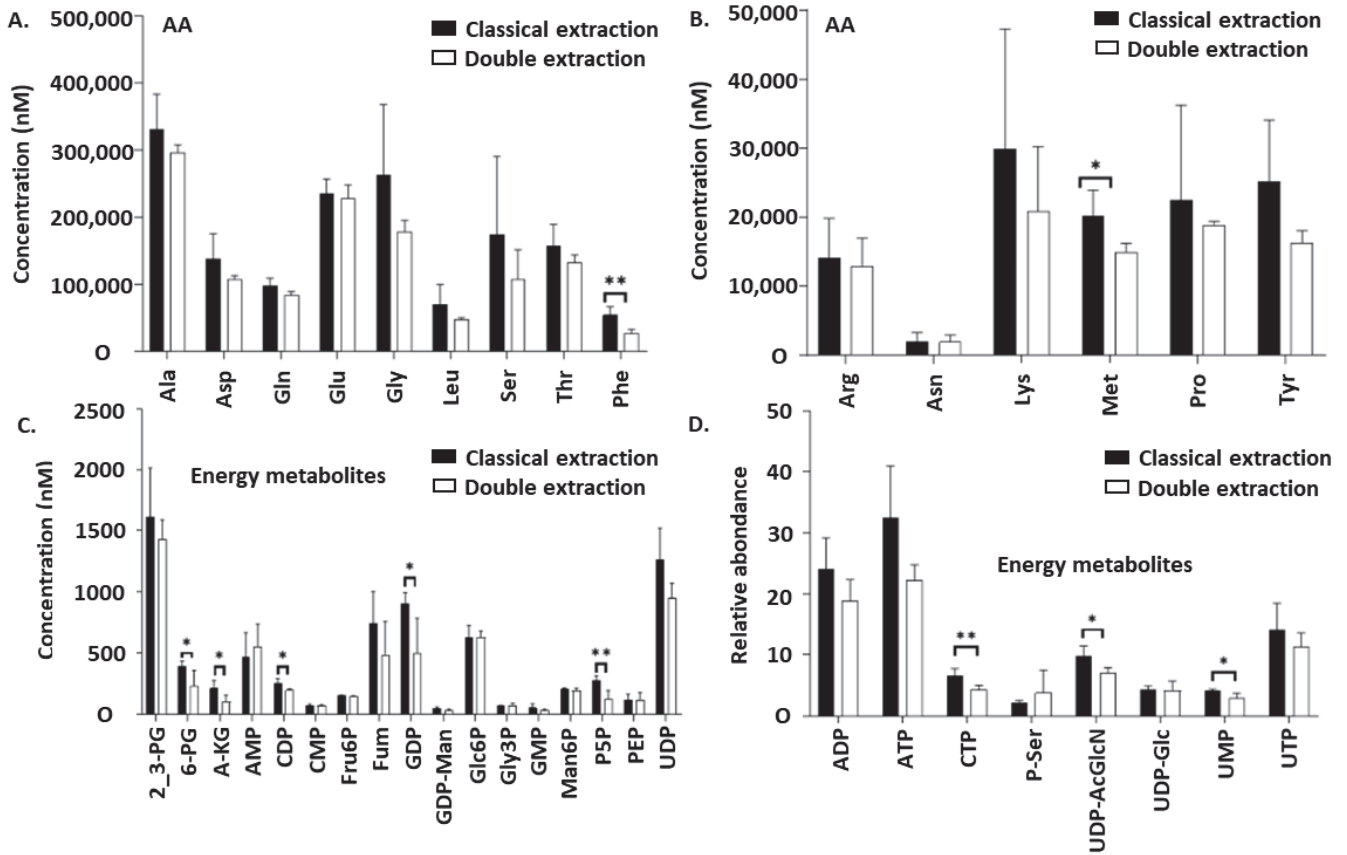
Villaret-Cazadamont *et al.* ont modifié le protocole MPLEX en rajoutant une étape permettant d'arrêter le métabolisme [155]. Pour cela, ils réalisent une première extraction en ajoutant un mélange d'ACN/MeOH/H<sub>2</sub>O (2/2/1; v/v/v) à 0.1% d'acide formique pour figer le système puis du

dichlorométhane est ajouté pour réaliser l'extraction de Bligh and Dyer, séparant les métabolites polaires des lipides (cf. figure n°25). Ces deux extractions successives expliquent le nom de « protocole à double extractions ».



**Figure 25 : Schéma expérimental des méthodes d'extraction classique et double extractions utilisées pour effectuer des analyses métabolomiques et lipidomiques.** Pour l'extraction classique (A), un mélange d'acétonitrile (ACN), méthanol (MeOH), eau acidifiée (H<sub>2</sub>O) avec de l'acide formique (FA) a été utilisé pour effectuer l'extraction des métabolites polaires. Un tampon salin phosphaté (PBS) associé au dichlorométhane (CH<sub>2</sub>Cl<sub>2</sub>) et au MeOH a été utilisé pour extraire les lipides. Pour la méthode à double extractions (B), le mélange d'ACN, de MeOH et d'eau acidifiée a été complété par du CH<sub>2</sub>Cl<sub>2</sub> pour extraire les métabolites polaires et les lipides à partir d'un échantillon unique. Ajout de standards internes isotopiquement enrichis pour effectuer une quantification absolue pour l'analyse métabolomique. En lipidomique la quantification est relative (Adapté de la publication de Villaret-Cazadamont *et al.* [155]).

Pour la majorité des métabolites polaires étudiés, la concentration obtenue avec le protocole à double extractions est similaire au protocole de référence en métabolomique comme le montre la figure n°26. Seuls quelques composés polaires présentent une extraction plus faible avec le protocole à double extractions par rapport à celui de référence. Cela peut être expliqué par l'ajout de la phase chloroformique qui induit une séparation de phase. En effet, les métabolites de polarité intermédiaire sont répartis entre les deux phases ce qui peut entraîner une sous-évaluation de leur concentration dans la phase aqueuse. Cependant, ces différences entre les deux protocoles sont mineures.



**Figure 26 : Comparaison de la récupération des métabolites polaires avec la méthode classique (barres noires) et la méthode double extractions (barres blanches).** Les acides aminés (A, B) et les métabolites énergétiques (C, D) ont été quantifiés en utilisant le rapport 12C/13C pour chaque métabolite. Les concentrations (A, B et C) ont été exprimées en nM en utilisant un étalonnage externe. L'abondance relative (D) a été obtenue en utilisant uniquement les rapports 12C/13C. Les barres représentent la moyenne et l'écart type (n=5). Des tests "t" ont été effectués pour chaque métabolite (\*p<0,05 ; \*\* p < 0,01) [155].

Concernant la récupération des lipides, il a été démontré, à travers les résultats de la figure n°27, que le protocole à double extractions permet d'obtenir un meilleur taux de récupération pour le cholestérol et les céramides. Inversement, une extraction significativement meilleure a été observée pour la procédure d'extraction classique pour la famille des TG, PC, PE et PI. Pour finir, les SM ont été extraites de manière similaire par les deux protocoles. Bien que certaines classes de lipides aient été extraites plus efficacement par l'une des deux méthodes, la distribution relative des espèces de lipides dans chaque famille n'a pas été influencée par le protocole d'extraction (exemple de distribution pour la famille des TG et des SM sur la figure n°28). Ces différences peuvent être expliquées par la différence de proportion entre les différents solvants utilisés lors de l'extraction liquide/liquide et l'ajout de l'acétonitrile dans la première partie du protocole à double extractions.



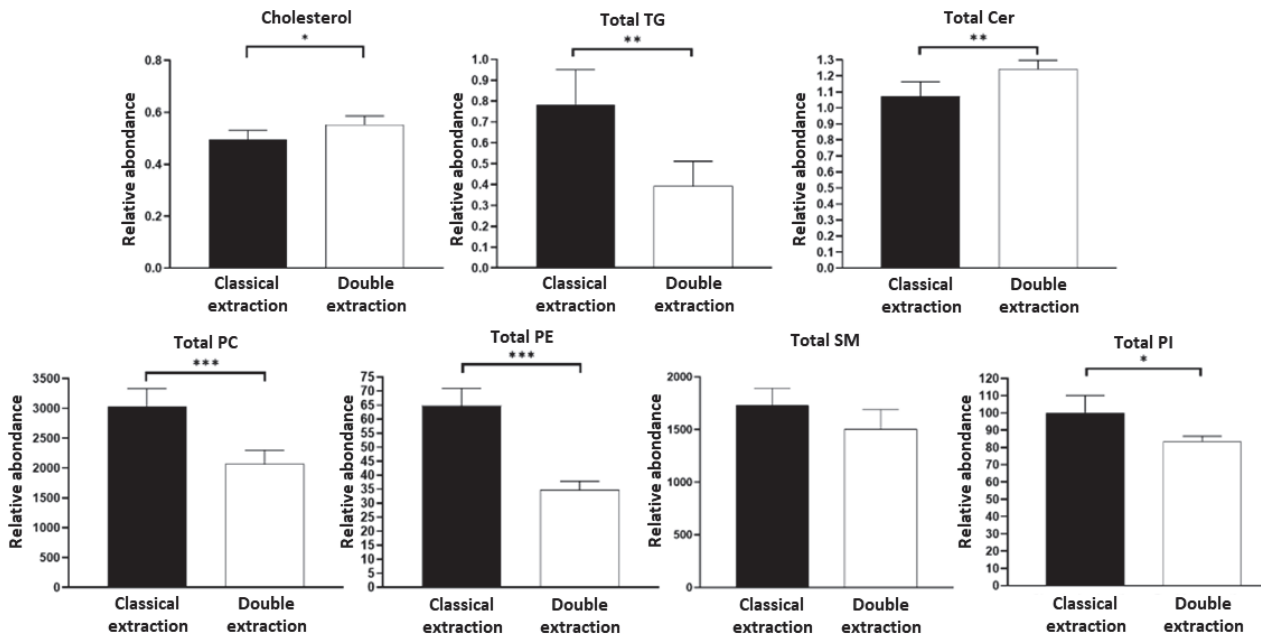


Figure 27 : Comparaison de la quantification des classes de lipides entre la méthode classique et la méthode de double extractions. Les classes de lipides totaux correspondent à une addition de toutes les espèces moléculaires (rapport avec leur standard interne) obtenues pour chaque classe. Les barres représentent la moyenne  $\pm$  l'écart type (n = 5). Des tests « t » ont été effectués pour chaque métabolite (\* p < 0,05 ; \*\* p < 0,01 ; \*\*\* p < 0,001) [155].

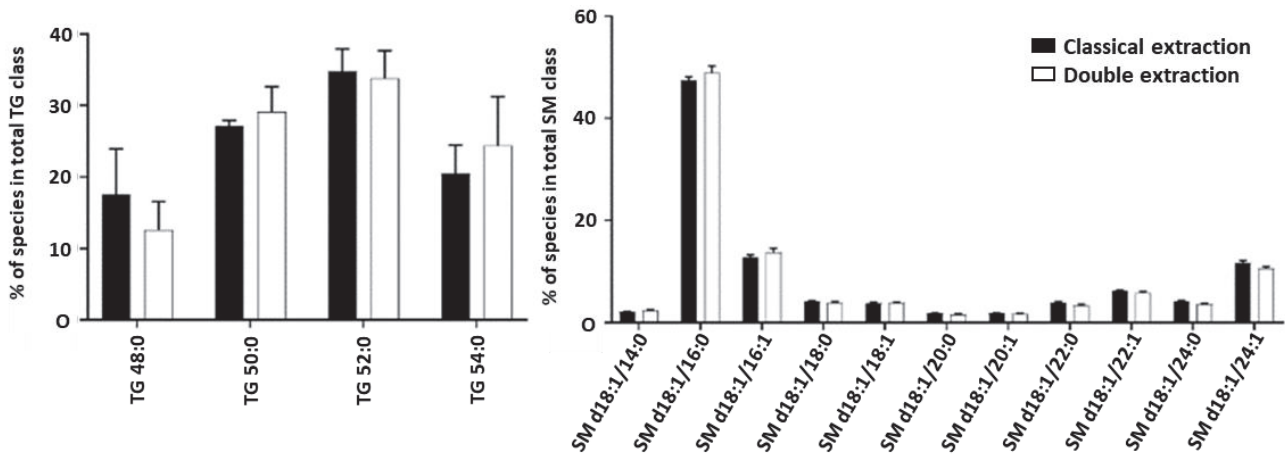


Figure 28 : Comparaison de la distribution des espèces de sphingomyélines (SM) et de triglycérides (TG) entre les méthodes d'extraction classique et double extractions. Chaque espèce est représentée comme une proportion (%) de sa classe totale. Les classes de lipides totales sont considérées comme 100%. Les barres représentent la moyenne des écarts types (n=5) [155].

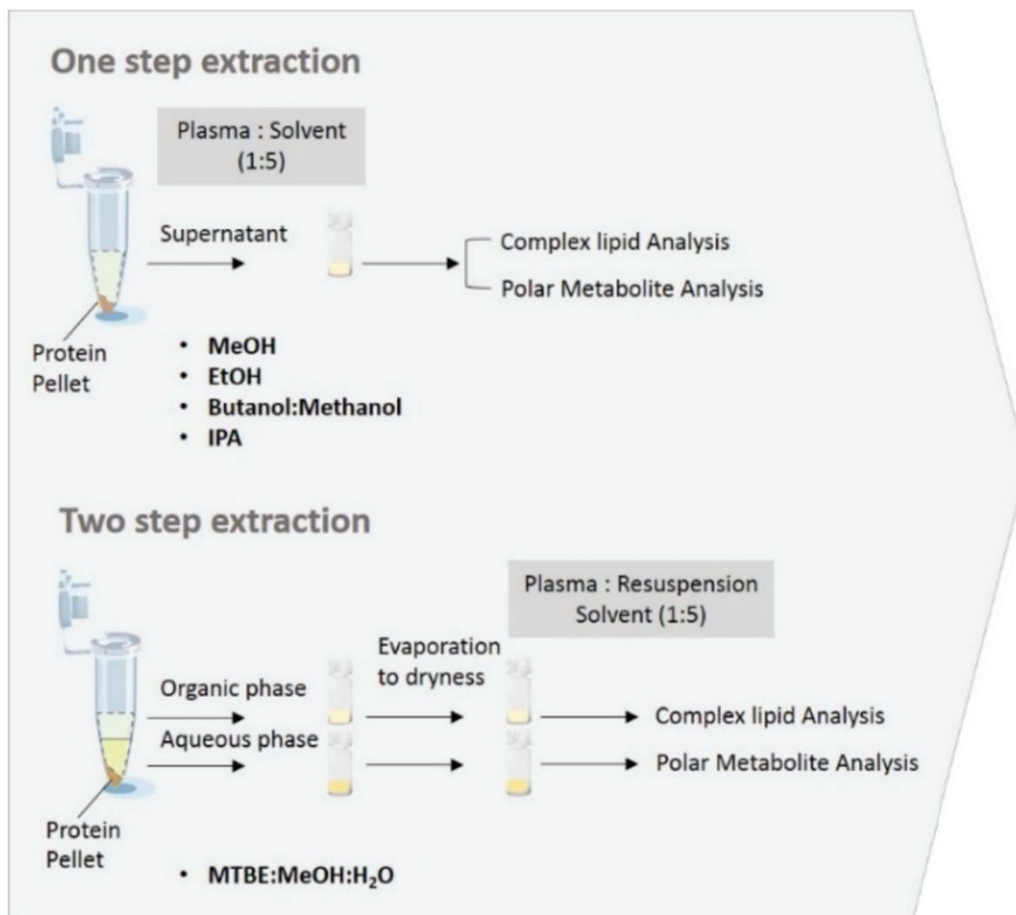
Dans cette étude, les protéines ne sont pas étudiées, mais ce protocole est adapté pour ce type d'omique puisque les solvants chloroforme/méthanol sont compatibles avec l'extraction de composés protéiques comme démontré par le protocole MPLEx.

Dans le cas des protocoles SIMPLEX, MPLEx et à double extractions on peut s'attendre à ce que l'extraction ne soit pas optimale pour les lipides hydrophiles, les lipoprotéines où les métabolites amphiphiles. En effet, ces molécules risquent d'être distribuées entre les deux phases ce qui rendra leurs analyses difficiles [165].



### 5.4.3. Protocole « Single-Step extraction » avec séparation chromatographique en HILIC

Dans le cadre de l'étude de Medina *et al.* en 2020 [166], l'équipe de Julijana Ivanisevic a évalué les performances de plusieurs protocoles utilisant une seule étape d'extraction simultanée de lipides complexes et de métabolites polaires à partir d'un échantillon unique comme indiqué dans la figure n°29. Les extractions ont été réalisées avec des solvants couvrant une gamme de polarité importante méthanol (MeOH) > éthanol (EtOH) > isopropanol (IPA) > BUME > MTBE, le mélange Butanol/méthanol (BUME) se situant au milieu de cette gamme.



**Figure 29 : One step extraction versus extraction MTBE : Préparation des échantillons.** Les performances des méthodes d'extraction en une seule étape, y compris le méthanol (MeOH), l'éthanol (EtOH), le BUTanol:MEthanol (BUME) et l'isopropanol (IPA), ont été comparées au mélange de solvants considéré comme le meilleur compromis, c'est-à-dire MTBE:MeOH:H<sub>2</sub>O pour l'extraction des lipides complexes et MeOH pour l'extraction des métabolites polaires (Adapté de la publication de Medina *et al.* [166]).

Dans un premier temps, l'efficacité de précipitation des protéines de ces quatre solvants a été évaluée en mesurant la teneur en protéines totales dans le surnageant et dans le culot. L'équivalent de 95 % d'élimination des protéines a été atteint avec les quatre méthodes démontrant ainsi la capacité du protocole *Single-step* à séparer les protéines de la matrice.

Ensuite, l'abondance relative du signal de l'ensemble du panel d'espèces lipidiques détectées a été évaluée dans les quatre extraits de plasma. Le méthanol a démontré la plus faible performance pour l'extraction des lipides complexes, à l'exception des lysophosphatidylcholines (LPC) et des lysophosphatidylinositols (LPI). Un signal significativement plus faible pour les sphingomyélines (SM), les acides gras et les lysophospholipides a également été obtenu à partir des extraits d'éthanol, par rapport à l'IPA et au BUME. Par conséquent, le MeOH et l'EtOH ont été exclus de la suite des évaluations, car ils sont moins performants pour l'extraction de lipides complexes. Les meilleurs candidats, BUME et IPA, ont été évalués par rapport à l'extraction biphasique de Matyash (MTBE), pour déterminer leur capacité à extraire de manière reproductible les lipides et les métabolites polaires. Les espèces lipidiques représentatives des différentes classes de lipides ont été extraites avec succès par les trois solvants testés, à l'exception des LPI, pour lesquels le signal dans l'extrait MTBE était proche de la limite de détection (LOD), comme le montre la figure n°30. Parmi les sphingolipides, le signal le plus élevé a été obtenu avec les extraits d'IPA et de MTBE. Dans le cas des glycérolipides (mono, di et triacylglycérols), les signaux les plus élevés ont été obtenus après une extraction en une seule étape avec le BUME et l'IPA, bien que le signal relatif n'ait pas été significativement différent du signal de référence de l'extrait MTBE. Parmi les glycérophospholipides, le signal le plus élevé a été obtenu avec l'IPA pour les PS. Pour les autres sous-classes, la différence entre les trois solvants n'était pas significative. Pour les lysophospholipides, l'abondance relative du signal était en général significativement plus élevée pour les extraits IPA et BUME par rapport à la phase MTBE.

Dans le cas des métabolites, des signaux significativement plus élevés pour les acides aminés, les acides carboxyliques et les hydrates de carbone ont été observés après l'extraction avec la méthode MeOH de référence. Au contraire, les nucléosides, les alkylamines, les acylcarnitines, et d'autres métabolites polaires ont montré des signaux significativement plus élevés après l'extraction avec IPA, par rapport à BUME et MeOH. Pour les dérivés d'acides aminés et les acides nucléiques, l'intensité du signal n'a pas varié de manière significative entre les extraits MeOH, BUME et IPA. Les résultats sont présentés sur la figure n°31.

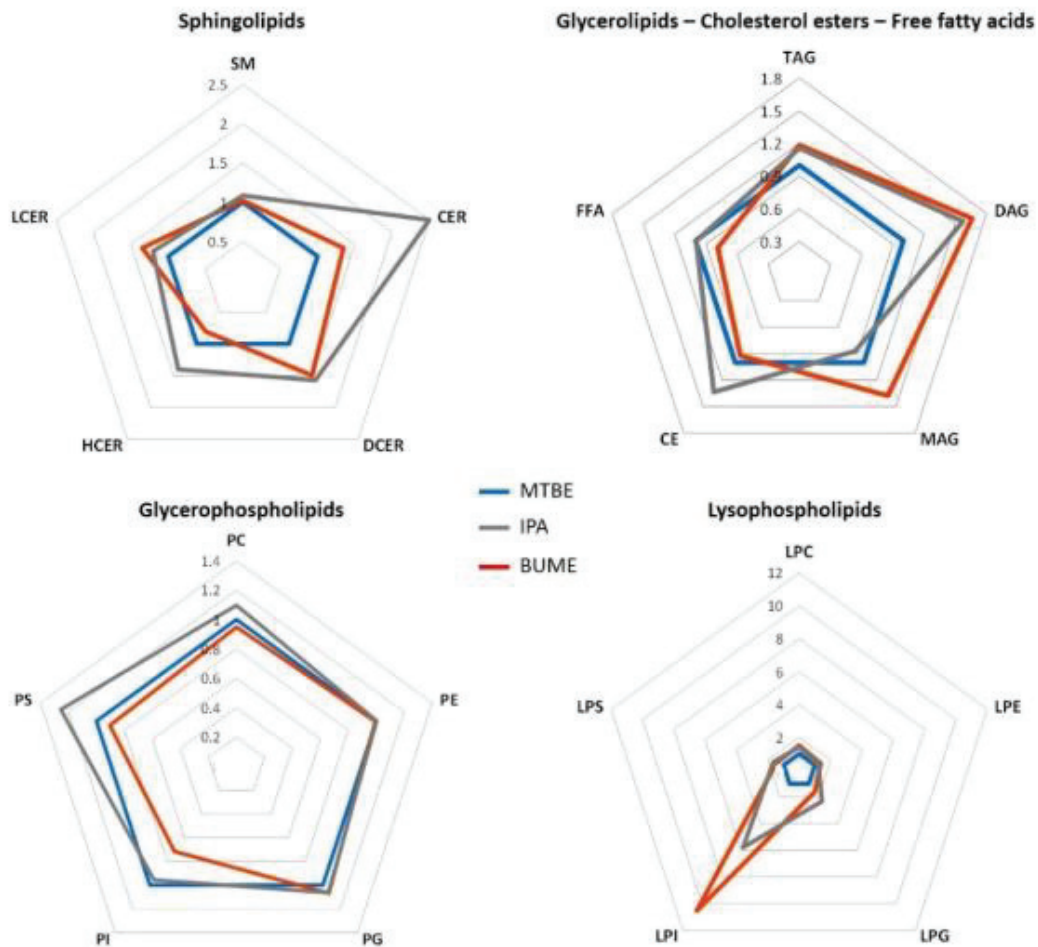


Figure 30 : Abondance relative du signal (par classe de lipides) par rapport au signal récupéré avec l'extraction biphasique à l'éther méthyl-tert-butylrique (MTBE). Les valeurs indiquées sur les graphiques en araignée représentent l'abondance relative du signal pour chaque méthode en une seule étape, en utilisant l'IPA ou le BUME, par rapport à l'extrait MTBE de référence [166].

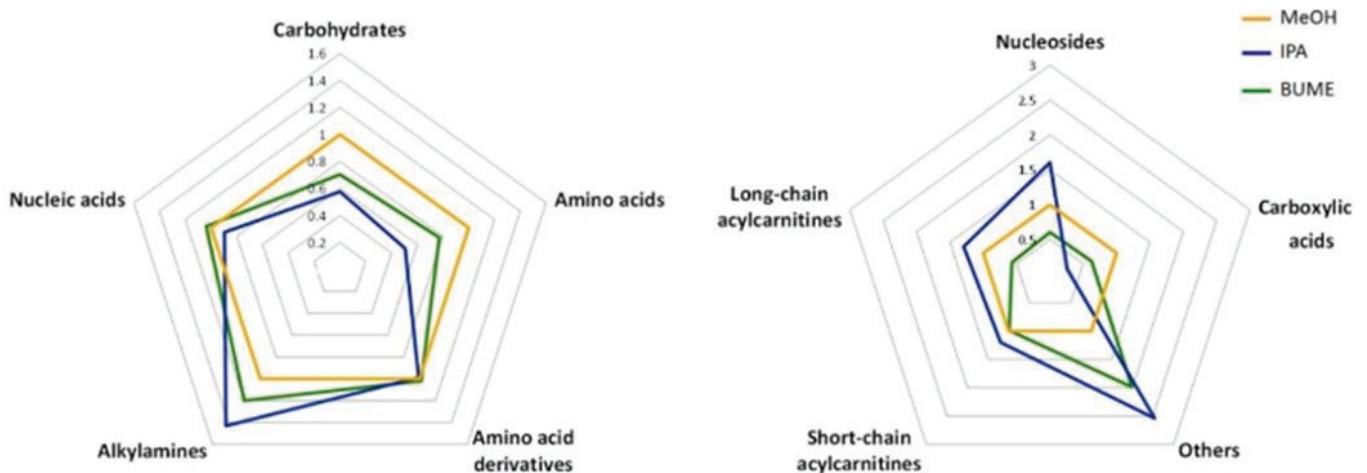


Figure 31 : L'abondance relative des classes de métabolites polaires dans les extraits IPA et BUME par rapport à l'extrait plasmatique MeOH. Les valeurs indiquées sur les graphiques en araignée représentent l'abondance relative du signal pour chaque méthode en une seule étape, en utilisant IPA ou BUME, par rapport à l'extrait MeOH de référence. La classe "Autres", c'est-à-dire les autres métabolites polaires, comprend le glycocholate, l'hydroquinone, l'hydroxyphényllactate, la pyridoxine, le salsolinol, la trigonelline et la tryptamine. Les acides carboxyliques comprennent les acides mono-, di- et tri-carboxyliques [166].

Les résultats présentés dans cette étude démontrent les avantages d'utiliser le protocole Single-step par rapport au protocole de Matyash. Celui-ci est plus adapté à l'extraction des lipides et des métabolites polaires, car il permet une meilleure récupération des familles lipidiques et classes métaboliques en plus de la simplification du protocole. Cependant, cette approche, qui a été mise en place avec une séparation chromatographique HILIC pour les deux familles omiques, n'est pas envisageable dans le cas de la chromatographie liquide de phase inverse. Le protocole Single-step induit que les métabolites polaires et les lipides se retrouvent dans le même solvant d'injection et donc que le type de chromatographie soit identique. Bien que les lipides se comportent très bien en RPLC, ce n'est pas le cas de tous les métabolites qui présentent peu de rétention dans ce mode chromatographique.

### 5.4.4. Protocole de Vorreiter

Le dernier protocole présenté dans ce rapport bibliographique est basé sur l'ajout d'un mélange méthanol/chloroforme en amont de l'ajout d'un mélange phénol/chloroforme. Il s'agit de la procédure la plus complète, mais aussi la plus complexe à mettre en œuvre, car elle permet d'extraire de manière simultanée l'ADN, l'ARN, les protéines et les métabolites [167] comme présenté sur la figure n°32. L'ajout du mélange H<sub>2</sub>O/MeOH/CHCl<sub>3</sub> (50/45/5 ; v/v/v) permet de faire précipiter les protéines, l'ADN et l'ARN pour ne garder dans la phase monophasique que les métabolites. Le culot est ensuite extrait avec un mélange phénol/CHCl<sub>3</sub> qui permet d'obtenir un mélange triphasique avec l'ARN qui se trouve dans la phase aqueuse supérieure, l'ADN au niveau de l'interphase et les protéines au sein de la phase organique supérieure.

Bien que Vorreiter *et al.* [167] n'étudie pas les lipides dans leur méthode, ce protocole semble être adapté pour l'analyse des lipides puisque les solvants chloroforme/méthanol sont utilisés. Il suffit d'augmenter le ratio de chloroforme après avoir récupéré la phase métabolomique pour réaliser l'extraction de Bligh and Dyer [144].

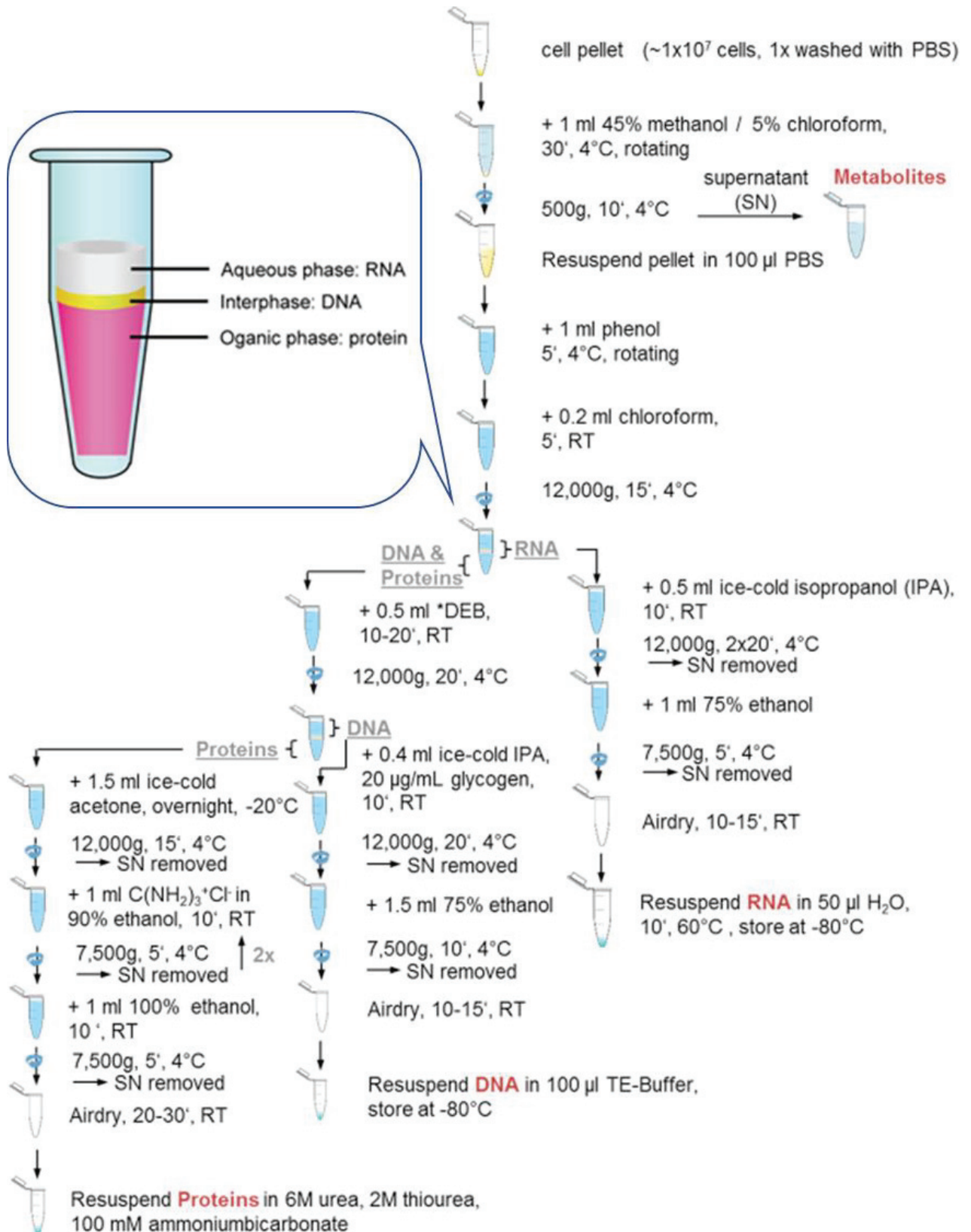


Figure 32 : Schéma de la procédure d'extraction permettant d'extraire et de séparer les protéines/métabolites/ADN/ARN à partir d'un extrait cellulaire unique (Adapté de la publication de Vorreiter *et al.* [167]).

## 6. Les sciences omiques en écotoxicologie

Les outils analytiques de type LC-MS/MS dans le cadre des analyses omiques sont d'ores et déjà utilisés dans les études cliniques pour la détermination de biomarqueurs. Depuis quelques années, l'ensemble des méthodologies mises en place dans le domaine de la santé a été utilisé par les écotoxicologues afin d'évaluer la dégradation des écosystèmes. De nombreuses innovations d'un point de vue analytique restent à mettre en place dans ce domaine vis-à-vis des sciences omiques ce qui explique que ces travaux de thèses ont été réalisés dans un contexte écotoxicologique.

### 6.1. Écotoxicologie et biomarqueur

#### 6.1.1. Qu'est-ce que l'écotoxicologie ?

Pour bien comprendre le concept de l'écotoxicologie, il est important de remonter à l'origine des deux termes qui le composent, à savoir la toxicologie et l'écologie. L'évaluation du danger et de la probabilité d'exposition à une substance dangereuse est l'objectif de la toxicologie. L'écologie est une science qui étudie l'environnement à travers l'étude d'un écosystème et d'une population présent dans un biome en particulier. Pour cela, des paramètres physiologiques pouvant impacter la démographie ou la santé d'une population sont mesurés comme, par exemple, les niveaux de mortalité. L'association de l'écologie et de la toxicologie traite du devenir des contaminants dans les différents compartiments environnementaux (eau, air, sol et sédiments) et de la façon dont ils affectent les organismes vivants à court et à long terme [168]. Toutefois, cette science est de plus en plus utilisée pour mettre en évidence des activités anthropiques sur l'environnement afin de répondre à des questions sociétales majeures [169].

Actuellement, les techniques analytiques de dosage de composés moléculaires sont suffisamment performantes et précises pour mesurer et détecter la présence d'une substance dans les milieux aquatiques ou dans l'air. Toutefois, ces types d'analyses ne permettent pas de couvrir l'ensemble des causes à effets dans les milieux contaminés. Pour cela, il faut développer une approche complémentaire permettant de relier l'apparition d'effets néfastes à la présence d'un contaminant. La double approche consistant à utiliser l'approche chimique et biologique lors de l'évaluation du risque en écotoxicologie repose sur la caractérisation de la contamination des organismes et l'apparition des effets toxiques associés comme le montre la figure n°33 [170]. L'analyse biologique du milieu utilise le biote pour évaluer la présence et l'incidence des contaminants dans l'écosystème. Cette approche est appelée biosurveillance ou *biomonitoring* et elle permet d'évaluer la fraction biodisponible qui est plus pertinente pour la compréhension des effets toxiques que celle obtenue à partir des matrices physiques [171].



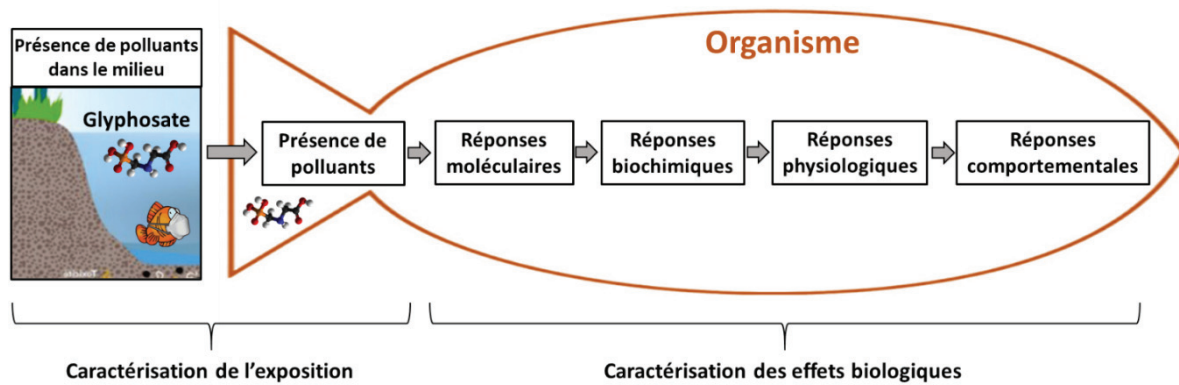


Figure 33 : Évaluation du risque écotoxicologique et représentation des réponses induites au sein d'un organisme en réponse à une contamination (adapté de la thèse de Judith Trapp [170]).

### 6.1.2. Biomarqueur : outil de prédilection en écotoxicologie

Les outils principaux en écotoxicologie sont les biomarqueurs qui sont couramment utilisés dans le domaine de la médecine et qui ont été récupérés par les écotoxicologues afin d'évaluer la dégradation d'un écosystème. Ces indicateurs permettent d'obtenir des informations sur les effets biologiques d'un contaminant environnemental [172]. Il existe de nombreuses définitions pour expliquer le rôle d'un biomarqueur. L'une des définitions les plus répandues dans la littérature est celle de Depledge qui indique qu'un biomarqueur correspond à "une variation biochimique, cellulaire, physiologique ou comportementale qui peut être mesurée dans des échantillons de tissus ou de fluides corporels ainsi qu'au niveau d'organismes entiers et qui fournit la preuve de l'exposition à un ou plusieurs polluants chimiques et/ou de leurs effets" [173]. Les biomarqueurs peuvent indiquer les mécanismes de régulation du métabolisme mis en place dans l'organisme lors d'un stress toxique. Cela peut confirmer l'exposition à un polluant en particulier et ainsi relier l'exposition dans le milieu à l'introduction de ce composé dans l'organisme. On parle ainsi de biomarqueurs de défense [174]. À l'inverse, on peut avoir des biomarqueurs de dommage qui indiquent un dépassement des capacités de régulation de l'organisme avec l'observation d'effets néfastes pour l'espèce contaminée [175]. Les biomarqueurs sont utilisés comme outils de diagnostic et/ou de prédiction, car bien qu'ils n'indiquent pas précisément la nature des contaminants, ils facilitent l'analyse chimique des matrices environnementales en réduisant le nombre de substances à suivre [175]. L'une des principales limitations des biomarqueurs est la capacité de discrimination entre une réponse à stress chimique et une réponse adaptative qualifiée de naturelle [176].

## 6.2. Utilisation d'organismes sentinelles en écotoxicologie

### 6.2.1. Qu'est-ce qu'une espèce sentinelle ?

L'écotoxicologie a pour but le suivi et la détection de polluant dans l'environnement à travers l'analyse d'organismes vivants appelés biomarqueurs. Cette biosurveillance réalisée sur des espèces qualifiées de sentinelles est possible, car ces organismes donnent des réponses biochimiques et physiologiques qui reflètent l'état de la pollution dans leur environnement [168]. Toutefois, cette approche nécessite des sites de référence non pollués permettant de comparer les niveaux des contaminants ou des biomarqueurs d'effet par rapport à ceux qualifiés de contaminés [177]. Les espèces sentinelles sont utilisées depuis de nombreuses années et cela même avant la mise en place de l'écotoxicologie moderne. Autrefois, des canaris étaient amenés dans les mines de charbon afin de détecter la présence de monoxyde de carbone. Leur mort indiquait aux mineurs qu'il fallait remonter à la surface [178]. Aujourd'hui, l'écotoxicologie utilise toujours les mêmes principes à travers l'utilisation des animaux pour détecter des dangers.

Une espèce sentinelle idéale doit [177,179] :

- Être sensible aux changements environnementaux (ex. contamination chimique, agents pathogènes, biotoxines).
- Accumuler les agents stressants dans ses tissus.
- Produire des effets sur la santé mesurables et interprétables à la suite d'une contamination.
- Être maîtrisé par les scientifiques (aspect physiologique, anatomique et écologique) afin d'identifier tout signal qui n'est pas causé par l'exposition aux agents stressants environnementaux.
- Pouvoir être surveillés, capturés et tolérer la manipulation ainsi que l'échantillonnage.
- Jouer un rôle trophique et écologique clé dans son habitat afin de pouvoir fournir des informations pertinentes.

Les hydrosystèmes sont les écosystèmes les plus impactés par les activités anthropiques puisque ce sont les réservoirs finaux dans lesquels on peut retrouver de nombreux contaminants provenant des activités agricoles, industrielles et domestiques [180]. De ce fait, la suite de cet état de l'art sur les espèces sentinelles sera appliquée aux organismes utilisés dans les écosystèmes aquatiques.

### 6.2.2. Les espèces sentinelles dans les écosystèmes aquatiques

Les espèces sentinelles présentes dans le milieu marin peuvent être réparties dans deux catégories en fonction de la mobilité des organismes : Les espèces très mobiles qui sont des sentinelles des



environnements marins pélagiques (relatif à la pleine mer) et les moins mobiles qui sont des sentinelles des environnements côtiers [179].

### 6.2.2.A. Les grands vertébrés marins, espèces sentinelles de l'environnement marin pélagique

Les effets de l'exposition à de nombreux stress et polluants dans l'environnement marin peuvent s'exprimer à différents niveaux trophiques de l'écosystème étudié. Cependant, l'attention des scientifiques s'est récemment focalisée sur les grands vertébrés marins à travers les oiseaux, les grands poissons pélagiques et les mammifères marins. Dans le passé, l'utilisation des grands vertébrés marins en tant qu'espèce sentinelle a été critiquée à cause de leur grande mobilité et de leur caractère pélagique [179]. Elles ne seraient pas des espèces indicatrices de grand intérêt. De plus, les organismes marins peuvent avoir un statut d'espèce protégée ce qui rend difficile l'obtention d'échantillon de tissus. Toutefois, ces points de vue ont évolué au cours des dernières années grâce à la compréhension des différences de distribution et de comportement de ces espèces dans l'espace et le temps. De plus, depuis les années 2000, plusieurs groupes de recherche ont développé des approches avec des biomarqueurs non létaux [179]. La mise en place de biopsie cutanée sur des animaux en liberté a été validée comme technique d'échantillonnage puissante et relativement non invasive [179,181,182]. Il est également possible de réaliser des tests *in vitro* en utilisant des cultures de cellules fibroblastes et des lignées cellulaires organotypiques exposées à différentes doses de contaminants obtenus à partir des biopsies [179,181,182].

Les grands vertébrés marins sont particulièrement adaptés pour l'étude des maladies chroniques. En effet, puisque ce sont des prédateurs et qu'ils ont une durée de vie relativement longue, les niveaux de contaminations anthropiques sont généralement élevés grâce à la bioamplification. Cela permet de mettre en lumière des polluants environnementaux hérités du passé comme le mercure ou les polychlorobiphényles (PCB). De plus, ils peuvent également alerter sur les nouveaux contaminants émergents (composés perfluorés, nanomatériaux, microplastiques, etc.) [179,183]. Les espèces sentinelles, comme les mammifères marins, dont la physiologie est suffisamment similaire à celui de l'Homme, peuvent donner une indication précoce des effets néfastes potentiels sur la santé et fournir un aperçu des mécanismes de toxicité d'un composé dangereux donné [179,184].

### 6.2.2.B. Les invertébrés : espèces sentinelles de l'environnement côtier et des rivières

Depuis les années 1990, l'approche des biomarqueurs chez les invertébrés pour l'évaluation des effets toxicologiques des contaminants dans l'environnement côtier et d'eau douce a été de plus en plus utilisée. Les invertébrés sont privilégiés pour l'évaluation des risques écotoxicologiques dans les écosystèmes marins, car ils constituent 95 % de toutes les espèces animales. Ce sont donc des

composants majeurs de nombreux écosystèmes. De plus, leurs populations sont souvent nombreuses, de sorte que des échantillons peuvent être prélevés sans affecter de manière significative la dynamique des populations [179,185]. En raison de leur faible mobilité, plusieurs espèces de moules ont également été utilisées comme sentinelles de l'environnement côtier [179]. Les moules sont des espèces idéales pour la biosurveillance des milieux marins, car elles ont une taille appropriée, une large distribution dans le monde et une activité de filtration de l'eau qui induit une bioaccumulation des contaminants dans leurs tissus. Les moules *Mytilus edulis* [186,187] et *Mytilus galloprovincialis* [188,189] sont les espèces les plus communes pour ces études écotoxicologiques. Parmi le large choix d'espèces sentinelles utilisées en écotoxicologie aquatique, quelques espèces ont montré de plus en plus d'intérêt auprès de la communauté des écotoxicologues au cours de la dernière décennie, à savoir l'amphipode d'eau douce *Gammarus fossarum*.

### 6.2.3. L'espèce sentinelle *Gammarus fossarum*

*Gammarus fossarum*, représenté sur la figure n°34, appartient à la famille des Gammaridaea comprenant plus de 1500 espèces appartenant à environ 1000 genres.



Figure 34 : Organisme *Gammarus fossarum* utilisé comme espèce sentinelle pour le développement d'indicateur de la contamination chimique des milieux aquatiques.

Le genre *Gammarus*, dont fait partie *G. fossarum*, contient plus de cent espèces d'eau douce. Cette espèce de gammare est l'un des amphipodes les plus représentés et les plus répandus en Europe et au nord de l'Asie Mineure [190–192]. Les gammares détritivores ont commencé à être utilisés en écotoxicologie à partir des années 1970, avec principalement deux espèces, à savoir *Gammarus pulex* et *Gammarus fossarum*. L'utilisation croissante des espèces d'amphipodes pour les études de toxicité s'explique par leur statut d'espèces clés dans les écosystèmes aquatiques, mais aussi par leur large distribution, leur identification facile, leur échantillonnage et leur facilité à être maintenu en laboratoire. De plus, ils sont sensibles, mais résistants aux substances toxiques [193]. Dans cette thèse, l'organisme utilisé comme matrice pour développer de nouvelles approches par chromatographie liquide couplée à la spectrométrie de masse est l'espèce *G. fossarum*. Deux applications biologiques

ont été mises en place à partir de cette espèce sentinelle : une étude de biosurveillance active lors d'une pollution agricole (cf. chapitre 2) et l'étude multi-omiques (lipides, métabolites et protéines) chez la femelle lors des différents stades du cycle reproducteur (cf. chapitre 4).

#### 6.2.4. Cycle reproducteur et cycle de mue chez l'espèce *G. fossarum* femelle

Le cycle de reproduction des gammares femelles est étroitement synchronisé avec celui de la mue. Ces deux cycles ont été caractérisés précisément par Geffard *et al.* [194–196]. Le cycle de mue dure environ 30 jours et se compose de 3 phases (post-mue, inter-mue et pré-mue) qui sont subdivisées en 6 stades distincts (A, B, C1, C2, D1, D2). L'observation microscopique du décollement de l'épiderme et de l'apparition du nouvel exosquelette au niveau de la griffe des périopodes des 3<sup>èmes</sup> et 4<sup>èmes</sup> paires permet de déterminer les différents stades (cf. figure n°35 et n°36).

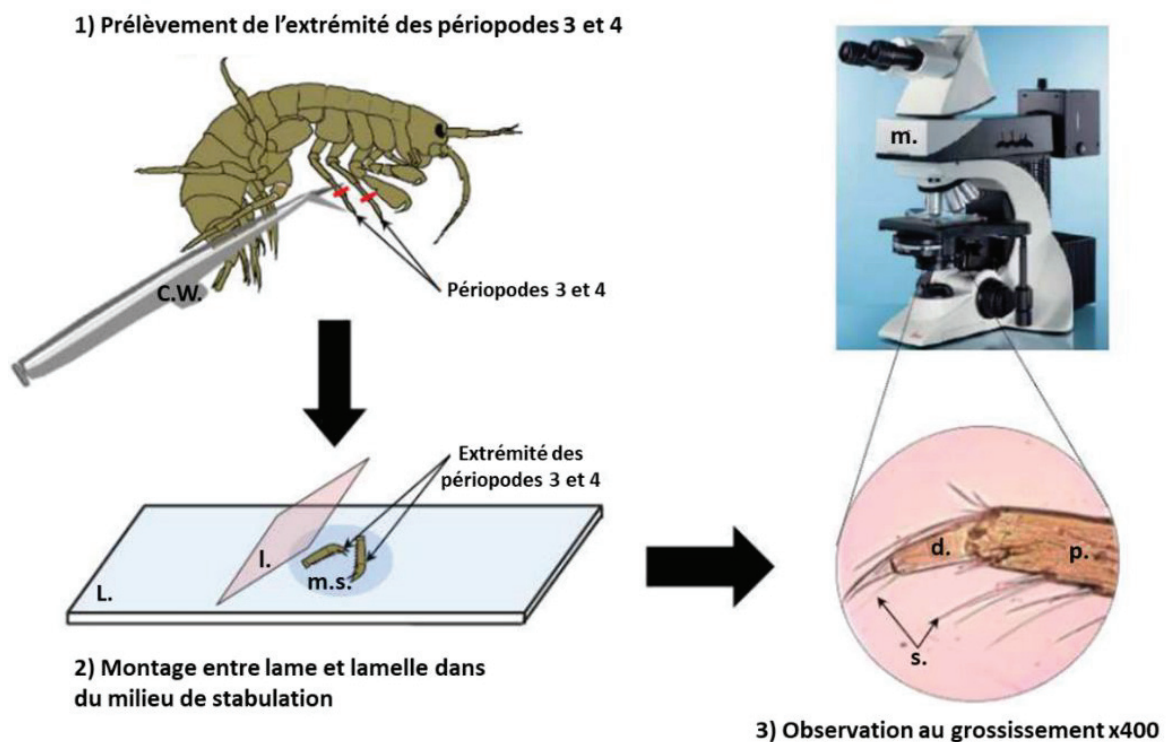


Figure 35 : Schéma représentant la procédure utilisée pour déterminer le stade d'inter-mue des organismes *G. fossarum* [170].

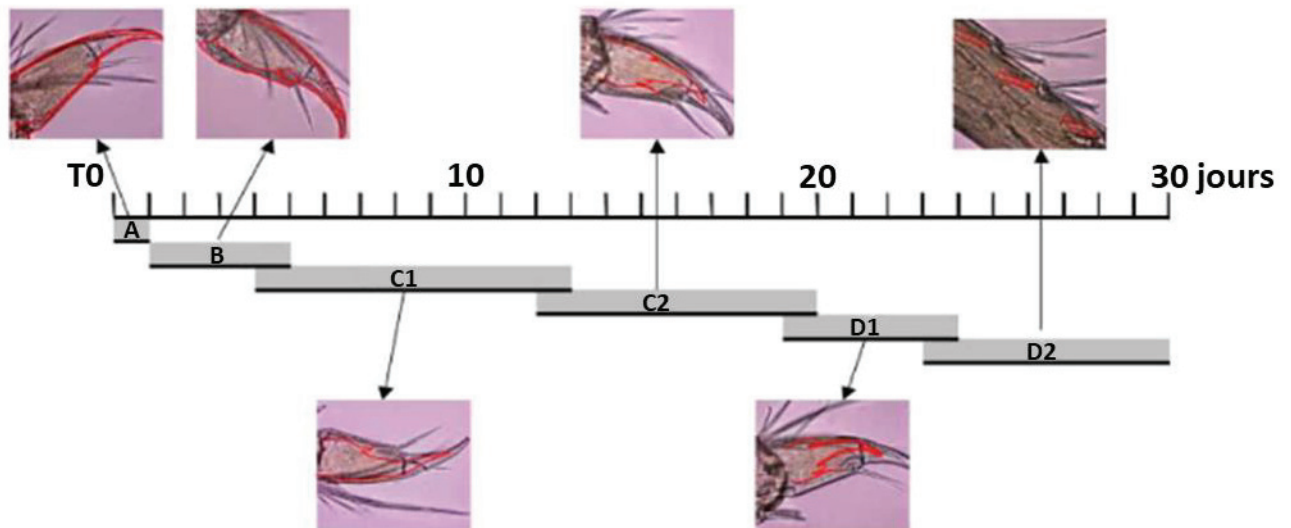


Figure 36 : Changements intertégumentaires à l'extrémité des périopodes 3 et 4 chez *G. fossarum* au cours du cycle de mue [195].

La phase de post-mue correspond au stade A et B. Le stade A est très court, il dure environ un jour et peut être regroupé avec le stade B qui dure 4 jours pour former un seul stade nommé AB. Ce stade est caractérisé par un durcissement de la cuticule. Lors de cette phase, les ovocytes présentent peu de globules lipidiques et de vésicules vitellines (structure embryonnaire sur laquelle se développe l'embryon). La phase d'inter-mue, associée aux stades C1 et C2, est caractérisée par la rétractation de l'épiderme qui dure 17 jours. La transition entre les stades C1 et C2 marque le début de la vitellogenèse secondaire, ainsi les vésicules vitellines et les globules lipidiques des ovocytes augmentent drastiquement en nombre et en taille. Enfin, lors des stades D1 et D2 de la phase pré-mue, une nouvelle cuticule est générée avant la mue. Les caractéristiques histologiques des ovocytes à ces stades sont très similaires à celles du stade C2 [194–196]. Parallèlement au développement phénotypique, une accumulation de vitellogénine, la protéine clé impliquée dans la reproduction, a été observée tout au long du cycle de reproduction. Les augmentations de concentration les plus importantes de cette protéine se produisent lors du passage du stade C1 à C2 et de C2 à D1 [197]. La figure n°37 présente le cycle de mue parfaitement corrélé avec celui de la reproduction. En effet, la maturation des gonades et le développement des embryons chez les gammare femelles sexuellement actives sont synchronisés avec chaque cycle d'inter-mue (passage du stade B à C1). Le rejet de l'ancienne cuticule ou exuviation est le point de départ et l'aboutissement des deux cycles. De ce fait, les différents stades du cycle de mue des gammare peuvent être utilisés pour qualifier les différents stades de la reproduction [170].

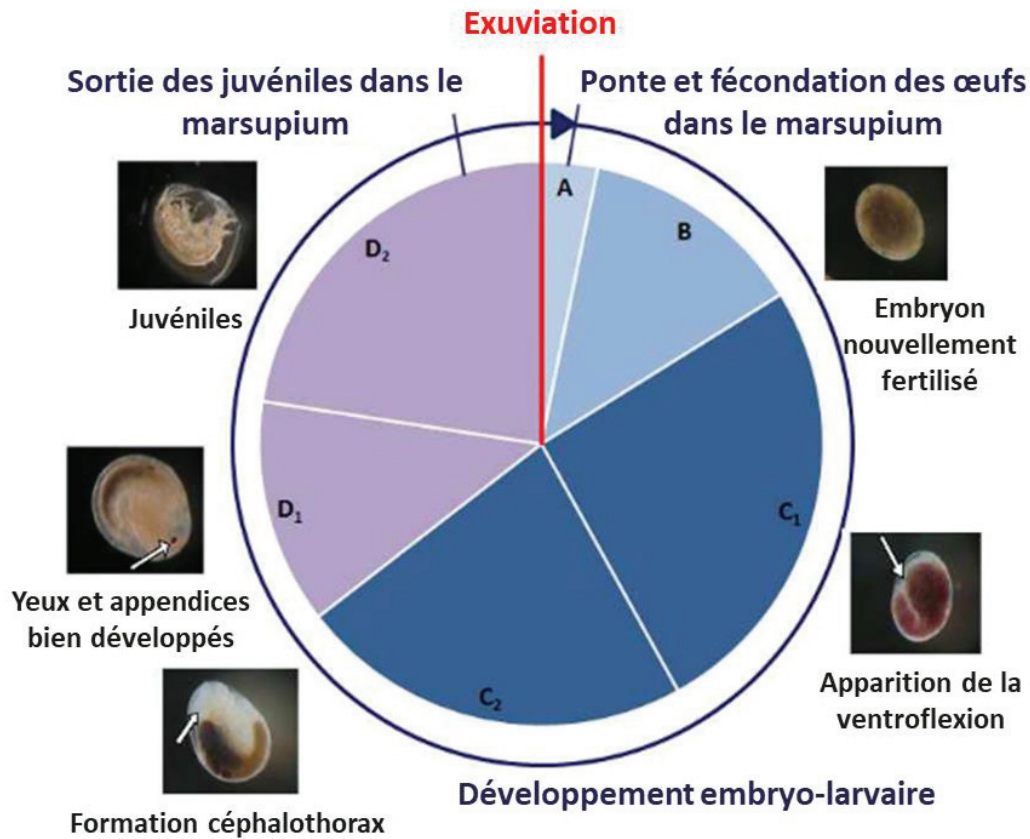


Figure 37 : Stades de mue corrélés aux stades embryonnaires chez la femelle *G. fossarum* [170].

Contrairement aux femelles, la spermatogenèse chez les gammarès mâles n'est pas liée au cycle de mue. De plus, les paramètres morphologiques ne sont pas disponibles pour représenter avec précision les organismes à différents stades de la spermatogenèse [198].

## 6.3. Objectifs et enjeux des omiques pour l'écotoxicologie

### 6.3.1. Apport de la protéomique pour l'écotoxicologie

Le terme écotoxicoprotéomique est apparu pour la première fois dans des revues scientifiques à la fin des années 2000 [168,186,199–202]. Cette approche étudie les variations dynamiques des protéines dans les cellules, tissus ou autres liquides corporels. Les empreintes protéiques qui en résultent peuvent être comparées afin de comprendre les modes d'action des contaminants et ainsi déchiffrer les mécanismes d'adaptation des organismes. Cette comparaison identifie également les biomarqueurs environnementaux spécifiques [168]. L'écotoxicoprotéomique a toujours été guidée par l'impact que la protéomique a eu sur la gestion de la santé humaine, offrant ainsi de nombreux outils de diagnostic. Comme indiqué sur la figure n°38, les approches protéomiques en écotoxicologie étaient réalisées à l'origine à partir de l'électrophorèse bidimensionnelle sur gel couplée à un spectromètre de masse [203].

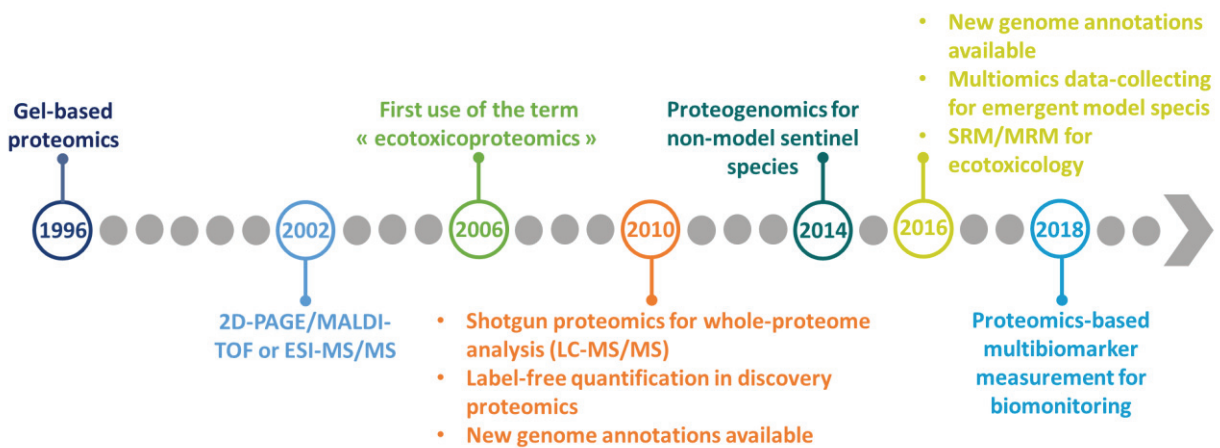


Figure 38 : Évolution, au cours du temps, des techniques de protéomique utilisées dans les approches écotoxicologiques (Adapté de la thèse de Duarte Domingos Gouveia [204]).

Cette technique basée sur la séparation des composés en fonction du poids moléculaire et de leurs points isoélectriques permet seulement de détecter les protéines solubles les plus abondantes [203,205]. Plus récemment, les approches utilisant la chromatographie liquide couplée à des spectromètres de masse de haute résolution se sont imposées comme méthode de référence en écotoxicoprotéomique afin d'étudier l'ensemble d'un protéome en une seule analyse sans *a priori* [206]. D'un point de vue écotoxicologique, il est très important que la méthodologie ne soit pas fondée sur des hypothèses en raison de la grande variété de produits chimiques existants et de leurs effets très contrastés sur les différentes espèces présentes dans un écosystème [168]. Par exemple, dans le cas des perturbateurs endocriniens, les mécanismes d'action sont bien connus pour des groupes phylogénétiques restreints comme pour les œstrogènes de synthèse qui provoquent des dysfonctionnements chez les vertébrés mâles [207]. Cependant, chez certaines espèces, leurs effets diffèrent de ceux prédits, soit parce que le mode d'action a changé, soit parce que nous connaissons mal l'endocrinologie de base des espèces non modèles. Parce qu'elles permettent une quantification précise, les approches protéomiques ciblées avec le mode d'acquisition MRM ont été proposées comme méthode de validation des biomarqueurs, après la phase de découverte [208]. Au fil des années, la quantification des protéines basée sur le mode MRM s'est imposée comme une alternative viable aux immuno-essais classiques (test ELISA), ayant déjà démontré des performances comparables ou supérieures [209]. En ajoutant des peptides isotopiquement enrichis dans les échantillons, on peut obtenir une quantification précise, rapide et reproductible utilisable en routine [50,86].

Ces stratégies couplées aux spectrométries de masse de haute et basse résolution ont été proposées comme outil de diagnostic pour évaluer l'état de santé de l'organisme sentinelle *G. fossarum*. Les réponses protéiques des gammars à plusieurs perturbateurs endocriniens ont été étudiées par une approche comparative de type shotgun [210,211]. Cette étude a permis de proposer plusieurs



protéines comme biomarqueurs des troubles de la reproduction et des perturbations endocriniennes chez le mâle. Les réponses de *G. fossarum* à une exposition aux métaux ont été récemment rapportées par Gismondi *et al.* [212]. Grâce à leur étude, ils ont pu identifier des profils de dérégulation de protéines spécifiques au cadmium, au plomb et au cuivre [212]. Les études menées par les laboratoires de Arnaud Chaumot (INRAE), Jean Armengaud (Li2D) et Arnaud Salvador (Anabio-MS) ont permis de découvrir des séquences de protéines spécifiques liées à des processus physiologiques clés, ainsi que des dizaines de biomarqueurs candidats. La majorité des biomarqueurs identifiés chez les organismes du genre *Gammarus* sont des protéines qui dans 20% des cas répondent à une activité de défense oxydative, 10% des cas à une activité de détoxification et dans 22% des cas à une réponse générale au stress subit [170]. Grâce à l'expertise du laboratoire Anabio-MS, des stratégies de quantification ciblées et précises des multibiomarqueurs par spectrométrie de masse en mode MRM ont été réalisées pour valider ces biomarqueurs.

La première étude en mode ciblé a été réalisée par Romain Simon et Guillaume Jubeaux en 2010. Ils ont développé et validé une méthodologie visant à quantifier la vitellogénine chez *G. fossarum* par LC-MS/MS avec le mode MRM comme alternative aux immunodosages conventionnels [213]. La vitellogénine (VTG), présente chez les ovipares, est une protéine précurseur du jaune d'œuf. Elle est produite par les femelles matures en réponse aux hormones stéroïdiennes. Le gène codant pour la VTG est également présent chez les mâles, mais il est inexprimé sauf en cas de dérèglement hormonal [213]. Lors d'une exposition à des perturbateurs endocriniens, le gène de la VTG est activé et la protéine est synthétisée. Par conséquent, la mesure de la VTG chez les femelles et les mâles immatures est considérée comme un biomarqueur d'exposition aux perturbateurs endocriniens [214]. Dans cette étude, des gammares mâles et femelles ont été exposés en laboratoire à des hormones (20-hydroxyecdysone et méthyl-farnésolate). Cela a permis d'évaluer la fiabilité de l'approche en vérifiant que la quantification de la VTG avec la méthode LC-MRM est un indicateur de la qualité des ovocytes chez les femelles et un biomarqueur spécifique de la perturbation endocrinienne chez les mâles [197]. Il est important de noter que la méthode a également été appliquée avec succès à des organismes prélevés dans le cadre d'un programme régional de biosurveillance des rivières, validant ainsi sa robustesse et pour son application sur le terrain [215]. Cette étude a également démontré la pertinence de l'encagement des gammares témoins en constituant un plan expérimental permettant de réduire la variabilité interspèce. En effet, les facteurs tels que le bagage génétique, l'âge, le sexe, le statut reproductif, les ressources alimentaires et la durée d'exposition ont pu être contrôlés [215,216]. Pour finir, Jubeaux *et al.* ont prouvé que la spectrométrie de masse en protéomique peut fournir des biomarqueurs applicables pour la surveillance environnementale sans qu'il soit nécessaire de procéder à de nouveaux développements pour chaque espèce sentinelle. En effet, ils ont montré qu'une méthode validée pour une espèce qui réalise la vitellogenèse peut être appliquée à des espèces

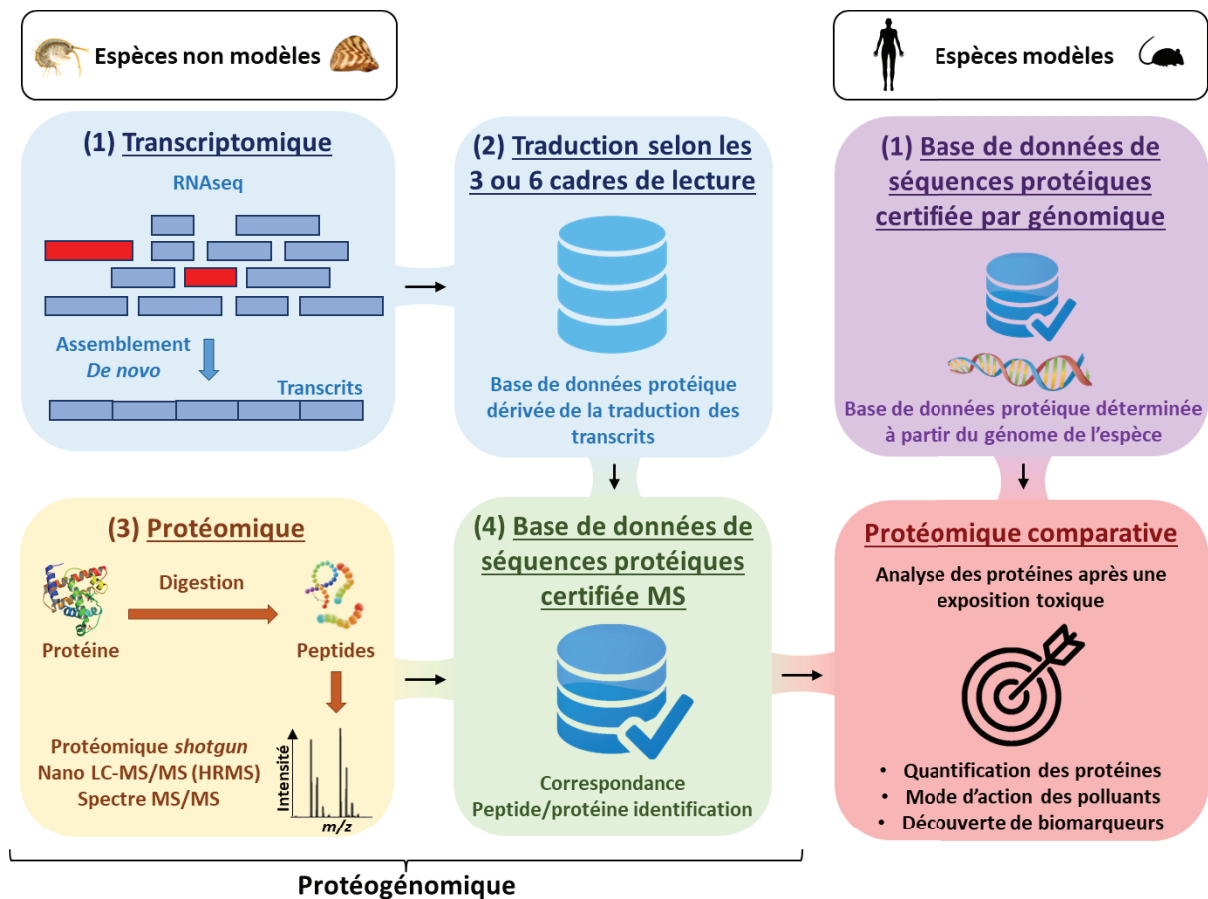
phylogénétiques proches grâce à la conservation évolutive des motifs peptidiques de la vitellogénine [217]. À la suite des travaux de Simon *et al.*, Aurore Charnot et Duarte Gouveia ont mis au point le premier multiplexe en écotoxicologie par LC-MS/MS en protéomique ciblée permettant de quantifier simultanément 40 protéines présentes chez *G. fossarum* [50,216]. La stratégie utilisant la sMRM permet de surmonter les problèmes rencontrés lors de la multiplication des protocoles employés pour l'analyse individuelle de chaque biomarqueur. Outre le fait qu'elle nécessite des volumes d'échantillons plus faibles, l'utilisation du multiplexe diminue le temps nécessaire à la réalisation de l'essai [50,216]. À la suite de ces travaux, l'application *in situ* de la protéomique ciblée pour mesurer des dizaines de biomarqueurs simultanément dans un même organisme est une stratégie prometteuse pour le développement d'outils de surveillance fonctionnels à des fins réglementaires [216].

Contrairement à la protéomique de découverte, qui analyse l'ensemble du protéome, le mode MRM nécessite certaines connaissances *a priori*, telles que les séquences des protéines obtenues par protéogénomique dans le cas des espèces sentinelles non modèles.

### 6.3.2. La protéogénomique : outil de prédilection pour les organismes non modèle

L'écotoxicologie est basée sur l'utilisation des espèces sentinelles pour évaluer la qualité d'un milieu. Cependant, cela sous-entend que ces organismes soient suffisamment connus pour être utilisés. À ce jour, il existe une faible quantité d'organismes dont le génome est séquencé ce qui rend l'étude des mécanismes moléculaires complexes. Cela est particulièrement vrai pour l'identification de potentiels biomarqueurs protéiques. Dans le cas d'une analyse protéomique, lors d'une approche de découverte en mode non ciblée, l'interprétation des données LC-MS/MS est fondée sur la recherche de séquences peptidiques homologues dans des bases de données spécialisées (*ex.* Ensembl, RefSeq ou UniProtKB) (*cf.* figure n°39) [218,219]. Ces bases de données sont élaborées à partir de données génomiques, triées et annotées à partir d'organismes séquencés. Si les organismes humains et les organismes modèles ont fait l'objet d'une grande attention en ce qui concerne l'annotation de leur génome, ce n'est manifestement pas le cas des organismes non modèles comme le gammare ou la dreissène dont le génome n'a pas été séquencé ou l'a été que partiellement [218,219]. Par conséquent, les bases de données de séquences protéiques de référence pour ces organismes sont incomplètes et mal annotées.





**Figure 39 : Identification des protéines par protéogénomique.** Les données génomiques et transcriptomiques sont utilisées pour générer des bases de données de séquences protéiques permettant l'interprétation de spectres de masse issus d'analyses nano LC-MS/MS avec une approche *shotgun*.

Plus la diversité de la population d'un milieu est importante, plus il sera facile de révéler des mécanismes moléculaires permettant la découverte de biomarqueurs en réponse aux différents polluants [219]. L'utilisation de ces espèces sentinelles non modèles est donc primordiale, mais pour cela il faut être capable d'obtenir des informations sur leur génome. La protéogénomique qui allie la protéomique et la transcriptomique apparaît comme une solution de choix pour les organismes non séquencés en écotoxicologie. En effet, les protéines sont issues de l'ADN présent dans les cellules qui est lu par l'ARN polymérase afin de transcrire les brins d'ADN en ARN messager (ARNm). Cet ARNm est ensuite traduit par les ribosomes au sein du cytoplasme pour former des protéines grâce à l'association d'acides aminés reliés entre eux par des liaisons peptidiques. Les protéines sont donc étroitement liées à l'ADN, mais aussi aux transcrits qui désignent les ARNm issus de la transcription des gènes. Comme indiqué sur la figure n°39, le principe de cette nouvelle science omique appelée protéogénomique consiste à séquencer les ARNm des organismes non modèles par RNAseq pour générer une première base de données de transcrits. Cette première analyse permet d'annoter ces derniers afin d'obtenir une orientation sur les potentielles fonctions des transcrits grâce aux similarités avec d'autres organismes modèles. Pour finir, la traduction de ces ARNm ou transcrits grâce aux cadres de lecture

permet de réaliser une première base de données protéomique pour faciliter l'interprétation des protéomes des organismes non modèles avant ou après exposition à un stress chimique [218,219]. De plus, les protéines certifiées par spectrométrie de masse grâce au spectre MS/MS de haute résolution peuvent faire l'objet d'une caractérisation structurale et fonctionnelle plus poussée [218,219].

En 2017, l'application de l'approche protéogénomique pour déchiffrer le protéome de *G. fossarum* a généré une base de données comprenant 1873 protéines certifiées par spectrométrie de masse [50,220,221]. L'ensemble de ces données a fourni les outils nécessaires pour réaliser des études protéomiques sur cette espèce sentinelle [50,216,222].

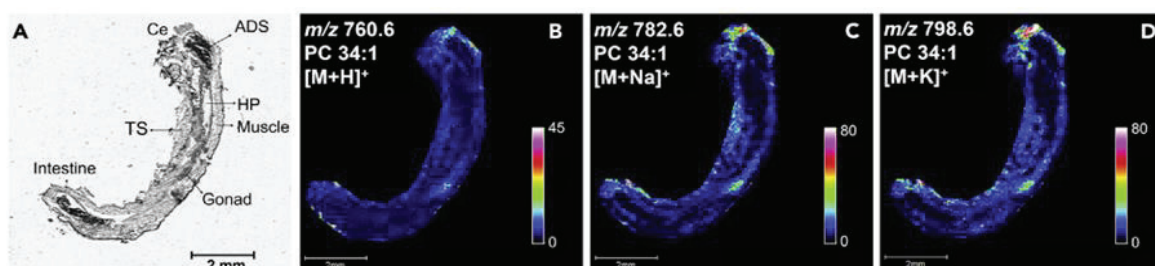
### 6.3.3. Apport de la lipidomique pour l'écotoxicologie

Les progrès en lipidomique ont permis l'émergence de nouvelles applications en toxicologie avec par exemple l'étude de l'oxydation des lipides comme biomarqueur pour identifier le dysfonctionnement des mitochondries, le stress oxydatif et de possibles inflammations [223]. Malgré cette amélioration de la lipidomique en toxicologie, il reste des domaines qui ont été négligés, notamment en écotoxicologie. C'est pour cela que les lipides sont de plus en plus étudiés dans ce domaine puisqu'il est évident que ces biomolécules sont vitales pour de nombreux processus physiologiques. Il est important de comprendre comment ces molécules sont perturbées dans leur synthèse, leur métabolisme et leur fonctionnement par des facteurs de stress environnementaux comme les polluants chimiques. Récemment, la lipidomique a été déployée sur le terrain pour soutenir la surveillance environnementale. En effet, des profils lipidiques de moustiques vivant dans des zones contaminées par des métaux ont révélé une dérégulation des lipides membranaires accompagnés par une augmentation des lipides responsables des réserves énergétiques [224]. Plusieurs études ont été élaborées pour mettre en évidence que les facteurs de stress chimique altèrent les réseaux liés à la signalisation et au métabolisme des lipides. Par exemple, Greer *et al.* [225] ont démontré qu'une exposition des truites arc-en-ciel juvéniles à une concentration de 10 à 40 µg/L de chlorpyrifos, un pesticide, pendant 7 jours, a altéré le métabolisme lipidique via l'inhibition de la production de lipase de phospholipide. Cela a induit une diminution critique des lysophosphatidyléthanolamines (LPE), des lipides utilisés dans la signalisation cellulaire, accompagnée d'une augmentation de leurs phosphatidyléthanolamines (PE) respectives. Marqueño *et al.* ont montré que le profil des PC permet de mettre en évidence la pollution d'eaux usées et industrielles puisque la mesure des lipides dans des cyprinidés méditerranéens indique clairement une différence de profil entre les sites pollués et celui de référence. En effet, la diminution des PC, des PC plasmogènes ainsi que l'augmentation des PC plasmanyles, LPC et des esters de cholestérol suggèrent un stress oxydatif et une altération de l'homéostasie du cholestérol [226]. À travers ces exemples, il est compréhensible que l'obtention de profils rétrospectifs et prospectifs complets de lipidome chez les organismes sentinelles est d'une

haute importance. Cela permet de comprendre quelles espèces lipidiques peuvent être altérées par un polluant, comment est effectuée cette altération et comment évaluer le risque écologique encouru par l'organisme contaminé.

Dans le cas de l'espèce sentinelle *G. fossarum*, les connaissances sur la composition lipidique de cet organisme sont peu nombreuses. Une évaluation de la perturbation lipidique de ces gammares exposés à un insecticide régulateur de croissance a été récemment rapportée par Arambourou *et al.* en 2018 [227]. Toutefois, un nombre limité de classes lipidiques et d'espèces moléculaires a été décrit [227]. C'est pour cela qu'en 2020, Fu *et al.* [198] ont réalisé la caractérisation exhaustive du lipidome de *G. fossarum* grâce à une approche *shotgun* par spectrométrie de masse de haute résolution. Plus de 200 espèces moléculaires appartenant à 11 classes lipidiques (TG, DG, PC, PC-O, PE, PE-O, PI, LPC, LPE, SM, cholestérol) ont été quantifiées chez le gammare mâle et femelle. Les acides gras totaux ont été étudiés par chromatographie en phase gaz couplée à un détecteur à ionisation de flamme (GC-FID) pour établir les longueurs de chaînes prédominantes et savoir si ces acides gras sont saturés ou non. L'acide gras (AG ou FA pour *fatty acid*) majoritaire chez les gammares est le AG 18:1 à 33%. Les autres acides gras dont la teneur est élevée sont le AG 16:0 à 16%, le AG 20:5 à 14%, le AG 16:1 à 13%, le AG 18:3 et le AG 18:2 à 6%. Leurs résultats indiquent que les acides gras du gammare sont majoritairement insaturés.

Malgré l'identification de centaines d'espèces moléculaires permettant la caractérisation du lipidome, la distribution spatiale des espèces lipidiques analysées est manquante alors qu'elle peut s'avérer cruciale pour comprendre leurs modes d'action [228–232]. De ce fait, Fu *et al.* ont également révélé la localisation *in situ* de lipides présents chez *G. fossarum* grâce à des cryocoupes de gammares analysées par imagerie MALDI-MS et TOF-SIMS [198]. Un exemple de la localisation du PC 34:1 dans le gammare est représenté sur la figure n°40.



**Figure 40 : Localisation du PC 34:1 dans une coupe verticale de *G. fossarum* par MALDI MSI.** (A) Image optique de la section de tissu de *G. fossarum* mâle. Ce : céphalon ; ADS : système digestif antérieur ; HP : hépatopancréas ; TS : segments du thorax. (B) Image ionique du PC 34:1 protoné. (C) Image ionique de l'adduit de sodium du PC 34:1. (D) Image ionique de l'adduit de potassium du PC 34:1 [198].

Dans l'ensemble, les résultats de Tingting Fu et Sophie Aycirix fournissent des informations à la fois sur la composition et la répartition spatiale des lipides chez cette espèce de crustacé. Pour compléter

ces travaux, il est nécessaire de développer des méthodes de quantification hautement multiplexées par spectrométrie de masse ciblée pour mesurer précisément des variations lipidiques dans des échantillons de *G. fossarum* de référence par rapport à des échantillons ayant subi un stress environnemental.

### 6.3.4. Apport de la métabolomique pour l'écotoxicologie

Compte tenu de l'augmentation des études en métabolomique, cette science omique ne devrait plus être qualifiée de nouvelle [19,233]. Viant *et al.* [234] ont montré que les méthodes métabolomiques par résonance magnétique nucléaire (RMN) étaient 20 fois plus sensibles que les paramètres traditionnels (anomalies du développement, succès d'éclosion et mortalité) pour mesurer des effets dus à la contamination des embryons de medaka, un petit poisson d'Asie du Sud-est, au trichloréthylène. L'écotoxicométabolomique permet de relier des réponses métaboliques anormales à des variations environnementales afin de valider ou non des hypothèses sur la présence de facteurs de stress environnementaux et à élucider des modes d'action sous-jacents de l'organisme impacté [19]. Les stress étudiés en métabolomique peuvent être dus à des contaminants chimiques, à des changements nutritionnels et environnementaux (humidité, température, *etc.*) dans un large éventail d'organismes sentinelles [19]. Jusqu'à récemment, la RMN était le principal outil d'analyse en métabolomique, cependant bien qu'elle présente certains avantages, la spectrométrie de masse a révolutionné la capacité à sonder les métabolomes environnementaux grâce à sa grande sensibilité [233]. Suite à l'augmentation de la rapidité d'analyse des méthodes métabolomiques environnementales par rapport aux autres méthodes conventionnelles en écotoxicologie, des méthodes d'alerte précoce de la perturbation d'un écosystème peuvent être envisagées. Dans ce cas de figure, les futurs défis à surmonter sont la normalisation ainsi que le développement de méthodes pour les études métabolomiques directement sur le terrain [19]. L'échantillonnage direct sur le terrain est un paramètre primordial en métabolomique environnementale. Il permet de mieux observer les changements phénotypiques induits par un stress externe en diminuant la variabilité métabolique grâce à une stabilisation de l'organisme plus rapide que dans des conditions de laboratoire [6]. L'un des besoins importants en écotoxicologie est la capacité de collecter des échantillons non invasifs afin de pouvoir collecter plusieurs échantillons d'un même individu au cours d'une exposition. De ce fait, le même organisme peut servir de témoin, ce qui élimine la variabilité due aux différences entre les individus. Dans ce cas de figure, la métabolomique est très performante, car elle permet d'analyser les fluides corporels comme l'urine ou le sang qui sont de bons indicateurs de la perturbation du métabolisme de l'organisme [6].

À ce jour, les connaissances sur le métabolisme de l'espèce sentinelle *G. fossarum* sont peu nombreuses, car il n'existe qu'une seule publication réalisée par Bonnefoy *et al.* [65]. L'objectif de cette

étude était d'examiner l'effet d'une exposition à un mélange de carbamazépine (antiépileptique) et d'oxazépam (anxiolytique) sur le métabolisme de *G. fossarum*. Pour cela des analyses métabolomiques non ciblées ont été réalisées par nano LC-HRMS après une extraction  $\mu$ -QuEChERS. Les résultats obtenus ont montré une différence dans le comportement des mâles et des femelles de *G. fossarum* exposés au mélange de composés pharmaceutiques. En effet, des variables discriminantes ont été mises en évidence pour expliquer une partie des effets de l'exposition en fonction du sexe. Toutefois, aucune de ces variables métaboliques n'a pu être identifiée et annotée avec certitude. D'autres stratégies analytiques, basées sur la fragmentation MS/MS et l'interrogation de bases de données spectrales en masse devraient être envisagées pour réaliser cette identification de biomarqueurs potentiels. Deux autres études ont été réalisées sur une espèce sentinelle proche de *Gammarus fossarum*, à savoir *Gammarus pulex*. La première a été conduite par Gómez-Canela *et al.* [235] dont l'objectif est de développer une méthode ciblée pour analyser 29 métabolites appartenant à différentes classes biochimiques (acides aminés, acides organiques, nucléosides, nucléotides et sucres) par LC-HRMS. Parmi 26 produits pharmaceutiques, 3 ont été sélectionnés (propranolol, triclosan et nimésulide) pour réaliser des expositions. À la suite de ces contaminations, des variations des niveaux de concentration en métabolites impliquées dans des voies liées à la synthèse des protéines, au stress oxydatif et aux cascades de signalisation ont été observées. Des analyses supplémentaires sont nécessaires pour caractériser et comprendre pleinement les effets de ces contaminants sur les voies spécifiques notamment par la réalisation d'approche globale pour identifier plus de métabolites. C'est pour cela que Sheikholeslami *et al.* [66] ont poursuivi les expérimentations de Gómez-Canela *et al.* [235] en identifiant 190 métabolites avec une approche non ciblée et les mêmes conditions d'analyse et d'exposition sur *G. pulex*. Comme l'étude précédente, leurs travaux ont indiqué des changements significatifs dans les concentrations de plusieurs métabolites ayant des effets négatifs sur la santé du gammare provoqué par l'exposition aux différents produits pharmaceutiques. Les données obtenues ont fourni des informations plus précises pour comprendre les effets au niveau mécanistique de la biologie moléculaire de cet organisme. Les effets les plus importants sont le stress oxydatif qui provoque un déséquilibre entre les espèces réactives de l'oxygène (radicaux libres) et les défenses antioxydantes, ce qui peut entraîner des dommages aux tissus adipeux, à l'ADN et aux protéines.

# **Chapitre 2 : Scout-MRM, un nouveau mode d'acquisition pour augmenter les capacités de multiplexage en écotoxicoprotéomique.**

## **1. Contexte de l'étude**

La biologie des systèmes ou biologie intégrative traite des interactions complexes entre les différents compartiments qui composent les organismes biologiques. L'étude de la biologie des systèmes est réalisée à travers la compréhension de la structure de ces systèmes, de leur dynamique et des processus de régulation les contrôlant [21]. Les termes omiques définissent les méthodes analytiques qui permettent d'explorer l'ensemble des différents constituants (génomique, protéomique, métabolome, lipidome, *etc.*) de ces systèmes. Ces approches en génomique, protéomique, lipidomique et métabolomique offrent une vision intégrative des processus moléculaires interagissant à différents niveaux. Par exemple, pour avoir une idée de la fonction d'un gène, il faut maintenant non seulement en connaître la séquence, mais également les transcrits, les protéines, et les métabolites qui en découlent [21]. En écotoxicologie, l'intégration de ces différentes omiques permet d'étudier et de renseigner des mécanismes d'action et des effets toxiques [189]. La spectrométrie de masse est un outil analytique qui tient une place centrale dans le cadre des analyses omiques puisqu'elle permet d'identifier avec une grande précision et de quantifier avec une grande sensibilité la présence de biomolécules dans un mélange complexe. Depuis une dizaine d'années, de nombreux progrès ont été réalisés dans chacune des omiques pour contribuer à la description de ces systèmes, à travers des innovations en bio-informatique, l'augmentation du haut débit d'analyse et de la couverture de molécules analysables par spectrométrie de masse. Plus une technique analytique est capable d'analyser des composés en simultanément, plus elle sera qualifiée de multiplexée. Cette capacité à réaliser des multiplexes de plus en plus importants permet une meilleure caractérisation des différents compartiments biologiques d'un organisme en limitant le risque de lacunes dû à un manque d'information qui pourrait altérer le contexte biologique dans un système donné [24].

Parmi les différentes méthodologies analytiques déjà existantes en spectrométrie de masse, l'approche dite ascendante « *shotgun* » est l'approche de protéogénomique la plus pertinente pour la découverte de nouvelles protéines pour les organismes non modèles. Elle repose sur l'étude complète d'un protéome après digestion enzymatique. La spectrométrie de masse non ciblée (DDA et DIA) est adaptée aux études à haut débit d'analyse lors des approches globales grâce à sa capacité d'analyse de milliers de composés en simultanément et de sa très grande résolution. Toutefois, cette méthodologie

présente certaines limitations concernant sa faible reproductibilité et son incapacité à identifier et quantifier des protéines peu abondantes [45–47]. Les méthodes d'acquisition ciblées, telles que le mode MRM, améliorent la reproductibilité et la sensibilité par rapport aux approches *shotgun*, en se focalisant sur des peptides cibles prédéterminés. En revanche, elles nécessitent une connaissance *a priori* des séquences peptidiques obtenues expérimentalement ou *in silico* pour définir les transitions MRM [220,221]. De plus, l'une des limitations majeures des approches ciblées est le faible nombre de composés suivis par méthode. En effet, le nombre de composés suivis dépend du temps de cycle du spectromètre de masse et des largeurs des pics chromatographiques. Ce temps de cycle est réparti entre les différents composés pour suivre une ou plusieurs transitions par composé, tout en ayant suffisamment de points d'acquisitions pour bien définir chaque pic chromatographique. Ainsi, il existe un nombre maximal de transitions MRM par méthode en spectrométrie de masse ciblée (*cf.* chapitre 1 partie 2.3). En général, les conditions analytiques adoptées (*dwell time* minimum de l'appareil, largeur moyenne des pics chromatographiques et nombre de points par pic) permettent d'obtenir entre 100 et 150 transitions MRM par méthode [39]. En protéomique quantitative ciblée, le dosage d'une protéine est souvent réalisé à partir de plusieurs peptides rapporteurs (3 en général) afin d'augmenter le degré de confiance de l'analyse [236]. De plus, il est souhaitable d'utiliser au moins trois transitions MRM pour un même peptide afin d'avoir une transition quantifiante et deux qualifiantes. Ainsi, une méthode MRM ne peut suivre que 10 à 15 protéines au maximum.

En écotoxicologie, Romain Simon *et al.* [213] ont été les premiers à montrer que la spectrométrie de masse était une très bonne alternative aux méthodes indirectes et immunoenzymatiques pour doser des biomarqueurs protéiques potentiels chez l'espèce sentinelle *G. fossarum*. Une méthodologie pour quantifier la vitellogénine a ainsi été développée et validée. Pour augmenter les capacités de multiplexage, il est courant d'utiliser des méthodes MRM où la détection des composés est programmée dans le spectromètre de masse autour de leurs temps de rétention (sMRM) grâce à l'utilisation de fenêtres de détection [45–47]. Cela a permis à Aurore Charnot *et al.*, en 2017, de mettre au point le premier dosage multiplexé en écotoxicologie par LC-MS/MS en mode ciblée qui permet de quantifier simultanément 40 protéines présentes chez *G. fossarum* [50,216,237]. Bien que déjà important, ce nombre de protéines demeure faible dans un contexte de biologie des systèmes. Être capable de doser simultanément plusieurs centaines de protéines avec précision demeure un enjeu analytique essentiel. Cependant en MRM, le suivi hautement multiplexé des composés nécessite des segmentations temporelles très étroites autour des temps de rétention. En procédant de cette manière, la plus petite variation des temps de rétention peut conduire à la non-détection d'un composé qui se retrouve en dehors de la fenêtre de détection ou à un signal tronqué. Pour surmonter cette problématique, une nouvelle méthode ciblée basée sur la MRM a été développée dans notre



laboratoire. Cette approche, intitulée Scout-MRM [45–47], augmente les capacités de multiplexage des méthodes MRM sans être dépendante des temps de rétention.

Le premier défi décrit dans ce chapitre de thèse est de prouver que l'approche Scout-MRM permet de repousser les limites en termes de capacité de multiplexage sans faire de concession sur la robustesse de la méthode. L'objectif est donc de développer une méthode capable d'analyser simultanément plusieurs centaines de protéines chez l'espèce sentinelle *G. fossarum* et ainsi obtenir le plus grand multiplexe proposé à ce jour pour de futures études en écotoxicologie. Comme nous le verrons, le Scout-MRM a également pour but de faciliter le transfert de méthodes MRM hautement multiplexées entre laboratoires en s'affranchissant des temps de rétention. Ces travaux ont conduit à une publication dans *Journal of proteomics* en juillet 2020. L'intégralité de cette publication intitulée "*High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics*" [86] a été retranscrite dans la suite de ce chapitre.

## **2. Résumé de la méthodologie pour l'élaboration de la méthode Scout-MRM chez l'espèce sentinelle *G. fossarum***

Les différentes étapes clés de la mise en place de la méthode Scout-MRM sont résumées sur la figure n°41 correspondant au résumé graphique de la publication intitulée *High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics* [86]. Dans un premier temps, une étude protéogénomique par nano LC de phase inverse couplée à la spectrométrie de masse en mode DDA a été réalisée chez *G. fossarum*. Cela a permis d'identifier 5298 séquences de protéines qui ont été fonctionnellement annotées. Ainsi une base de données a été élaborée pour cette espèce sentinelle dont le génome n'est pas caractérisé. 263 protéines d'intérêts, impliquées dans des fonctions physiologiques clés (reproduction, immunité, homéostasie, mécanisme de détoxification et de défense, *etc.*) ont été choisies pour la mise en place de la méthode Scout-MRM. Ces protéines ont ensuite été analysées sur un nouveau système analytique constitué d'une chaîne chromatographique couplée à un Qtrap en mode MRM classique dédié aux analyses ciblées. Parmi les 263 protéines, 182 protéines ont été détectées sur cette nouvelle configuration. La méthode Scout-MRM a ensuite été développée en déterminant les composés scouts puis en répartissant les transitions MRM dans les différents groupes de détection. Par la suite, la méthode a été validée en évaluant les domaines de linéarité, les limites de quantification et la répétabilité. L'étude des rapports de transitions pour un même peptide a permis de ne conserver dans la méthode que celles qui sont le moins interférées. Après avoir appliqué les critères de validation, la méthode Scout-MRM contient 157 protéines ce qui est une première en écotoxicoprotéomique. Pour



terminer, cette méthodologie a été appliquée dans une étude de biosurveillance active dans un contexte de pollution agricole.

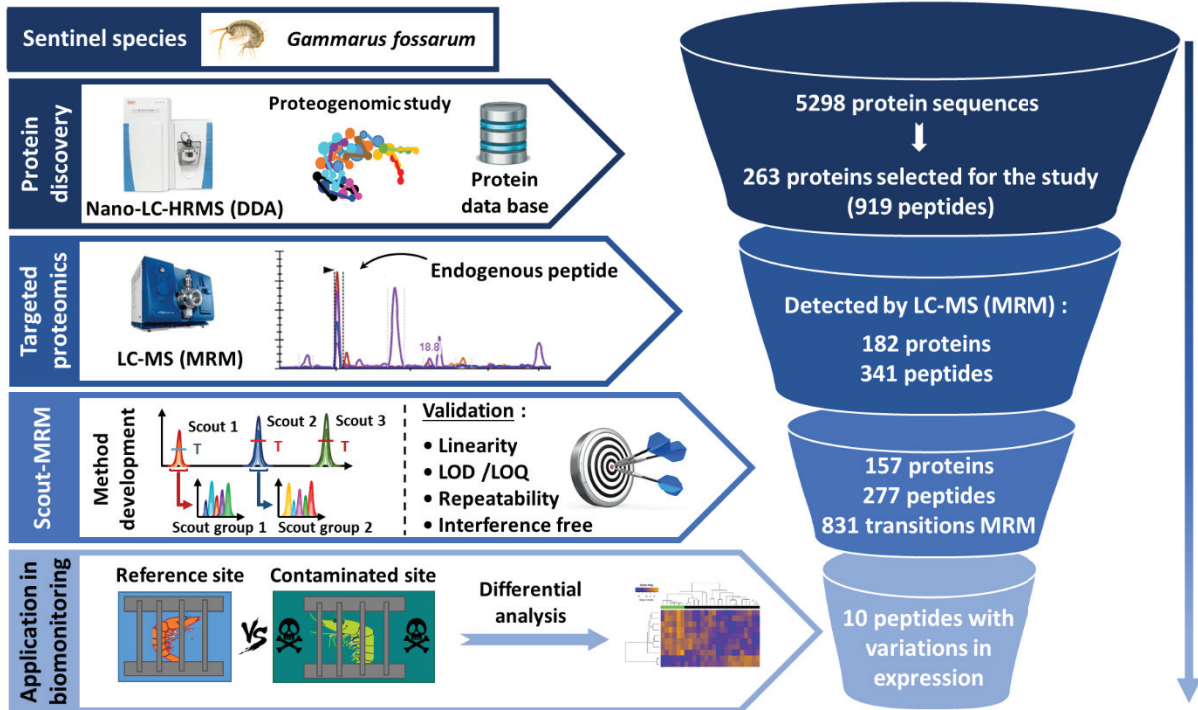


Figure 41 : Résumé de la méthodologie Scout-MRM dans le cadre de la publication *High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics*. 1<sup>ère</sup> étape : Étude protéogénomique par nano-RPLC/MS en mode DDA chez *G. fossarum* pour élaborer une base de données. 2<sup>ème</sup> étape : Analyses en mode MRM classique pour déterminer les peptides détectés dans le nouveau système constitué de la RPLC couplée à un Qtrap. 3<sup>ème</sup> étape : Choix des composés scouts et répartition des transitions MRM dans les différents groupes de détection. Validation de la méthode en évaluant les domaines de linéarité, les limites de quantification, la répétabilité et la sélection des transitions non interféré. Dernière étape : Application de la méthode de biosurveillance protéique pour l'évaluation de la pollution des eaux douces provenant de différents sites agricoles [86].

### 3. Publication n°1 : High-multiplexed monitoring of protein biomarkers in the sentinel *G.fossarum* by targeted Scout-MRM assay, a new vision for ecotoxicoproteomics

Journal of Proteomics 226 (2020) 103901



Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: [www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)



High-multiplexed monitoring of protein biomarkers in the sentinel *Gammarus fossarum* by targeted scout-MRM assay, a new vision for ecotoxicoproteomics



Julien Faugere<sup>a,1</sup>, Duarte Gouveia<sup>b,c,1</sup>, Sophie Ayciriex<sup>a</sup>, Arnaud Chaumot<sup>b</sup>, Christine Almunia<sup>c</sup>, Adeline François<sup>b</sup>, Jean Armengaud<sup>c</sup>, Jérôme Lemoine<sup>a</sup>, Olivier Geffard<sup>b</sup>, Davide Degli-Esposti<sup>b</sup>, Arnaud Salvador<sup>a,\*</sup>

<sup>a</sup> Université de Lyon, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, CNRS UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France

<sup>b</sup> INRAE, UR RiverLy, Laboratoire d'écotoxicologie, F-69625 Villeurbanne, France

<sup>c</sup> Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (LI2D), Service de Pharmacologie et Immunoanalyse (SPI), Commissariat à l'Energie Atomique, Bagnols-sur-Cèze, France

#### ARTICLE INFO

##### Keywords:

Mass spectrometry  
Protein quantification  
Scout-MRM  
Ecotoxicology  
Biomarker  
Multiplex approach

\* Corresponding author.

E-mail address: [arnaud.salvador@univ-lyon1.fr](mailto:arnaud.salvador@univ-lyon1.fr) (A. Salvador).

<sup>1</sup> Equal contribution.

<https://doi.org/10.1016/j.jprot.2020.103901>

Received 7 February 2020; Received in revised form 2 June 2020; Accepted 8 July 2020

Available online 12 July 2020

1874-3919/ © 2020 Elsevier B.V. All rights reserved.

#### Highlights

- Scout-MRM approach was used for the development of a large multiplex protein assay in *G. fossarum*.
- This represents the most important multiplex validated in ecotoxicology field that includes 157 proteins encompassing 277 peptides and 831 transitions, in male and female gammarids at distinct reproductive stages.
- 9 key proteins from caged *G. fossarum* enabled to discriminate agricultural contaminated sites.

#### 3.1. Abstract

Ecotoxicoproteomics employs mass spectrometry-based approaches centered on proteins of sentinel organisms to assess for instance, chemical toxicity in fresh water. In this study, we combined proteogenomics experiments and a novel targeted proteomics approach free from retention time scheduling called Scout-MRM. This methodology will enable the measurement of simultaneously changes in the relative abundance of multiple proteins involved in key physiological processes and

potentially impacted by contaminants in the freshwater sentinel *Gammarus fossarum*. The development and validation of the assay were performed to target 157 protein biomarkers of this non-model organism. We carefully chose and validated the transitions to monitor using conventional parameters (linearity, repeatability, LOD, LOQ). Finally, the potential of the methodology is illustrated by measuring 277-peptide-plex assay (831 transitions) in sentinel animals exposed *in natura* to different agricultural sites potentially exposed to pesticide contamination. Multivariate data analyses highlighted the modulation of several key proteins involved in feeding and molting. This multiplex-targeted proteomics assay paves the way for the discovery and the use of a large panel of novel protein biomarkers in emergent ecotoxicological models for environmental monitoring in the future.

### 3.2. Biological significance

The study contributed to the development of Scout-MRM for the high-throughput quantitation of a large panel of proteins in *Gammarus fossarum* freshwater sentinel. Increasing the number of markers in ecotoxicoproteomics is of most interest to assess the impact of pollutants in freshwater organisms. The development and validation of the assay enabled the monitoring of a large panel of reporter peptides of exposed gammarids. To illustrate the applicability of the methodology, animals from different agricultural sites were analysed. The application of the assay highlighted the modulation of some biomarker proteins involved in key physiological pathways, such as molting, feeding and general stress response. Increasing multiplexing capabilities and field test will provide the development of diagnostic protein biomarkers for emergent ecotoxicological models in the future environmental biomonitoring programmes.

### 3.3. Introduction

Ecotoxicoproteomics is being increasingly used in environmental hazard identification, through the monitoring of protein expression in sentinel organisms exposed to environmental pollutants in both laboratory and field studies [168,201,210,238–241]. Dynamic changes in the molecular machinery of an organism subjected to a toxic stress are the starting points of its physiological response [242]. Molecular biomarkers are therefore able to provide us with early diagnostics of adverse effects in comparison with other higher-level endpoints such as reproductive impairments or other physiological biomarkers.

Among the different mass spectrometry (MS) based strategies available for protein analysis, shotgun proteomics using data-dependent acquisition (DDA) remains the most popular for proteome discovery in ecotoxicological models. In this acquisition mode, a fixed number of precursor ions is selected in the MS<sup>1</sup> survey scan, followed by a sequential isolation and fragmentation of the *N* most intense precursors. This acquisition mode is adapted for high-throughput studies for protein discovery but

presents some limitations regarding its low reproducibility and inability to identify and quantify low-abundant proteins. DDA-based studies have recently proved to be extremely useful for assessing the impact of several model pollutants in the molecular machinery of sentinel organisms from aquatic ecosystems, identifying exposure fingerprints that inform about the pollutants modes of action, and highlighting potential toxicity biomarkers [206,210,212,238,240,241,243–245]. Nevertheless, despite the increasing number of publications proposing new toxicity biomarkers, few are being considered for use in routine environmental biomonitoring. This is mainly due to the lack of high-throughput quantitative assays available for their verification and validation before its implementation in operational monitoring programs.

Alternatively, targeted acquisition methods such as Multiple Reaction Monitoring (MRM) or Parallel Reaction Monitoring (PRM), avoid the lack of reproducibility and limited quantitative power of DDA methods, by focusing the MS/MS scans on a subset of predetermined target peptides. These methods require *a priori* knowledge of the elution time windows of the targeted peptides and the precursor-product ion transitions obtained from a spectral library (subsequent to DDA analysis). MRM assays were proposed in recent years as a promising tool for specific multi-biomarker measurements in environmental biomonitoring [50,215,216]. However, targeted data acquisition experiments such as MRM or PRM exhibit some restraints. One of the limitations of this approach is the number of transitions monitored per peptide restricted by the duty cycle to keep an acceptable signal-to-noise ratio. In addition, the development of large multiplexed assays becomes rather complex because of the RT reliance or unwanted retention time shift due to sample matrix effects. More recently, a new MRM-based targeted method, namely Scout-MRM has been proposed to increase the multiplexing capability and the robustness of classic targeted approaches [45–47]. Briefly, Scout-MRM is based on the successive monitoring of complex transition groups triggered by scout peptide signals distributed along the chromatogram. This method is completely independent from RT and consequently of time scheduling, thereby increasing the multiplexing capability and facilitating the analytical transfer between laboratories.

Herein, we develop for the first time in aquatic ecotoxicology Scout-MRM for the high-throughput quantitation of a list of key proteins resulting from a proteogenomics study in *Gammarus fossarum*, an aquatic model organism used as sentinel to assess freshwater pollution [193,195]. Increasing multiplexing capabilities is of great importance for the development of biomarkers in ecotoxicology since it allows monitoring a broader list of candidate biomarkers and validating a higher number of reliable surrogate peptide biomarkers for developing absolute quantification assays. If robust enough, the simultaneous quantification of hundreds of peptides and proteins also allows performing shotgun-like protein network and/or co-expression analysis, with the advantage of targeting a sub-proteome

covering only the functions of interest. The development and validation of Scout-MRM assay to monitor a large panel of protein's reported peptides from the emergent ecotoxicological model *G. fossarum* is presented. We also demonstrate the interest of the methodology in ecotoxicological studies through an application with active biomonitoring in an agricultural pollution context, followed by a concise discussion around the advantages and innovations of this methodology for environmental monitoring.

### 3.4. Materials and Methods

#### 3.4.1. Reagents and chemicals

Water, acetonitrile (ACN) and methanol (MeOH) were obtained from Fisher Scientific (LC-MS grade, Strasbourg, France). EDTA, triton X-100, iodoacetamide (IAM), dithiothreitol (DTT), formic acid, sodium chloride, aprotinin, ammonium bicarbonate (AMBIC), leupeptin, trypsin (treated TPCK from bovine pancreas), ethylic ether and absolute ethanol were obtained from Sigma Aldrich (St Quentin-Fallavier, France). Isotopically labelled peptides containing either a C-terminal [<sup>15</sup>N<sub>2</sub> and <sup>13</sup>C<sub>6</sub>] lysine or arginine were synthesized by Thermo Fisher Scientific (purity > 97 %) and stored at -20°C until use.

#### 3.4.2. Collection and maintenance of *G. fossarum* organisms

Gammarids were collected by kick sampling from the Pollon river in France, and acclimatized to laboratory conditions, as previously described [171,215,216,246]. This sampling site contains a gammarid population frequently used as a source of organisms for active biomonitoring studies by the laboratory of ecotoxicology in INRAe [246,247]. Before experiments, organisms were kept for two weeks in 30 L tanks continuously supplied with drilled groundwater, which was adjusted with osmotic water to the same conductivity and pH values as the sampling site. The temperature of the water was maintained at 12 ± 1°C, with 16/8h light/dark photoperiods. Organisms were fed *ad libitum* with water-conditioned alder leaves (*Alnus glutinosa*) and tubifex were added once a week.

