



Développement d'approches protéomiques pour l'identification, la détection spécifique et la quantification de pathogènes

Liz Paola Cantero Mendieta

► To cite this version:

Liz Paola Cantero Mendieta. Développement d'approches protéomiques pour l'identification, la détection spécifique et la quantification de pathogènes. Chimie analytique. Université de Strasbourg, 2020. Français. NNT : 2020STRAF048 . tel-03636477

HAL Id: tel-03636477

<https://theses.hal.science/tel-03636477>

Submitted on 10 Apr 2022

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.

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

UMR 7178

THÈSE présentée par :

Liz Paola CANTERO MENDIETA

soutenue le **20 octobre 2020**

pour obtenir le grade de **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Chimie Analytique

Développement d'approches protéomiques pour l'identification, la détection spécifique et la quantification de pathogènes

THÈSE dirigée par :

Mme SABATIER Laurence

Professeur, Université de Strasbourg

RAPPORTEURS :

Mme PFLIEGER Delphine

Chargée de recherche, CNRS, Grenoble

M. BECHER François

Chercheur, CEA, Gif sur Yvette

AUTRES MEMBRES DU JURY :

M. MARIN Philippe

Directeur de recherche, CNRS, Montpellier

M. HANSMANN Yves

Professeur, Université de Strasbourg

*A las personas que amo,
Mes enfants Valentina & Rodrigo
Mon mari
Mis padres*

« Les efforts de l'intelligence et les affections du cœur sont pour nous la meilleure source d'où le bonheur puisse jaillir. Faire une découverte utile, obtenir un succès littéraire ou scientifique, sauver son semblable, affranchir son pays ; voilà des bonheurs qui n'incombent pas à tout le monde et que tout le monde envie. Et pourquoi ? Parce qu'ils naissent de la conscience d'avoir bien fait. »

Louis-Auguste Martin

Remerciements

Premièrement je tiens à remercier à **Laurence Sabatier**, pour votre encadrement, votre soutien et pour m'avoir accordé votre confiance pendant ces années de collaboration. Ah et aussi merci beaucoup pour votre patience et pour vos nombreux conseils (scientifiques ou personnels). Comme je vous l'ai déjà dit, je suis venue au labo suite à la passion que vous montriez dans vos cours et j'avais envie de travailler avec vous, et c'était un vrai plaisir « la plupart de temps » ☺.

Je souhaite également remercier **Sarah Cianferani** pour m'avoir donné la possibilité de faire partie du LSMBO, également pour les conditions idéales de travail au sein du laboratoire et la bonne volonté, l'écoute et la joie que vous transmettiez qui font que le labo soit un endroit plaisant pour le travail.

Je remercie également **Delphine Pflieger**, **François Becher**, **Philippe Marin** et **Yves Hansmann** d'avoir accepté d'évaluer ces travaux de thèse.

Un grand merci à l'ensemble des collaborateurs avec qui j'ai pu interagir. A **Nathalie Boulanger**, pour votre disponibilité et pour m'avoir transmis une partie de vos connaissances au sujet des maladies vectorielles. **Isabelle Schalk** et **Quentin Perraud** pour votre collaboration sur le sujet de *P. aeruginosa*. **Quentin** ! la personne la plus silencieuse et surprenante que j'ai pu connaître, avec une passion énorme pour son sujet de thèse ! Merci d'avoir fait que les longues manip *Pseudo* soient agréables et de m'avoir fait part de tes connaissances en bio.....

Aux personnes avec qui j'ai eu l'opportunité de travailler, **Ludivine**, **Vildan** et **Josiane**.

Aux membres du LSMBO, pour ces longues années vécues avec vous ! Aux bureaux des filles « mon bureau », **Juju** !! merci pour tous les fous rires que nous avons eus, les chansons chantées même si parfois tu aimes bien mettre « la musique dans la poche » ☺ ! **Leslie**...j'aurais fait quoi sans tes corrections de français !! orales et écrites..., la fin est arrivée.... et il y a encore du boulot mais je compte bien t'embêter même du Paraguay.... Ah **Juju** les mangues, fruits de la passion, ananas... t'attendront aussi ☺ **Benoît**... tu n'es pas resté longtemps au bureau « des filles » mais tu rigolais bien avec nos histoires ! A vous trois, je vous remercie de m'avoir transmis vos connaissances, et les filles un grand merci pour votre aide scientifique. Marie ! tu fais toujours partie de notre bureau, même si tu es partie !!! merci pour les petits trophées apportés (d'ailleurs, je veux le mien) et pour ton aide dans les stats... **Chloé**, arrivée un peu plus tard, merci de partager avec moi « les soirs » au labo... et

très bonne continuation... **Bastien**, bienvenue dans le team Lyme...et bonne continuation pour la suite.

Le groupe **CC, Joanna...** une personne pleine de surprises, que je remercie énormément pour les moments partagés... et jojo « lorsque tu veux vraiment une chose, tout l'univers conspire à te permettre de réaliser ton désir » !! Le petit **Nico...** je t'ai enregistré avec ce nom partout !! mail, tel... merci pour les divers échanges et bon courage tu es le suivant à quitter le LSMBO... **Aurélie**, tu es tellement gentille, tu es toujours prête à aider tout le monde et disponible quand nous en avons besoin, marrante je te reçois avec plaisir au Paraguay si tu veux venir mais il faudra que tu apportes ton pot de nutella, ça n'existe pas là-bas...**Steve** !! nous avons fait le master ensemble, mais c'est au labo que j'ai appris à te connaître !! enfin je crois... parce que je n'arrive toujours pas à savoir quand tu blagues ou pas... « toujours sérieux » bonne continuation !!!

Magali et **Christine** !!! un grand merci pour l'aide que vous m'avez apportée au niveau scientifique, toutes les deux m'ont toujours donné une réponse à mes diverses questions ! et un grand merci pour votre aide et conseils pour le nouvel univers que j'ai connu pendant ma thèse, celui « d'être maman ». **Alex** et **Patrick** !!!! l'équipe informatique, qui sûrement va se reposer avec mon départ. **Patrick** merci pour les diverses banques de données et **Alex** !!! tu étais toujours dispo pour mes divers problèmes, merci pour ta patience et disponibilité.

Au labo R5, il n'y a pas une personne à qui je ne suis pas allée demander de l'aide pendant ces 4 ans de thèse !!! à chacun d'entre vous merci... ! **Jean Marc**... indispensable !!! et toujours dispo pour mes problème de spectro ! **Agnès**... merci pour ton aide avec mon HPLC ! **Hélène**, toujours dispo pour des questions « qualité » ou Triple Tof... Les garçons, **Stéphane**, **Max**, **Thomas**, **Antho**, merci pour les bons moments passés pendant les repas de midi. **Alfred** !!! hopla ... toujours dispo et dans la bonne humeur ! **François**, **Christine**, les deux **Fabrice**, **Oscar** c'était un plaisir d'échanger avec vous ! **Blandine**, bientôt nous serons à nouveau plus proches, au cas où t'aurais besoin d'une maman temporaire... c'était un plaisir de partager des moments avec toi prési... Pour ceux qui vont rester encore un petit moment. **Evolène**, **Charlotte**, **Marie** et **Marie**... bonne continuation !!!

En dernier je voudrais remercier mes enfants, **Rodrigo** pour sa patience pendant ces années d'études, ma petite **Valentina** pour la force qu'elle me donne avec ses pleurs et ses rires, **Mon mari** pour m'avoir soutenue dans toutes mes aventures, d'être toujours présent dans toutes les situations, et pour me faire rire même dans les pires situations !!! à **mes parents** même au loin, vous étiez toujours présents pour moi !!! (a **mis padres**, mismo a la distancia siempre estuvieron presentes para mi

Sommaire

Remerciements

Sommaire	7
Liste des abréviations	15
Publications et communications.....	21
Introduction générale	23
Chapitre I. Stratégies et outils pour l'analyse protéomique	25
I- Identification des protéines	25
1. Préparation d'échantillon.....	26
1.1. Extraction des protéines.....	26
1.2. Prétraitement des protéines.....	27
1.3. Digestion des protéines.....	31
2. Analyse par LC-MS/MS.....	32
2.1. Séparation des peptides par chromatographie liquide	32
2.2. Spectrométrie de masse en tandem	33
2.2.1. Différents modes d'acquisition	34
2.2.1.1. Modes d'acquisition non ciblée	34
2.2.1.2. Modes d'acquisition ciblée	36
2.2.2. Fragmentation des peptides	38
2.2.3. Nature des spectromètres de masse.....	39
3. Traitements des données et identification des protéines	40
3.1.1. Banques protéiques	41
3.1.2. Interprétation des données	43
3.1.3. Validation des identifications	46
II- Protéomique ciblée.....	47
1. Analyse LC-SRM.....	47
1.1. Principe de l'analyse LC-SRM	47
1.2. Stratégies de quantification par LC-SRM	49

1.3.	Développement d'une méthode LC-SRM.....	51
1.3.1.	Sélection des protéines et peptides.....	51
1.3.2.	Sélection, validation et optimisation des transitions.....	52
2.	Analyse LC-PRM.....	55
2.1.	Principe de l'analyse LC-PRM	55
2.2.	Développement d'une méthode LC-PRM.....	57
3.	Analyse DIA.....	59
4.	Traitement des données méthodes ciblées (SRM et PRM)	60
4.1.	Validation de la détection d'un peptide	60
4.2.	Quantification absolue.....	63
III-	Quantification relative des protéines.....	64
1.	Analyse différentielle avec marquage.....	65
2.	Analyse différentielle sans marquage.....	67
2.1.	Comptage des spectres MS/MS	67
2.2.	Extraction des courants d'ions	68
	Chapitre II. Application pour l'étude de la borréliose de Lyme.	71
I-	Borréliose de Lyme	71
1.	Epidémiologie	71
2.	Manifestations cliniques	73
2.1.	Phase précoce localisée	73
2.2.	Phase précoce disséminée	73
2.3.	Phase tardive disséminée	74
3.	L'interface cutanée	75
4.	Prévention et vaccination.	76
5.	Diagnostic et traitement.	78
6.	Objectifs.....	81
II-	Approche vaccinale	81
1.	Analyse non ciblée des protéines présentes dans les peaux de chiens infectés ...	82
1.1.	Echantillons	82

1.2.	Stratégie analytique.....	84
1.2.1.	Extraction des peaux de chiens et dosage des protéines.....	84
1.2.2.	Préfractionnement sur gel SDS-PAGE.....	85
1.2.3.	Analyse LC-MS/MS et bases de données.....	85
1.3.	Résultats d'identification des protéines	87
2.	Analyse ciblée des candidats vaccinaux dans les peaux de chiens infectés	90
2.1.	Développement de la méthode LC-SRM.....	91
2.1.1.	Sélection des protéines	91
2.1.2.	Sélection des peptides.....	91
2.1.3.	Sélection, optimisation et validation des transitions.....	92
2.2.	Application aux peaux de chiens infectés.....	96
3.	Production des protéines recombinantes	99
4.	Essais vaccinaux chez le chien	100
III-	Approche diagnostique	101
1.	Diagnostic de la phase précoce de la maladie de Lyme	101
1.1.	Stratégie analytique.....	102
1.2.	Analyse par protéomique ciblée.....	102
1.2.1.	Description de la méthode LC-SRM.....	102
1.2.2.	Analyse de la cohorte de patients : problèmes rencontrés.....	103
1.2.2.1.	Problèmes de décalage de Tr	104
1.2.2.2.	Problèmes de sensibilité	109
1.2.2.3.	Problèmes d'échantillonnage	116
1.2.3.	Analyse de la cohorte de patients : résultats	117
1.3.	Stratégies alternatives.....	129
1.3.1.	Préparation tube-gel	129
1.3.2.	Mode DIA	131
2.	Diagnostic de la phase tardive disséminée de la maladie de Lyme	133
2.1.	Développement et optimisation de la méthode ciblée.....	134
2.2.	Analyse de peaux de souris 40 jours après une infection.....	136

2.3.	Modifications de la méthode ciblée.....	139
2.4.	Comparaison de deux dermocorticoïdes.....	142
2.5.	Effet de la lidocaïne.....	144
IV-	Conclusion.....	145
Chapitre III. Etude des agents pathogènes transmis par des tiques.....		147
I-	Maladies vectorielles à tiques : acquisition et transmission de pathogènes.....	147
1.	Vecteur tique	147
1.1.	Cycle de vie de <i>I. ricinus</i>	148
1.2.	Interactions hôte-tique-pathogènes	149
1.3.	Répartition géographique et habitat.....	151
2.	Pathogènes transmis à l'hôte et maladies associées.....	152
II-	Objectifs.....	155
III-	Stratégie analytique d'identification des protéines	156
1.	Nature des échantillons.....	156
2.	Stratégie analytique mise en œuvre	157
2.1.	Extraction et dosage des protéines	158
2.2.	Traitements bio-informatiques	159
2.2.1.	Création des banques de données.....	159
2.2.2.	Algorithmes de recherche dans les banques de données.....	163
2.2.3.	Actualisation des banques de données.....	165
IV-	Résultats d'identification des protéines de pathogènes	169
1.	Interprétation des données issues de la validation	169
2.	Analyse des tiques <i>I. ricinus</i>	172
3.	Analyse des peaux de souris	180
3.1.	Peaux de souris infectées expérimentalement.....	180
3.1.1.	<i>Anaplasma phagocytophilum</i>	181
3.1.2.	<i>Babesia microti</i>	183
3.2.	Peaux de souris infectées via des tiques collectées dans la nature	185
V-	Conclusion.....	187

Chapitre IV. Analyse protéomique de <i>Pseudomonas aeruginosa</i>	189
I- Introduction	189
II- Développements méthodologiques	191
1.1. Etude de faisabilité de la détection de protéines de <i>P. aeruginosa</i> dans des cellules épithéliales pulmonaires infectées	191
1.2. Optimisation du gradient chromatographique	193
1.3. Sélection de la banque de données et de l'échantillon biologique	194
1.4. Sélection du moteur de recherche	196
1.5. Optimisation du protocole de préparation d'échantillon	197
1.6. Choix des réplicats	199
III- Réalisation des études protéomiques différentielles	203
1. Adaptation phénotypique de <i>P. aeruginosa</i> à la présence d'exosidérophores	203
1.1. Contexte	203
1.2. Stratégie analytique	204
1.3. Suivi de la stabilité instrumentale	205
1.3.1. Contrôle qualité externe	206
1.3.2. Contrôle qualité interne	207
1.4. Résultats	208
2. Adaptation phénotypique de <i>P. aeruginosa</i> à la présence de différentes conditions	216
2.1. Contexte	216
2.2. Stratégie analytique	216
2.3. Résultats	217
IV- Conclusion	221
Conclusion générale	223
Partie Expérimentale	226
Chapitre II. Application pour l'étude de la borréliose de Lyme	226
1. Paragraphe II-1.1 : Peaux de chiens infectés	226
2. Paragraphe II-1.2.1 : Extraction des peaux de chiens et dosage des protéines	226
3. Paragraphe II-1.2.2: Préfractionnement sur gel SDS-PAGE	226

4.	Paragraphe II-1.2.3 Analyse LC-MS/MS et bases de données	227
5.	Paragraphe .II-2.2 Application aux peaux de chiens infectés	229
6.	Paragraphe .III-1 Diagnostic de la phase précoce de la maladie de Lyme	229
7.	Paragraphe III-1.2.1 : Description de la méthode LC-SRM.....	230
8.	Paragraphe .III-1.2.2.1 Problèmes de décalage de Tr	230
9.	Paragraphe III-1.2.2.2 Problèmes de sensibilité.....	231
10.	Paragraphe III-1.3.1 Préparation de tube-gels (TG).....	232
11.	Paragraphe III-1.3.2 Mode DIA	233
12.	Paragraphe III-2 Diagnostic de la phase tardive disséminée de la maladie de Lyme	235
13.	Paragraphe III-2.2 Analyse de peaux de souris 40 jours après une infection ..	236
14.	Paragraphe III-2.3 Modifications de la méthode ciblée	236
15.	Paragraphe III-2.4 Comparaison de deux dermocorticoïdes	237
16.	Paragraphe III-2.5 Effet de la lidocaïne	238
	Chapitre III. Etude des agents pathogènes transmis par des tiques.....	239
17.	Paragraphe III-1 Nature des échantillons	239
18.	Paragraphe III-2 Extraction et dosage des protéines	239
19.	Paragraphe V-2- Résultats d'identification des protéines de pathogènes	240
20.	Paragraphe IV-3.1 Peaux de souris infectées expérimentalement.....	241
21.	Paragraphe IV-3.1.1 <i>Anaplasma phagocytophilum</i>	241
22.	paragraphe IV-3.1.2 <i>Babesia microti</i> et le paragraphe IV-3.2 Peaux de souris infectées via des tiques collectées dans la nature	242
	Chapitre IV. Analyse protéomique de <i>Pseudomonas aeruginosa</i>	242
23.	Paragraphe IV.II- Développements méthodologiques.....	242
24.	Paragraphe III- Réalisation des études protéomiques différentielles	245
	Bibliographie.....	248
	Annexes.....	264
	Annexe 1 : Liste des protéines et peptides utilisés dans les méthodes de quantification ciblée.....	264

Annexe 2 : Liste des protéines et peptides détectés dans les analyses LC-SRM et LC-PRM pour les patients de la cohorte (phase précoce de la borréliose de Lyme)	268
Annexe 3 : Tableau récapitulatif des banques de données utilisées dans le chapitre III .	270

Liste des abréviations

ACA : Acrodermatite chronique atrophiante

ACN: Acétonitrile

ADN: Acide désoxyribonucleéique

AGC : Automatic Gain Control

AQUA™: Absolute QUAntification

ARN: Acide ribonucléique

AUC: Aires sous la courbe

BOPs : Microbead-Based and Organic-media-assisted Protéolysis strategy

BSA : Albumine de sérum bovin

CAA : Casamino acid medium

C-C : Carbone-carbone

CFP: ChipFilter Proteolysis

CHC: Carcinome hépatocellulaire

CID: Collision Induced Dissociation

CQ: Contrôles qualité

CV : Coefficient of variation

CXCL13 : chimioattractant des lymphocytes B

DC : Détergent compatible

DDA : Data-Dependent Acquisition

DDBJ : DNA Data Bank of Japan

DIA : Data-Independent Acquisition

DS: Digestion en solution

DTT: Dithiothréitol

EC: Energie de collision

ECD: Electron Capture Dissociation

ELISA : Enzyme-linked immunosorbent assay

EM: Erythema Migrans

EMBL : European Molecular Biology Laboratory

EMM : Erythème migrant multiple

ESI: Ionisation électrospray

ETD: Electron-Transfer Dissociation

FA: Acide formique

FASP: Filter Aided Sample Preparation

FDR: False Discovery Rate

FLEXIQuant: Full-length expressed stable isotope-labelled proteins for quantification)

GPMDDB : Global Proteome Machine Data Base

HCD: Higher-energy Collision Dissociation

HILIC: Hydrophilic Interaction Liquid Chromatography

HR/AM : High Resolution/Accurate Mass

IAA: Iodoacétamide

IDA: Information-Dependent Acquisition

iRT: indexed Retention Time

iSRM : intelligent Selected Reaction Monitoring

iST : in-Stage Tip digestion

LB : Lysogeny Broth ou Luria-Bertani

LC-MS/MS : Chromatographie liquide couplée à la spectrométrie de masse en tandem

LLOQ: Low Limit Of Quantification

m/z: rapport masse sur charge

MeOH: Méthanol

MM: Masse Moléculaire

MRM : Multiple Reaction Monitoring

MRM-HR: Multiple Reaction Monitoring - High Resolution

MS: Spectrométrie de masse

MuDPIT: Multi-Dimensional Protein Identification Technology

NAT : natural peptide

NCBI: National Center for Biotechnology Information

OMSSA: Open Mass-Spectrometry Search Algorithm

Osp : Outer surface protein

PAGE : Electrophorèse sur gel de polyacrylamide

PBS : Phosphate buffered saline

PCR : Polymerase Chain Reaction

PDB : Protein Data Bank

PFF : Peptide Fragment Fingerprint

PIR : Protein Information Resource

PIVKA-II : Protein Induced by Vitamin K Absence

PrEST : Protein Epitope Signature Tag

PRIDE : PRoteome IDentification database

PRM: Parallel Reaction Monitoring

ProFI: Infrastructure nationale de protéomique

PSAQ™ : Protein Standard Absolute Quantification

PSM: Peptide Spectrum Match

PTMs: Modifications post-traductionnelle

Q: quadripôle

QconCAT : Quantification conCATamer

rdotp: Rdot product

RefSeq : Reference Sequence

RP: Phase inverse ou Reverse phase

RPMI : Roswell Park Memorial Institute medium

Salp15 : salivary protein 15

SAX: Chromatographie d'échange d'anions ou Strong Anion Exchanger

SCX : Chromatographie d'échange de cations ou Strong Cation Exchange

SDC: déoxycholate de sodium

SDS : Dodécylsulfate de sodium

SEC: Chromatographie d'exclusion stérique ou Size-exclusion chromatography

SG : Gel stacking ou de concentration

SIB: Institut Suisse de Bioinformatique

SIS: Stable Isotope-labeled Standard

SLSA : Structured least square adaptative

SP3 : Single-Pot Solid Phase-enhanced Sample Preparation

SPE : Extraction sur phase solide ou « Solid-Phase Extraction »

SRM : Selected Reaction Monitoring

STrap : Suspension Trapping

SWATH-MS: Sequential Windowed Acquisition of all THeoretical fragment ion Mass Spectra

TCA: Acide trichloroacétique

TFA: Acide trifluoroacétique

TG : Tube-Gel

TOF: Temps de vol ou Time of flight

TPA : Third Party Annotation

USP³ : Universal Solid-Phase Protein Preparation

UVPD : UltraViolet PhotoDissociation

WAX : Weak anion-exchange

WCX: weak-cation exchange chromatography

Publications et communications

Publications

Identification of Borrelia protein candidates in mouse skin for potential diagnosis of disseminated Lyme borreliosis

A. Grillon, B. Westermann, **P. Cantero**, B. Jaulhac, M. J. Voordouw, D. Kapps, E. Collin, C. Barthel, L. Ehret-Sabatier, N. Boulanger.

Scientific Reports, 2017, 7: 16719. DOI:10.1038/s41598-017-16749-9

Phenotypic adaption of Pseudomonas aeruginosa by hacking siderophores produced by other microorganisms

Q. Perraud, **P. Cantero**, B. Roche, V. Gasser, V. P. Normant, L. Kuhn, P. Hammann, G. L. A. Mislin, L. Ehret-Sabatier, I. J. Schalk.

Molecular and Cellular Proteomics, 2020, 19(4), 589-607. DOI:10.1074/mcp.RA119.001829

Effects of topical corticosteroids and lidocaine on Borrelia burgdorferi sensu lato in mouse skin: potential impact to human clinical trials

B. Lefeuvre, **P. Cantero**, L. Ehret-Sabatier, C. Lenormand, C. Barthel, C. Po, N. Parveen, A. Grillon, B. Jaulhac, N. Boulanger.

Scientific Reports 2020 10(1):10552. DOI: 10.1038/s41598-020-67440-5.

Proteomic as an Exploratory Approach to Develop Vaccines against Tick-Borne Diseases Using Lyme Borreliosis as a Test Case

E. Talagrand-Reboul, B. Westermann, M. A. Raess, G. Schnell, **P. Cantero**, C. Barthel, L. Ehret-Sabatier, B. Jaulhac, N. Boulanger.

Vaccines 2020 8(3):E463. DOI: 10.3390/vaccines8030463.

Phenotypical adaptation of Pseudomonas aeruginosa in the presence of siderophore-antibiotic conjugates during epithelial cell infection

Q. Perraud, **P. Cantero**, M. Munier, F. Hoegy, N. Zill, V. Gasser, G. L. A. Mislin, L. Ehret-Sabatier, I.J. Schalk. soumise à *PLOS Pathogens* le 25/08/2020.

Communications orales

Development of quantitative targeted proteomics for the diagnosis of Lyme borreliosis

P. Cantero, V. Gulsari, B. Jaulhac, N. Boulanger, Y. Hansmann, C. Lenormand, L. Zilliox, P. Kieffer, L. Ehret-Sabatier

Congrès de Spectrométrie de Masse et Analyse Protéomique (SMAP), Strasbourg, France, 16 au 19 septembre 2019

Communications par affiche

Proteomic approaches for the diagnosis of vector-borne diseases

P. Cantero, B. Westermann, A. Grillon, B. Jaulhac, M. Kazimirova, N. Boulanger, L. Sabatier
11th European Summer School- Advanced Proteomics, Brixen/Bressanone, Italie, 30 juillet au 5 août 2017

Proteomic approaches for the diagnosis of early and disseminated Lyme borreliosis

P. Cantero, B. Westermann, A. Grillon, L. Esteves-Gloria, B. Jaulhac, N. Boulanger, L. Ehret-Sabatier

35^{ème} Congrès de la Société Française d'Electrophorèse et Analyse Protéomique (SFEAP), Albi, France, 09 au 12 octobre 2018

Development of quantitative targeted proteomics for the diagnosis of Lyme borreliosis

P. Cantero, B. Jaulhac, N. Boulanger, Y. Hansmann, C. Lenormand, L. Zilliox, P. Kieffer, L. Ehret-Sabatier

XIII. Annual Congress of the European Proteomics Association: From Genes via proteins and their Interactions to Functions, Potsdam, Allemagne, 24 au 28 mars 2019

Proteomic approaches for the epidemiology and diagnosis of tick-borne diseases

P. Cantero, P. Boyer, J. Saade, N. Boulanger, B. Jaulhac, M. Kazimirova, J. de la Fuente, M. Jalovecka, L. Ehret-Sabatier

68th Conference on Mass Spectrometry and Allied Topics (ASMS), Houston, USA, 31 mai au 04 juin 2020

Introduction générale

La protéomique est la science qui étudie les protéines exprimées dans les organismes, et qui permet d'identifier, de détecter et de quantifier les protéines présentes dans une cellule, un tissu ou un fluide biologique à un instant donné. Une de ses applications est l'étude des protéines impliquées dans des pathologies, pour mieux comprendre ces pathologies, s'en protéger et les soigner.

Dans le cas des maladies infectieuses, de nombreuses problématiques existent telles que (i) les agents pathogènes multi-résistants, (ii) la propagation rapide de maladies émergentes, comme nous l'avons constaté avec la pandémie de la COVID-19 ou (iii) l'augmentation et la globalisation des maladies à transmission vectorielle, en lien avec les changements climatiques. Grâce aux progrès réalisés au niveau de la préparation de l'échantillon, des méthodes séparatives, de la spectrométrie de masse et de l'interprétation des données, la protéomique peut jouer un rôle clef pour la compréhension de ces maladies, ainsi que pour la conception de nouvelles stratégies vaccinales, diagnostiques ou thérapeutiques.

Mes travaux de thèse sont centrés sur le développement d'approches protéomiques pour étudier différents agents pathogènes en contexte d'infection, c'est-à-dire au sein d'un hôte infecté. Différents pathogènes ont été étudiés, ainsi que différents modèles d'infection, et les approches protéomiques mises en œuvre ont été, soit non ciblées pour étudier toutes les protéines présentes dans l'échantillon, soit ciblées sur quelques protéines d'intérêt.

Le manuscrit est structuré en quatre chapitres :

Le **premier chapitre** résume les stratégies actuelles dans le domaine de la protéomique, et décrit de manière détaillée les méthodes et outils utilisés pendant ma thèse.

Le **deuxième chapitre** est centré sur l'étude de la borréliose de Lyme, une maladie vectorielle à tique, et le développement de différentes approches protéomiques pour (i) identifier et valider des candidats-vaccin dans un modèle canin pour une application vétérinaire, (ii) valider une méthode alternative de diagnostic direct de la phase précoce de la borréliose de Lyme dans des biopsies cutanées de patients, (iii) développer une méthode de diagnostic de la phase tardive disséminée de la borréliose de Lyme, et (iv) envisager des alternatives et améliorations dans les schémas analytiques.

Le **troisième chapitre** s'intéresse aux autres pathogènes transmis par la tique. Des études protéomiques ont été réalisées sur des tiques et sur des peaux de souris infectées, afin d'identifier des marqueurs d'infection de ces différents pathogènes. Ces études ont notamment nécessité un travail important sur les banques de données qui sera détaillé dans ce chapitre.

Le **quatrième chapitre** est consacré à l'étude de *Pseudomonas aeruginosa*, bactérie opportuniste, ubiquitaire et résistance à de nombreux antibiotiques. Nous avons réalisé plusieurs études protéomiques de cette bactérie au sein de cellules épithéliales infectées, afin de mieux comprendre les voies d'acquisition du fer par *P. aeruginosa* et de pouvoir utiliser ces voies pour le développement de nouveaux antibiotiques.

La **partie expérimentale**, située en la fin du manuscrit, décrit les conditions opératoires des différentes manipulations

Chapitre I. Stratégies et outils pour l'analyse protéomique

L'analyse protéomique s'intéresse à l'identification, la caractérisation, la détection et la quantification des protéines contenues dans un échantillon biologique (tissus, cellules, fluides biologiques, cultures microbiennes, etc.). Il existe trois types d'approches (L. C. Gillet *et al.*, 2016) :

La stratégie **bottom-up**, qui consiste en l'étude des peptides issus de la digestion enzymatique pour remonter à la protéine (R. Matthiesen *et al.*, 2008).

La stratégie **top-down**, qui consiste en l'étude de protéines entières (dénaturées ou natives) et qui permet la caractérisation de différentes protéoformes (T. K. Toby *et al.*, 2016).

La stratégie **middle-down**, qui combine les avantages des approches *bottom-up* et *top-down*, et qui consiste en l'étude de peptides dont la taille est plus importante que dans les approches *bottom-up* (2 et 20 kDa) (A. Moradian *et al.*, 2014; S. Sidoli *et al.*, 2017).

Au cours de ce chapitre, nous présenterons uniquement la stratégie **bottom-up**, utilisée dans le cadre de ces travaux de thèse.

I- Identification des protéines

L'analyse protéomique par approche *bottom-up* permet l'identification et la quantification des protéines à partir des peptides qui sont obtenus par digestion enzymatique (Y. Zhang *et al.*, 2013). Cette stratégie comporte trois étapes (Figure I-1) : la préparation d'échantillon, l'analyse LC-MS/MS et le traitement de données réalisé à l'aide d'outils bio-informatiques. Ces trois étapes peuvent être optimisées afin de mieux répondre à la question biologique posée. Dans la suite de ce manuscrit, les différentes étapes seront décrites en détail, ainsi que les optimisations réalisées durant mes travaux de thèse.

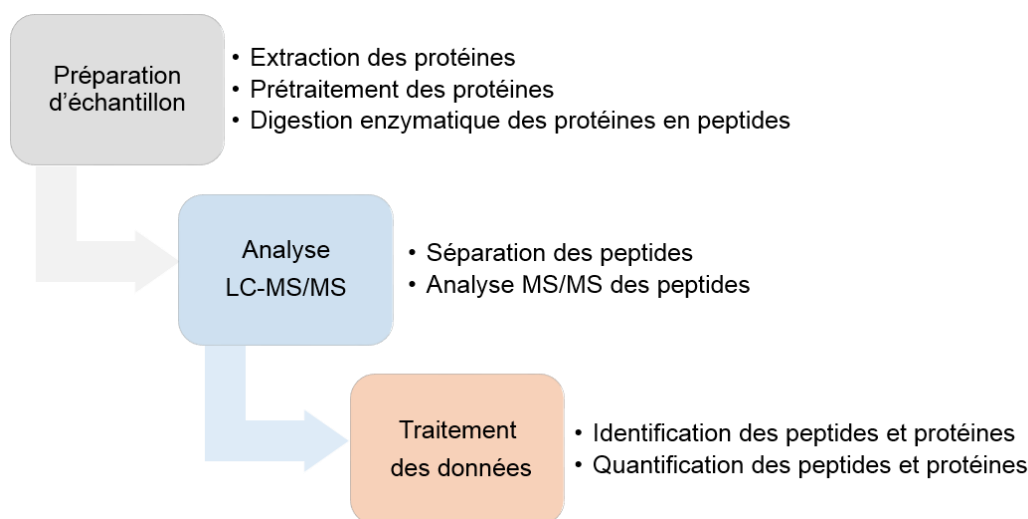


Figure I-1. Description des différentes étapes de l'analyse protéomique de type *bottom-up*. (1) Préparation d'échantillon, (2) Analyse LC-MS/MS, (m/z) correspond au rapport masse sur charge, (3) Traitement des données.

1. Préparation d'échantillon

1.1. Extraction des protéines

L'extraction des protéines est effectuée à partir de tissus, de cellules, de fluides biologiques, etc. L'objectif principal est d'extraire le maximum de protéines, tout en limitant la présence d'autres constituants biologiques contaminants. Le protocole d'extraction des protéines doit être choisi en fonction de la nature des protéines à extraire et de l'approche analytique utilisée.

Il existe de nombreuses techniques d'extraction, mécanique ou chimique. L'extraction mécanique peut être effectuée par broyage de l'échantillon, ultrasonication, cycles de congélation/décongélation (A. Dhabaria *et al.*, 2015). L'extraction chimique est généralement réalisée à l'aide de :

Détergents, qui aident à solubiliser les protéines, notamment les protéines membranaires. Ils sont classés en détergents ioniques (ex. SDS, SDC), non-ioniques (ex : Triton X-100, octylglucoside), ou zwitterioniques (ex. CHAPS) (M. le Maire *et al.*, 2000; M. Orwick-Rydmark *et al.*, 2016; A. E. Speers *et al.*, 2007). Certains sont compatibles avec la spectrométrie de masse (ex. RapiGest™) (E. I. Chen *et al.*, 2007; F. Meng *et al.*, 2002)

Chaotropes, qui perturbent les interactions entre les protéines et permettent de détruire la structure tridimensionnelle des édifices biologiques (ex : urée, thiourée) (A. E. Speers *et al.*, 2007)

Acides organiques, qui détériorent les membranes et solubilisent les protéines membranaires (ex : acide formique (FA) et acide trifluoroacétique (TFA)) (A. E. Speers *et al.*, 2007)

Mélanges solvant organique-eau, qui facilitent la rupture des membranes et dénaturent les protéines (ex : acétonitrile-eau et méthanol-eau) (A. E. Speers *et al.*, 2007).

La préparation d'échantillon doit être optimisée selon la nature de l'échantillon, les méthodes analytiques qui seront utilisées et l'objectif final de l'étude. Dans le cadre de ma thèse, pour la quasi-totalité des projets, les protéines ont été obtenues par une combinaison d'extraction mécanique et chimique en réalisant un broyage dans du tampon Laemmli qui contient 2% de SDS (U. K. Laemmli, 1970). Une optimisation de cette étape a été effectuée, afin d'adapter le volume du tampon d'extraction (cf. chapitre III).

1.2. Prétraitement des protéines

Suite à l'extraction des protéines, nous obtenons un mélange protéique complexe, qui peut ne pas être compatible avec la digestion enzymatique et l'analyse LC-MS/MS selon le protocole d'extraction utilisé et les réactifs utilisés. Une étape de prétraitement peut permettre d'éliminer les contaminants et/ou les détergents.

La purification de l'échantillon ou élimination des contaminants (lipides, sels, etc) peut être effectuée par précipitation des protéines (acide trichloroacétique (TCA), acétone, mélange chloroforme-méthanol ou acétate d'éthyle) (D. Channaveerappa *et al.*, 2019), par dialyse (S. M. Smith, 2017) ou ultrafiltration (M. Chevallet *et al.*, 2007; P. Feist *et al.*, 2015; J. R. Wisniewski *et al.*, 2009). Cette étape est nécessaire pour limiter la suppression de signal, les interférences et ainsi faciliter l'identification des protéines. Cependant, ces méthodes peuvent avoir comme conséquence la perte de protéines (M. Chevallet *et al.*, 2007).

Une autre technique, très courante en protéomique, est l'électrophorèse sur gel de polyacrylamide en présence de SDS (SDS-PAGE), qui permet de fractionner les protéines, et d'éliminer les sels et les contaminants (U. K. Laemmli, 1970). Les protéines sont réduites par un agent réducteur et dénaturées par du SDS, puis elles sont soumises à un champ électrique. Grâce à la charge négative apportée par le SDS, la vitesse de migration des protéines ne dépend plus que de leur masse moléculaire (MM). Le pourcentage d'acrylamide (%T) dans les gels de polyacrylamide est choisi en fonction de la MM des protéines à séparer : pourcentage plus important pour les faibles MM, pourcentage plus faible pour les MM élevées.

De manière générale, le gel d'électrophorèse est constitué de deux parties : un gel de concentration (4 à 5%T) et un gel de séparation (8 à 20%T) (Figure I-2). Le fractionnement des protéines permet de réduire la complexité de l'échantillon et augmente les chances d'identifier les protéines faiblement abondantes. Par exemple, il est très utile lorsque l'organisme d'intérêt est présent en quantité très faible dans un échantillon contenant différents protéomes (A. Shevchenko *et al.*, 2007).

Ce fractionnement est d'autant plus performant que la migration sur le gel de séparation est faite sur une grande distance (Figure I-2A). A la fin de la migration les bandes de gel sont excisées (en bandes de 2 mm durant ma thèse) et prêtes pour la digestion enzymatique et l'analyse LC-MS/MS. Cependant, le fractionnement SDS-PAGE est chronophage par le temps nécessaire à l'électrophorèse, mais surtout par le temps d'analyse LC-MS/MS qui est multiplié par le nombre de bandes excisées.

Il existe d'autres techniques qui peuvent être utilisées pour rendre l'échantillon compatible avec la digestion et l'analyse LC-MS/MS. Ces techniques, qui s'affranchissent du fractionnement, peuvent être envisagées lorsque l'échantillon et la question biologique posée permettent de se passer de cette étape, mais aussi grâce aux avancées en termes de performances instrumentales (efficacité chromatographique, vitesse d'acquisition et résolution des spectromètres de masse). Durant ma thèse, pour l'un des projets (cf. chapitre IV), j'ai été amenée à comparer trois procédés de prétraitement sans fractionnement (*stacking gel*, tube gel et digestion en solution).

Le procédé ***stacking gel (SG)*** est basé sur l'approche SDS-PAGE, et consiste à faire migrer les protéines uniquement dans le gel de concentration (Figure I-2B), d'où le nom de *stacking gel*. La totalité de l'échantillon se retrouve concentré dans une seule bande. La bande est découpée avant l'étape de digestion *in gel*.

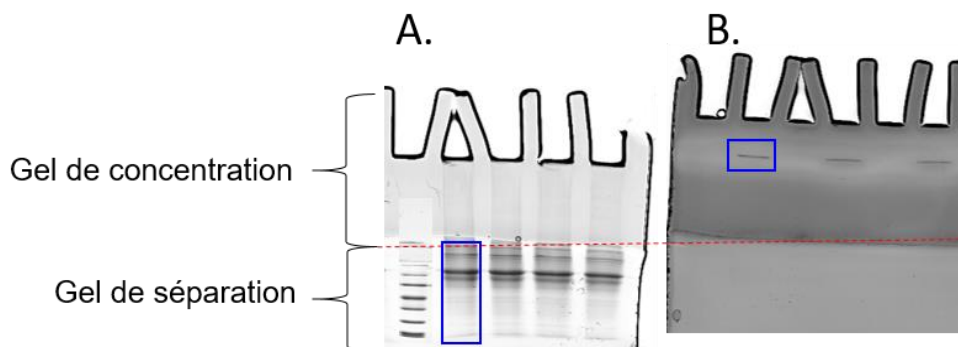


Figure I-2. Exemple de gel d'électrophorèse SDS-PAGE. La ligne en pointillé rouge indique la limite entre le gel de concentration (4%T) et le gel de séparation (12%T). **A.** Exemple de fractionnement d'une biopsie cutanée étudiée dans le chapitre III. Le carré bleu montre la zone des 10 bandes excisées. La piste de gauche correspond aux marqueurs de MM. **B.** Exemple d'un *stacking* gel d'un échantillon de cellules infectées étudié dans le chapitre IV. Le carré bleu montre l'échantillon concentré dans une seule bande.

Le procédé **tube-gel (TG)** avait été proposé il y a quelques années (R. Fischer *et al.*, 2015; X. Lu *et al.*, 2005) et a été évalué au laboratoire pour des approches quantitatives (L. Muller *et al.*, 2019). L'échantillon est piégé dans le gel de polyacrylamide au moment de la polymérisation, ce qui est compatible avec l'utilisation du SDS très efficace pour l'extraction des protéines. Le TG offre les mêmes avantages que le SG, tout en limitant la manipulation de l'échantillon. Les gels sont découpés en morceaux de taille homogène avant l'étape de digestion *in gel*.

La **digestion en solution (DS)** (N. Nagaraj *et al.*, 2008) évite le passage par un gel de polyacrylamide mais nécessite d'avoir une compatibilité des réactifs/détergents, utilisés pour l'extraction, avec la digestion enzymatique. Par exemple l'urée, qui est utilisée couramment comme chaotrope, compatible à condition d'être dilué. La DS est généralement suivie par une étape d'extraction sur phase solide (SPE) pour éliminer les sels avant l'analyse LC-MS/MS. Cela peut être chronophage selon le nombre d'échantillons à traiter et peut engendrer des pertes d'échantillon quand la quantité de départ est très faible (R. M. Tubaon *et al.*, 2017). Les détergents qui peuvent être utilisés avant une DS sont par exemple l'octylglucoside, le SDC, le SDS et le RapiGest™ détergent acido-clivable compatible avec l'analyse LC-MS/MS (F. Meng *et al.*, 2002).

D'autres techniques de préparation d'échantillon ont été proposées. Une des plus utilisées est le **FASP** (*Filter Aided Sample Preparation*) compatible avec une extraction des protéines avec du SDS (J. R. Wisniewski *et al.*, 2009). Les protéines sont retenues par une membrane filtrante (seuil de coupure 10 ou 30 kDa) tandis que le détergent est éliminé. Les protéines peuvent

ensuite être digérées et les peptides sont ne sont plus retenus par le filtre. Une étape de SPE sur phase inverse C18 peut compléter le protocole FASP avant l'analyse LC-MS/MS.

En 2014, trois autres techniques ont été proposées afin de limiter les étapes de manipulation de l'échantillon. La première permet de réaliser la lyse cellulaire jusqu'à l'élution des peptides dans un seul volume, il s'agit de l'**iST** (*in-Stage Tip digestion*) (N. A. Kulak *et al.*, 2014). Cette technique permet la digestion en solution comme le FASP, mais utilise un disque de phase inverse C18 qui sert de barrière physique et permet de nettoyer directement les peptides avant l'analyse LC-MS/MS. L'autre technique, nommée **STrap** (*Suspension Trapping*) (A. Zougman *et al.*, 2014) est compatible avec une extraction des protéines avec du SDS, et permet d'effectuer l'élimination du détergent, la digestion des protéines et le nettoyage de peptides « tout en un ». Cette technique combine l'avantage du FASP dans un temps plus court, notamment lors de la digestion (30 min). Enfin, une autre technique nommée **SP3** (*Single-Pot Solid Phase-enhanced Sample Preparation*) (C. S. Hughes *et al.*, 2014) permet de piéger les protéines à l'aide de billes magnétiques, d'éliminer les contaminants à l'aide de solvants organiques, et de réaliser la digestion enzymatique et la récupération des peptides dans un seul tube.

Plus récemment, d'autres techniques basées sur l'utilisation de billes ont été proposées, telles que le **BOPs** (*microbead-Based and Organic-media-assisted Proteolysis strategy*) (M. Taoka *et al.*, 2017) qui combine les avantages de l'utilisation de phase inverse pour la rétention des protéines et l'efficacité d'une digestion dans un milieu aqueux en présence d'acétonitrile (ACN), et le tout dans un seul tube. Dans la technique **USP³** (*Universal Solid-Phase Protein Preparation*) (L. F. Dagley *et al.*, 2019) inspirée du SP3, une étape d'hydrolyse acide de l'ADN et l'ARN pendant l'extraction des protéines a été incorporée.

Enfin, une technique basée sur un dispositif microfluidique contenant une membrane de filtration, nommé **CFP** (*ChipFilter Proteolysis*), a été récemment proposé (M. M. Ndiaye *et al.*, 2020). Cette technique permet de réaliser de la lyse cellulaire jusqu'à la digestion enzymatique avec l'avantage de l'automatisation et de la miniaturisation.

Tous ces procédés incluent une étape de réduction des ponts disulfure puis d'alkylation des thiols pour éviter la reformation des ponts disulfure. Dans le cadre de ma thèse, nous avons utilisé du dithiothréitol (DTT) et de l'iodoacétamide (IAA). Nous avons effectué ces traitements de manière automatisée (robot de prédigestion MassPrep) dans le cas du préfractionnement SDS-PAGE (Figure I-3), et de manière manuelle dans le cas de SG, TG et DS.

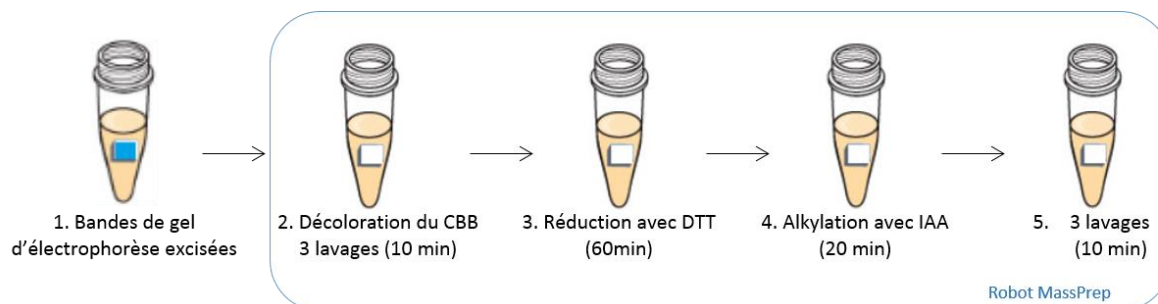


Figure I-3. Traitement des bandes de gel avant la digestion enzymatique. Les bandes de gel sont excisées après coloration au CBB. L'encadré représente les étapes effectuées par le robot MassPrep : élimination du CBB et des contaminants par 3 lavages dans 25 mM NH_4HCO_3 , puis réduction des ponts disulfure par 10 mM DTT et alkylation des thiols par 55 mM IAA. Après 3 lavages, les échantillons sont prêts pour la digestion enzymatique.

Au cours de ma thèse, la plupart des échantillons ont été soumis à un fractionnement SDS-PAGE (Chapitre II, Chapitre III). Quelques échantillons ont été traités par TG, SG et DS (Chapitre IV).

1.3. Digestion des protéines

La stratégie *bottom-up* passe par une étape primordiale de digestion enzymatique des protéines, en solution ou *in gel*. Plusieurs enzymes ayant une bonne spécificité de clivage (ex : trypsine, Lys-C, Asp-N, chymotrypsine) sont disponibles, ainsi que des enzymes « non spécifiques » (ex : pepsine, protéinase K). Malgré l'efficacité de ces enzymes, elles ne permettent pas d'obtenir une digestion complète de toutes les protéines d'un échantillon. En effet, lors de la digestion enzymatique, des clivages manqués peuvent apparaître (L. H. Betancourt *et al.*, 2018). C'est pourquoi, dans certains cas, une combinaison de plusieurs enzymes peut être utilisée pour augmenter la couverture de la protéine ou pour limiter les clivages manqués, dans le cas de la combinaison Lys-C/trypsine.

La trypsine reste l'enzyme de choix pour les analyses protéomiques avec des coupures au niveau des résidus lysine (K) et arginine (R) sauf s'ils sont suivis par un résidu proline (P). La taille de la trypsine (23 kDa) lui permet de pénétrer dans le gel de polyacrylamide, et la rend compatible avec une digestion *in gel* (L. C. Gillet *et al.*, 2016; L. Tsiatsiani *et al.*, 2015; E. Vandermarliere *et al.*, 2013). La fréquence moyenne des résidus K et R dans les protéines permet d'obtenir des peptides dans une gamme de masse comprise entre 500 et 3000 Da, adaptés aux méthodes de séparation chromatographique qui suivent. De plus, les peptides

générés avec K ou R en position C-terminale sont des peptides facilement ionisables en milieu acide pour les analyses par spectrométrie de masse.

Dans le cadre de ces travaux de thèse, la quasi-totalité des digestions a été réalisée *in gel* par la trypsine. Les peptides ont été extraits avec une solution d'ACN en milieu acide, puis analysés par LC-MS/MS.

2. Analyse par LC-MS/MS

2.1. Séparation des peptides par chromatographie liquide

L'utilisation de la chromatographie liquide est essentielle avant les analyses par spectrométrie de masse afin de séparer les peptides issus de la digestion. Il existe plusieurs techniques de chromatographie liquide, telles que la chromatographie en phase inverse (RP), la chromatographie d'interaction hydrophile (HILIC), la chromatographie par échange d'ions (SCX ou WCX pour les cations et SAX ou WAX pour les anions) et la chromatographie d'exclusion stérique (SEC).

Les différents modes de chromatographie liquide peuvent être combinés pour augmenter le fractionnement des peptides. Cette approche, connue sous le nom de MuDPIT (*Multi-Dimensional Protein Identification Technology*), a été décrite pour la première fois en 1999 en combinant SCX et RP (A. Link *et al.*, 1999; D. A. Wolters *et al.*, 2001). D'autres combinaisons ont ensuite été proposées (HILIC-RP, RP à différents pH, SEC-RP et d'autres) (Z. Zhang *et al.*, 2014). L'approche MuDPIT [SCX-RP] a été récemment réévaluée avec différents spectromètres de masse de type Orbitrap (Y. Zhang *et al.*, 2019). Elle a permis une identification plus profonde du protéome et des informations supplémentaires au niveau des modifications post-traductionnelles (PTMs). Néanmoins, l'analyse d'un seul échantillon peut prendre jusqu'à 24 heures.

La chromatographie en phase inverse est la plus utilisée en protéomique pour la séparation des mélanges complexes de peptides en amont du spectromètre de masse. En effet, elle est performante pour séparer les peptides tryptiques et la phase mobile est directement compatible avec la spectrométrie de masse (L. C. Gillet *et al.*, 2016). Les peptides sont séparés en fonction de leur polarité, grâce à la combinaison d'une phase stationnaire apolaire (silice greffée par des chaînes hydrophobes de type C18) et d'une phase mobile polaire, à base d'eau et d'ACN ou de MeOH, dont la polarité diminue au cours de l'analyse. La qualité de la séparation dépend de plusieurs paramètres dont la qualité de la phase stationnaire, le

diamètre des particules (granulométrie), la longueur de la colonne, et le gradient de polarité de la phase mobile.

Au cours de ces travaux, différents équipements de chromatographie liquide en phase inverse ont été utilisés (Tableau I-1) :

- MicroLC : microHPLC 1100 Series HPLC system (Agilent technologies)
- nanoUPLC : nanoAcquity (Waters)

Tableau I-1. Principales caractéristiques des systèmes de chromatographie liquide utilisés au cours de ma thèse. Deux systèmes chromatographiques ont été utilisés : un système « micro » en mode HPLC et un système « nano » en mode UPLC.

Système	Greffage de la phase stationnaire	Longueur de la colonne (mm)	Diamètre de la colonne (μm)	Taille des particules (μm)	Porosité (Å)	Débit (μL/min)
MicroLC	C18	150	300	3,5	300	5,00
nanoUPLC	C18	250	75	1,7	130	0,450

Les deux équipements ont des avantages et des inconvénients. Le système nanoUPLC permet d'avoir une meilleure résolution (granulométrie de 1,7 μm, colonne plus longue) et une meilleure sensibilité (diamètre interne plus petit) appréciable lorsque la quantité d'échantillon est faible. Cependant la quantité injectable reste limitée et la pression est plus importante. Le système microLC est plus robuste que le système nanoUPLC et permet d'injecter une quantité de matériel plus importante. Cependant la résolution chromatographique est moindre (granulométrie de 3,5 μm, colonne plus courte).

Les équipements ont été choisis en fonction du type d'analyse, du spectromètre de masse couplé à la chromatographie et de la quantité d'échantillon.

2.2. Spectrométrie de masse en tandem

En sortie de chromatographie liquide, les peptides sont analysés par spectrométrie de masse en tandem selon les processus suivants (B. Domon *et al.*, 2010) :

- Ionisation des analytes dans la source électrospray (ESI).
- Séparation des ions en fonction de leur rapport masse/charge (m/z), pour générer un spectre MS représentant l'intensité en fonction de m/z .

- Sélection d'un ion précurseur, ou de plusieurs ions précurseurs dans une gamme de m/z .
- Fragmentation de l'ion précurseur sélectionné ou des ions co-sélectionnés.
- Séparation des ions fragments en fonction de leur rapport m/z pour générer les spectres MS/MS.

2.2.1. Différents modes d'acquisition

La spectrométrie de masse en tandem peut être réalisée selon deux modes d'acquisition, non ciblée ou ciblée. Ces deux modes d'acquisition diffèrent par la manière de faire l'acquisition des données relatives aux précurseurs et aux fragments ainsi que par le traitement de données (L. C. Gillet *et al.*, 2016).

2.2.1.1. Modes d'acquisition non ciblée

L'acquisition non ciblée est utilisée principalement quand on souhaite identifier ou éventuellement quantifier un maximum de protéines contenues dans un échantillon. On distingue deux modes : le mode DDA (*Data-Dependent Acquisition*) et le mode DIA (*Data-Independent Acquisition*) (Figure I-4)

Le **mode DDA** est le mode le plus utilisé en protéomique. Une analyse est effectuée par une succession de cycles d'acquisition. Chaque cycle démarre par l'acquisition d'un spectre MS (mesure des ions précurseurs générés dans un cycle de temps), à partir duquel certains précurseurs sont séquentiellement sélectionnés, selon des critères prédéfinis généralement basés sur l'intensité du signal, pour être isolés, fragmentés et finalement obtenir les spectres MS/MS correspondants (L. C. Gillet *et al.*, 2016; Y. Zhang *et al.*, 2013). La sélection des ions précurseurs pour la fragmentation limite le nombre de peptides séquencés au cours d'une analyse, et par conséquent le nombre de protéines identifiées. De plus, le mode DDA souffre du caractère stochastique de la sélection des ions précurseurs, causé par la variabilité des temps de rétention lors des analyses chromatographiques, qui ne sont jamais absolument identiques d'une analyse à l'autre (L. C. Gillet *et al.*, 2016; A. Michalski *et al.*, 2011). Des études ont montré que, dans un échantillon complexe, 84% des peptides ne sont pas identifiés, et que 30% des peptides peuvent varier entre des répliquats d'analyses (J. D. Egertson *et al.*, 2013).

Afin de palier le problème de sélection de l'ion précurseur pour sa fragmentation, le **mode DIA** a été développé (J. D. Egertson *et al.*, 2013; T. Geiger *et al.*, 2010a). Ce mode DIA a pour but

de fragmenter tous les précurseurs présents dans une certaine gamme de masse. Il existe deux techniques principales :

La technique nommée MS^E consiste à alterner les acquisitions à basse et haute énergies. Les spectres à basse énergie servent à cartographier les précurseurs, puis les acquisitions à haute énergie vont permettre la fragmentation de l'ensemble de précurseurs (S. J. Geromanos *et al.*, 2009).

La deuxième technique nommée SWATH™ consiste à fragmenter tous les précurseurs co-sélectionnés dans de larges fenêtres d'isolement (entre 10 et 25 m/z). Cette fenêtre d'isolement est répétée tout au long de l'analyse afin de couvrir l'ensemble de la gamme de masse (L. C. Gillet *et al.*, 2012). Pour la sélection des précurseurs, il est possible d'appliquer des fenêtres fixes et variables. Les fenêtres variables permettent de diminuer la complexité des spectres MS/MS grâce à l'optimisation de leur taille en fonction de la distribution en nombre ou en intensité des signaux des ions précurseurs en fonction du ratio m/z, ce qui permet donc d'augmenter la spécificité (C. Ludwig *et al.*, 2018). L'extraction des données et *in fine* la détection des peptides sont réalisées à partir des informations contenues dans des bibliothèques spectrales, généralement générées par des analyses DDA (L. C. Gillet *et al.*, 2012) ou directement à partir des banques de données avec des logiciels spécialisés (L. C. Gillet *et al.*, 2016).

Le mode DIA combine les avantages du mode DDA et des modes de quantification ciblée en étant capable de fournir une couverture du protéome similaire ou supérieure à celle du mode DDA avec une sensibilité, spécificité et robustesse comparables aux modes de quantification ciblée (L. C. Gillet *et al.*, 2012).

Cependant, les spectres générés en mode DIA sont plus complexes que ceux obtenus en mode DDA, ce qui rend impossible l'utilisation des moteurs de recherche classiquement utilisés pour l'analyse des données DDA (J. D. Egerton *et al.*, 2013). Pour les deux techniques MS^E et SWATH™, la relation entre les fragments et les précurseurs est retrouvée post-acquisition à l'aide de logiciels dédiés, ce qui rend le travail d'interprétation des données assez important. Quant à la technique SWATH™, la nécessité d'utiliser des bibliothèques spectrales générées à partir d'analyses acquises en mode DDA limite également la détection des peptides/protéines. Pour contourner ce problème, de nouvelles approches ont été introduites récemment, dans laquelle les bibliothèques spectrales sont générées à partir d'outils de prédiction de spectres, de l'utilisation de banques de données et/ou d'injections effectuées directement en mode DIA (B. C. Searle *et al.*, 2020).

Les deux modes d'acquisition non ciblée, DDA et DIA, permettent également de quantifier les peptides/protéines, mais pour des raisons informatiques les données acquises en mode DDA restent à l'heure actuelle les plus utilisées.

2.2.1.2. Modes d'acquisition ciblée

L'acquisition ciblée est utilisée pour détecter et/ou quantifier un nombre limité de protéines avec une très bonne sensibilité et spécificité. Deux mode différents peuvent être distingués, la **SRM** (*Selected Reaction Monitoring*) (B. Domon *et al.*, 2006; P. Picotti *et al.*, 2012) et la **PRM** (*Parallel Reaction Monitoring*) (A. Bourmaud *et al.*, 2016; S. Gallien *et al.*, 2012; A. C. Peterson *et al.*, 2012) (Figure I-4). L'analyse de type SRM est effectuée sur des instruments de type triple quadripôle (QqQ). Son principe repose sur la sélection d'un ion précurseur (peptide) d'un m/z prédéfini, qui est isolé dans le premier quadripôle, puis fragmenté pour générer des ions fragments qui sont isolés successivement dans le dernier quadripôle. Le couple formé par un ion précurseur et un ion fragment est appelé une transition. La sélection de plusieurs transitions, propres à chaque peptide, rend la SRM très spécifique. Cependant, le développement d'une méthode SRM est un processus qui prend du temps. De plus, lors de l'analyse d'échantillons très complexes, des interférences peuvent perturber la bonne détection des peptides.

Pour pallier ces limitations, l'analyse de type PRM plus récente est réalisée sur d'autres instruments où le dernier quadripôle est remplacé par un analyseur doté d'une meilleure résolution et meilleure précision de masse de type Orbitrap ou TOF. Cette technique est également appelée HR/AM (*High Resolution/Accurate Mass*). En effet, elle permet de réaliser une analyse à haute résolution et sans sélection préalable des ions fragments : une fois l'ion précurseur sélectionné et fragmenté, tous les fragments sont analysés, et les transitions sont extraites *a posteriori*. La PRM offre une plus grande spécificité et une sensibilité proche de la SRM. Ces deux modes d'acquisition ciblée seront plus amplement détaillés dans la partie II de ce chapitre.

Dans le cadre de ma thèse, l'identification des protéines a été réalisée en utilisant un mode non ciblé DDA (Chapitre II, Chapitre III et Chapitre IV), hormis une étude préliminaire réalisée en mode non ciblé DIA (Chapitre II.III-1.3.2). En ce qui concerne l'acquisition ciblée, les deux modes SRM et PRM ont été utilisés pour la détection spécifique et la quantification de protéines d'intérêt (Chapitre II).

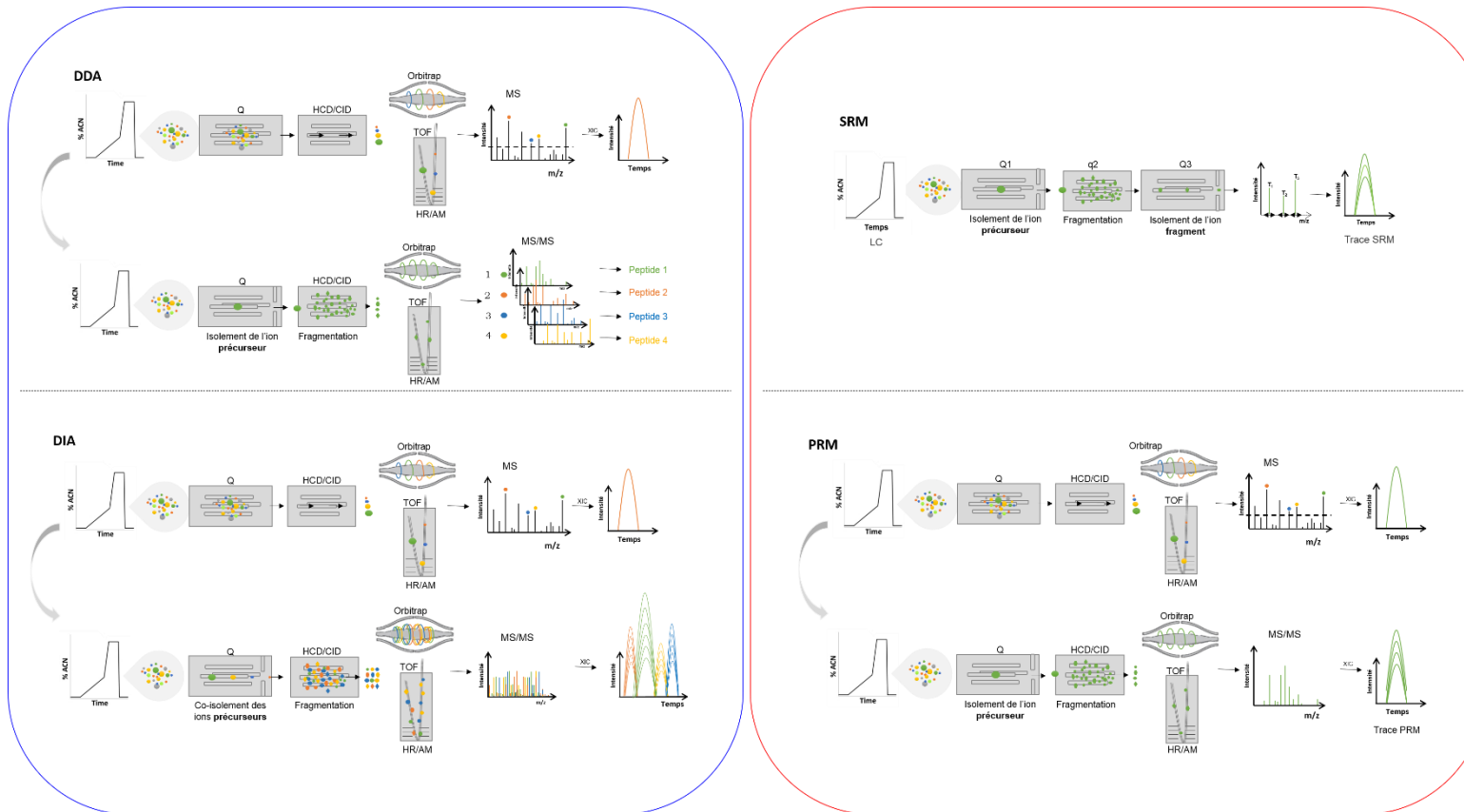


Figure I-4. Principaux modes d'acquisition MS utilisés en protéomique. Le mode DDA consiste à réaliser un spectre MS puis, à partir du spectre MS, à sélectionner un nombre limité d'ions précurseurs et à les fragmenter pour obtenir les spectres MS/MS correspondants. Similaire au mode DDA pour l'acquisition des spectres MS, le mode DIA repose sur la sélection de l'ensemble des précurseurs contenus dans une fenêtre de masse, puis leur fragmentation simultanée, ce qui génère des spectres MS/MS multiplexés. Les modes d'acquisition ciblée (SRM et PRM) permettent de détecter spécifiquement des peptides d'intérêt. En mode SRM, les ions précurseurs d'intérêt sont sélectionnés dans le 1^{er} quadripôle, puis fragmentés, et un nombre limité de fragments prédéfinis sont sélectionnés dans le dernier quadripôle pour obtenir la trace SRM des différentes transitions. En mode PRM, les ions précurseurs d'intérêt sont sélectionnés dans le 1^{er} quadripôle, puis fragmentés, et l'ensemble des fragments sont analysés dans l'analyseur à haute résolution (TOF, Orbitrap). Les transitions sont extraites des spectres MS/MS.

2.2.2. Fragmentation des peptides

La détermination de la séquence peptidique pour l'analyse protéomique dépend de l'étape de fragmentation. Il existe plusieurs types de fragmentation : CID (*Collision Induced Dissociation*) (L. Sleno *et al.*, 2004), HCD (*Higher-energy Collision Dissociation*) (J. V. Olsen *et al.*, 2007), ETD (*Electron-Transfer Dissociation*) (J. E. Syka *et al.*, 2004), ECD (*Electron Capture Dissociation*) (R. A. Zubarev *et al.*, 2000), ETHcd (C. K. Frese *et al.*, 2013) qui résulte d'une combinaison de l'ETD et la HCD, et UVPD (*UltraViolet PhotoDissociation*) (J. A. Madsen *et al.*, 2010). Les modes de fragmentation CID et HCD sont similaires et sont les plus couramment utilisés en protéomique. Au cours de ma thèse, seuls les modes CID et HCD ont été utilisés et seront par conséquent les seuls à être présentés.

Comme indiqué précédemment, l'enzyme la plus couramment utilisée pour l'analyse protéomique est la trypsine. Cette enzyme permet la formation majoritaire de précurseurs doublement chargés, dont les charges sont portées par l'extrémité C-terminale basique (K ou R) et par l'amine N-terminale. Ces caractéristiques sont mises à profit lors de la fragmentation des peptides en modes CID et HCD. La fragmentation CID est réalisée grâce à la collision successive des ions avec des molécules neutres de gaz (He, Ar, N₂). La fragmentation a lieu quand l'énergie de collision est assez élevée ce qui permet l'augmentation de leur énergie interne et sa conversion en partie en énergie vibrationnelle permettant la délocalisation du proton le long du squelette. La fragmentation CID est dite dirigée par la charge et suit le modèle du proton mobile (A. R. Dongre *et al.*, 1996; V. H. Wysocki *et al.*, 2000). En mode CID on observe principalement la rupture de la liaison peptidique et la formation de fragments où la charge peut être retenue soit du côté N-terminal (ions de type b), soit du côté C-terminal (ions de type y) (Figure I-5). En mode HCD, une radiofréquence plus élevée est appliquée afin de piéger l'ensemble des fragments dans la *C-trap*, avant leur transfert dans l'Orbitrap. L'appellation *Higher energy*, ne traduit pas une plus grande énergie utilisée pour la fragmentation, le mode HCD reste une dissociation à basse énergie comme la CID. Le mode HCD produit principalement des ions y et b comme en CID.

La nomenclature de la fragmentation peptidique a été introduite dans les années 1980 par Roepstorff (P. Roepstorff *et al.*, 1984) avant d'être simplifiée par Biemann (K. Biemann, 1990) (Figure I-5). De manière générale cette fragmentation peut intervenir au niveau de 3 liaisons : la rupture de la liaison C-C produit les fragments a et x, la rupture de la liaison peptidique

produit les fragments b et y (obtenus en CID et HCD), la rupture de la liaison N-C produit les fragments c et z (obtenus en ECD et ETD).

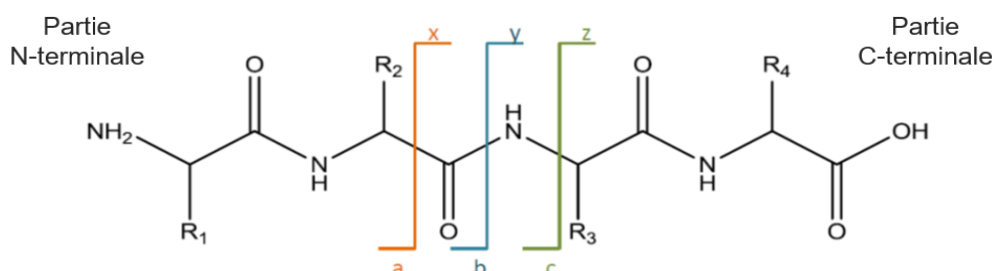


Figure I-5. Nomenclature de fragmentation peptidique selon Biemann. R1 à R4 représentent les chaînes latérales de quatre acides aminés. Les ions des séries a, b et c contiennent l'extrémité N-terminale alors que les ions des séries x, y et z contiennent l'extrémité C-terminale.

La différence de masse entre deux ions d'une même série (ex. y_1 et y_2 ou b_1 et b_2) permet d'identifier l'acide aminé et d'obtenir des informations de séquence. Un exemple est donné dans la Figure I-6

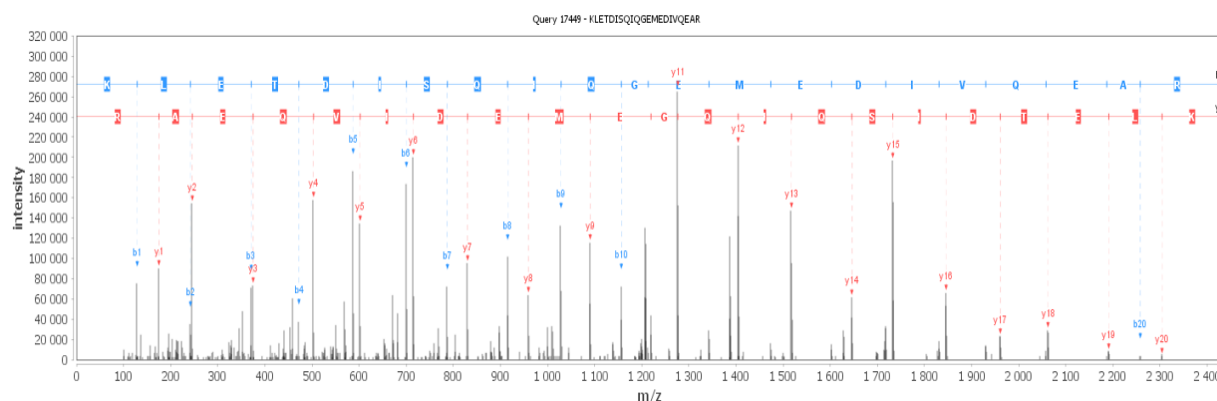


Figure I-6. Exemple d'un spectre MS/MS obtenu par fragmentation HCD d'un peptide. Deux séries d'ions y et b, correspondant à la rupture de la liaison peptidique, sont observées. La série y majoritaire permet de déduire la séquence peptidique K(I/L)ETD(I/L)SQ(I/L)QGEMEDIVQEAR. Les acides aminés isobares I et L ne peuvent pas être distingués.

2.2.3. Nature des spectromètres de masse

Dans le cadre de mes travaux de thèse, plusieurs spectromètres de masse couplés à la nanoUPLC ou la microLC ont été utilisés. Le choix de ces instruments a été dicté par les caractéristiques de l'instrument et le but de l'analyse (identification de protéines, détection spécifique, quantification). Les caractéristiques des spectromètres de masse, et les types d'analyse pour lesquels ils ont été utilisés, sont détaillés dans le (Tableau I-2).

Tableau I-2. Principales caractéristiques des instruments utilisées au cours de ma thèse. SRM: *Selected Reaction Monitoring*. PRM: *Parallel Reaction Monitoring*. DDA: *Data-Dependent Acquisition* DIA: *Data-Independent Acquisition*. Dans le cas du Triple TOF 5600, la nomenclature est spécifique à la société Sciex. IDA: *Information-Dependent Acquisition*. MRM-HR: *Multiple Reaction Monitoring - High Resolution*. SWATH: *Sequential Windowed Acquisition of all Theoretical spectra*. Les acronymes suivants désignent le même mode d'acquisition : DDA=IDA ; MRM-HR=PRM et SWATH-MS=DIA).

Machine	QqQ-6490	Triple TOF 5600	Q-Exactive +	Q-Exactive HF-x
Constructeur	Agilent Technologies	Sciex	Thermo Fisher Scientific	Thermo Fisher Scientific
Type d'analyseur	QqQ	Q-TOF	Q-Orbitrap	Q-Orbitrap
Gamme de masse (m/z)	1 400	40 000	6 000	6 000
Vitesse d'acquisition (Hz)	50	100	12	40
Précision de masse	0,7 Da	10 ppm	5 ppm	5 ppm
Résolution	Unitaire	30 000 (MS) 15 000 (MS/MS)	17 500 à 140 000 (à 200 m/z)	7 500 à 240 000 (à 200 m/z)
Mode d'acquisition	SRM	IDA, MRM-HR, SWATH-MS	DDA, PRM, DIA	DDA, PRM, DIA
Couplage chromatographique	microLC	nanoUPLC	nanoUPLC	nanoUPLC
Application	Détection spécifique Quantification ciblée	Identification	Identification Quantification ciblé Quantification relative	Identification Quantification ciblé
Année d'installation	2014	2014	2014	2017

3. Traitements des données et identification des protéines

L'interprétation des données brutes issues des analyses MS/MS nécessite une série de traitements informatiques pour identifier les protéines d'intérêt. Trois approches sont utilisées pour l'identification des protéines à partir des spectres MS/MS obtenus par fragmentation des peptides (L. C. Gillet *et al.*, 2016; L. McHugh *et al.*, 2008) : (i) approche PFF (*Peptide Fragment Fingerprint*), qui consiste à comparer les spectres d'ions fragments à ceux issus des banques protéiques, (ii) séquençage *de novo*, qui consiste en l'interprétation des spectres MS/MS pour obtenir une séquence peptidique, et (iii) utilisation de bibliothèques spectrales, qui permettent de comparer les spectres obtenus avec une liste de spectres assignés avec confiance. Dans cette partie, nous allons détailler l'approche PFF majoritairement utilisée durant ma thèse, avec la recherche dans les banques protéiques et la validation des données.

3.1.1. Banques protéiques

La première étape du traitement est la sélection de la banque protéique adaptée. La composition de la banque de données est importante car les identifications dépendent des données présentes dans la banque utilisée. En outre, il est impossible d'identifier une protéine si elle n'est pas présente dans la banque de données sélectionnée.

De nombreuses banques protéiques sont disponibles, mais sont variables en termes de nombre de protéines, de qualité d'annotation des séquences et de degré de redondance (A. Nesvizhskii, 2007). En matière de qualité d'annotation et de redondance, il existe deux types de banques : les banques non corrigées et les banques corrigées (R. Apweiler *et al.*, 2004).

Les **banques non corrigées** contiennent des données peu ou pas vérifiées. Selon le niveau de vérification, elles peuvent être volumineuses et plus ou moins redondantes. Parmi ces banques, nous pouvons distinguer :

NCBI Entrez (D. Benson *et al.*, 1990; E. W. Sayers *et al.*, 2020a; G. D. Schuler *et al.*, 1996), créée par le NCBI (*National Center for Biotechnology Information*). Cette banque contient des séquences protéiques provenant de séquences nucléotidiques de DDBJ (*DNA Data Bank of Japan*) (S. Miyazaki *et al.*, 2003), EMBL (*European Molecular Biology Laboratory*) (G. Stoesser *et al.*, 2003) et GenBank (E. W. Sayers *et al.*, 2020b) regroupés dans la banque TPA (*Third Party Annotation*). Elle contient également des séquences provenant de SwissProt (B. Boeckmann *et al.*, 2003), PIR (*Protein Information Resource*) (C. H. Wu *et al.*, 2003), RefSeq (*Reference Sequence*) (N. A. O'Leary *et al.*, 2016) et PDB (*Protein Data Bank*) (H. M. Berman *et al.*, 2000). Cette banque est redondante, du fait des différentes sources d'entrées. Néanmoins, elle reste une option très utile pour des organismes peu séquencés.

RefSeq (N. A. O'Leary *et al.*, 2016), créée par le NCBI, avec un travail supplémentaire de l'équipe de NCBI pour éliminer les redondances des données de GenBank. Les séquences et annotations sont générées par une combinaison de collaborations externes, prédiction automatisée et vérification manuelle (C. Chen *et al.*, 2017). Cependant, la vérification des annotations est effectuée de façon partielle par un groupe de collaborateurs externes, sans validation supplémentaire par l'équipe de NCBI. Cette banque est en constante amélioration, avec des mises à jour publiées tous les ans dans la revue *Nucleic Acids Research* (E. W. Sayers *et al.*, 2020a) et accessibles sur le site de NCBI. On dénombre un total de 181 567 562 d'entrées protéiques au 13 août 2020.

Les **banques corrigées** contiennent des séquences protéiques non redondantes et corrigées. La fiabilité de ces données provient d'une vérification et d'une validation des annotations par des experts biologistes avant leur publication. La banque de données couramment utilisée est UniProtKB.

UniProtKB (*United Protein Knowledge Base*)(C. UniProt, 2019) a été créée en 2002. Elle correspond à la fusion de trois banques corrigées (R. Apweiler *et al.*, 2004) :

- **PIR-PSD** (*Protein Information Resource - Protein Sequence Database*), la plus ancienne banque de données corrigée située au centre médical de l'Université de Georgetown à Washington.
- **Swiss-Prot**, principale banque de séquences protéiques corrigées, développée par l'Institut Suisse de Bioinformatique (SIB). Cette banque est assez restreinte, en raison du grand travail mené pour fournir des informations fiables et complètes.
- **TrEMBL** (*Translation from EMBL*), supplément créé pour déposer des séquences pas encore intégrées dans la banque Swiss-Prot, afin de rendre ces nouvelles séquences disponibles.

A l'heure actuelle, UniProtKB contient des séquences, des informations sur les fonctions des protéines, sur leur localisation, sur leurs PTMs, etc. Elle est divisée en deux parties (C. Chen *et al.*, 2017; K. Verheggen *et al.*, 2020) :

UniProtKB/Swiss-Prot : qui contient des données extraites de la littérature ou des analyses informatiques, toutes deux étant vérifiées et annotées manuellement. On y dénombre un total de 563 082 d'entrées protéiques au 13 aout 2020.

UniProtKB/TrEMBL : qui contient des données obtenues par traitement informatique avec une vérification et une annotation automatiques. On y dénombre un total de 188 961 949 d'entrées protéiques au 13 aout 2020.

Les banques de données UniProtKB sont mises à jour mensuellement (K. Verheggen *et al.*, 2020). C'est pourquoi il est recommandé d'utiliser les banques de données les plus actualisées possibles ou, dans le cas d'un projet à long terme, d'utiliser la même version afin d'éviter les biais dans les résultats qui seraient dus aux mises à jour.

Comme nous venons de le voir, diverses banques protéiques avec des caractéristiques différentes existent. Bien choisir la banque et l'adapter aux questions biologiques posées est primordial pour l'obtention de données de qualité. Un mauvais choix de banque peut entraîner

une mauvaise assignation des protéines et par conséquent une mauvaise interprétation des données.

Il est également important de préciser qu'une « grande » banque protéique peut diminuer la spécificité de la recherche, du fait du plus grand nombre de possibilités d'assigner des peptides aux spectres expérimentaux (K. Verheggen *et al.*, 2020). En effet, plus on augmente la taille de la banque, plus il y a de possibilités d'assigner des peptides à un même spectre MS/MS, ce qui a pour conséquence une identification moins spécifique et une augmentation de faux positifs (S. Chatterjee *et al.*, 2016).

Nous pouvons aussi faire mention des banques de données qui servent de stockage et d'exploitation des données acquises par MS (Y. Perez-Riverol *et al.*, 2015) (banques de dépôt), telles que **GPMD** (*Global Proteome Machine Data Base*) (R. Craig *et al.*, 2004b) qui permet le stockage et le retraitement des données présentes dans d'autres banques de dépôt et utilisant le moteur de recherche X! Tandem, **PRIDE** (*PRoteome IDentification database*) (P. Jones *et al.*, 2006), qui stocke des données expérimentales et **Peptide Atlas** (F. Desiere *et al.*, 2006), utilisée pour constituer des bibliothèques spectrales (H. Lam *et al.*, 2011), et qui contient d'autres outils pour les approches quantitatives ciblées (E. W. Deutsch *et al.*, 2008).