#### 3.4.3. Sample preparation for shotgun proteomics

Whole-body of five male and five female gammarids were disrupted in extraction buffer (50 mM Tris-Base, 100 mM NaCl, 1 mM EDTA, 0.1% v/v Triton X-100, 6 M Urea, protease inhibitor cocktail) with a Tissue ruptor device (Qiagen). Homogenates were centrifuged at 10000g for 7 minutes, and the supernatant transferred to a new tube. 30 µL of homogenate were subjected to a short SDS-PAGE migration at 200 V (roughly 3 minutes). The whole protein content from each sample in the gel was cut and processed with Trypsin Gold (Promega) and 0.011% ProteaseMax surfactant (Promega) as previously described [248].

#### 3.4.4. Sample preparation for Scout-MRM assay

Specimens (whole bodies) were homogenized in Tris buffer (Tris-HCL 50 mM, 100 mM NaCl, 1 mM EDTA, Triton X100 0.1% v/v, adjusted to pH 7.8 plus 10 µg/mL of each leupeptin and aprotinin), in a volume of 600 µL, with a bead homogenizer. All the homogenates were centrifuged at 10000 g at 4°C for 15 min. A volume of 250 µL of clear supernatant was collected and delipidated by adding 750 µL of ethanol/diethylether solution (1:1 ; v/v). The mixture was vortexed and kept on ice for 10 min. Then, samples were centrifuged at 10000 g at 4 °C for 10 min. Clear supernatant was removed and the pellet was resuspended with 250 µL Tris buffer. A volume of 3 mL of ammonium bicarbonate 50 mM and DTT at a final concentration of 15 mM was added to the supernatant. Samples were incubated for 40 min at 60°C. After cooling to room temperature, iodoacetamide was added (final concentration of 15 mM) and the samples were placed in the dark at room temperature for 40 min. Then, 300 µg of trypsin was added and samples were incubated for 1h at 37°C. Finally, 40 µL of formic acid was added to stop the reaction. 10 µL of heavy peptides at 4 µg/mL and 10 µL of scout peptides at 5 µg/mL were then added. Afterwards, a solid phase extraction was carried out using the 60 mg hydrophilic lipophilic balance (HLB) cartridge Oasis from Waters (Millford, MA). Cartridges were preconditioned with 1 mL of methanol and 1 mL of water acidified with 0.5% of formic acid. Samples (3 mL) were then deposited on the cartridges. Afterwards, cartridges were rinsed with 1 mL of solution of water/methanol (95:5 ; v/v) acidified at 0.5% of formic acid. The analytes were eluted using 1 mL of methanol acidified at 0.5% of formic acid into an Eppendorf tube. 100 µL of MeOH/glycerol mixture (90:10 ; v/v) is added to the sample. Then the solvent was evaporated under nitrogen flow until a volume of 10 µL, which was diluted back with 90 µL of a H<sub>2</sub>O/ACN mixture (90:10 ; v/v) with 0.5% of formic acid, and centrifuged at 12,000 rpm for 5 min at room temperature, before the LC-MS/MS analysis [50,237].

#### 3.4.5. Standards (SIL peptides, scout peptides) preparation

Isotopically enriched peptide stock solution was prepared from lyophilized peptides dissolved in H<sub>2</sub>O/ACN mixture (50:50 ; v/v) with 0.5% of formic acid to obtain a 400 µg/mL stock solution. Scout peptides used in this study are stable isotope labelled peptides, selected to trigger the analysis of MRM transition groups with Scout-MRM. Solutions containing isotopically enriched peptides were prepared from stock solutions and diluted to obtain the desired concentrations. The list of the labelled peptides and scout peptides are shown in supplementary information Table S1.

#### 3.4.6. Nano LC-MS/MS analysis on high resolution mass spectrometry

Extracted and digested proteins were analysed through data-dependent acquisition mode on a Q-Exactive HF mass spectrometer (Thermo Fisher) coupled to an Ultimate 3000 LC system (Dionex-LC Packings). Tryptic peptides were first desalted on a reversed-phase PepMap 100 C<sub>18</sub> µ-precolumn and



separated on a nanoscale PepMap 100 C<sub>18</sub> nanoLC column (3 mm, 100 Å, 75 µm i.d. 50 cm, Thermo Fisher Scientific) with a 90 min gradient of ACN with 0.1% formic acid, and a flow rate of 0.3 µL/min. Full MS were acquired from 350 to 1800 *m/z* and the 20 most abundant precursor ions were selected for fragmentation with 10 s dynamic exclusion time. Only ions with +2 or +3 charges were selected for MS/MS analysis.

### 3.4.7. Liquid chromatography and targeted mass spectrometry

LC-MS/MS analysis was performed on an 1290 series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to a QTRAP® 5500 LC-MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems / MDS Analytical Technologies, Foster City, CA, USA) equipped with a Turbo V™ ion source. The LC separation of the 20 µL injected sample was carried out on an Xbridge C<sub>18</sub> column (100 mm × 2.1 mm, particle size 3.5 µm) with a symmetry C<sub>18</sub> guard column (2.1 mm x 10 mm, particle size 3.5 µm) from Waters (Milford, MA, USA). Elution was performed at a flow rate of 300 µL/min with water containing 0.1% of formic acid as eluent A and acetonitrile containing 0.1% of formic acid as eluent B, employing an isocratic gradient from the beginning at 5% B to for 2 min, followed by a linear gradient from 5% B to 35% B in 36 min. Column was washed at 100% B for 5 min and re-equilibrated at 5% B for 5 min. The injection duty cycle was 48 min, considering the column equilibration time. Instrument control, data acquisition, and processing were performed using a modified Analyst 1.6.2 software®. For academic research, Scout-MRM provisional software patch is available on request from Sciex company (contact: yves.leblanc@sciex.com). MS analysis was performed in positive ionization mode using an ion spray voltage of 5500 V. The curtain gas flows and the nebulizer were set at 50 psi using nitrogen. The TurboV™ ion source was set at 550°C with the auxiliary gas flow (nitrogen) set at 40 psi. The software Skyline v4.1 (MacCoss Lab Software, USA) was used to product a list of suitable SRM transitions. The mass in Q1 and Q3 as well as the collision energy (CE) and the declustering potential (DP) values were predicted using Skyline for endogenous peptides when the associated heavy peptides were not available. The duty cycle is set at 1.52 s in order to attain fifteen points per chromatographic peak for each MRM transition in one of the thirteen scout groups. For peptides with corresponding associated heavy peptides, their mass parameters have been optimized by direct infusion into the mass spectrometer. From the initial set of candidates SRM transitions, three transitions by peptides were selected for the final assay. Details of parameters are reported in Table S2.

### 3.4.8. MS/MS spectra interpretation from GFOSS database

MS/MS spectra were assigned to peptide sequences by searching against a customized RNAseq-derived database GFOSS, which was previously published [220]. This database contains 1 311 444

putative protein sequences and 289 084 257 residues. The algorithm from the MASCOT Daemon v2.3.2 search engine (Matrix Science) was used for database search and spectral matching. The following parameters were used: 5 ppm peptide tolerance and 0.02 Da MS/MS fragment tolerance, +2 or +3 peptide charge, a maximum of two missed cleavages, carbamidomethylation of cysteine as fixed modification, oxidation of methionine as variable modification, and trypsin as proteolytic enzyme. Mascot results were parsed with IRMa 1.31.1c software [249]. Peptide-spectrum matches presenting a MASCOT peptide score with a p-value of < 0.05 were filtered and assigned to a protein according to the parsimony principle. The data have been uploaded to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD017017 (<https://doi.org/10.6019/PXD017017>) [250].

### 3.4.9. Candidate list for Scout-MRM assay development

From the peptide and protein lists obtained in the DDA mode, peptides that did not match the requirements needed for the targeted approach were filtered out of the final list of candidates. At the precursor level, the parameters chosen for selection were the following: idotp > 0.92 (isotope dot product), mass error  $\pm$  4 ppm,  $\geq$  4 peptides per protein. Only the four most intense precursors (based on MS peak area) per protein were kept. At the MS/MS level, all product ions with  $m/z$  higher than 1050 were eliminated, and a maximum of 4 product ions per precursor were kept (the most intense ones). The list was further reduced by removing proteins with whose functional annotations suggested little interest to be used as toxicity biomarkers (mostly housekeeping and orphan proteins). A final list of 2942 transitions, 919 peptides, and 263 proteins was obtained (Table S3).

### 3.4.10. Method validation for Scout-MRM assay

Verification of MRM transitions free from interferences and the correlation of peptide ratios were performed using a correlation study of transition ratios per peptide. The correlation study was carried out on 46 male and 24 female gammarids at different stages of reproduction. A peptide measured with three MRM transitions is considered as non-interfered if at least one of the transition ratios has a RSD < 20%. The remaining analytical parameters were performed from a mixture of male and female *G. fossarum* in order to have samples containing the same amount of protein digest as background matrix. The repeatability of the method including SPE, evaporation under nitrogen flow and MS measurement were evaluated by performing analyses of independent *G. fossarum* samples (n=3) spiked with 500 ng/mL of heavy peptides. The repeatability of the entire analytical protocol including the digestion step for endogenous peptides was evaluated by analysing independent samples (n=10) from the same gammarid pool. Matrix effects, SPE recoveries and evaporation recoveries were evaluated by preparing *G. fossarum* sample spiked with heavy peptides at different protocol steps at



a concentration of 500 ng/mL (n=3). The different recoveries (after SPE and evaporation) and matrix effect are calculated according to the following formulas:

- $SPE\ recovery\ (\%) = (Heavy\ peptide\ area\ add\ before\ SPE / Heavy\ peptide\ area\ add\ after\ SPE) \times 100$
- $Evaporation\ recovery\ (\%) = (Heavy\ peptide\ area\ add\ after\ SPE / Heavy\ peptide\ area\ add\ after\ evaporation) \times 100$
- $Matrice\ effect\ (\%) = [(Heavy\ peptide\ area\ in\ matrix / Heavy\ peptide\ area\ add\ in\ solvent) - 1] \times 100$

In order to carry out the linearity study of heavy peptides, as well as the evaluation of detection limits (LOD) and quantification limits (LOQ), 9 pools of *G. fossarum* peptide solutions were spiked with an increasing quantity of heavy peptides from 1 to 10000 ng/mL. Each calibration curve has at least 5 levels of concentration (between 1 and 10000 ng/mL, 10 levels of concentrations were investigated) with an accuracy between 80 and 120%. In addition, a weighted least-square linear regression was used ( $1/x^2$ ). To determine the LOD, we measured the signal-to-noise ratio in the detection area of the chromatographic peak and considered the 3:1 ratio as acceptable to define it. The LOQ is defined as being equal to three times the LOD. Protein extraction and digestion kinetics in *G. fossarum* have already been optimized and described in a previous publication [50].

### 3.4.11. Data analysis

The integration of chromatographic peaks and data reprocessing was performed with MultiQuant™ software (version 2.1.1, Sciex®). Peak areas for each peptide were log2 transformed to perform multivariate analyses. In order to identify possible outliers, sample clustering was performed using the hierarchical clustering function implemented in the lumi R package [251]. Principal Component Analysis (PCA) was used to analyse and identify the variables explaining the maximum variance associated to the proteomic data in male gammarids caged in the contaminated or the reference sites. Differential protein abundance analysis was performed using linear models and empirical Bayes methods implemented in the limma R package [252,253]. Peptides showing differences with a BH adjusted p-value (FDR) < 0.1 are considered significantly different. Targeted proteomics data have been uploaded to the PeptideAtlas SRM Experiment Library (PASSEL) under dataset identifier PASS01501.

### 3.4.12. Application of the assay

#### 3.4.12.A. Selection and exposure of organisms

For each experiment, organisms were collected from the water tanks at specific reproductive stages. For method validation, mature male and female organisms were sampled based on visual observation

of couples having a female in an advanced stage of the reproductive cycle (well-developed embryos in the brood pouch as described in [195]). For the field studies, couples of *G. fossarum* were sorted in with female in the final molting stages D2 and placed by 7 in punctured polypropylene cylinders with alder leaves as food supply. The day after, four cages of 7 couples were deployed in each study site during three weeks following the reproductive bioassay protocol described in [254]. The organisms were then brought back to the laboratory in water from the study site. Seven males by sites were weighted and directly frozen in liquid nitrogen and stored at -80°C for the proteomic analysis.

#### 3.4.12.B. Study sites

Organisms were caged in four distinct study sites in the “Jura” region in France for a three week exposure. Three sites (A, I, Av) were localized along the river « La Madeleine » (N 46°42'15"; E 5°31'9") which drains a mixed agricultural watershed (wine, corn and wheat crops) before joining drinking water catchments of the City of Lons-le-Saunier. According to the pesticide program commissioned by the water authorities of Lons-le-Saunier between 2011 and 2017, this watershed is at risk for pesticide contamination. A fourth study site situated upstream of the “Seille” river (N 46°42'58"; E 5°37'54") was used as a reference site. This station is located at less than 10 km of the east of the “Madeleine” watershed and is in a good chemical and ecological status among the National Reference Network implemented for the European Water Framework Directive.

### 3.5. Results and discussion

#### 3.5.1. Selection of proteins of interest from proteogenomics and DDA analysis

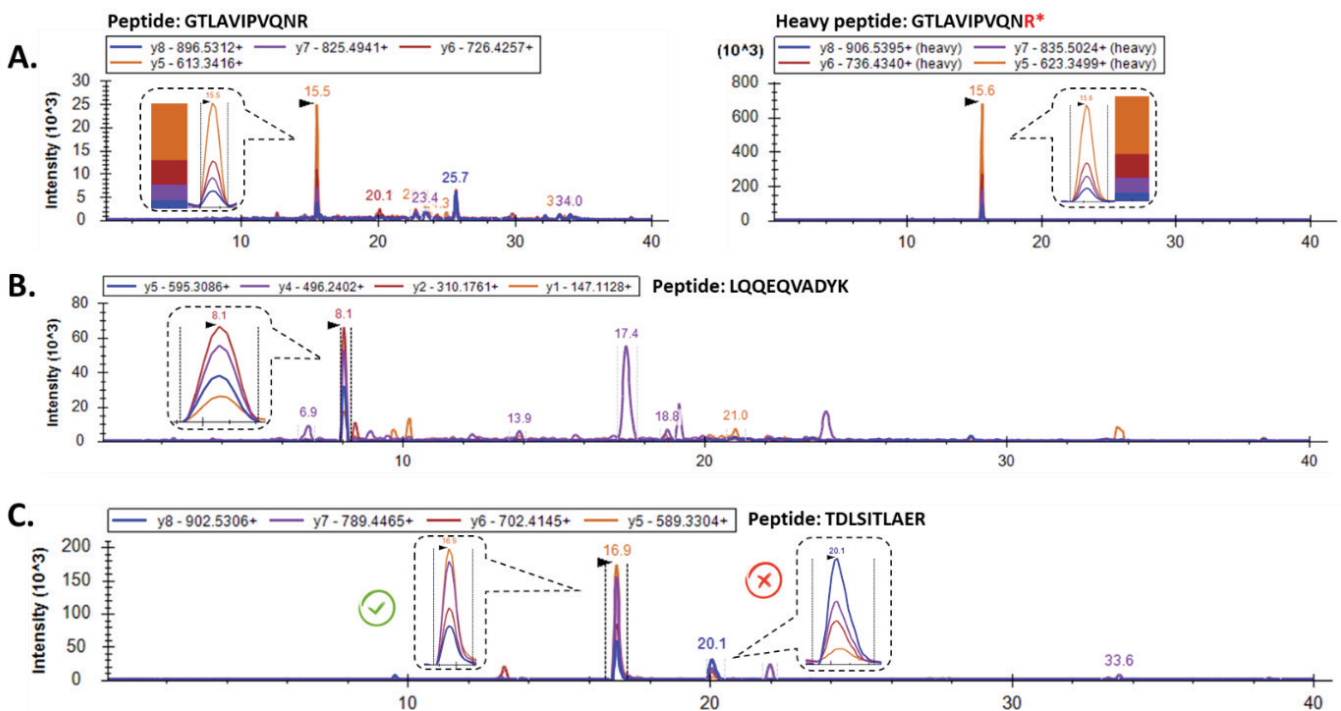
In the shotgun experiment, an average of 46068 MS/MS spectra were recorded per sample, from which 27-29% were attributed to peptide sequences after querying the GFOSS proteogenomics database. These peptide-to-spectrum matches (PSMs) allowed the identification of 5298 protein sequences, which were functionally annotated through BLAST searches and used as the starting point for the selection of the candidates for the targeted assay. The *in silico* mining of this large list of proteins allowed to select a panel of 263 proteins according to the criteria described in section 3.4.9. These proteins are involved in key physiological functions such as reproduction, immunity, homeostasis, and detoxification / defence mechanisms, and are therefore susceptible to be disrupted by exposure to toxic compounds. The list of the proteins is shown in Table S3.

#### 3.5.2. Reporter peptide identification and MS optimization in targeted MS approach

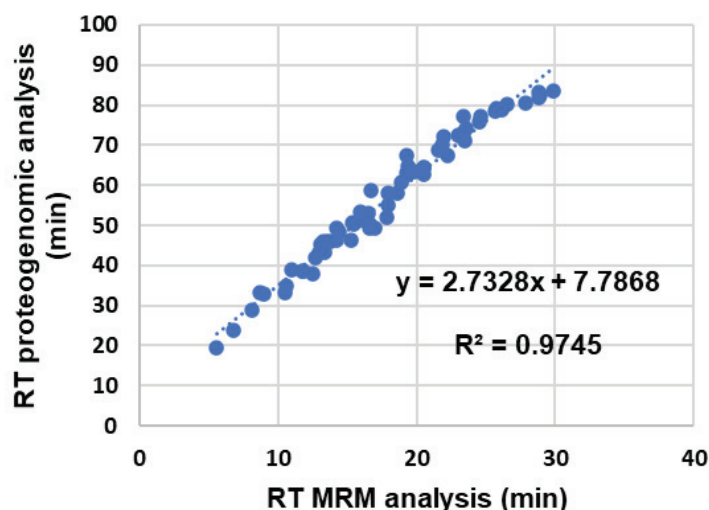
The list of 263 proteins was used to develop the Scout-MRM method on a triple quadrupole mass spectrometer. This type of acquisition method is more suitable for quantification because it provides more robustness. The specificity and sensitivity of targeted-based assays depend on the suitable

selection of the proteotypic peptides. For the unambiguous peptide identification in low resolution targeted MS, four MRM transitions corresponding to the most intense MS/MS fragment ion peptides were selected from proteogenomic experiments, which correspond to 3676 transitions among 27 MRM methods. The chromatographic conditions of the proteogenomic study and the Scout-MRM method are different. Indeed, nano-LC configuration was not used for the development of Scout-MRM method. It was therefore necessary to redefine the retention times (RT) of each peptide. RT information was used as a reference for peptide detection in our system. A retention time was assigned to a peptide, when the monitored transitions were perfectly aligned.

As shown in figure n°42 A, the extracted ion chromatograms showed four overlapped and identical MRM transitions for GTLAVIPVQNR and for their corresponding heavy peptide GTLAVIPVQNR\*. Figure n°42 B shows extracted ion chromatogram of LQQEQVADYK at 8.1 min with the 4 overlapped transitions that confirms the elution time of this peptide. In certain cases, difficulties can be encountered for the determination of the RT of a proxy peptide that exhibits 4 aligned transitions when an interference occurred and the corresponding labelled peptide is not available. As shown in figure n°42 C, the extracted ion chromatogram exhibits several peaks for which the 4 transitions for TDLSTITLAER are perfectly aligned. To overcome this problem *i.e.* to correctly identify the RT area of the target peptides, we established a correlation curve between peptide RT obtained in targeted-MRM based assay and high-resolution-MS<sup>2</sup> based experiment (Figure n°43 and table S4).



**Figure 42 : Identification of endogenous peptides from *G. fossarum* by MRM and selection of transitions.** (A) Detection of endogenous peptides when labelled peptides are available. Retention times and transition ratios must be identical. (B) Detection of endogenous peptides when isotopically enriched peptides are not available. The four MRM transitions must be aligned. (C) Detection and selection of peptide with an isobaric interference also having an alignment of the four MRM transitions. The RT area in MRM mode was predicted based on proteogenomics data.



**Figure 43 : Correlation curve plotting the retention time of peptides identified by proteogenomics to the peptide retention time obtained in MRM.** Retention times of 66 peptides obtained in nanoLC-MS/MS (reversed-phase PepMap 100 C18  $\mu$ -precolumn, nanoscale PepMap 100 C18 nanoLC column, 3 mm, 100 Å, 75  $\mu$ m i.d. 50 cm and Q-Exactive HF) and LC-MS/MS (symmetry C18 guard column, 2.1 mm x 10 mm, particle size 3.5  $\mu$ m, Xbridge C18 LC column, 2.1 mm, 10 cm, particle size 3.5  $\mu$ m and Qtrap 5500) have been reported (Table S4).

Even if peptide separation conditions were very different (nanoLC *versus* microLC configuration, different stationary phase), the RT of the target peptides have been easily identified from other interfering isobaric compounds by using the formula of the regression curve. For 187 peptides, we had also to corroborate our results by performing additional MRM runs by adding supplemental transitions, *i.e.* 1879 transitions. Finally, after the optimization of the precursor and transition selection, 341 peptides proxy of 182 proteins have been perfectly identified on the QTrap system.

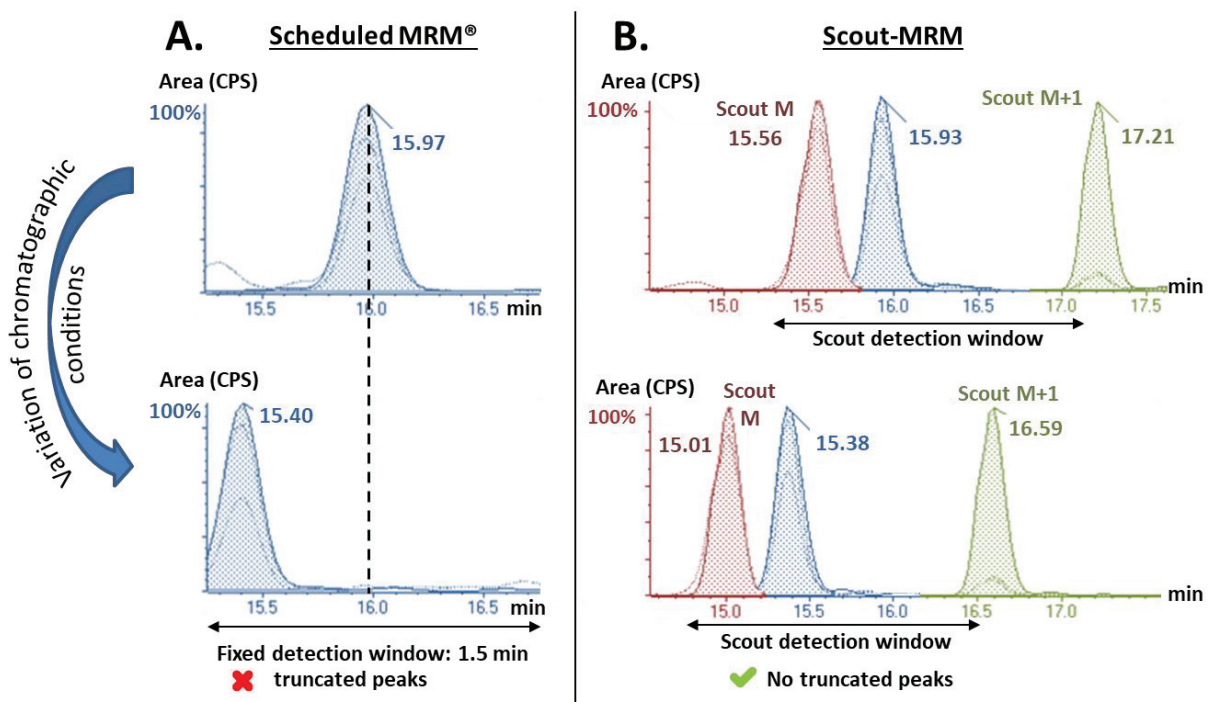
### 3.5.3. Improvement of multiplexing capacity

In a second step, we evaluated the capacity to configure a 182-plex MRM protein assay. This assay requires following 1155 transitions including 44 heavy labelled peptides (up to 3 peptides per protein, at least 3 transitions per peptides). Considering the large number of transitions, it was not possible to develop a single MRM method. To solve this problem, an acquisition method capable of segmenting the chromatogram is needed to focus the mass spectrometer on specific areas associated with retention time of compounds of interest. For this purpose, we can use a single time fragmentation method (called Scheduled MRM<sup>®</sup>, Time MRM<sup>®</sup> or Dynamic MRM<sup>®</sup> according to the MS supplier).

#### 3.5.3.A. Principle and limitations of MRM methods with time window.

To guarantee a constant acquisition setting of 10-15 points per chromatographic peak in time window scheduling experiment, MRM transitions of each peptide are centered on their expected retention times. A constant cycle time is maintained with the simultaneous adjustment of the dwell time according to the number of peptides co-eluted in the same window. To ensure highly multiplexed

analysis, RT scheduling methods must monitor the transitions across small time segments. However, any unexpected RT shift out of the scheduled RT window (modification in chromatographic conditions, change of volume delay with distinct pumping systems during a method transfer, injection overload of the column due to samples with different protein contents) can compromise assay robustness and its implementation for large-scale analysis. As shown in figure n° 44 A, an artificial modification in HPLC conditions, *i.e.* decrease of the isocratic step at the beginning of the gradient, leads as expected to a RT shift of QFQYTWR peptide during retention time scheduling method. The chromatographic peak is truncated, and the quantification is difficult to achieve. In this case, reducing the RT window detection proves to be detrimental to the multiplexed analysis. In some cases, peptides may no longer be detected because they are outside the detection windows.

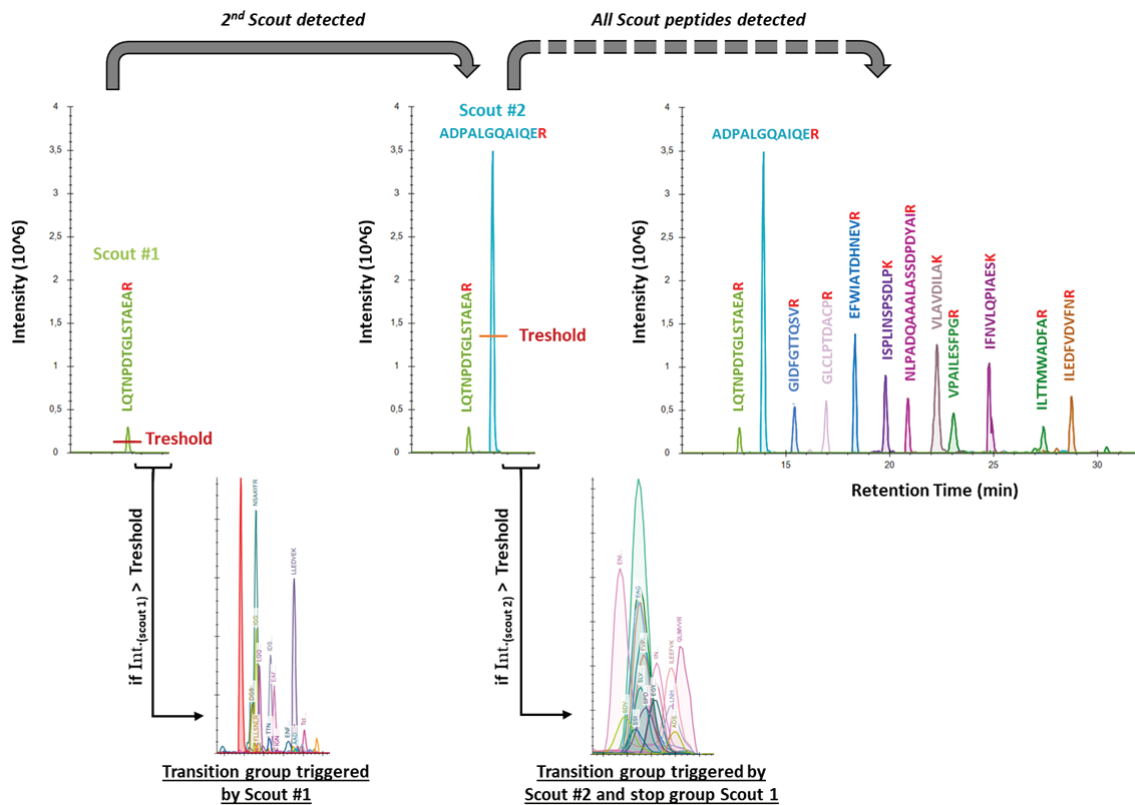


**Figure 44 : Effects of RT shifts for targeted peptides monitored by Scheduled-MRM® versus Scout-MRM method.** (A) When the chromatographic conditions change (decrease in the isocratic level at the beginning of the gradient) a RT shift occurs and induces an offsetting of the chromatographic peak from the detection window. (B) If any RT shift occurs in Scout-MRM method, scout peptides automatically realigns the detection window.

### 3.5.3.B. Principle and advantages of Scout-MRM

To avoid partial or no peak detection, we introduced a new mode of targeted data acquisition called Scout-MRM in order to rationalize the development of targeted proteomics assay and to facilitate dissemination of ready-to-use methods [45,47]. Scout-MRM relies on the monitoring of complex transitions successively triggered under the detection of scout peptides. Compared with scheduled methods, the acquisition was triggered by scout peptides for each segment, instead of by pre-defined scheduled time windows (here in this case, segment of 1.5 min). No extra adjustment for acquisition windows is needed. As we observed, the peptide is correctly eluted during chromatographic variations

(Figure n°44 B). In practice, the scout peptides chosen in this study are 12 stable isotope labelled peptides that can be used for relative quantification (Table S1). Scout peptides triggering groups of transitions (13 groups), are regularly dispatched all along the chromatogram (Figure n°45). When the intensity of the MRM transition of the first scout exceeds a defined threshold, the monitoring of a transition group is triggered. The follow-up of the group stops when the second scout peptide is detected and consequently the second group is triggered and so on [45,47] (Figure n°45).



**Figure 45 : Scout-MRM concept.** This acquisition mode consists in the detection of a first compound called scout. When the intensity of the scout MRM transition exceeds a threshold defined by the operator, the monitoring of an unordered transition group as well as the monitoring of a new scout transition and so on until the last group is triggered by scout n. The triggering of a new group ends the follow-up of the previous one.

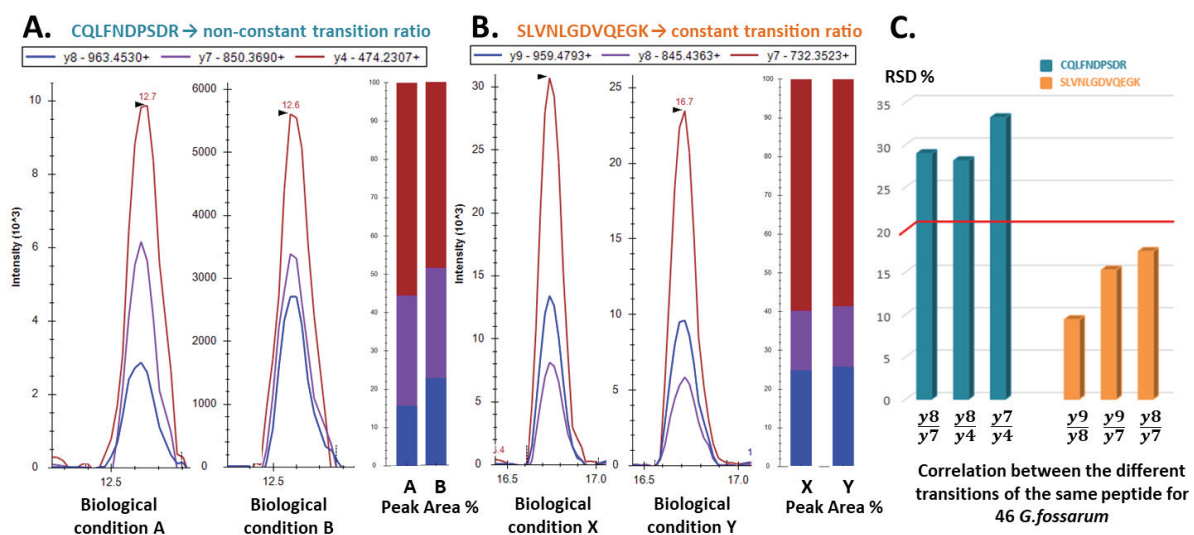
Any incidental RT shift is completely without consequence on the target detection of peptides. To ensure highly multiplexed analysis, RT scheduling method monitors the transitions across large segments generally set to less than 1 min that ensures a compromise between multiplexing capacity and robustness of the method [45]. In Scout-MRM it is sufficient to distribute scouts with an interval of less than one minute to have a higher gain in multiplexing capacity without loss of robustness. To increase multiplexing capacity for future analysis, it is therefore enough to increase the number of scout groups in the method.



### 3.5.4. Scout-MRM method optimisation

#### 3.5.4.A. Biological validation of MRM transitions

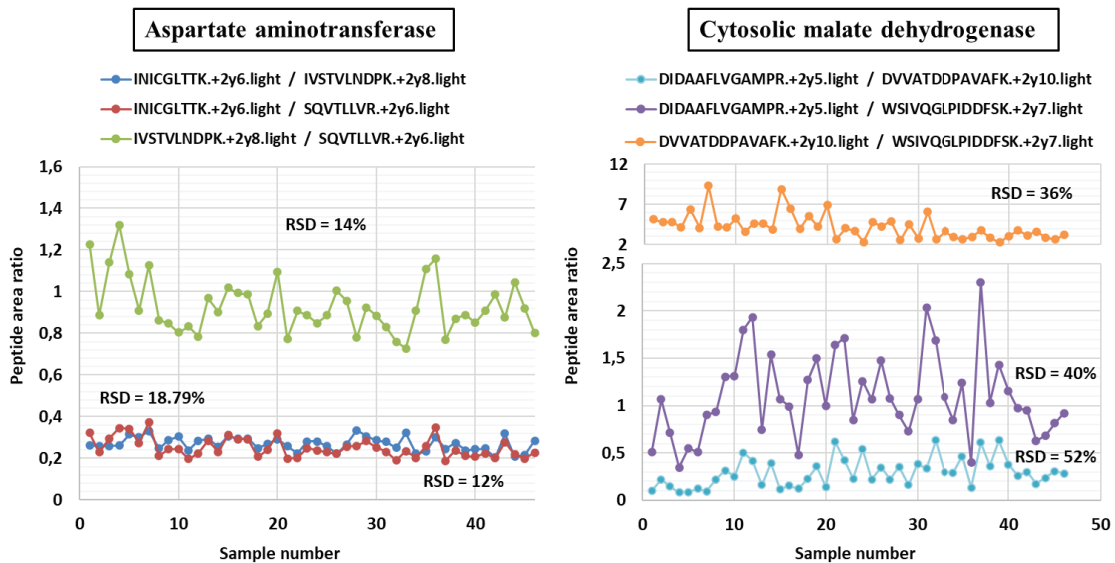
To ensure that the selected MRM transitions were not interfered in the different biological conditions, we verified that transition ratios from a specific peptide are constant. In order to be most relevant several samples need to be analysed. From the sample available at the laboratory at the time of the study, 46 adult male sampled from different locations and 24 adult female organisms at distinct reproductive stages were used to include as much variability as possible. MRM transitions were considered as non-interfered when the ratio between 3 transitions remains constant regardless the sample. From our data we have performed transition area ratio calculation for all samples and estimation of relative standard deviation values for each peptide. A threshold above 20% of MRM ratio is considered as interfered. As shown in figure n°46, the peptide CQLFNDRPSDR exhibits different transition ratios between  $y_8/y_7$ , ions with a RSD > 20%, in two different biological conditions (different male gammarids from different sites). On the contrary, the tryptic peptide SLVNLGDVQEGK conserves the transition ratio between the different ions  $y_9/y_8$ ,  $y_8/y_7$  and  $y_9/y_7$  with a RSD < 20% [255]. After evaluation of all the MRM ratio transitions (341 peptides proxy for 182 proteins), we kept 277 peptides proxy for 157 proteins (Table S2). As a result, some proteins are only reported by one or two peptides in the method. More precisely, 67 proteins with 1 peptide, 42 proteins with 2 peptides and 30 proteins with 3 peptides were followed in male gammarids. For female gammarids, 76 proteins with 1 peptide, 47 protein with 2 peptides and 25 proteins with 3 peptides per protein were monitored.



**Figure 46 : Correlation study between 3 MRM transitions of the same peptide.** (A) Non-constant transition ratio obtained for the 3 MRM transitions of CQLFNDRPSDR peptide under two different biological conditions. (B) Constant transition ratio obtained for 3 MRM transitions the peptide SLVNLGDVQEGK under two different biological conditions. (C) RSD obtained for the mean transition ratios of CQLFNDRPSDR and SLVNLGDVQEGK peptides in 46 different male *G. fossarum* samples. Only the 3 transitions of the peptide SLVNLGDVQEGK are considered as non-interfered (RSD < 20%).

### 3.5.4.B. Biological validation of peptides

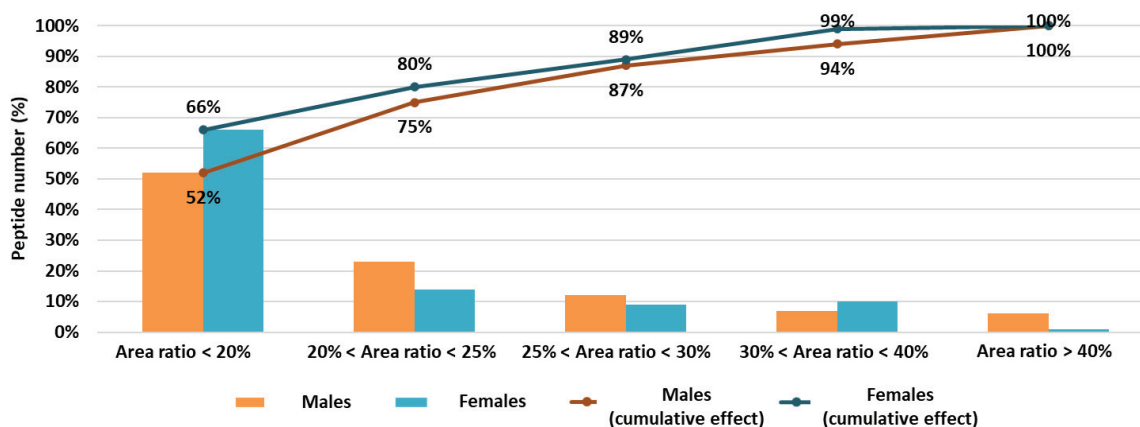
A second study between the different peptides of the same protein was carried out in male and female *G. fossarum*. Indeed, peptides from the same protein must have constant area ratios between the different samples analysed, as shown in figure n°47. This ratio is calculated from the most intense and least interfering transition areas determined in the first correlation study. If the ratios are not respected, this may be due for example to the presence of a modification on one of the peptides of the protein or different matrix effects. Peptide ratios for the same protein are more uncertain than transition ratios of a peptide. Indeed, the different transitions of the same peptide will undergo the same matrix effects because they are eluted at the same retention time while the peptides of the same protein will have different retention times and can therefore undergo different matrix effects between different samples. This criterion was therefore not used to remove a transition from the method. Rather, it has been used as an indicative value to know which peptides within the protein correlate best with each other.



**Figure 47 : Area ratio between 3 peptides of the same protein.** Peptide area ratio between different peptides of the same protein, aspartate aminotransferase and cytosolic malate dehydrogenase, in 46 different male *G. fossarum* samples. For the aspartate aminotransferase, the correlation of the 3 peptides shows good results (RSD < 20%) and for the cytosolic malate dehydrogenase protein, the peptide area ratios show a lower correlation (RSD > 40%).

The results obtained for our correlation study between the different peptides of the same protein are summarized in figure n°48. These results show that for most proteins we have a good correlation between the different peptides.





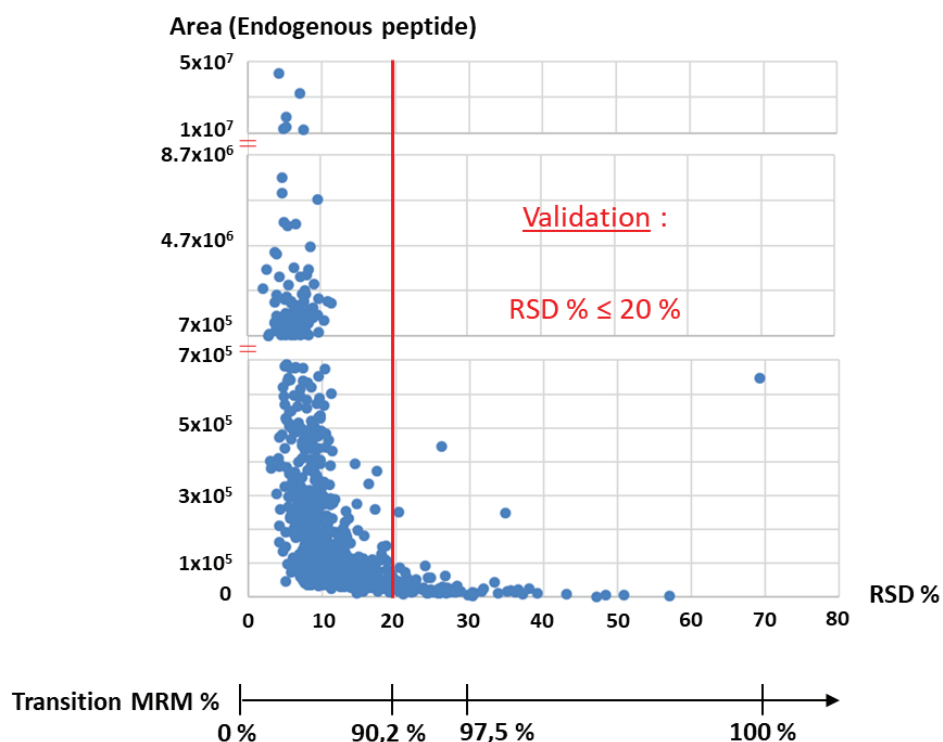
**Figure 48 : Assessment of the correlation study of peptide ratios for the same protein.** For each gender are represented the percentages of protein that have at least one peptide ratio with the indicated correlation coefficient. The orange and blue curves represent the cumulative protein percentages in function of peptide area ratios.

### 3.5.5. Analytical performance evaluation of the assay for quantification

As the multiplexed protein assay will be used to support future ecotoxicology studies, the analytical performances of the assay for quantification were assessed. With our Scout-MRM method, an absolute or relative quantification can be considered. Absolute quantitation is achieved with incorporation for each peptide of isotopically labelled synthetic peptide internal standards. Disposing of all the peptides labelled with high purity results in substantial costs that are often incompatible with the financial means of ecotoxicological studies. In a first step, a relative quantification is performed. Only 44 labelled peptides have been synthesized and will be used for performance evaluation. Since relative quantification method is based on sample comparison, it must first be ensured that the analytical protocol is repeatable. Therefore, a repeatability study was carried out.

#### 3.5.5.A. Repeatability

Intraday repeatability was assessed by spiking heavy peptides to a pool of gammarids at a concentration of 500 ng/mL. All the results are presented in Table S1 as supplementary data. The relative standard deviations (RSD) for this repeatability study is between 1 and 12% that shows good repeatability. However, since we do not have the counterparts marked for each peptide, we have carried out another repeatability study by analysing endogenous peptides from ten independent extractions to consider all the peptides. Furthermore, this study also takes into account the reproducibility of protein digestion into peptides. All the results are presented in Table S5 as supplementary data. Figure n°49 shows that 90.2% of MRM transitions have an RSD of less than 20%. This threshold represents the limit values that have been set to define whether a transition is repeatable. The RSD obtained for all previously selected qualifying MRM transitions are less than or equal to 20%. When the RSD is greater than 20%, this corresponds to the lowest MRM transitions that generate more difficult integrations and are more easily interfered. These results indicate that the digestion step and extraction step are controlled and reproducible.



**Figure 49 : Repeatability of endogenous peptides : Average intensity of the MRM transitions presents in the Scout-MRM method according to their repeatability.** Results obtained after independent extraction from the same pool of male and female *G. fossarum* (n=10). The repeatability validation criterion is set at 20% which corresponds to 90.2% of the MRM transitions of the method. The entire protocol method is therefore well repeatable. When the criterion is not satisfied, this corresponds to the least intense MRM transition not used as a qualifying transition (more easily interferable and more difficult to integrate) (Table S5).

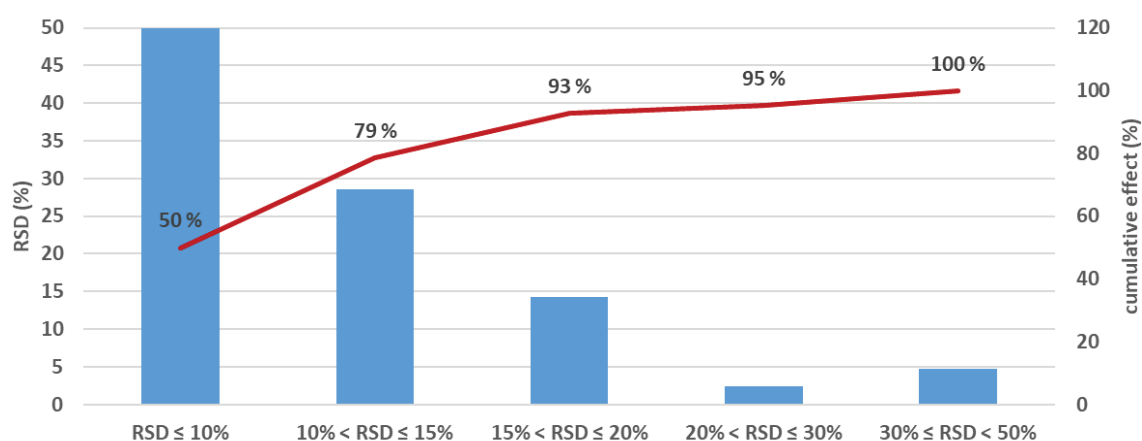
### 3.5.5.B. Extraction recoveries

It was shown in a previous paper that peptides can be lost during evaporation [237]. Therefore, this step was also specifically evaluated in addition to the SPE extraction recovery. To estimate extraction recoveries, digested protein extracts from gammarids were spiked with labelled peptides before and after SPE, and after evaporation at a concentration of 500 ng/mL (n=3) and compared. The results are presented in Table S1. It can be observed that SPE recoveries are between 72 and 122% with an average of 96% for the 44 heavy peptides. Evaporation recovery range from 67 to 106% with an average of 94% for the 44 heavy peptides. These results showed a good extraction recovery.

### 3.5.5.C. Matrix effect evaluation

Matrices effects in MS correspond to the influence of co-eluted compounds during their simultaneous introduction into the source due to ionization competition. This can result in an increase or a decrease of their intensity. To measure these matrix effects, sample extracts were spiked with heavy peptides added just before injection at a concentration of 500 ng/mL (n=3) which were compared to samples containing the same quantity of heavy peptides spiked in reconstitution solvent (n=3). The results are presented in Table S1. The 44 heavy peptides undergo matrix effects that induce a loss or gain of the

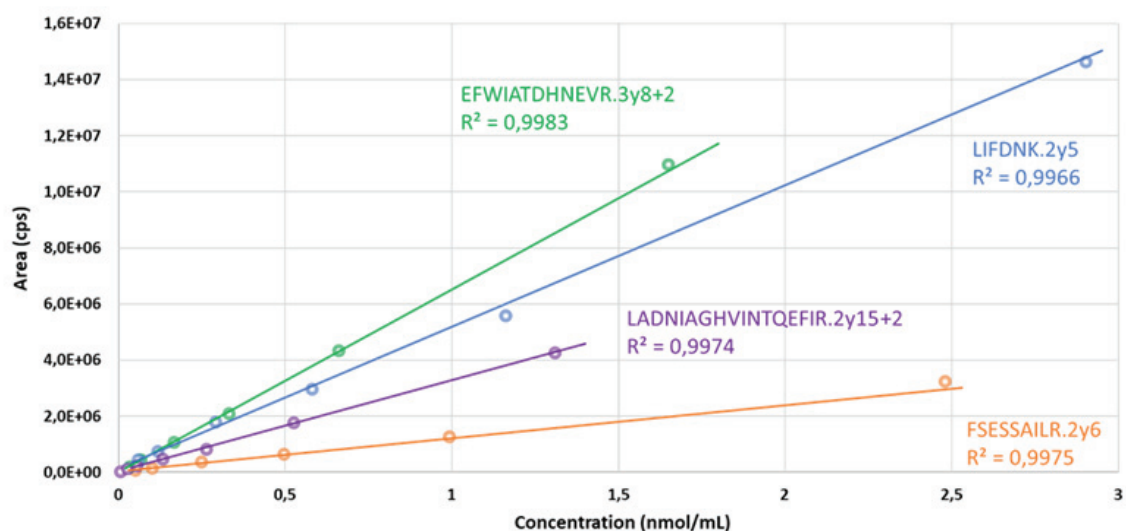
signal between -67% and 51% with an average loss of 28%. As expected, matrix effects occur. However, the case presented here illustrates the most extreme case where a very complex mixture is compared to compounds in solution. The analysis of heavy peptides spiked at the same concentration in biologically different *G. fossarum* extracts showed that even if matrix effects affect the peptides analysed, these effects are comparable from one sample to another. Indeed, the results presented in figure n°50. obtained from the data in table S6 show that for 93% of heavy peptides the RSD are less than 20%. To correct different matrix effects between different samples, internal standards (heavy peptides) similar to endogenous peptides must be used, which will undergo the same matrix effects as the compounds of interest. In our case, we do not have all the heavy peptides corresponding to the 277 peptides selected in the method. Our study aims to highlight trends by identifying potential biomarkers. Once the candidates have been found, it will be necessary to have their counterparts marked to carry out the most accurate quantification possible, considering the matrix effects. This correction of matrix effects will also allow to get peptide ratios for the same protein as precise as possible.



**Figure 50 :** Variation of the areas of heavy peptides spiked in different gammarid extracts. RSD (%) performed on areas of 42 heavy peptides that were spiked before SPE in biologically different male *G. fossarum* extracts to assess whether the matrix effects for the same peptide between different samples are variable (Table S6).

#### 3.5.5.D. Linearity and limit of quantification (LOQ)

Although this study is based on a relative quantification, we have determined the linearity domains and LOQ of our heavy peptides in order to evaluate the performance of the method. A weighted least-square linear regression ( $1/x^2$ ) was used for the calibration. All calibration curves corresponding to heavy peptides spiked into the *G. fossarum* matrix at different levels of concentration confirmed the high degree of linearity ( $0.991 < r^2 < 0.999$ ) (Table S1 and figure n°51). The LOD was determined to be 3 times greater than the background noise ( $S/N=3$ ) and the LOQ is determined as 3 times the LOD. The LOQ are between 1.12 pmol/mL and 113 pmol/mL (Table S1).



**Figure 51 : Example of calibration curves for 4 isotopically enriched peptides.** All calibration curves corresponding to heavy peptides spiked into the *G. fossarum* matrix at different level of concentration with an accuracy between 80 and 120%. In addition, a weighted least-square linear regression was used ( $1/x^2$ ).

Finally, the method developed with Scout-MRM enabled to monitor 157 proteins in male and female species of *G. fossarum* following the validation criteria (157 proteins, 277 peptides, and 831 MRM transitions) among the targeted 263 proteins considered initially.

### 3.5.6. Application of Scout-MRM assay : comparative proteomic analysis using active biomonitoring in an agriculture watershed

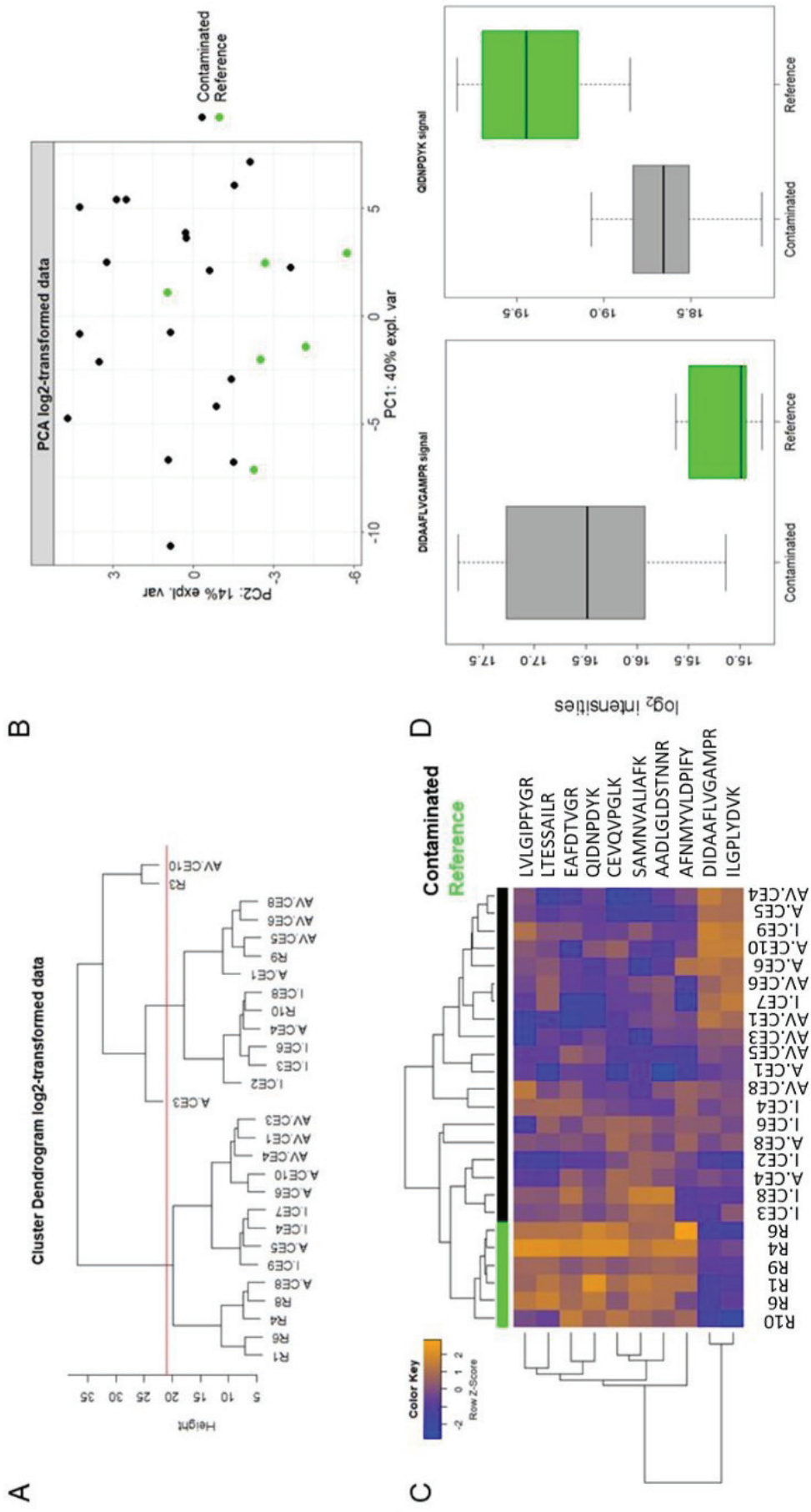
We applied Scout-MRM method to assess its interest and feasibility in the context of active biomonitoring in freshwater streams. Twenty-eight calibrated male gammarids (7 from a non-contaminated reference site, 7 from each of the three contaminated sites) were used for the proteomics analysis. Two hundred sixty-five peptides were detected for each male organism.

To compare the global proteomes of the organisms caged in the contaminated sites or the reference site, we performed clustering and principal component analyses. The cluster analysis identified three organisms as potential outliers, based on their distance with the rest of the samples (Figure n°52 A). These outliers (an organism in the reference site, R3 ; and two organisms in two different contaminated sites, A CE3 and AV CE10) were excluded from the following analyses. The PCA analysis showed that organisms are quite dispersed, even though there is a tendency of the organisms caged in the reference site to cluster more closely than those clustered in the contaminated sites (Figure n°52 B).

In order to identify peptides whose abundance is able to distinguish the organisms caged in the different type of sites, and thus could be identified as potential biomarkers of exposure to agriculture pollutants, a differential analysis was performed. The differential analysis identified 10 peptides as differently expressed ( $FDR < 0.1$ ) (Figure n°52 C). The peptide DIDAAFLVGAMPR, annotated as cytosolic

malate dehydrogenase or malic enzyme (ME) was the most upregulated in organisms exposed in the contaminated sites (logFC=1.38, FDR=0.038) (Figure n°52 D). The *Drosophila* ME activity can be induced by the juvenile hormone (JH) by both a direct effect on the enzyme in the short term and the activation of its gene (*Men*) transcription [256]. Since JH analogs are among the most used insecticides in agriculture, the increased expression of a homolog of ME in *G. fossarum* might suggest a response of these non-target organisms to JH analogs contamination in the observed sites. Among the downregulated peptides, the most significant was QIDNPDYK (logFC=-0.86, FDR=7.9x10<sup>-5</sup>) (Figure n°52 D), belonging to a homolog of calreticulin. Calreticulin is a highly conserved endoplasmic reticulum protein of the lectin family. It is involved in osmoregulation, molecular chaperoning and immune response in crustaceans, with most data available from decapods models [257–259]. Calreticulin has been reported as a general stress biomarker in *in vitro* and vertebrate models [260] and its downregulation may thus suggest a stress effects in amphipods caged in the contaminated sites.

Among the downregulated peptides found in organisms caged in the contaminated site, we found 2 peptides belonging to an homolog of endochitinase (EAFDTVGR, logFC=-0.61, FDR=0.059; LVLGIPFYGR, logFC=-0.55, FDR=0.072) and 2 peptides belonging to a protein annotated as endoglucanase (or cellulase A). (SAMNVALIAFK logFC=-0.76, FDR=0.063; AADLGLDSTNNR, logFC=-1.28, FDR=0.06). Chitinases are key enzymes for successfully complete molting cycle in arthropods [261]. We have previously observed its decreased detection in *G. fossarum* exposed to contaminated sites [216]. Similarly, other crustacean chitinases have been reported as sensitive biomarker to insecticides and fungicides [261]. Cellulases are essential for digestion in *G. fossarum* and their decreased activity and/or abundances may affect its reproductive capacity [259]. Moreover, cellulase activity were reported to be very sensitive to insecticides and fungicides [259]. Overall, these results suggest that multiple protein biomarkers in *G. fossarum* are useful to detect a biological response to a contamination of agricultural origin. The limited number of modulated peptides in the contaminated sites might suggest a certain specificity of the impact of the chemicals present in the aquatic environment to which the gammarids were exposed. However, due to the limited number of investigated sites, it will be crucial to extend the use of these multiplexed protein biomarker approach to a larger spatial scale with contrasted contamination profiles.



**Figure 52 : Comparative proteomic analysis using active biomonitoring in an agriculture watershed.** Analyses were performed using log2 transformed peak areas for each peptide identified in caged organisms. (A) Hierarchical clustering samples over the red line show a higher distance compared with the rest of the organisms, thus they were excluded from the datasets. (B) Principal Component Analysis. Samples caged in the reference site (green) tend to cluster together. (C) Heatmap of differentially expressed proteins in male gammarids caged for three weeks in a reference site (R) and three sites under agriculture pressure (A, I, Av). Colours represent abundance fold changes (orange over detection and blue down regulation). (D) Boxplot representing the differential expression of the most deregulated peptides (DIDAAFLVGAMPR; QIDNPDYK).



### 3.6. Conclusion

A robust LC-MS/MS method for the simultaneous analysis of many potential protein biomarkers in the sentinel species *G. fossarum* has been successfully developed and validated. We applied a new acquisition mode called Scout-MRM recently developed in our laboratory to significantly increase the multiplexing capacity through the implementation of a 157-protein multiplex (277 peptides, 831 MRM transitions) in adult gammarids. Scout-MRM provides a more robust method than acquisition mode using time window scheduling. Scout-MRM is free from retention times, thus limiting the loss of information due to potential RT shifts. The RT independency of Scout-MRM opens up the perspective to build large multiplex by adding more scout peptides and MRM transitions if one wants to follow more proteins of interest in the future. Indeed, since *G. fossarum* genome is still not fully characterized, new discovered proteins can be implemented.

The first application of our method for biomonitoring key proteins to assess freshwater pollution from different agricultural sites demonstrated the potential value of this methodology in ecotoxicology studies. Indeed, the detection of protein reporter peptides with modulation in response to stress shows that the Scout-MRM method is a relevant method to detect biological responses due to contamination. Further studies will allow to identify reference values for the investigated peptides and eventually fine-tuning the choice of more specific reporter peptides of proteins related with different mode of actions involved in the adverse outcome observed in the ecotoxicological bioassays used in the field [254,262]. In conclusion, Scout-MRM streamlines the development of targeted proteomics method in an ecotoxicology study and simplifies dissemination of ready-to-use assays as they are easily transferable from one laboratory to another.

### 3.7. Supplementary material of the first publication

The additional data are too voluminous for this manuscript so additional data documents are available by following the link below : <https://doi.org/10.1016/j.jprot.2020.103901>

Table S1: Results of the validation steps of Scout-MRM assay performed on heavy peptides.

Table S2: Transitions list of endogenous and heavy peptides that has been validated and followed in Scout-MRM method.

Table S3: Final list of 263 proteins selected after shotgun proteomics.

Table S4: Correlation curve plotting the retention time of peptides identified by proteogenomics to the peptide retention time obtained in MRM.



Table S5: Tables representing the average areas of endogenous peptides as a function of measurement repeatability (n=10).

Table S6: Tables representing the areas of heavy peptides spiked before SPE in biologically different male *G. fossarum* extracts to assess whether the matrix effects for the same peptide between different samples are variable.

Table S7: Final Scout-MRM method.

## 3.8. Application de la méthode multiplexée pour l'analyse de cohortes d'échantillons dans le cadre d'études écotoxicologiques

### 3.8.1. Transfert de la méthode de dosage protéique appliquée à *G. fossarum* entre le laboratoire Anabio-MS et l'INRAe

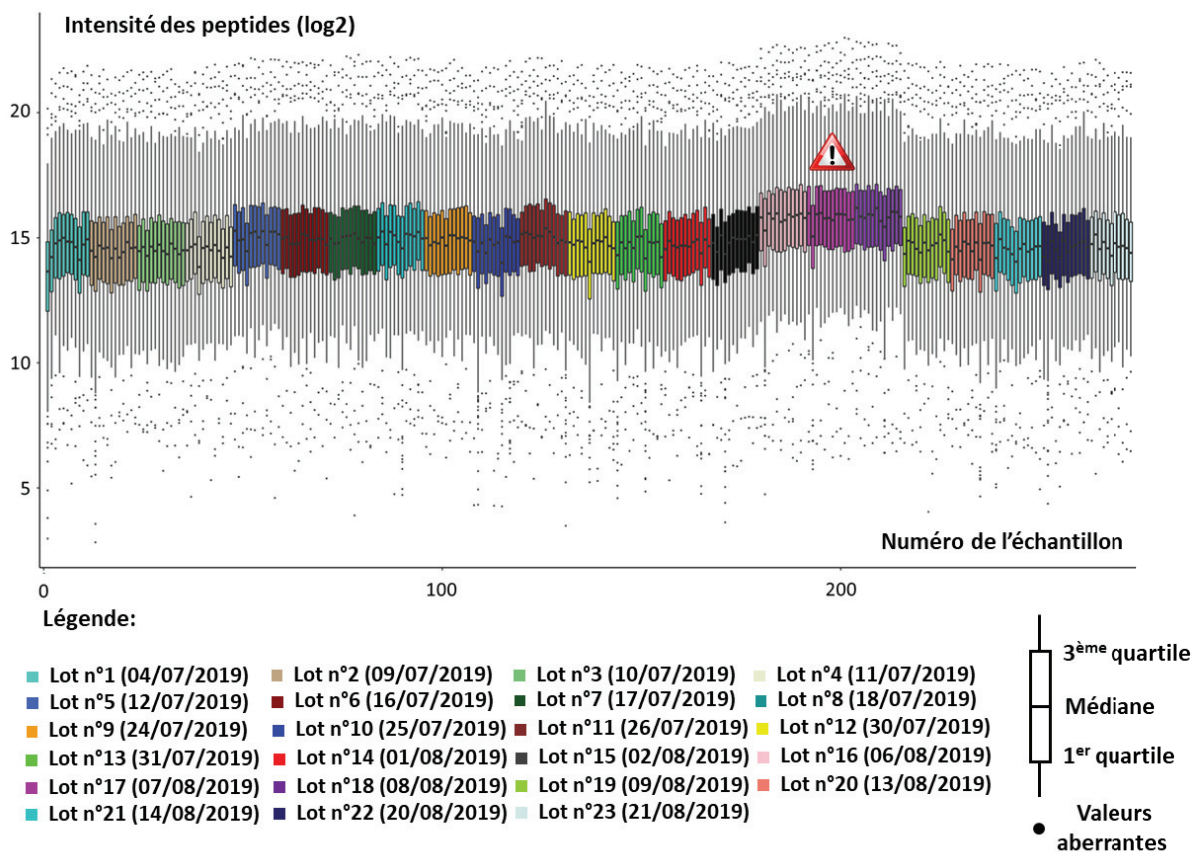
Les études permettant d'affiner le choix des peptides liés aux différents modes d'action impliqués dans les bio-essais écotoxicologiques ainsi que leurs valeurs de références seront conduites par le laboratoire Écotoxicologie de l'INRAe-Villeurbanne. De ce fait, il est nécessaire de réaliser un transfert de méthode entre leur laboratoire et celui de l'équipe Anabio-MS.

Actuellement, la méthodologie Scout-MRM n'est possible que sur les instruments de la société Sciex® (Brevet « Sentinel signal for adaptive retention time in targeted method »). Toutefois, les transitions sélectionnées pour l'analyse de protéines chez *G. fossarum* peuvent être utilisées sur n'importe quel spectromètre de masse de type triple quadripôle en MRM classique ou avec fenêtres de temps en acceptant une limitation des capacités de multiplexage. Ainsi, le transfert de méthode a été réalisé avec succès sur le QQQ 6495 (Agilent®) de l'équipe Écotoxicologie de l'INRAe-Villeurbanne dirigée par Arnaud Chaumot en optant pour un mode analogue à la sMRM (nommé dynamic-MRM). Lors de ce transfert de méthode, il a fallu adapter les paramètres de masse (énergie de collision et paramètres d'ionisation) et vérifier les fenêtres d'acquisition du fait d'un volume de délais différent entre les deux systèmes LC. Ainsi, de nouvelles fenêtres de temps ont pu être mises en place. Cela n'aurait pas été nécessaire si le mode d'acquisition Scout-MRM avait été disponible sur les appareils Agilent. La méthode transférée fonctionne en routine depuis plus d'un an et a permis d'acquérir de nombreuses données. Cela démontre que la quantification multiplexée de protéines est envisageable en routine pour l'écotoxicologie.

Ces nombreuses séries d'analyses réalisées sur des périodes importantes ont soulevé des questions sur la normalisation des données pour la comparaison d'échantillons dans le temps lorsque des effets de lots sont rencontrés.

### 3.8.2. Mise en évidence d'effets de lots lors d'études comparatives avec des approches protéomiques

Les résultats présentés sur la figure n°53 correspondent à l'analyse de 275 échantillons de *G. fossarum* mâles engagés dans 55 sites différents. Cette étude a été réalisée par l'équipe Écotoxicologie de l'INRAE-Villeurbanne dans le but d'identifier des biomarqueurs potentiels de contamination environnementale sur des espèces sentinelles engagées. Au total, 23 séries d'analyses ont été réalisées sur une durée de 48 jours. À partir de ces données, il a été mis en évidence des variations de mesures d'abondances peptidiques entre séries d'analyses qui perturbent la comparaison d'échantillon dans le temps. Les diagrammes en boîte indiquent plusieurs données statistiques, dont la médiane (centre des rectangles colorés). Cette valeur sépare la moitié inférieure de la moitié supérieure de l'ensemble des intensités peptidiques (log 2) pour chaque échantillon.



**Figure 53 : Mise en évidence de l'effet de lots.** Analyse de 275 échantillons de *G. fossarum* par LC-MS/MS provenant de 55 sites différents (n=5). La médiane globale issue de la réponse de l'ensemble des peptides de la méthode est représentée pour chaque échantillon. La réponse en intensité pour chaque peptide est transformée en log2.

Chaque lot correspond à une série d'échantillons analysés le même jour. Les lots n°16, 17 et 18 présentent un effet de lot qui se traduit par une forte augmentation du signal MS global par rapport aux autres séries d'analyses (augmentation de la valeur de la médiane). Les gammars proviennent de nombreux sites avec des conditions expérimentales différentes et chaque lot analysé correspond à une

répartition aléatoire de ces organismes. Il est peu envisageable que l'ensemble des gammes d'un même lot présente une augmentation de signal global, d'autant plus que la méthode comprend de nombreuses protéines impliquées dans des processus biologiques différents. Ainsi, la probabilité est faible qu'une variation environnementale agisse sur l'ensemble des fonctions biologiques des individus au même moment. La cause de cette variabilité n'est donc pas liée à des variations biologiques. Il est plus probable que cela soit dû à l'instabilité du spectromètre de masse. Cette constatation soulève la question de la normalisation des données afin de comparer des résultats entre eux lors de l'analyse de grande cohorte d'échantillons.

Pour résoudre ces problèmes d'effets de lots, plusieurs types de normalisation des données en protéomiques sont en cours d'application pour évaluer leurs avantages et inconvénients. Le premier type de normalisation envisagé est basé sur le signal global avec notamment la normalisation par la médiane [263,264]. Elle utilise l'ensemble des données mesurées pour chaque échantillon afin de déterminer une médiane moyenne. L'utilisation de cette médiane moyenne est soustraite à la médiane spécifique de chaque échantillon pour normaliser les données [265–268]. Bien que ce type de normalisation soit la plus simple à mettre en place, elle suppose que la distribution globale du signal MS est similaire entre les échantillons. Cette normalisation n'est donc pas compatible avec les études quantifiant exclusivement des biomarqueurs puisque la définition même d'un biomarqueur est de présenter une variation lorsque l'on change de conditions expérimentales. Elle ne peut être appliquée que sur des jeux de données exhaustives permettant la découverte de biomarqueurs parmi un nombre de composés important où l'on s'attend à ce que la plupart d'entre eux ne présentent pas de variabilité. Le deuxième type de normalisation utilise un ou plusieurs composés spécifiques dont la concentration ne varie pas, quelle que soit la variation biologique étudiée telle que des protéines/peptides isotopiquement enrichies ainsi que des protéines de ménage. En biologie moléculaire, les protéines de ménage sont issues généralement des gènes constitutifs qui sont nécessaires au maintien de la fonction cellulaire de base (rôle structural, énergétique, *etc.*) et qui sont exprimées à des taux relativement stables dans l'organisme [269,270]. Ces composés corrigent l'ensemble des variations issues de la préparation d'échantillon et de l'analyse par spectrométrie de masse. Toutefois, la normalisation des données permettra de corriger seulement les variations associées aux étapes ultérieures à l'ajout des étalons internes. Les protéines de ménage corrigent l'intégralité de la procédure analytique, car ils sont présents nativement dans l'échantillon. Les protéines et les peptides isotopiquement enrichis corrigent respectivement les variations après l'étape de broyage et après l'étape de digestion. Ces travaux donneront lieu à de futures publications sous la direction de l'équipe d'Arnaud Chaumot (INRAE).

Dans le futur, si l'on souhaite rajouter des protéines et n'avoir qu'une seule méthode d'analyse, il faudra réaliser les expérimentations sur les instruments de la société Sciex®. En effet, seule la méthodologie Scout-MRM permet d'atteindre ces plus grandes capacités de multiplexages en une seule analyse. Lors de l'élaboration de méthodes Scout-MRM nous avons utilisé comme scout des peptides marqués de protéines d'intérêts (cf. 3.5.3.B). Il est également possible d'utiliser comme scout des peptides synthétiques, exogènes et non marqués ou isotopiquement enrichis. Dans ce cas, comme décrit dans le paragraphe suivant, cela permet non seulement de construire de grands multiplexes, mais également de faciliter un transfert rapide entre les approches *shotgun* de découverte et la protéomique ciblée par MRM.

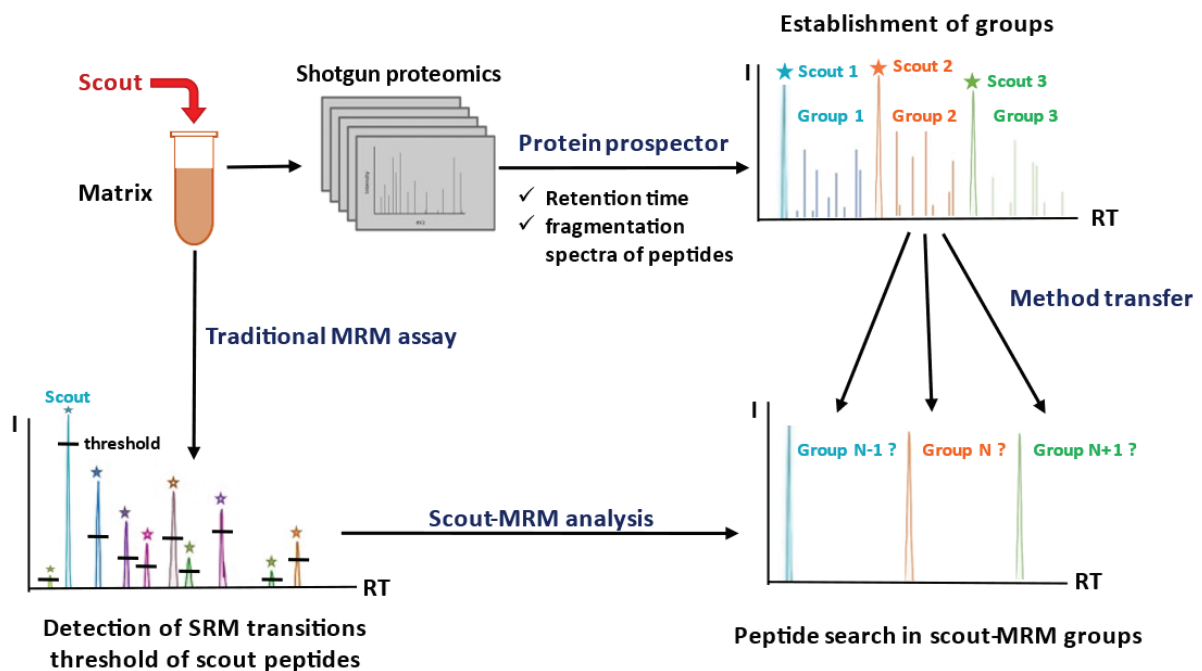
### 3.9. Méthode de transfert rapide entre les approches *shotgun* et ciblée via l'utilisation du Scout-MRM

Le développement de méthodes de dosages protéomiques appliquées à des espèces sentinelles dont le génome n'est pas séquencé est souvent réalisé en deux étapes [86] : (i) identification des protéines d'intérêts grâce à des analyses DDA en haute résolution (ii) suivie par la mise en place de méthodes quantitatives de ces molécules par spectrométrie de masse ciblée. Comme démontré précédemment [86], l'élaboration d'une méthode hautement multiplexée en mode MRM à partir de données issues de la haute résolution en mode DDA n'est pas immédiate et peut être chronophage. En effet, les paramètres LC entre les deux configurations peuvent être différents du fait des chimies des phases stationnaires et des dimensionnements des colonnes. La nano LC est classiquement utilisée lors des approches globales alors que la micro LC est plus souvent utilisée pour les approches ciblées pour des raisons de robustesse. Dans le cas de l'élaboration de la méthode Scout-MRM chez l'espèce sentinelle *G. fossarum*, il a été nécessaire de réaliser 27 méthodes MRM pour pouvoir étudier les 263 protéines issues de notre base de données [86]. Ce transfert aurait pu être plus rapide si les peptides scouts avaient été ajoutés lors de la phase d'identification par protéogénomique. Ainsi, Maxime Lepretre, doctorant en cotutelle entre le laboratoire Anabio-MS et le laboratoire SEBIO, a pu bénéficier dans le cadre de son doctorat de notre recul et notre expertise du Scout-MRM. Cela a permis l'élaboration d'une stratégie permettant le transfert rapide de méthodes entre les approches *shotgun* et ciblées grâce à l'utilisation du Scout-MRM. Ces travaux dans lesquels j'ai été impliqué ont été publiés dans le journal *Analytical and Bioanalytical Chemistry* en août 2020 [271] et sont présentés dans l'annexe n°1.

Comme discuté précédemment, le Scout-MRM est un mode d'acquisition indépendant des temps de rétention. Cette indépendance dans la programmation de fenêtres temporelles rend cette méthodologie particulièrement attractive pour faciliter le développement de méthodes ciblées basées sur des données de protéomique *shotgun*. D'autant plus lorsque différentes configurations de chromatographie liquide sont utilisées. Le but de cet article est de proposer une nouvelle application

du Scout-MRM pour le développement de méthodes protéomiques hautement multiplexées à partir de données issues de la spectrométrie de masse de haute résolution. Cette stratégie a été développée pour la quantification de biomarqueurs protéiques relatifs au système immunitaire de l'espèce sentinelle *Dreissena polymorpha*. Toutefois, elle pourrait être appliquée facilement à d'autres espèces sentinelles comme *G. fossarum*.

Pour réaliser ce transfert de méthode rapide, il faut introduire dans un premier temps les peptides scouts dans les échantillons analysés lors de l'étape de découverte en HRMS comme indiqué dans la figure n°54.



**Figure 54 : Stratégie pour convertir une méthode de quantification en protéomique shotgun vers de la protéomique ciblée, en utilisant la méthodologie Scout-MRM.** (1) Des composés scouts sont ajoutés dans les échantillons et analysés par protéomique shotgun et ciblée. (2) Protein Prospector est utilisé pour extraire les trois ions fragments les plus intenses et le temps de rétention des peptides rapporteurs afin d'établir les transitions MRM et de définir des groupes de peptides rapporteurs encadrés par des scouts. (3) Les transitions MRM des peptides scouts sont analysées pour définir les seuils de détection. (4) Les transitions MRM des peptides rapporteurs sont recherchées dans différents groupes (N, N+1, N-1, et autres) pour leur quantification en protéomique ciblée [271].

Les peptides scouts sélectionnés doivent avoir une séquence en acides aminés unique. Ils doivent n'appartenir à aucune séquence connue à ce jour dans les différentes bases de données pour éviter de polluer les échantillons à analyser. Une première analyse en haute résolution (*Shotgun*) est réalisée pour obtenir le temps de rétention des peptides scouts par rapport aux peptides d'intérêts. Cela permet d'établir les différents groupes scouts dans lesquels placer les transitions MRM cibles. La présence des composés scouts dans l'échantillon analysé en HRMS permet d'utiliser le Scout-MRM sans l'étude au préalable des temps de rétention des peptides endogènes. Dans une deuxième analyse, les composés scouts enrichis dans la matrice sont détectés selon l'approche MRM classique. Cette

étape permet de déterminer les seuils de déclenchement de l'ensemble des groupes scouts dans les nouvelles conditions associées au mode ciblé. Pour finir, une analyse est ensuite réalisée en Scout-MRM sur la matrice enrichie pour identifier les peptides endogènes à partir de la protéomique par spectrométrie de masse ciblée.

Des variations de sélectivité lors du passage du système LC utilisé en approche *shotgun* à celui du mode ciblé peuvent entraîner des changements de groupe scout pour certains composés. De ce fait, certaines transitions risquent de ne pas être détectées. Il faut donc réaliser une injection supplémentaire en ajoutant les composés non détectés dans les groupes scouts N+1 et N-1.

Dans le cas de la publication *High-multiplexed monitoring of protein biomarkers in the sentinel Gammarus fossarum by targeted Scout-MRM assay* [86], l'application de cette stratégie pour la mise en place de la méthode Scout-MRM aurait été limitée à 2 ou 3 analyses au lieu de 27.

## **4. Conclusion générale du chapitre**

Dans ce chapitre, nous avons démontré que la quantification hautement multiplexée de protéines est désormais possible grâce au dosage de 157 protéines en simultanément. Ce dosage hautement multiplexé de protéines est à ce jour le plus grand proposé dans le domaine de l'écotoxicologie. Le mode d'acquisition Scout-MRM est un outil performant non seulement pour des études de biosurveillance environnementales, mais également pour le transfert rapide de méthodes de la protéomique de découverte vers la protéomique quantitative. En effet, il offre d'excellentes performances analytiques en termes de robustesse, de répétabilité, de transférabilité et de multiplexage. Grâce à l'approche Scout-MRM, il est envisageable de suivre l'état de santé d'organismes non modèles en dosant simultanément plusieurs centaines de protéines. La protéomique environnementale peut être ainsi envisagée dans le futur comme un outil supplémentaire de biosurveillance.

L'ajout de peptides scouts supplémentaires donne la possibilité de construire des multiplexes de plus en plus grands en implémentant de nouvelles transitions MRM si l'on souhaite étudier de nouvelles protéines d'intérêts. En effet, comme le génome de *G. fossarum* n'est pas encore complètement caractérisé et que les connaissances sur cet organisme progressent de jour en jour, de nouvelles protéines pourront être ajoutées dans la méthode.

Grâce à la méthodologie Scout-MRM nous avons repoussé les limites des capacités de multiplexage en MRM. Cela ouvre de nouvelles perspectives de recherche comme la possibilité de faire de la « découverte ciblée » à partir de grands jeux de données *in silico* et de la construction de grandes listes de transitions MRM. Tout en conservant l'idée de proposer de nouveaux outils au service de la biologie des systèmes, nous avons souhaité explorer cette approche dans le cadre de l'analyse de lipides. Bien

que les transitions MRM peuvent être déterminées à partir des bases de données et que des règles de fragmentation sont disponibles, la connaissance des temps de rétention est toujours un prérequis indispensable pour la construction des méthodes. Il est nécessaire de connaître le temps de rétention d'un composé par rapport à ceux des scouts pour savoir dans quel groupe placer les transitions MRM issues de la base de données. Nous allons montrer dans le chapitre suivant que la lipidomique se prête particulièrement bien à la possibilité de réaliser de la découverte ciblée avec le Scout-MRM. En effet, de nombreuses approches existent pour prédire les temps de rétention des lipides (*ex.* modèles des nombres de carbones équivalents [272]) et pour aider à la construction des groupes. De plus, il existe à ce jour peu de méthodes MRM capables d'analyser et quantifier précisément de nombreux lipides. C'est pour cela que le 3<sup>ème</sup> chapitre de ce manuscrit traitera du développement d'une méthode d'analyse ciblée unique pour un profilage rapide de lipidomes par RPLC-MS/MS utilisant le Scout-MRM.



## **Chapitre 3 : Profilage rapide de lipidome par Scout-MRM et modélisation des temps de rétention**

### **1. Contexte de l'étude**

La caractérisation exhaustive du lipidome chez des organismes non modèles est le plus souvent réalisée en approche globale par spectrométrie de masse haute résolution (HRMS). À l'inverse, l'approche par spectrométrie de masse ciblée est la technique de choix pour une quantification précise en raison d'une meilleure sensibilité et reproductibilité [29]. Afin de maintenir un rapport signal/bruit et un nombre de points par pic chromatographique acceptable, le temps de cycle de l'appareil impose le suivi d'un nombre limité de composés. Comme démontré dans le chapitre précédent, l'utilisation du mode d'acquisition Scout-MRM a permis de repousser les limites en capacité de multiplexage [86,271]. Ainsi, il est possible de réaliser des expériences de découverte en mode ciblé. Cela implique la construction de méthodes avec de grandes listes de transitions MRM issues de bases de données *in silico* qui seront recherchées au sein d'échantillons. Le gain en capacité de multiplexage de l'approche Scout-MRM par rapport à la MRM classique permet de réduire drastiquement le nombre de méthodes nécessaires pour l'identification de composés potentiellement présents dans la matrice à caractériser. Nous avons donc souhaité explorer tout le potentiel de cette approche en lipidomique. Lors de la construction d'une méthode Scout-MRM, il est nécessaire de connaître l'ordre d'élution des composés à analyser par rapport aux composés scouts responsables du déclenchement des groupes de transitions MRM. Cependant, la détermination de l'ordre d'élution des lipides nécessiterait l'utilisation de standards lipidiques de référence pour chaque composé. Ceci n'est pas envisageable au vu du grand nombre de lipides pouvant être présents au sein d'une matrice biologique et du nombre limité de standards disponibles. Il est donc primordial d'avoir des outils de prédiction pour déterminer l'ordre d'élution des composés tout au long du gradient chromatographique. On retrouve dans la littérature, en lipidomique, différents outils permettant de prédire les temps de rétention comme l'utilisation des coefficients de partage [273,274] ou de modèles basés sur le nombre de carbones équivalents [272,275,276].

Dans le cadre de ce 3<sup>ème</sup> chapitre de thèse, nous avons développé des modèles de prédiction à partir du modèle des nombres de carbones équivalents (ECN) pour concevoir une base de données qui référence les ordres d'élution de centaines de lipides. Cette base de données pourra ainsi être utilisée dans l'élaboration de méthodes Scout-MRM lors d'études de caractérisation exhaustive en mode ciblé. Pour que l'utilisation de cette base de données soit pertinente, il faudra démontrer qu'elle est

compatible avec des conditions chromatographiques très diverses. En effet, il n'est pas concevable de développer une nouvelle base de données en ordre d'éluion dès lors qu'une variation dans le processus analytique est observée (changement de matrices, de phases mobiles, de gradient ou de colonne chromatographique, *etc.*). C'est pour cela que la méthode Scout-MRM s'avère être l'outil de prédilection pour rendre la base de données compatible avec les variations analytiques susmentionnées. Nous avons donc évalué les changements de sélectivité rencontrés lors d'une analyse lipidomique en faisant varier la nature de la phase stationnaire tout en conservant la même méthode Scout-MRM. Les lipides sont classés en 8 catégories présentant une grande diversité structurale. On dénombre 25220 lipides annotés dans la base de données Lipid Maps [277]. Cela impacte considérablement les capacités de multiplexage d'une méthode ciblée. C'est pour cela que nous avons défini des règles de sélection lors du choix des transitions MRM à ajouter dans la méthode Scout-MRM. Pour finir, nous avons évalué les performances analytiques de la méthode Scout-MRM (répétabilité, limite de quantification et domaine de linéarité) pour prouver que la méthodologie que nous avons mise en place permet de réaliser de façon simultanée la caractérisation et la quantification des lipides détectés.

## 2. Matériels et méthodes

### 2.1. Réactifs et produits chimiques

L'eau, l'acétonitrile (ACN), le méthanol (MeOH), l'isopropanol (IPA) ont été obtenus auprès de Fisher Scientific (qualité LC-MS, Strasbourg, France). L'EDTA, le triton X-100, le formiate d'ammonium, le dichlorométhane (DCM), le tert-butyl méthyl éther (MTBE), l'acide formique et le chlorure de sodium ont été achetés auprès de Sigma Aldrich (St Quentin-Fallavier, France). Les standards internes lipidiques (PC 11:0/11:0, LPC 13:0, PE 12:0/13:0, PG 12:0/13:0, PS 12:0/13:0, PI 17:0/14:1, DG 17:0/17:0 d5, TG 17:0/17:1/17:0 d5, Cer d18:1/12:0 et SM d18:1/18:1 d9) ont été obtenus auprès de Sigma Aldrich (St Quentin-Fallavier, France, pureté > 99%). Le mélange de standards SPLASH LIPIDOMIX® mass spec a été acheté auprès de Sigma Aldrich (St Quentin-Fallavier, France).

### 2.2. Préparation des composés scouts

Pour le développement de la méthode Scout-MRM, des standards lipidiques synthétiques achetés auprès de Avanti Polar Lipids Inc. (Alabaster, AL) ont été utilisés comme composés scouts. Ils sont au nombre de 9 lipides scouts en mode ion positif et 5 lipides scouts en mode ion négatif (*cf.* tableau n°3). 10 µL du mélange scout ont été ajoutés à l'échantillon ( $V_{\text{total}} = 100 \mu\text{L}$ ) juste avant l'injection dans le système LC-MS/MS.

Mode d'ionisation	Numéro des scouts	Lipide	Référence (Avanti)	Concentration finale dans l'échantillon injecté ( $\mu\text{g}/\text{mL}$ )
Positif	Scout-1	PC 9:0/9:0	850320P-25MG	0.012
	Scout-2	PC 12:0/13:0	LM1000-1EA	0.019
	Scout-3	PC 14:0/14:0 d4	860341P-10MG	0.015
	Scout-4	PC 15:0/15:0	850350P-25MG	0.012
	Scout-5	Cer d18:1/17:0	860517P-5MG	0.05
	Scout-6	PC 19:0/19:0	850367C-25MG	0.023
	Scout-7	TG 14:0/16:1/14:0 d5	860900C-1mg	0.019
	Scout-8	TG 15:0/18:1/15:0 d5	860901P-1mg	0.017
	Scout-9	TG 16:0/18:0/16:0 d5	860902P-1mg	0.06
Négatif	Scout-1	PC 9:0/9:0	850320P-25MG	0.15
	Scout-2	PC 12:0/13:0	LM1000-1EA	0.15
	Scout-3	PC 14:0/14:0 d4	860341P-10MG	0.15
	Scout-4	PC 15:0/15:0	850350P-25MG	0.15
	Scout-5	Cer d18:1/17:0	860517P-5MG	0.15

Tableau 3 : Liste des lipides scouts utilisés dans la méthode.

### 2.3. Préparation des échantillons

**Extraction des lipides du plasma humain :** Des échantillons de plasma humain provenant de volontaires ont été collectés par aphérèse auprès de l'Établissement Français du Sang (EFS). Les échantillons de plasma sont conservés à  $-80^{\circ}\text{C}$  avant utilisation. Pour la préparation de ce type d'échantillon, 20  $\mu\text{L}$  de plasma sont dilués avec 280  $\mu\text{L}$  d'eau ( $4^{\circ}\text{C}$ ) puis dopés avec 10  $\mu\text{L}$  de standards internes. 360  $\mu\text{L}$  de méthanol ( $4^{\circ}\text{C}$ ) sont ajoutés à l'échantillon de plasma dilué et agités pendant 60 min à  $4^{\circ}\text{C}$ . Ensuite, 1200  $\mu\text{L}$  de MTBE ( $4^{\circ}\text{C}$ ) sont ajoutés et le mélange est agité pendant 10 minutes à  $4^{\circ}\text{C}$  avant d'être séparé par centrifugation (10000 g, 5 minutes,  $4^{\circ}\text{C}$ ). Après centrifugation, 1240  $\mu\text{L}$  de la phase organique supérieure sont recueillis. Les extraits lipidiques sont séchés sous un flux d'azote pour ensuite être reconstitués dans 20  $\mu\text{L}$  de DCM/MeOH (1:1, v/v) et 180  $\mu\text{L}$  d'IPA/ACN/ $\text{H}_2\text{O}$  (2:1:1, v/v/v). L'échantillon est dilué 5 fois pour l'analyse en mode ESI (-) et 50 fois pour l'analyse ESI (+) avec le mélange de solvants d'injection.

**Extraction des lipides de *Gammarus fossarum* :** La collecte et l'entretien des organismes de *G. fossarum* avant l'extraction des lipides sont effectués selon la méthode rapportée précédemment [50,237]. Pour la préparation de ce type d'échantillon, un gammaré adulte entier est homogénéisé dans 600  $\mu\text{L}$  de tampon de lyse (Tris-base 50 mM, EDTA 1 mM, Triton X-100 0.1% v/v, NaCl 100 mM et pH ajusté à 7.8) avec une bille en acier inoxydable (diamètre 4 mm). Ensuite, 360  $\mu\text{L}$  de méthanol ( $4^{\circ}\text{C}$ ) et 10  $\mu\text{L}$  de mélange de standards internes sont ajoutés à 300  $\mu\text{L}$  d'homogénat de gammars. Après agitation pendant 60 min à  $4^{\circ}\text{C}$ , 1200  $\mu\text{L}$  de MTBE ( $4^{\circ}\text{C}$ ) sont ajoutés. Le mélange est agité pendant 10 minutes à  $4^{\circ}\text{C}$  avant d'être séparé par centrifugation à 10000 g pendant 5 minutes à  $4^{\circ}\text{C}$ . Après centrifugation, 1240  $\mu\text{L}$  de la phase organique supérieure sont recueillis. Les extraits lipidiques sont séchés sous un flux d'azote puis reconstitués dans 20  $\mu\text{L}$  de DCM/MeOH (1:1, v/v) et 180  $\mu\text{L}$

d'IPA/ACN/H<sub>2</sub>O (2:1:1, v/v/v). L'échantillon est dilué 10 fois pour l'analyse en mode ESI (-) et 100 fois pour l'analyse en mode ESI (+) avec le mélange de solvants d'injection.

**Extraction des lipides de la drosophile :** Les souches de *Drosophila melanogaster* Oregon-RS (#4269) (Bloomington Stock Center, Indiana, IN) ont été élevées à 25°C avec un cycle lumière/obscurité de 12/12 h sur un milieu à base de farine de maïs et de levure. Les mouches âgées de 5 jours ont été collectées, congelées dans de l'azote liquide et stockées à -80°C avant utilisation. Pour la préparation de ce type d'échantillon, les drosophiles (n=3) sont homogénéisées dans 450 µL de tampon de lyse comme décrit ci-dessus, avec 30 billes de verre (diamètre 0.5 mm). 360 µL de méthanol (4°C) et 10 µL de mélange de standards internes sont ajoutés à 300 µL d'homogénat de mouche. Après agitation pendant 60 min à 4°C, 1200 µL de MTBE (4°C) sont ajoutés. Le mélange est agité pendant 10 minutes à 4°C avant d'être séparé par centrifugation à 10000 g pendant 5 minutes à 4°C. Après centrifugation, 1240 µL de la phase organique supérieure sont recueillis. Les extraits lipidiques sont séchés sous un flux d'azote puis reconstitués dans 20 µL de DCM/MeOH (1:1, v/v) et 180 µL d'IPA/ACN/H<sub>2</sub>O (2:1:1, v/v/v). L'échantillon est dilué 5 fois pour l'analyse en mode ESI (-) et 20 fois pour l'analyse en mode ESI (+) avec le mélange de solvants d'injection.

## 2.4. Analyse lipidomique par LC-MS/MS

L'analyse LC-MS/MS a été réalisée sur un spectromètre de masse hybride triple quadripôle/piège à ions linéaire du système QTRAP® 5500 LC-MS/MS (Sciex, Foster City, CA, USA) équipé d'une source d'ions Turbo V™ couplée à un dispositif HPLC de la série 1290 (Agilent Technologies, Waldbronn, Allemagne). La séparation LC des lipides a été réalisée sur une colonne de phase inverse Xselect CSH C18 (100 mm × 2.1 mm, taille des particules 3.5 µm) avec un volume d'injection de 5 µL. L'élution a été réalisée à un débit de 500 µL/min avec H<sub>2</sub>O/ACN (60:40, v/v) contenant 0.1 mM de formiate d'ammonium et 0,1 % d'acide formique comme éluant A et IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) contenant 10 mM de formiate d'ammonium et 0.1 % d'acide formique en tant qu'éluant B. Un gradient linéaire a été utilisé de 1 % à 99% de B en 24 min. La colonne a été rincée avec 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 min. Compte tenu du temps d'équilibrage de la colonne, le cycle d'injection était de 30 min.

Pour la détection par spectrométrie de masse, la tension de pulvérisation d'ions a été fixée à 5500 V pour l'ionisation positive et à -4500 V pour l'ionisation négative. Le nébuliseur et le débit de gaz rideau ont été réglés respectivement à 25 et 50 psi en utilisant de l'azote. La source d'ions TurboV™ a été réglée à 450 °C avec le débit de gaz auxiliaire (azote) réglé à 40 psi. L'acquisition des données est réalisée avec le mode Scout-MRM. Les transitions MRM ainsi que les valeurs du potentiel de déclusterisation (DP) et de l'énergie de collision (CE) ont été optimisées par infusion directe dans le

spectromètre de masse à partir du SPLASH® LIPIDOMIX® qui contient un standard par famille de lipides (PC, PE PS, PG, PI, LPC, LPE, LPS, DG, TG, SM). Les paramètres de masse des céramides ont été optimisés avec le standard Cer d18:1/12:0. Le temps de cycle est fixé à 1.6 seconde afin d'atteindre dix points par pic chromatographique pour chaque transition MRM.

## 2.5. Comparaison des modes Scheduled-MRM versus Scout-MRM

### 2.5.1. Variation du gradient chromatographique

**Condition 1** : Les lipides ont été séparés à l'aide d'une colonne Xselect CSH C18 (100 mm × 2.1 mm, taille des particules 3.5 µm). Les solvants de la phase mobile sont constitués de formiate d'ammonium à 10 mM et d'acide formique à 0.1 % dans H<sub>2</sub>O/ACN (60:40, v/v) (A) et de formiate d'ammonium à 10 mM et d'acide formique à 0,1 % dans IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) (B). Un gradient linéaire a été utilisé de 1 à 99% de B pendant 24 min. La colonne a été rincée à 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 min. Le débit a été réglé à 500 µL/min.

**Condition 2** : Les lipides ont été séparés à l'aide d'une colonne Xselect CSH C18 (100 mm × 2.1 mm, taille des particules 3.5 µm). Les solvants de la phase mobile sont constitués de formiate d'ammonium à 10 mM et d'acide formique à 0,1% dans H<sub>2</sub>O/ACN (60:40, v/v) (A) et de formiate d'ammonium à 10 mM et d'acide formique à 0,1% dans IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) (B). **Une séparation isocratique a été appliquée pendant 1 min à 1% de B** et un gradient linéaire a été réalisé de 1 à 99% de B en 24 min. La colonne a été rincée à 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 min. Le débit a été réglé à 500 µL/min.

### 2.5.2. Modification de la composition de la phase mobile B

**Condition 1** : Les lipides ont été séparés à l'aide d'une colonne Xselect CSH C18 (100 mm × 2.1 mm, taille des particules 3.5 µm). Les solvants de la phase mobile sont constitués de formiate d'ammonium à 10 mM et d'acide formique à 0.1 % dans H<sub>2</sub>O/ACN (**60:40**, v/v) (A) et de formiate d'ammonium à 10 mM et d'acide formique à 0,1 % dans IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) (B). Un gradient linéaire a été utilisé de 1 à 99% de B pendant 24 min. La colonne a été rincée à 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 min. Le débit a été réglé à 500 µL/min.

**Condition 2** : Les lipides ont été séparés à l'aide d'une colonne Xselect CSH C18 (100 mm × 2.1 mm, taille des particules 3.5 µm). Les solvants de la phase mobile sont constitués de formiate d'ammonium à 10 mM et d'acide formique à 0,1 % dans H<sub>2</sub>O/ACN (**40:60**, v/v) (A) et de formiate d'ammonium 10 mM et d'acide formique 0,1 % dans IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) (B). Un gradient linéaire a été utilisé de 1 à 99% de B pendant 24 min. La colonne a été rincée à 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 min. Le débit a été réglé à 500 µL/min.

### 2.5.3. Étude du *dwell time* et du nombre de points par pic chromatographique en fonction de la répétabilité de mesure

L'évaluation de la variation du *dwell time* et du nombre de points par pic chromatographique en fonction de la répétabilité de mesure a été réalisée sur un mélange de standards lipidiques. Les concentrations de ces standards ont été fixées de manière à ce que les transitions MRM représentent une large gamme de S/B représentative d'un échantillon réel ( $4.8 < S/B < 466$ ). La liste des standards lipidiques et les concentrations utilisées sont indiquées dans le tableau n°4.

Six conditions ont été étudiées (n=3) : *dwell time* = 5 ms et nombre de points / pic = 10 ; *dwell time* = 5 ms et nombre de points / pic = 8 ; *dwell time* = 3 ms et nombre de points / pic = 10 ; *dwell time* = 3 ms et nombre de points / pic = 8 ; *dwell time* = 1 ms et nombre de points / pic = 10 et *dwell time* = 1 ms et nombre de points / pic = 8.

Standards lipidiques	Concentration (µg/ml)	Standards lipidiques	Concentration (µg/ml)
LPC 13:0	0,005	PC 9:0/9:0	0,0025
TG 17:0/17:1/17:0 d5	0,0025	PC 19:0/19:0	0,0025
LPS 13:0	0,01	TG 14:0/16:1/14:0 d5	0,0025
SM d18 :1/18 :1 d9	0,0025	PC 12:0/13:0	0,0025
LPE 17:1	0,01	PC 15:0/15:0	0,0025
PE 12:0/13:0	0,005	TG 15:0/18:1/15:0 d5	0,0025
PI 17:0/14:1	0,01	TG 16:0/18:0/16:0 d5	0,005
PS 12:0/13:0	0,005	PC 14:0/14:0 d4	0,0025
Cer d18:1/12:0	0,0025	Cer (35:1)	0,0025
PC 11:0/11:0	0,0025		

Tableau 4 : Liste et concentrations des standards lipidiques utilisés pour l'étude du nombre de points par pic et du *dwell time* en fonction de la répétabilité de mesure en Scout-MRM.

## 2.6. Balayage de perte de neutre pour l'analyse des triacylglycérols

L'évaluation des chaînes d'acides gras majoritaires au sein des triacylglycérols a été réalisée sur un extrait lipidique contenant du plasma humain, des drosophiles et du *G. fassarum* (n=3). L'analyse LC-MS/MS a été réalisée sur un spectromètre de masse hybride triple quadripôle/piège à ions linéaire du système QTRAP® 5500 LC-MS/MS (Sciex, Foster City, CA, USA) équipé d'une source d'ions Turbo V™ couplée à un dispositif HPLC de la série 1290 (Agilent Technologies, Waldbronn, Allemagne). Les conditions chromatographiques décrites dans la partie 2.4. ont été utilisées. Pour la détection par spectrométrie de masse, la tension de pulvérisation d'ions a été fixée à 5500 V en ionisation positive. Le nébuliseur et le débit de gaz rideau ont été réglés respectivement à 25 et 50 psi en utilisant de l'azote. La source d'ions TurboV™ a été réglée à 450 °C avec le débit de gaz auxiliaire (azote) réglé à 40 psi. L'acquisition des données est réalisée avec le mode de balayage de perte de neutre. La gamme de masse en Q1 a été fixée de 600 à 1050 Da avec des pertes de masse qui sont décrites dans le tableau n°5. La vitesse de balayage a été fixée à 1000 Da/s, le potentiel de déclusterisation (DP) à 50 eV et

l'énergie de collision (CE) à 36 eV. Les chromatogrammes ont été intégrés de 20 min à 23 min pour l'ensemble des pertes de neutre considérées.

Perte de neutre (Da)	Chaîne d'acide gras	Perte de neutre (Da)	Chaîne d'acide gras
245.24	AG 14:0	295.26	AG 18:3
243.22	AG 14:1	323.25	AG 20:3
273.27	AG 16:0	321.27	AG 20:4
271.25	AG 16:1	319.25	AG 20:5
269.24	AG 16:2	349.3	AG 22:4
301.3	AG 18:0	347.28	AG 22:5
299.28	AG 18:1	345.27	AG 22:6
297.27	AG 18:2		

Tableau 5 : Liste des pertes de neutre lors de l'analyse des triacylglycérols.

## 2.7. Validation de la méthode Scout-MRM

La validation de la méthode Scout-MRM a été réalisée dans des matrices plasma, drosophiles et gammares, en mode d'ionisation positif et négatif. La linéarité, la limite de quantification (LOQ), la répétabilité et l'étude de sélectivité ont été évaluées en se référant aux « *guidelines* » de la Food and Drug administration (FDA) pour la validation des méthodes bioanalytiques.

**Linéarité :** La linéarité a été évaluée pour 11 standards lipidiques dopés dans l'homogénat de gammares à 16 niveaux de concentration (*cf.* annexe n°2) après extraction des lipides. Chaque niveau de concentration a été préparé en triplicat. Une régression linéaire pondérée en  $1/x^2$  ou  $1/x$  a été utilisée. Un niveau de concentration est validé si l'exactitude se situe entre 80 et 120% et si l'écart entre les triplicats est inférieur à 20%. Le coefficient de détermination ( $R^2$ ) mesuré pour chaque courbe d'étalonnage doit être supérieur à 0,99.

**Sélectivité :** L'étude des changements de sélectivité des lipides en RPLC a été réalisée sur 5 colonnes RPLC différentes. La colonne Xselect CSH C18 2.1x100 mm 3.5 $\mu$ m (Waters), la colonne Zorbax eclipse plus C18 2.1x100 mm 3.5  $\mu$ m (Agilent), la colonne Accucore C18 2.1x100 mm 2.6  $\mu$ m (Fisher scientific), la colonne Xselect HSS T3 2.1x150 mm 2.5  $\mu$ m (Waters) et la colonne Xbridge BEH premier 2.1x100 2.5  $\mu$ m (Waters). Cette étude a été réalisée en utilisant les mêmes extraits lipidiques de *G. fassarum*, de drosophiles et de plasma humain analysés avec la même méthode LC-MS/MS en mode Scout-MRM.

**Répétabilité :** La répétabilité de la méthode, comprenant la préparation d'échantillon et la mesure MS, a été évaluée en analysant des échantillons indépendants (n=10) provenant d'un mélange de plasma humain, de drosophiles et de *G. fassarum*. La réponse obtenue pour chaque lipide est corrigée grâce à l'étalon interne ajouté avant l'étape de LLE pour chaque famille de lipides (*cf.* tableau n°6). Le critère de répétabilité fixé dans cette étude est CV < 20%.



Lipide	Référence (Avanti)	Masse ajoutée durant l'étape de LLE (µg)
LPC 13:0	855476C	3.08
TG 17:0/17:1/17:0 d5	860903P	1.64
SM d18:1/18:1 d9	860740L	1.11
LPE 17:1	856707C	50
PE 12:0/13:0	LM1100-1EA	8.89
PS 12:0/13:0	LM1300-1EA	13.89
Cer d18:1/12:0	860512P	2.35
DG 17:0/17:0 d5	800854P	233.33
PI 17:0/14:1	LM1504-1EA	28.24
PC 11:0/11:0	850330P	1.11

Tableau 6 : Liste des standards lipidiques utilisés lors de l'étude de répétabilité.

## 2.8. Acquisition et retraitement des données

Le retraitement, l'acquisition des données et le contrôle de l'instrument ont été effectués à l'aide du logiciel Analyst 1.6.3 modifié. Les fichiers « .wiff » ont été traités avec le logiciel MultiQuant 2.1.1. Pour la recherche universitaire, le patch provisoire du logiciel Scout-MRM est disponible sur demande auprès de la société Sciex (contact : [yves.leblanc@sciex.com](mailto:yves.leblanc@sciex.com)).

## 3. Résultats

Les règles de nomenclature utilisées lors de la mise en place de la base de données sont décrites dans la première partie des résultats. En effet, les lipides présentent une grande diversité structurale pouvant générer un nombre important de combinaisons possibles ce qui nécessite une codification stricte.

### 3.1. Nomenclature et profils de fragmentation des lipides dans la base de données associée à la méthode Scout-MRM

#### 3.1.1. Nomenclature des lipides

Les lipides sont des composés qui possèdent une nomenclature codifiée où chaque élément a une signification bien précise en fonction du niveau de précision souhaité. Liebisch *et al.* ont uniformisé les règles d'écritures qui ont été adoptées dans la base de données Lipid Maps [278]. La figure n°55 résume l'ensemble des règles nécessaires pour comprendre la nomenclature utilisée dans ce manuscrit. La structure de chaque famille de lipides est présentée dans l'annexe n°3.

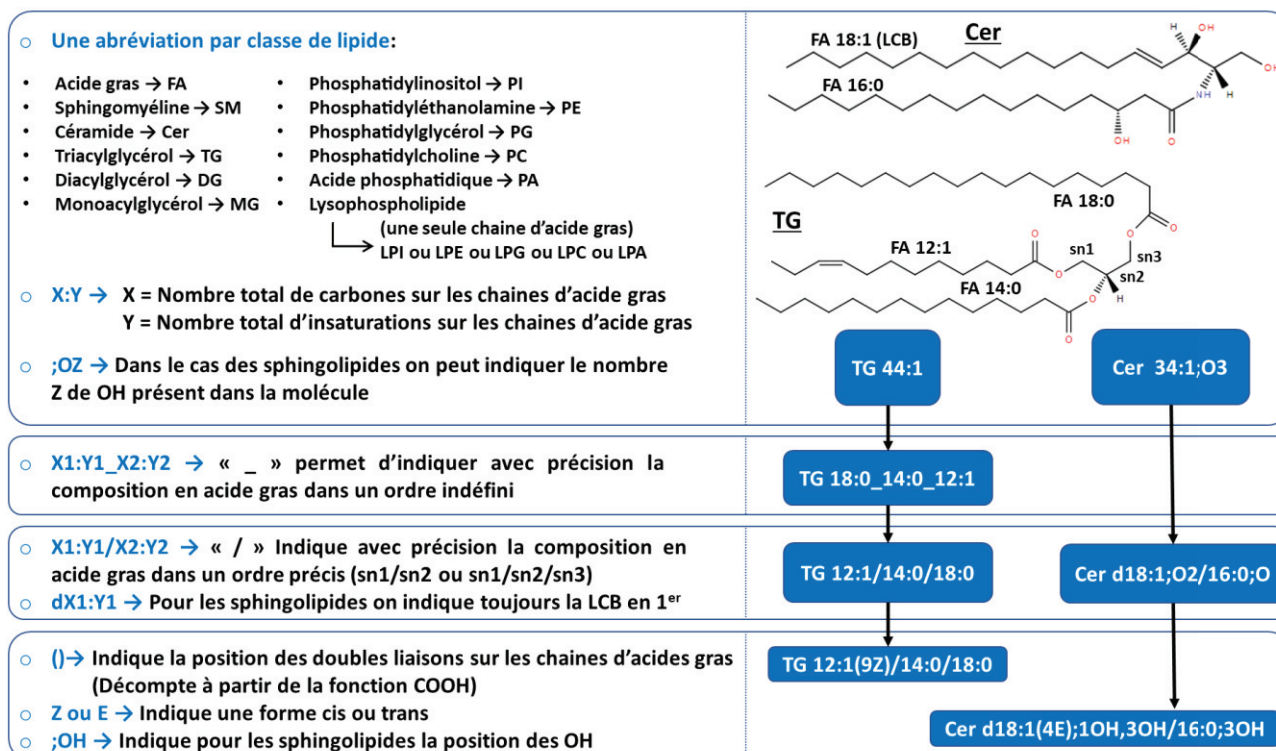


Figure 55 : Règles de nomenclature des lipides en accord avec celles utilisées par la base de données Lipid maps [278]. Deux exemples illustrés à partir d'un triglycéride et d'un céramide permettent de mettre en application cette nomenclature en fonction des différents niveaux de précision. LCB indique la longue chaîne de base de la sphingosine qui a servi à former le céramide et sn est l'abréviation de numérotation stéréospécifique.

Les lipides tels que les glycérolipides, les glycérophospholipides et les sphingolipides sont des composés présentant un très grand nombre de combinaisons structurales. Cette complexité est d'autant plus importante qu'elle résulte de la prise en compte de lipides isomères différant par la position et la composition des chaînes d'acides gras (AG ou FA pour *fatty acid*) ou de la stéréochimie des doubles liaisons (ex. doubles liaisons cis et trans). Les différentes structures de PC 34:1 présentées dans la figure n°56 illustrent cette diversité moléculaire avec 3 combinaisons d'acides gras possibles pour les deux chaînes carbonées.

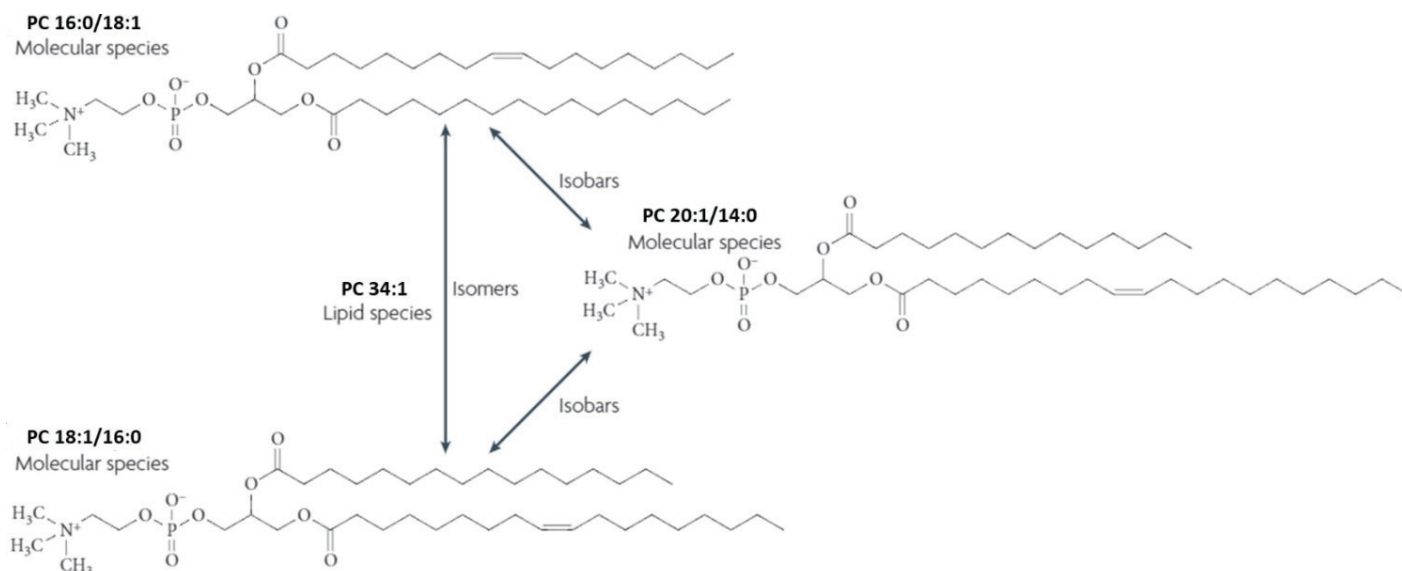


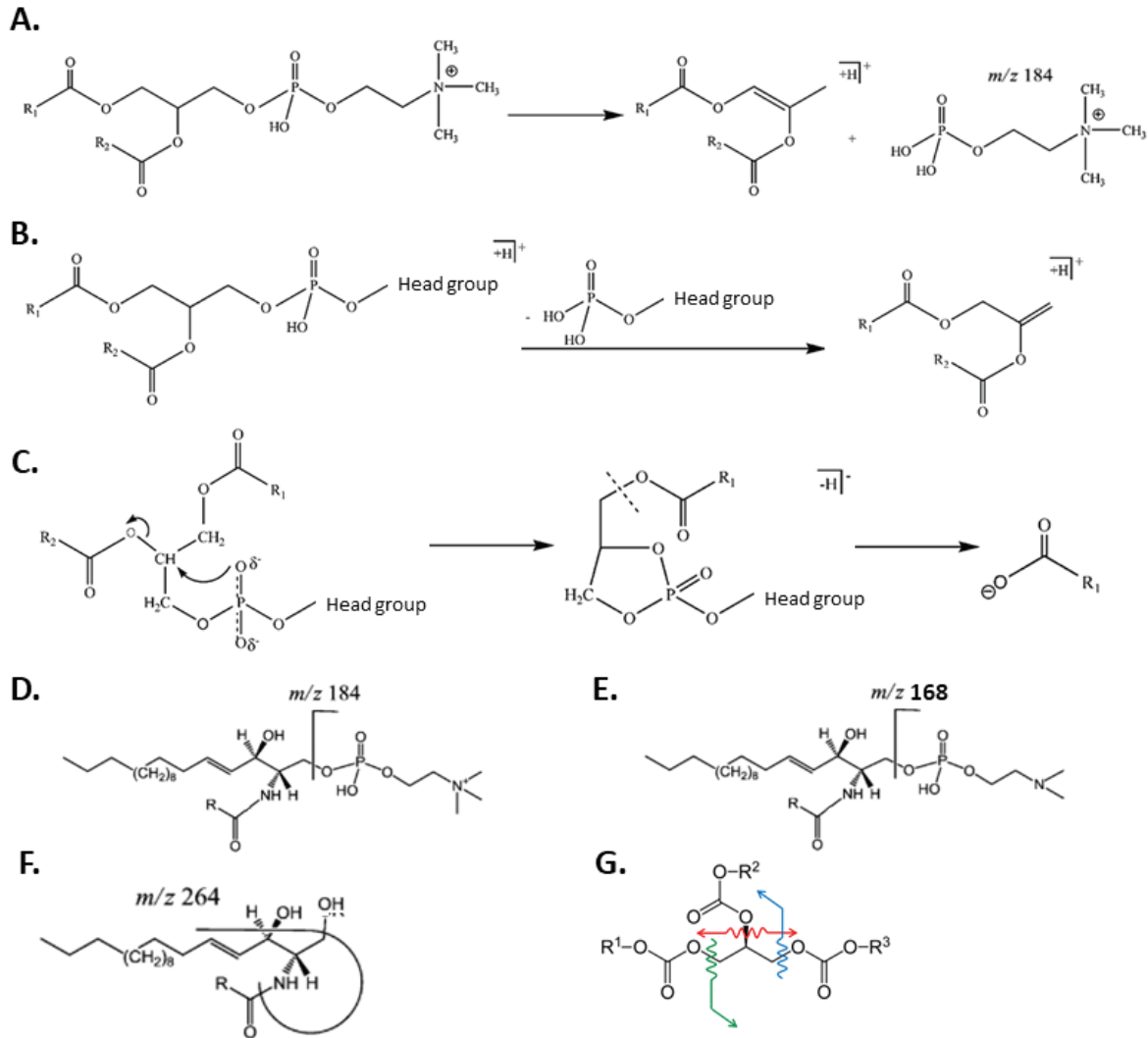
Figure 56 : Diversité moléculaire du phospholipide PC 34:1 [279].

L'analyse directe en infusion par spectrométrie de masse (MS simple) ne permet pas de différencier les molécules ayant le même nombre d'atomes de carbone et de doubles liaisons pour une même famille de lipides. Chacun des signaux obtenus représente un groupe de lipides isobariques avec des AG différents où il n'est pas possible de différencier un PC 16:0/18:1 d'un PC 18:1/16:0 et d'un PC 20:1/14:0 par exemple. L'identification et la quantification d'espèces moléculaires particulières comme les PC 18:1\_16:0 et les PC 20:1\_14:0, nécessitent des analyses par spectrométrie de masse en tandem (MS/MS). Cela permet de différencier les d'acides gras de manière individuels et de déterminer leurs compositions au sein de la structure du lipide.

### 3.1.2. Fragmentation des lipides

L'analyse par spectrométrie de masse en tandem avec l'utilisation d'une cellule de collision induit des fragments permettant la caractérisation structurale des lipides en tant qu'espèce individuelle et non plus par famille. Dans le cas des phospholipides, deux types de fragmentations majoritaires peuvent être détectés en mode ESI (+). La fragmentation des PC contenant une tête polaire choline donne lieu à un fragment caractéristique de cette famille de lipides avec un  $m/z = 184$  correspondant au groupement phosphocholine (*cf.* figure n°57 A). Dans le cas des autres familles de phospholipides (PE, PS, PI et PG), la fragmentation en mode ESI (+) conduit à des fragments majoritaires résultant de la perte de neutre du groupement associé à la tête polaire (PE [M+H-141], PS [M+H-185], PI [M+H-260] et PG [M+H-172]) comme indiqué sur la figure n°57 B. La fragmentation en mode d'ionisation positive de ces espèces permet l'obtention d'informations caractéristiques de la classe lipidique [280,281]. Pour l'ensemble des familles de phospholipides (PC, PE, PS, PI, PG et PA) la fragmentation en mode ESI (-) conduit à la formation de fragments correspondant aux acides gras qui composent le lipide (*cf.* figure n°57 C). La fragmentation en mode d'ionisation négative de ces espèces permet l'obtention

d'informations structurales concernant la composition des chaînes d'acides gras [280,281]. Les variantes lyso de ces phospholipides (une seule chaîne d'acide gras) suivent les mêmes règles de fragmentation.



**Figure 57 : Schémas de fragmentation des lipides.** (A) Fragmentation des PC en ESI (+)[281]. (B) Fragmentation des phospholipides (PE, PI, PS, PG et PA) autre que PC en ESI (+) [281]. (C) Fragmentation des phospholipides (PC, PE, PI, PS, PG et PA) en ESI (-) [281]. (D) Fragmentation des SM en ESI (+) [282]. (E) Fragmentation des SM en ESI (-). (F) Fragmentation des Cer d18:1/X:Y en ESI (+) [282]. (G) Fragmentation des TG en ESI (+).

Les sphingomyélines se fragmentent majoritairement en un fragment de *m/z* = 184 qui correspond au groupement de la tête phosphocholine en mode d'ionisation positive (cf. figure n°57 D) [282]. Les sphingomyélines peuvent être analysées en mode ESI (-), avec la formation de fragments majoritaires associés à la perte d'un groupement CH<sub>3</sub> [M-H-15] ou à l'apparition d'un fragment correspondant à la tête phosphocholine et d'une perte du groupement CH<sub>3</sub> (*m/z* = 168) (cf. figure n°57 E). Que ce soit en mode d'ionisation positif ou négatif, il n'est pas possible de remonter à l'information de la composition de la LCB et de l'AG qui compose les SM. En mode ESI (+), les céramides possèdent un fragment caractéristique de la composition de la longue chaîne de base (LCB) du sphingosine qui a servi à leur

formation (perte de l'acide gras accompagnée de la perte des deux fonctions alcools pour former une base de sphingosine doublement déshydratée ;  $m/z = 264$  pour les Cer d18:1 [282] ; cf. figure n°57 F). Les sphingomyélines et les céramides sont majoritairement analysés en mode d'ionisation positive, car leur sensibilité est beaucoup plus faible en mode d'ionisation négative.

Les triacylglycérols (squelette glycérol avec 3 chaînes d'acides gras) sont des lipides neutres non ionisables en mode d'ionisation positive qui nécessitent la présence d'un contre-ion pour être détectés en mode ESI (+) (ex.  $[M+NH_4^+]$ ). Lors de la fragmentation des TG, il y a une perte de neutre correspondant à l'un des acides gras (cf. figure n°57 G) et du contre-ion ce qui permet d'obtenir un fragment correspondant à un ion diacylglycérols (squelette glycérol avec 2 chaînes d'acides gras). La différence de masse entre l'ion précurseur  $[M+NH_4^+]$  et l'ion produit permet d'identifier la chaîne d'acide gras perdue à partir du nombre total d'atomes de carbone et de doubles liaisons [283]. Cette règle de fragmentation est également valable pour les diacylglycérols qui se fragmentent en ions monoacylglycérols (squelette glycérol avec 1 chaîne d'acide gras) [283]. Les triacylglycérols ne sont pas analysables en ESI (-), car ils ne sont pas ionisables en mode négatif. Les diacylglycérols sont majoritairement analysés en mode d'ionisation positive, car leur sensibilité est beaucoup plus faible en mode d'ionisation négative.

Il est important de noter que les schémas de fragmentation présentés sur la figure n°57 ne permettent pas d'identifier la position de la double liaison sur les chaînes d'acides gras ainsi que la position de chacune des chaînes d'acides gras dans la structure du lipide.

À partir des schémas de fragmentation décrits précédemment il est possible de créer les listes de transitions MRM correspondant aux couples d'ions précurseurs et d'ions produits caractéristiques utilisés lors de l'élaboration de la méthode Scout-MRM. Par exemple pour le PE 18:1/18:1, de masse 743.5 Da, nous avons ajouté la transition MRM du couple 744.6/603.5 en ESI (+) pour le fragment issu de la perte de neutre de la tête polaire  $[M+H-141]$ . En ESI (-), nous avons ajouté le couple 742.5/281.2 qui correspond au fragment de l'AG 18:1 ( $m/z = 281.2$ ).

Les principaux paramètres du spectromètre de masse (potentiel de désolvatation, DP, et de l'énergie de collision, CE) ont été optimisés par infusion directe du mélange SPLASH® LIPIDOMIX® contenant un standard par famille de lipides (PC, PE, PS, PG, PI, PG, PA, LPC, LPE, DG, TG, SM). Les paramètres de masse des céramides ont été optimisés avec le standard Cer 30:1. Ces valeurs optimisées sont indiquées dans l'annexe n°4.

Une fois l'optimisation des transitions MRM réalisée, il a fallu optimiser les conditions chromatographiques pour la mise en place de la méthode Scout-MRM en lipidomique. En effet, pour l'analyse des lipides, il existe de nombreuses interférences isobariques qu'il est essentiel de séparer

par chromatographie. Par exemple, les transitions des PC peuvent interférer sur celle des SM puisqu'ils possèdent le même fragment ( $m/z = 184$ ) et présentent un recouvrement de leurs massifs isotopiques lors de la détection de l'ion précurseur en ESI (+) (cf. 3.2.2).

## 3.2. Optimisation du gradient chromatographique

### 3.2.1. Choix de la phase stationnaire et des phases mobiles

Comme indiqué dans le 1<sup>er</sup> chapitre de ce manuscrit, deux types de chromatographies sont envisageables pour l'analyse des lipides : RPLC et HILIC. Cependant, ces deux types de phases stationnaires ne séparent pas les lipides selon les mêmes critères. En HILIC, les lipides ne possédant pas de tête polaire tels que les DG, les TG et les Cer ne sont pas retenus sur la phase stationnaire à interaction hydrophile et sont élués au temps mort. Ceux ayant une tête polaire sont élués par familles lipidiques. En effet, les têtes polaires ne varient pas au sein d'une même famille et la phase stationnaire HILIC induit peu de séparation selon la longueur des chaînes d'acide gras [58]. Ainsi, dans ce mode de séparation, l'ensemble des lipides issus d'une même classe sont élués à des temps de rétention très proche. Du fait de son manque de résolution entre l'ensemble des lipides appartenant à une même famille de composés, la chromatographie HILIC n'a pas été retenue.

Les phases stationnaires des colonnes C18 ont des capacités de séparation beaucoup plus importantes que la HILIC (cf. chapitre 1 partie 3.2). Ainsi, tous les lipides vont être répartis tout au long du chromatogramme ce qui permet d'analyser plus de lipides lors de la mise en place de méthodes hautement multiplexées en limitant les co-élutions entre lipides. C'est pour cela que la méthode Scout-MRM a été développée en mode de chromatographie de phase inverse. D'autant plus que la RP est la phase stationnaire la plus utilisée pour la séparation des lipides [55,58,284,285].

De nombreuses publications [51,55,123,286] utilisent comme phase mobile A un mélange acétonitrile/H<sub>2</sub>O (60:40 ; v/v) contenant 10 mmol/L de formiate d'ammonium et 0.1% d'acide formique. La phase mobile B contient un mélange Isopropanol/ACN (90:10 ; v/v) avec 10 mmol/L de formiate d'ammonium et 0.1% d'acide formique. Les lipides sont des composés très apolaires. Donc pour avoir des phases mobiles suffisamment éluantes, la phase A contient initialement une forte proportion d'ACN et la phase B de l'isopropanol qui est un solvant plus éluant que l'ACN. Il est nécessaire d'ajouter des sels dans les phases mobiles afin d'ioniser les lipides neutres comme les TG en formant des adduits (ex.  $[M+NH_4^+]$ ). Cependant, cette phase mobile A (H<sub>2</sub>O/ACN ; 40/60 ; v/v) ne permet pas d'obtenir une séparation optimale pour les lysophospholipides dont font partie les LPC qui éluent en début de chromatogramme. Ces conditions ne différencient pas les deux isomères concernant la position de l'unique chaîne d'acide gras sur le squelette glycérol (cf. figure n°58 et 59 A) [287,288].

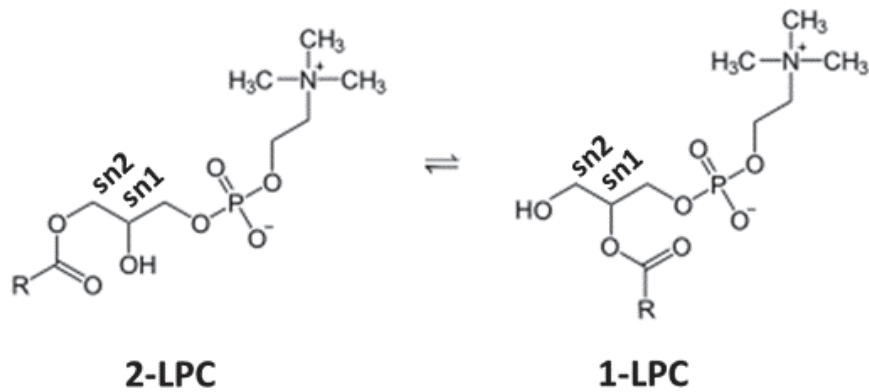


Figure 58 : Différence de structures entre sn-2 et sn-1 pour les LPC.

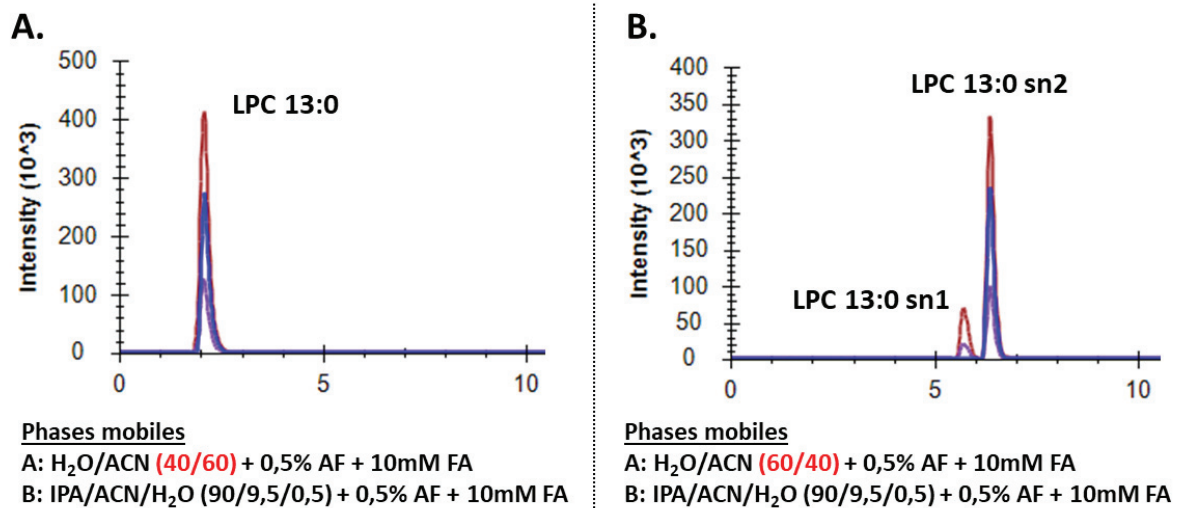


Figure 59 : Séparation des LPC en chromatographie de phase inverse avec deux conditions chromatographiques différentes. Les compositions des phases mobiles sont différentes mais les pentes sont indiquées. sn est l'abréviation de numérotation stéréospécifique.

Afin d'effectuer cette séparation, il a été nécessaire d'inverser les proportions du mélange qui constituent la phase mobile A en utilisant 60% d'eau et 40% d'acétonitrile comme indiqué dans la figure n°59 B. Ainsi, nous observons la séparation entre les LPC dont la chaîne d'acide gras est positionnée en sn1, des LPC dont la chaîne d'acide gras est positionnée en sn2.

### 3.2.2. Validation du gradient chromatographique

Lors de la mise en place d'une méthode de séparation chromatographique en lipidomique, il est nécessaire que la méthode permette de séparer les composés qui ne diffèrent que d'une seule double liaison. En effet, comme la résolution des spectromètres de masse de type triple quadripôles est faible, il y a un risque important d'avoir de nombreuses transitions interférées du fait de recouvrement des massifs isotopiques entre les masses monoisotopiques et les contributions des carbones 13 de ces lipides. Cela peut provoquer des biais de quantification [289] comme représentés dans la figure n°60.



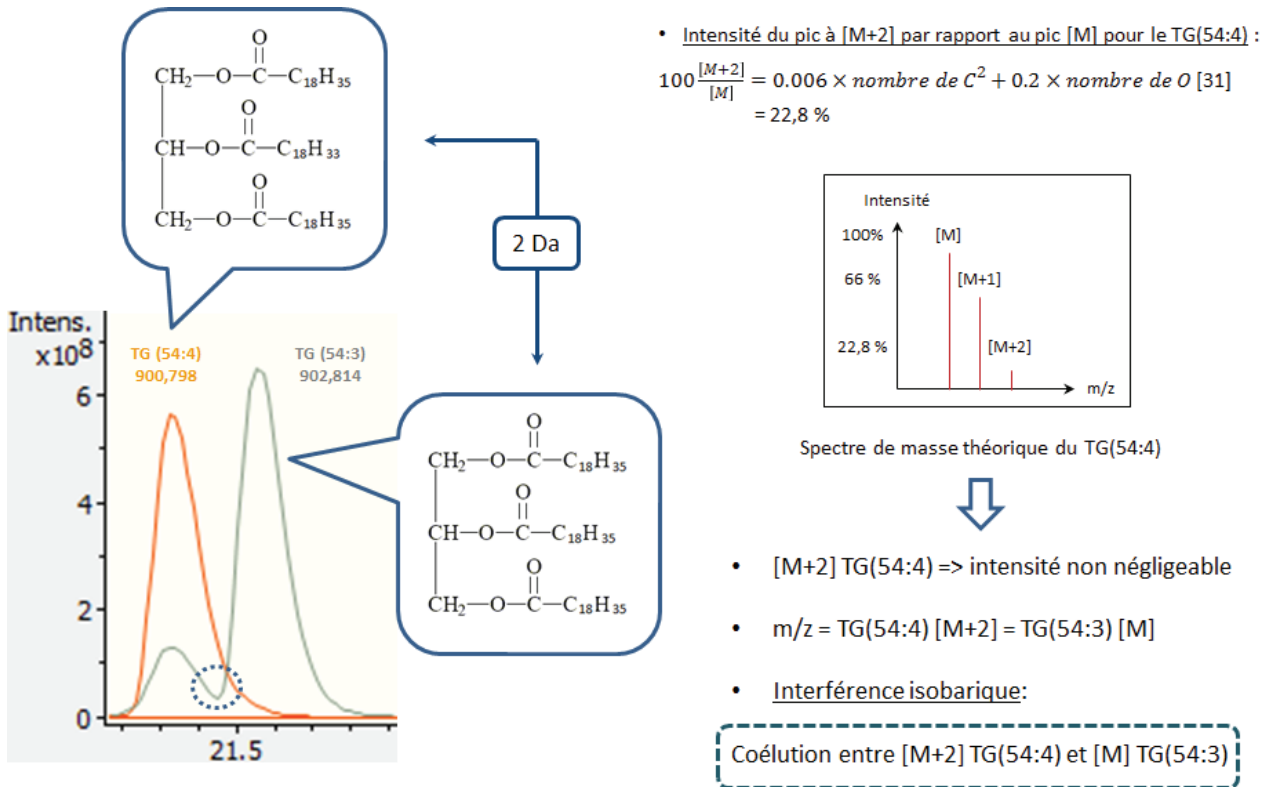


Figure 60 : Problèmes d'interférences isobariques entre deux lipides qui ne diffèrent que d'une insaturation. La contribution isotopique à M+2 du TG 54:4 induit une co-élution avec le pic monoisotopique du TG 54:3 qui est non résolu en spectrométrie de masse de basse résolution.

Par exemple, la transition MRM d'un TG 54:3 peut être co-éluee avec la contribution isotopique de la transition MRM d'un TG 54:4 à M+2, car la masse de ces deux TG ne diffère que de 2 Da (cf. figure n°61). Lorsqu'on utilise un spectromètre de masse dont la résolution en masse est trop faible (dans l'exemple de la figure n°60,  $R \geq 285000$ ) pour lever l'interférence isobarique ou que la résolution chromatographique est insuffisante, une correction isotopique est nécessaire pour limiter les biais de quantification des composés qui ne diffèrent que d'une seule double liaison [289].

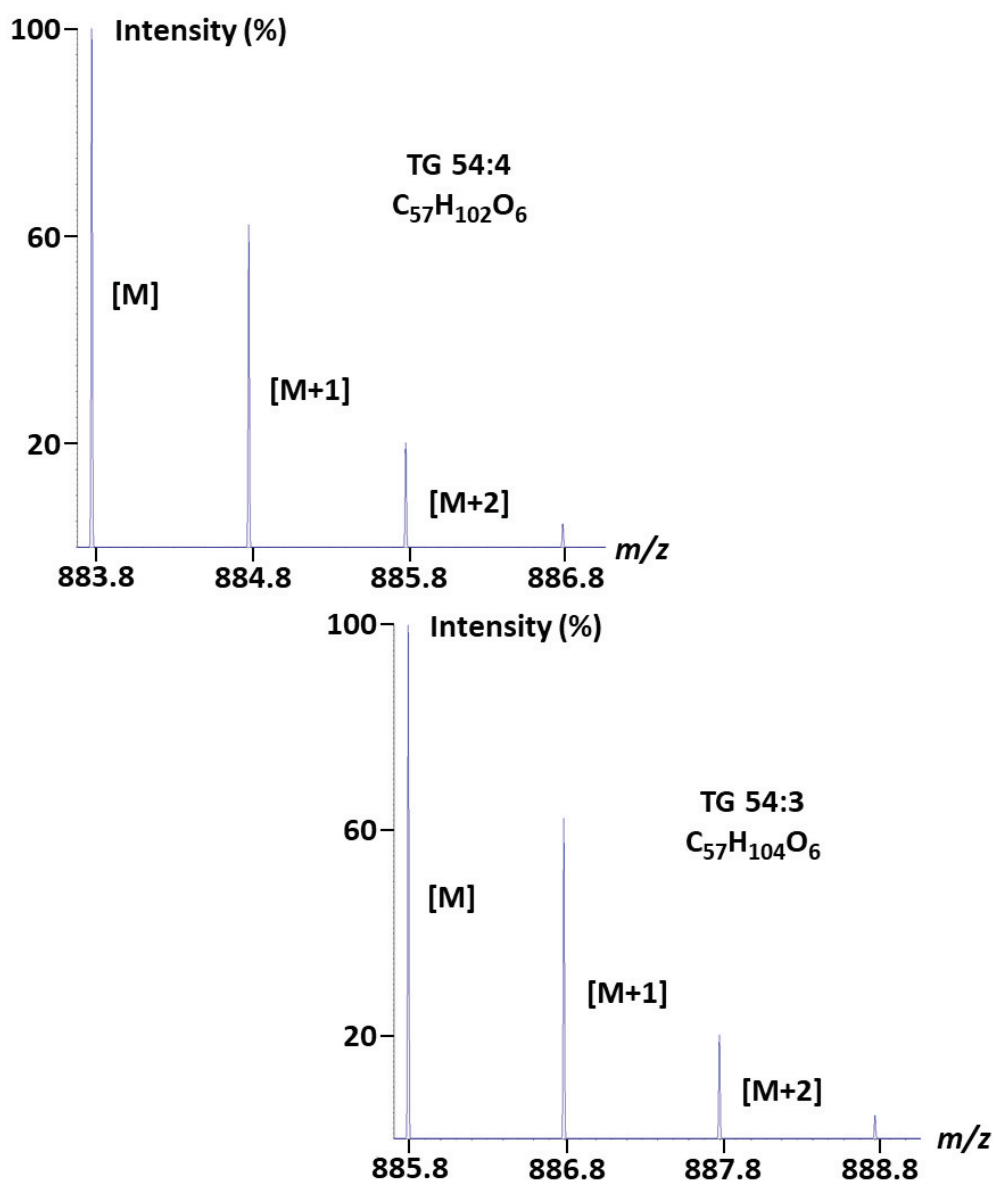


Figure 61 : Recouvrement du massif isotopique du TG 54:3 et du TG 54:4.

Les recouvrements de massifs isotopiques sont observés au sein des composés de même classe de lipides. Toutefois, ils peuvent être également rencontrés entre différentes familles de lipides. Par exemple, l'ensemble des PC et des SM possèdent le même ion précurseur lors de leur fragmentation en ionisation positive (tête phosphocholine ;  $m/z = 184$ ). Dès lors que les masses des précurseurs PC et SM sont proches, des recouvrements isotopiques peuvent survenir et provoquer des interférences isobariques entre le pic monoisotopique du PC X:Y et le pic correspondant à l'abondance isotopique à [M+1] du SM X+4:Y-1. (ex. PC 34:2 → 758.56/184 et SM 38:1 → 759.63/184).

Dans le but d'éviter les recouvrements de massif isotopique, les paramètres chromatographiques (composition des phases mobiles, gradient, débit, etc.) ont été optimisés. Afin de conserver des capacités analytiques convenables en termes de nombre d'échantillons analysés, le temps total de

l'analyse ne devait pas dépasser 30 minutes. Les conditions après optimisation sont référencées dans le tableau n°7 ci-dessous. Ces conditions permettent d'obtenir la séparation des composés lipidiques de l'espèce sentinelle *G. fossarum* comme représentée sur la figure n°62. On peut observer sur ce chromatogramme 3 zones d'élutions. La 1<sup>ère</sup> représente la zone des lysophospholipides, la 2<sup>ème</sup> la zone des phospholipides, sphingomyélines, céramides et diacylglycérols. La 3<sup>ème</sup> zone est celle des triacylglycérols.

<b>Colonne</b>	XSelect CSH C18 2.1x100 mm, 3.5 µm (taille de particule)
<b>Température de colonne, débit et volume d'injection</b>	70°C, 500 µL/min et 5 µl injecté
<b>Phase mobile A</b>	H <sub>2</sub> O/ACN (60/40 ; v/v) + 10 mM formiate d'ammonium + 0.1% acide formique
<b>Phase mobile B</b>	IPA/ACN/H <sub>2</sub> O (90/9.5/0.5 ; v/v/v) + 10 mM formiate d'ammonium + 0.1% acide formique
<b>Solvant d'injection</b>	IPA/ACN/H <sub>2</sub> O (2/1/1 ; v/v/v)
<b>Gradient</b>	

Tableau 7 : Paramètres chromatographiques utilisés pour le développement de la méthode Scout-MRM pour l'analyse lipidique.

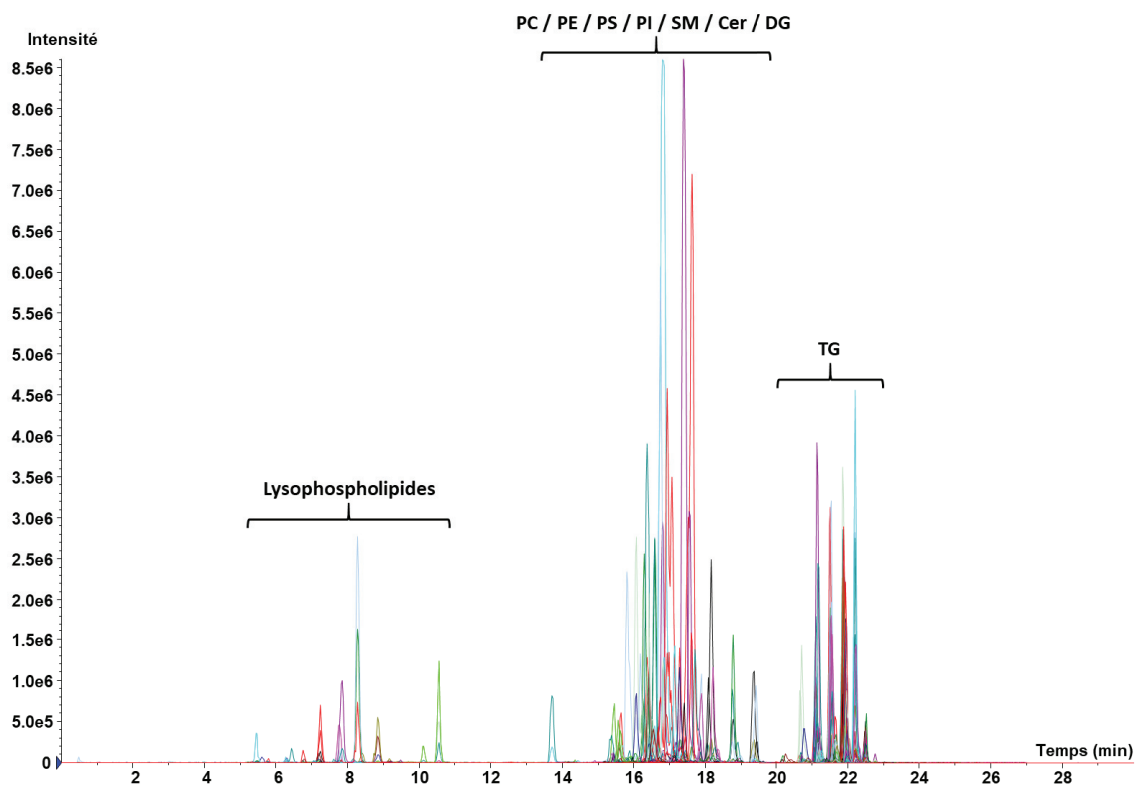


Figure 62 : Séparation chromatographique de lipides de *G. fossarum* selon les conditions chromatographiques décrites dans le tableau n°7.

La zone du chromatogramme la plus critique lors des recouvrements isotopiques est celle des triacylglycérols, car il s'agit des composés présentant le plus de rétention et qui sont les moins bien séparés au sein de la même famille lipidique. Dans notre cas, les conditions chromatographiques (cf. tableau n°7) permettent d'obtenir une résolution suffisante entre deux TG qui ne diffèrent que d'une insaturation puisqu'aucune co-élution n'est observée comme le montre la figure n°63. La séparation chromatographique issue de notre optimisation permet d'éviter les recouvrements des profils isotopiques pour chacune des familles lipidiques comprises dans notre méthode.

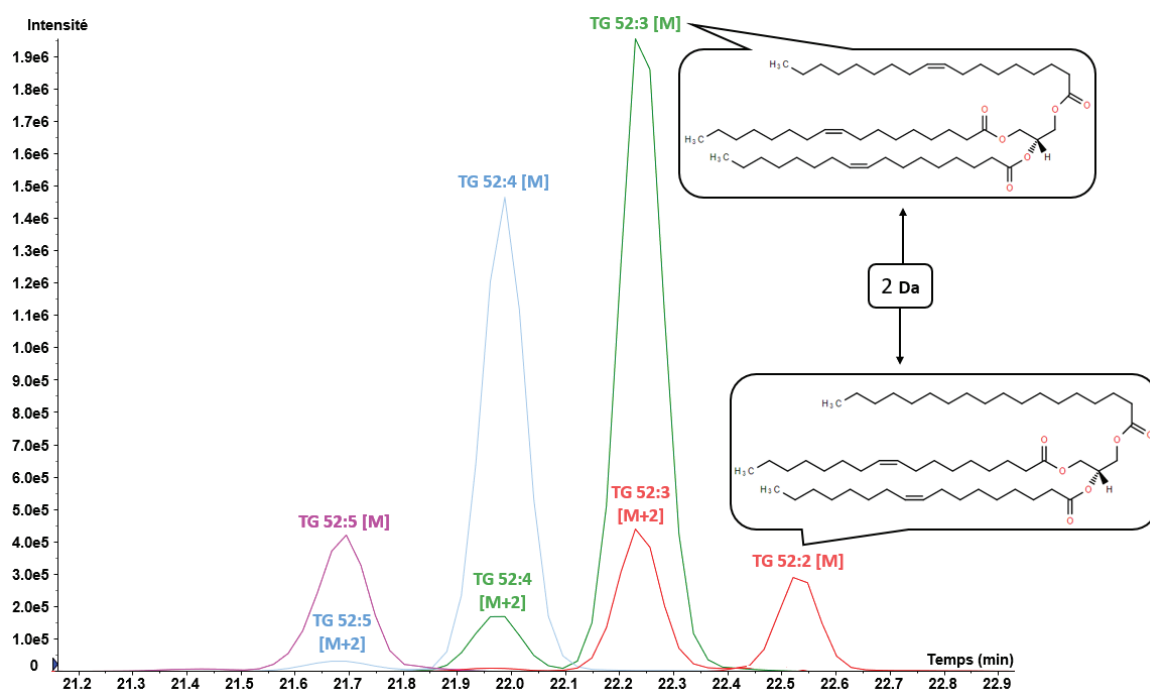


Figure 63 : Exemple de séparation chromatographique de TG de *G. fossarum* selon les conditions chromatographiques décrites dans le tableau n°7.

Le gradient chromatographique décrit dans le tableau n°7 ne possède pas de palier isocratique qui est d'ordinaire utilisé pour corriger de possible volume de délais. En effet, cela n'est plus nécessaire lorsque le mode d'acquisition Scout-MRM est utilisé puisqu'il est indépendant des variations en temps de rétention (cf. chapitre 2 partie 3.5.3.B).

Une fois que les conditions chromatographiques ont été définies et optimisées, nous nous sommes focalisés sur la construction de la méthode Scout-MRM en débutant par le choix des composés scouts.

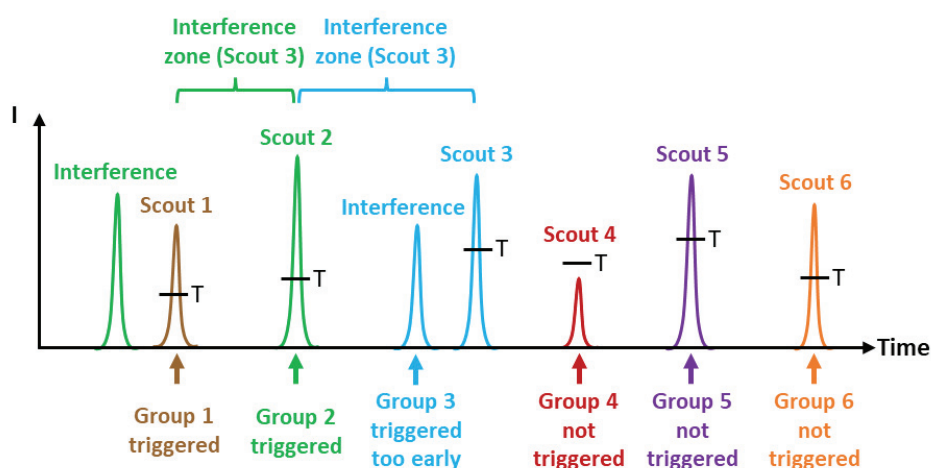
### 3.3. Construction de la méthode Scout-MRM

La première étape de la construction de la méthode Scout-MRM est de sélectionner les composés scouts qui déclenchent les différents groupes. Le choix de ces composés est une étape cruciale dans le développement d'une méthode Scout-MRM. Pour cela, il est nécessaire de respecter un certain nombre de critères afin de développer une méthode robuste avec un très haut degré de multiplexage.

### 3.3.1. Critères de sélection d'un scout idéal

Afin d'avoir la méthode scout-MRM optimale, il est nécessaire que les composés scouts soient répartis tout le long du gradient chromatographique pour maximiser le nombre de groupes scout et donc le nombre maximum de transitions MRM dans la méthode. Pour remplir ce critère, il est conseillé d'utiliser des composés scouts qui ont des structures chimiques de même nature que les composés d'intérêt. Cela assure une comptabilité entre les conditions chromatographiques et les composés scouts (rétention sur la phase stationnaire, pas de précipitation en contact des phases mobiles, etc.).

Pour rappel, un groupe de transitions est déclenché si et seulement si l'ensemble des transitions du scout associé dépasse un seuil fixé par l'expérimentateur au même moment. Cela a pour but d'éviter que des groupes ne se déclenchent de manière prématurée. De plus, un scout « n » ne peut être déclenché que si le scout « n-1 » a déjà été déclenché au préalable. Ainsi, la zone d'interférence durant laquelle un scout « n » peut être déclenché prématurément est fortement réduite (domaine d'interférence → [RT scout « n-1 » à RT scout « n »]). Dans l'exemple de la figure n°64, le scout 2 est déclenché normalement, car son interférent est élué avant le scout 1. Le scout 3 en revanche est déclenché trop tôt, car son interférent est dans sa zone d'interférence (entre le scout 2 et le scout 3). Le déclenchement d'un groupe de manière anticipée a pour conséquence l'arrêt prématuré du groupe « n-1 » pour commencer l'acquisition du groupe « n » plus tôt que prévu. L'autre difficulté qui peut être rencontrée est que si un groupe ne se déclenche pas, les autres qui le suivent ne le seront pas non plus. Dans l'exemple de la figure n°64, les scouts 5 et 6 ne sont pas déclenchés même s'ils dépassent les seuils fixés, car le scout 4 n'a pas rempli les conditions de déclenchement.



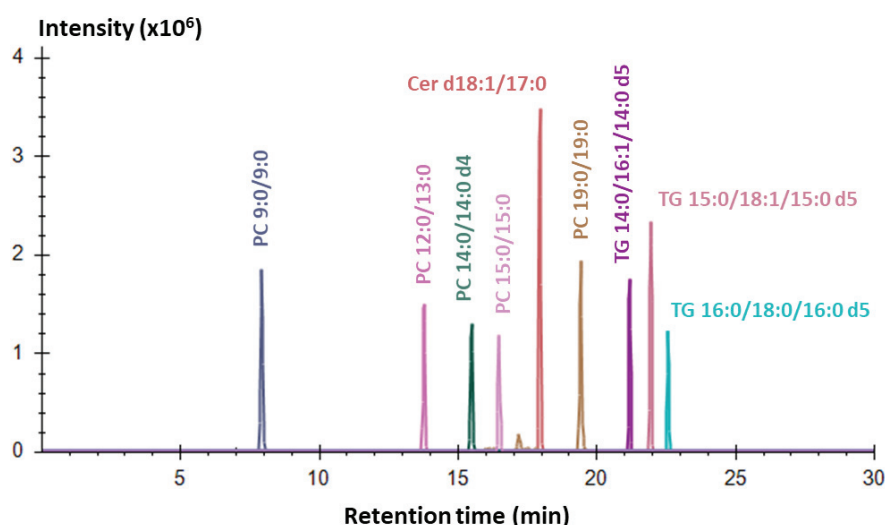
**Figure 64 : Fonctionnement du déclenchement des groupes scouts en fonction de l'intensité des scouts et de la présence d'interférents.** Les pics représentés avec la même couleur sont issus de la même transition MRM. Le scout 1 est bien déclenché, car le signal de ce scout dépasse le seuil fixé. Le scout 2 se déclenche au bon moment, car il dépasse le seuil fixé et que l'interférent issu du scout 2 est élué avant le scout 1. Le scout 3 est déclenché trop tôt, car un interférent présent dans la zone d'interférence du scout 3 dépasse le seuil fixé. Le groupe 4 n'est pas déclenché, car le seuil fixé n'est pas atteint. Les groupes 5 et 6 ne sont pas déclenchés, car le groupe 4 avant eux ne s'est pas déclenché. « T » représente le seuil des scouts (*threshold*).

Pour limiter les problèmes cités ci-dessus, il est préférable de choisir des scouts exogènes afin de les ajouter dans la matrice à une forte concentration et ainsi être certain de dépasser les seuils fixés. Cela permet également de fixer des seuils plus importants diminuant le risque qu'un interférent déclenche un groupe prématurément. Dans le cas où les scouts sont des composés exogènes, il est nécessaire de s'assurer au préalable qu'ils ne soient pas présents dans l'échantillon.

### 3.3.2. Choix des composés scouts en lipidomique

Dans le cas des analyses lipidomiques il n'est pas obligatoire d'avoir une répartition des scouts tout le long du gradient chromatographique (cf. figure n°65) puisque comme indiqué dans la figure n°62, peu de lipides sont élués en début d'analyse. Dans cette portion de chromatogramme ne figurent que les lysophospholipides qui sont moins nombreux que les autres classes de lipides du fait qu'ils n'ont qu'un seul acide gras et donc un nombre de possibilités isomériques plus limité. Le besoin en capacité de multiplexage est donc moins important que pour la zone d'élution des phospholipides (milieu du chromatogramme) et des glycérolipides (fin du chromatogramme).

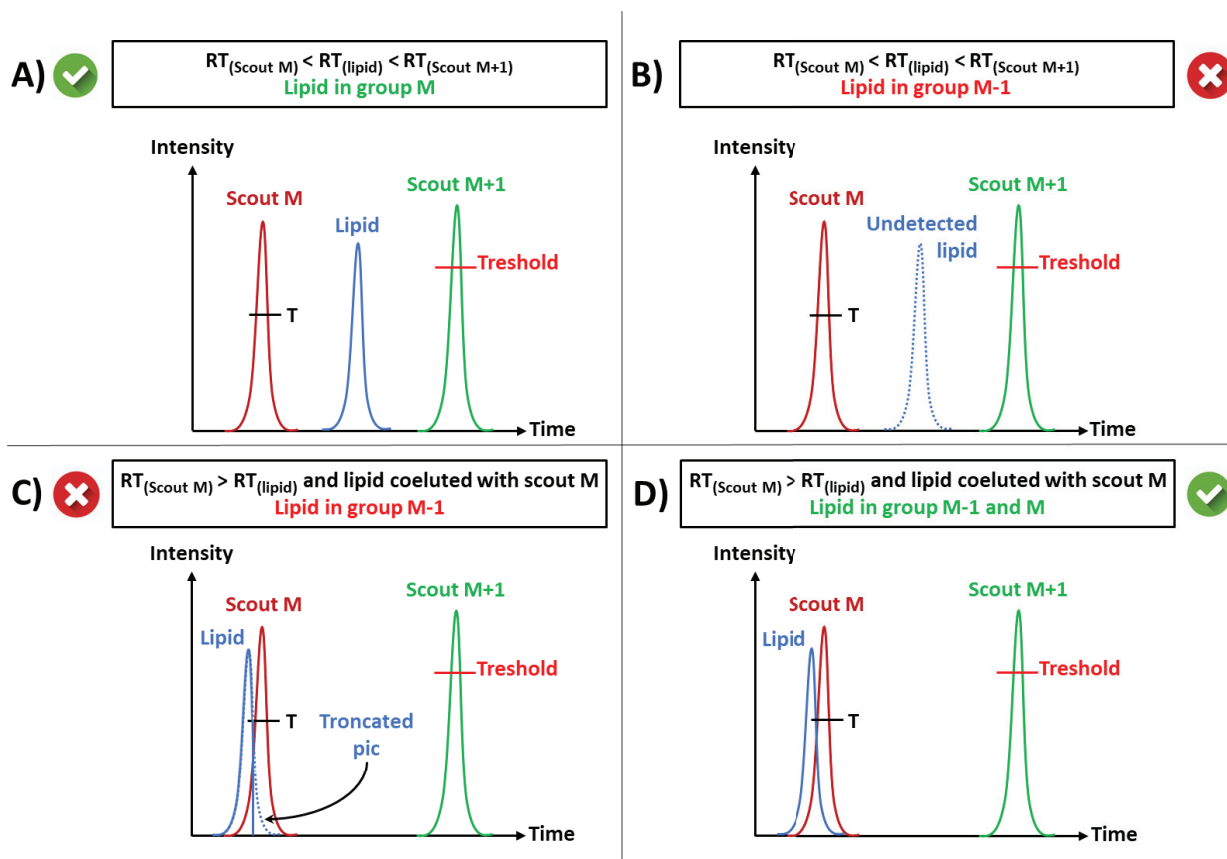
Lors du choix de la nature des composés scouts, nous avons sélectionné des standards de lipides. Afin d'être certains qu'ils ne soient pas présents nativement dans la matrice analysée, nous avons choisi des lipides isotopiquement enrichis (deutérys). Toutefois, tous les lipides isotopiquement enrichis ne sont pas disponibles commercialement. Dans ce cas de figure, les lipides scouts peuvent être choisis parmi les lipides à chaînes d'acides gras impaires, car ces lipides sont très peu présents dans les échantillons biologiques. Les lipides scouts sélectionnés dans notre méthode Scout-MRM en lipidomique sont référencés sur la figure n°65. Deux transitions MRM par composés scouts sont utilisées pour le déclenchement des groupes.



**Figure 65 : Distribution des lipides scouts le long du gradient chromatographique.** Les scouts PC 9:0/9:0, PC 12:0/13:0, PC 14:0/14:0 d4, PC 15:0/15:0 et Cer d18:1/17:0 sont utilisés en ESI (+) et en ESI (-). Les scouts PC 19:0/19:0, TG 14:0/16:1/14:0 d5, TG 15:0/18:1/15:0 d5, TG 16:0/18:0/16:0 d5 ne sont utilisés qu'en ESI (+).

### 3.3.3. Attribution des composés aux différents groupes de scouts

Pour ajouter les transitions MRM dans les différents groupes de scouts, il est nécessaire de connaître l'ordre d'éluion des composés d'intérêts et des composés scouts pour savoir dans quel groupe placer les transitions MRM. Lorsqu'une transition est placée dans le mauvais groupe, cela entraîne la non-détection de ce composé (cf. figure n°66 A et B).



**Figure 66 :** Remplissage des groupes scouts avec les transitions MRM d'intérêts selon la présence ou l'absence d'une co-élution avec un composé scout. (A) Situation où la transition MRM du lipide est bien positionnée. Si  $RT_{\text{scout M+1}} > RT_{\text{lipide}} > RT_{\text{scout M}}$ , il faut placer le lipide dans le groupe M. (B) Situation où la transition MRM du lipide est mal positionnée. Si un lipide est placé dans le mauvais groupe (M-1 au lieu de M), il n'est pas détecté. (C) Situation où le pic chromatographique est tronqué, car la transition MRM du lipide est co-éluee avec celle du scout. (D) Situation où le pic chromatographique n'est pas tronqué, car la transition MRM du lipide est placée dans les 2 groupes scouts (par exemple M-1 et M) lorsqu'elle est co-éluee avec un scout.

Lorsqu'une transition MRM est co-éluee avec un composé scout, il est nécessaire de placer cette transition dans les deux groupes (avant et après le scout co-élue) pour éviter que le pic chromatographique ne soit tronqué (cf. figure n°66 C et D). Ce phénomène de troncature peut avoir deux origines expliquées ci-dessous :

- Cas où le lipide est co-éluee avec le scout et que  $RT_{\text{lipide}} < RT_{\text{scout co-élue}}$  :

Cette troncature a lieu lors de l'arrêt de la détection du groupe auquel le composé d'intérêt appartient et le début de la détection du groupe suivant alors que le pic d'intérêt n'est pas encore revenu à la ligne de base (cf. figure n°66 C).



- Cas où le lipide est co-élué avec le scout et que  $RT_{lipide} > RT_{scout}$  co-élué :

Cette troncation a lieu puisque la détection de la transition MRM du lipide d'intérêt débute alors que le composé a déjà commencé à être élué.

L'ajout d'une même transition dans deux groupes scouts différents a pour conséquence de diminuer les capacités de multiplexage de la méthode Scout-MRM.

### 3.3.4. Quelles sont les limites de la capacité de multiplexage en Scout-MRM ?

Les conditions chromatographiques associées à la méthode Scout-MRM que nous avons développée permettent d'obtenir une largeur de pic moyenne de 16 secondes. Dans le cas où nous souhaitons avoir 10 points par pic chromatographique pour assurer une intégration répétable de l'aire de ce pic, il est nécessaire d'avoir un temps de cycle de 1.6 seconde soit 1600 ms. Le nombre de transitions par groupe scout sera donc dicté par le *dwell time* appliqué sur chaque transition MRM et le temps de repos entre deux transitions MRM. Dans le cas de notre Qtrap 5500, Sciex® conseille de maintenir un *dwell time* et un temps de repos minimum de 5 ms pour assurer un rapport S/B de qualité. Dans le cas où le *dwell time* est de 5 ms, il est possible de placer 160 transitions MRM par groupe scout. Deux de ces transitions seront réservées au scout déclenchant le groupe suivant. De ce fait, dans notre configuration nous pouvons placer au maximum 158 transitions MRM par groupe. Sachant que nous avons 9 composés scouts, cela représente 10 groupes (le premier étant déclenché à  $t = 0$  min) soit 1580 potentielles transitions MRM qui peuvent être ajoutées dans notre méthode. Pour améliorer davantage les capacités de multiplexage, il suffit d'ajouter de nouveaux composés scouts dans la méthode pour ainsi augmenter le nombre de groupes. Toutefois, il n'est pas possible d'augmenter ce nombre de composés scouts indéfiniment. En effet, avoir des composés scouts avec des différences de temps de rétention trop faible entrainerait trop de transitions MRM co-éluées avec celles des scouts (cf. 3.3.3). Le gain en capacité de multiplexage dû à l'ajout de ces très nombreux scouts ne serait pas pertinent, puisque la majorité des transitions MRM seraient dédoublées dans ces conditions.

Si l'on souhaite augmenter davantage les capacités de multiplexage de la méthode Scout-MRM sans ajouter de nouveaux composés scouts ou si l'un des groupes est complet, il faut diminuer le *dwell time* et/ou le nombre de points par pic pour ajouter de nouvelles transitions (cf. tableau n°8 ; cf. chapitre 1 partie 2.3.2). Par exemple, à nombre de points identiques, si l'on diminue le *dwell time* de 5 à 3 ms on augmente de 25% les capacités de multiplexage. Si l'on souhaite conserver le même *dwell time* et avoir 25 % de capacité de multiplexage en plus, cela se fait avec un compromis de 2 points d'acquisition en moins pour décrire le pic chromatographique.

	DT = 5 ms Points/pic = 10	DT = 5 ms Points/pic = 8	DT = 3 ms Points/pic = 10	DT = 3 ms Points/pic = 8	DT = 1 ms Points/pic = 10	DT = 1 ms Points/pic = 8
Transitions / groupe	160	200	200	250	266	332
Gain en nombre de transitions	Référence	25%	25%	56%	66%	108%

Tableau 8 : Variation du nombre de transitions/groupe scout en fonction du *dwell time* (DT = 5 ms ou 3 ms ou 1 ms) et du nombre de points/pic chromatographique (10 ou 8 points) dans le cas de pic chromatographique d'une largeur moyenne de 16 secondes.

Cependant, ce gain en capacité de multiplexage s'accompagne d'une perte de répétabilité de l'intégration des aires des pics chromatographiques comme indiqué sur la figure n°67. En effet, en diminuant le nombre de points, on décrit moins bien le profil gaussien du pic chromatographique et cela conduit à des différences d'aires. De même, la diminution du *dwell time* conduit à une augmentation du bruit de fond analytique et donc également du risque d'une mauvaise intégration des pics chromatographiques.

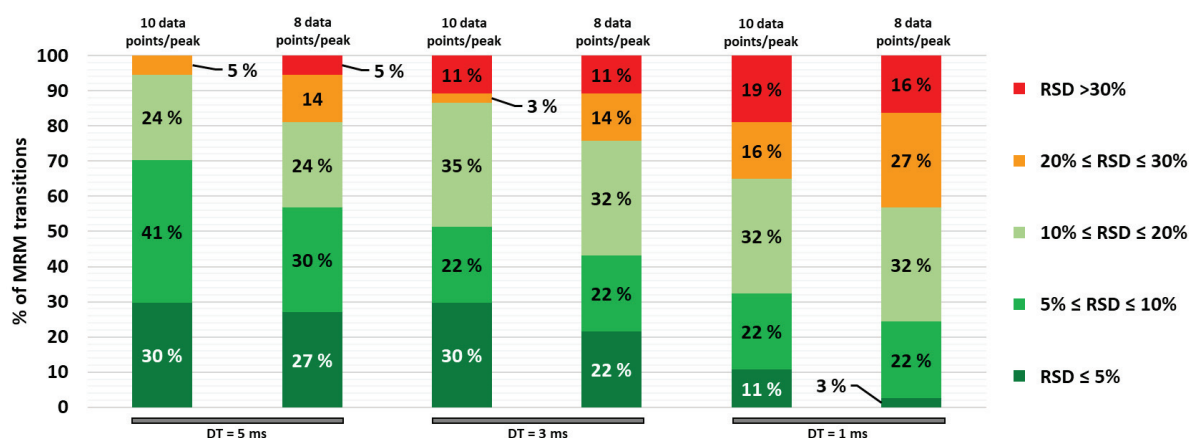
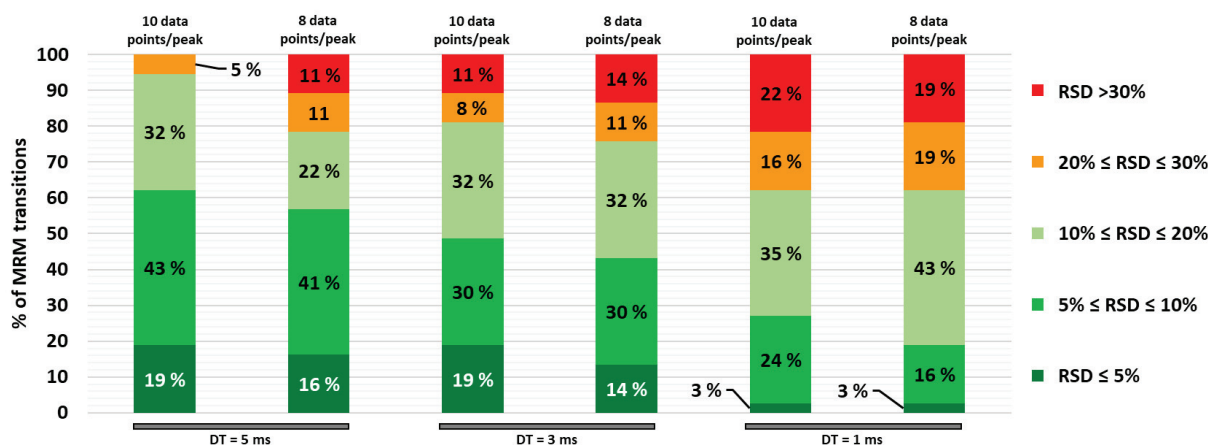


Figure 67 : Impact de la diminution du *dwell time* et du nombre de points par pic chromatographique sur la répétabilité de l'intégration des aires d'une transition en Scout-MRM. Cette étude est basée sur la réponse en aire du signal de 19 standards et 37 transitions MRM avec une large gamme de sensibilité ( $4.8 < S/B < 466$ ) et 3 réplicats par expérimentation. Le RSDle plus important pour chaque condition est respectivement 28%, 49%, 66%, 73%, 64% et 55%.

Il est également important de souligner que les résultats sont répétables aussi bien sur les aires que les hauteurs des pics chromatographiques (cf. figure n°67 et 68). Ainsi, il est possible de quantifier sur les aires ou les hauteurs. En quantifiant sur les hauteurs, il n'est donc plus nécessaire de se soucier des co-élutions entre composés lipidiques et les scouts. L'obtention d'un pic partiel est acceptable du moment que l'on détecte l'apex de ce pic. Ainsi, la capacité de multiplexage ne sera plus impactée par l'obligation de placer des transitions identiques dans deux groupes différents.



**Figure 68** : Impact de la diminution du *dwell time* et du nombre de points par pic chromatographique sur la répétabilité de l'intensité du signal d'une transition en Scout-MRM. Cette étude est basée sur la réponse en intensité du signal de 19 standards et 37 transitions MRM avec une large gamme de sensibilité ( $4.8 < S/B < 466$ ) et 3 réplicats par expérimentation. Le RSD le plus important pour chaque condition est respectivement 26%, 47%, 62%, 73%, 62% et 56%.

Les lipides présentent une grande diversité structurale ce qui implique que le nombre de composés à suivre dans la méthode Scout-MRM doit être conséquent, puisque plus de 25220 structures lipidiques sont référencées dans la base de données Lipid Maps. Même si le mode d'acquisition Scout-MRM permet de très grandes capacités de multiplexage, il n'est pas possible de tendre vers l'analyse en simultanée de dizaines de milliers de lipides. D'autant plus qu'ils ne présentent pas tous un intérêt biologique.

### 3.4. Règles de sélection des transitions MRM à ajouter dans la méthode Scout-MRM

En fonction des matrices biologiques considérées, certaines combinaisons n'ont aucune pertinence biologique. En effet, dans la nature la majorité des acides gras, liés ou libres, possèdent un nombre pair de carbones. Ceci résulte du mode de synthèse de ces composés qui consiste en la condensation successive d'unités de deux carbones [290]. Dans le cas du lipidome humain, plus de 99% de la concentration plasmatique totale en acide gras présente des chaînes paires de carbones [291,292]. Toutefois, cette hypothèse du nombre pair de carbones peut ne pas être vérifiée notamment dans le lipidome des bactéries et des protozoaires. Par exemple, les acides gras à chaîne impaire du lipidome des bactéries et des protozoaires au sein du rumen de vaches représentent respectivement 16.5% et 11% des acides gras totaux [293]. La diversification en nombre de carbones et de doubles liaisons en lipidomique n'est pas aussi importante que l'on pourrait l'imaginer puisque les matrices biologiques sont majoritairement constituées des 15 chaînes d'acides gras suivantes : **AG 14:0, AG 14:1, AG 16:0, AG 16:1, AG 18:0, AG 18:1, AG 18:2, AG 18:3, AG 20:2, AG 20:3, AG 20:4, AG 20:5, AG 22:4, AG 22:5 et AG 22:6** [198,294–298]. Cette liste d'acides gras est appelée liste majoritaire de référence dans la suite de cette thèse. Dans le cadre de notre approche de découverte en mode ciblé, nous conseillons

donc de mettre en priorité dans la méthode Scout-MRM les lipides avec des combinaisons en acide gras citées précédemment. Ainsi, cela limite le champ des possibilités en réduisant le nombre de transitions MRM pour être compatible avec les capacités de multiplexages de la méthode Scout-MRM.

### 3.4.1. Choix des transitions pour les lysophospholipides et les sphingomyélines

Lorsque l'on souhaite investiguer le lipidome des lysophospholipides, il n'est pas nécessaire de tenir compte des règles susmentionnées. En effet, ces lipides ne possèdent qu'une seule chaîne d'acide gras ce qui représente peu de transitions MRM à ajouter dans la méthode pour chaque LPL suivi. Par exemple, il faut ajouter dans la méthode la transition MRM associée à la tête polaire en ESI (+) et la transition MRM associée à l'unique chaîne d'acide gras en ESI (-) pour chaque LPC.

Les sphingomyélines se trouvent dans le même cas de figure que les lysophospholipides. Comme indiqué précédemment (*cf.* 3.1.2), il n'est pas possible de déterminer la composition de la LCB et de la chaîne d'acides gras des sphingomyélines. Ainsi, il existe une seule combinaison possible de transition MRM pour chacune des sphingomyélines (SM X:Y) avec les fragments associés à la tête polaire phosphocholine dans les deux modes d'ionisation ( $m/z = 184$  en ESI (+) et  $168$  en ESI (-)).

Ces deux familles nécessitent peu de transitions MRM par lipide pour pouvoir être analysées. Ainsi, il n'est pas nécessaire de limiter le nombre de composés à suivre puisque ce ne sont pas ces classes lipidiques qui limiteront les capacités de multiplexage en lipidomique. Toutefois, ce n'est pas le cas des autres familles de lipides

### 3.4.2. Choix des transitions pour les phospholipides, les céramides et les glycérolipides autres que les triacylglycérols.

Pour l'analyse des phospholipides en ESI (+), une seule transition MRM est possible pour chaque structure, car le fragment associé est issu de la tête polaire. Par exemple, pour le PE 36:1, il faudra utiliser le fragment 746.6/605.6 correspondant au couple  $[M+H^+]/[M+H-141]$ . Pour les mêmes raisons énoncées précédemment, les règles de sélection peuvent être ignorées.

Lorsque l'on recherche des phospholipides en ESI (-) ou des diacylglycérols en ESI (+), il est nécessaire d'ajouter deux transitions MRM par lipide issues de la présence des deux chaînes d'acides gras dans la structure des lipides. Dans le cas où aucune donnée n'est disponible concernant la composition de ces deux chaînes d'acides gras, il faut ajouter dans la méthode toutes les combinaisons possibles pour ces lipides. Par exemple, pour le PE 36:3, il faut ajouter les transitions MRM associées aux PE 16:1\_20:2, PE 16:0\_20:3, PE 18:1\_18:2, PE 19:1\_17:2, *etc.* En tout, cela représente 53 combinaisons de chaînes d'acides gras différentes pour le PE 36 :3 soit 106 transitions MRM à ajouter dans la méthode. Dans nos conditions de multiplexage optimales (*dwell time* = 1ms et nombre de points par pic = 8) il est

possible de suivre 332 transitions MRM par groupe. Ainsi, il n'est pas envisageable de réaliser une telle étude, car 30% des capacités de multiplexage d'un groupe scout seraient occupées par un seul lipide. Les règles de sélection prennent ici tous leurs sens, car cela ramène le nombre de combinaisons à 6 (soit 12 transitions MRM) : PE 18:0\_18:3, PE 18:1\_18:2, PE 16:0\_20:3, PE 16:1\_20:4, PE 14:0\_22:3 et PE 14:1\_22:2.

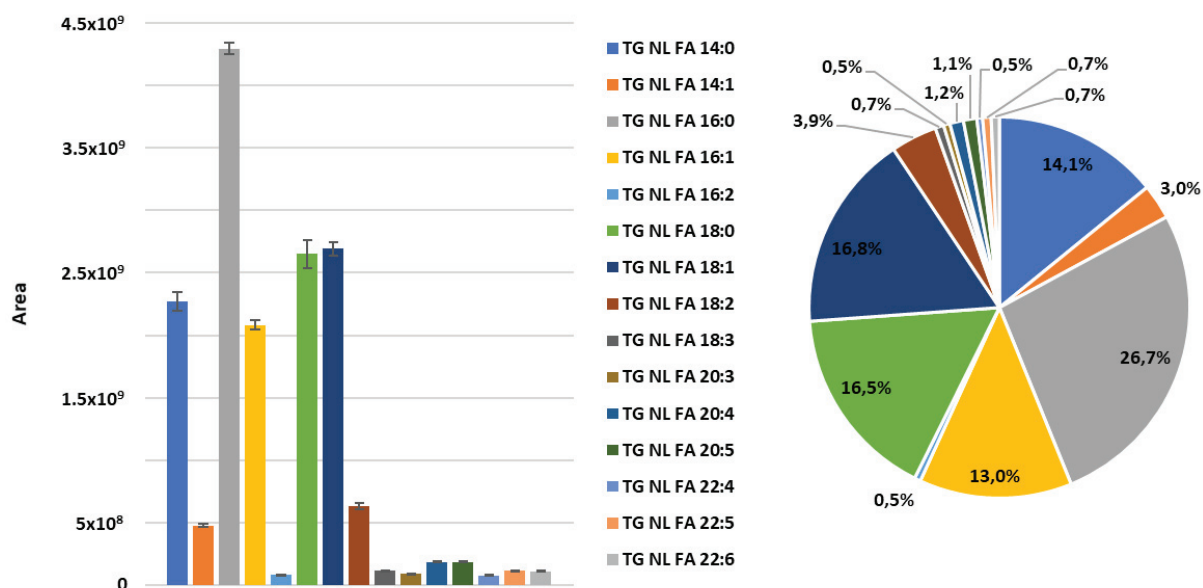
Pour les céramides, une seule transition MRM en ESI (+) est nécessaire pour caractériser leur structure. En effet, à partir du fragment caractéristique de la LCB, on peut remonter à l'information de la composition de leur AG. Par exemple, si un signal est détecté pour le Cer 34:1 avec la transition 536.5/264 [M+H+]/[fragment caractéristique LCB d18:1], on peut en déduire que la chaîne d'acides gras est l'AG 16:0. Comme pour les PL et DG, un même Cer 34:1 possède plusieurs combinaisons de LCB et de chaîne d'acide gras. Par exemple on peut avoir des Cer d18:1/16:0, Cer d16:0/18:1, etc. Toutefois, les céramides sont essentiellement formés d'une LCB d18:1. Ainsi, nous proposons d'utiliser les combinaisons correspondant à la LCB d18:1 et des acides gras qui composent la liste des AG majoritaires.

### 3.4.3. Choix des transitions pour les triacylglycérols

Les triacylglycérols sont un cas particulier lorsqu'aucune donnée n'est disponible concernant leur composition en acide gras au sein de leurs chaînes carbonées. En effet, ce sont des lipides qui possèdent trois chaînes d'acides gras et il n'est pas envisageable de rechercher la totalité des combinaisons. Pour un TG 54:1, cela représente 1014 combinaisons de chaînes d'acides gras différentes soit environ 3000 transitions MRM. Même l'utilisation de la liste des AG majoritaires ne permettrait pas de diminuer suffisamment ce nombre de combinaisons tant les possibilités sont importantes dans un système à 3 chaînes carbonées.

Pour résoudre cette problématique, nous proposons la solution suivante qui consiste à trouver dans un premier temps la composition de l'une des chaînes d'acides gras pour obtenir un système à deux chaînes carbonées inconnues plutôt que trois :

Pour cela, nous avons évalué parmi la liste des chaînes d'acides gras majoritaires celles étant les plus répandues au sein des triglycérides. Ainsi des analyses en mode balayage de perte de neutre ont été réalisées sur des extraits lipidiques de plasma humain, drosophile et *G. fassarum*. Ces matrices ont été sélectionnées pour obtenir un grand nombre de TG de structures différentes afin d'être représentatives d'un panel de TG pouvant être retrouvés dans une matrice biologique. Les résultats obtenus ont indiqué que l'AG 16:0 est présent dans 26.7% des cas, l'AG 18:1 dans 16.8% des cas, l'AG 18:0 dans 16.5% des cas, l'AG 14:0 dans 14.1% des cas et l'AG 16:1 dans 13% des cas (cf. figure n°69).



**Figure 69 : Proportion des acides gras les plus souvent retrouvés dans la composition des TG.** Analyse en mode balayage de perte de neutre (NL, pour *neutral loss*) pour un extrait protéique contenant du plasma humain, de la drosophile et du *G. fossarum*.

Cela nous a permis de définir la liste des 5 chaînes d'acides gras majoritaires spécifiques aux triglycérides : **AG 16:0, l'AG 18:1, l'AG 18:0, AG 14:0 et l'AG 16:1.**

Lorsqu'une étude de caractérisation est réalisée en Scout-MRM pour l'analyse des TG, nous conseillons d'ajouter pour chaque TG X:Y à suivre, les transitions MRM associées à la liste des 5 AG majoritaires spécifiques au TG. Ainsi, si le TG X:Y est bien présent dans la matrice, la probabilité qu'il contienne au moins l'un de ces 5 AG majoritaires est élevée. L'obtention d'un signal pour l'une de ces 5 transitions permet de caractériser l'une des chaînes d'acides gras du TG détecté et d'obtenir un système non plus à 3 chaînes d'acides gras inconnues, mais à 2. Cette analyse permet également de déterminer les AG majoritaires absents du TG à caractériser. Cette information est nécessaire pour la suite de l'identification des structures du TG.

Suite à l'identification d'un premier AG constituant le TG, il est possible de déduire les combinaisons probables des deux AG inconnus. En effet, la somme des carbones et insaturations des 3 AG doit correspondre à la structure du TG à caractériser. Les AG les plus répandus sont répertoriés dans la liste des 15 AG majoritaires de référence. Il est donc possible de déterminer les deux derniers AG constituant le TG permettant d'obtenir le bon nombre d'insaturations et de carbones à partir de cette liste. Il faut ensuite confirmer ces déductions par une seconde analyse Scout-MRM permettant de suivre les transitions spécifiques aux AG envisagés. Un exemple détaillé de la méthode à appliquer pour étudier le TG 54:1 est présenté ci-dessous :

- Pour l'étude du TG 54:1, il faut ajouter dans la première méthode Scout-MRM les transitions associées à une perte de l'AG 16:0, l'AG 18:1, l'AG 18:0, AG 14:0 et l'AG 16:1 soit 5 transitions MRM issues de la liste des 5 AG majoritaires spécifiques des TG.
- Suite à cette analyse, un signal est obtenu seulement pour la perte de l'AG 18:0, l'AG 18:1 et l'AG 16:0. Cela indique que les TG 54:1 présents dans l'échantillon doivent posséder au moins un de ces acides gras (AG 18:0, AG 18:1 et AG 16:0). L'autre information que l'on peut tirer de cette analyse est que les TG 54:1 ne possèdent pas les chaînes d'acides gras AG 14:0 et AG 16:1.
- Il faut ensuite déduire les compositions des deux derniers AG à partir de la liste des 15 AG majoritaires de référence. La somme des carbones et des insaturations doit correspondre aux caractéristiques du TG étudié. De plus, la première analyse en Scout-MRM a permis de déterminer l'absence de deux AG majoritaires AG 14:0 et AG 16:1 qu'il faudra enlever des combinaisons probables. D'après les informations disponibles, on peut déduire des structures probables pour les TG 54:1. Par exemple : TG 18:1\_18:0\_18:0, du TG 16:0\_20:1\_18:0 et du TG 16:0\_20:0\_18:1. Les déductions sont ensuite confirmées par une seconde analyse en Scout-MRM. Pour cela il faut ajouter les transitions associées à une perte de l'AG 16:0, l'AG 18:0, l'AG 18:1, l'AG 20:0 et l'AG 20:1 soit 5 transitions MRM.

La méthodologie décrite dans ce chapitre de thèse est qualifiée d'étude de découverte en mode ciblé et non pas d'étude de découverte au sens d'approche globale. En effet, certaines hypothèses sont incontournables comme le nombre pair de carbones par chaîne d'acide gras et la liste des AG majoritaires pour la composition des chaînes carbonées de chaque lipide. Sans ces règles, le nombre de combinaisons par lipide serait trop important, et cela malgré les grandes capacités de multiplexage de l'approche Scout-MRM. Il n'est pas possible de couvrir toutes les possibilités, en revanche nous sommes capables d'en cibler un très grand nombre et surtout celles qui ont le plus de probabilité d'être présentes au sein de nos échantillons.

En termes de caractérisation d'échantillon complexe, la spectrométrie de masse de haute résolution reste l'outil de prédilection pour l'analyse d'échantillon sans *a priori*. Toutefois, la méthodologie Scout-MRM en lipidomique peut être une bonne alternative dans le cas où l'on n'a pas accès à la HRMS puisqu'elle augmente considérablement les capacités de multiplexage et le nombre de lipides ciblés. D'autant plus qu'à ce jour les appareils de types triple quadripôles sont déjà implantés dans la plupart des laboratoires.

Les études de découverte en mode ciblé nécessitent d'être capable d'analyser un grand nombre de composés au sein d'un échantillon. C'est pour cela que nous avons envisagé le mode d'acquisition Scout-MRM pour ces grandes capacités de multiplexage réduisant de ce fait le nombre de méthodes.



Toutefois, dans le cadre d'échantillon inconnu, l'un des verrous rencontrés en Scout-MRM est de savoir dans quel groupe placer un composé pour lequel nous n'avons aucune information préalable sur son élution par rapport aux scouts sélectionnés. L'un des intérêts en lipidomique est que les différences de nombre de carbones et de doubles liaisons au sein d'une même famille changent la polarité de façon homogène permettant de créer des modèles de rétention et de définir facilement les ordres d'élution.

### 3.5. Construction d'une base de données à l'aide des modèles de rétention

#### 3.5.1. Modèle du nombre de carbones équivalents (ECN)

Afin de prédire le temps de rétention des composés d'intérêts, nous avons sélectionné le modèle du nombre de carbones équivalents (ECN) [272]. Grâce à ces modèles, il est possible de prédire par rapport à un jeu de données les temps de rétention d'autres lipides. Pour l'élaboration des groupes scouts il n'est pas nécessaire de prédire les temps de rétention avec une extrême précision puisque l'information qui nous intéresse est l'élution d'un composé relative aux scouts. Ce modèle repose sur la corrélation entre les propriétés de rétention des composés d'une même famille lipidique et le nombre de carbones au sein des chaînes d'acides gras. En effet, la courbe de corrélation du nombre de carbones en fonction du RT des lipides X:Y, où X représente la somme **variable** des atomes de carbone et Y le nombre **fixe** de double liaisons se traduit par une fonction polynomiale d'ordre 2 (cf. figure n°70 avec l'exemple des PC X:0).

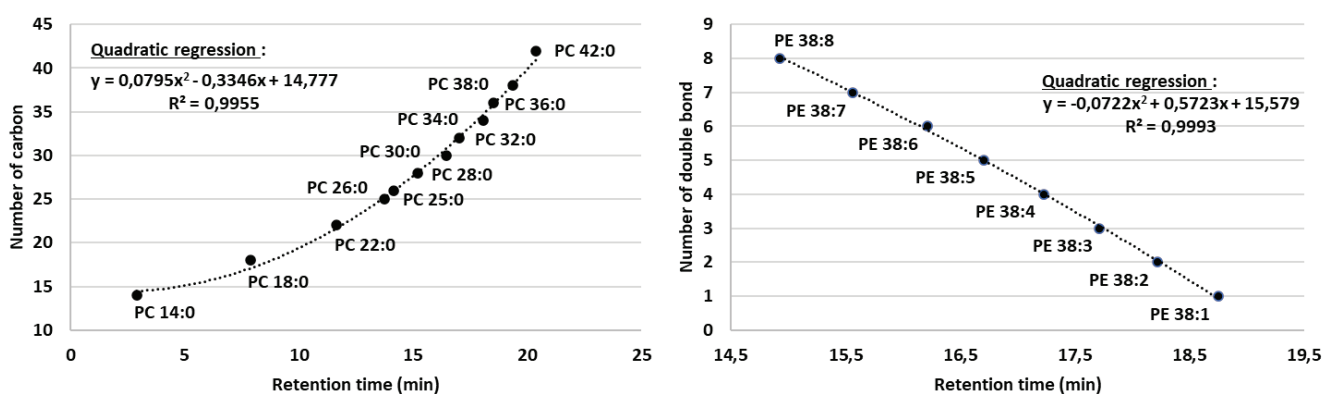


Figure 70 : Exemples de courbes de corrélation correspondant au nombre de carbones en fonction du temps de rétention et au nombre de doubles liaisons en fonction du temps de rétention pour des lipides d'une même famille. L'exemple de courbe de corrélation correspondant au nombre de carbones (à gauche) est réalisé pour le PC X:0 où X représente la somme des atomes de carbone qui composent les chaînes d'acides gras du PC. L'exemple de courbe de corrélation correspondant au nombre de doubles liaisons (à droite) est réalisé pour le PE 38:Y où Y représente la somme des doubles liaisons qui composent les chaînes d'acides gras du PE.

Cette fonction permet de corréler de la même façon le nombre de doubles liaisons et le RT des lipides X:Y, où X représente le nombre de carbones **fixe** et Y la somme des doubles liaisons **variables** qui

composent les chaînes d'acide gras (cf. figure n°70 avec l'exemple des PE 38:Y). Ce phénomène est observé pour l'ensemble des familles lipidiques et grâce à ces équations issues des fonctions polynomiales d'ordre 2, il est très facile et rapide de prédire le temps de rétention des lipides. La multiplication de ces courbes de corrélation permet la mise en place de bases de données lipidiques dans lesquelles figurent des temps de rétention expérimentaux et théoriques issus des modèles de prédiction.

### 3.5.2. Développement de la base de données

Pour mettre en place cette base de données, il a été nécessaire de déterminer les temps de rétention de nombreux lipides. D'une matrice à une autre, la composition du lipidome peut être très variable. C'est pour cela que nous avons utilisé les matrices *G. fossarum*, plasma humain et drosophile pour détecter un grand nombre de lipides et ainsi obtenir une base de données la plus exhaustive. En tout, plus de 400 lipides différents ont été détectés dans l'ensemble de nos trois matrices parmi les familles lipidiques suivantes : PC, PE, PI, PS, LPC, LPE, SM, Cer, DG et TG (cf. annexe n°5). Les familles des PA, PG et MG ne figurent pas dans notre base de données, car ils étaient détectés en trop faible quantité dans les matrices considérées (plasma humain, drosophile et *G. fossarum*).

L'obtention des temps de rétention expérimentaux de ces lipides par rapport aux composés scouts nous a permis d'obtenir deux niveaux d'informations appelés V.0 et V.1.1 qui composent notre base de données (cf. tableau n°9). Le niveau V.0 consiste à identifier à quel groupe scout appartient le composé d'intérêt (ex.  $RT_{\text{scout } 5} < RT_{(\text{DG } 14:0_{16:1})}$  et  $RT_{(\text{DG } 14:1_{16:0})} < RT_{\text{scout } 6}$ ). Cette information est valable pour plusieurs conditions chromatographiques de phase inverse. Le niveau d'information V.1.1 de notre base de données répertorie les valeurs chiffrées des temps de rétention expérimentaux (ex. dans les conditions LC V1.1 :  $RT_{\text{scout } 5} = 16.19 \text{ min} < RT_{(\text{DG } 14:0_{16:1})} = 17.16 \text{ min}$  et  $RT_{(\text{DG } 14:1_{16:0})} = 17.21 \text{ min} < RT_{\text{scout } 6} = 17.68 \text{ min}$ ). Ces temps de rétention sont valables uniquement dans des conditions chromatographiques spécifiques.

Lipides	Composition AG	1 <sup>er</sup> niveau d'informations V.0	2 <sup>ème</sup> niveau d'informations V.1.1	
		Groupe scout	Moyenne RT isomères (min)	RT lipides (min)
DG 30:1	DG 14:0_16:1	Groupe 5	17,18	17,16
	DG 14:1_16:0			17,21

**Tableau 9 : Exemple de base de données en temps de rétention pour les DG.** Le niveau V.0 indique dans quel groupe scout placer le lipide. Le niveau V.1.1 indique une valeur chiffrée de RT pour un lipide dans les conditions LC V1.1. La moyenne des RT des isomères est la valeur qui est utilisée dans les modèles de prédiction.

Par la suite, les modèles de prédiction ont été établis grâce aux courbes de corrélation pour permettre la prédiction des temps de rétention de lipides non détectés dans nos 3 matrices. Les valeurs de temps de rétention prédites peuvent être directement implémentées dans la base de données en ajoutant

les RT théoriques dans le niveau d'information V.1.1 et en indiquant à quel groupe scout appartient théoriquement le composé issu de la prédiction dans le niveau V.0.

Une fois la prédiction des temps de rétention obtenue il a fallu leur associer une incertitude de mesure qui dépend de la façon dont les modèles de prédiction ont été conçus.

### 3.5.3. Incertitude associée à la séparation d'isomères

Les conditions chromatographiques utilisées lors de l'analyse des matrices *G. fossarum*, plasma humain et drosophile lors de la mise en place de la base de données permettent la séparation d'isomères d'un même lipide. Par exemple, il est possible de séparer le PE 18:1\_18:3 du PE 18:2\_18:2 et du PE 16:0\_20:4 (cf. figure n°71 A).

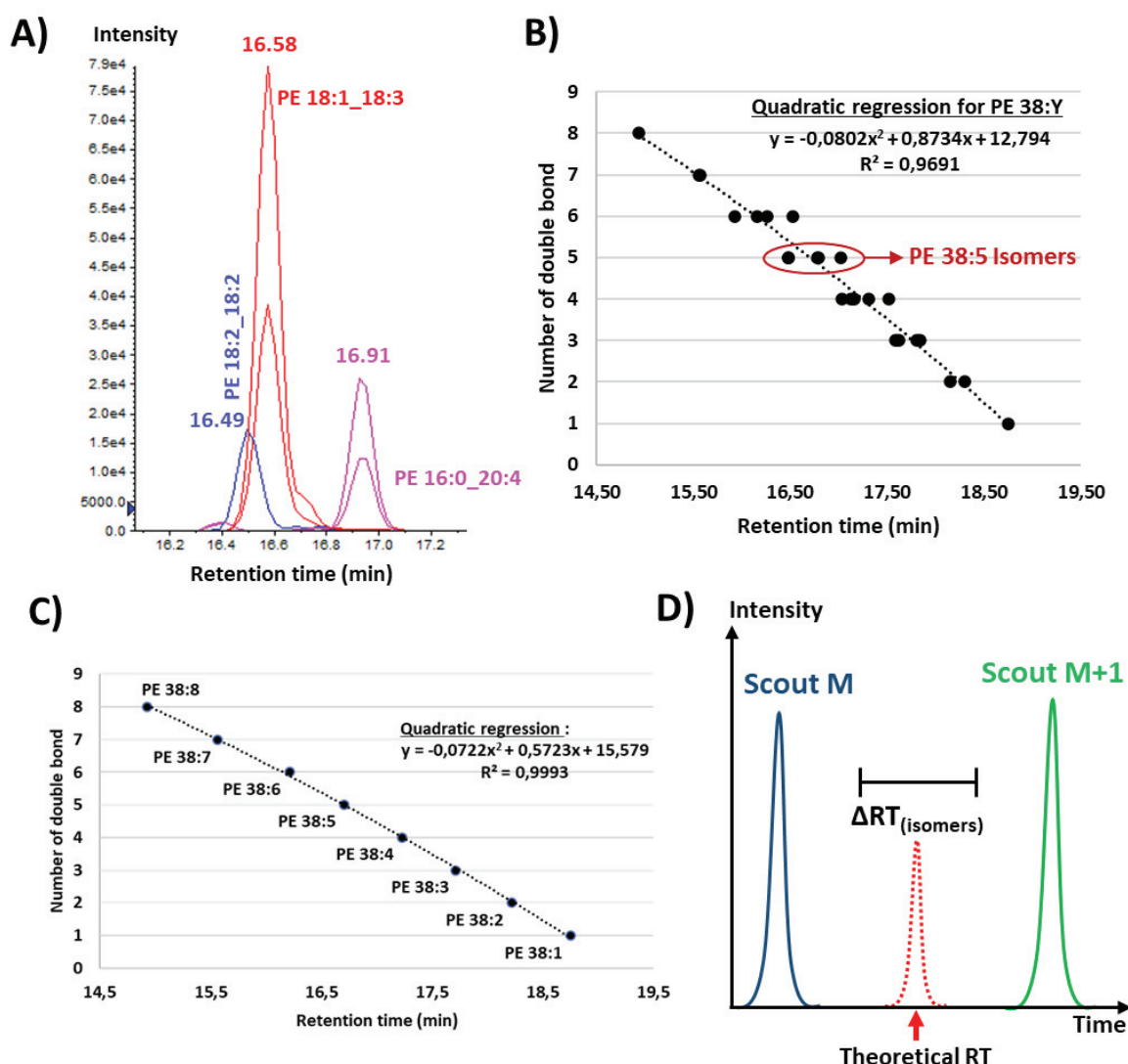


Figure 71 : Impact de la séparation de composés isomères sur la prédiction des temps de rétention. (A) Les conditions RPLC sont capables de séparer les isomères au sein d'une même famille de lipides grâce à leur composition en AG (ex. PE 18:2\_18:2, PE 18:1\_18:3 et PE 16:0\_20:4). (B) Modèles de prédiction issus du RT des isomères indépendants avec une mauvaise corrélation. (C) Modèles de prédiction issus du RT moyen de l'ensemble des isomères avec une excellente corrélation. (D) L'utilisation des RT moyens issus de l'ensemble des isomères, nécessite que les RT théoriques moyens des lipides prédits à partir des modèles de prédiction soient associés d'une incertitude appelée  $\Delta RT_{(isomères)}$ .

Cependant, si on utilise le temps de rétention de chacun des isomères pour la construction des modèles de rétention on observe une mauvaise corrélation des données ( $R^2 = 0.969$ ) comme indiqué sur la figure n°71 B. Pour améliorer cette corrélation, nous avons utilisé un temps de rétention moyen pour chacun des isomères détectés. Cela permet d'obtenir le modèle de rétention décrit sur la figure n°71 C qui présente une meilleure corrélation des données ( $R^2 = 0.999$ ) que sur la figure n°71 B. Les temps de rétention utilisés pour l'élaboration des modèles de prédiction sont donc des moyennes obtenues pour l'ensemble des lipides isomériques détectés. De ce fait, les temps de rétention issus des modèles de prédiction doivent être associés à des incertitudes appelées  $\Delta RT_{(isomères)}$  (cf. figure n°71 D). Les  $\Delta RT_{(isomères)}$  sont calculés à partir de la différence maximale entre les temps de rétention de deux isomères pour une même famille lipidique. Ces valeurs sont référencées dans le tableau n°10.

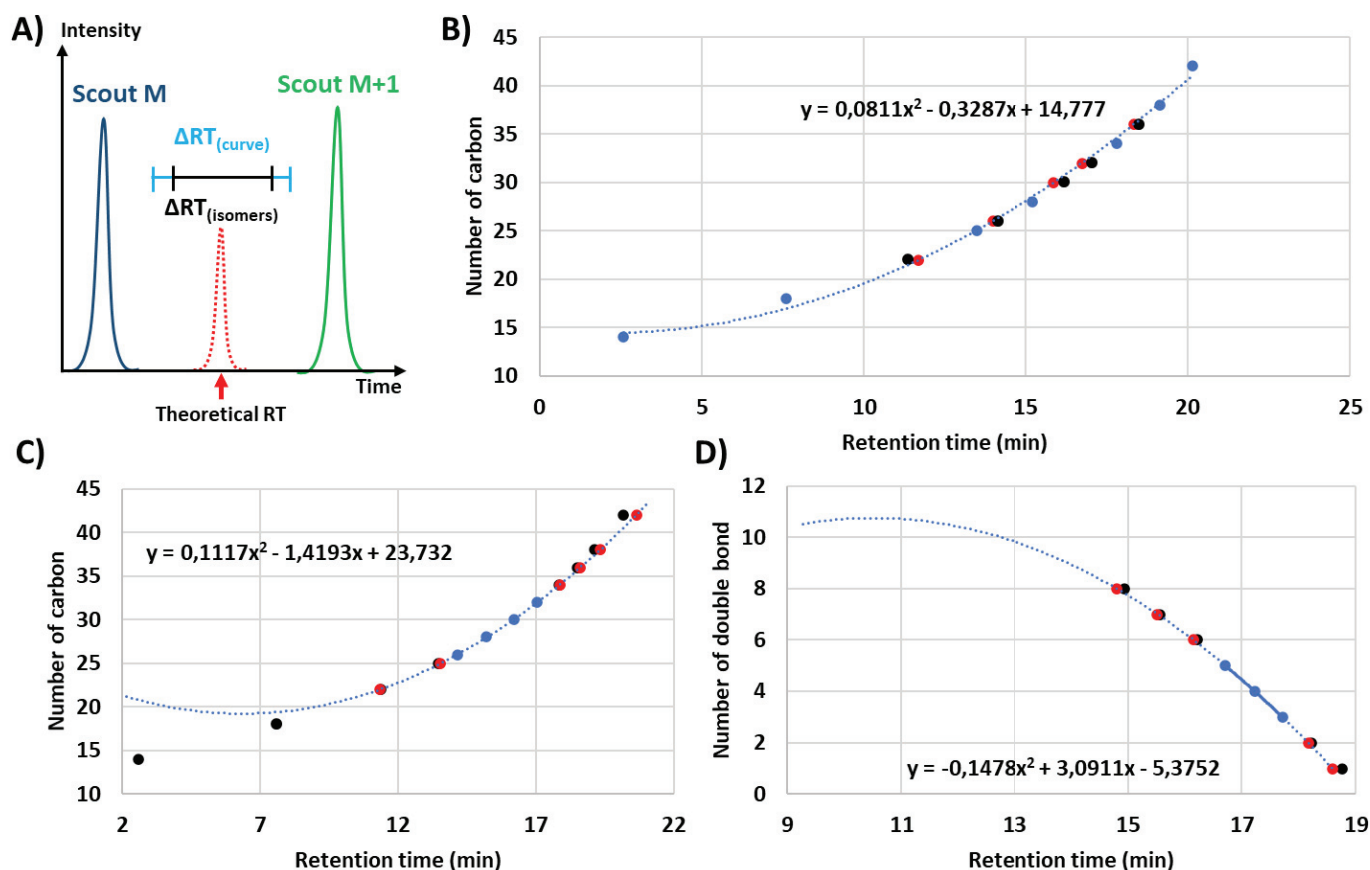
	$\Delta RT_{(isomères)}$
PC	$\pm 0.3$ min
PE	$\pm 0.3$ min
PI	$\pm 0.4$ min
PC	$\pm 0.3$ min
PS	$\pm 0.25$ min
TAG, DAG, Cer et SM	Pas de variation de RT observée pour les différents isomères
LPC et LPE	Pas de $\Delta RT_{(isomères)}$ , car qu'une seule chaîne d'acide gras

Tableau 10 :  $\Delta RT_{(isomères)}$  associée à chaque famille de lipides.

### 3.5.4. Incertitude en dehors des modèles de prédiction

Une deuxième incertitude de mesure peut être associée à la prédiction d'un temps de rétention à travers le  $\Delta RT_{(courbe)}$  qui doit être ajouté à l'incertitude  $\Delta RT_{(isomères)}$  (cf. figure n°72 A). Pour éviter que l'incertitude ne soit trop grande et que des transitions MRM issues de prédictions soient placées dans plusieurs groupes scouts diminuant le degré de multiplexage, il faut éviter de réaliser des prédictions en dehors des modèles.

Le  $\Delta RT_{(courbe)}$  correspond à l'incertitude de mesure associée à la différence entre le RT expérimental (points noirs sur la figure n°72 B, C et D) et le RT théorique (points rouges sur la figure n°72 B, C et D). Lorsque la prédiction du temps de rétention est réalisée sur des lipides présents au sein de la courbe de corrélation, aucune dérive n'est observée et la prédiction est précise (cf. figure n°72 B). En revanche, dans le cas où la prédiction est réalisée en extrapolant en dehors de la courbe de corrélation (cf. figure n°72 C et D), une dérive entre la droite polynomiale d'ordre 2 et le temps de rétention expérimental est observée. Plus l'extrapolation sera importante, plus cette dérive le sera également. Il est difficile de donner une valeur chiffrée à l'incertitude  $\Delta RT_{(courbe)}$ , car elle est dépendante de l'alignement des points expérimentaux qui ont servi à construire la fonction polynomiale d'ordre 2.



**Figure 72 : Incertitudes associées à la prédiction des temps de rétention.** (A) L'extrapolation en dehors du modèle de prédiction implique une incertitude supplémentaire appelée  $\Delta RT_{(courbe)}$ . (B) Cette figure met en évidence les RT théoriques (rouge) calculés à partir du modèle de prédiction (bleu) lorsque les lipides prédits sont à l'intérieur des courbes. Les RT en noir correspondent aux RT expérimentaux pour ces mêmes lipides prédits. (C et D) La figure D (nombre de carbonnes) et E (nombre d'insaturations) mettent en évidence les RT théoriques (rouge) calculés à partir du modèle de prédiction (bleu) lorsque les lipides prédits sont à l'extérieur des courbes. Les RT en noir correspondent aux RT expérimentaux pour ces mêmes lipides prédits.

Puisqu'il n'est pas recommandé de réaliser des prédictions en dehors des domaines de corrélation, il est nécessaire d'analyser d'autres matrices que celles qui ont été employées pour enrichir la base de données. Cela permettra de détecter de nouveaux lipides qui pourront être ajoutés à la base de données afin d'améliorer les modèles prédictifs en agrandissant les domaines de corrélation.

### 3.6. Méthodologie pour améliorer la base de données

Afin d'ajouter de nouveaux lipides dans la base de données en temps de rétention deux méthodologies peuvent être utilisées. La 1<sup>ère</sup> méthode est plus contraignante à mettre en place, car elle impose des conditions chromatographiques particulières, mais permet d'améliorer de manière significative la base de données par rapport à la 2<sup>ème</sup> méthode.

### 3.6.1. 1<sup>ère</sup> méthodologie : Amélioration de la base de données en temps de rétention à partir de conditions LC imposées

La 1<sup>ère</sup> méthodologie pour implémenter de nouveaux lipides et améliorer la base de données en temps de rétention consiste à réaliser les étapes suivantes qui sont décrites dans la figure n°73 :

- Dans un premier temps, il faut ajouter les composés scouts dans l'extrait lipidique à analyser qui contient les nouveaux lipides à ajouter dans la base de données (cf. figure n°73 A). Les conditions chromatographiques pour analyser l'extrait lipidique doivent être les mêmes que celles qui ont servi à développer la base de données V.1.1 (cf. figure n°73 B). Par exemple, si la colonne Xselect CSH C18 (2.1x100 mm, 3.5 µm) a été utilisée avec des phases mobiles et un gradient spécifique pour la mise en place de la base de données V1.1, il faudra se placer dans les mêmes conditions. Le mode d'acquisition MS/MS utilisé doit être le mode MRM classique, car aucune information concernant les temps de rétention de ces nouveaux lipides n'est disponible (ils ne sont pas dans notre base de données).

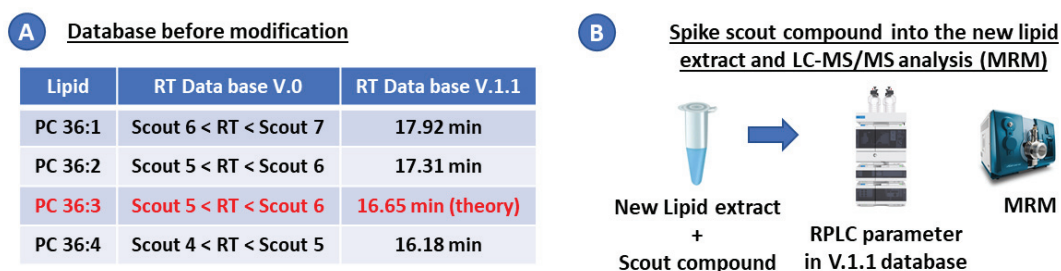


Figure 73 A et B : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration.

- La détection des nouveaux lipides (ex. PC 36:3 et PC 36:6) et des composés scouts permet d'améliorer la base de données V.0 (cf. figure n°73 C). En effet, la comparaison des temps de rétention des nouveaux lipides par rapport aux scouts permet de savoir dans quel groupe scout placer ces nouveaux lipides (cf. figure n°73 D). Dans l'exemple, la détection du PC 36:3 permet également de valider des informations théoriques issues d'une précédente prédiction.