Au cours de ma thèse, j'ai utilisé des banques corrigées et non corrigées. L'ensemble des banques utilisées contient une liste de protéines appelées contaminants, basée sur la liste proposée par cRAP (*common Repository of Adventitious Proteins*) (J. Shin *et al.*, 2019) et complétée, dans notre laboratoire, par d'autres protéines. Les différentes banques ont été générées soit par le logiciel MSDA (C. Carapito *et al.*, 2014) soit par le logiciel Galaxy (E. Afgan *et al.*, 2018).

3.1.2. Interprétation des données

Dans l'approche protéomique *bottom-up*, les protéines sont identifiées à partir des peptides identifiés. Cela peut cependant rendre difficile l'interprétation des données lorsque :

- les organismes analysés ont des protéomes très proches, telles que les différentes espèces d'un organisme ou encore les différentes souches d'une espèce,
- la banque contient des redondances,
- les peptides contiennent des acides aminés isobares (Ile/Leu),
- la digestion des protéines n'a pas été complète ou spécifique.

De plus, un peptide peut appartenir à plusieurs protéines ou isoformes protéiques, ce qui est désigné par le terme de **peptides partagés**. Ces peptides partagés peuvent conduire à des ambiguïtés pour l'identification des protéines (A. I. Nesvizhskii *et al.*, 2005). Cela n'est pas maîtrisable, c'est pourquoi il est important de considérer l'unicité des peptides à une protéine donnée. Un **peptide** est considéré comme étant **unique** à une protéine lorsque celui-ci appartient exclusivement à la protéine à laquelle il a été attribué dans la banque de données utilisée. Comprendre ces difficultés dans l'interprétation de données permet de choisir le traitement de données le plus adapté ou d'effectuer une validation supplémentaire.

L'approche PFF que j'ai utilisée pour identifier les peptides repose sur la comparaison, à l'aide d'algorithmes dédiés ou moteurs de recherche, d'une liste de **masses expérimentales** obtenues par l'analyse LC-MS/MS avec une liste des **masses théoriques** calculées à partir des protéines présentes dans les banques de données. Cette comparaison permet d'identifier les peptides et par inférence d'identifier les protéines. Le spectre MS/MS attribué à une séquence peptidique qui a permis l'identification d'une protéine donnée est nommé *Peptide Spectrum Match* (PSM). La qualité de l'identification est évaluée à l'aide de scores spécifiques calculés par les moteurs de recherche. Plusieurs peptides peuvent être attribués à un spectre donné, un score est alors attribué à chacun afin d'évaluer la qualité de la assignation de la séquence du peptide au spectre (K. Verheggen *et al.*, 2020). Plus ce score est élevé, meilleure est la qualité de l'identification.

Les données brutes expérimentales contiennent toutes les informations acquises par le spectromètre de masse dans un format spécifique à chaque constructeur. Pour que ces données puissent être utilisées par le moteur de recherche, elles sont transformées dans un format standardisé, qui contient les m/z des précurseurs et des fragments associés, ainsi que leurs intensités et charges respectives.

Il existe à l'heure actuelle un grand nombre de moteurs de recherche, qu'ils soient libres ou commerciaux. Verheggen *et al.* ont publié une liste de 39 différentes moteurs de recherche avec notamment le nombre de citations comme indicateur de leur utilisation (K. Verheggen *et al.*, 2020). Les moteurs de recherche les plus utilisés sont (K. Verheggen *et al.*, 2020): SEQUEST (J. K. Eng *et al.*, 1994), Mascot (D. N. Perkins *et al.*, 1999), OMSSA (L. Y. Geer *et al.*, 2004), X! Tandem (R. Craig *et al.*, 2004a) et Andromeda (J. Cox *et al.*, 2011). L'utilisation de ces algorithmes est basée sur des paramètres dépendant des conditions expérimentales. Les principaux paramètres à définir par l'utilisateur sont (A. Nesvizhskii, 2007) :

- La banque protéique,

- L'enzyme utilisée pour la digestion des protéines et le nombre autorisé de coupures manquées (*miscleavage*),
- La charge des précurseurs,
- Les types d'ions fragments, qui dépendent de la fragmentation réalisée,
- Les tolérances en masse sur les précurseurs et les fragments, qui dépendent de la résolution et de la précision de masse avec lesquelles les spectres MS et MS/MS sont acquis,
- Les PTMs (ex. phosphorylation, acétylation) potentiellement présentes,
- Les modifications chimiques, fixes ou variables, qui peuvent être présentes (ex. oxydation de méthionines, carbamidométhylation des cystéines).

D'autres paramètres sont accessibles selon les moteurs de recherche, mais ils doivent être utilisés avec précaution. En effet, l'ensemble de ces paramètres définit l'espace de recherche de l'algorithme, et l'augmentation de cet espace de recherche augmente le nombre de correspondances possibles entre les spectres expérimentaux et théoriques. Ceci peut conduire à une identification erronée.

En moyenne 25% des spectres expérimentaux MS/MS réussissent à être attribués (J. Griss *et al.*, 2016). Un espace de recherche inadapté peut diminuer ce pourcentage d'identification. Les raisons les plus courantes de cette mauvaise attribution sont (A. Nesvizhskii, 2007) :

- Une mauvaise assignation de peptide lors de la comparaison des données théoriques et expérimentales,
- Une mauvaise qualité des spectres MS/MS, avec un rapport signal sur bruit faible qui peut induire une mauvaise interprétation,
- La génération de spectres MS/MS chimères qui résultent de la co-fragmentation de deux ou plusieurs précurseurs (K. Verheggen *et al.*, 2020),
- La présence de peptides partagés dans la banque de données utilisée,
- La présence d'acides aminés isobares,
- La sous-évaluation des PTMs lors des recherches (J. M. Chick *et al.*, 2015),
- L'utilisation de banques inaptes ou incomplètes ou inversement l'utilisation de banques très larges qui augmentent l'espace de recherche (L. C. Gillet *et al.*, 2016).

Au cours de ma thèse, j'ai utilisé principalement l'algorithme **Mascot**. Pour certains projets, **OMSSA** et **Mascot** ont été utilisés en parallèle afin d'augmenter la confiance accordée aux identifications ou de maximiser le nombre d'identifications (Chapitre II). **Andromeda** a

également été utilisé, cet algorithme étant intégré au logiciel MaxQuant (Chapitre IV). **Pulsar**, algorithme de recherche implémenté par Biognosys dans le logiciel Spectronaut, a été utilisé pour les analyses en mode DIA (Chapitre II).

3.1.3. Validation des identifications

La validation des identifications obtenues via ces algorithmes est une étape essentielle. En effet, comme discuté précédemment, la comparaison des spectres expérimentaux et théoriques peut générer des faux positifs, c'est-à-dire une mauvaise attribution d'un spectre MS/MS à une séquence (K. Verheggen *et al.*, 2020). La stratégie la plus couramment utilisée pour évaluer les faux positifs est appelée *target-decoy* (J. E. Elias *et al.*, 2007). Elle repose sur une banque composée de la banque protéique cible (*target*) et de la banque contenant des protéines leurres (*decoy*). Les séquences leurres peuvent être générées par randomisation ou inversion des séquences afin de préserver la fréquence des acides aminés et la taille des peptides et protéines. L'hypothèse est une assignation aléatoire de la présence d'un peptide dans la banque protéique cible ou dans la banque leurre. Ainsi, le nombre de séquences leurres assignées donne une estimation du nombre d'assignements incorrects, ou faux positifs, qu'on peut obtenir dans la banque protéique cible (M. M. Savitski *et al.*, 2015). Cette hypothèse permet le calcul du FDR (*False Discovery Rate*) à partir de l'équation suivante (P. Navarro *et al.*, 2009) :

$$FDR = 2 \times \frac{\text{Nombre de séquences leurres assignées}}{\text{Nombre de séquences leurres} + \text{Nombre de séquences cibles assignées}} \times 100 \quad \text{Equation 1}$$

Cette hypothèse peut être utilisée pour contrôler le FDR au niveau des PSMs, des peptides et des protéines en fonction d'un score attribué. Il existe à l'heure actuelle des avis partagés sur l'utilisation du FDR et il n'y a pas de valeur optimale (P. Navarro *et al.*, 2009; A. I. Nesvizhskii *et al.*, 2007). De plus, l'utilisation de la stratégie *target-decoy* peut présenter des limitations lors de l'analyse de très grands jeux de données, du fait d'une augmentation du nombre de séquences leurres (M. M. Savitski *et al.*, 2015), étant donné qu'elle sont déterminées à partir des banques de données utilisées. En d'autres termes, plus nous augmentons la taille de la banque, plus un nombre important de séquences leurres sont présentes dans l'espace de recherche. Des filtres peuvent être appliqués pour limiter le nombre de faux positifs, tels que la longueur minimale de la séquence peptidique (nombre d'acides aminés) et le score d'identification (ex. 25 pour Mascot et 9 pour OMSSA) qui assure une bonne qualité d'assignation spectrale.

Dans le cadre de ma thèse, les recherches ont été validées à l'aide des filtres et de l'application d'un FDR au niveau protéique inférieur à 1%. Cette étape a été réalisée avec le logiciel Proline (D. Bouyssie *et al.*, 2020), développé en partie au sein du laboratoire dans le cadre de l'infrastructure nationale de protéomique (ProFI). De plus, selon les projets, nous avons choisi de travailler soit avec uniquement des peptides uniques, soit avec des peptides partagés (cf. Chapitre III).

II- Protéomique ciblée

1. Analyse LC-SRM

1.1. Principe de l'analyse LC-SRM

L'analyse LC-SRM (*Liquid Chromatography-Selected Reaction Monitoring*) est une **analyse ciblée** sur des protéines d'intérêt. Autrement dit, elle ne permet pas d'identifier toutes les protéines présentes dans l'échantillon, mais elle se focalise sur une détection spécifique d'un nombre limité de protéines d'intérêt. Les protéines d'intérêt sont détectées via des peptides qui n'appartiennent qu'à ces protéines.

L'instrument utilisé pour une analyse LC-SRM est un spectromètre de masse de type QqQ constitué de trois analyseurs de type quadripôle, ou deux quadripôles et un multipôle central. En mode SRM, un **ion précurseur** de m/z donné est sélectionné dans le premier quadripôle (Q1), puis il est **fragmenté** dans la cellule de collision (q2), et un ou plusieurs **ions fragments** sont successivement sélectionnés dans le troisième quadripôle (Q3) (Figure I-7) (P. Picotti *et al.*, 2012). Un couple précurseur/fragment est appelé une **transition**. Le suivi de cette transition, c'est-à-dire la double sélection de l'ion précurseur et de l'ion fragment, permet une détection spécifique du précurseur d'intérêt. Cette spécificité augmente lorsqu'on suit plusieurs transitions pour un même précurseur. De plus, le fait de consacrer tout le temps d'acquisition au suivi des précurseurs d'intérêt augmente la sensibilité. *In fine* on obtient un profil chromatographique pour chaque précurseur désigné par le terme de trace SRM. En toute rigueur, le terme MRM (*Multiple Reaction Monitoring*) devrait être utilisé lorsque plusieurs transitions sont suivies. On utilisera cependant dans ce manuscrit le terme SRM qui est communément utilisé dans la littérature pour désigner une analyse SRM ou MRM.

SRM

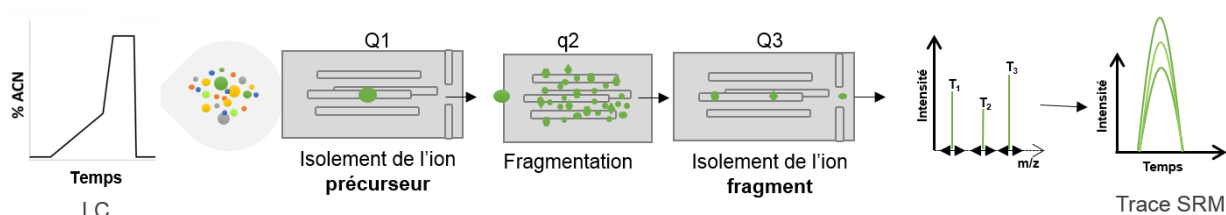


Figure I-7. Schéma d'une acquisition LC-SRM dans un spectromètre de masse de type QqQ. Les peptides séparés par chromatographie liquide sont ionisés dans la source ESI. Au niveau des analyseurs, un ion précurseur est isolé dans le premier quadripôle (Q1), puis fragmenté dans la cellule de collision (q2), et un ion fragment est isolé dans le dernier quadripôle (Q3). Cette double sélection dans Q1 et Q3 correspond au suivi d'une transition [précurseur→fragment]. Si plusieurs transitions T1, T2, T3 sont suivies, on obtient la trace SRM du précurseur, avec la superposition des transitions suivies pour ce précurseur.

Il est possible de choisir plusieurs peptides par protéine, ainsi que plusieurs protéines différentes. Cette capacité de multiplexage est un atout important de l'analyse LC-SRM et permet de rechercher plusieurs protéines d'intérêt en une seule analyse. Grâce à sa spécificité et sa sensibilité, l'analyse LC-SRM est appropriée pour la détection de protéines d'intérêt en concentration faible dans des matrices complexes, et elle offre une bonne reproductibilité (T. A. Addona *et al.*, 2009).

L'intérêt de cette approche pour la détection et la quantification de biomarqueurs de différentes pathologies a été amplement montré, notamment dans le domaine du cancer (R. Huttenhain *et al.*, 2012). Par exemple la détection de biomarqueurs du cancer de la prostate a été réalisée dans des tissus tumoraux (A. K. Yocum *et al.*, 2010) ou dans l'urine (T. Shi *et al.*, 2017). Plus récemment une méthode SRM a été utilisée pour la quantification de PIVKA-II dans le sérum humain afin de diagnostiquer le carcinome hépatocellulaire (CHC) à un stade précoce (A. Sohn *et al.*, 2018). La SRM a été également utilisé pour l'étude de biomarqueurs dans d'autres types d'échantillon, liquide synovial (D. Cretu *et al.*, 2014), liquide céphalorachidien (Y. Jia *et al.*, 2012) et peau (J. C. Williamson *et al.*, 2013).

Dans le domaine des maladies infectieuses, la SRM a été utilisée pour détecter quelques fmoles de protéines de *Mycobacterium tuberculosis* dans le sérum humain (C. Mehaffy *et al.*, 2017). Une autre étude a été réalisée sur des protéines sanguines de l'hôte qui varient dans le cas de la borréliose de Lyme (Y. Zhou *et al.*, 2020).

La LC-SRM permet la détection spécifique mais également la quantification des protéines détectées, comme nous allons l'expliquer ci-dessous.

1.2. Stratégies de quantification par LC-SRM

De nombreuses stratégies de quantification par LC-SRM sont basées sur l'utilisation de standards internes marqués aux isotopes stables. Le principe est d'ajouter en quantité connue des protéines ou peptides marqués dans l'échantillon à analyser. La comparaison des signaux entre analytes d'intérêt et standards internes (protéines/peptides marqués) permet de quantifier les analytes d'intérêt. Ces stratégies sont considérées comme méthodes de référence pour la quantification absolue de protéines/peptides d'intérêt.

La méthode la plus courante en protéomique quantitative consiste à ajouter dans l'échantillon un **peptide** synthétique, de séquence identique à celle du peptide d'intérêt mais **marqué aux isotopes stables**. Ce peptide marqué est nommé peptide **SIS** (*Stable Isotope-labeled Standard*). Les peptides SIS les plus courants contiennent des isotopes ^{13}C et ^{15}N . Ces isotopes sont incorporés sur les lysines et arginines C-terminales, ce qui représente une différence de masse de +8 Da et +10 Da, par rapport aux lysines et arginines non marquées, respectivement.

Les peptides SIS synthétiques sont disponibles sous deux formes :

- Des peptides avec une pureté élevée, dont la concentration est connue et certifiée. Ce sont les peptides **AQUA™** (*Absolute QUAntification*), utilisés quand une quantification juste et précise est souhaitée (S. A. Gerber *et al.*, 2003).
- Des peptides **crude** avec une pureté plus faible, dont la quantité n'est pas certifiée. Ces peptides ont permis de rendre la quantification via les peptides SIS plus courante, en raison de leur coût plus faible que les AQUA™ (en général inférieur à 40€ selon la séquence, les PTMs, le fournisseur, etc, à comparer au coût d'environ 500€ d'un peptide AQUA™).

Cette méthode présente toutefois des limitations, principalement liées à la préparation d'échantillon. Par exemple, une digestion enzymatique incomplète biaisera la quantification dans le cas de l'ajout des peptides SIS en aval de la digestion, juste avant l'analyse LC-SRM. Pour pallier ce problème, certains fournisseurs proposent des peptides SIS à ajouter en amont de la digestion, et qui contiennent une étiquette (*tag*) clivable lors de la digestion (Figure I-8). Les peptides SIS sont ainsi soumis à la même étape de digestion que les protéines de

l'échantillon. Toutefois, cela ne permet pas de s'affranchir des biais dus aux pertes de protéines en amont de la digestion.



Figure I-8. Représentation d'un peptide SIS avec une étiquette dans la lysine C-terminale. Le peptide GPNLTEISK possède une étiquette dans la lysine C-terminale. L'étiquette est clivable lors de la digestion trypsique pour obtenir le peptide SIS d'intérêt.

D'autres biais peuvent apparaître mais qui peuvent être minimisés par l'opérateur, tels qu'une solubilisation incomplète des peptides SIS (dans le cas des peptides lyophilisés), ou une adsorption des peptides SIS hydrophobes sur les parois des eppendorfs, surtout pour les concentrations très faibles.

Ces limitations ont conduit au développement d'autres stratégies, telles que la protéine **QconCAT** (*Quantification conCATamer*) (R. J. Beynon *et al.*, 2005). Les peptides d'intérêt sont concaténés, exprimés dans *E. coli*, dans un milieu enrichi en arginine et lysine marquées, constituant ainsi une protéine artificielle marquée. La protéine QconCAT est ajoutée à l'échantillon et digérée en même temps, générant des peptides en quantité équimolaire. Cette quantité ne peut pas être adaptée aux peptides endogènes générés. Bien que l'ajout soit effectué avant la digestion enzymatique, il ne permet pas de limiter les pertes liées aux étapes qui précèdent à cette digestion.

D'autres stratégies sont basées sur l'emploi de **protéines marquées aux isotopes stables**. Elles visent à réduire les biais liés à la préparation d'échantillon. La première stratégie, proposée en 2007, est appelée **PSAQ™** (*Protein Standard Absolute Quantification*) (V. Brun *et al.*, 2007; G. Picard *et al.*, 2012) : la protéine est marquée *in vitro* et a l'avantage de pouvoir être ajoutée le plus tôt possible dans l'étape de préparation d'échantillon afin d'être soumise au même traitement que la protéine endogène. Toutefois, le coût d'une protéine marquée PSAQ™ reste très important. Une autre stratégie, nommée **absolute SILAC** (*Stable Isotope Labeling by Amino acids in Cell culture*), est une adaptation de l'approche originelle SILAC (S.-E. Ong *et al.*, 2002) : la protéine standard est obtenue par expression dans *E. coli* dans un milieu enrichi en arginine et lysine marquées (S. Hanke *et al.*, 2008.). Dans la stratégie **FLEXIQuant** (*Full-length expressed stable isotope-labelled proteins for quantification*) la protéine standard est produite dans des extraits de germes de blé au lieu de *E. coli*, et marquée avec un « peptide FLEX », qui est utilisé pour effectuer la quantification (S. Singh *et al.*, 2009).

Enfin, la stratégie **PrEST** (Protein Epitope Signature Tag) (M. Zeiler *et al.*, 2012) a été développée dans le cadre du projet *Human Protein Atlas* et utilise des épitopes connus de protéines humaines.

Des études ont été menées pour comparer les stratégies utilisant des protéines marquées ou des peptides marqués. Elles suggèrent que les approches utilisant des protéines marquées sont plus adaptées pour la quantification, à condition que la protéine standard synthétisée se comporte exactement comme la protéine endogène ciblée (A. Hober *et al.*, 2019; C. Huillet *et al.*, 2012).

Au cours de ma thèse, la quantification des peptides d'intérêt a été réalisée à l'aide de peptides synthétiques marqués SIS. Certains ont été ajoutés avant la digestion, d'autres après la digestion juste avant l'analyse LC-SRM. Nous avons également utilisé des peptides synthétiques non marqués **NAT** (*natural form*) pour déterminer des limites de quantification.

1.3. Développement d'une méthode LC-SRM

1.3.1. Sélection des protéines et peptides

L'analyse de type SRM nécessite d'avoir sélectionné en amont les protéines d'intérêt, ou protéines cibles.

A partir de la liste des protéines cibles, l'étape suivante est la sélection des peptides protéotypiques. Un peptide est défini comme **protéotypique** quand il est unique à la protéine d'intérêt et qu'il est détectable par spectrométrie de masse (V. Lange *et al.*, 2008b; P. Mallick *et al.*, 2007). Pour une meilleure quantification, il est conseillé d'appliquer des critères supplémentaires pour la sélection des peptides d'intérêt :

- Favoriser les peptides sans clivage manqués,
- Favoriser les peptides d'une longueur comprise entre 8 et 25 acides aminés,
- Favoriser les précurseurs chargés 2+ et 3+ (accessibles dans la gamme de masse de l'instrument),
- Limiter les peptides tryptiques avec deux acides aminés basiques voisins (KK, KR, RK et RR),
- Limiter les peptides qui peuvent subir des modifications chimiques (ex. oxydation de méthionines, carbamidométhylation des cystéines),
- Limiter les peptides susceptibles de porter des PTMs.

Les peptides protéotypiques qui remplissent ces critères sont nommés peptides **quantotypiques** (P. Brownridge *et al.*, 2011).

De nombreux outils ont été développés pour aider au choix des peptides. Ils sont basés sur des données expérimentales générées par analyse LC-MS/MS ou prédictives en prenant en compte leur propriétés physico-chimiques. Plus particulièrement, en absence de données LC-MS/MS collectées sur l'échantillon d'intérêt, il existe des banques protéiques de dépôt dans lesquelles ces informations peuvent être accessibles. Des outils bioinformatiques peuvent également prédire les temps de rétention des peptides d'intérêt. Ces outils ne sont pas adaptés à tous les types d'échantillon et conditions d'analyse, néanmoins ils peuvent s'avérer très utiles en cas d'absence de données expérimentales. Lorsqu'une analyse non ciblée LC-MS/MS a été réalisée sur la même matrice biologique, le choix des peptides cibles peut être facilité par la liste de ceux qui ont déjà été détectés par MS.

Dans le cadre de ma thèse, nous verrons que les protéines cibles sont des protéines bactériennes choisies à partir d'analyses protéomiques non ciblées de peaux de souris infectées. Pour chaque protéine, au minimum deux peptides cibles ont été sélectionnés.

1.3.2. Sélection, validation et optimisation des transitions

Lorsque les peptides cibles ont été sélectionnés, l'étape suivante est la **sélection des transitions** (ion précurseur/ion fragment). Cette étape permet de trouver le meilleur compromis entre la spécificité et la sensibilité. La spécificité peut être obtenue en sélectionnant des transitions pour lesquelles aucune interférence n'est observée. La sensibilité dépend du nombre de transitions suivies. Dans l'analyse SRM, le nombre de transitions est lié à deux paramètres (Equation 2) :

- Le temps de balayage (*dwell time* ou *scan time*, selon le constructeur), qui correspond au temps d'acquisition d'une transition.
- Le temps de cycle (*cycle time*), qui correspond au temps d'acquisition de toutes les transitions.

$$\text{Temps de cycle} = \text{nombre de transitions} \times \text{temps de balayage} \quad \text{Equation 2}$$

Comme nous pouvons le constater, le temps de cycle dépend directement du nombre de transitions suivies par peptide, et du temps de balayage de chaque transition. Pour augmenter la sensibilité, il est possible d'augmenter le temps de balayage. Cela permet d'accumuler un

nombre plus important d'ions pour une transition donnée, mais c'est au détriment de la résolution (nombre de points par pics). Ainsi, il faut trouver un compromis entre le nombre de transitions, le temps de cycle et le temps de balayage.

Il est recommandé une acquisition de huit à dix points par pic chromatographique et un temps de balayage entre 10 et 100 ms (V. Lange *et al.*, 2008b). Par conséquent, le bon compromis va également dépendre de la largeur du pic, qui peut aller de 15 à 30s. Il est important de préciser que la largeur du pic est variable selon le système chromatographique utilisé, la concentration du peptide et le gradient utilisé. Ainsi le temps de cycle peut aller de 1s à 3s. Ainsi pour 3s de temps de cycle et 20 ms de temps de balayage nous pouvons suivre un maximum de 150 transitions (Equation 3), ce qui correspond à 50 peptides avec 3 transitions par peptide et environ 17 protéines si nous choisissons de suivre trois peptides par protéine.

$$\text{Nombre de transitions} = \frac{3000 \text{ ms}}{20 \text{ ms}} = 150 \text{ transitions} \quad \text{Equation 3}$$

Il faut également noter que le nombre total de transitions suivies dans la méthode est multiplié par deux, pour suivre les peptides SIS en même temps que les peptides cibles. Cela est également à prendre en considération dans le choix des paramètres.

Lorsque les transitions potentielles à suivre ont été choisies, une étape de validation dans la matrice biologique est primordiale afin de limiter les interférences dues à cette matrice. Dans le cadre de ma thèse la matrice biologique a été de la peau de chien, de souris ou d'homme, matrice complexe susceptible de nombreuses interférences (Chapitre II).

La **validation des transitions** peut être effectuée de différentes manières. L'une d'entre elles est la méthode *SRM-triggered* : la détection du peptide peut être confirmée par l'acquisition d'un spectre MS/MS déclenché au cours de l'analyse puis soumis à une recherche en banque de données pour vérifier que le peptide détecté correspond bien au peptide ciblé (V. Lange *et al.*, 2008a; V. Lange *et al.*, 2008b; R. D. Unwin *et al.*, 2005). Une autre méthode est la *iSRM* (*intelligent Selected Reaction Monitoring*), avec dans un premier temps le suivi d'un ensemble de transitions (≈ 3 transitions) afin de réaliser une quantification, puis dans un second temps, d'autres transitions « secondaires » sont mesurées pour générer des spectres MS/MS qui vont servir à valider les premières transitions (R. Kiyonami *et al.*, 2011). Les outils informatiques utilisés pour la sélection des peptides sont également applicables pour la sélection des transitions. Cependant, comme pour la sélection des peptides, la méthode la plus fiable reste l'utilisation de données expérimentales.

La meilleure approche pour la validation des transitions est l'utilisation des peptides SIS qui ont les mêmes caractéristiques physico-chimiques que les peptides cibles mais avec une légère variation de masse. C'est la méthode qui a été utilisée durant ma thèse, en dopant des peaux saines non infectées par des mélanges de peptides SIS, et en évaluant chacune des transitions.

Une analyse SRM peut être réalisée selon deux modes d'acquisition : acquisition non programmée (*unscheduled*) et acquisition programmée (*scheduled*). Le **mode d'acquisition non programmé** est utilisé quand le temps de rétention des peptides est inconnu et/ou que le nombre de transitions suivies est faible. Toutes les transitions de tous les peptides sont suivies tout au long de l'analyse (Figure I-9). Au contraire, le **mode d'acquisition programmée** nécessite une connaissance préalable des temps de rétention des peptides. Pour chaque peptide, les transitions ne sont suivies que durant une fenêtre de temps limitée autour du temps de rétention de ce peptide. Les fenêtres d'acquisition sont à définir en fonction du nombre de peptides et de transitions suivies, et de la reproductibilité des temps de rétention.

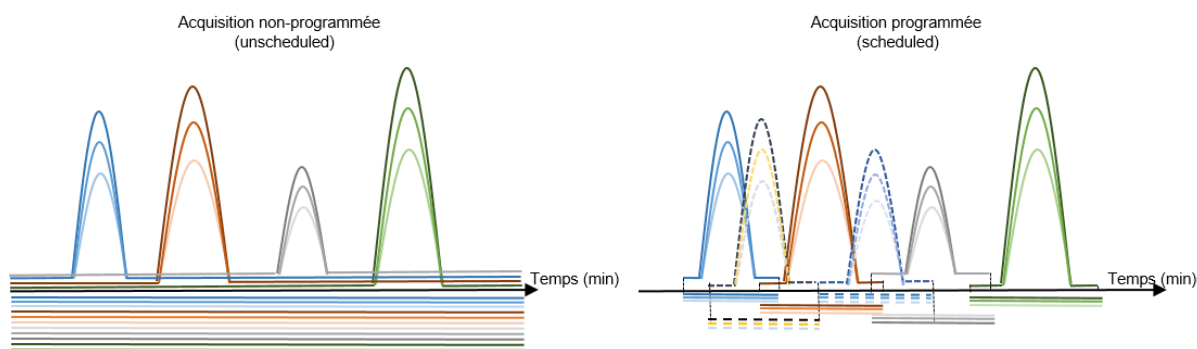


Figure I-9. Représentation schématique d'une acquisition SRM non programmée (*unscheduled*) et programmée (*scheduled*). En mode d'acquisition non programmée, les trois transitions des trois peptides sont suivies du début à la fin de l'analyse. En mode d'acquisition programmée, les trois transitions d'un peptide donné sont suivies uniquement dans la fenêtre de temps fixée autour du T_r de ce peptide, libérant ainsi du temps d'analyse pour de nouveaux peptides. Les nouveaux peptides sont représentés en pointillés

Le mode d'acquisition programmée permet d'augmenter le nombre de peptides ciblés, ou le nombre de transitions suivies, sans diminuer le temps de balayage, ce qui améliore le multiplexage. Selon les caractéristiques chromatographiques (durée de l'analyse LC, largeur du pic et reproductibilité), le nombre de transitions peut augmenter d'un facteur 5 à 20 sans diminution du temps de balayage en mode d'acquisition programmée (J. Stahl-Zeng *et al.*, 2007).

Au cours de ma thèse, j'ai utilisé ce mode d'acquisition programmée qui m'a permis de suivre un nombre de transitions très supérieur à 150, comme nous le verrons dans les Chapitre II.II-2, Chapitre II.III-1.2 et Chapitre II.III-2.1 .

L'étape suivante est l'**optimisation des transitions** qui vise à augmenter la sensibilité en modifiant certains paramètres instrumentaux. La nature et la pression du gaz de collision sont modifiées uniquement dans des cas exceptionnels (S. Gallien *et al.*, 2011). En revanche, l'énergie de collision (EC) est plus couramment effectuée lorsqu'une sensibilité élevée est désirée (P. Picotti *et al.*, 2012). Cette optimisation peut donner des résultats variables pour les transitions d'un même peptide, ou de donner aucune variation (C. A. Sherwood *et al.*, 2009).

Les optimisations des EC sont effectuées à partir des valeurs préconisées par les constructeurs (Equation 4), où a et b représentent des constantes. Cette équation est caractéristique de chaque instrument et déterminée de manière empirique en fonction de la masse et l'état de charge du précurseur.

$$EC = a \times \frac{m}{z} + b \quad \text{Equation 4}$$

Dans le cadre de ma thèse, nous avons systématiquement optimisé les EC, en faisant varier la valeur, par incrément de 2V (Chapitre II.II-2.1) ou 3V (Chapitre II.III-2.1) autour de la valeur préconisée par le constructeur.

La dernière étape consiste en l'**optimisation des dilutions des peptides SIS**. En effet tous les peptides n'ont pas le même facteur de réponse en MS et il est préférable d'avoir des intensités du même ordre de grandeur pour les différents peptides ainsi que des signaux des peptides SIS comparables à ceux des peptides endogènes détectés, ce qui n'est pas facile à obtenir puisque la quantité à détecter est inconnue.

2. Analyse LC-PRM

2.1. Principe de l'analyse LC-PRM

L'analyse ciblée de peptides d'intérêt peut également être réalisée par LC-PRM (*Liquid Chromatography - Parallel Reaction Monitoring*). Le concept de PRM a été proposé en 2012 par Peterson *et al.* en mettant à profit la haute résolution et la précision de masse (HR/AM, pour *high resolution/accurate mass*) des instruments hybrides de type Q-TOF ou Q-Orbitrap (S. Gallien *et al.*, 2012; A. C. Peterson *et al.*, 2012; B. Schilling *et al.*, 2015). Sur ces

instruments la sélection de l'ion précurseur est réalisée dans un quadripôle comme en SRM. Il est ensuite fragmenté et tous les fragments sont analysés dans l'analyseur HR/AM, contrairement à la SRM où seuls certains ions fragments prédéfinis sont suivis (Figure I-10). En mode PRM, l'analyse de tous les fragments génère un spectre MS/MS pour chaque ion précurseur. Ainsi, il n'est pas nécessaire d'avoir sélectionné au préalable les transitions, elles sont extraites *a posteriori* lors du traitement de données (E. Borrás *et al.*, 2017; S. Gallien *et al.*, 2012). Cela représente un gain de temps au niveau de l'optimisation de la méthode, mais le traitement de données est plus long que celui réalisé après une analyse SRM.

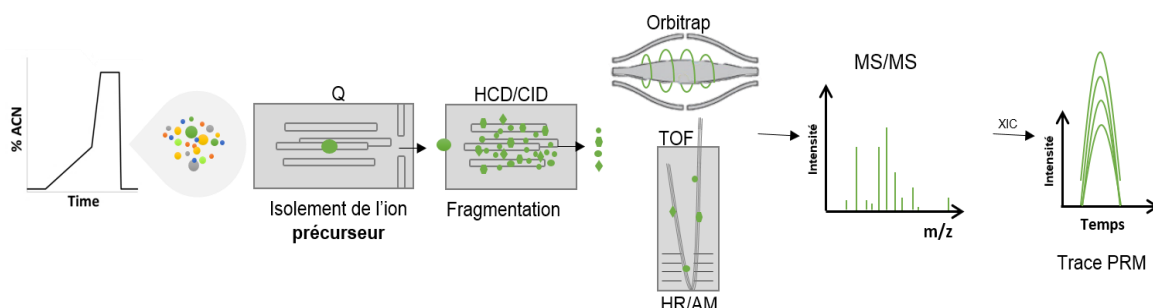


Figure I-10. Représentation schématique d'une acquisition LC-PRM dans un instrument hybride de type Q-Orbitrap ou Q-TOF. Les peptides séparés par chromatographie liquide sont ionisés au niveau de la source ESI. Un ion précurseur est isolé dans le premier quadripôle puis fragmenté dans la cellule de collision. Tous les ions fragments sont analysés dans un analyseur TOF ou Orbitrap pour générer le spectre MS/MS du précurseur. La trace PRM du précurseur est obtenue en extrayant *a posteriori* les transitions dans le spectre MS/MS.

Des études comparatives entre LC-PRM et LC-SRM ont montré des résultats similaires en termes de sensibilité, de reproductibilité, de précision, (M. A. Hoffman *et al.*, 2018; G. E. Ronsein *et al.*, 2015; C. Schiffmann *et al.*, 2014). Une complémentarité a également été observée pour la quantification de peptides dans les études menées par Schiffmann *et al.* et Hoffman *et al.*). Au laboratoire, des études réalisées par Joanna Bons pendant sa thèse, ont également montré la similarité en termes de reproductibilité et de précision, avec toutefois des quantifications différentes selon les techniques (J. Bons, 2019).

La PRM commence à être utilisée pour l'étude des maladies infectieuses. Des travaux récents réalisés sur des gouttes de sang séché (DBS, *Dried Blood Spot*) ont permis la quantification de deux protéines de *Yersinia pestis* en utilisant une technique d'immunocapture suivie d'une analyse PRM (A. Rifflet *et al.*, 2019). D'autres travaux ont été réalisés par PRM pour le diagnostic et le typage de flavivirus (virus de la dengue, Zika, fièvre jaune, virus de kunjin) dans le sérum humain. (S. Wee *et al.*, 2019)

La PRM peut présenter un avantage en terme de spécificité grâce au plus grand nombre de transitions disponibles pour confirmer la présence d'un peptide. La SRM peut, quant à elle, offrir une meilleure spécificité due à l'utilisation du triple quadripôle et à la double sélection de masse.

2.2. Développement d'une méthode LC-PRM

La sélection des peptides est effectuée comme décrit précédemment pour l'analyse LC-SRM (Chapitre I.II-1.3.1), mais la sélection des transitions n'est ensuite pas nécessaire pour la LC-PRM. Cela simplifie le développement de méthode mais génère un travail supplémentaire lors du traitement, notamment pour éliminer les interférences dues à la matrice. En mode PRM, l'extraction des transitions *a posteriori* permet d'effectuer une sélection des meilleures transitions du peptide ciblé.

Les analyses PRM ayant été réalisées durant ma thèse sur un instrument de type Q-Orbitrap (QExactive+, Thermo Fisher), la suite de ce paragraphe concernera cet instrument. Les paramètres importants pour une analyse PRM sur cet équipement sont les suivants (A. Bourmaud *et al.*, 2016; S. Gallien *et al.*, 2012).

Le temps transitoire (*transient time*) durant lequel les ions sont détectés dans l'Orbitrap et la **résolution** sont directement reliés. Ainsi, pour le spectromètre QExactive+, les résolutions de 17 500, 35 000 et 70 000 sont associées à des temps transitoires de 64, 128 et 256 ms, respectivement.

Le temps maximal d'injection (*maximum fill time*) est le temps maximal d'accumulation des ions dans la *C-trap* avant d'être envoyés dans l'Orbitrap. L'augmentation de ce temps d'injection augmente la sensibilité.

Le gain de contrôle automatique AGC (*Automatic Gain Control*) définit le nombre d'ion accumulés dans la *C-trap* avant d'être envoyés dans l'Orbitrap. Le temps maximal d'injection et le gain de contrôle automatique ont le même objectif, ainsi lorsqu'un des deux atteint la valeur maximale imposée, les ions sont envoyés dans l'Orbitrap pour y être analysés.

Comme pour la SRM, **le temps de cycle** doit permettre d'acquérir 8 à 10 points par pic chromatographique. Pour optimiser le temps de cycle, le temps maximal d'injection de la *C-trap* doit être inférieur au temps transitoire dans l'Orbitrap, les analyses sont ainsi réalisées en mode générique (Figure I-11,A). Ainsi le temps de cycle est calculé à partir de l'équation 5 (S. Gallien *et al.*, 2012) :

Condition : *Temps maximal d'injection* < *Temps transitoire*

$$\text{Temps de cycle} = \text{Nombre de précurseurs} \times \text{temps transitoire} \quad \text{Equation 5}$$

Les instruments de type Q-Orbitrap ont la capacité d'effectuer des analyses multiplexées, dans lesquelles des précurseurs sont isolés séquentiellement, puis fragmentés et accumulés dans la cellule HCD, pour finalement être analysés simultanément dans l'Orbitrap (Figure I-11, B). Dans ce cas, le temps maximal d'injection doit être inférieur au rapport entre le temps transitoire et le nombre de précurseurs analysés simultanément (degré de multiplexage). Le temps de cycle est alors calculé à partir de l'équation 6 (S. Gallien *et al.*, 2012).

$$\text{Condition : } \text{Temps maximal d'injection} < \frac{\text{Temps transitoire}}{\text{degré de multiplexage}}$$

$$\text{Temps de cycle} = \text{Nombre de précurseurs} \times \frac{\text{Temps transitoire}}{\text{Degré de multiplexage}} \quad \text{Equation 6}$$

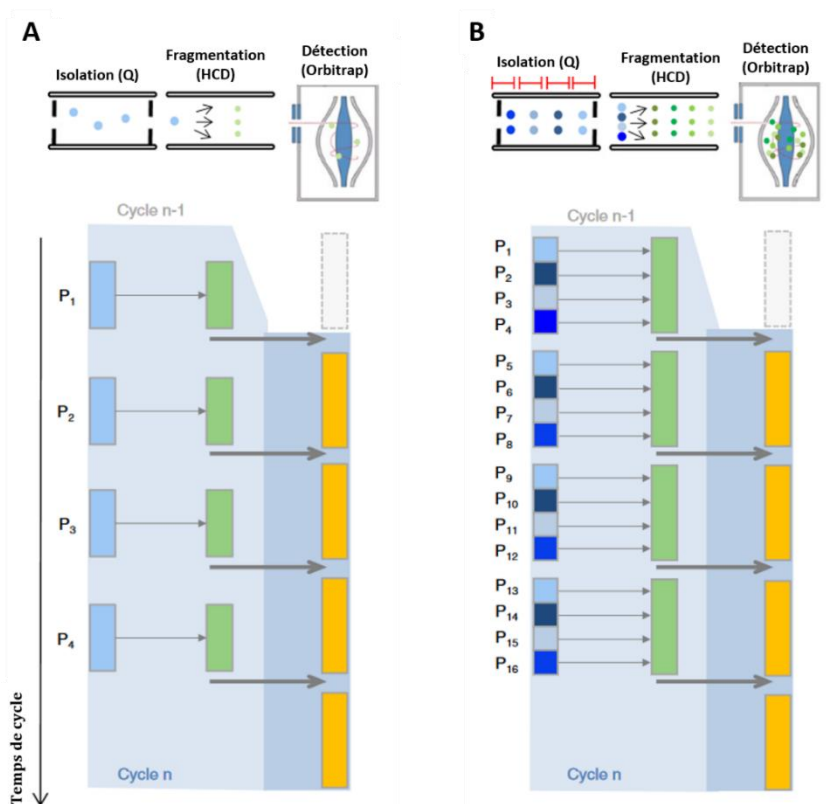


Figure I-11. Représentation de l'analyse PRM dans un Q-Orbitrap. Figure adaptée de (S. Gallien *et al.*, 2014) **A.** En mode générique, les peptides sont analysés de manière séquentielle. Ce mode consiste à isoler un précurseur à la fois (isolement du P1 dans le quadripôle), puis à le fragmenter dans la cellule HCD, et à analyser les ions fragments dans l'Orbitrap. **B.** En mode multiplexé, les précurseurs sont isolés séquentiellement dans le quadripôle (en rouge sont représentés les fenêtres d'isolement pour chaque ion), puis ils sont fragmentés et l'ensemble des fragments générés sont analysés dans une seule acquisition dans l'Orbitrap (P1-P4).

Les analyses PRM peuvent être réalisées dans ces différents modes d'acquisition selon l'objectif de l'étude. Cependant, en mode multiplexé, le temps maximal d'injection et le temps transitoire sont diminués pour augmenter le nombre de peptides analysés, engendrant une perte de sensibilité et sélectivité. (A. Bourmaud *et al.*, 2016) (Figure I-12).

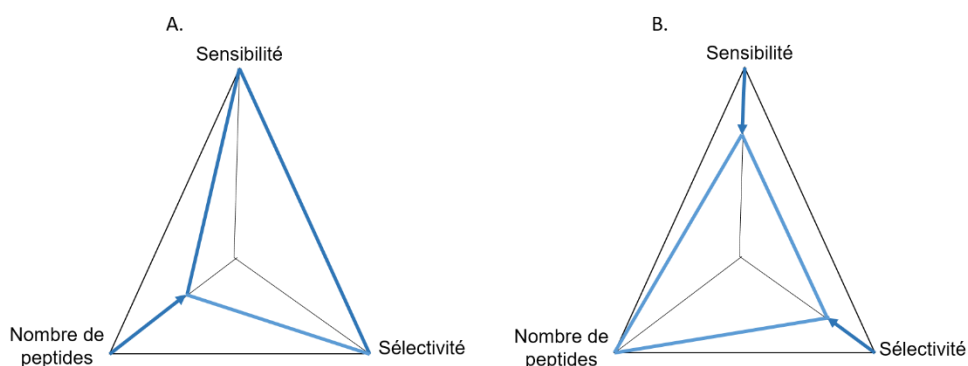


Figure I-12. Avantages et inconvénients des deux modes d'acquisition PRM. A. Acquisition non multiplexée dans laquelle la sélectivité et la sensibilité sont prioritaires par rapport au nombre de peptides analysés. B. Acquisition multiplexée dans laquelle le nombre de peptides analysés est favorisé au détriment de la sensibilité et la sélectivité. Figure adaptée (S. Gallien *et al.*, 2014)

Enfin, comme pour l'analyse LC-SRM, l'analyse LC-PRM peut également s'effectuer selon un mode d'acquisition programmée (*scheduled PRM*), avec des fenêtres temporelles déterminées pour chaque peptide pour augmenter le nombre de précurseurs suivis (A. Bourmaud *et al.*, 2016; S. Gallien *et al.*, 2012).

Au cours de ma thèse, seul le mode générique avec une fenêtre fixe a été utilisé, en mode d'acquisition programmée (Chapitre I.II-1.3.2). Les paramètres utilisés seront détaillés dans Chapitre II.III-1.2.2.2.

Les stratégies de quantification par LC-PRM sont les mêmes que celles décrites pour la LC-SRM internes et reposent sur des standards marqués aux isotopes stables ajoutés en quantité connue dans l'échantillon.

3. Analyse DIA

La troisième approche ciblée consiste à utiliser l'approche DIA présentée dans le paragraphe II-3 (Figure I-4) sur des instruments HR/AM de type Q-TOF ou Q-Orbitrap. Comme décrit précédemment, en mode DIA, un ensemble de précurseurs contenus dans une fenêtre de masse sont sélectionnés, puis ils sont fragmentés simultanément, et les spectres MS/MS

multiplexés sont interprétés. L'intérêt de cette approche est que toutes les informations concernant les ions précurseurs et les ions fragments, dans la limite de détection de l'instrument, peuvent être extraites *a posteriori*. L'avantage principal du mode DIA pour une approche ciblée est qu'aucune information sur les peptides ou protéines d'intérêt n'est nécessaire préalablement à l'analyse. L'analyse ciblée consiste alors à aller rechercher *a posteriori* dans l'ensemble des données acquises celles qui correspondent aux peptides d'intérêt. L'ajout de peptides SIS dans le mélange analysé, permet de confirmer la détection des peptides ciblés et de les quantifier (J. D. Egertson *et al.*, 2013; L. C. Gillet *et al.*, 2012).. Cependant, bien que la quantification reste entre 3 à 10 fois moins sensible que les approches SRM ou PRM (C. Ludwig *et al.*, 2018), le mode DIA reste une option intéressante pour la quantification de protéines peu abondantes. Cette technique a permis la quantification de HCP (*Host Cell Proteins*) au laboratoire (J. Bons, 2019; G. Husson *et al.*, 2018).

Durant ma thèse, une étude préliminaire de cette approche DIA pour la détection de protéines bactériennes dans des biopsies cutanées a été réalisée dans le cadre du projet de diagnostic de la borréliose de Lyme (Chapitre II.III-1.3.2).

4. Traitement des données méthodes ciblées (SRM et PRM)

4.1. Validation de la détection d'un peptide

Les traitements de données SRM et PRM nécessitent la plupart du temps une inspection visuelle et un traitement manuel. Le logiciel le plus couramment utilisé pour ces traitements est le logiciel en accès libre Skyline, développé par le laboratoire de MacCoss (B. MacLean *et al.*, 2010). Ce logiciel permet de sélectionner le bon pic en cas de mauvaise assignation, d'améliorer l'intégration des pics et de simplifier la visualisation des données. Il a été utilisé dans le cadre de mes travaux de thèse.

La **validation de la détection** d'un peptide endogène est effectuée grâce au peptide SIS qui a été ajouté. Plusieurs critères doivent être respectés pour cette validation (Figure I-13) :

1. La co-élution des transitions du peptide endogène d'une part, et du peptide SIS d'autre part (Figure I-13, panel C).
2. La co-élution du peptide endogène et du peptide SIS (Figure I-13, panel E)
3. Le respect des rapports de transitions entre le peptide endogène et le peptide SIS (Figure I-13, panel D). Une valeur est fournie pour mesurer le niveau de similitude entre

les peptides endogènes et peptides SIS (*Rdot product* : rdotp) (Figure I-13, panel A) : plus la valeur est proche de 1, meilleure est la similitude.

4. Lors de l'utilisation de librairies spectrales, la bonne correspondance entre les transitions détectées et la librairie spectrale, avec une mesure de la similitude donnée par le *dot product* (dotp) (Figure I-13, panel A et B).
5. Le temps de rétention observé du peptide endogène par rapport au temps attendu pour ce peptide.
6. Une bonne forme de pic.

Les points 1 à 3 sont les critères principaux pour valider la détection d'un peptide cible. La librairie spectrale apporte une confirmation supplémentaire, mais n'est pas indispensable si le développement a été effectué à l'aide des peptides SIS. Le point 5 est vérifié lors de l'interprétation des données et corrigé manuellement. Enfin, le point 6 est un critère d'exclusion du peptide.

Ces critères ont été utilisés pour la validation des résultats présentés dans le Chapitre II de ce manuscrit. L'ensemble de méthodes ciblé ont été validés à l'aide de rdotp.

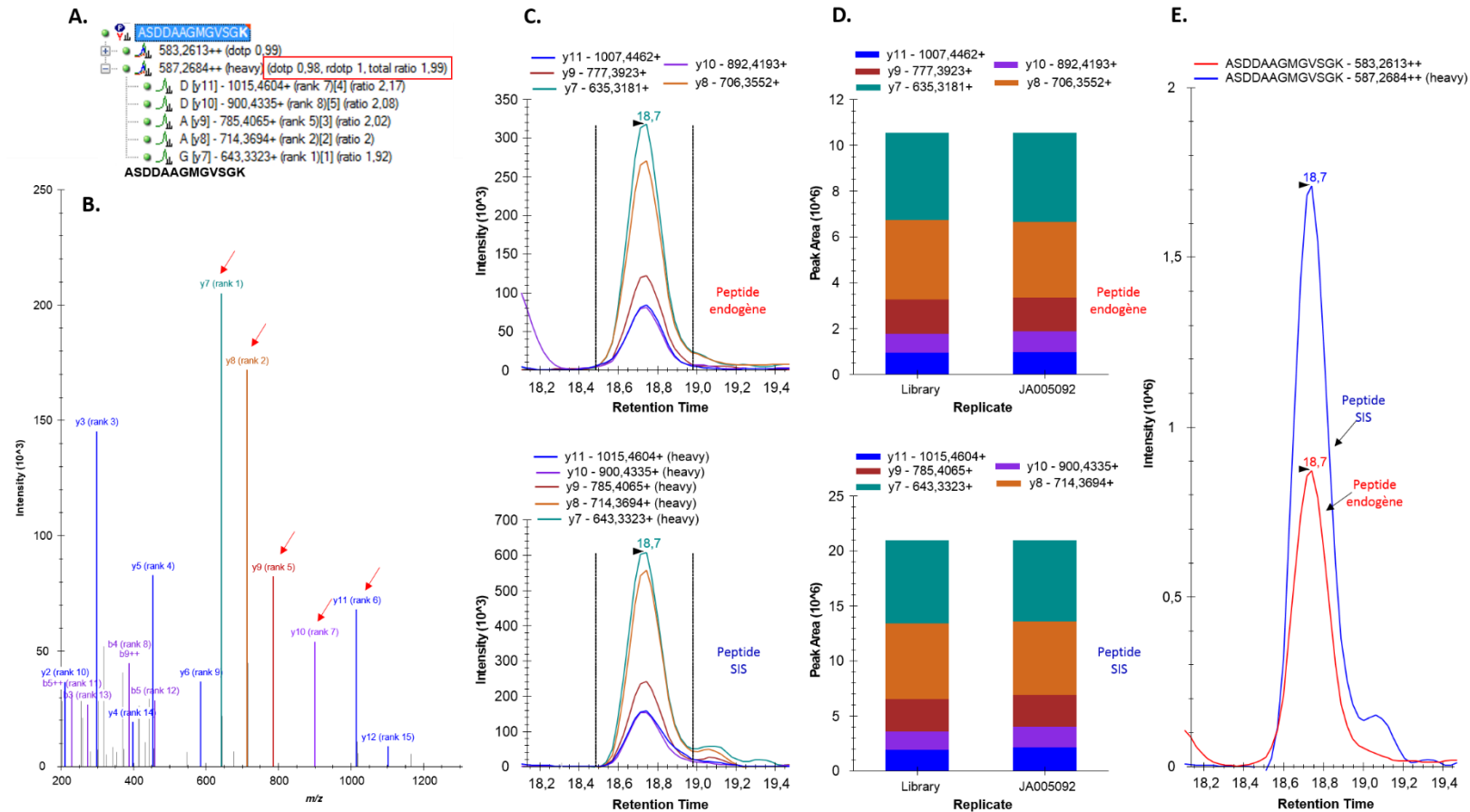


Figure I-13. Exemple de validation de la détection d'un peptide cible (peptide endogène) avec le logiciel Skyline. **A.** Séquence peptidique du peptide, valeurs m/z du peptide endogène et du peptide SIS (heavy), et différents ions fragments b,y. Dans le carré rouge se trouvent les valeurs de dotp (librairie spectrale), rdotp (similitude entre peptides endogène et SIS) et le ratio qui permettra la quantification du peptide ; **B.** Spectre MS/MS du peptide ASDDAAGMGVSGK issu de la librairie spectrale. Les flèches rouges indiquent les transitions détectées ; **C.** Co-élution des 5 transitions détectées pour le peptide endogène et le peptide SIS ; **D.** Respect du ratio des transitions pour le peptide endogène et le peptide SIS. **E.** Co-élution du peptide endogène et du peptide SIS. Les signaux sont obtenus en sommant les 5 transitions pour chaque peptide.

4.2. Quantification absolue

La quantification absolue des peptides d'intérêt est possible grâce aux peptides SIS qui ont été ajoutés en quantité connue dans l'échantillon. En effet, les peptides endogène et SIS sont semblables en tout point, hormis la présence des isotopes, et ils ont le même facteur de réponse en MS. On peut alors utiliser le rapport entre les aires sous la courbe (AUC) mesurées pour le peptide endogène et pour le peptide SIS. Pour chaque peptide, l'AUC est obtenue en sommant les signaux de toutes les transitions détectées. Ce rapport est donné par le logiciel Skyline (par ex ratio 1,99 dans l'exemple de la Figure I-13, panel A).

La linéarité entre l'intensité du signal et la concentration dépend de la gamme dynamique de chaque spectromètre de masse pour une matrice donnée. Ainsi, il est recommandé pour chaque peptide SIS d'adapter sa dilution et d'ajouter une quantité proche de celle attendue pour le peptide endogène ciblé (S. Gallien *et al.*, 2011). Ceci n'est pas facile dans le cas d'échantillons biologiques où la quantité attendue est inconnue. La quantification d'une protéine est faite par l'inférence des peptides quantifiés, et étant donné que les peptides présentent des comportements différents lors des analyses (P. Picotti *et al.*, 2007), une différence de quantification d'une protéine à partir de différents peptides peut être observée. Ce comportement a été constaté dans différentes études (J. Bons, 2019; S. Gallien *et al.*, 2012; C. Schiffmann *et al.*, 2014)

Enfin, selon l'objectif de l'étude, les limites de quantification (haute et basse) peuvent être déterminées pour s'assurer que l'analyse quantitative est faite dans la gamme de linéarité.

Au cours de mes travaux de thèse, la limite basse de quantification (LLOQ, pour *Low Limit Of Quantification*) a été déterminée pour les peptides ciblés dans la méthode SRM de diagnostic de la phase précoce de la borréliose de Lyme (Chapitre II.III-1.2.2.2). Pour cela nous avons utilisé les peptides synthétiques non marqués (NAT) et nous les avons analysés à différentes concentrations dans notre matrice biologique, à savoir une biopsie cutanée humaine.

La LLOQ est la quantité la plus faible pour laquelle la justesse, la linéarité et la répétabilité sont respectées. Pour chaque droite de calibration, ces trois critères (justesse, linéarité et répétabilité) ont été évalués en appliquant des critères de validation couramment utilisés (D. Domanski *et al.*, 2012; G. E. Ronsein *et al.*, 2015). Le premier critère est la répétabilité des analyses - déterminée par le coefficient de variation (CV) des aires de triplicats d'injection pour chaque point de la gamme. Une valeur de CV $\leq 20\%$ a été retenue (Figure I-14). Si la répétabilité des analyses est vérifiée, la linéarité et la justesse sont calculées. La linéarité de

la courbe de calibration est évaluée par le coefficient de détermination R^2 , dans notre cas une valeur minimale de 0,98 a été choisie. La justesse de mesure est définie par l'écart entre la valeur expérimentale et la valeur attendue, dans notre cas les valeurs fixées pour la validation devaient être comprises entre 80 et 120% et un nombre de points égal ou supérieur à trois ($N \geq 3$).

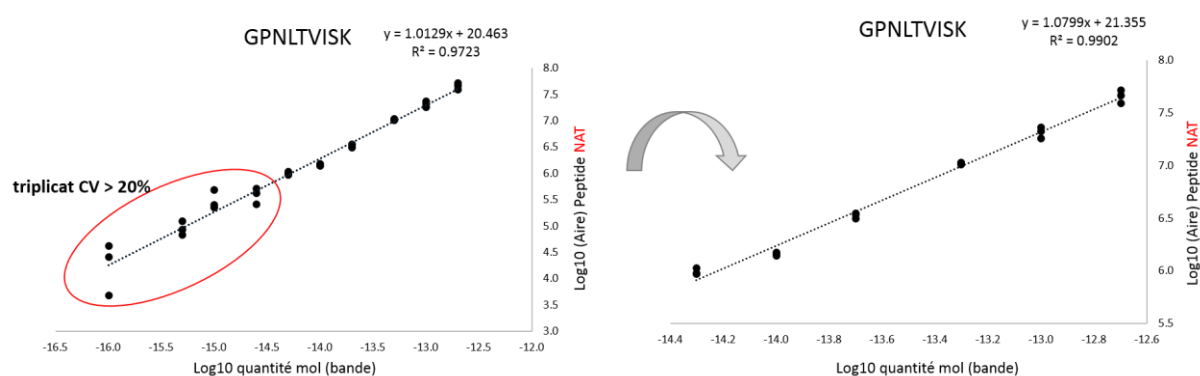


Figure I-14. Courbe de calibration pour la détermination de la LLOQ du peptide GPNLTVISK (OspC). Logarithme de l'aire sous la courbe du peptide NAT en fonction du logarithme de la quantité de peptide déposé dans la bande de gel. A gauche : Courbe de calibration avec 10 points, dont 4 points ne satisfont pas le critère de $CV \leq 20\%$. A droite : Courbe de calibration avec 6 points restants, qui satisfont le critère de $CV \leq 20\%$, $R^2 \geq 0.98$, $N \geq 3$ et justesse comprise entre 80 et 120%. La LLOQ déterminée à partir de cette courbe est 5 fmoles.

III- Quantification relative des protéines

Comme nous venons de le voir, les approches ciblées permettent de quantifier spécifiquement et de manière « absolue » des protéines d'intérêt dans un mélange complexe. Il existe d'autres approches protéomiques quantitatives, basées sur une **étude différentielle des contenus protéiques**. Ces approches permettent de quantifier l'ensemble des protéines contenues dans un échantillon. Ces approches ont été mises en œuvre dès le début de l'analyse protéomique pour étudier par exemple les variations d'abondance des protéines en réponse à une perturbation biologique, pour identifier des marqueurs protéiques corrélés à certaines pathologies, etc. Elles ont évolué en même temps que les stratégies d'identification des protéines. Actuellement nous pouvons distinguer deux stratégies de protéomique quantitative différentielle : les approches avec marquage (*label-based*) et les approches sans marquage (*label-free*) (Figure I-15).

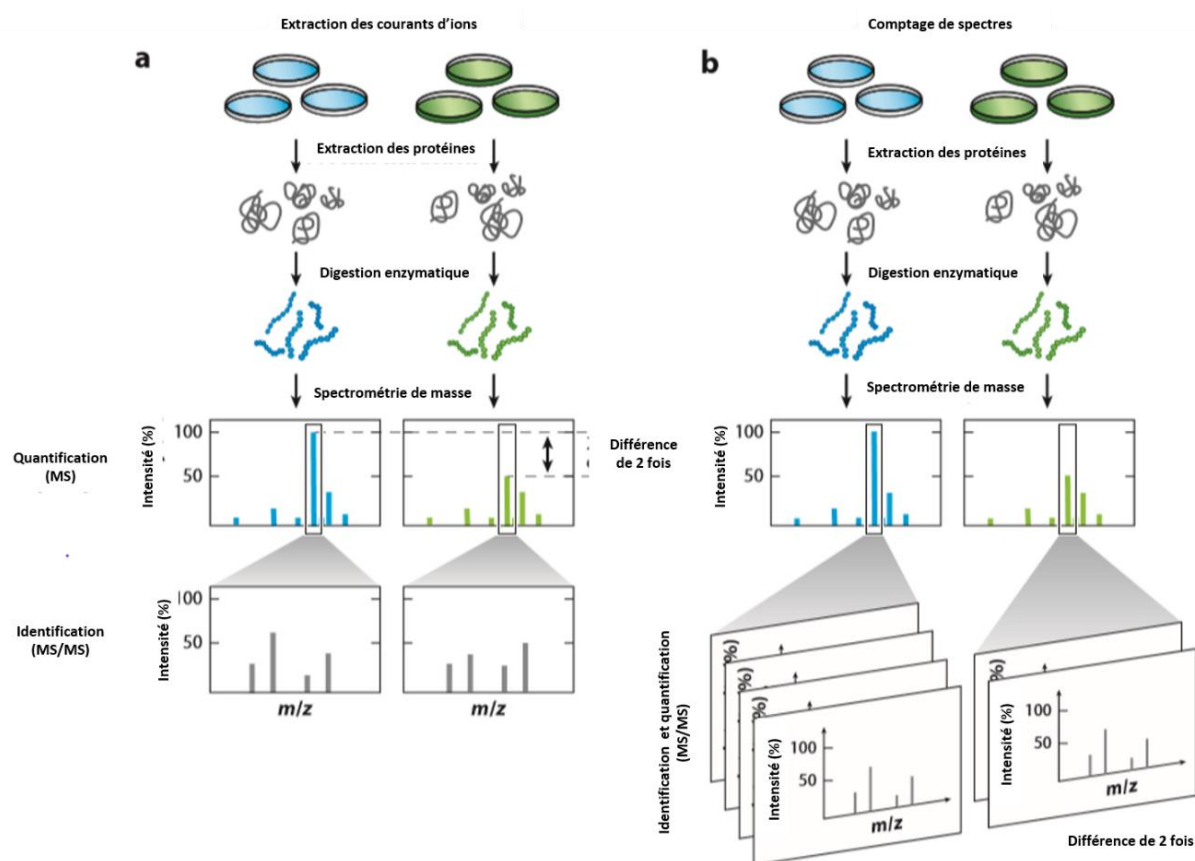


Figure I-15.. a. Approche sans marquage et avec extraction de courant d'ions : les échantillons sont digérés séparément, puis analysés par spectrométrie de masse. La quantification relative est obtenue en comparant les signaux MS pour chaque peptide (quantification MS). b. Approches sans marquage et avec comptage de spectres : les échantillons sont digérés séparément, puis analysés par spectrométrie de masse. La quantification relative est obtenue en comparant le nombre de spectres MS/MS pour chaque protéine. Figure adaptée de (J. A. Ankney *et al.*, 2018)

1. Analyse différentielle avec marquage

Le principe d'une analyse différentielle avec marquage (*label-based*) consiste à comparer une protéine présente dans différents échantillons, après l'avoir marqué par différents isotopes stables (Figure I-15, panel a).

Pour ce faire les protéines d'une condition A sont marquées par un isotope stable, tandis que les protéines d'une condition B sont marquées par un autre isotope. Les deux échantillons sont combinés et analysés simultanément par LC-MS/MS. Cette combinaison pallie les problèmes de variation entre les analyses chromatographiques, une seule analyse étant réalisée. Les peptides marqués par un isotope stable (issus de la condition A) ont les mêmes propriétés que leurs équivalents marqués par un autre isotope (issus de la condition B). Ils

sont co-élués en LC mais sont distinguables en MS et MS/MS. La quantification relative est obtenue en comparant les signaux des peptides marqués par les différents isotopes, pour une séquence donnée. Parallèlement les spectres MS/MS permettent d'identifier la séquence peptidique et de remonter à l'identification de la protéine. Dans certains cas, un multiplexage est possible pour comparer, en une seule analyse, un nombre d'échantillons supérieur à deux. Plusieurs stratégies de marquage ont été proposées au fil des années.

Le marquage métabolique consiste à marquer les protéines intactes par incorporation *in vivo* dans un milieu de culture. Les cellules sont cultivées dans un milieu contenant soit des acides aminés naturels, soit leurs équivalents isotopiquement marqués (^{13}C et ^{15}N). Cette approche a été nommée SILAC pour *Stable Isotope Labeling by Amino acids in Cell culture* (S.-E. Ong *et al.*, 2002). Une alternative Super SILAC a été proposée ultérieurement (T. Geiger *et al.*, 2010b) qui permet d'effectuer la quantification des protéines par l'ajout à l'échantillon d'une culture de cellules marquées représentative de l'échantillon. D'autres marquages *in vivo* ont été proposés comme le SILAM (*Stable Isotope Labeling in Mammals*) (C. C. Wu *et al.*, 2004) et NeuCode (Neutron enCoding) SILAC (C. M. Rose *et al.*, 2013).

Le marquage chimique consiste à marquer les protéines ou les peptides, à l'aide de réactifs chimiques qui réagissent classiquement avec les cystéines ou les lysines, et qui contiennent les isotopes stables. Ces marquages sont applicables aux lysats cellulaires, tissus et fluides biologiques, ce qui représente un avantage par rapport au marquage métabolique

Nous pouvons distinguer :

- Les marquages isotopiques comme le marquage ICAT (*Isotope-Coded Affinity Tag*), qui a été le précurseur dans sa catégorie (S. P. Gygi *et al.*, 1999). Dans l'approche originelle, le réactif ICAT contenait 8 deutériums, ce qui induisait un léger décalage de temps de rétention entre les peptides marqués au ^2D et au ^1H . Pour réduire ce décalage, le marquage au ^2D a été remplacé par un marquage au ^{13}C (Y. Shiio *et al.*, 2006).
- Le marquage qui consiste à remplacer le ^{16}O par le ^{18}O lors de la digestion enzymatique en présence de H_2^{18}O . La quantification est réalisée entre l'échantillon marqué (^{18}O) et non marqué (^{16}O) (K. L. Johnson *et al.*, 2004; X. Yao *et al.*, 2001).
- Les marquages isobariques, qui sont actuellement les plus utilisés, avec des réactifs iTRAQ (*Isobaric Tag for Relative and Absolute Quantification*) (P. L. Ross *et al.*, 2004) ou TMT (*Tandem Mass Tag*) (A. Thompson *et al.*, 2003). Ces réactifs ont l'avantage de pouvoir sélectionner simultanément les peptides marqués au niveau de la MS (car

ils sont tous la même masse), puis de réaliser la quantification relative au niveau de la MS/MS. La capacité de multiplexage est supérieure aux techniques citées précédemment : jusqu'à huit échantillons pour l'iTRAQ et 16 échantillons pour le TMT (G. C. McAlister *et al.*, 2012). D'autres marquages isobariques ont été proposés, tels que le DiLeu (*Dimethylated LEUcine*) (F. Xiang *et al.*, 2010), DiART (*Deuterium isobaric Amine Reactive Tag*) (J. Zhang *et al.*, 2010).

2. Analyse différentielle sans marquage

Avec l'amélioration de la robustesse des systèmes chromatographiques, il est devenu possible de réaliser une analyse différentielle des protéines sans marquage (*label-free*). Dans ce cas les échantillons à comparer sont traités séparément dans des conditions strictement identiques, puis sont analysés individuellement par LC-MS/MS dans des conditions chromatographiques strictement identiques. Cette approche *label-free* nécessite une très bonne reproductibilité du couplage analytique pour pouvoir aligner les chromatogrammes (J. A. Ankney *et al.*, 2018; S. E. Ong *et al.*, 2005). La quantification relative des protéines peut être réalisée au niveau des peptides (MS) ou au niveau des fragments des peptides (MS/MS), et deux stratégies ont été proposées : le comptage des spectres MS/MS et l'extraction des courants d'ions.

2.1. Comptage des spectres MS/MS

Le comptage des spectres MS/MS (*spectral count*) a été la première méthode proposée pour la quantification relative sans marquage en mode d'acquisition DDA (Figure I-15, panel b). Cette méthode est basée sur l'hypothèse que le nombre de spectres MS/MS pour une protéine donnée est corrélé à l'abondance de cette protéine. Plus une protéine sera abondante, plus les peptides générés par digestion seront abondants et plus un nombre important de spectres MS/MS pourront être acquis (H. Liu *et al.*, 2004). Cette stratégie facilite l'étape de traitement de données, puisque que le comptage de spectres se fait directement à partir des données utilisées pour l'identification des protéines. De plus, elle est adaptée aux instruments à basse résolution qui ne permettent pas d'extraire les courants d'ions avec une précision importante (D. H. Lundgren *et al.*, 2010). Cependant, cette approche n'est pas compatible avec l'exclusion dynamique des ions qui permet d'identifier un plus grand nombre de protéines, principalement les protéines les moins abondantes, en excluant, pour une durée définie, les précurseurs une fois qu'ils ont été fragmentés (W. M. Old *et al.*, 2005; Y. Zhang *et al.*, 2009). De plus, l'absence

d'un spectre MS/MS dans une des conditions n'est pas synonyme d'absence de la protéine. En effet, elle peut être due au caractère stochastique du mode DDA et à la non sélection du précurseur correspondant dans une des conditions. Enfin, un autre paramètre à prendre en compte est la présence de peptides partagés, ce qui, avec cette stratégie, implique l'attribution de ces spectres à toutes les protéines possibles et a une influence directe sur la précision de la quantification (Y. Zhang *et al.*, 2010).

2.2. Extraction des courants d'ions

Cette stratégie consiste à extraire les courants d'ions de chaque peptide (précurseur) au niveau de la MS dans les analyses DDA. Cette stratégie est également appelée *quantification MS1* (Figure I-15, panel a). Elle est basée sur le fait que l'intensité mesurée pour un précurseur individuel est corrélée linéairement avec sa concentration (D. Chelius *et al.*, 2002). L'extraction des courants d'ions s'effectue via des logiciels adaptés, tels que Skyline (B. MacLean *et al.*, 2010), Maxquant (J. Cox *et al.*, 2008) et Proline (D. Bouyssie *et al.*, 2020). Cette approche a été rendue possible par l'acquisition de spectres de masse à haute résolution qui permettent de minimiser les interférences au cours de l'extraction des courants d'ions (J. A. Ankney *et al.*, 2018; M. Bantscheff *et al.*, 2007; S. E. Ong *et al.*, 2005). La quantification relative repose sur la mesure de l'AUC d'un peptide donné à un m/z et un temps de rétention donnés entre les échantillons à comparer (J. A. Ankney *et al.*, 2018; S. Cappadona *et al.*, 2012). Les mesures obtenues pour un ion précurseur dans un échantillon peuvent être utilisées pour rechercher le même ion précurseur dans les autres échantillons à comparer. Ceci permet de pallier l'effet stochastique des analyses DDA et représente un avantage par rapport à la stratégie de comptage de spectres MS/MS.

Les approches DIA permettent également d'effectuer des analyses différentielles par extraction des courants d'ions au niveau des fragments (MS/MS). Elles présentent l'avantage d'être plus sensibles et spécifiques du fait d'un meilleur rapport signal sur bruit et d'une plus grande gamme dynamique (J. D. Venable *et al.*, 2004). Cependant, les approches DIA restent actuellement moins utilisées que les approches DDA.

Durant ma thèse, des analyses protéomiques différentielles *label-free* ont été réalisées pour l'étude de *Pseudomonas aeruginosa* dans différentes conditions (chapitre IV). L'abondance relative des protéines a été déterminée par extraction des courants d'ions (quantification MS1)

Le tableau ci-dessous récapitule les approches mises en œuvre pendant ma thèse, ainsi que les instruments et les logiciels utilisés pour les différents projets.

Tableau I-3 Présentation des chapitres et applications de la stratégie *bottom-up* employée. Un résumé des chapitres de ce manuscrit est présenté avec les applications principales des trois étapes du schéma analytique employé. Le choix de chaque technique a été adapté au type d'échantillon analysé et à l'objectif du projet.

Chapitre	Objectif	Echantillon	Préparation d'échantillon	Analyse par LC-MS/MS	Instrument	Traitement de données
II- Application pour l'étude de la borréliose de Lyme	Identification de candidats vaccinaux	Peaux de chiens infectés	Préfractionnement	DDA	nanoAcquity/Q-Exactive+	Mascot OMSSA Proline
	Validation des candidats vaccinaux		Préfractionnement	SRM	microHPLC/QqQ-6490	Skyline
	Diagnostic de la phase précoce	Peaux humaines	Préfractionnement Tube gel	SRM	microHPLC/QqQ-6490	Skyline
				PRM DIA	nanoAcquity/Q-Exactive+ nanoAcquity/Q-Exactive HFX	
	Diagnostic de la phase tardive disséminée	Peaux de souris infectées	Préfractionnement	PRM	nanoAcquity/Q-Exactive+	Skyline
	Effet des dermocorticoïdes et de la lidocaïne		Préfractionnement	DDA	nanoAcquity/Q-Exactive+	Mascot
III-Etude des agents pathogènes transmis par des tiques	Identification des protéines de pathogènes	Tiques Peaux de souris infectées	Préfractionnement	DDA	nanoAcquity/TripleTOF 5600	Mascot OMSSA Proline
IV- Analyse protéomique de <i>Pseudomonas aeruginosa</i>	Etude de l'adaptation phénotypique de <i>P. aeruginosa</i>	Cellules épithéliales infectées	Tube gel Stacking gel Digestion en solution	DDA	nanoAcquity/Q-Exactive+	Mascot Proline ProStar Maxquant Perseus

Chapitre II. Application pour l'étude de la borréliose de Lyme.

I- Borréliose de Lyme

1. Epidémiologie

La borréliose de Lyme, appelée également maladie de Lyme, est la maladie vectorielle la plus répandue dans l'hémisphère nord (Z. Hubálek, 2009; G. Stanek *et al.*, 2012; A. C. Steere *et al.*, 2016). Cette pathologie est causée par des bactéries du complexe *Borrelia burgdorferi* sensu lato (sl) transmises par les tiques dures du genre *Ixodes* : principalement *I. scapularis* et *I. pacificus* en Amérique du Nord, *I. ricinus* en Europe et *I. persulcatus* en Asie (A. C. Steere *et al.*, 2016).

La maladie de Lyme est également recherchée dans l'hémisphère sud. Une revue publiée en 2018 montre l'augmentation de l'intérêt porté à l'étude de la borréliose de Lyme dans des tiques et chez l'homme dans certains pays d'Amérique du Sud (Argentine, Chili, Brésil, Pérou et Uruguay) (A. Robles *et al.*, 2018). Une autre revue, sur l'étude de maladies humaines transmises par les tiques en Australie, évoque cependant l'absence d'une preuve convaincante de la transmission de la borréliose de Lyme dans ce pays (M. Dehhaghi *et al.*, 2019).

L'augmentation des tiques est en partie liée aux modifications climatiques ; elles se retrouvent alors dans un environnement de plus en plus proche de l'homme. L'adaptation de la tique à ces changements pourrait expliquer l'augmentation des cas de borréliose de Lyme et l'intérêt croissant porté à cette maladie.

Depuis quelques années, le genre *Borrelia* est divisé en deux groupes (M. Adeolu *et al.*, 2014; S. J. Cutler *et al.*, 2017; R. S. Gupta, 2019) : (i) le groupe des *Borrelia*, constitué par les espèces responsables de fièvres récurrentes, (ii) le groupe des *Borrelia*, constitué par les espèces responsables de la borréliose de Lyme, complexe *B. burgdorferi* sl. La taxonomie de ces deux genres est bien différenciée, néanmoins cette nomenclature est encore controversée, et les deux nomenclatures sont toujours utilisées.

Les principales espèces de bactéries responsables de la borréliose de Lyme sont (K. D. McCoy *et al.*, 2017; A. C. Steere *et al.*, 2016) :

- *B. burgdorferi* sensu stricto (ss), principalement en Amérique du Nord,
- *B. burgdorferi* ss, *B. afzelii* ou *B. garinii*, et plus rarement *B. spielmanii* et *B. bavariensis* principalement en Europe,
- *B. afzelii* ou *B. garinii* principalement en Asie.

Aux Etats-Unis, l'incidence de cette maladie est suivie par le CDC (*Centers for Disease Control and Prevention*, <http://www.cdc.gov/lyme>), mais probablement largement sous-estimée. Selon les dernières données de décembre 2018, le nombre de personnes diagnostiquées avec la maladie de Lyme est d'environ 300 000 par an.

En France, depuis 2009, le réseau Sentinelles, INSERM/Sorbonne Université (<https://www.sentiweb.fr>) a inclus la maladie de Lyme dans la liste des maladies surveillées. Le taux d'incidence par 100 000 habitants est représenté dans la Figure II-1. En 2018, l'incidence annuelle a été estimée à 68 530 cas, soit 20 000 cas de plus qu'en 2017.

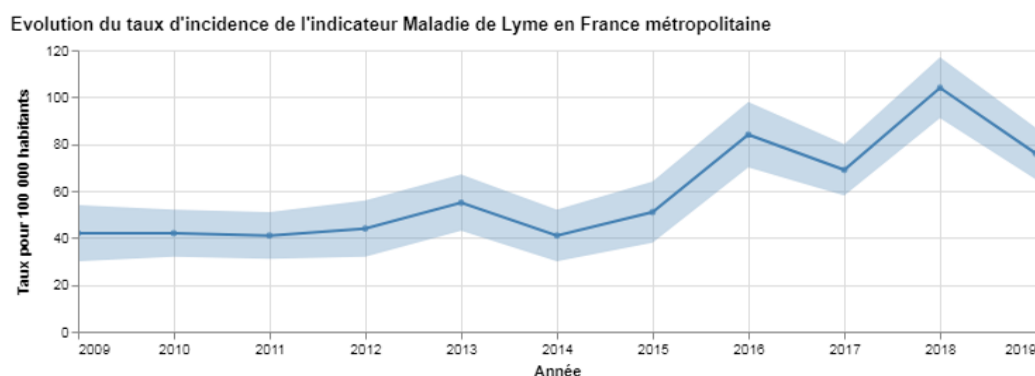


Figure II-1. Evolution de l'incidence de la borréliose de Lyme en France métropolitaine de 2009 à 2019. (réseau Sentinelles, INSERM/Sorbonne Université, <https://www.sentiweb.fr/>)

Ces données montrent l'augmentation du nombre de cas recensés chaque année, probablement suite à la mise en place du plan Lyme gouvernemental en 2016, qui a sensibilisé la population et le domaine médical à cette pathologie (https://solidarites-sante.gouv.fr/IMG/pdf/plan_lyme_180117.pdf). Cependant, les cas sont probablement encore sous-évalués étant donné qu'ils sont issus exclusivement des consultations effectuées en médecine générale en France métropolitaine.

La région Grand-Est est une des régions les plus touchées d'après les dernières données de 2018 (avant le regroupement des régions), les cinq anciennes régions qui la composent faisant partie des sept régions aux taux d'incidence pour 100 000 habitants les plus élevés.

2. Manifestations cliniques

La borréliose de Lyme est une maladie multisystémique, caractérisée par un polymorphisme clinique important. Cette maladie peut se manifester par des atteintes cutanées, neurologiques, articulaires ou cardiaques. La variété de ces manifestations peut s'expliquer par la diversité des espèces de *Borrelia* pathogènes pour l'homme. Il existe des divergences entre les symptômes de la borréliose de Lyme contractée en Europe et aux Etats-Unis (M. E. Aguero-Rosenfeld *et al.*, 2005; G. Stanek *et al.*, 2011a; G. Stanek *et al.*, 2012; A. C. Steere *et al.*, 2016). Sans aucun traitement, les manifestations cliniques de la maladie se déroulent en trois phases : la **phase précoce localisée**, la **phase précoce disséminée** et la **phase tardive disséminée**. Les manifestations cliniques sont spécifiques à chacune de ces phases.

2.1. Phase précoce localisée

Cette phase est caractérisée dans 80% des cas (M. E. Embers *et al.*, 2017; A. C. Steere *et al.*, 2016) par l'apparition d'un érythème migrant (*erythema migrans*, EM), inflammation plus ou moins circulaire d'au moins 5 cm et pouvant atteindre jusqu'à 30 cm de diamètre (C. Eldin *et al.*, 2019; J. Figoni *et al.*, 2019) entre 3 à 30 jours après la piqûre de la tique (Figure II-2). D'autres symptômes comme la fièvre, les maux de tête, la fatigue et les douleurs musculaires et articulaires peuvent apparaître (M. E. Aguero-Rosenfeld *et al.*, 2005; S. Madison-Antenucci *et al.*, 2020). Ces symptômes sont plus fréquemment observés aux Etats-Unis et moins en Europe.



Figure II-2. Exemples d'érythèmes migrants centrés autour du point d'inoculation et pouvant apparaître quelques jours après la piqûre de tique (A. T. Borchers *et al.*, 2015; G. Stanek *et al.*, 2012).

2.2. Phase précoce disséminée

La phase précoce disséminée intervient quelques semaines à quelques mois après la piqûre de la tique. Une propagation de la bactérie peut être observée par l'apparition d'EM multiples.

En Europe, une manifestation cutanée rare est le lymphocytome borrélien, généralement situé dans le lobe de l'oreille chez les enfants (Figure II-3) ou dans la plaque aréolaire chez les adultes, et plutôt due à une infection par l'espèce *B. afzelii* (A. C. Steere *et al.*, 2016). A ce stade, d'autres organes peuvent être atteints, tels que le système nerveux central et/ou périphérique (neuroborréliose), le cœur ou les articulations (mono-arthrite).



Figure II-3. Lymphocytome borrélien dans le lobe de l'oreille d'un enfant
(G. Stanek *et al.*, 2012).

2.3. Phase tardive disséminée

A ce stade les symptômes se développent quelques mois ou années après la piqûre de la tique. Les différences entre les symptômes observés aux Etats-Unis et en Europe sont plus accentuées, principalement en raison des différences entre espèces de *Borrelia* infectantes. L'arthrite chronique est plutôt attribuée à *B. burgdorferi* ss, la neuroborréliose à *B. garinii*, et l'acrodermatite chronique atrophiante (ACA) à *B. afzelii*. Les deux dernières atteintes sont plus observées en Europe (A. C. Steere *et al.*, 2016) (Figure II-4)



Figure II-4. Arthrite, atteinte articulaire des genoux (droite). Acrodermatite chronique atrophiante dans les mains (gauche)
(G. Stanek *et al.*, 2012).

3. L'interface cutanée

L'interface cutanée constitue un site d'interaction privilégié entre l'hôte et la bactérie. C'est d'abord le site d'inoculation de la bactérie au moment de la piqûre de tique. La peau de l'hôte et la salive jouent un rôle important dans la transmission du pathogène. Lors de la piqûre par la tique infectée, la bactérie est injectée via la salive lors du repas sanguin. La protéine Salp15 (*salivary protein 15*), présente au niveau des glandes salivaires se fixe alors sur la bactérie et facilite la transmission et la multiplication des bactéries dans la peau (N. Ramamoorthi *et al.*, 2005). Cette protéine interagit avec la protéine OspC (*Outer surface protein C*) de la bactérie, ce qui facilite la survie de la bactérie dans la peau de l'hôte lors de la transmission (Q. Bernard *et al.*, 2020).

La peau est également un site de multiplication et de persistance de *Borrelia*. Nos collaborateurs ont mis en évidence un pic de multiplication de *Borrelia* dans des biopsies cutanées de souris 5-7 jours après une inoculation des bactéries à la seringue (Figure II-5). Lorsque l'infection est faite via une piqûre de tiques, ce pic est décalé vers 15 jours.

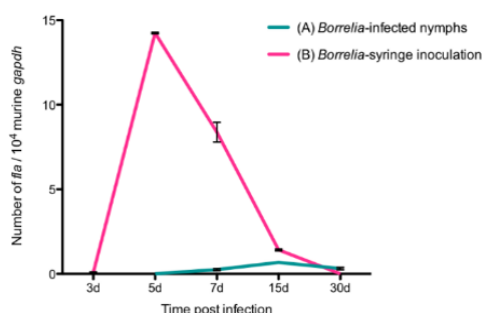


Figure II-5. Quantification par PCR de *B. burgdorferi* dans des biopsies cutanées de souris après une infection via des tiques ou inoculation à la seringue (A. Kern *et al.*, 2011)

Il y a plusieurs années, la persistance de *Borrelia* a été suggérée par Barthold *et al.* dans des biopsies cutanées murines (S. W. Barthold *et al.*, 1993). Des études menées avec nos collaborateurs ont montré une persistance 40 à 90 jours après l'inoculation de la bactérie dans des biopsies cutanées murines (A. Grillon *et al.*, 2017). Il a également été montré dans le modèle murin, une persistance de bactéries vivantes dans le derme jusqu'à 800 jours en absence de toute inflammation (N. Boulanger- M. Wooten : communication personnelle). La persistance de *Borrelia* a été également observée dans des biopsies cutanées chez le chien, traités avec des antibiotiques jusqu'à 30 jours après le traitement (R. K. Straubinger *et al.*, 1997).

Chez l'homme, les EM multiples et le lymphocytome borrélien, qui apparaissent à des endroits éloignés de la piqûre, reflètent l'importance de la peau dans la dissémination et la persistance de *Borrelia* (Q. Bernard *et al.*, 2020).

La multiplication précoce dans la peau et la persistance de *Borrelia* à ce niveau justifient notre choix de travailler sur des biopsies cutanées pour aller rechercher par analyse protéomique des protéines de *Borrelia*, que ce soit pour notre approche vaccinale (Chapitre II. II) ou notre approche diagnostique (Chapitre II. III).

4. Prévention et vaccination.

Le pouvoir d'adaptation des tiques du complexe *I. ricinus* à différents hôtes et différents environnements facilite leur déplacement et leur expansion géographique. A l'heure actuelle, les meilleurs moyens de prévention contre la borréliose de Lyme sont le port de vêtements couvrants et le fait d'éviter les zones de forte incidence de tiques. Dans le cas d'une exposition, la meilleure option reste une vérification minutieuse et le retrait mécanique rapide (pince fine) de la tique en cas de piqûre.

Un autre moyen de prévention est la mise en place de panneaux indiquant la zone de forte influence des tiques afin de sensibiliser les personnes et d'acquérir les bonnes pratiques. Il existe également une application et un site internet « Signalement Tique » (<https://www.citique.fr/signalement-tique/>) développée par l'INRAE, l'ANSES, et le Ministère de la santé dans le cadre du plan Lyme, qui sert principalement à signaler des piqûres des tiques et qui contient des informations diverses telles qu'un guide de prévention, les mesures à prendre en cas de piqûre et une infographie (carte issue des signalements de tiques) qui peut servir de guide pour connaître les zones les plus touchées.

Les répulsifs peuvent être utilisés pour prévenir les piqûres de tiques. Le plus utilisé est le N,N-diéthyl-3-méthylbenzamide (DEET) à appliquer directement sur la peau ou des insecticides comme la perméthrine qui peut être utilisée pour traiter les vêtements (N. Boulanger *et al.*, 2015; S. Madison-Antenucci *et al.*, 2020).

En ce qui concerne la vaccination, un vaccin a déjà été commercialisé aux Etats-Unis (LYMErix™) mais a été retiré du marché en 2002 en raison de controverses et de la nécessité de vaccination régulière. Ce vaccin était basé sur la lipoprotéine de surface OspA et consistait à bloquer la migration de *Borrelia* par des anticorps anti-OspA au niveau de l'intestin de la

tique avant la migration vers les glandes salivaires, inhibant ainsi la transmission à l'hôte vertébré (A. M. de Silva *et al.*, 1996).

Actuellement une nouvelle formulation (VLA15), toujours basée sur OspA, mais sur différentes espèces de *Borrelia* (*B. afzelii*, *B. garinii*, *B. burgdorferi* et *B. bavariensis*) est en cours d'évaluation (P. Comstedt *et al.*, 2017). Valneva et Pfizer ont annoncé une collaboration pour co-développer et commercialiser le vaccin VLA15. En juillet 2020, Valneva a annoncé des résultats initiaux positifs pour la première des deux études de Phase 2 (VLA15-201). Des études complémentaires de Phase 2 sont en cours.

Sur le plan animal, il existe plusieurs vaccins canins pour la prévention de la borréliose de Lyme : les vaccins conçus à partir de protéines recombinantes hautement purifiées VANGUARD[®]crLyme (J. R. Izac *et al.*, 2019) (OspA et OspC), Recombitek[®]Lyme (J. R. Izac *et al.*, 2019; J. A. Rice Conlon *et al.*, 2000) (OspA), et ceux conçus à partir de lysat bactérien : Novibac[®]Lyme (J. R. Izac *et al.*, 2019; R. L. LaFleur *et al.*, 2009), LymeVax[®] (J. R. Izac *et al.*, 2019) et Ultra[™] Duramune Lyme (J. R. Izac *et al.*, 2019; S. A. Levy, 2002), Borrelym 3 et Merilym 3. La composition des souches utilisées pour ces différents vaccins commerciaux n'est pas accessible au grand public.

L'inconvénient des vaccins basés sur le seul antigène OspA d'une espèce donnée est de réduire l'efficacité du vaccin dans un contexte où plusieurs espèces de *Borrelia* circulent comme en Europe. Par conséquent, pour améliorer l'efficacité, il serait intéressant de cibler plusieurs espèces. Comme indiqué précédemment, c'est la stratégie développée par Valneva avec la formulation VLA15. Une combinaison de différents d'antigènes semble également prometteuse.

C'est dans ce contexte que notre laboratoire a démarré en 2014 un projet d'identification de nouveaux candidats vaccinaux, en collaboration avec le groupe de N. Boulanger et B. Jaulhac. Des candidats ont été identifiés par une approche *in vitro* (G. Schnell *et al.*, 2015a; B. N. WO2015022470A2, 2015), puis l'étude s'est poursuivie par une étude *in vivo* dans un modèle murin (projet de prématuration SATT Conectus, Strasbourg). A mon arrivée au laboratoire, un programme de maturation débutait qui avait pour but : (i) de valider chez le chien les candidats vaccinaux identifiés chez la souris, et/ou d'en identifier de nouveaux (ii) de sélectionner les candidats vaccins et de les produire par voie recombinante, (iii) de réaliser les premiers essais vaccinaux chez le chien. Les travaux de ma thèse relatifs à ce projet seront détaillés dans le paragraphe III de ce chapitre.

5. Diagnostic et traitement.

En ce qui concerne le diagnostic de la maladie de Lyme, il repose principalement dans la phase précoce sur l'observation d'un EM et la vérification de l'exposition aux tiques. En effet, la présence d'un EM est suffisante pour poser le diagnostic et le traitement (N. Boulanger *et al.*, 2015; S. J. Cutler *et al.*, 2017; G. Stanek *et al.*, 2012; A. C. Steere *et al.*, 2016).

En cas d'absence d'EM, le diagnostic est plus difficile à réaliser, à cause de la diversité des manifestations cliniques et des symptômes communs à d'autres pathologies. Des tests biologiques peuvent alors s'avérer nécessaires pour confirmer le diagnostic. Les tests diagnostiques de la borréliose de Lyme reposent soit sur la **détection directe**, soit sur la **détection indirecte** de la bactérie.

Les tests de **détection indirecte** consistent à détecter des anticorps dirigés contre la bactérie. Cette détection est réalisée en deux temps (A. T. Borchers *et al.*, 2015; B. Jaulhac *et al.*, 2019; G. Stanek *et al.*, 2011a; A. C. Steere *et al.*, 2016) : un test immuno-enzymatique (**ELISA**) qui s'il est positif ou douteux, est suivi d'une confirmation par test d'immuno-empreinte (**Western Blot**) (Figure II-6). Ces techniques ne sont pas adaptées pour le diagnostic de phase précoce localisée, car la réponse immunitaire apparaît entre 2 à 6 semaines après l'inoculation de la bactérie. De plus, ces tests ne sont pas une preuve d'une infection active. Les anticorps peuvent persister pendant des mois voire des années, même après un traitement (G. Stanek *et al.*, 2011a). La sensibilité de cette technique dans la phase précoce localisée est très faible (30-40%), après six semaines (phase précoce disséminée) la sensibilité peut être supérieure à 90% (B. Jaulhac *et al.*, 2019).

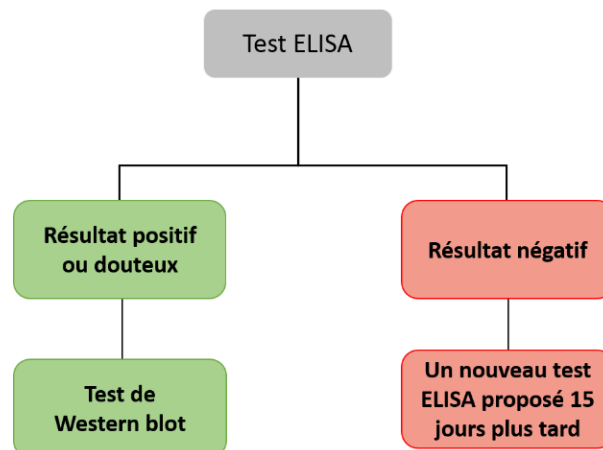


Figure II-6. Protocole de test sérologique pour diagnostiquer la borréliose de Lyme. Un premier test (ELISA) est réalisé. En cas de résultat douteux ou positif, un test par Western blot est réalisé. Dans le cas d'un premier test ELISA négatif, un deuxième test ELISA est proposé 15 jours plus tard.

Les tests de **détection directe** de la bactérie reposent soit sur une analyse microscopique de la culture bactérienne, soit sur l'amplification de l'ADN bactérien par PCR (*Polymerase Chain Reaction*). Ces tests sont réservés à des laboratoires spécialisés. La **PCR**, présente une sensibilité très dépendante de la forme clinique de la maladie et du prélèvement effectué. Les gènes ciblés sont en général *p66*, *16S ANRr*, *fla*, *23S ANRr*, *5S/23S ANRr*, *recA* et *OspA* (M. E. Aguero-Rosenfeld *et al.*, 2005; A. T. Borchers *et al.*, 2015). Dans le cas d'atteintes cutanées, la sensibilité du test réalisé dans des biopsies cutanées est de 69% dans la phase précoce (B. Jaulhac *et al.*, 2019), et plus variable (entre 16 à 92%) dans le cas d'ACA (B. Jaulhac *et al.*, 2019). Une sensibilité variable de ce test a été également observé pour l'arthrite allant de 36 à 85% dans le liquide synovial (B. Jaulhac *et al.*, 2019). Pour le liquide cébrospinal, la sensibilité de ce test est plus faible que pour les deux précédents, avec des valeurs situées entre 15 à 30% (A. T. Borchers *et al.*, 2015). Les tests PCR ne sont pas recommandés pour des échantillons d'urine ou de sang, par manque de sensibilité ou de spécificité (B. Jaulhac *et al.*, 2019). La **culture bactérienne** est considérée comme la méthode de référence (*gold standard*) pour la détection de *Borrelia* avec une analyse PCR comme test de confirmation (A. T. Borchers *et al.*, 2015; G. Stanek *et al.*, 2011a). Cependant, la culture n'est pas utilisée en routine pour le diagnostic du fait de sa variabilité en fonction de l'échantillon biologique, du faible nombre de bactéries présentes initialement dans les tissus ou le liquide biologique, du besoin de matériel spécifique et d'une bonne expertise. La détection de spirochètes ne peut pas être réalisée par une coloration de Gram lors d'un examen de microscopique direct ; un microscope à fond noir est nécessaire pour leur détection. Les bactéries *Borrelia* ont une

croissance lente, et cela peut prendre jusqu'à 12 semaines avant de considérer un résultat négatif (M. E. Aguero-Rosenfeld *et al.*, 2005). Ce temps peut être réduit entre 2 à 8 semaines en cas de résultat positif. La sensibilité de cette technique est de 40 à 70% à partir d'un EM, de 3-17% pour le liquide cébrospinal (A. T. Borchers *et al.*, 2015) et moins de 1% pour le liquide synovial testé lors des arthrites (G. Stanek *et al.*, 2011a).

De nouveaux tests ont fait l'objet d'études mais ne font pas partie des tests recommandés pour le diagnostic de la maladie de Lyme, à cause de leurs sensibilités faibles ou très variables (S. J. Cutler *et al.*, 2017; A. Raffetin *et al.*, 2020). Le test à la chimiokine CXCL13 (chimioattractant des lymphocytes B), qui mesure le niveau de CXCL13 dans le liquide cébrospinal, a été proposé comme indicateur de la neuroborréliose (U. Koedel *et al.*, 2015; T. A. Rupprecht *et al.*, 2018). Ce test présente une sensibilité satisfaisante (89-97%) et une spécificité de 92-98% dans le cas de la neuroborréliose (B. Jaulhac *et al.*, 2019).

En ce qui concerne le **traitement** de la maladie de Lyme, l'antibiothérapie est systématiquement mise en place après l'apparition de l'EM. Le traitement des patients au stade précoce de la maladie entraîne généralement un rétablissement complet et la prévention d'autres manifestations cliniques (S. Madison-Antenucci *et al.*, 2020). Un traitement antibiotique au stade de la phase tardive présente une efficacité plus faible. La durée du traitement est adaptée à la manifestation clinique pouvant aller de 14 à 30 jours, traitement administré la plupart de temps par voie orale, mais pouvant également être administré par voie intraveineuse (Y. Hansmann, 2009; A. C. Steere *et al.*, 2016). Malgré les traitements antibiotiques, certains symptômes peuvent persister. C'est pourquoi il peut être intéressant de développer d'autres approches diagnostiques avec une spécificité, une sensibilité et une fiabilité satisfaisantes, qui pourraient permettre de mettre en évidence une infection active et ainsi prescrire un traitement approprié aux différents stade de la borréliose de Lyme. La difficulté du diagnostic de la borréliose de Lyme réside dans la phase tardive disséminée où les symptômes sont très variés et peuvent mimer d'autres affections.

C'est dans ce contexte que notre laboratoire s'intéresse depuis plusieurs années au développement d'une approche alternative de détection directe des *Borrelia* par spectrométrie de masse, à différents stades de développement de la maladie. Les travaux réalisés par Amandine Bœuf (thèse 2010-2013) (A. Boeuf, 2013) et Gilles Schnell (thèse 2011-2014) (G. Schnell, 2014) ont montré la faisabilité d'une détection spécifique, par préfractionnement SDS-PAGE et LC-SRM, de protéines de *Borrelia* dans la phase précoce de la maladie, d'abord dans la peau de souris puis dans la peau de quatre patients (G. Schnell *et al.*, 2015b). Grâce

à cette preuve de concept, un financement a été obtenu pour la validation de cette méthode sur une cohorte de 70 patients dans la phase précoce (Projet Hospitalier de Recherche Clinique (PHRC, HUS N°6084). Une grande partie de mes travaux de thèse ont porté sur cette validation et seront détaillés dans le paragraphe IV de ce chapitre. Mon principal objectif était de comparer dans cette cohorte de patients, l'approche protéomique ciblée aux deux autres méthodes actuelles de diagnostic direct (culture et PCR). En parallèle, un autre objectif de ma thèse était de poursuivre le développement d'une méthode de détection de marqueurs d'infection dans la phase tardive disséminée (travaux initiés par Benoit Westermann, thèse 2014-2016) (B. Westermann, 2016) dans la perspective d'améliorer le diagnostic de la maladie de Lyme à ce stade.

6. Objectifs

Dans ce contexte les objectifs de mes travaux de thèse étaient de :

- Valider des candidats vaccinaux dans un modèle canin, afin de sélectionner les protéines les plus pertinentes pour le développement d'un vaccin animal.
- Valider à l'échelle d'une cohorte humaine une approche protéomique ciblée comme méthode alternative de diagnostic direct de la phase précoce de la borréliose de Lyme.
- Poursuivre le développement de l'approche protéomique ciblée pour le diagnostic de la phase tardive disséminée de la borréliose de Lyme.

Tous ces travaux ont été effectués en collaboration avec le Dr N. Boulanger et le Pr B. Jaulhac (UR7290, Institut de bactériologie, Strasbourg et Centre National de Référence des *Borrelia*, CHRU Strasbourg). Les cliniciens de la région Grand-Est ont également été sollicités pour les essais cliniques (Prof. Y. Hansmann, Dr. C. Lenormand, Dr. P. Kieffer).

II- Approche vaccinale

Ce travail a été réalisé dans le cadre d'un programme de maturation (SATT Conectus Alsace), en partenariat avec l'entreprise vétérinaire Virbac.

Comme indiqué précédemment les vaccins actuellement commercialisés reposent sur un antigène unique (OspA) ou sur des lysats bactériens. La stratégie mise en place au laboratoire a consisté à identifier les protéines de *Borrelia* exprimées dans la peau d'un hôte infecté, pour proposer de nouveaux candidats vaccinaux.

Des protéines candidates-vaccin optimales sont celles qui satisfont les trois critères suivants (M. R. Kenedy *et al.*, 2012) :

- Protéines exposées au niveau de la surface,
- Protéines communes entre différentes espèces /souches,
- Protéines impliquées dans la transmission du pathogène à l'hôte.

Notre hypothèse était que des protéines bactériennes exprimées dans la peau, de manière précoce après l'inoculation, pouvaient être de bons candidats-vaccin et permettraient de neutraliser les bactéries lors la transmission du pathogène et ainsi bloquer sa dissémination.

Les résultats obtenus avant le début de ma thèse avaient permis d'identifier des protéines exprimées dans la peau de souris au pic de multiplication des bactéries (5-7 jours après inoculation). La suite consistait à mettre en œuvre une approche similaire chez le chien. Trois étapes ont alors été planifiées dans le programme de maturation (Figure II-7).

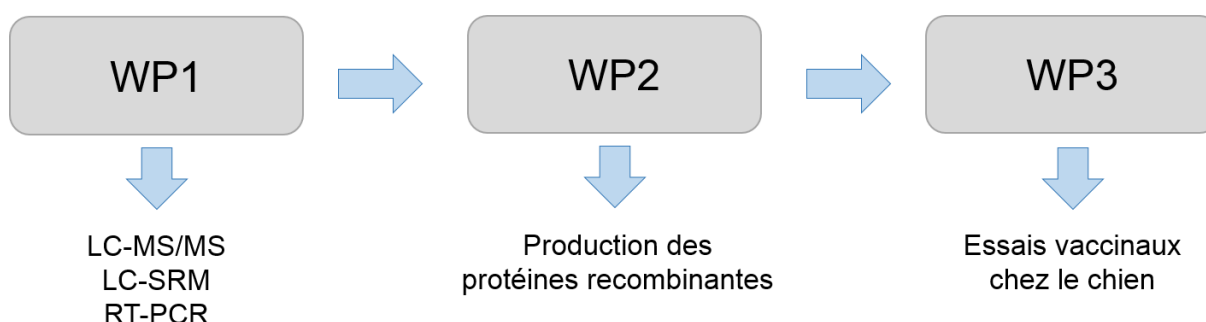


Figure II-7. Représentation des trois étapes (WP) du projet de maturation. **WP1:** Identification de candidats-vaccin chez le chien. **LC-MS/MS** : analyse non ciblée des protéines présentes dans la peau de chiens infectés ; **LC-SRM** : analyse ciblée de certaines protéines bactériennes ; **RT-PCR** : analyse ciblée de gènes codant certaines protéines bactériennes. **WP2** : production des protéines bactériennes d'intérêt. **WP3** : tests vaccinaux chez le chien.

1. Analyse non ciblée des protéines présentes dans les peaux de chiens infectés

1.1. Echantillons

Les expérimentations sur les chiens Beagle (dépôts des tiques infectées, prélèvement des biopsies) ont été réalisées par le Pr Straubinger (Département vétérinaire de l'Université de Munich, Allemagne), dans des conditions validées par un comité d'éthique. Des tiques *I. scapularis* collectées dans l'état du Rhodes Island (USA) ont été déposées sur 5 chiens (A-E),

à raison de 20 tiques par chien, puis des biopsies cutanées ont été prélevées à différents jours (8,14,21) après le dépôt des tiques (Figure II-8).



Figure II-8. Processus d'infection des chiens effectué par le Pr. Straubinger. a. Tiques adultes collectées à Rhodes Island, USA b. Infection d'un chien en déposant les 20 tiques infectées c. Chien infecté, visualisation de la zone d'infection d. Parc de chiens infectés en attente de collecte des peaux

Après les derniers prélèvements (J21) les chiens ont été traités et adoptés par des particuliers. Pour l'ensemble des biopsies 1/3 de la surface a été conservé par le Pr. Straubinger pour la mesure du taux d'infection de la zone prélevée, 2/3 ont été envoyés à Strasbourg pour les analyses protéomiques (LC-MS/MS et LC-PRM) et RT-PCR.

Le Tableau II-1 montre les taux d'infection déterminés par PCR quantitative à l'Institut de bactériologie pour six biopsies. Toutes les autres biopsies étaient négatives. On peut observer qu'ils étaient très faibles par rapport aux résultats déjà obtenus dans le modèle murin, ce qui allait compliquer l'identification des protéines bactériennes.

Tableau II-1. Mesure par PCR quantitative du taux d'infection pour les six biopsies de peau de chien positives

Chien	Jour de prélèvement	PCR <i>fla</i> / 10^4 <i>gapdh</i>
D	8	<0,012
C	14	<0,05
D	14	16,2
D	21	1,5
C	21	5,5
B	21	6,0

1.2. Stratégie analytique

Le défi analytique de ce projet était de pouvoir détecter des protéines de *Borrelia* en concentration très minoritaire par rapport aux protéines de chien. De plus, des études antérieures avaient montré que l'identification de protéines bactériennes dans des biopsies infectées dans les conditions naturelles d'infection (*via* tiques) est plus difficile que lorsque l'inoculation se fait à la seringue avec des quantités injectées plus élevées.

La stratégie mise en place est décrite dans la Figure II-9. Les protéines sont extraites de la peau, puis fractionnées sur gel SDS-PAGE avant d'être digérées pour l'analyse par spectrométrie de masse.

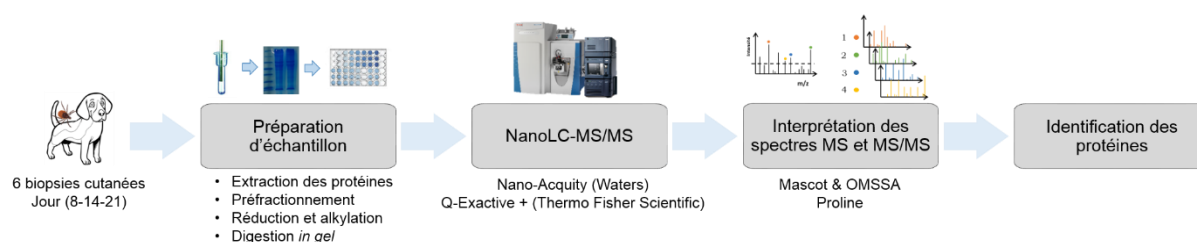


Figure II-9. Stratégie analytique d'identification des protéines dans les peaux de chiens infectés. Les protéines sont extraites de la peau par broyage manuel puis séparées sur gel SDS-PAGE. Les pistes de gel sont découpées en bandes de 2 mm. Après réduction, alkylation et digestion par la trypsine, les peptides sont analysés par nanoLC-MS/MS, les données de masse sont traitées avec deux algorithmes de recherche (Mascot & OMSSA), puis les protéines sont validées par le logiciel Proline.

1.2.1. Extraction des peaux de chiens et dosage des protéines

Pour cette étude, un lavage des peaux canines dans du tampon PBS pH 7 a été réalisé afin d'éliminer le sang, puis l'extraction des protéines a été effectuée avec le protocole de broyage manuel établi au laboratoire pour les peaux de souris (5 mg de peau max dans 185 μ L solution de Laemmli).

L'élimination du sang est importante pour augmenter les chances d'identifier les bactéries de *Borrelia*. En effet, celles-ci sont très minoritaires face aux protéines de l'hôte, principalement la kératine, et toute autre source de protéines telle que les poils de l'hôte ou le sang doivent être évités. Le contenu protéique des échantillons a été ensuite estimé par un dosage DC (*Detergent Compatible*, Biorad) en utilisant la BSA comme référence. Les rendements d'extraction obtenus pour les peaux canines étaient de 125 à 185 μ g protéine / mg peau, comparables aux rendements observés pour les peaux murines (150 à 270 μ g/mg).

1.2.2. Préfractionnement sur gel SDS-PAGE

Lorsque les protéines qu'on cherche à identifier sont très minoritaires dans un échantillon, une méthode robuste est le préfractionnement SDS-PAGE, qui permet de décomplexifier l'échantillon avant les analyses par spectrométrie de masse. Nous avons opté pour cette technique en réalisant un gel 12% SDS-PAGE et un fractionnement en 20 bandes (4 cm de migration). Ce choix a été dicté par les quantifications très faibles obtenues par PCR quantitative (max $16,2 \text{ fla}/10^4 \text{ gapdh}$, Tableau II-1) sur ces biopsies. Les études antérieures faites sur les biopsies murines avaient été faites avec un fractionnement de 10 bandes mais avec un taux d'infection beaucoup plus élevé ($100\text{-}500 \text{ fla}/10^4 \text{ gapdh}$).

1.2.3. Analyse LC-MS/MS et bases de données

Les peptides issus de la digestion enzymatique ont été analysés par nanoLC-MS/MS sur le couplage nanoAcquity/ Q-Exactive+.

La constitution de la banque protéique est essentielle pour l'interprétation des données de MS et MS/MS. Les banques utilisées pour les travaux antérieurs sur souris ayant été créées en 2013 et 2015, une mise à jour était nécessaire. De plus, ces banques ne contenaient qu'une seule souche de *Borrelia* (*B. burgdorferi* B31). Or pour cette approche vaccinale, il était intéressant d'identifier des protéines communes à différentes souches/espèces. Les banques ont donc été modifiées en ajoutant d'autres souches/espèces, puis deux comparaisons ont été faites (Tableau II-2) :

- Première comparaison des résultats obtenus avec la banque A (ancienne banque, NCBI) et la banque B mise à jour (nouvelle banque, NCBI) contenant les autres souches/espèces.

- Deuxième comparaison des résultats obtenus avec les deux nouvelles banques: banque B (NCBI) et banque C (UniProtKB).

Tableau II-2. Banques de données utilisées pour l'identification des protéines dans les peaux canines. Sont indiqués la taxonomie et le nombre de protéines des différents organismes.

		Ancienne banque de données (NCBI) Banque A	Nouvelle banque de données (NCBI) Banque B	Nouvelle banque de données (UniProtKB) Banque C
Espèce	Taxonomie	Nombre de protéines	Nombre de protéines	Nombre de protéines
<i>Canis lupus</i>	9615	38262	58747	28932
<i>B. burgdorferi</i> B31	224326	1758	1533	1327
<i>B. burgdorferi</i> N40	521007		1474	75
<i>B. burgdorferi</i> 297	521009		712	493
<i>B. afzelii</i>	390236		2608	1435
<i>B. garinii</i>	290434		1331	1308

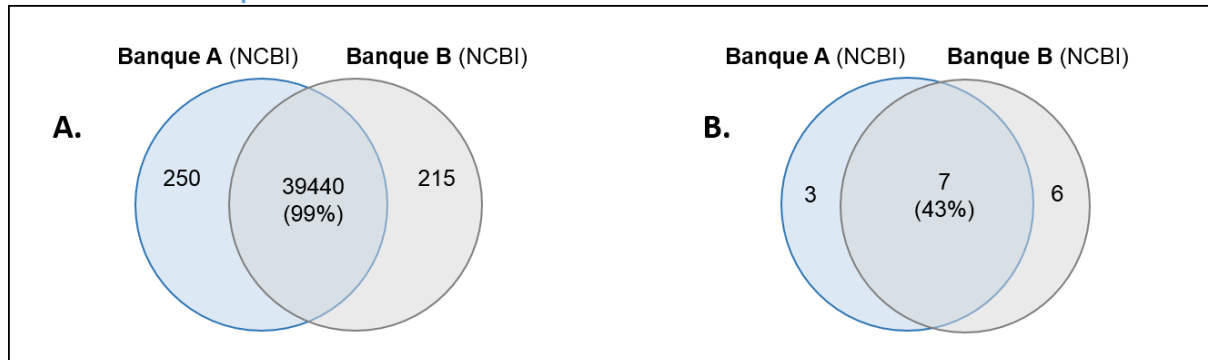
Nouvelles espèces/souches ajoutées

Première comparaison

Deuxième comparaison

La comparaison a été effectuée sur les séquences peptidiques identifiées dans une biopsie canine prélevée 14 jours après le dépôt des tiques. Les résultats utilisés pour la comparaison sont issus d'un seul moteur de recherche (Mascot) (Figure II-10).

Première comparaison



Deuxième comparaison

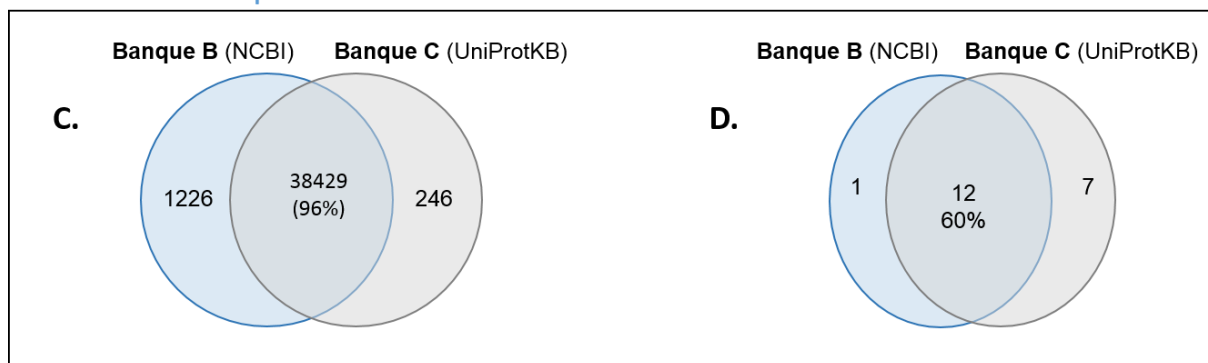


Figure II-10. . Comparaison du nombre de peptides identifiés dans une biopsie canine infectée (J14) en fonction de la banque utilisée. **Première comparaison** : A. Nombre total de peptides identifiés avec la banque A (ancienne banque, NCBI) et la banque B (nouvelle banque, NCBI). B. Nombre de peptides de *Borrelia* identifiés avec la banque A (ancienne banque, NCBI) et la banque B (nouvelle banque, NCBI). **Deuxième comparaison** : C. Nombre total de peptides identifiés les deux nouvelles banques : banque B (NCBI) et banque C (UniProtKB). D. Nombre de peptides de *Borrelia* identifiés les deux nouvelles banques : banque B (NCBI) et banque C (UniProtKB).

Nous constatons que l'impact de la modification des banques est important pour l'identification des peptides de *Borrelia*. Les résultats montrent également la faible quantité de peptides de *Borrelia* présents dans cette biopsie. Au vu de ces résultats, nous avons décidé de travailler avec les deux nouvelles banques UniProtKB et NCBI pour les analyses des biopsies de chien.

1.3. Résultats d'identification des protéines

Nous avons analysé six biopsies de chien, certaines en duplicat, et nous avons identifié au total 92 protéines de *Borrelia* parmi plus de 5600 protéines canines (Tableau II-3). Malgré les taux d'infection très faibles, notre stratégie de préfractionnement sur 20 bandes s'est donc avérée efficace. Vingt-huit protéines de *Borrelia* ont été identifiées dans au moins 2/6 des biopsies analysées, ou avec les deux moteurs de recherche ou dans les deux banques de

données (Tableau II-4). Sans surprise, un nombre plus important de protéines bactériennes a été identifié dans le chien le plus infecté (chien D). On peut aussi observer que la plupart des protéines ont été identifiées avec un seul peptide. Le recouvrement entre les deux algorithmes Mascot et OMSSA a été supérieur dans les biopsies les plus infectées. Les protéines identifiées dans les biopsies les moins infectées l'ont été par un seul moteur de recherche.

Tableau II-3. Nombre de protéines et peptides identifiés dans les six biopsies canines. Les deux panels indiquent les nombres de protéines et peptides de *Borrelia* en rouge et les nombres totaux de protéines et peptides en noir. Panel supérieur : chien B - prélèvement à J21 (2 réplicats) ; chien C - prélèvement à J14 (2 réplicats) ; chien C – prélèvement à J21. Panel inférieur : chien D – prélèvement à J8 ; chien D – prélèvement à J14 (2 réplicats) ; chien D – prélèvement à J21.

	Chien B								Chien C							
	J21_1				J21_2				J14_1		J14_2		J21			
	6.0 flal/10 ⁴ gapdh				6.0 flal/10 ⁴ gapdh				<0,05 flal/10 ⁴ gapdh				5.5 flal/10 ⁴ gapdh			
	UniProtKB		NCBI		UniProtKB		NCBI		UniProtKB				UniProtKB		NCBI	
	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA
Nombre total de protéines/peptides de <i>Borrelia</i> identifiés	3/4	2/2	7/8	2/2	8/8	4/4	2/2	1/1	6/6	7/8	6/6	4/4	11/15	12/16	10/14	9/12
Nombre total de protéines/peptides identifiés	4920/44175	4658/37257	5071/46301	4645/36879	4308/35449	4147/30526	4440/37098	4198/30735	5465/51844	4916/36897	5620/54500	5606/47516	5545/51326	5543/46965	5665/53385	5549/46566

	Chien D															
	J8				J14_1				J14_2				J21			
	0.012 flal/10 ⁴ gapdh				16.2 flal/10 ⁴ gapdh				16.2 flal/10 ⁴ gapdh				1.5 flal/10 ⁴ gapdh			
	UniProtKB		NCBI		UniProtKB		NCBI		UniProtKB		NCBI		UniProtKB		NCBI	
	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA	Mascot	OMSSA
Nombre total de protéines/peptides de <i>Borrelia</i> identifiés	8/8	6/6	8/8	0/0	14/19	14/20	8/13	11/16	12/18	15/18	7/10	14/18	8/9	9/10	8/9	7/8
Nombre total de protéines/peptides identifiés	5226/46979	5024/39912	5358/49263	4177/31233	5252/46030	5015/38209	5347/47731	5326/44022	3984/32269	3965/27432	4098/33461	4198/31138	5481/50968	5359/44184	5632/40319	5322/43919

Afin d'affiner la liste des candidats vaccinaux, j'ai comparé les 28 protéines identifiées chez le chien aux protéines précédemment identifiées au laboratoire dans des peaux de souris infectées, par piqûre de tique ou inoculation à la seringue (Tableau II-4). Neuf protéines identifiées chez le chien avaient déjà été identifiées chez la souris.

Tableau II-4. Liste des protéines identifiées dans les peaux de chiens infectés. Les 28 protéines identifiées dans au moins 2 biopsies, ou avec les deux moteurs de recherche ou dans les deux banques de données ont été comparées aux protéines identifiées dans des peaux de souris infectées (via tique ou via seringue). Les neuf protéines en rouge avaient déjà été identifiées chez la souris.

Protéine	Chien infecté via tique (<i>B. burgdorferi</i> ss) Nb de chien infecté / jour de prélèvement	Souris infectée via seringue (<i>B. burgdorferi</i> ss)	Souris infectée via tique (<i>B. afzelii</i>)	Souris infectée via tique (<i>B. burgdorferi</i> ss)	Souris infectée via tique (<i>B. garinii</i>)
OspC	1/3 à J14	X	X	X	X
Flagellar protéine FlbA flagellar protéine	1/3 à J14				
Flagellar filament outer layer protein	1/3 à J21		X		
Flagellin, Flagellar filament 41 kDa core protéine,	3/3 à J14 - J21	X	X		
hypothetical protéine BB_0758 chez <i>B. burgdorferi</i> B31	3/3 à J8 - J14 - J21			X	
Triosephosphate isomerase	2/3 à J21			X	
Decorin-binding protéine A	2/3 à J14 - J21	X			
lipoprotéine gij365823350	1/3 à J14 - J21	X			
phosphoglycolate phosphatase	2/3 à J8 - J14 - J21				
hypothetical protein BB_O05	1/3 à J14				
Glycerol-3-phosphate dehydrogenase	1/3 à J21				
GAPDH Glyceraldehyde-3-phosphate dehydrogenase	1/3 à J14	X			
50S ribosomal protéine L4	1/3 à J21				
30S ribosomal protéine S7	1/3 à J14				
aminoacyl-histidine dipeptidase	1/3 à J21				
mevalonate kinase	2/3 à J21				
cell division protéine B	1/3 à J21				
Isoleucine tRNA ligase	1/3 à J8				
Transcription termination factor Rho	1/3 à J14				
DNA-directed RNA polymerase subunit beta	1/3 à J14				
RNA polymerase sigma factor RpoD	1/3 à J14				
ATP-dependent Clp protease ATP-binding subunit ClpX	1/3 à J14				
Phosphate import ATP-binding protéine PstB	1/3 à J14				
Putative N-acetylmannosamine-6-phosphate 2-epimerase	1/3 à J21				
Surface lipoprotéine P27 (plasmid)	3/3 à J8 - J14 - J21		X		
Borrelia REV protéine	2/3 à J14 - J21				
OppB protéine, oligopeptide ABC transporter, permease protéine	2/3 à J8 - J14 - J21				
Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protéine	3/3 à J14 - J21				

2. Analyse ciblée des candidats vaccinaux dans les peaux de chiens infectés

Après avoir identifié des protéines de *Borrelia* dans les peaux de chiens infectés, nous avons voulu confirmer leur présence avant de les sélectionner pour le WP2. Nos collaborateurs ont essayé de valider par RT-PCR les protéines OspC, flagelline, DbpA, lipoprotéine gij365823350, lipoprotéine P27, PGP, HP BB0758, REV protéine, OppB mais aucune détection n'a été possible, sans doute à cause du faible niveau d'infection. Nous avons alors entrepris de les valider par une approche protéomique ciblée en utilisant la spectrométrie de masse SRM présentée dans le chapitre I.

2.1. Développement de la méthode LC-SRM

2.1.1. Sélection des protéines

Parmi les 28 protéines du Tableau II-4 nous en avons sélectionné certaines selon les critères suivants:

- Protéines communes entre différentes espèces /souches,
- Protéines identifiées chez la souris et le chien,
- Protéines décrites dans la littérature comme jouant un rôle dans la transmission et/ou la virulence.

Nous avons ainsi obtenu une liste restreinte de neuf protéines (Tableau II-5) dont six ont été identifiées chez la souris et le chien, et trois uniquement chez le chien.

Tableau II-5. Protéines cibles pour les analyses LC-SRM. Les protéines ont été sélectionnées parmi les 28 protéines identifiées dans les biopsies canines.

Protéine	Communes entre espèces/souches	Identifiées dans les biopsies de souris	Identifiées dans les biopsies de chien	Impliquées dans la transmission et/ou virulence
Flagelline	X	X	X	X
OspC	X	X	X	X
HP BB_0758	X	X	X	
DbpA	X	X	X	X
Lipoprotein gij365823350	X	X	X	X
PGP	X		X	X
Lipoprotein P27	X	X	X	X
REV protein	X		X	X
OppB	X		X	X

Pour ces protéines, nous avons aussi recherché les séquences des protéines homologues dans différentes espèces ou souches de *Borrelia*. Cette étape est importante pour générer une liste de protéines la plus exhaustive possible pour la création de la méthode SRM. Un fichier .fasta contenant les séquences des neuf protéines a été créé.

2.1.2. Sélection des peptides

Pour la sélection des peptides, nous avons gardé tous ceux qui avaient été identifiés dans les analyses non ciblées LC-MS/MS (Chapitre II.II-1.3) à condition qu'ils soient protéotypiques

mais pas forcément quantotypiques. Parallèlement, nous avons effectué une digestion *in silico* du fichier .fasta des neuf protéines, pour obtenir une liste de tous les peptides tryptiques. A ce stade, le nombre de peptides était de 162.

L'unicité de ces peptides a été évaluée à l'aide du logiciel Skyline (B. MacLean *et al.*, 2010) dans les cinq banques suivantes :

- Fasta 1 : séquences des protéines de chien et des différentes espèces de *Borrelia* (banque utilisée pour l'identification des protéines Tableau II-2)
- Fasta 2 : séquences des protéines humaines et des différentes espèces de *Borrelia*
- Fasta 3 : séquences des protéines de tique *I. ricinus*, de souris et des différents pathogènes transmis par la tique (Chapitre III, Figure III-5)
- Fasta 4 : séquences des protéines de la famille *Borreliaceae*
- Fasta 5 : banque SwissProt complète

A l'issue de ces vérifications, il est resté 92 peptides d'intérêt. Ces peptides ont été achetés sous forme de peptides marqués par des isotopes stables (peptides SIS, 79 peptides *crude* et 13 peptides AQUA™).

2.1.3. Sélection, optimisation et validation des transitions

Pour des raisons de délai de synthèse et d'obtention des peptides SIS, nous avons développé successivement deux méthodes SRM : une méthode M1 ciblant 6 protéines et une méthode M2 ciblant 4 protéines, avec une protéine commune entre les deux méthodes (Tableau II-6).

Tableau II-6. Protéines ciblées dans les deux méthodes LC-SRM. Deux méthodes LC-SRM ont été développées successivement. M1 : 6 protéines ciblées. M2 : 4 protéines ciblées. La lipoprotéine gi|365823350 est ciblée dans les deux méthodes.

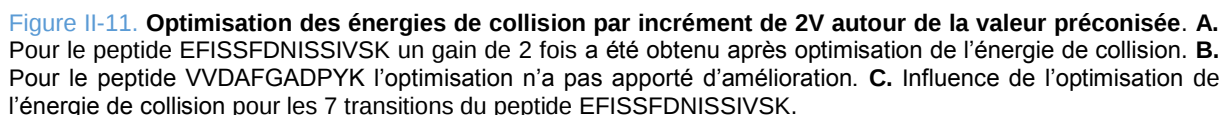
Méthode LC-SRM	Protéines					
M1	Lipoprotéine gi 365823350	OspC	Flagelline	HP BB0758	PGP	DbpA
M2	Lipoprotéine gi 365823350	OppB	Lipoprotéine P27	<i>Borrellia</i> REV		

Les transitions théoriques ont été déterminées à l'aide du logiciel Skyline. Pour le choix des transitions, nous avons conservé au départ toutes les transitions des précurseurs doublement et triplement chargés, en considérant les ions fragments y et b. Les ions b sont très peu utilisés, étant donné qu'ils ont tendance à induire des fragmentations secondaires à la détection (K. W.

Lau *et al.*, 2009), néanmoins ils peuvent être utilisés pour des analyses SRM, dont les paramètres comme l'énergie de collision sont optimisés (C. A. Holstein-Sherwood *et al.*, 2011).

Une fois la méthode créée, le mélange de peptides SIS a été analysé sur un couplage microHPLC 1100/QqQ-6490 (Agilent Technologies) afin de procéder à la sélection des transitions. Les analyses ont été effectuées en mode d'acquisition non-programmé (*unscheduled*), ce qui a nécessité la création de 14 méthodes pour la méthode M1 et six méthodes pour la méthode M2 avec un temps de cycle fixe de 3 s et un temps de balayage de 20 ms, pour obtenir un maximum de 150 transitions par méthode. Un traitement manuel a été effectué pour sélectionner les meilleures transitions, afin de créer la méthode en mode d'acquisition programmé (*scheduled*) (Figure II-12, étape N° 3).

Une fois les transitions sélectionnées, leur optimisation a été effectuée dans des bandes de gel SDS-PAGE riches en protéines issues d'une biopsie canine non infectée. Cette biopsie a été préparée de la même façon que les biopsies canines infectées. Les peptides SIS ont été ajoutés aux bandes de gel juste avant la digestion trypsique, car ils possédaient l'étiquette devant être ôtée par cette digestion (Figure I-8). Cette étape a permis de supprimer les transitions interférentes dues à la matrice puis d'optimiser les énergies de collision en testant 3 valeurs à ± 2 V autour de la valeur préconisée par le constructeur (Figure II-12, étape N°4). Nous avons observé que l'optimisation est peptide-dépendante, conformément à ce qui a déjà été observé au laboratoire et décrit dans la littérature (G. Schnell, 2014). Pour un peptide donné, on voit que le changement de l'énergie de collision a un impact différent selon les transitions (Figure II-11, B).



94

	M1	M2
1 Sélection des protéines	6 protéines	4 protéines
2 Sélection des peptides	62 peptides	30 peptides
3 Sélection des transitions	124 précurseurs 1591 transitions	60 précurseurs 790 transitions
4 Validation de transitions	57 précurseurs 368 transitions	28 précurseurs 164 transitions
5 Méthode finale	6 protéines 52 peptides 112 précurseurs 568 transitions	4 protéines 26 peptides 52 précurseurs 292 transitions

Figure II-12. Stratégie utilisée pour mettre au point les deux méthodes ciblées, M1 et M2. Le mode d'acquisition non-programmé a été choisi pour sélectionner et optimiser les transitions. Les méthodes finales M1 et M2 ciblent respectivement 52 et 26 peptides, ce qui correspond à 568 et 292 transitions suivies.

Les deux méthodes ont été validées en analysant les 20 bandes de gel obtenues par fractionnement d'un extrait de biopsie canine non infectée. Cela nous a permis d'étudier l'effet matrice dans les différentes bandes ainsi que la répétabilité du système chromatographique (exemple d'un peptide de flagelline en Figure II-13). Les variations de temps de rétention pour les peptides SIS dans les différentes bandes étaient inférieures à 0,3 % ce qui montre la bonne répétabilité du système chromatographique.

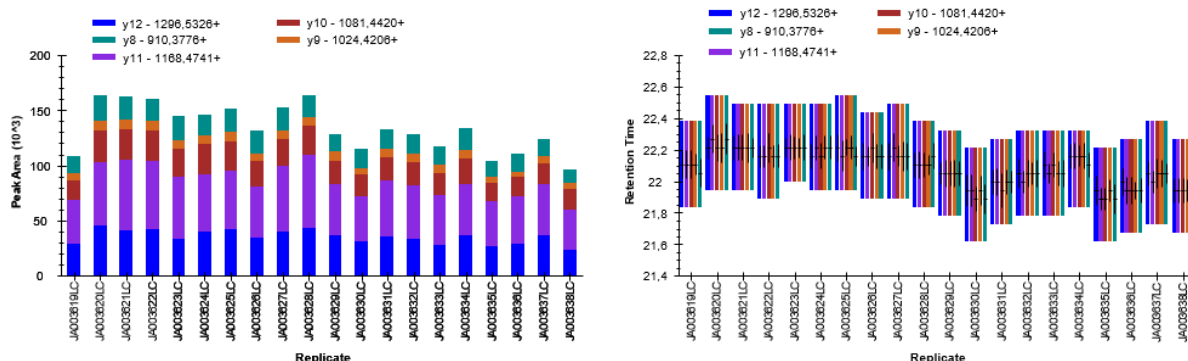


Figure II-13. Exemple de variabilité observée pour le peptide ELAVQSGNGTYSDSDR de flagelline dans différentes bandes de gel. Une biopsie cutanée canine a été fractionnée en 20 bandes, et le mélange de peptides SIS a été ajouté à chaque extrait peptidique issu de chaque bande. A gauche sont représentées les intensités du peptide ELAVQSGNGTYSDSDR dans les 20 analyses LC-SRM, à droite les temps de rétention du même peptide.

2.2. Application aux peaux de chiens infectés

La Figure II-14 montre le traitement qui a été appliqué aux peaux de chiens infectés. Cinq biopsies ont été analysées ce qui a généré une centaine d'analyses LC-SRM.

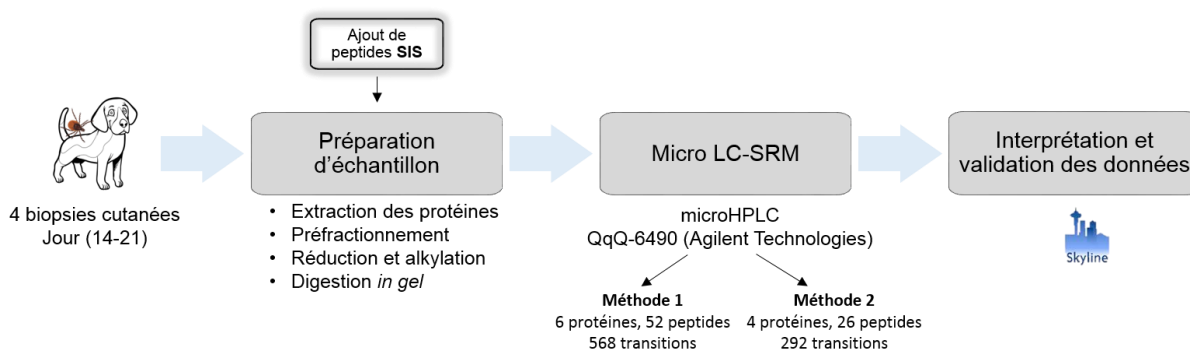


Figure II-14. Analyse des peaux de chiens infectés en ciblant neuf protéines de *Borrelia*. Les protéines sont extraites de la peau par broyage manuel puis préfractionnées sur gel SDS-PAGE (20 bandes). Après réduction/alkylation des protéines, les peptides SIS sont ajoutés aux bandes, puis la digestion par la trypsine est effectuée. Les peptides sont analysés par LC-SRM avec les deux méthodes M1 et M2 (1 protéine commune aux deux méthodes), et les données sont interprétées à l'aide du logiciel Skyline.

Les bandes ont été dopées avec un mélange de 52 et 26 peptides marqués (SIS) pour les méthodes M1 et M2 respectivement.

Un point à souligner est l'importance de l'optimisation de la dilution des peptides SIS lors de ce dopage. C'était une étape cruciale pour la détection des peptides de *Borrelia* dans les

biopsies canines compte tenu des quantités très faibles attendues (cf taux d'infection très faibles mesurés par PCR, Tableau II-1). Les mélanges de peptides SIS ont été dilués et adaptés pour chaque peptide afin d'obtenir un signal relativement faible proche de celui attendu pour les peptides endogènes, et de limiter les effets de suppression de signal (S. E. Abbatiello *et al.*, 2010).

Compte tenu de la quantité de protéines déposées sur le gel (100 µg), et du nombre de bandes (20), nous avons estimé à 5 µg la quantité de protéines dans chaque bande injectée sur le système microLC-QqQ6490 pour chaque analyse (M1 et M2). Les analyses LC-SRM ont été réalisées en mode d'acquisition programmé (*scheduled*) avec une fenêtre de 6 min (M1) et une fenêtre de 4 min (M2) et un temps de cycle de 3 s.

Un mélange de quatre peptides AQUA™ dopé dans un extrait peptidique issu de la digestion de *S. cerevisiae* a été injecté avant l'analyse de chaque biopsie pour vérifier la répétabilité du système chromatographique : un coefficient de variation satisfaisant de 0,3 à 1,0% sur les temps de rétention de ces quatre peptides a été déterminé.

Pour valider les transitions observées, nous avons utilisé les trois critères explicités dans le chapitre I :

- Co-élution des transitions pour le peptide endogène et le peptide SIS,
- Co-élution entre le peptide endogène et le peptide SIS,
- Respect des rapports de transitions entre le peptide endogène et le peptide SIS, $rdotp \geq 0,95$.
- Nous avons également examiné si l'identification du peptide était faite dans la bande attendue.

La Figure II-15 montre un exemple de détection d'un peptide de *Borrelia* appartenant à la flagelline, avec la satisfaction de ces trois critères. On peut aussi observer que le signal du peptide endogène (en rouge) est faible devant le signal du peptide SIS (en bleu).

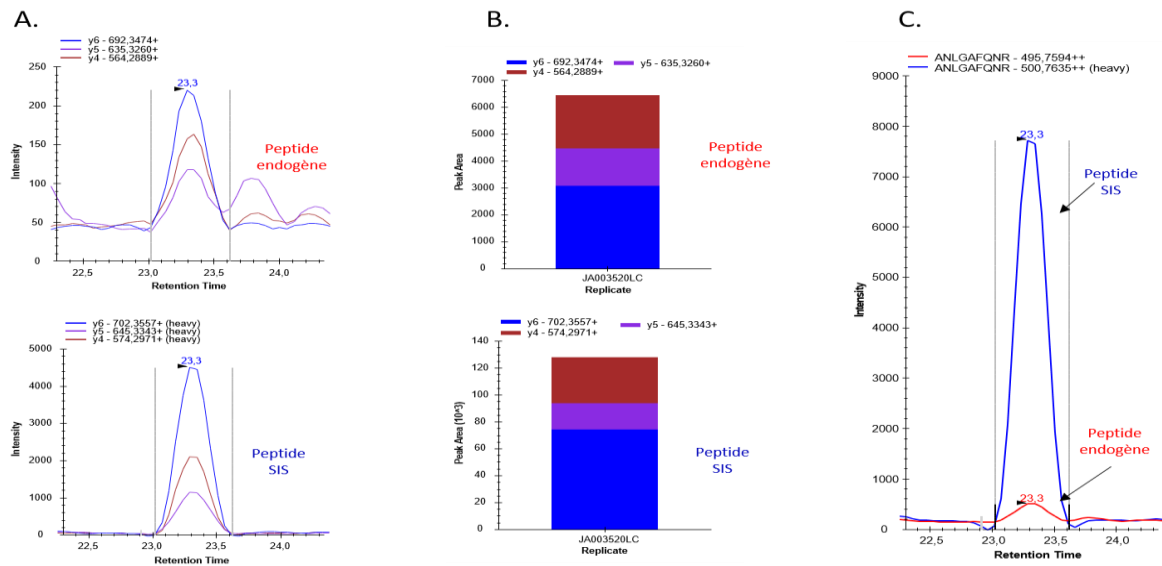


Figure II-15. Exemple de trace SRM obtenue pour le peptide ANLGAFQNR (flagelline) dans une biopsie cutanée de chien infecté. **A.** Co-élution des transitions pour le peptide endogène et le peptide marqué (SIS). **B.** Rapport de transitions du peptide endogène et du peptide SIS. **C.** Co-élution entre le peptide endogène et le peptide SIS.

Le Tableau II-7 montre les résultats obtenus sur quatre biopsies canines infectées. Trois protéines de *Borrelia* parmi les neuf ciblées ont été détectées : flagelline, lipoprotéine gjl365823350 et la lipoprotéine 27. Les deux biopsies qui présentaient un taux d'infection très faible de $<0,012$ et $<0,05$ *fla/10⁴ gapdh* n'ont pas été analysées par LC-SRM. Le tableau permet également de comparer aux résultats obtenus par analyse non ciblée LC-MS/MS pour ces trois protéines.

Tableau II-7. Détection de trois protéines de *Borrelia* dans les peaux de chiens infectés, par analyse non ciblée (LC-MS/MS) ou ciblée (LC-SRM). Dans les deux cas, un préfractionnement SDS-PAGE (20 bandes) a été effectué. Le nombre entre parenthèse indique le nombre de peptides identifiés pour chaque analyse. La lipoprotéine gjl365823350 est une protéine commune aux 2 méthodes M1 et M2.

Protéines	Chien D J21 (1,5 <i>fla/10⁴ gapdh</i>)			Chien C J21 (5,5 <i>fla/10⁴ gapdh</i>)			Chien B J21 (6,0 <i>fla/10⁴ gapdh</i>)			Chien D J14 (16,2 <i>fla/10⁴ gapdh</i>)		
	LC-MS/MS	LC-SRM		LC-MS/MS	LC-SRM		LC-MS/MS	LC-SRM		LC-MS/MS	LC-SRM	
		M1	M2		M1	M2		M1	M2		M1	M2
flagelline		X (1)		X (4)	X (3)		X (1)	X (1)		X (3)	X (2)	
Lipoprotéine gjl365823350	X (2)				X (1)	X (1)				X (3)	X (1)	X (1)
Lipoprotéine P27	X (1)			X (1)		X (1)	X (1)		X (1)	X (1)		X (1)

Ces analyses LC-SRM ont confirmé la présence de ces trois protéines dans les peaux de chiens infectés, avec des peptides supplémentaires qui n'avaient pas été détectés en analyse non ciblée.

Toutefois, il faut préciser qu'aucune quantification absolue n'a été faite dans les peaux de chiens infectés, la présence des peptides ayant été validée avec des peptides SIS de qualité *crude*, c'est-à-dire non quantifiés avec exactitude. Cette quantification n'était pas indispensable pour sélectionner les protéines candidates pour le WP2.

Au final, sur la base des analyses protéomiques, mais également des caractéristiques des protéines (masse moléculaire, membranaire), d'une analyse de la littérature et du potentiel de brevetabilité des candidats-vaccin, les protéines retenues pour le WP2 du projet de maturation ont été les suivantes : **lipoprotéine P27, lipoprotéine gi|365823350, HP BB0758, DbpA et REV protéine**. Les cinq protéines sélectionnées sont des protéines membranaires, dont trois sont des lipoprotéines (P27, gi|365823350, HP BB0758). Ces trois lipoprotéines présentent des caractéristiques immunogènes chez la souris (F. T. Liang *et al.*, 2002). La DbpA, présente dans la membrane externe de la bactérie est connue comme facteur de virulence, permettant la propagation de *Borrelia* chez l'hôte (B. P. Guo *et al.*, 1998). La REV protéine a montré un caractère immunogène chez la souris lors de l'infection suite à la piqûre de tiques (R. D. Gilmore Jr *et al.*, 1998).

3. Production des protéines recombinantes

Cette partie du projet a été réalisée par la société prestataire Bioaster (<https://www.bioaster.org/fr/>). Une des difficultés était d'obtenir des protéines lipidées, ce qui a nécessité l'adaptation du système d'expression. Une première étude de faisabilité a été effectuée selon les étapes décrites dans la Figure II-16.

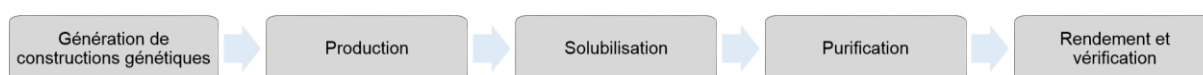


Figure II-16. Etapes de production des protéines recombinantes au sein de l'entreprise Bioaster.

Au cours de cette étude de faisabilité, deux protéines ont présenté une bonne solubilité, deux autres une faible solubilité et une protéine s'est avérée insoluble (la protéine HP BB0758). Le rendement de purification de la lipoprotéine P27 n'a pas été satisfaisant.

Suite à cette étude de faisabilité, ce sont donc trois protéines (**DbpA**, **lipoprotéine gI****365823350** et **REV protéine**) qui ont pu être produites avec une pureté >80%, une lipidation confirmée par LC-MS, et une quantité suffisante (entre 5 et 10 mg). Ces trois protéines ont été formulées par l'entreprise Virbac partenaire du projet pour les essais vaccinaux (WP3).

4. Essais vaccinaux chez le chien

Le but était de tester l'efficacité d'une formulation vaccinale contenant les trois protéines sélectionnées. Cet essai a été réalisé entre septembre 2019 et février 2020 par une société spécialisée (Clinvet, <https://www.clinvet.com/>, USA) sur deux groupes de chien : un groupe de 8 chiens témoins et un groupe de 8 chiens vaccinés (par deux injections à 28 jours d'intervalle). Après 3 semaines, les deux groupes ont été infectés par le dépôt de 50 tiques (25 femelles et 25 mâles) *I. scapularis* infectées par *B. burgdorferi* ss (taux d'infection des tiques de 45%). Un suivi clinique des chiens (poids, température, mesure d'inflammation, etc.) a été réalisé pendant 133 jours, ainsi que des prélèvements pour des analyses histologiques, PCR et sérologiques. Malheureusement, aucune différence n'a été vue entre les deux groupes, au niveau de la détection de *Borrelia* (PCR), de la détection (sérologie) ou de lésions dans les tissus synoviaux. Les chiens ont été infectés par *Borrelia* de manière similaire dans les deux groupes. La formulation utilisée dans cet essai n'a donc pas permis de protéger efficacement le groupe de chiens contre l'infection par *Borrelia*.

L'ensemble des données collectées pendant cet essai vaccinal sont encore en cours d'examen par Clinvet, Virbac et nos collaborateurs, mais plusieurs hypothèses sont actuellement proposées :

- Un nombre de tiques déposées sur les chiens trop important, ayant peut-être conduit à infection trop massive et un débordement du système immunitaire. Une amélioration pourrait être de fractionner le dépôt de tiques sur le chien et de ne déposer que 10 tiques en répétant l'opération à plusieurs jours d'intervalle.
- Deux injections n'ont peut-être pas été suffisantes, une 3^{ème} injection étant couramment réalisée dans ce type d'essais vaccinaux.
- Une combinaison de 3 antigènes dans une seule formulation, qui n'est peut-être pas optimale. Le procédé de formulation ou l'ordre d'administration des antigènes peut avoir une influence sur la réponse immunitaire. Une amélioration pourrait être d'injecter une formule monovalente, et de réaliser soit 3 injections à trois endroits différents.

Même si ce projet de maturation s'est achevé sur un échec de l'essai vaccinal, l'ensemble des approches analytiques mises en œuvre ont montré leur intérêt et leur efficacité dans la détection d'antigènes potentiels au sein de biopsies cutanées d'hôtes infectés. La sensibilité des approches protéomiques a notamment été meilleure que la détection de gènes de *Borrelia* par PCR. Cette sensibilité sera à nouveau abordée dans le paragraphe suivant consacrée à l'approche diagnostique.

Les protéines candidats-vaccins mise en évidence dans le model murin ont fait l'objet de la publication : *Proteomic as an Exploratory Approach to Develop Vaccines against Tick-Borne Diseases Using Lyme Borreliosis as a Test Case*, E. Talagrand-Reboul, B. Westermann, M. A. Raess, G. Schnell, P. Cantero, C. Barthel, L. Ehret-Sabatier, B. Jaulhac, N. Boulanger, *Vaccines* 2020 8(3): E463. DOI: 10.3390/vaccines8030463.

III- Approche diagnostique

Comme nous l'avons vu dans le paragraphe I-5 de ce chapitre, il serait intéressant d'améliorer le diagnostic de la maladie de Lyme, que ce soit au niveau de la phase précoce, mais encore plus au niveau de la phase tardive disséminée. Les méthodes actuelles manquent de sensibilité et de spécificité, et pour pallier ces limitations le laboratoire travaille depuis plusieurs années sur le développement d'une méthode de détection directe des bactéries *Borrelia* dans la peau par spectrométrie de masse ciblée afin de mettre en évidence une infection active.

Durant ma thèse, j'ai eu en charge la validation du diagnostic de la phase précoce par spectrométrie de masse ciblée, sur une cohorte de 70 patients présentant un EM, et la comparaison de cette approche avec les méthodes diagnostiques directes actuelles (PCR et culture bactérienne).

Parallèlement, j'ai poursuivi le développement de l'approche protéomique ciblée pour le diagnostic de la phase tardive disséminée et j'ai optimisé une méthode pour la détection spécifique de protéines de *Borrelia* dans un modèle murin 40 jours après l'infection.

1. Diagnostic de la phase précoce de la maladie de Lyme

Le Projet Hospitalier de Recherche Clinique (PHRC 2015, HUS N° 6084) a été effectué en collaboration avec des médecins de 4 centres hospitaliers dans la région Grand-Est (Strasbourg, Colmar, Mulhouse, Metz). Les prélèvements se sont échelonnés entre 2016 et 2019.

1.1. Stratégie analytique

Le protocole analytique avait été mis en place avant mon arrivée au laboratoire et avait été utilisé pour les premières biopsies humaines (G. Schnell, 2014; G. Schnell *et al.*, 2015b; B. Westermann, 2016). Brièvement, les protéines sont extraites par broyage manuel dans une solution de Laemmli, puis préfractionnées en une dizaine de bandes par SDS-PAGE (12% acrylamide), réduites, alkylées et digérées par la trypsine. Les peptides SIS ne contenant pas d'étiquette, contrairement à ceux utilisés pour l'approche vaccinale, sont ajoutés juste avant l'analyse microLC-SRM et les données sont traitées avec le logiciel Skyline (Figure II-17).

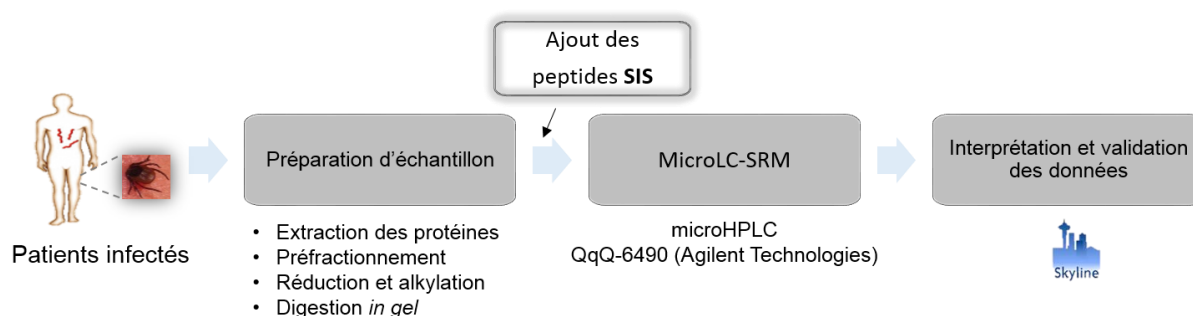


Figure II-17. Stratégie analytique de détection ciblée des protéines de *Borrelia* dans les biopsies cutanées de patients. Les protéines sont extraites des biopsies cutanées par broyage manuel puis préfractionnées sur gel 12% SDS-PAGE. Les pistes de gel sont découpées en bandes de 2 mm. Le contenu protéique de ces bandes est réduit par du DTT, alkylé par de l'IAA, puis digéré *in gel* avec la trypsine. Les peptides SIS sont ajoutés avant l'analyse par microLC-SRM et les données sont interprétées à l'aide du logiciel Skyline.

Les prélèvements ont été effectués au niveau de l'EM par le médecin réalisant l'examen clinique. Certains prélèvements ont été faits sur le tronc, le bras ou la jambe. Deux prélèvements rapprochés ont été effectués à l'aide d'un trépan circulaire de 3 mm (*punch biopsy*, selon la dénomination utilisée en dermatologie). Un prélèvement était destiné à l'analyse PCR et l'observation microscopique de la culture, le deuxième à l'analyse protéomique. Tous les prélèvements ont été codifiés par 6 chiffres (ex 001-001).

1.2. Analyse par protéomique ciblée

1.2.1. Description de la méthode LC-SRM

La méthode LC-SRM avait été mise au point avant mon arrivée au laboratoire (G. Schnell *et al.*, 2015b). Cinq souches différentes appartenant à trois espèces de *Borrelia* (*B. burgdorferi*

ss, *B. afzelii*, *B. garinii*) avaient été sélectionnées. Six protéines (OspC, flagelline, DbpA, GAPDH, énoïase, lipoprotéine gi|365823350) étaient ciblées, correspondant à 56 peptides (Figure II-18) et 458 transitions (229 transitions pour les peptides endogènes + 229 transitions pour les peptides SIS) suivies en mode programmé dans des fenêtres de 10 min. Ces protéines avaient été choisies à partir des analyses protéomiques globales (non ciblées) de peaux de souris infectées. Dans cette méthode LC-SRM, la protéine la plus représentée, avec 31/56 peptides, est OspC, protéine essentielle dans la transmission (D. Grimm *et al.*, 2004) mais très variable selon les espèces/souches de *Borrelia*. On peut voir sur la Figure II-18 que certains peptides d'OspC ne diffèrent que par un seul acide aminé. Pour les 56 peptides SIS, 5 sont de qualité AQUA™, 51 de qualité *crude* (quantification moins exacte). Le couplage utilisé est le même que celui utilisé pour l'approche vaccinale à savoir une chaîne microHPLC Agilent1100 couplée à un spectromètre de masse ESI-triple quadrupole (QqQ-6490, Agilent Technologies).

6 protéines, 56 peptides cibles

OspC (31)		Flagellin (12)	Enolase (2)	GAPDH (2)	DbpA (5)	lipo gi 365823350 (4)
EVEALLSSIDEIAAK	AAQAALTNVSK	AINFIQTTEGNLNEVEK	IQLVGDDLVFTNTSFLK	GGYLDHVNHAGAK	DITDEIDAIK	EFFDWLSK
EVEALLSSIDELSK	AAQVALTNSVK	ANLGAFONR	SGETEDTTIADLVVALGTGQIK	GIDIVAINDLTPK	GVNFDFAK	GEALSFFQK
EVETLLASIDELATK	AILITDAAK	ASDDAAGMGVSGK			TTANGIIEIVK	SLTEIDSGNGIPLVSDVVK
EVETLVLSIDELAK	AQLGVQNGGATNDHAK	ELAVQSGNGTYSDDSR			TVTDAAEQHPPTTAEIGILEIAK	VLTESESNNELK
GGDSASTNPADSAK	AVENLAK	GSIQIEIQLTDEINR			VSENSFILEAK	
GPNLTEISK	AVENLSK	IADQAQYNQMHLNKN				
GPNLTVISK	DGNASANSADSVK	INTPASLSGSQASWTLR				
GVEELEK	DLFESVEGLLK	MIINHNTSAINASR				
IGNNGLEANQSK	DSHAELGIQSVQDDNAK	NNAINAANLSK				
ITDSNAFLAVK	EALNSVK	NNGINAANLSK				
ITDSNAFVLAVK	ELTNPVVAETPK	NSTYAIENLK				
ITDSNAFVLAVK	ELTSPVVAEAPK	TAEELGMQPAK				
LFESVEVLSK	ELTSPVVAESPK					
LSGSLESLSK	ELTSPVVAESPK					
LSVLNSEEK						
NESLIAGAYEISK						
NTSLLAGAYSISTLTK						

Spécifique à une espèce

Figure II-18. Protéines et peptides ciblés dans la méthode LC-SRM pour le diagnostic de la borréliose de Lyme au stade précoce. Six protéines sont ciblées en suivant 56 peptides. Pour OspC, certains peptides sont spécifiques d'une espèce : les différences de séquence sont indiquées en rouge. Les peptides encadrés en bleu correspondent aux séquences pour lesquels les peptides SIS sont de qualité AQUA™.

1.2.2. Analyse de la cohorte de patients : problèmes rencontrés

L'analyse de la cohorte s'est déroulée entre 2016 et 2019 en plusieurs séries d'analyses. Afin de s'assurer du bon fonctionnement du système microLC-QqQ sur cette longue période, nous avons mis en place plusieurs tests de performance : (i) un premier test avec le calibrant fourni par le constructeur (2 ions suivis) (ESI-L Tuning Mix, Agilent Technologies), (ii) un deuxième

test avec un digestat de BSA analysé en triplicat, (iii) un troisième test avec un digestat de *Saccharomyces cerevisiae* dopé avec les 5 peptides AQUA™ de notre méthode SRM. Ces tests ont été faits aussi souvent que nécessaire, avant chaque série d'analyses (i, ii) ou avant chaque biopsie (iii).

Pour valider les transitions observées, nous avons utilisé les trois critères explicités dans le chapitre I :

- Co-élution des transitions pour le peptide endogène et le peptide SIS,
- Co-élution entre le peptide endogène et le peptide SIS,
- Respect du rapport de transitions entre le peptide endogène et le peptide SIS, $rdotp \geq 0,97$,
- Nous avons également examiné si l'identification du peptide était faite dans la bande attendue.

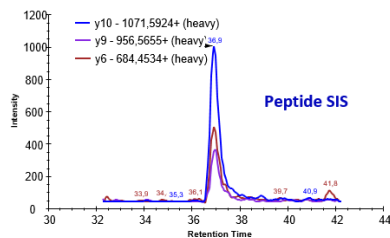
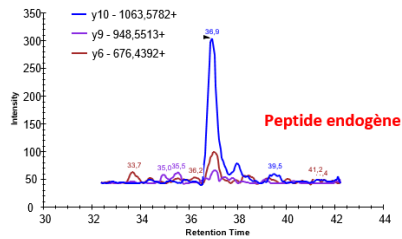
Au cours de l'analyse de la cohorte, je me suis heurtée à de nombreux problèmes dont les plus importants vont être présentés.

1.2.2.1. Problèmes de décalage de Tr

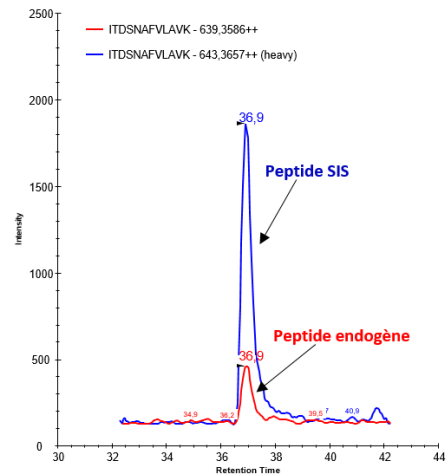
Lors de l'analyse des premières biopsies, un peptide d'OspC de *B. afzelii* (ITDSNAFVLAVK) a passé tous les critères de validation. Il a pu être quantifié grâce au peptide SIS de qualité AQUA™ (14 à 87 fmol d'OspC /mg biopsie selon les biopsies). La présence de ce peptide était en accord avec la positivité par PCR et l'observation microscopique de ces biopsies. Malheureusement suite à des problèmes survenus sur le spectromètre de masse, les analyses ont été interrompues, et lors de la reprise nous avons observé un décalage du temps de rétention (Tr) pour ce peptide.

La Figure II-19 illustre ce problème avec deux répliquats d'analyse d'une même biopsie (001-001). Pour le 1^{er} répliquat (analysé avant l'interruption), on observe une bonne co-élution des transitions ainsi qu'une bonne co-élution du peptide SIS et du peptide endogène : la présence du peptide ITDSNAFVLAVK est validée. Par contre, pour le 2^{ème} répliquat (analysé après l'interruption), on observe un léger décalage entre le Tr du peptide SIS et celui du peptide endogène, ce qui ne permet pas de valider la présence du peptide.

Biopsy 001-001: Replicat 1 ITDSNAFVLAVK_OspC

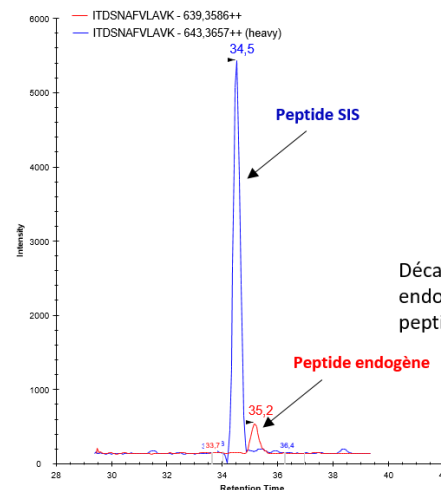
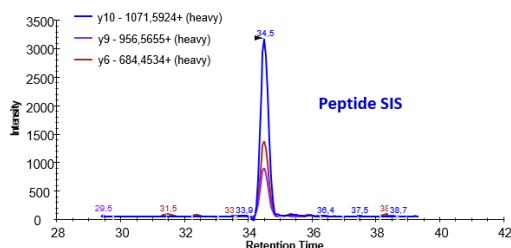
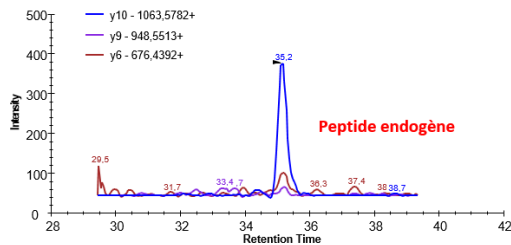


1. Co-élution de transitions



2. Co-élution entre le peptide endogène et le peptide SIS

Biopsy 001-001: Replicat 2 ITDSNAFVLAVK_OspC



Décalage de Tr du peptide endogène par rapport au peptide marqué

Figure II-19. Illustration du décalage de Tr pour deux répliquats d'analyse de la biopsie 001-001. Tous les critères de validation sont satisfaits pour le répliquat 1 (co-élution des transitions, respect du rapport des transitions, co-élution des peptides SIS et endogène). Par contre pour le répliquat 2, les peptides SIS et endogène ne sont pas bien co-élus.

L'analyse d'autres biopsies a confirmé que nous n'arrivons plus à valider la présence de ce peptide ITDSNAFVLAVK, mais que nous observons un signal avec un Tr légèrement décalé par rapport au peptide SIS. Nous avons alors mené différentes investigations.

- Vérification de la qualité et de la séquence du peptide marqué : aucun problème de pureté ou séquence n'a pu être mis en évidence.

- Analyse sur un autre équipement ESI-triple quadrupole : 3 biopsies ont été analysées sur un couplage microLC-ESI-QqQ (TSQ Vantage) et un décalage similaire de 0,5 min entre les Tr des peptides SIS et endogène a également été observé.