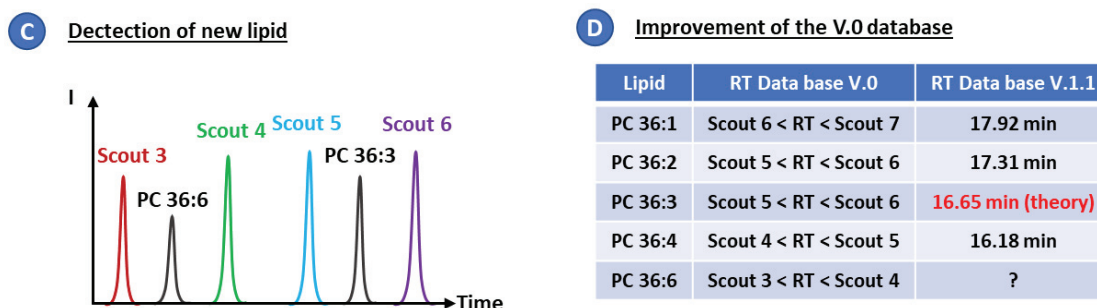


Figure 73 C et D : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration.

- Même si les conditions analytiques sont les mêmes que celles employées pour élaborer la base de données V1.1, il n'est pas possible d'ajouter directement les valeurs chiffrées des temps de rétention des nouveaux lipides analysés. Des variations en temps de rétention peuvent être observées lors du transfert de méthode issue des paramètres LC de la base de données V.1.1. Ces variations peuvent être dues à des effets matrices issus de la nature de l'échantillon analysé, à des volumes de délais différents, des numéros de lots de colonnes chromatographiques différents, etc. Il est donc indispensable de normaliser les nouveaux temps de rétention avant de les ajouter à la base de données. Pour cela il faut comparer les temps de rétention des scouts de la base de données à ceux obtenus expérimentalement. Il faut ensuite réaliser une droite de corrélation entre les temps de rétention des scouts issus de la base de données V1.1 en fonction des temps de rétention des scouts dans la nouvelle analyse. Grâce à l'équation de cette droite, il est possible de normaliser les temps de rétention des nouveaux lipides détectés dans le référentiel de la base de données V1.1 (cf. figure n°73 E).

**E** Normalization of RT according to the V.1.1 database

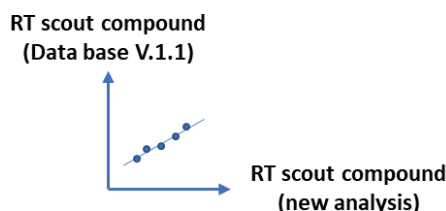


Figure 73 E : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration.

- L'ajout des RT des nouveaux lipides dans la base de données V1.1 entraîne le passage de la version V1.1 à la version V1.2 (cf. figure n°73 F).

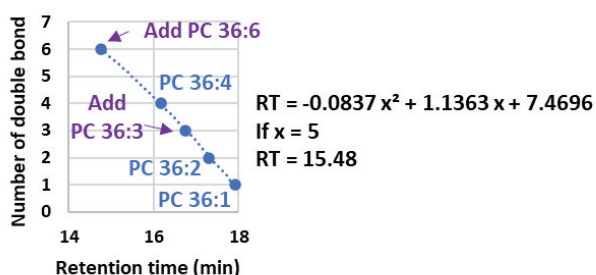
**F** Improvement of the V.1.1 database (V1.1 → V1.2)

Lipid	RT Data base V.0	RT Data base V.1.2
PC 36:1	Scout 6 < RT < Scout 7	17.92 min
PC 36:2	Scout 5 < RT < Scout 6	17.31 min
PC 36:3	Scout 5 < RT < Scout 6	16.76 min
PC 36:4	Scout 4 < RT < Scout 5	16.18 min
PC 36:6	Scout 3 < RT < Scout 4	14.76 min

Figure 73 F : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration.

- Suite à l'ajout des nouveaux lipides dans les modèles, les prédictions des temps de rétention peuvent être mises à jour.. Cela permet d'améliorer la précision de ces modèles et de prédire le temps de rétention de nouveaux lipides (cf. figure n°73 G). Amélioration de la base de données V1.2 en ajoutant les temps de rétention issus des modèles de prédiction (cf. figure n°73 H).



**G** Improvement of prediction models (V.1.2 database)

**H** Improvement of the V.1.2 database with prediction

Lipid	RT Data base V.0	RT Data base V.1.2
PC 36:1	Scout 6 < RT < Scout 7	17.92 min
PC 36:2	Scout 5 < RT < Scout 6	17.31 min
PC 36:3	Scout 5 < RT < Scout 6	16.76 min
PC 36:4	Scout 4 < RT < Scout 5	16.18 min
PC 36:5	Scout 4 < RT < Scout 5	15.48 min (theory)
PC 36:6	Scout 3 < RT < Scout 4	14.76 min

Figure 73 G et H : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidiques en temps de rétention lorsque les conditions chromatographiques utilisées sont les mêmes que celles employées pour son élaboration.

### 3.6.2. 2<sup>ème</sup> méthodologie : Amélioration de la base de données en temps de rétention à partir de conditions LC variables

La 2<sup>ème</sup> méthodologie pour implémenter de nouveaux lipides et améliorer la base de données en temps de rétention consiste à réaliser les étapes suivantes qui sont décrites dans la figure n°74 :

- Dans un premier temps, il faut ajouter les composés scouts dans l'extrait lipidique à analyser qui contient les nouveaux lipides à ajouter dans la base de données (cf. figure n°74 A). Les conditions chromatographiques pour analyser l'extrait lipidique sont différentes de celles qui ont servi à développer la base de données V.1.1 (cf. figure n°74 B). Le mode d'acquisition MS/MS utilisé doit être le mode MRM classique, car aucune information concernant les temps de rétention de ces lipides n'est disponible (ils ne sont pas dans notre base de données).

**A** Database before modification

Lipid	RT Data base V.0	RT Data base V.1.1
PC 36:1	Scout 6 < RT < Scout 7	17.92 min
PC 36:2	Scout 5 < RT < Scout 6	17.31 min
PC 36:3	Scout 5 < RT < Scout 6	16.65 min (theory)
PC 36:4	Scout 4 < RT < Scout 5	16.18 min

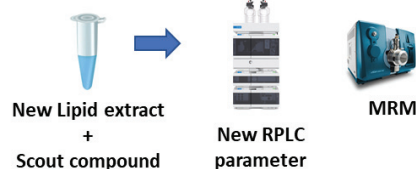
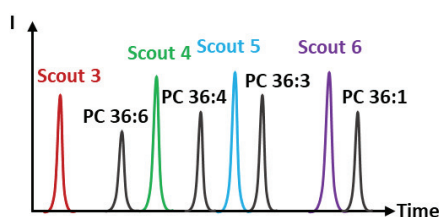
**B** Spike scout compound into the new lipid extract and LC-MS/MS analysis (MRM)


Figure 74 A et B : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidique en temps de rétention lorsque les conditions chromatographiques utilisées sont différentes de celles employées pour son élaboration.

- La détection des nouveaux lipides (ex. PC 36:3 et PC 36:6) et des composés scouts permet d'améliorer la base de données V.0 (cf. figure n°74 C). En effet, la comparaison des temps de rétention des nouveaux lipides par rapport aux scouts permet de savoir dans quel groupe scout placer ces nouveaux lipides (cf. figure n°74 D). Dans l'exemple, la détection du PC 36:3 permet également de valider des informations théoriques issues d'une précédente prédiction. Puisque les conditions chromatographiques sont différentes de celles employées pour l'élaboration de la base de données V1.1, il n'est pas possible d'utiliser cette base de données. Toutefois, les PC 36:1, PC 36:3, PC 36:4 et PC 36:6 sont détectés dans la matrice lipidique. Cela permet de créer une nouvelle base de données nommée V.2.1 en répertoriant les nouveaux temps de rétention de ces lipides

(cf. figure n°74 D). Les bases de données V.2.1 et V.1.1 sont complémentaires, car elles fournissent des informations pour compléter la base de données V.0.

**C** Detection of new lipid



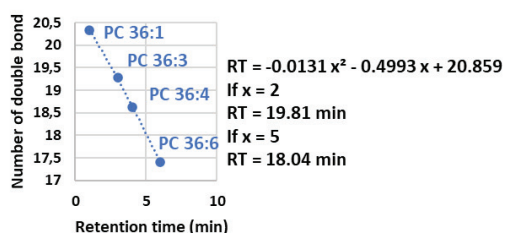
**D** Improvement of the V.0 database and generation of a V2.1 database

Lipid	RT Data base V.0	RT Data base V.1.1	RT Data base V.2.1
PC 36:1	Scout 6 < RT < Scout 7	17.92 min	20.34 min
PC 36:2	Scout 5 < RT < Scout 6	17.31 min	?
PC 36:3	Scout 5 < RT < Scout 6	16.65 min (theory)	19.27 min
PC 36:4	Scout 4 < RT < Scout 5	16.18 min	18.62 min
PC 36:6	Scout 3 < RT < Scout 4	?	17.4 min

Figure 74 C et D : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidique en temps de rétention lorsque les conditions chromatographiques utilisées sont différentes de celles employées pour son élaboration.

- Grâce aux temps de rétention répertoriés dans la base de données V.2.1, il est possible d’élaborer de nouveaux modèles de prédiction et de déterminer le temps de rétention de nouveaux lipides (cf. figure n°74 E). La base de données V2.1 peut être améliorée en ajoutant les temps de rétention issus des nouveaux modèles de prédiction (cf. figure n°74 F).

**E** New prediction models (V.2.1 database)



**F** Improvement of the V.2.1 database with prediction

Lipid	RT Data base V.0	RT Data base V.1.1	RT Data base V.2.1
PC 36:1	Scout 6 < RT < Scout 7	17.92 min	20.34 min
PC 36:2	Scout 5 < RT < Scout 6	17.31 min	19.81 min (theory)
PC 36:3	Scout 5 < RT < Scout 6	16.65 min (theory)	19.27 min
PC 36:4	Scout 4 < RT < Scout 5	16.18 min	18.62 min
PC 36:5	Scout 4 < RT < Scout 5	?	18.04 min (theory)
PC 36:6	Scout 3 < RT < Scout 4	?	17.4 min

Figure 74 E et F : Méthodologie pour implémenter de nouveaux lipides dans la base de données lipidique en temps de rétention lorsque les conditions chromatographiques utilisées sont différentes de celles employées pour son élaboration.

La 1<sup>ère</sup> méthodologie est plus contraignante à mettre en place que la 2<sup>ème</sup> car les conditions chromatographiques sont imposées, mais en contrepartie elle permet d’améliorer des modèles de prédictions déjà existants. De manière générale, plus la méthodologie Scout-MRM sera utilisée par la communauté des lipidomistes pour analyser de nouvelles matrices plus la base de données pourra être améliorée grâce à l’ajout de nouveaux lipides. L’enrichissement de la base de données a pour conséquence l’amélioration des capacités de découverte en mode ciblé de notre approche Scout-MRM en permettant de rechercher plus de lipides dans les échantillons non caractérisés.

À ce stade, nous avons démontré que nous avons à disposition une méthode à haut degré de multiplexage qui peut être utilisée pour réaliser des études de découverte ciblées sur des échantillons inconnus pour détecter la présence de composés issus de notre base de données. Toutefois, cette base de données a été obtenue à partir de conditions chromatographiques particulières qui sont décrites

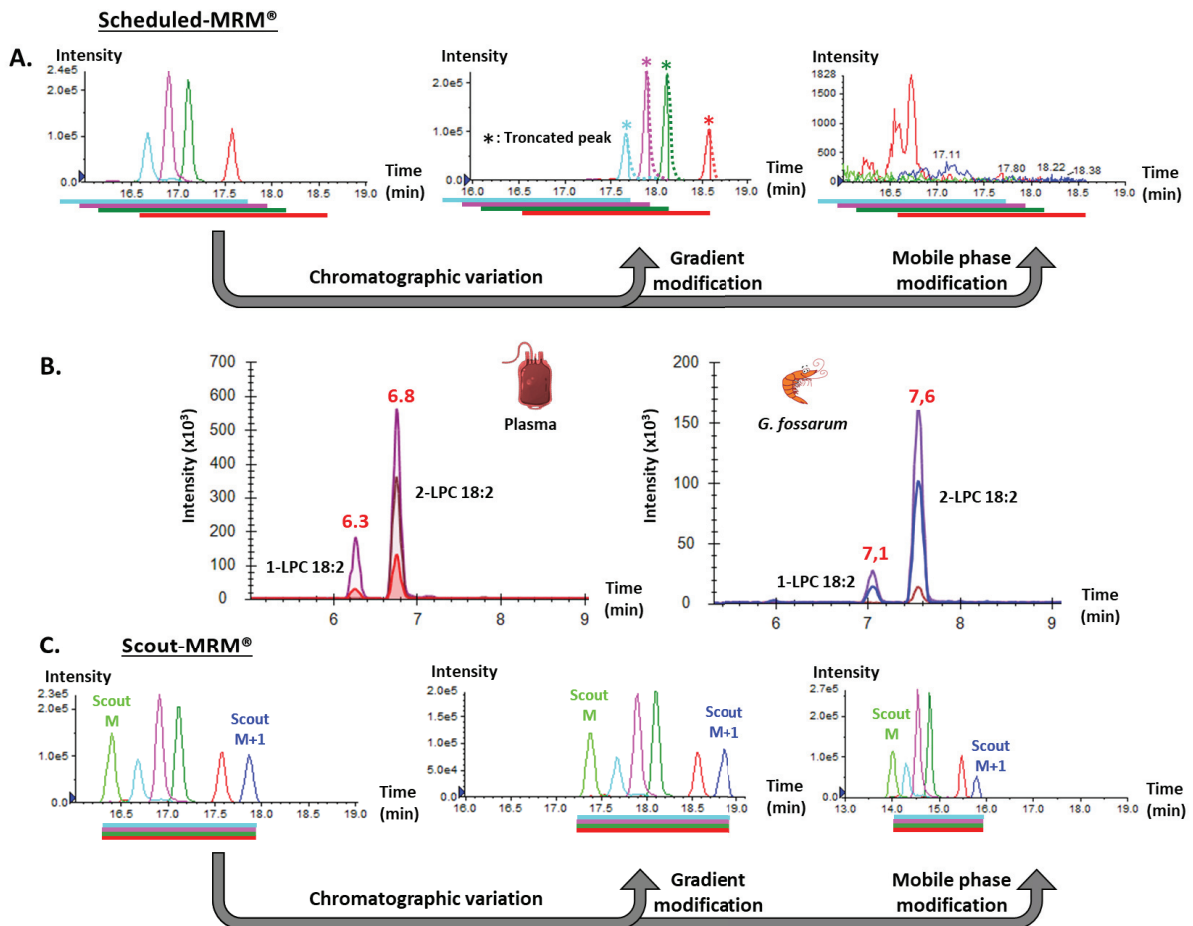
dans la partie 2.4 du matériels et méthodes. Pour que cette méthodologie de découverte en mode ciblé puisse être pertinente, il faut que la base de données puisse être utilisée, quelles que soient les conditions LC.

## 3.7. Transférabilité du Scout-MRM

### 3.7.1. Indépendance des temps de rétention de la méthodologie Scout-MRM

Pour pouvoir présenter un réel intérêt, les méthodes de découverte en mode ciblé doivent être facilement déployées en fonction des paramètres chromatographiques. Elles doivent présenter un caractère « prêt à l'emploi » aussi qualifié de « *plug and play* ». C'est-à-dire que la technologie développée doit permettre une mise en œuvre rapide et requérant un minimum d'intervention de la part de l'utilisateur. En Scout-MRM appliqué à la lipidomique, cela se traduit par l'utilisation d'une même méthode d'acquisition en spectrométrie de masse compatible avec une multitude de configurations chromatographiques de phase inverse différentes : Variation des lots de colonnes, des phases mobiles, des gradients, etc.

Il ne faut pas que les potentiels décalages en temps de rétention dus à des variations de paramètres analytiques impactent la détection des transitions MRM (pic tronqué ou pic non détecté) comme cela peut être le cas en sMRM (cf. chapitre 2 partie 3.5.3.A). Par exemple, si le gradient ou les phases mobiles diffèrent, les temps de rétention des composés varient et peuvent conduire à la détection partielle (pic tronqué) ou à la non-détection de pics chromatographiques lorsqu'ils se retrouvent en dehors des fenêtres de temps (cf. figure n°75 A). De plus, les méthodes de découverte en mode ciblé ont pour vocation d'être employées sur des échantillons de natures très diverses. Le passage d'un type d'échantillon à un autre entraîne souvent des effets matrices qui modifient les temps de rétention des composés lipidiques comme indiqué sur la figure n°75 B. Dans ce cas de figure, la méthode Scout-MRM est un outil de prédilection, car les composés scouts subissent les mêmes décalages en temps de rétention et corrigent automatiquement les fenêtres de détection lors de la variation des conditions analytiques (cf. figure n°75 C). La méthode Scout-MRM peut donc être employée pour des matrices et des conditions chromatographiques très diverses traduisant ainsi le caractère *plug and play* de la méthodologie de découverte en mode ciblé que nous avons développée.



**Figure 75 : Impact des décalages en temps de rétention des lipides lors d'une analyse en sMRM et en Scout-MRM.** (A) La variation de la pente du gradient d'éluion et le changement de la phase mobile B (ACN/H<sub>2</sub>O 40/60 v/v à ACN/H<sub>2</sub>O 60/40 v/v) impliquent un décalage du temps de rétention en dehors des fenêtres temporelles, ce qui entraîne une perte d'information en sMRM (pic tronqué, disparition du pic). (B) Ces décalages de temps de rétention peuvent également se produire lors du changement de matrice (plasma à *G. fassarum*). (C) Les mêmes variations que lors des analyses sMRM impliquent un décalage en temps de rétention qui a également un impact sur les composés scouts. En conséquence, les composés scouts corrigent automatiquement la position des fenêtres de détection.

Le Scout-MRM présente bien des avantages concernant son adaptabilité aux changements de paramètres LC (changement de matrice, de débit, de gradient, de phases mobiles, etc.). Le Scout-MRM corrige automatiquement les fenêtres de détection lors des variations analytiques. Toutefois, cette correction dynamique peut présenter des limites lorsque les variations de temps induisent des changements de sélectivité entre un composé d'intérêt et un scout. La conséquence est que le composé change de groupe conduisant ainsi à sa non-détection. Ces changements de sélectivité peuvent survenir lorsque l'on modifie la nature de la phase stationnaire des colonnes chromatographiques.

### 3.7.2. Changement de phase stationnaire et potentiels changements de sélectivité en Scout-MRM

Afin de nous assurer que notre méthodologie est compatible avec différentes phases stationnaires de type RPLC, nous avons évalué les changements de sélectivité des lipides par rapport aux composés

scouts. Pour cela, nous avons analysé le même extrait lipidique issu des matrices plasma humain, drosophile et *G. fossarum* avec la même méthode Scout-MRM, mais avec 5 colonnes de phase inverse différentes. Les 5 colonnes RPLC sélectionnées sont couramment référencées dans la littérature en lipidomique [55,299–301]. La 1<sup>ère</sup> colonne est la XSelect CSH C18 (2.1x100 mm, 3.5 µm de taille de particule) qui possède une phase apolaire chargée positivement. Il s'agit de notre colonne de référence qui a été utilisée pour développer la méthode Scout-MRM. Les 2<sup>ème</sup> et 3<sup>ème</sup> colonne sont la Zorbax eclipse plus C18 (2.1x100 mm, 3.5 µm de taille de particule) et l'Accucore C18 (2.1x100 mm, 2.6 µm de taille de particule) qui ont des phases apolaires neutres. La 4<sup>ème</sup> colonne est la Xbridge BEH C18 premier (2.1x100 mm, 2.5 µm de taille de particule) qui a une phase apolaire neutre. La BEH premier adopte la technologie *MaxPeak High Performance Surface* qui réduit considérablement les interactions indésirables entre l'analyte et la surface de la colonne qui peuvent entraîner un élargissement des pics chromatographiques et des pertes d'intensité du signal. La 5<sup>ème</sup> colonne est la XSelect HSS T3 (2.1x150 mm, 2.5 µm de taille de particule) qui a une phase stationnaire plus hydrophobe que les phases C18 classiques comme la CSH XSelect C18.

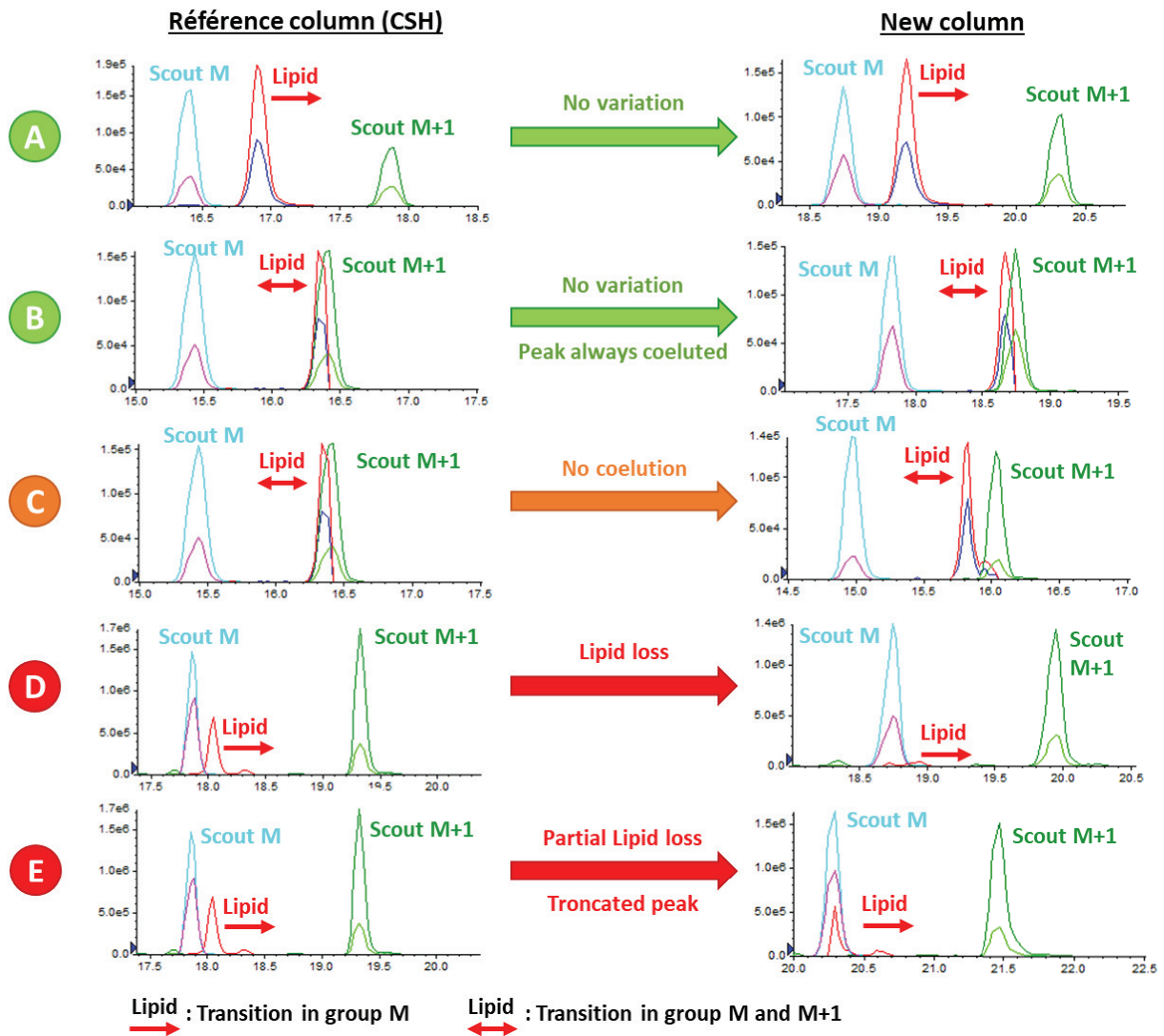
Lors du passage de la colonne de référence (CSH C18) à une autre colonne de phase inverse, 5 cas différents peuvent être observés comme indiqué dans la figure n°76 :

- **(A)** Un lipide élué entre deux scouts est toujours élué entre ces mêmes scouts lors du changement de colonne.
- **(B)** Un lipide co-élué avec un scout est toujours co-élué avec le même scout lors du changement de colonne.
- **(C)** Un lipide co-élué avec un scout n'est plus co-élué avec ce scout lors du changement de colonne.
- **(D)** Inversion entre le temps de rétention d'un composé scout et le temps de rétention d'un lipide d'intérêt non co-élué avec ce scout lors du changement de colonne.
- **(E)** Un lipide non co-élué avec un scout devient co-élué avec un scout lors du changement de colonne.

Les conditions **(A)** et **(B)** ne présentent aucun désavantage, car elles n'entraînent aucune perte d'information (pas de transition MRM non détectée ni de pic chromatographique tronqué) et aucune diminution de capacité de multiplexage par rapport à la condition de référence. La condition **(C)** ne présente aucune perte d'information. Toutefois, les capacités de multiplexage sont diminuées, car la transition MRM est placée dans deux groupes scouts successifs alors que cela n'est plus nécessaire. Il n'y a plus de co-élution entre le lipide et le composé scout lors du changement de colonne. La condition **(D)** est la plus problématique, car il y a une perte d'information suite à l'absence de détection d'un lipide. Ce cas de figure peut compromettre une étude de découverte en mode ciblé, car lors de l'absence de détection, il n'y a aucun indice qui indique que le composé est réellement présent dans l'échantillon, mais que sa transition MRM est placée dans le mauvais groupe scout. La condition **(E)** est



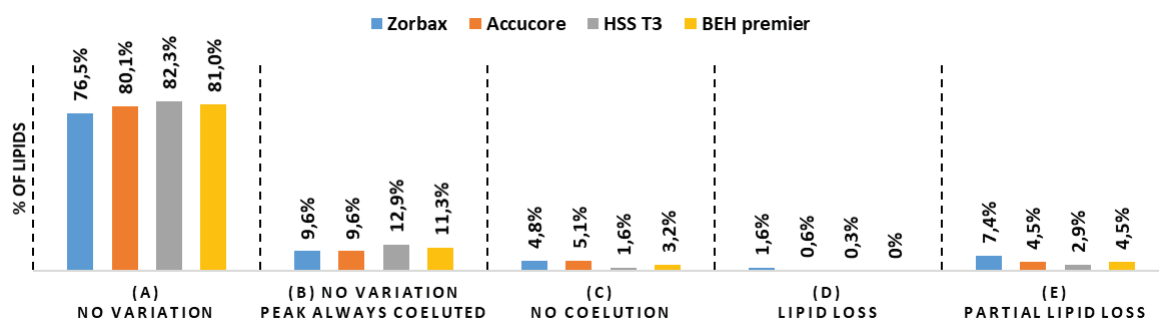
problématique, car il y a une perte d'information suite à l'obtention d'un pic chromatographique tronqué. Cependant, ce cas de figure est moins compromettant que le cas (D), car la visualisation d'une partie du pic indique que le composé est bien présent, mais qu'il manque une partie de l'information. Suite à cette observation, il suffit de placer la transition MRM dans les deux groupes scouts pour récupérer l'information manquante.



**Figure 76 :** Différents cas rencontrés lors de l'évaluation des changements de sélectivité en Scout-MRM lors du passage d'une colonne de phase inverse à une autre. La figure A montre le cas où un lipide élué entre deux scouts et élué toujours entre ces mêmes scouts lors du changement de colonne. La figure B montre le cas où un lipide co-élué avec un scout et co-élué toujours avec le même scout lors du changement de colonne. La figure C montre le cas où un lipide co-élué avec un scout n'est plus co-élué avec le même scout lors du changement de colonne. La figure D montre le cas où il y a une inversion entre le temps de rétention d'un composé scout et le temps de rétention d'un lipide d'intérêt lors du changement de colonne. La figure E montre le cas où un lipide non co-élué avec un scout devient co-élué avec un scout lors du changement de colonne.

Les résultats obtenus pour chacune des colonnes suite à l'analyse de 311 lipides issus de 10 familles (PC / PE / PS / PI / LPC / LPE / SM / Cer / DG / TG) présents dans notre extrait lipidique (plasma, drosophile et *G. fassarum*) sont présentés dans la figure n°77. Les données présentées indiquent des résultats similaires lors des changements de colonne pour l'ensemble des configurations testées. En

moyenne, le cas (A) est obtenu pour 80% des lipides de l'étude, le cas (B) pour 10%, le cas (C) pour 4%, le cas (D) pour 1% et le cas (E) pour 5%. En résumé dans 94% des cas ((A) + (B) + (C)), les lipides se retrouvent toujours dans le même groupe. De plus, le cas de figure (D) qui est le plus problématique ne représente que 1% des lipides.



**Figure 77 : Résultats de l'étude concernant l'évaluation des changements de sélectivité en Scout-MRM pour 4 colonnes en phase inverse par rapport à une colonne de référence.** Les quatre colonnes étudiées sont la colonne Zorbax eclipse plus C18 (2.1x100 mm, 3.5 µm), la colonne Accucore C18 (2.1x100 mm, 2.6 µm), la colonne XSelect HSS T3 (2.1x150 mm, 2.5 µm) et la colonne Xbridge BEH premier (2.1x100 mm, 2.5 µm). La colonne XSelect CSH C18 (2.1x100 mm, 3.5 µm) est la colonne de référence avec laquelle la méthode Scout-MRM a été développée. L'étude a été réalisée sur 311 lipides issus de 10 familles de lipides (PC / PE / PS / PI / LPC / LPE / SM / Cer / DG / TG). Les cas A, B, C, D et E font référence aux différents cas possibles présentés dans la figure n°76.

Ainsi, nous pouvons conclure qu'il y a peu de changement de sélectivité entre un lipide et un composé scout lipidique lors des analyses lipidomiques en RPLC. La méthodologie Scout-MRM pour la découverte ciblée de composés en lipidomique est donc pertinente grâce à son caractère *plug and play* qui a été prouvé expérimentalement. En effet, le Scout-MRM a rempli son rôle en corrigeant automatiquement les fenêtres de détection lors de la variation des conditions analytiques par le changement de colonne de phase inverse.

Cette étude a également permis d'obtenir les modèles de rétention pour l'ensemble de nos lipides avec les 5 colonnes décrites précédemment. De ce fait, nous disposons de plusieurs niveaux d'information dans notre base de données ( V1.1 pour la colonne CSH C18, V2.1 pour la colonne HSS T3, V3.1 pour la colonne zorbax C18, etc.).

Avant d'être un outil de découverte en mode ciblé, le Scout-MRM est avant tout un mode d'acquisition spécialisé dans les études de quantification. L'approche Scout-MRM lipidomique développée au cours de cette thèse pose la question sur la capacité de réaliser à la fois des études de caractérisation et de quantification de manière simultanée.



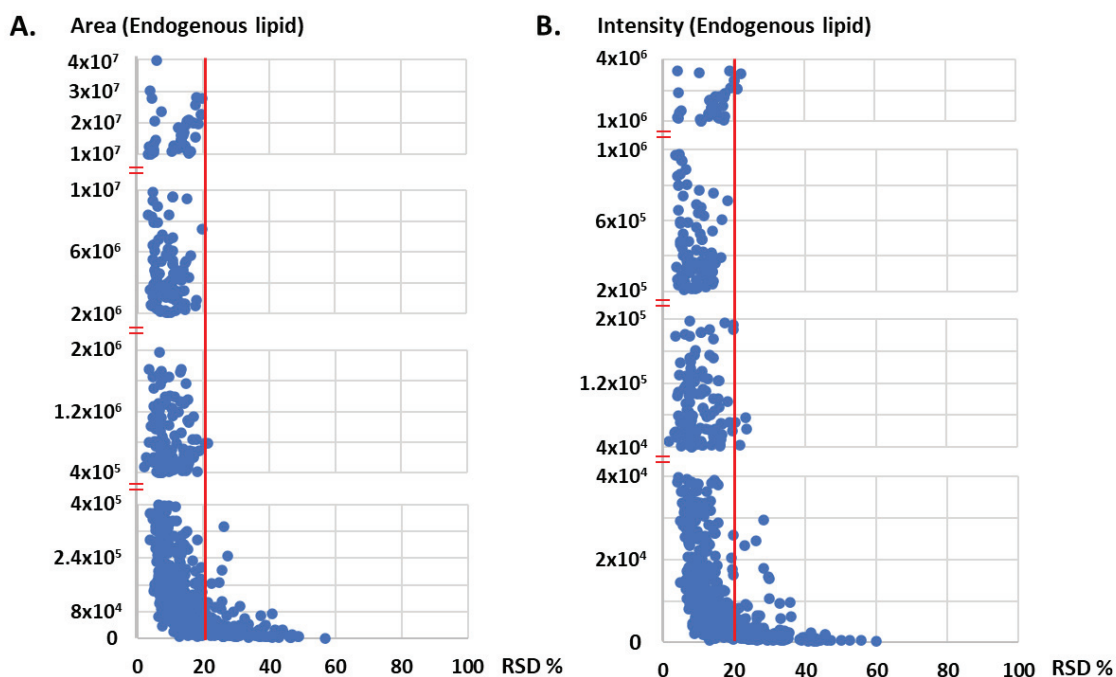
### 3.8. Évaluation du Scout-MRM en tant que méthode de caractérisation et de quantification simultanée pour l'analyse de lipides

La mise en place de méthodes de quantification en lipidomique par spectrométrie de masse nécessite de prendre en considération la variation de la réponse analytique des lipides d'intérêt qui peuvent être considérablement influencés par l'environnement d'ionisation (composition de la phase mobile et phénomène de suppression d'ion). L'ajout de standards de lipides isotopiquement enrichis corrige la variation du signal en spectrométrie masse puisque ces étalons réagiront de la même manière que leurs analogues qui diffèrent par leurs propriétés physiques ( $m/z$ ), mais pas par leurs propriétés chimiques (rétention, ionisation, fragmentation)[302,303]. La mise en place d'un étalonnage interne avec ajout de standards pendant le protocole analytique confère également l'avantage de corriger les pertes d'échantillon notamment lors de l'extraction liquide/liquide. Toutefois, la mise en place de cette méthodologie est limitée en lipidomique à cause de la faible diversité des standards lipidiques disponibles commercialement. C'est pour cela que les études en lipidomique utilisent un standard par famille de lipides [53,289].

Afin de démontrer les performances analytiques de la méthode Scout-MRM, nous avons réalisé plusieurs études pour évaluer la répétabilité, les domaines de linéarité et les limites de quantification (LOQ).

#### 3.8.1. Répétabilité de la méthode Scout-MRM

Dans le but d'évaluer la répétabilité de la méthode Scout-MRM nous avons analysé les lipides endogènes de 10 extractions indépendantes d'une matrice contenant du plasma humain, de la drosophile et du *G. fassarum*. Un standard lipidique par famille de lipides a été ajouté dans chacun des échantillons. La figure n°78 indique que la majorité des transitions MRM ont un coefficient de variation (CV) inférieur à 20%. Les résultats sont similaires dans le cas où la réponse MS est obtenue à partir de l'aire des pics chromatographiques ou de l'intensité du signal de ces mêmes pics. Ce seuil de 20% représente la valeur limite pour laquelle une transition MRM est estimée répétable. Les données pour lesquelles le CV est supérieur à 20% correspondent aux transitions MRM les moins intenses qui impliquent des intégrations plus difficiles et qui sont plus facilement interférées. Ces résultats indiquent que l'étape d'extraction et l'analyse en Scout-MRM sont reproductibles.



**Figure 78 : Répétabilité des lipides endogènes en fonction de l'aire ou de l'intensité du signal.** Résultats obtenus après 10 extractions indépendantes réalisées à partir du même mélange de *G. fassarum*, de plasma humain et de drosophile avec un standard interne pour chaque famille de lipides. Le critère de validation de la répétabilité est fixé à 20%.

### 3.8.2. Domaine de linéarité et LOQ de la méthode Scout-MRM

Afin de déterminer les domaines de linéarité pour chacun des standards lipidiques (un par famille de lipide) des régressions linéaires (pondération  $1/x$  ou  $1/x^2$  en fonction des standards) ont été réalisées à partir de courbes d'étalonnage en triplicat. Ces courbes respectent des critères stricts pour définir les domaines de linéarité (au minimum 5 niveaux de concentration,  $CV < 20\%$ ,  $R^2 > 0.99$  et  $80\% < \text{exactitude} < 120\%$ ). Toutes les courbes d'étalonnage correspondant aux standards de lipides dopés dans des extraits lipidiques de *G. fassarum* à différents niveaux de concentration ont confirmé le haut degré de linéarité de la méthode Scout-MRM ( $0.99 < R^2 < 0.999$  en aire et  $0.989 < R^2 < 0.999$  en intensité ; cf. tableau n°11). Les domaines de linéarité pour chacun des standards sont indiqués dans le tableau n°11. La LOQ a été déterminée comme étant le niveau le plus bas du domaine de linéarité pour lequel le CV est inférieur à 20%. Les LOQ répertoriées dans le tableau n°11 démontrent l'excellente sensibilité de la quantification de lipides par Scout-MRM puisque ces standards lipidiques peuvent être quantifiés à de très faibles concentrations au sein de matrices complexes (de l'ordre du ng/mL au pg/mL). Les résultats obtenus pour les domaines de linéarité et les LOQ sont similaires dans le cas où la réponse MS est obtenue à partir de l'aire des pics chromatographiques ou de l'intensité du signal de ces mêmes pics.

Lipid	POS				NEG			
	Area		Height		Area		Height	
	Linearity range (ng/mL)	LOQ (ng/mL)	Linearity range (ng/mL)	LOQ (ng/mL)	Linearity range (ng/mL)	LOQ (ng/mL)	Linearity range (ng/mL)	LOQ (ng/mL)
PC 11:0/11:0	0.005 – 30 $R^2 = 0.990 ; (1/x^2)$	0.005	0.005 – 30 $R^2 = 0.990 ; (1/x^2)$	0.005	2.9 - 435.16 $R^2 = 0.999 ; (1/x)$	2.9	2.9 - 435.16 $R^2 = 0.999 ; (1/x)$	2.9
PE 12:0/13:0	1.51 - 225.24 $R^2 = 0.993 ; (1/x)$	1.51	1.51 - 225.24 $R^2 = 0.992 ; (1/x)$	1.51	1.25 – 750 $R^2 = 0.998 ; (1/x)$	1.25	1.25 – 750 $R^2 = 0.993 ; (1/x^2)$	1.25
PS 12:0/13:0	3.02 - 449.46 $R^2 = 0.990 ; (1/x^2)$	3.02	3.02 - 449.46 $R^2 = 0.989 ; (1/x^2)$	3.02	3 - 900.43 $R^2 = 0.997 ; (1/x)$	3	6 - 900.43 $R^2 = 0.997 ; (1/x)$	6
PG 12:0/13:0	11.99 - 119.95 $R^2 = 0.992 ; (1/x)$	11.99	11.99 - 119.95 $R^2 = 0.991 ; (1/x)$	11.99	3 - 599.33 $R^2 = 0.997 ; (1/x)$	3	3 - 599.33 $R^2 = 0.997 ; (1/x)$	3
PI 17:0/14:1	15.05 - 300.1 $R^2 = 0.994 ; (1/x)$	15.05	15.05 - 300.1 $R^2 = 0.994 ; (1/x)$	15.05	22.53 - 4499.89 $R^2 = 0.991 ; (1/x^2)$	22.53	22.53 - 4499.89 $R^2 = 0.990 ; (1/x^2)$	22.53
2-LPC 13:0	0.22 - 53.61 $R^2 = 0.991 ; (1/x^2)$	0.22	0.22 - 53.61 $R^2 = 0.990 ; (1/x^2)$	0.22	0.6 - 899.37 $R^2 = 0.997 ; (1/x)$	0.6	0.6 - 899.37 $R^2 = 0.997 ; (1/x)$	0.6
2-LPE 17:1	7.86 - 975.23 $R^2 = 0.991 ; (1/x)$	7.86	9.1 - 1128.74 $R^2 = 0.991 ; (1/x)$	9.81	2.6 - 519.89 $R^2 = 0.996 ; (1/x)$	2.6	2.6 - 519.89 $R^2 = 0.996 ; (1/x)$	2.6
Cer d18:1/12:0	0.46 - 135.16 $R^2 = 0.996 ; (1/x)$	0.46	0.46 - 135.16 $R^2 = 0.996 ; (1/x)$	0.46	10.04 - 301.26 $R^2 = 0.995 ; (1/x)$	10.04	10.04 - 301.26 $R^2 = 0.994 ; (1/x)$	10.04
SM d18:1/18:1 d9	0.15 - 45.05 $R^2 = 0.997 ; (1/x)$	0.15	0.15 - 45.05 $R^2 = 0.997 ; (1/x^2)$	0.15				
DG 17:0/17:0 d5	101.22 - 6007.32 $R^2 = 0.993 ; (1/x)$	101.22	101.22 - 6007.32 $R^2 = 0.993 ; (1/x)$	101.22				
TG 17:0/17:1/17:0 d5	0.76 - 37.57 $R^2 = 0.990 ; (1/x^2)$	0.76	0.76 - 37.57 $R^2 = 0.990 ; (1/x^2)$	0.76				

**Tableau 11 : Domaines de linéarité et limites de quantification (LOQ) pour chaque standard lipidique en ESI (+) et ESI (-) obtenues grâce à l'aire ou l'intensité des signaux détectés.**

Les résultats obtenus lors des études de répétabilité, de linéarité et de limite de quantification mettent en évidence les qualités de la méthodologie Scout-MRM en tant que méthode de quantification. Les performances analytiques dans ces trois études sont très similaires, que l'on utilise l'aire ou l'intensité des pics chromatographiques. Cela confirme que le retraitement des données selon la hauteur du signal est réalisable en Scout-MRM et que dans ce cas de figure il est possible de gagner en capacité de multiplexage puisque les co-élutions entre un lipide d'intérêt et un scout ne sont plus un problème (cf. 3.3.3 et 3.3.4).

À travers les résultats présentés dans ce chapitre de thèse, la méthodologie Scout-MRM en lipidomique permet de réaliser des études de caractérisation d'échantillons complexes. Dans le cas où au moins un standard interne par famille de lipide est introduit dans l'échantillon lors de cette étape de découverte ciblée, il est possible de quantifier directement chacun des lipides découverts lors de cette même analyse. Le passage d'une approche de découverte à une étude de quantification est instantané. De plus, l'adaptabilité du Scout-MRM en fonction des paramètres chromatographiques facilite le transfert de méthodes entre laboratoires et donc la propagation rapide de méthodes de quantification.

### 3.9. Conclusion

Le développement de nouveaux outils innovants comme le mode d'acquisition Scout-MRM a ouvert de nouvelles perspectives de recherche comme la possibilité de réaliser des études de découverte en mode ciblé. Ainsi, nous avons exploré cette approche dans le cadre des analyses de lipides pour proposer de nouveaux outils au service de la biologie des systèmes. L'association du haut degré de multiplexage de la méthodologie Scout-MRM à l'utilisation de bases de données en temps de rétention autorise l'analyse en simultané d'un grand nombre de lipides pour la caractérisation d'échantillons inconnus.

Ces études de découverte en mode ciblé ne permettent pas de caractériser les échantillons avec le même niveau d'information que les approches sans *a priori* par HRMS. En effet, certaines hypothèses concernant la composition des chaînes d'acides gras sont incontournables pour limiter la liste des composés potentiels à rechercher dans un échantillon inconnu. Sans cela les capacités de multiplexage de la méthode Scout-MRM ne pourraient pas couvrir l'ensemble des structures envisagées. Toutefois, la méthodologie Scout-MRM en lipidomique peut être une bonne alternative dans le cas où l'on n'a pas accès à la HRMS pour obtenir rapidement des informations sur un lipidome complexe.

L'ajout de scouts dans les échantillons à analyser permet à la méthode d'être employée sur des matrices et des conditions RPLC variées puisqu'ils assurent le recalibrage automatique de la méthode en la rendant indépendante des variations de temps de rétention. Cela traduit le caractère *plug and play* de notre approche Scout-MRM qui garantit le bon fonctionnement de notre base de données en ordre d'élution dans des conditions autres que celles qui ont été employées pour sa réalisation.

Cette base de données est en perpétuelle évolution puisque plus la méthodologie Scout-MRM sera utilisée par la communauté des lipidomistes, plus il sera possible d'implémenter de nouveaux lipides (ajout de nouveaux RT expérimentaux et amélioration des modèles de prédiction). Cela permettra de réaliser des caractérisations de plus en plus complètes en augmentant la quantité de composés lipidiques pouvant être recherchés dans divers échantillons.

Ces travaux en lipidomique ont permis de mettre en évidence un nouveau type d'approche en Scout-MRM en démontrant la faisabilité de méthodes de caractérisation. Toutefois, ce mode d'acquisition a été développé avant tout pour réaliser des quantifications performantes (robustesse, linéarité et sensibilité). Cela signifie que la méthodologie Scout-MRM appliquée à l'analyse de lipides permet de réaliser en simultané l'étude de découverte en mode ciblé et la quantification des éléments constitutifs du système.

Dans le cadre de cette thèse, la méthodologie Scout-MRM a été appliquée avec succès aux domaines de la protéomique et de la lipidomique. Afin de compléter le panel de méthodes à haut degré de multiplexage dans le cadre des analyses omiques, une méthode Scout-MRM est en cours de développement dans le laboratoire pour l'analyse des métabolites polaires.

À l'issue du chapitre 2 et 3, il a été démontré la possibilité d'analyser en simultanément plusieurs centaines de composés grâce à la technologie Scout-MRM tout en facilitant le transfert de méthode entre laboratoires grâce à l'indépendance aux variations en temps de rétention. L'optimisation de ce mode d'acquisition en MRM a ainsi résolu la première problématique de ces travaux qui était de savoir comment augmenter les capacités de multiplexage en LC-MS/MS sans sacrifier la robustesse de la méthode. Concernant la problématique de réduction des effets matrices, tout en limitant les variations biologiques, l'utilisation du spectromètre de masse seul ne sera pas suffisant. C'est pour cela que nous avons envisagé d'optimiser l'étape de préparation d'échantillon en développant une extraction multi-omique capable d'extraire simultanément les protéines, les lipides et les métabolites à partir d'une prise d'essai unique.

# Chapitre 4 : Développement d'une méthode d'extraction multi-omique en écotoxicologie

## 1. Contexte de l'étude

Le développement des sciences omiques a pris de l'importance dans toutes les facettes des sciences de la vie, et notamment en écotoxicologie. En effet, l'approche omique introduit de nouveaux outils pour analyser les impacts des contaminants sur un écosystème. Parmi les omiques, la protéomique est une science qui analyse et quantifie l'abondance des protéines dans un organisme biologique en réponse à des facteurs de stress externes tels que les maladies, l'exposition à des contaminants ou des déséquilibres nutritionnels. Par rapport à la génomique, la protéomique peut fournir des informations fonctionnelles puisque seule une quantité limitée d'ARNm est traduite en protéines [18]. Toutefois, les protéines ne représentent pas l'ultime niveau de la cascade des omiques puisqu'il s'agit des métabolites. Ainsi, pour établir un lien plus étroit avec les processus physiologiques fonctionnels, il faut utiliser la métabolomique [304], qui est le domaine scientifique qui caractérise les biomolécules endogènes et exogènes de faible poids moléculaire appelées métabolites (inférieur à 1 KDa) dans un organisme. En effet, ces petites molécules sont dérivées des activités des protéines et sont essentielles aux processus de régulation cellulaire. La lipidomique est une sous-discipline de la métabolomique qui étudie plus spécifiquement les métabolites apolaires appelés lipides [18]. L'utilisation d'approches omiques complémentaires (protéomique, métabolomique et lipidomique) améliore l'acquisition d'informations essentielles en corrélant des données issues de plusieurs classes de molécules [24]. Cela assure une meilleure compréhension du phénotype exprimé grâce à des hypothèses fondées sur des analyses multi-omiques fiables. Les données protéiques, lipidiques ou métaboliques peuvent combler un manque d'information dans l'une de ces omiques pour valider ou non une hypothèse lors d'une contamination environnementale [25].

La compréhension globale de l'impact des contaminants environnementaux et/ou de l'état de santé des espèces sentinelles nécessite des interprétations complexes de variations moléculaires à des niveaux multiples tels que le protéome, le métabolome et le lipidome. Les approches multi-omiques, de par leur capacité à étudier un phénomène biologique dans son ensemble, ont la capacité d'améliorer la compréhension du flux d'informations pour identifier et relier les réseaux moléculaires. Ainsi, elles permettent de mieux prédire les résultats en écotoxicologie et de mieux comprendre les aspects mécanistiques des effets chimiques [305,306]. Ces approches multi-omiques au sein de la biologie des systèmes sont sous-exploitées dans le domaine de l'écotoxicologie, bien que l'acquisition

de données résultant de l'analyse d'une seule omique a été améliorée (analyse à haut débit, sensibilité, répétabilité, *etc.*), ce qui rend les études multi-omiques réalisables. Cependant, comme l'ont mentionné Robert Haas *et al.* [307], la combinaison d'ensembles de données issues d'une seule omique s'est avérée jusqu'à présent être un défi, et la découverte de nouvelles informations biologiques est en retard sur les attentes actuelles. L'une des raisons est que les expériences menées dans différents laboratoires ne peuvent généralement pas être combinées sans restriction. Deuxièmement, l'interprétation d'ensembles de données multi-omiques représente un défi important par nature, car les ensembles de données biologiques sont hétérogènes, non seulement pour des raisons techniques, mais aussi pour des raisons biologiques, chimiques et physiques [307]. Idéalement, les protocoles d'extractions multi-omiques doivent être réalisés à partir d'un échantillon unique afin de minimiser les variations interindividuelles. Cela permettra une meilleure corrélation des résultats, car les réponses analytiques ne seront pas lissées par un ensemble de réplicats d'expositions destinées à chacune des omiques. Les autres avantages de la mise en place d'une procédure d'extraction multi-omique unique sont la réduction du coût des solvants, des consommables et du temps de manipulation des échantillons [158], ce qui augmente les analyses à haut débit.

Dans ce 4<sup>ème</sup> chapitre sera décrite la procédure d'extraction multi-omique que nous avons développée pour l'extraction simultanée de composés polaires (protéines, métabolites) et apolaires (lipides) à partir d'un échantillon biologique unique. Le modèle d'étude, à partir duquel l'extraction multi-omique a été développée, est l'espèce sentinelle *G.fossarum*. Il s'agit d'une espèce sensible à un large éventail de polluants [193] qui a suscité un intérêt pour les études de biosurveillance basées sur les sciences omiques au cours de la dernière décennie [65,168,198,215,216,227,308]. En effet, Gouveia *et al.* [168] ont récemment mis en évidence le potentiel de cette espèce sentinelle pour les applications en écotoxicoprotéomique [168]. Par exemple, les réponses moléculaires dues à une exposition aux perturbateurs endocriniens [210,211] ou aux métaux [212] ont été rapportées et ont donné lieu à la découverte de biomarqueurs potentiels liés à des processus physiologiques clés. La métabolomique a également été utilisée pour étudier les changements moléculaires chez les gammares en réponse à l'exposition aux produits pharmaceutiques pour l'identification d'altérations possibles des voies métaboliques [66,235]. De plus, la composition lipidique chez le gammare femelle a été récemment caractérisée, élargissant ainsi notre compréhension des changements biochimiques pendant le développement embryonnaire [309].

Dans le but de prouver que notre méthodologie est capable de fournir une caractérisation exhaustive d'un échantillon à un niveau multi-échelle, l'extraction multi-omique à partir d'un échantillon unique a été appliquée pour l'étude des différents stades de reproduction de l'amphipode femelle *G.*



*fossarum*. Des analyses statistiques multivariées (PLS-DA) et factorielle multiple (AFM), ont été utilisées pour l'intégration des données dans le but de différencier les différents stades.

Les résultats présentés dans ce chapitre sont issus de la publication intitulée « Development of a multi-omics extraction method for ecotoxicology: in depth investigation of the reproductive cycle of *G. fossarum* » soumis dans le journal *Talanta*.

## 2. Publication n°2

### **Development of a multi-omics extraction method for ecotoxicology: in depth investigation of the reproductive cycle of *Gammarus fossarum***

Julien Faugere<sup>a,1</sup>, Thomas Alexandre Brunet<sup>a,1</sup>, Yohann Clément<sup>a</sup>, Anabelle Espeyte<sup>b</sup>, Olivier Geffard<sup>b</sup>, Jérôme Lemoine<sup>a</sup>, Arnaud Chaumot<sup>b</sup>, Davide Degli-Esposti<sup>b</sup>, Sophie Ayciriex<sup>a,1</sup> and Arnaud Salvador<sup>a,1\*</sup>

<sup>a</sup> Université de Lyon, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, CNRS UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France.

<sup>b</sup> INRAE, UR RiverLy, Laboratoire d'écotoxicologie, F-69625 Villeurbanne, France

<sup>1</sup> These authors contributed equally to this work

\* Correspondence may be addressed to [arnaud.salvador@univ-lyon1.fr](mailto:arnaud.salvador@univ-lyon1.fr)

#### 2.1. Abstract

Omics study exemplified by proteomics, lipidomics or metabolomics, provides the opportunity to get insight of the molecular modifications occurring in living organisms in response to contaminants or in different physiological conditions. However, individual omics discloses only a single layer of information leading to a partial image of the biological complexity. Multiplication of samples preparation and processing can generate analytical variations resulting from several extractions and instrumental runs. To get all the -omics information at the proteins, metabolites and lipids level coming from a unique sample, a specific sample preparation must be optimized. In this study, we streamlined a biphasic extraction procedure based on a MTBE/Methanol mixture to provide the simultaneous extraction of polar (proteins, metabolites) and apolar compounds (lipids) for multi-omics analyses from a unique biological sample by a liquid chromatography (LC)/mass spectrometry (MS)/MS-based targeted approach. We applied the methodology for the study of female amphipod *Gammarus fossarum* during the reproductive cycle. Multivariate data analyses including Partial Least Squares

Discriminant Analysis and multiple factor analysis were applied for the integration of the multi-omics data sets and highlighted molecular signatures, specific to the different stages.

### Keywords

Proteomics, Lipidomics, Metabolomics, extraction, LC-MS/MS, multi-omics, data integration, Partial Least Squares Discriminant Analysis (PLS-DA), Multiple Factor Analysis (MFA).

### Highlights

- Optimization of MTBE/MeOH biphasic extraction procedure for multi-omics analysis
- The cuticles affect the extraction efficiency of proteins
- Application of the multi-omics protocol to characterize gammarid reproductive stage
- 152 proteins, 290 lipid species and 79 metabolites were monitored by MRM approach
- PLS-DA identified proteins, lipids, and metabolites changes along ovarian maturation
- MFA was applied for the unsupervised integration of the multi-omics data sets

## 2.2. Introduction

The development of omics technologies is gaining prominence in all aspects of life science, especially in ecotoxicology. Indeed, omics give the opportunity to get a qualitative and quantitative overview of the cellular processes and to monitor in a high throughput manner the molecular changes occurring in living organisms in response to contaminants [6,18,216,304,310–312]. Among the omics panel in ecotoxicology, proteomics is of particular interest since it involves large-scale study of proteins, reflecting direct biological change of the organism's response towards micro-pollutants toxicity [310,313,314]. Applications of proteomics in ecotoxicology, also referred to Ecotoxicoproteomics, were emphasized in several research studies performed mainly in aquatic species, establishing the bases for new-generation proteomics in ecotoxicology [216]. Among the broad panel of sentinel species used in aquatic ecotoxicology, freshwater amphipods, such as *Gammarus fossarum* or *Gammarus pulex*, arouse interest in the last decade in biomonitoring studies. These organisms are recognized species widely used to monitor aquatic pollution in proteomics research studies [50,216,251,315–317]. Molecular responses to endocrine disruptor or metal exposures were reported revealing potential protein biomarkers associated with key physiological processes [210,211,315,318]. Proteomics was successfully applied in field condition experiments to assess quickly and quantitatively the health status of the animals by the simultaneous measurement of a set of proteins and for physiological studies [86,216]. However, these studies revealed only a single layer of information providing a skewed picture of the biological effects induced by different environmental conditions. To gain in-depth insights into the functional physiological processes, a complete examination demands a parallel coverage of a broad spectrum of molecules at other molecular levels such as metabolites and

lipids exemplified by metabolomics and lipidomics, respectively [198,227,319]. With the tremendous progress and improvement of sample preparation and mass spectrometry methods, thousands of proteins, metabolites, and lipids can be extracted and monitored [86,198,316]. However, various sample collection, preparation and processing are usually required for single comprehensive -omics study, leading to numerous caveats arising from the multiplication of experimental extractions and instrumental run. Thus, it would be highly meaningful to get most information at the proteins, metabolites and lipids level coming from the same sample. Multi-omics experiments settled from a unique sample will drive to a better correlation of results as the analytical responses will not be smoothed out by a mixture of samples. Other significant benefits are the conservation of critical samples, reduction of organic solvents, consumables cost and the reduction of sample handling time increasing the throughput of the analyses.

In ecotoxicoproteomics studies applied in wild organisms such as *G. fossarum*, the first step of sample preparation is the homogenization of individuals for protein extraction in presence of lysis buffer containing non-ionic detergent. A delipidation step employing ethanol/diethylether mixture (1/1, v:v) is then performed to remove lipid interference prior to reduction/alkylation and tryptic digestion steps [50,86,222,271,316]. This solvent combination has not been designed for lipid analysis but only for their removal. In a multi-omics context, the lipid extraction step is a key step and must be optimized. Several studies assessed the use of lipid extraction procedure based on chloroform/methanol extraction derived from gold standard protocol such as Folch [143] or Bligh and Dyer [320] recipes, or methyl-tert-butyl ether (MTBE)/methanol mixture [145] to perform multi-omics analysis from a unique biological sample [158,159,321]. The core principle relies on the analysis of the three layers obtained from these biphasic organic extractions: the organic phase (apolar metabolites, mostly lipids), the aqueous phase (polar metabolites) and the non-extractable insoluble matrix (precipitated proteins, DNA, and RNA). In the Folch or Bligh and Dyer procedure, a chloroform/methanol mixture is supplemented to samples resuspended in water or aqueous buffer, or directly to sample that contain enough water content like biofluids for instance [143,320]. In this two-phase partitioning system, the lipids and hydrophobic metabolites are mostly enriched in the lower organic phase because of the higher density of chloroform; the hydrophilic metabolites are contained in the upper aqueous phase. The interphase composed of a precipitated protein disk separates the two layers. The major bottleneck of the protocol resides in the collection of the chloroform fraction through the disk interphase that can eventually pollute the lower organic fraction of insoluble compounds. Conversely, the MTBE protocol is a simple and quick procedure in which the organic phase forms the upper layer because of the low-density of the MTBE solvent [145]. The collection of lipid extracts is greatly streamlined, and drip losses are reduced.

Since this extraction is adapted for high throughput analysis, we rationalized the MTBE procedure including the homogenization step conditions to provide the simultaneous extraction of proteins, metabolites, and lipids for ensuing multi-omics analyses from unique gammarid sample. As a case study, we integrated proteomics, lipidomics, and metabolomics analyses in the study of the reproductive cycles of female *G. fossarum*, using a liquid chromatography (LC)/mass spectrometry (MS)/MS-based targeted approach. Multivariate data analysis including unsupervised analysis (Principal Component Analysis, PCA) and supervised classification method (Partial Least Squares Discriminant Analysis, PLS-DA) were applied and highlighted molecular signatures specific to reproductive stages. Finally, a multiple factor analysis was applied for the integration of the multi-omics data sets to identify the link between the different biological features.

## 2.3. Materials and methods

### 2.3.1. Reagents and chemicals

Water, acetonitrile (ACN) methanol (MeOH), isopropanol (IPA) solvents, ethylic ether and trypsin (TPCK-treated from bovine pancreas) were obtained from Fisher Scientific (LC-MS grade, Strasbourg, France). EDTA, Trizma<sup>®</sup> hydrochloride, sodium hydroxide, glycerol, Triton X-100, iodoacetamide, ammonium formate, dichloromethane (DCM), tert-Butyl methyl ether (MTBE), dithiothreitol, formic acid (FA), sodium chloride, aprotinin, ammonium bicarbonate, and leupeptin hydrochloride were obtained from Sigma Aldrich (St Quentin-Fallavier, France). Absolute ethanol was obtained from CARLO ERBA (Val de Reuil, France). Isotopically labelled peptides containing either a C-terminal [15N2 and 13C6] lysine or arginine were synthesized by Thermo Fisher Scientific (purity > 97 %) and stored at -20°C until use. Isotopically labelled lipids were obtained from Sigma Aldrich (St Quentin-Fallavier, France, purity > 99%). Splash<sup>®</sup> lipidomix<sup>®</sup> mass spec standard, 1,2-diundecanoyl-sn-glycero-3-phosphocholine (11:0/11:0 PC), 1-tridecanoyl-sn-glycero-3-phosphocholine (LPC 13:0), 1-dodecanoyl-2-tridecanoyl-sn-glycero-3-phosphoethanolamine (12:0/13:0 PE), 1-dodecanoyl-2-tridecanoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (12:0/13:0 PG), 1-dodecanoyl-2-tridecanoyl-sn-glycero-3-phospho-L-serine (12:0/13:0 PS), 1-heptadecanoyl-2-(9Z-tetradecenoyl)-sn-glycero-3-phospho-(1'-myo-inositol) (17:0/14:1 PI), 1,3-diheptadecanoyl-glycerol (d5) (17:0/17:0-d5 DG), N-lauroyl-D-erythro-sphingosine (Cer d18:1/12:0), N-oleoyl(d9)-D-erythro-sphingylphosphorylcholine (SM d18:1/18:1-d9), 1,3(d5)-diheptadecanoyl-2-(10Z-heptadecenoyl)-glycerol (17:0/17:1/17:0-d5 TG), 1,3(d5)-dipentadecanoyl-2-(9Z-octadecenoyl)-glycerol (15:0/18:1/15:0-d5 TG) were obtained from Sigma Aldrich (St Quentin-Fallavier, France). Albumin from human serum (purity > 96 %) and Bradford reagent were obtained from Sigma Aldrich (St Quentin-Fallavier, France). Standard metabolites such as amino acids, nucleosides, nucleotides, sugars, amines, and organic acids were obtained from Sigma Aldrich (St Quentin-Fallavier, France, purity > 90-99%). QReSS kit containing labelled metabolites

(stable isotope-labelled molecules), deuterated L-arginine (4,4,5,5-d<sub>4</sub>, purity 97%) and L-serine (2,3,3-d<sub>3</sub>, purity 98%) were obtained from Cambridge Isotope Laboratories Inc. (Tewksbury, Massachusetts, USA).

### 2.3.2. Biological sample collection

Collection and maintenance of *G. fossarum* organisms used for this study were performed according to previous published works [86,198,216]. Briefly, female gammarids were collected and kept in 30L tanks supplied with continuous drilled groundwater under constant aeration and without food supply. The temperature was kept at 12±1°C. After 24h, female gammarids were sorted out at specific reproductive stages according to morphological criteria (n=5 per stage): A/B, C1, C2, D1 and D2 stages [195]. All the organisms were washed with deionized water, weighted, snap freezing in liquid nitrogen and stored at -80°C until use.

### 2.3.3. Evaluation of homogenization conditions

Different conditions were tested to evaluate the best homogenization procedure. The homogenization step was performed either in presence of methanol, lysis buffer with or without a centrifugation step, prior to lipid extraction with MTBE/MeOH.

Entire adult gammarids (n=20, ~0,55 g of crude homogenate) were grinded without solvent with one stainless steel bead (Ø 4 mm) and sample tubes were transferred on ice.

Methanol - crude homogenate (~28 mg) were homogenized with 360 µL of cold methanol with one steel bead per organism (Ø 4 mm).

Lysis buffer without centrifugation – crude homogenates (~28 mg) were homogenized in 300 µL of lysis buffer (Tris HCl 50 mM, 1mM EDTA, Triton X-100 0.1%, 100 mM NaCl, 10 µg/mL leupeptin and aprotinin, pH adjusted to 7.8) with one steel bead per organism (Ø 4 mm).

Lysis buffer with centrifugation – crude homogenates (~28 mg) were homogenized in 600 µL of lysis buffer (Tris HCl 50 mM, 1mM EDTA, Triton X-100 0.1%, 100 mM NaCl, 10 µg/mL leupeptin and aprotinin, pH adjusted to 7.8) with one steel bead per organism (Ø 4 mm). The homogenate was centrifuged at 10,000 g at 4°C during 15 min and 300 µL of the supernatant was used for the two-step extraction with MTBE/MeOH.

For all the conditions, cold MeOH or cold water was added to have a final volume of 660 µL MeOH/H<sub>2</sub>O with a ratio of 2:1.6 (v/v). The samples were agitated for 10 min at 4°C. 1 200 µL of MTBE were added to all the samples and mixed for 10 min at 4°C. After centrifugation at 10 000 g during 5 min at 4°C, the different fractions were prepared as described below.

Lipid analysis - 1 260  $\mu\text{L}$  of the upper organic phase were collected, dried under nitrogen flow.

The dried lipid film was resuspended with 20  $\mu\text{L}$  of DCM/MeOH (1:1, v/v) and afterwards with 180  $\mu\text{L}$  of IPA/ACN/H<sub>2</sub>O (2:1:1, v/v/v). Lipid-containing fractions were diluted 20-fold for negative ion mode and 200-fold for positive ion mode with IPA/ACN/H<sub>2</sub>O (2:1:1, v/v/v). For the tested condition (lysis buffer with centrifugation), the lipid—containing fractions were diluted 10-fold for negative ion mode and 100-fold for positive ion mode with IPA/ACN/H<sub>2</sub>O (2:1:1, v/v/v).

Metabolites analysis - 200  $\mu\text{L}$  of lower phase (aqueous) were collected and diluted 12.3-fold with H<sub>2</sub>O/ACN (10:90, v/v) prior to LC-MS/MS analysis. For the tested condition (lysis buffer with centrifugation), metabolite-containing fractions were diluted 6.15-fold with H<sub>2</sub>O/ACN (10:90, v/v) prior to LC-MS/MS analysis.

Protein analysis - the pellet was resuspended in 250  $\mu\text{L}$  of Tris buffer (Tris HCl 50 mM, 1mM EDTA, 100 mM NaCl, pH adjusted to 7.8, supplemented with leupeptin and aprotinin at 10  $\mu\text{g}/\text{mL}$ ). For the tested condition (methanol and lysis buffer with centrifugation), half of the protein extract was recovered in Tris buffer to perform the protein digestion step and the rest of the proteomics protocol. For the tested condition (lysis buffer without centrifugation), the totality of the protein extract was kept for proteomics.

#### **2.3.4. Reference protocol for proteomics sample preparation**

Proteins from whole gammarids were extracted according to previous published works [50,86,237]. After a SPE clean-up and evaporation, sample was resuspended with 90  $\mu\text{L}$  of a H<sub>2</sub>O/ACN mixture (90:10, v/v) with 0.5% FA, and centrifuged at 12 000 rpm for 5 min at RT, before the LC-MS/MS analysis.

#### **2.3.5. Evaluation of protein fraction after dual protein precipitation procedure**

After MTBE/MeOH, the upper organic phase was removed. 1200  $\mu\text{L}$  of cold MeOH were added to the remaining lower phase to get a ratio of 4:1 (v/v) MeOH/H<sub>2</sub>O and the samples were incubated for 1h at -20°C. Samples were centrifuged at 10 000g during 12 min at 4°C. The supernatant was removed. The remaining pellet was solubilized with 250  $\mu\text{L}$  of Tris buffer (Tris HCl 50 mM, 1mM EDTA, 100 mM NaCl, pH adjusted to 7.8) for further proteomics experiments (protein-containing fraction) as previously mentioned.

#### **2.3.6. Reference protocol for metabolomics sample preparation**

A single gammarid was homogenized with 800  $\mu\text{L}$  of cold MeOH/H<sub>2</sub>O (9:1, v/v) with one stainless steel bead ( $\varnothing$  4 mm) during two cycles (4 m/s for 25 sec) [66]. After a centrifugation step (13 300g, 25 min, 4°C), the supernatant was diluted 5-fold before injection.

### 2.3.7. Recovery and repeatability evaluation during multi-omics extraction procedure

Recovery and repeatability during the multi-omics extraction step was evaluated. Gammarids homogenate (32,5 µL) were spiked with 2 µL of the metabolomics QReSS™ Kit to obtain a 100-fold dilution factor before MTBE/MeOH extraction (n=5).

$$\text{Recovery (\%)} = \left( \frac{\text{ISTD area add before LLE}}{\text{ISTD area add after LLE}} \right) \times 100$$

Repeatability was evaluated during the extraction, and the MS measurement by the analysis of independent samples from the same gammarid pool (n=15 for metabolite, n=7 for protein, n=10 for lipid). Endogenous peptides, lipids and metabolites were monitored. For lipid analysis, one representative lipid standard per lipid class from an internal standards mix (PC 11:0/11:0, LPC 13:0, PE 12:0/13:0, PG 12:0/13:0, PS 12:0/13:0, PI 17:0/14:1, DG 17:0/17:0-d5, TG 17:0/17:1/17:0-d5, Cer d18:1/12:0 and SM d18:1/18:1-d9) was added to the sample before the extraction step.

### 2.3.8. Multi-omics extraction procedure for the biological application

Gammarids were homogenized in 600 µL of lysis buffer (Tris HCl 50 mM, 1mM EDTA, Triton X-100 0.1%, 100 mM NaCl, 10µg/mL leupeptin and aprotinin, pH adjusted to 7.8) with one steel bead per organism (Ø 4 mm). Protein determination assay from the homogenates was performed with BCA protein assay. The homogenate was centrifuged at 10 000 g at 4°C during 15 min and 300 µL of the supernatant was used for the liquid-liquid extraction step (MTBE/MeOH). 360 µL of cold MeOH and 5 µL of deuterated L-serine (120 µg/mL), deuterated L-arginine (6,4 mg/mL) and PC 11:0/11:0-d5 (40 µg/mL) were added and the sample was mixed for 1h at 4°C. Then, 1 200µL of cold MTBE were added and the sample was mixed for 10 min at 4°C. After centrifugation at 10 000 g during 5 min at 4°C, 1 260 µL of the upper organic phase were collected dried under nitrogen flow, resuspended with 20 µL of DCM/MeOH (1:1, v/v) and afterwards with 180 µL of IPA/ACN/H2O (2:1:1, v/v/v). Lipid-containing fractions were diluted 10-fold for negative ion mode and 100-fold for positive ion mode with IPA/ACN/H2O (2:1:1, v/v/v).

For metabolites analysis, 200 µL of lower phase (aqueous) were collected and diluted 6.15-fold with H2O/ACN (10:90, v/v) prior to LC-MS/MS analysis.

For protein analysis, the pellet was resuspended in 250 µL of Tris buffer (Tris HCl 50 mM, 1mM EDTA, 100 mM NaCl, pH adjusted to 7.8, supplemented with leupeptin and aprotinin at 10 µg/mL). Proteins were chemically modified (reduction and alkylation) according to previous published works [50,86,237]. After tryptic digestion, 10 µL of heavy labelled peptides from a stock solution of 4 µg/mL were added. After SPE clean-up, samples were resuspended with 90 µL of a H2O/ACN (90:10, v/v) with 0.5% FA, and centrifuged at 12 000 rpm for 5 min at room temperature [86].



### 2.3.9. Liquid chromatography - Multiple Reaction Monitoring- mass spectrometry (LC-MRM-MS)

For protein analysis, LC-MS/MS analysis was performed on a 6495B Triple Quad LC/MS equipped with a Jet Stream Technology ion source coupled to a 1290 Infinity II series HPLC device (Agilent Technologies, Waldbronn, Germany). The LC separation was carried out on an Xbridge C18 column (100 mm x 2.1 mm, particle size 3.5  $\mu\text{m}$ ) from Waters (Milford, MA, USA) with an injection volume of 20  $\mu\text{L}$ . Elution was performed at a flow rate of 300  $\mu\text{L}/\text{min}$  with H<sub>2</sub>O containing 0.1% formic acid as eluent A and ACN containing 0.1% formic acid as eluent B, employing an isocratic gradient from the beginning at 5% B for 2 min, followed by a linear gradient from 5% B to 35% B in 36 min. Column was washed at 100% B for 5 min and re-equilibrated at 5% B for 5 min. The injection duty cycle was 48 min, considering the column equilibration time. During the analysis, the column is heated to 40°C. Processing, data acquisition and instrument control were performed using MassHunter software®. MS analysis was performed in positive ionization mode using a capillary voltage of 4500 V and a nozzle voltage of 0V. The iFunnel parameters, high- and low-pressure RF were set to 145 V and 115 V respectively. The nebulizer was set at 40 psi. The gas flow and the sheath gas flow were set at 17 and 12 L/min respectively using nitrogen. The gas temperature and the sheath gas temperature were set at 210°C and 250°C respectively. The acquisition mode is the dynamic MRM mode with time windows between 2 and 4.57 min. The mass-to-charge ratio in Q1 and Q3 as well as the collision energy (CE) values were predicted using Skyline for endogenous peptides. The duty cycle is set at 1.5 s to get ten data points per chromatographic peak for each MRM transition.

For lipid analysis, LC-MS/MS analysis was performed on a QTRAP® 5500 LC-MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Sciex) equipped with a Turbo V™ ion source coupled to an 1290 series HPLC device (Agilent Technologies, Waldbronn, Germany). The LC separation was carried out on an XSelect CSH C18 column (100 mm x 2.1 mm, particle size 3.5  $\mu\text{m}$ ) with an injection volume of 5  $\mu\text{L}$  (Waters, Milford, MA, USA). Elution was performed at a flow rate of 500  $\mu\text{L}/\text{min}$  with H<sub>2</sub>O/ACN (60:40, v/v) containing 0.1 mM of ammonium formate and 0.1% formic acid as eluent A and IPA/ACN/H<sub>2</sub>O (90:9.5:0.5, v/v/v) containing 10 mM of ammonium formate and 0.1% formic acid as eluent B, employing a linear gradient from 1% B to 100% B in 24 min. Column was washed at 100% B for 2 min and re-equilibrated at 1% B for 3 min. Taking into consideration the column equilibration time, the injection duty cycle was 30 min. The column was heated to 70°C. MS analysis was performed in positive ionization mode using an ion spray voltage of 5500 V and in negative ionization mode using an ion spray voltage of -4500 V. The nebulizer and the curtain gas flow were set at 25 and 50 psi respectively using nitrogen. The TurboV™ ion source was set at 450°C with the auxiliary gas flow (nitrogen) set at 40 psi. The mass in Q1 and Q3 as well as the declustering potential (DP) and the

collision energy (CE) values have been optimized by direct infusion into the mass spectrometer carried out from Splash® Lipidomics®. The duty cycle is set at 1s to get 15 data points per chromatographic peak for each MRM transition.

For metabolite analysis, LC-MS/MS analysis was performed on a QTRAP® 5500 LC-MS/MS System hybrid triple quadrupole/linear ion trap mass spectrometer (Sciex) equipped with a Turbo V™ ion source coupled to an 1290 series HPLC device (Agilent Technologies, Waldbronn, Germany). The LC separation was carried out on an XBridge BEH amide column (100 mm × 2.1 mm, particle size 2.5 µm; Waters) with an injection volume of 5 µL in ESI+ and 10 µL in ESI-. Elution was performed at a flow rate of 300 µL/min with H<sub>2</sub>O containing 10 mM of ammonium formate and 0.15% formic acid as eluent A and ACN:H<sub>2</sub>O (90:10, v/v) containing 10 mM of ammonium formate and 0.15% formic acid as eluent B, employing a linear gradient from 0% A to 55.6% A in 11 min. Column was washed at 55.6% A for 2 min and re-equilibrated at 1% B for 6 min. Taking into account the column equilibration time (4 min), the injection duty cycle was 23 min. MS analysis was performed in positive ionization mode using an ion spray voltage of 5500 V and in negative ionization mode using an ion spray voltage of -4500 V. The nebulizer and the curtain gas flow were set at 25 and 50 psi respectively using nitrogen. The TurboV™ ion source was set at 450 °C with the auxiliary gas flow (nitrogen) set at 40 psi. Scheduled MRM mode was performed with time windows of 4 min in positive mode and 6 min in negative mode. The duty cycle is set at 0,8 s to get twelve points per chromatographic peak for each MRM transition. Processing, data acquisition and instrument control were performed using Analyst 1.6.3 software®.

### 2.3.10. Data analysis

MRM data processing for lipid and metabolites analyses were performed with MultiQuant™ analysis software (vs 2.1.1., Sciex). Targeted proteomics data were processed with MassHunter Quantitative Analysis software (vs B.09.00, Agilent).

Data were analyzed using function on ropls package for statistical analyses under R program [322]. Principal Components Analyses (PCA) were performed on all samples from the different omics dataset to visualize difference among the different sample condition. For data normalization, EigenMS was used [323]. Models were calculated using Partial Least Square Discriminant Analysis (PLS-DA). Variable selection was performed using VIP scores and S-plot. Statistical significance of models was assessed using confusion matrix and by permuting the validation set. Significance of variables was calculated by multiple testing correction of the p-values (Benjamini-Hochberg False Discovery Rate). Multiple Factor Analysis (MFA) was applied to all the matrices to determine the differences and similarities between the groups, the variables, and individuals. MFA was performed using the R package FactoMineR.

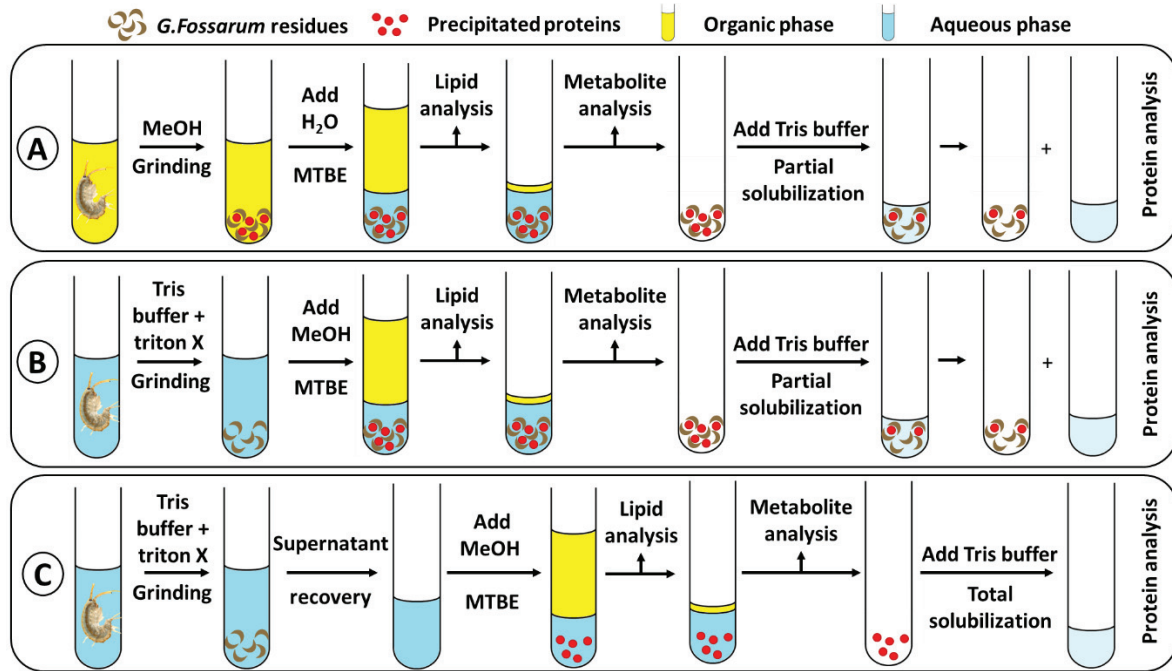
## 2.4. Results and discussion

To routinely perform multi-omics analysis in sentinel species widely used in ecotoxicology experiments, we evaluated and adapted the two-step liquid-liquid extraction protocol using MTBE/MeOH mixture, commonly used in lipidomics, to analyse the proteins, lipids, and metabolites fraction in a sentinel species *G. fossarum*. In this procedure, we carefully examined and optimized crucial step including homogenization step to extract the maximum of proteins, lipids, and metabolites in the presence or not of gammarid exoskeleton.

### 2.4.1. Comparison of protein extraction methods

One crucial point to evaluate in the multi-omics protocol is firstly the homogenization step and more specifically the protein extraction condition. Indeed, gammarids are crustacean amphipods protected by an exoskeleton composed of complex polysaccharide called chitin, and calcium. This exoskeleton confers hardness properties requiring the use of mechanical disruption and buffer containing detergent. After mechanical disruption, cuticle residues are mixed in the grinding raw materials and can interfere with proteins after their denaturation and precipitation with organic solvents.

We investigated three conditions prior to the two-step lipid extraction protocol (MeOH/MTBE/H<sub>2</sub>O) (Figure n°79). Gammarids were homogenized directly (i) in presence of methanol (MeOH) or (ii) in a lysis buffer supplemented with a non-ionic detergent with or without a centrifugation step prior to lipid extraction procedure. MeOH is an amphiprotic solvent commonly used for lipids and metabolites extraction [320,322]. The lysis buffer is composed of Tris buffer supplemented with Triton X-100 largely used in proteomics sample preparation [50,86,222,271,316]. In presence of MeOH during the homogenization step, the proteins are denatured, precipitate and thus mixed with non-solubilized residues. We have evaluated the impact of keeping or removing the cuticle residues on the molecules (proteins, lipids, metabolites) yield (Figure n°79).



**Figure 79 : Workflow of the different proteins, lipids, and metabolites extraction conditions from *Gammarus fossarum*.** The first condition corresponds to the gammarid homogenization with methanol (A), followed by water and MTBE extraction (two-step lipid extraction). The second condition corresponds to the sample homogenization with lysis buffer (B) followed by the addition of MTBE and methanol. The last condition (C) corresponds to the homogenization of the organism in lysis buffer followed by a centrifugation step prior to lipid extraction (MTBE/MeOH).

At the end of the three conditions, the protein fraction was firstly analyzed using a liquid chromatography/mass spectrometry (MS)/MS-based targeted proteomics method corresponding to the selective monitoring of 263 reporter peptides corresponding to 152 proteins of gammarids that have been validated in previous studies [50,86,222]. As shown in the heat map obtained from the peptide areas comparison (Figure n°80), the sample homogenization in a lysis buffer including a centrifugation step produced better results for protein extraction which is reflected by the number of monitored peptides. A poor extraction condition was noticed with the use of methanol and lysis buffer without centrifugation step (Appendix n°6). For only 10% of the reporter peptides, the sample homogenization in lysis buffer including a centrifugation step appeared to be less efficient for protein extraction (Appendix n°6). For 17% of the reporter peptides, this protocol allowed the extraction up to 5 times more reporter peptides and in consequence more proteins. In 73% of cases, the area ratio was greater than 5, and this ratio exceeded a factor of 50 in 18% of the cases. As expected, methanol is not the most efficient method as it precipitates the proteins together with cell residues including exoskeleton.



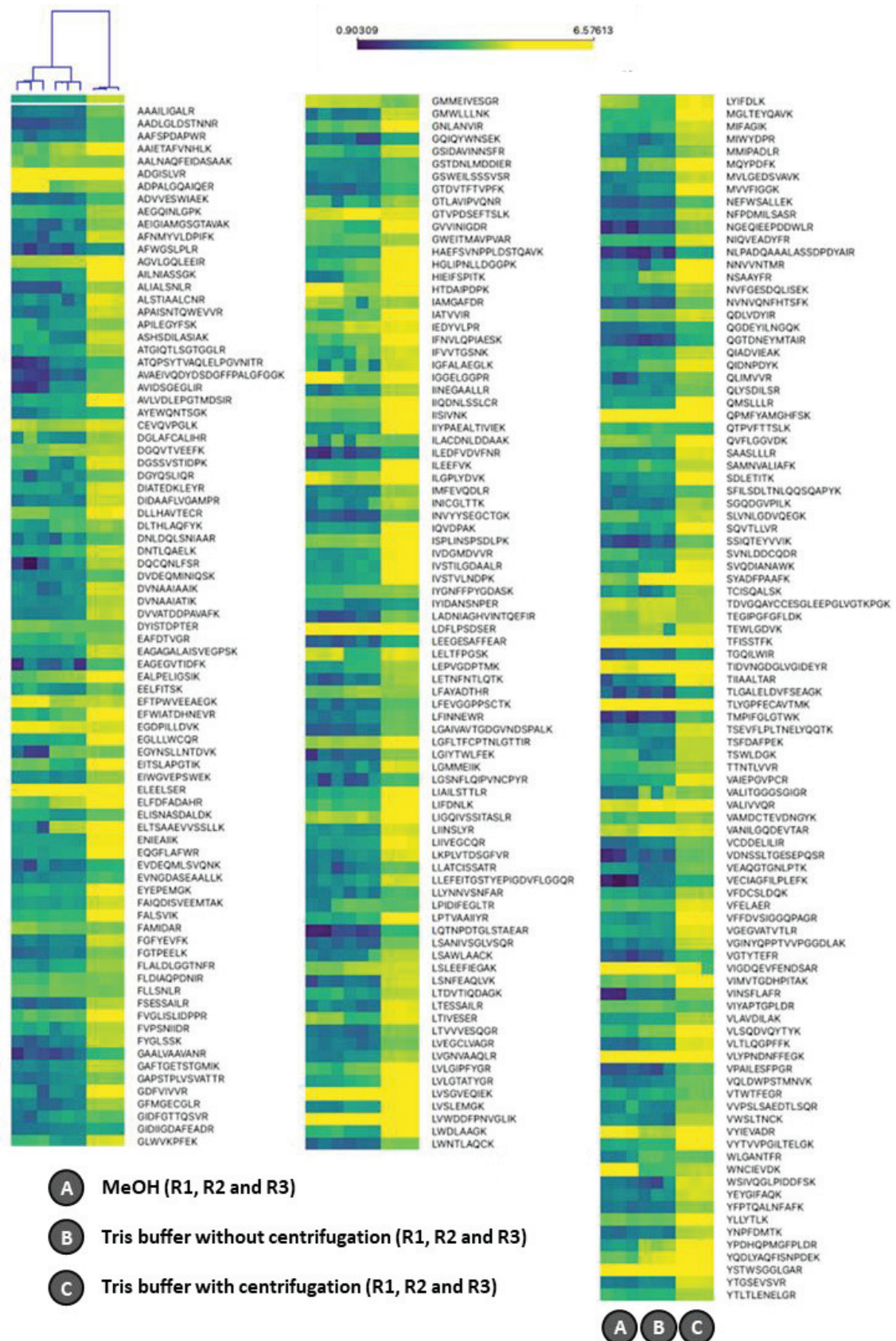
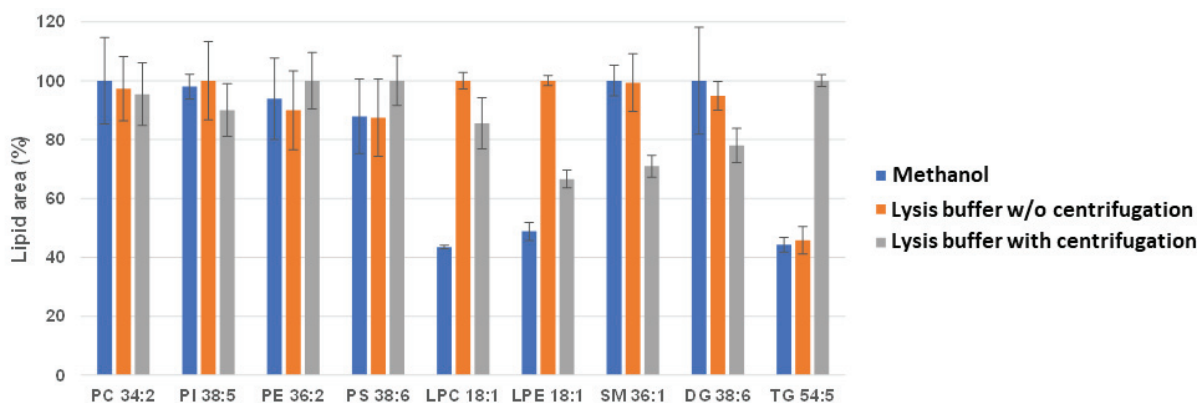


Figure 80 : Heatmap view of peptides monitored by Scheduled-MRM according to the different protein extraction conditions: (A) methanol (MeOH), lysis buffer including (C) or not a centrifugation step after the extraction (B). Yellow color indicates a higher abundance level. Dark blue indicates a lower abundance level. Displayed data are log10 transformed. Three biological replicate per condition have been analyzed. The monitored peptides are listed on the right.

For efficient protein extraction, the centrifugation step is required to get rid of gammarid residues or insoluble materials and for a better solubilization of proteins. A prolonged agitation didn't help to extract more proteins. In conclusion, the procedure with lysis buffer together with a centrifugation step was preferred. The impact of this procedure on the main gammarid lipid class was further evaluated.

#### 2.4.2. Evaluation of homogenization condition on lipids

Targeted LC-MS/MS analysis in positive and negative ionization mode was performed on the organic phase (lipid fraction) to evaluate the best condition for lipid extraction recovery. The main lipid classes i.e. phosphatidylcholine (PC), phosphatidylethanolamine (PE), lysophosphatidylcholine (LPC), lysophosphatidylethanolamine (LPE), phosphatidylinositol (PI), phosphatidylserine (PS), sphingomyelin (SM), diacylglycerol (DG), triacylglycerol (TG) were targeted by Scheduled-MRM including 290 lipid species (429 MRM transitions in positive ion mode and 294 MRM transitions in negative ion mode) based on their different fatty acid moieties or polar head group. The list has been established based on shotgun lipidomics data [198]. As illustrated in figure n°81, the following species PC 34:2, PI 38:5, PE 36:2, PS 38:6, SM 36:1 as well as DG 38:6 were similarly extracted regardless the homogenization conditions. This result was representative of the studied lipid species within a given class.



**Figure 81 : Comparison of the homogenization conditions on the lipid extraction yield. After homogenization of entire gammarids in methanol, lysis buffer with or without centrifugation, lipids were extracted according to the MTBE/MeOH protocol. The reconstituted lipid extracts were analyzed by LC-MS/MS (scheduled-MRM acquisition mode) and peak area for each lipid species was extracted. Data were expressed as % peak area ratio.**

Unsurprisingly, methanol proved to be suitable for lipids solubilization. The lysis buffer containing Triton X-100 improved the extraction of lysophospholipids (LPC 18:1, LPE 18:1). However, a slight decrease was observed when the lipid extraction was performed on a clear homogenate without chitin residues. Higher amount of triacylglycerols (TG 54:5) was recovered after homogenization in lysis buffer followed by a centrifugation step. To verify this observation, synthetic lipid standards mix containing PC 22:0 (PC 11:0/11:0) and TG 51:1-d5 (TG 17:0/17:1/17:0-d5) were spiked to the gammarid

homogenates in presence or absence of gammarid residues after the MTBE/MeOH extraction process (Appendix n°7). Our assumption is that during the liquid-liquid extraction, TG species can adsorb onto the cuticle residues leading to bad recovery. No significant signal variation was observed for PC standard. On the contrary, the signal of the synthetic TG standard and the TG lipid species was increased.

### 2.4.3. Evaluation of the metabolite fraction

We further explore the homogenization conditions on the analysis of the metabolite fraction (aqueous phase). 79 endogenous metabolites including labelled metabolites were specifically targeted in positive (247 MRM transitions) and negative (108 MRM transitions) ion mode, representing 10 categories of metabolites (Appendix n°8). 76 metabolites were detected among the samples. Depending on their physicochemical properties, some metabolites were efficiently extracted with MeOH or lysis buffer. About 47% of metabolites corresponding to polar metabolites such as acylcarnitine, amines, amino acids, indoles, nucleotides, and nucleosides, were more abundant in lysis buffer with a centrifugation step. Methanol and lysis buffer without centrifugation procedure exhibited extraction efficiency for about 32% and 21%, respectively. Concerning methanol extraction, 32% corresponded to apolar metabolites: fatty acids, carbohydrates, vitamins, and some low polar amines, amino acids and nucleotides/nucleosides. The signal decrease for 21% of the metabolites in favor of the protocol without centrifugation, that can be solely due to the additional centrifugation step where some of the compounds are lost when the gammarid residues are removed. We can speculate that some metabolites were adsorbed on the chitin surface.

A more detailed comparison (Figure n°82) between methanol and lysis buffer with centrifugation condition showed that the procedure with methanol is significantly better for only 24% of the metabolites (gain > 20%). Whereas for 76% of the metabolites, the extraction is similar (31%) or higher (45%) regarding the lysis buffer with centrifugation procedure. 88% of the metabolites were recovered in the lysis buffer condition including or not a centrifugation step. Only 12 % of metabolites were affected by the presence of the residues. From these results and the conclusion drawn on the proteins and lipids data, the lysis buffer condition including a centrifugation step before the two-step extraction, was retained for the rest of the study.



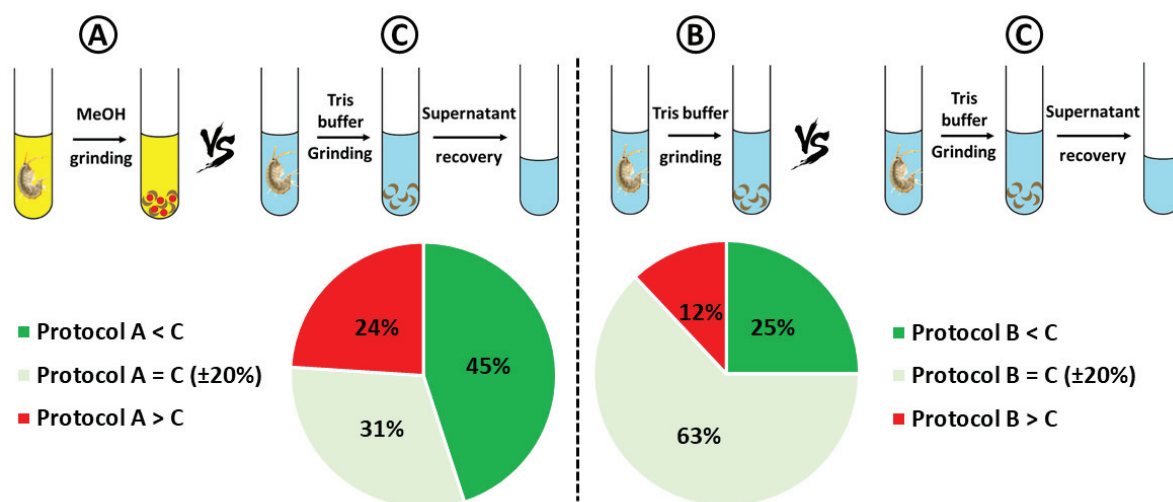


Figure 82 : Comparison of protocol “lysis buffer with centrifugation” (C) versus protocol “MeOH” (A) and “lysis buffer without centrifugation” (B) in metabolite analysis.

#### 2.4.4. Comparison of multi-omics extraction method versus single omics

To explore whether the biphasic extraction procedure based on a MTBE/MeOH mixture was compatible with proteins and metabolites analyses applied on *G. fossarum*, we compared classical protocols that we used routinely for proteomics and metabolomics with MTBE/MeOH multi-omics procedure.

##### 2.4.4.A. Protein's fraction

We measured the concentration of total proteins after MTBE/MeOH extraction and our reference protocol used for proteomics, using a dye-binding method based on Bradford assay. Briefly, after protein extraction with lysis buffer containing Triton X-100, a delipidation step with ethanol/diethylether mixture (1:1, v/v) is usually performed before denaturation, cysteine reduction and alkylation, and tryptic digestion. The composition solvent mixture is not optimal for accurate lipid analysis but only for rough lipid elimination.

1.78 mg/mL of total protein was estimated in the reference protocol compared to 1.32 mg/mL with the MTBE/MeOH extraction, corresponding to approximately 26% less proteins than the reference (Appendix n°9).

Although the amount of protein extracted is lower, a precise inspection is mandatory at the peptide levels. Indeed, the biphasic extraction process with MTBE/MeOH is more advanced reducing matrix effects and interference. We performed the extraction and analysis on the same pool of gammarid samples with both protocols, with the same scheduled-MRM analysis as aforementioned. The results indicated that 49% of the peptides areas didn't show any significant area variation between both protocols ( $\pm 20\%$ ) (Figure n°83 A). For 31% of peptides, a decrease of the signal variation was between

20 and 30%. For 15% of peptides, the signal decrease ranged from 30 and 50%. Only 5% of peptides exhibited a decrease of the signal by a factor 2 (> 50%) using MTBE/MeOH mixture (Figure n°83 A).

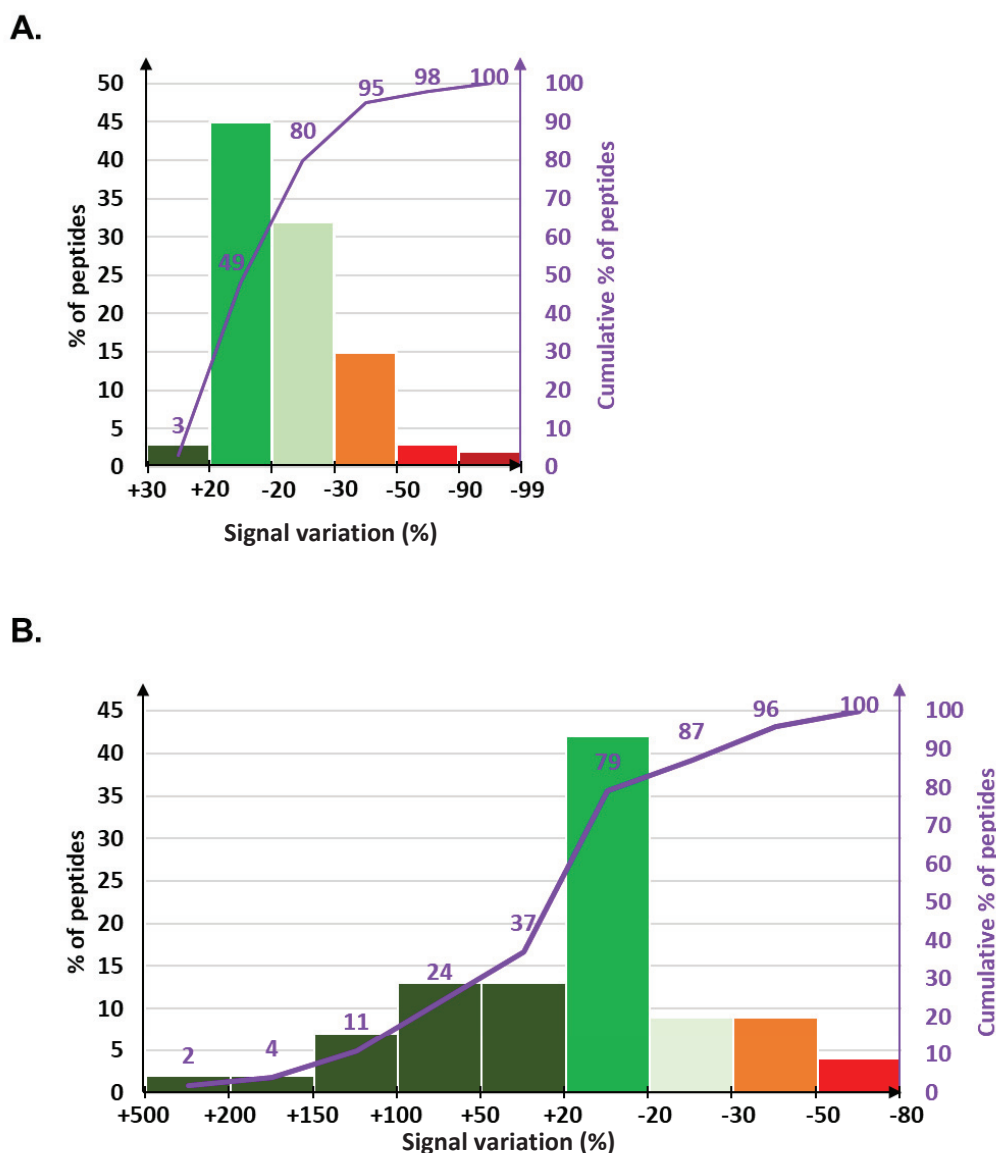


Figure 83 : Signal variations of peptides by LC-MS/MS when switching from the classical proteomic protocol to the multi-omic protocol. The purple curve indicates the cumulative % of peptide. (A) multi-omic vs classical proteomic. (B) multi-omic with dual precipitation vs classical proteomic.

To improve the amount of protein extracted, a second precipitation step with methanol was evaluated. Briefly, after the organic phase extraction, cold methanol was added to a final ratio of MeOH/H<sub>2</sub>O (4:1, v/v) and allowed a precipitation step during 1h at -20°C. The methanolic phase was centrifuged, evaporated for at least 5h and used for metabolites analysis. A total protein amount of 1.57 mg/mL was measured (12% less than our classical proteomics protocol). As reported in figure n°83 B, we observed that 79% of peptide exhibited similar or higher area. The gain in signal areas was not negligible since 37% of peptides exhibited a signal increase between 20% and 500% compared to our conventional protocol. The dual precipitation procedure together with MTBE/MeOH seems to

eliminate more interferences probably due to a more efficiency extraction of apolar compounds, alleviating matrix effects during electrospray ionization. Since additional precipitation step followed by sample evaporation lengthens the procedure, we did not pursue in this study.

Even if less proteins were extracted with MTBE/MeOH extraction protocol in comparison to the classical proteomics sample preparation, multi-omics procedure remains suitable for high throughput proteins analysis.

#### 2.4.4.B. Metabolite's fraction

Likewise, we evaluated the influence of the multi-omics protocol on the metabolite's extraction.

Depending on their polarities and partition coefficient, some of the targeted metabolites could distribute between the aqueous phase and the organic phase. To verify this hypothesis, labelled metabolites with different polarities were added before or after the extraction step. The recoveries results are presented in table n°12. Except, for two metabolites (Thymine  $^{15}\text{N}_2$  and Vitamin B3  $^{13}\text{C}_6$ ), the yields are satisfactory.

Metabolite	Metabolite chemical class (HMDB/KEGG)	Intraday precision (RSD %)	Recovery (RSD %)	LogP
Ethanolamine D <sub>4</sub>	Amines and derivatives	17	98	-1,31
Putrescine $^{13}\text{C}_4$	Amines and derivatives	18	109	-0,70
Alanine $^{13}\text{C}_3$	Amino acids, derivatives and peptides	16	76	-2,85
Creatinine D <sub>3</sub>	Amino acids, derivatives and peptides	3	86	-1,76
Leucine $^{13}\text{C}_6$	Amino acids, derivatives and peptides	7	81	-1,52
Phenylalanine $^{13}\text{C}_6$	Amino acids, derivatives and peptides	3	76	-1,38
Tryptophan $^{13}\text{C}_{11}$	Amino acids, derivatives and peptides	5	62	-1,05
Tyrosine $^{13}\text{C}_6$	Amino acids, derivatives and peptides	8	82	-2,26
Indole-3-acetic acid $^{13}\text{C}_6$	Indoles and derivatives	7	116	1,41
Guanosine $^{15}\text{N}_5$	Nucleotides, nucleosides and derivatives	18	89	-1,90
Hypoxanthine $^{13}\text{C}_5$	Nucleotides, nucleosides and derivatives	4	61	-1,11
Thymine $^{15}\text{N}_2$	Nucleotides, nucleosides and derivatives	5	40	-1,00
Vitamin B3 $^{13}\text{C}_6$	Vitamins and derivatives	4	42	-0,40

**Tableau 12 : Evaluation of QRESS metabolites recovery and extraction efficiency.** RSD, Relative Standard Deviation. HMDB, Human Metabolome Data Base. KEGG, KEGG compound database.

As previously described, the same pool of gammarids was used and extracted either with the multi-omics protocol or the classical metabolomics protocol. 76 metabolites were detected and certified by LC-MS/MS criteria. 72% of the metabolites exhibited equal or higher peak area with the MTBE/MeOH protocol (Figure n° 84). Among the metabolites showing an increased signal, 37% of them pointed an increase of at least 50% to over 200% for 13% of them. Only 28 % showed a significant decrease in intensity (more than 20%) with only 5% of them displaying a decrease between 50% and 80%.

Regardless of these results, less intense compounds were still measurable such as 6-phosphogluconic acid and docosahexaenoic acid which exhibited the most significant loss (approximately ~33-fold decrease). On the other hand, the sensitivity was increased for a significant number of metabolites such as serine, anserine and arginine which exhibited 6 to 9-fold increase. For less intense metabolites, our procedure will not completely supplant classical metabolomics extraction. However, this multi-omics protocol allows the collection of a relevant metabolomics profiles along with the proteomics and lipidomics ones from the same biological sample.

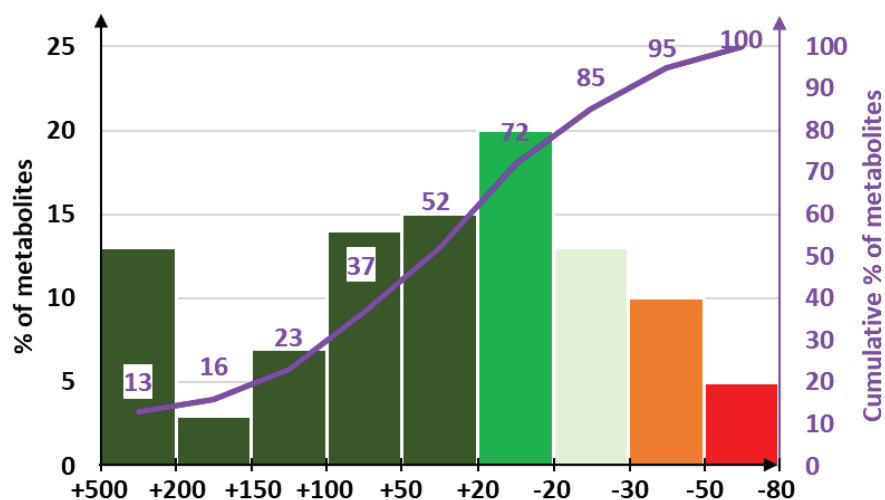


Figure 84 : Signal variations of metabolites by LC-MS/MS when switching from the classical metabolomic protocol to the multi-omic protocol. The purple curve indicates the cumulative % of metabolites.

#### 2.4.5. Repeatability

Assessing the repeatability of the whole analytical workflow (extraction, MS measurement) is a critical first step for improving the reliability of downstream analyses of targeted multi-omics experiments. The repeatability was evaluated by evaluating independent samples from the same pool of gammarids performed and using endogenous peptides, lipids, and metabolites from *G. fossarum* (Figure n°85). In the case of lipids, only one synthetic lipid standard per lipid class was added before the extraction step. As shown in figure n°85, the repeatability of most of the MRM transitions for peptides, lipids and metabolites were categorized acceptable since the % RSD is less than 20%. When the criterion is not met, the MRM transitions areas with RSD > 20% were considered as qualifier and not quantifier transition. These MRM transitions are more easily interfered and therefore more difficult to integrate.

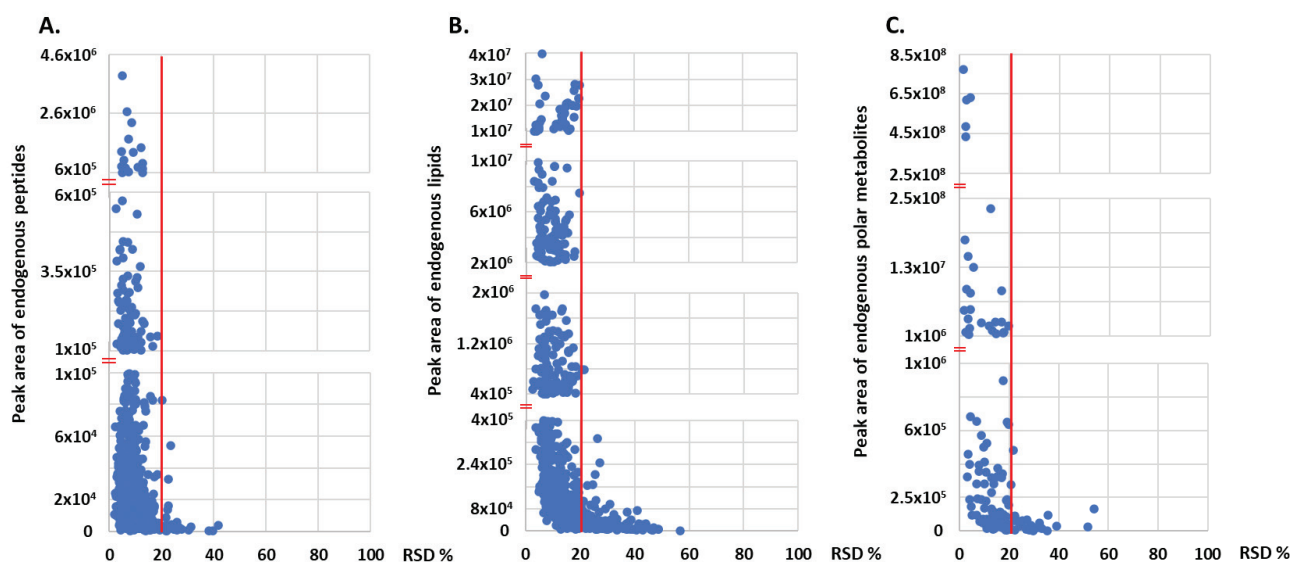
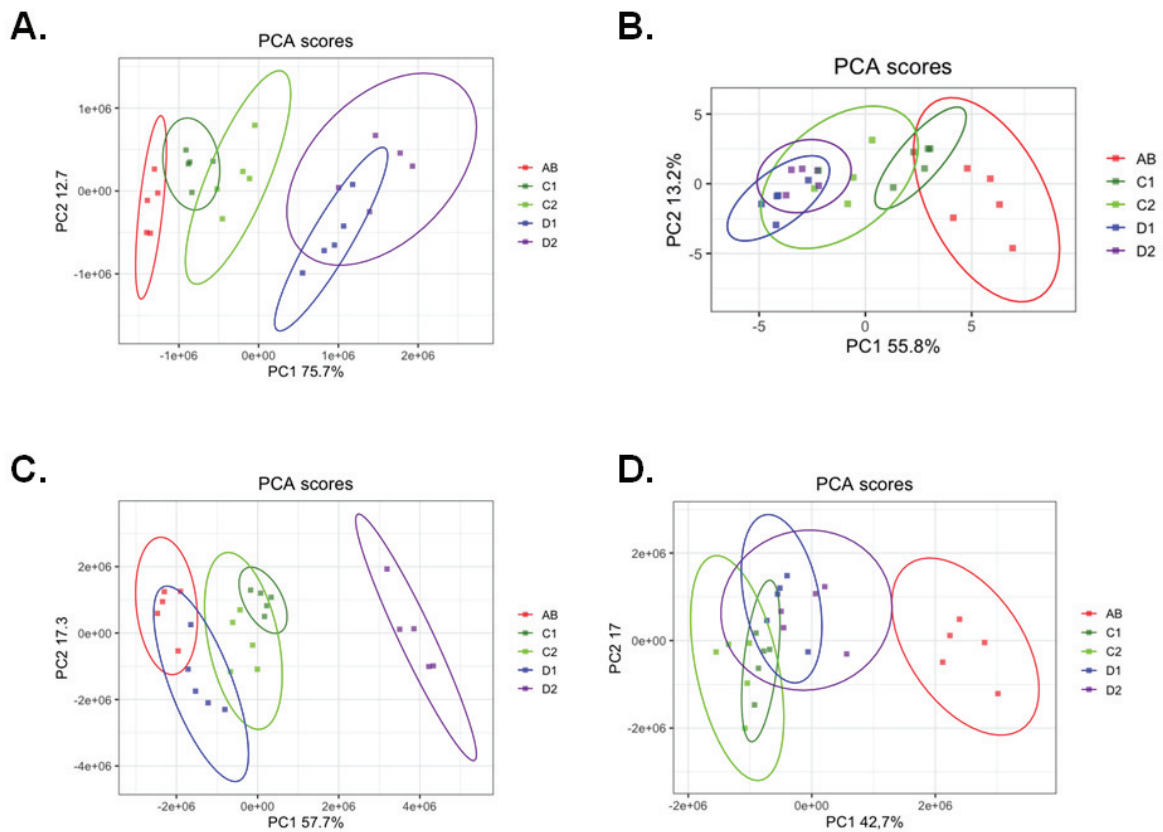


Figure 85 : Evaluation of repeatability by the measurement of endogenous peptides (A), lipids (B), and metabolites (C) from *G. fossarum*.

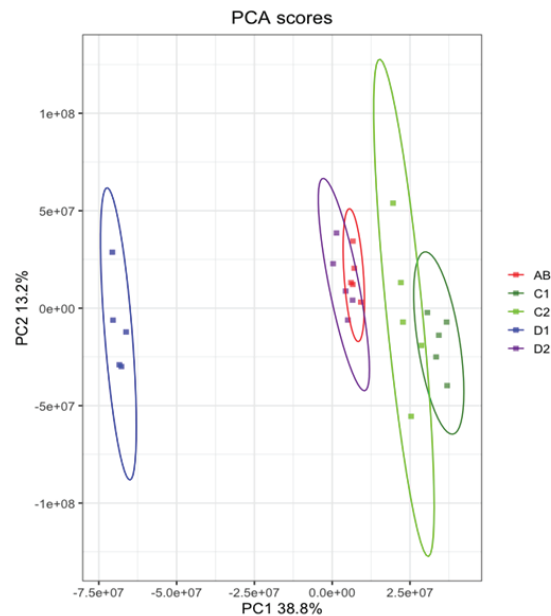
#### 2.4.6. Multi-omics method application

We investigated our multi-omics procedure to observe molecular differences during the reproductive stage of female gammarids. The general sequence of reproductive events in *G. fossarum* is well known and has been described by Geffard et al. [195]. Briefly, the reproductive and molting cycles in female gammarids are closely synchronized [194–196]. Six molt stages defined as postmolt (A, B), intermolt (C1, C2), and premolt (D1, D2) have been characterized. After sample collection of female gammarids at specific reproductive cycle, we applied our multi-omics procedure followed by targeted proteomics, lipidomics and metabolomics LC-ESI-MS/MS analysis (scheduled-MRM approach).

The omics data sets were processed under R using the *ropls* package for the multivariate data analysis. Before establishing quantitative model using Partial Least Square Discriminant Analysis (PLS-DA), descriptive method like Principal Component Analysis (PCA) was used to get information among samples such as sample clustering, differences, or outliers. As depicted by figure n°86, four PCA score plots were obtained by the different omics data sets. For proteomics, the PCA score plot revealed five clusters corresponding to the separation of the different A/B, C1, C2, D1 and D2 stages. Same trends were obtained for ESI(+) in lipidomics and ESI(-) in lipidomics and metabolomics. No striking differences among groups AB versus D2, have been obtained for metabolites in ESI(+) while clear separations were observed for the other groups (Figure n°87).

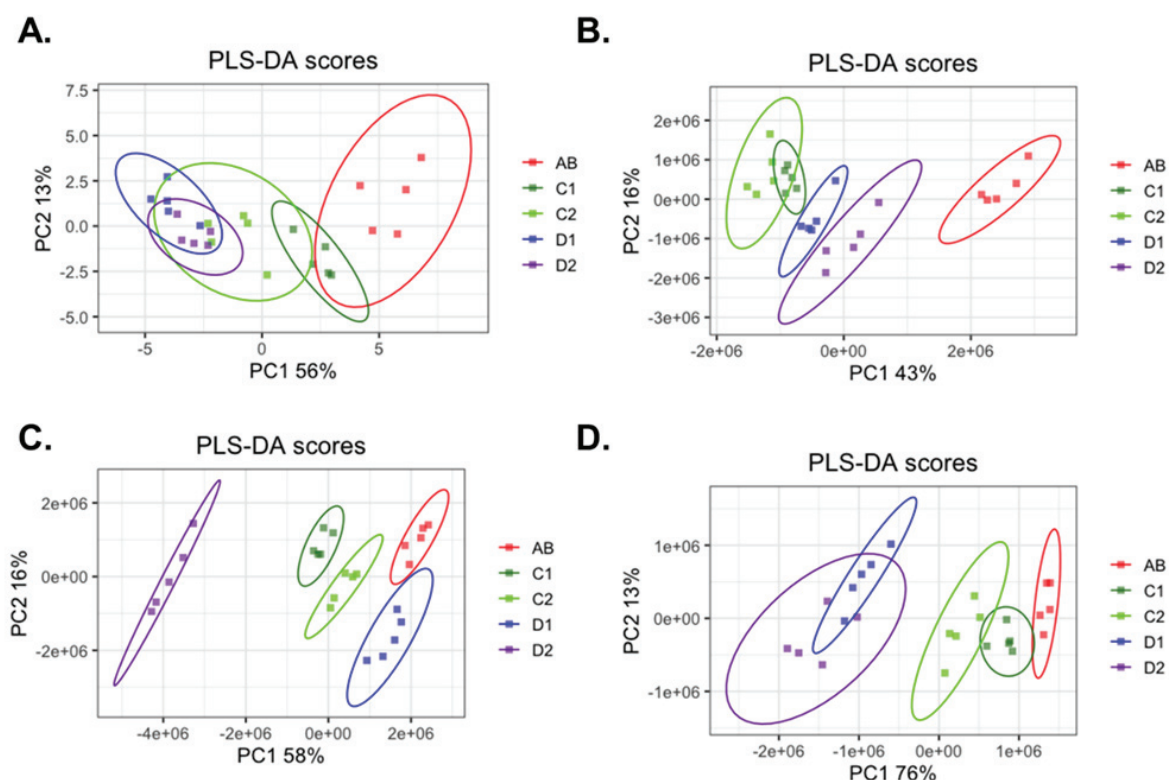


**Figure 86 :** PCA score scatter plots based from the multi-omics data set, namely proteomics data set acquired in ESI(+) (dynamic-MRM) (A), from metabolomics data set acquired in ESI(-) (B), from lipidomics data set acquired in ESI(+) (C), and ESI(-) (D) by LC-ESI-MS/MS (scheduled-MRM). Each dot represents a gammarid organism sample at a specific reproductive stage (AB stage in red, C1 in dark green, C2 in light green, D1 in blue and D2 in purple). The dot distribution shows clustering in each group.



**Figure 87 :** PCA score scatter plot from metabolomics data set acquired in ESI(+) by LC-ESI-MS/MS (scheduled MRM). Each dot represents a gammarid organism sample at a specific reproductive stage (AB stage in red, C1 in dark green, C2 in light green, D1 in blue and D2 in purple). The dot distribution shows clustering in each group. Group AB and D2 exhibits sample overlap. On the contrary, C1 and D1 exhibits clear separation. The PCA score plot corresponding to the first two principal components PC shows that PC1 and PC2 explain 38.8 and 13.2% of the dataset total variance, respectively.

To get the pertinent variables responsible for the separation, PLS-DA models were constructed. The quality of the models was assessed by bootstrap resampling. R2 values represents goodness-fit-measure and Q2 values estimated the predictive ability of the model. These values were used to assess model performance and to select the optimal number of principal components (PCs). Therefore, four PLS-DA models were built (Figure n°88).



**Figure 88** : PLS-DA score scatter plots from the multivariate analysis of the multi-omics data sets: proteomics ESI(+) (A), metabolomics ESI(-) (B), lipidomics ESI(+) (C) and ESI(-) data sets (D). Score plots show significant separation between the different reproductive stage and among the different multi-omics data ( $p < 0.05$ ). The ellipse for each group represented Hotelling's T2 95% confidence interval.

The performance of the models using two PCs was assessed by examining the R2 values ranging from 0.73 to 0.95, and their corresponding Q2 values ranging from 0.31 to 0.59. According to the paired comparison, the separation between the five groups was clear. The R2X, R2Y, Q2Y and RMSEE values in the PLS-DA models are listed in table 13.

Data set	R <sup>2</sup> X	R <sup>2</sup> Y	Q <sup>2</sup> Y	RMSEE*
Proteomics (ESI+)	0.883	0.377	0.307	0.337
Metabolomics (ESI-)	0.731	0.539	0.358	0.296
Lipidomics (ESI+)	0.949	0.839	0.594	0.189
Lipidomics (ESI-)	0.74	0.499	0.379	0.309

**Tableau 13** : Summary of the PLS-DA models performance parameters. \*RMSEE stands for Root Mean Square Error of Estimation



Variable Importance in Projection (VIP) score derived from PLS-DA was used for the selection of variables responsible of the developmental stage separation. VIP values greater than one within each data set, were used as variable selection criterion. The list of the VIP and the discriminant variables are listed in table 14-17. A list of 109 features were obtained (13 for proteomics, 24 for metabolomics and 72 for lipidomics).

13 discriminant peptides were reporter-peptides for clottable protein (6 peptides), vitellogenin-like protein (2 peptides), arginine kinase (2 peptides), tubulin  $\alpha$ -3 chain (1 peptide) and sarcoplasmic calcium binding protein (1 peptide) and a protein with an unknown function (1 peptide) (Table n°14, appendix n°10). These proteins are most likely involved in the female reproductive process since most of them are protein members of the large lipid transfer protein superfamily like vitellogenins and clottable proteins. These proteins are structural scaffolds for the assembly of lipoproteins and binds a variety of lipid species. These results are in perfect agreement with previous proteomics studies [213,220,251].

Peptides	VIP	P value	P value adjusted	Proteins ID GFOSS DB	Function
ISPLINSPDLPK	10.7	1.13E <sup>-06</sup>	1.27E <sup>-06</sup>	39606_fr3	Clotting protein precursor
LVSGVEQIEK	7.3	2.78E <sup>-19</sup>	3.33E <sup>-18</sup>	14404_fr1	Arginine kinase
AAIETAFVNHK	6.3	9.42E <sup>-08</sup>	1.08E <sup>-07</sup>	206469_fr3	Clotting protein precursor
IFNVLQPIAISK	5.5	1.67E <sup>-06</sup>	1.82E <sup>-06</sup>	17046_fr6	Hemolymph clottable protein
ELIEEAR	3.1	6.40E <sup>-08</sup>	7.37E <sup>-08</sup>	30490_fr1	Vitellogenin-like protein precursor
TSEVFLPLTNELYQQTK	2.8	1.20E <sup>-06</sup>	1.33E <sup>-06</sup>	17046_fr6	Hemolymph clottable protein
LGFLTFCPTNLGTTIR	2.3	6.21E <sup>-19</sup>	6.62E <sup>-18</sup>	14404_fr1	Arginine kinase
LQAIQGLSK	2.2	1.05E <sup>-06</sup>	1.18E <sup>-06</sup>	17421_fr1	Hemolymph clottable protein
IIYPAEALTIVIEK	1.5	2.92E <sup>-05</sup>	2.99E <sup>-05</sup>	276_fr4	Hemolymph clottable protein
SSALINLSTLIR	1.4	2.83E <sup>-06</sup>	3.02E <sup>-06</sup>	18251_fr1	unknown
LIGQIVSSITASLR	1.4	1.33E <sup>-16</sup>	4.60E <sup>-16</sup>	9002_fr1	Tubulin alpha-3 chain (Alpha-III tubulin)
TFISSTFK	1.4	3.86E <sup>-13</sup>	6.31E <sup>-13</sup>	38781_fr1	Sarcoplasmic calcium-binding protein
GPISVTR	1.1	1.08E <sup>-05</sup>	1.13E <sup>-05</sup>	277_fr1	vitellogenin

Tableau 14 : List of peptides and their corresponding proteins with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted.

Among the 8 metabolites discriminating the different reproductive stages, 3 metabolites are involved in pyrimidine metabolism (uridine diphosphate glucose, orotidine monophosphate, uracil), 2 metabolites in purine metabolism (urate, guanine), 1 in vitamin B3 metabolism (nicotinic acid), 1 in tyrosine and phenylalanine metabolism (cinnamoylglycine) and 1 in eicosanoid and resolving metabolism (13S-hydroxyoctadecadienoic acid) (Table n°15, appendix n°11). Further investigations must be undertaken to better understand the link between the metabolic variations and the ovarian cycle in gammarids.

Metabolites	Biological pathway	VIP	p-value	p-value adjusted
Guanine	Purine metabolism	1.34	2.29E <sup>-18</sup>	1.83E <sup>-17</sup>
Nicotinic acid	Vitamin B3 metabolism	1.33	3.29E <sup>-17</sup>	1.98E <sup>-16</sup>
Uridine diphosphate glucose	Pyrimidine metabolism	1.28	4.32E <sup>-07</sup>	4.94E <sup>-07</sup>
Orotidine monophosphate	Pyrimidine metabolism	1.24	2.61E <sup>-05</sup>	2.72E <sup>-05</sup>
Uracil	Pyrimidine metabolism	1.22	3.39E <sup>-11</sup>	7.40E <sup>-11</sup>
Cinnamoylglycine	Tyrosine and phenylalanine metabolism	1.14	2.75E <sup>-08</sup>	3.89E <sup>-08</sup>
13S-hydroxyoctadecadienoic acid	Eicosanoid and resolvin metabolism	1.08	1.74E <sup>-13</sup>	4.63E <sup>-13</sup>
Uric acid	Purine metabolism	1.05	2.29E <sup>-14</sup>	7.84E <sup>-14</sup>

Tableau 15 : List of metabolites with PLS-DA VIP scores (&gt;1) together with its p-value and p-value-adjusted.

As depicted in table n°16 and 17, 46 discriminant lipids in positive ion mode and 26 lipids in negative ion mode were identified. 8 lipid class variables (TG, SM, PC, PE, PI, PS LPC, LPE) were accounted to significant contribution (VIP score greater than 1) and can be used to differentiate the different reproductive time points. TG class is the most predominant lipid class in this amphipod [198]. During the reproductive cycle of female gammarids, oocytes mobilize energy and consequently require substantial amount of lipids to accomplish the vitellogenesis process [324–326]. TG serve as energy storage reservoir i.e., fatty acid store, that can be quickly mobilize on request during vitellogenesis [327]. Membrane lipids increase such as glycerophospholipids (PC, PE, PI and PS), lysophospholipids (LPC, LPE) and sphingolipids (SM) with long-chain polyunsaturated fatty acid, occurred at the end of ovarian maturation. These observations are in perfect agreement with previously published results [198].

Lipid species	MRM fragments	Molecular species	VIP	p-value	p-value adjusted
TG 52:2	NL 18:1	16:0-18:1-18:1	6.94	6.52E <sup>-15</sup>	3.82E <sup>-14</sup>
TG 54:3	NL 18:1	18:1-18:1-18:1	6.20	6.09E <sup>-13</sup>	2.16E <sup>-12</sup>
LPC 18:1	HG	0:0/18:1	4.13	1.02E <sup>-09</sup>	1.58E <sup>-09</sup>
TG 50:1	NL 16:0	18:1-16:0-16:0	3.89	5.40E <sup>-09</sup>	7.20E <sup>-09</sup>
TG 52:2	NL 16:0	16:0-18:1-18:1	3.83	3.57E <sup>-12</sup>	1.08E <sup>-11</sup>
TG 52:3	NL 18:1	16:1-18:1-18:1	3.26	5.44E <sup>-19</sup>	4.15E <sup>-17</sup>
TG 54:4	NL 18:1	18:1-18:1-18:2	3.26	9.69E <sup>-12</sup>	2.64E <sup>-11</sup>
TG 52:3	NL 18:1	16:0-18:2-18:1	3.17	1.67E <sup>-19</sup>	1.70E <sup>-17</sup>
LPC 18:1	Phosphate	0:0/18:1	3.04	1.49E <sup>-09</sup>	2.25E <sup>-09</sup>
TG 54:6	NL 20:5	20:5-18:1-16:0	2.75	1.05E <sup>-09</sup>	1.61E <sup>-09</sup>
TG 52:3	NL 18:2	16:0-18:2-18:1	2.75	4.94E <sup>-12</sup>	1.43E <sup>-11</sup>
TG 50:2	NL 18:1	18:1-16:1-16:0	2.62	4.37E <sup>-17</sup>	7.85E <sup>-16</sup>
LPC 16:0	HG	0:0/16:0	2.48	2.64E <sup>-10</sup>	4.63E <sup>-10</sup>
TG 50:2	NL 18:1	18:1-18:1-14:0	2.41	1.97E <sup>-17</sup>	4.64E <sup>-16</sup>
TG 52:4	NL 18:2	16:1-18:1-18:2	2.34	1.43E <sup>-09</sup>	2.18E <sup>-09</sup>
TG 52:4	NL 18:2	16:0-18:2-18:2	2.23	9.83E <sup>-10</sup>	1.54E <sup>-09</sup>
TG 52:4	NL 18:1	16:0-18:1-18:3	1.83	4.76E <sup>-17</sup>	8.06E <sup>-16</sup>
LPC 16:0	Phosphate	0:0/16:0	1.83	9.87E <sup>-10</sup>	1.54E <sup>-09</sup>

TG 54:5	NL 18:2	18:1-18:2-18:2	1.81	9.66E-07	1.02E-06
SM 36:1	HG	d18:1/18:0	1.78	2.37E-16	3.28E-15
LPC 18:2	HG	0:0/18:2	1.75	1.86E-08	2.25E-08
TG 52:6	NL 20:5	20:5-16:1-16:0	1.70	3.24E-10	5.52E-10
LPC 16:0	HG	16:0/0:0	1.69	4.57E-09	6.20E-09
TG 52:3	NL 16:0	16:0-18:2-18:1	1.68	7.91E-14	3.50E-13
TG 54:4	NL 18:2	18:1-18:1-18:2	1.60	6.19E-11	1.36E-10
TG 50:2	NL 16:1	18:1-16:1-16:0	1.56	5.39E-16	6.58E-15
TG 54:5	NL 18:1	18:1-18:1-18:3	1.56	2.38E-17	4.65E-16
TG 50:1	NL 18:1	18:1-16:0-16:0	1.48	2.26E-10	4.09E-10
TG 52:4	NL 18:1	16:1-18:1-18:2	1.46	1.11E-17	3.76E-16
TG 54:5	NL 18:1	18:1-18:2-18:2	1.45	2.44E-17	4.65E-16
LPC 18:1	HG	18:1/0:0	1.44	4.43E-11	1.00E-10
TG 50:2	NL 16:0	18:1-16:1-16:0	1.43	7.88E-16	8.90E-15
TG 48:1	NL 16:0	16:1-16:0-16:0	1.42	2.34E-11	5.71E-11
TG 48:1	NL 16:0	18:1-16:0-14:0	1.41	1.38E-11	3.54E-11
TG 54:7	NL 20:5	16:1-18:1-20:5	1.40	1.99E-10	3.70E-10
TG 54:2	NL 18:1	18:1-18:1-18:0	1.22	7.92E-09	1.01E-08
TG 52:5	NL 18:2	16:1-18:2-18:2	1.22	1.59E-10	3.00E-10
TG 52:4	NL 18:3	16:0-18:1-18:3	1.21	2.77E-12	8.52E-12
TG 54:2	NL 18:1	20:1-18:1-16:0	1.18	1.64E-09	2.44E-09
SM 36:2	HG	d18:1/18:1	1.14	1.67E-15	1.54E-14
LPC 18:1	NL HG	0:0/18:1	1.10	2.18E-09	3.19E-09
LPC 20:5	HG	0:0/20:5	1.10	4.89E-10	8.06E-10
TG 52:3	NL 16:1	16:1-18:1-18:1	1.08	1.20E-17	3.76E-16
TG 48:1	NL 14:0	18:1-16:0-14:0	1.07	5.02E-10	8.24E-10
TG 52:4	NL 16:0	16:0-18:1-18:3	1.04	6.80E-15	3.91E-14
TG 52:4	NL 16:0	16:0-18:2-18:2	1.03	4.14E-15	2.74E-14

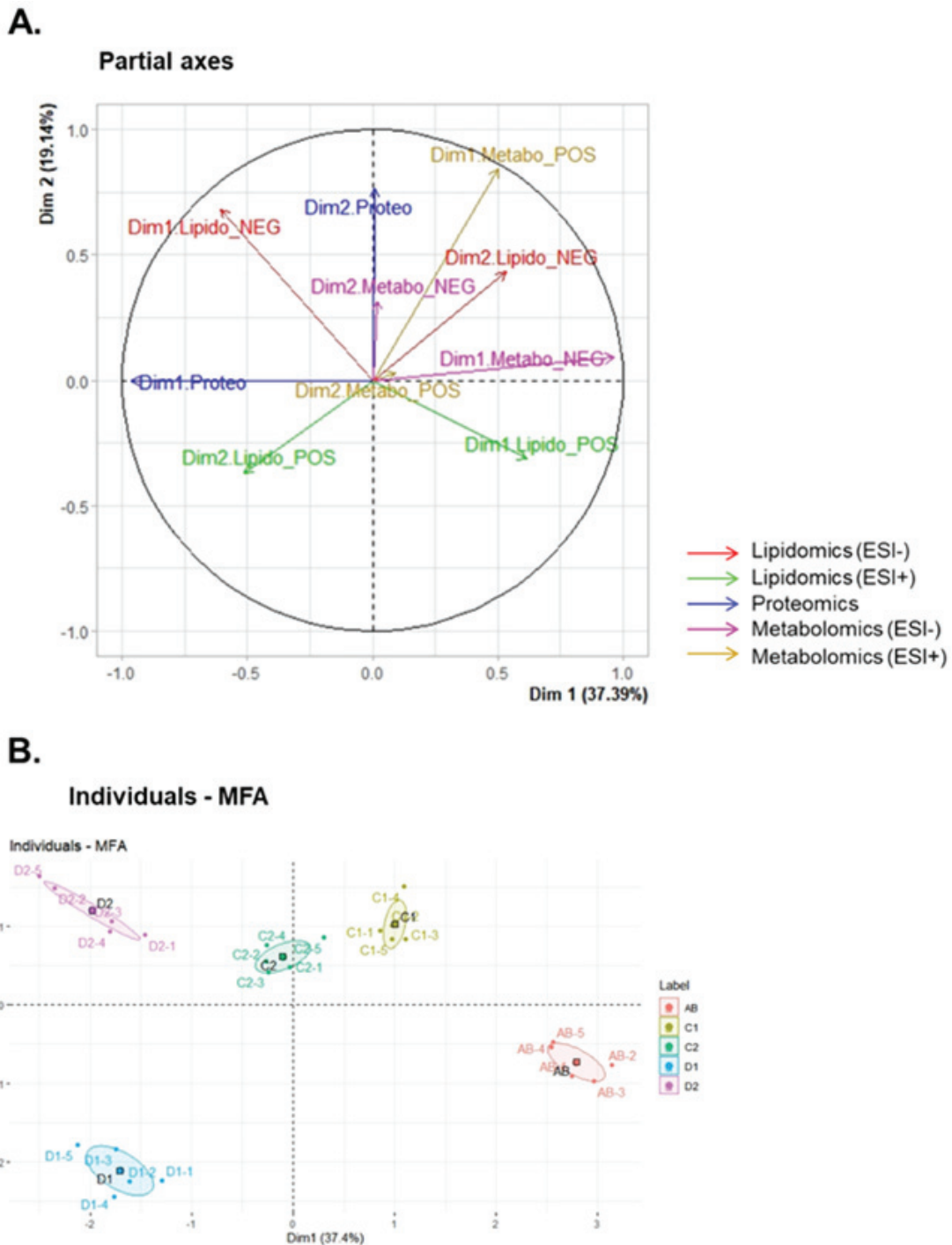
Tableau 16 : List of lipids with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted. The variables were obtained in positive ion mode. NL, neutral loss. HG, Head group loss (for LPC and PC, m/z 184.2)

Lipid species	MRM fragments	Molecular species	VIP	p-value	p-value adjusted
LPE 18:1	18:1	0:0/18:1	6.70	4.18E-12	5.15E-12
PC 34:1	16:0	16:0-18:1	6.44	5.38E-26	7.02E-24
PC 36:2	18:1	18:1-18:1	6.43	5.33E-16	1.29E-15
PC 36:3	18:2	18:2-18:1	4.84	2.31E-13	3.30E-13
PC 36:4	18:2	18:2-18:2	3.84	2.64E-07	2.72E-07
PC 34:1	18:1	18:1-16:0	3.15	3.29E-26	7.02E-24
PC 36:4	18:3	18:3-18:1	3.05	5.18E-16	1.26E-15
PC 36:3	18:1	18:1-18:2	2.47	6.01E-13	7.92E-13
PC 34:2	18:2	18:2-16:0	2.46	7.53E-16	1.77E-15
PC 34:2	16:0	16:0-18:2	2.34	2.05E-16	5.70E-16
PC 38:6	20:5	20:5-18:1	2.23	1.32E-22	2.87E-21
PE 36:3	18:2	18:2-18:1	2.08	9.75E-14	1.56E-13
PE 36:2	18:1	18:1-18:1	2.06	1.09E-20	1.23E-19
PC 38:6	18:1	18:1-20:5	2.02	1.79E-22	3.58E-21
PE 38:6	20:5	20:5-18:1	1.98	1.20E-24	7.84E-23
PC 34:3	18:3	18:3-16:0	1.91	3.82E-15	7.98E-15

PE 38:6	18:1	18:1-20:5	1.74	1.26E-23	4.70E-22
PC 34:2	18:1	18:1-16:1	1.50	6.34E-15	1.25E-14
PS 38:5	18:1	18:1-20:4	1.30	3.03E-19	2.03E-18
PC 36:4	18:1	18:1-18:3	1.27	3.34E-16	8.55E-16
PI 38:5	18:1	18:1-20:5	1.23	4.52E-17	1.64E-16
PC 38:7	20:5	20:5-18:2	1.21	7.96E-16	1.84E-15
PS 38:6	18:1	18:1-20:5	1.19	1.92E-24	1.00E-22
PE 38:5	20:4	20:4-18:1	1.17	3.15E-18	1.61E-17
PC 34:2	16:1	16:1-18:1	1.08	1.52E-15	3.26E-15
PC 34:3	16:0	16:0-18:3	1.08	6.95E-18	3.12E-17

Tableau 17 : List of lipids with PLS-DA VIP scores (>1) together with its p-value and p-value-adjusted. The variables were obtained in negative ion mode.

To deepen the data integration and interpretation, we performed a multi-block data analysis, more precisely, a multiple factor analysis (MFA) on the different omics data (Figure n°89 A and B) [328]. MFA was applied on normalized matrices and calculated from 5 components. The two first factorial axes were principally explained by the proteomics data (contribution matrix) and by the metabolomics data (negative) (Figure n°89 A). Lipidomics data (positive and negative) explained the 3rd and 4th factorial axes respectively. The representation of the groups revealed that the proteomics and metabolomics data were correlated, providing similar information. Lipidomics matrices were correlated. The representation of individuals with the five matrices (proteomics, metabolomics in positive and negative ion mode, and lipidomics in positive and negative ion mode) displayed a discrimination between the different reproductive stages on the two first axis (56.5%) (Figure n°89 B).



**Figure 89 : Multiple factor analysis (MFA) unveils proteomics, lipidomics and metabolomics variation during the different reproductive stage in female gammarid.** (A) Partial axes – MFA scatter plot. The partial axis representation reveals the link between the first two principal components of the MFA and those of each individual group. The red arrow represents the lipidomics data in ESI(-); the light green arrow represents the lipidomics data in ESI(+); the blue arrow represents the proteomics data; the pink arrow represents the metabolomics data in ESI(-) and the yellow arrow represents the metabolomics data in ESI(+). The scatter plot exhibits a reflection of the variables as regards their correlation with the factor axes. The arrows indicate the directions of growth of the variables in the factor space, which allows us to identify which omics table have the most weight in each component. The center of the circle of correlations represents the mean of all variables. (B) Individual factor map. MFA highlights a good separation between female gammarids sampled at different reproductive stages (A/B in red, C1 in yellow, C2 in green, D1 in blue, and D2 in purple). Organisms are presented by colored dots according to the reproductive stage on the scatter plot created with the first two main dimensions of MFA. Confidence ellipses for a 95% interval are plotted for each group to visualize the range of variability.

The LG-coefficient indicated that all the blocks are necessary for the interpretation from a global and individual point of view (Appendix n°12). The RV-coefficient highlighted that the data from metabolomics (negative) and proteomics provided most of the information for the reproductive stage differences (Appendix n°13).

### **3. Conclusion**

This study reports the development and optimization of a robust multi-omics extraction method applied in the freshwater amphipod *G. fossarum*. Our multi-omics procedure allows simultaneously extraction of proteins, lipids and metabolites from a unique sample, followed by a multiplex targeted approach (scheduled-MRM monitoring). We examined the influence of exoskeleton residues on the extraction efficiency. We applied the protocol to investigate the reproductive cycle of female gammarids at different stages. Molecular signatures were highlighted by multivariate data analysis (PCA, PLS-DA) and multiblock approach (MFA) revealing a concomitant increase of specific proteins involved in ovarian maturation, a lipid storage mobilization, and variations in specific metabolic pathways. To our knowledge, this is the first study that combines a multi-omics extraction method in conjunction of multivariate statistical analysis on a unique sample, to describe in detail the reproductive stage of female gammarids. Our study opens interesting perspectives for future ecotoxicology studies since analysis can be performed on a unique individual, reducing the number of exposures and therefore the number of samples to collect while allowing an exhaustive characterization of the sample at a multi-scale level.

# Chapitre 5 : Apport de la chromatographie bidimensionnelle pour la lipidomique

## 1. Pourquoi développer des méthodes bidimensionnelles en lipidomique ?

Réaliser les différentes études multi-omiques à partir d'un même échantillon permet de s'affranchir des variations biologiques que nous observons lorsqu'une expérience (ex. effet de l'exposition à un contaminant) est répétée sur plusieurs individus. Ainsi, une meilleure corrélation des informations entre les omiques sera observée [155]. De plus, cela permet d'avoir suffisamment de matériel biologique pour réaliser les différentes omiques. Comme nous l'avons vu dans le chapitre 4, cette préparation d'échantillon multi-omiques, bien que réduisant la complexité de l'échantillon, impose de faire des concessions en optant pour des conditions d'extraction compatibles entre les différentes omiques. Ainsi, le fractionnement de l'échantillon n'est pas aussi performant que si l'on avait envisagé les différentes omiques séparément. Les extraits sont donc plus complexes et peuvent être sujets aux phénomènes de suppression d'ions, également appelés effets matrices. Ces phénomènes, peuvent induire une diminution de la sensibilité lorsque des échantillons trop complexes sont analysés en LC-MS/MS avec les sources d'ionisation électrospray [62]. La présence de composés isobariques génère également des interférences spectrales. Le seul moyen de diminuer les effets matrices et/ou les interférences spectrales est de rajouter une étape de séparation entre la préparation d'échantillons et la détection. Même si la chromatographie permet de les diminuer, elle demeure limitée en termes de capacité de séparation pour les échantillons complexes [91,92]. De nos jours, de nouveaux outils de séparation plus avancés sont disponibles, comme la chromatographie liquide bidimensionnelle (2D-LC) qui dispose de deux dimensions de séparations orthogonales. Le dernier objectif de cette thèse est donc d'évaluer l'apport de la 2D-LC par rapport à la chromatographie monodimensionnelle (1D-LC) lors des analyses omiques.

L'état de l'art sur l'analyse des lipides a montré des besoins en termes de séparation. Par exemple, la chromatographie HILIC, qui utilise une phase stationnaire hydrophile, sépare les composés en fonction de leurs polarités. Dans le cas des lipides, cette séparation est due à la présence des têtes polaires qui sont communes à chaque famille lipidique. Cela implique une bonne séparation par classe de lipides, avec peu de séparation entre les lipides qui diffèrent uniquement par leurs chaînes d'acides gras (nombre de carbone et/ou d'insaturation) [329]. La phase inverse présente des mécanismes de séparations orthogonaux et complémentaires à la HILIC puisque la séparation se fait en fonction de l'hydrophobicité des composés. Dans le cas des lipides, cette hydrophobicité est due à la présence des



chaines d'acides gras, saturés ou non, avec des nombres de carbones plus ou moins importants. Cela implique une bonne séparation entre les différents lipides d'une même famille, mais peut engendrer des co-élutions entre lipides de classes différentes, notamment pour les familles de phospholipides qui sont co-élusés entre elles [329]. Lorsqu'elles sont utilisées indépendamment l'une de l'autre, ni la RPLC, ni la HILIC ne peuvent séparer l'ensemble des lipides. Toutefois, l'association de ces deux modes de chromatographie présentant des mécanismes de rétention orthogonaux devrait améliorer la séparation en lipidomique. C'est pour cette raison qu'une approche chromatographique bidimensionnelle couplée à la spectrométrie de masse a été envisagée pour l'analyse des lipides.

Comme il a été souligné dans le chapitre 1, plusieurs critères doivent être pris en considération si l'on souhaite obtenir une méthode analytique 2D-LC idéale avec les limites de détection les plus faibles possibles et un minimum d'interférences spectrales :

- La méthode doit être capable d'augmenter les capacités de séparation par rapport aux méthodes 1D-LC afin de limiter les effets matrices et les interférences isobariques. Pour cela, le système bidimensionnel doit séparer les classes lipidiques entre elles, mais aussi tous les lipides d'une même famille.
- Cette augmentation du pouvoir de séparation ne doit pas être accompagnée d'un temps d'analyse global trop important.
- L'approche 2D doit combiner des séparations chromatographiques orthogonales avec des phases mobiles compatibles afin d'éviter l'usage d'étapes d'évaporation entre les deux dimensions.
- Le passage entre les deux dimensions doit être totalement automatisable.
- Pour éviter la perte d'informations, la totalité des fractions issues de la 1<sup>ère</sup> dimension doit être envoyée dans la 2<sup>ème</sup> dimension.
- La méthode 2D doit être compatible avec la totalité des spectromètres de masse, que ce soit en haute résolution (TOF, Orbitrap) ou en basse résolution (QQQ).
- Le retraitement de données doit être le plus simple possible.

À ce jour, aucune configuration n'est capable de remplir la totalité de ces critères afin d'obtenir la méthode 2D-LC idéale pour l'analyse de lipides.

La configuration RPLCxHILIC décrite par Holcapek *et al.* [133] et Xu *et al.* [123] est celle qui s'en rapproche le plus. Pour rappel, le mode *comprehensive* (LCxLC) permet d'analyser l'ensemble des fractions issues de la 1<sup>ère</sup> dimension dans la 2<sup>ème</sup> dimension, ce qui limite la perte d'informations. Ce système est également automatisable, tout en permettant de séparer les lipides entre classes et à l'intérieur des familles lipidiques. Cela a pour effet d'augmenter la capacité de pic par rapport aux méthodes 1D. Les études de Holcapek *et al.* [133] et de Xu *et al.* [123] ont également mis en évidence

que la RPLCxHILIC présente un meilleur pouvoir de séparation que la HILICxRPLC lors d'une analyse lipidique. Cette différence peut être expliquée par le fait que la RPLC ne permet pas d'atteindre des capacités de séparation suffisantes pour la séparation des lipides en utilisant des temps de gradient inférieurs à la minute en seconde dimension, ce qui est le cas en mode HILIC [133].

Les travaux qui seront présentés dans ce chapitre se sont principalement intéressés au développement d'une méthode alternative au mode RPLCxHILIC. Cette méthode devra conserver les avantages de la LCxLC en termes d'automatisation, de qualité de séparation (diminution des effets matrices) et de quantité d'informations collectées (nombre de lipides détectés). Le premier choix que nous avons dû faire a été de sélectionner le type de phase stationnaire pour chacune des dimensions. Bien que les études précédentes [123,133] ont démontré de meilleurs résultats avec la phase RPLC en 1<sup>ère</sup> dimension et la phase HILIC en 2<sup>ème</sup> dimension, nous avons choisi d'inverser ces deux dimensions de séparation. Cela nous permettra de mettre en évidence de nouvelles possibilités, telles que la sélection d'une ou plusieurs familles de lipides à envoyer dans la 2<sup>ème</sup> dimension, afin d'utiliser la 1<sup>ère</sup> dimension HILIC comme une « chromatographie préparatrice ». Ainsi, il sera possible d'isoler les fractions minoritaires et d'injecter des extraits de différentes concentrations en fonction des différentes familles de lipides pour gagner en sensibilité.

Les premières expérimentations présentées dans ce chapitre évalueront et confirmeront les résultats de Holcapek *et al.* [133] et de Xu *et al.* [123] concernant la faible performance de la RPLC en tant que 2<sup>ème</sup> dimension dans la configuration HILICxRPLC lors d'analyses lipidomiques. Cela nous permettra d'observer et d'identifier la principale limitation de cette configuration : un temps de modulation très rapide entre les deux dimensions, qui impose un temps de gradient trop court en 2<sup>ème</sup> dimension pour effectuer une séparation chromatographique efficace en RPLC. Pour pouvoir conserver la colonne HILIC comme 1<sup>ère</sup> dimension nous montrerons que le mode *multiple heart cut* (LC-LC) est mieux adapté.

Les avantages et inconvénients de cette approche *multiple heart cut* HILIC-RPLC-MS seront comparés à l'approche *comprehensive* RPLCxHILIC-MS à travers l'évaluation (i) du temps d'analyse global, (ii) de la qualité du fractionnement entre les deux dimensions (nombre de fractions et effets à l'injection lors du passage de la 1<sup>ère</sup> à la 2<sup>ème</sup> dimension), (iii) de la compatibilité avec les spectromètres de masse et (iv) la facilité du retraitement des données.

Afin d'avoir la plus grande diversité de lipides, la méthode a été développée à partir d'extraits de deux matrices biologiques différentes : du plasma humain et l'organisme entier *G. fossarum*. Ainsi nous avons à disposition l'ensemble des familles lipidiques couramment rencontrées dans la littérature, avec une grande diversité au sein de ces familles (PC / PI / PE / PS / SM / Cer / DG / TG / LPE / LPC).

## 2. Matériels et méthodes

Les expérimentations visant à optimiser la méthode et les analyses 2D-LC en mode hors ligne ont été réalisées sur un spectromètre de masse de type triple quadripôles (Qtrap 5500 de chez Sciex®) utilisant le mode MRM pour des raisons de disponibilités d'appareillage. Les analyses 2D-LC en mode en ligne ont ensuite été réalisées sur un IM-QTOF 6560 de chez Agilent®. Ce système possède la mobilité ionique (IM), mais cette technologie n'a pas été utilisée pour ces travaux.

### 2.1. Réactifs et produits chimiques

L'eau, l'acétonitrile (ACN), le méthanol (MeOH), l'isopropanol (IPA) ont été obtenus auprès de Fisher Scientific (qualité LC-MS, Strasbourg, France). L'EDTA, le triton X-100, le formiate d'ammonium, le dichlorométhane (DCM), le tert-butyl méthyl éther (MTBE), l'acide formique et le chlorure de sodium ont été achetés auprès de Sigma Aldrich (St Quentin-Fallavier, France). Les lipides marqués isotopiquement (PC 12:0/13:0, PC 11:0/11:0, LPC 13:0, PE 12:0/13:0, PG 12:0/13:0, PS 12:0/13:0, PI 12:0/13:0, DG 17:0/17:0, TG 15:0/18:1/15:0 d5, TG 17:0/17:1/17:0 d5, Cer d18:1/12:0 et SM d18:1/18:1 d9) ont été obtenus auprès de Sigma Aldrich (St Quentin-Fallavier, France, pureté > 99%). Le mélange de standards SPLASH LIPIDOMIX® mass spec a été acheté auprès de Sigma Aldrich (St Quentin-Fallavier, France).

### 2.2. Collecte des échantillons

Les organismes de *G. fossarum* ont été prélevés à l'aide d'un filet à La Tour du Pin (Isère, France), en amont de la rivière La Bourbre (centre est de la France), et transportés rapidement au laboratoire (INRAe, Villeurbanne) dans de l'eau douce provenant de la station de collecte à température ambiante. Les organismes ont été maintenus pendant une période d'acclimatation d'au moins 10 jours dans des cuves de 30 L alimentées en continu par de l'eau souterraine ajustée à la conductivité du site d'échantillonnage ( $600 \mu\text{S}\cdot\text{cm}^{-1}$ ) et sous aération constante. Une photopériode lumière/obscurité de 16/8 heures a été maintenue, ainsi qu'une température de  $12 \pm 1$  °C. Les organismes ont été nourris *ad libitum* avec des feuilles d'aulne (*Alnus glutinosa*). Les feuilles ont été conditionnées dans l'eau pendant au moins  $6 \pm 1$  jours. Des vers Tubifex ont été donnés comme compléments alimentaires deux fois par semaine. Tous les organismes choisis ont été pesés et immédiatement congelés dans de l'azote liquide pour être stockés à  $80$  °C jusqu'à extraction [171,215,216,246]. Le plasma humain a été collecté par l'établissement français du sang par aphérèse. Après la collecte, le plasma est aliquoté puis stocké dans un congélateur à  $-80$ °C.

## 2.3. Extraction des lipides

### 2.3.1. Broyage et extraction liquide-liquide

L'échantillon entier de *G. fossarum* a été broyé dans 600 µL de tampon d'extraction (Tris HCL 50 mM, 1mM EDTA, Triton X-100 0.1%, 100 mM NaCl, pH ajusté à 7.8) en utilisant un broyeur à billes. Après avoir retiré les billes d'acier de 4 mm, l'échantillon est centrifugé pendant 15 min à 4°C et 10000g. Après centrifugation, 300 µL de surnageant ont été prélevés pour réaliser l'extraction liquide-liquide (LLE). Dans le cas du plasma, 50 µL de plasma sont prélevés et complétés avec 250 µL d'eau. Ensuite, 360 µL de MeOH froid (4°C) sont ajoutés et l'échantillon est agité pendant 60 minutes. Dans une deuxième étape, 1200 µL de MTBE froid (4°C) sont ajoutés et l'échantillon est agité une deuxième fois pendant 10 minutes. Une séparation de phases a lieu pour séparer la phase organique de la phase aqueuse. Après centrifugation de l'échantillon à 4°C pendant 5 minutes à 10000g, 1260 µL de la phase organique contenant les lipides (phase supérieure) sont récupérés.

### 2.3.2. Évaporation et dilution avant injection

Après la LLE, la fraction lipidique est évaporée à sec sous flux d'azote. Finalement, l'échantillon est resuspendu avec 20 µL de DCM/MeOH (1:1 ; v/v) pour solubiliser les lipides avant d'ajouter 180 µL d'IPA/ACN/H<sub>2</sub>O (2:1:1 ; v/v/v), qui correspond au solvant d'injection en RPLC. Dans le cas des analyses HILIC, 180 µL d'ACN/H<sub>2</sub>O/MeOH sont ajoutés comme solvant d'injection. Pour les analyses ESI (+) l'échantillon de gammare est dilué par 100 et par 10 en ESI (-). Pour les analyses ESI (+), l'échantillon de plasma est dilué par 50 et par 5 en ESI (-).

## 2.4. Étude des effets à l'injection en RPLC-HILIC hors ligne

L'étude des phénomènes de *breakthrough* avec la configuration RPLC en 1<sup>ère</sup> dimension (D1) et HILIC en 2<sup>ème</sup> dimension (D2) a été évaluée à partir d'un mélange de standards. Les standards de lipides sont mis en solution avec une composition correspondant à leur temps de rétention lors de la séparation chromatographique avec la colonne de type RPLC (prise en compte du temps de délais et du temps mort) et les conditions suivantes : utilisation d'une colonne Xselect CSH C<sub>18</sub> (100 mm × 2.1 mm, taille des particules 3.5 µm) avec un volume d'injection de 5 µL et une température de colonne de 70°C. L'élution est réalisée à un débit de 500 µL/min avec H<sub>2</sub>O/ACN (60/40 ; v/v) contenant 0,1 mM de formiate d'ammonium et 0,1 % d'acide formique comme éluant A. La phase B contient un mélange IPA/ACN/H<sub>2</sub>O (90/9.5/0.5 ; v/v/v) contenant 10 mM de formiate d'ammonium et 0,1 % d'acide formique. Le gradient appliqué dans la colonne RPLC commence avec une augmentation de 1% à 99% de B en 24 minutes puis à 99% de B pendant 2 min. Pour finir, la colonne est rééquilibrée à 1% de B pendant 3 minutes.

Les analyses LC-MS/MS de chacun des standards de lipides mis en solution selon les compositions de séparation précédentes ont été réalisées sur un spectromètre de masse hybride triple quadripôles/piège à ions linéaire du système QTRAP® 5500 LC-MS/MS (Applied Biosystems / MDS Analytical Technologies, Foster City, CA, USA). Ce système est équipé d'une source d'ions Turbo V™ couplée à un dispositif HPLC de la série 1290 (Agilent Technologies, Waldbronn, Allemagne). La séparation LC a été réalisée sur une colonne BEH HILIC (100 mm x 2.1 mm 1.7 µm) qui est utilisée comme 2<sup>ème</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'ACN/H<sub>2</sub>O (95/5 ; v/v) avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. Le gradient appliqué en D2 est le suivant : 1% à 20% de A en 6.3 minutes puis augmentation de 20% à 50% de A en 2 minutes. Le % de A est maintenu à 50% pendant 2 minutes puis il diminue à 1% de A en 1 minute. Pour finir, le gradient est maintenu à 1% de A pendant 3.5 minutes. Le débit en D2 est fixé à 600 µl/min avec une température de colonne de 40°C. Le volume injecté en 1<sup>ère</sup> dimension varie de 1 µL à 20 µL.

### 2.5. Utilisation du collecteur de fractions pour les analyses HILIC-RPLC en mode hors ligne

La collecte de fractions a été réalisée avec une chaîne chromatographique Acquity UPLC class 1 de chez Waters® équipée d'un collecteur de fractions automatique. Ce système a permis de récolter les fractions issues de la séparation chromatographique d'un extrait lipidique de gammare et de plasma, à partir de la colonne BEH HILIC (100 mm x 4.6 mm 3.5 µm) qui est utilisée comme 1<sup>ère</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'ACN/H<sub>2</sub>O (95/5 ; v/v) avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. Le gradient appliqué en D1 est le suivant : 1% à 20% de A en 13.9 minutes suivi d'un deuxième gradient linéaire jusqu'à 50% de A en 4.4 min. Ensuite, le pourcentage de phase A est maintenu à 50% pendant 4.4 min suivi d'une diminution du gradient à 1% de A en 2.2 min. Pour finir, le gradient est maintenu à 1% de A pendant 7.7 min. Le débit en D1 est fixé à 1.3 mL/min avec une température de colonne de 40°C. Le volume injecté en 1<sup>ère</sup> dimension est de 5 µL. Chaque collecte dure 0.3 min, ce qui permet de récupérer 390 µL de solution dans des flacons. Dans le cas des essais en mode hors ligne, le volume injecté en 2<sup>ème</sup> dimension varie de 5 à 40 µL. Les analyses LC-MS/MS de chacune des fractions issues de la séparation en HILIC ont été réalisées sur un spectromètre de masse hybride triple quadripôles/piège à ions linéaire du système QTRAP® 5500 LC-MS/MS (Applied Biosystems/ MDS Analytical Technologies, Foster City, CA, USA) équipé d'une source d'ions Turbo V™ couplée à un dispositif HPLC de la série 1290 (Agilent Technologies, Waldbronn, Allemagne). La séparation LC a été réalisée sur une colonne Xselect CSH C<sub>18</sub> (100 mm x 2.1 mm, taille des particules 3.5 µm) avec une température de colonne de 70°C. L'élution a été réalisée à un débit de 500 µL/min

avec H<sub>2</sub>O/ACN (60/40 ; v/v) contenant 0,1 mM de formiate d'ammonium et 0,1 % d'acide formique comme éluant A et IPA/ACN/H<sub>2</sub>O (90/9.5/0.5 ; v/v/v) contenant 10 mM de formiate d'ammonium et 0,1 % d'acide formique comme éluant B. Le gradient appliqué en D2 dans le cas des essais contenant les lysophospholipides est le suivant : 1% à 99% de B en 24 minutes. La colonne a ensuite été lavée à 99% de B pendant 2 min et rééquilibrée à 1% de B pendant 3 minutes. Le gradient appliqué pour les PL et les SM est le suivant : 1% à 55% de B en 1.3 minute puis 55% à 80% de B en 6.2 minutes. Pour terminer le pourcentage de B augmente de 80% à 99% en 1 minute. La colonne a ensuite été maintenue à 99% de B pendant 2.4 minutes et rééquilibrée à 1% de B pendant 3 minutes. Le gradient appliqué pour les TG, DG et Cer en 2D est le suivant : 1% à 60% de B en 1.4 minute puis 60% à 99% de B en 9.6 minutes. La colonne a ensuite été maintenue à 99% de B pendant 2.4 minutes et rééquilibrée à 1% de B pendant 3 minutes.

## 2.6. Paramètres de masse des analyses réalisées sur le système Qtrap 5500

Le traitement, l'acquisition des données et le contrôle de l'instrument ont été effectués à l'aide du logiciel Analyst<sup>®</sup> 1.6.3. L'analyse par spectrométrie de masse a été réalisée en mode d'ionisation positive avec une tension de spray de 5500 V et en mode d'ionisation négative avec une tension de spray de -4500 V. Le nébuliseur et le flux de gaz du rideau ont été réglés à 25 et 50 psi respectivement avec de l'azote. La source d'ions TurboV™ a été réglée à 450 °C avec le débit de gaz auxiliaire (azote) réglé à 40 psi. Le temps de cycle est fixé à 1 seconde afin d'atteindre quinze points par pic chromatographique pour chaque transition MRM.

## 2.7. 2D-LC *comprehensive* et spectrométrie de masse (QTOF)

Les analyses LCxLC en mode *comprehensive* ont été réalisées avec un IM-QTOF 6560 couplé à un système 2D-LC 1290 infinity 2 de chez Agilent<sup>®</sup>. Le spectromètre de masse est équipé d'une source d'ionisation électrospray Jetstream. L'interface qui connecte les deux dimensions est constituée d'une vanne 8 voies possédant 2 positions et équipée de deux boucles de 20 µl (*cf.* figure n°90).

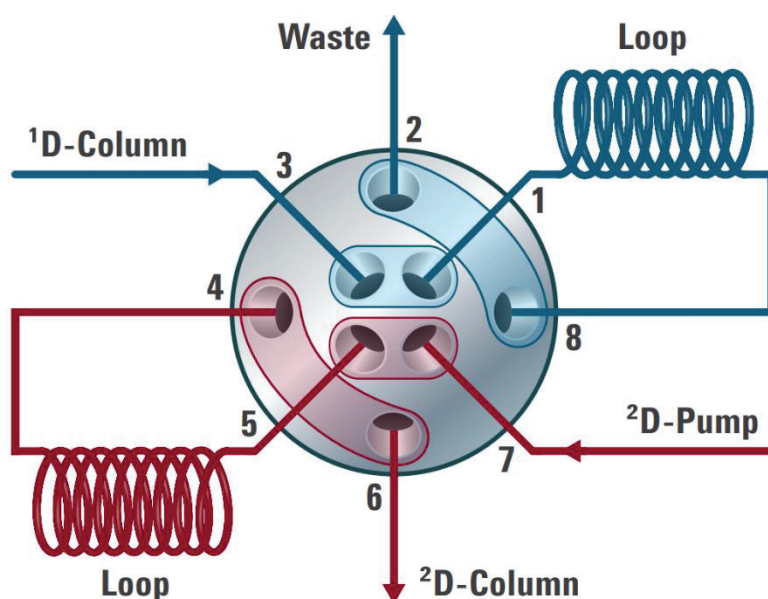


Figure 90 : Interface pour la chromatographie bidimensionnelle en mode *compréhensive* (LCxLC) [121].

Afin de diminuer la dispersion, la vanne a été configurée en mode contre-courant (*backflush*) pour que la vanne se charge dans un sens puis que l'élution soit réalisée dans le sens contraire. Une jonction en « T » est placée entre les deux dimensions (1:10) et à l'entrée du spectromètre de masse (1:3). La colonne BEH HILIC (50 mm x 2.1 mm et 3.5  $\mu\text{m}$  de taille de particule) est utilisée comme 1<sup>ère</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'ACN/H<sub>2</sub>O (95/5 ; v/v) avec 0.1% d'acide formique et 10mM de formiate d'ammonium. Le gradient appliqué en D1 est le suivant : 1% à 50% de A en 45 minutes puis diminution du gradient à 1% de A en 2.1 minutes. Pour finir, le gradient est maintenu à 1% de A pendant 2.4 minutes. La colonne Acquity C18 (30 mm x 2.1 mm 1.7  $\mu\text{m}$ ) est utilisée comme 2<sup>ème</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'IPA/ACN/H<sub>2</sub>O (90/9.5/0.5 ; v/v/v) avec 0.1% d'acide formique et 10mM de formiate d'ammonium. Le gradient appliqué en D2 est le suivant : 40% à 99% de B en 0.41 minute puis diminution du gradient à 40% de B en 0.04 minute. Pour finir, le gradient est maintenu à 40% de B pendant 0.02 minute. Le débit en D1 est fixé à 50  $\mu\text{L}/\text{min}$  et à 1.5 mL/min en D2. La température de colonne en D1 est fixée à 30°C et à 80°C en D2. Le volume injecté en 1<sup>ère</sup> dimension est de 5  $\mu\text{L}$ . Les données QTOF ont été acquises de 400  $m/z$  à 1000  $m/z$  avec un taux d'acquisition de 20 spectres/s. La tension du nébuliser, du *skimmer*, de l'octopole RF et du *nozzel* ont été fixés à 20, 65, 750 et 500 V respectivement. Le débit de *sheat gas* et de gaz séchant ont été fixés à 12 et 10 psi respectivement. La température de gaz séchant et de *sheat gas* ont été fixées à 200 et 350°C respectivement. Pour les analyses en mode ESI (+) la tension de capillaire est fixée à 3500 V et à -3000 V en mode ESI (-).



## 2.8. 2D-LC *multiple heart cut* en ligne et spectrométrie de masse (QTOF)

Les analyses LC-LC en mode *multiple heart cut* ont été réalisées avec un IM-QTOF 6560 couplé à un système 2D-LC 1290 infinity 2 de chez Agilent®. Le spectromètre de masse est équipé d'une source d'ionisation électrospray Jetstream. L'interface qui connecte les deux dimensions est constituée d'une vanne 8 voies possédant 2 positions et équipée de deux decks possédant chacun 6 boucles de 120 µl (cf. figure n°91). Afin de diminuer les phénomènes de dispersion, la vanne a été configurée en mode *backflush*. Une jonction en « T » est placée entre les deux dimensions (1:5).

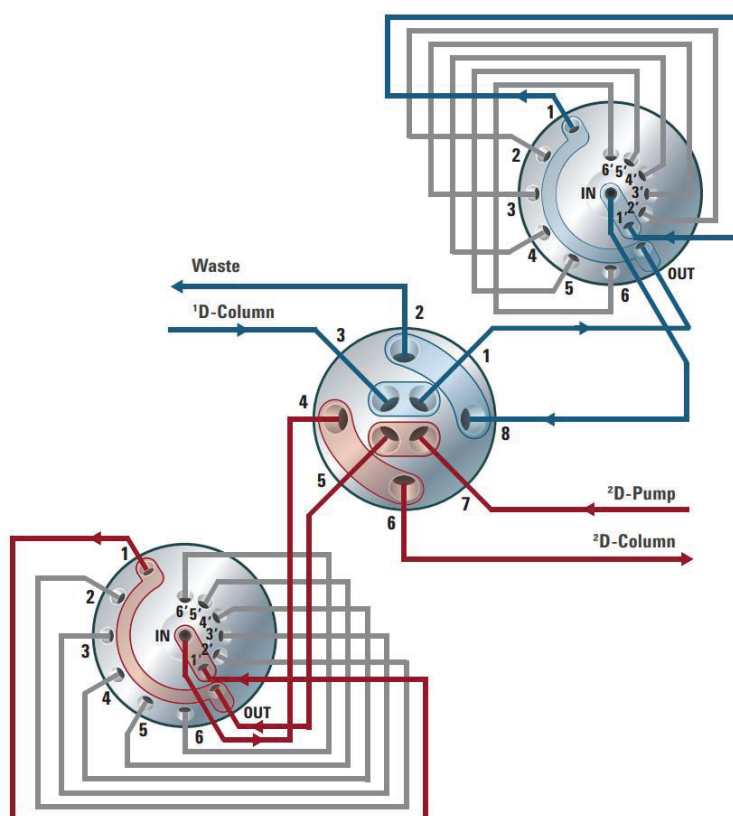


Figure 91 : Interface pour la chromatographie bidimensionnelle en mode *multiple heart cut* (LC-LC) [121].

La colonne BEH HILIC (100 mm x 2.1 mm et 1.7 µm de taille de particule) est utilisée comme 1<sup>ère</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'ACN/H<sub>2</sub>O (95/5 ; v/v) avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. Le gradient appliqué en D1 est le suivant : 1% à 20% de A en 6.3 minutes puis augmentation de 20% à 50% de A en 2 minutes. Le % de A est maintenu à 50% pendant 2 minutes puis il diminue à 1% en 1 minute. Pour finir, le gradient est maintenu à 1% de A jusqu'à la fin de l'analyse de l'ensemble des fractions dans la 2<sup>ème</sup> dimension. La fraction commune des TG, DG et Cer est collectée à 0.35 minute pendant 20 secondes. La fraction des PE est collectée à 2 minutes pendant 20 secondes. La fraction des PI est collectée à 2.35 minutes pendant 13 secondes. La fraction

des PC est collectée à 2.57 minutes pendant 20 secondes. La fraction des PS est collectée à 3 minutes pendant 25 secondes. La fraction des SM est collectée à 3.39 minutes pendant 30 secondes. La colonne Xselect CSH C18 (100 mm x 2.1 mm et 3.5 µm de taille de particule) est utilisée comme 2<sup>ème</sup> dimension. La phase mobile A est constituée de H<sub>2</sub>O/ACN (60/40 ; v/v) avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. La phase mobile B est constituée d'IPA/ACN/H<sub>2</sub>O (90/9.5/0.5 ; v/v/v) avec 0.1% d'acide formique et 10 mM de formiate d'ammonium. Le gradient appliqué pour les PL et SM en 2D est le suivant : 1% à 55% de B en 1.3 minute puis augmentation de 55% à 80% de B en 6.2 minutes suivie par une deuxième augmentation de 80% à 99% de B en 1 minute. Le pourcentage de B est maintenu à 99% pendant 2.4 minutes puis il diminue à 1% en 0.5 minute. Pour finir, le gradient est maintenu à 1% de B pendant 3 minutes. Le gradient appliqué pour les TG, DG et Cer en 2D est le suivant : 1% à 60% de B en 1.4 minute puis augmentation de 60% à 99% de B en 9.65 minutes. Le pourcentage de B est maintenu à 99% pendant 2.4 minutes puis il diminue à 1% en 0.5 minute. Pour finir, le gradient est maintenu à 1% de B pendant 3 minutes. Le débit en D1 est fixé à 600 µl/min et à 500 µl/min en D2. La température de colonne en D1 est fixée à 40°C et à 70°C en D2. Le volume injecté en 1<sup>ère</sup> dimension est de 5µL. Les données QTOF ont été acquises de 400 m/z à 1000 m/z à un taux d'acquisition de 20 spectres/s. Le nébuliseur, le *sheat gas* et le gaz séchant sont fixés respectivement à 20, 12 et 11 psi. La température du *sheat gas* et du gaz séchant est fixée respectivement à 300 et 200°C.

Les tensions optimisées pour chacune des fractions sont indiquées dans le tableau n°18. Lorsque l'ensemble des fractions d'un deck ont fini d'être analysées, la boucle de passage est nettoyée avec 100% de la phase B et un volume correspondant à 10 fois le volume de la boucle.

	TG / DG / Cer	PG		PE		PI		PC		PS		SM
Polarité de l'ionisation	(+)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Tension du <i>nozzel</i> (V)	1500	2000	-500	2000	-500	2000	-500	250	-1000	2000	-1000	500
Tension du capillaire (V)	5000	5000	-4000	5500	-4000	5000	-4000	4000	-4000	5000	-4500	5000
Énergie de collision (eV)	20	5	-30	10	-25	5	-45	20	-30	10	-35	30

Tableau 18 : Tensions appliquées pour les analyses en LC-LC pour chaque famille de lipides en fonction de la polarité d'ionisation.

## 2.9. Acquisition et retraitement des données 2D

Le contrôle des appareils et l'acquisition des données en LCxLC et LC-LC ont été réalisés avec le logiciel Agilent Openlab® pour la 2D-LC et le logiciel MassHunter® pour la partie spectrométrie de masse. Le retraitement des données a été réalisé avec le logiciel MassHunter qualitative analysis®.

### 3. Résultats et discussion

#### 3.1. Avantage de la chromatographie HILIC en tant que 1<sup>ère</sup> dimension lors des analyses 2D-LC pour la lipidomique

Placer la colonne HILIC en 2<sup>ème</sup> dimension n'est pas la condition idéale, car les temps d'équilibrage des phases stationnaires HILIC sont importants (2 à 4 fois plus long que la RPLC [73]). Cela est un désavantage lorsque l'on cherche à avoir un gradient le plus rapide possible dans la dernière dimension.

De plus, contrairement aux approches où la HILIC est placée en première position, la configuration *comprehensive* RPLCxHILIC ne permet pas d'isoler les fractions de famille de lipides minoritaires à l'issue de la 1<sup>ère</sup> dimension en éliminant des portions de chromatogramme. Il est donc impossible de préconcentrer les échantillons injectés en fonction des familles de lipides minoritaires sélectionnées pour pouvoir gagner en sensibilité. Ce sont les familles lipidiques majoritaires (ex. TG et PC) qui fixent la quantité maximale de matrices injectables pour éviter de saturer et d'encrasser prématurément le détecteur. Les capacités d'isolation des classes lipidiques en fonction de l'ordre des dimensions lors des analyses bidimensionnelles avec une colonne HILIC et RPLC sont illustrées dans la figure n°92. Dans le cas où la colonne HILIC est placée en 1<sup>ère</sup> dimension, comme dans la figure n°92 A, la séparation entre classes de lipides intervient avant la séparation des lipides d'une même famille. Par exemple, les familles de composés PE, PC et SM sont totalement séparées les unes des autres lors de la séparation effectuée en 1<sup>ère</sup> dimension. Il n'y a pas de superposition entre classes de lipide lors de la projection sur l'axe horizontal. Cela conduit à l'obtention de chromatogrammes dans lesquels se trouve une seule famille de lipides à l'issue des séparations dans les deux dimensions. Dans le cas où la colonne HILIC est placée en 2<sup>ème</sup> dimension, la séparation entre classes de lipides intervient après la séparation des différents lipides d'une même famille (cf. figure n°92 B). Par exemple, les familles de composés PE, PC et SM ne sont séparées les unes des autres que lors de la séparation effectuée en 2<sup>ème</sup> dimension. Lors de l'analyse en 1<sup>ère</sup> dimension, ces trois familles sont co-éluées, puisqu'il y a superposition des classes de lipides lors de la projection sur l'axe horizontal. Cela conduit à l'obtention de chromatogrammes dans lesquels figurent plusieurs familles de lipides à l'issue des séparations dans les deux dimensions.

C'est pour l'ensemble de ces raisons que nous avons choisi de conserver la colonne HILIC comme 1<sup>ère</sup> dimension et la colonne RPLC comme 2<sup>ème</sup> dimension. Une fois que le type de chromatographie a été sélectionné (HILIC ou RPLC) pour chacune des dimensions, il faut définir la nature précise des phases stationnaires.

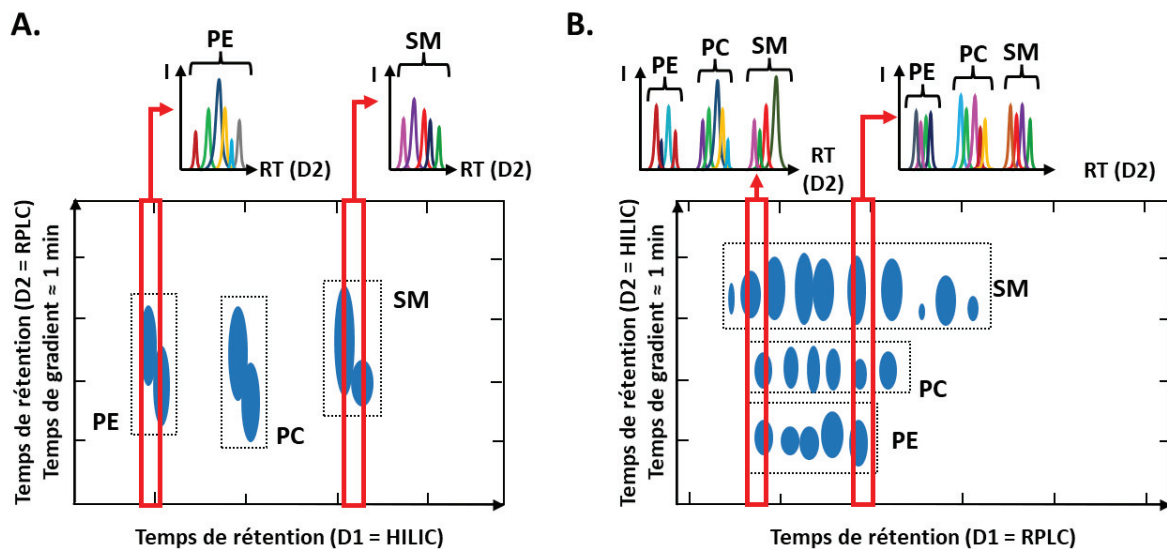
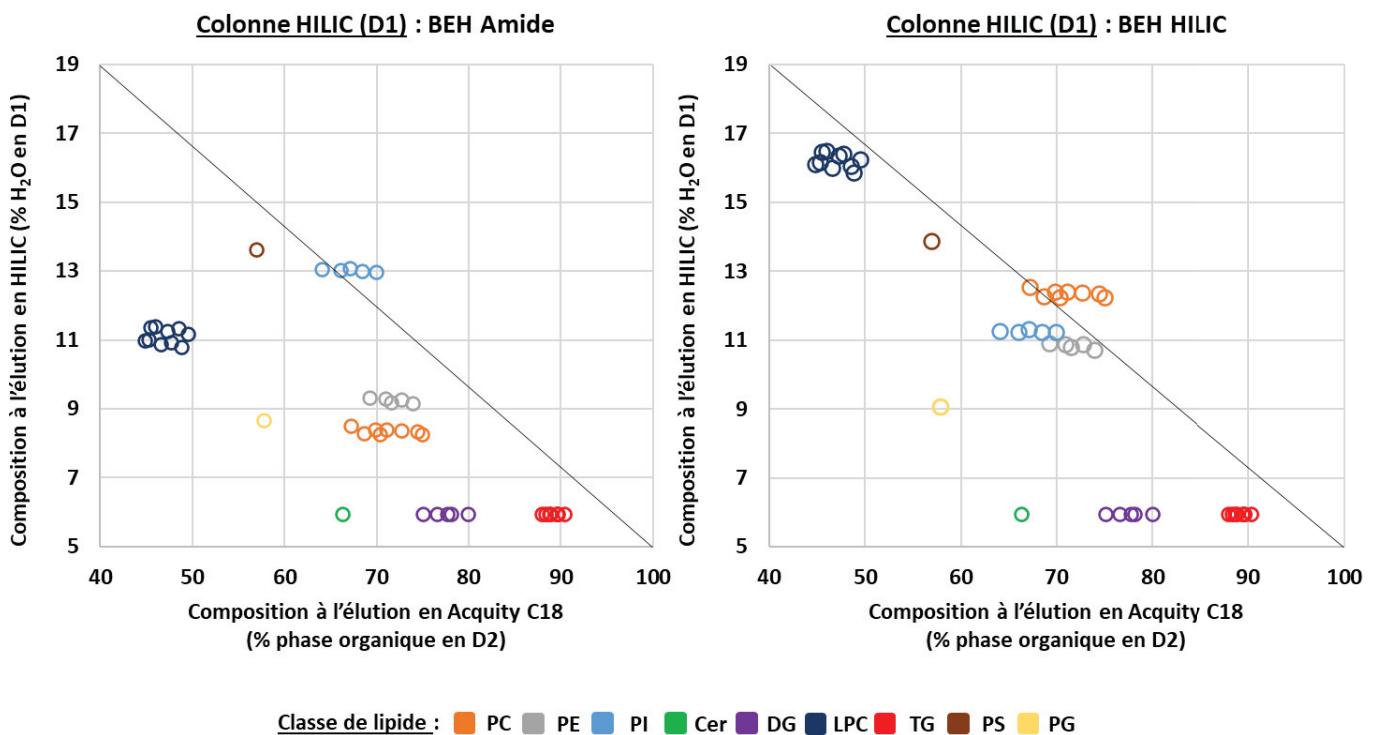


Figure 92 : Capacité d'isolation des classes lipidiques en fonction de l'ordre des dimensions lors des analyses bidimensionnelles avec une colonne HILIC et RPLC. Le rectangle rouge représente l'illustration d'un chromatogramme obtenu à un instant « t » après la séparation d'une fraction à l'issue de la D2. (A) Représentation schématisée de la séparation d'un échantillon contenant des SM, PC et PE avec un système HILICxRPLC. (B) Représentation schématisée de la séparation d'un échantillon contenant des SM, PC et PE avec un système RPLCxHILIC.

### 3.2. Choix de la phase stationnaire en 1<sup>ère</sup> dimension et 2<sup>ème</sup> dimension

Afin de couvrir au maximum l'espace de séparation, il est souhaitable d'avoir la plus grande orthogonalité entre les deux dimensions. Pour cela, des mécanismes de séparations différents et complémentaires permettent d'augmenter les capacités de séparation. Le degré d'orthogonalité correspond donc aux différences de sélectivité entre les deux dimensions. Il peut être évalué comme étant l'espace bidimensionnel occupé par les pics sur le diagramme de composition à l'élution (cf. figure n°93). Plus l'espace est occupé, plus le système chromatographique aura de chance de séparer des composés co-élus en 1<sup>ère</sup> dimension (D1). L'analyse des lipides par chromatographie hydrophile (HILIC) est souvent réalisée sur des colonnes micropores de 100 à 150 mm de longueur pour une taille de particules de 1.7 à 5  $\mu\text{m}$ , possédant comme phase stationnaire de la silice vierge ou de la silice greffée. Les silices greffées changent la sélectivité des colonnes HILIC : elles en modifient la nature des interactions de la phase stationnaire avec les analytes, en accentuant l'affinité avec les liaisons hydrogènes ou en engendrant des interactions électrostatiques fortes [330]. Cette modification de la silice peut être réalisée grâce à des fonctions amides ou diols, comme c'est le cas dans les colonnes chromatographiques Atlantis HILIC Silica, Acquity UPLC BEH amide, Nucleosil 100-5 OH et Spherisorb Si [55]. Afin de sélectionner le type de phase HILIC et ainsi confirmer les travaux de Holcapek *et al.* [133] et de Xu *et al.* [123], nous avons testé deux colonnes dont les propriétés physicochimiques correspondent à leurs études. Ces colonnes sont la colonne BEH Amide (silice greffée) pour les travaux de Xu *et al.* [123] et la colonne BEH HILIC (silice non greffée) du même type de phases que les travaux

2D-LC de Holcapek *et al.* [133]. Pour la phase stationnaire de type inverse qui sera en 2<sup>ème</sup> dimension (D2) nous avons sélectionné une phase Acquity C18, très utilisée pour les analyses lipidiques globales [55]. Afin de sélectionner la meilleure phase HILIC et d'évaluer le degré d'orthogonalité, nous avons comparé les pourcentages de composition à l'élution des 3 phases stationnaires. Ce pourcentage correspond à la quantité nécessaire de solvant possédant un pouvoir éluant élevé (phase aqueuse en HILIC et phase organique en RPLC) pour entrainer un composé hors de la colonne chromatographique. Les compositions à l'élution obtenues avec les phases HILIC en analyse monodimensionnelle ont été corrélées à celles obtenues en RPLC pour former le diagramme des compositions à l'élution présenté sur la figure n°93.



**Figure 93 :** Compositions à l'élution de composés lipidiques en HILIC en fonction des compositions à l'élution de ces mêmes composés en RPLC. Les analyses en HILIC ont été obtenues à partir d'analyses 1D-LC réalisées à partir de deux colonnes différentes : BEH HILIC (2.1x100 1.7µm) et BEH Amide (2.1x100 3.5µm). L'analyse en RPLC a été obtenue à partir d'analyse 1D-LC réalisé avec la colonne Acquity C18 (2.1x100 1.7µm). Les gradients chromatographiques sont normalisés en fonction des dimensions de colonnes.

Cette figure nous montre que nous avons une très mauvaise séparation des composés pour une même famille en chromatographie HILIC puisque tous les lipides d'une même classe se retrouvent élués avec le même pourcentage de phase aqueuse (alignement sur l'axe des ordonnées). Ce qui n'est pas le cas en RPLC où les différents lipides d'une même famille sont élués à des pourcentages de phase organique différents (répartition sur l'axe des abscisses). Lorsque l'on observe les différentes classes, on remarque qu'elles sont bien réparties par rapport à l'axe des ordonnées, ce qui indique une bonne séparation en HILIC. En revanche, cela n'est pas le cas sur l'axe des abscisses avec la superposition des différentes classes de phospholipides (PC, PE, PI, PS et PG). Ainsi la chromatographie de phase inverse

ne permet pas d'avoir une bonne séparation entre classes lipidiques. Comme attendu, ces résultats indiquent une orthogonalité de séparation entre les deux dimensions, car l'association des deux colonnes permet de séparer des composés co-élus lors de l'utilisation d'une seule dimension.

Les deux colonnes HILIC présentent des variations de sélectivités par rapport aux différentes familles de lipides, puisque les familles de composés ne sont pas éluées dans le même ordre et avec la même composition à l'élution (pourcentages de phase aqueuse différents). Nous avons sélectionné la colonne BEH HILIC à base de silice vierge, car elle présente plus de rétention. En effet, les compositions à l'élution pour chaque famille de lipides sont de manière générale plus importantes avec la colonne en silice non greffée (ex. LPC colonne BEH amide = 11% < LPC colonne BEH HILIC = 16%). Cette constatation est intéressante, car cela permet d'obtenir des compositions d'élution moins riches en phases organiques. Ainsi la phase mobile présente plus de compatibilité entre les dimensions. Lorsque l'on souhaite réaliser des analyses 2D-LC en ligne, il est nécessaire de prendre en considération les compatibilités de solvants entre chaque dimension (cf. chapitre 1 partie n°4.2.1). Même si la HILIC et la RPLC utilisent les mêmes solvants, ceux-ci ont des affinités contraires. En HILIC, la phase aqueuse est le solvant fort par rapport aux phases organiques, ce qui n'est pas le cas en RPLC. Donc plus le composé sera élué avec un pourcentage de phase organique élevé en HILIC, plus il y aura de risques que le composé subisse des effets de *breakthrough* dans la colonne de phase inverse. Pour rappel, le *breakthrough* se traduit par des déformations de pics chromatographiques. Cela est dû à une partie des composés qui sont entraînés plus rapidement par le solvant provenant de la 1<sup>ère</sup> dimension [114] comme défini dans le chapitre 1.

Lors des analyses 2D en ligne qui rencontrent ce problème de déséquilibre de la force du solvant, il est nécessaire de diminuer la force d'élution de la phase mobile qui provient de la 1<sup>ère</sup> dimension. Pour cela, il faut limiter la quantité de solvant transféré dans la 2<sup>ème</sup> dimension en divisant le débit avec un *split*, ou en diluant les fractions issues de la 1<sup>ère</sup> dimension en introduisant un co-solvant avec un pouvoir éluant plus faible entre les deux dimensions [114]. Plus le phénomène de *breakthrough* sera important, plus il faudra diviser le débit ou diluer l'échantillon avant injection, ce qui diminuera la sensibilité de la méthode. Étant donné sa simplicité de mise en place, nous avons sélectionné dans notre étude la jonction avec un « T » pour diviser le débit en sortie de 1<sup>ère</sup> dimension afin de ne récupérer qu'une partie de l'éluat. Nous avons donc par la suite évalué le meilleur ratio de *split* à appliquer pour éviter le phénomène de *breakthrough* tout en limitant la baisse de sensibilité.

### 3.3. Étude des effets à l'injection et mise en place du *split*

Afin d'évaluer l'intérêt de placer la colonne HILIC en 1<sup>ère</sup> dimension et la colonne RPLC en 2<sup>ème</sup> dimension, les effets de *breakthrough* ont été comparés selon les configurations RPLC-HILIC et HILIC-RPLC.



### 3.3.1. RPLC en 1<sup>ère</sup> dimension et HILIC en 2<sup>ème</sup> dimension

Pour pouvoir évaluer les phénomènes de *breakthrough*, il est nécessaire de récupérer chaque lipide dans un solvant correspondant à leur composition à l'élution issue de la séparation chromatographique de 1<sup>ère</sup> dimension. Cette étape peut être réalisée à l'aide d'un collecteur de fractions. Dans le cas où la RPLC est utilisée en 1<sup>ère</sup> dimension, l'ensemble des lipides d'une même classe lipidique est élué avec des pourcentages de compositions à l'élution différents. Ainsi, l'utilisation d'une matrice biologique dans le cadre de cette étude nécessiterait autant de collectes qu'il y a de lipides dans l'échantillon, ce qui est très complexe à mettre en place. Pour simplifier cette expérimentation et avoir une idée des effets de *breakthrough*, un standard par famille de lipides a été mis en solution avec une composition de phases correspondant à leur élution en phase inverse. Chaque fraction a ensuite été injectée dans un système HILIC-MS/MS en faisant varier le volume injecté de 1  $\mu\text{L}$  à 10  $\mu\text{L}$  (0.5%, 1%, 2.5% et 5% du volume mort de la colonne) comme présenté dans les figures n°94 et 95.

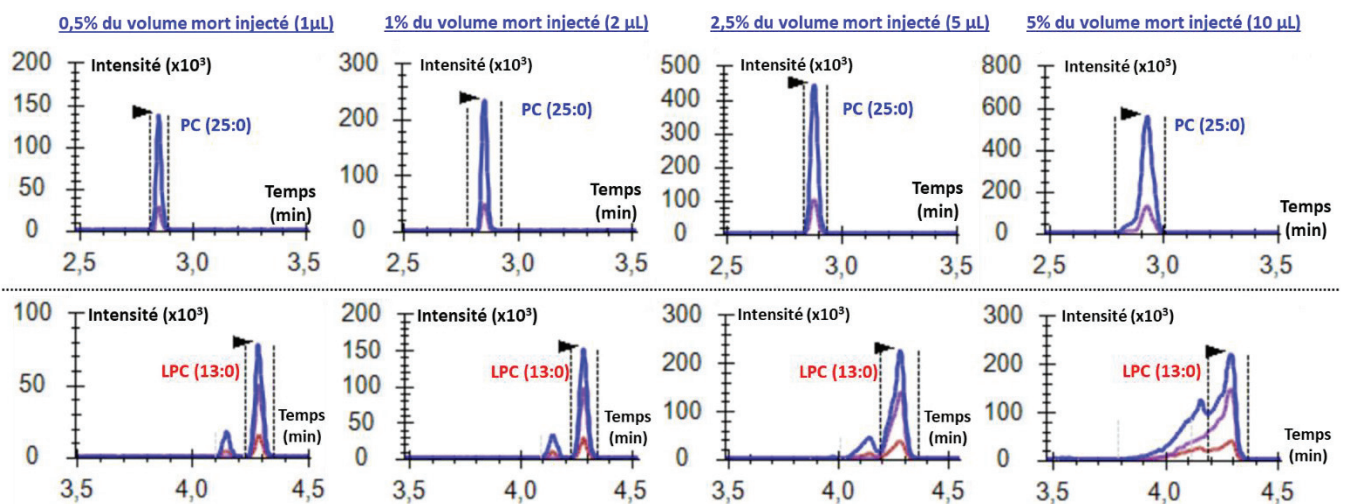


Figure 94 : Évaluation du phénomène de *breakthrough* pour les standards PC et LPC obtenus par RPLC-HILIC hors ligne en faisant varier les volumes d'injection dans la 2<sup>ème</sup> dimension. Les volumes investigués sont 1  $\mu\text{L}$  (0.5% du volume mort), 2  $\mu\text{L}$  (1% du volume mort), 5  $\mu\text{L}$  (2.5% du volume mort) et 10  $\mu\text{L}$  (5% du volume mort).



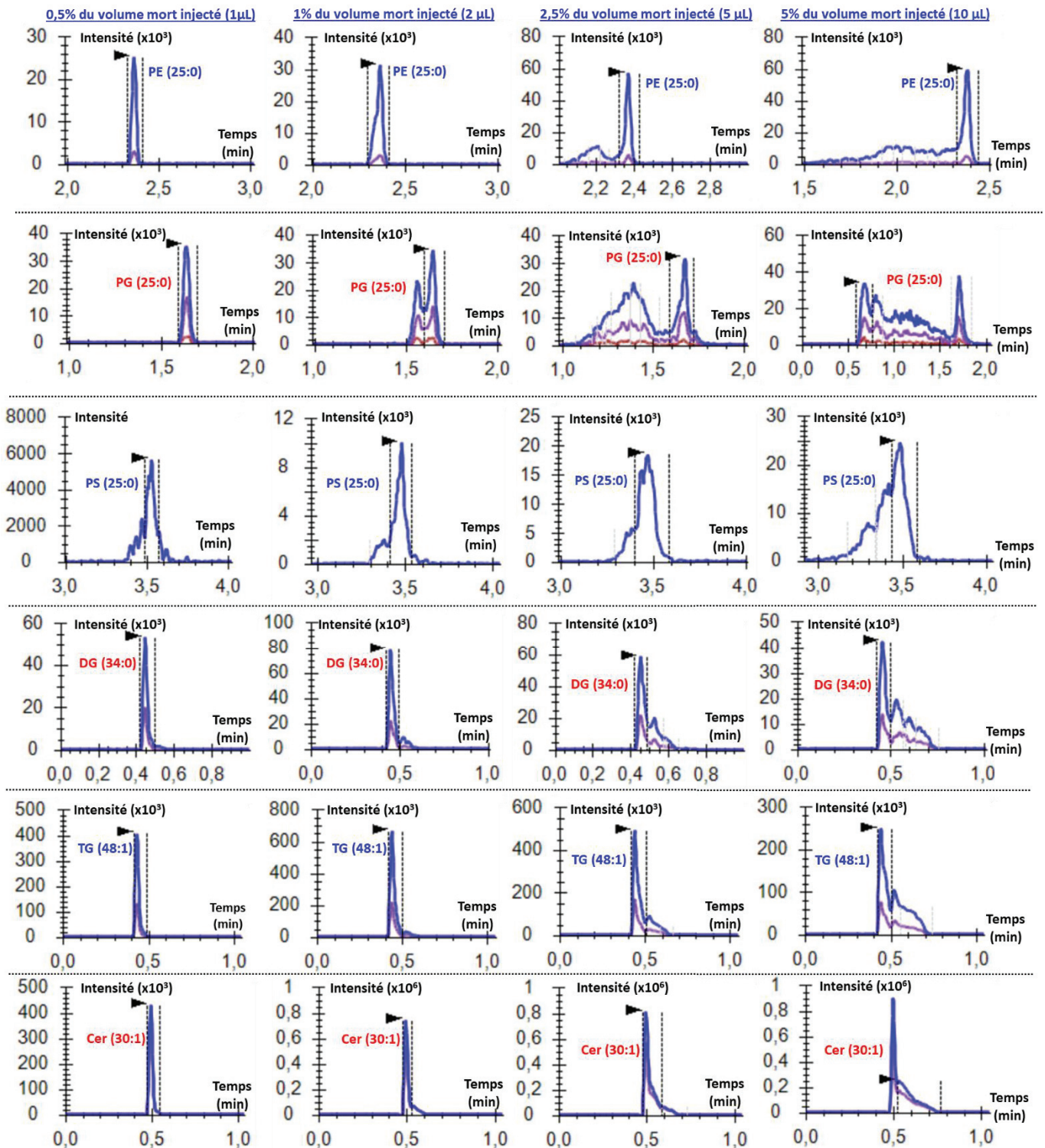


Figure 95 : Évaluation du phénomène de *breakthrough* pour les standards PE, PG, PS, DG, TG et Cer obtenus par RPLC-HILIC hors ligne en faisant varier les volumes d'injection dans la 2<sup>ème</sup> dimension. Les volumes investigués sont 2 µL (1% du volume mort), 5 µL (2.5% du volume mort), 10 µL (5% du volume mort) et 20 µL (10% du volume mort).

Les déformations de pic pour la configuration RPLC en D1 et HILIC en D2 commencent à partir de (i) 1% du volume mort pour les PG et les PE, (ii) 2.5% pour les LPC et PS, (iii) 5% pour les PC, Cer, DG et TG.

### 3.3.2. HILIC en 1<sup>ère</sup> dimension et RPLC en 2<sup>ème</sup> dimension

Dans le but d'évaluer le phénomène de *breakthrough* dans la configuration inverse, nous avons injecté un extrait d'échantillon de matrice lipidique de *G. fossarum* dans un système chromatographique avec une colonne BEH HILIC équipée d'un collecteur de fractions. La matrice de gammare a été utilisée afin d'avoir un mélange complexe contenant de nombreux lipides de familles différentes. Cette expérimentation nous a permis de récupérer chaque famille lipidique dans un solvant qui correspond à la composition à l'éluion issue de la séparation chromatographique de 1<sup>ère</sup> dimension. Chaque fraction a été ensuite injectée dans un système RPLC-MS/MS en faisant varier le volume injecté de 5  $\mu$ L à 40  $\mu$ L (2.5%, 5%, 10%, 15% et 20% du volume mort de la colonne). Les seules familles impactées sont les lysophospholipides dont font partie les LPC. Ils commencent à subir des effets de *breakthrough* de manière significative à partir de 10% du volume mort injecté comme indiqué sur la figure n°96. Ces effets se traduisent par une déformation à l'avant des pics chromatographiques et d'une diminution de la sensibilité d'environ 35% pour les LPC. Lorsque la quantité injectée est doublée entre le passage de 2.5% à 5% du volume mort, le signal augmente bien proportionnellement au volume injecté. Cela n'est plus le cas entre le passage de 5% à 10% du volume mort.

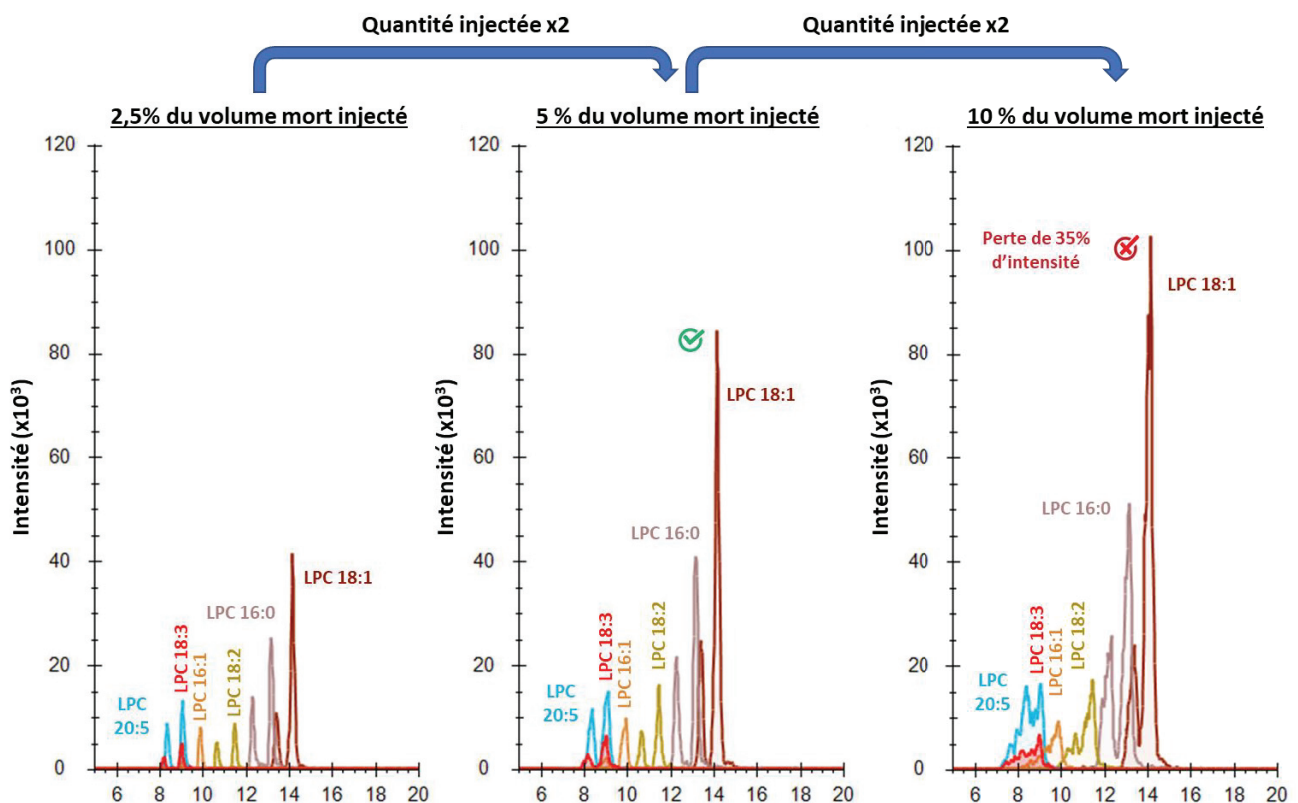


Figure 96 : Évaluation du phénomène de *breakthrough* pour la famille des LPC obtenu par HILIC-RPLC hors ligne en faisant varier les volumes d'injection dans la 2<sup>ème</sup> dimension. Les fractions injectées en RPLC (D2) ont été obtenues grâce à un collecteur de fractions à l'issue de la séparation en HILIC (D1). Les volumes injectés en D2 sont de 5  $\mu$ L (2.5% du volume mort), 10  $\mu$ L (5% du volume mort) et 20  $\mu$ L (10% du volume mort).

Pour le reste des familles testées (Cer, DG, TG, PG, PE, PC, PI, PS et SM), les volumes représentant jusqu'à 20% du volume mort n'entraînent aucune déformation de pic chromatographique (cf. figures annexes n°14 et 15).

Ce phénomène impactant particulièrement les LPC est expliqué par le fait qu'ils font partie des lipides les plus polaires. Ils ont besoin d'un plus faible pourcentage de phase organique pour être retenus en phase inverse. Les 84 % d'acétonitriles transférés de la 1<sup>ère</sup> dimension HILIC (cf. figure n°93) influencent donc plus fortement leurs rétentions. Les lysophospholipides sont ainsi plus sensibles au phénomène de *breakthrough*.

Dans la configuration RPLC en D1 et HILIC en D2, la limite du volume mort injectable est inférieure ou égale à 5%. Dans le cas de la configuration HILIC en D1 et RPLC en D2, la limite du volume mort injectable est de 5% pour les LPC et 20% pour les autres familles de lipides. Ces résultats impliquent que les configurations RPLC en D1 et HILIC en D2 doivent avoir un *split* plus important entre les deux dimensions que pour les configurations inverses. Cela conforte notre choix d'utiliser la HILIC en 1<sup>ère</sup> dimension afin de diminuer le *split* et ainsi conserver une meilleure sensibilité.

Le *split* entre la 1<sup>ère</sup> dimension HILIC et la 2<sup>ème</sup> dimension RPLC devra limiter la quantité de solvant arrivant en 2<sup>ème</sup> dimension. Le volume injecté devra être inférieur ou égal à 5% du volume mort injecté si l'on souhaite conserver les lysophospholipides, dont font partie les LPC. Si on écarte les lysophospholipides, ce pourcentage peut être augmenté jusqu'à 20%, voire même au-delà comme aucune déformation n'a été observée à ce pourcentage. Il serait judicieux de refaire ces analyses en augmentant davantage le volume injecté ( $V_{\text{injecté}} > 40\mu\text{L}$ ) en 2<sup>ème</sup> dimension. Faute de temps cela n'a pas été réalisé.

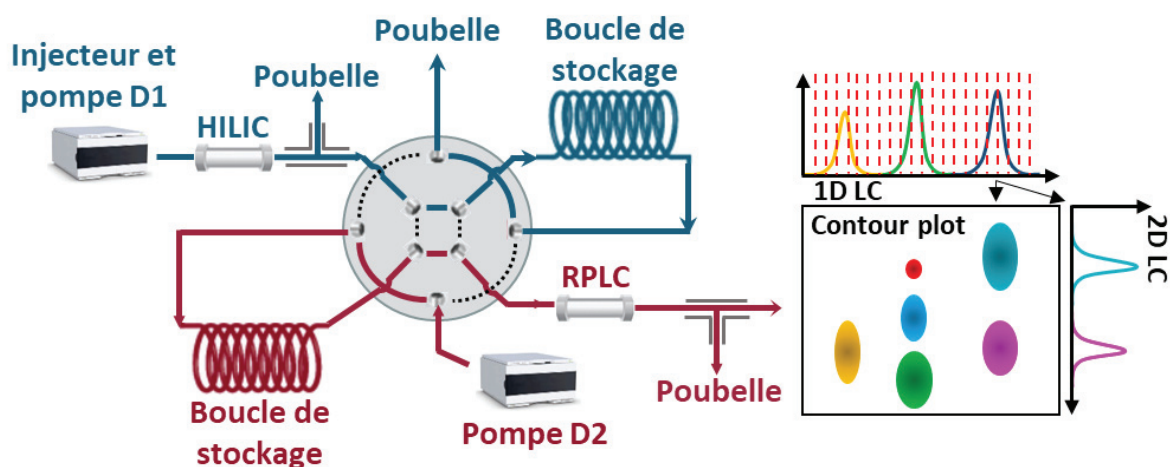
### 3.4. Étude du mode *comprehensive* pour des applications lipidomiques selon la configuration HILICxRPLC-MS/MS

#### 3.4.1. Présentation du système envisagé

Les premières analyses en mode 2D-LC en ligne ont été réalisées en mode *comprehensive* selon la configuration HILICxRPLC. Ces expérimentations ont pour objectif d'évaluer les limitations rencontrées en phase inverse lorsque celle-ci est employée en tant que 2<sup>ème</sup> dimension dans le cadre d'études en lipidomique. Ces informations nous indiqueront les verrous scientifiques qu'il faudra surmonter pour maintenir la phase HILIC en 1<sup>ère</sup> dimension, en rendant compatible la RPLC avec une utilisation en 2<sup>ème</sup> dimension.

Le schéma du montage HILICxRPLC mis en place est présenté dans la figure n°97. Un *split* est placé entre les deux dimensions pour limiter les effets de *breakthrough*, comme cela a été étudié dans le

paragraphe précédent. Le débit en D1 est de 50  $\mu\text{L}/\text{min}$  avec un temps de gradient total de 50 minutes. Le temps de gradients en D2 est de 0.47 minute. Ce temps correspond également au temps de collecte de chaque fraction en D1. Ainsi, le volume collecté par fraction est de 23.5  $\mu\text{L}$ . Sachant que le volume mort de la colonne en D2 a été évalué à 62  $\mu\text{L}$ , le volume maximal pour respecter le critère des 5% du volume mort injecté, induit par les lysophospholipides, est de 3.1  $\mu\text{L}$ . Le *split* appliqué entre les deux dimensions a donc un ratio de 1:10 ce qui permet d'envoyer 2.34  $\mu\text{L}$  en D2 en accord avec les résultats précédents. Le débit dans la 2<sup>ème</sup> dimension est de 1.5 mL/min, ce qui oblige l'ajout d'un second *split* en amont du détecteur en masse. Le ratio appliqué est 1:3 afin d'avoir un débit de 500  $\mu\text{L}/\text{min}$  compatibles avec la source d'ionisation électrospray.



**Figure 97** : Système HILICxRPLC-QTOF. Cette configuration possède un *split* (1:10) entre les deux dimensions et un *split* (1:3) entre la D2 et le détecteur.

Pour maintenir la résolution obtenue dans la 1<sup>ère</sup> dimension, lors du passage dans la dimension suivante, il est important de fractionner plusieurs fois le chromatogramme de 1<sup>ère</sup> dimension dans la largeur des pics chromatographique [331]. Murphy *et al.* ont montré qu'avec trois à quatre fractions par pic de 1<sup>ère</sup> dimension, la résolution obtenue en première dimension était maintenue [331]. Nos conditions chromatographiques nous permettent d'obtenir 4 fractions par pic chromatographique de 1<sup>ère</sup> dimension (*cf.* figure n°98 C).

Les pics chromatographiques de 2<sup>ème</sup> dimension en mode *comprehensive* ont des largeurs de pics de l'ordre de la seconde, ce qui impose d'avoir un détecteur possédant une vitesse d'acquisition importante pour obtenir entre 10 et 15 points par pic afin de bien les définir. De ce fait, le spectromètre de masse utilisé est l'IM-QTOF 6560 de chez Agilent®. Le dimensionnement des conditions chromatographiques (calcul des pentes de gradient, débit, *etc.*) a été réalisé grâce au logiciel non commercial développé par l'équipe chromatographie et techniques couplées dirigée par Sabine Heinisch au sein de l'Institut des Sciences Analytiques (ISA).



Après avoir choisi les deux colonnes chromatographiques ainsi que les *splits* à appliquer, nous avons réalisé les analyses *comprehensives* HILICxRPLC en ligne pour la matrice *G. fossarum*.

### 3.4.2. Mise en évidence d'un manque de résolution en 2<sup>ème</sup> dimension lors de la configuration HILICxRPLC

Afin d'évaluer la configuration HILICxRPLC, un extrait lipidique de l'espèce *G. fossarum* a été analysé. Une attention particulière a été accordée à la séparation entre deux classes lipidiques et à la séparation entre deux lipides d'une même famille qui ne diffèrent que d'une insaturation. Un exemple de séparation des composés PC 34:1, PC 34:2, PE 38:5 et PE 38:6 obtenus par chromatographie bidimensionnelle HILICxRPLC a été comparé à des analyses monodimensionnelles 1D-HILIC et 1D-RPLC dans la figure n°98.