Nous avons émis l'hypothèse de la présence d'une isoforme de la séquence **ITDSNAFVLAVK** ciblée. En effet, cette séquence contient une leucine et une isoleucine, et les quatre isoformes de ce peptide (**ITDSNAFVLAVK**, **LTDSNAFVLAVK**, **ITDSNAFVIAVK**, **LTDSNAFVIAVK**) auraient la même masse de précurseur et les mêmes transitions en SRM. Nous avons commandé les trois autres peptides sous forme marquée et avons effectivement observé des Tr légèrement différents pour les quatre isoformes tout en ayant les mêmes transitions (Tr entre 34,4 et 35,7 min, Figure II-20). L'ordre d'élution et le décalage entre les peptides **ITDSNAFVLAVK** et **LTDSNAFVLAVK** pouvait être cohérent avec la présence de l'isoforme **LTDSNAFVLAVK** dans les biopsies de patients (Figure II-20)

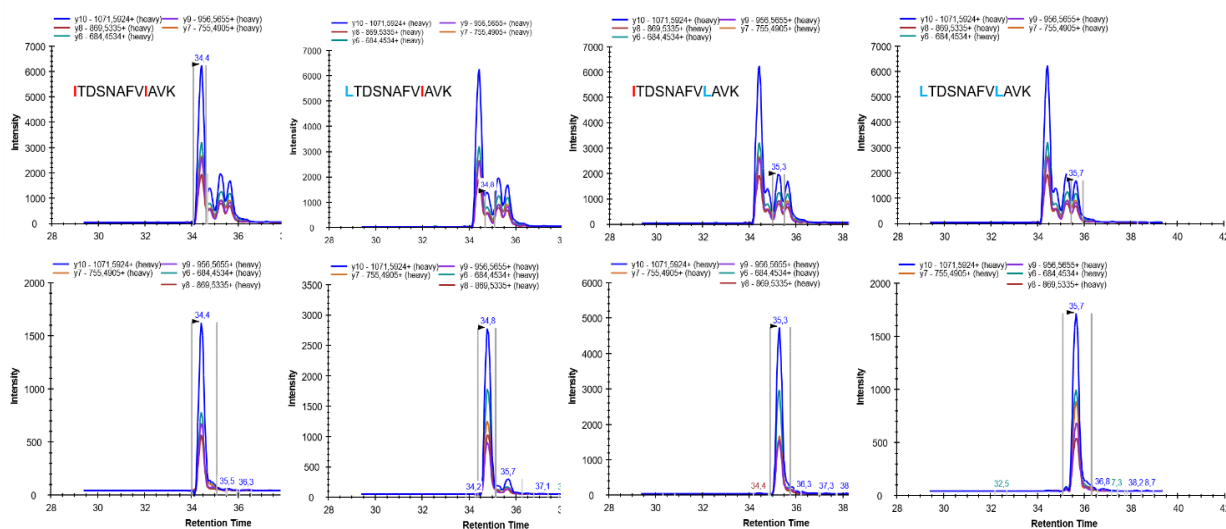


Figure II-20. Analyse des 4 isoformes du peptide (I/L)TDSNAFV(L/I)AVK. La partie supérieure correspond à l'analyse par LC-SRM d'un mélange des 4 isoformes SIS, la partie inférieure correspond aux 4 isoformes SIS analysées séparément.

Nous avons ré-analysé une biopsie humaine (patient 001-003), dans laquelle nous avons observé le décalage de Tr avec le peptide SIS **ITDSNAFVLAVK**, en y ajoutant cette fois le mélange des quatre isoformes SIS. Le résultat suggère que la séquence présente dans cette biopsie est **LTDSNAFVLAVK** (Figure II-21)

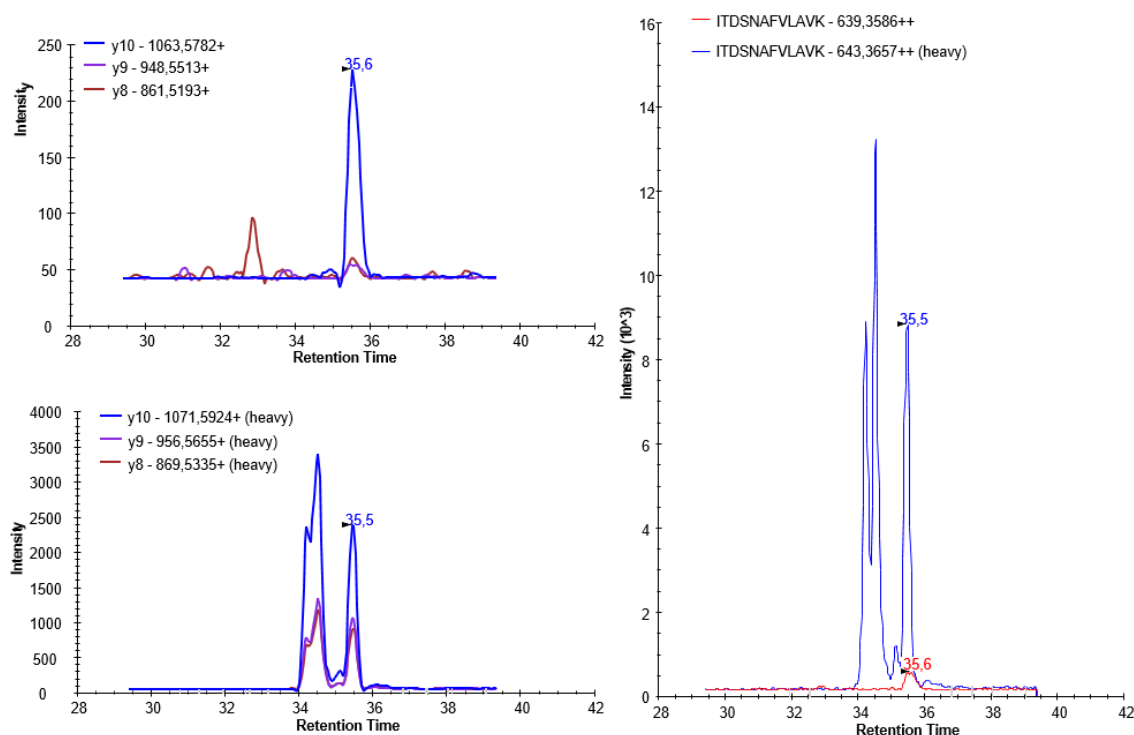


Figure II-21. Trace SRM de la biopsie cutanée 001-003. L'analyse de cette biopsie a été effectuée avec les 4 isoformes du peptide (I/L)TDSNAFV(L/I)AVK. En haut à gauche : trace SRM du peptide endogène présent dans cette biopsie. En bas à gauche : trace SRM des 4 peptides SIS. A droite : superposition des 4 peptides SIS et du peptide endogène. La co-élution du peptide endogène présente dans la biopsie cutanée est observée avec le dernier pic, correspondant à l'isoforme LTDSNAFVLAVK

Cette séquence LTDSNAFVLAVK n'appartenait à aucune protéine de *Borrelia* présente dans les banques de données, mais des variations de séquence dans certaines souches de *Borrelia* pouvaient être imaginées. Nous avons alors analysé deux isolats bactériens obtenus par nos collaborateurs à partir de deux biopsies cutanées de patients de la cohorte, dont celle pour laquelle nous observions le décalage de Tr.

Malgré une quantité faible de matériel biologique dans ces isolats (5 µg d'extrait protéique préfractionné), nous avons pu détecter 3 protéines et jusqu'à 15 peptides de *Borrelia* dans ces isolats (Tableau II-8).

Tableau II-8. Protéines et nombre de peptides détectés par LC-SRM dans les isolats bactériens obtenus à partir de deux patients de la cohorte. Les analyses ont été effectuées en présence des trois autres isoformes (LTDSNAFVLAVK, LTDSNAFVLAVK, ITDSNAFVLAVK) du peptide ciblé (ITDSNAFVLAVK)

Protéines	Nombre total de peptides ciblés dans la méthode	Nombre de peptides identifiés	
		001-001	003-005
OspC	31	3	1
Flagellin	12	10	7
DbpA	5	0	0
Enolase	2	2	0
GAPDH	2	0	0
Lipoprotein gi 365823350	4	0	0
TOTAL	56	15	8

En ce qui concerne OspC, la Figure II-22 montre clairement que c'est l'isoforme ITDSNAFVLAVK qui était présente dans l'isolat bactérien obtenu à partir du patient 001-001, et non pas l'isoforme LTDSNAFVLAVK. Pour l'isolat bactérien obtenu à partir du patient 003-005 pour la protéine OspC, nous n'avons pas détecté le peptide (I/L)TDSNAFV(L/I)AVK.

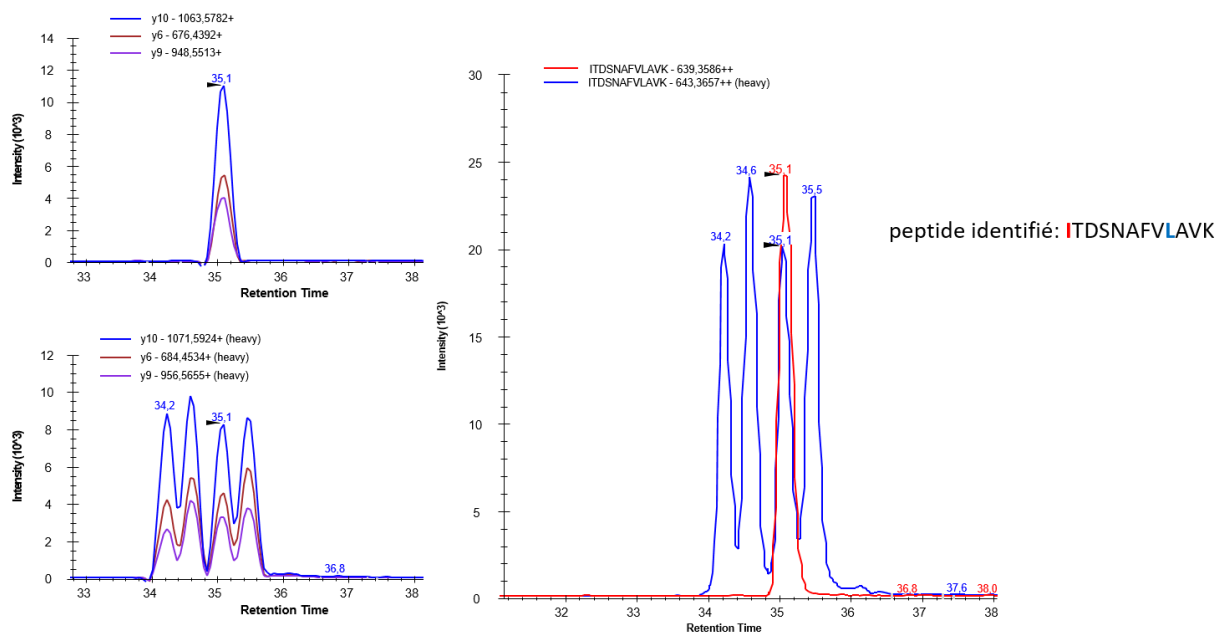


Figure II-22. Trace SRM de l'isolat bactérien obtenu à partir de la biopsie cutanée 001-001. L'isolat bactérien a été analysé en suivant le même protocole que pour les biopsies cutanées. En haut à gauche : trace SRM du peptide endogène présent dans cet isolat bactérien. En bas à gauche : trace SRM des 4 peptides SIS. A droite : superposition des 4 peptides SIS et du peptide endogène. La co-élution du peptide endogène est observée avec le 3^{ème} pic, correspondant à l'isoforme ITDSNAFVLAVK.

Enfin, nous avons également analysé un culot bactérien de *B. afzelli* fourni par nos collaborateurs, et là encore nous avons identifié l'isoforme ITDSNAFVLAVK, qui co-éluit parfaitement avec le peptide SIS correspondant (non montré).

L'ensemble de ces investigations n'ont donc pas permis de confirmer la présence d'une autre isoforme du peptide ITDSNAFVLAVK dans les biopsies de patients, ni de comprendre la raison du décalage entre le Tr du peptide marqué ITDSNAFVLAVK et le signal endogène observé dans certaines analyses.

Pour la suite de l'étude, nous avons décidé de ne pas considérer les biopsies présentant ce décalage de Tr comme positives par spectrométrie de masse et d'analyser la suite des biopsies systématiquement avec les 4 isoformes des peptides SIS.

1.2.2.2. Problèmes de sensibilité

Dans les premières séries d'analyse, nous avons constaté que pour plusieurs biopsies déclarées positives par PCR et par observation microscopique, aucun peptide de *Borrelia* n'était détectable par SRM. Nous avons alors remis en cause la sensibilité de notre approche SRM. Des limites basses de quantification (LLOQ) de certains peptides avaient été déterminées dans la peau de souris (G. Schnell, 2014; G. Schnell *et al.*, 2015b), mais nous avons décidé de déterminer les LLOQ des 56 peptides dans la peau humaine (LLOQ_{SRM}). Nous en avons profité pour évaluer une approche PRM (*Parallel Reaction Monitoring*) qui n'avait encore jamais été testée dans le cadre du projet de diagnostic de la maladie de Lyme. Pour la détermination des LLOQ_{PRM}, la configuration instrumentale était une chaîne nanoUPLC (nanoAcquity, Waters) couplée à un spectromètre de masse QOrbitrap (QExactive+, Thermo Scientific). La méthode PRM a été développée pour suivre les mêmes 56 peptides en mode séquentiel (non multiplexé) explicité dans le chapitre I (A. Bourmaud *et al.*, 2016; S. Gallien *et al.*, 2012). Contrairement au mode SRM, le mode PRM n'est pas limité en nombre de transitions, ce qui pouvait être un avantage pour l'extraction post-acquisition d'un nombre plus important de transitions de peptides de *Borrelia*. Les dilutions des peptides SIS et la quantité de matériel injecté ont été adaptées aux dimensions de la colonne et à la sensibilité de l'instrument.

Nous avons déterminé les LLOQ_{SRM} et LLOQ_{PRM} dans la peau humaine à l'aide des peptides synthétiques **non marqués (NAT)**. Les protéines d'une biopsie humaine saine (servant de matrice) ont été extraites puis préfractionnées par SDS-PAGE (Figure II-23). Puis nous avons sélectionné la bande la plus riche en protéines pour l'ajout avant la digestion des 56 peptides

NAT à différentes concentrations (gamme allant de 100 amoles à 2000 fmoles pour la SRM, et 100 amoles à 200 fmoles pour la PRM). Le mélange des peptides SIS a été ajouté après la digestion, juste avant l'analyse par spectrométrie de masse. Cette différence entre les peptides NAT et SIS provient du changement de fournisseur : peptides NAT (avec étiquette) fournis par JPT Technologies ; peptides SIS (sans étiquette) fournis par Thermo Scientific avant mon arrivée au laboratoire. Une liste des peptides utilisés, avec et sans étiquette, avec les spécifications des fournisseurs se trouve en annexe 1. A ce stade il faut souligner que cette étude a été réalisée avec les 56 peptides SIS en notre possession (5 AQUA™, 51 *crude*). Seules les LLOQ des 5 peptides pour lesquels nous avons les AQUA™ pourraient donc être déterminées avec justesse, les autres seraient estimées. L'utilisation d'une seule bande constituait également une approximation de la situation réelle où les protéines de *Borrelia* pouvaient être réparties dans différentes bandes. Pour la SRM, la totalité de l'extrait peptidique a été injectée (soit $\approx 5 \mu\text{g}/\text{bande}$) sur le couplage microLC-QqQ. Pour la PRM, 6% de l'extrait (soit $\approx 300 \text{ ng}/\text{bande}$) ont été injectés sur le couplage nanoLC-QOrbitrap.

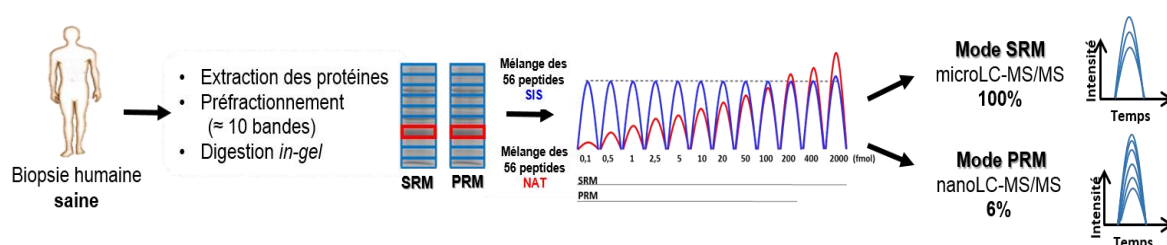


Figure II-23. Détermination des LLOQ des 56 peptides de *Borrelia* par microLC-SRM et nanoLC-PRM. Une biopsie saine humaine a été utilisée comme matrice. Deux extraits protéiques (50 μg) ont été préfractionnés par SDS-PAGE (≈ 10 bandes de 2 mm), puis une bande a été sélectionnée dans chaque piste, une pour la LC-SRM et l'autre pour la LC-PRM. Les peptides **SIS** ont été ajoutés en quantité constante et les peptides **NAT** en quantité variable pour chaque point de la gamme de concentration. Pour la LC-SRM, la totalité de la bande a été injectée (100%). Pour la LC-PRM, une fraction (1/16) de la bande a été injectée (6%). Les analyses ont été effectuées en triplicat d'injection par chaque point de la gamme.

Les critères que nous avons définis pour le calcul des LLOQ ont été explicités dans le chapitre I (D. Domanski *et al.*, 2012; G. E. Ronsein *et al.*, 2015) (Chapitre I.II-4.2) :

- $\text{CV} \leq 20\%$ sur les aires des signaux des peptides NAT pour les triplicats d'injection
- trois points minimum par courbe de calibration ($N \geq 3$)
- $R^2 \geq 0,98$
- justesse entre 80 et 120%

Les résultats des LLOQ_{SRM} et LLOQ_{PRM} sont rassemblés dans le Tableau II-9 ci-dessous.

Tableau II-9. LLOQ des 56 peptides de *Borrelia* par LC-SRM et LC-PRM. Les résultats présentés sont ceux qui respectent les critères définis pour au moins une des méthodes : CV ≤ 20% ; N ≥ 3 ; R² > 0,98 ; justesse entre 80 et 120%. Les 5 peptides AQUA™ sont en gras. Les autres sont de qualité *crude*. N.d : non détecté. Les LLOQ correspondent à la quantité de peptide NAT ajoutée dans la bande de gel (et non à la quantité de NAT injectée dans le spectromètre de masse).

Protéine	Peptide	SRM				PRM			
		Equation de la droite de régression linéaire	N	Coefficient R ²	LLOQ _{SRM} (fmol)	Equation de la droite de régression linéaire	N	Coefficient R ²	LLOQ _{PRM} (fmol)
OspC	GNLITEISK	y = 0,8441x + 15,749	4	0,9854	1	y = 1,2549x + 24,013	4	0,9820	10
	NESLIAGAYEISK	y = 0,9444x + 16,72	4	0,9903	1	y = 1,2242x + 23,073	5	0,9931	10
	LSVLNSEEK	y = 1,0945x + 19,34	9	0,9906	2,5	y = 1,2877x + 24,192	5	0,9954	10
	LSGSLESLSK	y = 1,1044x + 19,502	9	0,9920	2,5	y = 1,0758x + 21,371	7	0,9917	2,5
	LFESVEVLSK	y = 1,104x + 19,283	7	0,9865	2,5	y = 1,1582x + 22,276	4	0,9865	10
	ITDSNAVLLAVK	y = 0,9342x + 16,96	9	0,9868	2,5	y = 1,285x + 23,394	5	0,9901	10
	ITDSNAFLAVK	y = 0,8053x + 14,51	4	0,9842	2,5	y = 1,2562x + 23,320	5	0,9863	10
	GVEELEK	y = 0,7899x + 13,601	3	0,9959	50	y = 0,9142x + 19,198	6	0,9862	5
	GNLTVISK	y = 1,3129x + 21,873	6	0,9862	10	y = 1,0799x + 21,355	6	0,9902	5
	GGDSASTNPADESAK	y = 0,8064x + 15,225	9	0,9819	1	y = 0,9573x + 19,747	6	0,9936	5
	EVEALLSSIDELSK	y = 0,9925x + 17,499	6	0,9919	2,5	y = 1,3945x + 25,410	4	0,9822	10
	EVEALLSSIDEIAK	y = 0,9139x + 16,032	4	0,9819	5	y = 1,3312x + 24,863	3	0,9865	10
	ELTSPVVESPK	y = 0,8823x + 16,438	10	0,9839	1	y = 1,2647x + 23,923	5	0,9871	10
	ELTSPVVAESPK	y = 1,0559x + 18,568	8	0,9964	2,5	y = 1,0522x + 21,057	4	0,9808	2,5
	ELTSPVVAEAPK	y = 0,8246x + 15,415	6	0,9810	1	y = 1,2516x + 23,994	5	0,9890	10
	ELTNPVVAETPK	y = 0,8831x + 16,256	4	0,9913	2,5	y = 1,0644x + 21,853	6	0,9848	2,5
	AQLGVQNGGATNDHAK	y = 1,0499x + 18,121	4	0,9926	100	y = 0,9763x + 19,959	4	0,9967	5
	AILITDAK	y = 1,051x + 18,931	9	0,9968	2,5	y = 0,8809x + 19,459	6	0,9828	1
	DGNASANSADSVK	y = 1,0773x + 18,578	7	0,9931	2,5	y = 0,9422x + 19,685	6	0,9918	5
	AAQAALTNSVK	y = 1,0467x + 18,422	8	0,9967	2,5	y = 0,9918x + 20,965	7	0,9927	2,5
	AVENLAK	y = 0,918x + 16,797	8	0,9887	5	N.d	N.d	N.d	N.d
	AAQVALTNSVK	y = 1,1004x + 19,427	9	0,9923	2,5	N.d	N.d	N.d	N.d
	ITDSNAFLAVK	y = 1,1139x + 19,001	4	0,9845	2,5	N.d	N.d	N.d	N.d
	NTSLLAGAYSISTLITK	y = 1,5003x + 23,471	3	0,9903	20	N.d	N.d	N.d	N.d
	IGNNGLEANQSK	y = 1,1239x + 18,979	6	0,9985	1	N.d	N.d	N.d	N.d
	EVETLVLSIDELAK	y = 1,2633x + 20,138	3	0,9955	200	N.d	N.d	N.d	N.d
	EVETLLASIDELATK	y = 0,9282x + 15,985	4	0,9835	1	N.d	N.d	N.d	N.d
	EALSNSVK	y = 0,8989x + 16,591	6	0,9911	1	N.d	N.d	N.d	N.d
	DSHAELGIQSVQDDNAK	y = 1,1298x + 19,403	7	0,9898	10	N.d	N.d	N.d	N.d
Flagellin	ANLGAQNR	y = 1,0568x + 19,224	7	0,9967	1	y = 1,148x + 22,940	6	0,9817	5
	NSTEYAIENLK	y = 1,0146x + 17,593	4	0,9946	5	y = 1,2698x + 24,120	5	0,9959	10
	NNGINAAANLSK	y = 1,0308x + 17,951	4	0,9932	100	y = 1,0318x + 20,607	6	0,9931	5
	NNAINAAANLSK	y = 1,0572x + 18,637	9	0,9950	2,5	y = 0,9693x + 20,371	7	0,9836	1
	AINFIQTTEGNLNEVEK	y = 1,4266x + 22,736	4	0,9954	10	N.d	N.d	N.d	N.d
	INTPASLSGSQASWTLR	y = 1,405x + 22,658	3	0,9973	100	N.d	N.d	N.d	N.d
	TAEELGMQPAK	y = 0,9677x + 17,218	3	0,9939	5	N.d	N.d	N.d	N.d
	MIINHNTSAINASR	y = 1,0938x + 18,045	4	0,9967	100	N.d	N.d	N.d	N.d
	IADQAQYNQMHMLSNK	y = 0,9115x + 16,201	4	0,9875	20	N.d	N.d	N.d	N.d
	ELAVQSGNGTYSDDSR	y = 0,9808x + 15,967	3	0,9969	100	N.d	N.d	N.d	N.d
DbpA	ASDDAAGMGVSGK	y = 0,8854x + 15,867	4	0,9911	5	N.d	N.d	N.d	N.d
	VSENSFILEAK	y = 1,0392x + 18,212	9	0,9929	2,5	y = 1,2365x + 23,286	5	0,9932	10
	DITDEIDAIK	y = 0,668x + 12,396	6	0,9850	2,5	y = 1,2825x + 24,123	5	0,9944	10
	TVTDAAEQHPHTTAEGILEIAK	y = 1,1552x + 19,374	5	0,9979	2,5	N.d	N.d	N.d	N.d
	TTANGIEIVK	y = 1,2424x + 20,889	6	0,9876	10	N.d	N.d	N.d	N.d
GAPDH	GVNFDAFK	N.d	N.d	N.d	N.d	y = 1,1329x + 21,891	5	0,9844	10
	GIDIVAINDLTPK	y = 0,7423x + 14,037	7	0,9812	2,5	y = 1,2059x + 22,256	3	0,9866	10
Enolase	GGYLDHVNHAGAK	y = 1,0617x + 18,124	4	0,9961	100	y = 0,7275x + 16,104	3	0,9872	5
	SGETEDTTIADLVVALGTGQIK	y = 1,1405x + 19,449	3	0,9942	200	N.d	N.d	N.d	N.d
lipoprotein gij365823350	IQLVGDDLFTNTSFLK	y = 0,8351x + 14,865	4	0,9896	1	N.d	N.d	N.d	N.d
	EFFDWLSK	y = 1,0056x + 17,708	10	0,9930	1	y = 1,3298x + 24,421	4	0,9856	10
	VLTESESNELK	y = 1,0243x + 18,005	6	0,9949	1	y = 1,0294x + 21,363	7	0,9889	2,5
	GEALSFFQK	y = 1,0169x + 16,395	7	0,9818	5	N.d	N.d	N.d	N.d
	SLTEIDSGNGIPLVSDVVK	y = 1,1061x + 18,719	4	0,9981	2,5	N.d	N.d	N.d	N.d

Nous avons déterminé les LLOQ de 53 peptides sur les 56 ciblés : 52 par LC-SRM et 31 par LC-PRM (Figure II-24). Pour 30 peptides, les LLOQ ont été déterminées par les deux méthodes. Parmi ces peptides, 4 ont la même LLOQ par LC-SRM et LC-PRM, 7 sont quantifiables à des concentrations plus faibles par LC-PRM et 19 peptides sont quantifiables à des concentrations plus faibles par LC-SRM (Tableau II-9)

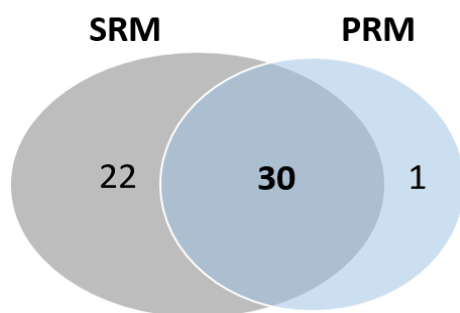


Figure II-24. Nombre de peptides pour lesquels la LLOQ par spectrométrie de masse ciblée a été déterminée. Les LLOQ de 53 peptides sur les 56 ciblés ont été déterminées par LC-SRM et/ou LC-PRM.

Ces résultats montrent la complémentarité des deux méthodes LC-SRM et LC-PRM pour la détection de *Borrelia* dans la peau humaine. La LC-SRM permet de quantifier les 6 protéines ciblées de *Borrelia*, tandis que la LC-PRM ne permet pas de quantifier l'énolase, mais permet de gagner en sensibilité pour certains peptides de 3 protéines (OspC, flagelline, GAPDH). L'absence de LLOQ_{PRM} pour de nombreux peptides peut être due à une concentration insuffisante de NAT pour la gamme PRM (max 200 fmoles). Des concentrations plus élevées auraient pu permettre d'augmenter le nombre de points de la courbe de calibration et de satisfaire le critère $N \geq 3$.

Un exemple de courbe de calibration est montré dans la figure Figure II-25, pour le peptide AAQAALTNSVK de la protéine OspC, avec les traces SRM et PRM obtenues aux LLOQ_{SRM} et LLOQ_{PRM}.

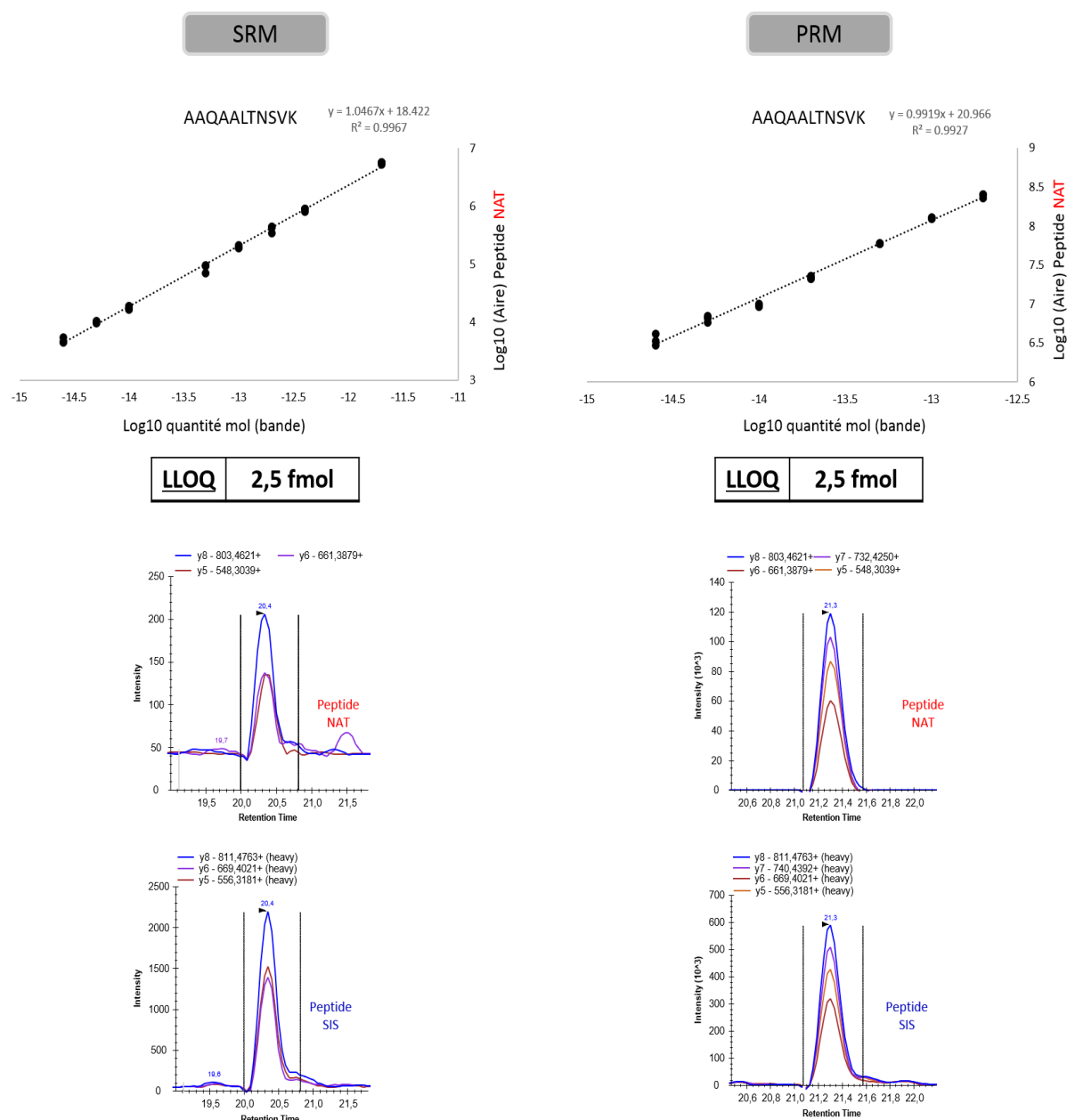


Figure II-25. Détermination de la LLOQ_{SRM} et LLOQ_{PRM} pour le peptide AAQAALTNSVK de la protéine OspC. Les deux courbes de calibration sont représentées ainsi que les traces microLC-SRM et nanoLC-PRM observées à une concentration de 2,5 fmoles de NAT, correspondant aux deux LLOQ_{SRM} et LLOQ_{PRM}. Trois transitions ont été conservées pour la SRM, 4 transitions pour la PRM.

Une autre manière de montrer la différence entre les deux méthodes est la représentation du nombre de peptides quantifiables aux différentes concentrations de la gamme (Figure II-26). On voit que 12 peptides de *Borrelia* (correspondant à 4 protéines) sont quantifiables par microLC-SRM à partir de 1 fmoles dans la bande de gel. Par nanoLC-PRM, seuls deux peptides

(correspondant à 2 protéines) sont quantifiables avec 1 fmole dans la bande. Cela s'explique par l'injection de seulement 6% de l'extrait peptidique sur le système nanoLC : pour 1 fmole de peptide dans la bande, la quantité réellement injectée est seulement de 60 amoles.

L'avantage de la nanoLC-PRM est également de pouvoir extraire plus de transitions que celles ciblées par microLC-SRM, et de pouvoir valider la présence d'un peptide avec plus de certitude.

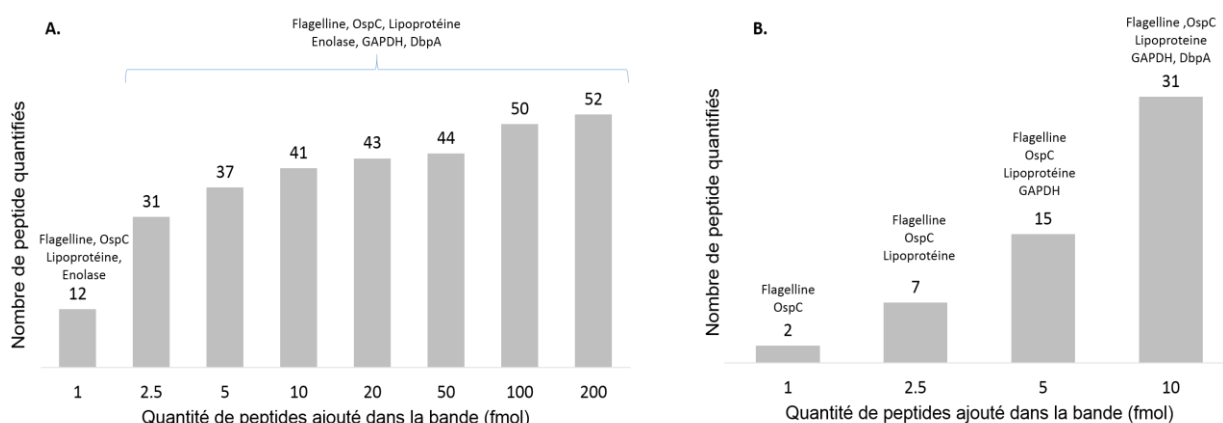


Figure II-26. Histogrammes représentant le nombre de peptides quantifiés par microLC-SRM et nanoLC-PRM. Le nombre de peptides quantifiés pour chaque point de la gamme de concentration des peptides NAT est indiqué pour les analyses microLC-SRM (A) et nanoLC-PRM (B). Les quantités injectées sont de 100% (injection de la totalité de la bande) en microLC-SRM et 6% (injection de 1/16 de la bande) en nanoLC-PRM. Les protéines quantifiées sont décrites sur chaque point de gamme, l'indication lipoprotéine correspond à la lipoprotéine gi|365823350.

La quantité de peptide NAT dans la bande de gel nous a permis d'estimer la quantité de peptide par mg de biopsie que chaque technique détecte. Pour déterminer cette quantité (X , fmol/mg) nous avons utilisé les données suivantes :

- quantité de peptide dans la bande (A , fmol)
- rendement d'extraction (η , μg protéine/mg biopsie)
- quantité de protéine déposée sur le gel SDS-PAGE pour le préfractionnement (Q , μg)

selon la formule suivante :

$$X = \frac{A \times \eta}{Q} \quad \text{Equation 7}$$

La quantité de protéine détectable par mg de biopsie est estimée à partir de la quantité de peptide, mais se traduit par des variations selon le peptide considéré. Ces variations sont dues

à la réponse variable des différents peptides d'une même protéine (P. Picotti *et al.*, 2007), comme cela a été constaté dans d'autres projets au laboratoire (J. Bons, 2019; B. Westermann, 2016).

Le tableau ci-dessous (Tableau II-10) récapitule les résultats obtenus par LC-SRM et LC-PRM avec les gammes de concentration de protéine par mg de biopsie qui ont pu être détectées par les deux méthodes.

Tableau II-10. Comparaison des résultats obtenus par microLC-SRM et nanoLC-PRM. Les quantités les plus faibles injectées ont été calculées à partir des quantités les plus faibles ajoutées dans la bande, et de la fraction injectée (100%, 6%). Les gammes de LLOQ (fmol/bande) proviennent du Tableau II-9. Les gammes de LLOQ (fmol protéine/mg biopsie) ont été calculées avec l'équation 7.

	MicroLC-SRM	NanoLC-PRM
Quantité injectée	100%	6%
Nombre de peptides quantifiés	52/56	31/56
Quantité plus faible injectée (fmol)	1	0.06
LLOQ (fmol/bande)	1-200	1-10
LLOQ (fmol protéine/mg biopsie)	3-600	3-31

Toute cette étude des LLOQ nous a montré que nous étions capables de détecter des peptides de *Borrelia* à des concentrations de quelques fmoles par mg de biopsie dans une peau humaine. L'absence de signal constatée dans plusieurs biopsies de la cohorte de patients était donc probablement due à des concentrations inférieures à ce seuil de détection.

De plus, ces résultats nous ont montré que nous pouvions mettre à profit la complémentarité de la SRM et de la PRM pour les analyses des biopsies de la cohorte. Nous avons donc poursuivi les analyses de la cohorte en effectuant deux dépôts de 50 µg pour le préfractionnement (≈ 10 bandes), une injection pour microLC-SRM (100%) et une injection pour nanoLC-PRM (6%) (Figure II-27).

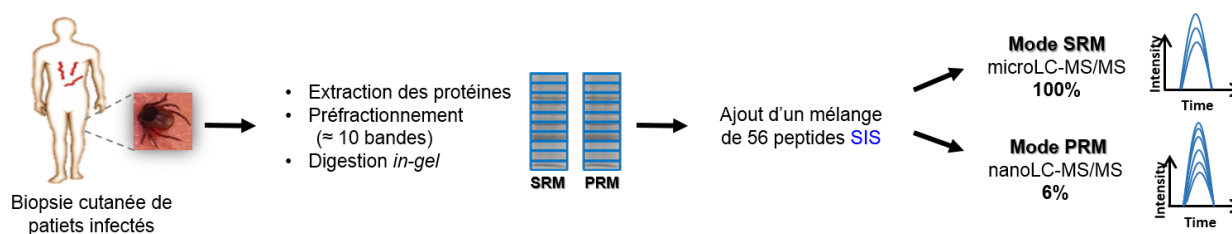


Figure II-27. Analyse des biopsies cutanées de patients en combinant la LC-SRM et la LC-PRM. Suite aux déterminations des LLOQ, le protocole d'analyse des biopsies cutanées a été modifié, deux dépôts de 50 µg ont été réalisés pour être analysés par microLC-SRM (100 %) et par nanoLC-PRM (6%).

1.2.2.3. Problèmes d'échantillonnage

Après avoir analysé plusieurs biopsies de patients, sans observer de signal par spectrométrie de masse malgré une positivité par PCR, nous avons réexaminé avec attention le protocole d'échantillonnage des biopsies pour tenter de comprendre ce manque de sensibilité.

Le premier point qui a retenu notre attention est l'utilisation de **lidocaïne**, un anesthésique injecté localement aux patients avant le prélèvement de la biopsie. En plus de ses propriétés anesthésiques, la lidocaïne possède des propriétés bactériostatiques et bactéricides contre plusieurs micro-organismes, ce qui a comme conséquence l'inhibition de la croissance de la bactérie (S. M. Johnson *et al.*, 2008; B. M. Razavi *et al.*, 2019).

Le deuxième point est la **comparaison des traitements des deux biopsies** pour l'analyse protéomique d'une part, et la PCR d'autre part :

- **Biopsie 1** : dépôt du prélèvement cutané dans un tube contenant le milieu de culture spécifique de *Borrelia* (BSK), puis envoi au CNR *Borrelia*, (Strasbourg) pour les analyses PCR et observation microscopique.
- **Biopsie 2** : dépôt du prélèvement cutané dans un eppendorf et mise à -80°C pour être transféré au LSMBO pour les analyses protéomiques.

Cette comparaison a révélé que, contrairement à ce qui était prévu au démarrage du projet, les analyses PCR et protéomique n'étaient pas strictement comparables puisque la PCR était réalisée après plusieurs semaines de culture (jusqu'à 12 semaines), alors que l'analyse protéomique était faite immédiatement sans culture préalable des bactéries. Cette observation pouvait expliquer le manque de sensibilité de la microLC-SRM constaté dans les premières séries d'analyse protéomique. Il a donc été décidé, en accord avec les médecins, de modifier

le protocole d'échantillonnage de manière à effectuer une comparaison plus stricte entre les trois méthodes (analyse microscopique du milieu de culture, PCR et microLC-SRM) (Figure II-28). Cette modification a été effective pour les biopsies prélevées à partir d'octobre 2017.

A.

2 punch de biopsie: 3 mm
Lidocaïne: 10 mg/ml



½ biopsie: culture en anaérobiose, jusqu'à 12 semaines → PCR

½ biopsie: culture en aérobie, jusqu'à 12 semaines → Stocké à -80°C

Stocké - 80°C → Protéomique ciblée

B.

2 punch de biopsie: 3 mm
Lidocaïne: 10 mg/ml



½ biopsie: culture en anaérobiose, jusqu'à 12 semaines → PCR

½ biopsie: culture en aérobie, jusqu'à 12 semaines → Stocké à -80°C

1 biopsie: culture en anaérobiose, jusqu'à 12 semaines → Protéomique ciblée

Figure II-28. Protocole de prélèvement des biopsies cutanées sur les patients. Pour les deux panels, un anesthésique local est appliqué (lidocaïne) puis deux prélèvements rapprochés sont effectués au niveau de l'EM. Le premier prélèvement est partagé en deux, une ½ biopsie cultivée en anaérobiose puis analysée par PCR en cas d'observation de *Borrelia*. La deuxième ½ biopsie est cultivée en aérobie, pendant la même durée que l'autre ½ biopsie, puis stockée à -80°C. Pour le panel **A**, correspondant au protocole initial, le deuxième prélèvement est stocké à -80°C jusqu'à l'analyse protéomique ciblée. Pour le panel **B**, correspondant au protocole modifié, le deuxième prélèvement est mis en culture pendant 8 à 12 semaines avant l'analyse protéomique ciblée.

1.2.3. Analyse de la cohorte de patients : résultats

Grâce à toutes les optimisations, l'implémentation de la nanoLC-PRM et la modification du protocole d'échantillonnage, nous avons pu analyser l'ensemble des biopsies de la cohorte.

Le **décalage de Tr** observé précédemment pour le peptide ITDSNAFVLAVK d'OspC n'a plus été observé dans les biopsies après culture. Ce problème a donc disparu avec des signaux plus intenses et une amélioration de la résolution chromatographique (pics plus fins). Un exemple d'une biopsie (003-044) analysée par microLC-SRM et nanoLC-PRM est représenté dans la Figure II-29 où nous pouvons constater que le peptide détecté correspond sans ambiguïté à la séquence ciblée dans la méthode et non à une isoforme.

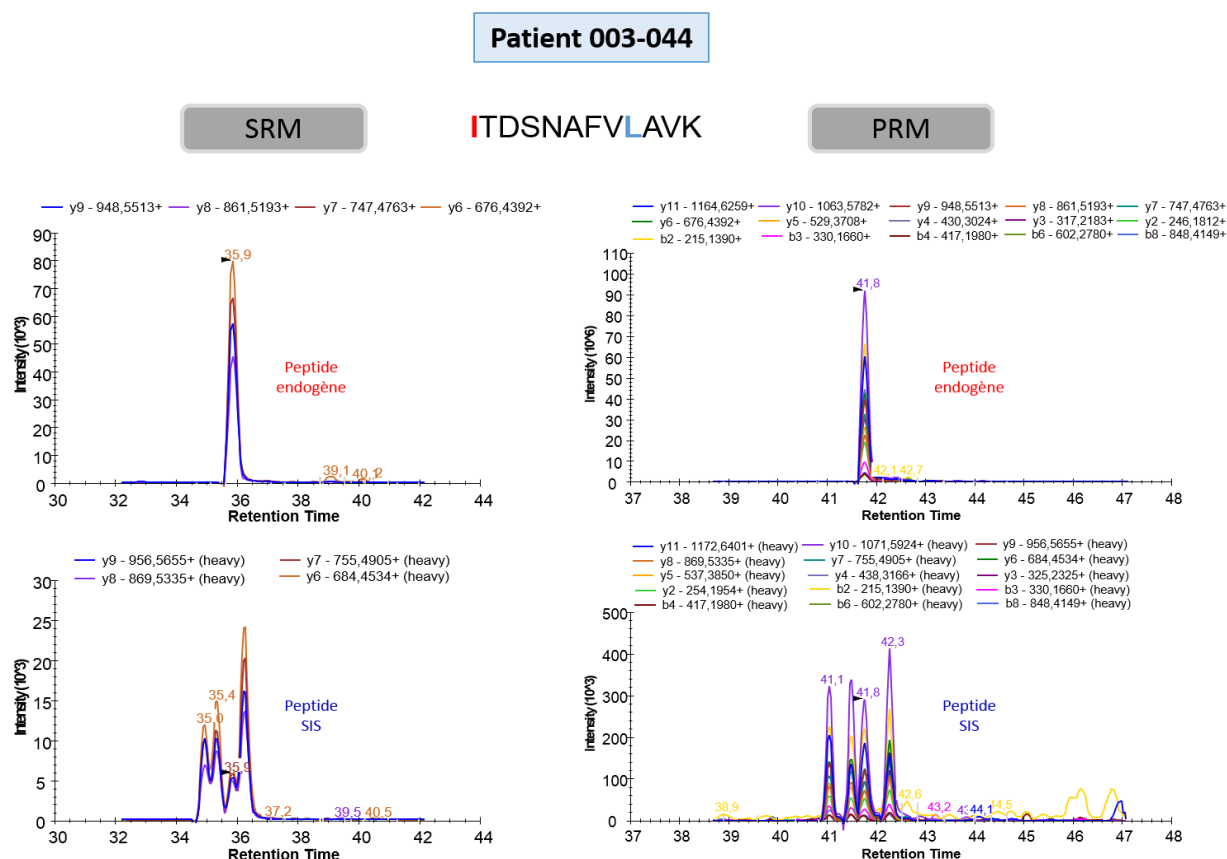


Figure II-29. Traces SRM et PRM du peptide ITDSNAFVLAVK (OspC) dans une biopsie cutanée de patient. La biopsie 003-044 a été analysée en parallèle par les deux méthodes. Dans les deux cas, les 4 isoformes du peptide SIS (I/L)TDSNAFV(L/I)AVK ont été ajoutées. Le peptide endogène est co-élué avec le troisième pic correspondant bien à la séquence ciblée.

La **complémentarité de la SRM et de la PRM** a été confirmée sur plusieurs biopsies, comme illustré sur la Figure II-30. Pour la biopsie 003-027, le peptide TAEELGMQPAK de la flagelline n'a pas pu être validé par nanoLC-PRM en raison d'un signal très faible et d'interférences. Par contre il a été validé par microLC-SRM. Pour la biopsie 003-031, le peptide AAQAALTNSVK d'OspC n'a pas pu être validé par microLC-SRM à cause d'une mauvaise co-élution des transitions, mais il a été validé par nanoLC-PRM. Cet exemple illustre également le gain de la PRM en termes de nombre de transitions pouvant être extraites *a posteriori* (7 transitions pour le peptide AAQAALTNSVK dans la biopsie 003-031).

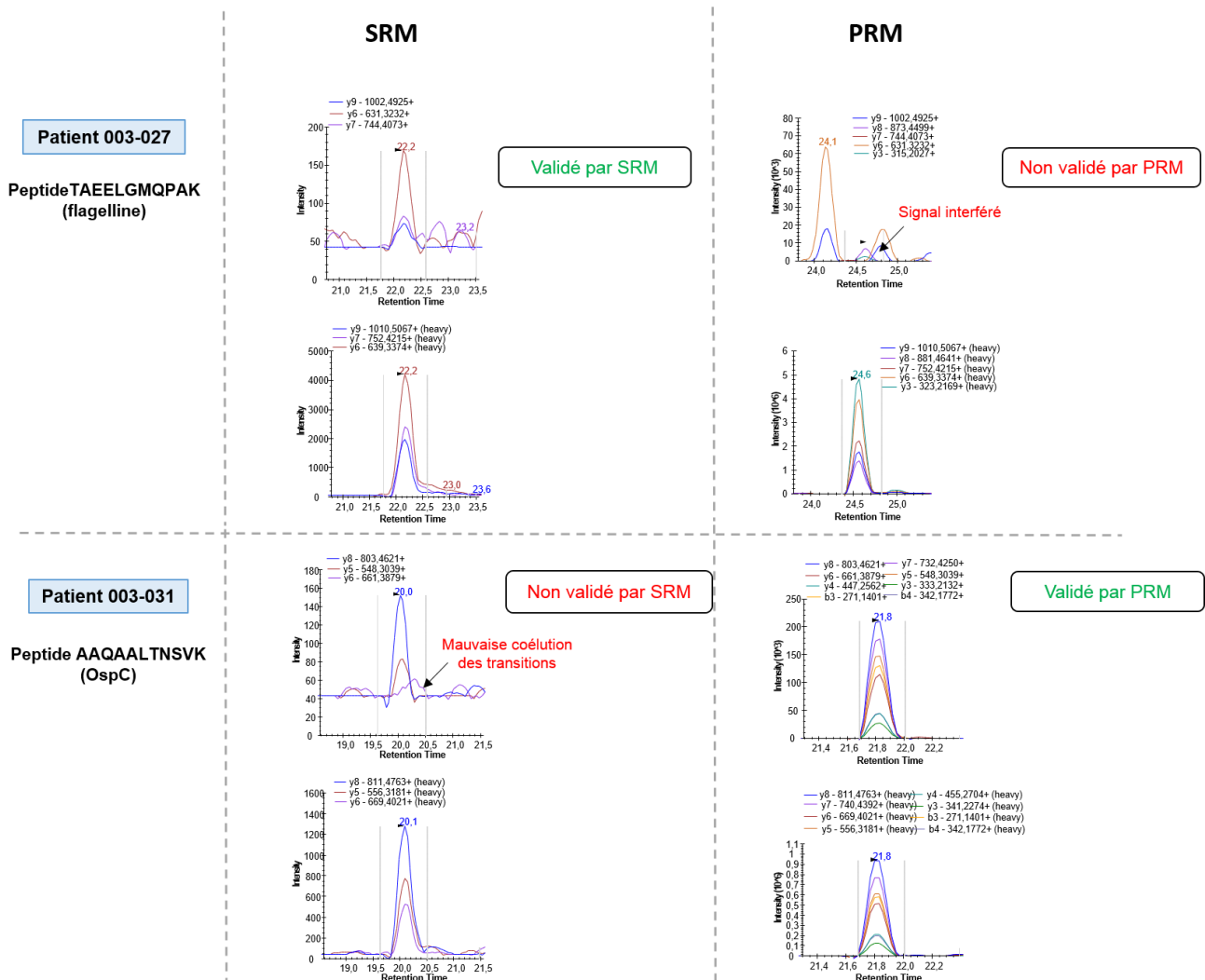


Figure II-30. Exemple de complémentarité de la microLC-SRM et nanoLC-PRM pour l'analyse des biopsies cutanées de patients. Les 2 biopsies 003-027 et 003-031 ont été analysées en parallèle par les deux méthodes. La présence du peptide de flagelline n'a pu être validée que par microLC-SRM dans la biopsie 003-027, tandis que la présence du peptide d'OspC n'a pu être validée que par nanoLC-PRM dans la biopsie 003-031.

Dans la suite de ce paragraphe, nous avons séparé les résultats d'analyse protéomique en deux groupes selon le protocole d'échantillonnage :

1. Groupe 1 : biopsies analysées sans culture préalable (protocole d'échantillonnage initial)
2. Groupe 2 : biopsies analysées après une culture de 8 à 12 semaines (protocole d'échantillonnage modifié)

Pour le **groupe 1**, 19 biopsies ont été analysées correspondant à 17 patients (deux patients présentaient des EM multiples). Une biopsie a donné un résultat négatif quelle que soit la

méthode utilisée (PCR, observation microscopique, protéomique). Les 18 autres biopsies ont donné un résultat positif par au moins une des trois méthodes (Figure II-31). 13/18 biopsies ont donné un signal positif par spectrométrie de masse (SRM et/ou PRM), avec un ou deux peptides par biopsie.

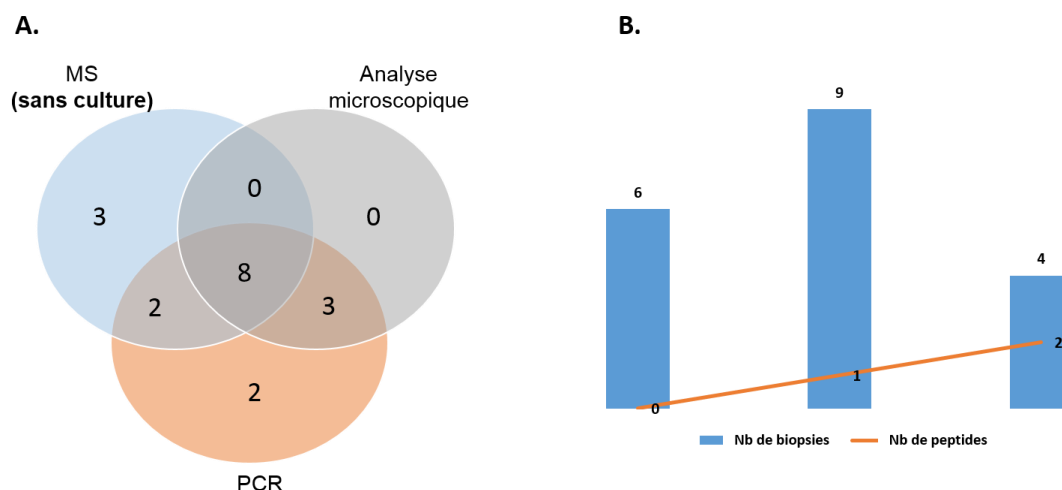


Figure II-31. Résultats de l'analyse des biopsies cutanées des patients du groupe 1 sans culture préalable avant l'analyse par spectrométrie de masse. A. Diagramme de Venn montrant la répartition des 18 biopsies positives par au moins une des trois méthodes. B. Représentation du nombre de peptides détectés dans les 19 biopsies de ce groupe 1.

Pour ce groupe de biopsies, nous observons que l'analyse microscopique de la culture et la PCR effectuée après culture sont globalement plus sensibles que la spectrométrie de masse réalisée sans culture préalable. On peut aussi noter que pour 3 biopsies, seule la spectrométrie de masse a donné un résultat positif, avec toutefois un nombre faible de peptides détectés, pour deux biopsies 1 seul peptide a été identifié et pour la troisième biopsie un seul peptide a été identifié.

Pour le **groupe 2**, constitué de biopsies mises en culture avant l'analyse protéomique, 53 biopsies ont été analysées. Cinq biopsies ont donné des résultats négatifs quelle que soit la méthode utilisée (PCR, observation microscopique, protéomique). Les 48 autres biopsies ont donné un résultat positif par au moins une des trois méthodes (Figure II-32).

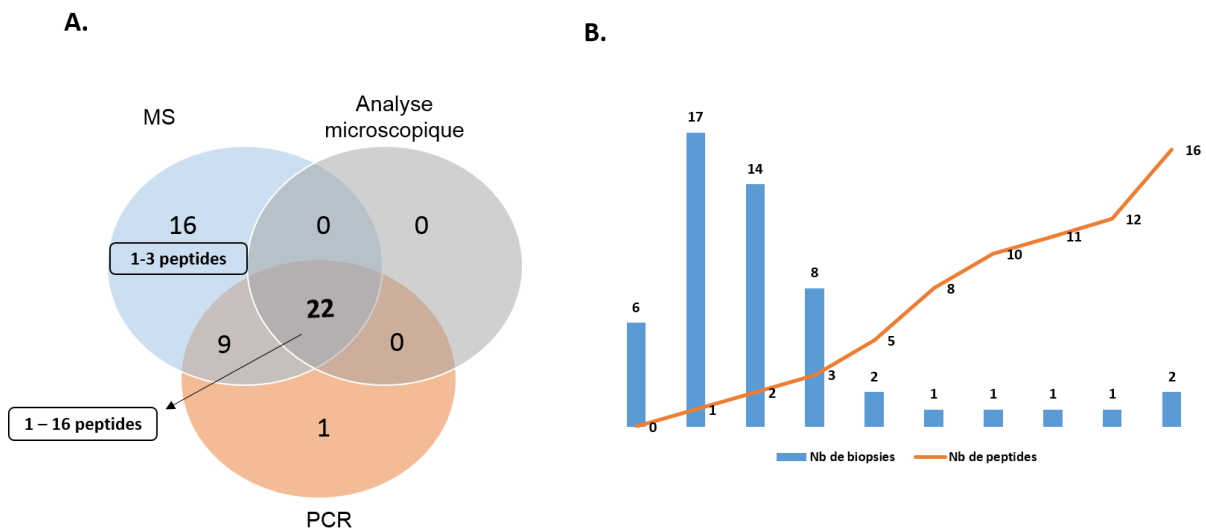


Figure II-32. Résultats de l'analyse des biopsies cutanées des patients du groupe 2 avec culture préalable avant l'analyse par spectrométrie de masse. A. Diagramme de Venn montrant la répartition des 48 biopsies positives par au moins une des trois méthodes. B. Représentation du nombre de peptides détectés dans les 53 biopsies de ce groupe 2.

Ces résultats montrent que la mise en culture a permis d'augmenter significativement la détection par spectrométrie de masse. 47/48 biopsies ont donné un signal positif par spectrométrie de masse (SRM et/ou PRM). Cela a permis de calculer une **sensibilité de 98% pour la spectrométrie de masse**. Pour ce groupe la sensibilité calculée pour la PCR est de 67% (32/48), celle de l'analyse microscopique de la culture est de 46% (22/48).

En termes de nombre de peptides détectés, on constate qu'il a été possible de détecter plusieurs peptides dans certaines biopsies, le maximum étant de 16 peptides détectés dans deux biopsies. Il faut quand même noter que pour 39/48 biopsies (81%), le nombre de peptides détectés a été au maximum de 3.

Dans quelques cas de la cohorte, les analyses protéomiques ont permis le **typage de l'espèce infectante** de la bactérie. Ce typage a été possible car grâce aux variations de séquence d'OspC dans différentes espèces de *Borrelia*, et à la présence de peptides spécifiques d'une espèce dans la liste des peptides ciblés (Figure II-18, paragraphe. 2.2.1). Dans la Figure II-33, on peut voir les traces PRM obtenues pour deux patients, avec la détection des peptides GPNLTVISK et AAQAALTNSVK spécifiques de *B. garinii* dans la biopsie 003-031, et la détection des peptides GPNLTEISK et AAQVALTNSVK spécifiques de *B. afzelii* dans la biopsie 003-028.

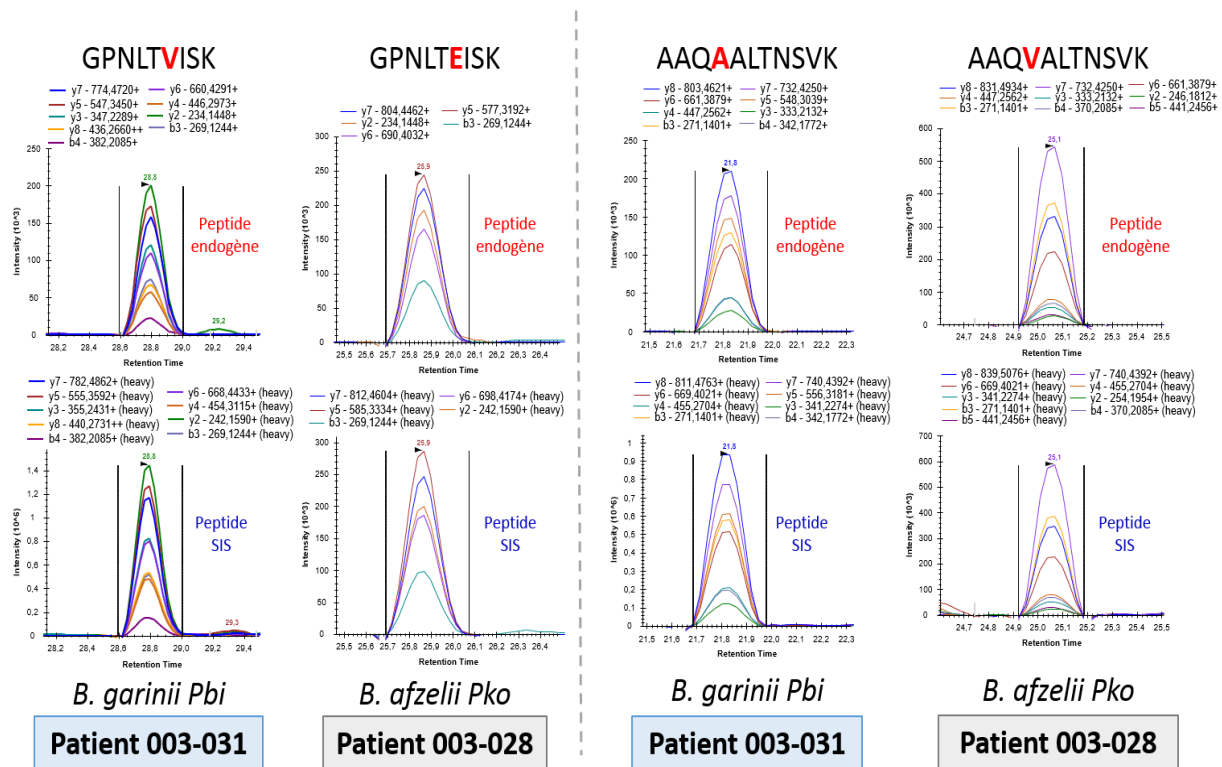


Figure II-33. Typage de l'espèce infectante dans deux biopsies cutanées de patients analysées par LC-PRM. Plusieurs peptides d'OspC contenus dans la liste des peptides ciblés sont des peptides spécifiques d'une espèce de *Borrelia*. Dans la biopsie 003-031, deux peptides spécifiques de *B. garinii Pbi* sont détectés, tandis que dans la biopsie 003-028, deux peptides spécifiques de *B. afzelii Pko* sont détectés.

Concernant les 16 biopsies pour lesquelles seule la spectrométrie de masse a permis de faire un diagnostic positif, il faut également souligner que cinq protéines de *Borrelia* ont pu être détectées (flagelline, OspC, DbpA, GAPDH et l'enolase), alors que la PCR ne cible que le gène *fla* codant la flagelline. Pour 13/16 biopsies, ce sont d'ailleurs des peptides des trois autres protéines (OspC, DbpA et GAPDH) qui ont été détectés, et non pas des peptides de flagelline. **Le multiplexage de la spectrométrie de masse ciblée, c'est-à-dire le ciblage de plusieurs protéines dans une seule méthode, apparaît ainsi comme un réel avantage.**

En considérant les deux groupes 1 et 2, le total des biopsies dans lesquelles nous avons détecté des peptides de *Borrelia* par spectrométrie de masse s'élève à 60 parmi les 72 biopsies analysées. Ces 60 biopsies sont réparties dans la Figure II-34 selon la méthode qui a permis la détection. Une biopsie s'est révélée positive uniquement par l'analyse microLC-SRM, c'est la biopsie 003-010 avec un seul peptide de la protéine DbpA (TTANGIIEIVK) (voir tableau d'annexe 2). Quarante-quatre biopsies se sont révélées positives uniquement par

l'analyse nanoLC-PRM, et 15 par les deux approches. Ce diagramme montre que l'implémentation de la PRM a été très bénéfique pour l'approche diagnostique par spectrométrie de masse ciblée.

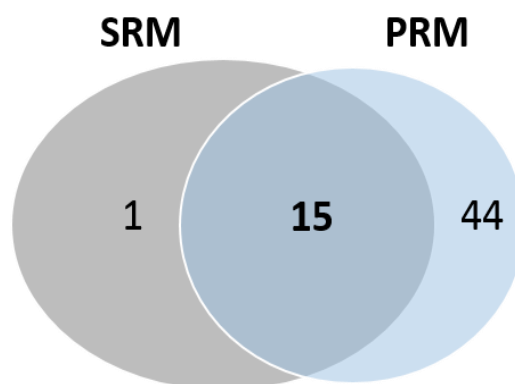


Figure II-34. Nombre de biopsies dans lesquelles des peptides de *Borrelia* ont été détectés par microLC-SRM et/ou nanoLC-PRM. Sur l'ensemble des deux groupes (avec et sans culture préalable), des peptides ont été détectés dans 60 biopsies sur les 72 biopsies analysées (19 pour groupe 1 + 53 pour groupe 2).

En termes de quantification, les signaux observés ont été majoritairement en dessous des LLOQ. Nous avons pu quantifier des peptides de *Borrelia* dans 20 biopsies, pour les autres nous ne pouvons parler que de détection mais pas de quantification.

Pour déterminer la quantité de protéines présente dans une biopsie (X , fmol de protéine/mg de biopsie) nous avons utilisé les données suivantes :

- concentration du peptide SIS (C , fmol/ μ L) et dilution (D) utilisée dans le mélange de peptides
- volume du mélange de peptides SIS (V , μ L)
- ratio obtenu entre le peptide endogène et le peptide SIS (ratio (L/H))
- rendement protéique (η , μ g protéine/mg biopsie)
- quantité de protéine déposée sur le gel SDS-PAGE pour le préfractionnement (Q , μ g)

selon la formule suivante

$$X = \frac{C}{D} \times V \times \text{ratio } (L/H) \times \frac{\eta}{Q}$$

Equation 8

Cette formule peut être appliquée pour la quantification par microLC-SRM, et donne directement la quantité de peptide en fmoles par mg de biopsie. Par contre pour la nanoLC-PRM, comme nous l'avons vu précédemment, seule une fraction de l'extrait (6%) est injectée sur le système nanoLC : il faut donc prendre en compte ce rapport et multiplier la quantité calculée partir de l'équation 8 par un facteur 100/6 soit 16,7.

Un exemple de biopsie (003-043) pour laquelle une quantification a pu être réalisée par les deux méthodes est montré dans la Figure II-35 : il s'agit de la quantification du peptide GPNLTEISK d'OspC. Dans ce cas le peptide SIS est de qualité AQUA™, ce qui permet une quantification plus juste. Pour la SRM, à partir du ratio L/H = 1,25, nous avons calculé une quantité de ≈ 8 fmol de peptide/mg biopsie, soit par extension ≈ 8 fmol OspC/mg biopsie. Pour la PRM, nous avons deux traces car le peptide a été détecté dans deux bandes contiguës du gel. Dans ce cas la quantité totale a été obtenue en sommant les quantités dans chaque bande de gel, soit ≈ 15 fmol OspC/mg biopsie.

Patient 003-043

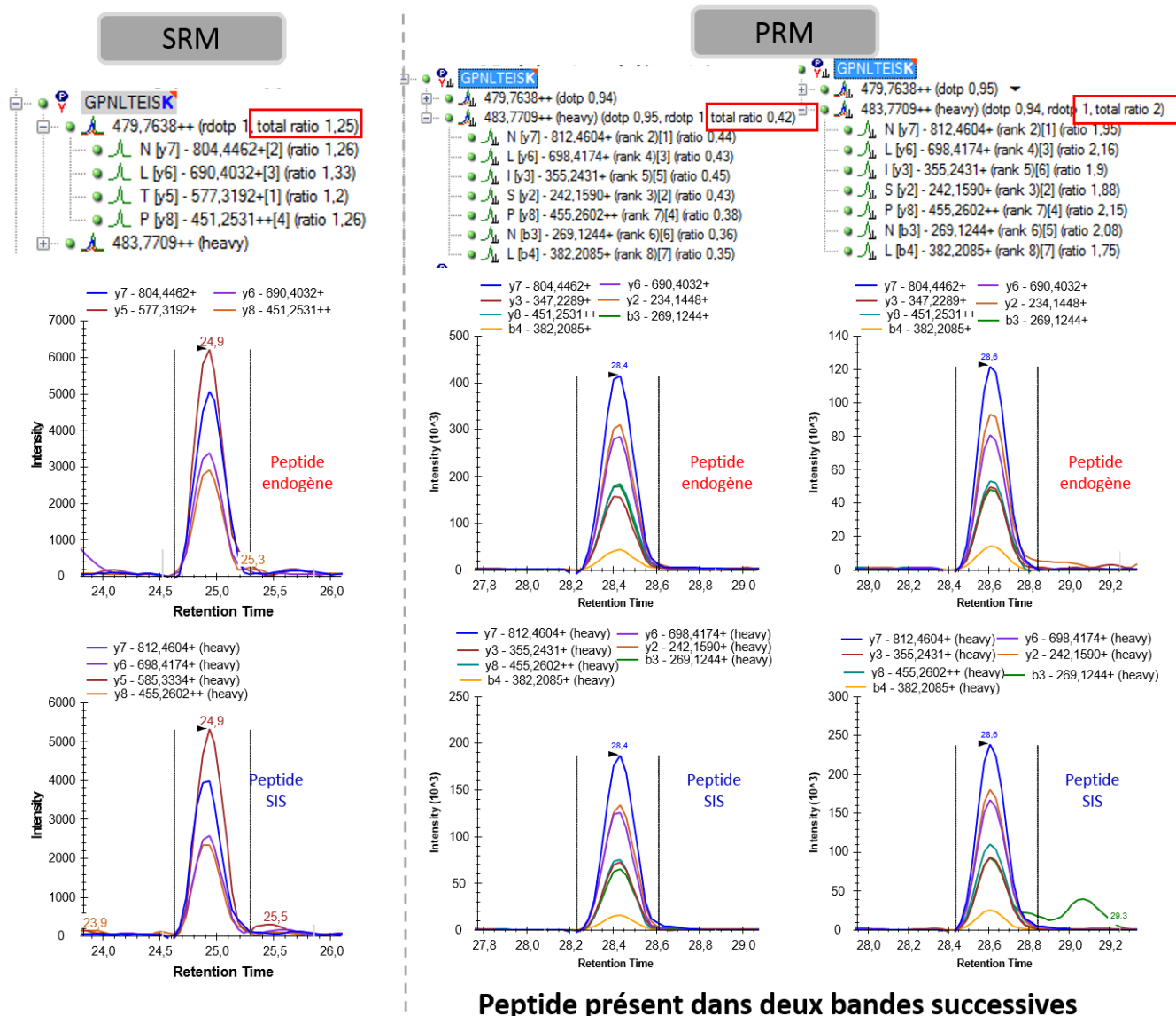


Figure II-35. Traces SRM et PRM du peptide GPNLTEISK (OspC) dans une biopsie cutanée de patient. La biopsie 003-043 a été analysée en parallèle par les deux méthodes. La quantification a été faite à partir du ratio entre le peptide endogène et le peptide SIS. Dans le cas de la microLC-SRM, le ratio endogène/SIS est encadré ($L/H=1,25$), dans le cas de la nanoLC-PRM, c'est le ratio SIS/endogène qui est indiqué ($H/L=0,42$ et $H/L=2$, pour les deux bandes de gel contiguës dans lesquelles le peptide a été détecté).

Parmi les 15 biopsies dans lesquelles des peptides ont été détectés par microLC-SRM et nanoLC-PRM (Figure II-34), la quantification par au moins l'une des deux méthodes a pu être faite pour 10 biopsies. Les résultats pour ces 10 biopsies sont donnés dans le Tableau II-11.

Tableau II-11. Résultats obtenus pour les 10 biopsies dans lesquelles des peptides de *Borrelia* ont été détectés par microLC-SRM et nanoLC-PRM, et quantifiés par au moins une des deux méthodes. Séquences en rouge : peptides pour lesquels les peptides SIS sont de qualité AQUA™. Séquences en noir : peptides pour lesquels les peptides SIS sont de qualité *crude*. Pour chaque biopsie, une colonne montre les résultats de PRM et une colonne montre les résultats de SRM. **N/A_LLOQ** : peptides détectés mais pour lesquels la LLOQ est inconnue. **<LLOQ** : peptides détectés mais pour lesquels les signaux sont inférieurs aux LLOQ. En **vert** : quantification en fmol protéine/mg biopsie. **N.d** : peptides non détectés. La dernière ligne indique le nombre total de peptides détectés (peptides quantifiés en **vert**, <LLOQ et N/A_LLOQ).

Protéines	Biopsie	003-044		003-027		003-043		002-003		003-028		001-020		001-023		003-031		002-002		003-032	
	Méthode d'analyse Peptide	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM	PRM	SRM
Flagelline	AINFIQTTEGNLNEVEK	N/A_LLOQ	827	72	21	N/A_LLOQ	<LLOQ	N/A_LLOQ	110	N/A_LLOQ	N.d	N/A_LLOQ	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d
	ASDDAAGMGVSGK	N/A_LLOQ	27	N.d	43	N.d	N.d	N/A_LLOQ	59	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	ELAVQSGNGTYSDDSR	N/A_LLOQ	3812	<LLOQ	N.d	N/A_LLOQ	<LLOQ	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	GSIQIEIQLTDEINR	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	IADQAQYNQMHLMSNK	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	INTPASLSGSQASWTLR	N/A_LLOQ	1167	<LLOQ	<LLOQ	N/A_LLOQ	N.d	N/A_LLOQ	115	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	MIINHNTSAINASR	N/A_LLOQ	718	<LLOQ	140	N/A_LLOQ	<LLOQ	N/A_LLOQ	316	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	NNAINAANLSK	11758	8455	131	118	171	77	N.d	N.d	68	58	10	N.d	2	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	NNGINAANLSK	N.d	N.d	N.d	N.d	N.d	N.d	276	188	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	NSTHEYAIENLK	8300	3223	50	54	18	233	3600	N.d	19	N.d	<LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
OspC	TAEELGMQPAK	N/A_LLOQ	<LLOQ	N.d	15	N.d	N.d	N/A_LLOQ	31	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	AAQAALNSVK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	126	N.d	N.d	N.d	N.d	N.d
	AAQVALNSVK	N.d	N.d	N/A_LLOQ	195	N.d	N.d	N.d	N.d	N/A_LLOQ	162	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	AILITDAAK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	14	N.d	N.d	N.d
	DLFESVEGLLK	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	ELTNPVVAETPK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	36	N.d	N.d	N.d	N.d	N.d
	ELTSPVVAESPK	117583	29343	158	269	N.d	N.d	N.d	N.d	270	199	37	235	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	EVEALLSSIDELSK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	138	9	N.d	N.d	N.d	N.d
	EVETLVLSIDELAK	N.d	N.d	1110	5408	N.d	N.d	N.d	N.d	N.d	6610	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	GGDSASTNPADSAK	192679	3480	51	N.d	60	29	N.d	N.d	40	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	GNPLTEISK	4323	916	<LLOQ	10	15	8	N.d	N.d	<LLOQ	9	<LLOQ	8	<LLOQ	3	N.d	N.d	N.d	N.d	N.d	N.d
	GNPLTVISK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	64	28	N.d	N.d	N.d	N.d
	ITDSNAFLAVK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	199	N.d	N.d	N.d	N.d	N.d
	ITDSNAFVLAVK	N/A_LLOQ	2055	N/A_LLOQ	26	N/A_LLOQ	15	N/A_LLOQ	N.d	N/A_LLOQ	23	N/A_LLOQ	23	N/A_LLOQ	7	N.d	N.d	N.d	<LLOQ	N.d	N.d
	LSVLNSEELK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	208	83	71	28	N.d	N.d	N.d	N.d
	NESLIAGAYEISK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	173	N.d	N.d	N.d	N.d	N.d
DbpA	TTANGIIEIVK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	TVTDAAEQHPTTAEGILEIAK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	22
	DITDEIDAIK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	<LLOQ	N.d
GAPDH	GGYLDHVNHAGAK	<LLOQ	<LLOQ	<LLOQ	N.d	<LLOQ	N.d	<LLOQ	<LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	GIDIVAINDLTDPK	N.d	N.d	N.d	N.d	48	12	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
Enolase	IQLVGDDLFVTNTSFLK	N/A_LLOQ	487	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
Total des peptides détectés		16	14	13	13	12	9	9	8	9	7	6	3	5	3	7	4	1	1	1	1

Les biopsies 003-044 et 003-027 sont les deux biopsies dans lesquelles 16 peptides ont été détectés.

Les quantifications les plus élevées ont été obtenues dans la biopsie 003-044. D'après les informations recueillies par le médecin, le patient avait indiqué une apparition de l'EM 10 jours avant le prélèvement. Ces quantifications élevées pourraient suggérer un pic de multiplication de *Borrelia*, à comparer au pic de multiplication observé chez la souris 7 jours après l'infection (A. Kern *et al.*, 2011).

On peut également constater que les quantifications d'une protéine sont très variables si on considère tous les peptides quantifiés (par exemple de 27 à 8455 fmol flagelline /mg biopsie pour la biopsie 003-044 analysée par microLC-SRM). Par contre les valeurs sont plus homogènes si on ne considère que les quantifications effectuées à partir de peptides AQUA™ (par exemple 827 et 1167 fmol flagelline /mg biopsie obtenues à partir des deux peptides AQUA™ de la flagelline). Cette observation peut s'expliquer par la moindre justesse des quantifications faites à partir des peptides *crude*. Une quantification différente d'une protéine à partir de différents peptides a déjà été observée (P. Picotti *et al.*, 2007).

Enfin, on peut relever que les quantifications obtenues par SRM et PRM sont comparables pour quelques biopsies (par exemple 50 et 54 fmol flagelline /mg biopsie pour le peptide NSTEYAIENLK dans la biopsie 003-027, par PRM et SRM respectivement). Mais dans la plupart des cas, il y a plus d'un facteur 2 entre les quantités déterminées par nanoLC-PRM et microLC-SRM. Ces différences pourraient s'expliquer par une plus grande dilution des peptides SIS pour la nanoLC-PRM conduisant à des rapports de signaux [endogène/SIS] supérieurs, comme illustré sur la Figure II-36 pour le peptide GGDSASTNPADESAK. Dans ce cas la valeur très importante de 192,7 pmol/mg biopsie déterminée par nanoLC-PRM semble surestimée par rapport à la quantité déterminée par microLC-SRM.

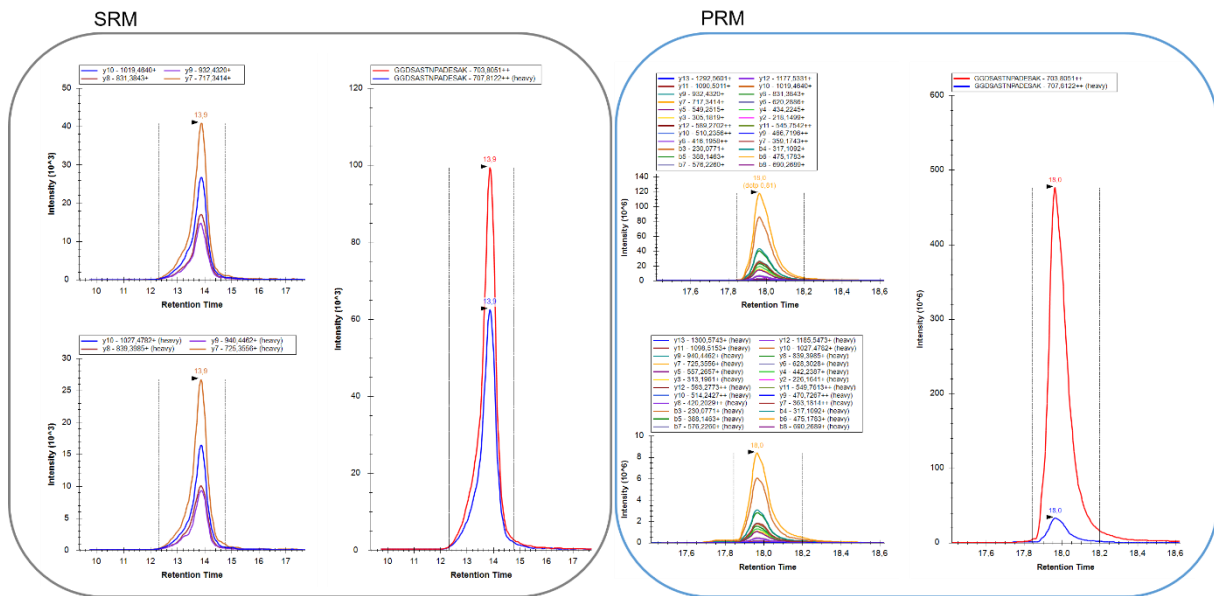


Figure II-36. Comparaison de la quantification d'un peptide d'OspC par microLC-SRM et nanoLC-PRM dans une biopsie cutanée de patient. Le peptide GGDSASTNPDESAK (OspC) a été quantifié dans la biopsie 003-044 par SRM et PRM avec des valeurs de 3,5 fmol/mg biopsie et 192,7 pmol/ mg biopsie, respectivement.

Parmi les 44 biopsies dans lesquelles des peptides ont été détectés uniquement par nanoLC-PRM (Figure II-34), la quantification a pu être faite dans 10 biopsies. Les résultats pour ces 10 biopsies sont donnés dans le Tableau II-12. Dans ces biopsies, le nombre de peptides détectés, et *a fortiori* quantifiés, est inférieur à ce qui a été observé dans les biopsies où des peptides ont été détectés par les deux méthodes.

Tableau II-12. Résultats obtenus pour les 10 biopsies dans lesquelles des peptides de *Borrelia* ont été détectés et quantifiés uniquement par nanoLC-PRM. Séquences en rouge : peptides pour lesquels les peptides SIS sont de qualité AQUA™. Séquences en noir : peptides pour lesquels les peptides SIS sont de qualité crude. **N/A_LLOQ** : peptides détectés mais pour lesquels la LLOQ est inconnue. **<LLOQ** : peptides détectés mais pour lesquels les signaux sont inférieurs aux LLOQ. En **vert** : quantification en fmol protéine/mg biopsie. **N.d** : peptides non détectés. La dernière ligne indique le nombre total de peptides détectés (peptides quantifiés en **vert**, **<LLOQ** et **N/A_LLOQ**).

Protéines	Biopsie	001-019	001-022	002-001	003-001	003-002	003-008	003-014	003-022	003-038	005-003
	Méthode d'analyse Peptide	PRM	PRM	PRM	PRM	PRM	PRM	PRM	PRM	PRM	PRM
OspC	AAQAALNSVK	N.d	N.d	N.d	N.d	17	N.d	N.d	33	N.d	N.d
	AILITDAK	24	N.d	N.d	N.d	N.d	10	N.d	N.d	N.d	N.d
	ELTNPVVAETPK	N.d	29	144	176	N.d	N.d	30	N.d	142	213
	EVETLLASIDELATK	N.d	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d
	GNPLTEISK	<LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N.d
	IGNNGLEANQSK	N.d	N.d	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d
	LFESVEVLSK	N.d	N.d	N.d	<LLOQ	N.d	N.d	N.d	N.d	N.d	N.d
Flagelline	AINFIQTTEGNLNEVEK	N.d	N.d	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d
	GSIQIEIEQLTDEINR	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N.d	N.d	N.d
DbpA	DITDEIDAIK	N.d	N.d	N.d	<LLOQ	<LLOQ	N.d	N.d	N.d	N.d	N.d
	TTANGIEIVK	N.d	N.d	N/A_LLOQ	N.d	N.d	N.d	N.d	N/A_LLOQ	N.d	N.d
Total des peptides détectés		2	1	3	3	2	1	3	3	1	1

En considérant l'ensemble des résultats, l'analyse de la cohorte humaine a été satisfaisante et a prouvé l'intérêt de l'approche protéomique ciblée pour détecter, voire quantifier, des protéines de *Borrelia* dans les biopsies cutanées de ces patients. Les points essentiels à retenir sont (i) l'excellente sensibilité de la protéomique ciblée (98%) pour les biopsies mises en culture avant l'analyse protéomique (groupe 2), (ii) la complémentarité des approches SRM et PRM, et (iii) l'intérêt du multiplexage de la protéomique ciblée. Toutefois il faut garder à l'esprit la faiblesse, voire l'absence des signaux, dans les biopsies sans culture préalable. La culture des biopsies pendant plusieurs semaines est donc encore nécessaire pour un diagnostic direct par protéomique.

Compte tenu du faible nombre de peptides détectés dans la majorité des biopsies (3 peptides max dans 81% des biopsies), et en concertation avec les cliniciens impliqués dans le projet, il est apparu souhaitable d'analyser une dizaine de biopsies de patients contrôles négatifs (sans EM) pour confirmer la spécificité de la détection. Le recrutement de ces patients devait se faire au cours de l'année 2020, mais la crise sanitaire liée au Covid a retardé la collecte de ces biopsies. Ces analyses devraient être effectuées au cours du dernier trimestre 2020.

1.3. Stratégies alternatives

1.3.1. Préparation tube-gel

Comme nous venons de le voir, l'analyse de la cohorte humaine a été réalisée avec un préfractionnement des protéines sur gel SDS-PAGE en une dizaine de bandes, puis une analyse par spectrométrie de masse de chaque bande. Ce fractionnement permet de réduire la complexité du mélange peptidique mais il prend du temps et multiplie les analyses LC-MS/MS. De plus il est source de variation entre réplicats, et source de biais si on cherche à quantifier les protéines d'intérêt (risque de répartition dans plusieurs bandes au moment de l'excision, comme l'exemple montré dans la Figure II-35). Dans le but de nous affranchir de cette étape de préfractionnement, nous avons évalué la préparation tube-gel (TG) (X. Lu *et al.*, 2005) pour des biopsies cutanées. Cette étude a été réalisée durant le stage de Vildan Gulsari que j'ai co-encadrée (stage Master M2 Sciences analytiques février-juin 2019). Comme expliqué dans le chapitre I, avec une préparation TG, l'échantillon est piégé et concentré dans un gel de polyacrylamide, puis la digestion est effectuée *in gel* sans avoir effectué de migration électrophorétique. Cette préparation TG a montré ses avantages en termes d'efficacité et de robustesse (L. Muller *et al.*, 2019; L. Muller *et al.*, 2016). Nous avons

déterminé les LLOQ_{PRM} des 56 peptides cibles de *Borrelia* avec une préparation TG, de manière analogue à ce qui a été fait après un préfractionnement (Figure II-37).

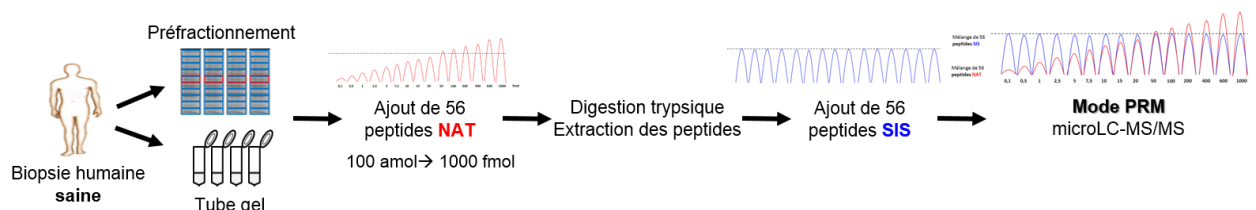


Figure II-37. Comparaison des LLOQ_{PRM} des 56 peptides de *Borrelia* avec ou sans préfractionnement. Une biopsie saine humaine a été utilisée comme matrice. L'extrait protéique a été préfractionné par SDS-PAGE (≈10 bandes de 2 mm), ou piégé dans un tube-gel. Les peptides **NAT** ont été ajoutés en quantité variable (100 amol-1000 fmol) et les peptides **SIS** en quantité constante. Les analyses nanoLC-PRM ont été effectuées en 4 réplicats d'injection. Une quantité équivalente de peptides a été injectée avec et sans préfractionnement.

Les LLOQ_{PRM} de 44 peptides correspondant à 5 protéines (OspC, flagelline, DbpA, GAPDH et lipoprotéine gi|365823350) ont pu être déterminées avec des valeurs comprises entre 10 et 100 fmol de peptide NAT présent dans le TG. La comparaison avec les LLOQ_{PRM} obtenues avec préfractionnement (Tableau II-9) a montré une baisse de sensibilité (d'un facteur 2 à 10 selon les peptides) avec la préparation TG, mais ces résultats étaient encourageants puisqu'ils ont montré qu'il était possible de détecter et de quantifier des protéines de *Borrelia* dans une biopsie cutanée, sans étape de préfractionnement. Nous avons alors appliqué cette stratégie à trois biopsies de la cohorte humaine (002-003, 003-040, 003-015) qui avaient déjà été analysées par préfractionnement et nanoLC-PRM : deux peptides de *Borrelia*, GVEELEK et TTANGIIEIVK, ont été détectés dans ces biopsies (Tableau II-13). On constate que le nombre de peptides détectés est inférieur en absence de préfractionnement, mais ce résultat nous incite à analyser d'autres biopsies et à poursuivre l'amélioration de la préparation d'échantillons.

Tableau II-13. Protéines et peptides de *Borrelia* détectés dans 3 biopsies de patients avec et sans préfractionnement. Les trois biopsies ont été traitées par préfractionnement SDS-PAGE puis nanoLC-PRM (comme décrit dans la Figure II-30) ou par une préparation tube-gel (TG) puis nanoLC-PRM. Ce tableau est extrait du rapport Master de V. Gulsari (2019)

Protéine	Peptide	002-003		003-040		003-015	
		Préfac	TG	Préfac	TG	Préfac	TG
OspC	GVEELEK		X		X		X
	ITDSNAFLAVK					X	
	ITDSNAFVLAVK	X					
Flagelline	AINFIQTTEGNLNEVEK	X					
	ASDDAAGMGVSGK	X					
	INTPASLSGSQASWTLR	X					
	MIINHNTSAINASR	X					
	NNGINAANLSK	X					
	NSTEYAIENLK	X					
	TAEELGMQPAK	X		X			
DbpA	TTANGIIEIVK				X	X	
GAPDH	GGYLDHVNHAGAK	X				X	
Total		9	1	1	2	3	1

1.3.2. Mode DIA

Dans le but d'améliorer la détection des protéines de *Borrelia* présentes en très faible quantité dans la peau, nous avons voulu évaluer un autre mode d'acquisition des données de MS/MS : le mode DIA (*Data Independent Acquisition*). Comme explicité dans le chapitre I, ce mode d'acquisition présente des avantages pour l'identification et dans certains cas la quantification des protéines. Le mode DIA permet la fragmentation de tous les précurseurs contenus dans une certaine gamme de masse. Il a l'avantage d'utiliser les fragments obtenus pour effectuer différents types d'interprétation de données sans avoir à effectuer de nouvelles acquisitions. Une détection ciblée sur des peptides d'intérêt peut être effectuée *a posteriori* lors du traitement de données.

En collaboration avec Nicolas Pythoud, dont le sujet de thèse est centré sur le développement de ce mode DIA au laboratoire, deux biopsies de la cohorte humaine (003-027 et 003-034) ont été analysées en mode DIA. Ces analyses ont été effectuées sur un couplage nanoAcquity/Q-Exactive HF-X, doté d'une résolution allant jusqu'à 240 000, une précision de masse de 5 ppm et une vitesse d'acquisition de 40 Hz (Tableau I-2, Chapitre I).

Les biopsies humaines ont été traitées comme pour les analyses SRM ou PRM avec un préfractionnement SDS-PAGE (10 bandes). Le mélange des 56 peptides SIS a été ajouté avant les analyses par spectrométrie de masse, comme montré dans la Figure II-17.

Les données brutes ont été traitées pour l'identification en mode DDA et DIA. Pour le mode DDA, une banque de données contenant les protéines humaines et les protéines de *Borrelia* a été utilisée pour l'assignation des spectres MS/MS. Pour le mode DIA, la librairie spectrale a été générée à l'aide du mélange de 56 peptides SIS, et les analyses ont été effectuées avec le logiciel Spectronaut™ Pulsar. Les analyses en mode DIA ont également été traitées avec directDIA™ implémenté dans le logiciel Spectronaut™ Pulsar, ce qui permet l'interprétation des données en utilisant uniquement une banque de données et pas de librairie spectrale (Tableau II-14).

Tableau II-14. Nombre de protéines et de peptides identifiés dans deux biopsies cutanées analysées en mode DDA, DIA et direct DIA. Deux biopsies cutanées ont été extraites et préfractionnées sur gel SDS-PAGE (≈10 bandes) puis analysées sur un couplage nanoAcquity/ Q-Exactive HF-X.

	DDA		DIA		directDIA	
	003-027	003-034	003-027	003-034	003-027	003-034
Nombre total de protéines identifiées	2387	3843	2837	4469	2371	4205
Nombre de protéines de <i>Borrelia</i> identifiées	6	4	7	8	9	6
Nombre total de peptides identifiés	11952	23159	18896	32263	12154	26926
Nombre de peptides de <i>Borrelia</i> identifiés	11	4	14	8	15	6

Ces résultats montrent l'avantage du mode DIA pour l'identification de peptides dans un mélange complexe : un nombre plus important de protéines et de peptides a été identifié en mode DIA. Un maximum de 9 protéines (15 peptides) de *Borrelia* a été détecté dans la biopsie 003-027. Dans cette biopsie, 5 peptides sont communs aux identifications par DDA, DIA et directDIA. Dans la biopsie 003-034, aucun peptide n'est commun aux 3 modes DDA, DIA et directDIA.

Les mêmes données ont été utilisées pour effectuer *a posteriori* une détection ciblée de 56 peptides de *Borrelia*, en se servant des 56 peptides SIS ajoutés à l'échantillon pour valider la détection des peptides d'intérêt. En comparant les peptides détectés par SRM, PRM et DIA (Tableau II-15), on observe un assez bon recouvrement pour la biopsie 003-027, avec néanmoins un peu moins de peptides détectés par DIA (10 vs 14). Pour la biopsie 003-034, la DIA n'a pas permis de détecter de peptide d'intérêt. Ces résultats sont en accord avec la littérature qui montre que la SRM et la PRM sont plus sensibles que la DIA (L. C. Gillet *et al.*, 2012; C. Ludwig *et al.*, 2018). La détection d'un nombre supérieur de peptides en PRM par rapport à la DIA a également été observée dans une autre étude au laboratoire (J. Bons, 2019).

Bien que préliminaire, et réalisée uniquement sur deux biopsies, cette évaluation du mode DIA est très encourageante et va être poursuivie pour la détection des protéines de *Borrelia* dans des biopsies cutanées, voire dans des fluides biologiques.

Tableau II-15. Comparaison des peptides de *Borrelia* détectés dans deux biopsies cutanées analysées par SRM, PRM et DIA. Les résultats pour la SRM et PRM sont issus de l'analyse de la cohorte humaine présentés dans le paragraphe 1.2.3 Les résultats pour la DIA sont issus de l'analyse de ces deux biopsies par DIA sur le couplage nanoAcquity/ Q-Exactive HF-X.

Protéine	Peptide	003-027			003-034		
		SRM	PRM	DIA	SRM	PRM	DIA
OspC	AAQVALTNSVK	X	X	X			
	DLFESVEGLLK	X					
	ELTSPVVAESPK	X	X	X			
	EVETLVLSIDELAK	X	X				
	GGDSASTNPADESAK		X	X			
	GNLITEISK	X	X	X			
	ITDSNAFVLAVK	X	X	X			
Flagelline	AINFIQTTEGNLNEVEK	X	X				
	ASDDAAGMGVSGK	X		X			
	ELAVQSGNGTYSDSDR		X				
	INTPASLSGSQASWTLR	X	X				
	MIINHNTSAINASR	X	X	X			
	NNAINAANLSK	X	X	X			
	NSTEYAIENLK	X	X	X			
	TAEELGMQPAK	X		X			
DbpA	DITDEIDAIK					X	
GAPDH	GGYLDHVNHAGAK		X				
TOTAL		14	14	10	0	1	0

2. Diagnostic de la phase tardive disséminée de la maladie de Lyme

Dans cette partie sera présentée ma contribution au développement d'une méthode de diagnostic de la phase tardive disséminée de la maladie de Lyme. Pour rappel, les outils diagnostiques sérologiques actuellement disponibles ont une mauvaise valeur prédictive positive pour les infections disséminées et l'analyse protéomique ciblée pourrait être une alternative intéressante.

Avant mon arrivée au laboratoire, nos collaborateurs avaient montré dans un modèle murin la persistance cutanée de *Borrelia* 40 jours après l'infection et la possibilité de réactiver la prolifération bactérienne par l'application locale d'un dermocorticoïde, le clobetasol (Figure II-38).

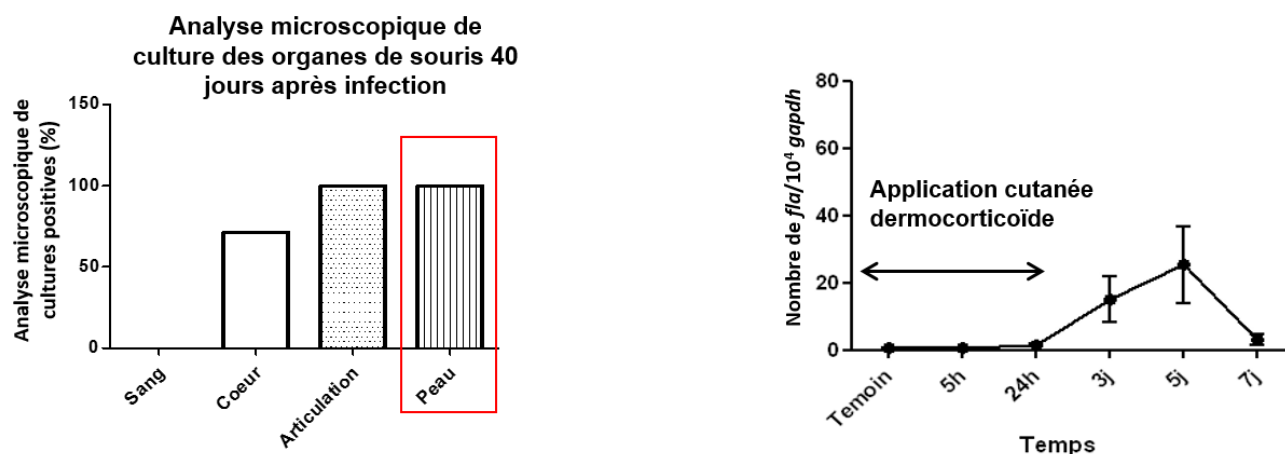


Figure II-38. Mise en évidence de la latence cutanée de *Borrelia* dans la peau de souris infectée par inoculation intra-dermique. Panel de gauche : Mesure de la présence de *Borrelia* dans différents organes 40 jours après une inoculation des bactéries. Panel de droite : Quantification PCR du gène codant la flagelline après l'application locale du clobetasol sur la peau de souris 40 jours après l'infection.

L'analyse protéomique non ciblée de biopsies de peaux de souris, prélevées au pic de multiplication 3-5 jours après l'application du dermocorticoïde (clobetasol), avait été réalisée par Benoit Westermann (B. Westermann, 2016). Cette analyse avait permis d'identifier des protéines de *Borrelia* présentes dans la peau, et avait généré une liste de protéines d'intérêt pour le développement d'une méthode de détection spécifique de marqueurs d'infection bactérienne au stade tardif disséminé. Quatorze protéines avaient été sélectionnées (Tableau II-16), sur la base de leur identification dans au moins 2 biopsies murines cutanées (7 jours ou 40 jours post-infection) et de leur identification par les deux moteurs de recherche Mascot et OMSSA. L'étape suivante était de développer une méthode de protéomique ciblée pour détecter de manière spécifique et sensible ces marqueurs d'infection active dans la peau.

2.1. Développement et optimisation de la méthode ciblée

La sélection des peptides (Tableau II-16) avait également été effectuée avant mon arrivée à partir de peptides identifiés au cours de l'analyse non ciblée. Soixante-huit peptides SIS avaient été achetés, correspondant aux 14 protéines cibles.

Tableau II-16. Liste des protéines et peptides protéotypiques pour la méthode de diagnostic de la phase tardive disséminée par microLC-SRM. Les 14 protéines cibles ont été sélectionnées à partir des analyses non ciblées des biopsies cutanées murines infectées. Pour chaque protéine, plusieurs peptides protéotypiques ont été choisis. La dernière colonne indique quels peptides ont été détectés par microLC-SRM et conservés pour le développement de la méthode ciblée.

Protéine	Nombre de peptides protéotypiques	Peptides protéotypiques sélectionnés	Détection LC-SRM
DbpA	5	DITDEIDAIK	Oui
		GVNFDAFK	Oui
		TTANGIIEIVK	Oui
		TVTDAAEQHPPTTAEIGILEIAK	Oui
		VSENSFILEAK	Oui
Flagelline	12	ANLGAFQNR	Oui
		AINFIQTTEGNLNEVEK	Oui
		INTPASLSGSQASWTLR	Oui
		ASDDAAGMGVSGK	Oui
		ELAVQSGNGTYSDSDR	Oui
		GSIQIEIEQLTDEINR	Oui
		IADQAQYNQMHMLSNK	Oui
		MIINHNTSAINASR	Oui
		NNAINAANLSK	Oui
		NNGINAANLSK	Oui
		NSTEYAIENLK	Oui
		TAEELGMQPAK	Oui
GADPH	3	GGYLDHVNHAGAK	Oui
		GIDIVAINDLTDPK	Oui
		VPVPTGSIVDLTVQLK	Oui
BB081	4	DLNYFFEK	Oui
		DSILISNQIK	Oui
		FNISGIIETGLK	Oui
		IDILIPNTK	Oui
BBP42	5	FEEEQTLTK	Non
		INLYGGDK	Non
		TDYIDNIATFK	Non
		LTNSSINFDK	Oui
		SVEVNVLGQFEK	Non
Sensory histidine kinase	5	FAILDFLYSK	Oui
		IAEAANAANK	Oui
		NIDLFSYSK	Oui
		YVASFEDAYK	Oui
		YVLDDNIQVLFSDLR	Oui

Protéine	Nombre de peptides protéotypiques	Peptides protéotypiques sélectionnés	Détection LC-SRM
HSP90	5	EIKPILELNPNNK	Oui
		ELISNASDAIDK	Non
		FINIINEFIEK	Oui
		FLSLTNEK	Oui
		YSNHINYPYIK	Oui
lipoprotéine gi 365823350	5	EFFDWLSK	Oui
		EFISSFDNISIVSK	Non
		GEALSLFFQK	Oui
		SLTEIDSGNGIPLVSDVVK	Oui
		VLTESESNNEK	Oui
VlsE	18	AAAAVTAVSGEQILK	Oui
		AAEEAAK	Oui
		AAGAVSAVSGEQILK	Oui
		AVSAVSGEQILK	Oui
		GIVAAAGK	Oui
		GIVDAAGK	Oui
		IAAAIVLR	Oui
		IGDSAANNAGK	Oui
		IGDSANQGAVADADSVK	Oui
		IGDVGADNK	Oui
		AAAAVSSVSGEQILK	Oui
		AVSSVSGEQILK	Oui
		GIVDAAGTAAGK	Oui
		GIVDAAGTAAGKK	Oui
		IGESADNGAAADADSVK	Oui
		VAAALVLR	Oui
		AAEAVSSVSGEQILK	Oui
		AAEEAIVGATGDGK	Oui
BAPK00593	1	LPLALNLAVSR	Oui
BAPK04515	1	ALINLFNQLEK	Oui
RNA Polymerase	1	LNPNTLEFK	Oui
HP gi 500023077	1	SNISQYFR	Oui
Enolase	2	IQLVGDDLFTNTSFLK	Oui
		SGETEDTTIADLVVALGTGQIK	Oui

Au cours de ma thèse, la méthode de détection ciblée par microLC-SRM a été développée conjointement avec B. Westermann. Les différentes étapes de ce développement sont décrites dans la Figure II-39. Nous avons commencé par analyser le mélange des 68 peptides SIS en mode d'acquisition non programmé (*unscheduled*) en utilisant 18 méthodes d'acquisition différentes et une centaine de transitions par méthode. Sur les 68 peptides, 6 peptides n'ont pas été détectés correctement et ont été éliminés de la méthode. Ces peptides sont indiqués dans la dernière colonne du Tableau II-16.

Suite à ces analyses, nous avons sélectionné 6 transitions par précurseur, soit un total de 408 transitions correspondant à 68 précurseurs pour 62 peptides. Le nombre de précurseurs est

supérieur au nombre de peptides en raison de plusieurs peptides à Met et de deux états de charge pour un des peptides. Puis une méthode d'acquisition programmée (fenêtre de 6 min) a été créée et nous avons optimisé les énergies de collision, comme dans le paragraphe II-2.1.3 en testant 4 valeurs à ± 3 V autour de la valeur préconisée par le constructeur. Ces optimisations ont conduit à une méthode finale ciblant 14 protéines, 62 peptides, 136 précurseurs (68 précurseurs SIS et 68 précurseurs endogènes) et 544 transitions (272 transitions pour les précurseurs SIS et 272 transitions pour les précurseurs endogènes). Cette méthode a été nommée **M4**, pour la différencier des méthodes M1, M2, M3 précédemment développées au laboratoire pour le diagnostic de la phase précoce (G. Schnell *et al.*, 2015b; B. Westermann, 2016).

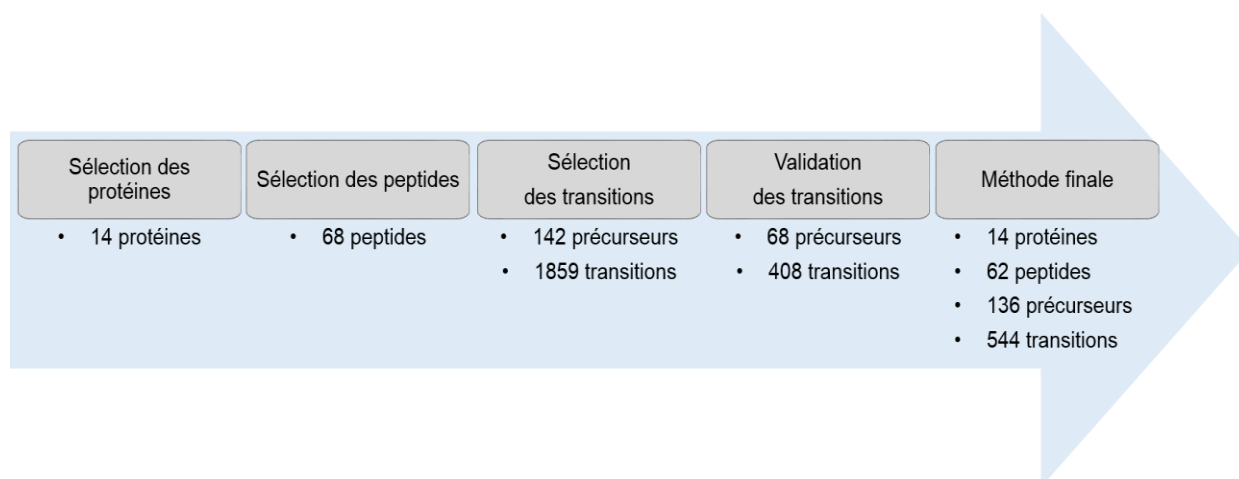


Figure II-39. Différentes étapes de l'optimisation de la méthode SRM pour la détection des protéines de *Borrelia* dans la phase tardive disséminée. Après optimisation cette méthode M4 cible 14 protéines de *Borrelia* par l'intermédiaire de 62 peptides, 136 précurseurs (68 précurseurs SIS et 68 précurseurs endogènes) et 544 transitions (272 transitions pour les précurseurs SIS et 272 transitions pour les précurseurs endogènes).

2.2. Analyse de peaux de souris 40 jours après une infection

Cette méthode microLC-SRM a été utilisée pour analyser des peaux de souris infectées par une inoculation de *Borrelia* à la seringue, puis traitées avec un dermocorticoïde 42 jours après l'inoculation. Les biopsies cutanées ont été prélevées 3-5 jours après l'application du dermocorticoïde, de manière à se trouver au pic de multiplication dans cette phase de réactivation (Figure II-38). Les protéines ont été extraites, préfractionnées sur gel SDS-PAGE, puis digérées et analysées par microLC-SRM.

Huit biopsies ont été analysées, dont quatre biopsies ont été infectées avec *B. burgdorferi* ss N40 et quatre avec *B. afzelii* NE4049. Certaines inoculations ont été faites par voie intra-dermique, d'autres par voie intra-péritonéale. Les taux d'infection ont été déterminés par PCR et variaient de 6 à 304 *fla*/10⁴ *gapdh*. Les échantillons ont été préparés comme décrit dans le Chapitre II.III-1.1 avec un préfractionnement sur gel SDS-PAGE, et un ajout des peptides marqués SIS avant l'injection.

Les critères de validation pour décider qu'un peptide endogène était présent ont été les mêmes que pour l'analyse de la cohorte humaine :

- Co-élution des transitions pour les peptides endogène et SIS,
- Co-élution des peptides endogène et SIS,
- Respect du ratio entre les transitions du peptide endogène et du peptide SIS,
- Identification du peptide dans la bande attendue.

Des peptides de *Borrelia* ont été détectés dans 5 biopsies (Tableau II-17). Certains de ces peptides ont pu être quantifiés car les peptides SIS correspondants étaient de qualité AQUA™. Aucun peptide de *Borrelia* n'a été identifié dans les biopsies présentant les taux d'infection les plus faibles.

Tableau II-17. Protéines et peptides de *Borrelia* détectés par microLC-SRM dans les peaux de souris 40 jours après inoculation et réactivation des bactéries. Des souris ont été infectées par injection intra-péritonéale (IP) ou intra-dermique (ID) de bactéries *Borrelia*, puis au bout de 40 jours les bactéries ont été réactivées par application locale du dermocorticoïde clobetasol. Les biopsies cutanées ont été prélevées et analysées par préfractionnement et microLC-SRM avec la méthode M4. **N.d.** : peptide non détecté. **D** : peptide détecté. **Q** : peptide quantifié (fmol protéine/mg biopsie) grâce au peptide marqué AQUA™.

		<i>B. burgdorferi</i> ss N40				<i>B. afzelii</i> NE4049			
		IP	ID	ID	ID	ID	ID	ID	ID
		Taux d'infection <i>fla</i> /10 ⁴ <i>gapdh</i>							
Protéines détectées	Peptides	Souris 1	Souris 2	Souris 3	Souris 4	Souris 1	Souris 2	Souris 3	Souris 4
DbpA	DITDEIDAIK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	GVNFDAFK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	TTANGIIEIVK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	TVTDAAEQHPTTTAEGILEIAK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	VSENSFILEAK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
Flagelline	ANLGAFQNR	Q (45)	Q (5)	N.d.	N.d.	Q (36)	Q (4)	N.d.	N.d.
	AINFIQTTEGNLNEVEK	Q (94)	N.d.	N.d.	N.d.	Q (52)	N.d.	N.d.	N.d.
	INTPASLSGSQASWTLR	Q (88)	Q (14)	N.d.	N.d.	Q (77)	N.d.	N.d.	N.d.
	ASDDAAGMGVSGK	D	D	N.d.	N.d.	D	N.d.	N.d.	N.d.
	ELAVQSGNGTYSDSDR	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	GSIQIEIQLTDEINR	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	IADQQAQYNQMHLMSNK	D	D	N.d.	N.d.	D	N.d.	N.d.	N.d.
	MIINHNTSAINASR	D	D	N.d.	N.d.	D	N.d.	N.d.	N.d.
	NNAINAANLSK	N.d.	N.d.	N.d.	N.d.	D	D	N.d.	N.d.
	NNGINAANLSK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	NSTEYAIENLK	N.d.	N.d.	N.d.	N.d.	D	N.d.	N.d.	N.d.
	TAEELGMQPAK	D	D	N.d.	N.d.	D	N.d.	N.d.	N.d.
HP BAPKO_0593	LPLALNLAVSR	N.d.	N.d.	N.d.	N.d.	D	N.d.	N.d.	N.d.
Lipoprotéine gi 365823350	EFFDWLSK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	GEALSLFFQK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	SLTEIDSGNGIPLVSDVVK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	VLTESESNNELK	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
Vls13	AAAAVSSVSGEQILK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	AVSSVSGEQILK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	GIVDAAGTAAGK	D	D	D	N.d.	D	N.d.	N.d.	N.d.
	GIVDAAGTAAGKK	D	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	IGESADNGAAADADSVK	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
Vls2	VAAALVLR	D	D	N.d.	N.d.	D	N.d.	N.d.	N.d.
	AAEAVSSVSGEQILK	D	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	AAEEAIVGATGDGK	D	D	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
	Nombre de peptides	18	11	1	0	12	2	0	0
	Typage de la bactérie	Oui	Oui	Oui	Non	Oui	Oui	Non	Non

On constate que la **flagelline** a été détectée dans quatre souris et grâce à plusieurs peptides. Elle a été quantifiée grâce à trois peptides AQUA™ et on peut voir que les trois valeurs sont du même ordre de grandeur pour une souris donnée (76 ± 27 fmol /mg biopsie pour la souris 1 infectée par *B.burgdorferi* ; 55 ± 21 fmol /mg biopsie pour la souris 1 infectée par *B. afzelii*). On peut aussi noter que c'est dans les deux biopsies les plus infectées (d'après PCR sur gène *fla*) que la flagelline a été la mieux détectée. La flagelline avait déjà été détectée dans les biopsies murines dans la phase précoce après l'infection (G. Schnell *et al.*, 2015b), et faisait partie de la liste des protéines ciblées dans les méthodes SRM/PRM du paragraphe 2.1.1. Elle a été détectée dans plusieurs biopsies de la cohorte humaine. Cette protéine apparait donc

également comme un marqueur d'infection active pour le diagnostic de la phase tardive disséminée.

Dans une biopsie, nous avons également identifié deux protéines (**DbpA**, **lipoprotéine gi|365823350**) qui faisaient partie des protéines ciblées par SRM/PRM pour la phase précoce.

Une protéine détectée dans trois souris est la **VlsE** (*Variable like protein sequence Expressed*) avec deux isoformes Vls13 et Vls2. Cette lipoprotéine est composée de régions variables qui sont en constante mutation ce qui permet à la bactérie d'échapper au système immunitaire de l'hôte (S. J. Norris, 2014). Le peptide C6 ou IR6 est une partie très conservée entre les différentes espèces de *Borrelia*, et a été identifié comme très immunogène. Cette protéine a été ajoutée au test sérologique ELISA de diagnostic de la maladie de Lyme, afin d'augmenter la sensibilité et la spécificité du test (F. T. Liang *et al.*, 2002). Dernièrement, il a été question de remplacer le test de Western blot par un second test ELISA dirigé contre VlsE (C6) dans le test diagnostique de la borréliose de Lyme réalisé en deux niveaux (Figure II-6) (J. J. Halperin, 2019). Cette protéine apparaît donc également comme un marqueur d'infection active pour le diagnostic de la phase tardive disséminée.

Enfin, l'étude de souris infectées par deux espèces différentes de *Borrelia* nous a permis de montrer l'efficacité de l'analyse microLC-SRM pour le **typage** de l'espèce infectante par la détection de peptides discriminant chaque espèce.

La démonstration de la latence cutanée de *Borrelia* dans la peau des souris, l'identification des protéines de *Borrelia* dans la peau des souris et le développement de la méthode LC-SRM (M4) ont fait l'objet de la publication : *Identification of Borrelia protein candidates in mouse skin for potential diagnosis of disseminated Lyme borreliosis*, Grillon, A., Westermann, B., Cantero, P., Jaulhac, B., Voordouw, M.J., Kapps, D., Collin, E., Barthel, C., Ehret-Sabatier, L., Boulanger, N. Sci Reports (2017) 7(1):16719.

2.3. Modifications de la méthode ciblée

Les analyses des huit peaux de souris ont montré que plusieurs peptides ciblés dans la méthode M4 n'avaient jamais été détectés, même dans les deux biopsies de souris les plus infectées (Tableau II-17). Nous avons alors décidé de poursuivre l'optimisation de la méthode SRM pour supprimer ces peptides et tenter d'augmenter la sensibilité. Cette optimisation a été effectuée en collaboration avec Ludivine Esteves-Gloria (CDD 2017-2018) et nous avons travaillé avec un mélange de peptides dopés dans des peaux de souris saines. Nous avons

supprimé 4 peptides ainsi que 34 transitions qui étaient faiblement observés (Figure II-40). Puis nous avons encore supprimé 8 peptides qui n'étaient jamais détectés, dont la protéine HP gi|500023077 (ciblée par un seul peptide), et 2 peptides qui ne satisfaisaient plus les critères de protéotypicité. En revanche, nous avons ajouté 5 peptides spécifiques de *B. afzelli* pour améliorer le typage de l'espèce infectante. Nous avons ainsi obtenu une méthode M6 ciblant 13 protéines, 53 peptides, 112 précurseurs (56 précurseurs SIS et 56 précurseurs endogènes) et 536 transitions (268 transitions pour les précurseurs SIS et 268 transitions pour les précurseurs endogènes). Nous avons également commandé 13 peptides AQUA™, afin d'avoir au moins un peptide AQUA™ par protéine pour la quantification.

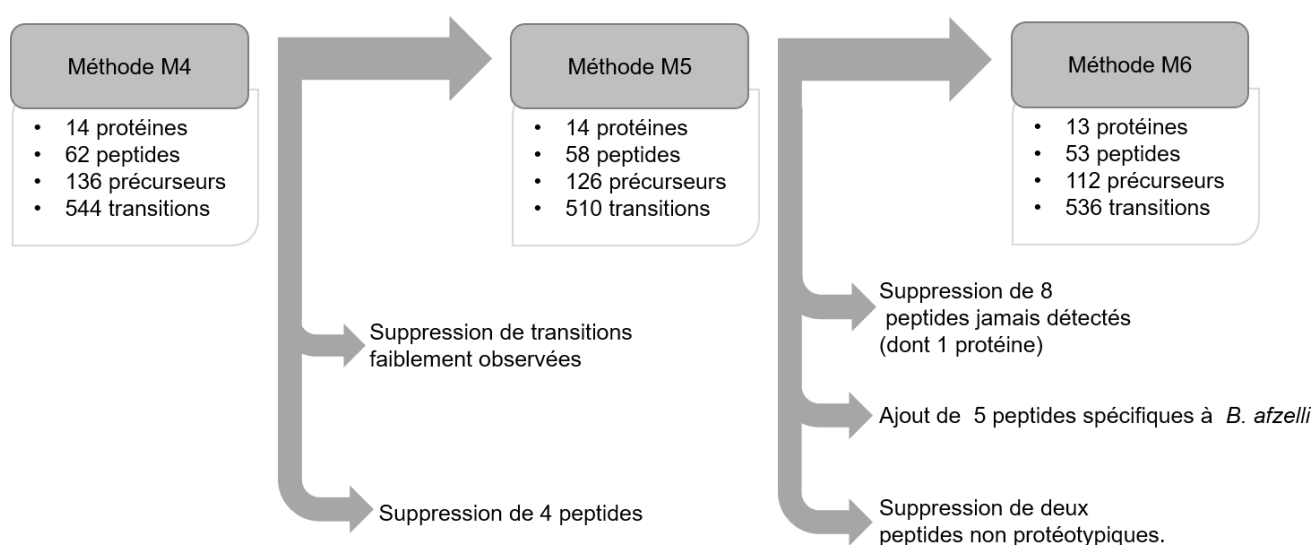


Figure II-40. Modifications apportées à la méthode SRM pour la détection des protéines de *Borrelia* dans la phase tardive disséminée. La méthode M4 a été modifiée en supprimant des transitions et peptides non détectés et en ajoutant des peptides spécifiques à *B. afzelli*. Après optimisation, la méthode M6 cible 13 protéines de *Borrelia* par l'intermédiaire de 53 peptides, 112 précurseurs et 536 transitions.

Afin d'évaluer cette méthode pour la détection de *Borrelia* dans la peau humaine, nous l'avons testée sur des biopsies cutanées humaines saines dopées avec deux quantités de *B. afzelli* (10^4 et 10^5 bactéries). Dans la peau dopée avec 10^4 bactéries, 3 peptides correspondant à deux protéines (flagelline et GAPDH) ont été détectés ; dans celle dopée avec 10^5 bactéries, 18 peptides correspondant à cinq protéines (flagelline, GAPDH, Enolase, Bapko_0593 et lipoprotéine gi|365823350|) ont été détectés. Un total de quatre protéines a pu être quantifié à l'aide des peptides marqués AQUA™ (Tableau II-18).

Tableau II-18. . Protéines et peptides de *Borrelia* détectés par LC-SRM dans les peaux humaines dopées avec *B. afzelii*. Des quantités variables de *B. afzelii* ont été utilisées pour doper les biopsies cutanées humaines (10^4 et 10^5 bactéries). Les biopsies cutanées ont été analysées par préfractionnement et microLC-SRM avec la méthode M6. Les séquences en rouge correspondent aux peptides de qualité AQUA™, les autres peptides sont de qualité *crude*. N.d. : peptide non détecté. D : peptide détecté. Q : peptide quantifié (fmol protéine) grâce au peptide marqué AQUA™.

Protéines détectées	Peptides	10 ⁴ bactéries	10 ⁵ bactéries
Enolase gi 3913583	SGETEDTTIADLVVALGTGQIK	N.d	Q (85)
Flagelline	ANLGAFAQNR	Q (2)	Q (163)
	AINFIQTTEGNLNEVEK	N.d	Q (165)
	INTPASLSGSQASWTLR	N.d	Q (255)
	ASDDAAGMGVSGK	N.d	D
	ELAVQSGNGTYSDDSR	N.d	D
	GSIQIEIEQLTDEINR	N.d	D
	IADQAQYNQMHLMSNK	N.d	D
	MIINHNTSAINASR	N.d	D
	MIINHNTSAINASR	N.d	D
	NNAINAANLSK	D	D
	NSTEYAIENLK	N.d	D
	TAEELGMQPAK	N.d	D
	NNGINAANLSK	N.d	N.d
GADPH	GGYLDHVNHAGAK	N.d	D
	GIDIVAINDLTPDK	N.d	Q (93)
	VPVPTGSIVDLTVQLK	D	D
HP BAPKO_0593	LPLALNLAVSR	N.d	Q (6)
HP BAPKO_4515	ALINLNFNQLLEK	N.d	N.d
Lipoprotéine gi 365823350	GEALSLFFQK	N.d	N.d
	SLTEIDSGNGIPLVSDVVK	N.d	N.d
	VLTESESNNELK	N.d	D

Ces résultats montrent que la méthode M6 pourra être utilisée pour analyser des biopsies de patients. Ceci est prévu dans le cadre d'une étude clinique qui est en cours d'acceptation par l'ANSM. Dans un premier temps, il est prévu d'analyser 10 patients atteints d'une borréliose de Lyme disséminée confirmée ou probable et définie selon les critères diagnostiques actuels de référence (neuroborréliose, arthrite de Lyme, acrodermatite chronique atrophique). Dans un second temps, le projet sera étendu en incluant 20 patients supplémentaires « avec borréliose de Lyme » et 50 patients « sans borréliose de Lyme » qui sont soit des patients symptomatiques (d'autres pathologies que Lyme) mais avec sérologie Lyme négative, soit des patients asymptomatiques avec sérologie Lyme négative.

Pour cette étude clinique, une réactivation par un dermocorticoïde sera nécessaire compte tenu de la faible quantité de *Borrelia* qui persiste dans la peau, comme cela a été montré dans

le modèle murin. Cette application est prévue de manière courte (48h) sur le site prévu pour les biopsies cutanées. C'est pourquoi, en vue de cette étude clinique, nous avons voulu comparer l'effet de deux dermocorticoïdes et l'étendue de la réactivation qui pourrait être problématique pour les patients. Compte tenu des problèmes rencontrés lors de l'analyse de la cohorte humaine (phase précoce, paragraphe 1.2.2), nous avons également voulu évaluer l'effet de la lidocaïne, anesthésique utilisé au moment de la collecte des biopsies, sur la multiplication des bactéries *Borrelia*.

2.4. Comparaison de deux dermocorticoïdes

Dans les expériences précédentes, le dermocorticoïde utilisé pour la réactivation était le clobetasol, un corticostéroïde utilisé pour traiter les affections cutanées telles que l'eczéma, la dermatite de contact, la dermatite séborrhéique et le psoriasis. Nous avons comparé l'effet du clobetasol à celui du locapred qui est un anti-inflammatoire plus modéré. Pour cela nous avons analysé deux peaux de souris infectées par *Borrelia*, puis après 40 jours, traitées par l'un ou l'autre de ces dermocorticoïdes pendant deux jours. Ces biopsies ont été analysées par LC-SRM avec les méthodes M5 et M6 décrites précédemment. Le Tableau II-19 montre les protéines de *Borrelia* détectées dans ces deux biopsies.

Tableau II-19. Comparaison des protéines détectées par microLC-SRM après l'application de deux dermocorticoïdes. Des souris ont été infectées par inoculation de *Borrelia*, puis au bout de 42 jours les bactéries ont été réactivées par application de clobetasol ou locapred pendant 2 jours. Les biopsies cutanées ont été analysées par préfractionnement et microLC-SRM en utilisant les méthodes M5 et M6. X : protéine détectée ; entre parenthèses : nombre de peptides détectés ; N.d : protéine non détectée.

	Méthode LC-SRM	Clobetasol		Locapred	
		M5	M6	M5	M6
Protéines	Flagellin	N.d	X(5)	X(1)	N.d
	GAPDH	N.d	X(1)	N.d	N.d
	Liproprotein gi 365823350	N.d	X(1)	N.d	N.d
	VLSE	N.d	X(2)	N.d	N.d
	VLS3	X(1)	X(1)	N.d	N.d

Ce tableau montre que le clobetasol permet une meilleure réactivation des bactéries *Borrelia* que le locapred. Cinq protéines de *Borrelia* ont été détectées avec le clobetasol, alors que seule la flagelline a été détectée - avec un seul peptide - avec le locapred. Ceci est en accord avec le pouvoir anti-inflammatoire modéré du locapred.

Si on regarde les résultats obtenus pour le clobetasol, ce tableau illustre également le bénéfice de la méthode M6 par rapport à la méthode M5, résultat de l'optimisation décrite précédemment. Pour le locapred, la non-détection du peptide de flagelline avec la méthode M6 peut s'expliquer par la suppression du précurseur correspondant en passant de la méthode M5 à la méthode M6.

En parallèle, nos collaborateurs ont réalisé une étude exhaustive de l'effet des dermocorticoïdes pour la réactivation des bactéries dans les biopsies cutanées. Ces différentes expériences ont montré, dans un modèle murin, qu'il est possible de réactiver la prolifération bactérienne en appliquant localement un dermocorticoïde, non seulement dans le site de l'inoculation de la bactérie, mais également à un site distant. De plus, ils ont montré qu'il y a un lien entre la réactivation de la bactérie et l'activité anti-inflammatoire du dermocorticoïde (Figure II-41). En effet, le clobetasol est considéré comme ayant une très forte activité anti-inflammatoire, la bétaméthasone et le désonide comme ayant une forte activité anti-inflammatoire et l'hydrocortisone comme ayant une faible activité anti-inflammatoire. Ces résultats désignent le clobetasol comme dermocorticoïde avec un meilleur effet pour la réactivation de la prolifération bactérienne dans les biopsies cutanées murines.

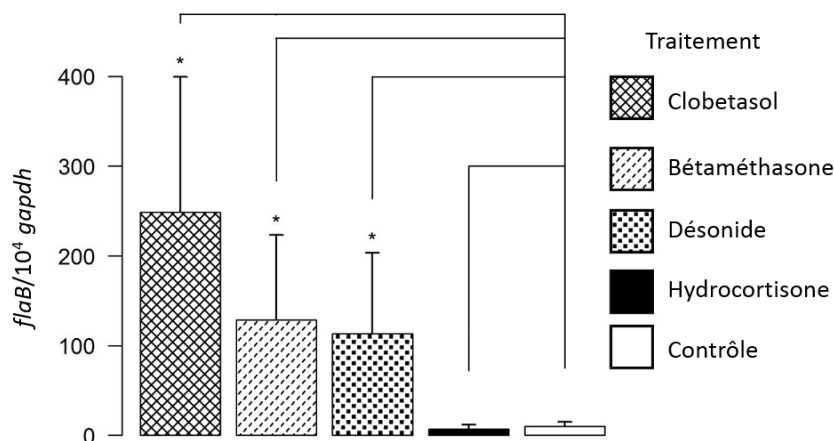


Figure II-41. Détermination de la réactivation de *B. afzelii* NE4049 par PCR. Application locale des dermocorticoïdes avec différentes activités anti-inflammatoires au site de l'inoculation 50 jours après l'inoculation. L'activité anti-inflammatoire des dermocorticoïdes va de très forte à faible : clobetasol (très forte), bétaméthasone et désonide (forte) et hydrocortisone (faible).

2.5. Effet de la lidocaïne

L'effet négatif de la lidocaïne sur les bactéries avait été suspecté pendant l'analyse de la cohorte humaine avec la difficulté de détecter des protéines de *Borrelia* dans les premières biopsies (sans culture préalable des bactéries). Par la suite nos collaborateurs ont montré un effet létal de la lidocaïne (5 mg/mL) *in vitro* sur *B. afzelli* et *B. burgdorferi*. Par observation au microscope ils ont également montré une viabilité plus faible des spirochètes dans une culture de peau infectée traitée par la lidocaïne.

Pour compléter ces observations, nous avons analysé des peaux de souris prélevées au pic de multiplication (J7) après une inoculation de *Borrelia* et

- traitées ou non par de la lidocaïne (10 mg/mL),
- mises en culture pendant 7 jours avant l'analyse ou analysées directement après prélèvement.

Quatre biopsies de peaux de souris ont ainsi été analysées selon une stratégie de préfractionnement SDS-PAGE puis analyse non ciblée LC-MS/MS (couplage nanoAcquity/ Q-Exactive+). Les résultats sont rassemblés dans le Tableau II-20. On observe qu'en absence de culture, les nombres de protéines et de peptides de *Borrelia* détectés dans la souris traitée à la lidocaïne sont proches de ceux détectés dans la souris non traitée (64 versus 70, et 186 versus 206). Après 7 jours de culture, le nombre de protéines de *Borrelia* détectées est nettement plus important (6-7x) en accord avec la multiplication des bactéries. Les nombres de protéines et de peptides de *Borrelia* détectés sont légèrement moins élevés dans la souris traitée à la lidocaïne que dans la souris non traitée (397 versus 474, et 1992 versus 2736), mais ces expériences n'ayant été faites qu'une seule fois, avec une seule analyse par biopsie, il est difficile de tirer une conclusion nette sur l'effet de la lidocaïne. Une hypothèse est également que la culture permette d'éliminer la lidocaïne présente sur la peau, ce qui réduit son impact sur la multiplication des *Borrelia*.

Tableau II-20. Etude de l'impact de la lidocaïne avant ou après multiplication des bactéries. Nombre de protéines et de peptides identifiés dans des peaux de souris infectées puis traitées ou non (contrôle) par de la lidocaïne avant le prélèvement. Jour 0 : analyses protéomiques réalisées sans mise en culture. Jour 7 : analyses protéomiques réalisées après 7 jours de mise en culture.

	<i>fla/10⁴ gapdh</i>	<i>fla/100ng</i>	Nombre total de protéines identifiées	Nombre total de peptide identifiés	Nombre de protéines de <i>Borrelia</i> identifiées	Nombre de peptides de <i>Borrelia</i> identifiés
Contrôle Jour 0	252	836	4274	31100	70	206
Lidocaïne Jour 0	272	838	4685	37300	64	186
Contrôle Jour 7	N.d	15200	3140	21733	474	2736
Lidocaïne Jour 7	N.d	22700	3691	25427	397	1992

Quant au fort impact d'une semaine de culture sur le nombre de protéines et de peptides de *Borrelia* détectés, il explique la différence de sensibilité observée lors de l'analyse de la cohorte humaine avant (sans culture) et après (8-12 semaines de culture) changement de protocole.

La démonstration de la réactivation de la prolifération de *Borrelia* dans la peau des souris et l'effet de la lidocaïne ont fait l'objet de la publication : *Effects of topical corticosteroids and lidocaine on Borrelia burgdorferi sensu lato in mouse skin: potential impact to human clinical trials*, Lefeuvre, B., Cantero, P., Ehret-Sabatier, L., Lenormand, C., Barthel, C., Po, C., Parveen, N., Grillon, A., Jaulhac, B., Boulanger, N. Sci Reports (2020) 10(1):10552.

IV- Conclusion

Cette partie de mes travaux de thèse avait pour objectif d'identifier des protéines de *Borrelia* dans la peau d'hôtes infectés, puis de mettre en place des approches ciblées pour les détecter de façon spécifique et les quantifier, et ce dans le cadre d'un premier projet vaccinal contre la borréliose canine, et d'un second projet de diagnostic direct de la borréliose humaine.

En ce qui concerne le développement d'un nouveau vaccin, l'approche protéomique a permis de détecter des protéines de *Borrelia* dans les peaux de chiens infectés malgré des taux d'infection extrêmement faibles. Nos analyses protéomiques ont été déterminantes pour le choix des protéines candidates-vaccin. Ces protéines ont été produites, associées dans une formulation, puis ont fait l'objet d'un essai vaccinal sur quelques chiens. Aucune protection n'a malheureusement pu être observée chez les chiens vaccinés, mais des pistes d'amélioration sont à l'étude. L'efficacité de l'analyse protéomique pour la détection des protéines bactériennes présentes en quantité très faible dans la peau canine n'en reste pas moins très marquante.

En ce qui concerne l'approche diagnostique, nous avons validé à l'échelle d'une cohorte humaine de 70 patients, avec des résultats très satisfaisants, l'approche protéomique ciblée comme méthode alternative de diagnostic direct au stade précoce de la borréliose de Lyme. La sensibilité de la protéomique ciblée s'est avérée meilleure que celle de la PCR et de l'observation des bactéries au microscope à fond noir. Les nombreux problèmes rencontrés nous ont conduits à implémenter le mode PRM en plus du mode SRM initialement prévu. La complémentarité de ces deux approches ciblées a été montrée, en termes de LLOQ mais également lorsque nous les avons appliquées en parallèle sur des biopsies de patients. Ce projet est presque arrivé à son terme. Même si nous avons déjà observé 10 biopsies négatives dans la cohorte, nous avons décidé de collecter une dizaine de biopsies de patients sains (non infectés) pour nous assurer parfaitement de la spécificité de l'analyse protéomique.

En ce qui concerne la phase tardive disséminée, nous avons optimisé une méthode de détection ciblée des protéines de *Borrelia* dans la peau de souris après réactivation des bactéries par un dermocorticoïde. Cette méthode est applicable à la peau humaine et devrait être utilisée à court terme pour des peaux de patients. Afin de préparer au mieux cette étude clinique nous avons également évalué différents dermocorticoïdes et l'impact de la lidocaïne. D'autres prélèvements biologiques tels que le liquide synovial et le liquide céphalo-rachidien sont envisagés et ont commencé à être étudiés au laboratoire.

Chapitre III. Etude des agents pathogènes transmis par des tiques

I- Maladies vectorielles à tiques : acquisition et transmission de pathogènes

Les pathogènes associés aux maladies vectorielles sont transmis par des vecteurs variés, insectes (moustique, puce, pou, etc.) ou acariens (tiques principalement). Ces vecteurs s'infectent sur un hôte vertébré, animal sauvage ou domestique ; ils peuvent ensuite transmettre ces pathogènes à d'autres animaux ou à l'homme suite à la piqûre de ce vecteur infecté (F. Dantas-Torres *et al.*, 2016). Les maladies à transmission vectorielle représentent un enjeu majeur en santé publique au niveau mondial. Selon les données de mars 2020 de l'Organisation Mondiale de la Santé (<https://www.who.int/fr/>), elles représentent plus de 17% des maladies infectieuses et sont responsables de plus de 700 000 décès par an.

Les tiques sont considérées comme le vecteur le plus important, avant les moustiques, dans le domaine médical et vétérinaire (F. Dantas-Torres *et al.*, 2016; A. Toledo *et al.*, 2009). Les tiques sont caractérisées par leur capacité à transmettre un grand nombre de pathogènes (F. Dantas-Torres *et al.*, 2016). Par conséquent, les maladies transmises par les tiques représentent un problème de santé publique, et dans le cas des animaux, elles peuvent induire des pertes économiques considérables en production animale, avec l'anaplasmose ou la babésiose (O. Hajdusek *et al.*, 2013).

Dans ces travaux de thèse, nous nous sommes intéressés aux maladies vectorielles à tiques et aux pathogènes associés dans le cadre de maladies bien connues mais aussi pour des pathologies émergentes.

1. Vecteur tique

Les tiques sont des arthropodes hématophages, ectoparasites de vertébrés. Dans l'ordre Ixodida, à l'heure actuelle approximativement 956 espèces de tiques ont été décrites. Les tiques sont regroupées en trois familles: celle des ixodidés ou « tiques dures » et celle des argasidés ou « tiques molles » et les nuttalliellidés (F. Dantas-Torres *et al.*, 2019; K. D. McCoy *et al.*, 2017). Parmi les espèces de tiques, *I. ricinus* appartenant à la famille des ixodidés est l'espèce prédominante en Europe et le principal vecteur de pathogènes pour l'homme et les

animaux (E. Lejal *et al.*, 2019). Ce travail de thèse porte uniquement sur l'espèce de tique *I. ricinus*.

1.1. Cycle de vie de *I. ricinus*

Au cours de son cycle de vie, l'espèce *I. ricinus* passe par trois stases après l'œuf: la larve, la nymphe et l'adulte, mâle et femelle. Chaque stase est caractérisée par un repas sanguin unique (Figure III-1). La tique *I. ricinus* étant une tique généraliste, cela lui permet d'accomplir son cycle en 2 ou 3 ans sur des hôtes très variés. Elle peut observer des diapauses développementales selon le climat et des diapauses comportementales si elle ne rencontre pas d'hôte pour son repas sanguin (K. D. McCoy *et al.*, 2017) .

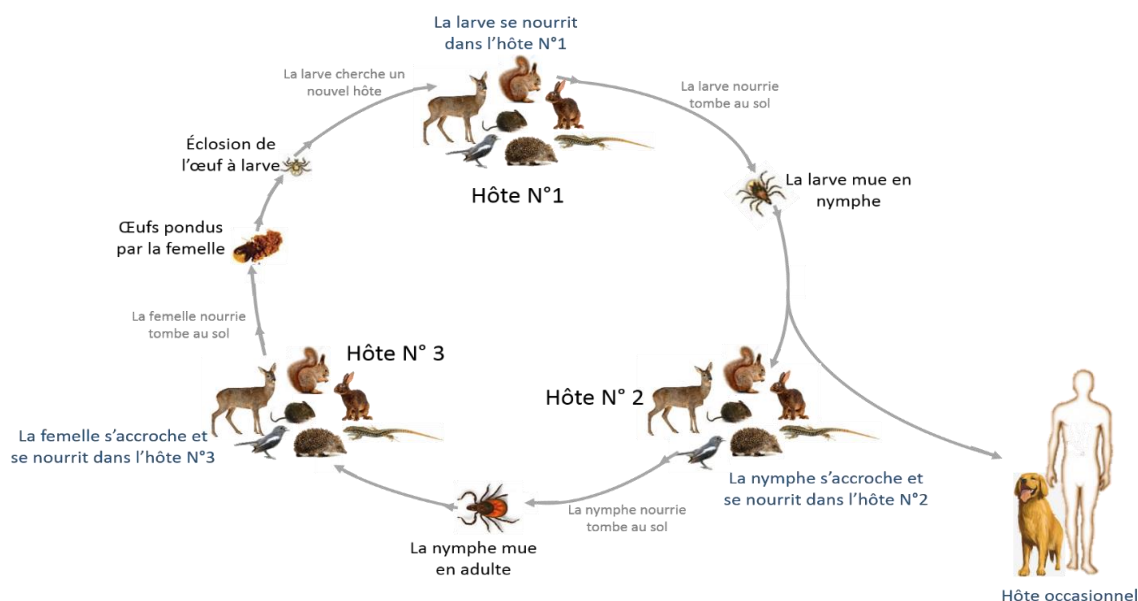


Figure III-1. Cycle de vie de la tique *Ixodes spp.* La larve se nourrit sur un premier hôte vertébré pendant 2 à 3 jours, tombe au sol et mue en nymphe. La nymphe recherche un deuxième hôte sur lequel elle prendra son deuxième repas sanguin qui dure 3 à 5 jours. Elle tombe au sol pour muer à la stase adulte. Un troisième hôte est nécessaire aux femelles pour ce dernier repas sanguin qui dure 10 jours environ et pour la ponte. L'accouplement a lieu sur l'hôte. Graphique adapté de (J. S. Gray *et al.*, 2016).

C'est au cours de ces différents repas sanguins que la tique est susceptible de s'infecter. La larve qui se nourrit sur un hôte infecté, le demeure toute sa vie. On parle de transmission transtadiale des pathogènes. La transmission transovarienne, de la femelle à la descendance, est rare. Les vertébrés que la tique rencontre au cours de son cycle de vie, peuvent être réservoir de plusieurs pathogènes, en conséquence les tiques à leur tour peuvent être infectées par plusieurs pathogènes. Les tiques ont aussi la particularité de pouvoir s'infecter

par « *co-feeding* » (transmission de pathogènes lors de repas commun entre deux tiques proches l'une de l'autre et se nourrissant sur le même hôte) (M. Kazimirova *et al.*, 2017)

1.2. Interactions hôte-tique-pathogènes

La tique *I. ricinus* est vectrice d'une grande diversité d'agents infectieux, tels que des bactéries (*B. burgdorferi sensu lato* (sl), *A. phagocytophilum*, *Neoehrlichia mikurensis*, *Rickettsia*....), des parasites (*Babesia*) et des virus (virus de l'encéphalite à tiques) (S. Moutailler *et al.*, 2016; P. Parola *et al.*, 2001; A. Rizzoli *et al.*, 2014). Cette espèce de tique est connue comme un vecteur généraliste, dû à sa capacité de se nourrir sur une grande diversité d'animaux ; plus de 300 espèces différentes de vertébrés (C. Herrmann *et al.*, 2015). Selon leurs stases, les tiques se nourrissent sur des animaux de tailles différentes. Les tiques adultes se nourrissent principalement sur des hôtes de moyenne et grande taille (lièvres, cerfs, hérissons, etc.) et les stases immatures peuvent se retrouver sur toute taille d'hôtes, principalement des rongeurs, des lézards et des oiseaux (J. S. Gray *et al.*, 2016) (Tableau III-1).

A cela s'ajoute, une circulation particulière des agents infectieux en fonction des hôtes. Ceci est particulièrement vrai pour les *Borrelia* : les rongeurs sont des hôtes privilégiés pour *B. afzelii*, les oiseaux pour *B. garinii*, par contre les cervidés ne sont pas des réservoirs de *Borrelia* (Tableau III-1) (A. Rizzoli *et al.*, 2014). Ceci est dû au système complément de ces hôtes qui va lyser certaines *Borrelia* (K. Kurtenbach *et al.*, 2006). De manière plus générale, les petits mammifères sont d'excellents réservoirs pour les pathogènes transmis par les tiques comme *B. microti*, *B. miyamotoi*, le virus de l'encéphalite à tique (TBEV) et *N. mikurensis*. *A. phagocytophilum* est un pathogène très ubiquiste.

Cette circulation des pathogènes explique également que des co-infections peuvent être observées chez les tiques *Ixodes* (S. Moutailler *et al.*, 2016). Cette co-infection s'accroît davantage chez les tiques adultes à cause du nombre de repas que la tique a pris au cours de sa vie (P. M. Jensen *et al.*, 2017). Cependant, d'autres facteurs peuvent influencer ces co-infections pour une stase en particulier, tels que la réceptivité de l'hôte réservoir au pathogène, le phénomène de co-repas et la compétence vectorielle qui désigne l'aptitude d'un vecteur à s'infecter et à transmettre un agent pathogène à un autre hôte (J. de la Fuente *et al.*, 2017a; M. Eder *et al.*, 2018).

Tableau III-1. Principaux hôtes et pathogènes transmis par la tique *I. ricinus*.

L : larve ; N : nymphe et A : adulte. Tableau adapté de (A. Rizzoli *et al.*, 2014). La larve et la nymphe ont une affinité pour des hôtes de toute taille, l'adulte a une affinité plus importante pour les vertébrés de moyenne et grande taille.

Hôte	Stase	Pathogènes
Campagnol	L, N	<i>B. afzelii</i> <i>B. burgdorferi</i> s.s <i>B. miyamotoi</i> <i>C. N. mikurensis</i> <i>A. phagocytophilum</i> <i>B. microti</i> TBEV
Mulot	L,N	<i>B. afzelii</i> <i>B. burgdorferi</i> s.s <i>B. miyamotoi</i> <i>B. spielmanii</i> <i>C. N. mikurensis</i> <i>A. phagocytophilum</i> <i>B. microti</i> TBEV
Rat	L,N	<i>B. afzelii</i> <i>B. Spielmanii</i> <i>A. phagocytophilum</i>
Ecureuil	L,N	<i>B. afzelii</i> <i>B. burgdorferi</i> s.s <i>B. garinii</i> TBEV
Lièvre	L,N,A	<i>B. burgdorferi</i> s.l <i>A. phagocytophilum</i> TBEV
Musaraigne	L,N	<i>B. burgdorferi</i> s.l <i>A. phagocytophilum</i> <i>B. microti</i>
Hérisson	L,N,A	<i>B. afzelii</i> <i>B. spielmanii</i> <i>B. bavariensis</i> <i>A. phagocytophilum</i> <i>B. C. N. mikurensis</i> TBEV
Chevreuil	L,N,A	<i>A. phagocytophilum</i>
Cerf	L,N,A	<i>A. phagocytophilum</i>
Renard	L,N,A	<i>B. burgdorferi</i> s.l <i>A. phagocytophilum</i>
Blaireau	L,N,A	<i>B. afzelii</i> <i>B. valaisiana</i>
Oiseaux	L,N	<i>B. garinii</i> <i>B. valaisiana</i> TBEV <i>Rickettsia</i> <i>A. phagocytophilum</i> <i>B. microti</i>

1.3. Répartition géographique et habitat

I. ricinus est l'espèce de tiques la plus répandue en Europe (G. Stanek *et al.*, 2011a; G. Stanek *et al.*, 2012). Cette espèce est également connue comme le vecteur le plus important de maladies humaines et animales en Europe (M. Di Venere *et al.*, 2015). La dernière mise à jour de janvier 2019 du Centre Européen de prévention et de Contrôle (ECDC), montre la répartition d'*I. ricinus* en Europe (Figure III-2).

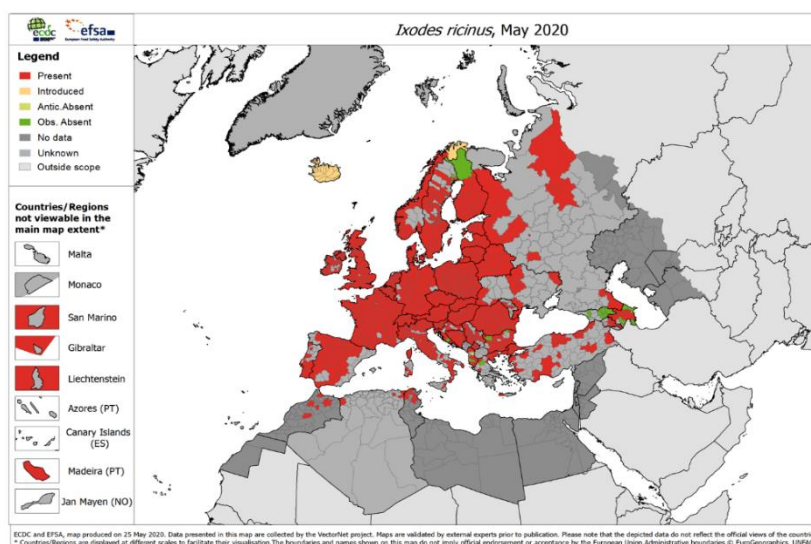


Figure III-2. Distribution d'*I. ricinus* en Europe (source ECDC,2020)

Ces dernières années, une augmentation de la population de tiques a été observée. Les conséquences directes de cette augmentation sont une plus grande répartition géographique et donc une adaptation des tiques à de nouveaux environnements (S. Madison-Antenucci *et al.*, 2020). La caractéristique généraliste pour le choix de l'hôte permet à cette tique de se disperser facilement dans la nature, par le biais des hôtes qui se déplacent. Les oiseaux ont un rôle important dans cette dispersion (A. M. Buczek *et al.*, 2020).

La tique *I. ricinus* étant hygrophile, elle occupe des habitats avec une humidité élevée pouvant atteindre 80% (G. Stanek *et al.*, 2012) et une température comprise entre 8 et 24°C (E. Lejal *et al.*, 2019). Elle se retrouve dans des zones boisées, lisières de forêt et prairies (C. Perez-Eid *et al.*, 1998). Suite aux changements environnementaux, les tiques ont montré leur capacité à s'adapter à de nouvelles conditions. Cette caractéristique pourrait expliquer le nombre croissant des piqûres de tiques dans les parcs urbains ou les forêts suburbaines (A. Grochowska *et al.*, 2020; A. Rizzoli *et al.*, 2014).

2. Pathogènes transmis à l'hôte et maladies associées

Comme indiqué précédemment, les tiques *I. ricinus* peuvent se nourrir sur une grande variété d'hôtes. A chaque repas la tique peut infecter l'hôte, s'infecter à partir d'un hôte infecté (dont certains peuvent être réservoirs de multiples pathogènes) ou effectuer les deux à la fois. Une fois que le vecteur est infecté, il est capable de transmettre l'agent pathogène lors du repas de sang ultérieur.

L'homme et les animaux de compagnie représentent des hôtes accidentels qui peuvent être infectés par différents pathogènes. Ceci dépendra de la compétence vectorielle de la tique à transmettre le pathogène. Cependant, le risque d'être infecté dépend aussi d'un certain nombre de paramètres tels que la densité vectorielle (abondance de vecteurs dans l'environnement), la durée du repas sanguin et la réceptivité de l'hôte (la sensibilité de l'hôte à l'infection et la faculté des pathogènes à échapper aux réponses immunitaires)(J. de la Fuente *et al.*, 2017a; O. Hajdusek *et al.*, 2013; K. D. McCoy *et al.*, 2017). Les pathogènes transmis par *I. ricinus* peuvent générer des maladies chez l'homme telles que la borréliose de Lyme, l'anaplasmose humaine, la rickettsiose, la babésiose, l'encéphalite à tique, etc. (S. Jahfari *et al.*, 2016; A. Portillo *et al.*, 2018). La durée du repas sanguin de la tique augmente le risque de contracter une maladie à tique. Pour les bactéries et les parasites, il y a un délai de 12 à 24 heures en Europe. Les agents infectieux doivent d'abord mûrir avant d'être transmis. Par contre pour les virus, ils sont transmis immédiatement, les virus se trouvant dans les glandes salivaires de la tique (J. de la Fuente *et al.*, 2017b). Les pathogènes étudiés lors de ces travaux de thèse ainsi que les maladies associées sont résumés dans le Tableau III-2

La **borréliose de Lyme** a été amplement détaillée dans le chapitre II, à noter que seules les espèces pathogènes étudiées au cours de mes travaux de thèse sont représentées dans le Tableau III-2; d'autres espèces sont également responsables de pathologies humaines, mais plus rarement. Ce sont *B. spielmanii* et d'autres espèces observées occasionnellement lors de diagnostic de patients (*B. bissetti* et *B. lusitaniae*) (J. D. Radolf *et al.*, 2012; G. Stanek *et al.*, 2011b)

L'infection due à la bactérie ***N. mikurensis*** a été principalement observée chez des patients immunodéprimés (A. Portillo *et al.*, 2018; V. Rar *et al.*, 2011).

L'infection due au parasite ***B. microti*** est observée principalement suite à une piqûre de la tique, cependant la transmission peut se produire également lors d'une transplantation

d'organes ou lors d'une transfusion sanguine (S. Madison-Antenucci *et al.*, 2020; L. Schnittger *et al.*, 2012).

Les infections dues au **virus de l'encéphalite à tique** sont asymptomatiques dans 70 à 95% de cas (S. Madison-Antenucci *et al.*, 2020).

Des cas de co-infection de tiques avec *Babesia spp.*, *Rickettsia spp.*, *A. phagocytophilum* et *Borrelia spp.* ont été observés dans différents pays européens (A. Portillo *et al.*, 2018). Ceci peut conduire à des co-infections chez les patients, phénomène qui reste cependant peu fréquent (S. Jahfari *et al.*, 2016).

Tableau III-2. Pathogènes et maladies associées, susceptibles d'être transmis par la tique *Ixodes* et étudiés au cours de ces travaux de thèse.

Agent infectieux	Maladies	Espèce pathogène	Manifestation Clinique	Diagnostic	Traitement	Source
Bactérie	Borréliose de Lyme	<i>Borrelia bavariensis</i> <i>Borrelia garinii</i> <i>Borrelia afzelii</i> <i>Borrelia valaisiana</i> <i>Borrelia burgdorferi</i> (297, N40, B31)	EM* Lymphocytome borrélien ACA* Neuroborréliose (paralyse faciale, méningite lymphocytaire) Arthralgies Monoarthrite Oligoarthrite (moins fréquente)	Sérologie PCR Culture	Antibiotiques (doxycycline, ceftriaxone, amoxicillin, azitromycine)	(Stanek <i>et al.</i> , 2012) (Jaulhac <i>et al.</i> , 2019) (Madison-Antenucci <i>et al.</i> , 2020)
	Fièvre récurrente	<i>Borrelia miyamotoi</i>	Fièvre Syndrome pseudo-grippal (myalgie, maux de tête, fatigue, frissons) Nausée	Microscopie à fond noir Sérologie PCR Culture	Antibiotique (pénicilline, doxycycline, érythromycine)	(Cutler <i>et al.</i> , 2017) (McCoy <i>et al.</i> , 2017) (Talagrand-Reboul <i>et al.</i> , 2018) (Madison-Antenucci <i>et al.</i> , 2020)
	Rickettsiose	<i>Rickettsia helvetica</i>	Escarre d'inoculation Fièvre Frissons Myalgies Céphalées	Sérologie PCR Culture	Antibiotiques (doxycycline, josamycine, clarithromycine, azithromycine)	(Parola <i>et al.</i> , 2013) (Fang <i>et al.</i> , 2017) (McCoy <i>et al.</i> , 2017)
	Anaplasmose humaine	<i>Anaplasma phagocytophilum</i>	Fièvre Frissons Myalgies Céphalées Plus rarement (signes digestifs, pneumopathie)	Frottis sanguin colorée (MGG*) PCR* Culture Sérologie Eruption cutanée Anomalies cliniques (CRP*, leucocytes, plaquettes, lymphocytes, entre autres)	Antibiotiques (doxycycline, gentamicine (1), ceftriaxone (2), cefixime) Combinaison de (1) et (2)	(Koebel <i>et al.</i> , 2012) (McCoy <i>et al.</i> , 2017) (Madison-Antenucci <i>et al.</i> , 2020)
	Neoehrlichiose	<i>Neoehrlichia mikurensis</i>	Fièvre prolongée Eruption cutanée (érysipèle) Picotements de membres Maux de tête Myalgies Arthralgie	PCR	Antibiotique (doxycycline) (1) Antibactérien (rifampine) (2) Combinaison de (1) et (2)	(Rar <i>et al.</i> , 2011) (Portillo <i>et al.</i> , 2018)
Virus	Encéphalite à tiques	TBEV	Première phase : syndrome pseudo-grippal Période asymptomatique Deuxième phase : méningite, encéphalite, méningo-encéphalite, myélite	Sérologie PCR	Traitement non-spécifique (adapté aux symptômes des patients) Prévention: vaccin anti TBEV (Encepur®, FSME-IMMUN®, TBE-Moscow, EnceVir)	(OMS, 2011) (Kaiser, 2012) (Jaulhac <i>et al.</i> , 2019) (Madison-Antenucci <i>et al.</i> , 2020)
Parasite	Babésiose	<i>Babesia divergens</i> <i>Babesia microti</i>	Fièvre Frissons Myalgies Cas aigus (détresse respiratoire, insuffisance rénale ou cardiaque)	Frottis sanguin colorée (Giemsa) PCR Sérologie (IFA)	Antiparasitaire (atovaquone) (1) Antibiotique (azithromycine) (2) Combinaison de (1) et (2) Antipaludique (quinine) (3) Antibiotique (clindamycine) (4) Combinaison de (3) et (4), formes graves	(Madison-Antenucci <i>et al.</i> , 2020)

II- Objectifs

Les tiques *Ixodes* peuvent être co-infectées donc vectrices de différents pathogènes susceptibles d'induire des maladies ayant une symptomatologie très similaire. Ainsi lors de co-infection, il est important d'identifier les différents pathogènes (W. Berghoff, 2012; E. Lejal *et al.*, 2019). Cette multiplicité d'agents infectieux nécessite d'avoir à disposition des outils diagnostiques afin d'établir le bon diagnostic et donc la mise en place d'un traitement adéquat. De plus, après piqûre de tique, les manifestations cliniques peuvent être assez polymorphes, on parle souvent d'un syndrome pseudo-grippal, ce qui peut compliquer le diagnostic (K. Klitgaard *et al.*, 2019). Quant aux diagnostics actuellement utilisés, ils reposent principalement sur la sérologie, qui peut présenter des réactions croisées (S. Madison-Antenucci *et al.*, 2020) et la PCR qui varie en terme de spécificité et sensibilité selon le pathogène transmis (B. Jaulhac *et al.*, 2019).

Plusieurs études ont été réalisées afin d'identifier ces pathogènes ou de mieux comprendre le phénomène de co-infection. Certaines ont été faites chez l'hôte vertébré infecté (W. Berghoff, 2012; M. A. Diuk-Wasser *et al.*, 2016; S. Jahfari *et al.*, 2016; A. Moniuszko *et al.*, 2014; Z. H. Svitalkova *et al.*, 2016), d'autres chez des tiques collectées en milieu naturel (A. Grochowska *et al.*, 2020; P. M. Jensen *et al.*, 2017; K. Klitgaard *et al.*, 2019; E. Lejal *et al.*, 2019; S. Moutailler *et al.*, 2016; A. Nebbak *et al.*, 2019). Il est donc nécessaire de mieux comprendre les différentes interactions lors de la transmission du pathogène à l'hôte vertébré (J. de la Fuente *et al.*, 2017a; M. Popara *et al.*, 2015), lors de l'interaction des tiques avec les pathogènes qu'elles sont susceptibles de transmettre, mais aussi de mesurer l'impact de l'environnement sur ces interactions (O. Hajdusek *et al.*, 2013; C. Raileanu *et al.*, 2017; A. Rizzoli *et al.*, 2014; Z. H. Svitalkova *et al.*, 2016).

Dans ce contexte, l'objectif de mes travaux de thèse a été de développer des méthodes analytiques par spectrométrie de masse afin :

- d'analyser des tiques collectées dans leur milieu naturel et d'identifier des marqueurs protéiques des pathogènes présents dans les tiques.
- d'identifier les pathogènes transmis à un hôte infecté (souris), via l'identification des protéines présentes dans la peau de cet hôte, afin d'avoir des informations sur la compétence vectorielle des tiques et la persistance cutanée.

- de sélectionner des marqueurs d'infection précoce dans la peau pour les autres pathogènes transmis par les tiques pour le développement d'une méthode de diagnostic par protéomique ciblée
- de mettre au point une méthode de diagnostic précoce et multiplexée par protéomique ciblant des pathogènes transmis par les tiques dans des biopsies cutanées (souris).

Cette étude vient compléter les analyses protéomiques d'identification des marqueurs d'infection active de *Borrelia* dans la peau de l'hôte infecté (A. Grillon *et al.*, 2017; G. Schnell *et al.*, 2015b) qui ont conduit à l'approche diagnostique de la borréliose de Lyme décrite dans le Chapitre II.

Comme pour la borréliose de Lyme ces travaux ont été mené en collaboration avec le groupe du Dr Nathalie Boulanger pour (i) l'étude des pathogènes circulant dans des tiques collectées dans la région Grand Est, et (ii) l'étude des peaux des souris infectées par piqûre de ces tiques afin d'analyser la compétence vectorielle. D'autres collaborations ont été mises en place pour l'étude des biopsies de souris infectées expérimentalement par une inoculation à la seringue d'*A. phagocytophilum* (Dr M. Kazimirova, Slovaquie) et de *B. microti* (Dr M. Jalovecka, République tchèque). Une partie de ces travaux a été réalisée au laboratoire en collaboration avec le Dr J. Saade (ATER ECPM 2019/2020).