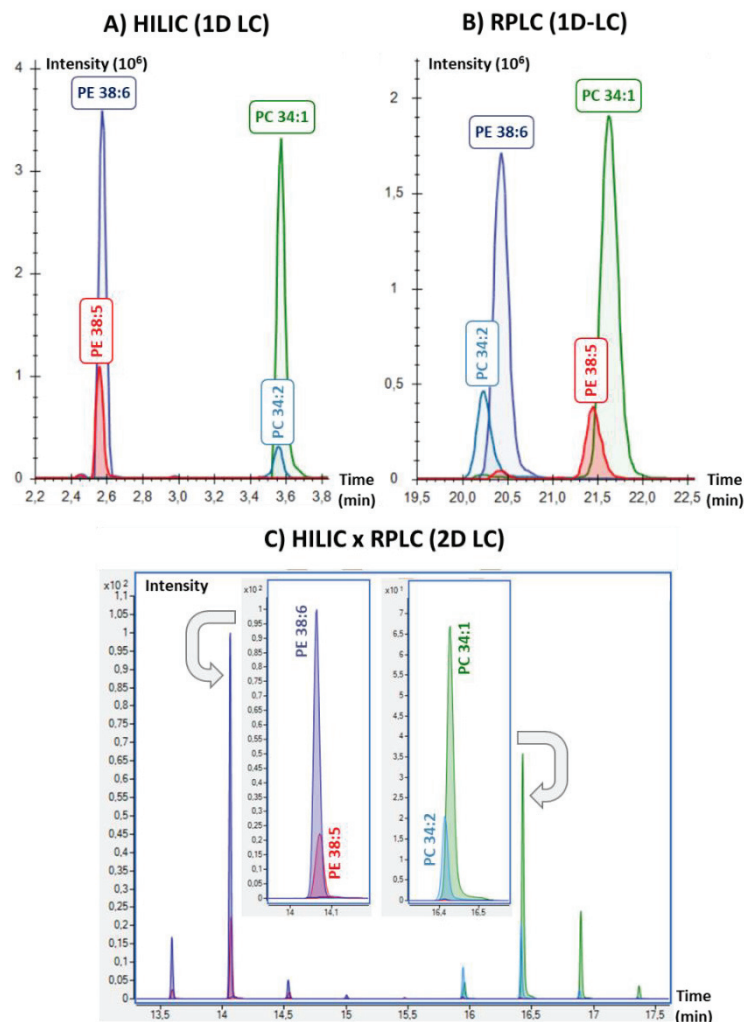


Figure 98 : Comparaison de la séparation (A) HILIC-MS/MS versus (B) RPLC-MS/MS versus (C) HILICxRPLC-MS/MS lors de l'analyse de 4 lipides (PE 38:6, PE 38:5, PC 34:1, PC 34:2) provenant de deux familles lipidiques différentes.

Comme on peut s'y attendre, la colonne HILIC utilisée comme 1<sup>ère</sup> dimension permet bien de séparer les différentes classes lipidiques lors de la séparation bidimensionnelle (*cf.* figures 98 A et 98 C) puisque

la famille des PE est bien séparée des PC. Les résultats en HILIC et HILICxRPLC sont identiques. L'analyse en chromatographie classique 1D-RPLC (cf. figure n°98 B) permet une bonne séparation des lipides d'une même famille. Par contre, lorsque la RPLC est utilisée comme 2<sup>ème</sup> dimension, on observe une mauvaise résolution puisqu'il n'y a pas de séparation entre les lipides d'une même classe (cf. figure n°98 C). Cela indique que le temps de gradient appliqué dans la 2<sup>ème</sup> dimension n'est pas assez long pour pouvoir séparer des lipides qui ne diffèrent que d'une insaturation lors d'une séparation en RPLC. Cette étude nous a permis de confirmer les résultats de Xu *et al.* [123] ainsi que l'affirmation formulée par Holcapek *et al.* [133] concernant les mauvaises performances des colonnes RPLC en tant que 2<sup>ème</sup> dimension lors des analyses LCxLC pour les lipides.

Afin de conserver la HILIC en tant que 1<sup>ère</sup> dimension, il est nécessaire d'allonger le temps de gradient dans la 2<sup>ème</sup> dimension RPLC. Ainsi, il n'est plus envisageable de rester en mode *comprehensive* (LCxLC). Pour cela, nous avons décidé de changer de mode de chromatographie bidimensionnelle en utilisant le mode *multiple heart cut* (LC-LC). Contrairement au mode *comprehensive*, ce mode ne sélectionne que les fractions d'intérêt de la D1 qui sont transférées dans la D2 (cf. chapitre 1 partie 4.1.2.A). Cette configuration possède de nombreuses boucles qui permettent de stocker l'ensemble des fractions contenant les familles lipidiques. Cela permet de ne plus avoir de contrainte de temps dans la 2<sup>ème</sup> dimension. Le temps de gradient en D2 pourra être choisi indépendamment de la D1. Toutefois, le changement du mode LCxLC pour le mode LC-LC soulève la question de la perte d'informations, puisque la totalité des fractions ne subit pas la séparation en 2<sup>ème</sup> dimension et n'est pas envoyée dans le spectromètre de masse. Avant de réaliser le changement de mode 2D-LC, la problématique concernant l'analyse des lysophospholipides (LPL) en chromatographie HILIC en 1<sup>ère</sup> dimension a dû être prise en considération.

### 3.5. Problématiques rencontrées lors de l'analyse des lysophospholipides en chromatographie HILIC de 1<sup>ère</sup> dimension

Avant de développer la méthode HILIC-RPLC-MS/MS, nous nous sommes concentrés sur le cas particulier des lysophospholipides dont font partie les LPC et les LPE, puisqu'il s'agit des familles présentant le plus de désavantages lors de leurs analyses en mode 2D. En effet, comme nous l'avons montré, ils subissent le plus d'effet à l'injection lors du passage de la HILIC vers la RPLC. De plus, comme indiqué sur la figure n°99, les LPC et LPE possèdent de larges fenêtres de fractionnement à cause du dédoublement de chaque pic chromatographique. Cela est dû notamment à la présence d'isomères de position correspondant aux deux positions possibles pour l'unique chaîne d'acide gras sur leur squelette glycérol (cf. chapitre 3 partie 3.2.1). Ce cas de figure en HILIC n'a pas été rencontré pour les autres familles de lipides en dehors des lysophospholipides.

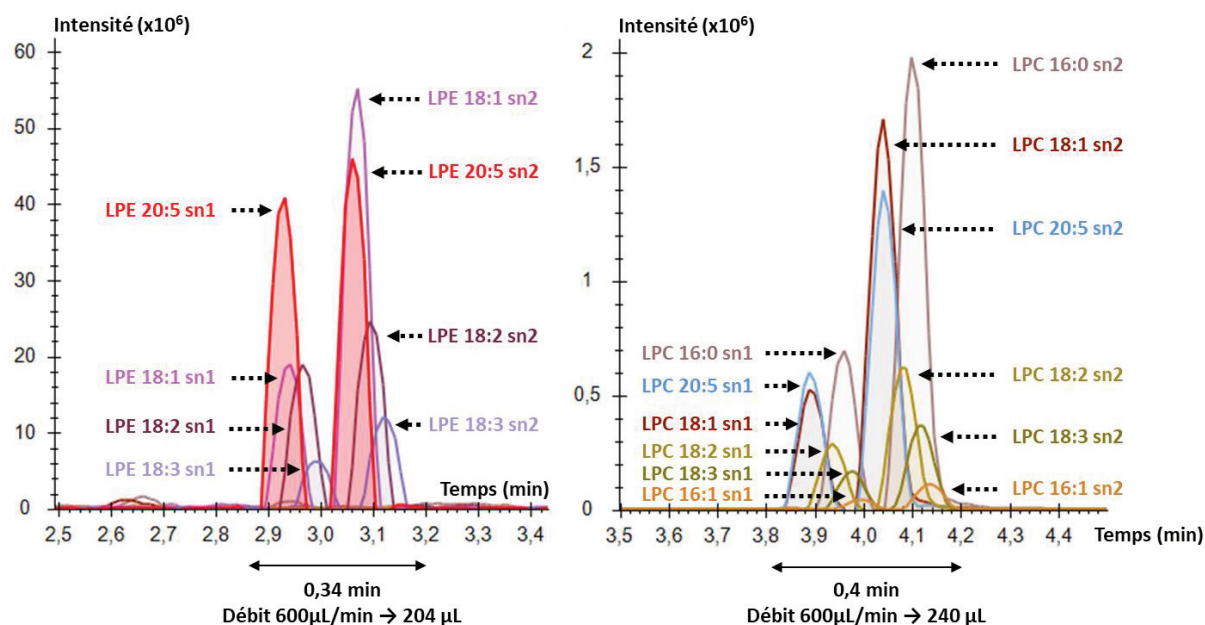


Figure 99 : Mise en évidence de larges fenêtres de fractionnement pour les familles de lysophospholipides (ex. LPE et LPC) lors de leur analyse en HILIC-MS/MS.

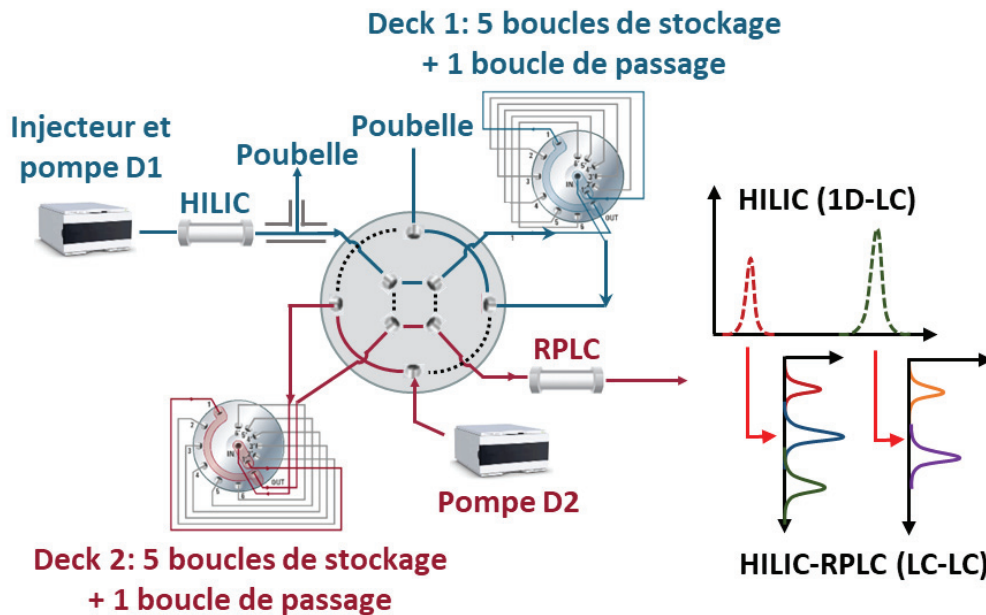
La colonne RPLC (2.1x100 mm et 3.5 µm de taille de particule) utilisée en 2<sup>ème</sup> dimension possède un volume mort de 208 µL. Pour préserver la forme des pics, le volume injecté dans la D2 ne doit pas excéder les 5% du volume mort de la colonne, soit 10.4 µL. Ainsi, avec un débit de 600 µL/min en D1 associé à des fractions larges de 0.4 min pour les LPC et 0.34 min pour les LPE (cf. figure n°99), cela correspond à des fractions respectives de 240 µL et 204 µL. Pour respecter le volume maximal à injecter dans la D2, il faudrait mettre en place un *split* supérieur au ratio 1:20. Cela n'est pas souhaitable et risque d'affecter considérablement la sensibilité. De plus, d'un point de vue pratique il est difficile de mettre en place un *split* supérieur au facteur 10 avec une grande précision. L'autre solution serait de mettre en place un *split* moins important, mais en fractionnant plusieurs fois au sein des familles de lysophospholipides. Cette option n'a pas été retenue, car cela demanderait trop de temps d'analyse. En effet, en appliquant un *split* de ratio de 1:10, il sera nécessaire de subdiviser la famille des LPC en 3 fractions et la famille des LPE en 2 fractions. Dans le cas où le gradient de 2<sup>ème</sup> dimension (RPLC) dure 30 minutes (cf. chapitre 3 partie 3.2.1), cela rajouterait 90 minutes d'analyses. L'analyse des LPC et des LPE en mode 2D-LC induit trop de contraintes. Par ailleurs, leur séparation est satisfaisante lors de l'analyse en 1D-HILIC et 1D-RPLC puisqu'ils sont élués en dehors de la zone d'éluion des phospholipides et des glycérolipides, limitant les effets matrices entre les différentes familles de lipides. De plus, la HILIC ainsi que la RPLC permettent de séparer les isomères de positions au sein des lysophospholipides. Dans le cas de notre méthode 2D-LC, les lysophospholipides ne seront pas considérés et ils seront analysés avec des méthodes 1D classiques indépendantes. Cela permettra d'utiliser un *split* moins important entre les deux dimensions, car les effets à l'injection sont moins importants pour les autres familles lipidiques.



### 3.6. Optimisation de la méthode *multiple heart cut* HILIC-RPLC-MS/MS

#### 3.6.1. Présentation du système envisagé

Théoriquement, le passage du mode LCxLC au mode LC-LC ne fait pas perdre d'information lors des analyses lipidomiques puisque l'utilisation de la colonne HILIC en tant que 1<sup>ère</sup> dimension permet de regrouper l'ensemble des lipides d'une même famille sous un seul pic chromatographique. Ainsi la récupération des lipides est facilitée en récoltant une fraction (ou *cut*) par famille lipidique. Cela permet de réaliser des analyses pouvant être assimilées à du mode *comprehensive*, où les *cuts* ne contenant pas de lipides ne sont pas analysés. Toutefois, ce changement de mode 2D s'accompagne d'une augmentation du temps d'analyse, puisque cette fois-ci le temps total correspond au temps d'analyse de la 1<sup>ère</sup> dimension accompagné de la somme des temps d'analyse de chaque fraction envoyée en 2<sup>ème</sup> dimension. Le système mis en place pour les analyses *multiple heart cut* est présenté sur la figure n°100. Il possède deux decks qui contiennent chacun cinq boucles de stockages, pouvant récolter jusqu'à 120 µL d'échantillon, et une boucle de passage. Le rôle de cette boucle de passage est de laisser passer les portions de chromatogramme qui ne présentent pas d'intérêt. Il est donc possible de récupérer jusqu'à 10 fractions en une seule injection.



**Figure 100 : Système HILIC-RPLC/IM-QTOF.** Cette configuration possède un split (1:5) entre les deux dimensions. Les 2 decks possèdent chacune 5 boucles de stockages et 1 de boucle de passage pouvant stocker jusqu'à 120 µL.

Comme nous allons le voir, il y a à ce jour un certain nombre de contraintes concernant le pilotage du logiciel 2D-LC qui perturbent la collecte des fractions et le gradient de 1<sup>ère</sup> dimension.

### 3.6.2. Mise au point du gradient HILIC en 1<sup>ère</sup> dimension

#### 3.6.2.A. Fonctionnement de la collecte des fractions et application à l'approche HILIC-RPLC pour l'analyse lipidomique

Les conditions chromatographiques que nous avons initialement mises au point pour la 2D en mode *comprehensive* peuvent être réoptimisées pour la 2D-LC en mode *multiple heart cut*. En effet, le mode LCxLC utilisait un gradient de 50 minutes avec un faible débit (50 µL/min) qui induisait des pics assez larges d'environ 1.8 minute (cf. 3.4.1). Dans le cas des approches *multiple heart cut* il est nécessaire d'avoir des pics plus fins pour faciliter la récolte des fractions. Avant d'optimiser les conditions chromatographiques de la 1<sup>ère</sup> dimension, les conditions de fractionnement ont été étudiées, notamment avec le fonctionnement des decks et des boucles de stockages. L'exemple détaillé ci-dessous va expliquer le fonctionnement des decks de stockages via l'utilisation du logiciel de pilotage 2D Openlab de chez Agilent®. Dans le cas du fractionnement envisagé pour les lipides, il sera réalisé avec une fraction unique par classe de lipides et une pause entre chaque fraction (temps pendant lequel aucune fraction n'est collectée), comme présenté sur l'exemple de la figure n°101.

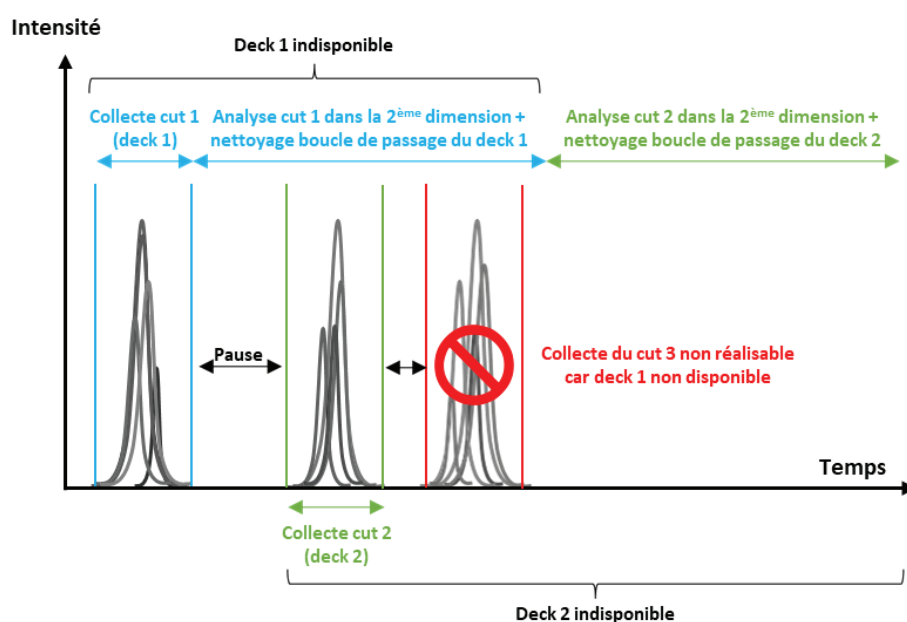


Figure 101 : Fonctionnement du fractionnement en LC-LC lorsque trois fractions uniques sont séparées par une pause.

Il serait possible de supprimer ces pauses entre chaque fraction en réalisant un fractionnement continu, pour avoir un mode *multiple heart cut* pouvant s'apparenter à un mode *comprehensive* avec des fenêtres de fractionnement larges. Toutefois, le logiciel de pilotage 2D impose le même gradient en 2<sup>ème</sup> dimension pour l'ensemble des fractions issues d'une même injection. Ainsi le temps d'analyse en D2 des fractions correspondant aux familles lipidiques et celles se situant entre ces classes de lipides seraient identiques. L'analyse des fractions se situant entre 2 familles de lipides engendrerait des

analyses en 2<sup>ème</sup> dimension inutiles, puisqu'aucun lipide d'intérêt n'est collecté, tout en augmentant drastiquement le temps d'analyse. De plus, le nombre de fractions collectées simultanément est limité à 10 boucles, ce qui conduirait à une saturation prématurée des boucles, empêchant le stockage de fractions ultérieures. Cette solution n'a pas été retenue pour la suite du développement de la méthode HILIC-RPLC, et les pauses entre les fractions lipidiques ont été maintenues.

La programmation du logiciel pilotant la vanne de fractionnement considère qu'un deck est complet dès que l'ensemble des boucles sont remplies ou qu'une pause entre deux collectes est appliquée. De plus, les analyses en 2<sup>ème</sup> dimension ne peuvent commencer que lorsqu'un deck est rempli.

Lorsque 3 fractions sont récoltées avec deux decks possédant chacun 5 boucles de stockage et 1 boucle de passage (cf. figure n°101), le pilotage des vannes est réalisé de la façon suivante :

- (1) la 1<sup>ère</sup> fraction est récoltée dans la boucle 1 du deck 1.
- (2) Dès que la 1<sup>ère</sup> fraction est collectée, son analyse dans la 2<sup>ème</sup> dimension débute directement puisqu'une pause entre deux fractionnements est rencontrée. Le deck 1 est considéré occupé tant que l'analyse dans la D2 n'est pas terminée.
- (2') En parallèle de l'analyse de la 1<sup>ère</sup> fraction dans la D2, la 2<sup>ème</sup> fraction est collectée dans la boucle de stockage 1 du deck 2 en attendant d'être analysée.
- (3) Lorsque la fraction stockée dans le deck 1 a été analysée, une procédure de nettoyage de la boucle de passage du deck 1 est réalisée.
- (3') En parallèle de ce nettoyage, la 2<sup>ème</sup> fraction stockée dans le deck 2 est analysée dans la D2. Le deck 2 est considéré occupé tant que l'analyse dans la D2 n'est pas terminée.
- (4) Lorsque la 2<sup>ème</sup> fraction stockée dans le deck 2 a été analysée, une procédure de nettoyage de la boucle de passage du deck 2 est effectuée.
- (4') En parallèle, la 3<sup>ème</sup> fraction est collectée dans la boucle de stockage 1 du deck 1 en attendant d'être analysée.
- (5) Dès que la 3<sup>ème</sup> fraction est collectée, son analyse dans la D2 débute puis lorsque la fraction a été analysée, une procédure de nettoyage de la boucle de passage du deck 1 est réalisée.

Dans les conditions où une pause est rencontrée entre chaque collecte, une seule boucle est utilisable par deck, ce qui limite drastiquement le nombre de fractions à collecter. D'autant plus que la 3<sup>ème</sup> fraction ne peut être collectée que si son temps de début du fractionnement est supérieur au temps de fin d'analyse de la 1<sup>ère</sup> fraction dans la 2<sup>ème</sup> dimension, auquel il faut associer le temps de nettoyage de la boucle de passage du deck 1. Avant de collecter dans une nouvelle boucle, il faut être sûr qu'elle soit de nouveau disponible, comme expliquée dans l'exemple de la figure n°101.

Dans notre étude, le temps de gradient total de la 2<sup>ème</sup> dimension doit être supérieur à 10 minutes, pour laisser suffisamment de temps aux lipides d'une même classe qui ne diffèrent que d'une insaturation d'être résolus (*cf.* chapitre 3 partie 3.3.2). Cependant, la séparation entre deux familles de lipides dans la D1 est bien inférieure au temps de gradient global de la D2. Ainsi le premier deck ne sera pas disponible lors de l'élution de la 3<sup>ème</sup> fraction ainsi que pour les fractions suivantes.

Pour pallier à ce problème de programmation de la vanne de fractionnement, nous avons choisi d'utiliser une 1<sup>ère</sup> dimension extrêmement rapide pour réaliser des injections multiples. Chaque injection permettra de collecter une ou deux fractions au maximum, qui seront analysées en D2. Il faudra donc plusieurs injections pour analyser la totalité des fractions en 2<sup>ème</sup> dimension. Grâce au gradient rapide en D1, le temps d'analyse global ne sera pas trop allongé comparé à l'injection unique. Toutefois, cette stratégie ne sera pas compatible avec les échantillons « précieux », où l'on a de faibles quantités à disposition.

### 3.6.2.B. Mise en place du gradient de séparation rapide en 1<sup>ère</sup> dimension

Afin de réduire le temps d'analyse en 1<sup>ère</sup> dimension, les conditions chromatographiques de la colonne BEH HILIC (2.1x100 mm et 1.7µm de taille de particules) utilisée lors de l'approche LCxLC, ont été optimisées (débit et pente du gradient). Les nouveaux paramètres (pente de 1% et débit de 600 µL/min) permettent de séparer les différentes classes de lipides avec un gradient d'élution de 15 minutes, équilibrage compris. Ces conditions sont compatibles avec notre stratégie de multiples injections puisque la totalité des lipides est éluée en moins de 4 minutes (*cf.* figure n°102).

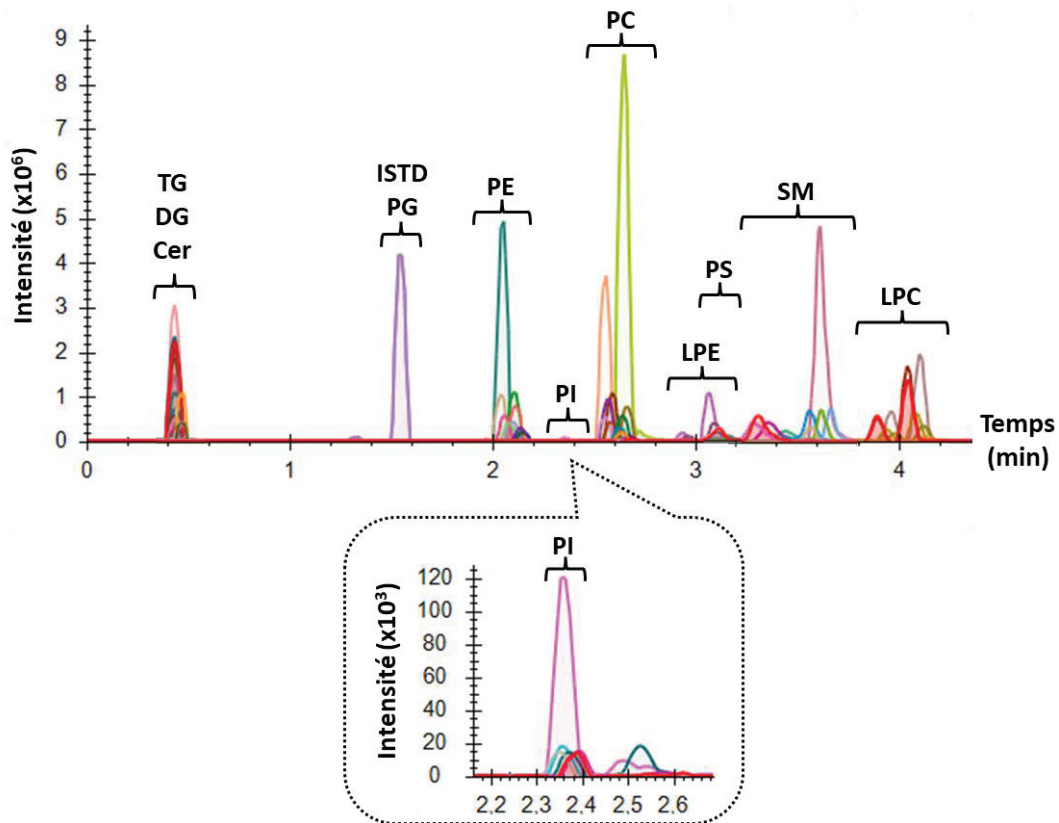


Figure 102 : Séparation des différentes familles lipidiques dans la 1<sup>ère</sup> dimension en mode HILIC lors de l'analyse d'un extrait de *G. fossarum*.

Même si les LPL (ex. LPC et LPE) ne sont plus considérés dans notre stratégie analytique, il est important de vérifier qu'ils n'interfèrent pas avec le fractionnement des autres classes lipidiques. Comme on peut le voir sur la figure n°102, les familles des TG, DG et Cer ne présentent pas de rétention, comme attendu puisqu'ils n'ont pas de tête polaire. Ils seront donc récoltés dans la même fraction. Chacune des autres fractions pourra être récoltée indépendamment les unes des autres, sauf pour les PS et LPE qui devront être récoltés dans la même fraction. La superposition des familles de TG, DG et Cer ainsi que la superposition des LPE et PS n'est pas problématique puisque la RPLC qui est utilisée en 2<sup>ème</sup> dimension est capable de les séparer.

Pour la suite de ces travaux, il est également important de s'assurer que tous les lipides appartenant à une même classe (nombre de carbones et d'insaturations différents) sont bien élués sous un seul et même pic chromatographique.

### 3.6.2.C. Influence du nombre de carbone sur les chaînes d'acides gras des lipides lors de la séparation en chromatographie HILIC

Dans le cadre des essais réalisés sur *G. fossarum*, l'ensemble des lipides d'une même famille, composée d'une combinaison d'acides gras compris entre 14 et 20 carbones, sont élués dans la même zone d'éluion lors de leurs séparations en HILIC. La littérature en lipidomique conforte ces résultats, car il

est référencé qu'en HILIC il est possible d'avoir un seul standard interne (ISTD) par famille, puisque l'ensemble des lipides d'une même classe est élué dans la même zone d'éluion avec le même environnement d'ionisation [53,58]. Cependant, il faut nuancer cette affirmation, étant donné que cela est vrai seulement pour une certaine gamme de lipides qui présentent un certain nombre de carbones. Par exemple, lors de l'analyse de la matrice de gammare enrichie avec des standards de lipides à chaîne courte (combinaison de chaîne d'acides gras inférieurs à 14 carbones), une différence de temps de rétention non négligeable par rapport à la zone d'éluion attendue a été observée. Comme on peut le voir sur la figure n°103, les standards de lipide possédant des chaînes d'acide gras plus courtes, représentés en couleurs, ne sont pas élués dans la zone correspondant à leur famille lipidique, dans laquelle se trouvent les lipides endogènes représentés en noir et blanc. Des co-élutions entre familles sont observées, avec par exemple le standard PE 25:0 co-élué avec la famille des PC et le standard PC 25:0 co-élué avec la famille des LPE.

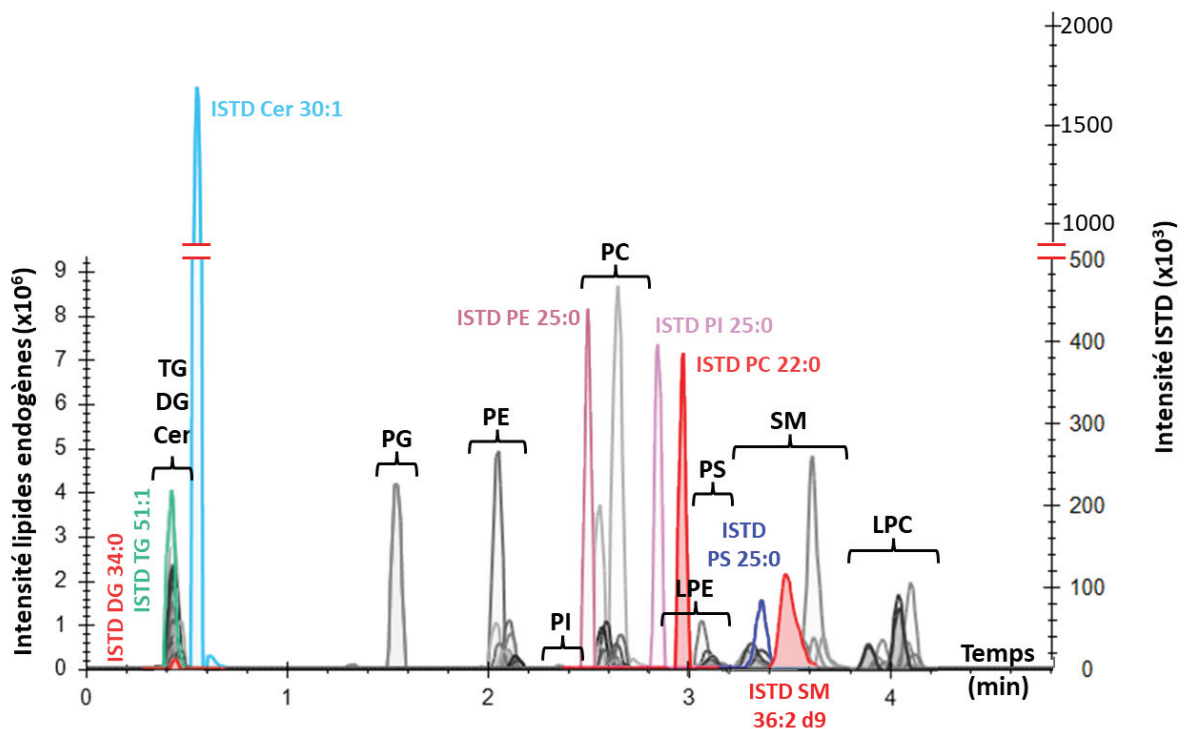


Figure 103 : Séparation des différentes familles lipidiques du gammare dans la première dimension en mode HILIC avec ajout de standards à chaîne d'acides gras courts par classe de lipides. Composition des standards lipidiques : DG 17:0/17:0, TG 17:0/17:1/17:0 d5, Cer d18:1/12:0, PE 12:0/13:0, PI 12:0/13:0, PC 11:0/11:0, PS 12:0/13:0, SM d18:1/18:1 d9.

Ce phénomène pourrait être une véritable limitation pour l'approche 2D-LC que nous développons. En effet, si l'échantillon présente des lipides avec des nombres de carbones très différents pour une même famille, on risque de ne plus les retrouver dans la même fraction. Ils seront co-élusés avec d'autres classes de lipides. Néanmoins, cette diversification en nombre de carbone au sein d'une même famille lipidique n'est pas couramment rencontrée, puisque les matrices biologiques sont

majoritairement constituées de lipides à chaîne d'acides gras comportant entre 14 et 22 atomes de carbone [198,294–298].

En conclusion, ces résultats montrent que les standards internes doivent être judicieusement sélectionnés pour être élués dans la même fraction que les lipides endogènes de la même classe lipidique. Par exemple, il faut choisir des ISTD possédant des longueurs de chaîne « standard » comprises entre 14 et 22 carbones. Il faut également que ce soit des ISTD isotopiquement enrichies (ex. PC 18:0-18:0 d9) ou des lipides avec au moins une chaîne d'acides gras possédant un nombre impair d'atomes de carbone (ex. PC 17:0-17:0), puisque dans la nature les chaînes d'acides gras paires sont prédominantes [290].

Après avoir optimisé la 1<sup>ère</sup> dimension de séparation HILIC, nous avons optimisé les séparations chromatographiques de la 2<sup>ème</sup> dimension ainsi que les paramètres du spectromètre de masse.

### 3.6.3. Mise au point du gradient RPLC en 2<sup>ème</sup> dimension

L'un des désavantages de l'approche *multiple heart cut* comparé aux approches *comprehensives* est le temps d'analyse, qui est drastiquement augmenté en multipliant le nombre de fractions analysées en 2<sup>ème</sup> dimension. Pour réduire le temps global, le gradient de 2<sup>ème</sup> dimension a été optimisé et adapté à chaque classe de lipides. Initialement, le gradient utilisé était le même que la méthode Scout-MRM (cf. chapitre 3). Il avait pour but d'étudier l'ensemble des classes lipidiques en une seule analyse. Le temps total était de 30 minutes comme rappelé sur la figure n°104. Il est important de noter que le logiciel d'acquisition (Agilent® Openlab 2D), permettant le pilotage des méthodes bidimensionnelles, oblige l'utilisation d'un seul et même gradient d'élution pour l'ensemble des fractions collectées lors d'une même injection dans la 1<sup>ère</sup> dimension. Toutefois, comme décrit précédemment (cf. partie 3.6.2), la méthode LC-LC développée dans cette étude utilisera une approche avec injections multiples. Ainsi cette limitation sera surmontée en adaptant le gradient pour chaque injection correspondant à une fraction spécifique. De plus, les LPL n'étant plus considérés dans cette approche 2D, il sera possible de gagner en temps d'analyse, car le gradient d'élution pourra démarrer avec un pourcentage de solvant éluant plus important.



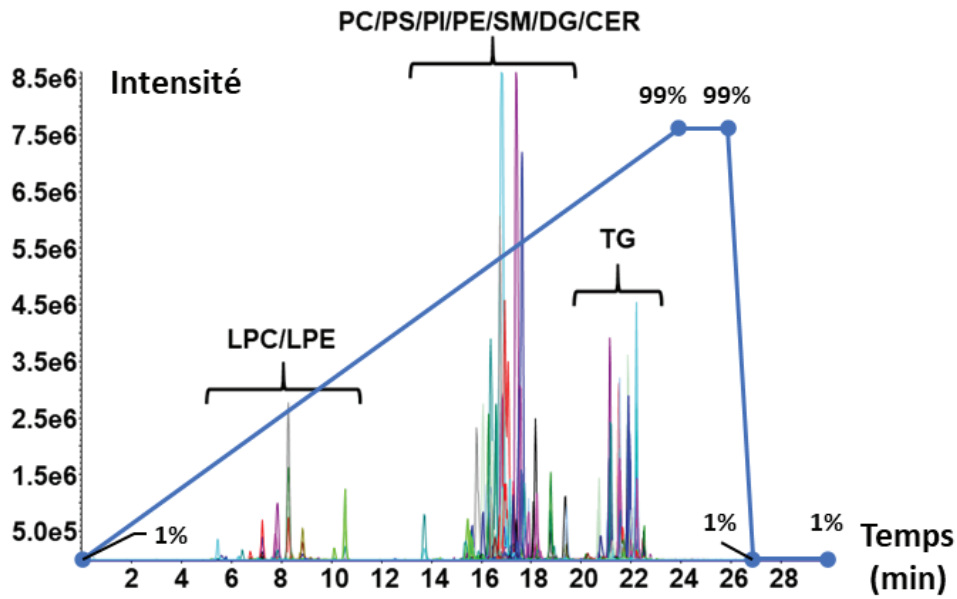


Figure 104 : Séparation chromatographique d'origine pour les familles lipidiques (LPL / PL / SM / Cer / DG / TG) avec la méthode RPLC-MS/MS en mode ESI (+). Le tracé en bleu représente le gradient chromatographique avec le pourcentage de phase B (IPA/H<sub>2</sub>O/ACN 90/9.5/0.5 ; v/v/v + 0.1% acide formique + 10 mM formiate d'ammonium) au cours du temps. La phase A est constitué du mélange H<sub>2</sub>O/ACN 60/40 ; v/v + 0.1% acide formique + 10 mM formiate d'ammonium.

Deux gradients types avec un temps total inférieur à 15 minutes ont été mis en place. Le 1<sup>er</sup> gradient possède une pente allant de 55% à 80% de phase B pour les phospholipides et pour les sphingomyélines (cf. figure n°105).

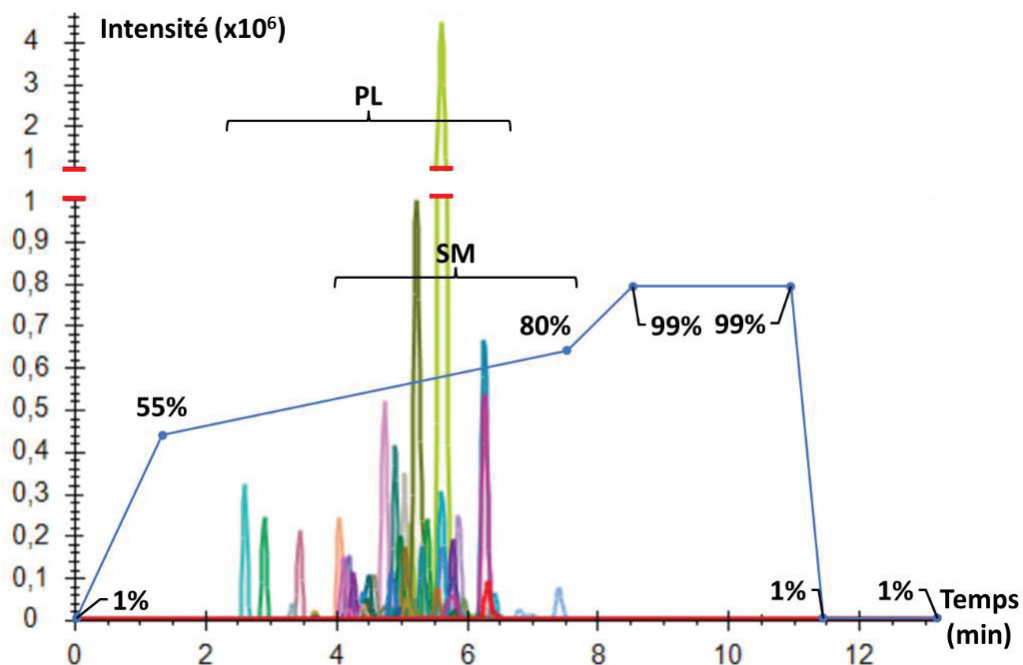
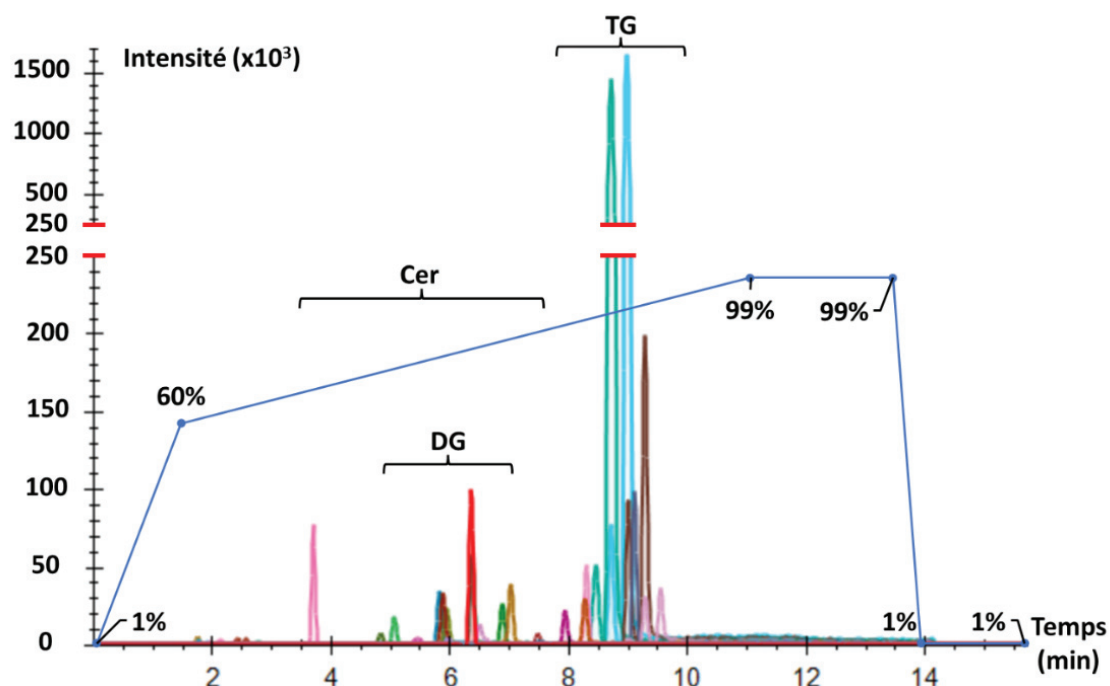


Figure 105 : Séparation chromatographique optimisée pour les familles lipidiques (PL / SM) avec la méthode RPLC-MS/MS utilisée en deuxième dimension en mode ESI (-). Le tracé en bleu représente le gradient chromatographique avec le pourcentage de phase B (IPA/H<sub>2</sub>O/ACN 90/9.5/0.5 ; v/v/v + 0.1% acide formique + 10 mM formiate d'ammonium) au cours du temps. La phase A est constitué du mélange H<sub>2</sub>O/ACN 60/40 ; v/v + 0.1% acide formique + 10 mM formiate d'ammonium.

Le deuxième gradient a une pente allant de 60% à 99% de phase B pour les triglycérides, diglycérides et les céramides (cf. figure n°106).

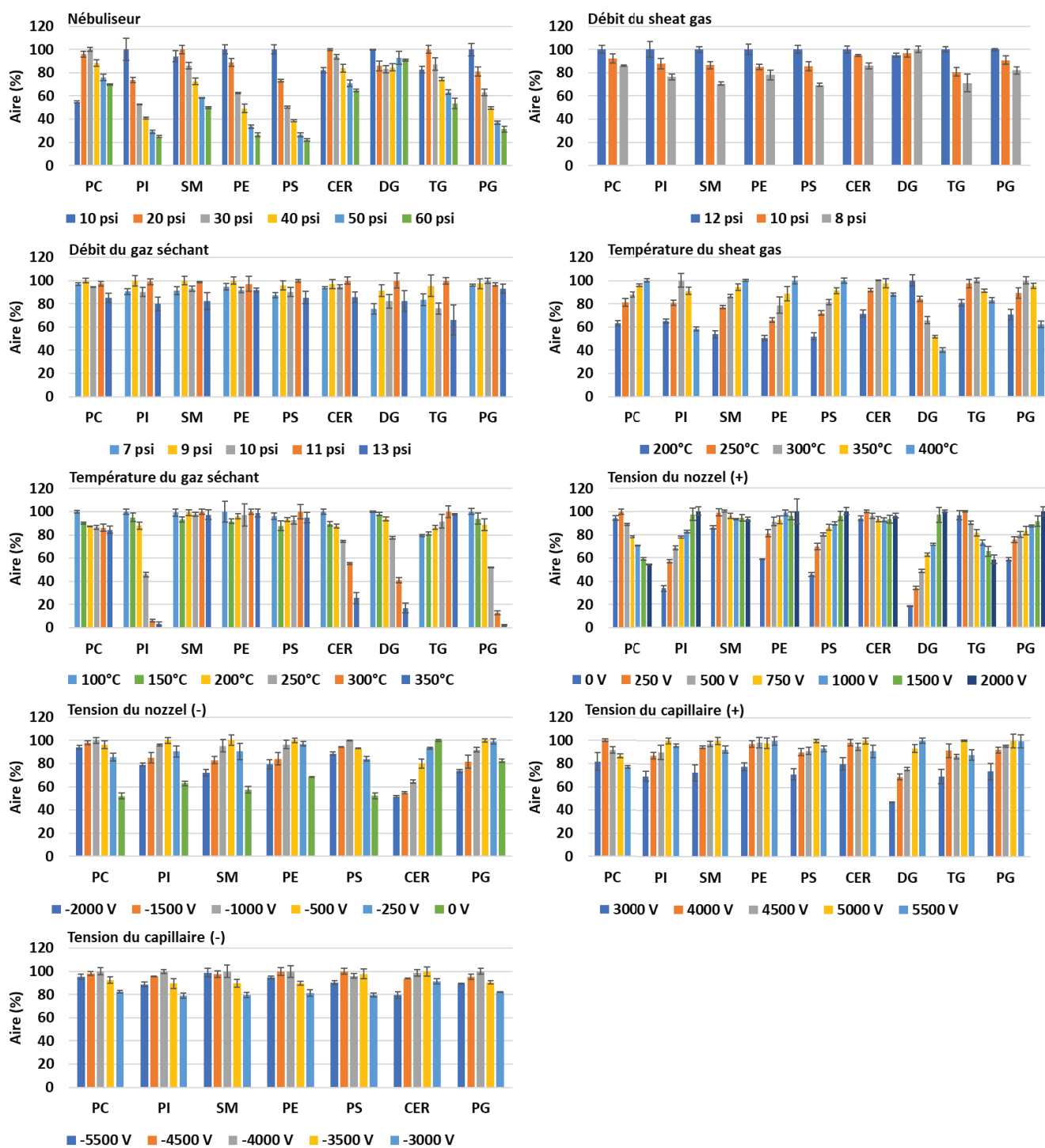


**Figure 106 : Séparation chromatographique optimisée pour les familles lipidiques (TG / DG / Cer) avec la méthode RPLC-MS/MS utilisée en deuxième dimension en mode ESI (+).** Le tracé en bleu représente le gradient chromatographique avec le pourcentage de phase B (IPA/H<sub>2</sub>O/ACN 90/9.5/0.5 ; v/v/v + 0.1% acide formique + 10 mM formiate d'ammonium) au cours du temps. La phase A est constituée du mélange H<sub>2</sub>O/ACN 60/40 ; v/v + 0.1% acide formique + 10 mM formiate d'ammonium.

Les deux gradients ont les mêmes compositions initiales et finales en phase B afin de faciliter le passage de l'un à l'autre sans temps d'équilibrage supplémentaire. Les nouveaux gradients utilisés pour la séparation des composés dans la 2<sup>ème</sup> dimension respectent les critères listés dans le chapitre 3 partie 3.2.2 de ce manuscrit, à savoir la bonne séparation entre deux composés d'une même famille qui diffèrent d'une insaturation.

### 3.6.4. Optimisation des paramètres de masse du QTOF

Le système permettant de réaliser les analyses *multiple heart cut* en ligne est couplé à un spectromètre de masse de type QTOF (IM-QTOF 6560 de chez Agilent®). Cet instrument étant récent au laboratoire, une étude rapide des conditions d'ionisation et de fragmentation a été nécessaire. Les paramètres du spectromètre de masse pour chacune des familles de lipides ont été optimisés. Pour cela, un mélange de standards a été injecté dans le spectromètre de masse avec le système chromatographique en mode FIA (*flow injection analysis*), c'est-à-dire sans colonne chromatographique. Le débit est fixé à 500 µL/min comme pour les conditions de 2<sup>ème</sup> dimension. Les résultats pour l'ensemble des paramètres de masse sont représentés sur les figures n°107 et 108.



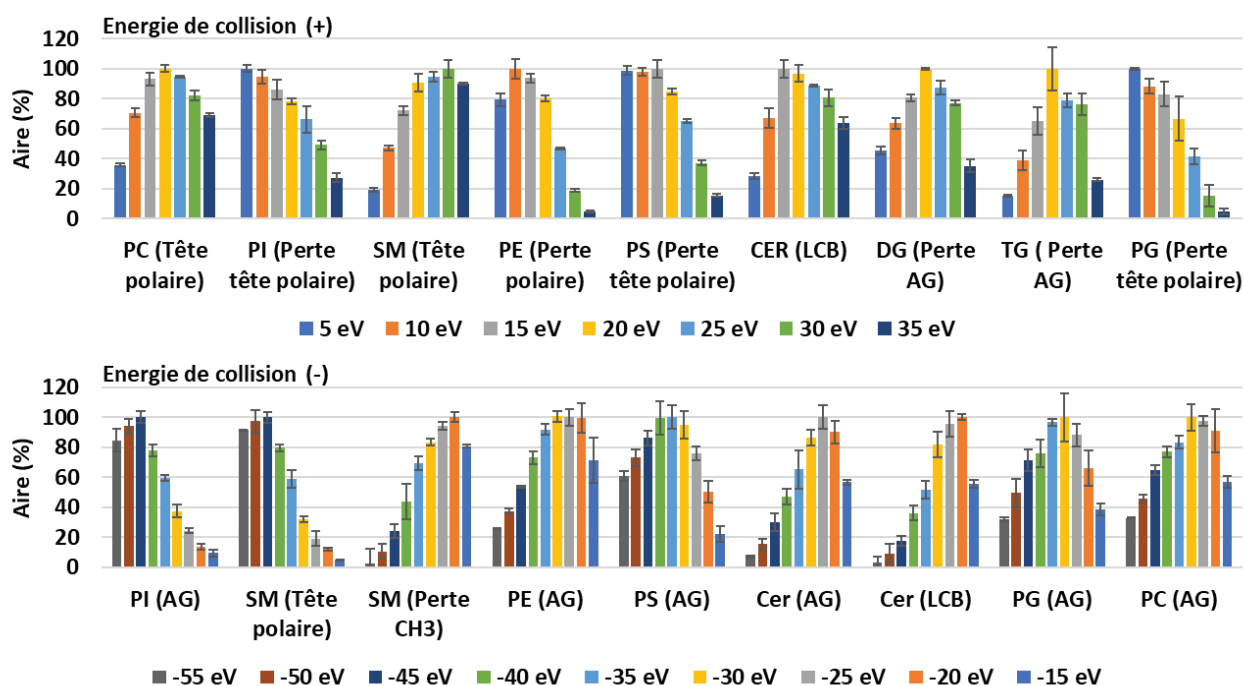


Figure 108 : Optimisation des énergies de collision en mode ESI (+) et ESI (-) par famille lipidique sur un système IM-QTOF 6560 de chez Agilent® avec un débit de 500 µL/min lors de l'ionisation (n=3). Les indications entre parenthèses indiquent le type de fragment.

Comme cela est illustré, les valeurs optimales pour chacun de ces paramètres peuvent être très différentes d'une famille lipidique à une autre. Il n'existe aucune condition unique qui soit favorable à l'ensemble des lipides. Par exemple, l'augmentation de la tension du *nozzle voltage* entre 1500 V et 2000 V en ESI (+) permet d'améliorer la sensibilité pour la majorité des lipides, et notamment celles des DG qui gagnent 70% du signal par rapport à la condition 250V. Toutefois, cela se fait au détriment des PC et des TG qui perdent 40% de leur signal par rapport à leur condition optimale, qui est de 250V. Ces résultats démontrent le besoin d'optimiser les paramètres de masse pour chacune des classes lipidiques pour atteindre les meilleures sensibilités. De la même manière que la mise en place du gradient d'élution, le logiciel de pilotage du système 2D oblige l'expérimentateur à utiliser une méthode de masse unique pour l'ensemble des fractions analysées par injection. L'approche utilisant les injections multiples permet de surmonter le problème des gradients d'élution uniques, mais aussi d'adapter les paramètres de masse pour chaque fraction collectée. Les paramètres de masse utilisés pour la mise en place de la méthode *multiple heart cut* HILIC-RPLC-QTOF en ligne sont résumés dans le tableau n°19.

	TG / DG / Cer			PG		PE		PI		PC		PS		SM
Polarité de l'ionisation	(+)			(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Nébuliseur (psi)	20													
Débit du <i>sheat gas</i> (psi)	12													
Débit du gaz séchant (psi)	11													
Température du <i>sheat gas</i> (°C)	300													
Température du gaz séchant (°C)	200													
Tension du <i>nozzel</i> (V)	1500	2000	-500	2000	-500	2000	-500	250	-1000	2000	-1000	500		
Tension du capillaire (V)	5000	5000	-4000	5500	-4000	5000	-4000	4000	-4000	5000	-4500	5000		
Énergie de collision (eV)	20	5	-30	10	-25	5	-45	20	-30	10	-35	30		

Tableau 19 : Paramètres de masse pour les analyses en LC-LC pour chaque famille de lipide en fonction de la polarité d'ionisation.

### 3.7. Application de la méthode LC-LC en ligne

Afin de vérifier l'automatisation des analyses utilisant la méthode HILIC-RPLC-QTOF selon le système présenté dans la figure n°100, des essais ont été réalisés pour s'assurer du bon fonctionnement de la programmation des différents *cuts*. Pour cela, la fraction TG/DG/Cer a été récupérée entre 0.35 min et 0.68 min. La fraction des PE entre 2 min et 2.33 min. La fraction des PI entre 2.35 min et 2.56 min. La fraction des PC entre 2.57 min et 2.9 min. La fraction des PS entre 3 min et 3.38 min. Pour finir, la fraction SM a été récupérée entre 3.39 min et 3.89 min. Chacune des fractions a été récupérée sans rencontrer de problème et sans effet de *breakthrough*. Un exemple pour la famille des PC est présenté sur la figure n°109.

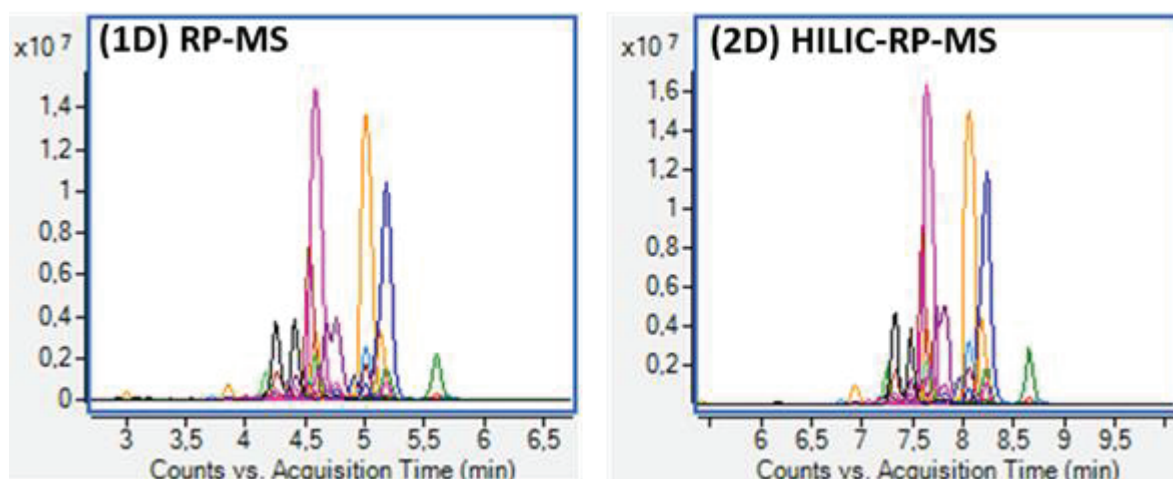


Figure 109 : Analyse de la fraction des PC avec l'approche classique RPLC-MS versus *multiple heart cut* HILIC-RPLC-MS. L'échantillon analysé en 1D-LC est dilué par un facteur 5 pour tenir compte de split de ratio 1:5 placé entre les deux dimensions en 2D-LC. Les chromatogrammes sont constitués de la superposition des SIM (single ion monitoring) extraits pour chaque PC contenu dans l'échantillon de plasma/gammare en ESI (+).

La comparaison des chromatogrammes des PC obtenus en RPLC-MS et HILIC-RPLC-MS en ligne présente des profils identiques puisque l'ensemble des lipides détectés en mode 1D le sont également en 2D et dans le même ordre d'éluion. Cela indique que la programmation du fractionnement des

familles lipidiques est un succès. La différence de temps de rétention observée entre les deux chromatogrammes n'est pas due à une différence de gradient ou de volume de délai. En effet, les analyses en mode 2D comprennent les temps de rétention des lipides en 2<sup>ème</sup> dimension, auxquels il faut ajouter le temps de rétention de la fin de la collecte de la famille lipidique associée dans la 1<sup>ère</sup> dimension (2.9 min pour les PC).

Un premier essai a été réalisé afin de démontrer les capacités de décomplexification d'un échantillon grâce à une approche 2D par rapport à une approche classique. Pour cela, un même échantillon a été analysé sur les deux systèmes afin de comparer la sensibilité de la réponse en masse pour chaque famille lipidique. Dans le cas des analyses 1D, l'échantillon a été dilué par un facteur 5 pour tenir compte du *split* 1:5 placé entre les deux dimensions en 2D. De manière générale, un léger gain de sensibilité en 2D par rapport aux analyses 1D est observé en ESI (+). Par exemple, pour la famille des PC le gain moyen est de 25%. Cependant, ce résultat reste à confirmer, car des résultats opposés sont obtenus en ESI (-) avec une diminution du signal moyen de 24%. Cette tendance est observée pour l'ensemble des familles lipidiques. Cette différence entre les deux modes pourrait provenir de la stabilité du *split* lors des analyses successives des différentes fractions. Dans le cas où une partie du système, placée après le *split*, subit une augmentation de pression, à cause d'un raccord qui se bouche par exemple, cela a une incidence sur le ratio du *split* appliqué et donc sur la quantité de lipides transférés de la 1<sup>ère</sup> dimension vers la 2<sup>ème</sup> dimension. À cause d'un manque de temps lié au contexte sanitaire du COVID 19, ces essais n'ont pas été reconduits et devront être réalisés lors d'une prochaine étude.

L'optimisation de la méthode *multiple heart cut* HILIC-RPLC en ligne avec un système d'injection multiple rapide permet d'obtenir une méthode LC-LC présentant de nombreux avantages et quelques inconvénients par rapport aux méthodes LCxLC rencontrées en lipidomique, qui seront détaillés ci-dessous.

### 3.8. Bilan de l'approche LC-LC en lipidomique comparée à l'approche LCxLC

#### 3.8.1. Inconvénients du mode LC-LC vis-à-vis du mode LCxLC

##### 3.8.1.A. Nombre de fractions analysées en 2<sup>ème</sup> dimension

L'un des inconvénients du mode *multiple heart cut* vis-à-vis du mode *comprehensive* est la perte d'informations possible du fait que seule une partie de l'échantillon est analysée en 2<sup>ème</sup> dimension. Pour limiter cette perte d'informations, il faut utiliser la chromatographie de phase HILIC en 1<sup>ère</sup> dimension, qui possède une affinité pour les composés polaires afin de regrouper chaque famille

lipidique sous un seul pic chromatographique, facilitant la récupération de la totalité des lipides. Cependant, ce constat n'est vrai que pour les lipides possédant un nombre de carbones total du même ordre de grandeur. Bien que dans la nature le nombre de carbones composant les chaînes d'acides gras des lipides est majoritairement compris entre 14 et 22 carbones [198,294–298], il n'est pas impossible que des lipides à chaîne courte soient présents. Les portions non collectées du chromatogramme en D1, entre chaque fraction lipidique, peuvent donc comprendre des composés d'intérêts limitant ainsi la quantité d'informations pouvant être acquises. De plus, la méthode développée ne prend pas en compte la famille des lysophospholipides (LPL).

### 3.8.1.B. Temps d'analyse global

L'autre inconvénient de l'approche LC-LC comparé à l'approche LCxLC est le temps d'analyse qui est plus important pour ce premier mode. Cela est dû au temps d'analyse global qui correspond à la somme des temps d'analyse de chacune des fractions envoyées dans la 2<sup>ème</sup> dimension pour le mode *multiple heart cut*, comme rappelé dans les équations suivantes :

- **Temps d'analyse global (LC-LC) = temps de collecte de la 1<sup>ère</sup> fraction en D1 + (temps de gradient en D2) x nombre de fraction**
- **Temps d'analyse global (LCxLC) = temps de gradient en D1**

Pour diminuer le temps d'analyse global de la méthode LC-LC, nous avons optimisé le gradient d'éluion en D1 pour permettre d'éluier l'ensemble des lipides en moins de 4 minutes. En effet, il faut attendre 0.6 minute après l'injection pour collecter la fraction des TG/DG/Cer, 1.7 minute pour celle des PG, 2.2 minutes pour celle des PE, 2.5 minutes pour celle des PI, 2.8 minutes pour celle des PC, 3.2 minutes pour celle des PS et 3.8 minutes pour celle des SM (cf. figure n°102). Cela permet de réaliser des analyses avec des injections multiples sans perdre de temps. Ces injections multiples rendent possible l'optimisation du gradient en D2 pour chaque fraction analysée, diminuant ainsi le temps d'analyse (temps de gradient fraction TG/DG/Cer = 15.5 min et temps de gradient pour les autres fractions = 13 min). De plus, la capacité à isoler les fractions dans la première dimension permet d'avoir une méthode à façon où il est possible d'analyser seulement les familles lipidiques d'intérêt, diminuant ainsi le temps d'analyse global.

Dans le cas de la méthode développée dans cette étude, l'analyse des familles TG, DG, Cer, PG, PE, PI, PC, PS et SM en HILIC-RPLC possède un temps d'analyse global de 110.3 minutes, comme indiqué dans l'équation suivante qui tient compte des injections multiples.



➤ **Temps d'analyse global =  $\Sigma$  (temps de collecte d'une fraction en D1 + temps de gradient en D2)**

$$= (0.6+15.5) + (1.7+13) + (2.2+13) + (2.5+13) + (2.8+13) + (3.2+13) + (3.8+13)$$
$$= 110.3 \text{ min}$$

Cela est plus faible que le temps d'analyse des méthodes RPLCxHILIC développé par Xu *et al.* [123] et Holcapek *et al.* [133] qui est de 160 minutes. Toutefois, le temps d'analyse de ces deux méthodes complètes est à nuancer, car contrairement à notre méthodologie, elles prennent en considération l'ensemble des lipides, dont les LPL.

La possibilité d'avoir une méthode à façon permet également de sélectionner directement le mode d'ionisation, ESI (+) ou ESI (-), le plus favorable à chaque famille lipidique. Dans le cas du mode RPLCxHILIC, il faut réaliser deux analyses en fonction de la polarité, ce qui double le temps global. Le changement de polarité alternée pourrait être utilisé en RPLCxHILIC, mais cela diminuera par deux le nombre de points par pic.

Un autre inconvénient de l'approche *multiple heart cut* à injections multiples est l'analyse d'échantillons « précieux », qui est rendue complexe à cause de la quantité d'échantillons nécessaire pour réaliser de nombreuses injections.

### 3.8.2. Avantage du mode LC-LC en lipidomique

#### 3.8.2.A. Élimination du *split* avant le détecteur

Lors de la mise en place d'une méthode LCxLC, toutes les fractions issues de la 1<sup>ère</sup> dimension sont envoyées dans la 2<sup>ème</sup> dimension, ce qui implique que le temps d'analyse en 2<sup>ème</sup> dimension doit être plus court que le temps de remplissage de la boucle de stockage. Ce temps est inférieur à la minute. Pour remplir ce critère, il est nécessaire d'utiliser des débits importants dans la 2<sup>ème</sup> dimension, pouvant dépasser plusieurs millilitres par minute. Ces débits nécessitent de réaliser un *split* en amont du détecteur. En LC-LC, le débit utilisé en seconde dimension est moins important (inférieurs ou égaux à 500  $\mu\text{L}/\text{min}$ ). Ce qui permet d'éliminer ce *split* placé avant le détecteur (*cf.* figure n°97 vs figure n°100). L'élimination de ce *split* permet d'éviter une potentielle perte d'intensité supplémentaire.

#### 3.8.2.B. Compatibilité des spectromètres de masse

L'autre conséquence de la mise en place d'un gradient d'élution rapide en 2<sup>ème</sup> dimension est que les pics chromatographiques sont plus fins avec des largeurs de pic inférieures à 5 secondes, comme le montre la figure n°98. En effet, la largeur de pic obtenu en HILICxRPLC pour le lipide PE (38:6) est de 3.6 secondes contre 15 secondes en HILIC-RPLC. L'augmentation des largeurs de pics chromatographiques n'oblige plus l'utilisation d'analyseurs ultra rapides comme le TOF pour obtenir

une quinzaine de points par pic dans le cas de méthode hautement multiplexée. Il est maintenant possible d'envisager les expériences MS/MS en 2D-LC, à la fois avec un TOF et avec un QQQ utilisant le mode MRM lorsque l'approche *multiple heart cut* est utilisée (cf. annexe n°16). Cette augmentation des largeurs de pics permettant d'obtenir des temps de cycle plus importants autorise également plus d'expériences en MS/MS avec les TOF.

### 3.8.2.C. Retraitement des données

Les données obtenues en mode *comprehensive* sont complexes à interpréter puisqu'elles se trouvent sous la forme de données tridimensionnelles avec le temps de rétention en D1, le temps de rétention en D2 et l'intensité. De plus, les pics chromatographiques larges issus de la 1<sup>ère</sup> dimension sont fractionnés en plusieurs pics en 2<sup>ème</sup> dimensions. Ces données peuvent devenir encore plus complexes si l'on rajoute une 4<sup>ème</sup> dimension avec la fragmentation des composés. Ainsi, il est nécessaire d'utiliser des logiciels spécifiques à l'analyse 2D permettant de représenter les données sous forme de cartographie à 3 dimensions. Dans le cas des analyses en mode *multiple heart cut* traitées dans cette étude, le retraitement des données est facilité, car il s'effectuera de la même manière que pour les analyses 1D. En effet, les données obtenues sont exprimées par une succession de chromatogrammes de type 1D puisqu'il n'est pas nécessaire de faire intervenir le temps de rétention de la séparation en 1<sup>ère</sup> dimension.

### 3.8.2.D. Optimisation individuelle de chaque famille de lipides

Que ce soit en mode LC-LC ou LCxLC, le logiciel de pilotage de l'appareillage 2D impose d'utiliser le même gradient d'élution en 2<sup>ème</sup> dimension et les mêmes paramètres de masse pour l'ensemble des fractions analysées lors d'une même injection. Ce problème est surmonté avec le gradient rapide qui est mis en place dans la 1<sup>ère</sup> dimension de notre méthode, car il permet d'éluer l'ensemble des fractions lipidiques en moins de 4 minutes. Cela permet de réaliser des injections multiples afin d'optimiser le gradient et les paramètres de masses individuellement pour chaque fraction de lipides sans faire de compromis.

### 3.8.2.E. Préconcentration des familles minoritaires avec moins de risques d'encrassement du spectromètre de masse

L'utilisation d'une séparation chromatographique très rapide en 1<sup>ère</sup> dimension, associée à la capacité de sélectionner une ou plusieurs familles de lipides à envoyer dans la 2<sup>ème</sup> dimension, nous a permis d'utiliser la colonne HILIC comme chromatographie préparatrice. Les avantages qui en découlent sont qu'il est désormais possible d'isoler les fractions minoritaires, et de faire varier la concentration des échantillons en fonction des familles de lipides considérées pour pouvoir gagner en sensibilité. L'élimination des familles majoritaires (ex. PC, TG, etc.) lors de la préconcentration des familles

minoritaire (ex. PI) permet d'éviter un encrassement précoce du détecteur. En comparaison, le mode LCxLC avec la configuration RPLCxHILIC impose d'analyser l'ensemble des fractions issues de la 1<sup>ère</sup> dimension, sans séparation des familles lipidiques en amont du détecteur. Ainsi, ce sont les familles de lipides majoritaires qui fixeront la quantité maximale injectable dans le système afin d'éviter la saturation du signal (cf. 3.1).

## 4. Conclusion et perspectives

L'état de l'art de l'analyse des lipides avait montré un besoin d'amélioration de la séparation chromatographique afin de limiter les effets matrices au sein et entre classes lipidiques. C'est pour cette raison que la chromatographie bidimensionnelle a été envisagée. À travers ces travaux de thèses, les limitations en termes de temps de séparation dans la 2<sup>ème</sup> dimension avec la configuration HILICxRPLC ont pu être vérifiées. Pour résoudre ce problème, le système 2D-LC en mode *multiple heart-cut* a été sélectionné, car il permet de conserver dans des boucles de stockage certaines fractions cibles éluées de la 1<sup>ère</sup> dimension. Cela laisse davantage de temps pour séparer les composés analysés dans la 2<sup>ème</sup> dimension. Cette méthodologie rend plus adaptée l'utilisation de la colonne HILIC en tant que 1<sup>ère</sup> dimension dans un système 2D totalement automatisable (HILIC-RPLC-MS/MS) et compatible avec tous types de spectromètres de masse.

Après avoir obtenu des résultats en mode hors ligne très encourageants, le développement de la méthode en ligne a été initié. Certaines familles de lipides ne supportent pas le passage de la phase HILIC vers la phase inverse. De ce fait, les phénomènes de *breakthrough* ont été évalués et un *split* a été mis en place pour contrebalancer ces effets. Chacune des dimensions chromatographiques a été optimisée pour réduire le temps d'analyse globale. L'optimisation de la 1<sup>ère</sup> dimension a rendu le fractionnement des familles lipidiques compatible avec le fonctionnement des boucles grâce au gradient rapide, permettant de réaliser des injections multiples sans perte de temps. Ces injections multiples permettent d'optimiser la quantité injectée, le gradient et les paramètres de masse individuellement pour chaque fraction de lipides. Cette optimisation peut être réalisée sans faire de compromis, en surmontant les limitations du logiciel de pilotage de l'appareillage 2D qui impose des paramètres uniques pour l'ensemble des fractions analysées lors d'une même injection. Cette capacité à adapter les concentrations de lipides injectés est un véritable avantage en lipidomique puisqu'il existe une gamme dynamique de concentrations lipidiques importante entre les différentes classes. À ce jour, la méthodologie mise en place a été automatisée avec succès, mais des paramètres restent encore à optimiser. En effet, faute de temps la méthode 2D-LC n'a pas pu être finalisée à l'issue de cette thèse qui a été impactée par les conditions sanitaires exceptionnelles.

En perspective, des étalons internes éluants dans les zones de fractionnement associées à leur classe lipidique respective en HILIC devront être sélectionnés. De plus, il faut évaluer à partir de quel nombre de carbones un lipide n'est plus élué dans la zone de fractionnement correspondant à sa classe lipidique en HILIC. Des investigations plus poussées devront être réalisées pour résoudre ce problème de co-élution entre classes lipidiques due à la variabilité des longueurs de chaîne des acides gras. Par exemple, la réalisation de tests sur différents types de phases HILIC qui seraient moins résolutive au sein d'une même classe lipidique pourrait être une solution. Il serait également judicieux de refaire le test des effets à l'injection lors du passage de la D1 à la D2 en évaluant des volumes supérieurs à 20% du volume mort injecté en D2. Cela permettra de vérifier s'il est possible de mettre en place un *split* moins important entre les deux dimensions pour augmenter la sensibilité de la méthode. Le *total breakthrough* pourrait être une solution pour se passer de l'utilisation d'un *split*, mais cela représente un très grand risque concernant la perte en sensibilité. D'autres paramètres liés à la robustesse de la méthode devront être vérifiés, comme la robustesse du *split* et la répétabilité du fractionnement des lipides dans la 1<sup>ère</sup> dimension. Il est courant, lors des analyses 2D-LC, de placer un détecteur UV non destructif entre les deux dimensions. Cela permet de s'assurer que chaque fraction est bien collectée en vérifiant le temps de rétention des composés en 1<sup>ère</sup> dimension. Les lipides ne présentant aucune réponse avec ce type de détecteur, il n'est pas possible d'utiliser cette méthodologie. Pour surmonter cette problématique, il faudrait rajouter dans l'échantillon un ou plusieurs composés qui soient détectables en UV et qui soient compatibles avec les conditions chromatographiques utilisées en lipidomique. Pour finir certaines performances analytiques sont à évaluer, notamment le gain de sensibilité grâce à la limitation des effets matrices intra et extra famille associée à la modulation de la quantité maximale injectée par classe lipidique. Ce gain de sensibilité devra être supérieur à la quantité de lipides perdus à cause du *split* placée entre les deux dimensions pour que l'utilisation de cette approche 2D-LC soit pertinente par rapport aux approches 1D-LC. Actuellement, on ne peut pas encore conclure sur la diminution des effets matrices apportée par la 2D-LC.

Bien que les perspectives de la chromatographie bidimensionnelle soient très attractives, les applications dans un contexte d'analyses de routines sont encore peu nombreuses à ce jour. Cela est probablement lié au développement des approches 2D-LC puisqu'il s'agit d'un exercice long et délicat. De nombreux paramètres sont à prendre en considération, notamment pour le dimensionnement des colonnes et les effets à l'injection entre les deux dimensions, *etc.* Cette technologie est probablement encore trop complexe à mettre en œuvre pour être appréhendée facilement par des non-initiés et transférée dans des laboratoires d'analyses en écotoxicologie. Des développements technologiques, notamment sur le logiciel de pilotage 2D-LC, sont encore nécessaires pour démocratiser et propager la chromatographie bidimensionnelle en dehors des laboratoires de recherche en chimie analytique. De plus, les aspects quantitatifs ont à ce jour été peu étudiés pour les approches 2D-LC en lipidomique.

## Conclusion générale

Cette thèse avait pour objectifs de développer de nouveaux outils analytiques dans le cadre d'analyses multi-omiques afin d'améliorer l'acquisition de données multi-échelles pour mieux caractériser de façon exhaustive un échantillon biologique. Les travaux décrits dans ce manuscrit se sont concentrés principalement sur des développements analytiques en chromatographie liquide couplée à la spectrométrie de masse pour des analyses ciblées en protéomique, lipidomique et/ou métabolomique pour l'étude de l'espèce sentinelle *G. fossarum*. Cette espèce d'intérêt en écotoxicologie est étudiée depuis plusieurs années au sein de notre laboratoire en étroite collaboration avec l'équipe Écotoxicologie de l'INRAE-Villeurbanne. La réglementation dans le cadre de la directive cadre sur l'eau préconise l'utilisation d'espèces vivantes pour le suivi de la contamination environnementale *via* l'utilisation de bio-essais. Mes travaux vont permettre dans le futur de répondre à cette demande en transférant ces nouveaux outils analytiques aux différents acteurs du domaine de l'écotoxicologie.

Dans un premier temps, à travers l'étude bibliographique réalisée dans le 1<sup>er</sup> chapitre de ce manuscrit, différents verrous analytiques ont été relevés.

Le premier verrou analytique était de savoir comment augmenter le nombre de composés à analyser de façon simultanée dans une méthode LC-MS/MS ciblée sans avoir à sacrifier sa robustesse. Afin de construire une méthode hautement multiplexée en protéomique, un mode d'acquisition Scout-MRM, a été développé et appliqué pour le dosage simultané de 157 protéines en tant que potentiels biomarqueurs d'expositions chez *G. fossarum*. À ce jour, il s'agit du plus grand multiplexe en écotoxicologie. L'utilisation de la méthodologie Scout-MRM va faciliter le transfert de méthodes entre laboratoires en proposant des méthodes prêtes à l'emploi, car indépendantes des variations en temps de rétention et des configurations chromatographiques.

Grâce à ce mode d'acquisition ciblé hautement multiplexé et adaptatif, il est possible de surmonter les limitations des capacités de multiplexage des méthodes MRM classiques. Cela offre de nouvelles possibilités de recherche en proposant une approche de «découverte ciblée» *via* la construction de grandes listes de transitions MRM issues de bases de données. L'augmentation des capacités de multiplexage avec le Scout-MRM diminue drastiquement le nombre de méthodes et d'injections pour une caractérisation exhaustive d'un échantillon en mode ciblé. La méthodologie Scout-MRM peut être une alternative aux approches globales par HRMS lorsque des instruments de haute résolution ne sont pas accessibles. La lipidomique est une science omique qui se prête bien à ces approches, car il est possible de modéliser le temps de rétention des lipides et de réaliser des modèles de prédiction. Cela permet d'avoir accès à une base de données conséquente permettant la construction facilitée de

méthodes Scout-MRM. C'est pour cela que le 3<sup>ème</sup> chapitre de ce manuscrit a décrit le développement d'une méthode d'analyse ciblée pour un profilage rapide de lipidomes par RPLC-MS/MS utilisant le Scout-MRM. La prédiction des temps de rétention développée lors de ces travaux de thèse a été basée sur le modèle du nombre de carbones équivalent (ECN) et a permis d'élaborer une base de données comportant plus de 400 lipides. Ce nombre devrait rapidement augmenter dans le futur. En effet, plus la méthode Scout-MRM sera utilisée par d'autres chercheurs et dans d'autres domaines d'application, plus il sera possible de rajouter de nouveaux composés et d'affiner les modèles de prédiction pour de nouveaux lipides. Ainsi, la qualité de cette base de données ne peut qu'augmenter au fil du temps. L'adaptabilité du Scout-MRM, grâce à sa faculté à corriger les variations en temps de rétention, autorise l'application de cette méthodologie sur n'importe quel extrait lipidique ou système chromatographique à condition d'utiliser une colonne chromatographique de phase inverse. Pour finir, en plus de l'utilisation de l'approche Scout-MRM en tant qu'outil de découverte ciblée pour des études exploratoires, un transfert vers un mode de quantification ciblée est instantané.

Pour compléter le panel d'outils à disposition pour l'étude des différentes omiques (protéomique et lipidomique), une méthode Scout-MRM appliquée au domaine de la métabolomique est en cours de développement dans le cadre de la thèse de Thomas Brunet au sein de l'équipe Anabio-MS. Cette méthode compte déjà plusieurs centaines de métabolites.

Les 2<sup>ème</sup> et 3<sup>ème</sup> chapitres de cette thèse ont prouvé que l'on est désormais capable d'utiliser des méthodes hautement multiplexées pour chacune des omiques. Toutefois, avant de pouvoir réaliser ces analyses, il faut être capable d'extraire et de fractionner ces différentes classes de biomolécules à partir du même échantillon.

Le deuxième verrou identifié était de rendre compatible la préparation d'échantillon avec l'ensemble des omiques cibles. De plus, il était nécessaire que l'extraction de chaque omique ne se fasse pas au détriment des autres. Ainsi, une méthode d'extraction multi-omique a été développée dans le 4<sup>ème</sup> chapitre de cette thèse. La procédure mise en place avait également pour but de limiter les variations biologiques en diminuant le nombre de réplicats d'exposition pour chacune des omiques. C'est pourquoi l'extraction a été conçue à partir d'une prise d'essai unique et d'une méthodologie capable de récupérer simultanément les protéines, lipides et métabolites. Cette extraction a été adaptée pour l'analyse de matrices complexes solides comme l'espèce sentinelle *G. fassarum* à travers l'optimisation des conditions d'homogénéisation de l'échantillon. Les différentes expérimentations ont montré que l'utilisation de tampon d'extraction aqueux est obligatoire si l'on souhaite conserver des rendements d'extraction acceptables en protéomique. Il a été montré que les résidus solides lors de l'étape de broyage avaient une grande importance sur les rendements d'extraction en protéomique. Nous avons également démontré que le tampon aqueux contenant du triton-X nécessaire à l'extraction des

protéines était compatible pour l'extraction des autres omiques (lipides et métabolites). Afin de démontrer le potentiel de notre protocole multi-omique, une étude du cycle reproductif des gammarès femelles a été réalisée. Cela a mis en évidence des signatures moléculaires révélant une augmentation de l'expression de protéines spécifiques impliquées dans la maturation ovarienne. Cette augmentation en protéomique est en corrélation avec la mobilisation du stockage des lipides et des variations observées dans des voies métaboliques spécifiques. Ces résultats ouvrent la voie à de futures études écotoxicologiques puisque l'analyse peut être réalisée sur un organisme unique, réduisant le nombre de réplicats d'exposition et donc le nombre d'échantillons à collecter tout en permettant une caractérisation exhaustive de l'échantillon à un niveau multi-échelle.

L'extraction multi-omique implique de faire des compromis en adoptant une procédure compatible avec toutes les familles omiques considérées. Ainsi, la décomplexification de l'échantillon n'est pas aussi importante qu'elle ne le serait si une seule omique avait été étudiée. Cela peut engendrer des effets matrices pouvant perturber la détection, notamment avec les interférences isobariques et les phénomènes de suppression d'ions qui peuvent se produire dans la source électrospray. Afin de solutionner cela, il peut être intéressant d'employer une dimension de séparation supplémentaire en amont de la source d'ionisation. Cela peut être réalisé grâce à l'apport de la chromatographie liquide bidimensionnelle (2D-LC) qui présente des capacités de séparation plus importantes que les méthodes monodimensionnelles (1D-LC).

Le troisième verrou analytique de cette thèse concernait l'évaluation des outils chromatographiques avancés tels que la 2D-LC pour l'analyse d'extraits multi-omiques complexes. Ainsi, le développement d'une méthode 2D-LC a été initié pour analyser les lipides. Suite à l'étude bibliographique, il avait été démontré des besoins en termes de séparation chromatographique pour cette famille de composés. L'approche *multiple heart cut* (HILIC-RPLC-MS/MS) décrit dans le 5<sup>ème</sup> et dernier chapitre de cette thèse a été préférée à l'approche *comprehensive*. La mise au point d'un gradient extrêmement rapide dans la 1<sup>ère</sup> dimension HILIC a permis de surmonter les problèmes de pilotage du logiciel d'acquisition 2D lors de la récolte des fractions d'intérêts grâce aux injections multiples. Cette multiplication des injections permet de faire varier la quantité injectée, le gradient et les paramètres de masses individuellement pour chaque fraction de lipides. La capacité à adapter les concentrations injectées par classe de lipides est un avantage en lipidomique puisqu'il existe une gamme dynamique de concentrations lipidiques importante entre les différentes classes. Cela a été rendu possible grâce à la conservation de la colonne HILIC comme 1<sup>ère</sup> dimension. Ce qui n'était pas envisageable en mode *comprehensive*. À ce jour, la méthodologie a été automatisée avec succès. Cependant des paramètres restent encore à évaluer. Notamment le gain de sensibilité grâce à la limitation des effets matrices au sein et entre familles de lipides associée à la modulation de la quantité maximale injectée par classe



lipidique. Ce gain de sensibilité devra être supérieur à la quantité de lipides perdus lors de la division du débit qui est placée entre les deux dimensions dans le but de contrebalancer le phénomène de *breakthrough*. Il est capital que ce critère soit respecté pour que l'utilisation de cette approche 2D-LC soit pertinente par rapport aux approches 1D-LC.

# Annexes

**Annexe 1 : Publication “from shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential immunomarkers in *Dreissena polymorpha*”.**

RESEARCH PAPER



## From shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential immunomarkers in *Dreissena polymorpha*

Maxime Leprêtre<sup>1,2</sup> · Mélissa Palos-Ladeiro<sup>2</sup> · Julien Faugere<sup>1</sup> · Christine Almunia<sup>3</sup> · Jérôme Lemoine<sup>1</sup> · Jean Armengaud<sup>3</sup> · Alain Geffard<sup>2</sup> · Arnaud Salvador<sup>1</sup>

Received: 30 April 2020 / Revised: 27 July 2020 / Accepted: 6 August 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

### Abstract

A highly multiplexed liquid chromatography mass spectrometry–multiple reaction monitoring (MRM)-based assay has been developed for evaluating 107 candidate immune biomarkers in both hemocytes and plasma of the zebra mussel *Dreissena polymorpha*. The Scout-MRM strategy was employed for the first time, shortening the implementation of a targeted MRM bottom-up proteomics assay using selected immune protein-related peptides identified by shotgun discovery proteogenomics. This strategy relies on spiking scout peptides during the discovery phase and using them to build and deploy the MRM targeted proteomics method. It proved to be highly relevant, since about 90% of the targeted peptides and proteins were monitored and rapidly measured in both hemocyte and plasma samples. The sample preparation protocol was optimized by evaluating the digestion efficiency of tryptic peptides over time. The accuracy and precision of 50 stable isotope-labeled peptides were evaluated for use as internal standards. Finally, the specificity of the transitions was thoroughly assessed to ensure the reliable measurement of protein biomarkers. Several analytical and biological validation criteria were evaluated across hemocytes and plasma samples exposed *ex vivo* to biological contaminants, resulting in the validation of two Scout-MRM assays for the relative quantitation of 85 and 89 proteins in hemocytes and plasma, respectively.

**Keywords** Hemolymph · Bivalves · Immunomarkers · Multiplexed assay · Scout-MRM

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00216-020-02868-2>) contains supplementary material, which is available to authorized users.

✉ Arnaud Salvador  
arnaud.salvador@univ-lyon1.fr

<sup>1</sup> Université de Lyon, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, CNRS UMR 5280, 69100 Villeurbanne, France

<sup>2</sup> Université de Reims Champagne-Ardenne, UMR-I 02 INERIS-URCA-ULH SEBIO Unité Stress Environnementaux et BIOSurveillance des milieux aquatiques, UFR Sciences Exactes et Naturelles, Campus du Moulin de la Housse, BP 1039, 51687 Reims, Cedex, France

<sup>3</sup> Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, 30200 Bagnols-sur-Cèze, France

### Introduction

Water quality evaluation in the framework of maintaining the sustainability of freshwater ecosystems and protecting biodiversity and ecosystem services represents a major challenge. These issues have led to the need to develop sensitive and predictive tools capable of assessing the toxicity of various anthropogenic contaminants in order to evaluate their associated risk to the ecological maintenance of aquatic ecosystems. In this context, the biological responses of sentinel organisms have been proposed as an interesting alternative to current regulatory actions based on chemical and biocenotic analyses [1, 2]. Proteomics presents several advantages for assessing the effects of contaminants on the health status of organisms [3]. Proteins are key compounds in biosynthesis and cell, tissue, and organ signaling, and provide cell and tissue structural stability in living organisms. As key functional components of cells, proteins are known to be sensitive to environmental

Published online: 17 August 2020

Springer



contamination. They enable the assessment of biological responses of organisms under stress, the prediction of contaminant toxicity, and understanding of their mechanisms of action on the physiology of organisms. Therefore, proteomics are increasingly used in the field of ecotoxicology [3–5]. As a result, the concept of “ecotoxicoproteomics” has emerged in several scientific journals. Proteomic analysis has been performed on several aquatic sentinel species including marine bivalves, crustaceans and fishes to: (i) assess the toxicity of contaminants, (ii) identify molecular mechanisms impacted by environmental contamination, and (iii) identify new ecotoxicological biomarkers [6–11].

Due to its high bioaccumulation potential, the zebra mussel (*Dreissena polymorpha*) is a relevant sentinel species used in aquatic ecotoxicology and freshwater quality monitoring. Toxic substances and microorganisms in the environment can generate adverse effects on *Dreissena polymorpha*, from the molecular to the community level. These effects negatively impact physiological functions such as energy metabolism, reproduction, and the immune system [12]. Among these biological functions, the immune system of *Dreissena polymorpha* is of great interest, as it reflects the health status of this organism during stressful conditions and actively participates in the elimination of foreign particles. In this case, immune protein quantitation is a powerful tool used to observe early molecular events involved in contamination responses, and which are responsible for the adverse effects observed at higher levels of biological organization. Recently, a proteogenomic approach was conducted on both hemocytes and plasma to identify the immune proteins of *Dreissena polymorpha* [13]. About 3020 proteins were identified, representing the largest protein inventory for this non-model organism. To identify immune proteins involved in host–microbial interactions and potential immunomarkers, differential proteomics was conducted on hemolymph samples exposed *ex vivo* to biological stress (the protozoa *Cryptosporidium parvum*, and RNA polyinosinic-polycytidylic acid (poly I:C)) [14]. At present, the validation of these potential immunomarkers as reliable indicators of biological stress requires extensive evaluation across the analysis of hundreds of extracted *Dreissena polymorpha* samples. Tackling such a challenge can only be done through the implementation of a sensitive, robust, quantitative, and high-throughput assay.

To this end, targeted mass spectrometry-based strategies such as multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM) offer an excellent means of accurately monitoring the abundance of proteins identified by proteogenomics and thus bridging the gap between protein discovery and the validation of candidate biomarkers [15–17]. MRM and PRM are often compared and show comparable performance. Both enable the accurate and specific simultaneous evaluation of multiple biomarkers in omics projects [18–21] and especially in sentinel species [10, 20, 22]. In

complex matrices, PRM provides better specificity [19], while the sensitivity provided by triple quadrupole mass spectrometers leads the research community to favor MRM [23], although the gains in sensitivity obtained are often marginal. As most environmental analysis laboratories are already well equipped with triple quadrupole devices, we have chosen the MRM approach. This may facilitate the transfer of our method for future analysis.

As highlighted by Chen and co-workers [24], transitioning from the discovery experiment to the quantitative analyses of hundreds of samples often represents the bottleneck of proteomics. Such a change in scale is indeed time-consuming and requires significant resources, not only during the assay development stage but also along the course of sample analyses. To solve this issue, these authors suggested a workflow using data acquired from discovery proteomic experiments, retention time prediction, and standard-flow chromatography to rapidly develop targeted proteomic assays. The authors use retention time predictions from discovery data to rapidly construct an MRM assay. For large MRM or PRM assays, targeted peptides are monitored only around their expected elution time. In this case, the development of large-multiplexed assays becomes rather complex, since the retention times of peptides could shift. To solve this issue, “on-the-fly” corrections of the acquisition windows time methods have been proposed [25, 26]. These methods are very promising for the development of large-multiplexed assays, but alternative strategies could also be used. Recently, a new mode of targeted analysis called “Scout-MRM” has been proposed, enabling the building of highly multiplexed assays without the requirement of time scheduling [27, 28]. This independence in the retention time dimension makes Scout-MRM especially attractive for shortening the development phase of targeted assays based on shotgun proteomics data, especially when different liquid chromatography (LC) configurations are used (a different  $C_{18}$  chemistry for the stationary phase or nano versus microbore liquid chromatography).

The aim of this study was to propose a new application of a “Scout-MRM” independency retention time method and to use the “Scout” approach to quickly develop a multiplexed protein assay from discovery experiments. These assays of about 100 potential immunomarkers in the hemolymph of *D. polymorpha* are necessary to simultaneously measure the abundance of many immune protein candidates in different toxicological contexts from both the hemocyte and plasma compartments. While this strategy was successfully developed for the relative quantification of immune markers of *D. polymorpha*, it could be easily extrapolated for other species used in ecotoxicology or other research fields. This new workflow could be applied to a broad set of proteomic analyses and simplify the validation of biomarkers identified from shotgun proteomics, with minimal development time. Moreover, since it is independent of peptide retention times,



the Scout-MRM method offers a significant advantage when transitions from shotgun to targeted proteomics are performed with different chromatographic systems, which is often the case when analyses are conducted in different laboratories.

## Materials and methods

### Chemicals and materials

Acetonitrile, methanol and water (LC-MS grade), and trypsin (TPCK-treated from bovine pancreas) were obtained from Fisher Scientific™ (Strasbourg, France). Dithiothreitol (DTT), iodoacetamide (IAM), formic acid (FA) (LC-MS grade), urea and Tris, EDTA, Triton X, sodium chloride, ammonium bicarbonate, and TRI Reagent® were purchased from Sigma-Aldrich (St Quentin-Fallavier, France). Absolute ethanol and ethyl ether were obtained from Carlo Erba (Val de Reuil, France). An XBridge C<sub>18</sub> column (2.1 × 100 mm, 3.5 μm) and hydrophilic-lipophilic balance (HLB) column (60 mg Oasis sorbent) were purchased from Waters (Milford, MA, USA). The scout peptide-coated RT-BEADS™ (AZ-011-015-A469) kit was purchased from Anaquant (Lyon, France).

### Selection of candidate proteins, reporter peptides, and MRM transitions

Protein sequences were extracted from a protein database previously created with a proteogenomics approach combining shotgun proteomics and RNAseq data acquired from *Dreissena polymorpha* [13]. This database comprises the mass-spectrometry-certified sequences of 3020 proteins. Their putative functions were derived from sequence similarity searches against the National Center for Biotechnology Information (NCBI) database of non-redundant protein sequences (nr) and the Swiss-Prot database.

Based on shotgun proteomic results [13, 14], 114 candidate proteins were selected for their potential function in zebra mussel immunity (Electronic Supplementary Material [ESM] Table S1). Ideally, three reporter peptides per protein were then selected for their qualification and/or relative quantification by targeted proteomics. According to our previous proteogenomic results [13, 14], these peptides were preferentially selected to (i) be specific to the target protein, (ii) not exhibit chemical modifications such as spontaneous deamidation (94% of peptides) or uncontrolled oxidation (80% of peptides), and (iii) not exhibit missed cleavage phenomena (97% of peptides) that could induce quantitation bias [29] (see ESM Fig. S1). We also avoided short (<5 residues) and long (>25 residues) peptides, and whenever possible, peptide sequences containing methionine were excluded, since they are known to be unsuitable for quantitative analysis due

to their oxidative modification [30]. Finally, raw data from shotgun proteomics were analyzed by Protein Prospector software (<http://prospector.ucsf.edu>) to select the three most intense fragment ions for the MRM transition settings of targeted peptides [31].

### Isotope-labeled peptides

Fifty stable isotope-labeled peptides (1 peptide per protein), corresponding to 50 proteins selected for their major role in the immunity of *D. polymorpha*, were purchased from Fisher Scientific™ (Strasbourg, France). These peptides were synthesized as isotopically labeled C-terminal lysine [<sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>2</sub>] or arginine [<sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>4</sub>] with purity greater than 97%, and carbamidomethylated modifications on cysteine residues. The accurately weighed lyophilized heavy peptides were first individually reconstituted with 1 mL of water/methanol (50/50, v/v) mixture containing 0.5% FA. Then, a dilution to yield a concentration of 500 μg/mL was performed in water/methanol (90/10, v/v) for each peptide. A stock solution mixture (S0) containing all the labeled peptides (*n* = 50) was prepared at 5000 ng/mL, aliquoted, and stored at -20 °C until use to avoid freeze-thaw cycles. The aliquots were used for both hemocyte and plasma samples, and 38 and 40 labeled peptides were monitored by MRM in hemocytes and plasma samples, respectively. Solutions containing the labeled peptides were stored at -20 °C for a maximum of 2 months.

### Sampling and maintenance of animals

The sampling and maintenance of animals were performed as described in Leprêtre et al. [13]. More than 200 zebra mussels about 2 cm long were sampled at the “Lac du Der” (Grand Est, France) in April 2018. A 2-week laboratory acclimatization period was ensured by placing the mussels in 20-L aquariums filled with Cristaline Aurèle spring water (Jandun, France), under a constant temperature (14 °C) and controlled aeration. Zebra mussels were fed twice a week with a mixture of two species of microalgae (*Chlorella pyrenoidosa* and *Scenedesmus obliquus*) at a concentration of two million microalgae per day per mussel. The aquarium water was changed weekly.

### Sample preparation and protein extraction

The preparation and protein extraction of hemocytes and plasma samples were performed manually, as previously described in Leprêtre et al. [13]. Briefly, 200 μL of hemolymph per mussel was collected from the adductor muscle with a 30 g × 0.3 mL syringe and pooled to reach 1 mL hemolymph per replicate, which represents about five mussels per pool of hemolymph. A total of 10 replicates, representing 10 mL of hemolymph collected from about 50 mussels, were used for



all the analyses performed in this study, except for the biological validation of MRM methods and the identification of the proteins modulated during immunological challenges (see section 2.12).

Once collected, the hemolymph samples were centrifuged at  $800\times g$  for 15 min at  $4\text{ }^{\circ}\text{C}$  to separate hemocytes from plasma. To precipitate proteins from plasma, three volumes of ice-cold acetone/water mixture (80/20 v/v) were added to the plasma samples and left overnight at  $-20\text{ }^{\circ}\text{C}$ . After washing the protein pellet twice with 1 mL of 80% cold acetone, centrifugation, and drying, the proteins were resuspended in 80  $\mu\text{L}$  of a resuspension buffer (6 M urea; 2 M thiourea; 4% 3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate (CHAPS); 65 mM dithiothreitol (DTT); 0.1% Triton X100 (v/v) at pH 7.8).

For the hemocytes, 500  $\mu\text{L}$  of TRI-Reagent was added to hemocyte pellets and stored at  $-20\text{ }^{\circ}\text{C}$  until protein extraction. Hemocyte lysis was performed by passing the cells at least 10 times through a syringe equipped with a 25-gauge needle, and an initial cleanup was performed by centrifugation at  $12,000\times g$  for 2 min. RNAs were removed from the resulting supernatant by adding 100  $\mu\text{L}$  of chloroform, incubating for 15 min, centrifugation at  $12,000\times g$  and  $4\text{ }^{\circ}\text{C}$  for 15 min, and removing the colorless aqueous phase. To discard DNA, 150  $\mu\text{L}$  of absolute ethanol was added to the color phase and incubated for 3 min at room temperature, followed by centrifugation at  $2000\times g$  and  $4\text{ }^{\circ}\text{C}$  for 5 min. To precipitate the proteins from the resulting supernatant, three volumes of ice-cold acetone were added and left for 10 min at room temperature. Then the samples were centrifuged for 10 min at  $12,000\times g$  and  $4\text{ }^{\circ}\text{C}$  and the remaining protein pellets were washed three times for 10 min in 1 mL of 0.3 M guanidine hydrochloride diluted in 95% ethanol (v/v), followed by centrifugation at  $8000\times g$  and  $4\text{ }^{\circ}\text{C}$  for 5 min and the removal of the supernatant. A final wash was performed by adding 1 mL of 95% ethanol to the protein pellets. After drying at room temperature, the protein pellets were resuspended in 80  $\mu\text{L}$  of resuspension buffer. For each sample, the protein concentration was assessed by the Bradford assay before trypsin proteolysis and MS/MS analysis [32].

### Trypsin proteolysis

After protein extraction, all the sample preparation steps were performed in 1.5 or 2 mL protein LoBind tubes (Eppendorf) to avoid unpredictable peptide loss [33, 34]. A volume equivalent of 90  $\mu\text{g}$  of proteins was diluted with 1.2 mL of a 50 mM ammonium bicarbonate (AMBIC) buffer and 90  $\mu\text{L}$  of DTT at 150 mM prepared in AMBIC buffer (50 mM). Then the samples were incubated for 40 min at  $60\text{ }^{\circ}\text{C}$  in a water bath. After cooling to room temperature (10 min), 200  $\mu\text{L}$  of IAM at

150 mM, diluted in AMBIC buffer (50 mM), was added to obtain a final concentration of 20 mM, and the samples were placed in the dark at room temperature for 40 min. A volume of 30  $\mu\text{L}$  of trypsin at 0.1 mg/mL in 50 mM AMBIC was added to give a substrate:enzyme ratio of 30:1 (w/w). The samples were then incubated for 15 h at  $37\text{ }^{\circ}\text{C}$ . Finally, 20  $\mu\text{L}$  of formic acid (FA) was added to tryptic digestion. The samples were centrifuged at  $10,000\times g$  at  $4\text{ }^{\circ}\text{C}$  for 10 min before solid-phase extraction.

### Solid-phase extraction (SPE)

Solid-phase extraction was carried out manually using the 60 mg Oasis HLB cartridge (Waters, Milford, MA, USA). Cartridges were preconditioned with 1 mL of methanol and equilibrated with 1 mL of water containing 0.5% FA. Then the samples (2 mL) were deposited on the cartridges, which were then rinsed with a 2 mL water/methanol (95/5 v/v) solution containing 0.5% FA. Analytes were eluted with 1 mL of methanol acidified with 0.5% of FA. A volume of 100  $\mu\text{L}$  of a mixture of glycerol at 10% in methanol was added in each sample. Then the solvent was evaporated under nitrogen flow to a volume of 10  $\mu\text{L}$ , which was diluted back with 10  $\mu\text{L}$  of water/methanol (90/10, v/v) with 0.5% of FA, 10  $\mu\text{L}$  of one READYBEADS™ (Anaqant, France) diluted in 100  $\mu\text{L}$  water/methanol (90/10, v/v) with 0.5% of FA and 30  $\mu\text{L}$  of the peptide mix solution at 1500 ng/mL to obtain a final concentration at 500 ng/mL of labeled peptides. The tubes were then vortexed for several seconds and the contents were transferred to glass autosampler vials.

### Liquid chromatography and mass spectrometry

The analyses were carried out on an Agilent 1200 Series HPLC device (Agilent Technologies, Waldbronn, Germany) coupled to a QTRAP® 5500 hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems/MDS Analytical Technologies, Foster City, CA, USA). LC separation of the 20  $\mu\text{L}$  sample injected was carried out on an XBridge C<sub>18</sub> column (100 mm  $\times$  2.1 mm, particle size 3.5  $\mu\text{m}$ ) from Waters (Milford, MA, USA). The samples were separated at 300  $\mu\text{L}/\text{min}$  using the following gradient (time in min, %B): (0, 2%), (1, 2%), (30, 40%), (30.5, 100%), (35.5, 2%), (40, 2%), with water containing 0.1% FA (v/v) as eluent A and acetonitrile containing 0.1% FA (v/v) as eluent B. The temperatures of the column and autosampler were  $40\text{ }^{\circ}\text{C}$  and room temperature, respectively.

The mass spectrometer was interfaced with an electrospray Turbo V™ ion source operated in positive ionization mode using an ion spray voltage of 5500 V. The nebulizer and the curtain gas flows were set at 50 psi using nitrogen (N<sub>2</sub>). The Turbo V™ ion source was set at  $550\text{ }^{\circ}\text{C}$ , with the auxiliary gas flow (nitrogen) set at 40 psi. The collision gas (N<sub>2</sub>) was set to a



high value with a collision cell exit potential (CXP) set at 15 V. Q1 and Q3 were set to unit resolution (0.7 Da). Instrument control, data acquisition, and processing were performed using a modified Analyst<sup>®</sup> 1.6.2 software application. Implementing the Scout-MRM tool in Analyst<sup>®</sup> 1.6.2 software required computer developments by David Cox at the Sciex R&D laboratory.

Skyline v3.1 software (MacCoss Lab Software, USA) was used to produce the list of appropriate MRM transitions used to screen the precursor/product ions of peptides. The masses Q1 and Q3 as well as the collision energy (CE) and the declustering potential (DP) values were predicted by Skyline [35, 36]. From the initial set of candidates, between one and four reporter peptides per protein were selected, with three MRM transitions per peptide. Labeled peptides were also monitored with three MRM transitions, while scout peptides were monitored with two transitions. For the development of MRM assays, 931 MRM transitions (92 proteins, 263 peptides, 38 labeled peptides) were monitored in the hemocyte samples and 973 MRM transitions (92 proteins; 275 peptides; 40 labeled peptides) were monitored in the plasma, taking into account the transitions of the 14 scout peptides (28 MRM transitions per sample type) (ESM Fig. S1). A maximum of 150 transitions were monitored per triggered group to ensure a minimal dwell time of 3 ms and a duty cycle set to 1.52 s in order to obtain between 10 and 15 points per chromatographic peak.

Scout-MRM data were deposited at the PeptideAtlas SRM Experiment Library (PASSEL) [37] with the dataset identifier PASS01600.

### Detection of targeted peptides by a Scout-MRM approach

As described by the manufacturer's protocol (Anaquant), one bead containing 14 scout peptides was dissolved in 100  $\mu$ L of 0.5% FA in water/methanol (90:10, v/v). Then, 10  $\mu$ L of the scout peptide solutions was added to the final solution of peptides extracted from hemocytes or plasma, just before the shotgun and targeted LC-MRM/MS analysis.

Shotgun proteomic analyses were performed on hemocytes and plasma samples, which contained scout peptides, as previously described in Leprêtre et al. [13]. The resulting raw data from the shotgun proteomics including scout peptides were then analyzed by the online Protein Prospector software tool, which provides details on the retention times of peptides and their fragmentation spectra [31]. Retention times of scout peptides were used to create several groups of MRM transition acquisitions to facilitate the detection of targeted peptides during MRM experiments. Each target peptide was located in a group framed by two scout peptides.

### Standard solutions and quality controls

One milliliter of water/methanol (50/50, v/v) mixture containing 0.5% FA was added in each lyophilized tube containing accurately weighed isotopically labeled peptides. Then, a dilution to yield a concentration of 500  $\mu$ g/mL was mixed in water/methanol (90/10, v/v). A mixture stock solution (S0) containing all the labeled peptides was prepared at 5000 ng/mL. Solutions at 3000, 2000, 1000, 500, 200, 100, and 50 ng/mL were prepared in water/methanol (90/10, v/v) from the stock mixture solution and further diluted by 2 with extracted proteins from hemocytes or plasma samples to reach the desired concentrations of 25, 50, 100, 250, 500, 1000, 1500, and 2500 ng/mL. Two levels of quality control (QC) were prepared from the stock solution at concentrations of 375 ng/mL (i.e. QC1) and 2000 ng/mL (i.e. QC2).