III- Stratégie analytique d'identification des protéines

1. Nature des échantillons

Pour cette étude, nous nous sommes intéressés à différents types d'échantillons :

- des tiques collectées dans leur milieu naturel en Alsace : des tiques mâles collectées à Murbach ; des nymphes et des tiques femelles collectées à Niedermunster.
- des peaux de souris prélevées :
 - après avoir déposé 5 tiques femelles par souris et après avoir attendu 3 jours de repas sanguin,
 - après avoir déposé 10 nymphes par souris et après avoir attendu la complétion du repas sanguin (entre 4 à 6 jours),
 - à des jours différents (1, 3, 7, 14 et 21 jours), après inoculation intra-péritonéale avec des cellules infectées par *A. phagocytophilum*.

- à des jours différents (1, 3 et 5 jours) après inoculation à la seringue avec *B. microti*.

A ces échantillons se sont ajoutés des échantillons témoin :

- des tiques d'élevage (non infectées),
- des peaux de souris saines (non infectées),
- des peaux de souris saines piquées par des tiques d'élevage (non infectées).

2. Stratégie analytique mise en œuvre

Afin d'identifier le maximum de protéines présentes dans ces échantillons, nous avons mis en œuvre la stratégie analytique déjà utilisée avec succès pour identifier des protéines de *Borrelia* dans des peaux de souris (A. Grillon *et al.*, 2017; G. Schnell *et al.*, 2015b). Les différentes étapes de cette stratégie sont décrites dans la Figure III-3. Les protéines sont extraites, préfractionnées par SDS-PAGE, puis identifiées par nanoLC-MS/MS.

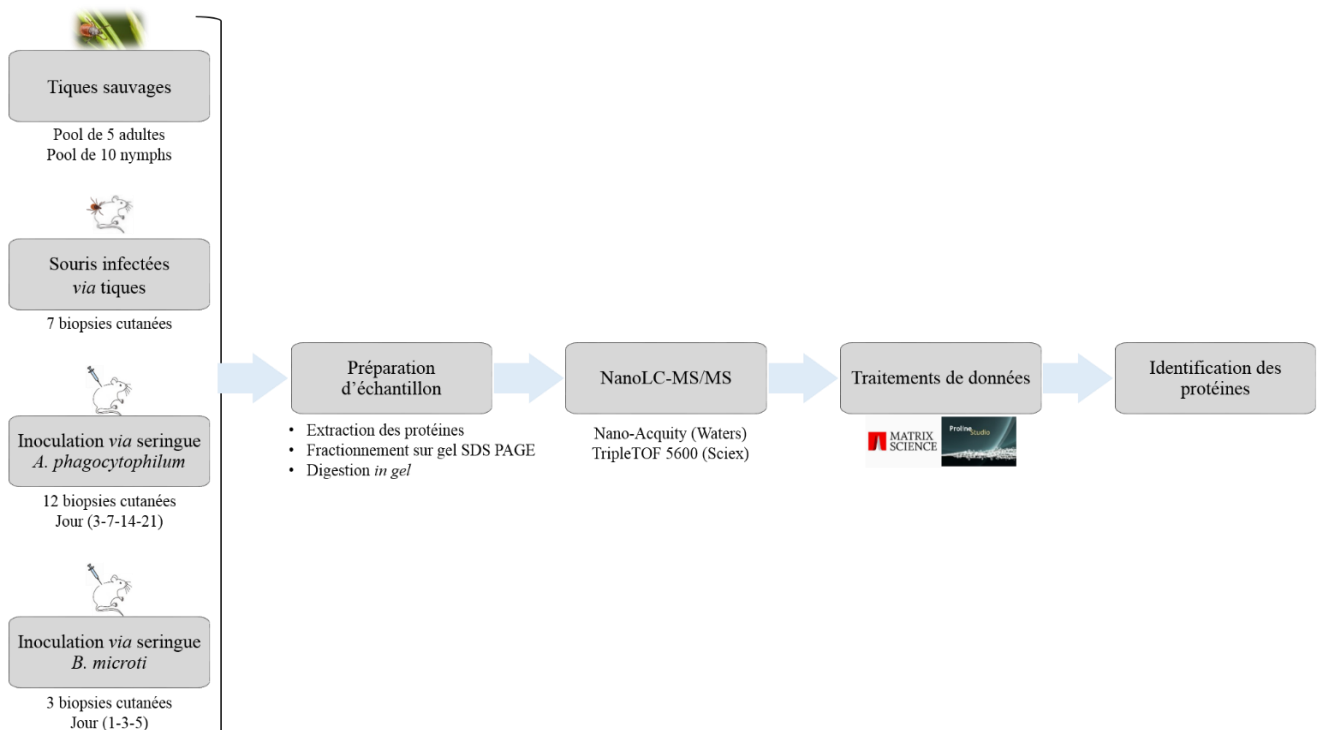


Figure III-3. Stratégie analytique d'identification des protéines dans les tiques et peaux de souris. Les protéines sont extraites et séparées sur gel SDS-PAGE. Les pistes de gel sont découpées en bandes de 2 mm. La digestion *in gel* est effectuée par la trypsine. Les peptides générés sont analysés par nanoLC-MS/MS. Les données générées sont traitées par l'algorithme de recherche Mascot et la validation des protéines est effectuée par le logiciel Proline.

Les défis analytiques étaient la complexité des échantillons, la variété des pathogènes potentiellement présents au sein des tiques puis transmis à l'hôte, et la présence minoritaire des protéines de pathogènes au milieu de toutes les protéines de tiques ou de souris. Comme nous le verrons plus loin, une étape cruciale a été la construction de banques de données adaptées à chaque type d'échantillon, pour optimiser l'identification des protéines de pathogènes.

2.1. Extraction et dosage des protéines

Les protéines ont été extraites des peaux de souris selon le protocole précédemment développé au laboratoire : un broyage manuel de 5 mg de biopsie cutanée dans 185 μ L de solution de Laemmli contenant 2% de SDS. Toutefois les travaux antérieurs menés au laboratoire avaient montré une variabilité importante de la présence des bactéries *Borrelia* autour du point d'inoculation ($CV \approx 50\%$) (G. Schnell *et al.*, 2015b). C'est pourquoi, pour pallier cette variabilité, nous avons voulu évaluer l'extraction de la totalité de la biopsie. Nous avons déterminé les rendements d'extraction pour deux quantités de biopsie (15 mg, gros potter *versus* 5 mg, petit potter) et pour différents volumes d'extraction (Figure III-4.A). Les concentrations en protéines ont été déterminées par un dosage DC (*Detergent compatible*, Biorad), compatible avec le SDS, en utilisant la BSA comme référence.

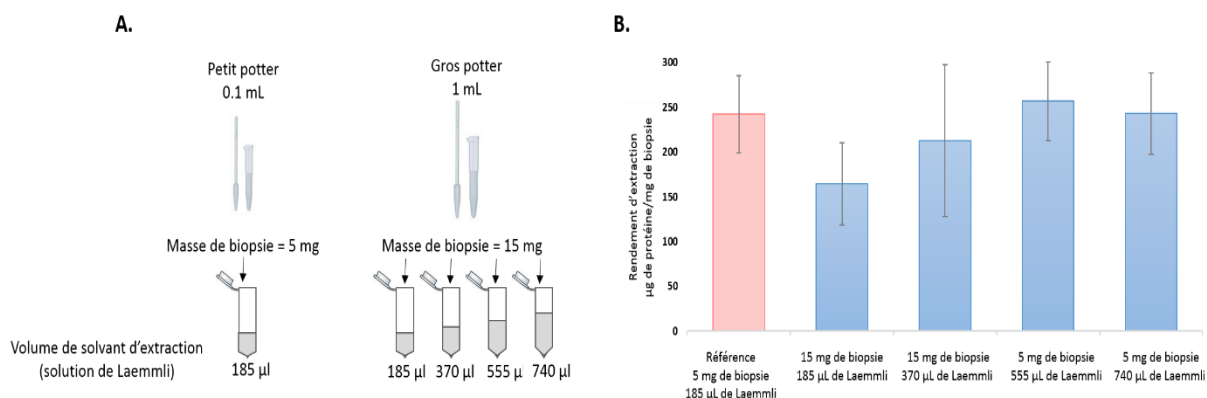


Figure III-4. Optimisation de l'extraction de protéines par broyage manuel. A. Etude de l'extraction des protéines de peau de souris réalisée en triplicat sur une biopsie de 5 mg broyée dans 185 μ L de Laemmli et sur des biopsies de 15 mg broyées dans des volumes différents de Laemmli (185 μ L, 370 μ L, 555 μ L, 740 μ L). B. Rendement d'extraction des protéines (μ g protéine/mg biopsie).

Pour la biopsie de 15 mg, quatre volumes de solution d'extraction ont été testés : 185 µL, 370 µL et 555 µL et 740 µL. Malgré une variation assez grande entre les réplicats, nous observons sur la Figure III-4.B que le rendement augmente avec le volume d'extraction excepté pour 740 µL où le rendement n'a pas été amélioré. Nous observons également qu'un volume de 555 µL permet de retrouver un rendement d'extraction similaire à celui obtenu avec 185 µL pour 5 mg de biopsie, soit 257 ± 44 µg de protéines extraites par mg de biopsie ($n=3$). Suite à ces résultats, il a été décidé que l'extraction des biopsies cutanées serait dorénavant effectuée à partir de morceaux de 15 mg avec un volume d'extraction de 555 µL, de manière à augmenter la surface prélevée autour du point d'inoculation des pathogènes, lorsque la quantité de biopsie le permettra.

En ce qui concerne les tiques, elles ont d'abord été lavées à l'éthanol 70% puis broyées dans un potter avec la solution de Laemmli. Pour le pool de tiques, l'extraction des protéines a été effectuée par broyage avec 185 µL de solution de Laemmli, plus adapté à la masse obtenue lors de la pesée des différents pools (pool de cinq nymphes : moyenne de 1,8 mg ; pool de cinq adultes : moyenne de 3,8 mg, soit une moyenne de 3 mg pour les pools de tiques).

Tous les échantillons ont été décomplexifiés à l'aide d'un préfractionnement sur gel SDS-PAGE (12% acrylamide). Les peptides issus de l'ensemble de ces échantillons ont été analysés par nanoLC-MS/MS sur un couplage nanoLC (NanoAcquity-Waters) - QTOF (TripleTOF 5600 – Sciex).

2.2. Traitements bio-informatiques

2.2.1. Création des banques de données

Comme expliqué dans le Chapitre I, la composition de la banque de données est essentielle pour l'identification des protéines par une approche *bottom-up*. Dans notre cas, où nous recherchons des protéines de divers pathogènes présents dans des peaux de souris ou des tiques, il était essentiel que les banques utilisées soient adaptées à cette multitude de microorganismes.

Ainsi pour ce projet, nous avons sélectionné le vecteur (tique *Ixodes ricinus*), l'hôte (souris *Mus musculus*) et différents pathogènes (C. Aubry *et al.*, 2016; B. Jaulhac *et al.*, 2019; P. M. Jensen *et al.*, 2017; E. Lejal *et al.*, 2019; A. Nebbak *et al.*, 2019; C. Raileanu *et al.*, 2017). Pour les pathogènes, nous avons sélectionné ceux avec un pouvoir pathogène reconnu chez l'humain. Pour la création de cette banque personnalisée, nous avons également effectué une

comparaison entre les protéines contenues dans les banques protéiques UniprotKB et NCBI. Ainsi, pour certains pathogènes, la banque finale contient des protéines soit uniquement d'UniprotKB, soit uniquement de NCBI, soit des deux, pour avoir le nombre d'entrées le plus complet possible. Le Tableau III-3 récapitule les organismes sélectionnés pour la création de la banque, et le nombre final de protéines pour chaque espèce au moment de sa création en janvier 2018.

Tableau III-3. Liste des pathogènes, hôte et vecteur associés avec leurs abréviations et nombre d'entrées dans la banque de données finale au moment de sa création (Janvier 2018). Le vecteur (*Ixodes ricinus*), l'hôte (*Mus musculus*) et les 14 pathogènes étudiés. La liste est composée principalement de bactéries (9 espèces différents dont 3 souches pour *B. burgdorferi* ss, 2 espèces de parasites (*B. microti* et *B. divergens*) et un virus (TBEV).

Vecteur\ Hôte\Pathogènes	Abrev.	Nombre de protéines
<i>Ixodes ricinus</i>	Ixori	25431
<i>Mus musculus</i>	Mouse	84937
<i>Anaplasma phagocytophilum</i>	ANAPH	1061
<i>Neoehrlichia mikurensis</i>	CNEOMI	12
<i>Rickettsia helvetica</i>	RICHE	1229
<i>Borrelia bavariensis</i>	BORBP	857
<i>Borrelia garinii</i>	BORGR	932
<i>Borrelia afzelii</i>	BORAP	1126
<i>Borrelia valaisiana</i>	BORVA	1571
<i>Borrelia burgdorferi</i> 297	Bbur-297	541
<i>Borrelia burgdorferi</i> N40	BORBN40	1233
<i>Borrelia burgdorferi</i> B31	BORBU-B31	1391
<i>Borrelia miyamotoi</i>	BORMIYA	776
<i>Babesia microti</i>	BABMI	3601
<i>Babesia divergens</i>	BABDI	55
<i>Tick-borne encephalitis virus</i>	TickBEnc	1

Comme nous pouvons le constater les différents pathogènes sont représentés par un nombre d'entrées protéiques très variable. Ceci reflète parfaitement la problématique des microorganismes peu séquencés, comme c'est le cas pour le virus TBE ou *Neoehrlichia mikurensis*, représentés dans notre banque finale par 1 et 12 protéines respectivement. Ce nombre faible de protéines peut impacter l'identification de ces pathogènes.

Au cours de ce projet, nous avons travaillé avec quatre banques différentes (Figure III-5) adaptées aux différents échantillons traités. Ces quatre banques ont été générées à partir de la même banque de départ.

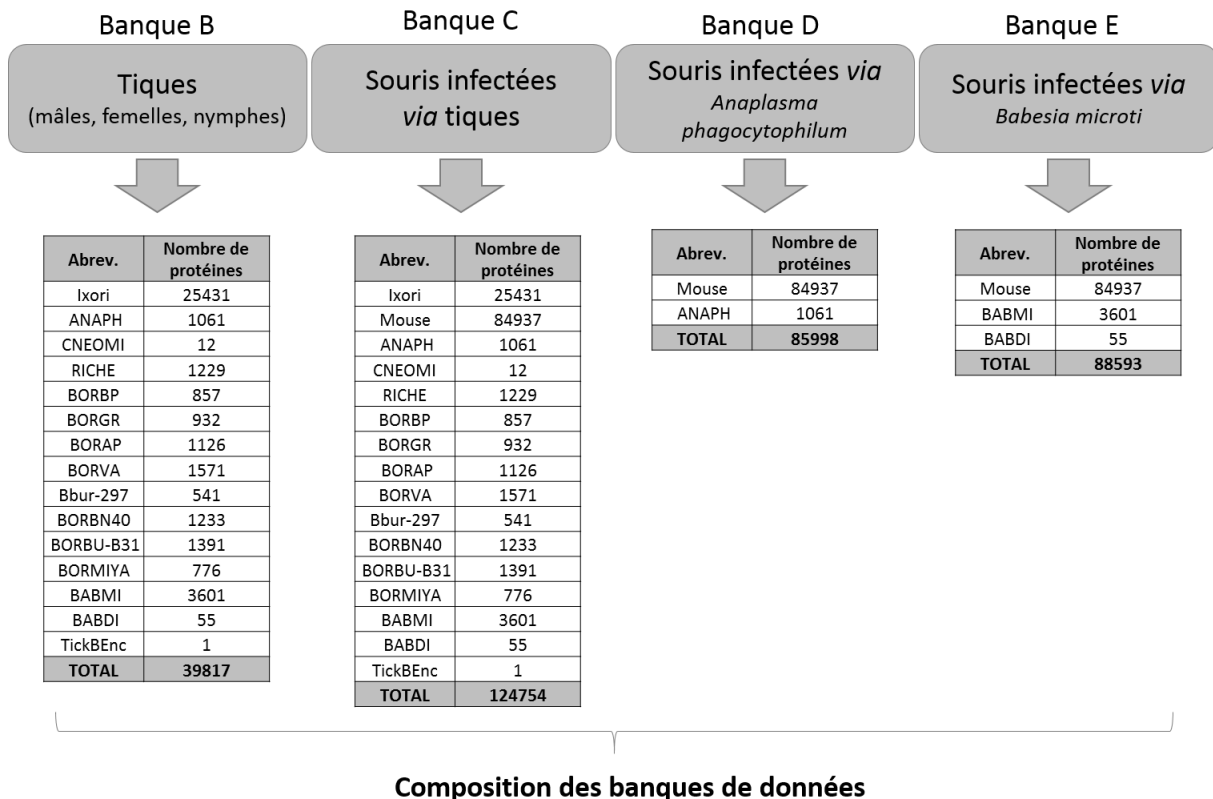


Figure III-5. Composition des banques de données utilisées pour les différents échantillons. Pour chaque organisme est indiqué le nombre de protéines dans la banque de données.

Afin de faciliter la différenciation entre les banques utilisées lors de mes travaux, nous avons identifié chaque banque par une lettre et un récapitulatif détaillé se trouve dans l'annexe 3

- **Banque B** pour l'analyse des **tiques** : constituée par *I. ricinus*, l'espèce de tique prédominante en Europe, et les 14 pathogènes susceptibles d'être présents dans les tiques (sélectionnés d'après la littérature et en accord avec le Dr N. Boulanger).
- **Banque C** pour l'analyse des **peaux de souris saines ou piquées par des tiques** : constituée par la souris *M. musculus*, la tique *I. ricinus*, et les 14 pathogènes.

- **Banque D** pour l'analyse des **peaux de souris infectées à la seringue par *A. phagocytophilum*** : constituées par la souris *M. musculus* et la bactérie *A. phagocytophilum*.
- **Banque E** pour l'analyse des **peaux de souris infectées à la seringue par *B. microti*** : constituée par la souris *M. musculus* et le parasite *B. microti*.

Afin d'améliorer les banques protéiques sélectionnées au départ, des modifications ont été apportées pour constituer une banque de données personnalisée plus adaptée, ce qui m'a permis de générer les banques finales présentées en Figure III-5. Ces modifications servent principalement : (i) à diminuer le nombre d'entrées en cas de redondances qui n'apportent pas d'information supplémentaire et qui sont gênantes lors de l'interprétation de données, (ii) à privilégier les séquences vérifiées et annotées pour chaque espèce étudiée afin d'augmenter la fiabilité des résultats et (iii) à diminuer les nombres d'entrées dans la banque finale pour limiter le nombre de faux positifs. Comme indiqué dans le Chapitre I, plus on augmente la taille de la banque, plus il y a des possibilités d'assigner des peptides à un spectre MS/MS : l'identification est moins spécifique et le taux de faux positifs augmente (S. Chatterjee *et al.*, 2016) (Chapitre I.I-3.1.3).

Ainsi nous avons privilégié des protéines annotées à celles prédites par des traitements informatiques, comme c'était le cas pour la protéine Rab family de l'espèce *B. microti* (Figure III-6).

RefSeq	XP_021338491.1	1	MRSLEKFWSKSKENPTVPASSTKDSATYQYKLVLLGETAVGKSLTTRFCKDVFLDYQDST	60	
UniProtKB/TrEMBL	A0A1R4ABC7	A0A1R4ABC7_BABMR	1	MRSLEKFWSKSKENPTVPASSTKDSATYQYKLVLLGETAVGKSLTTRFCKDVFLDYQDST	60

	XP_021338491.1	61	IGAAFMTKTINLESSVVKFEIWDTAGQERYRSLAPMYRGASAALIVDITSADTFEQAR	120	
	A0A1R4ABC7	A0A1R4ABC7_BABMR	61	IGAAFMTKTINLESSVVKFEIWDTAGQERYRSLAPMYRGASAALIVDITSADTFEQAR	120

	XP_021338491.1	121	SWINELKAVSRADVIIALAGNKVDLERNRSVDIETAQNFALNNCLFMETSAKTGHNVQE	180	
	A0A1R4ABC7	A0A1R4ABC7_BABMR	121	SWINELKAVSRADVIIALAGNKVDLERNRSVDIETAQNFALNNCLFMETSAKTGHNVQE	180

	XP_021338491.1	181	MFMKIANLLPRDNNPTLESVHFDQPTQAFSNCCNQ	215	
	A0A1R4ABC7	A0A1R4ABC7_BABMR	181	MFMKIANLLPRDNNPTLESVHFDQPTQAFSNCCNQ	215

Figure III-6 Alignement de séquences de la protéine Rab family de l'espèce pathogène *B. microti*. Les séquences alignées proviennent des banques protéiques RefSeq et UniProtKB/TrEMBL. Les séquences provenant de la banque protéique RefSeq ont été privilégiées.

Nous avons supprimé les séquences protéiques strictement identiques (protéines avec des noms de protéines similaires et numéros d'accension différents), ainsi que les sous-séquences ou séquences tronquées, en conservant seulement les séquences les plus complètes.

A titre d'exemple, nous avons représenté dans la Figure III-7, quatre protéines d' *A. phagocytophilum*, identifiées à partir du même peptide, toutes nommées *uncharacterized protein*, et présentant de grandes homologues de séquence. Seule la première protéine a été supprimée car sa séquence est strictement identique à la deuxième.

```

CLUSTAL O(1.2.4) multiple sequence alignment

TR|A0A0F3NDE5|A0A0F3NDE5_ANAPH MFALGCNYEGCIKVVYVPRVIASMLILHYVLSIKCRIVLNFLLEL 45
TR|A0A0F3PX34|A0A0F3PX34_ANAPH MFALGCNYEGCIKVVYVPRVIASMLILHYVLSIKCRIVLNFLLEL 45
TR|A0A0F3Q409|A0A0F3Q409_ANAPH MFALGCNYEGCIKVCYAPRVIASMLILHYVLSIKCRIVLNFLLEL 45
TR|A0A0F3NKH3|A0A0F3NKH3_ANAPH MFALGCNYEGCIKVCYAPRVIASMLILHYVLSIKCRIVLNFLLEL 45
*****

```

Figure III-7. Alignement de 4 protéines d'*A. phagocytophilum* identifiées par le peptide IVLNFLLEL encadré en bleu. La 1ère et la 2ème protéine ont des séquences identiques, mais des numéros d'accèsion différents (en vert). Les 3èmes et 4èmes protéines présentent des différences de séquences en orange. Seule la 1ère protéine a été supprimée.

Il est important de préciser que ces suppressions ont été effectuées dans les banques protéiques individuelles de chaque espèce avant d'effectuer la concaténation des banques pour générer la banque finale.

L'ordre suivi pour générer les banques de données finales est important, pour que les protéines homologues entre différentes espèces pathogènes soient conservées et pour ne pas perdre les informations des protéines partagées entre différentes espèces.

2.2.2. Algorithmes de recherche dans les banques de données

De nombreux algorithmes existent pour effectuer la comparaison entre les masses théoriques et les masses expérimentales (Chapitre I.I-3.1.2). Au début de ma thèse, deux moteurs de recherche étaient utilisés couramment au laboratoire pour identifier les protéines : Mascot (D. N. Perkins *et al.*, 1999) et OMSSA (L. Y. Geer *et al.*, 2004). Dans le cadre des travaux antérieurs sur la borréliose de Lyme, il avait été observé quelques différences entre les protéines identifiées par ces deux algorithmes (A. Grillon *et al.*, 2017; G. Schnell *et al.*, 2015a). Dans le chapitre II (approche vaccinale) nous avons également montré des différences entre ces deux algorithmes. Utiliser plusieurs algorithmes présente l'avantage de pouvoir augmenter le nombre de protéines identifiées, ou de pouvoir privilégier les protéines identifiées par les deux algorithmes, mais l'inconvénient réside dans le temps nécessaire au double traitement. Afin d'évaluer l'intérêt de ce double traitement pour ce projet, nous avons effectué une

comparaison des résultats d'identification des protéines d'*A. phagocytophilum* dans 12 biopsies de peaux de souris infectées (Figure III-8).

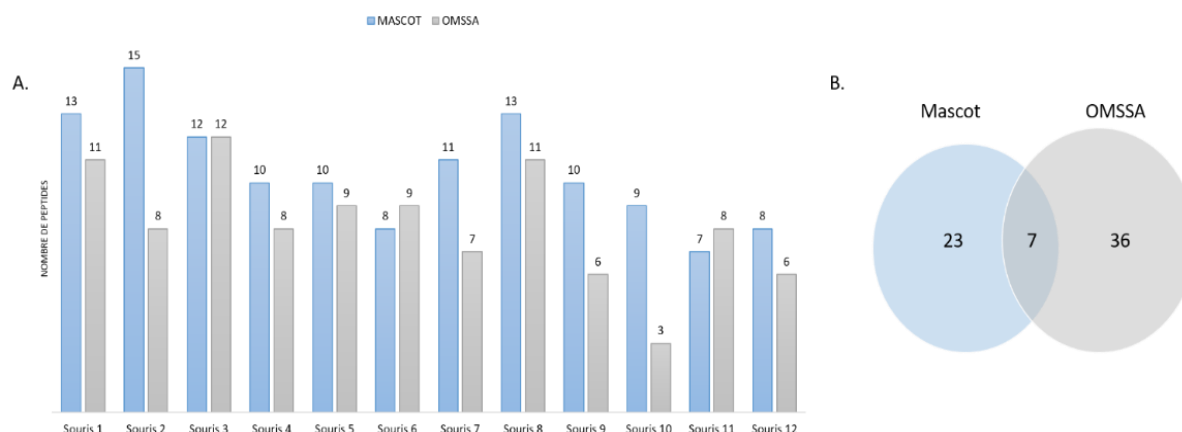


Figure III-8. Biopsies de peaux de souris infectées par inoculation avec *A. phagocytophilum*, puis analysées par LC-MS/MS. Données traitées soit avec Mascot, soit avec OMSSA avec la même banque de données. Comparaison des peptides d'*A. phagocytophilum* identifiés en utilisant les deux algorithmes de recherche (Mascot et OMSSA) **A.** Nombre de peptides d'*A. phagocytophilum* identifiés dans 12 biopsies. **B.** Nombre total de peptides différents d'*A. phagocytophilum* identifiés.

Cette étude a permis de détecter majoritairement des protéines de souris et un nombre plus faible de protéines de la bactérie avec un cœur en commun de plus de 89% de protéines totales identifiées. La Figure III-8, A montre la comparaison pour les peptides d'intérêt d'*A. phagocytophilum*. Dans la majorité des cas (9/12), plus de peptides ont été identifiés par Mascot. Dans deux biopsies, c'est l'inverse avec plus de peptides identifiés par OMSSA (souris 6 et 11). Pour une biopsie (souris 3), le nombre de peptides identifiés est identique avec les deux algorithmes de recherche. Sur la Figure 8,B, nous constatons que le recouvrement des peptides identifiés par les deux algorithmes est faible. Enfin, si on considère le nombre total de peptides identifiés, c'est OMSSA qui a permis d'identifier le plus grand nombre de peptides différents (43 *versus* 30).

Nous avons ensuite regardé le nombre de fois que les peptides ont été identifiés dans les 12 biopsies. Pour Mascot, 19/30 peptides (63%) ont été identifiés dans deux biopsies ou plus. Pour OMSSA, 17/43 peptides (39%) ont été identifiés dans deux biopsies ou plus. Cette plus grande occurrence des peptides identifiés avec Mascot peut expliquer le nombre plus faible de peptides différents identifiés avec Mascot (30 par Mascot *versus* 43 par OMSSA). En résumé, Mascot a conduit à l'identification de moins de peptides d'*A. phagocytophilum* mais de manière plus reproductible au sein des 12 biopsies.

Un autre élément pris en considération a été le temps nécessaire aux traitements. Les recherches effectuées avec OMSSA utilisent des ressources informatiques très importantes lors de l'étape de validation. A titre d'exemple, pour un même échantillon, le temps nécessaire à la validation des données est d'environ 15 min pour les données issues de la recherche Mascot, alors qu'il est de plus d'une heure avec les données issues de la recherche OMSSA.

Au vu de ces résultats, nous avons décidé de sélectionner Mascot comme moteur de recherche pour la suite de ces travaux.

2.2.3. Actualisation des banques de données

La première banque de données a été créée en 2016 (banque A) au début de ma thèse, pour identifier des protéines d'*A. phagocytophilum* dans des biopsies de souris infectées, puis j'ai été amenée à effectuer des actualisations au fur et à mesure de l'avancement du projet. Ces actualisations ne sont pas anodines, comme on peut le voir dans le Tableau III-4. Nous constatons une diminution très importante entre la banque d'*A. phagocytophilum* créée en 2016 (banque A) et celle de 2018 (banque D), due à l'évolution de la banque, à la sélection de protéines annotées, et principalement à l'adaptation nécessaire afin de créer une unique banque généralisée (banques B et C). Une diminution du nombre d'entrées des protéines d'*A. phagocytophilum* a permis de gagner de la place pour l'ajout d'autres pathogènes transmis par la tique.

Tableau III-4. Données des deux banques créées pour *A. phagocytophilum*. Première banque créée en 2016 à partir des entrées UniProtKB. Deuxième banque créée en 2018 à partir des entrées RefSeq, suite à la suppression de protéines redondantes et de protéines tronquées.

Banque	Année de création	Nombre de protéines	Banque de référence
A	2016	19223	UniProtKB
D	2018	1061	RefSeq

La finalité était de diminuer le nombre d'entrées d'*A. phagocytophilum*, critère indispensable pour la création de la banque contenant plusieurs pathogènes, afin de diminuer l'espace de recherche et par conséquent limiter le nombre de faux positifs.

Les données issues des analyses de biopsies d'*A. phagocytophilum* ont été traitées avec la banque A, puis ultérieurement avec la banque D. L'utilisation de cette nouvelle banque D allège de façon conséquente le traitement de données. Par exemple, le peptide AVATDLTK de la protéine P44 est présent 129 fois dans la banque A contre seulement 6 fois dans la banque D (Tableau III-5). Le nombre de fois où le peptide est présent dans la banque est important, car seuls les peptides uniques d'une protéine sont conservés pour l'identification.

Tableau III-5. Comparaison de la présence du peptide AVATDLTK dans les deux banques de données. L'extrait d'export de résultats après validation par Proline montre le nombre de fois où le peptide se trouve dans la banque de données.

Banque	Année de création	Peptide	N° d'accèsion	Description	Nombre de protéines dans la banque de données
A	2016	AVATDLTK	tr A0A098EGQ2 A0A098EGQ2_ANAPH	P44-80 outer membrane protein OS=Anaplasma phagocytophilum	129
D	2018	AVATDLTK	lc WP_020849070.1	ANAPH - WP_020849070.1 P44/Msp2 family outer membrane protein	6

Comme la banque A contenait un nombre très important d'entrées, se limiter à un seul peptide par protéine identifiée réduisait le nombre de protéines d'*Anaplasma* identifiées. Ces résultats m'ont conduite à vérifier manuellement les peptides présents plus d'une fois dans la banque de données.

A titre d'exemple, pour le peptide AVATDLTK présent dans 129 protéines (Tableau III-5), nous avons vérifié si les protéines validées correspondaient seulement à l'espèce d'intérêt ou si le peptide était partagé avec l'hôte. Suite à cette vérification, il s'est avéré que les 129 protéines appartenaient bien à *A. phagocytophilum*. Parmi ces 129 protéines, nous avons observé des protéines avec des N° d'accèsion différents mais avec des séquences protéiques strictement identiques, des protéines tronquées, ainsi que des protéines différentes contenant cette séquence commune. Dans la banque D, le peptide AVATDLTK était présent six fois dans six protéines différentes (Figure III-9) La vérification manuelle était plus facile à effectuer grâce au nombre de protéines plus faible dans la banque D.

Figure III-9. Protéines contenant le peptide AVATDLTK. Banque de donnée créée en 2018 (RefSec).

```

AVATDLTK

>|c|WP_020849070.1 ANAPH - WP_020849070.1 P44/Msp2 family outer membrane protein
[Anaplasma phagocytophilum]
MKISSSLKNTGKYAHYLESRAIPGAYTITLIKKRKTGIRDGSGSKEADTVYLLAKELAYDVVTGQTDNLAAALAKTSG
KDIVQFAKAVEISAPNIDKKVCNGTHAKGEDNNSPTAYSADMKTSATNKTAQCSGLNAPGGQKFSSFKVGLPDNKHWP
TGSYSKSGSTPTQNEQNSNAKAVATDLTKLTPEEKTIVAGLLAKTIEGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPI
ACVGLGGNFVGVVDGHITNHSISDPVCTSK

>|c|WP_020849075.1 ANAPH - WP_020849075.1 hypothetical protein [Anaplasma phagocytophilum]
MKISAPNIDKQVCTGSHAKGGDTTATTFSTNLASNTDNNTAQCSGLTKGENKFSTFAKEVVLQDNKHWP TGSYTGSSAA
KENEQNSNAKAVATDLTKLTSTDEKTIVAGLLAKTIEGGEVVEIRAVSSTSVMALELWVCWLL

>|c|WP_020849091.1 ANAPH - WP_020849091.1 hypothetical protein [Anaplasma phagocytophilum]
MEISNPNIDKKVCTGTHAKDARDSSGSPASYKEEPSGNNNTAQCSGFSKQEDHPFSEFARVLGLREGKNWPTGRYYES
GVKDGAPNSNAKAVATDLTKLTTEEKTIVAGLLAKTIEGGDN

>|c|WP_020849537.1 ANAPH - WP_020849537.1 P44/Msp2 family outer membrane protein
[Anaplasma phagocytophilum]
MRSRSKLFGLSVMMSLAIVMAGNEVRAHDDVSALETGGAGYFYVGLDYSPAFSKIRDFSIRESNGETKAVYPYLKDGKSV
KLESHKFDWNTPDPRI GFKDNMLVAMEGSGVYGIGGARVELEIGYERFKTKGIRDGSGSKEADTVYLLAKELAYDVVTG
QTDNLAAALAKTSGKDIVQFAKAVEISHSEIDKKVCNGGHSIAKAGSNTAGKYYAAEPVDGTNKETAQCSGLGESGEEKS
HKFLSKFVLDLTKIEGKNWPTGRAAQNDSSSGSVKIVEGPTNSNAKAVATDLTKLTAEKKTIVAGLLAKTIEGGEVVEIR
AVSSTSVMVNACYDLLSEGLGVVPIACVGLGGNFVGVVDGSRRTIRF

>|c|WP_020849666.1 ANAPH - WP_020849666.1 P44/Msp2 family outer membrane protein
[Anaplasma phagocytophilum]
MRSRSKLFGLSVMMSLAIVMAGNDVRAHDDVSALETGGAGYFYVGLDYSPAFSKIRDFSIRESNGETKAVYPYLKDGKSV
KLESHKFDWNTPDPRI GFKDNMLVAMEGSGVYGIGGARVELEIGYERFKTKGIRDGSGSKEADTVYLLAKELAYDVVTG
QTDNLAAALAKTSGKDIVQFANAVKISSPEIDGKVCNGDHAAIKATKGKAYVAELTSTYSNEETTQCSGLGNTSQATGPK
SLSGFVNTVKVGEKNWPRGRASDGSSKNIEGDPNSNAKAVATDLTKLTSTDEKTIVAGLLTKTIEGGEVVEIRAVSSTSV
MVNACYDLLSEGLGVVPIACVGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHQVVAAVSSRRY

>|c|WP_021799128.1 ANAPH - WP_021799128.1 P44/Msp2 family outer membrane protein
[Anaplasma phagocytophilum]
MMSMAIVMAGNDVRAHDDVSALETGGAGYFYVGLDYSPAFSKIRDFSIRESNGETKAVYPYLKDGKSVKLESNKFDWNTP
DPRIGFKDNMLVAMEGSGVYGIGGARVELEIGYERFKTKGIRDGSGSKEADTVYLLAKELAYDVVSGQTDNLAAALAKT
SGKDIVQFAKAVEISHPNIDSKVCSGDHAAITKAKGKKYVAELSTTHSNEETTQCSGLGNTTGATGPKTLSGFVNTVKVG
EGKNWPRGRASDGSSKNIEGDPNSNAKAVATDLTKLTSTDEKTIVAGLLTKTIEGGEVVEIRSVSSTSVMVNACYDLLSEG
LGVPVPIACVGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHVRVVDGVDLPEPMQ

```

L'exemple du peptide AVATDLTK illustre bien l'influence des redondances dans les banques de données, dues à des protéines tronquées ou des protéines avec des séquences strictement identiques, mais avec des noms et/ou numéros d'accès différents. Ces redondances résultent de l'annotation des protéines et des informations fournies par les auteurs lors de la soumission des séquences.

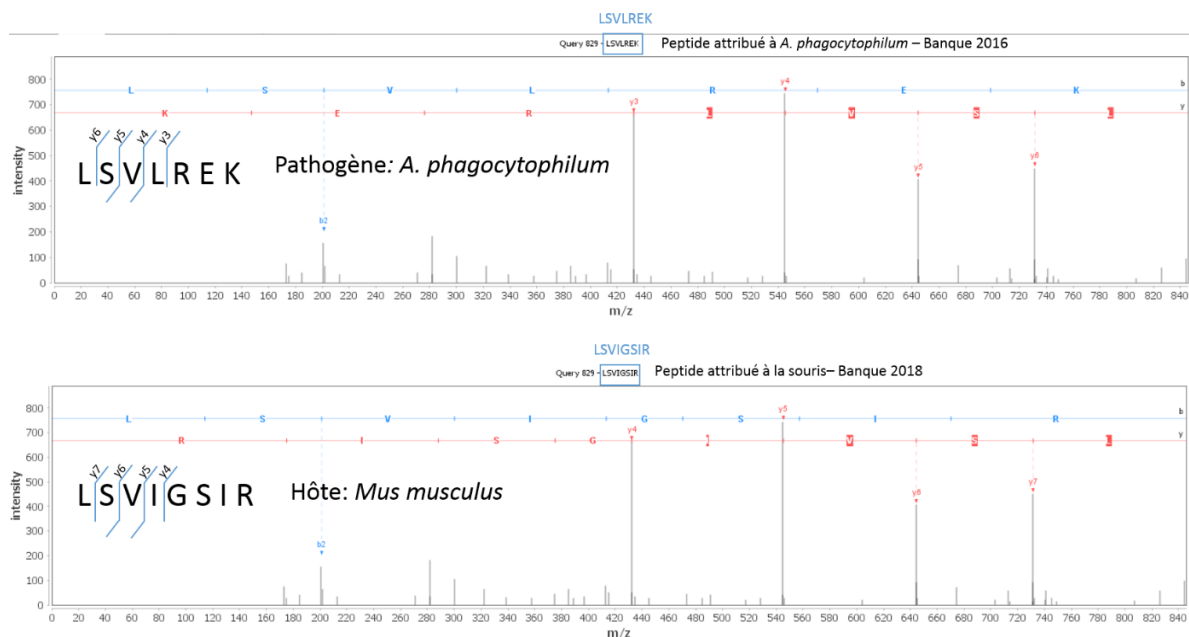
La banque D restreinte d'*A. phagocytophilum* présentait un intérêt pour la construction de la banque contenant plusieurs pathogènes. En effet, si l'espace de recherche est grand, et si les séquences de la banque sont similaires, il est difficile de séparer les assignations correctes des spectres à un peptide des assignations incorrectes (G. Gonnelli *et al.*, 2015). Or cela peut arriver pour des banques contenant des séquences d'organismes proches, tels que les différents pathogènes transmis par la tique.

L'évolution des banques de données peut également se traduire par la suppression ou l'ajout de protéines, ou par un changement de nom d'une protéine. Un tel exemple est montré dans le Tableau III-6

Tableau III-6. Exemple de mise à jour d'une protéine d'*A. phagocytophilum*, dans la banque RefSec (NCBI)

N° d'accèsion	Nom de la protéine	Date de téléchargement	Banque de référence
WP_044104104.1	hypothetical protein	Janvier 2018	RefSec
WP_044104104.1	peptidylprolyl isomerase	Juillet 2019	RefSec

L'actualisation de la banque D a également entraîné la perte de certains peptides identifiés dans la banque A. Cela a été observé pour le peptide LSVLREK attribué à *A. phagocytophilum* dans la banque A, mais plus dans la banque D, alors qu'une vérification manuelle dans la banque protéique a confirmé sa présence dans la seconde banque. En réalité, on peut observer que le même spectre a été attribué à deux séquences différentes, selon la banque dans laquelle la recherche est effectuée (Figure III-10).



Cette attribution est due à la similitude des deux séquences (Figure III-10), à la qualité du spectre et à la taille des deux banques. Ce type de spectre, malgré un bon score d'identification, doit être exclu pour éviter des identifications peu fiables. Durant ces travaux, j'ai donc pris soin de vérifier les spectres en cas d'identification douteuse de protéine de pathogènes.

IV- Résultats d'identification des protéines de pathogènes

1. Interprétation des données issues de la validation

Une fois les données de MS et MS/MS interprétées avec ces différentes banques de données, nous avons effectué d'ultimes vérifications : nous avons vérifié l'unicité des peptides dans la banque de données ainsi que les peptides partagés. Pour les peptides partagés, ceux appartenant à différentes protéines d'une même espèce pathogène, ainsi qu'à différentes espèces ont été conservés. Les peptides partagés entre plusieurs ordres ont été gardés (exemple, peptide de flagelline partagé entre plusieurs espèces de *Borrelia*, Figure III-11).

```
>|c| BORAP - WP_004790292.1 flagellin [Borrelia afzelii]
MIINHNTSAINASRNNGINAANLSKTQEKLSGGYRINRASDDAAGMGVSGKINAIQIRGLSQASRNTSKAINFIQTTEGNL
NEVEKVLVRMKELAVQSGNGTYSDDRGSIQIEIEQLTDEINRIADQAQYNQMHLMSNKSASQNVKTAEEELGMQPAKINT
PASLSGSQASWTLRVHVGANQDEIAIVNIYANVANLFSGEGAAQAAPVQEGAAQQAQPTAPATAPSGGGVNSPVNV
TTTVDANTSLAKIENAIMISDQRANLGAQNRLESIKNSTEYAIENLKASYAQIKDATMTDEVVAATTNSILTQSAMAM
IAQANQVPQYVLSLLR

>|g|1194982| BORBU-B31 - NP_212281.1 flagellin [Borrelia burgdorferi B31]
MIINHNTSAINASRNNGINAANLSKTQEKLSGGYRINRASDDAAGMGVSGKINAIQIRGLSQASRNTSKAINFIQTTEGNL
NEVEKVLVRMKELAVQSGNGTYSADRGSIQIEIEQLTDEINRIADQAQYNQMHLMSNKSASQNVRTAEELGMQPAKINT
PASLSGSQASWTLRVHVGANQDEIAIVNIYAANVANLFSGEGAAQAAPVQEGVQQAQQAQAPATAPSGGGVNSPVNV
TTTVDANTSLAKIENAIMISDQRANLGAQNRLESIKNSTEYAIENLKASYAQIKDATMTDEVVAATTNSILTQSAMAM
IAQANQVPQYVLSLLR

>|c| BORBN40 - WP_002556748.1 flagellin [Borrelia burgdorferi]
MIINHNTSAINASRNNGINAANLSKTQEKLSGGYRINRASDDAAGMGVSGKINAIQIRGLSQASRNTSKAINFIQTTEGNL
NEVEKVLVRMKELAVQSGNGTYSADRGSIQIEIEQLTDEINRIADQAQYNQMHLMSNKSASQNVRTAEELGMQPAKINT
PASLSGSQASWTLRVHVGANQDEIAIVNIYAANVANLFSGEGAAQAAPVQEGVQQAQQAQAPATAPSGGGVNSPVNV
TTTVDANTSLAKIENAIMISDQRANLGAQNRLESIKNSTEYAIENLKASYAQIKDATMTDEVVAATTNSILTQSAMAM
IAQANQVPQYVLSLLR

>|c| BORBP - WP_011193498.1 flagellin [Borrelia baviensis]
MIINHNTSAINASRNNGINAANLSKTQEKLSGGYRINRASDDAAGMGVSGKINAIQIRGLSQASRNTSKAINFIQTTEGNL
NEVEKVLVRMKELAVQSGNGTYSADRGSIQIEIEQLTDEINRIADQAQYNQMHLMSNKSASQNVRTAEELGMQPAKINT
PASLSGSQASWTLRVHVGANQDEIAIVNIYAANVANLFSGEGAAQAAPVQEGVQQAQQAQAPATAPSGGGVNSPVNV
TTTVDANTSLAKIENAIMISDQRANLGAQNRLESIKNSTEYAIENLKASYAQIKDATMTDEVVAATTNSILTQSAMAM
IAQANQVPQYVLSLLR

>|c| BORVA - WP_006068485.1 flagellin [Borrelia valaisiana]
MIINHNTSAINASRNNGINAANLSKTQEKLSGGYRINRASDDAAGMGVSGKINAIQIRGLSQASRNTSKAINFIQTTEGNL
NEVEKVLVRMKELAVQSGNGTYSDDRGSIQIEIEQLTDEINRIADQAQYNQMHLMSNKSAAQNVKTAEEELGMQPAKINT
PASLSGSQASWTLRVHVGANQDEIAIVNIYAANVANLFSGEGAAQAAPVQEGVQQAQQAQAPATAPSGGGVNSPVNV
TTTVDANTSLAKIENAIMISDQRANLGAQNRLESIKNSTEYAIENLKASYAQIKDATMTDEVVAATTNSILTQSAMAM
IAQANQVPQYVLSLLR
```

Figure III-11. Représentation de la séquence protéique de la protéine flagelline pour cinq espèces différentes de *Borrelia* responsables de la borréliose de Lyme. Pour ces cinq protéines la séquence du peptide partagé est représenté en bleu **ASDDAAGMGVSGK**

Par contre, nous avons exclu les peptides partagés avec l'hôte (souris) ou le vecteur (tique), en considérant qu'ils avaient plus de chance d'appartenir à ces deux organismes compte tenu de la prépondérance des protéines de souris et de tique devant les protéines de pathogènes.

Nous avons vérifié pour les peptides attribués à un pathogène, qu'avec une permutation Ile/Leu dans leur séquence, ces peptides ne pouvaient appartenir ni à la souris, ni à la tique.

Enfin, nous avons tenu compte des peptides identifiés dans les 3 échantillons contrôle, à savoir des peaux de souris saines (négatives), des tiques d'élevage (non infectées) et des souris saines piquées par des tiques d'élevage non infectées. Ces peptides ont été considérés comme non spécifiques des pathogènes (puisque présents dans des échantillons non infectés) et ils ont été supprimés de la liste finale des peptides d'intérêt.

La Figure III-12 montre l'exemple d'un échantillon de tiques (pool de 5 tiques femelles) où un total de 3120 protéines a été identifié. Parmi ces protéines, 61 ont été dans un premier temps attribuées à l'un des 14 pathogènes présents dans la banque B, parmi lesquelles 41 protéines sont restées après la vérification Ile/Leu.

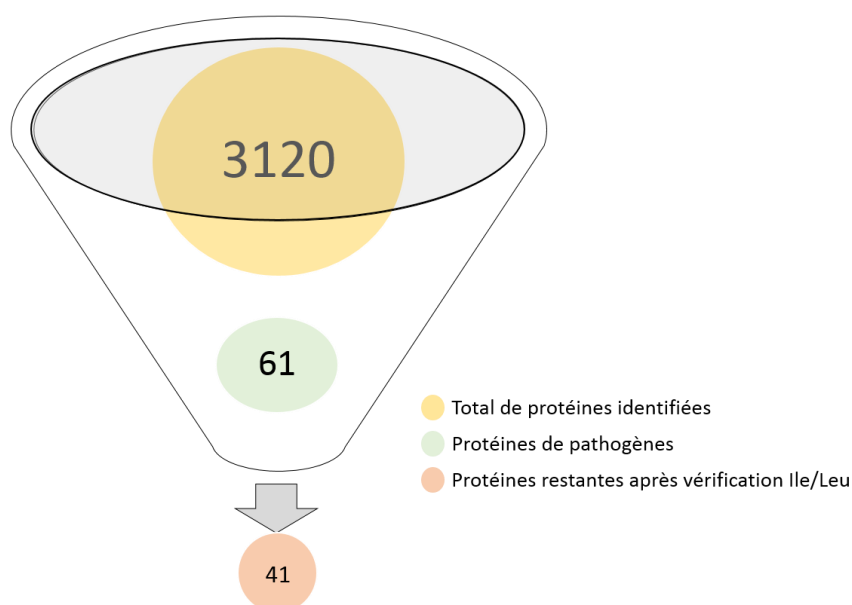


Figure III-12. Protéines identifiées dans un pool de 5 tiques *I. ricinus* femelles sauvages. Les différentes vérifications post-traitement des données MS et MS/MS conduisent à une réduction du nombre de protéines de pathogènes identifiées à 41 (avant comparaison aux échantillons contrôle).

Comme indiqué précédemment, nous avons analysé trois types d'échantillons contrôle auxquels les différentes biopsies ont été comparées pour écarter les peptides partagés avec l'hôte ou le vecteur. Pour cet exemple, 27 protéines sur les 41 ont été conservées car pouvant être attribuées à un pathogène avec plus de confiance suite à la suppression des peptides partagés soit avec l'hôte, soit avec le vecteur.

Nous pouvons voir dans cet exemple que **les protéines de pathogènes représentent moins de 1% des protéines totales identifiées**. Cela confirme que les protéines de la tique sont très largement majoritaires devant les protéines du pathogène.

Ces 27 protéines correspondent à 76 peptides, et parmi ces 76 peptides, 15 (20 %) sont des peptides partagés entre deux ou plus espèces pathogènes (Figure III-13). Dans le cadre de la sélection des protéines pour une méthode de diagnostic pathogène-spécifique, seules les protéines identifiées avec au moins un peptide spécifique pourront être conservées.

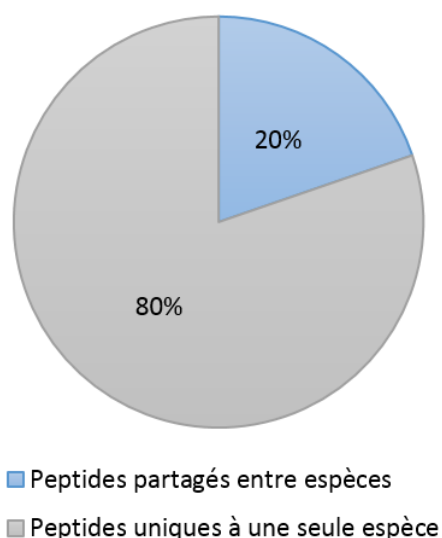


Figure III-13. Répartition de l'unicité des 76 peptides de pathogènes identifiés dans les tiques femelles. 15/76 peptides sont partagés entre différents pathogènes. 61/76 peptides sont uniques à un seul pathogène.

Enfin, si on considère le nombre de peptides identifiés par protéine, on observe que parmi les 27 protéines, 18 protéines (67%) ont été identifiées avec un seul peptide et les neuf restant avec deux peptides et plus (Figure III-14). Ces résultats sont en accord avec ceux précédemment obtenus pour des peaux de souris infectées par *Borrelia* (A. Grillon *et al.*, 2017; G. Schnell *et al.*, 2015b), et sont dus à la faible abondance des protéines de pathogènes parmi les protéines majoritaires de souris ou de tique. Sans surprise, les protéines de tiques, qui sont

majoritaires dans l'échantillon, ont été identifiées avec plus d'un peptide par protéine (1 à 394 peptides par protéine, moyenne de 22).

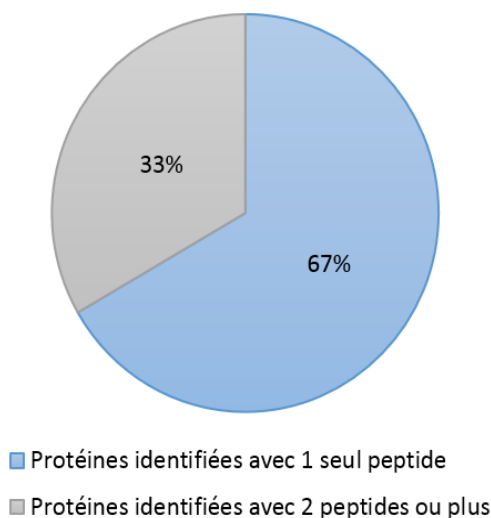


Figure III-14. Représentation des protéines identifiées avec des peptides uniques à l'espèce identifiée dans la banque de données théorique utilisée. Parmi les 27 protéines, 18 protéines ont été identifiées avec 1 seul peptide et les 9 protéines restantes avec 2 peptides ou plus.

Dans le cadre de mes travaux, nous avons conservé les protéines de pathogènes identifiées avec un seul peptide, même si les protéines identifiées avec au moins 2 peptides le sont avec plus de certitude, en gardant à l'esprit qu'une validation biologique ou une étude bibliographique approfondie des protéines retenues est nécessaire.

2. Analyse des tiques *I. ricinus*

Cette analyse avait pour objectif d'identifier les pathogènes circulant au sein des tiques *I. ricinus* collectées sur la végétation dans la région Grand Est. Nous avons ainsi analysé des tiques mâles, des tiques femelles et des nymphes.

Pour les tiques adultes (mâles et femelles), nous avons analysé des pools de 5 individus. Pour les nymphes, nous avons analysé des pools de 10 tiques. Les résultats ont été divisés en deux groupes : peptides uniques et peptides partagés. En moyenne 16300 peptides de tiques, correspondant à environ 2250 protéines, ont été identifiés dans ces échantillons. La totalité des peptides attribués aux pathogènes sont représentés dans le Tableau III-7

Tableau III-7. Protéines et peptides identifiés dans l'ensemble des échantillons: contrôles, tiques collectées dans leur milieu naturel, peaux piquées par des tiques sauvages et peaux infectées par inoculation à la seringue d'*A. phagocytophilum* et *B. microti*. En rouge sont indiqués les peptides partagés entre les espèces ou souches présentes dans la banque de données. Les peptides grisés indiquent les peptides identifiés dans au moins un échantillon contrôle. Les peptides en noir (sans aucun marquage) représentent les peptides uniques à une seule espèce pathogène dans la banque de données et qui n'ont pas été identifiés dans les échantillons contrôle. Les croix indiquent les peptides identifiés par biopsie. Le nombre de peptides identifiés dans les différentes types d'échantillon est indiqué dans la dernière ligne de chaque page, la somme de ces valeurs donne le nombre total des peptides identifiés par biopsie.

			Contrôle (Tiques neg, souris neg et peaux de souris piquées par de tiques neg)				Tiques collectées sur la végétation							Peaux infectées via seringue par <i>A. phagocytophilum</i>							Peaux infectées via seringue par <i>B. microti</i>			Peaux piquées par des tiques sauvages												
Peptides	N° d'accession	Description	Tique neg 1	Tique neg 2	(Souris neg + tique neg) 1	(Souris neg + tique neg) 2	Souris neg 1	Souris neg 2	Femelle	Mâle 1	Mâle 2	Mâle 3	Mâle 4	Nymphe 1	Nymphe 2	Nymphe 3	Souris 1_(J3)	Souris 2_(J3)	Souris 3_(J3)	Souris 4_(J3)	Souris 5_(J7)	Souris 6_(J7)	Souris 7_(J7)	Souris 9_(J14)	Souris 1_(J21)	Souris 1_(J1)	Souris 2_(J3)	Souris 3_(J5)	Souris 1	Souris 2	Souris 3	Souris 4	Souris 5	Souris 6	Souris 7	
CHTIMNCTKCPK	lc WP_011450309.1	ANAPH - WP_011450309.1 succinate dehydrogenase iron-sulfur subunit [Anaplasma phagocytophilum]																							X											
IDEKDLR	lc WP_011450546.1	ANAPH - WP_011450546.1 peptide chain release factor 1 [Anaplasma phagocytophilum]																				X														
IIVGELR	lc WP_011451401.1	ANAPH - WP_011451401.1 type VI secretion protein [Anaplasma phagocytophilum]															X	X				X		X												
IDTELLK	lc WP_021798967.1	ANAPH - WP_021798967.1 DNA polymerase I [Anaplasma phagocytophilum]																X	X					X												
NELAELEK	lc WP_044104104.1	ANAPH - WP_044104104.1 hypothetical protein [Anaplasma phagocytophilum]															X																			
LEDLGYK	lc WP_011450753.1	ANAPH - WP_011450753.1 hypothetical protein [Anaplasma phagocytophilum]								X																										
ILSTLAR	lc WP_020848748.1	ANAPH - WP_020848748.1 hypothetical protein [Anaplasma phagocytophilum]							X																											
ALFSDALK	lc WP_020849328.1	ANAPH - WP_020849328.1 hypothetical protein [Anaplasma phagocytophilum]							X						X	X																				
ISAFNAR	lc WP_020849594.1	ANAPH - WP_020849594.1 hypothetical protein [Anaplasma phagocytophilum]													X																					
YVVNDELAVK	lc WP_011450380.1											X																								
LSGGVAVLK	lc WP_011450380.1	ANAPH - WP_011450380.1 molecular chaperone GroEL [Anaplasma phagocytophilum]										X																								
VGSSEVEVK	lc WP_011450380.1											X																								
ALSDAGIKNSK	lc WP_044104082.1	ANAPH - WP_044104082.1 molecular chaperone DnaK [Anaplasma phagocytophilum]		X				X							X				X											X		X				
TTPSVVAFTDNER	lc WP_044104082.1															X																				
NTMMIAQK	lc WP_011450708.1	ANAPH - WP_011450708.1 type I DNA topoisomerase [Anaplasma phagocytophilum]										X																								
AVATDLTK	lc WP_020849070.1	ANAPH - WP_020849070.1 P44/Msp2 family outer membrane protein [Anaplasma phagocytophilum]				X											X	X	X	X	X		X	X												
ELAYDVVTGQTDK	lc WP_020849489.1											X																								
TIVAGLLAK	lc WP_020849489.1	ANAPH - WP_020849489.1 P44/Msp2 family outer membrane protein [Anaplasma phagocytophilum]										X																								
VVGDGVYDDLPAQR	lc WP_020849489.1											X																								
SGETEDTFIADLVGLR	tr A0A0D6DQG0	BABDI - A0A0D6DQG0 BABDI Enolase OS=Babesia divergens GN=enol PE=2 SV=1						X																												
IDNGIAR	tr P90652	BABDI - P90652 BABDI Bdi60B OS=Babesia divergens PE=4 SV=1					X		X	X	X	X	X		X																					
FWEVISDEHGIDPSPGTYHGDSDLQLER	tr Q9NFI3	BABDI - Q9NFI3 BABDI Beta tubulin OS=Babesia divergens GN=beta tubulin PE=4 SV=1					X	X		X	X		X	X	X																					
AIEFAIK	tr Q0VY85	BABDI - Q0VY85 BABDI Heat shock protein 70 OS=Babesia divergens GN=hsp70 PE=3 SV=1																											X							
LNQEFIK	gi 24425719	BABMI - XP_021337634.1 Cullin family [Babesia microti strain RJ]																																	X	
LVVVTAGAR	gi 24423087	BABMI - XP_012647088.1 LGT, LDH, L-lactate dehydrogenase [Babesia microti strain RJ]																									X									X
FIIPNVVK	gi 24423087						X	X	X						X																					
VIGSGCNLDSAR	gi 24423087																																			
ILNSACCTTNCLAPLAK	gi 24423126	BABMI - XP_012647126.1 glyceraldehyde 3-phosphate dehydrogenase, GAPDH [Babesia microti strain RJ]	X	X	X	X																					X	X	X	X	X	X	X	X	X	X
DVNAAVATIK	gi 24426516					X	X	X	X																											
FEGALNVDITEFQTNLVPYPR	gi 24426516	BABMI - XP_012650470.1 tubulin alpha [Babesia microti strain RJ]				X	X	X	X						X	X	X																		X	
YMACCLMYR	gi 24426516								X	X	X	X	X	X	X	X																				
GHYTEGAELIDSVLDVVR	gi 24423287	BABMI - XP_012647284.1 tubulin beta [Babesia microti strain RJ]	X	X	X	X	X	X	X			X	X	X	X	X											X	X	X	X	X	X	X	X	X	
GHYTEGAELIDSVLDVVRK	gi 24423287						X																													
SIFVDLEPTVVDVVR	gi 24423972	BABMI - XP_012647957.1 Tubulin C-terminal domain [Babesia microti strain RJ]							X																											
FDENFTIK	gi 24423476	BABMI - XP_012647471.1 Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD [Babesia microti strain RJ]									X																									
SVGPIMK	gi 24423946	BABMI - XP_012647931.1 prefoldin beta subunit [Babesia microti strain RJ]						X																												
EIAILNR	gi 24424018	BABMI - XP_012648003.1 mitogen-activated protein kinase [Babesia microti strain RJ]												X																						
VALTGLTVAEYFR	gi 24424036	BABMI - XP_012648021.1 F-type H+-transporting ATPase subunit beta [Babesia microti strain RJ]					X	X	X	X		X		X	X	X																				
NIYPPINVLPSLR	gi 24424261	BABMI - XP_012648242.1 V-type H+-transporting ATPase subunit B [Babesia microti strain RJ]							X					X	X	X																				
IVVGLDNAGK	gi 24424416	BABMI - XP_012648396.1 ADP-ribosylation factor family [Babesia microti strain RJ]				X																														
YSVWVGGSILSSLSTFQQMWITK	gi 24425120																										X	X	X							
VNPEEHSVLLTEAPLNPK	gi 24425120	BABMI - XP_012649089.1 Actin [Babesia microti strain RJ]											X		X												X	X	X							
SOIADLR	gi 24424574	BABMI - XP_012648551.1 DDX5, DBP2, ATP-dependent RNA helicase DDX5/DBP2 [Babesia microti strain RJ]									X	X	X	X																						
LALSIAK	gi 24424720	BABMI - XP_012648696.1 UTP24, FCF1, U3 small nucleolar RNA-associated protein 24 [Babesia microti strain RJ]																																		
TFTDCFNCLPVAIIIDDK	gi 24425094	BABMI - XP_012649062.1 protein phosphatase 1, catalytic subunit [Babesia microti strain RJ]					X		X	X	X	X	X																							

[illegible]

				Contrôle (Tiques neg, souris neg et peaux de souris piquées par de tiques neg)		Tiques collectées sur la végétation							Peaux infectées via seringue par <i>A. phagocytophilum</i>							Peaux infectées via seringue par <i>B. microti</i>			Peaux piquées par des tiques sauvages													
Peptides	N° d'accession	Description	Tique neg 1	Tique neg 2	(Souris neg + tique neg) 1 (Souris neg + tique neg) 2	Souris neg 1	Souris neg 2	Femelle	Mâle 1	Mâle 2	Mâle 3	Mâle 4	Nymph 1	Nymph 2	Nymph 3	Souris 1_(J3)	Souris 2_(J3)	Souris 3_(J3)	Souris 4_(J3)	Souris 5_(J7)	Souris 6_(J7)	Souris 7_(J7)	Souris 9_(J14)	Souris 1_(J21)	Souris 1_(J1)	Souris 2_(J3)	Souris 3_(J5)	Souris 1	Souris 2	Souris 3	Souris 4	Souris 5	Souris 6	Souris 7		
AIELEK	lc WP_006068606.1	BORVA - WP_006068606.1 asparagine-tRNA ligase [Borrelia valaisiana]					X				X		X		X																					
QDLDEKLK	lc WP_040351410.1	BORVA - WP_040351410.1 Bdr family protein [Borrelia valaisiana]																															X			
STLIVEGK	gi 1195496	BORBU-B31 - NP_212778.1 N-acetylmannosamine-6-phosphate 2-epimerase [Borrelia burgdorferi B31]	X																																	
GTVGLIGTK	lc WP_010420474.1	RICHE - WP_010420474.1 porin [Rickettsia helvetica]						X																												
ITGSQVTAGTGGWCR	lc WP_010420474.1							X																												
TINDVIAGAK	lc WP_010420474.1							X																												
VELGSPADVS AK	lc WP_010420474.1							X																												
VVEVLNTYK	lc WP_010420474.1							X																												
LLKAGMK	lc WP_010420656.1	RICHE - WP_010420656.1 DUF2748 domain-containing protein [Rickettsia helvetica]			X	X																								X		X			X	
ADLTLGTIIGK	lc WP_010421007.1	RICHE - WP_010421007.1 porin family protein [Rickettsia helvetica]						X					X	X	X																					
AELAYSWIDDGR	lc WP_010421007.1							X						X																						
IDAGAAMFNK	lc WP_010421007.1							X						X	X	X																				
ITYNGITGLSSNTK	lc WP_010421007.1							X						X	X	X																				
LTLGTSAQIADGVK	lc WP_010421007.1							X						X	X	X																				
NIYQGTSTVPTGGMR	lc WP_010421007.1							X							X																					
SNMTVPVDLIGIYYISK	lc WP_010421007.1														X																					
TATNAPFTGTNISASHKPTITR	lc WP_010421007.1								X					X	X																					
YQSHNLTVGIR	lc WP_010421007.1								X					X	X	X																				
AELVYSWTDYGK	lc WP_010421009.1															X																				
LNAGGVIFDK	lc WP_010421009.1								X					X	X																					
TVANNNVISSNFSGTTK	lc WP_010421009.1								X					X																						
VFAGAGVGAA LVK	lc WP_010421009.1															X																				
LTVARAR	lc WP_010421107.1	RICHE - WP_010421107.1 F0F1 ATP synthase subunit beta [Rickettsia helvetica]						X																												
IFASIDK	lc WP_010421116.1	RICHE - WP_010421116.1 NADH-quinone oxidoreductase subunit G [Rickettsia helvetica]						X																												
DIQAQLGNR	lc WP_010421481.1	RICHE - WP_010421481.1 autotransporter outer membrane beta-barrel domain-containing protein [Rickettsia helvetica]						X					X		X																					
FTGPAGSLEK	lc WP_010421481.1							X																												
FVNYGLIR	lc WP_010421481.1							X						X		X																				
NSADSANFVGTTTDDTSAAVTNAQLDI AK	lc WP_010421481.1							X								X																				
SSDENYKETGTTVANK	lc WP_010421481.1							X								X																				
STVGNNGPEGDFNVPR	lc WP_010421481.1							X						X		X																				
TDLIVGAK	lc WP_010421481.1							X						X	X	X																				
TSYNLGLSASIRPDAK	lc WP_010421481.1													X		X																				
YYTLIEGGAR	lc WP_010421481.1								X																											
ALALEALR	lc WP_010421508.1	RICHE - WP_010421508.1 toxin-antitoxin system HicB family antitoxin [Rickettsia helvetica]								X																										
VGEEIEIVGLK	lc WP_010421650.1	RICHE - WP_010421650.1 elongation factor Tu [Rickettsia helvetica]						X																												
IEIPSDIK	lc WP_010421760.1	RICHE - WP_010421760.1 50S ribosomal protein L6 [Rickettsia helvetica]							X																											
AAEVAGDGTITATVLAR	lc WP_010421840.1	RICHE - WP_010421840.1 chaperonin GroEL [Rickettsia helvetica]						X																												
AAVEEGVAVGGGVTLTHASQTLTK	lc WP_010421840.1							X						X	X	X																				
AGIIDPAK	lc WP_010421840.1							X						X	X	X																				
DQQAGIEIVIEALKDPLK	lc WP_010421840.1							X						X	X	X																				
EGVITVEEAK	lc WP_010421840.1							X								X																				
ENTVIVDGS GDKK	lc WP_010421840.1							X						X	X	X																				
GELITEDLGMK	lc WP_010421840.1							X						X		X																				
GYLSPYFVTNSEK	lc WP_010421840.1							X						X		X																				
LENVSIK	lc WP_010421840.1							X							X	X																				
LVAAGYNPMDLK	lc WP_010421840.1							X						X	X	X																				
MVVELENPFILLFEK	lc WP_010421840.1							X						X	X	X																				
NAGAQLLK	lc WP_010421840.1							X						X	X	X																				
NVLIEQSFSGSPK	lc WP_010421840.1							X								X																				
SQIAETTS DYDKEK	lc WP_010421840.1							X						X	X	X																				

L'ensemble de ces résultats résultent des traitements explicités dans la Tableau III-7 Suite à ces traitements, les peptides identifiées dans les 3 échantillons contrôle, à savoir des peaux de souris saines (négatives), des tiques d'élevage (non infectées) et des souris saines piquées par des tiques d'élevage non infectées, ont été écartées.

Pour ce qui concerne l'analyse des tiques, 61/88 (69%) des protéines de pathogènes identifiées dans les tiques sauvages ont été conservées, ce qui correspond aux lignes non grisées dans le Tableau III-7

La Figure III-15 montre la variété des pathogènes identifiés **à partir de peptides uniques** (un total de 100 peptides). Deux ordres bactériens ont été identifiés : des *Rickettsiales* (*A. phagocytophilum* et *R. helvetica*) et des *Spirochètales* (différentes espèces de *Borrelia* responsables de la borréliose de Lyme et une espèce de *Borrelia* responsable de fièvre récurrente). Pour le parasite, nous avons identifié l'espèce *B. microti*. Aucun peptide de *B. divergens* n'a été identifié.

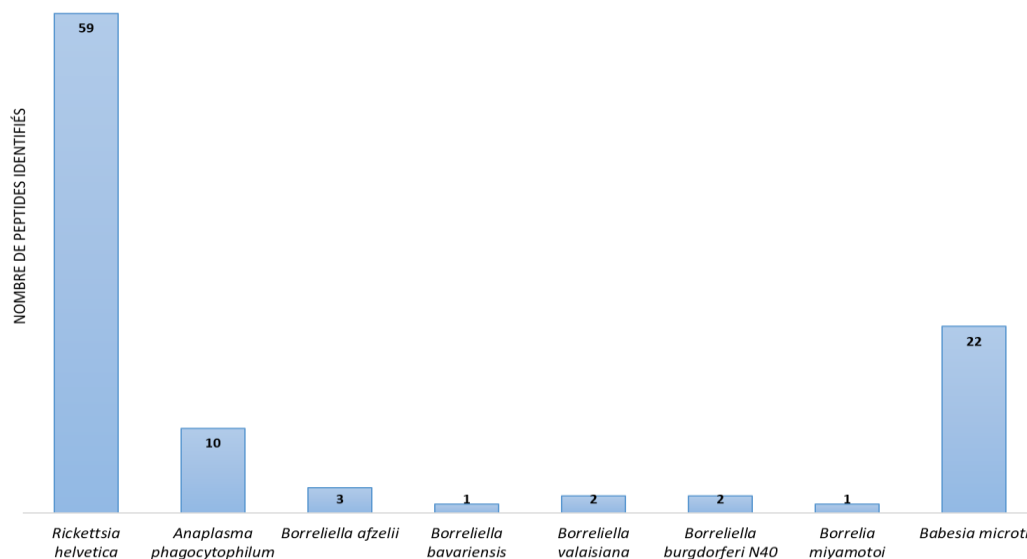


Figure III-15. Nombre de peptides uniques identifiés par pathogène dans les tiques *I. ricinus* collectées dans leur milieu naturel. Les protéines ont été extraites des tiques puis analysées par préfractionnement et LC-MS/MS. Les données ont été interprétées avec la banque B et les peptides ont été sélectionnés comme explicité dans Figure III-12, puis comparées avec les échantillons contrôles.

Pour les bactéries, on constate qu'un grand nombre de peptides appartient à *R. helvetica* (59% des peptides identifiés) et à *A. phagocytophilum* (10%), ces deux espèces faisant partie de l'ordre des *Rickettsiales* (A. Toledo *et al.*, 2009). Plusieurs peptides de différentes espèces de *Borrelia* (8%), responsables de la borréliose de Lyme (*B. afzelii*, *B. bavariensis*, *B. valaisiana*, *B. burgdorferi* N40) et un peptide (1%) à *B. miyamotoi*, agent responsable de fièvres récurrentes ont également été détectés. Le second pathogène très représenté dans ces tiques sauvages est le parasite *B. microti* avec 22 peptides (22% des peptides identifiés).

Ces résultats sont en accord avec d'autres études réalisées dans des tiques collectées dans la nature. Une étude menée par Toledo *et al.*, avec des tiques collectées dans le centre de l'Espagne et analysées par PCR, a montré que les pathogènes les plus fréquemment identifiés étaient *Rickettsia spp.*, suivie d'*A. phagocytophilum* et en troisième position *Borrelia spp.* (A. Toledo *et al.*, 2009). Une autre étude a analysé la co-infection des tiques collectées dans le département des Ardennes, France (S. Moutailler *et al.*, 2016). La technique de PCR utilisée reposait sur la multi détection d'ADN de pathogènes par une puce haut débit (Fluidigm). L'analyse des tiques femelles (267) analysées individuellement a montré que 45% étaient infectées par au moins un pathogène, avec une prévalence importante pour *Rickettsia spp.*, et moins importante pour d'autres espèces telles que *B. miyamotoi*, *A. phagocytophilum*, *Neorlichia mikurensis*, *B. divergens*. Le virus TBE et le parasite *B. microti* n'ont pas été détectés. Les co-infections les plus courante sont parmi les espèces responsables de la borréliose de Lyme (*B. burgdorferi* ss, *B. garinii*, *B. afzelii* et *B. spielmanii*) (S. Moutailler *et al.*, 2016).

D'autres études publiées récemment et réalisées par PCR ont mis en évidence un grand nombre de pathogènes étudiés dans mes travaux de thèse. Une étude réalisée avec des tiques collectées pendant trois ans dans la forêt de Sénart au Sud de Paris, a révélé la présence d'*A. phagocytophilum*, *R. helvetica*, *Borrelia spp.* (*B. burgdorferi* ss, *B. garinii*, *B. valaisiana*, *B. afzelii*, *B. miyamotoi*) et *B. microti* (E. Lejal *et al.*, 2019). Une autre étude réalisée avec des tiques collectées dans la forêt de Gribskov au Danemark a mis en évidence différentes espèces de *Borrelia* (*B. afzelii*, *B. valaisiana*, *B. miyamotoi*, *B. burgdorferi*, *B. garinii*), *R. helvetica*, *A. phagocytophilum* et *Neorlichia mikurensis*. Quant aux co-infections, des espèces de genres différents ont été détectées : *Borrelia/R. helvetica*, *B. burgdorferi/A. phagocytophilum*, *R. helvetica/A. phagocytophilum*, *N. mikurensis/Borrelia* (K. Klitgaard *et al.*, 2019). La troisième étude a été réalisée avec des tiques collectées dans la région de Murbach en Alsace (site de collecte des tiques femelles étudiées pendant mes travaux de thèse). Cette dernière étude a été effectuée par PCR et spectrométrie de masse MALDI-TOF

et a permis la détection de plusieurs espèces de *Borrelia* (*B. burgdorferi* ss, *B. afzelii*, *B. garinii*, *B. valaisiana*, *B. miyamotoi*) et *R. helvetica*. Des co-infections, telles que *Borrelia* spp./*Anaplasmataceae* et *Rickettsia* spp./*Anaplasmataceae* ont été observées (A. Nebbak et al., 2019).

Ces différentes études, réalisées par biologie moléculaire, montrent la prévalence d'un grand nombre d'espèces de pathogènes dans les tiques collectées dans leur milieu naturel. D'autres espèces de pathogènes ont été également étudiées mais nous n'avons cité que les pathogènes communs avec les études menées pendant ces travaux de thèse et qui se trouvent parmi les plus couramment observés.

Les résultats obtenus pendant mes travaux de thèse par protéomique confirment la présence de ces pathogènes dans la tique et la co-infection possible des tiques.

L'analyse des 21 **peptides partagés** a mis en évidence d'autres espèces pathogènes (Tableau III-8). Pour l'ordre des *Rickettsiales*, les peptides identifiés pour les trois espèces pathogènes identifiées (*R. helvetica*, *A. phagocytophilum* et *Neorlichia mikurensis*) a permis de mettre en évidence la présence de l'espèce *Neorlichia mikurensis*. Pour l'ordre des *Spirochétales*, d'autres espèces de *Borrelia* ont été identifiées telles que *B. garinii* et *B. burgdorferi* ss.

Tableau III-8. Peptides partagés identifiés dans les différentes espèces de pathogènes. Les trois premières colonnes correspondent aux espèces appartenant à l'ordre des Rickettsiales, puis les différentes espèces de *Borrelia* et la dernière colonne correspond au parasite *B. microti*.

Peptides	Protéine	<i>Anaplasma phagocytophilum</i>	<i>Neoehrlichia mikurensis</i>	<i>Rickettsia helvetica</i>	<i>Borrelia afzelii</i>	<i>Borrelia garinii</i>	<i>Borrelia valaisiana</i>	<i>Borrelia bavariensis</i>	<i>Borrelia burgdorferi</i> 297	<i>Borrelia burgdorferi</i> N40	<i>Borrelia burgdorferi</i> B31	<i>Borrelia miyamotoi</i>	<i>Babesia microti</i>
EGTVTLISK	outer surface protein A				X				X	X	X		
GTSDKNNGSGVLEGVK									X	X	X		
GYVLEGLTAEK									X	X	X		
LTISDDLQTTLEVFK									X	X	X		
LTISDDLQTTLEVFREDGK									X	X	X		
SSTEEFNEK									X	X	X		
YDLIATVDK									X	X	X		
YDLIATVDKLELK									X	X	X		
ANLGAFQNR					X	X	X	X	X	X	X	X	
ASDDAAGMGVSGK	Flagellin				X		X	X		X	X		
INTPASLSGSQASWTLR					X	X	X	X	X	X	X		
MIINHNTSAINASR					X	X	X	X		X	X	X	
APGFGDR		X	X	X								X	
VGGSSSEVEVK	molecular chaperone GroEL	X	X										
LSGGVAVLK		X	X	X									
IIDIVAER	neutrophil activating A (napA) / DNA starvation/stationary phase protection					X		X		X	X		
IIENIAK	ATP-dependent Clp protease subunit C									X	X		
ISDIISK	DNA polymerase III subunit alpha				X	X	X	X		X	X		
IAGLEVLR	molecular chaperone DnaK			X									X
LPLNDGK	S2/P23 family protein					X	X	X					
IQSILDK	proline--tRNA ligase				X	X	X	X		X	X		
Nombre de peptides identifiés		3	3	3	7	7	7	8	10	16	16	3	1

Ce tableau montre que les peptides partagés se trouvent principalement regroupés pour l'ordre des *Rickettsiales* et pour l'ordre des *Spirochétales*. Ces résultats mettent en évidence la forte homologie entre les différentes espèces de *Borrelia* responsables de la borréliose de Lyme. Si ces peptides sont uniques à ces espèces de *Borrelia*, ils pourraient être utilisés comme peptides cibles pour le diagnostic de la borréliose de Lyme.

Ces résultats illustrent bien la complexité d'une étude comportant des espèces proches avec des séquences très similaires, comme c'est le cas pour certains pathogènes retrouvés chez *I. ricinus*.

3. Analyse des peaux de souris

Après avoir identifié les pathogènes présents dans les tiques, nous nous sommes intéressés aux pathogènes transmis à l'hôte. En effet, la présence d'un pathogène au sein du vecteur n'entraîne pas nécessairement une transmission à l'hôte, d'où la notion de compétence vectorielle (cf. paragraphe 1.2) de ce chapitre. Nous avons ainsi analysé des peaux de souris infectées ou non, et nous avons cherché à détecter les pathogènes présents dans ces peaux.

La peau joue un rôle important dans les maladies à transmission vectorielle, elle constitue le premier contact avec le vecteur, il a été également mis en évidence qu'elle représente le site de multiplication et de persistance de la bactérie *Borrelia* (Q. Bernard *et al.*, 2020; Q. Bernard *et al.*, 2014; A. Kern *et al.*, 2011). Nous avons vu dans le chapitre II comment l'identification des protéines de *Borrelia* dans la peau avaient conduit aux approches vaccinale et diagnostique de la borréliose de Lyme. Nous avons utilisé la même approche pour détecter des protéines d'autres pathogènes dans la peau d'un hôte infecté.

Les biopsies de peaux de souris ont été analysées en deux groupes distincts :

- Le premier groupe composé des peaux de souris infectées par inoculation des bactéries *A. phagocytophilum* ou parasites *B. microti*.
- Le deuxième groupe constitué de peaux de souris infectées « naturellement » via des piqûres de tiques collectées dans la nature.