### Method performance assessment

To evaluate the linearity of the assays, a standard curve was produced based on 16 samples containing equal amounts of plasma or hemocytes from *D. polymorpha* protein extract digests as background matrix. Each sample was spiked with an increasing amount of labeled peptide between 25 and 2500 ng/mL, spanning a 100-fold range. To assess linearity, each standard sample was analyzed three times. Each run consisted of one set of calibration standards and three intra-batch replicates of each QC concentration (375, and 2000 ng/mL). The analytical batch consisted of blank solvents, calibration curve samples, and QC samples. Blank solvents were introduced before and after the calibration curve and between QC samples to evaluate carryover. The curve for the standards was plotted according to least squares linear regression with a weighting factor of  $1/x^2$ . For each concentration level, the three replicates were taken into consideration. The calibration curves were qualified with a minimum of six of the eight concentration levels. For some labeled peptides ( $N=13$ ), the lowest levels (25 and/or 50 ng/mL) were removed to qualify the calibration curve.

For the validation criteria of the methodology, we adopted the same criteria used for the multiplexed quantification assay of *Gammarus fossarum*, as there is no guidance for analytical method validation in ecotoxicology [20]. In particular, we followed the Food and Drug Administration (FDA) bioanalytical method validation guidance for industry, with minor relaxation of performance criteria for precision and accuracy [38]. A criterion of precision determined at each concentration level should not exceed a percent CV above 20% and average accuracy in determining the expected concentration within 80–120%. If standard points for any level fell outside these ranges, the entire level was removed from the curve and the linear regression equation recalculated.



Calibration standards and the final calibration line contained at least five calibration concentrations.

### Biological validation of MRM methodology

To validate the specificity of reporter peptides, we monitored their signal and area ratios in hemocytes and plasma samples with contrasting immunological status. To do this, 150  $\mu\text{L}$  of hemolymph per mussel was withdrawn from the posterior adductor muscle by gentle aspiration with a  $30\text{ g} \times 0.3\text{ mL}$  syringe. The quality of each sample was checked under a microscope. The samples were then pooled to obtain 3 mL of hemolymph. From this pool (3 mL), 1 mL of hemolymph was distributed to ensure three exposure conditions: (i) the control condition where the hemolymph was not exposed to any immunomodulators, (ii) the anti-protozoa challenge where the hemolymph was exposed to *Cryptosporidium parvum* oocysts (5:1 protozoa:hemocyte; INRA, Val de Loire research center, France) as described in Le Guemic et al. [39], and (iii) the antiviral challenge where the hemolymph was exposed to 100  $\mu\text{g}/\text{mL}$  of RNA poly I:C (Sigma-Aldrich). For all the conditions, the samples were incubated at 16 °C for 4 h and the experiment was carried out five times to obtain five replicates per exposure condition. Then the hemocytes were separated from the plasma by centrifugation. After protein extraction and peptide digestion, the MRM methodology was performed to monitor peptides related to immune proteins in both the hemocytes and the plasma samples.

### Identification of modulated proteins

Samples used for the biological validation of MRM methods (see section 2.12) were also used for the identification of proteins modulated during immunological challenges. In order to avoid analytical bias and ensure comparability between the different biological samples, the abundance of peptides and proteins was normalized using the ratio between the peak areas of the endogenous peptides and their corresponding heavy labeled forms. For targeted peptides without internal standards, abundance values were normalized by the sum of the peak areas of all the heavy peptides present in the sample. In the hemocyte samples, the intensity of peptides without associated internal standards was normalized by the intensity of peptides derived from a reference protein, the elongation factor 1 alpha (EF1- $\alpha$ ). This protein was chosen as the reference protein for several reasons. Commonly used as a reference gene in quantitative PCR gene expression assays, this housekeeping gene has been shown to have consistent expression in the hemocytes of the flat oyster *Ostrea edulis* [40]. Furthermore, in previous differential proteomic analyses performed on zebra mussel hemocytes exposed to various biological contaminants [14], EF1- $\alpha$  appeared to be the most stable protein in the hemocyte replicates analyzed. Fold

changes were calculated for each protein by comparing their signal variations between challenged and control samples. Nonparametric Kruskal–Wallis tests with Dunn’s corrections were used to determine significant modulation, with a  $p$  value set to  $<0.05$ .

## Results and discussion

### Rapid transfer from proteogenomic data to a Scout-MRM assay

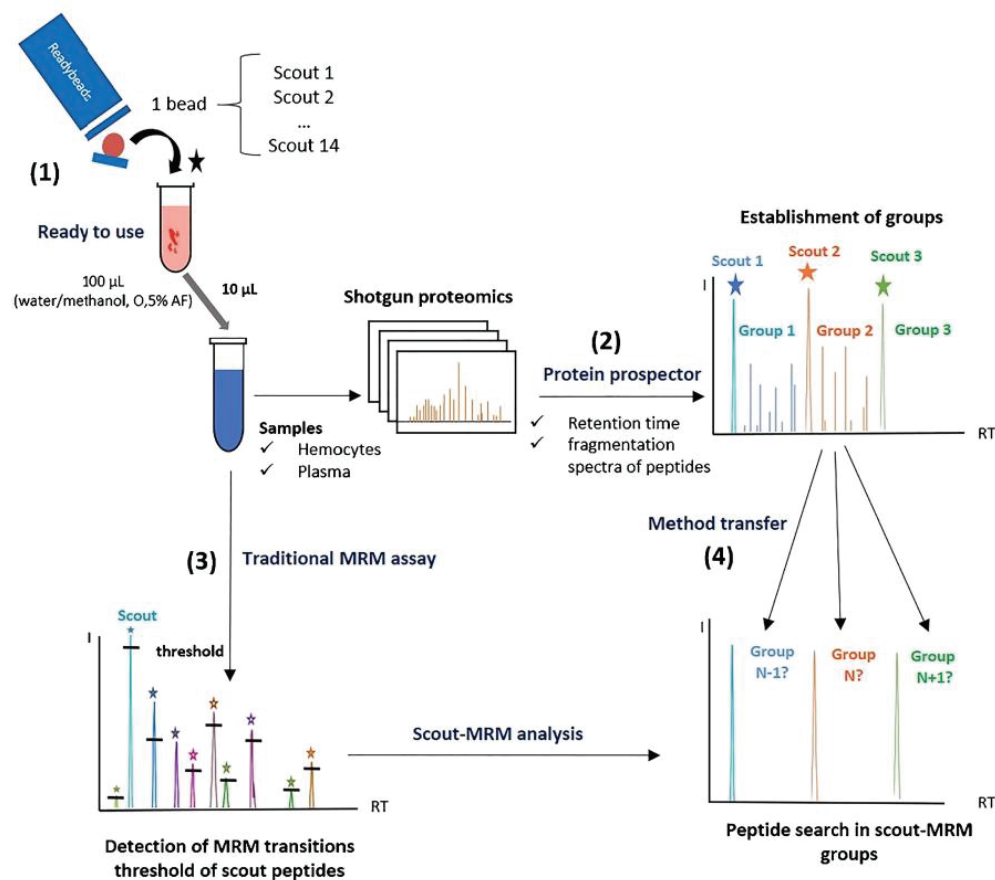
Previously, a proteogenomic analysis performed on the hemolymph of *D. polymorpha* allowed the mapping of immune elements in both hemocytes and plasma [13]. Differential proteomic analyses were then performed on pools of hemolymph exposed ex vivo to *C. parvum* oocysts and a synthetic double-stranded RNA (RNA poly I:C) to identify immune proteins modulated during interaction with microorganisms [14]. Based on these results, potential immunomarkers reflecting a wide range of immune mechanisms were selected for assessment by targeted proteomics approaches. Overall, a list of 114 proteins from the hemolymph, which include the hemocytes and the plasma, were initially selected as target candidates for MRM assays (ESM Fig. S1). Whenever possible, up to three reporter peptides per protein were selected to report the abundance of these targeted proteins. Indeed, for highly concentrated proteins with long polypeptide sequences, several peptides were identified by shotgun proteomics, allowing the selection of three to four reporter peptides per protein. On the contrary, for low-abundance proteins and/or with short polypeptide sequences, only one to two reporter peptides per protein could be selected. In total, 329 peptide surrogates of the panel of 114 proteins were selected as candidates for MRM assays (ESM Table S1). The three most intense fragment ions were then chosen for MRM transition settings, resulting in the selection of 987 MRM transitions.

Building a multiplexed MRM assay from shotgun proteomic experiments is not straightforward, as chromatographic parameters are often different between both the experimental set-up (i.e. nanoLC versus  $\mu\text{LC}$ ) and/or the chemistry of the stationary phase. Therefore, the retention times of each peptide must be redetermined. Without knowledge of retention times, we had to build several MRM methods to identify the best fragments and retention times for each peptide. The number of methods depends on the number of MRM transitions [28]. Indeed, liquid chromatography peaks must be defined by at least 15 data points, and the duty cycle of the mass spectrometer is divided into all transitions. For example, with a 15 s chromatographic peak width and a transition dwell time of 10 ms, only 100 transitions could be monitored in the same run. Eighteen transitions are required to quantify one protein with three surrogate peptides with each corresponding labeled



peptide and three fragments per peptide. Thus, in our example, for 114 proteins, 20 analytical runs should be necessary. For the rapid development of MRM assays, a new analytical approach derived from the Scout-MRM methodology and using “scout peptides” from Anaquant beads was developed to extract and select the targeted peptides for relative protein quantification from shotgun mass spectrometry runs (Fig. 1). To do this, peptides of known sequences (see ESM Tables S2 and S3) designed by Anaquant and named scout peptides were previously introduced in the hemocytes and plasma samples during the discovery phase (Fig. 1). Their sequences are unique and do not belong in any sequence known to date in the different species databases. These peptides were selected to be eluted throughout a 90 min chromatography run with a C<sub>18</sub> column. They were introduced into the samples via READYBEADS technology. This innovative technique,

based on a water-soluble biopolymer coated with 14 peptides, allows the rapid dissolution of the same amount of scout peptides in the samples. Thanks to these scout peptides, we can define for each peptide a group supervised by two scout peptides. Knowing the group for each peptide enables the rapid transfer from discovery proteomics to quantitative proteomics (Fig. 1). Indeed, a major breakthrough of the Scout approach relies on the fact that it releases acquisition from scheduling retention windows [27, 28]. The principle of the Scout-MRM method was described by Rougemont et al. [27]. Briefly, scout peptides are dispatched all along the chromatogram to successively manage the monitoring of transition lists (Fig. 2). Once the first scout peptide (scout1) is detected above a defined threshold, it triggers the monitoring of the first group of peptides (group1), independently of their respective retention times. As soon as the second scout is detected, the monitoring



**Fig. 1** Strategy to convert the quantification method from shotgun proteomics into targeted proteomics, using the Scout-MRM methodology. (1) READYBEADS containing 14 scout peptides are dissolved in hemocyte and plasma samples analyzed by shotgun and targeted proteomics. (2) Protein Prospector is used for extracting the three most intense fragment ions and the retention time of reporter peptides in order to set

MRM transitions and define reporter peptide search groups framed by scout peptides. (3) MRM transitions of scout peptides are monitored to define detection thresholds. (4) MRM transitions of reporter peptides are searched in different groups (N, N+1, N-1, and others) for their quantification by targeted proteomics

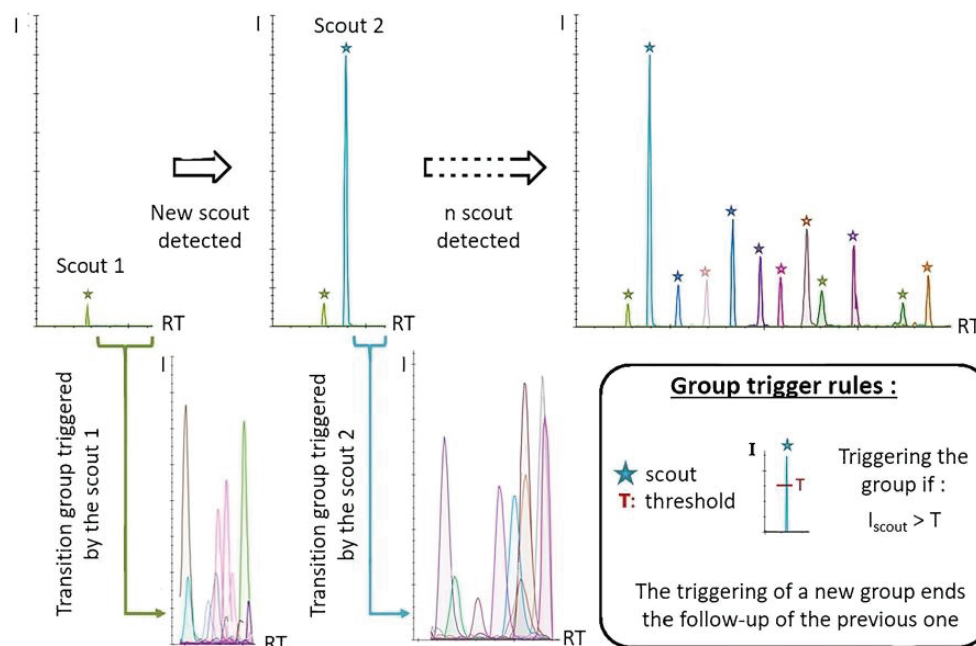


Fig. 2 Principle of Scout-MRM acquisition mode and development of a multiplexed targeted assay

of group 1 is switched off, while the monitoring of the MRM transitions defined in group 2 begins, and so on. Hence, exhaustive detection and analysis remain effective even when voluntary or random retention drifts occur.

The first step was performed to monitor transitions of scout peptides added to samples, so as to define thresholds (Fig. 1). These thresholds allow the detection and trigger the acquisition of MRM transitions groups during the application of the Scout-MRM methodology [27, 28]. Only the two major MRM transitions per scout protein-reporter-peptide were selected for further analysis. Scout-MRM analyses were then performed on both hemocytes and plasma samples containing scout peptides, to identify reporter peptides selected from shotgun proteomics by MRM mass spectrometry.

In total, 15 groups monitored by Scout-MRM were created, according to the retention times of the scout and targeted peptides observed during the discovery phase. As seen in Table 1, 139 out of 263 peptides from hemocyte samples and 151 out of 275 peptides from the plasma samples were detected in a single Scout-MRM analysis, with no dependence on their retention times being necessary. Several peptides were not detected directly in the N group, probably because of a change in chromatographic selectivity. Indeed, LC gradients, the stationary phase, and diameters of the chromatographic columns were considerably different between MRM and shotgun analyses (microscale XBridge C<sub>18</sub> column 2.1 mm versus nano-scale 500  $\mu$ m C<sub>18</sub> PepMap<sup>TM</sup> 100 from LC Packings). The use of the same type of stationary phase would make it possible to

recover almost all the compounds, as we demonstrated for pesticide effect studies using the Scout approach [28]. Then, a second Scout-MRM method was built in which missing peptides were assigned to the groups preceding (N-1) and following (N+1) group N. In all, 87 and 91 additional peptides in hemocyte and plasma samples were recovered, respectively, thanks to this second run (Table 1). Finally, 12 additional peptides were also confirmed by considering other groups such as N+2 and N-2. Overall, only 36 peptides (= 11.2%) out of the 329 candidates to be assessed in the hemolymph of *D. polymorpha* were never detected (Table 1). According to these results, two Scout-MRM methods were developed (ESM Fig. S1). The first one (ESM Table S2) allows the detection of 235 peptides and 58 proteins from hemocyte samples, while the second method includes the detection of 250 peptides and 89 proteins from plasma of *D. polymorpha* (ESM Table S3). Once the MRM transitions were detected, the protocol of sample preparation was optimized and the MRM methods were validated to ensure the efficient detection and further evaluation of targeted peptides and proteins.

## Optimization and validation of MRM assays

### Optimization of protein digestion

Protein digestion step optimization is required for developing a successful protein assay with a shorter preparation time and higher reproducibility and sensitivity. To investigate the rate



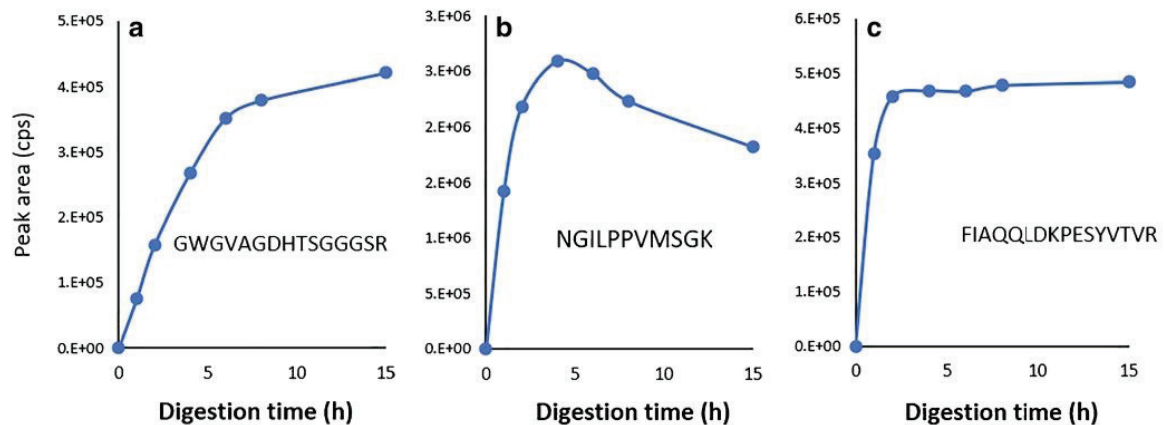
From shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential...

**Table 1** Number of peptides and proteins identified during method transfer from shotgun to targeted proteomics

	Hemolymph (hemocytes+plasma)		Hemocytes		Plasma	
	Proteins	Peptides	Proteins	Peptides	Proteins	Peptides
Selected proteins and peptides from shotgun proteomic data	114	329	92	263	92	275
Proteins and peptides detected by Scout- MRM groups						
N	64	177	49	139	54	151
N-1	21	58	19	48	16	50
N+1	17	46	14	39	16	41
Other	5	12	3	9	3	8
Total number of proteins and peptides detected by Scout-MRM	107	293	85	235	89	250
Non-detected	7	36	7	28	3	25

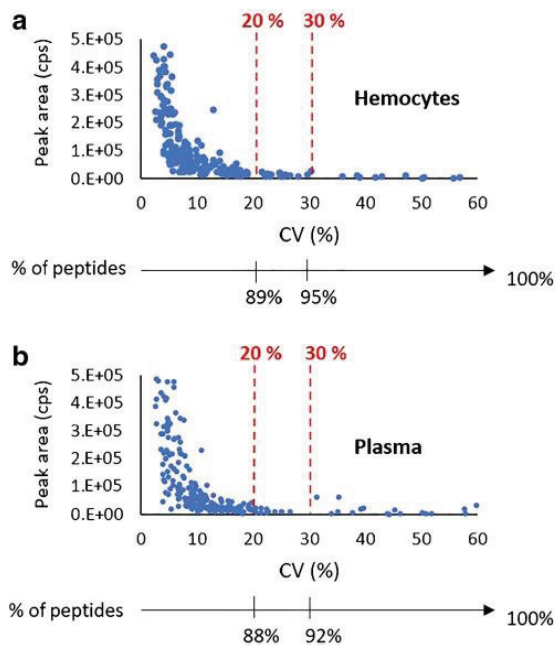
of protein digestion, a time course experiment was performed for all peptides. Protein digestion was conducted with TPCK trypsin to reduce contaminating chymotryptic activity while maintaining a reasonable cost per assay. The best protein-to-trypsin ratio was fixed at 30:1, and digestion time varied between 1 h and 15 h. As reported in Fig. 3, three types of profiles can be observed, corresponding to 60% (Fig. 3a), 14% (Fig. 3b), and 26% (Fig. 3c), respectively, of the 293 peptides monitored in the hemolymph (which includes the hemocytes and plasma). A digestion time of 14 h was required to reach a plateau for the peptide area, except for profile B. In this group, we observed a decrease in the signal after 6 h. This decrease may indicate the degradation and/or the chemical modification of the peptide once it is released from the protein. As for the multiplexed assay, it was necessary to reach a compromise, so a digestion time of 15 h was selected for practical reasons (overnight digestion), during which most of the peptides are well digested.

After defining a compromise for the trypsin digestion duration, the repeatability of peptide abundance measurements (biological sample preparation, digestion, Scout-MRM analysis) was assessed, as it undoubtedly represents the key performance parameter that ensures reliable comparisons of the protein expression level. This criterion was assessed by analyzing, in singleton, 10 aliquots of pooled plasma and 10 aliquots of pooled hemocytes. As shown in Fig. 4, good reproducibility was obtained for the peptides in the hemocyte (Fig. 4a) and plasma samples (Fig. 4b), as more than 88% of them were measured with a CV better than 20%. As expected, low-intensity peptides (about 5% in hemocyte and plasma samples) were associated with poorer CV values (> 30%), and their assays were therefore considered non-reproducible (Fig. 4). Once the reproducibility of the endogenous peptides was assessed, we proceeded to the evaluation of the analytical performance of the MRM method developed in this study.



**Fig. 3** Digestion efficiency at six different incubation times (1, 2, 4, 6, 8, and 15 h) of the 293 peptides monitored in the hemolymph (which includes the hemocytes and the plasma). Three different profiles were obtained. The profile (a) was observed for 175 peptides, including the

peptide ADPALGQAIQER, the profile (b) was observed for 41 peptides, including the peptide NGILPPVMSGK, and other peptides (77) followed the profile (c), as shown for the peptide FIAQQLDKPESYVTVR



**Fig. 4** Repeatability of peptide abundance measurements by Scout-MRM approach. %CV was calculated as the ratio of the standard deviation to the mean of peptide abundance from 10 pooled hemocyte (a) or plasma (b) samples analyzed in singleton

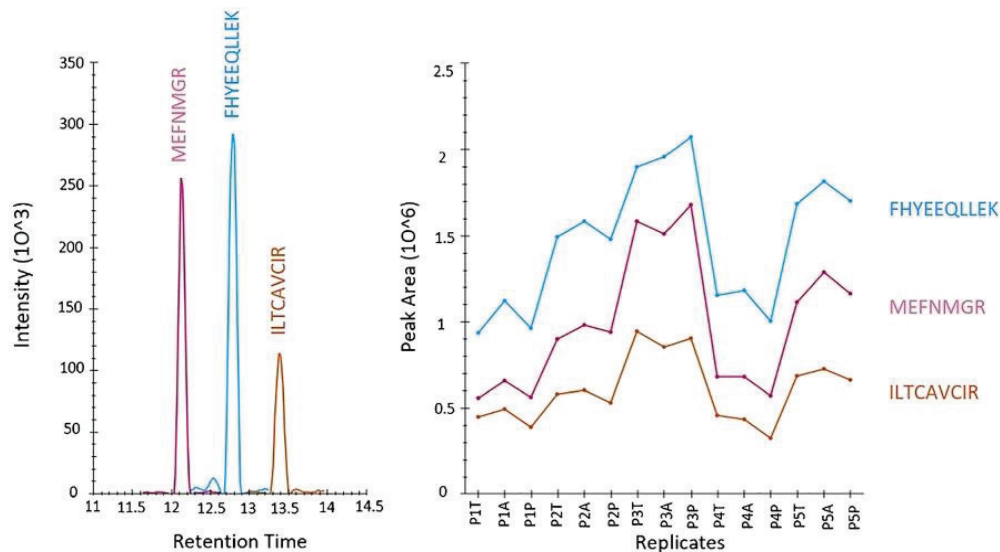
#### Validation of MRM transitions and peptides

High MRM transition specificity is essential to consistently reflect the true abundance of the peptides targeted, and hence to ensure the reliable comparison of protein expression levels across a biological sample collection. Given the large number of analytes present in biological samples, the probability that a transition has been subject to interference is relatively high [41]. Additionally, low-intensity transitions may also be susceptible to interference from the background noise of the mass spectrometer. The use of transitions subject to interference could lead to the inaccurate area determination of their corresponding peptides. Therefore, interference detection is a crucial step in ensuring that transitions are free of interference and reflect the abundance of the targeted peptides. In our study, transitions were screened for interference by monitoring the three MRM transitions selected to identify and possibly assess the abundance of peptides targeted in the hemocytes (235 peptides) and plasma (250 peptides) samples, as done in several studies [20, 42, 43]. To this end, peak area ratios between transitions were evaluated for 15 hemocyte and 15 plasma samples exposed to different stressful conditions for biological variability, with a sufficient number of samples to be representative. Transitions were considered as interference-free when a coefficient of variation (CV) below 20% of the average relative ratio of the MRM transitions was recorded. In

addition, when an internal standard was available, it was also verified that retention times and transition ratios were exactly the same between exogenous and endogenous peptides. According to these criteria, 118 transitions in the hemocyte samples and 132 transitions in the plasma samples were considered to have been subject to interference (ESM Fig. S1). From these results, the most intense transitions, free of interference, were selected to assess the abundance of peptides.

Once the transitions have been validated, the peptides used for the protein qualification and/or quantification must be approved. Indeed, to be representative of the protein abundance from which they are derived, signature peptides have to be specific to this protein, and their abundance must be correlated regardless of the biological sample [20]. As illustrated in Fig. 5, peptides derived from a protein have to follow the same dynamics in different biological samples, in order to validate their use as a reporter of the targeted protein abundance. For proteins identified with multiple peptides, ratios of responses between peptides were calculated for the different samples. When a CV of less than 20% of the relative mean of the peptide areas was recorded, the peptides were validated to estimate the abundance of the protein from which they were derived. While most of the proteins were qualified with several peptides, some were still qualified with a single peptide. Overall, 18 and 24 peptides were considered as not suitable for assessing the protein abundance and were removed from MRM methods of hemocyte and plasma samples, respectively (ESM Fig. S1). Among the 85 proteins qualified for the hemocyte samples, 14 proteins were qualified according to four signature peptides, 30 proteins with three peptides, 30 proteins with two peptides, and 11 proteins with only one peptide. For the 89 proteins monitored in plasma samples, 14 proteins were qualified according to four signature peptides, 30 proteins with three peptides, 35 proteins with two peptides, and 10 proteins with one peptide. As for the validation of the transitions, the ratios between each reporter peptide of a protein in the different samples were calculated to select the best peptide for assessing the abundance of the targeted proteins. Peptides were selected as a good abundance reporter of proteins when CV values between the peptide ratio were lower than 20%. For example, as shown in Fig. 5, the three peptides derived from the laccase-like protein (FHYEEQLLEK, MEFNMGR, and ILTCAVCIR) with a CV ratio lower than 20% followed the same dynamics in the different plasma samples of *D. polymorpha*, and their ratios were constant across the different biological samples. In this specific case, the FHYEEQLLEK was selected as the reporter of the laccase-like protein, since its intensity was higher than the other peptides. When CV values between peptide ratios were higher than 20% and for proteins with a single





**Fig. 5** Correlation of the chromatographic peak area between three peptides of the laccase-like protein (FHVEEQLLEK, MEFNMGR, and ILTCAVCIR) in plasma samples

reporter peptide, the values were still used only if the MRM transition ratios were validated.

#### Evaluation of analytical performance assay

Once the transitions and peptides had been selected, the performance of the assay was evaluated. As internal standard, 50 synthesized peptides were used, corresponding to 50 proteins selected for their putative major role in zebra mussel immunity (Table S1 in blue). A mix solution, containing all heavy peptides at a known concentration, was prepared and added to the hemocyte and plasma samples analyzed by LC-MRM/MS. Three transitions per heavy peptide were then monitored in the samples by Scout-MRM. Overall, 38 and 40 labeled peptides were monitored in the hemocytes and plasma, respectively (Table 2).

Linearity assessments were evaluated by a standard curve prepared in equal amounts of digested protein extract. Calibration curves were generated by spiking labeled peptides in the samples. Ideally, the concentration range expected should be known before carrying out the validation of the analytical method. This range should be covered by the calibration curve range. Unfortunately, no information on the expected concentrations was available. Compared to our experiment with other sentinel species like *Gammarus fossarum* [20], we therefore arbitrarily chose to cover a 100-fold range at a final concentration ranging from 25 to 2500 ng/mL. Then, a weighted (1/×2) least squares linear regression of area response versus concentration was used for the calibration. As reported in Table 2, the calibration curves were linear, as the coefficient of correlation  $r^2 > 0.9$  was observed for most of the

labeled peptides. No carryover was observed after the upper limit of quantification. After evaluating the linearity, assay accuracy (mean determined concentration/nominal concentration × 100) and precision (SD/mean concentration × 100) were evaluated by analyzing two quality controls (QC), at medium (375 ng/mL) and high concentration (2000 ng/mL), in triplicate. As shown in Table 2, in both the hemocyte and plasma samples, the intra-day accuracy of QC1 and QC2 ranged from 84.3 to 114.1%, except for the peptide “SVNCVSVNDPFVMDAFGK” in the plasma samples (QC1: 127.1%), and precision ranged from 0.6 to 16.9%. As the performance of the assay was in the range of acceptance, labeled peptides were used to spike the hemocyte and plasma samples at a concentration of 500 ng/mL (midpoint of calibration curves) for further investigations. The results are summarized in Table 2, and examples of calibration curves as well as extracted-ion chromatograms (XICs) of the labeled peptides at 500 ng/mL are shown in the ESM Fig. S2.

#### Application of the assay

As described above, optimized Scout-MRM assays were developed to investigate the immune response of *D. polymorpha* hemolymph during stressful conditions. In total, 15 hemocyte samples and 15 plasma samples were analyzed according to the two Scout-MRM methods developed for each sample type (ESM Fig. S1). A control peptide mixture was injected throughout the analytical runs to ensure the stability of the mass spectrometer. This reference was always stable for series of up to 15 samples analyzed. The relative abundance of 85 immune proteins in the hemocyte and 89 proteins in the

**Table 2** MRM assay linearity and inter-sample accuracy and precision of labeled peptides, in hemocyte and plasma samples

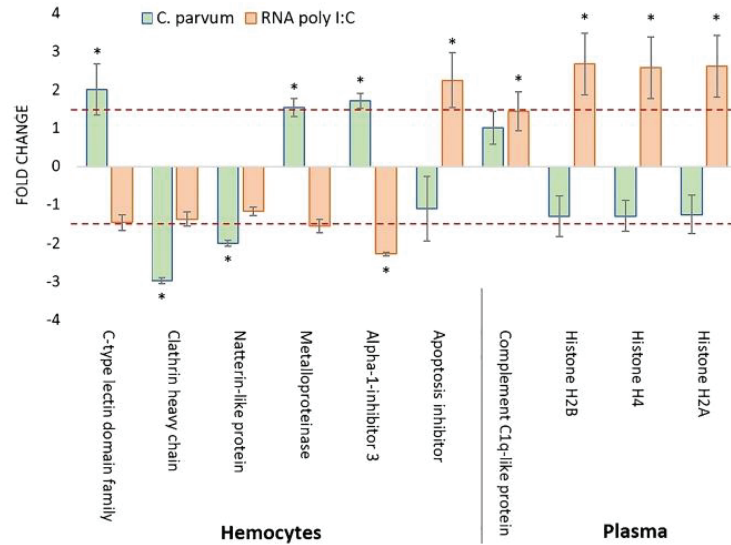
Labeled peptides	Hemocyte samples					Plasma samples				
	Correlation coefficient ( $r^2$ )	QC 1 (375 ng/mL)		QC 2 (2000 ng/mL)		Correlation coefficient ( $r^2$ )	QC 1 (375 ng/mL)		QC 2 (2000 ng/mL)	
		Accuracy (%)	Precision (RSD %)	Accuracy (%)	Precision (RSD %)		Accuracy (%)	Precision (RSD %)	Accuracy (%)	Precision (RSD %)
APSSPSALR	0.991	105.5	5.9	104.3	4.9	0.993	98.9	9.6	103.4	9.1
FNFGYDR	0.994	99.2	2.2	100.0	2.4	0.997	111.2	4.7	104.2	4.0
FSIYLQR	0.991	105.9	4.3	103.5	2.9	0.996	111.2	5.4	106.2	3.7
LGAIWFQINIK	0.996	105.6	9.8	106.7	3.8	0.993	98.9	4.5	96.9	1.4
NGILPPVMSGK	0.996	112.6	9.5	109.2	2.5	0.994	110.7	10.4	102.2	4.0
LITGIATR	–	–	–	–	–	0.994	105.9	2.9	100.6	2.6
ITQLNMWSR	–	–	–	–	–	0.991	109.6	5.2	101.8	8.2
SLVNAVIR	–	–	–	–	–	0.997	105.7	4.2	102.3	1.3
VVFTALK	0.995	104.7	2.6	105.7	0.8	0.999	104.9	4.4	101.6	2.1
LNIPODR	0.999	98.7	1.4	97.5	2.7	0.997	108.6	2.6	105.4	3.9
ICLDILK	0.994	105.8	5.0	108.6	3.0	0.995	107.5	4.1	106.1	3.1
VALQALK	0.997	106.1	2.7	105.0	1.8	–	–	–	–	–
TPCNVECFR	0.992	110.3	6.5	106.4	7.8	–	–	–	–	–
AIIGEYIYSHSK	0.998	111.2	5.1	110.3	3.2	0.996	104.5	4.1	99.0	2.4
YLECSALTQQLK	0.995	113.3	2.9	111.4	4.6	–	–	–	–	–
LSAEFLK	–	–	–	–	–	0.995	108.3	2.7	105.9	2.2
GISNIPINLR	0.996	110.0	2.1	112.3	1.8	–	–	–	–	–
EMGFEIR	0.995	98.9	4.0	100.0	1.6	–	–	–	–	–
YVECSALTQK	0.996	99.5	6.7	103.2	1.9	0.985	100.4	6.4	106.5	5.5
GLAVFISDIR	0.992	98.7	2.9	100.7	1.6	–	–	–	–	–
LLLPLWLEMR	0.991	92.8	3.9	100.8	4.7	0.987	106.8	7.2	102.3	7.4
SLPSGGGGGLAGNPHVK	0.994	113.0	2.9	114.2	5.2	0.995	101.3	5.2	98.9	10.3
LQVPLVYSTLTK	0.997	102.2	5.1	102.7	4.8	0.992	96.9	4.0	106.3	5.0
VGEVVTILGYALQR	0.992	104.6	5.9	104.7	4.5	–	–	–	–	–
LTLSTMLEGLSAGGR	0.996	106.4	13.1	111.9	5.2	0.991	105.8	2.9	106.6	6.9
FVGLDSGFR	0.997	109.2	4.2	109.4	2.0	0.993	107.3	4.7	106.6	3.4
VYVGDLYGAAK	0.992	99.7	2.2	100.6	4.3	–	–	–	–	–
FLYCPSGYR	0.994	109.6	2.6	109.1	1.3	0.996	108.5	4.0	100.1	0.6
WTWVPHWWSGR	0.992	103.5	7.8	106.2	4.6	0.984	96.6	15.4	94.3	4.1
AVNQALLDLHK	0.994	109.2	4.0	109.7	2.3	–	–	–	–	–
FIINVCR	–	–	–	–	–	0.996	101.8	4.1	102.6	2.4
ILLIGPVGSGK	0.993	105.9	4.6	106.2	1.8	0.994	100.6	6.8	96.6	3.3
LAPDSVAASEMFVR	0.992	91.6	8.2	95.6	6.6	–	–	–	–	–
DTDILSFILPHTAR	–	–	–	–	–	0.994	108.6	1.6	109.4	2.4
GNQVLDWVFK	0.997	101.5	8.5	104.3	2.8	0.991	109.3	5.3	108.6	4.6
QLLIEPEGQPHVATR	0.993	94.3	3.3	88.2	1.9	0.995	106.2	3.0	98.5	2.6
IFSPLPSK	0.995	111.3	5.7	112.8	1.7	0.996	101.2	2.8	99.4	2.2
IGTEQHFFAIDINK	–	–	–	–	–	0.991	102.0	8.0	100.3	7.3
NFFADPR	–	–	–	–	–	0.993	114.3	5.7	108.1	4.6
WVDNINCKPDR	0.990	111.1	3.3	112.5	4.9	0.995	105.0	4.1	101.1	4.0
TQWPNCPSTIK	0.999	109.7	5.9	110.0	2.8	0.997	105.7	4.0	102.0	1.5
DVYLPAAAR	–	–	–	–	–	0.997	114.2	3.0	105.9	7.3
SLLPVLIDIEPK	0.988	84.3	16.9	101.8	4.0	0.996	88.4	7.7	105.7	5.4
ISEAISK	–	–	–	–	–	0.997	104.0	2.9	103.6	4.6
SVNCVSVNDPFVMDAFGK	0.986	97.5	8.2	106.2	7.6	0.970	127.1	14.3	107.9	9.8
LVGALPNAIK	0.997	105.7	2.7	103.7	1.0	0.996	–	–	–	–
AIFDIANWEEVGNR	0.995	114.9	2.2	113.4	2.6	0.993	108.5	13.6	101.0	3.4
ELGTVMR	0.997	110.5	4.6	112.2	2.3	0.996	107.4	4.1	102.5	5.5
VSGVLTIPR	–	–	–	–	–	0.995	108.0	10.1	104.6	5.4
AIAEQFAK	–	–	–	–	–	0.992	108.0	5.6	109.8	5.2

plasma samples was compared between control conditions and samples exposed to *C. parvum* oocysts or RNA poly I:C. According to Dunn's test, only the protein expressions modulated with a fold change greater than 1.5 were considered. As shown in Fig. 6, different modulation patterns were observed depending on whether the samples were exposed or not (control) to *C. parvum* or RNA poly I:C. In the case of hemolymph challenged with *C. parvum*, proteins related to pattern recognition receptors (PRRs), proteases, and protease

inhibitors were significantly affected. Among the PRRs, a C-type lectin was modulated by a twofold change. Lectins are a group of PRRs that recognize carbohydrates on the surface of a microorganism's wall. C-type lectins have been shown to be modulated in the hemolymph of bivalves exposed to several kinds of microorganisms, including bacteria, protozoa, and viruses [44]. A metalloproteinase and a serine protease inhibitor (alpha-1 inhibitor) were also modulated in hemolymph exposed to *C. parvum*. Proteases and serine protease inhibitors



**Fig. 6** Example of significantly modulated proteins in hemocytes exposed to RNA poly I:C and *C. parvum*. The figure shows the relative abundance in comparison to the control condition. The red dotted lines correspond to a fold change of 1.5. Stars show significant fold changes according to Dunn's test with an error risk lower than 0.05%



are well known to participate in the immune defense of bivalves by cleaving proteins involved in immune signaling proteolytic cascades or by degrading microorganism components or proteases [45–47]. In the hemolymph exposed to RNA poly I:C, an apoptosis inhibitor protein was up-modulated in hemocytes with a fold change higher than 2.2 (Fig. 6). It is well established that apoptosis processes are involved during host interactions with viruses [48]. Green et al. [49] also hypothesized that apoptosis may represent an important antiviral response of mollusks. Modulation of key antiapoptotic factors regulating the apoptotic signaling pathway has been observed during viral infections of several mollusk species, including the abalone *Haliotis diversicolor*, the oyster *Crassostrea gigas*, and the clam *Scapharca broughtonii* [50–52]. In plasma, a C1q containing domain protein was significantly modulated by the RNA poly I:C. C1q protein acts as a PRR with a broad recognition spectrum in bivalves and has been modulated during exposure to RNA poly I:C in the oyster *C. gigas* [53, 54]. Histones H2B, H4, and H2A were also highly modulated, with a fold change greater than 2.5, in the plasma of *D. polymorpha* challenged with RNA poly I:C. In the innate immune system, histones are considered powerful antimicrobial agents and have shown a wide range of actions, including the permeabilization of the bacterial cell membrane and binding to bacterial DNA and/or RNA [55]. It was recently reported that the release of histones in the extracellular fraction of the oyster *C. gigas* is associated with DNA extracellular trapping mechanisms, which allow the entrapment and destruction of microorganisms [56]. The up-regulation in the plasma of *D. polymorpha* exposed to RNA poly I:C may indicate that DNA extracellular trapping occurs to neutralize RNA double strand breaks.

## Conclusion

Multiplexed LC-MRM/MS assays were developed for measuring, for the first, 217 and, for the second, 226 proteotypic peptides used as surrogates of 85 and 89 immune proteins purified from either the hemocytes or plasma of *Dreissena polymorpha*. To our knowledge, this is one of the most comprehensive MRM methods developed for sentinel species used in ecotoxicology. After further validation (inter-batch, stability, precisely determination of the limits of quantification for each peptide, etc.), these multiplexed targeted assays will be used in future *Dreissena polymorpha* investigations to evaluate the expression dynamics of immune proteins during exposure to various environmental contaminants. Moreover, we proposed a new application of Scout-MRM, which allows very rapid development of an MRM analysis method for a non-model species. In our case, the Scout workflow made possible the quick development of MRM assays from proteogenomic data acquired from zebra mussels. This methodology could easily be extrapolated to other organisms provided that scout peptides are introduced during the discovery phase. In addition, by overcoming the issue of peptide retention times, the Scout-MRM methodology facilitates the transition between shotgun and targeted proteomics performed with different chromatographic systems.

**Acknowledgments** We thank the University of Reims Champagne-Ardenne for its financial support (project PROMETHEE). This work benefitted from the French GDR “Aquatic Ecotoxicology” framework which is aimed at fostering stimulating scientific discussions and collaborations for more integrative approaches.



### Compliance with ethical standards

The authors declare the following financial interests: Jérôme Lemoine filed a patent describing the Scout-MRM concept for targeted analysis by mass spectrometry. The authors declare that they have no conflict of interest.

### References

- Oikari A, Boon JP, Fairbrother A, Fossi MC, Rand G, Renzoni A, et al. Development and Validation of Biomarkers. *Biomarkers*. Springer Berlin Heidelberg 1993;63–77. [https://doi.org/10.1007/978-3-642-84631-1\\_5](https://doi.org/10.1007/978-3-642-84631-1_5).
- Gosset A, Ferro Y, Durieu C. Methods for evaluating the pollution impact of urban wet weather discharges on biocenosis: a review. *Water Res*. 2016;89:330–54. <https://doi.org/10.1016/j.watres.2015.11.020>.
- Gouveia D, Almunia C, Cogne Y, Pible O, Degli-Esposti D, Salvador A, et al. Ecotoxicoproteomics: a decade of progress in our understanding of anthropogenic impact on the environment. *J Proteome*. 2019;198:66–77. <https://doi.org/10.1016/j.jprot.2018.12.001>.
- Lemos MFL, Soares AMVM, Correia AC, Esteves AC. Proteins in ecotoxicology – how, why and why not? *Proteomics*. 2010;10:873–87. <https://doi.org/10.1002/pmic.200900470>.
- Monsinjon T, Knigge T. Proteomic applications in ecotoxicology. *Proteomics*. 2007;7:2997–3009. <https://doi.org/10.1002/pmic.200700101>.
- Bjørnstad A, Larsen BK, Skadsheim A, Jones MB, Andersen OK. The potential of ecotoxicoproteomics in environmental monitoring: biomarker profiling in mussel plasma using ProteinChip array technology. *J Toxicol Environ Health Part A*. 2006;69:77–96. <https://doi.org/10.1080/15287390500259277>.
- Apraiz I, Cajaraville MP, Cristobal S. Peroxisomal proteomics: biomonitoring in mussels after the Prestige's oil spill. *Mar Pollut Bull*. 2009;58:1815–26. <https://doi.org/10.1016/j.marpolbul.2009.08.004>.
- Wang J, Wei Y, Wang D, Chan LL, Dai J. Proteomic study of the effects of complex environmental stresses in the livers of goldfish (*Carassius auratus*) that inhabit Gaobeidian Lake in Beijing, China. *Ecotoxicology*. 2008;17:213–20. <https://doi.org/10.1007/s10646-007-0187-4>.
- Trapp J, Armengaud J, Pible O, Gaillard J-C, Abbaci K, Habtoul Y, et al. Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors. *J Proteome Res*. 2015;14:292–303. <https://doi.org/10.1021/pr500984z>.
- Gouveia D, Chaumot A, Charnot A, Queau H, Armengaud J, Almunia C, et al. Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species *Gammarus fossarum*: a physiological and ecotoxicological study. *Aquat Toxicol*. 2017;190:199–209. <https://doi.org/10.1016/j.aquatox.2017.07.007>.
- Borcier E, Artigaud S, Gaillard J-C, Armengaud J, Charrier G, Couteau J, et al. Coupling caging and proteomics on the European flounder (*Platichthys flesus*) to assess the estuarine water quality at micro scale. *Sci Total Environ*. 2019;695:133760. <https://doi.org/10.1016/j.scitotenv.2019.133760>.
- Palos-Ladeiro MP, Barjhoux I, Bigot-Clivot A, Bonnard M, David E, Dedourge-Geffard O, et al. Mussel as a Tool to Define Continental Watershed Quality. In: Ray S, editor. *Organismal and Molecular Malacology*, InTech. 2017. <https://doi.org/10.5772/67995>.
- 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 Proteome*. 2019;202:103366. <https://doi.org/10.1016/j.jprot.2019.04.016>.
- Leprêtre M, Almunia C, Armengaud J, Le Guemec A, Salvador A, Geffard A, et al. Identification of immune-related proteins of *Dreissena polymorpha* hemocytes and plasma involved in host-microbe interactions by differential proteomics. *Sci Rep*. 2020;10:1–18. <https://doi.org/10.1038/s41598-020-63321-z>.
- Kockmann T, Trachsel C, Panse C, Wahlander A, Selevsek N, Grossmann J, et al. Targeted proteomics coming of age – SRM, PRM and DIA performance evaluated from a core facility perspective. *Proteomics*. 2016;16:2183–92. <https://doi.org/10.1002/pmic.201500502>.
- Bourmaud A, Gallien S, Domon B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: principle and applications. *Proteomics*. 2016;16:2146–59. <https://doi.org/10.1002/pmic.201500543>.
- 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>.
- Schiffmann C, Hansen R, Baumann S, Kublik A, Nielsen PH, Adrian L, et al. Comparison of targeted peptide quantification assays for reductive dehalogenases by selective reaction monitoring (SRM) and precursor reaction monitoring (PRM). *Anal Bioanal Chem*. 2014;406:283–91. <https://doi.org/10.1007/s00216-013-7451-7>.
- Ronsein GE, Pamir N, von Haller PD, Kim DS, Oda MN, Jarvik GP, et al. Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J Proteome*. 2015;113:388–99. <https://doi.org/10.1016/j.jprot.2014.10.017>.
- Charnot A, Gouveia D, Armengaud J, Almunia C, Chaumot A, Lemoine J, et al. Multiplexed assay for protein quantification in the invertebrate *Gammarus fossarum* by liquid chromatography coupled to tandem mass spectrometry. *Anal Bioanal Chem*. 2017;409:3969–91. <https://doi.org/10.1007/s00216-017-0348-0>.
- Saleh S, Staes A, Deborgraeve S, Gevaert K. Targeted proteomics for studying pathogenic Bacteria. *Proteomics*. 2019;19:1800435. <https://doi.org/10.1002/pmic.201800435>.
- Gouveia D, Chaumot A, Charnot A, Almunia C, François A, Navarro L, et al. Ecotoxicoproteomics for aquatic environmental monitoring: first in situ application of a new proteomics-based multibiomarker assay using caged amphipods. *Environ Sci Technol*. 2017;51:13417–26. <https://doi.org/10.1021/acs.est.7b03736>.
- Hoffman MA, Fang B, Haura EB, Rix U, Koomen JM. Comparison of quantitative mass spectrometry platforms for monitoring kinase ATP probe uptake in lung Cancer. *J Proteome Res*. 2018;17:63–75. <https://doi.org/10.1021/acs.jproteome.7b00329>.
- Chen Y, Vu J, Thompson MG, Sharpless WA, Chan LJJ, Gin JW, et al. A rapid methods development workflow for high-throughput quantitative proteomic applications. *PLoS One*. 2019;14:e0211582. <https://doi.org/10.1371/journal.pone.0211582>.
- Gallien S, Peterman S, Kiyonami R, Souady J, Duriez E, Schoen A, et al. Highly multiplexed targeted proteomics using precise control of peptide retention time. *Proteomics*. 2012;12:1122–33. <https://doi.org/10.1002/pmic.201100533>.
- Gallien S, Kim SY, Domon B. Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring (IS-PRM). *Mol Cell Proteomics*. 2015;14:1630–44. <https://doi.org/10.1074/mcp.O114.043968>.
- Rougemont B, Bontemps Gallo S, Aycirieux S, Carrière R, Hondermarck H, Lacroix JM, et al. Scout-MRM: multiplexed



- targeted mass spectrometry-based assay without retention time scheduling exemplified by *Dickeya dadantii* proteomic analysis during plant infection. *Anal Chem.* 2017;89:1421–6. <https://doi.org/10.1021/acs.analchem.6b03201>.
28. 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>.
  29. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *PNAS.* 2003;100:6940–5. <https://doi.org/10.1073/pnas.0832254100>.
  30. Chiva C, Sabido E. Peptide selection for targeted protein quantitation. *J Proteome Res.* 2017;16:1376–80. <https://doi.org/10.1021/acs.jproteome.6b00115>.
  31. Chalkley RJ, Baker PR, Huang L, Hansen KC, Allen NP, Rexach M, et al. Comprehensive Analysis of a Multidimensional Liquid Chromatography Mass Spectrometry Dataset Acquired on a Quadrupole Selecting, Quadrupole Collision Cell, Time-of-flight Mass Spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol Cell Proteomics.* 2005;4:1194–1204.
  32. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976;72:248–54.
  33. Weiß N, Wente W, Müller P. Eppendorf LoBind®: evaluation of protein recovery in Eppendorf protein LoBind tubes and plates. *Eppendorf Appl Note.* 2010;180:5.
  34. Goebel-Stengel M, Stengel A, Taché Y, Reeve JR. The importance of using the optimal plastic and glassware in studies involving peptides. *Anal Biochem.* 2011;414:38–46. <https://doi.org/10.1016/j.ab.2011.02.009>.
  35. MacLean B, Tomazela DM, Abbatiello SE, Zhang S, Whiteaker JR, Paulovich AG, et al. Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem.* 2010;82:10116–24. <https://doi.org/10.1021/ac102179j>.
  36. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26:966–8. <https://doi.org/10.1093/bioinformatics/btq054>.
  37. Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, et al. PASSEL: the PeptideAtlas SRM Experiment Library. *Proteomics.* 2012;12:1170–5. <https://doi.org/10.1002/pmic.201100515>.
  38. Kadian N, Raju KSR, Rashid M, Malik MY, Taneja I, Wahajuddin M. Comparative assessment of bioanalytical method validation guidelines for pharmaceutical industry. *J Pharm Biomed Anal.* 2016;126:83–97. <https://doi.org/10.1016/j.jpba.2016.03.052>.
  39. Le Guernic A, Geffard A, Rioult D, Bonnard I, Le Foll F, Palos LM. First evidence of cytotoxic effects of human protozoan parasites on zebra mussel (*Dreissena polymorpha*) haemocytes. *Environ Microbiol Rep.* 2019;11:414–8. <https://doi.org/10.1111/1758-2229.12720>.
  40. Morga B, Arzul I, Faury N, Renault T. Identification of genes from flat oyster *Ostrea edulis* as suitable housekeeping genes for quantitative real time PCR. *Fish Shellfish Immunol.* 2010;29:937–45. <https://doi.org/10.1016/j.fsi.2010.07.028>.
  41. Bao Y, Waldemarson S, Zhang G, Wahlander A, Ueberheide B, Myung S, et al. Detection and correction of interference in SRM analysis. *Methods.* 2013;61:299–303. <https://doi.org/10.1016/j.ymeth.2013.05.008>.
  42. Chen Y-T, Chen H-W, Domanski D, Smith DS, Liang K-H, Wu C-C, et al. Multiplexed quantification of 63 proteins in human urine by multiple reaction monitoring-based mass spectrometry for discovery of potential bladder cancer biomarkers. *J Proteome.* 2012;75:3529–45. <https://doi.org/10.1016/j.jprot.2011.12.031>.
  43. Percy AJ, Chambers AG, Yang J, Borchers CH. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics.* 2013;13:2202–15. <https://doi.org/10.1002/pmic.201200316>.
  44. Gerdol M, Gomez-Chiarri M, Castillo MG, Figueras A, Fiorito G, Moreira R, et al. Immunity in Molluscs: recognition and effector mechanisms, with a focus on Bivalvia. *Advances in comparative immunology.* Cham: Springer; 2018. p. 225–341. [https://doi.org/10.1007/978-3-319-76768-0\\_11](https://doi.org/10.1007/978-3-319-76768-0_11).
  45. Zhang D, Jiang S, Ma J, Jiang J, Pan D, Xu X. Molecular cloning, characterization and expression analysis of a clip-domain serine protease from pearl oyster *Pinctada fucata*. *Fish Shellfish Immunol.* 2009;26:662–8. <https://doi.org/10.1016/j.fsi.2009.02.015>.
  46. Allam B, Raftos D. Immune responses to infectious diseases in bivalves. *J Invertebr Pathol.* 2015;131:121–36. <https://doi.org/10.1016/j.jip.2015.05.005>.
  47. Xue Q, Itoh N, Schey KL, Cooper RK, La Peyre JF. Evidence indicating the existence of a novel family of serine protease inhibitors that may be involved in marine invertebrate immunity. *Fish Shellfish Immunol.* 2009;27:250–9. <https://doi.org/10.1016/j.fsi.2009.05.006>.
  48. Kvensakul M. Viral Infection and Apoptosis. *Viruses.* 2017;356. <https://doi.org/10.3390/v9120356>.
  49. Green TJ, Raftos D, Speck P, Montagnani C. Antiviral immunity in marine molluscs. *J Gen Virol.* 2015;96:2471–82. <https://doi.org/10.1099/jgv.0.000244>.
  50. Renault T, Faury N, Barbosa-Solomieu V, Moreau K. Suppression subtractive hybridisation (SSH) and real time PCR reveal differential gene expression in the Pacific cupped oyster, *Crassostrea gigas*, challenged with *Ostreid herpesvirus 1*. *Dev Comp Immunol.* 2011;35:725–35. <https://doi.org/10.1016/j.dci.2011.02.004>.
  51. Bai C-M, Rosani U, Xin L-S, Li G-Y, Li C, Wang Q-C, et al. Dual transcriptomic analysis of *Ostreid herpesvirus 1* infected *Scapharca broughtonii* with an emphasis on viral anti-apoptosis activities and host oxidative bursts. *Fish Shellfish Immunol.* 2018;82:554–64. <https://doi.org/10.1016/j.fsi.2018.08.054>.
  52. Bai C-M, Zhang S-M, Li Y-N, Xin L-S, Rosani U, Wang C-M. Dual transcriptomic analysis reveals a delayed antiviral response of *Haliotis diversicolor supertexta* against *Haliotid Herpesvirus-1*. *Viruses.* 2019;11:383. <https://doi.org/10.3390/v11040383>.
  53. Green TJ, Chataway T, Melwani AR, Raftos DA. Proteomic analysis of hemolymph from poly(I:C)-stimulated *Crassostrea gigas*. *Fish Shellfish Immunol.* 2016;48:39–42. <https://doi.org/10.1016/j.fsi.2015.11.018>.
  54. Zong Y, Liu Z, Wu Z, Han Z, Wang L, Song L. A novel globular C1q domain containing protein (C1qDC-7) from *Crassostrea gigas* acts as pattern recognition receptor with broad recognition spectrum. *Fish Shellfish Immunol.* 2019;84:920–6. <https://doi.org/10.1016/j.fsi.2018.10.079>.
  55. Kawasaki H, Iwamuro S. Potential roles of histones in host defense as antimicrobial agents. *Infect Disord Drug Targets.* 2008;8:195–205.
  56. Poirier AC, Schmitt P, Rosa RD, Vanhove AS, Kieffer-Jaquinod S, Rubio TP, et al. Antimicrobial histones and DNA traps in invertebrate immunity: EVIDENCES IN *CRASSOSTREA GIGAS*. *J Biol Chem.* 2014;289:24821–31. <https://doi.org/10.1074/jbc.M114.576546>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Annexe 2 : Concentrations des ISTD utilisés pour réaliser la gamme de linéarité en mode ESI (+) et ESI (-).

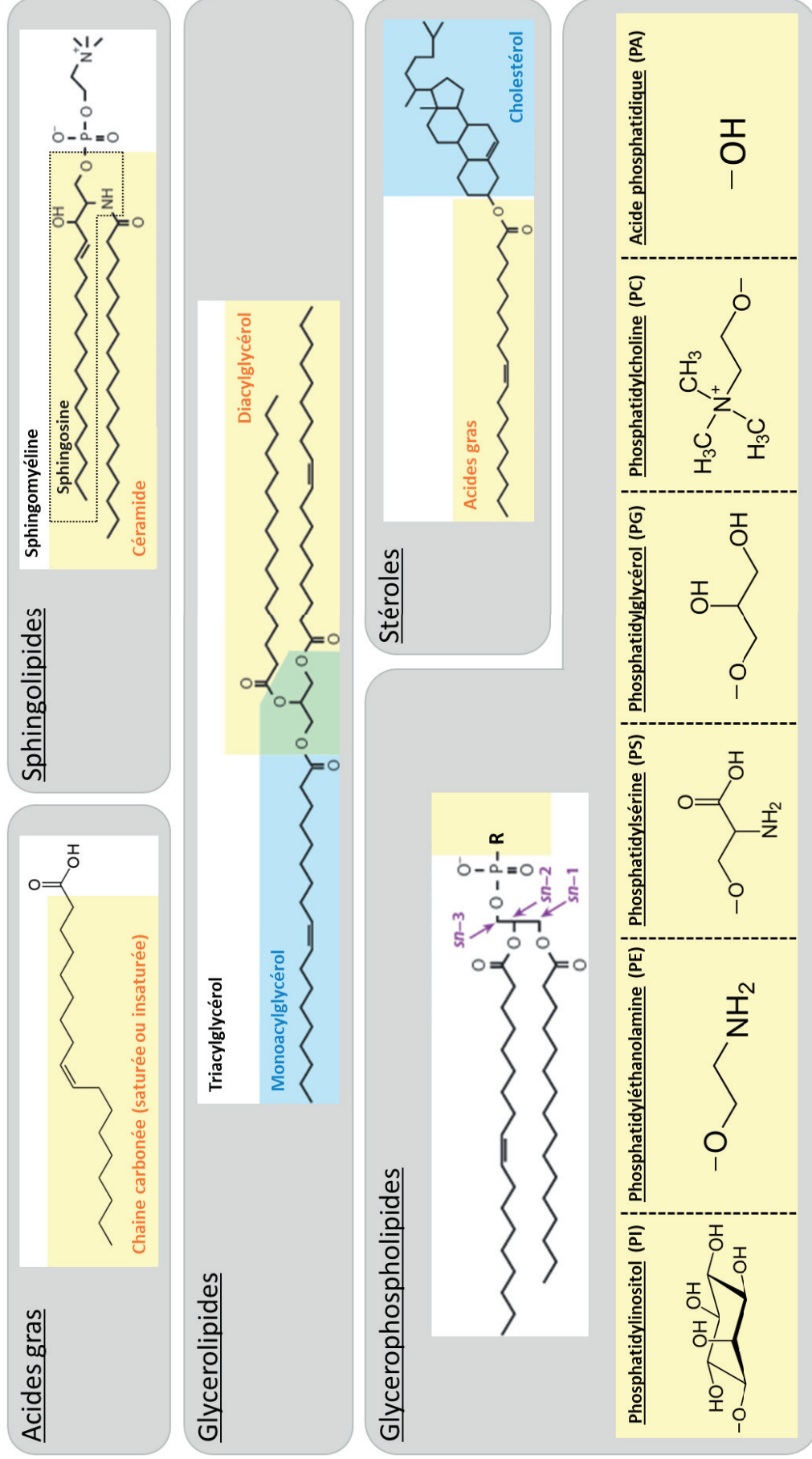
ISTD	point 1 (µg/mL)	point 2 (µg/mL)	point 3 (µg/mL)	point 4 (µg/mL)	point 5 (µg/mL)	point 6 (µg/mL)	point 7 (µg/mL)	point 8 (µg/mL)
LPC 13:0	0,000025	0,00005	0,000125	0,00025	0,000375	0,0005	0,00125	0,0025
TAG 51:1	0,000015	0,00003	0,000075	0,00015	0,000225	0,0003	0,00075	0,0015
SM 36:2	0,000015	0,00003	0,000075	0,00015	0,000225	0,0003	0,00075	0,0015
LPE 17:1	0,000450	0,0009	0,00225	0,0045	0,00675	0,009	0,0225	0,045
PE 25:0	0,000075	0,00015	0,000375	0,00075	0,001125	0,0015	0,00375	0,0075
PS 25:0	0,000150	0,0003	0,00075	0,0015	0,00225	0,003	0,0075	0,015
Cer 30:1	0,000045	0,00009	0,000225	0,00045	0,000675	0,0009	0,00225	0,0045
DG 34:0	0,002000	0,004	0,01	0,02	0,03	0,04	0,1	0,2
PI 31:1	0,001500	0,003	0,0075	0,015	0,0225	0,03	0,075	0,15
PC 22:0	0,000010	0,00002	0,00005	0,0001	0,00015	0,0002	0,0005	0,001
PG 25:0	0,000600	0,0012	0,003	0,006	0,009	0,012	0,03	0,06

ISTD	point 9 (µg/mL)	point 10 (µg/mL)	point 11 (µg/mL)	point 12 (µg/mL)	point 13 (µg/mL)	point 14 (µg/mL)	point 15 (µg/mL)	point 16 (µg/mL)
LPC 13:0	0,00375	0,005	0,0125	0,025	0,0375	0,05	0,0625	0,075
TAG 51:1	0,00225	0,003	0,0075	0,015	0,0225	0,03	0,0375	0,045
SM 36:2	0,00225	0,003	0,0075	0,015	0,0225	0,03	0,0375	0,045
LPE 17:1	0,0675	0,09	0,225	0,45	0,675	0,9	1,125	1,35
PE 25:0	0,01125	0,015	0,0375	0,075	0,1125	0,15	0,1875	0,225
PS 25:0	0,0225	0,03	0,075	0,15	0,225	0,3	0,375	0,45
Cer 30:1	0,00675	0,009	0,0225	0,045	0,0675	0,09	0,1125	0,135
DG 34:0	0,3	0,4	1	2	3	4	5	6
PI 31:1	0,225	0,3	x	x	x	x	x	x
PC 22:0	0,0015	0,002	0,005	0,01	0,015	0,02	0,025	0,03
PG 25:0	0,09	0,12	0,3	0,6	0,9	1,2	1,5	1,8

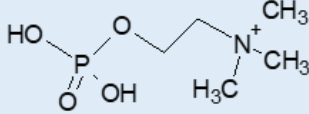

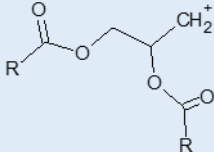

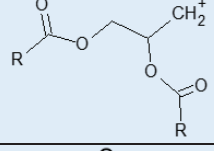

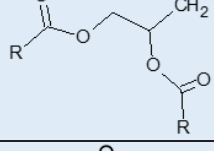

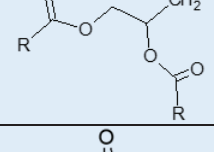

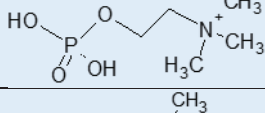
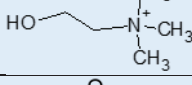

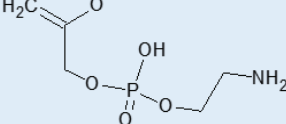
ISTD	point 1 (µg/mL)	point 2 (µg/mL)	point 3 (µg/mL)	point 4 (µg/mL)	point 5 (µg/mL)	point 6 (µg/mL)	point 7 (µg/mL)	point 8 (µg/mL)
LPC 13:0	0,00035	0,0007	0,00175	0,0035	0,00525	0,007	0,0175	0,035
SM 36:2	0,0035	0,007	0,0175	0,035	0,0525	0,07	0,175	0,35
LPE 17:1	0,0002	0,0004	0,001	0,002	0,003	0,004	0,01	0,02
PE 25:0	0,00025	0,0005	0,00125	0,0025	0,00375	0,005	0,0125	0,025
PS 25:0	0,0006	0,0012	0,003	0,006	0,009	0,012	0,03	0,06
Cer 30:1	0,0002	0,0004	0,001	0,002	0,003	0,004	0,01	0,02
PI 31:1	0,0015	0,003	0,0075	0,015	0,0225	0,03	0,075	0,15
PC 22:0	0,00029	0,00058	0,00145	0,0029	0,00435	0,0058	0,0145	0,029
PG 25:0	0,0002	0,0004	0,001	0,002	0,003	0,004	0,01	0,02

ISTD	point 9 (µg/mL)	point 10 (µg/mL)	point 11 (µg/mL)	point 12 (µg/mL)	point 13 (µg/mL)	point 14 (µg/mL)	point 15 (µg/mL)	point 16 (µg/mL)
LPC 13:0	0,0525	0,07	0,175	0,35	0,525	0,7	0,875	1,05
SM 36:2	0,525	0,7	1,75	3,5	5,25	7	8,75	10,5
LPE 17:1	0,03	0,04	0,1	0,2	0,3	0,4	0,5	0,6
PE 25:0	0,0375	0,05	0,125	0,25	0,375	0,5	0,625	0,75
PS 25:0	0,09	0,12	0,3	0,6	0,9	1,2	1,5	1,8
Cer 30:1	0,03	0,04	0,1	0,2	0,3	0,4	0,5	0,6
PI 31:1	0,225	0,3	0,75	1,5	2,25	3	3,75	4,5
PC 22:0	0,0435	0,058	0,145	0,29	0,435	0,58	0,725	0,87
PG 25:0	0,03	0,04	0,1	0,2	0,3	0,4	0,5	0,6

Annexe 3 : Structures des lipides.



Annexe 4 : Masse en Q1, Q3, DP et CE en fonction du mode d'ionisation pour les transitions MRM au sein de la méthode Scout-MRM.

Lipide	Précurseur	Type de fragment	Structure du fragment	m/z du fragment	DP (eV)	CE (eV)
PC	[M+H <sup>+</sup> ]	Tête polaire		184	191	39
	[M+HCOO <sup>-</sup> ]	Acide gras		Variable	-175	48
PS	[M+H <sup>+</sup> ]	Perte de la tête polaire		M-184	46	27
	[M-H <sup>+</sup> ]	Acide gras		Variable	-65	-53
PE	[M+H <sup>+</sup> ]	Perte de la tête polaire		M-140	91	29
	[M-H <sup>+</sup> ]	Acide gras		Variable	-115	-44
PI	[M+H <sup>+</sup> ]	Perte de la tête polaire		M-259	40	33
	[M-H <sup>+</sup> ]	Acide gras		Variable	-55	-59
PG	[M+H <sup>+</sup> ]	Perte de la tête polaire		M-171	181	23
	[M-H <sup>+</sup> ]	Acide gras		Variable	-20	-43
LPC	[M+H <sup>+</sup> ]	Tête polaire		184	96	33
		Fragmentation de la tête polaire		104	96	29
	[M-H <sup>+</sup> ]	Acide gras		Variable	-50	-38
		Perte de la tête polaire		224	-50	-38



LPE	[M+H <sup>+</sup> ]	Perte de la tête polaire		M-140	56	17
		Perte de OH		M-17	56	25
	[M-H <sup>+</sup> ]	Acide gras		Variable	-30	-36
		Perte de la tête polaire		196	-30	-36
SM	[M+H <sup>+</sup> ]	Tête polaire		184	116	33
	[M+HCOO <sup>-</sup> ]	Perte de CH <sub>3</sub>		M-17	-100	-60
		Fragmentation de la tête polaire		168	-100	-36
DG	[M+NH <sub>4</sub> <sup>+</sup> ]	Perte d'un acide gras		Variable	116	16
		Perte de OH		M-17	?	?
TG	[M+NH <sub>4</sub> <sup>+</sup> ]	Perte d'un acide gras		Variable	50	36
CER d18:1	[M+H <sup>+</sup> ]	Perte de l'acide gras et de 2 OH		264	46	37
		Perte de OH		M-17	46	15
	[M-H <sup>+</sup> ]	Perte du squelette sphingosine		M-257,4	-240	-38

## Annexe 5 : Base de données en temps de rétention pour l'élaboration de méthode Scout-MRM en lipidomique.

Composés scouts	RT Lipide V1.1 (min)
SCOUT PC 9:0/9:0	7,59
SCOUT PC 12:0/13:0	13,50
SCOUT PC 14:0/14:0 d4	15,22
SCOUT PC 15:0/15:0	16,19
SCOUT Cer d18:1/17:0	17,68
SCOUT PC 19:0/19:0	19,19
SCOUT TG 14:0/16:1/14:0 d5	20,94
SCOUT TG 15:0/18:1/15:0 d5	21,71
SCOUT TG 16:0/18:0/16:0 d5	22,33

ISTD lipidique	Groupe scout ESI (+)	Groupe scout ESI (-)	RT Lipide V1.1 (min)
PI 17:0/14:1	Groupe 3 (PC 25:0)	Groupe 3 (PC 25:0)	15,03
PE 12:0/13:0	Groupe 3 (PC 25:0)	Groupe 3 (PC 25:0)	13,74
PS 12:0/13:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	12,56
PC 11:0/11:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	11,36
1-LPC 13:0	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	3,21
2-LPC 13:0	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	3,75
1-LPE 17:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,99
2-LPE 17:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	7,48
SM d18:1:18:1 d9	Groupe 5 (PC 30:0)	Groupe 5 (PC 30:0)	16,34
Cer d18:1:12:0	Groupe 4 (PC 28:0)	x	15,39
DG 17:0/17:0 d5	Groupe 7 (PC 38:0)	x	19,27
TG 17:0/17:1/17:0 d5	Groupe 9 (TG 48:1)	x	22,18

Lipides	Composition AG	Groupe scout ESI (+) V0		Groupe scout ESI (-) V0		Moyenne RT isomères V1.1 (min)	RT lipides V1.1 (min)
PC 26:0	PC 12:0_14:0	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,15	14,15
PC 28:0	PC 14:0_14:0	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,21	15,21
PC 30:0	PC 14:0_16:0			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,18	16,18
PC 30:1	PC 14:1_16:0	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,44	15,46
	PC 14:0_16:1			Groupe 4 (PC 28:0)			15,41
PC 32:0	PC 16:0_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,05	17,05
	PC 14:0_18:0			Groupe 5 (PC 30:0)			17,05
PC 32:1	PC 16:0_16:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,35	16,36
	PC 14:0_18:1			Groupe 5 (PC 30:0)			16,34
PC 32:2	PC 16:1_16:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,60	15,55
	PC 14:0_18:2			Groupe 4 (PC 28:0)			15,64
PC 32:3	PC 14:0_18:3	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		15,01	15,01
PC 34:0	PC 16:0_18:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,83	17,83
PC 34:1	PC 16:0_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,20	17,18
	PC 16:1_18:0			Groupe 5 (PC 30:0)			17,22
PC 34:2	PC 16:1_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,52	16,47
	PC 16:0_18:2			Groupe 5 (PC 30:0)			16,57
PC 34:3	PC 16:0_18:3	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	15,89	16,00
	PC 16:1_18:2			Groupe 4 (PC 28:0)			15,81
	PC 16:2_18:1			Groupe 4 (PC 28:0)			15,88
PC 34:4	PC 16:1_18:3	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,31	15,18
	PC 16:0_18:4			Groupe 4 (PC 28:0)			15,56
	PC 16:2_18:2			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,14
	PC 16:3_18:1			Groupe 4 (PC 28:0)			15,34
PC 34:5	PC 16:1_18:4	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,84	14,70
	PC 14:0_20:5			Groupe 3 (PC 25:0)			14,97
PC 36:0	PC 18:0_18:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,51	18,51
PC 36:1	PC 18:0_18:1	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,92	17,94
	PC 16:0_20:1			Groupe 6 (Cer 35:1)			17,86
	PC 16:1_20:0			Groupe 6 (Cer 35:1)			17,97

PC 36:2	PC 18:1_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,31	17,28
	PC 18:0_18:2			Groupe 5 (PC 30:0)			17,40
	PC 16:1_20:1			Groupe 5 (PC 30:0)			17,24
	PC 16:0_20:2			Groupe 5 (PC 30:0)			17,31
PC 36:3	PC 18:0_18:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,76	16,89
	PC 18:1_18:2			Groupe 5 (PC 30:0)			16,69
	PC 16:0_20:3			Groupe 5 (PC 30:0)			16,84
	PC 16:1_20:2			Groupe 5 (PC 30:0)			16,61
PC 36:4	PC 16:1_20:3	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,19	16,08
	PC 18:1_18:3			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,15
	PC 16:0_20:4			Groupe 5 (PC 30:0)			16,48
	PC 18:2_18:2			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,05
PC 36:5	PC 16:0_20:5	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,70	15,94
	PC 18:2_18:3			Groupe 4 (PC 28:0)			15,43
	PC 16:1_20:4			Groupe 4 (PC 28:0)			15,72
PC 36:6	PC 18:3_18:3	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,76	14,76
PC 38:2	PC 18:2_20:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,06	18,14
	PC 18:1_20:1			Groupe 6 (Cer 35:1)			17,99
	PC 18:0_20:2			Groupe 6 (Cer 35:1)			18,06
	PC 16:0_22:2			Groupe 6 (Cer 35:1)			18,07
PC 38:4	PC 18:1_20:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,05	16,93
	PC 18:0_20:4			Groupe 5 (PC 30:0)			17,32
	PC 18:2_20:2			Groupe 5 (PC 30:0)			16,87
	PC 16:0_22:4			Groupe 5 (PC 30:0)			17,09
PC 38:5	PC 18:0_20:5	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,58	16,81
	PC 18:1_20:4			Groupe 5 (PC 30:0)			16,63
	PC 18:2_20:3			Groupe 5 (PC 30:0)			16,32
	PC 16:0_22:5			Groupe 5 (PC 30:0)			16,57
PC 38:6	PC 18:1_20:5	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,00	16,06
	PC 18:2_20:4			Groupe 4 (PC 28:0)			15,95
PC 38:7	PC 18:2_20:5	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,39	15,39
PC 40:3	PC 18:2_22:1	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,12	18,12
PC 40:4	PC 18:0_22:4	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,86	17,86
PC 40:5	PC 18:0_22:5	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,40	17,40
PC 40:6	PC 18:0_22:6	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,91	17,16
	PC 20:1_20:5			Groupe 5 (PC 30:0)			16,88
	PC 18:1_22:5			Groupe 5 (PC 30:0)			16,69
PC 40:7	PC 20:3_20:4	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,22	16,21
	PC 18:1_22:6			Groupe 5 (PC 30:0)			16,42
	PC 18:2_22:5			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,04
PC 40:8	PC 20:3_20:5	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,61	15,63
	PC 18:3_22:5			Groupe 4 (PC 28:0)			15,43
	PC 18:2_22:6			Groupe 4 (PC 28:0)			15,78

Lipides	Composition AG	Groupe scout ESI (+) V0		Groupe scout ESI (-) V0		Moyenne RT isomères V1.1 (min)	RT lipides V1.1 (min)
PE 26:0	PE 12:0_14:0	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,40	14,40
PE 28:0	PE 14:0_14:0	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,48	15,48
PE 30:0	PE 14:0_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,44	16,44
PE 30:1	PE 14:0_16:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,69	15,65
	PE 14:1_16:0			Groupe 4 (PC 28:0)			15,74
PE 30:2	PE 14:0_16:2	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,01	15,01
PE 32:0	PE 14:0_18:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,27	17,27
	PE 16:0_16:0			Groupe 5 (PC 30:0)			17,28
PE 32:1	PE 16:1_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,54	16,57
	PE 18:1_14:0			Groupe 5 (PC 30:0)			16,50
PE 32:2	PE 16:0_16:2	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	15,89	16,04
	PE 16:1_16:1			Groupe 4 (PC 28:0)			15,79
	PE 14:1_18:1			Groupe 4 (PC 28:0)			15,84

	PE 14:0_18:2			Groupe 4 (PC 28:0)			15,90
PE 32:3	PE 16:1_16:2	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,23	15,20
	PE 14:0_18:3			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,26
PE 34:0	PE 18:0_16:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,03	18,03
PE 34:1	PE 18:1_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,39	17,37
	PE 18:0_16:1			Groupe 5 (PC 30:0)			17,40
PE 34:2	PE 14:0_20:2	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,71	16,68
	PE 18:1_16:1			Groupe 5 (PC 30:0)			16,66
	PE 18:2_16:0			Groupe 5 (PC 30:0)			16,79
PE 34:3	PE 16:2_18:1	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,14	16,14
	PE 16:1_18:2			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,03
	PE 18:3_16:0			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,25
PE 34:4	PE 16:2_18:2	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,56	15,44
	PE 16:1_18:3			Groupe 4 (PC 28:0)			15,42
	PE 16:0_18:4			Groupe 4 (PC 28:0)			15,81
PE 36:1	PE 16:1_20:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,11	18,14
	PE 16:0_20:1			Groupe 6 (Cer 35:1)			18,08
	PE 18:1_18:0			Groupe 6 (Cer 35:1)			18,12
PE 36:2	PE 16:0_20:2	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,50	17,51
	PE 16:1_20:1			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,43
	PE 18:1_18:1			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,47
	PE 18:0_18:2			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,60
PE 36:3	PE 16:1_20:2	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,96	16,81
	PE 18:3_18:0			Groupe 5 (PC 30:0)			17,09
	PE 18:2_18:1			Groupe 5 (PC 30:0)			16,90
	PE 20:3_16:0			Groupe 5 (PC 30:0)			17,04
PE 36:4	PE 18:2_18:2	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,44	16,26
	PE 18:1_18:3			Groupe 5 (PC 30:0)			16,36
	PE 20:4_16:0			Groupe 5 (PC 30:0)			16,72
PE 36:5	PE 18:2_18:3	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)		15,92	15,66
	PE 18:1_18:4			Groupe 4 (PC 28:0)			15,91
	PE 16:1_20:4			Groupe 4 (PC 28:0)			15,97
	PE 20:5_16:0			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,15
PE 36:6	PE 18:3_18:3	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,17	15,00
	PE 16:1_20:5			Groupe 4 (PC 28:0)			15,34
PE 38:1	PE 18:1_20:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,75	18,75
PE 38:2	PE 18:2_20:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,22	18,30
	PE 18:1_20:1			Groupe 6 (Cer 35:1)			18,15
PE 38:3	PE 18:3_20:0	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	Groupe 6 (Cer 35:1)		17,71	17,83
	PE 18:1_20:2			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,58
	PE 18:0_20:3			Groupe 6 (Cer 35:1)			17,80
	PE 18:2_20:1			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,63
PE 38:4	PE 18:3_20:1	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	Groupe 5 (PC 30:0)		17,23	17,12
	PE 18:2_20:2			Groupe 5 (PC 30:0)			17,03
	PE 20:3_18:1			Groupe 5 (PC 30:0)			17,16
	PE 20:4_18:0			Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)		17,52
	PE 16:0_22:4			Groupe 5 (PC 30:0)			17,31
PE 38:5	PE 18:2_20:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,71	16,48
	PE 18:3_20:2			Groupe 5 (PC 30:0)			16,48
	PE 16:0_22:5			Groupe 5 (PC 30:0)			16,78
	PE 20:5_18:0			Groupe 5 (PC 30:0)			17,02
	PE 20:4_18:1			Groupe 5 (PC 30:0)			16,80
PE 38:6	PE 18:1_20:5	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,22	16,26
	PE 18:3_20:3			Groupe 4 (PC 28:0)			15,92
	PE 16:0_22:6			Groupe 5 (PC 30:0)			16,52
	PE 20:4_18:2			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,16
PE 38:7	PE 18:2_20:5	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,56	15,57
	PE 18:3_20:4			Groupe 4 (PC 28:0)			15,55
PE 38:8	PE 18:3_20:5	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,93	14,93

PE 40:4	PE 18:0_22:4	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,03	18,03
PE 40:5	PE 18:0_22:5	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,55	17,55
PE 40:6	PE 18:1_22:5	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,19	17,16
	PE 20:1_20:5			Groupe 5 (PC 30:0)			17,06
	PE 18:0_22:6			Groupe 5 (PC 30:0)			17,35
PE 40:7	PE 20:2_20:5	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,54	16,43
	PE 18:1_22:6			Groupe 5 (PC 30:0)			16,64
PE 40:8	PE 18:2_22:6	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	15,91	15,99
	PE 20:3_20:5			Groupe 4 (PC 28:0)			15,83
PE 40:9	PE 18:3_22:6	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,42	15,39
	PE 20:4_20:5			Groupe 4 (PC 28:0)			15,45

Lipides	Composition AG	Groupe scout ESI (+) V0		Groupe scout ESI (-) V0		Moyenne RT isomères V1.1 (min)	RT lipides V1.1 (min)
PI 32:1	PI 16:1_16:0	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,42	15,45
	PI 18:1_14:0			Groupe 4 (PC 28:0)			15,39
PI 34:0	PI 18:0_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,99	16,99
PI 34:1	PI 18:1_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,33	16,32
	PI 18:0_16:1			Groupe 5 (PC 30:0)			16,35
PI 34:2	PI 18:1_16:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,64	15,58
	PI 18:2_16:0			Groupe 4 (PC 28:0)			15,69
PI 36:0	PI 18:0_18:0	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,73	17,73
PI 36:1	PI 18:1_18:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,12	17,12
PI 36:2	PI 18:1_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,50	16,44
	PI 18:0_18:2			Groupe 5 (PC 30:0)			16,56
PI 36:3	PI 18:1_18:2	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)		15,90	15,81
	PI 20:3_16:0			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		15,99
PI 36:4	PI 18:1_18:3	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 4 (PC 28:0)		15,36	15,26
	PI 18:2_18:2			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,18
	PI 20:4_16:0			Groupe 4 (PC 28:0)			15,64
PI 36:5	PI 20:5_16:0	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,01	15,10
	PI 20:4_16:1			Groupe 3 (PC 25:0)			14,92
PI 38:2	PI 18:0_20:2	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,25	17,25
PI 38:3	PI 18:0_20:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,80	16,80
PI 38:4	PI 20:3_18:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		16,18	15,75
	PI 20:4_18:0			Groupe 5 (PC 30:0)			16,51
	PI 16:0_22:4			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,28
PI 38:5	PI 20:5_18:0	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,81	15,96
	PI 20:4_18:1			Groupe 4 (PC 28:0)			15,74
	PI 16:0_22:5			Groupe 4 (PC 28:0)			15,72
PI 38:6	PI 20:5_18:1	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,27	15,20
	PI 20:4_18:2			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,12
	PI 22:6_16:0			Groupe 4 (PC 28:0)			15,47
PI 38:7	PI 20:5_18:2	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,54	14,56
	PI 20:4_18:3			Groupe 3 (PC 25:0)			14,53
PI 40:4	PI 18:0_22:4	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,07	17,07
PI 40:6	PI 22:5_18:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		16,05	15,81
	PI 18:0_22:6			Groupe 5 (PC 30:0)			16,34
	PI 20:1_20:5			Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)		16,00
PI 40:7	PI 22:6_18:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,47	15,57
	PI 20:2_20:5			Groupe 4 (PC 28:0)			15,38

Lipides	Composition AG	Groupe scout ESI (+) V0		Groupe scout ESI (-) V0		Moyenne RT isomères V1.1 (min)	RT lipides V1.1 (min)
PS 30:1	PS 14:0_16:1	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,59	14,59
PS 32:0	PS 16:0_16:0	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,33	16,33
PS 32:1	PS 16:0_16:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,58	15,59
	PS 14:0_18:1			Groupe 4 (PC 28:0)			15,57
PS 32:2	PS 16:1_16:1	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,81	14,81

PS 34:1	PS 16:0_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,42	16,42
PS 34:2	PS 16:0_18:2	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,75	15,80
	PS 16:1_18:1			Groupe 4 (PC 28:0)			15,71
PS 34:3	PS 16:0_18:3	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,13	15,23
	PS 16:1_18:2			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,03
PS 36:1	PS 18:0_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,23	17,20
	PS 16:1_20:0			Groupe 5 (PC 30:0)			17,26
PS 36:2	PS 18:1_18:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,61	16,55
	PS 18:0_18:2			Groupe 5 (PC 30:0)			16,67
PS 36:3	PS 18:0_18:3	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,03	16,13
	PS 18:1_18:2			Groupe 4 (PC 28:0)			15,93
PS 36:4	PS 18:1_18:3	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,30	15,34
	PS 18:2_18:2			Groupe 4 (PC 28:0)			15,25
PS 36:5	PS 18:2_18:3	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)		14,91	14,66
	PS 20:5_16:0			Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)		15,16
PS 36:6	PS 18:3_18:3	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,02	14,02
PS 38:1	PS 18:1_20:0	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,92	17,92
PS 38:5	PS 20:5_18:0	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	15,98	16,09
	PS 20:4_18:1			Groupe 4 (PC 28:0)			15,86
PS 38:6	PS 20:5_18:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,25	15,30
	PS 20:4_18:2			Groupe 4 (PC 28:0)			15,19
PS 40:6	PS 20:1_20:5	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,13	16,13
PS 40:7	PS 22:6_18:1	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,58	15,68
	PS 20:2_20:5			Groupe 4 (PC 28:0)			15,48
PS 40:8	PS 20:3_20:5	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,85	14,85
PS 40:9	PS 20:4_20:5	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,44	14,44
PS 40:10	PS 20:5_20:5	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		13,81	13,81

Lipides	Groupe scout ESI (+) V0	Groupe scout ESI (-) V0	RT lipides V1.1 (min)
1-LPC 14:0	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,38
2-LPC 14:0	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,97
1-LPC 16:0	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	7,25
2-LPC 16:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,79
1-LPC 16:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,27
2-LPC 16:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,80
1-LPC 18:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	9,68
2-LPC 18:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	10,11
1-LPC 18:1	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,93
2-LPC 18:1	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,39
1-LPC 18:2	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,26
2-LPC 18:2	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,75
1-LPC 18:3	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,74
2-LPC 18:3	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,26
1-LPC 20:2	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,58
2-LPC 20:2	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	9,01
1-LPC 20:3	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	7,25
2-LPC 20:3	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,69
1-LPC 20:4	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,29
2-LPC 20:4	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,71
1-LPC 20:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,87
2-LPC 20:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,30
1-LPC 22:4	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,23
2-LPC 22:4	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,59
1-LPC 22:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,90
2-LPC 22:5	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,30
1-LPC 22:6	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,20
2-LPC 22:6	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,59



Lipides	Groupe scout ESI (+) V0	Groupe scout ESI (-) V0	RT lipides V1.1 (min)
1-LPE 16:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,43
2-LPE 16:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,96
1-LPE 16:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,34
2-LPE 16:1	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,87
1-LPE 18:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	9,92
2-LPE 18:0	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	10,34
1-LPE 18:1	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,13
2-LPE 18:1	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	8,56
1-LPE 18:2	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,43
2-LPE 18:2	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,88
1-LPE 18:3	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,77
2-LPE 18:3	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,28
1-LPE 20:4	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,45
2-LPE 20:4	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,85
1-LPE 20:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	4,88
2-LPE 20:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	5,35
1-LPE 22:5	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	7,07
2-LPE 22:5	Groupe 2 (PC 18:0)	Groupe 2 (PC 18:0)	7,45
1-LPE 22:6	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,34
2-LPE 22:6	Groupe 1 (T=0 min)	Groupe 1 (T=0 min)	6,73

Lipides	Groupe scout ESI (+) V0		Groupe scout ESI (-) V0		RT lipides V1.1 (min)
SM 32:1	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	Groupe 3 (PC 25:0)	Groupe 4 (PC 28:0)	15,06
SM 32:2	Groupe 3 (PC 25:0)		Groupe 3 (PC 25:0)		14,13
SM 34:1	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,09
SM 34:2	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,28
SM 36:1	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,03
SM 36:2	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,30
SM 36:3	Groupe 4 (PC 28:0)		Groupe 4 (PC 28:0)		15,55
SM 38:1	Groupe 6 (Cer 35:1)	Groupe 7 (PC 38:0)	Groupe 6 (Cer 35:1)		17,82
SM 38:2	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,18
SM 38:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		16,37
SM 40:1	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,53
SM 40:2	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,90
SM 40:3	Groupe 5 (PC 30:0)		Groupe 5 (PC 30:0)		17,20
SM 42:1	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		19,16
SM 42:2	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		18,59
SM 42:3	Groupe 6 (Cer 35:1)		Groupe 6 (Cer 35:1)		17,95

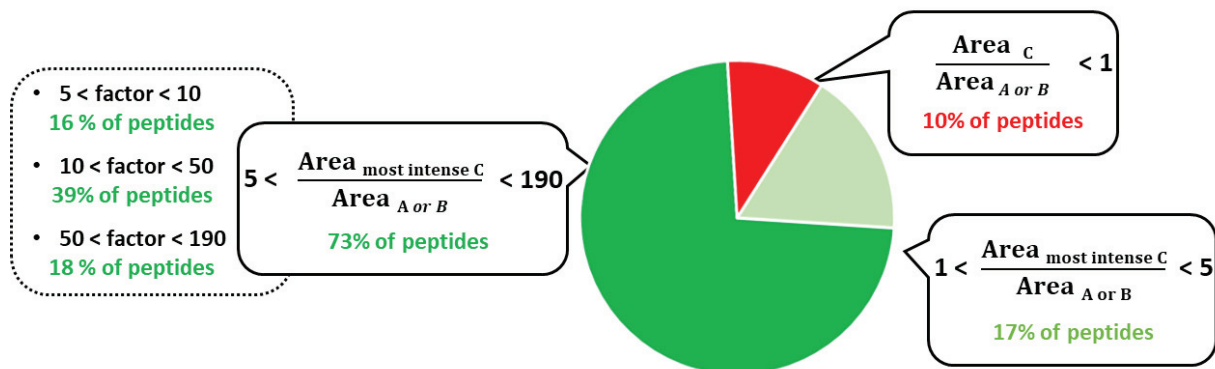
Lipides	Composition AG	Groupe scout ESI (+) V0		Moyenne RT isomères V1.1 (min)	RT lipides V1.1 (min)
DG 38:6	DG 18:1_20:5	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,61	17,61
DG 26:0	DG 12:0_14:0	Groupe 4 (PC 28:0)	Groupe 5 (PC 30:0)	16,13	16,13
DG 28:0	DG 14:0_14:0	Groupe 5 (PC 30:0)		17,03	17,03
DG 28:1	DG 14:0_14:1	Groupe 5 (PC 30:0)		16,32	16,32
DG 28:2	DG 14:0_14:2	Groupe 4 (PC 28:0)		15,86	15,86
DG 30:0	DG 14:0_16:0	Groupe 6 (Cer 35:1)		17,84	17,84
DG 30:1	DG 14:0_16:1	Groupe 5 (PC 30:0)		17,18	17,16
	DG 14:1_16:0				17,21
DG 30:2	DG 14:1_16:1	Groupe 5 (PC 30:0)		16,43	16,43
DG 32:0	DG 16:0_16:0	Groupe 6 (Cer 35:1)		18,58	18,57
	DG 14:0_18:0				18,59
DG 32:1	DG 16:0_16:1	Groupe 6 (Cer 35:1)		17,94	17,96
	DG 18:1_14:0				17,93
DG 32:2	DG 16:1_16:1	Groupe 5 (PC 30:0)		17,30	17,26
	DG 18:2_14:0				17,33
	DG 14:1_18:1				17,31
DG 32:3	DG 16:1_16:2	Groupe 5 (PC 30:0)		16,73	16,72
	DG 14:1_18:2				16,69

	DG 14:0_18:3			16,78
DG 34:0	DG 18:0_16:0	Groupe 7 (PC 38:0)		19,21
DG 34:1	DG 18:1_16:0	Groupe 6 (Cer 35:1)		18,66
	DG 16:1_18:0			18,67
DG 34:2	DG 18:2_16:0	Groupe 6 (Cer 35:1)		18,13
	DG 16:1_18:1			18,04
DG 34:3	DG 16:0_18:3	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,62
	DG 16:1_18:2			17,46
DG 34:4	DG 16:1_18:3	Groupe 5 (PC 30:0)		16,90
	DG 16:2_18:2			16,90
DG 36:0	DG 18:0_18:0	Groupe 7 (PC 38:0)		19,78
DG 36:1	DG 18:1_18:0	Groupe 7 (PC 38:0)		19,28
DG 36:2	DG 18:1_18:1	Groupe 6 (Cer 35:1)		18,71
	DG 18:0_18:2			18,81
DG 36:3	DG 18:0_18:3	Groupe 6 (Cer 35:1)		18,37
	DG 18:2_18:1			18,21
DG 36:4	DG 18:1_18:3	Groupe 5 (PC 30:0)	Groupe 6 (Cer 35:1)	17,74
	DG 18:2_18:2			17,63
DG 36:5	DG 18:2_18:3	Groupe 5 (PC 30:0)		17,09
DG 36:6	DG 18:3_18:3	Groupe 5 (PC 30:0)		16,55

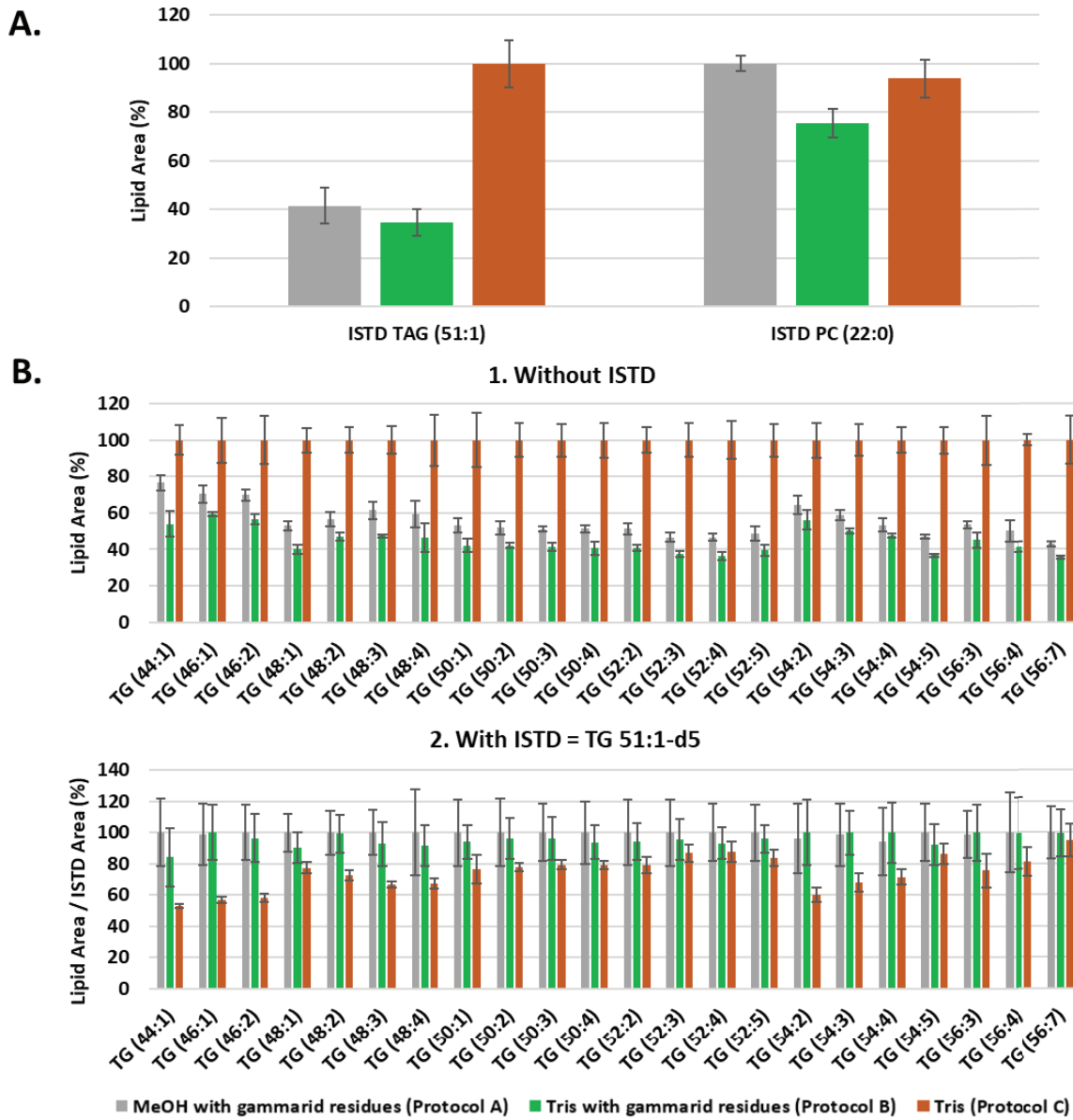
Lipide	Groupe scout ESI (+) V0	RT lipides V1.1 (min)
TG 40:0	Groupe 7 (PC 38:0)	20,51
TG 40:1	Groupe 7 (PC 38:0)	20,06
TG 40:2	Groupe 7 (PC 38:0)	19,70
TG 42:0	Groupe 7 (PC 38:0)   Groupe 8 (TG 44:1)	20,94
TG 42:1	Groupe 7 (PC 38:0)	20,54
TG 42:2	Groupe 7 (PC 38:0)	20,13
TG 42:3	Groupe 7 (PC 38:0)	19,86
TG 44:0	Groupe 8 (TG 44:1)	21,34
TG 44:1	Groupe 7 (PC 38:0)   Groupe 8 (TG 44:1)	20,94
TG 44:2	Groupe 7 (PC 38:0)	20,57
TG 44:3	Groupe 7 (PC 38:0)	20,22
TG 46:0	Groupe 8 (TG 44:1)   Groupe 9 (TG 48:1)	21,69
TG 46:1	Groupe 8 (TG 44:1)	21,34
TG 46:2	Groupe 7 (PC 38:0)   Groupe 8 (TG 44:1)	20,98
TG 46:3	Groupe 7 (PC 38:0)	20,65
TG 46:4	Groupe 7 (PC 38:0)	20,33
TG 48:0	Groupe 9 (TG 48:1)	21,99
TG 48:1	Groupe 8 (TG 44:1)   Groupe 9 (TG 48:1)	21,69
TG 48:2	Groupe 8 (TG 44:1)	21,37
TG 48:3	Groupe 8 (TG 44:1)	21,05
TG 48:4	Groupe 7 (PC 38:0)   Groupe 8 (TG 44:1)	20,74
TG 48:5	Groupe 7 (PC 38:0)	20,47
TG 49:1	Groupe 9 (TG 48:1)	21,85
TG 50:0	Groupe 9 (TG 48:1)   Groupe 10 (TG 50:0)	22,30
TG 50:1	Groupe 9 (TG 48:1)	22,01
TG 50:2	Groupe 8 (TG 44:1)   Groupe 9 (TG 48:1)	21,72
TG 50:3	Groupe 8 (TG 44:1)	21,41
TG 50:4	Groupe 8 (TG 44:1)	21,11
TG 50:5	Groupe 7 (PC 38:0)   Groupe 8 (TG 44:1)	20,79
TG 50:6	Groupe 7 (PC 38:0)	20,50
TG 50:9	Groupe 7 (PC 38:0)	19,57
TG 51:0	Groupe 10 (TG 50:0)	22,43
TG 51:1	Groupe 9 (TG 48:1)	22,15
TG 51:2	Groupe 9 (TG 48:1)	21,88
TG 52:1	Groupe 9 (TG 48:1)   Groupe 10 (TG 50:0)	22,31
TG 52:2	Groupe 9 (TG 48:1)	22,03
TG 52:3	Groupe 8 (TG 44:1)   Groupe 9 (TG 48:1)	21,76

TG 52:4	Groupe 8 (TG 44:1)		21,46
TG 52:5	Groupe 8 (TG 44:1)		21,16
TG 52:6	Groupe 7 (PC 38:0)	Groupe 8 (TG 44:1)	20,88
TG 53:2	Groupe 9 (TG 48:1)	Groupe 10 (TG 50:0)	22,16
TG 54:1	Groupe 10 (TG 50:0)		22,60
TG 54:2	Groupe 9 (TG 48:1)	Groupe 10 (TG 50:0)	22,33
TG 54:3	Groupe 9 (TG 48:1)		22,05
TG 54:4	Groupe 9 (TG 48:1)		21,78
TG 54:5	Groupe 8 (TG 44:1)		21,49
TG 54:6	Groupe 8 (TG 44:1)		21,20
TG 54:7	Groupe 7 (PC 38:0)	Groupe 8 (TG 44:1)	20,97
TG 56:1	Groupe 10 (TG 50:0)		22,84
TG 56:2	Groupe 10 (TG 50:0)		22,60
TG 56:3	Groupe 9 (TG 48:1)	Groupe 10 (TG 50:0)	22,32
TG 56:4	Groupe 9 (TG 48:1)		22,08
TG 56:7	Groupe 8 (TG 44:1)		21,46
TG 58:1	Groupe 10 (TG 50:0)		23,06
TG 58:2	Groupe 10 (TG 50:0)		22,81
TG 58:3	Groupe 10 (TG 50:0)		22,58
TG 58:4	Groupe 9 (TG 48:1)	Groupe 10 (TG 50:0)	22,35
TG 60:2	Groupe 10 (TG 50:0)		23,06
TG 60:3	Groupe 10 (TG 50:0)		22,82
TG 60:4	Groupe 10 (TG 50:0)		22,59
TG 60:5	Groupe 9 (TG 48:1)	Groupe 10 (TG 50:0)	22,40

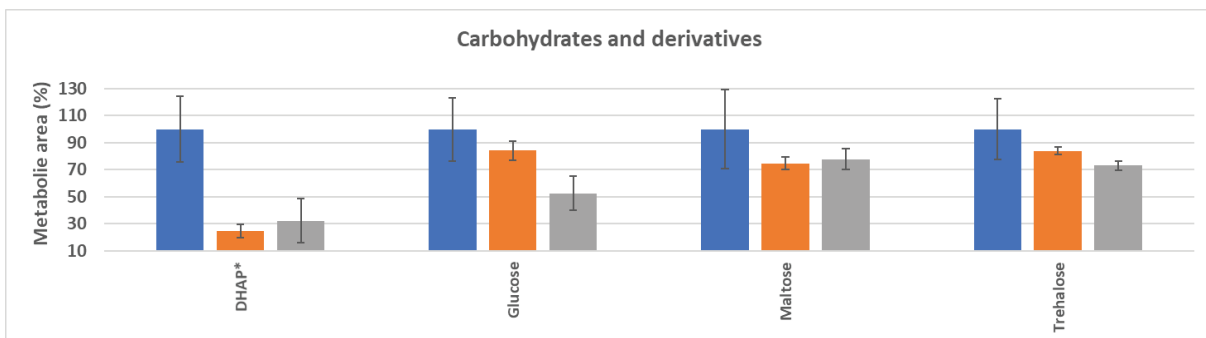
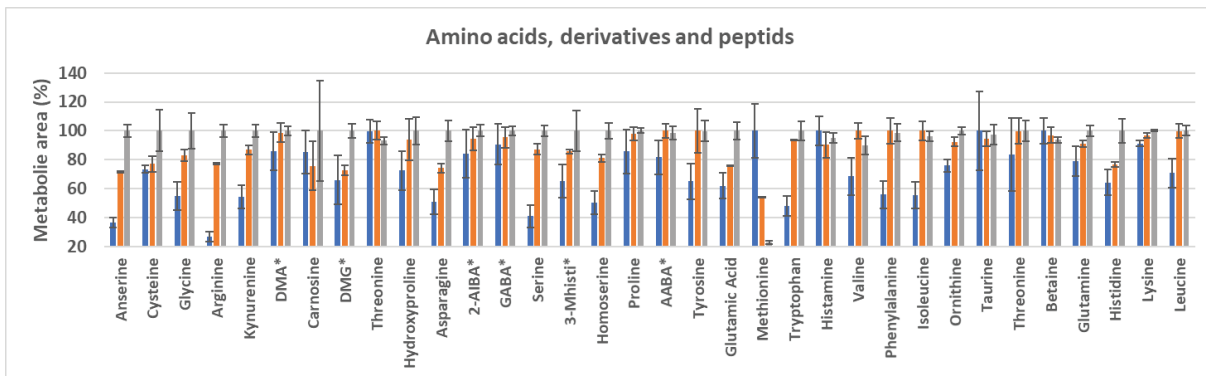
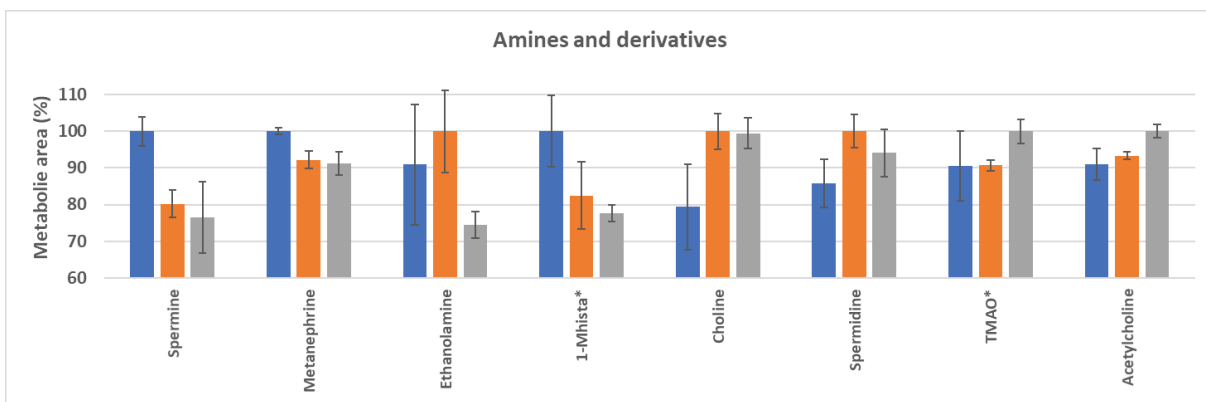
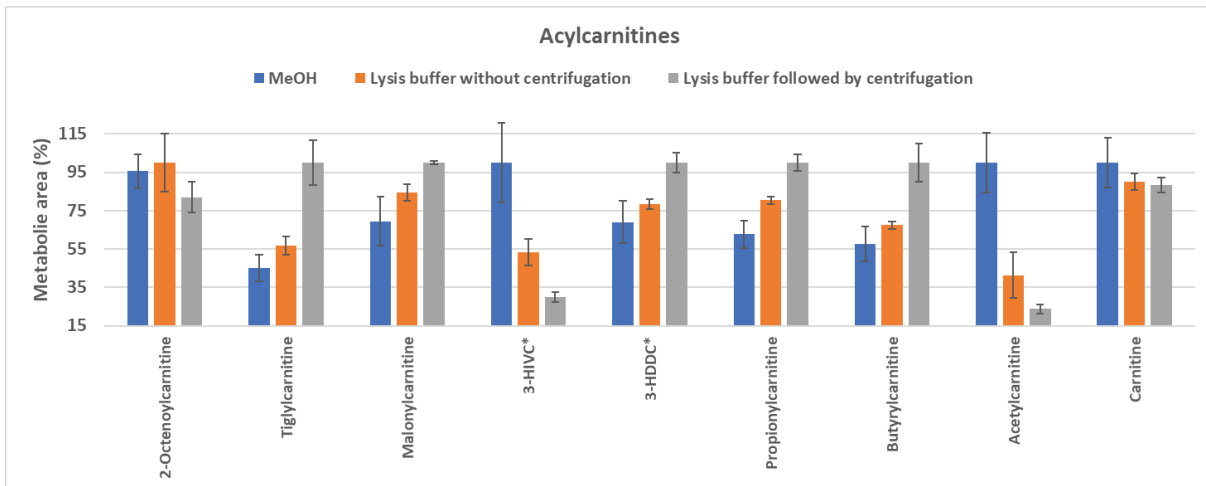
**Annexe 6 : Evaluation of the protein extraction protocols on the peptide's abundance.** Comparison of area ratios between the different conditions (methanol, lysis buffer with and without centrifugation). The experiments were performed in triplicate. For more readability, the different conditions have been annotated and shortened as follows: A- methanol, B- lysis buffer without centrifugation and C- lysis buffer with centrifugation.

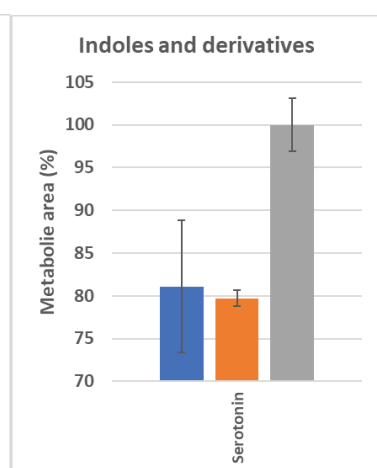
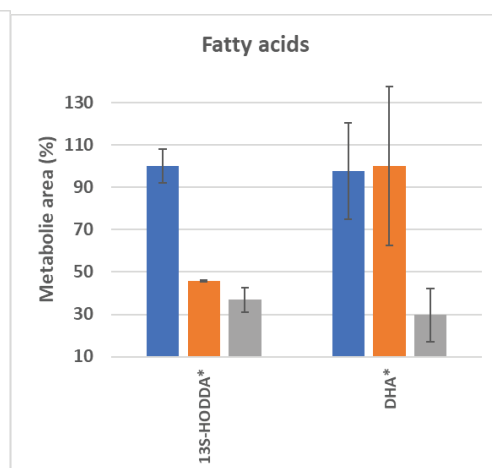
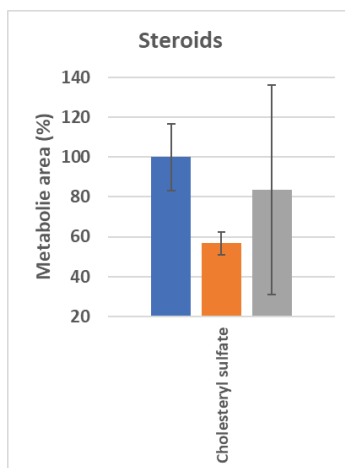
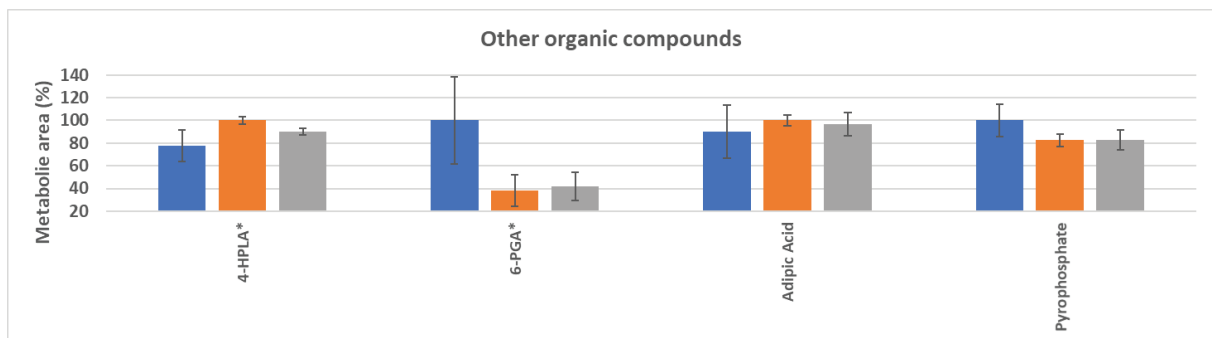
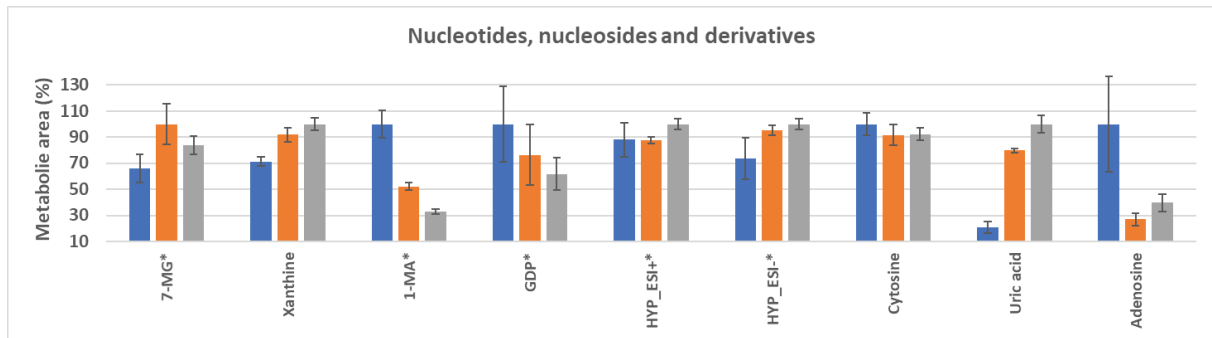
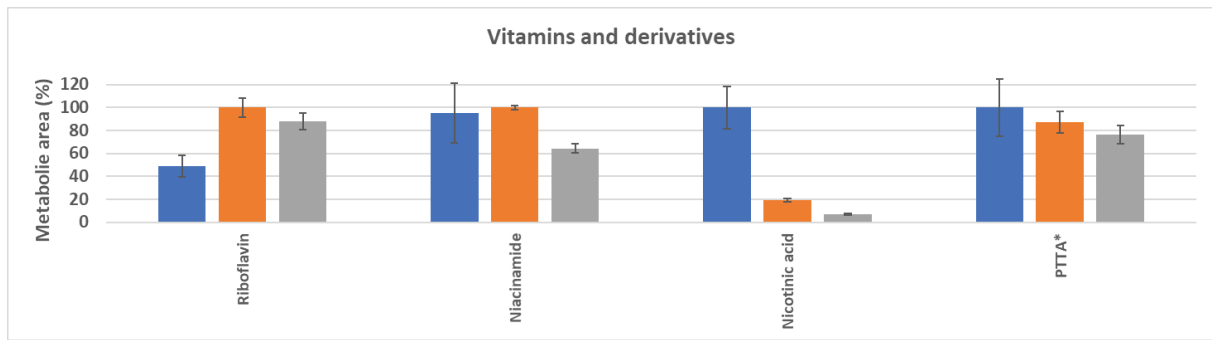


Annexe 7 : Comparison of the homogenization conditions on TG 51:1-d5 and PC 11:0/11:0 internal standard (A) and TG species before and after normalization (B).



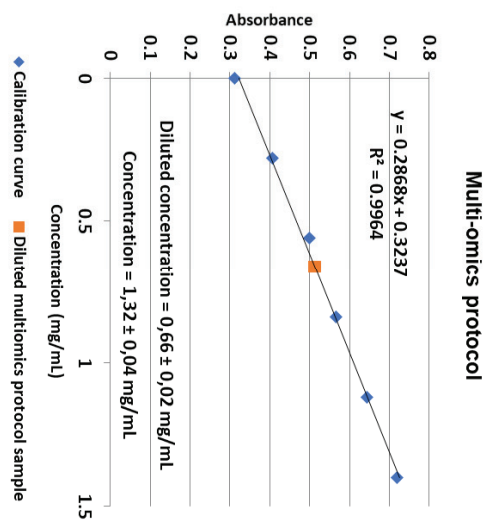
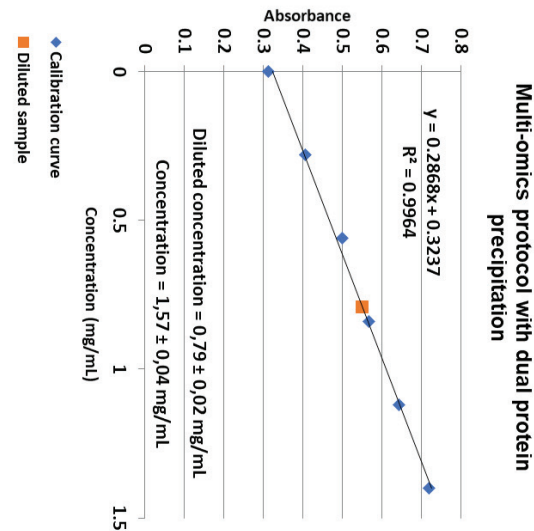
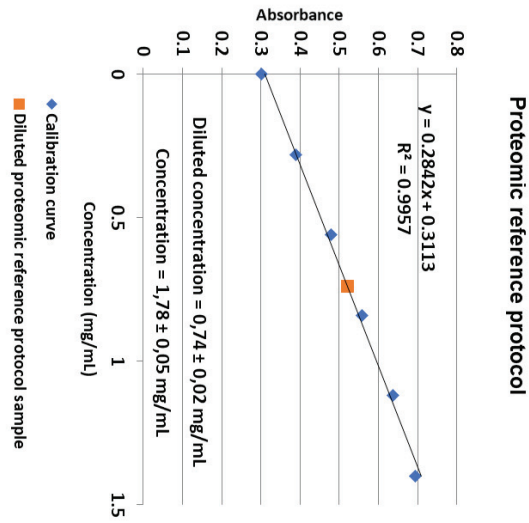
Annexe 8 : Comparison of the different homogenization conditions on gammarid metabolites.



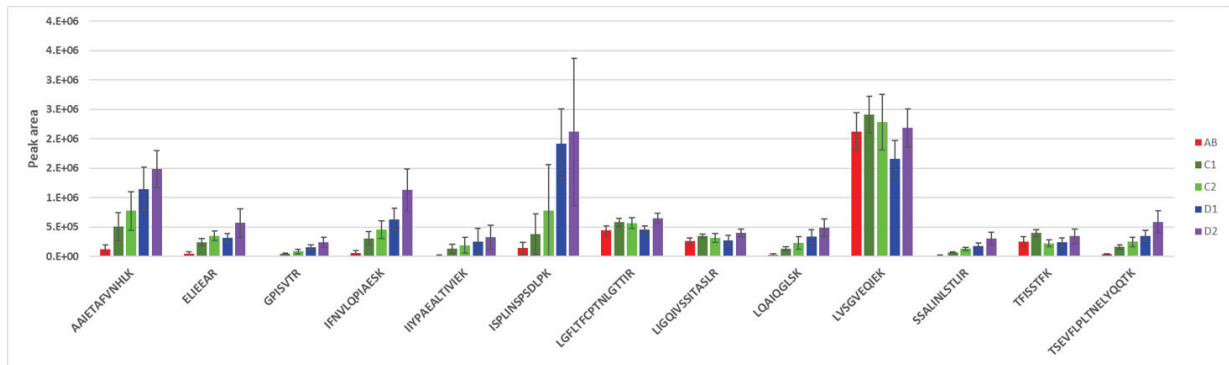




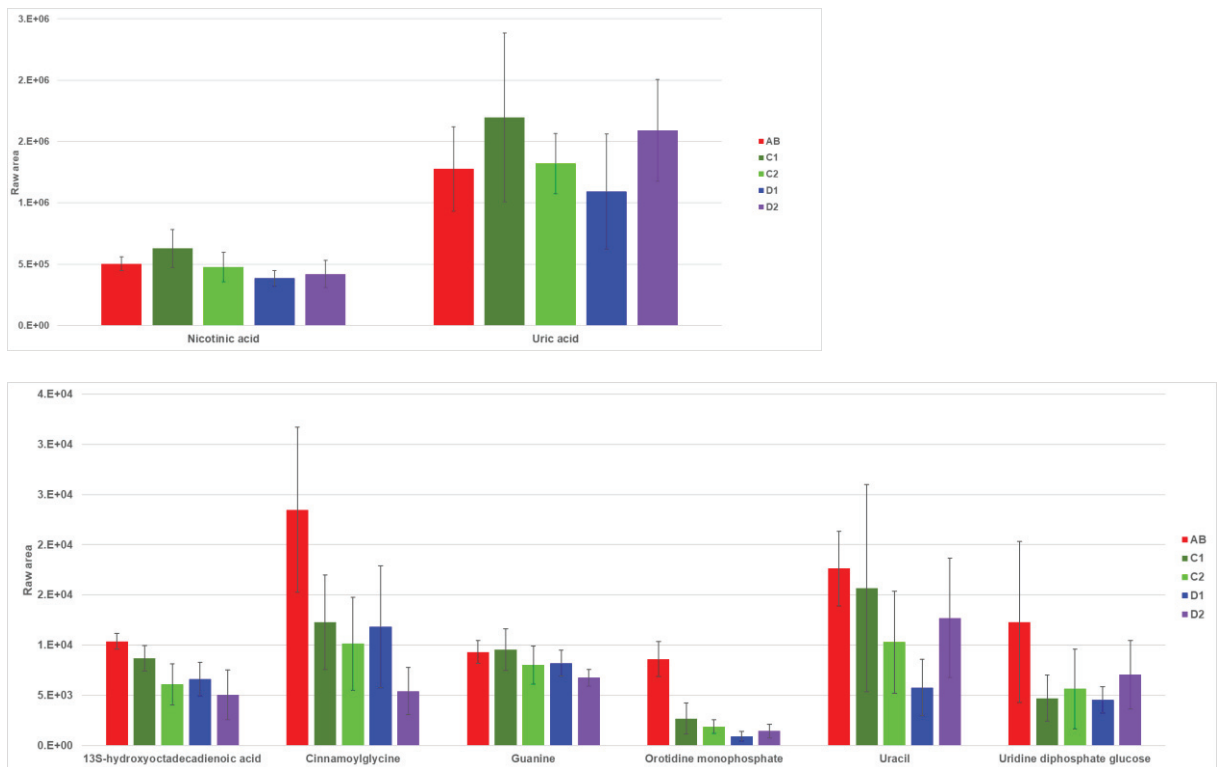
**Annexe 9: Determination of protein concentration using Bradford assay with classical Proteomics protocol, multi-omics with dual protein precipitation and multi-omics protocol.** Total amount of protein from entire gammarids was determined by Bradford assay calibration curve. BSA (Bovine Serum Albumin) standard curve recorded at 595 nm by spectrophotometry. Calibration curve: absorbance in function of BSA concentrations (mg/L).



**Annexe 10 : Comparison of peak area of the peptides with PLS-DA VIP scores (VIP >1) according to the different reproductive stages of female gammarids.**



**Annexe 11 : Comparison of peak area of metabolites with PLS-DA VIP scores (VIP >1) according to the different reproductive stages of female gammarids.**



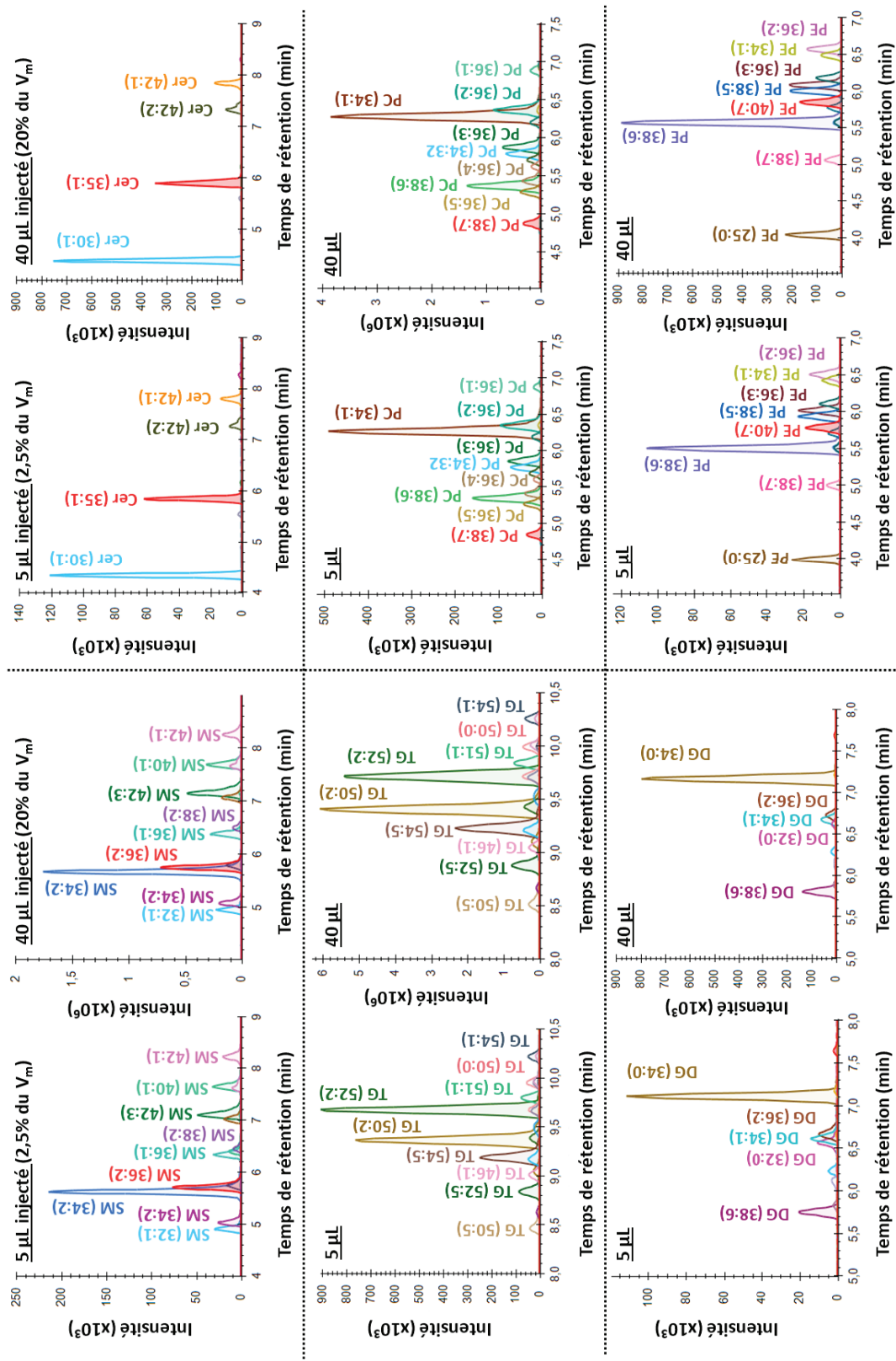
**Annexe 12 : Lg factors for MFA.**

	Lipidomics (ESI-)	Lipidomics (ESI+)	Proteomics	Metabolomics (ESI-)	Metabolomics (ESI+)	MFA
Lipidomics (ESI-)	2.09	1.13	0.90	0.66	0.62	<b>1.53</b>
Lipidomics (ESI+)	1.13	2.00	0.76	0.88	0.85	<b>1.59</b>
Proteomics	0.90	0.76	1.31	0.51	0.61	<b>1.16</b>
Metabolomics (ESI-)	0.66	0.88	0.51	1.09	0.64	<b>1.07</b>
Metabolomics (ESI+)	0.62	0.85	0.61	0.64	1.26	<b>1.13</b>
MFA	<b>1.53</b>	<b>1.59</b>	<b>1.16</b>	<b>1.07</b>	<b>1.13</b>	<b>1.83</b>

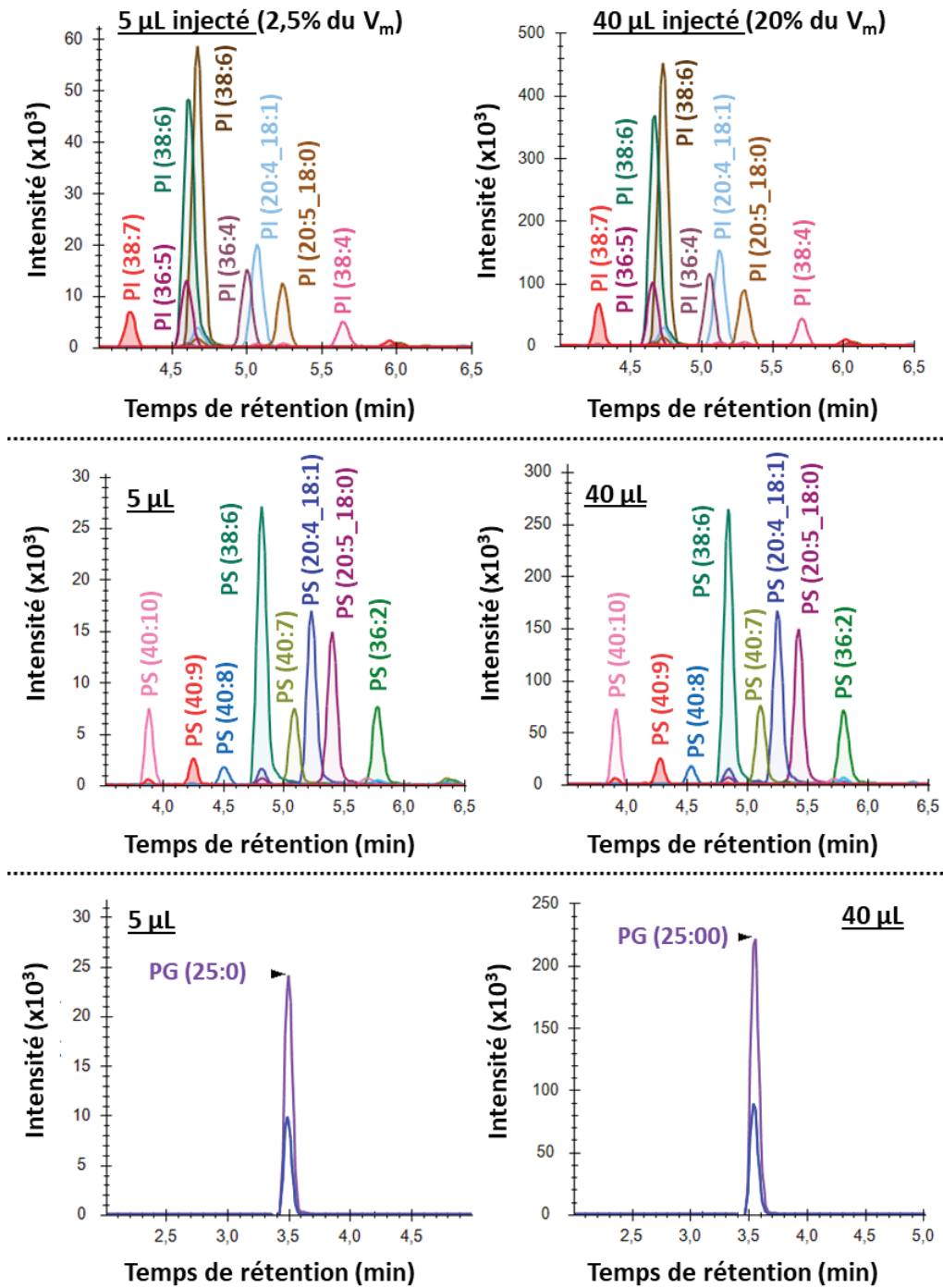
## Annexe 13 : RV factors for MFA.

	Lipidomics (ESI-)	Lipidomics (ESI+)	Proteomics	Metabolomics (ESI-)	Metabolomics (ESI+)	MFA
Lipidomics (ESI-)	1.00	0.55	0.55	0.44	0.38	<b>0.78</b>
Lipidomics (ESI+)	0.55	1.00	0.47	0.60	0.54	<b>0.83</b>
Proteomics	0.55	0.47	1.00	0.43	0.48	<b>0.75</b>
Metabolomics (ESI-)	0.44	0.60	0.43	1.00	0.55	<b>0.76</b>
Metabolomics (ESI+)	0.38	0.54	0.48	0.55	1.00	<b>0.74</b>
MFA	<b>0.78</b>	<b>0.83</b>	<b>0.75</b>	<b>0.76</b>	<b>0.74</b>	<b>1.00</b>

Annexe 14 : Évaluation du phénomène de *breakthrough* pour les familles lipidiques (SM, Cer, TG, PC, DG et PE) obtenues par HILIC-RPLC hors ligne en faisant varier les volumes d'injection de 5µL à 40µL dans la 2<sup>ème</sup> dimension, partie 1.

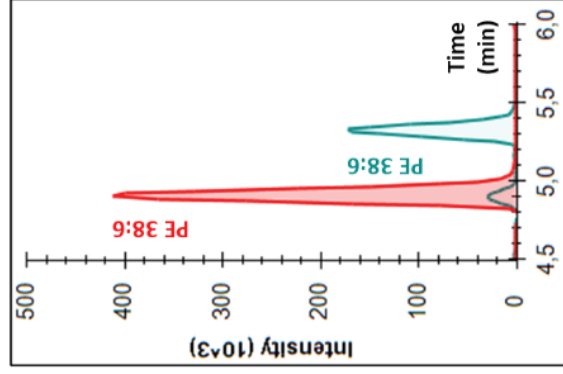


Annexe 15 : Évaluation du phénomène de *breakthrough* pour les familles lipidiques (PI, PS et PG) obtenues par HILIC-RPLC hors ligne en faisant varier les volumes d'injection de 5µL à 40µL dans la 2<sup>ème</sup> dimension, partie 2.



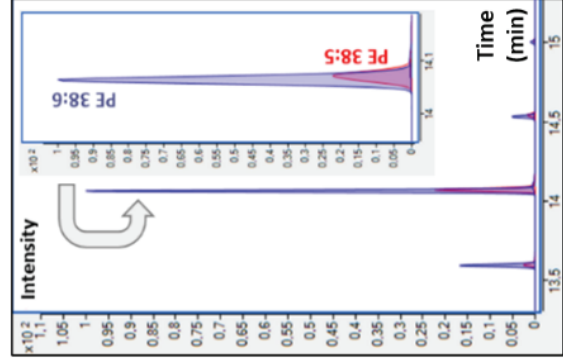
### 2<sup>ème</sup> dimension en HILIC - RPLC

Largueur de pic = 15 s



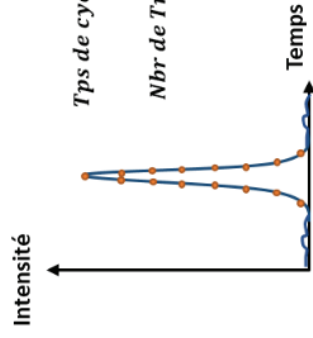
### 2<sup>ème</sup> dimension en HILIC x RPLC

Largueur de pic = 3,6 s



### Conditions de quantification en mode ciblée :

- Nombre de points par pic :



$$\text{Tps de cycle} = \frac{\text{Largueur des pics}}{\text{Nbr de points}}$$

$$\text{Nbr de Tr max} = \frac{\text{Tps de cycle}}{\text{dwell time} + \text{tps de repos}}$$

15 points/pic

- Performance analytique (Qtrap 5500) :
  - Dwell time min = 5 ms
  - Temps de repos min = 5 ms
- Nombre de transition max par méthode MRM selon le mode 2D :
  - **HILIC x RPLC** : Temps de cycle max = 0,24 s ; Nbr de Tr max = 24
  - **HILIC - RPLC** : Temps de cycle max = 1 s ; Nbr de Tr max = 100



## Références bibliographiques

1. D'Adamo, C.R., D'Urso, A., Ryan, K.A., Yerges-Armstrong, L.M., Semba, R.D., Steinle, N.I., Mitchell, B.D., Shuldiner, A.R., and McArdle, P.F. (2016). A common variant in the SETD7 gene predicts serum lycopene concentrations. *Nutrients* 8, 82.
2. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
3. Barbier Saint Hilaire, P. (2020). Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse.
4. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the yeast transcriptome. *Cell* 88, 243–251.
5. Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16, 1090–1094.
6. Garcia-Reyero, N., and Perkins, E.J. (2011). Systems biology: leading the revolution in ecotoxicology. *Environ. Toxicol. Chem.* 30, 265–273.
7. Patterson, S.D., and Aebersold, R.H. (2003). Proteomics: the first decade and beyond. *Nat. Genet.* 33 Suppl, 311–323.
8. Gregorich, Z.R., Chang, Y.-H., and Ge, Y. (2014). Proteomics in heart failure: top-down or bottom-up? *Pflugers Arch.* 466, 1199–1209.
9. Wright, E.P., Partridge, M.A., Padula, M.P., Gauci, V.J., Malladi, C.S., and Coorsen, J.R. (2014). Top-down proteomics: enhancing 2D gel electrophoresis from tissue processing to high-sensitivity protein detection. *Proteomics* 14, 872–889.
10. Chait, B.T. (2006). Chemistry. Mass spectrometry: bottom-up or top-down? *Science* 314, 65–66.
11. Catherman, A.D., Durbin, K.R., Ahlf, D.R., Early, B.P., Fellers, R.T., Tran, J.C., Thomas, P.M., and Kelleher, N.L. (2013). Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol. Cell. Proteomics* 12, 3465–3473.
12. Catherman, A.D., Skinner, O.S., and Kelleher, N.L. (2014). Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* 445, 683–693.
13. Ntai, I., LeDuc, R.D., Fellers, R.T., Erdmann-Gilmore, P., Davies, S.R., Rumsey, J., Early, B.P., Thomas, P.M., Li, S., Compton, P.D., *et al.* (2016). Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol. Cell. Proteomics* 15, 45–56.
14. Holmes, E., Wilson, I.D., and Nicholson, J.K. (2008). Metabolic phenotyping in health and disease. *Cell* 134, 714–717.
15. Junot, C., Fenaille, F., Colsch, B., and Bécher, F. (2014). High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev.* 33, 471–500.
16. Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., *et al.* (2007). HMDB: the human metabolome database. *Nucleic Acids Res.* 35, D521-6.
17. Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., *et al.* (2005). A comprehensive classification system for lipids. *J. Lipid Res.* 46, 839–861.
18. Sepúlveda, M.S., Ralston-Hooper, K.J., Sanchez, B.C., Hopf-Jannasch, A., Baker, S.D., Diaz, N., and Adamec, J. (2009). Use of proteomic and metabolomic techniques in ecotoxicological research. In *General, applied*

- and systems toxicology, B. Ballantyne, T. C. Marrs, T. Syversen, D. A. Casciano, and S. C. Sahu, eds. (Chichester, UK: John Wiley & Sons, Ltd).
19. Lankadurai, B.P., Nagato, E.G., and Simpson, M.J. (2013). Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environ. Rev.* *21*, 180–205.
  20. Vidal, M. (2009). A unifying view of 21st century systems biology. *FEBS Lett.* *583*, 3891–3894.
  21. Potters, G. (2010). *Systems Biology of the Cell*. Nature Education.
  22. Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H., and Tsai, P. (2008). The perils of plenty: what are we going to do with all these genes? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *363*, 3893–3902.
  23. Cavazzana-Calvo, M., and Debais, D. (2011). *Les biomédicaments* (Presses Universitaires de France).
  24. Odenkirk, M.T., Reif, D.M., and Baker, E.S. (2021). Multiomic big data analysis challenges: increasing confidence in the interpretation of artificial intelligence assessments. *Anal. Chem.* *93*, 7763–7773.
  25. Kopczynski, D., Coman, C., Zahedi, R.P., Lorenz, K., Sickmann, A., and Ahrends, R. (2017). Multi-OMICS: a critical technical perspective on integrative lipidomics approaches. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* *1862*, 808–811.
  26. Arrigoni-Martelli, E., and Caso, V. (2001). Carnitine protects mitochondria and removes toxic acyls from xenobiotics. *Drugs Exp. Clin. Res.* *27*, 27–49.
  27. Reddy, J.K., and Hashimoto, T. (2001). Peroxisomal beta-oxidation and peroxisome proliferator-activated receptor alpha: an adaptive metabolic system. *Annu. Rev. Nutr.* *21*, 193–230.
  28. Ciocan-Cartita, C.A., Jurj, A., Buse, M., Gulei, D., Braicu, C., Raduly, L., Cojocneanu, R., Pruteanu, L.L., Iuga, C.A., Coza, O., *et al.* (2019). The Relevance of Mass Spectrometry Analysis for Personalized Medicine through Its Successful Application in Cancer “Omics”. *Int. J. Mol. Sci.* *20*.
  29. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* *14*, e8126.
  30. Venable, J.D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., and Yates, J.R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* *1*, 39–45.
  31. Hu, A., Noble, W.S., and Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. [version 1; peer review: 3 approved]. *F1000Res.* *5*.
  32. Lam, H., and Aebersold, R. (2011). Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods* *54*, 424–431.
  33. Fortin, T., Salvador, A., Charrier, J.P., Lenz, C., Bettsworth, F., Lacoux, X., Choquet-Kastylevsky, G., and Lemoine, J. (2009). Multiple reaction monitoring cubed for protein quantification at the low nanogram/milliliter level in nondepleted human serum. *Anal. Chem.* *81*, 9343–9352.
  34. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (2012). Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteomics* *11*, 1709–1723.
  35. Gallien, S., Bourmaud, A., Kim, S.Y., and Domon, B. (2014). Technical considerations for large-scale parallel reaction monitoring analysis. *J. Proteomics* *100*, 147–159.
  36. Schiffmann, C., Hansen, R., Baumann, S., Kublik, A., Nielsen, P.H., Adrian, L., von Bergen, M., Jehmlich, N., and Seifert, J. (2014). Comparison of targeted peptide quantification assays for reductive dehalogenases by selective reaction monitoring (SRM) and precursor reaction monitoring (PRM). *Anal. Bioanal. Chem.* *406*, 283–291.
  37. Ronsein, G.E., Pamir, N., von Haller, P.D., Kim, D.S., Oda, M.N., Jarvik, G.P., Vaisar, T., and Heinecke, J.W. (2015). Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteomics* *113*, 388–399.

38. Hoffman, M.A., Fang, B., Haura, E.B., Rix, U., and Koomen, J.M. (2018). Comparison of quantitative mass spectrometry platforms for monitoring kinase ATP probe uptake in lung cancer. *J. Proteome Res.* *17*, 63–75.
39. Rougemont, B. (2016). Quantification de protéines dans des matrices complexes par spectrométrie de masse : nouveaux outils et applications.
40. Zhou, J., Liu, H., Liu, Y., Liu, J., Zhao, X., and Yin, Y. (2016). Development and Evaluation of a Parallel Reaction Monitoring Strategy for Large-Scale Targeted Metabolomics Quantification. *Anal. Chem.* *88*, 4478–4486.
41. Andrews, G.L., Simons, B.L., Young, J.B., Hawkrigde, A.M., and Muddiman, D.C. (2011). Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* *83*, 5442–5446.
42. Cajka, T., and Fiehn, O. (2016). Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* *88*, 524–545.
43. Tang, H., Fang, H., Yin, E., Brasier, A.R., Sowers, L.C., and Zhang, K. (2014). Multiplexed parallel reaction monitoring targeting histone modifications on the QExactive mass spectrometer. *Anal. Chem.* *86*, 5526–5534.
44. Yu, D., Rupasinghe, T.W.T., Boughton, B.A., Natera, S.H.A., Hill, C.B., Tarazona, P., Feussner, I., and Roessner, U. (2018). A high-resolution HPLC-QqTOF platform using parallel reaction monitoring for in-depth lipid discovery and rapid profiling. *Anal. Chim. Acta* *1026*, 87–100.
45. Aycirieux, S., Carrière, R., Bardet, C., Blanc, J.C.Y.L., Salvador, A., Fortin, T., and Lemoine, J. (2020). Streamlined Development of Targeted Mass Spectrometry-Based Method Combining Scout-MRM and a Web-Based Tool Indexed with Scout Peptides. *Proteomics* *20*, e1900254.
46. Salvador, A., Carrière, R., Aycirieux, S., Margoum, C., Leblanc, Y., and Lemoine, J. (2020). Scout-multiple reaction monitoring: A liquid chromatography tandem mass spectrometry approach for multi-residue pesticide analysis without time scheduling. *J. Chromatogr. A* *1621*, 461046.
47. Rougemont, B., Bontemps Gallo, S., Aycirieux, S., Carrière, R., Hondermarck, H., Lacroix, J.M., Le Blanc, J.C.Y., and Lemoine, J. (2017). Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by *Dickeya dadantii* Proteomic Analysis during Plant Infection. *Anal. Chem.* *89*, 1421–1426.
48. Schilling, B., MacLean, B., Held, J.M., Sahu, A.K., Rardin, M.J., Sorensen, D.J., Peters, T., Wolfe, A.J., Hunter, C.L., MacCoss, M.J., *et al.* (2015). Multiplexed, Scheduled, High-Resolution Parallel Reaction Monitoring on a Full Scan QqTOF Instrument with Integrated Data-Dependent and Targeted Mass Spectrometric Workflows. *Anal. Chem.* *87*, 10222–10229.
49. Burgess, M.W., Keshishian, H., Mani, D.R., Gillette, M.A., and Carr, S.A. (2014). Simplified and efficient quantification of low-abundance proteins at very high multiplex via targeted mass spectrometry. *Mol. Cell. Proteomics* *13*, 1137–1149.
50. Charnot, A., Gouveia, D., Armengaud, J., Almunia, C., Chaumot, A., Lemoine, J., Geffard, O., and Salvador, A. (2017). Multiplexed assay for protein quantitation in the invertebrate *Gammarus fossarum* by liquid chromatography coupled to tandem mass spectrometry. *Anal. Bioanal. Chem.* *409*, 3969–3991.
51. Zhou, J., Liu, C., Si, D., Jia, B., Zhong, L., and Yin, Y. (2017). Workflow development for targeted lipidomic quantification using parallel reaction monitoring on a quadrupole-time of flight mass spectrometry. *Anal. Chim. Acta* *972*, 62–72.
52. Weir, J.M., Wong, G., Barlow, C.K., Greeve, M.A., Kowalczyk, A., Almasy, L., Comuzzie, A.G., Mahaney, M.C., Jowett, J.B.M., Shaw, J., *et al.* (2013). Plasma lipid profiling in a large population-based cohort. *J. Lipid Res.* *54*, 2898–2908.
53. Züllig, T., Trötz Müller, M., and Köfeler, H.C. (2020). Lipidomics from sample preparation to data analysis: a primer. *Anal. Bioanal. Chem.* *412*, 2191–2209.
54. Quehenberger, O., Armando, A.M., Brown, A.H., Milne, S.B., Myers, D.S., Merrill, A.H., Bandyopadhyay, S., Jones, K.N., Kelly, S., Shaner, R.L., *et al.* (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.* *51*, 3299–3305.

55. Cajka, T., and Fiehn, O. (2014). Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *Trends Analyt. Chem.* *61*, 192–206.
56. González-Domínguez, R., Jáuregui, O., Queipo-Ortuño, M.I., and Andrés-Lacueva, C. (2020). Characterization of the Human Exposome by a Comprehensive and Quantitative Large-Scale Multianalyte Metabolomics Platform. *Anal. Chem.* *92*, 13767–13775.
57. Yuan, M., Breitkopf, S.B., and Asara, J.M. (2017). Serial-omics characterization of equine urine. *PLoS ONE* *12*, e0186258.
58. Züllig, T., and Köfeler, H.C. (2020). High resolution mass spectrometry in lipidomics. *Mass Spectrom. Rev.*
59. Cajka, T., Smilowitz, J.T., and Fiehn, O. (2017). Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. *Anal. Chem.* *89*, 12360–12368.
60. Antignac, J.-P., de Wasch, K., Monteau, F., De Brabander, H., Andre, F., and Le Bizec, B. (2005). The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. *Anal. Chim. Acta* *529*, 129–136.
61. King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., and Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *J. Am. Soc. Mass Spectrom.* *11*, 942–950.
62. Simon, R. (2012). La quantification ciblée de protéines et peptides par chromatographie liquide couplée à la spectrométrie de masse en tandem : développements analytiques et applications.
63. Koistinen, K.M., Suoniemi, M., Simolin, H., and Ekroos, K. (2015). Quantitative lysophospholipidomics in human plasma and skin by LC-MS/MS. *Anal. Bioanal. Chem.* *407*, 5091–5099.
64. Gika, H.G., Theodoridis, G.A., Plumb, R.S., and Wilson, I.D. (2014). Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* *87*, 12–25.
65. Bonnefoy, C., Fildier, A., Buleté, A., Bordes, C., Garric, J., and Vulliet, E. (2019). Untargeted analysis of nanoLC-HRMS data by ANOVA-PCA to highlight metabolites in *Gammarus fossarum* after in vivo exposure to pharmaceuticals. *Talanta* *202*, 221–229.
66. Sheikholeslami, M.N., Gómez-Canela, C., Barron, L.P., Barata, C., Vosough, M., and Tauler, R. (2020). Untargeted metabolomics changes on *Gammarus pulex* induced by propranolol, triclosan, and nimesulide pharmaceutical drugs. *Chemosphere* *260*, 127479.
67. Cui, L., Lu, H., and Lee, Y.H. (2018). Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases. *Mass Spectrom. Rev.* *37*, 772–792.
68. Liu, Z., Tu, M.-J., Zhang, C., Jilek, J.L., Zhang, Q.-Y., and Yu, A.-M. (2019). A reliable LC-MS/MS method for the quantification of natural amino acids in mouse plasma: Method validation and application to a study on amino acid dynamics during hepatocellular carcinoma progression. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* *1124*, 72–81.
69. Li, K., Naviaux, J.C., Bright, A.T., Wang, L., and Naviaux, R.K. (2017). A robust, single-injection method for targeted, broad-spectrum plasma metabolomics. *Metabolomics* *13*, 122.
70. Jiang, P., and Lucy, C.A. (2016). Coupling normal phase liquid chromatography with electrospray ionization mass spectrometry: strategies and applications. *Anal. Methods* *8*, 6478–6488.
71. Tang, D.-Q., Zou, L., Yin, X.-X., and Ong, C.N. (2016). HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass Spectrom. Rev.* *35*, 574–600.
72. Buszewski, B., and Noga, S. (2012). Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal. Bioanal. Chem.* *402*, 231–247.
73. Periat, A.C. (2015). Utilisation de la chromatographie d'interaction hydrophile pour l'analyse de composés d'intérêt pharmaceutique. Université de Genève.

74. Heaton, J.C., Russell, J.J., Underwood, T., Boughtflower, R., and McCalley, D.V. (2014). Comparison of peak shape in hydrophilic interaction chromatography using acidic salt buffers and simple acid solutions. *J. Chromatogr. A* *1347*, 39–48.
75. McCalley, D.V. (2007). Is hydrophilic interaction chromatography with silica columns a viable alternative to reversed-phase liquid chromatography for the analysis of ionisable compounds? *J. Chromatogr. A* *1171*, 46–55.
76. Gallego, S.F., Hermansson, M., Liebisch, G., Hodson, L., and Ejsing, C.S. (2018). Total Fatty Acid Analysis of Human Blood Samples in One Minute by High-Resolution Mass Spectrometry. *Biomolecules* *9*.
77. Topolska, M., Martínez-Montañés, F., and Ejsing, C.S. (2020). A Simple and Direct Assay for Monitoring Fatty Acid Synthase Activity and Product-Specificity by High-Resolution Mass Spectrometry. *Biomolecules* *10*.
78. Carrasco-Pancorbo, A., Navas-Iglesias, N., and Cuadros-Rodríguez, L. (2009). From lipid analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part I: Modern lipid analysis. *TrAC Trends in Analytical Chemistry* *28*, 263–278.
79. Holcapek, M., Jandera, P., Fischer, J., and Prokes, B. (1999). Analytical monitoring of the production of biodiesel by high-performance liquid chromatography with various detection methods. *J. Chromatogr. A* *858*, 13–31.
80. Mal, M., and Wong, S. (2011). Application Note: A HILIC-Based UPLC/MS Method for the Separation of Lipid Classes from Plasma. .
81. Sandra, K., Moshir, M., D’hondt, F., Verleysen, K., Kas, K., and Sandra, P. (2008). Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* *866*, 48–63.
82. Köcher, T., Swart, R., and Mechtler, K. (2011). Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal. Chem.* *83*, 2699–2704.
83. Foreman, R.E., George, A.L., Reimann, F., Gribble, F.M., and Kay, R.G. (2021). Peptidomics: A review of clinical applications and methodologies. *J. Proteome Res.* *20*, 3782–3797.
84. Lardeux, H., Duivelshof, B.L., Colas, O., Beck, A., McCalley, D.V., Guillarme, D., and D’Atri, V. (2021). Alternative mobile phase additives for the characterization of protein biopharmaceuticals in liquid chromatography - Mass spectrometry. *Anal. Chim. Acta* *1156*, 338347.
85. Nshanian, M., Lakshmanan, R., Chen, H., Loo, R.R.O., and Loo, J.A. (2017). Enhancing sensitivity of liquid chromatography–mass spectrometry of peptides and proteins using supercharging agents. *Int. J. Mass Spectrom.*
86. Faugere, J., Gouveia, D., Ayciriex, S., Chaumot, A., Almunia, C., François, A., Armengaud, J., Lemoine, J., Geffard, O., Degli-Esposti, D., *et al.* (2020). High-multiplexed monitoring of protein biomarkers in the sentinel *Gammarus fossarum* by targeted scout-MRM assay, a new vision for ecotoxicoproteomics. *J. Proteomics* *226*, 103901.
87. Simon, R., Enjalbert, Q., Biarc, J., Lemoine, J., and Salvador, A. (2012). Evaluation of hydrophilic interaction chromatography (HILIC) versus C<sub>18</sub> reversed-phase chromatography for targeted quantification of peptides by mass spectrometry. *J. Chromatogr. A* *1264*, 31–39.
88. Jones, J.L., Dongre, A.R., Somogyi, A., and Wysocki, V.H. (1994). Sequence Dependence of Peptide Fragmentation Efficiency Curves Determined by Electrospray Ionization/Surface-Induced Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* *116*, 8368–8369.
89. Tang, X.J., Thibault, P., and Boyd, R.K. (1993). Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal. Chem.* *65*, 2824–2834.
90. Burlet, O., Orkiszewski, R.S., Ballard, K.D., and Gaskell, S.J. (1992). Charge promotion of low-energy fragmentations of peptide ions. *Rapid Commun. Mass Spectrom.* *6*, 658–662.

91. Guiochon, G., Marchetti, N., Mriziq, K., and Shalliker, R.A. (2008). Implementations of two-dimensional liquid chromatography. *J. Chromatogr. A* *1189*, 109–168.
92. Matos, J.T.V., Duarte, R.M.B.O., and Duarte, A.C. (2012). Trends in data processing of comprehensive two-dimensional chromatography: state of the art. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* *910*, 31–45.
93. Giddings, J.C. (1987). Concepts and comparisons in multidimensional separation. *J. High Resol. Chromatogr.* *10*, 319–323.
94. Iguiniz, M., and Heinisch, S. (2017). Two-dimensional liquid chromatography in pharmaceutical analysis. Instrumental aspects, trends and applications. *J. Pharm. Biomed. Anal.* *145*, 482–503.
95. Giddings, J. Calvin. (1967). Maximum number of components resolvable by gel filtration and other elution chromatographic methods. *Anal. Chem.* *39*, 1027–1028.
96. Horvath, C.G., and Lipsky, S.R. (1967). Peak capacity in chromatography. *Anal. Chem.* *39*, 1893–1893.
97. Stoll, D.R., Li, X., Wang, X., Carr, P.W., Porter, S.E.G., and Rutan, S.C. (2007). Fast, comprehensive two-dimensional liquid chromatography. *J. Chromatogr. A* *1168*, 3–43; discussion 2.
98. Guiochon, G. (2006). The limits of the separation power of unidimensional column liquid chromatography. *J. Chromatogr. A* *1126*, 6–49.
99. Gilar, M., Daly, A.E., Kele, M., Neue, U.D., and Gebler, J.C. (2004). Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *J. Chromatogr. A* *1061*, 183–192.
100. Karger, B.L., 1939- (1973). *An Introduction To Separation Science* 1st ed. (New York: Wiley).
101. Giddings, J.C. (1984). Two-dimensional separations: concept and promise. *Anal. Chem.* *56*, 1258A-1270A.
102. Guiochon, G., Beaver, L.A., Gonnord, M.F., Siouffi, A.M., and Zakaria, M. (1983). Theoretical investigation of the potentialities of the use of a multidimensional column in chromatography. *Journal of Chromatography A* *255*, 415–437.
103. Xing, Q., Liang, T., Shen, G., Wang, X., Jin, Y., and Liang, X. (2012). Comprehensive HILIC × RPLC with mass spectrometry detection for the analysis of saponins in *Panax notoginseng*. *Analyst* *137*, 2239–2249.
104. Di Palma, S., Stange, D., van de Wetering, M., Clevers, H., Heck, A.J.R., and Mohammed, S. (2011). Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. *J. Proteome Res.* *10*, 3814–3819.
105. Schoenmakers, P., Marriott, P., and Beens, J. (2003). Nomenclature and conventions in comprehensive multidimensional chromatography. *LC•GC Europe*.
106. Wang, X., Buckenmaier, S., and Stoll, D. (2017). The growing role of two-dimensional LC in the biopharmaceutical industry. *J. Appl. Bioanal.* *3*, 120–126.
107. D’attoma, A. (2014). Développement de méthodes bidimensionnelles en ligne LCxLC-MS pour l’analyse de composés chargés.
108. Sommella, E., Cacciola, F., Donato, P., Dugo, P., Campiglia, P., and Mondello, L. (2012). Development of an online capillary comprehensive 2D-LC system for the analysis of proteome samples. *J. Sep. Sci.* *35*, 530–533.
109. Beelders, T., Kalili, K.M., Joubert, E., de Beer, D., and de Villiers, A. (2012). Comprehensive two-dimensional liquid chromatographic analysis of rooibos (*Aspalathus linearis*) phenolics. *J. Sep. Sci.* *35*, 1808–1820.
110. D’Attoma, A., and Heinisch, S. (2013). On-line comprehensive two dimensional separations of charged compounds using reversed-phase high performance liquid chromatography and hydrophilic interaction chromatography. Part II: application to the separation of peptides. *J. Chromatogr. A* *1306*, 27–36.
111. Stevenson, P.G., Bassanese, D.N., Conlan, X.A., and Barnett, N.W. (2014). Improving peak shapes with counter gradients in two-dimensional high performance liquid chromatography. *J. Chromatogr. A* *1337*, 147–154.



112. Chen, Y., Vu, J., Thompson, M.G., Sharpless, W.A., Chan, L.J.G., Gin, J.W., Keasling, J.D., Adams, P.D., and Petzold, C.J. (2019). A rapid methods development workflow for high-throughput quantitative proteomic applications. *PLoS ONE* *14*, e0211582.
113. Stoll, D.R., Harmes, D.C., Staples, G.O., Potter, O.G., Dammann, C.T., Guillaume, D., and Beck, A. (2018). Development of Comprehensive Online Two-Dimensional Liquid Chromatography/Mass Spectrometry Using Hydrophilic Interaction and Reversed-Phase Separations for Rapid and Deep Profiling of Therapeutic Antibodies. *Anal. Chem.* *90*, 5923–5929.
114. Chapel, S., Rouvière, F., and Heinisch, S. (2020). Pushing the limits of resolving power and analysis time in on-line comprehensive hydrophilic interaction x reversed phase liquid chromatography for the analysis of complex peptide samples. *J. Chromatogr. A* *1615*, 460753.
115. Montero, L., Herrero, M., Ibáñez, E., and Cifuentes, A. (2014). Separation and characterization of phlorotannins from brown algae *Cystoseira abies-marina* by comprehensive two-dimensional liquid chromatography. *Electrophoresis* *35*, 1644–1651.
116. Willemse, C.M., Stander, M.A., Tredoux, A.G.J., and de Villiers, A. (2014). Comprehensive two-dimensional liquid chromatographic analysis of anthocyanins. *J. Chromatogr. A* *1359*, 189–201.
117. Zheng, X.-X., Du, Y., Xu, B.-J., Wang, T.-Y., Zhong, Q.-Q., Li, Z., Ji, S., Guo, M.-Z., Yang, D.-Z., and Tang, D.-Q. (2019). Off-line two-dimensional liquid chromatography coupled with diode array detection and quadrupole-time of flight mass spectrometry for the biotransformation kinetics of Ginkgo biloba leaves extract by diabetic rat liver microsomes. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* *1109*, 1–9.
118. Bruins, A.P. (1998). Mechanistic aspects of electrospray ionization. *Journal of Chromatography A* *794*, 345–357.
119. Manisali, I., Chen, D.D.Y., and Schneider, B.B. (2006). Electrospray ionization source geometry for mass spectrometry: past, present, and future. *TrAC Trends in Analytical Chemistry* *25*, 243–256.
120. Nguyen, D.T.-T., Guillaume, D., Heinisch, S., Barrioulet, M.-P., Rocca, J.-L., Rudaz, S., and Veuthey, J.-L. (2007). High throughput liquid chromatography with sub-2 microm particles at high pressure and high temperature. *J. Chromatogr. A* *1167*, 76–84.
121. Agilent (2019). Simplify Complex Separations. Fast.
122. Grübner, M., Dunkel, A., Steiner, F., and Hofmann, T. (2021). Systematic Evaluation of Liquid Chromatography (LC) Column Combinations for Application in Two-Dimensional LC Metabolomic Studies. *Anal. Chem.* *93*, 12565–12573.
123. Xu, M., Legradi, J., and Leonards, P. (2020). Evaluation of LC-MS and LC×LC-MS in analysis of zebrafish embryo samples for comprehensive lipid profiling. *Anal. Bioanal. Chem.* *412*, 4313–4325.
124. Holcapek, M., Velínská, H., Lída, M., and Cesla, P. (2009). Orthogonality of silver-ion and non-aqueous reversed-phase HPLC/MS in the analysis of complex natural mixtures of triacylglycerols. *J. Sep. Sci.* *32*, 3672–3680.
125. Lída, M., Cífková, E., and Holcapek, M. (2011). Lipidomic profiling of biological tissues using off-line two-dimensional high-performance liquid chromatography-mass spectrometry. *J. Chromatogr. A* *1218*, 5146–5156.
126. Nie, H., Liu, R., Yang, Y., Bai, Y., Guan, Y., Qian, D., Wang, T., and Liu, H. (2010). Lipid profiling of rat peritoneal surface layers by online normal- and reversed-phase 2D LC QToF-MS. *J. Lipid Res.* *51*, 2833–2844.
127. Li, M., Tong, X., Lv, P., Feng, B., Yang, L., Wu, Z., Cui, X., Bai, Y., Huang, Y., and Liu, H. (2014). A not-stop-flow online normal-/reversed-phase two-dimensional liquid chromatography-quadrupole time-of-flight mass spectrometry method for comprehensive lipid profiling of human plasma from atherosclerosis patients. *J. Chromatogr. A* *1372C*, 110–119.
128. Ling, Y.S., Liang, H.-J., Lin, M.-H., Tang, C.-H., Wu, K.-Y., Kuo, M.-L., and Lin, C.Y. (2014). Two-dimensional LC-MS/MS to enhance ceramide and phosphatidylcholine species profiling in mouse liver. *Biomed. Chromatogr.* *28*, 1284–1293.

129. Sun, C., Zhao, Y.-Y., and Curtis, J.M. (2014). Elucidation of phosphatidylcholine isomers using two dimensional liquid chromatography coupled in-line with ozonolysis mass spectrometry. *J. Chromatogr. A* *1351*, 37–45.
130. Li, M., Feng, B., Liang, Y., Zhang, W., Bai, Y., Tang, W., Wang, T., and Liu, H. (2013). Lipid profiling of human plasma from peritoneal dialysis patients using an improved 2D (NP/RP) LC-QToF MS method. *Anal. Bioanal. Chem.* *405*, 6629–6638.
131. Dugo, P., Fawzy, N., Cichello, F., Cacciola, F., Donato, P., and Mondello, L. (2013). Stop-flow comprehensive two-dimensional liquid chromatography combined with mass spectrometric detection for phospholipid analysis. *J. Chromatogr. A* *1278*, 46–53.
132. Wang, S., Li, J., Shi, X., Qiao, L., Lu, X., and Xu, G. (2013). A novel stop-flow two-dimensional liquid chromatography-mass spectrometry method for lipid analysis. *J. Chromatogr. A* *1321*, 65–72.
133. Holčápek, M., Ovčáčiková, M., Lísa, M., Cífková, E., and Hájek, T. (2015). Continuous comprehensive two-dimensional liquid chromatography-electrospray ionization mass spectrometry of complex lipidomic samples. *Anal. Bioanal. Chem.* *407*, 5033–5043.
134. Lovric, J. (2011). *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis* 1st ed. (Chichester, West Sussex: Wiley).
135. Dupree, E.J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B.A., and Darie, C.C. (2020). A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. *Proteomes* *8*.
136. Feist, P., and Hummon, A.B. (2015). Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.* *16*, 3537–3563.
137. Kim, B., Araujo, R., Howard, M., Magni, R., Liotta, L.A., and Luchini, A. (2018). Affinity enrichment for mass spectrometry: improving the yield of low abundance biomarkers. *Expert Rev. Proteomics* *15*, 353–366.
138. ten Have, S., Boulon, S., Ahmad, Y., and Lamond, A.I. (2011). Mass spectrometry-based immunoprecipitation proteomics - the user's guide. *Proteomics* *11*, 1153–1159.
139. Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* *422*, 198–207.
140. Olsen, J.V., Ong, S.-E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* *3*, 608–614.
141. Gershon, P.D. (2014). Cleaved and missed sites for trypsin, lys-C, and lys-N can be predicted with high confidence on the basis of sequence context. *J. Proteome Res.* *13*, 702–709.
142. Siepen, J.A., Keevil, E.-J., Knight, D., and Hubbard, S.J. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J. Proteome Res.* *6*, 399–408.
143. Folch, J., Lees, M., and Sloane Stanley, G.H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* *226*, 497–509.
144. Bligh, E.G., and Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* *37*, 911–917.
145. Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A., and Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* *49*, 1137–1146.
146. Zhou, B., Xiao, J.F., Tuli, L., and Ressom, H.W. (2012). LC-MS-based metabolomics. *Mol. Biosyst.* *8*, 470–481.
147. Pinu, F.R., Villas-Boas, S.G., and Aggio, R. (2017). Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols. *Metabolites* *7*.
148. de Koning, W., and van Dam, K. (1992). A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal. Biochem.* *204*, 118–123.
149. Navarro-Reig, M., Jaumot, J., Piña, B., Moyano, E., Galceran, M.T., and Tauler, R. (2017). Metabolomic analysis of the effects of cadmium and copper treatment in *Oryza sativa* L. using untargeted liquid

- chromatography coupled to high resolution mass spectrometry and all-ion fragmentation. *Metallomics* **9**, 660–675.
150. Jorge, T.F., Mata, A.T., and António, C. (2016). Mass spectrometry as a quantitative tool in plant metabolomics. *Philos. Transact. A Math. Phys. Eng. Sci.* **374**.
  151. D'Alessandro, A. ed. (2019). *High-Throughput Metabolomics: Methods and Protocols* (New York, NY: Springer New York).
  152. Khamis, M.M., Adamko, D.J., and El-Aneed, A. (2017). Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom. Rev.* **36**, 115–134.
  153. Fernández-Peralbo, M.A., and Luque de Castro, M.D. (2012). Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. *TrAC Trends in Analytical Chemistry* **41**, 75–85.
  154. Stevens, V.L., Hoover, E., Wang, Y., and Zanetti, K.A. (2019). Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. *Metabolites* **9**.
  155. Villaret-Cazadamont, J., Poupin, N., Tournadre, A., Batut, A., Gales, L., Zalko, D., Cabaton, N.J., Bellvert, F., and Bertrand-Michel, J. (2020). An optimized dual extraction method for the simultaneous and accurate analysis of polar metabolites and lipids carried out on single biological samples. *Metabolites* **10**.
  156. Zhang, J., Hong, Y., Jiang, L., Yi, X., Chen, Y., Liu, L., Chen, Z., Wu, Y., and Cai, Z. (2020). Global Metabolomic and Lipidomic Analysis Reveal the Synergistic Effect of Bufalin in Combination with Cinobufagin against HepG2 Cells. *J. Proteome Res.* **19**, 873–883.
  157. Wawrzyniak, R., Kosnowska, A., Macioszek, S., Bartoszewski, R., and Jan Markuszewski, M. (2018). New plasma preparation approach to enrich metabolome coverage in untargeted metabolomics: plasma protein bound hydrophobic metabolite release with proteinase K. *Sci. Rep.* **8**, 9541.
  158. Nakayasu, E.S., Nicora, C.D., Sims, A.C., Burnum-Johnson, K.E., Kim, Y.-M., Kyle, J.E., Matzke, M.M., Shukla, A.K., Chu, R.K., Schepmoes, A.A., *et al.* (2016). MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. *mSystems* **1**.
  159. Coman, C., Solari, F.A., Hentschel, A., Sickmann, A., Zahedi, R.P., and Ahrends, R. (2016). Simultaneous metabolite, protein, lipid extraction (SIMPLEX): A combinatorial multimolecular omics approach for systems biology. *Mol. Cell. Proteomics* **15**, 1453–1466.
  160. Nicora, C.D., Burnum-Johnson, K.E., Nakayasu, E.S., Casey, C.P., White, R.A., Roy Chowdhury, T., Kyle, J.E., Kim, Y.-M., Smith, R.D., Metz, T.O., *et al.* (2018). The MPLEx Protocol for Multi-omic Analyses of Soil Samples. *J. Vis. Exp.*
  161. Burnum-Johnson, K.E., Kyle, J.E., Eisfeld, A.J., Casey, C.P., Stratton, K.G., Gonzalez, J.F., Habyarimana, F., Negretti, N.M., Sims, A.C., Chauhan, S., *et al.* (2017). MPLEx: a method for simultaneous pathogen inactivation and extraction of samples for multi-omics profiling. *Analyst* **142**, 442–448.
  162. Rampler, E., Abiead, Y.E., Schoeny, H., Rusz, M., Hildebrand, F., Fitz, V., and Koellensperger, G. (2021). Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics-Standardization, Coverage, and Throughput. *Anal. Chem.* **93**, 519–545.
  163. López-Bascón, M.A., Calderón-Santiago, M., Sánchez-Ceinos, J., Fernández-Vega, A., Guzmán-Ruiz, R., López-Miranda, J., Malagon, M.M., and Priego-Capote, F. (2018). Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue. *Talanta* **177**, 86–93.
  164. León, I.R., Schwämmle, V., Jensen, O.N., and Sprenger, R.R. (2013). Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol. Cell. Proteomics* **12**, 2992–3005.
  165. Hentschel, A., Zahedi, R.P., and Ahrends, R. (2016). Protein lipid modifications--More than just a greasy ballast. *Proteomics* **16**, 759–782.
  166. Medina, J., van der Velpen, V., Teav, T., Guitton, Y., Gallart-Ayala, H., and Ivanisevic, J. (2020). Single-Step Extraction Coupled with Targeted HILIC-MS/MS Approach for Comprehensive Analysis of Human Plasma Lipidome and Polar Metabolome. *Metabolites* **10**.

167. Vorreiter, F., Richter, S., Peter, M., Baumann, S., von Bergen, M., and Tomm, J.M. (2016). Comparison and optimization of methods for the simultaneous extraction of DNA, RNA, proteins, and metabolites. *Anal. Biochem.* *508*, 25–33.
168. Gouveia, D., Almunia, C., Cogne, Y., Pible, O., Degli-Esposti, D., Salvador, A., Cristobal, S., Sheehan, D., Chaumot, A., Geffard, O., *et al.* (2018). Ecotoxicoproteomics: A decade of progress in our understanding of anthropogenic impact on the environment. *J. Proteomics* *198*, 66–77.
169. Cogne, Y. (2019). Bioinformatique pour l'exploration de la diversité inter-espèce et inter-populations: hétérogénéité et données multi-omiques.
170. Trapp, J. (2014). Approches protéomiques pour le développement de biomarqueurs chez l'amphipode d'eau douce *Gammarus fossarum* : découverte, caractérisation et quantification de protéines impliquées dans la fonction reproductrice .
171. Besse, J.-P., Coquery, M., Lopes, C., Chaumot, A., Budzinski, H., Labadie, P., and Geffard, O. (2013). Caged *Gammarus fossarum* (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values. *Water Res.* *47*, 650–660.
172. Castilhos Ghisi, N. de (2012). Relationship between biomarkers and pesticide exposure in fishes: A review. In *Pesticides - Advances in Chemical and Botanical Pesticides*, R. P. Soundararajan, ed. (InTech).
173. Depledge, M.H. (1994). The rational basis for the use of biomarkers as ecotoxicological tools. In *Nondestructive biomarkers in vertebrates*, M. C. Fossi and C. Leonzio, eds. (Lewis Publishers), pp. 261–285.
174. McCarthy, J.F., and Shugart, L.R. (1990). *Biomarkers of environmental contamination* (Lewis Publishers).
175. Lagadic, L. (2002). Biomarkers: Useful tools for the monitoring of aquatic environments. *Revue de medecine veterinaire* *153*, 581–588.
176. Garric, J., Morin, S., and Vincent-Hubert, F. (2010). *Les biomarqueurs en écotoxicologie : définition, intérêt, limite, usage*. IRSTEA.
177. García-Fernández, A.J., Espín, S., Gómez-Ramírez, P., Martínez-López, E., and Navas, I. (2020). Wildlife sentinels for human and environmental health hazards in ecotoxicological risk assessment. In *Ecotoxicological QSARs Methods in pharmacology and toxicology.*, K. Roy, ed. (New York, NY: Springer US), pp. 77–94.
178. Beck, A.C., Lash, E.M., and Hack, J.B. (2020). Environmental toxic exposures using companion animals as an indicator of human toxicity: A case report and discussion. *J. Emerg. Med.* *59*, e1–e7.
179. Fossi, M.C., and Panti, C. (2017). Sentinel species of marine ecosystems. In *Oxford research encyclopedia of environmental science* (Oxford University Press).
180. Kennish, M.J. (2018). Estuaries: Anthropogenic Impacts. In *Encyclopedia of engineering geology Encyclopedia of earth sciences series.*, P. Bobrowsky and B. Marker, eds. (Cham: Springer International Publishing), pp. 1–9.
181. Godard-Codding, C.A.J., Clark, R., Fossi, M.C., Marsili, L., Maltese, S., West, A.G., Valenzuela, L., Rowntree, V., Polyak, I., Cannon, J.C., *et al.* (2011). Pacific Ocean-wide profile of CYP1A1 expression, stable carbon and nitrogen isotope ratios, and organic contaminant burden in sperm whale skin biopsies. *Environ. Health Perspect.* *119*, 337–343.
182. M Cristina Fossi And Letizia Marsili (1997). The use of non destructive biomarkers in the study of marine mammals. *Biomarkers* *2*, 205–216.
183. Fossi, M.C., and Depledge, M.H. (2014). Exploring the potential of large vertebrates as early warning sentinels of threats to marine ecosystems, human health and wellbeing. *Mar. Environ. Res.* *100*, 1–2.
184. Schwacke, L.H., Gulland, F.M., and White, S. (2013). Sentinel species in oceans and human health. In *Environmental Toxicology*, E. A. Laws, ed. (New York, NY: Springer New York), pp. 503–528.
185. Depledge, M.H., and Fossi, M.C. (1994). The role of biomarkers in environmental assessment (2). *Invertebrates. Ecotoxicology* *3*, 161–172.

186. Bjørnstad, A., Larsen, B.K., Skadsheim, A., Jones, M.B., and Andersen, O.K. (2006). The potential of ecotoxicoproteomics in environmental monitoring: biomarker profiling in mussel plasma using ProteinChip array technology. *J. Toxicol. Environ. Health Part A* *69*, 77–96.
187. Knigge, T., Monsinjon, T., and Andersen, O.-K. (2004). Surface-enhanced laser desorption/ionization-time of flight-mass spectrometry approach to biomarker discovery in blue mussels (*Mytilus edulis*) exposed to polyaromatic hydrocarbons and heavy metals under field conditions. *Proteomics* *4*, 2722–2727.
188. Dumas, T., Courant, F., Almunia, C., Boccard, J., Rosain, D., Duporté, G., Armengaud, J., Fenet, H., and Gomez, E. (2021). An integrated metabolomics and proteogenomics approach reveals molecular alterations following carbamazepine exposure in the male mussel *Mytilus galloprovincialis*. *Chemosphere* *286*, 131793.
189. Dumas, T. (2020). Les approches –omiques, métabolomique et protéomique, pour l'étude de la relation de cause à effet entre contaminants émergents, produits pharmaceutiques et organismes marins, *Mytilus galloprovincialis*.
190. Karaman, G.S., and Pinkster, S. (1977). Freshwater Gammarus Species from Europe, North Africa and Adjacent Regions of Asia (Crustacea-Amphipoda). *BTD* *47*, 1–97.
191. Marković, Z., and Živić, I. (2007). Distribution of the species *Gammarus balcanicus* and *Gammarus fossarum* on the territory of Serbia (central part of the Balkan peninsula). *Crustaceana* *80*, 67–76.
192. Barnard, J.L., and Barnard, C.M. (1983). Freshwater Amphipoda of the world, I. Evolutionary patterns and II. Handbook and bibliography. Hayfield Associates, Mt. Vermon, Virginia.
193. Kunz, P.Y., Kienle, C., and Gerhardt, A. (2010). *Gammarus* spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests. *Rev. Environ. Contam. Toxicol.* *205*, 1–76.
194. Chaumot, A., Coulaud, R., Adam, O., Quéau, H., Lopes, C., and Geffard, O. (2020). In Situ Reproductive Bioassay with Caged *Gammarus fossarum* (Crustacea): Part 1-Gauging the Confounding Influence of Temperature and Water Hardness. *Environ. Toxicol. Chem.* *39*, 667–677.
195. Geffard, O., Xuereb, B., Chaumot, A., Geffard, A., Biagianti, S., Noël, C., Abbaci, K., Garric, J., Charmantier, G., and Charmantier-Daures, M. (2010). Ovarian cycle and embryonic development in *Gammarus fossarum*: application for reproductive toxicity assessment. *Environ. Toxicol. Chem.* *29*, 2249–2259.
196. Schirling, M., Triebkorn, R., and Köhler, H.-R. (2004). Variation in stress protein levels (hsp70 and hsp90) in relation to oocyte development in *Gammarus fossarum* (Koch 1835). *Invertebr. Reprod. Dev.* *45*, 161–167.
197. Jubeaux, G., Simon, R., Salvador, A., Quéau, H., Chaumot, A., and Geffard, O. (2012). Vitellogenin-like proteins in the freshwater amphipod *Gammarus fossarum* (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. *Aquat. Toxicol.* *112–113*, 72–82.
198. Fu, T., Knittelfelder, O., Geffard, O., Clément, Y., Testet, E., Elie, N., Touboul, D., Abbaci, K., Shevchenko, A., Lemoine, J., *et al.* (2021). Shotgun lipidomics and mass spectrometry imaging unveil diversity and dynamics in *Gammarus fossarum* lipid composition. *iScience* *24*, 102115.
199. Calzolari, L., Ansorge, W., Calabrese, E., Denslow, N., Part, P., and Lettieri, T. (2007). Transcriptomics and proteomics. Applications to ecotoxicology. *Comp. Biochem. Physiol. Part D Genomics Proteomics* *2*, 245–249.
200. Monsinjon, T., and Knigge, T. (2007). Proteomic applications in ecotoxicology. *Proteomics* *7*, 2997–3009.
201. Lemos, M.F.L., Soares, A.M.V.M., Correia, A.C., and Esteves, A.C. (2010). Proteins in ecotoxicology - how, why and why not? *Proteomics* *10*, 873–887.
202. Sanchez, B.C., Ralston-Hooper, K., and Sepúlveda, M.S. (2011). Review of recent proteomic applications in aquatic toxicology. *Environ. Toxicol. Chem.* *30*, 274–282.
203. Shepard, J.L., Olsson, B., Tedengren, M., and Bradley, B.P. (2000). Protein expression signatures identified in *Mytilus edulis* exposed to PCBs, copper and salinity stress. *Mar. Environ. Res.* *50*, 337–340.



204. Gouveia, D.D. (2017). Approches moléculaires pour la découverte, le développement et l'application de biomarqueurs de toxicité chez les gammaridés .
205. Unlü, M., Morgan, M.E., and Minden, J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–2077.
206. Borgatta, M., Hernandez, C., Decosterd, L.A., Chèvre, N., and Waridel, P. (2015). Shotgun ecotoxicoproteomics of *Daphnia pulex*: biochemical effects of the anticancer drug tamoxifen. *J. Proteome Res.* **14**, 279–291.
207. Jobling, S., and Tyler, C.R. (2003). Endocrine disruption in wild freshwater fish. *Pure Appl. Chem.* **75**, 2219–2234.
208. Chen, Y.-T., Chen, H.-W., Wu, C.-F., Chu, L.J., Chiang, W.-F., Wu, C.-C., Yu, J.-S., Tsai, C.-H., Liang, K.-H., Chang, Y.-S., *et al.* (2017). Development of a Multiplexed Liquid Chromatography Multiple-Reaction-Monitoring Mass Spectrometry (LC-MRM/MS) Method for Evaluation of Salivary Proteins as Oral Cancer Biomarkers. *Mol. Cell. Proteomics* **16**, 799–811.
209. Hoofnagle, A.N., Becker, J.O., Oda, M.N., Cavigliolo, G., Mayer, P., and Vaisar, T. (2012). Multiple-reaction monitoring-mass spectrometric assays can accurately measure the relative protein abundance in complex mixtures. *Clin. Chem.* **58**, 777–781.
210. Trapp, J., Armengaud, J., Pible, O., Gaillard, J.-C., Abbaci, K., Habtoul, Y., Chaumot, A., and Geffard, O. (2015). Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors. *J. Proteome Res.* **14**, 292–303.
211. Trapp, J., Gouveia, D., Almunia, C., Pible, O., Degli Esposti, D., Gaillard, J.-C., Chaumot, A., Geffard, O., and Armengaud, J. (2018). Digging Deeper Into the Pyriproxyfen-Response of the Amphipod *Gammarus fossarum* With a Next-Generation Ultra-High-Field Orbitrap Analyser: New Perspectives for Environmental Toxicoproteomics. *Front. Environ. Sci.* **6**.
212. Gismondi, E., Thomé, J.P., Urien, N., Uher, E., Baiwir, D., Mazzucchelli, G., De Pauw, E., Fechner, L.C., and Lebrun, J.D. (2017). Ecotoxicoproteomic assessment of the functional alterations caused by chronic metallic exposures in gammarids. *Environ. Pollut.* **225**, 428–438.
213. Simon, R., Jubeaux, G., Chaumot, A., Lemoine, J., Geffard, O., and Salvador, A. (2010). Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (*Gammarus fossarum*). *J. Chromatogr. A* **1217**, 5109–5115.
214. Sumpter, J.P., and Jobling, S. (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environ. Health Perspect.* **103 Suppl 7**, 173–178.
215. Jubeaux, G., Simon, R., Salvador, A., Lopes, C., Lacaze, E., Quéau, H., Chaumot, A., and Geffard, O. (2012). Vitellogenin-like protein measurement in caged *Gammarus fossarum* males as a biomarker of endocrine disruptor exposure: inconclusive experience. *Aquat. Toxicol.* **122–123**, 9–18.
216. Gouveia, D., Chaumot, A., Charlot, A., Almunia, C., François, A., Navarro, L., Armengaud, J., Salvador, A., and Geffard, O. (2017). Ecotoxicoproteomics for Aquatic Environmental Monitoring: First in Situ Application of a New Proteomics-Based Multibiomarker Assay Using Caged Amphipods. *Environ. Sci. Technol.* **51**, 13417–13426.
217. Jubeaux, G., Audouard-Combe, F., Simon, R., Tutundjian, R., Salvador, A., Geffard, O., and Chaumot, A. (2012). Vitellogenin-like proteins among invertebrate species diversity: potential of proteomic mass spectrometry for biomarker development. *Environ. Sci. Technol.* **46**, 6315–6323.
218. Nesvizhskii, A.I. (2014). Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125.
219. Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., and Hartmann, E.M. (2014). Non-model organisms, a species endangered by proteogenomics. *J. Proteomics* **105**, 5–18.
220. Trapp, J., Geffard, O., Imbert, G., Gaillard, J.-C., Davin, A.-H., Chaumot, A., and Armengaud, J. (2014). Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods. *Mol. Cell. Proteomics* **13**, 3612–3625.



221. Trapp, J., Armengaud, J., Gaillard, J.-C., Pible, O., Chaumot, A., and Geffard, O. (2016). High-throughput proteome dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins diversity in the crustacean *Gammarus fossarum*. *J. Proteomics* *146*, 207–214.
222. Gouveia, D., Chaumot, A., Charnot, A., Queau, H., Armengaud, J., Almunia, C., Salvador, A., and Geffard, O. (2017). Assessing the relevance of a multiplexed methodology for proteomic biomarker measurement in the invertebrate species *Gammarus fossarum*: A physiological and ecotoxicological study. *Aquat. Toxicol.* *190*, 199–209.
223. Dreier, D.A., Bowden, J.A., Aristizabal-Henao, J.J., Denslow, N.D., and Martyniuk, C.J. (2020). Ecotoxicolipidomics: An emerging concept to understand chemical-metabolic relationships in comparative fish models. *Comp. Biochem. Physiol. Part D Genomics Proteomics* *36*, 100742.
224. Melvin, S.D., Lanctôt, C.M., Doriean, N.J.C., Bennett, W.W., and Carroll, A.R. (2019). NMR-based lipidomics of fish from a metal(loid) contaminated wetland show differences consistent with effects on cellular membranes and energy storage. *Sci. Total Environ.* *654*, 284–291.
225. Greer, J.B., Magnuson, J.T., Hester, K., Giroux, M., Pope, C., Anderson, T., Liu, J., Dang, V., Denslow, N.D., and Schlenk, D. (2019). Effects of Chlorpyrifos on Cholinesterase and Serine Lipase Activities and Lipid Metabolism in Brains of Rainbow Trout (*Oncorhynchus mykiss*). *Toxicol. Sci.*
226. Marqueño, A., Blanco, M., Maceda-Veiga, A., and Porte, C. (2019). Skeletal muscle lipidomics as a new tool to determine altered lipid homeostasis in fish exposed to urban and industrial wastewaters. *Environ. Sci. Technol.* *53*, 8416–8425.
227. Arambourou, H., Fuertes, I., Vulliet, E., Daniele, G., Noury, P., Delorme, N., Abbaci, K., and Barata, C. (2018). Fenoxycarb exposure disrupted the reproductive success of the amphipod *Gammarus fossarum* with limited effects on the lipid profile. *PLoS ONE* *13*, e0196461.
228. Ayciriex, S., Djelti, F., Alves, S., Regazzetti, A., Gaudin, M., Varin, J., Langui, D., Bièche, I., Hudry, E., Dargère, D., *et al.* (2017). Neuronal Cholesterol Accumulation Induced by Cyp46a1 Down-Regulation in Mouse Hippocampus Disrupts Brain Lipid Homeostasis. *Front. Mol. Neurosci.* *10*, 211.
229. Carvalho, M., Sampaio, J.L., Palm, W., Brankatschk, M., Eaton, S., and Shevchenko, A. (2012). Effects of diet and development on the *Drosophila* lipidome. *Mol. Syst. Biol.* *8*, 600.
230. Guan, X.L., Cestra, G., Shui, G., Kuhrs, A., Schittenhelm, R.B., Hafen, E., van der Goot, F.G., Robinett, C.C., Gatti, M., Gonzalez-Gaitan, M., *et al.* (2013). Biochemical membrane lipidomics during *Drosophila* development. *Dev. Cell* *24*, 98–111.
231. Knittelfelder, O., Traikov, S., Vvedenskaya, O., Schuhmann, A., Segeletz, S., Shevchenko, A., and Shevchenko, A. (2018). Shotgun Lipidomics Combined with Laser Capture Microdissection: A Tool To Analyze Histological Zones in Cryosections of Tissues. *Anal. Chem.* *90*, 9868–9878.
232. Wang, Y., Hinz, S., Uckermann, O., Hönscheid, P., von Schönfels, W., Burmeister, G., Hendricks, A., Ackerman, J.M., Baretton, G.B., Hampe, J., *et al.* (2020). Shotgun lipidomics-based characterization of the landscape of lipid metabolism in colorectal cancer. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* *1865*, 158579.
233. Viant, M.R., and Sommer, U. (2013). Mass spectrometry based environmental metabolomics: a primer and review. *Metabolomics* *9*, 144–158.
234. Viant, M.R., Bundy, J.G., Pincetich, C.A., de Ropp, J.S., and Tjeerdema, R.S. (2005). NMR-derived developmental metabolic trajectories: an approach for visualizing the toxic actions of trichloroethylene during embryogenesis. *Metabolomics* *1*, 149–158.
235. Gómez-Canela, C., Miller, T.H., Bury, N.R., Tauler, R., and Barron, L.P. (2016). Targeted metabolomics of *Gammarus pulex* following controlled exposures to selected pharmaceuticals in water. *Sci. Total Environ.* *562*, 777–788.
236. Mouton, N. (2020). Développement d'une nouvelle approche par spectrométrie de masse ciblée pour la caractérisation de la virulence et des résistances aux antibiotiques chez les bactéries.
237. Charnot, A., Gouveia, D., Ayciriex, S., Lemoine, J., Armengaud, J., Almunia, C., Chaumot, A., Geffard, O., and Salvador, A. (2018). On-Line Solid Phase Extraction Liquid Chromatography-Mass Spectrometry Method for

- Multiplexed Proteins Quantitation in an Ecotoxicology Test Specie: *Gammarus fossarum*. *J. Appl. Bioanal.* **4**, 81–101.
238. Lafontaine, A., Baiwir, D., Joaquim-Justo, C., De Pauw, E., Lemoine, S., Boulangé-Lecomte, C., Forget-Leray, J., Thomé, J.-P., and Gismondi, E. (2017). Proteomic response of *Macrobrachium rosenbergii* hepatopancreas exposed to chlordecone: Identification of endocrine disruption biomarkers? *Ecotoxicol. Environ. Saf.* **141**, 306–314.
239. Schmidt, W., Rainville, L.-C., McEneff, G., Sheehan, D., and Quinn, B. (2014). A proteomic evaluation of the effects of the pharmaceuticals diclofenac and gemfibrozil on marine mussels (*Mytilus* spp.): evidence for chronic sublethal effects on stress-response proteins. *Drug Test. Anal.* **6**, 210–219.
240. Dowling, V.A., and Sheehan, D. (2006). Proteomics as a route to identification of toxicity targets in environmental toxicology. *Proteomics* **6**, 5597–5604.
241. Silvestre, F., Dierick, J.-F., Dumont, V., Dieu, M., Raes, M., and Devos, P. (2006). Differential protein expression profiles in anterior gills of *Eriocheir sinensis* during acclimation to cadmium. *Aquat. Toxicol.* **76**, 46–58.
242. Allen, T.E.H., Goodman, J.M., Gutsell, S., and Russell, P.J. (2014). Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chem. Res. Toxicol.* **27**, 2100–2112.
243. Hu, W., Tedesco, S., McDonagh, B., and Sheehan, D. (2010). Shotgun redox proteomics in sub-proteomes trapped on functionalised beads: Identification of proteins targeted by oxidative stress. *Mar. Environ. Res.* **69 Suppl**, S25-7.
244. Leprêtre, M., Almunia, C., Armengaud, J., Salvador, A., Geffard, A., and Palos-Ladeiro, M. (2019). The immune system of the freshwater zebra mussel, *Dreissena polymorpha*, decrypted by proteogenomics of hemocytes and plasma compartments. *J. Proteomics* **202**, 103366.
245. Martyniuk, C.J., Kroll, K.J., Doperalski, N.J., Barber, D.S., and Denslow, N.D. (2010). Environmentally relevant exposure to 17 $\alpha$ -ethinylestradiol affects the telencephalic proteome of male fathead minnows. *Aquat. Toxicol.* **98**, 344–353.
246. Ciliberti, A., Chaumot, A., Recoura-Massaquant, R., Chandesris, A., François, A., Coquery, M., Ferréol, M., and Geffard, O. (2017). Caged *Gammarus* as biomonitors identifying thresholds of toxic metal bioavailability that affect gammarid densities at the French national scale. *Water Res.* **118**, 131–140.
247. Alric, B., Geffard, O., Chandesris, A., Ferréol, M., François, A., Perceval, O., Piffady, J., Villeneuve, B., and Chaumot, A. (2019). Multisubstance Indicators Based on Caged *Gammarus* Bioaccumulation Reveal the Influence of Chemical Contamination on Stream Macroinvertebrate Abundances across France. *Environ. Sci. Technol.* **53**, 5906–5915.
248. Hartmann, E.M., Allain, F., Gaillard, J.-C., Pible, O., and Armengaud, J. (2014). Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol. Biol.* **1197**, 275–285.
249. Dupierris, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., and Bruley, C. (2009). A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* **25**, 1980–1981.
250. Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., *et al.* (2013). The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063-9.
251. Degli Esposti, D., Almunia, C., Guery, M.-A., Koenig, N., Armengaud, J., Chaumot, A., and Geffard, O. (2019). Co-expression network analysis identifies gonad- and embryo-associated protein modules in the sentinel species *Gammarus fossarum*. *Sci. Rep.* **9**, 7862.
252. Kammers, K., Cole, R.N., Tiengwe, C., and Ruczinski, I. (2015). Detecting significant changes in protein abundance. *EuPA Open Proteom.* **7**, 11–19.
253. Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3.

254. Lopes, C., Chaumot, A., Xuereb, B., Coulaud, R., Jubeaux, G., Quéau, H., François, A., and Geffard, O. (2020). In Situ Reproductive Bioassay with Caged *Gammarus fossarum* (Crustacea): Part 2-Evaluating the Relevance of Using a Molt Cycle Temperature-Dependent Model as a Reference to Assess Toxicity in Freshwater Monitoring. *Environ. Toxicol. Chem.* **39**, 678–691.
255. Percy, A.J., Chambers, A.G., Yang, J., and Borchers, C.H. (2013). Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics* **13**, 2202–2215.
256. Farkas, R., Danis, P., Medved'ová, L., Mechler, B.M., and Knopp, J. (2002). Regulation of cytosolic malate dehydrogenase by juvenile hormone in *Drosophila melanogaster*. *Cell Biochem. Biophys.* **37**, 37–52.
257. Duan, Y., Liu, P., Li, J., Wang, Y., Li, J., and Chen, P. (2014). Molecular responses of calreticulin gene to *Vibrio anguillarum* and WSSV challenge in the ridgetail white prawn *Exopalaemon carinicauda*. *Fish Shellfish Immunol.* **36**, 164–171.
258. Huang, Y., Hui, K., Jin, M., Yin, S., Wang, W., and Ren, Q. (2016). Two endoplasmic reticulum proteins (calnexin and calreticulin) are involved in innate immunity in Chinese mitten crab (*Eriocheir sinensis*). *Sci. Rep.* **6**, 27578.
259. Moshtaghi, A., Rahi, M.L., Mather, P.B., and Hurwood, D.A. (2017). Understanding the Genomic Basis of Adaptive Response to Variable Osmotic Niches in Freshwater Prawns: A Comparative Intraspecific RNA-Seq Analysis of *Macrobrachium australiense*. *J. Hered.* **108**, 544–552.
260. Groebe, K., Hayess, K., Klemm-Manns, M., Schwall, G., Wozny, W., Steemans, M., Peters, A.K., Sastri, C., Jaeckel, P., Stegmann, W., *et al.* (2010). Protein biomarkers for in vitro testing of embryotoxicity. *J. Proteome Res.* **9**, 5727–5738.
261. Muthukrishnan, S., Merzendorfer, H., Arakane, Y., and Yang, Q. (2019). Chitin organizing and modifying enzymes and proteins involved in remodeling of the insect cuticle. *Adv. Exp. Med. Biol.* **1142**, 83–114.
262. Coulaud, R., Geffard, O., Xuereb, B., Lacaze, E., Quéau, H., Garric, J., Charles, S., and Chaumot, A. (2011). In situ feeding assay with *Gammarus fossarum* (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. *Water Res.* **45**, 6417–6429.
263. Chang, C.-Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., Jovanovic, M., Aebersold, R., and Vitek, O. (2012). Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol. Cell. Proteomics* **11**, M111.014662.
264. Chung, L.M., Colangelo, C.M., and Zhao, H. (2014). Data Pre-Processing for Label-Free Multiple Reaction Monitoring (MRM) Experiments. *Biology (Basel)* **3**, 383–402.
265. Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193.
266. Smyth, G.K., and Speed, T. (2003). Normalization of cDNA microarray data. *Methods* **31**, 265–273.
267. Ballman, K.V., Grill, D.E., Oberg, A.L., and Therneau, T.M. (2004). Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **20**, 2778–2786.
268. Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15.
269. Hu, X., Du, S., Yu, J., Yang, X., Yang, C., Zhou, D., Wang, Q., Qin, S., Yan, X., He, L., *et al.* (2016). Common housekeeping proteins are upregulated in colorectal adenocarcinoma and hepatocellular carcinoma, making the total protein a better “housekeeper”. *Oncotarget* **7**, 66679–66688.
270. Ferguson, R.E., Carroll, H.P., Harris, A., Maher, E.R., Selby, P.J., and Banks, R.E. (2005). Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. *Proteomics* **5**, 566–571.
271. Leprêtre, M., Palos-Ladeiro, M., Faugere, J., Almunia, C., Lemoine, J., Armengaud, J., Geffard, A., and Salvador, A. (2020). From shotgun to targeted proteomics: rapid Scout-MRM assay development for monitoring potential immunomarkers in *Dreissena polymorpha*. *Anal. Bioanal. Chem.* **412**, 7333–7347.

272. Ovčáčíková, M., Lísa, M., Cífková, E., and Holčapek, M. (2016). Retention behavior of lipids in reversed-phase ultrahigh-performance liquid chromatography-electrospray ionization mass spectrometry. *J. Chromatogr. A* *1450*, 76–85.
273. Lin, J.-T., Snyder, L.R., and McKeon, T.A. (1998). Prediction of relative retention times of triacylglycerols in non-aqueous reversed-phase high-performance liquid chromatography. *Journal of Chromatography A* *808*, 43–49.
274. Tsugawa, H., Ikeda, K., Tanaka, W., Senoo, Y., Arita, M., and Arita, M. (2017). Comprehensive identification of sphingolipid species by in silico retention time and tandem mass spectral library. *J. Cheminform.* *9*, 19.
275. Holcapek, M., Jandera, P., Zderadicka, P., and Hrubá, L. (2003). Characterization of triacylglycerol and diacylglycerol composition of plant oils using high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *J. Chromatogr. A* *1010*, 195–215.
276. Lísa, M., and Holcapek, M. (2008). Triacylglycerols profiling in plant oils important in food industry, dietetics and cosmetics using high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *J. Chromatogr. A* *1198–1199*, 115–130.
277. LIPID MAPS LIPID MAPS® Structure Database (LMSD). lipidmaps. Available at: <https://www.lipidmaps.org/databases/lmsd/overview> [Accessed February 7, 2022].
278. Liebisch, G., Fahy, E., Aoki, J., Dennis, E.A., Durand, T., Ejsing, C.S., Fedorova, M., Feussner, I., Griffiths, W.J., Köfeler, H., *et al.* (2020). Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J. Lipid Res.* *61*, 1539–1555.
279. Shevchenko, A., and Simons, K. (2010). Lipidomics: coming to grips with lipid diversity. *Nat. Rev. Mol. Cell Biol.* *11*, 593–598.
280. Ivanova, P.T., Milne, S.B., Byrne, M.O., Xiang, Y., and Brown, H.A. (2007). Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry. In *Lipidomics and bioactive lipids: mass-spectrometry-based lipid analysis Methods in Enzymology*. (Elsevier), pp. 21–57.
281. Pi, J., Wu, X., and Feng, Y. (2016). Fragmentation patterns of five types of phospholipids by ultra-high-performance liquid chromatography electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Anal. Methods* *8*, 1319–1332.
282. Sullards, M.C., Liu, Y., Chen, Y., and Merrill, A.H. (2011). Analysis of mammalian sphingolipids by liquid chromatography tandem mass spectrometry (LC-MS/MS) and tissue imaging mass spectrometry (TIMS). *Biochim. Biophys. Acta* *1811*, 838–853.
283. Krank, J., Murphy, R.C., Barkley, R.M., Duchoslav, E., and McAnoy, A. (2007). Qualitative analysis and quantitative assessment of changes in neutral glycerol lipid molecular species within cells. *Meth. Enzymol.* *432*, 1–20.
284. Wörmer, L., Lipp, J.S., Schröder, J.M., and Hinrichs, K.-U. (2013). Application of two new LC-ESI-MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. *Org. Geochem.* *59*, 10–21.
285. Abreu, S., Solgadi, A., and Chaminade, P. (2017). Optimization of normal phase chromatographic conditions for lipid analysis and comparison of associated detection techniques. *J. Chromatogr. A* *1514*, 54–71.
286. Peng, B., Weintraub, S.T., Coman, C., Ponnaiyan, S., Sharma, R., Tews, B., Winter, D., and Ahrends, R. (2017). A Comprehensive High-Resolution Targeted Workflow for the Deep Profiling of Sphingolipids. *Anal. Chem.* *89*, 12480–12487.
287. Furukawa, T., Fuda, H., Miyanaga, S., Watanabe, C., Chiba, H., and Hui, S.-P. (2016). Rapid tin-mediated access to a lysophosphatidylethanolamine (LPE) library: Application to positional LC/MS analysis for hepatic LPEs in non-alcoholic steatohepatitis model mice. *Chem. Phys. Lipids* *200*, 133–138.
288. Fang, N., Yu, S., and Badger, T.M. (2003). LC-MS/MS analysis of lysophospholipids associated with soy protein isolate. *J. Agric. Food Chem.* *51*, 6676–6682.

289. Wang, M., Wang, C., and Han, X. (2016). Selection of internal standards for accurate quantification of complex lipid species in biological extracts by electrospray ionization mass spectrometry-What, how and why? *Mass Spectrom. Rev.*
290. Nelson, D.L., and Cox, M.M. (2008). *Lehninger Principles of Biochemistry* 5th ed. (New York: W. H. Freeman).
291. Hodson, L., Skeaff, C.M., and Fielding, B.A. (2008). Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog. Lipid Res.* *47*, 348–380.
292. Khaw, K.-T., Friesen, M.D., Riboli, E., Luben, R., and Wareham, N. (2012). Plasma phospholipid fatty acid concentration and incident coronary heart disease in men and women: the EPIC-Norfolk prospective study. *PLoS Med.* *9*, e1001255.
293. Or-Rashid, M.M., Odongo, N.E., and McBride, B.W. (2007). Fatty acid composition of ruminal bacteria and protozoa, with emphasis on conjugated linoleic acid, vaccenic acid, and odd-chain and branched-chain fatty acids. *J. Anim. Sci.* *85*, 1228–1234.
294. Zhukova, N. (1995). Fatty acid composition of 15 species of marine microalgae. *Phytochemistry* *39*, 351–356.
295. Tanakol, R., Yazici, Z., Sener, E., and Sencer, E. (1999). Fatty acid composition of 19 species of fish from the Black Sea and the Marmara Sea. *Lipids* *34*, 291–297.
296. Prato, E., and Biondolino, F. (2012). Total lipid content and fatty acid composition of commercially important fish species from the Mediterranean, Mar Grande Sea. *Food Chem.* *131*, 1233–1239.
297. Rocquelin, G., Guenot, L., Justrabo, E., Grynberg, A., and David, M. (1985). Fatty acid composition of human heart phospholipids: data from 53 biopsy specimens. *J. Mol. Cell. Cardiol.* *17*, 769–773.
298. Nussbaum, J.L., Neskovic, N., and Mandel, P. (1971). The fatty acid composition of phospholipids and glycolipids in Jimpy mouse brain. *J. Neurochem.* *18*, 1529–1543.
299. Tarazona, P., Feussner, K., and Feussner, I. (2015). An enhanced plant lipidomics method based on multiplexed liquid chromatography-mass spectrometry reveals additional insights into cold- and drought-induced membrane remodeling. *Plant J.* *84*, 621–633.
300. Huynh, K., Barlow, C.K., Jayawardana, K.S., Weir, J.M., Mellett, N.A., Cinel, M., Magliano, D.J., Shaw, J.E., Drew, B.G., and Meikle, P.J. (2019). High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem. Biol.* *26*, 71-84.e4.
301. Lange, M., and Fedorova, M. (2020). Evaluation of lipid quantification accuracy using HILIC and RPLC MS on the example of NIST® SRM® 1950 metabolites in human plasma. *Anal. Bioanal. Chem.* *412*, 3573–3584.
302. Krautbauer, S., Büchler, C., and Liebisch, G. (2016). Relevance in the Use of Appropriate Internal Standards for Accurate Quantification Using LC-MS/MS: Tauro-Conjugated Bile Acids as an Example. *Anal. Chem.* *88*, 10957–10961.
303. Kauhanen, D., Sysi-Aho, M., Koistinen, K.M., Laaksonen, R., Sinisalo, J., and Ekroos, K. (2016). Development and validation of a high-throughput LC-MS/MS assay for routine measurement of molecular ceramides. *Anal. Bioanal. Chem.* *408*, 3475–3483.
304. Kim, H.M., and Kang, J.S. (2021). Metabolomic studies for the evaluation of toxicity induced by environmental toxicants on model organisms. *Metabolites* *11*.
305. Cavill, R., Jennen, D., Kleinjans, J., and Briedé, J.J. (2016). Transcriptomic and metabolomic data integration. *Brief. Bioinformatics* *17*, 891–901.
306. Santos, E.M., Ball, J.S., Williams, T.D., Wu, H., Ortega, F., van Aerle, R., Katsiadaki, I., Falciani, F., Viant, M.R., Chipman, J.K., *et al.* (2010). Identifying health impacts of exposure to copper using transcriptomics and metabolomics in a fish model. *Environ. Sci. Technol.* *44*, 820–826.
307. Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., and Ralser, M. (2017). Designing and interpreting “multi-omic” experiments that may change our understanding of biology. *Current Opinion in Systems Biology* *6*, 37–45.



308. Cogne, Y., Degli-Esposti, D., Pible, O., Gouveia, D., François, A., Bouchez, O., Eché, C., Ford, A., Geffard, O., Armengaud, J., *et al.* (2019). De novo transcriptomes of 14 gammarid individuals for proteogenomic analysis of seven taxonomic groups. *Sci. Data* 6, 184.
309. Fu, T., Oetjen, J., Chapelle, M., Verdu, A., Szesny, M., Chaumot, A., Degli-Esposti, D., Geffard, O., Clément, Y., Salvador, A., *et al.* (2020). In situ isobaric lipid mapping by MALDI-ion mobility separation-mass spectrometry imaging. *J. Mass Spectrom.* 55, e4531.
310. Liang, X., Martyniuk, C.J., and Simmons, D.B.D. (2020). Are we forgetting the “proteomics” in multi-omics ecotoxicology? *Comp. Biochem. Physiol. Part D Genomics Proteomics* 36, 100751.
311. Ebner, J.N. (2021). Trends in the application of “omics” to ecotoxicology and stress ecology. *Genes (Basel)* 12.
312. Simões, T., Novais, S.C., Natal-da-Luz, T., Devreese, B., de Boer, T., Roelofs, D., Sousa, J.P., van Straalen, N.M., and Lemos, M.F.L. (2018). An integrative omics approach to unravel toxicity mechanisms of environmental chemicals: effects of a formulated herbicide. *Sci. Rep.* 8, 11376.
313. Guo, H., Wang, L., Deng, Y., and Ye, J. (2021). Novel perspectives of environmental proteomics. *Sci. Total Environ.* 788, 147588.
314. Armengaud, J. (2016). Next-generation proteomics faces new challenges in environmental biotechnology. *Curr. Opin. Biotechnol.* 38, 174–182.
315. Koenig, N., Almunia, C., Bonnal-Conduzorgues, A., Armengaud, J., Chaumot, A., Geffard, O., and Esposti, D.D. (2021). Co-expression network analysis identifies novel molecular pathways associated with cadmium and pyriproxyfen testicular toxicity in *Gammarus fossarum*. *Aquat. Toxicol.* 235, 105816.
316. Gouveia, D., Cogne, Y., Gaillard, J.-C., Almunia, C., Pible, O., François, A., Degli-Esposti, D., Geffard, O., Chaumot, A., and Armengaud, J. (2019). Shotgun proteomics datasets acquired on *Gammarus pulex* animals sampled from the wild. *Data Brief* 27, 104650.
317. Cogne, Y., Almunia, C., Gouveia, D., Pible, O., François, A., Degli-Esposti, D., Geffard, O., Armengaud, J., and Chaumot, A. (2019). Comparative proteomics in the wild: Accounting for intrapopulation variability improves describing proteome response in a *Gammarus pulex* field population exposed to cadmium. *Aquat. Toxicol.* 214, 105244.
318. Cosio, C., Degli-Esposti, D., Almunia, C., Gaillet, V., Sartelet, H., Armengaud, J., Chaumot, A., Geffard, O., and Geffard, A. (2021). Subcellular Distribution of Dietary Methyl-Mercury in *Gammarus fossarum* and Its Impact on the Amphipod Proteome. *Environ. Sci. Technol.* 55, 10514–10523.
319. Arambourou, H., Vulliet, E., Daniele, G., Noury, P., Delorme, N., Abbaci, K., Forcellini, M., Tutundjian, R., and Barata, C. (2019). Comparison in the response of three European Gammarid species exposed to the growth regulator insecticide fenoxycarb. *Environ. Sci. Pollut. Res. Int.* 26, 11496–11502.
320. Bligh, E.G., and Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37.
321. Borgmeyer, M., Coman, C., Has, C., Schött, H.-F., Li, T., Westhoff, P., Cheung, Y.F.H., Hoffmann, N., Yuanxiang, P., Behnisch, T., *et al.* (2021). Multiomics of synaptic junctions reveals altered lipid metabolism and signaling following environmental enrichment. *Cell Rep.* 37, 109797.
322. Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13, e1005752.
323. Karpievitch, Y.V., Taverner, T., Adkins, J.N., Callister, S.J., Anderson, G.A., Smith, R.D., and Dabney, A.R. (2009). Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. *Bioinformatics* 25, 2573–2580.
324. Alava, V.R., Qunitio, E.T., de Pedro, J.B., Priolo, F.M.P., Orozco, Z.G.A., and Wille, M. (2007). Lipids and fatty acids in wild and pond-reared mud crab *Scylla serrata* (Forsskål) during ovarian maturation and spawning. *Aquac. Res.* 38, 1468–1477.
325. Lee, R.F., and Walker, A. (1995). Lipovitellin and lipid droplet accumulation in oocytes during ovarian maturation in the blue crab, *Callinectes sapidus*. *J. Exp. Zool.* 271, 401–412.



326. Ravid, T., Tietz, A., Khayat, M., Boehm, E., Michelis, R., and Lubzens, E. (1999). Lipid accumulation in the ovaries of a marine shrimp *penaeus semisulcatus* (de haan). *J. Exp. Biol.* *202* (Pt 13), 1819–1829.
327. Subramoniam, T. (2011). Mechanisms and control of vitellogenesis in crustaceans. *Fish. Sci.* *77*, 1–21.
328. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR : An R package for multivariate analysis. *J. Stat. Softw.* *25*.
329. Buré, C., Ayciriex, S., Testet, E., and Schmitter, J.-M. (2013). A single run LC-MS/MS method for phospholipidomics. *Anal. Bioanal. Chem.* *405*, 203–213.
330. Beuvier, L. (2015). Développement d'une méthode de séparation chromatographique couplée aux spectrométries de masse à source d'ionisation électrospray (ESI-MS) et à source plasma à couplage inductif (ICP-MS): application à l'analyse de spéciation des lanthanides.
331. Murphy, R.E., Schure, M.R., and Foley, J.P. (1998). Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography. *Anal. Chem.* *70*, 1585–1594.