3.1. Peaux de souris infectées expérimentalement

Pour deux pathogènes (bactérie *A. phagocytophilum* et parasite *B. microti*), nous avons pu analyser des biopsies infectées expérimentalement. Ces biopsies facilitent la détection et

l'identification de protéines des pathogènes grâce d'une part à l'inoculation d'une quantité connue et d'autre part, grâce à l'utilisation d'une banque adaptée uniquement au pathogène d'intérêt, ce qui augmente la possibilité d'identification de protéines de pathogènes faiblement présentes dans les échantillons. De plus, l'étude des peaux infectées à la seringue permet de réaliser une cinétique d'infection. Une telle étude cinétique, dans le cas de la borréliose de Lyme, avait permis de mettre en évidence par PCR un pic de multiplication des bactéries *Borrelia* dans la peau 7 jours après l'inoculation (A. Kern *et al.*, 2015) et ainsi de faciliter l'identification de protéines de *Borrelia* : les peaux avec les quantifications PCR les plus élevées avaient été retenues pour les analyses en protéomique (A. Grillon *et al.*, 2017; G. Schnell *et al.*, 2015b).

3.1.1. *Anaplasma phagocytophilum*

Comme nous ne savions pas si la bactérie *Anaplasma* se comportait comme *Borrelia*, les collaborateurs (M. Kazimirova, Slovaquie) ont d'abord réalisé une cinétique d'infection à 1, 3, 7, 14, 21 jours après inoculation intra-péritonéale avec des cellules infectées par *A. phagocytophilum*. Douze biopsies cutanées présentant les taux d'infection les plus élevés (déterminés par analyse PCR) ont été sélectionnées pour l'analyse protéomique. Comme précédemment, la majorité des protéines identifiées appartenait à la souris, avec une moyenne de 3500 protéines par biopsie, mais sept protéines ont été attribuées à *A. phagocytophilum* (Tableau III-9). Pour trois biopsies, aucune protéine de pathogène n'a été identifiée. Pour les neuf autres, trois protéines d'*A. phagocytophilum* au maximum ont été identifiées par biopsie, et toutes les protéines ont été identifiées avec un seul peptide. La détection de protéines bactériennes 14, voire 21 jours après l'inoculation suggère qu'*A. phagocytophilum* persiste dans la peau de l'hôte murin, comme cela a été observé pour *Borrelia* (Q. Bernard *et al.*, 2020).

Tableau III-9. Peptides identifiés dans des biopsies murines infectées par inoculation bactérienne et dont les protéines ont été attribuées à *A. phagocytophilum*. Les peptides grisés sont les peptides partagés avec les échantillons contrôles. La nature des échantillons dans lesquels les protéines ont été identifiées est indiquée par **a** : tiques collectées dans leur milieu naturel **b** : peaux de souris piquées par des tiques **c** : peaux de souris infectées par inoculation avec *A. phagocytophilum*.

		Nature des échantillons	Prélèvement après l'inoculation								
Peptides	Protéine Numéro d'accesion		Taux d'infection								
			J+3 31	J+3 36	J+3 27	J+3 29	J+7 26	J+7 24	J+7 23	J+14 29	J+21 31
CHTIMNCTKCPK	succinate dehydrogenase iron-sulfur subunit WP_011450309.1	c									X
IIVGELR	type IV secretion protein WP_011451401.1	c	X	X			X			X	
IDEKDLR	peptide chain release factor 1 WP_011450546.1	c						X			
NELAELEK	hypothetical protein WP_044104104.1	c	X								
IDTELLK	DNA polymerase I WP_021798967.1	c		X							
ALSDAGIKDNSK	molecular chaperone DnaK WP_044104082.1	a, b, c			X						
AVATDLTK	P44/Msp2 family outer membrane protein WP_020849070.1	a, c	X	X	X	X	X		X	X	
Nombre de peptides d' <i>A. phagocytophilum</i> identifiés			3	3	2	1	2	1	1	2	1

La protéine « type IV secretion protein » a été identifiée dans plusieurs biopsies (Tableau III-9) et d'après la littérature pourrait jouer un rôle lors de l'infection par *A. phagocytophilum* : elle est utilisée par *A. phagocytophilum* pour sécréter de l'ADN ou des protéines dans la cellule eucaryote (V. Foulongne *et al.*, 2002), mécanisme connu pour l'ordre des *Rickettsiales* pour favoriser l'infection des cellules hôtes (M. S. Severo *et al.*, 2012; M. Villar *et al.*, 2015). Elle est également suspectée d'être impliquée dans la sécrétion de protéines de réponse au stress telles que HSP70 lors d'une infection par des rickettsies (M. Villar *et al.*, 2015). Cette protéine apparait donc comme marqueur d'infection active par *Anaplasma*, et candidat intéressant pour une méthode protéomique ciblée.

Bien que le peptide AVATDLTK ait été identifié dans des échantillons contrôle, la protéine **P44/Msp2** a été conservée dans notre liste d'intérêt en raison des données de la littérature. En effet, la protéine membranaire P44/Msp2 appartient à la famille P44/Msp2 riche en glycoprotéines et est impliquée dans un large panel de processus biologiques : (i) la variation antigénique qui peut permettre aux bactéries d'échapper à la réponse immunitaire de l'hôte ; (ii) l'adaptation à l'environnement de l'hôte, ce qui renforce l'idée de considérer *A. phagocytophilum* comme un « caméléon microbien » ; (iii) l'adhésion bactérienne ; (iv) l'activité de porines (Y. Rikihisa, 2010; M. S. Severo *et al.*, 2012). Cette protéine présente une forte

variation antigénique qui favorise la persistance de l'infection par *A. phagocytophilum* (Y. Rikihisa, 2011).

Cinq autres protéines ont été identifiées, dont l'hypothetical protein (WP_044104104.1), qui est nommée peptidylpropyl isomerase depuis la mise à jour de la base de données RefSeq (NCBI) en juillet 2019 (Tableau III-6). Ces protéines n'ont pas été décrites dans la littérature comme ayant un lien avec l'infection par *A. phagocytophilum*.

3.1.2. *Babesia microti*

Comme pour *Anaplasma*, des biopsies des peaux de souris ont été prélevées à différents jours (1,3,5 jours) après une injection à la seringue du parasite *B. microti*. Trois biopsies présentant les taux d'infection les plus élevés (déterminés par analyse PCR) ont été sélectionnées pour l'analyse protéomique. En moyenne 1700 protéines murines ont été identifiées, ainsi que neuf protéines attribuées à *B. microti* et une protéine attribuée à *B. divergens* (Tableau III-10). Les protéines attribuées à *B. microti* ont toutes été identifiées avec un seul peptide.

Tableau III-10. Protéines identifiées dans des biopsies murines infectées par inoculation bactérienne et attribuées à *B. microti*. Les peptides grisés sont les peptides partagés avec les échantillons contrôles. Les séquences en rouge indiquent les peptides partagés avec d'autres pathogènes, l'hôte ou le vecteur. La nature des échantillons dans lesquels les protéines ont été identifiées est indiquée par **a** : tiques collectées dans leur milieu naturel **b** : peaux de souris piquées par des tiques **c** : peaux de souris infectées par inoculation à la seringue avec *B. microti*. Les analyses ont été faites en duplicat d'injection (R1, R2).

Pathogène	Peptides	Protéines / Numéro d'accension	Nature des échantillons	Prélèvement après l'inoculation					
				J+1		J+3		J+5	
				R1	R2	R1	R2	R1	R2
<i>B. microti</i>	FEQALNVDITEFQTNLVPYPR	tubulin alpha / XP_012650470.1	a,b,c	X					
	GHYTEGAELIDSVLDVVR	tubulin beta / XP_012647284.1	a,b,c	X	X	X	X	X	X
	ILSNASCTTNCLAPLAK	GAPDH / XP_012647126.1	b,c	X	X	X	X	X	X
	LVLLGETAVGK	Rab family / XP_021338491.1	a,b,c	X	X	X	X	X	X
	IILIGDATVGK	Rab family / XP_021338236.1	b,c				X		
	LLPGELSK	Histone H2B / XP_012649280.1	a,b,c	X	X		X		
	LVLPGELSK	Histone H2B / XP_012647666.1	c				X		
	QLEVQHAAK	T-complex protein 1 subunit alpha / XP_021338144.1	b,c		X				
	AGISVLR	T-complex protein 1 subunit gamma / XP_012650473.2	b,c				X		
	YSVWIGGSILSSLSTFQQMWITK	Actin / XP_012649089.1	a,b,c		X	X	X		X
	DFTVSSMHGMDQK	translation initiation factor eIF-4A / XP_021337506.1	a,c				X		
<i>B. divergens</i>	FIIPNVVK	LGT, LDH, L-lactate dehydrogenase / XP_012647088.1	a,b,c		X				
	AIEEAIK	Heat shock protein 70 / Q0WY85	c						X
Nombre total de peptides identifiés				5	7	4	9	3	5

Comme pour *A. phagocytophilum*, les peptides grisés sont ceux qui ont été attribués à *B. microti* mais partagés avec les échantillons contrôles. Dans ce cas, ces peptides ne sont pas pris en compte, mais cela n'exclut pas que les protéines soient sélectionnées comme cible potentielle. Comme nous pouvons le constater, 10 protéines ont été identifiées dans les peaux

piquées par des tiques sauvages et dans les peaux inoculées à la seringue (nature d'échantillon b et c). Ces protéines pourraient être des marqueurs intéressants pour la détection de *B. microti* dans la peau. Pour effectuer une sélection, une étude bibliographique de ces protéines a été menée.

La protéine **GAPDH**, est une enzyme impliquée dans la conversion du glucose en pyruvate. Cette protéine pourrait être utilisée comme un potentiel marqueur diagnostique d'une infection active à *B. microti* chez l'hôte (R. Magni *et al.*, 2019). La GAPDH de *Borrelia* fait partie des protéines ciblées dans la méthode diagnostique de la borréliose de Lyme et a été détectée dans des biopsies de patients. Cette protéine est connue pour être impliquée dans la transmission et la survie de la bactérie (Cf chapitre II).

Les **sous-unités alpha et beta de la tubuline** sont les protéines majeures des microtubules, fibres constitutives du cytosquelette des cellules eucaryotes. Le gène codant la sous-unité beta de la tubuline est utilisé comme amorce spécifique de *B. microti* pour l'analyse PCR (S. H. Hong *et al.*, 2019; L. Schnittger *et al.*, 2012).

L'**actine** est fortement conservée entre différentes espèces, ce qui pourrait expliquer l'identification d'un peptide commun entre la tique et *B. microti*. Les pathogènes se servent de l'actine lors de l'invasion de la cellule hôte, ce qui garantit leur survie (S. Kuhn *et al.*, 2017).

La protéine **Rab** compose la famille de GTPases qui ont pour fonction le transport vésiculaire. Elle intervient également lors de la signalisation, la division, la survie et la migration des cellules. L'altération de ces protéines est associée à des infections et sert aux pathogènes à échapper à leur dégradation dans la cellule de l'hôte (N. A. Guadagno *et al.*, 2019; R. Magni *et al.*, 2019).

L'**histone H2B** impliquée dans le métabolisme de l'ADN et de l'ARN (R. Magni *et al.*, 2019). Elle a été associée à l'infection avec la bactérie *Rickettsia* transmis par la tique (C. Thepparit *et al.*, 2010), cependant aucune information n'a été trouvée pour cette protéine en lien avec l'infection par *B. microti*.

Les **sous-unités alpha et beta du complexe T-complex 1** sont des protéines connues pour agir comme un chaperon moléculaire pour l'actine, la tubuline et probablement d'autres protéines (K. I. Brackley *et al.*, 2009), en accord avec l'identification de l'actine et la tubuline dans ces échantillons.

La **protéine d'initiation de translation eIF-4A** est connue pour participer au cycle de répllication de certains virus. Cette caractéristique est utilisée pour contrôler les infections associées au virus (H. Montero *et al.*, 2019).

La protéine **HSP70** est une protéine qui est observée dans la phase aiguë de la babésiose humaine (B. Carcy *et al.*, 1991) et a été également utilisée comme amorce pour effectuer la PCR dans une étude afin de discriminer *B. divergens* entre différents espèces du genre *Babesia* dans des analyse de tiques (M. Blaschitz *et al.*, 2008).

Aucun lien n'a été décrit entre la protéine **L-lactate dehydrogenase** et l'infection par *B. microti*. Cependant, cette protéine a été sélectionnée comme cible intéressante, en raison de son identification dans les trois types d'échantillon étudiés (tiques, peaux infectées via tiques ou via seringue).

Parmi les trois peptides partagés (en rouge dans le tableau III-10), le peptide YSVWIGGSILSSLSTFQQMWITK de l'actine, bien que partagé entre *B. microti* et la tique (*I. ricinus*), a été considéré comme spécifique de *B. microti* puisque ces souris n'avaient jamais été en contact avec une tique lors de l'infection. En ce qui concerne la β -tubuline, malgré la détection du même peptide dans les échantillons contrôle, elle a été conservée comme protéine d'intérêt en raison de son utilisation courante comme amorce pour les analyses PCR.

3.2. Peaux de souris infectées via des tiques collectées dans la nature

Nous avons également eu accès à des peaux de souris infectées naturellement *via* des piqûres de tiques. Le nombre de bactéries détectées par PCR suite à la piqûre de tiques est plus faible qu'avec l'inoculation à la seringue. Dans le cas de *Borrelia*, il a été montré par biologie moléculaire que l'inoculation après la piqûre correspond à environ 100 *Borrelia* par tique et l'inoculation à la seringue à 1000 *Borrelia* (A. Kern *et al.*, 2011; C. M. R. LIMA *et al.*, 2005). Le challenge analytique était donc encore plus grand mais permettrait de valider la compétence vectorielle des tiques pour les pathogènes identifiés.

Sept peaux de souris infectées « naturellement » ont été analysées, parmi lesquelles six étaient positives par analyse PCR pour l'espèce *Borrelia*, et une était faiblement positive pour *A. phagocytophilum*.

En utilisant la même approche protéomique que précédemment, nous avons identifié différents pathogènes dans ces biopsies. Suite à la comparaison avec les échantillons contrôle, 12/41 (29%) des protéines de pathogènes ont été conservées. Comme attendu, les protéines de

souris sont largement majoritaires devant les protéines de pathogènes (> 3500 protéines totales identifiées, 12 protéines de pathogènes).

Les résultats obtenus à partir de(s) **peptide(s) unique(s)** montrent la présence du parasite *B. microti* et de différentes espèces de la bactérie *Borrelia* (Tableau III-11). Une co-infection est observée pour la souris 3, avec l'identification de 3 pathogènes différents : *B. bavariensis*, *B. miyamotoi* et *R. helvetica*.

Tableau III-11. Protéines identifiées à partir de peptides uniques dans des biopsies murines infectées par des tiques collectées dans la nature. Ces protéines ont été identifiées à partir d'un seul peptide par protéine. Les résultats de PCR sont donnés pour *Borrelia* et *Anaplasma* (Pos ou Neg).

Pathogène	Peptides	Protéine / Numéro d'accension	Résultats PCR / protéomique						
			Souris 1 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Neg)	Souris 2 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Neg)	Souris 3 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Neg)	Souris 4 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Neg)	Souris 5 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Neg)	Souris 6 <i>Borrelia</i> (Pos) <i>Anaplasma</i> (Pos)	Souris 7 <i>Borrelia</i> (Neg) <i>Anaplasma</i> (Neg)
<i>B. microti</i>	LNQEFIK	Cullin family / XP_021337634.1						X	
<i>B. microti</i>	LVVVITAGAR	LGT, LDH, L-lactate dehydrogenase / XP_012647088.1	X						X
<i>B. microti</i>	AVITVPAYFNDQSR	molecular chaperone DnaK / XP_021337201.1	X						
<i>B. microti</i>	VAFIQDIK	minichromosome maintenance protein 7 (cell division control protein) / XP_021337353.1		X					
<i>B. microti</i>	DVKMLEK	structural maintenance of chromosome 4 / XP_021337511.1				X			
<i>B. bavariensis</i>	ISIVQTK	NAD(+) synthase / WP_011193832.1			X				
<i>B. valaisiana</i>	QDLDEKLK	Bdr family protein / WP_040351410.1					X		
<i>B. miyamotoi</i>	EINISLR	ABC transporter ATP-binding protein / WP_020955116.1			X				
<i>R. helvetica</i>	IFAGVDVTK	succinate dehydrogenase flavoprotein subunit / WP_010423907.1			X				

Le tableau montre des résultats positifs pour toutes les biopsies analysées. L'ensemble des protéines ont été identifiées à partir d'un seul peptide par protéine (et peptide unique au pathogène).

Le parasite *B. microti* a été identifié dans 5 biopsies analysées, résultats qui n'ont pas pu être comparés à la PCR non réalisée pour ce pathogène. Les PCR effectuées par nos collaborateurs ciblaient seulement les bactéries *Borrelia* et *Anaplasma*. La bactérie *R. helvetica* a été identifiée dans une biopsie (souris 3), sans comparaison possible à la PCR. La bactérie *A. phagocytophilum* n'a été identifiée dans aucune biopsie ; seule une biopsie était positive en PCR. Enfin, des bactéries *Borrelia* ont été identifiées dans deux biopsies (souris 3 et souris 5) ; 6 biopsies étaient positives par PCR.

En considérant les **peptides partagés**, nous avons identifié trois protéines supplémentaires par rapport au Tableau III-11. La **3-hydroxy-3-methylglutaryl-CoA synthase** est identifiée avec un peptide partagé aux quatre espèces de *Borrelia*. Elle peut donc être considérée comme une cible protéique pour la détection de la borréliose de Lyme. *A contrario*, les peptides appartenant à l'**actine** et la « **subunit ribosomal protein L10Ae** » partagés entre le parasite *B. microti* et la tique *I. ricinus* ne sont pas pris en compte car nous ne pouvons pas discriminer entre le vecteur et le pathogène.

V- Conclusion

En conclusion, l'analyse protéomique non ciblée des tiques et peaux de souris, réalisée par préfractionnement SDS-PAGE et nanoLC-MS/MS, nous a permis de mettre en évidence :

- la variété des pathogènes circulant dans la nature et présents dans des tiques collectées dans la région Grand-Est,
- la présence de plusieurs pathogènes dans des peaux de souris infectées, confirmant ainsi la compétence vectorielle des tiques et leur capacité à transmettre ces pathogènes à un hôte,
- la co-infection possible d'une souris, avec plusieurs pathogènes au sein d'une même biopsie,
- des protéines candidates pour le développement d'une méthode diagnostique multiplexée ciblant plusieurs pathogènes transmis par les tiques (Tableau III-12).

Tableau III-12. . Liste des protéines sélectionnées pour le développement de la méthode ciblant *A. phagocytophilum* et *B. microti*.

Pathogène	Protéine
<i>A. phagocytophilum</i>	AnkA
	DNA polymerase I
	Type IV secretion protein
	Peptidylpropyl isomerase
	Peptide chain release factor 1
	AcrB/AcrD/AcrF family protein
	P44 outer membrane protein
	Pentapeptide repeats family protein
	Succinate dehydrogenase iron-sulfur subunit
	DNA polymerase III subunit gamma/tau
<i>B. microti</i>	Actine
	Beta tubuline
	Culline family
	Molecular chaperone DnaK
	LGT, LDH, L-lactate dehydrogenase
	Large subunit ribosomal protein L10Ae
	Structural maintenance of chromosome 4
	minichromosome maintenance protein 7 (cell division control protein)

En vue de développer une méthode protéomique ciblée, des peptides protéotypiques de ces protéines ont été sélectionnés, et les peptides SIS ont été commandés récemment. La suite du travail consistera à mettre au point une méthode multiplexée pour trois pathogènes, ciblant

à la fois des peptides de *Borrelia* (cf chapitre II), des peptides d'*Anaplasma* et des peptides de *Babesia*.

Il est également prévu d'analyser des peaux de souris infectées par d'autres pathogènes, tels que *Rickettsia* et le virus TBE. Ces analyses nécessiteront de nouvelles optimisations pour les recherches dans les banques de données. Si des protéines d'autres pathogènes sont identifiées, elles pourront être implémentées dans la méthode diagnostique de manière à augmenter le nombre de pathogènes ciblés.

A terme, cette approche diagnostique par protéomique ciblée pourra être appliquée à des biopsies humaines. Une bonne sélection des patients pour la validation de la méthode sera un point clé, en raison de la diversité des symptômes observés pour les différentes pathologies. Les patients qui sont susceptibles d'être testés sont des patients qui auront été piqués par une tique et qui présenteront un syndrome pseudo-fébrile, en général 7 à 10 jours après piqure de tique.

Chapitre IV. Analyse protéomique de *Pseudomonas aeruginosa*

Ce projet d'analyse protéomique de *Pseudomonas aeruginosa* a été mené en étroite collaboration avec Quentin Perraud qui effectuait sa thèse dans l'équipe du Dr Isabelle Schalk (UMR 7242, ESBS, Strasbourg). Cette équipe s'intéresse depuis longtemps aux mécanismes moléculaires d'assimilation du fer chez *P. aeruginosa* et cherche à développer de nouveaux antibiotiques capables d'exploiter ces voies d'assimilation du fer pour s'introduire dans la bactérie.

I- Introduction

P. aeruginosa est une bactérie à Gram-négatif, pathogène et opportuniste, responsable de nombreuses et sévères infections chez l'homme. Son caractère **ubiquitaire** fait qu'on la retrouve dans divers environnements tels que l'eau, les plantes, d'autres zones humides et chez l'homme (M. Abdallah *et al.*, 2014; M. D. Parkins *et al.*, 2018)

Cette bactérie est dite **opportuniste**, car elle infecte les personnes déjà atteintes par d'autres pathologies, telles que les personnes immunodéprimées et les grands brûlés (J. B. Lyczaka *et al.*, 2000). Très répandu dans les milieux hospitaliers, ce pathogène est l'une des principales causes d'infections nosocomiales (G. L. Mislin *et al.*, 2014; M. D. Obritsch *et al.*, 2012).

Les infections nosocomiales causées par *P. aeruginosa* représentent un problème de santé publique, du fait de la résistance de la bactérie aux antibiotiques. Depuis 2017, l'Organisation Mondiale de la Santé a classé cette bactérie parmi la liste des pathogènes prioritaires pour la recherche et le développement de nouveaux antibiotiques (A. Savoldi *et al.*, 2019).

P. aeruginosa a la capacité d'acquérir rapidement de nouveaux mécanismes de résistance (G. L. Mislin *et al.*, 2014; Y. Morita *et al.*, 2014). Par exemple chez les patients atteints de mucoviscidose et souffrant d'infections chroniques à *P. aeruginosa*, l'administration en continue d'antibiotiques provoque une évolution adaptative de la bactérie (A. Folkesson *et al.*, 2012). *P. aeruginosa* présente également une résistance intrinsèque aux antibiotiques, due, au moins en partie, à la faible perméabilité de ses membranes qui empêche aux antibiotiques d'atteindre leurs cibles à l'intérieur de la bactérie (G. L. Mislin *et al.*, 2014).

Une des stratégies envisagées pour permettre aux antibiotiques de pénétrer dans la bactérie est celle du cheval de Troie (A. Gorska *et al.*, 2014; G. L. Mislin *et al.*, 2014; S. Rebuffat, 2012).

Dans ce cadre, l'équipe d'I. Schalk développe une stratégie « cheval de Troie » basée sur l'acquisition du fer par *P. aeruginosa*. Le fer est un nutriment essentiel pour la croissance des bactéries. Pour l'acquérir, la plupart des bactéries utilisent des chélateurs du fer appelés **sidérophores** (H. Boukhalfa *et al.*, 2002; S. Vaulont *et al.*, 2015). Ces sidérophores sont synthétisés par les bactéries et sécrétés dans leur environnement pour puiser le fer essentiel à leur développement (Figure IV-1).

L'idée est de vectoriser des antibiotiques incapables de diffuser à travers la paroi bactérienne en exploitant ces voies d'acquisition du fer (G. L. Mislin *et al.*, 2014) Figure IV-1. Le développement de conjugués sidérophore-antibiotiques devrait permettre d'améliorer l'efficacité des traitements.

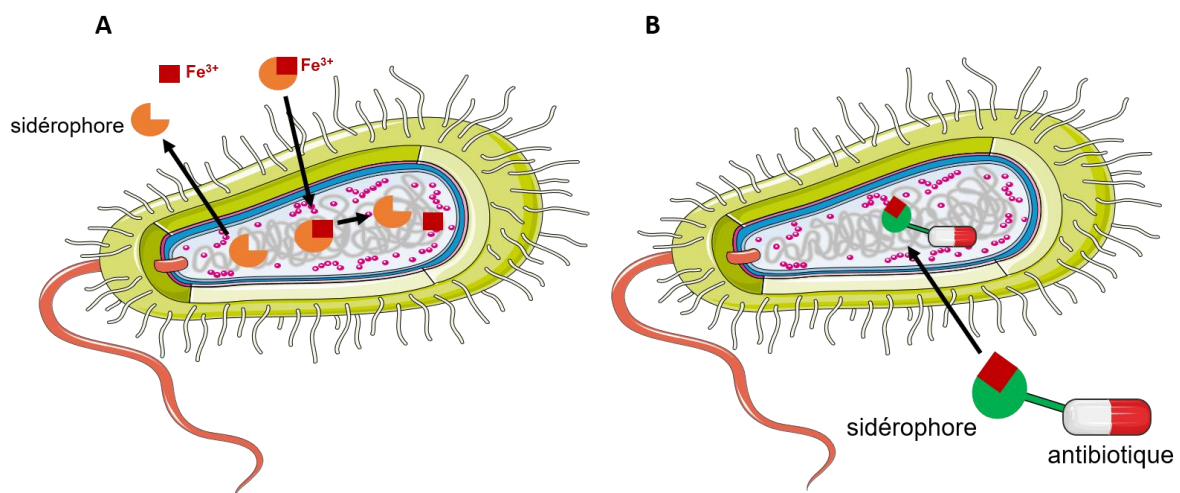


Figure IV-1. Stratégie antibiotique dite du cheval de Troie. **A.** Représentation de la bactérie *P. aeruginosa* et de l'action des sidérophores, qui sont sécrétés pour capter le fer à l'extérieur et l'internaliser dans la bactérie. **B.** Représentation de la vectorisation d'un conjugué sidérophore-antibiotique, en utilisant les voies d'acquisition du sidérophore pour faire pénétrer l'antibiotique dans la bactérie.

Dans ce contexte, nous avons mené différentes études protéomiques différentielles pour comprendre comment *P. aeruginosa* adapte l'expression de ses différentes voies d'acquisition du fer en fonction des composés présents dans le milieu extérieur. Ces informations peuvent aider au développement des conjugués sidérophore-antibiotiques.

L'originalité de cette étude, par rapport à des travaux antérieurs effectués sur des cultures de *P. aeruginosa* (O. Cunrath *et al.*, 2016) réside dans l'analyse de ***P. aeruginosa* dans un modèle d'infection**. Ce sont des cellules épithéliales pulmonaires infectées par *P. aeruginosa*.

qui ont été analysées, où l'environnement de la bactérie est plus représentatif des conditions réelles d'infection dans un organisme. Cette situation a nécessité beaucoup d'optimisations, d'une part sur le modèle d'infection (travail réalisé par Q. Perraud), d'autre part sur la préparation d'échantillon et les conditions d'analyse protéomique.

Nous avons ainsi étudié **l'adaptation phénotypique de *P. aeruginosa*** à la présence (i) de différents exosidérophores, des sidérophores bactériens ou fongiques non produits par *P. aeruginosa* mais que le pathogène peut utiliser pour accéder au fer, (ii) de deux conjugués sidérophore-antibiotiques : l'albomycine, et le TCV-L6. Pour cela nous avons mis en œuvre une approche différentielle *label-free*, en étudiant les variations d'abondance des protéines de *P. aeruginosa* en présence de toutes les autres protéines majoritaires provenant des cellules épithéliales humaines.

II- Développements méthodologiques

Comme expliqué dans le chapitre I, les différentes étapes d'une analyse protéomique différentielle sont :

1. La préparation de l'échantillon
2. L'analyse par LC-MS/MS
3. L'interprétation qualitative (identification des protéines) et quantitative (variation d'abondance) des données

Pour ce projet, nous avons commencé par étudier différents paramètres pour nous assurer de la faisabilité de l'étude, pour optimiser la préparation des échantillons, et pour optimiser le traitement des données.

1.1. Etude de faisabilité de la détection de protéines de *P. aeruginosa* dans des cellules épithéliales pulmonaires infectées

Une étude préliminaire avait été effectuée, avant mon arrivée au laboratoire, pour savoir s'il était possible de détecter des protéines de *P. aeruginosa* dans des cellules épithéliales pulmonaires A549 (ATCC® CCL-185™) infectées. Un extrait de cellules A549 infectées par *P. aeruginosa* avait été analysé selon la stratégie décrite sur la Figure II-9, par préfractionnement SDS-PAGE (9 bandes) puis analyse nanoLC-MS/MS sur un couplage nanoAcquity-QExactive+. Les recherches effectuées à l'aide de Mascot et OMSSA avaient permis

d'identifier 669 protéines de *P. aeruginosa* et plus de 5530 protéines humaines. Ce résultat montrait la faisabilité de la détection de protéines de *P. aeruginosa* au milieu des protéines humaines.

Durant ma thèse, nous avons commencé par évaluer la possibilité de nous affranchir de l'étape de préfractionnement SDS-PAGE. En effet, une approche sans préfractionnement est préférable dans une stratégie de protéomique différentielle *label-free*, où la reproductibilité est essentielle et où on cherche à éviter tout biais de quantification. Nous avons ainsi testé une préparation de l'échantillon via un *stacking gel* (SG), en concentrant dans une seule bande de gel toutes les protéines de l'échantillon (Chapitre I.1-1.2). La Figure IV-2 montre le nombre de protéines identifiées avec un préfractionnement SDS-PAGE (9 bandes) ou avec un SG. Il faut néanmoins préciser que cette expérience n'a pas été réalisée sur le même lot de cellules infectées (lot de 2015 pour le préfractionnement, lot de 2017 pour le SG).

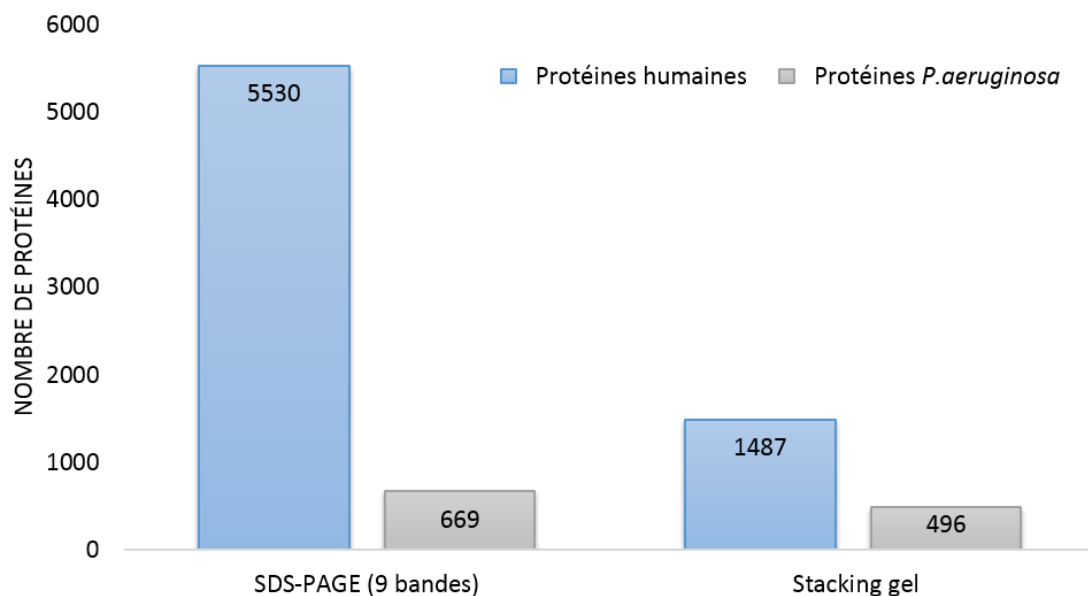


Figure IV-2. Nombre de protéines identifiées dans des cellules A549 infectées par *P. aeruginosa*. Les protéines extraites de cellules infectées par *P. aeruginosa* ont été extraites puis séparées par SDS-PAGE (à gauche) ou concentrées par SG (à droite), avant d'être digérées et analysées par nanoLC-MS/MS. Les gradients et les quantités injectées ont été adaptés à la technique de préparation d'échantillon.

Pour le SG, nous avons observé une baisse du nombre de protéines identifiées cohérente avec l'absence de fractionnement. La diminution du nombre de protéines bactériennes est moins prononcée que pour les protéines humaines, probablement en raison de la différence de lot de cellules infectées.

Les protéines de *P. aeruginosa* identifiées dans les deux analyses ont été comparées et un recouvrement de 75% a été observé. Nos collaborateurs ont vérifié que les protéines d'intérêt, qui sont impliquées dans les voies d'acquisition du fer, faisaient partie des 370 protéines communes.

Au vu de ces résultats, nous avons décidé de retenir une **approche SG** sans préfractionnement pour la suite de nos optimisations. Le nombre de protéines de *P. aeruginosa* identifiées semblait suffisant pour pouvoir réaliser une étude différentielle sur les protéines bactériennes, tout en évitant les biais de reproductibilité.

1.2. Optimisation du gradient chromatographique

Nous avons évalué deux gradients chromatographiques au niveau de l'étape de nanoLC-MS/MS. Cette étude a été menée sur une fraction surnageante de cellules A549 infectées par *P. aeruginosa* (A549 CAA sur) en comparant un gradient de 120 ou de 180 min. L'analyse des données a été réalisée avec l'outil *MS Diag* développé au laboratoire, qui permet d'avoir une estimation du nombre de spectres assignés avec le score associé en fonction du temps. Comme le montre la Figure IV-3 nous avons observé des profils similaires.

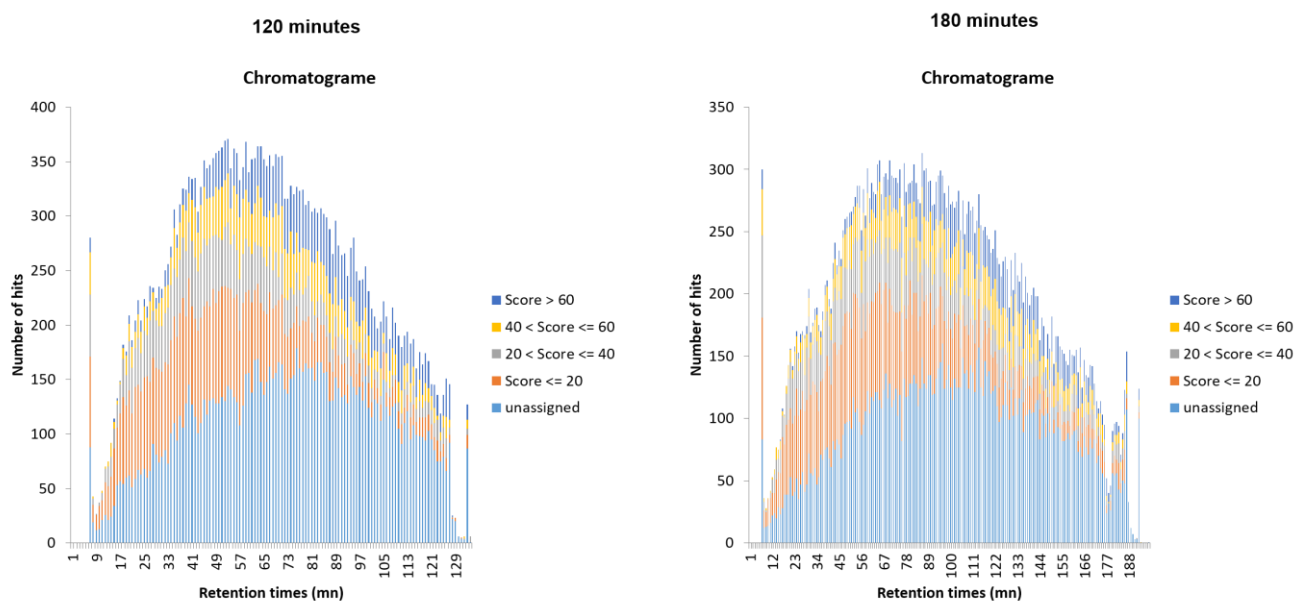


Figure IV-3. Comparaison de deux gradients de 120 min et 180 pour l'analyse des protéines contenues dans un surnageant de cellules A549 infectées par *P. aeruginosa* (A549 CAA sur). Injection 750 ng.

Nous avons également comparé le nombre de peptides et de protéines de *P. aeruginosa* identifiés, ainsi que le nombre de spectres MS/MS assignés (Tableau IV-1)

Tableau IV-1. Nombre de peptides/protéines identifiés et de spectres MS/MS assignés (score Mascot>20) pour les deux gradients de 120 min et 180 min utilisés pour analyser les protéines contenues dans un surnageant de cellules A549 infectées par *P. aeruginosa*

Gradient	Nombre de protéines	Nombre de peptides	Spectres MS/MS score > 20
120 min	684	6628	35%
180 min	684	6986	33%

Nous avons décidé de privilégier un maximum d'identification en un temps raisonnable et un maximum de spectres MS/MS assignés avec un score > 20. Nous avons donc retenu un **gradient de 120 min** pour la suite de notre étude.

1.3. Sélection de la banque de données et de l'échantillon biologique

L'étude préliminaire de faisabilité avait été faite en utilisant, pour les recherches bioinformatiques, une banque protéique contenant toutes les souches de *P. aeruginosa* issues

de UniProtKB/SwissProt (séquences vérifiées et annotées, Chapitre I.I-3.1.1). Comme l'étude différentielle allait être réalisée sur une souche PAO1 de *P. aeruginosa*, nous avons évalué une nouvelle banque générée avec les séquences *P. aeruginosa* PAO1 dans UniProtKB. L'«ancienne» banque et la «nouvelle» banque contenaient respectivement 2992 et 5677 entrées protéiques.

L'influence de la banque de données a été étudiée sur deux échantillons : une fraction surnageante de cellules A549 infectées par *P. aeruginosa* (A549 CAA sur) et une fraction adhérente de cellules A549 infectées par *P. aeruginosa* (A549 CAA). Les résultats sont montrés dans la Figure IV-4. Nous avons observé que la nouvelle banque permettait d'identifier deux fois plus de protéines bactériennes que la banque utilisée pour l'étude préliminaire. Ces résultats illustrent une nouvelle fois l'importance de bien choisir la banque protéique avant une étude d'envergure de manière à être dans les conditions optimales pour l'identification et la quantification des protéines d'intérêt.

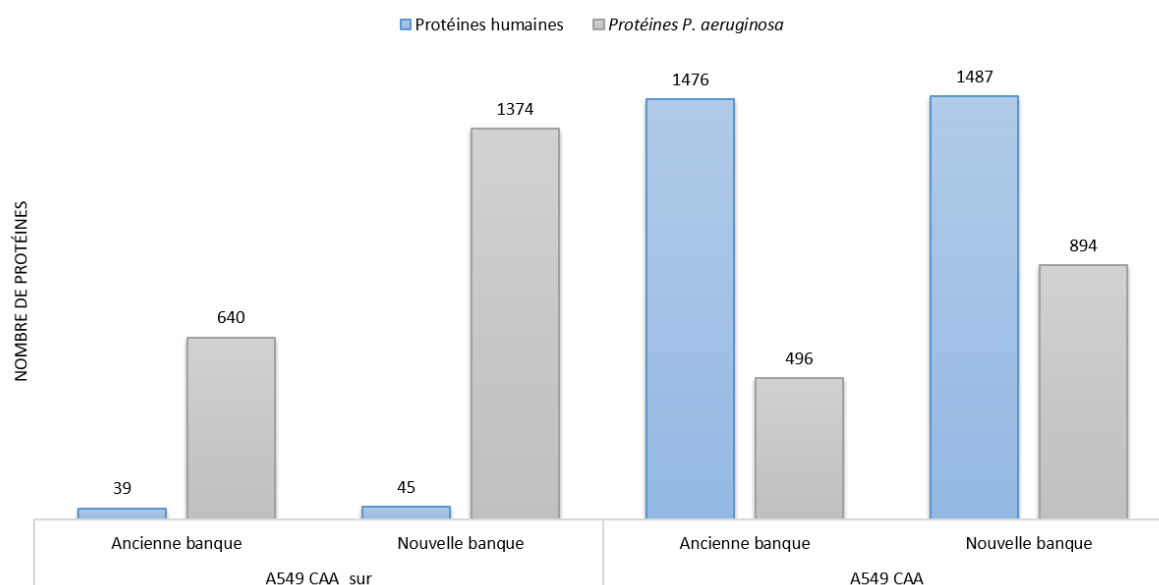


Figure IV-4. Influence de la banque de données sur le nombre de protéines identifiées. Les histogrammes montrent les nombres de protéines humaines (bleu) et bactériennes (gris) identifiées dans des cellules A549 infectées par *P. aeruginosa*. A549 CAA-sur : fraction surnageante ; A549 CAA : fraction adhérente. L'« ancienne » banque contient les protéines humaines et les protéines de toutes les souches de *P. aeruginosa* d'UniProtKB/SwissProt ; la « nouvelle » banque contient les protéines humaines et les protéines de *P. aeruginosa* PAO1 d'UniProtKB.

Au vu de ces résultats, nous avons décidé de retenir la **nouvelle banque** pour la suite de notre étude. Nous avons également décidé de travailler avec la fraction adhérente des cellules infectées (A549 CAA), car elle est plus représentative de l'infection.

1.4. Sélection du moteur de recherche

L'étude préliminaire de faisabilité avait été faite en utilisant, pour les recherches bioinformatiques, deux moteurs de recherche Mascot et OMSSA. L'utilisation de ces deux moteurs permet d'augmenter le nombre de protéines identifiées mais est chronophage dans l'étape de traitement des données (Chapitre I.I-1.2). C'est pourquoi nous avons évalué l'utilisation d'un seul moteur de recherche pour traiter les données acquises lors de l'expérience précédente. Nous avons observé qu'il n'y a pas de différence significative entre le nombre de protéines identifiées par Mascot ou OMSSA (Figure IV-5). Au vu de ces résultats, et pour gagner du temps dans l'étape de traitement de données, c'est l'algorithme Mascot qui a été retenu pour la suite de notre étude. Un tel choix avait déjà été fait dans le Chapitre III.III-2.2.2

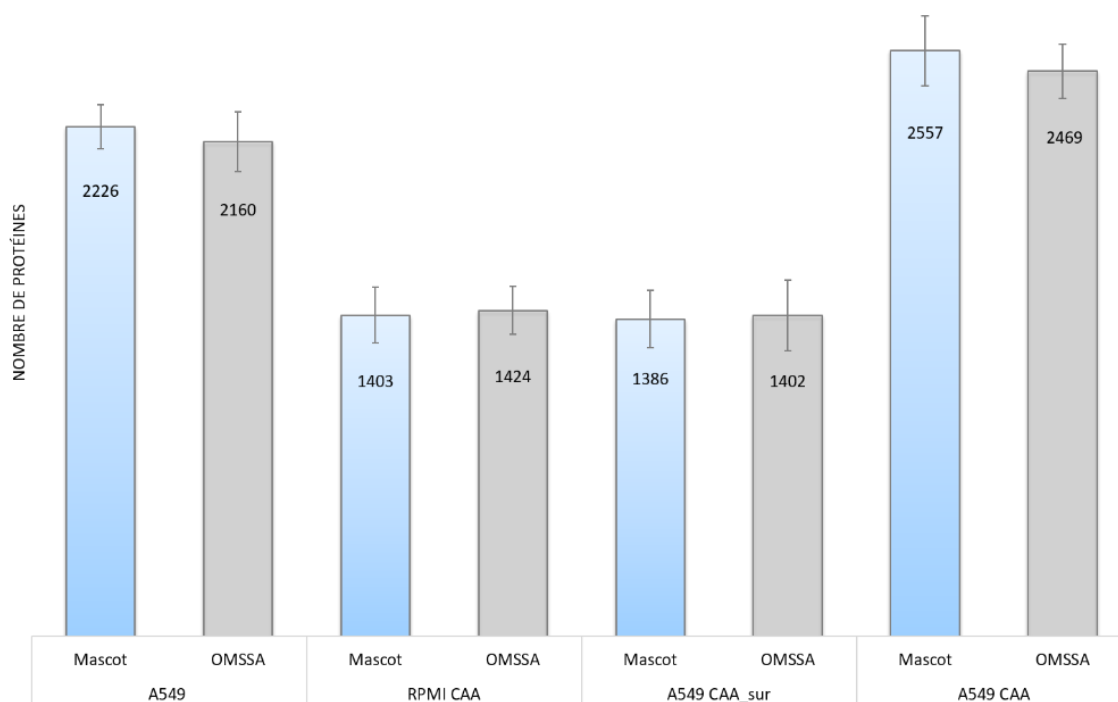


Figure IV-5. Nombre moyen de protéines totales identifiées selon le moteur de recherche utilisé. Les histogrammes montrent les nombres moyens (n=3) de protéines totales identifiées par Mascot (bleu) ou par OMSSA (gris) dans des cellules A549 infectées par *P. aeruginosa* (fraction surnageante ; fraction adhérente), dans des cellules A549 non infectées et dans une culture de *P. aeruginosa*.

1.5. Optimisation du protocole de préparation d'échantillon

Suite à ces premières optimisations, nous nous sommes ensuite intéressés au protocole de préparation des échantillons pour notre étude différentielle. La préparation d'échantillon est cruciale pour la réussite d'une approche protéomique quantitative *label-free*, elle doit être robuste et reproductible. En nous basant sur l'expérience acquise au laboratoire, nous avons comparé trois protocoles de préparation d'échantillon : stacking gel (SG), tube gel (TG) et digestion en solution (DS). Ces trois façons de préparer un échantillon ont été présentées dans le Chapitre I.I-1.2

Pour cette étude, nous avons utilisé des cellules A549 infectées par *P. aeruginosa* cultivées en milieu Lysogeny Broth (ou Luria-Bertani, LB). Ce milieu est différent de celui utilisé pour les optimisations précédentes, mais a été choisi suite à des modifications faites entre-temps par nos collaborateurs.

Ces cellules infectées ont été traitées par SG, TG, DS, en triplicat pour chaque protocole (Figure IV-6). Les 9 mélanges peptidiques issus de ces traitements ont été analysés par nanoLC-MS/MS avec un gradient de 120 min. Les données ont été traitées avec Mascot et la banque de donnée choisie précédemment.

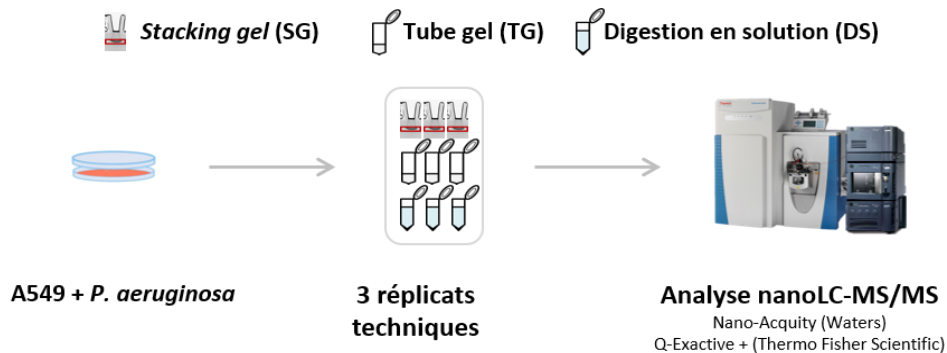


Figure IV-6. Comparaison de 3 protocoles de préparation des cellules infectées. Des cellules A549 infectées par *P. aeruginosa* ont été préparées par une approche SG, TG ou DS. Chaque expérience a été réalisée en triplicat. Après digestion, les peptides ont été analysés par nanoLC-MS/MS (750 ng injectés) et les données ont été traitées pour identifier les protéines humaines et de *P. aeruginosa*.

La reproductibilité de chaque protocole a été déterminée en comparant les résultats des triplicats (Figure IV-7). Au sein de chaque protocole, on observe une très bonne reproductibilité du nombre de protéines humaines identifiées (CV compris entre 2 et 4%). En ce qui concerne le nombre de protéines de *P. aeruginosa*, le protocole SG est légèrement plus reproductible (CV 8%) que les protocoles TG et DS (CV 12% et 11%, respectivement). Par contre, le protocole TG permet d'identifier plus de protéines que les deux autres protocoles, que ce soit des protéines humaines ou des protéines de *P. aeruginosa*. De plus, le protocole TG permet un gain de temps par rapport au temps nécessaire pour faire un SG, il est bien compatible avec une quantification *label-free* (L. Muller *et al.*, 2019) et il est robuste. Nous avons donc décidé de privilégier ce protocole TG pour la suite de nos travaux.

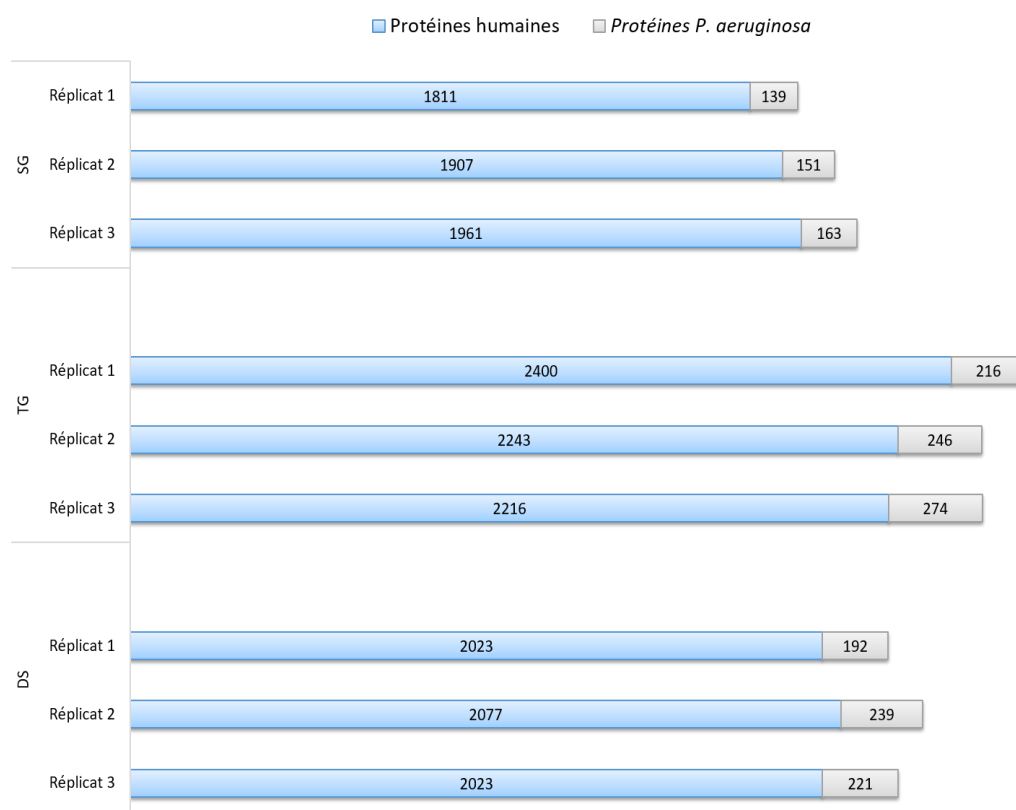


Figure IV-7. Nombre de protéines identifiées en fonction du protocole de préparation. Des cellules A549 infectées par *P. aeruginosa* ont été préparées par une approche SG, TG ou DS. Chaque expérience a été réalisée en triplicat. En bleu, le nombre de protéines humaines identifiées. En gris, le nombre de protéines de *P. aeruginosa* identifiées.

Globalement le nombre de protéines de *P. aeruginosa* identifiées est plus faible dans cette expérience que celui obtenu dans les expériences précédentes (par exemple Figure IV-4, A549 CAA). Cette baisse a été attribuée au changement de milieu de culture de *P. aeruginosa* (LB au lieu de CAA) et/ou à une croissance moins forte des bactéries. Cela a conduit nos collaborateurs à mettre en place un contrôle qualité en amont des analyses protéomiques pour s'assurer de la bonne croissance de *P. aeruginosa* dans les manipulations ultérieures.

1.6. Choix des réplicats

Dans une étude protéomique quantitative *label-free*, il est indispensable de faire des réplicats de chacune des conditions de l'analyse différentielle. En effet, dans le traitement des données, une analyse statistique est réalisée pour déterminer les variations d'abondance significatives

entre différentes conditions. Ces réplicats peuvent être des réplicats biologiques et/ou des réplicats techniques.

La dernière optimisation a consisté à déterminer s'il était préférable de faire des réplicats biologiques – dans notre cas des échantillons de cellules infectées – ou des réplicats techniques – dans notre cas des analyses nanoLC-MS/MS répétées du même échantillon. Trois lots de cellules A549 infectées par *P. aeruginosa* ont été traitées par TG, et chaque mélange peptidique a été injecté trois fois en nanoLC-MS/MS (Figure IV-8)

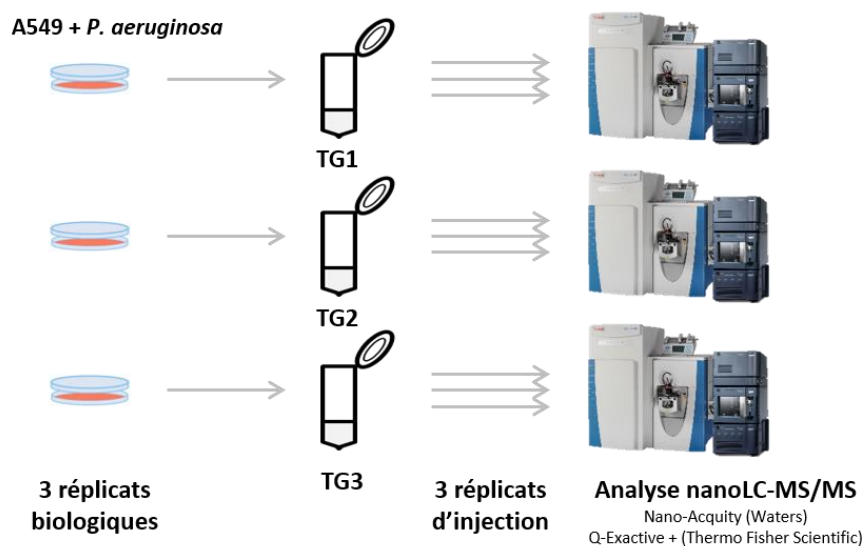


Figure IV-8. Analyse de la reproductibilité biologique et de la répétabilité technique. Trois lots de cellules A549 infectées par *P. aeruginosa* ont été préparés par TG (TG1, TG2, TG3), puis chaque mélange peptidique a été analysé 3 fois (Inj1, Inj2, Inj3) par nanoLC-MS/MS (750 ng injectés).

La Figure IV-9 montre les résultats de cette comparaison, en termes de nombre de protéines humaines et bactériennes identifiées. Nous pouvons d'abord constater que le contrôle qualité mis en place par nos collaborateurs pour s'assurer de la croissance de *P. aeruginosa*, a permis de retrouver un nombre de protéines bactériennes supérieur à 700 protéines.

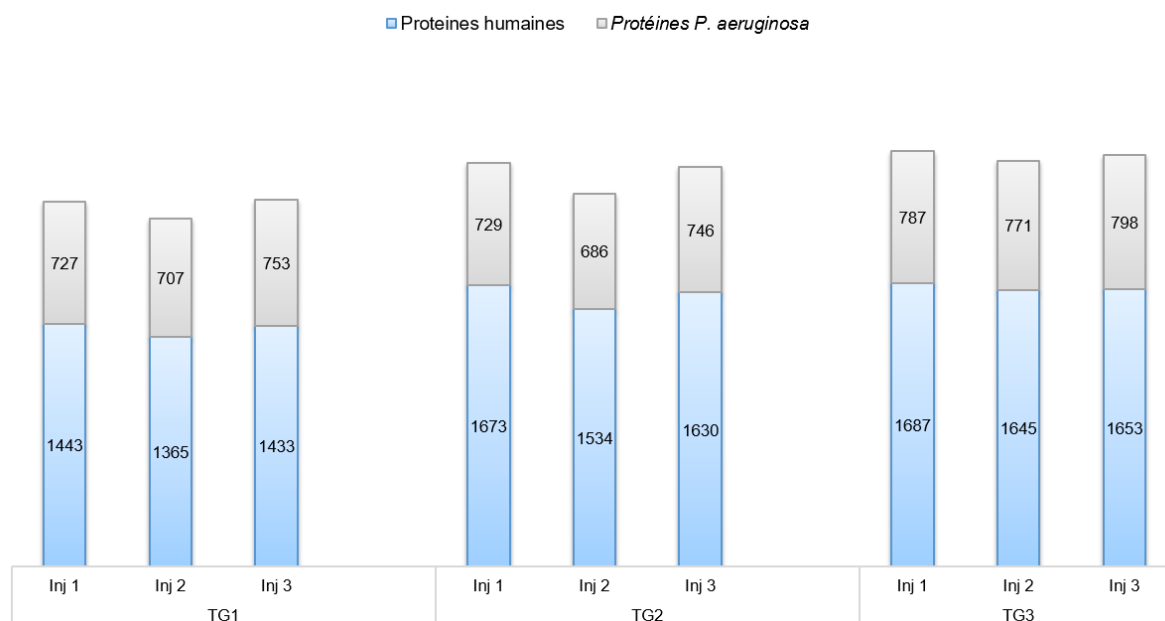


Figure IV-9. Nombre de protéines humaines et de *P. aeruginosa* identifiées dans trois réplicats biologiques et trois réplicats d'injection. Trois lots ces cellules A549 infectées par *P. aeruginosa* ont été préparés par TG (TG1, TG2, TG3), puis chaque mélange peptidique a été analysé 3 fois (Inj1, Inj2, Inj3) par nanoLC-MS/MS (750 ng injectés).

Nous avons évalué la **répétabilité** technique (ou répétabilité analytique) à l'aide des réplicats d'injection (Inj1, Inj2, Inj3) pour chaque TG. Les diagrammes de Venn sur la (Figure IV-10) montrent un recouvrement de 66,8% en moyenne pour les protéines de *P. aeruginosa* identifiées dans les trois réplicats d'injection. Cette valeur est conforme aux résultats classiquement obtenus au laboratoire sur ce couplage analytique, pour des analyses réalisées en mode DDA dont le caractère stochastique a été discuté dans le Chapitre I.III-2.2.

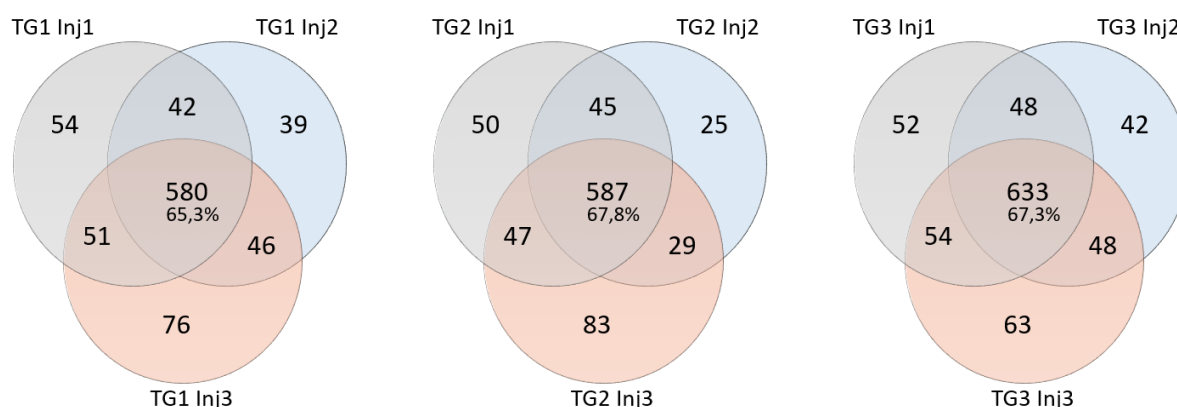


Figure IV-10. Recouvrement des protéines de *P. aeruginosa* identifiées dans 3 réplicats d'injections (Inj1, Inj2, Inj3). A gauche : trois injections issues du TG1 ; au centre : 3 injections issues du TG2 ; à droite : 3 injections du TG3. Les diagrammes de Venn ont été obtenus à partir des nombres de protéines donnés dans la Figure IV-9.

En parallèle, les réplicats biologiques ont permis d'évaluer **la reproductibilité** de la préparation des cellules infectées. Nous avons observé un recouvrement de 66,7% en moyenne pour les protéines de *P. aeruginosa* identifiées dans les trois réplicats biologiques (Figure IV-11). Cette valeur est très satisfaisante car elle englobe, non seulement la préparation des cellules infectées mais également tout le traitement analytique qui suit, y compris le TG et les injections nanoLC-MS/MS.

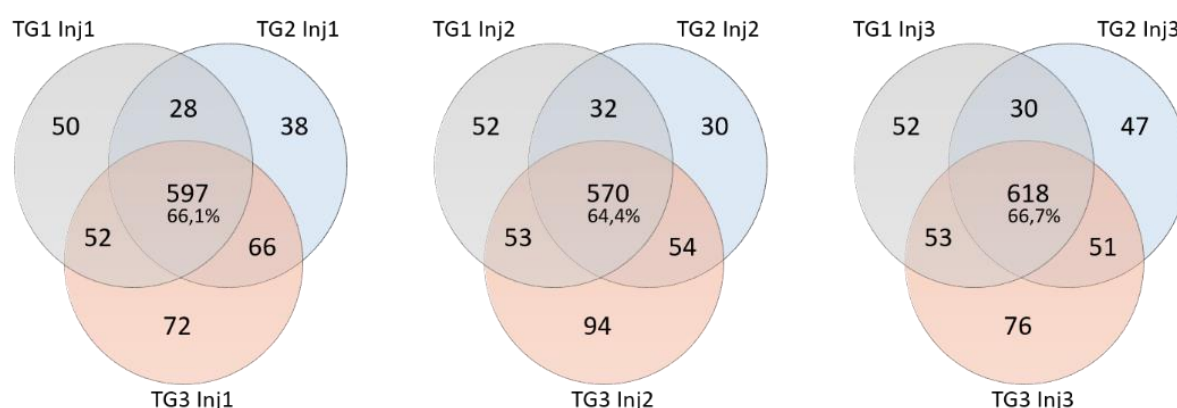


Figure IV-11. Recouvrement des protéines de *P. aeruginosa* identifiées dans 3 réplicats biologiques (TG1, TG2, TG3). A gauche : injections 1 issues des 3 réplicats biologiques ; au centre : injections 2 issues des 3 réplicats biologiques ; à droite : injections 3 issues des 3 réplicats biologiques. Les diagrammes de Venn ont été obtenus à partir des nombres de protéines donnés dans la Figure IV-9.

Nous avons alors décidé de privilégier des réplicats biologiques, qui nous paraissaient plus pertinents que des réplicats d'injection pour l'analyse statistique des études différentielles. En termes de nombre de réplicats, nous avons décidé de travailler avec **5 réplicats biologiques** par condition pour les deux études différentielles de *P. aeruginosa* qui vont maintenant être présentées.

III- Réalisation des études protéomiques différentielles

1. Adaptation phénotypique de *P. aeruginosa* à la présence d'exosidérophores

1.1. Contexte

Comme indiqué au début de ce chapitre, le fer est un nutriment essentiel pour les bactéries. Pour l'acquérir de façon efficace, elles produisent et sécrètent des sidérophores, molécules de faible poids moléculaire, sécrétées par les bactéries dans leur environnement et capables de chélater avec une très haute affinité le fer ferrique (S. Vaulont *et al.*, 2015). Très souvent, les bactéries peuvent produire et sécréter un ou plusieurs sidérophores de structures chimiques différentes, mais elles peuvent aussi utiliser des sidérophores exogènes (**exosidérophores**), produits par d'autres microorganismes, ce qui leur évite de les synthétiser. Il s'agit là d'une stratégie connue sous le nom de piratage de sidérophores. Dans le cas de *P. aeruginosa*, cette bactérie produit deux sidérophores spécifiques, la pyoverdine (PVD) et la pyochéline (PCH), mais elle peut également acquérir du fer en piratant les sidérophores d'autres organismes (P. Cornelis *et al.*, 2013).

Dans ce contexte, nous avons étudié l'adaptation phénotypique de *P. aeruginosa* dans un modèle d'infection (cellules épithéliales pulmonaires infectées), en présence d'exosidérophores produits par des organismes avec lesquels *P. aeruginosa* peut être en contact dans le cas d'une infection.

Les exosidérophores choisis par nos collaborateurs sont la yersiniabactine (YER), le ferrichrome (FERRI), la vibriobactine (VIB) et l'entérobactine (ENT), la mycobactine (MYCO) et la sidéromycine naturelle albomycine (ALBO). Ils ont été choisis car ils présentent tous des

modes distincts de chélation du fer et ils sont produits par des microorganismes qui peuvent partager la même niche écologique que *P. aeruginosa*.

Cette étude pouvait permettre :

- d'avoir une meilleure compréhension de la dynamique inter-espèces existant dans un environnement complexe,
- d'avoir une meilleure vue d'ensemble de l'acquisition du fer chez *P. aeruginosa*,
- de déterminer quel sidérophore serait plus intéressant dans le cadre d'une approche rationnelle de conception de sidéromycines synthétiques.

1.2. Stratégie analytique

Suite à toutes les optimisations présentées précédemment, nous avons pu mettre en œuvre une approche protéomique quantitative *label-free* sur des cellules épithéliales pulmonaires infectées par *P. aeruginosa*, en présence des différents composés.

Tous les échantillons ont été préparés par Q. Perraud (UMR 7242) de la façon suivante. Les bactéries *P. aeruginosa* PAO1 (50×10^6) ont été précultivées en milieu LB (milieu riche en fer). Les cellules épithéliales A549 ont été cultivées en milieu RPMI (milieu riche en fer). Les bactéries précultivées ont été ajoutées aux cellules épithéliales (10^6) en absence ou en présence des différents composés ($10 \mu\text{M}$). Les parties adhérentes des cellules ont été récoltées après 3h d'infection à 37°C . Les culots cellulaires ont été repris dans une solution directement compatible avec la préparation du TG.

Sept échantillons ont été préparés, chacun en 5 réplicats biologiques :

- A549 + *P. aeruginosa* PAO1 : Témoin
- A549 + *P. aeruginosa* PAO1 + ENT $10\mu\text{M}$
- A549 + *P. aeruginosa* PAO1 + VIB $10\mu\text{M}$
- A549 + *P. aeruginosa* PAO1 + MYC0 $10 \mu\text{M}$
- A549 + *P. aeruginosa* PAO1 + FERRI $10 \mu\text{M}$
- A549 + *P. aeruginosa* PAO1 + YER $10\mu\text{M}$
- A549 + *P. aeruginosa* PAO1 + ALBO $10\mu\text{M}$

La Figure IV-12 récapitule le protocole expérimental qui a été appliqué à ces échantillons.

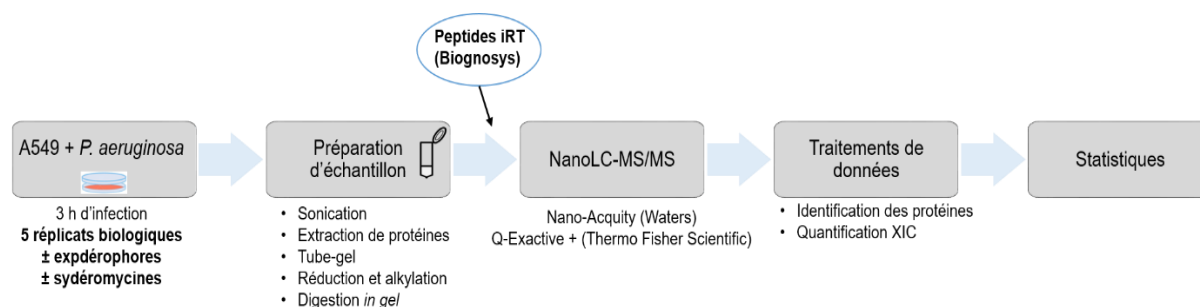


Figure IV-12. Schéma analytique utilisé pour l'étude différentielle des protéines de *P. aeruginosa* dans des cellules épithéliales pulmonaires infectées, en présence des différents exosidérophores. Chaque échantillon a été préparé indépendamment 5 fois, puis traité dans les mêmes conditions par approche TG et analyse nanoLC-MS/MS (gradient 120 min, 750 ng injectés). Des peptides iRT ont été ajoutés avant l'injection nanoLC-MS/MS. La quantification *label-free* a été réalisée en extrayant les signaux MS1 (quantification XIC). La banque de données utilisée pour le traitement des données est une banque générée à partir des protéines humaines et des protéines de *P. aeruginosa* PAO1 de UniProtKB.

A partir de l'extraction des cellules infectées par *P. aeruginosa*, 12,5 µg de protéines ont été préparés en TG puis analysés par nanoLC-MS/MS (750 ng injectés) sur le couplage [nanoAcquity/Q-Exactive+] avec un gradient de 120 min. Les 35 échantillons (sept conditions différentes en cinq réplicats) ont été analysés, dans une seule série d'analyse, selon la séquence [Levure/Cond1R1/Cond2R1/.../Cond7R1/Levure/Cond1R2/Cond2R2/Cond3R2..... Cond7R5/Levure]. De plus les échantillons ont été dopés avec des peptides iRT (*indexed Retention Time*, Biognosys) pour contrôler la stabilité des temps de rétention. Enfin, nous avons inséré des analyses de digestat tryptique de *Saccharomyces cerevisiae* (échantillon de référence au laboratoire) pour suivre les performances du couplage instrumental tout au long de la série d'analyse. Les données ont été traitées à l'aide d'une banque de données constituée des protéines humaines et des protéines de *P. aeruginosa* PAO1 issues de UniProtKB. Les traitements de données seront présentés plus loin dans le paragraphe 1.4.

1.3. Suivi de la stabilité instrumentale

Cette expérience, composée 35 échantillons, représentait un temps d'analyse du couplage [nanoAcquity/Q-Exactive+] de six jours. Nous avons évalué la stabilité du système grâce à des contrôles qualité (CQ). Le digestat de *S. cerevisiae* a été utilisé comme CQ externe, et a été analysé 6 fois à intervalles réguliers. Les iRTs ont été ajoutés à tous les échantillons juste avant l'analyse nanoLC-MS/MS et ont été utilisés comme CQ interne.

1.3.1. Contrôle qualité externe

Pour le digestat de *S. cerevisiae*, nous avons suivi le nombre de protéines identifiées dans les six analyses (Figure IV-13)

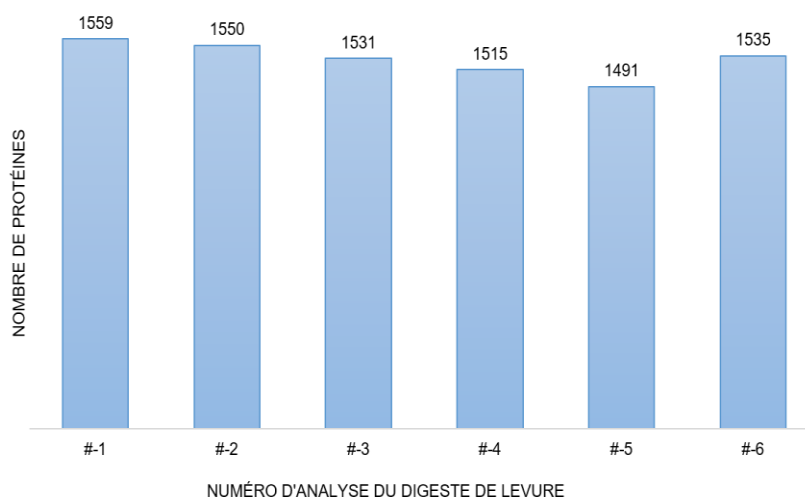


Figure IV-13. Nombre de protéines identifiées dans les 6 injections du digestat de levure *S. cerevisiae* utilisé comme CQ externe. Un digestat de levure a été analysé à six reprises durant la série d'analyse des 35 échantillons de cellules infectées par *P. aeruginosa*.

Nous pouvons constater que le nombre de protéines identifiées est resté stable durant les six jours d'analyse, avec un CV de 2%. Parmi les peptides tryptiques, nous en avons suivi quatre qui servent classiquement au contrôle des performances du couplage [nanoAcquity/Q-Exactive+]. Nous avons suivi les temps de rétention (Figure IV-14, A) et les largeurs à mi-hauteur (Figure IV-14, B). Pour les temps de rétention, les CV étaient très faibles, allant de 0 (peptide 3) à 3%. Pour les largeurs à mi-hauteur, les CV étaient plus importants, allant de 7 à 13%, mais toutes les valeurs obtenues se situaient sous le seuil fixé au laboratoire. Ce CQ externe était donc satisfaisant.

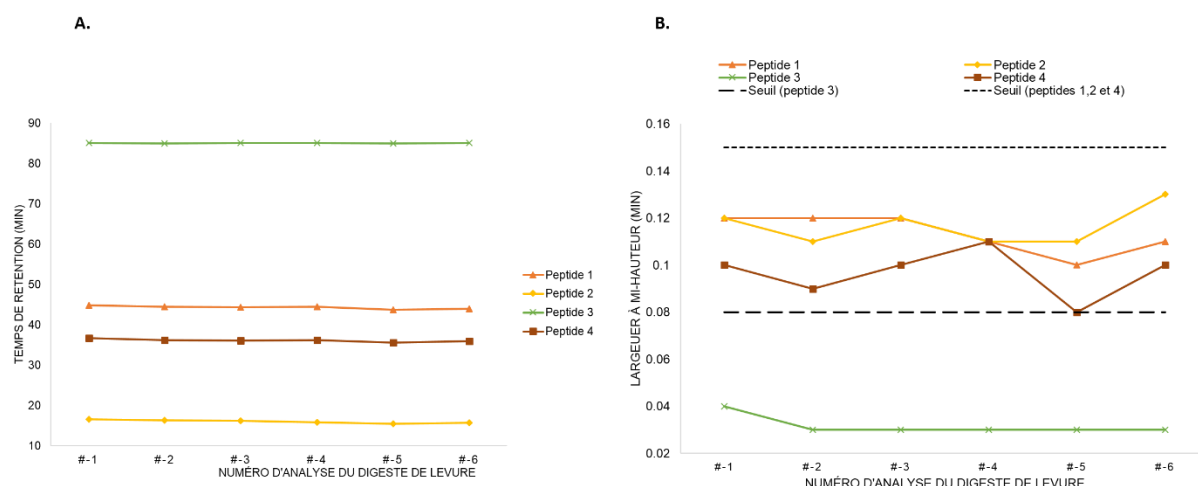


Figure IV-14. Suivi de 4 peptides dans les 6 injections du digestat de levure *S. cerevisiae* utilisé comme CQ externe. A. Suivi des temps de rétention des 4 peptides **B.** Suivi des largeurs à mi-hauteur des pics chromatographiques des 4 peptides Rajouter ici la correspondance N° peptides m/z 878,478 (peptide 1); m/z 580,301 (peptide 2); m/z 1086,544 (peptide 3); m/z 814,383 (peptide 4).

1.3.2. Contrôle qualité interne

Les peptides iRT ont été conçus pour être répartis sur l'ensemble du gradient chromatographique (C. Escher *et al.*, 2012). L'ajout de ces peptides iRT dans tous les échantillons a permis de suivre au plus près la stabilité du système chromatographique. Les CV mesurés sur les temps de rétention de ces 11 peptides sont représentés dans la Figure IV-15, A. La médiane est inférieure à 1%, montrant ainsi que le système chromatographique est resté stable tout au long de la série des 35 échantillons.

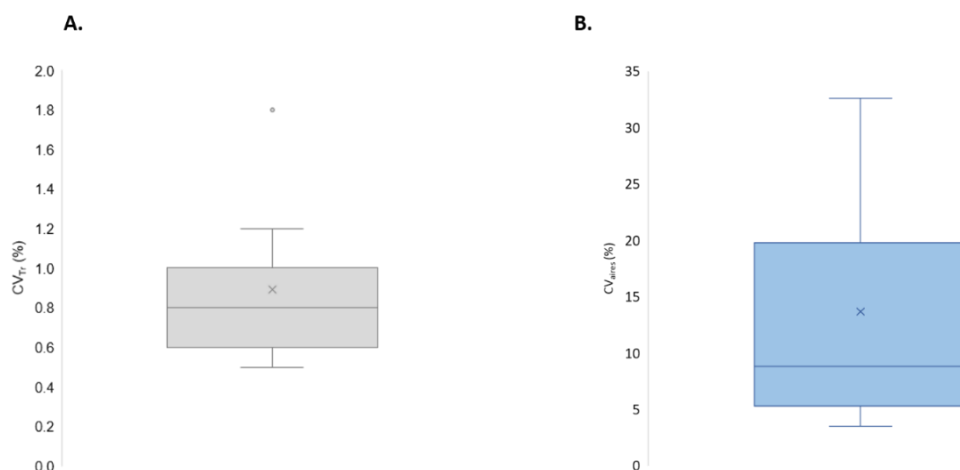


Figure IV-15. Suivi des 11 peptides iRT dans les 35 échantillons de cellules infectées par *P. aeruginosa*. Les peptides iRT ont été ajoutés à tous les échantillons avant l'analyse nanoLC-MS/MS. **A.** Boîte à moustache représentant la distribution des CV sur les temps de rétention des 11 iRT. **B.** Boîte à moustache représentant la distribution des CV sur les aires des 11 iRT.

Les aires et Tr des 11 peptides suivis ont été également extraits à l'aide du logiciel Skyline, ce qui nous a permis d'établir une boîte à moustache représentant les CV sur l'ensemble des peptides. (Figure IV-15, B). Nous avons constaté deux peptides (GTFIIDPAAVIR et LFLQFGAQGSPFLK) présentent des fortes variations, donnant des CV supérieurs à 20%, 27,5% et 32,6% respectivement. La valeur médiane, est quant à elle inférieure à 13,7%. Ces deux peptides ont présenté le même comportement pour d'autres projets au laboratoire, nous avons donc décidé d'ôter ces deux peptides. Ainsi notre nouvelle médiane avec les neuf peptides iRT restant était de 10,2%, résultat satisfaisant, en regardant l'ensemble des contrôles qualité. Nous pouvons donc conclure que le couplage est resté stable tout au long de la série des 35 échantillons.

Grâce à ces différents CQ (externe et interne) qui ont montré la répétabilité et la robustesse du couplage nanoLC-MS/MS, nous avons pu effectuer la quantification XIC des protéines de *P. aeruginosa* dans ces 35 échantillons.

1.4. Résultats

En combinant les identifications obtenues dans les 35 échantillons, ce sont plus de 2500 protéines de *P. aeruginosa* qui ont pu être identifiées, avec une moyenne de 830 protéines de *P. aeruginosa* (Figure IV-16.) Ce nombre est comparable à celui obtenu lors des optimisations, par exemple dans la Figure IV-9 de ce chapitre.

Le nombre de protéines de *P. aeruginosa* identifiées est variable (entre 528 et 1114) selon l'échantillon. Nous pouvons néanmoins constater une bonne reproductibilité biologique avec des CV compris entre 9 et 14% pour les 5 réplicats, excepté pour FERRI qui présente un CV de 20%.

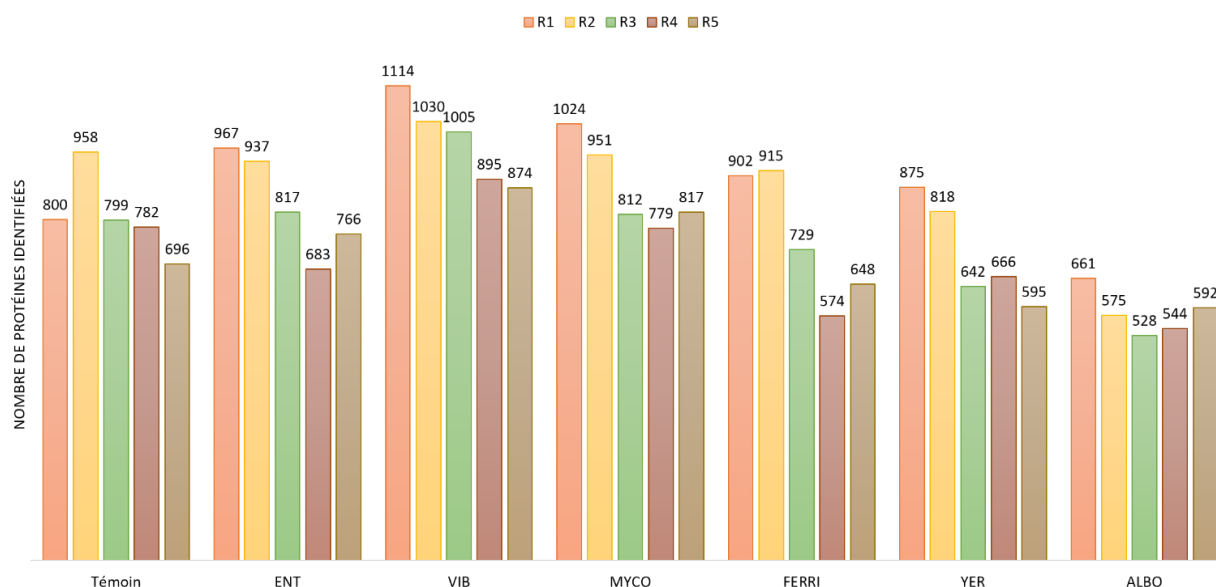


Figure IV-16. Nombre de protéines de *P. aeruginosa* identifiées dans les cellules infectées par *P. aeruginosa* en présence des différents exosidérophores. La condition Témoin correspond aux cellules infectées par *P. aeruginosa* en absence de d'exosidérophore. Les autres conditions correspondent à la présence des différents exosidérophores (10 μ M). Le protocole analytique est celui décrit dans la Figure IV-12. R1, R2, R3, R4, R5 correspondent aux 5 réplicats biologiques pour chaque condition.

Pour les analyses différentielles, nous avons comparé deux à deux les échantillons (présence exosidérophore *versus* témoin), ce qui a généré six comparaisons (ENT vs Témoin ; VIB vs Témoin ; MYCO vs Témoin ; FERRI vs Témoin ; YER vs Témoin ; ALBO vs Témoin). Un premier traitement (Traitement 1) a été effectué avec **Andromeda**, algorithme de recherche intégré au logiciel **MaxQuant** (J. Cox *et al.*, 2008) utilisé pour effectuer la normalisation et la quantification. Les comparaisons ont été faites à l'aide du logiciel **Perseus** (S. Tyanova *et al.*, 2016), utilisé couramment en protéomique quantitative. Le nombre de protéines quantifiées après normalisation pour les six comparaisons est donné dans la Figure IV-17.

Nous pouvons constater un nombre de protéines de *P. aeruginosa* plus important, grâce au paramétrage des logiciels utilisés pour la quantification. Ce paramétrage permet de réaliser un transfert de l'identification d'un échantillon aux autres échantillons en utilisant les m/z, les

Tr et la normalisation. En d'autres termes, il est possible d'extraire le courant d'ions d'un peptide dans une analyse dans lequel il n'a pas été fragmenté, permettant ainsi l'identification du peptide. Ce terme est nommé « *Match between run* » pour le logiciel **Maxquant** et « cross-assignement » dans le logiciel **Proline**.

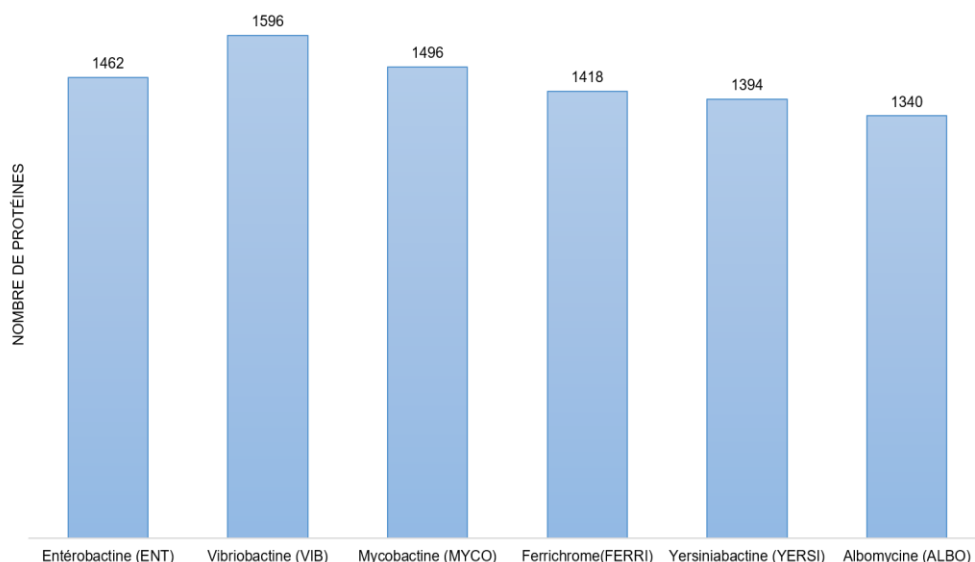


Figure IV-17. Nombre de protéines de *P. aeruginosa* quantifiées dans les cellules infectées par *P. aeruginosa* en présence des six exosidérophores. Le nombre de protéines quantifiées pour chaque comparaison a été obtenu après traitement Andromeda/MaxQuant/Perseus (Traitement 1).

A partir des données de quantification MS1, nous avons effectué un test statistique t de Welch pour établir la liste des protéines dont l'abondance varie dans chaque condition. Cela nous a permis d'obtenir six Volcano plot, dont un exemple est montré dans la Figure IV-18. Les protéines différentiellement exprimées ($p < 0,05$) dans le Volcano plot ont été examinées par nos collaborateurs, et les protéines impliquées dans les voies d'acquisition du fer ont fait l'objet d'une attention particulière.

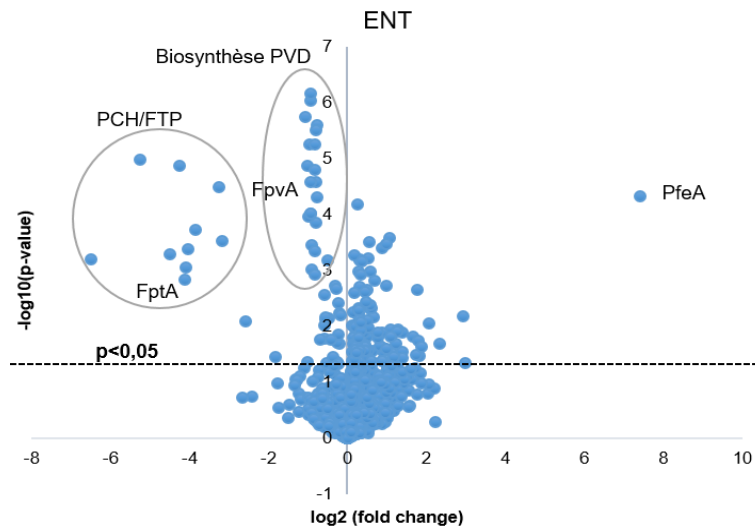


Figure IV-18. Volcano plot représentant les variations d'abondance des protéines de *P. aeruginosa* dans des cellules infectées en absence (Témoin) ou en présence de 10 μ M ENT. La statistique a été faite avec un test t de Welch. Sur les 1462 quantifiées, 195 sont différentiellement exprimées ($p < 0,05$). Les cercles indiquent les protéines impliquées dans les voies pyochéline (PCH) et pyoverdine (PVD).

Sur l'exemple de la Figure IV-18 qui montre la comparaison ENT vs Témoin, 195 protéines sont différentiellement exprimées ($p < 0,05$) parmi les 1462 protéines identifiées après normalisation. On observe une surexpression du transporteur de fer spécifique à l'ENT (**PfeA**) ($\log_2$ fold change = 7,4), concomitante avec une répression des voies d'acquisition du fer utilisant les sidérophores endogènes produits par *P. aeruginosa*, la pyoverdine PVD (**FpvA**) ou la pyochéline PCH (**FptA**). Nos collaborateurs ont ensuite confirmé ces résultats au niveau transcriptionnel par RT-qPCR en étudiant l'expression des gènes codants les protéines impliquées dans ces voies (Figure IV-19).

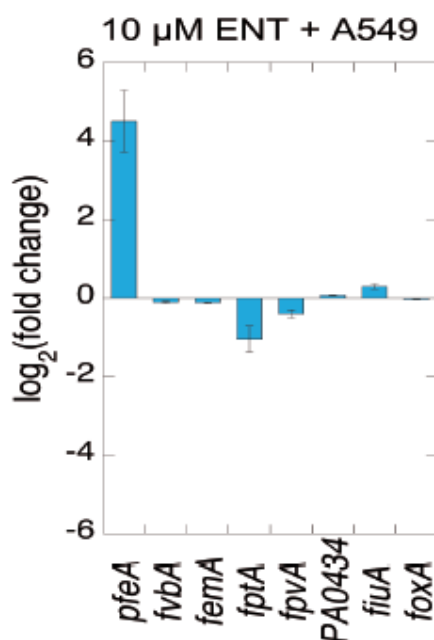


Figure IV-19. Histogramme représentant les variations d'abondance des gènes codant des protéines de *P. aeruginosa* dans des cellules infectées en présence de 10 μM ENT. La RT-qPCR a été réalisée sur des cellules A549 infectées par *P. aeruginosa* après 3 d'incubation. Les données ont été normalisées avec le gène de référence *uvrD* pour les triplicats biologiques.

En présence de vibriobactine (VIB), 321 (Tableau IV-2) protéines sont différentiellement exprimées ($p < 0,05$), avec une surexpression du transporteur spécifique de VIB (**FvbA**) et de deux protéines impliquées dans le transport du fer (**FemA** et **PiuA**), concomitante à une répression de la voie pyochéline.

En présence de ferrichrome (FERRI), 104 protéines sont différentiellement exprimées ($p < 0,05$) (Tableau IV-2), avec une surexpression de **FiuA**, un transporteur spécifique des sidérophores de type hydroxamates tels que FERRI, concomitante à une répression de la voie pyochéline.

En présence de yersiniabactine (YER), 261 protéines sont différentiellement exprimées ($p < 0,05$) (Tableau IV-2), avec une surexpression de trois transporteurs du fer (**FemA**, **OprC** et **PA0434**) et une répression de la voie pyoverdine.

En présence d'albomycine (ALBO), 306 protéines sont différentiellement exprimées ($p < 0,05$), avec une surexpression de **FiuA** et **HasR** (transporteur d'hème) et une répression des voies pyochéline et pyoverdine.

En présence de mycobactine (MYCO), 156 protéines sont différentiellement exprimées ($p < 0,05$), avec une surexpression de **FemA** et une répression de la voie pyoverdine.

De manière générale, les résultats de protéomique différentielle ont été confirmés au niveau transcriptionnel par RT-qPCR, et nos collaborateurs ont pu réaliser des expériences complémentaires en mélangeant par exemple des exosidérophores et en observant une induction de transporteurs de catéchols plus importante que celle des transporteurs d'hydroxamate.

L'ensemble de ces résultats ont montré une adaptation complexe de *P. aeruginosa* à la présence des exosidérophores qui ne se limite pas à surexpression du transporteur correspondant. La biosynthèse des sidérophores pyochéline et de pyoverdine de *P. aeruginosa* est délaissée lorsque des sidérophores « gratuits » sont déjà présents dans le milieu. L'approche protéomique a également permis de mettre en évidence des variations d'expression d'autres protéines que les transporteurs TonB-dépendant, impliquées également dans l'acquisition du fer, comme par exemple les oxydoréductases prédites PiuB ou PA4155. Nous avons aussi pu observer des variations d'expression de transporteurs peu caractérisés comme OprC ou PA0434., ce qui peut aider à comprendre le rôle de ces protéines dans l'acquisition du fer.

Dans le cadre de l'implémentation récente de la quantification XIC dans Proline au laboratoire (D. Bouyssie *et al.*, 2020), les données obtenues pour quatre exosidérophores (ENT, FERRI, VIB, YER) ont été soumises à un deuxième traitement (Traitement 2), en utilisant l'algorithme de recherche Mascot pour l'identification des protéines, **Proline** pour la validation et la quantification XIC, et **ProStar** (S. Wiczorek *et al.*, 2017) pour les traitements statistiques. Nous avons ici comparé des schémas de traitement de données bien distincts (traitement 1 et traitement 2), et non les logiciels deux à deux. Nous avons constaté que les nombres de protéines différentielles étaient très variables selon le traitement appliqué aux mêmes données brutes (Tableau IV-2).

Tableau IV-2. Comparaison du nombre de protéines différentiellement exprimées ($p < 0,05$) selon le traitement de données utilisé. **Traitement 1** : Andromeda/MaxQuant / Perseus ; **Traitement 2** : Mascot/Proline/Prostar.

	Protéines différentiellement exprimées ($p < 0,05$)			
	ENT	VIB	FERRI	YER
Traitement 1	195	321	104	261
Traitement 2	238	394	101	413

Comme nous pouvons constater, les résultats sont échantillon-dépendants. Si nous prenons l'exemple d'ENT, un cœur commun de 35% a été observé entre les deux traitements. Dans le cas de FERRI, un cœur commun de 26% a été observé malgré le nombre proche de protéines différentielles obtenus avec les deux traitements. Cependant, nous avons constaté que ce sont majoritairement les protéines proches du seuil de p-value (proches de 1,3 en log de p-value ($p < 0,05$)) qui sont en dehors du cœur commun. Cette faible différence est rassurante, car elle suggère que ce sont des variations de traitement qui font basculer une protéine en-dessous ou au-dessus de la valeur seuil fixée.

Cependant, ces différences n'ont eu aucune influence sur la variation des protéines impliquées dans les voies d'acquisition du fer. Nous avons retrouvé les surexpressions des transporteurs spécifiques aux exosidérophores ajoutés dans le milieu et les répressions des voies pyochéline et pyoverdine de *P. aeruginosa*.

Les volcano plot issus du traitement 2 sont présentés dans la Figure IV-20. Nous avons réduit le nombre de protéines différentiellement exprimées en abaissant le seuil de la p-value ($p < 0,001$). Les FDR obtenus sont 2,34% pour ENT, 1,97% pour VIB, 1,96% pour YER et 4,24% pour FERRI.

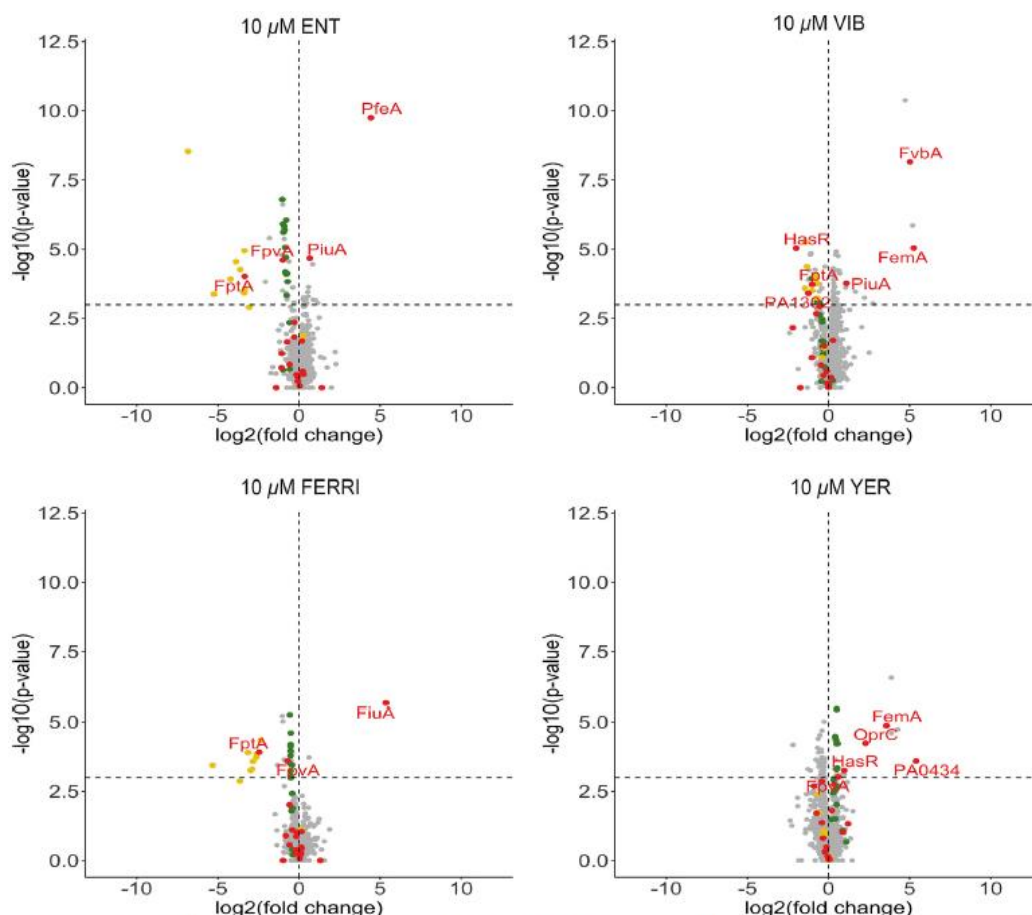


Figure IV-20. Volcano plot représentant les variations d'abondance des protéines de *P. aeruginosa* dans des cellules infectées en absence (Témoin) ou en présence de quatre exosidérophores (ENT, VIB, FERRI, YER). Les données ont été traitées avec Mascot/Proline/ProStar (traitement 2) et les statistiques ont été faites avec un test t de Welch. En vert les protéines de la voie pyoverdine, en orange les protéines de la voie pyochéline

Si nous considérons les volcano plot obtenus pour l'ENT avec les deux traitements (Figure IV-18 et Figure IV-20), nous pouvons constater que la protéine PfeA est surexprimée en présence d'ENT quel que soit le traitement, mais que les facteurs de variation et les p-values sont différents pour le traitement 1 (log de FC = 7,4 et log de p-value de 4,3) et pour le traitement 2 (log de FC = 4,4 et log de p-value de 9,8).

Ces résultats ont fait l'objet d'une publication : *Phenotypic adaption of Pseudomonas aeruginosa by hacking siderophores produced by other microorganisms*, Perraud, Q, Cantero, P., Roche, B., Gasser, V., Normant, V. P., Kuhn, L., Hammann, P., Mislin, G. L. A., Ehret-Sabatier, L., Schalk, I. J. Mol. Cell. Proteomics (2020), 19(4), 589-607.

En ce qui concerne l'utilisation des logiciels, le fait que Proline soit développé en partie par notre laboratoire, permet d'avoir une interaction directe avec les développeurs pour résoudre des problèmes ou proposer des idées d'amélioration. Quant aux logiciels de traitement statistique, Perseus a été développé pour l'interprétation de données pour la quantification, et pour les PTMs. ProStar propose des outils de représentation graphique des données à chaque étape de l'analyse statistique. Globalement les deux logiciels permettent d'obtenir des résultats similaires, lorsque le même traitement est appliqué. Selon moi, ProStar, présente l'avantage de pouvoir effectuer des choix plus adaptés aux données en cours de traitement, d'autant plus que ce logiciel a été développé pour la protéomique et ne propose que des traitements statistiques adaptés à ce type de jeux de données. Prostar a été développé par des bio-informaticiens et des statisticiens de manière à rendre les statistiques plus accessibles aux protéomistes.

2. Adaptation phénotypique de *P. aeruginosa* à la présence de différentes conditions

2.1. Contexte

Après avoir montré l'adaptation phénotypique de *P. aeruginosa* à la présence d'exosidérophores, une seconde étude protéomique différentielle a été effectuée avec plusieurs autres composés ayant un lien avec les voies d'acquisition du fer et une autre souche de *P. aeruginosa* (**Delta PfeA**) n'exprimant pas le transporteur PfeA (transporteur spécifique de l'ENT). Dans cette série ont été inclus : un conjugué sidérophore-antibiotique (**TCV-L6**), synthétisé par nos collaborateurs, et ayant comme partie chélatrice du fer un analogue catécholé de l'entérobactine et comme partie antibiotique le linezolid ; le **TCV**, vecteur catécholé du composé TCV-L6 ; la **transferrine**, protéine humaine de stockage du fer ; et finalement le **2,3 DHBS** produit de dégradation de l'entérobactine par l'estérase PfeE (Q. Perraud et al., 2018).

2.2. Stratégie analytique

L'analyse protéomique quantitative a été mise en œuvre comme précédemment décrit sur des cellules épithéliales pulmonaires A549 infectées par delta PfeA ou par *P. aeruginosa* PAO1 en présence ou non des différents composés. Les cellules A549 et les bactéries ont été préparées comme précédemment.

Sept échantillons ont été préparés, chacun en 5 réplicats biologiques :

- A549 + *P. aeruginosa* PAO1 : Témoin
- A549 + *P. aeruginosa* Delta PfeA
- A549 + *P. aeruginosa* PAO1 + ENT 10µM
- A549 + *P. aeruginosa* PAO1 + 2,3-DHBS 10µM
- A549 + *P. aeruginosa* PAO1 + TCV 10µM
- A549 + *P. aeruginosa* PAO1 + TCV-L6 10µM
- A549 + *P. aeruginosa* PAO1 + Transferrine 20µM

Les deux échantillons Témoin et ENT ont été re préparés de manière indépendante à la première série d'analyse présentée précédemment.

Ces échantillons ont été traités en suivant exactement le même protocole que la 1^{ère} étude (Figure IV-12). Les contrôles qualité ont été les mêmes que précédemment. Un CV de 2% a été calculé pour le nombre de protéines de *S. cerevisiae* identifiées dans les 6 injections de levure ; un CV médian inférieur à 1% a été obtenu pour les temps de rétention des peptides iRT, et un CV de 10% sur les aires des peptides iRT. L'analyse quantitative a été réalisée à partir de l'extraction des signaux MS1 (quantification XIC). Les analyses différentielles ont été réalisées en comparant chaque condition au témoin. Les données ont été interprétées avec Mascot pour l'identification des protéines, Proline pour la validation et la quantification XIC, et ProStar pour les traitements statistiques avec test t de Welch.

2.3. Résultats

En termes d'identification, plus de 2500 protéines totales ont pu être identifiées avec une moyenne de 630 protéines de *P. aeruginosa* par échantillon.

Une comparaison du nombre de protéines de *P. aeruginosa* a été effectuée entre les deux échantillons communs entre les deux études (Figure IV-21). De manière rassurante, les nombres de protéines identifiées sont comparables. Ceci confirme la reproductibilité de la préparation des cellules infectées mise au point par nos collaborateurs. Toutefois il est important de toujours remettre un témoin dans la même série d'analyse pour effectuer la comparaison d'une condition donnée à un témoin analysé dans la même série.

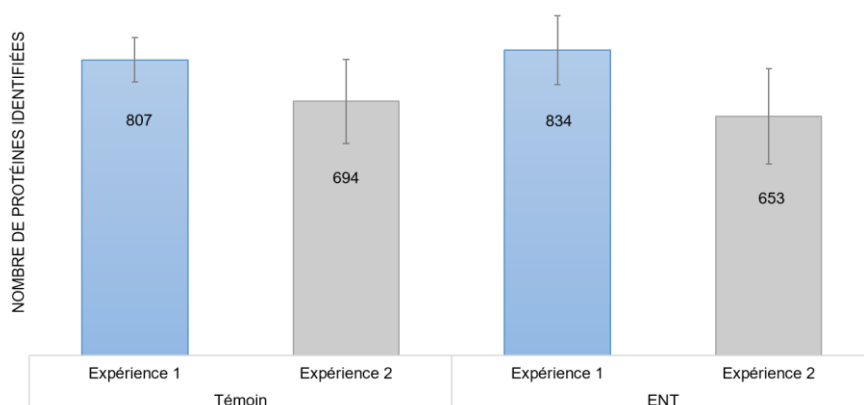


Figure IV-21. Nombre moyen de protéines de *P. aeruginosa* identifiées dans les cellules infectées en absence (Témoin) et en présence d'ENT (10 µM) dans les deux expériences de protéomique quantitative *label-free*. Les échantillons pour les deux expériences ont été préparés de manière indépendante à plusieurs mois d'intervalle. Ils ont été analysés en suivant le même protocole analytique décrit dans la Figure IV-12.

Après normalisation, nous avons obtenu un total de 1163 protéines différentiellement exprimées dans cette série d'analyse.

En présence d'ENT, nous avons identifié 48 protéines différentiellement exprimées avec un FDR de 2,35% ($p < 0,001$), ce qui est en accord avec les résultats obtenus pour lors de la 1^{ère} étude (FDR 2,34% ($p < 0,001$), pour 58 protéines différentielles) (Tableau IV-3).

Tableau IV-3. Nombre de protéines différentielles et FDR obtenus pour la comparaison ENT vs Témoin lors des deux études différentielles avec $p < 0,001$ - Benjamini-Hochberg. Le nombre total de protéines correspond au nombre de protéines obtenu après la normalisation entre les échantillons.

Entérobactine (ENT)				
	Total protéines	Protéines différentielles	p-value	FDR
Première étude	1470	58	0,1%	2,34
Deuxième étude	1163	48	0,1%	2,35

Ces deux études ont un cœur commun de 30 protéines (40%). Parmi ces protéines communes nous retrouvons les protéines impliquées dans les voies d'acquisition du fer. Dans les deux cas, on observe une surexpression du transporteur de fer spécifique à l'ENT (PfeA), ainsi qu'une répression des voies d'acquisition du fer utilisant les sidérophores endogènes produits par *P. aeruginosa*, la pyoverdine PVD (FpvA) ou la pyochéline PCH (FptA) (Tableau IV-4).

Nous pouvons observer que les facteurs de variation (log fold change) et les valeurs de p-value sont très similaires pour ces trois protéines.

Tableau IV-4. Comparaison des trois protéines différenciellement exprimées en présence d'ENT dans les deux études différentielles. Les facteurs de variation (log FC) et valeurs de p-value sont indiqués pour les deux études indépendantes réalisées sur des cellules infectées par *P. aeruginosa* en présence de 10 μ M ENT.

Protéine	Première étude		Deuxième étude	
	log FC	log p-value	log FC	log p-value
PfeA	4,4	9,8	3,4	7,9
FptA	-3,4	4,0	-4,1	4,5
FpvA	-1,0	4,6	-1,2	3,4

Comme observé précédemment pour les deux traitements 1 et 2, les protéines non communes aux deux études se trouvent très proches du seuil sélectionné ($p < 0,001$).

Pour trois conditions, contenant le 2,3 DHBS, la transferrine et la delta pfeA, aucune protéine impliquée dans les voies d'acquisition du fer n'a été trouvée comme différenciellement exprimée. En revanche, des protéines impliquées dans les voies d'acquisition du fer ont été trouvées comme différentielles en présence d'ENT, de TCV et de TCV-L6. Ces trois composés ont en commun un motif tris-catéchol. Nous avons retrouvé la surexpression de **PfeA**, transporteur spécifique de l'ENT, déjà observée lors de la précédente étude, ainsi que la répression des voies pyochéline (**FptA**) et pyoverdine (**FpvA**) (Figure IV-22).

En présence du vecteur TCV et du conjugué sidérophore-antibiotique TCV-L6, 46 et 14 protéines différenciellement exprimées ont été filtrées avec un FDR de 2,26% et 7,33% ($p < 0,001$), respectivement (Figure IV-22).

Le nombre faible de protéines différenciellement exprimées en présence de TCV-L6 est dû à la p-value très stringente ($p < 0,001$). Il est possible d'augmenter la valeur de p-value (0,01) pour obtenir un nombre de protéines différentielles plus important, mais cela entraîne l'augmentation du FDR. Les deux possibilités sont représentées dans la Figure IV-22. Malgré le plus faible nombre de protéines différentielles avec un FDR de 0.1%, nous avons considéré que ce choix assure une meilleure fiabilité des résultats et permet toujours de montrer la régulation de protéines impliquées dans l'acquisition du fer.

En présence de TCV-L6, nous pouvons observer la répression des voies pyochéline et pyoverdine et la surexpression de PfeA, ce qui est conforme aux homologues de structure entre l'entérobactine, le vecteur catécholé TCV et le conjugué TCV-L6. Nos collaborateurs ont confirmé ces résultats au niveau transcriptionnel par RT-qPCR, ainsi que par des expériences de croissance de *P. aeruginosa* et de transport du fer avec des souches mutantes.

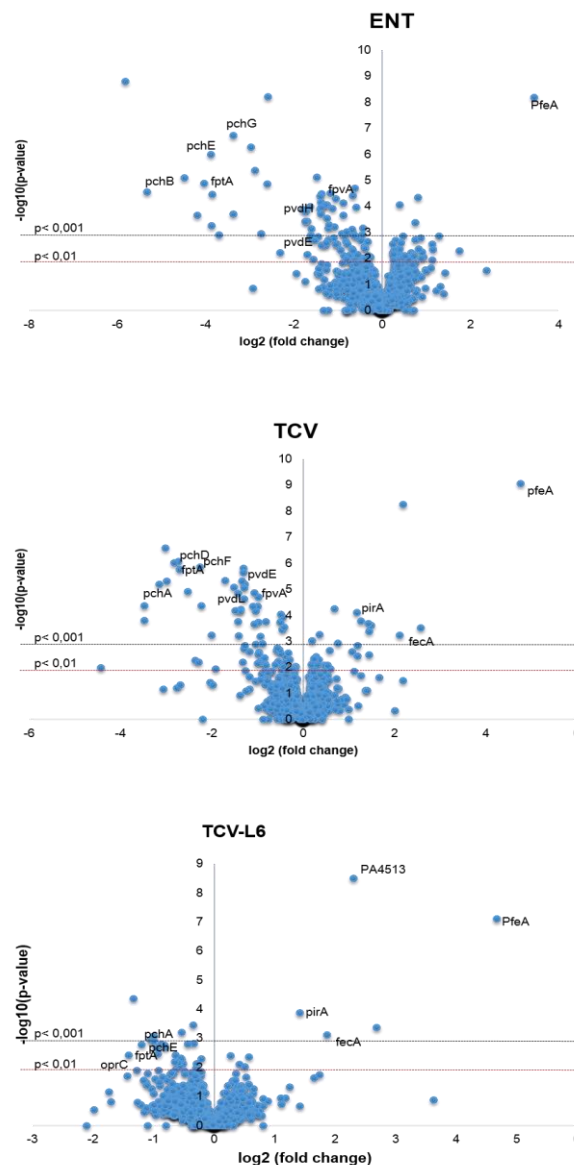


Figure IV-22. Volcano plot représentant les variations d'abondance des protéines de *P. aeruginosa* dans des cellules infectées en absence ou en présence de ENT, TCV, TCV-L6. Les données ont été traitées avec Mascot/Proline/ProStar et les statistiques ont été faites avec un test t de Welch. Les p-value à 0,1% (noir) et 1% (rouge) sont représentées.

En considérant l'ensemble des résultats obtenus dans les deux études, pour la sidéromycine synthétique TCVL6 et pour la sidéromycine naturelle albomycine (ALBO), et les expériences complémentaires des collaborateurs, nous avons pu en conclure que :

- Le fer présente une plus forte affinité pour les sidérophores de type catécholés,
- Les conjugués sidérophores-antibiotiques tels que TCV-L6 sont de bons candidats pour la stratégie du cheval de Troie.

Ces résultats viennent d'être soumis pour publication à PLOS Pathogens *Phenotypical adaptation of Pseudomonas aeruginosa in the presence of siderophore-antibiotic conjugates during epithelial cell infection*, Perraud, Q., Cantero, P., Munier, M., Hoegy, F., Zill, N., Gasser, V., Mislin, G. L. A., Ehret-Sabatier, L., Schalk, I. J.

IV- Conclusion

En conclusion, les différentes optimisations ont permis de se placer dans de très bonnes conditions pour réaliser les deux séries d'analyses différentielles *label-free* avec quantification MS1. Une adaptation de la banque de données a permis d'augmenter le nombre d'identification de protéines d'intérêt. La sélection du TG comme protocole de préparation d'échantillon a permis de mettre en place un protocole robuste et compatible avec des séries de 35 échantillons. Les contrôles de qualité externe et interne ont assuré la fiabilité des résultats tout au long des séries d'analyses *label-free*.

réplicats Nous avons ainsi pu identifier et quantifier les protéines de *P. aeruginosa* en contexte d'infection, au sein de cellules épithéliales pulmonaires humaines. Les différentes analyses différentielles ont permis :

- de mettre en évidence l'adaptation phénotypique de *P. aeruginosa* à son environnement,
- de montrer que *P. aeruginosa* est capable d'utiliser préférentiellement des exosidérophores, au lieu de ses sidérophores endogènes, pour l'acquisition du fer essentiel à sa croissance,
- de montrer que les conjugués sidérophore-antibiotiques tels que TCV-L6 sont de bons candidats pour la stratégie du cheval de Troie.

Ces travaux ont également été l'occasion de comparer différents logiciels de traitement des données, d'identifier leurs avantages et inconvénients, et de mieux appréhender les traitements statistiques sous-jacents.

Conclusion générale

Au travers des différents projets sur lesquels j'ai travaillé, nous avons démontré que nous étions capables de détecter des protéines de différents pathogènes, présentes en très faible quantité dans des tissus ou cellules infectés.

Cette détection a été possible grâce à de **multiples optimisations** : au niveau de la préparation d'échantillon, tout particulièrement pour l'étude de *P. aeruginosa* ; au niveau du développement des méthodes LC-MS/MS, LC-SRM et LC-PRM ; au niveau du traitement de données avec la création de banques protéiques adaptées à l'étude simultanée de plusieurs agents pathogènes et la comparaison de différents logiciels de traitement statistique des données d'analyse différentielle *label-free*.

Dans le cadre du projet vaccinal contre la borréliose de Lyme, mes travaux de thèse ont permis de réaliser le WP1 du projet de maturation et d'identifier des protéines de *Borrelia* dans la peau de chiens infectés. Cela a permis de choisir les protéines à produire par voie recombinante, et de tester l'efficacité d'une formulation vaccinale innovante contenant trois protéines bactériennes. Même si cet essai vaccinal n'a pas été un succès, l'ensemble des approches analytiques mises en œuvre ont montré leur intérêt et leur efficacité dans la détection d'antigènes potentiels au sein de biopsies cutanées d'hôtes infectés.

Dans le cadre du projet de diagnostic direct de la borréliose de Lyme par spectrométrie de masse, mes travaux de thèse ont permis de valider l'approche de protéomique ciblée à l'échelle d'une cohorte de 70 patients au stade précoce. La détection de protéines de *Borrelia* par protéomique ciblée a été comparée aux méthodes actuelles de diagnostic direct : PCR et observation des bactéries au microscope à fond noir. La sensibilité de la protéomique ciblée s'est avérée meilleure que les deux autres méthodes. Nous avons également montré la complémentarité de la LC-SRM et LC-PRM et l'intérêt du multiplexage, c'est-à-dire du ciblage simultané de plusieurs protéines de *Borrelia* au sein d'une seule méthode.

Cette étude de la cohorte humaine ne s'est pas faite sans avoir rencontré plusieurs problèmes, et les optimisations réalisées pour cette étude seront utiles pour le projet de diagnostic de la phase tardive disséminée de la borréliose de Lyme, qui représente à l'heure actuelle la difficulté principale en termes de diagnostic.

En ce qui concerne cette phase tardive disséminée, nous avons optimisé une méthode ciblée des protéines de *Borrelia* dans la peau de souris après réactivation des bactéries par un dermocorticoïde, et nous avons montré qu'elle était applicable à la peau humaine. Nous avons également évalué différents dermocorticoïdes et l'impact de l'anesthésique appliqué sur la peau au moment du prélèvement. Une étude clinique sur quelques peaux de patients devrait être réalisée prochainement, ainsi que l'analyse d'autres prélèvements biologiques tels que le liquide synovial et le liquide céphalo-rachidien.

Nous nous sommes également intéressés aux autres maladies vectorielles à tique, et nous avons étendu notre approche protéomique à d'**autres agents pathogènes transmis par les tiques**. Mes travaux de thèse ont permis d'identifier les pathogènes présents dans les tiques dans leur milieu naturel, et ceux qui sont transmis à l'hôte et persistent dans la peau. Ces résultats ont été obtenus grâce à un travail important sur les banques de données essentielles pour les recherches bio-informatiques. Nous avons sélectionné des protéines pour le développement d'une méthode diagnostique multiplexée, en ciblant plusieurs agents infectieux qui peuvent être pathogènes pour l'homme.

Enfin, nous avons mis à profit la sensibilité de la protéomique pour étudier ***P. aeruginosa*** au sein de cellules épithéliales pulmonaires, et comprendre comment cette bactérie s'adapte à son environnement. Après avoir optimisé la préparation des échantillons, nous avons réalisé plusieurs analyses différentielles, et nous avons montré que *P. aeruginosa* module l'expression de ses différentes voies d'acquisition du fer en fonction des composés présents dans le milieu extérieur. Ces études ont contribué à montrer que des conjugués sidérophore-antibiotiques tels que TCV-L6 sont de bons candidats pour le développement de nouveaux antibiotiques contre *P. aeruginosa*.

Dans leur ensemble, mes travaux de thèse ont montré comment la protéomique peut contribuer au diagnostic et à la lutte contre des maladies infectieuses. De par sa sensibilité et spécificité, la protéomique ciblée apparaît comme très prometteuse pour le diagnostic de ces maladies en détectant directement des protéines du(des) pathogène(s) dans des tissus ou fluides de l'hôte. La validation clinique de ces approches nécessite une collaboration étroite entre scientifiques de différents horizons (chimistes, biologistes, bio-informaticiens, cliniciens, etc).

Malgré les résultats positifs obtenus pour chaque projet, plusieurs limitations persistent et des améliorations sont envisageables.

Pour le **diagnostic de la borréliose de Lyme**, la problématique principale reste la faible quantité des protéines de *Borrelia* par rapport aux protéines majoritaires présentes dans l'échantillon. Le couplage d'un immuno-enrichissement des protéines ou peptides cibles avec la spectrométrie de masse pourrait permettre d'améliorer la détection des protéines bactériennes. La déplétion des protéines majoritaires est aussi une piste d'amélioration, et elle est actuellement à l'étude pour du liquide synovial de patient.

Le préfractionnement des protéines par SDS-PAGE avant les analyses ciblées est une étape chronophage qui multiplie le nombre d'analyses par spectrométrie de masse. Une simplification de ce protocole sera nécessaire pour une application à grande échelle de cette méthode diagnostique.

L'utilisation du mode d'acquisition DIA s'est révélée très prometteuse et mérite d'être poursuivie.

Le traitement de données pour la quantification ciblée n'est pas complètement automatisé, ce qui implique une validation manuelle et augmente les risques d'erreurs. Automatiser ce traitement est d'autant plus nécessaire dans le cas d'un mode d'acquisition qui conduit à l'obtention d'un nombre très élevé de données. (DIA>PRM>SRM).

Pour l'**étude des pathogènes transmis par les tiques**, la difficulté principale réside dans les banques de données, leur taille variable selon le pathogène, et leurs redondances.

Pour l'étude de *P. aeruginosa*, nous avons constaté que le traitement de données est un point clé et pas facilement automatisable. Les critères sélectionnés pour la validation ou le traitement statistique peuvent impacter les résultats.

Pour conclure, ces travaux de thèse m'ont également permis d'acquérir différentes compétences : **humaines** par la pluridisciplinarité demandée par chaque projet ; **analytique** pour les optimisations apportées ; **instrumentales** pour les problématiques rencontrées et **théoriques** par la richesse des différents sujets étudiés

Partie Expérimentale

Chapitre II. Application pour l'étude de la borréliose de Lyme

1. Chapitre II.II-1.1 : Peaux de chiens infectés

Des chiens Beagle ont été infectés en plaçant 20 tiques collectées dans la nature (Rhode Island, USA) sur une partie du dos. Les prélèvements des biopsies de 3 mm de diamètre (*punch biopsy*) sur la totalité de l'épaisseur de la peau ont été réalisés à des jours différents (8,14,21) après le dépôt des tiques. Ces biopsies ont été conservées à -80°C dans des eppendorfs jusqu'à utilisation.

2. Chapitre II.II-1.2.1 : Extraction des peaux de chiens et dosage des protéines

L'extraction des protéines a été réalisée à + 4°C par broyage manuel d'une biopsie ($5,2 \pm 0,4$ mg) dans un potter en verre de 100 μ L (Wheaton, Milville, USA) et un volume total de 185 μ L de solution de Laemmli (62,5 mM Tris-HCl pH 6,8, 10 % (v/v) glycérol, 2 % SDS (p/v)) sans bleu de bromophénol et sans DTT. L'extraction a été réalisée en deux étapes successives avec d'abord 100 μ L de solution d'extraction puis 85 μ L. Les protéines extraites ont été ensuite dosées par la méthode de Lowry (Bio-rad DC Assay, Munich, Allemagne) en utilisant la BSA comme étalon.

3. Chapitre II.II-1.2.2: Préfractionnement sur gel SDS-PAGE

L'électrophorèse SDS-PAGE a été réalisée sur des gels de polyacrylamide constitués d'un gel de séparation (12% acrylamide, pH 8,8) et d'un gel de concentration (4% acrylamide, pH 6,8), avec un tampon de migration constitué de 25 mM de Tris base, 192 mM de glycine et 0,1 % (p/v) de SDS.

100 μ g de protéines extraites ont été chauffées 5 min à 95°C en présence de 0,05 mM DTT et 0,001 % (p/v) bleu de bromophénol. La totalité a été déposée sur gel. 10 μ L d'une solution de marqueurs de masse moléculaire Precision Plus Protein™ Dual Color (Bio-Rad) ont également été déposés sur le gel.

La migration a été effectuée à l'aide d'un MiniProtean III (Bio-Rad, Munich, Allemagne) sous 50V pendant 30 min puis 150V jusqu'à atteindre 4 cm de migration.

Les protéines ont été fixées 30 min dans une solution aqueuse à 45 % (v/v) méthanol, 5 % (v/v) acide acétique, puis colorées à l'aide d'une solution colloïdale de bleu de Coomassie (0,12 % (p/v), 10 % (v/v) acide phosphorique, 10 % (p/v) sulfate d'ammonium, 20 % (v/v) méthanol) pendant une nuit (G. Candiano *et al.*, 2004). Puis les gels ont été décolorés par une succession de bains dans l'eau ultrapure. Des bandes de 2 mm ont été systématiquement excisées (20 bandes). Une étape de prétraitement a été réalisée à l'aide d'un robot de pré-digestion (MassPREP Station, Waters, Manchester, Angleterre). Les bandes ont été lavées 3 fois 10 min dans un mélange ACN/ NH_4HCO_3 25 mM (1:1) afin d'éliminer le colorant. Les bandes ont ensuite été déshydratées par l'ACN pendant 5 min et les protéines ont été réduites par 50 μL de DTT (10 mM dans une solution de NH_4HCO_3 25 mM, 30 min, 60°C, puis 30 min à T_{amb}) puis alkylées par 50 μL d'une solution d'iodoacétamide (55 mM dans une solution de NH_4HCO_3 25 mM, 20 min, T_{amb}). Les bandes ont ensuite été lavées par 100 μL de NH_4HCO_3 25 mM et déshydratées par de l'ACN.

20 μL de trypsine à 6,7 ng/ μL ont été ajoutés et la digestion enzymatique *in gel* a été effectuée à 37°C durant une nuit.

Les peptides générés ont été extraits par ajout de 40 μL d'une solution aqueuse de 80 % (v/v) ACN et 1 % (v/v) FA, sous agitation pendant 2h à T_{amb} . Les solvants ont été évaporés (SpeedVac) jusqu'à évaporation totale. Puis les peptides ont été repris dans un volume 30 μL avec une solution aqueuse de FA à 0,1 % (v/v) avant d'être injectés en nanoLC-MS/MS.

4. Chapitre II.II-1.2.3 Analyse LC-MS/MS et bases de données

Analyse LC-MS/MS

Les analyses nanoLC-MS/MS ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive+ (Thermo Scientific, Waltham, MA, USA). Pour chaque bande de gel, 3 μL d'extrait peptidique ont été injectés. Les peptides ont d'abord été piégés sur une pré-colonne Symmetry C18 (0,18 mm X 20 mm, 5 μm , Waters) puis séparés sur une colonne Acquity UPLC® BEH130 C18 (75 μm X 250 mm, 1,7 μm , Waters). Les solvants étaient constitués d'une solution aqueuse à 0,1 % (v/v) FA (solvant A) et d'une solution d'ACN à 0,1 % (v/v) FA (solvant B). Le piégeage des peptides a été effectué à un débit de 5 $\mu\text{L}/\text{min}$ (solvant A : 99 / solvant B : 1) pendant 3 min et l'élution

des peptides a été effectuée à un débit de 450 nL/min en utilisant un gradient de solvant B de 5 à 35 % sur 60 min.

La tension du capillaire était fixée à 1,8 kVolts et la température de la source était fixée à 250 °C. L'analyse des ions précurseurs a été effectuée en mode positif sur une gamme de masse de 300 à 1 800 m/z avec une résolution de 70 000 @ 200 m/z, une AGC (*Automatic Gain Control*) fixée à 3.10^6 et un temps d'accumulation maximum de 50 ms. Pour chaque spectre MS, les 10 précurseurs les plus intenses avec un état de charge minimum de 2 ont été sélectionnés pour être fragmentés dans la cellule HCD. L'acquisition des spectres MS/MS a été effectuée sur une gamme de masse de 200 à 2 000 m/z avec une résolution de 17 500 @ 200 m/z, une AGC fixée à 1.10^5 et un temps d'accumulation maximum de 100 ms. L'exclusion dynamique des précurseurs a été fixée à 60 s. L'ensemble du couplage a été piloté par le logiciel Xcalibur 3.1.66.10 (Thermo Scientific).

Interprétation des données

Les données de masse (.raw) collectées au cours des analyses ont été converties en listes de masses (.mgf) par le logiciel MSConvert (Proteowizard, version 3.0.6090), et interprétées par les algorithmes Mascot 2.5.1 (Matrix Science, London, UK) et OMSSA 2.1.9. Les recherches ont été effectuées dans une banque générée manuellement comprenant les séquences des protéines *Canis lupus* et *Borrelia* (Tableau II-2) extraites de UniprotKB et NCBIInr respectivement. Les contaminants connus tels que les kératines humaines ont été ajoutés. Toutes les séquences des protéines ont été rassemblées avec les copies inverses de chaque séquence (*decoy*). La trypsine a été sélectionnée comme enzyme, la carbamidométhylation des résidus cystéines (+57 Da) a été définie comme modification fixe et l'oxydation des résidus méthionines (+16 Da) a été définie comme modification variable. Pour les spectres MS/MS, les tolérances des ions parents et des fragments ont été fixées à 5 ppm et 0,07 Da respectivement avec un maximum d'une coupure manquée. Les résultats d'identification des algorithmes Mascot et OMSSA ont été chargés indépendamment dans le logiciel Proline (Proline Studio Release 1.0). Pour la validation des protéines, seuls les spectres respectant des pré-filtres ont été gardés : Score d'ion Mascot > 25, $-\log(e\text{-value})$ OMSSA > 7, longueur du peptide minimale de 7 résidus, rang du spectre = 1. Les protéines identifiées ont ensuite été validées en utilisant un taux de faux positif protéique (FDR) inférieur ou égal à 1 %.

5. Chapitre II.II-2.2 Application aux peaux de chiens infectés

Avant injection, les extraits peptidiques (env 5 µg, 100 µg déposés pour 20 bandes) ont été évaporés à sec et repris dans 5 µL d'un mélange des peptides isotopiquement marqués SIS. Pour chaque bande, la totalité du mélange de peptides (5 µL) a été injectée. Les méthodes finales M1 et M2 sont décrites dans la Figure II-12. Pour chaque méthode (M1 et M2) une piste de 20 bandes a été réalisée.

Les analyses LC-SRM ont été réalisées sur une chaîne de chromatographie Agilent 1100 (Agilent Technologies) couplée à un spectromètre de masse QQQ-6490 (Agilent Technologies). La totalité du système était pilotée par le logiciel MassHunter (Agilent Technologies). La colonne de séparation était une ZORBAX 300SB C18 (150 mm x 300 µm, 3,5 µm) précédée d'une colonne d'enrichissement ZORBAX 300SB C18 MicroBore Guard (17 mm x 1 mm, 5 µm). La température de la colonne de séparation était réglée à 23°C. Les solvants étaient constitués, d'une part, d'une solution aqueuse à 0,1 % (v/v) FA (solvant A) et d'une solution ACN à 0,1 % (v/v) FA (solvant B). La totalité de chaque extrait (5 µL) a été chargée sur la colonne d'enrichissement à un débit de 50 µL/min avec un solvant ACN/H₂O (2/98, v/v). L'élution a été effectuée à un débit de 5 µL/min avec un gradient linéaire de 5-45 % de solvant B en 57 min.

Les paramètres du spectromètre de masse étaient les suivants : température de source fixée à 200°C avec un débit de gaz au niveau du cône de 11 L/min et tension de nébulisation de 2000 V. L'acquisition a été réalisée en mode positif. La fragmentation a été effectuée en utilisant du diazote comme gaz de collision. La fenêtre d'isolement pour le Q1 et Q3 était fixée à 0,7 m/z. Les méthodes SRM étaient en mode d'acquisition programmé, avec des fenêtres de temps de rétention de ±3 min (6 min au total, M1) et ± 2 min (4 min au total, M2), et un temps de cycle de 3000 ms. Les données collectées ont été traitées manuellement avec le logiciel Skyline version 3.6.9.10875.

6. Chapitre II.III-1 Diagnostic de la phase précoce de la maladie de Lyme

Deux biopsies de 3 mm (*punch biopsy*) de diamètre ont été prélevées, après anesthésie locale à la lidocaïne, chez des patients présentant des érythèmes migrants après une piqûre de tique. Le prélèvement de ces biopsies a fait l'objet d'un consentement du patient. Une biopsie a été utilisée pour des analyses de biologie moléculaire (1/2 biopsie pour l'analyse microscopique de la culture (anaérobiose, 8 à 12 semaines)) et la PCR, et 1/2 biopsie mise en culture

(aérobiose, 8 à 12 semaines) puis stockage). La deuxième biopsie était destinée à l'analyse protéomique. Deux protocoles différents ont été utilisés au cours de l'étude de la cohorte : 19 biopsies stockées à -80°C jusqu'à l'analyse protéomique (protocole initial), puis 53 biopsies mises en culture (anaérobiose, 8 à 12 semaines) avant l'analyse protéomique (protocole modifié).

Les protocoles d'extraction des protéines et de préfractionnement SDS-PAGE étaient les mêmes que pour les biopsies canines (points 2 et 3), excepté une moyenne de 3,9 mg de biopsies, un dépôt de 50 µg de protéines, une migration sur 10 bandes (2 cm) et une digestion trypsique de 12h.

7. Chapitre II.III-1.2.1 : Description de la méthode LC-SRM

Les analyses LC-SRM ont été réalisées sur le même couplage Agilent 1100 - QQQ-6490 (Agilent Technologies) que dans le point 5. L'ensemble des paramètres de la chaîne chromatographique et du spectromètre de masse, ainsi que le gradient utilisé pour ces analyses sont décrits dans le point 5.

Avant injection, les extraits peptidiques (env 5 µg, 50 µg déposés pour 10 bandes) ont été évaporés à sec et repris dans 5 µL d'un mélange de peptides SIS (51 peptides PEPotec et 5 peptides AQUA™, Thermo Fisher Scientific). Pour chaque bande, la totalité du mélange de peptides (5 µL) a été injectée. La méthode LC-SRM ciblait 56 peptides correspondant à 6 protéines (Figure II-18). Elle a été opérée en mode d'acquisition programmé, avec des fenêtres de temps de rétention de ±5 min (10 min au total) et un temps de cycle de 3000 ms. Les données collectées ont été traitées manuellement avec le logiciel Skyline version 3.6.1.10615 (année 2016) ; version 4.0.9.11664 (année 2017) ; version 4.2.1.19004 (année 2018) et version 19.1.1.248 (année 2019).

8. Chapitre II.III-1.2.2.1 Problèmes de décalage de Tr

La vérification de l'isoforme présente dans les échantillons a été réalisée à l'aide de 3 peptides SIS de qualité crude (LTDSNAFVLAVK, ITDSNAFVIAVK, LTDSNAFVIAVK). Quatre injections de 5 µL de chaque peptide ont été réalisées, puis un mélange des quatre isoformes a été injecté. L'identification de l'isoforme a été réalisée par les Tr à l'aide du logiciel Skyline version 4.0.9.11664.

Pour la vérification de l'isoforme présente dans les isolats bactériens, les isolats de deux patients ont été réceptionnés dans des tubes stériles contenant du PBS. Les isolats ont été centrifugés à 14 000 g pendant 5 min à 20°C., puis les surnageant ont été éliminés et les culots ont été remis en suspension avec 185 µL de solution de Laemmli. Les protéines ont été dosées et préfractionnées sur SDS-PAGE comme décrit précédemment, excepté un dépôt de 5 µg de protéines. Avant injection, les extraits peptidiques ont été évaporés à sec et repris dans 5 µL d'un mélange de peptides SIS, comme décrit précédemment. Les traitements de données ont été réalisés à l'aide du logiciel Skyline version 4.2.1.19004.

9. Chapitre II.III-1.2.2.2 Problèmes de sensibilité

Détermination des LLOQ

Une biopsie cutanée humaine cutanée a été prélevée durant une abdominoplastie et conservée à -80°C jusqu'à utilisation. Les protocoles d'extraction des protéines et de préfractionnement SDS-PAGE sont décrits dans le point 6. Pour l'analyse LC-SRM, les paramètres sont décrits dans le point 7. Pour cette expérience un fractionnement SDS-PAGE de 10 bandes a été réalisé pour chaque point de gamme.

Un mélange de 56 peptides NAT (51 Spike Tides™ TL et 5 Spike Tides™ TQL, JPT Technologies, Berlin, Germany) a été réalisé. Les points de gamme suivants ont été préparés dans de l'eau acidifiée à 0,1% (v/v) FA : 100 amol- 500 amol- 1 fmol- 2,5 fmol- 5 fmol- 10 fmol- 20 fmol- 50 fmol, 100 fmol et 200 fmol pour la PRM et la SRM. Pour la SRM les points à 400 fmol et 2000 fmol ont été ajoutés (Figure II-23). Un volume de 20 µL de ces mélanges de peptides NAT a été ajouté aux bandes de gel, suivi d'une agitation de 20 min. La digestion enzymatique *in gel* a été réalisée avec 80 µL de trypsine avec un ratio enzyme:substrat (1:37) pendant une nuit à 37°C. L'extraction des peptides a été réalisée par ajout d'une solution à 80 % ACN, 0,1% FA. Ces extraits peptidiques ont ensuite été évaporés à sec (SpeedVac) puis repris avec un mélange optimisé de peptides SIS (51 peptides PEPotec et 5 peptides AQUA™, Thermo Fisher Scientific) dans une solution aqueuse acidifiée à 0,1% FA avant injection.

Analyse LC-SRM

Les analyses LC-SRM ont été réalisées sur le même couplage Agilent 1100 - QQQ-6490 (Agilent Technologies) que dans le point 5. L'ensemble des paramètres de la chaîne chromatographique et du spectromètre de masse, ainsi que le gradient utilisé pour ces analyses sont décrits dans le point 5.

Analyse LC-PRM

Les analyses nanoLC-PRM ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive+ (Thermo Scientific, Waltham, MA, USA). Pour chaque bande de gel, 5 μ L d'extrait peptidique ont été injectés. Les peptides ont d'abord été piégés sur une pré-colonne Symmetry C18 (0,18 mm X 20 mm, 5 μ m, Waters) puis séparés sur une colonne Acquity UPLC® BEH130 C18 (75 μ m X 250 mm, 1,7 μ m, Waters). Les solvants étaient constitués d'une solution aqueuse à 0,1 % (v/v) FA (solvant A) et d'une solution d'ACN à 0,1 % (v/v) FA (solvant B). Le piégeage des peptides a été effectué à un débit de 5 μ L/min (solvant A : 99 /solvant B : 1) pendant 3 min et l'élution des peptides a été effectuée à un débit de 450 nL/min en utilisant un gradient de solvant B de 5 à 35 % sur 60 min.

La méthode PRM a été développée avec une acquisition en full scan MS1 et 10 scans MS/MS. L'acquisition des spectres de masse a été effectuée en mode positif sur une gamme de masse de 300 à 1800 m/z, avec une résolution de 17 500 @ 200 m/z, une AGC maximum de $3 \cdot 10^6$ dans la C-trap et un temps d'accumulation maximum de 50 ms. Les spectres MS/MS ont été acquis avec une résolution de 35 000 @ 100 m/z. L'accumulation maximale d'ions utilisée est de $1 \cdot 10^6$ avec un temps d'accumulation maximal de 128 ms et une fenêtre d'acquisition programmée (*scheduled*) de ± 3 min (6 min, au total). Les 56 précurseurs ont été sélectionnés avec une fenêtre d'isolement de 2 m/z et fragmentés avec une énergie de collision normalisée (NCE) de 27. L'ensemble du couplage a été piloté par le logiciel Xcalibur 3.1.66.10 (Thermo Scientific).

Les données obtenues ont été traitées à l'aide du logiciel Skyline. LC-SRM (version 4.0.9.11664) et LC-PRM (version 4.2.1.19004).

10. Chapitre II.III-1.3.1 Préparation de tube-gels (TG)

La biopsie cutanée utilisée est la même que celle utilisée dans le point 9, et les protocoles d'extraction et dosage ont été décrits dans le point 6.

Les TG ont été préparés dans un tube Eppendorf de 0,5 mL, contenant 4,9 μ L d'extrait protéique (12,5 μ g de protéines). Pour la polymérisation du gel, 67,35 μ L d'eau ultrapure, 25 μ L d'acrylamide/bis-acrylamide à 30 %, 0,25 μ L de TEMED et 2,5 μ L de persulfate d'ammonium à 10 % (p/v) ont été ajoutés. Les solutions ainsi générées ont été vortexées et centrifugées de manière à retirer l'ensemble des bulles. Les TG résultants ont été fixés avec

une solution contenant 45 % de méthanol et 5 % d'acide acétique, avant d'être découpés en morceaux d'environ 2 mm puis d'effectuer les étapes de réduction, alkylation, lavage et déshydratation.

Comme pour le préfractionnement, différents points de gamme ont été réalisés avec un mélange de peptides NAT (51 Spike Tides™ TL et 5 Spike Tides™ TQL, JPT Technologies, Berlin, Germany) 500 amol- 2,5 fmol- 5 fmol- 7,5 fmol- 10 fmol-15 fmol- 20 fmol- 50 fmol- 100 fmol- 200 fmol- 400 fmol- 600 fmol et 1000 fmol. Un volume de 100 µL a été ajouté dans les échantillons avant l'étape de digestion (processus réalisé en triplicat).

La digestion enzymatique a été réalisée avec 350 µL de trypsine pendant la nuit à 37°C avec un ratio enzyme:substrat (1:37). L'extraction des peptides a été réalisée par ajout d'une solution à 80% ACN, 0,1% FA. Les extraits peptidiques ont été évaporés à sec (SpeedVac) puis repris dans un volume de 208 µL du mélange de peptides SIS (51 peptides PEPotec et 5 peptides AQUA™, Thermo Fisher Scientific) avant injection.

Les analyses nanoLC-PRM ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive+ (Thermo Scientific, Waltham, MA, USA), comme décrit précédemment dans le point 9. Les données obtenues ont été traitées à l'aide du logiciel Skyline (version 4.2.1.19004).

11. Chapitre II.III-1.3.2 Mode DIA

Deux biopsies de la cohorte de patients ont été analysées. La préparation d'échantillon a été réalisée comme cité précédemment (point 6).

Les analyses nanoLC-MS/MS ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive HF-X (Thermo Scientific, Waltham, MA, USA). La séparation des peptides a été effectuée comme décrit dans le point 4 avec les modifications suivantes : élution des peptides effectuée à un débit de 400 nL/min en utilisant un gradient de solvant B de 1 à 35 % sur 60 min.

La tension du capillaire était fixée à 2 kVolts et la température de la source était fixée à 250 °C. L'analyse en mode DDA a été effectuée avec un changement automatique entre les modes MS (gamme de masse de 300 à 1800 m/z avec une résolution de 60 000 @ 200 m/z, un AGC de $3 \cdot 10^6$ et un temps d'accumulation maximum de 50 ms. Pour chaque spectre MS, les 20

précurseurs les plus intenses ont été sélectionnés pour être fragmentés (NCE 27) dans la cellule HCD. Le temps d'exclusion dynamique a été fixé à 40 s.

L'acquisition des spectres MS/MS a été effectuée sur une gamme de masse allant de 200 à 2000 m/z avec une résolution de 15 000 @ 200 m/z, un AGC fixé à 1.10^5 et un temps d'accumulation maximum de 50 ms.

L'analyse en mode DIA a été effectuée avec 50 fenêtres fixes. La fenêtre d'isolement du précurseur a été fixée dans une gamme de masse 15,4 Da. La résolution utilisée était de 60 000 @ 200 m/z, avec un AGC de 3.10^6 et un temps d'accumulation maximum de 60 ms en mode MS. Les fragments générés à partir de 200 m/z ont été analysés en mode MS/MS avec une résolution de 15 000 @ 200 m/z, un AGC de 1.10^5 et un temps d'accumulation maximum de 50 et NCE de 27. L'ensemble du couplage a été piloté par le logiciel Xcalibur 4.0.27.19 (Thermo Scientific).

Les données de masse générées en mode DDA (.raw) ont été converties en listes de masses (.mgf) par le logiciel MSConvert (Proteowizard, version 3.0.6090), et interprétées par les algorithmes Mascot 2.5.1 (Matrix Science, London, UK). Les recherches ont été effectuées dans une banque générée manuellement, comprenant les séquences des protéines humaines (20 149 entrées, en novembre 2016 extraites de UniprotKB/SwissProt) ainsi que de *Borrelia* pour qui quatre banques différentes extraites de NCBI ont été utilisées : *B. burgorferi* ss B31 (1758 entrées en mai 2013), *B. burgdorferi* ss N40 (1480 entrées en janvier 2015), *B. afzelii* PKo (2186 entrées en octobre 2014), et *B. garinii* PBi (1720 entrées en août 2013). Les contaminants connus, tels que les kératines humaines, ont été ajoutés. L'ensemble des séquences protéiques a été rassemblé avec les copies inverses de chaque séquence (*decoy*). La trypsine a été sélectionnée comme enzyme, la carbamidométhylation des résidus cystéines (+57 Da) a été définie comme modification fixe, et l'oxydation des résidus méthionines (+16 Da) et label (13C(6)15N(2) (K) label 13C(6)15N(4) (R)) pour les peptides marqués ont été définie comme modification variable. Pour les spectres MS/MS, les tolérances des ions parents et des fragments ont été fixées à 5 ppm et 0,05 Da respectivement, avec un maximum d'une coupure manquée. Les résultats d'identification des algorithmes Mascot et OMSSA ont été chargés indépendamment dans le logiciel Proline (Proline Studio Release 2.0). Pour la validation des protéines, seuls les spectres respectant des pré-filtres ont été gardés : longueur du peptide minimale de 7 résidus, rang du spectre = 1. Les protéines identifiées ont ensuite été validées en utilisant un FDR peptidique (adjusted e-value) et protéique (modified Mudpit) inférieur ou égal à 1 %.

Pour le suivi global des protéines de *Borrelia*, la librairie spectrale a été générée avec le moteur de recherche Pulsar disponible dans Spectronaut™, en important les 22 fichiers .raw des analyses DDA, et en utilisant la même banque de données et les mêmes paramètres de recherches que décrits précédemment. La librairie spectrale était composée de 4753 groupes de protéines et 36044 peptides, générés avec les paramètres par défaut. Les extractions des données DIA ont été réalisées avec Spectronaut™ avec la librairie spectrale (DIA) et sans librairie spectrale (directDIA) en appliquant les paramètres par défaut (filtre q-value défini à 1% et profilage des temps de rétention (iRT profiling)). Les séquences peptidiques de *Borrelia* ne pouvant être distinguées des séquences humaines du fait de la présence d'une leucine ou d'une isoleucine ont été supprimées.

La détection ciblée des peptides de *Borrelia* a été réalisée avec le logiciel Skyline. La librairie spectrale a été générée avec le même logiciel, en fixant un FDR à 5%, à partir des recherches Mascot effectuées pour les 22 analyses DDA décrites précédemment. La librairie contenait finalement 95920 peptides. L'extraction des données a été effectuée en utilisant les paramètres appropriés. Brièvement, 6 ions produits (ions mono- et bi-chargés y et b) de l'ion 3 jusqu'au dernier ion -1 par ion précurseur ont été extraits. Un pouvoir de résolution égal à 30000 à 200 m/z a été appliqué pour filtrer au niveau MS/MS et seuls les signaux à moins de 5 min des temps de rétention espérés ont été utilisés pour le filtre sur le temps de rétention. Enfin, les pics chromatographiques ont été vérifiés manuellement afin d'ajuster les bornes d'intégration et éliminer les transitions interférées.

12. Chapitre II.III-2 Diagnostic de la phase tardive disséminée de la maladie de Lyme

Prélèvement des biopsies

Les biopsies cutanées de souris C3H/HeN ont été infectées par inoculation des bactéries à la seringue (1 000 *Borrelia*) ou par des tiques infectées (estimation ≈ 100 *Borrelia* par tique) sur des souris. Quatre souches différentes correspondant à trois espèces pathogènes de la bactérie ont été utilisées : *B. burgdorferi* ss N40 et 297, *B. garinii* PBi et *B. afzelii* Pko. Au bout de 40 jours, un dermocorticoïde (clobetasol) a été appliqué localement pendant 48 h. Les prélèvements des biopsies ont été effectués 5 jours après application du dermocorticoïde. Les biopsies cutanées ont été coupées en deux morceaux : un morceau a été utilisé pour vérifier la présence des bactéries grâce à une quantification par RT-PCR du gène de la flagelline (*fla*) de *Borrelia*, et l'autre morceau a été congelé puis transmis à notre laboratoire. La quantification

par RT-PCR nous a permis de sélectionner les biopsies les plus infectées pour les analyses protéomiques.

Les protocoles d'extraction des protéines, dosage et préfractionnement SDS-PAGE ont été les mêmes que décrits dans les points 2 et 3. La migration des bandes a été effectuée sur un gel SDS-PAGE et la digestion a été réalisée *in gel* avec la trypsine avec un ratio enzyme:substrat (1:37) pendant une nuit à 37°C.

Analyse LC-SRM

Les analyses LC-SRM ont été réalisées sur le même couplage Agilent 1100 - QQQ-6490 (Agilent Technologies) que dans le point 5. L'ensemble des paramètres de la chaîne chromatographique et du spectromètre de masse, ainsi que le gradient utilisé pour ces analyses sont décrits dans le point 5.

13. Chapitre II.III-2.2 Analyse de peaux de souris 40 jours après une infection

Pour le préfractionnement, 100 et 75 µg de protéines ont été déposés sur gel, correspondant à 16 et 11 bandes respectivement.

Analyse LC-SRM

Avant injection, les extraits peptidiques (équivalents à environ 6,25 µg de protéine par bande) ont été évaporés à sec et repris dans 6 µL d'un mélange de peptides SIS (59 peptides PEPotec et 3 peptides AQUA™, Thermo Fisher Scientific). Pour chaque bande, la totalité du mélange de peptides (6 µL) a été injectée. La méthode LC-SRM ciblait 62 peptides correspondant à 14 protéines (méthode M4). Elle a été opérée en mode d'acquisition programmé, avec des fenêtres de temps de rétention de ±3 min (6 min au total) et un temps de cycle de 3000 ms. Les données collectées ont été traitées avec le logiciel Skyline version 3.5.9

14. Chapitre II.III-2.3 Modifications de la méthode ciblée

L'évaluation de la méthode M6 pour la détection de *Borrelia* a été réalisée dans une peau humaine saine (décrit dans le point 9) dopée avec deux quantités de *B. afzelii* (10⁴ et 10⁵ bactéries).

Un volume de 1 mL de culture de *B. afzelii* (5x10⁶ bactéries/mL) a été centrifugé à 14 000 x g à 20°C pendant 5 min. Le surnageant a été éliminé et le culot a été remis en suspension dans

1 mL de PBS. Après centrifugation à 14 000 x g pendant 5 min à 20°C, le surnageant a été éliminé. Après le dernier lavage, le culot a été resuspendu dans 250 µL d'une solution de Laemmli (cf point 2) pour obtenir une culture *B. afzelli* à 1x10⁴ bactéries/mL.

Pour le dopage, deux morceaux de 5 mg de biopsie cutanée saine ont été pesés.

Dopage de biopsie cutanée saine avec 10⁴ bactéries : ajout de 1 µL de culture de bactérie sur la peau, puis 184 µL de solution de Laemmli, puis extraction des protéines par broyage comme décrit précédemment dans le point 2.

Dopage de biopsie cutanée saine avec 10⁵ bactéries : ajout de 10 µL de culture de bactérie sur la peau, puis 175 µL de solution de Laemmli, puis extraction des protéines par broyage comme décrit précédemment dans le point 2.

Les dosages et préfractionnement ont été réalisés comme décrit précédemment dans le point 2 et 3, excepté pour le dépôt de 50 µg de protéines préfractionnées en 10 bandes. La digestion a été réalisée comme citée dans le point 12.

Avant injection, les extraits peptidiques (équivalents à environ 5 µg de protéine par bande) ont été évaporés à sec et repris dans 5 µL d'un mélange de peptides SIS (40 peptides PEPotec et 13 peptides AQUA™ (3 peptides Thermo Fisher Scientific et 10 peptides JPT Technologies)). Pour chaque bande, la totalité du mélange de peptides (5 µL) a été injectée. La méthode LC-SRM ciblait 53 peptides correspondant à 13 protéines (méthode M6). Elle a été opérée en mode d'acquisition programmé, avec des fenêtres de temps de rétention de ±2 min (4 min au total) et un temps de cycle de 3000 ms. Les données collectées ont été traitées avec le logiciel Skyline version 4.2.1.19004.

15. Chapitre II.III-2.4 Comparaison de deux dermocorticoïdes

Les prélèvements des biopsies ont été réalisés comme décrit dans le point 13, excepté le fait qu'ici deux dermocorticoïdes différents ont été utilisés (clobetasol et locapred). Les analyses LC-SRM ont été réalisées comme décrit dans le point 12 avec les modifications suivantes :

Méthode LC-SRM (méthode M5) : Avant injection, les extraits peptidiques (équivalents à environ 5 µg de protéines par bande) ont été évaporés à sec et repris dans 5 µL d'un mélange de peptides SIS (55 peptides PEPotec et 3 peptides AQUA™ Thermo Fisher Scientific). Pour chaque bande, la totalité du mélange de peptides (5 µL) a été injectée. La méthode LC-SRM ciblait 58 peptides correspondant à 14 protéines.

Méthode LC-SRM (méthode M6) : Avant injection, les extraits peptidiques (équivalents à environ 5 µg de protéines par bande) ont été évaporés à sec et repris dans 5 µL d'un mélange de peptides SIS (40 peptides PEPotec et 13 peptides AQUA™ (3 peptides Thermo Fisher Scientific et 10 peptides JPT Technologies)). Pour chaque bande, la totalité du mélange de peptides a été injectée. La méthode LC-SRM ciblait 53 peptides correspondant à 13 protéines.

Pour les deux méthodes, les analyses ont été réalisées en mode d'acquisition programmé, avec des fenêtres de temps de rétention de ± 2 min (4 min au total) et un temps de cycle de 3000 ms. Les données collectées ont été traitées avec le logiciel Skyline version 4.2.1.19004.

16. Chapitre II.III-2.5 Effet de la lidocaïne

Les biopsies cutanées de souris C3H/HeN ont été infectées par inoculation des bactéries à la seringue. Trois espèces différentes de *Borrelia* ont été testées : *B. burgdorferi* ss, *B. garinii* et *B. afzelii*. Différentes souches appartenant à ces trois espèces principales ont été utilisées pour les expériences d'infection : la souche N40 de *B. burgdorferi* ss (isolée d'une tique) et la souche 297 (isolée d'un patient atteint de neuroborréliose). La souche NE4049 de *B. afzelii* (isolée d'une tique), la souche IBS 106 (isolée d'un patient atteint d'érythème migrant), et la souche PBi de *B. garinii* (isolée sur un patient atteint de neuroborréliose). Ces souches ont été cultivées dans le milieu complet BSK-H (Sigma) à 33 °C. Les animaux ont été infectés par voie intradermique dans la partie inférieure du dos.

Pour l'évaluation de l'effet de la lidocaïne sur la peau infectée par *Borrelia*, 50 µl d'une solution de 10 mg/mL a été injectée sur le site d'inoculation des bactéries. Le prélèvement a été effectué 30 min après inoculation, de manière à permettre la diffusion de l'anesthésique. Une partie du morceau prélevé a été conservée à -80°C avant les analyses (jour 0), le deuxième morceau a été mis en culture pendant 7 jours (jour 7).

Les protocoles d'extraction des protéines, de dosage et de préfractionnement SDS-PAGE ont été décrits précédemment, avec les modifications suivantes : extraction avec 555 µL de solution de Laemmli avec un potter en verre de 1 mL pour un broyage de 15 mg de biopsies cutanées.

La migration des bandes sur gel SDS-PAGE ainsi que la digestion ont été réalisées comme décrit précédemment dans le point 3. Les analyses nanoLC-MS/MS ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive+ (Thermo Scientific, Waltham, MA, USA). L'ensemble des paramètres de la

chaîne chromatographique et du spectromètre de masse, le gradient utilisé pour ces analyses ainsi que l'interprétation des données sont décrits dans le point 4. La seule modification apportée est la banque de données utilisée pour les recherches. Cette dernière a été générée manuellement avec les séquences protéiques de souris *Mus musculus* (16 872 entrées en décembre 2016) extraites de UniprotKB/SwissProt ainsi que les séquences protéiques de *Borrelia* décrites dans le point 11.

Chapitre III. Etude des agents pathogènes transmis par des tiques

17. Chapitre III.III-1 Nature des échantillons

Tiques : Les tiques ont été collectées dans leur milieu naturel en Alsace. Pour les tiques mâles à Murbach, pour les nymphes et femelles, à Niedermunster. Des tiques femelles d'élevage ont été utilisées comme échantillon témoin.

Peaux de souris : Les biopsies cutanées de souris C3H/HeN non infectées ont été utilisées comme échantillon témoin.

Les biopsies cutanées de souris C3H/HeN ont été infectées par 5 tiques femelles collectées dans leur milieu naturel (prélevées trois jours après le repas) et 10 nymphes (prélevées après la complétion du repas sanguin entre 4 à 6 jours).

Les biopsies cutanées de souris C3H infectées par inoculation intra-péritonéale avec des cellules infectées par *A. phagocytophilum* ont été prélevées à 1, 3, 7, 14, 21 jours après l'infection.

Les biopsies cutanées de souris infectées par inoculation à la seringue avec *B. microti* ont été prélevées à 1, 3 et 5 jours après l'infection.

18. Chapitre III.III-2 Extraction et dosage des protéines

Protocole d'extraction 1 : A mon arrivée au laboratoire, l'extraction des protéines était réalisé par broyage manuel d'une biopsie de 5 mg dans un potter en verre de 100 µL et un volume total de 185 µL de solution de Laemmli (cf point 2). L'extraction était réalisée en deux étapes successives avec d'abord 100 µL de solution d'extraction puis 85 µL. Les protéines extraites étaient ensuite dosées par la méthode de Lowry (Bio-rad DC Assay, Munich, Allemagne) en utilisant la BSA comme étalon.

Protocole d'extraction 2 : Les biopsies (15 mg) ont été broyées dans un potter en verre de 1 mL avec différents volumes de solution de Laemmli (185 μ L, 370 μ L, 555 μ L, 740 μ L). L'extraction des protéines a été réalisée en deux étapes successives avec dans un premier temps la moitié de la solution de Laemmli puis l'autre moitié. Les protéines ont ensuite été dosées par la méthode de Lowry en utilisant la BSA comme étalon.

Le protocole de préfractionnement utilisé a été décrit dans le point 2, avec une modification apportée lors de la digestion. Un ratio enzyme:substrat (1:25) a été utilisé pour l'ensemble des échantillons.

19. Chapitre III.IV- Résultats d'identification des protéines de pathogènes

Les tiques males (pool de 5) collectées dans leur milieu naturel ont été réceptionnées vivantes. Une étape de lavage avec l'éthanol 70% pendant 15 min a été réalisée. Puis, le protocole 2 d'extraction décrit précédemment a été appliqué. Pour les tiques nymphes (pool de 10) et femelles (pool de 5) collectées dans leur milieu naturel, ainsi que pour les tiques témoins (femelles, pool de 5), le protocole d'extraction 1 a été utilisé sans aucune modification.

Analyse LC-MS/MS

L'ensemble des échantillons dans ce chapitre ont été analysés par nanoLC-MS/MS sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, USA) couplé à un spectromètre de masse QqTOF (Triple TOF 5600, Sciex, Framingham, MA, USA). Pour chaque bande de gel, 2 μ L d'extrait peptidique ont été injectés. Les peptides ont d'abord été piégés sur une pré-colonne Symmetry C18 (0,18 mm X 20 mm, 5 μ m, Waters) puis séparés sur une colonne Acquity UPLC® BEH130 C18 (75 μ m X 250 mm, 1,7 μ m, Waters). Les solvants étaient constitués d'une solution aqueuse à 0,1 % (v/v) FA (solvant A) et d'une solution d'ACN à 0,1 % (v/v) FA (solvant B). Le piégeage des peptides a été effectué à un débit de 1 μ L/min (solvant A : 99 /solvant B : 1) pendant 3 min et l'élution des peptides a été effectuée à un débit de 300 nL/min en utilisant un gradient de solvant B de 3 à 40 % sur 79 min.

La tension du capillaire était fixée à 2,6 kVolts et la température de la source était fixée à 100°C. Le spectromètre de masse opérait en mode IDA (*Information Dependant Acquisition*) mode positif, piloté par le logiciel Analyst TF 1.7.1 (ABSciex). Les analyses de masse MS ont été effectuées sur une gamme de masse de 350-1250 m/z et 100-1800 m/z en mode MS/MS.

Pour l'acquisition des spectres MS/MS, les 30 ions les plus intenses ont été sélectionnés (top 30) pour être fragmentés.

Les données de masse collectées pendant l'analyse nanoLC-MS/MS ont été converties en fichier «. mgf » (pour mascot generic files) par le logiciel ProteinPilot™ (version 5.0 ABSciex). Les données ont ensuite été interprétées en utilisant les algorithmes Mascot 2.5.1 (Matrix Science, London, UK). Les recherches ont été effectuées dans une banque de données contenant les protéines des différents pathogènes, les protéines de tiques (*Ixodes ricinus*) et/ou les protéines de souris (*Mus musculus*) (Tableau III-3). Toutes les séquences protéiques ont été rassemblées avec les copies inverses de chaque séquence (*decoy*). La trypsine a été sélectionnée comme enzyme, la carbamidométhylation des résidus cystéines (+57 Da) a été définie comme modification fixe et l'oxydation des résidus méthionines (+16 Da) a été définie comme modification variable. Pour les spectres MS/MS, les tolérances des ions parents et des fragments ont été fixées à 10 ppm et 0,05 Da respectivement, avec un maximum d'une coupure manquée. Les résultats d'identification des algorithmes Mascot et OMSSA ont été chargés indépendamment dans le logiciel Proline (Proline Studio Release 2.0). Pour la validation des protéines, seuls les spectres respectant les pré-filtres ont été gardés : Score d'ion Mascot > 25, longueur minimale du peptide de 7 résidus, rang du spectre = 1. Les protéines identifiées ont ensuite été validées en utilisant un FDR protéique inférieur ou égal à 1 %.

20. Chapitre III.IV-3.1 Peaux de souris infectées expérimentalement

Pour l'ensemble des biopsies cutanées, le protocole d'extraction 2 décrit précédemment a été utilisé. Les étapes de préfractionnement jusqu'à l'interprétation des données ont été décrites dans le point 3.

21. Chapitre III.IV-3.1.1 *Anaplasma phagocytophilum*

L'étape de préparation d'échantillon, d'analyse LC-MS/MS et d'interprétation des données ont été décrites précédemment.

Les traitements supplémentaires suivants ont été réalisés avec ces échantillons. Les recherches ont été effectuées dans une banque de données contenant les protéines de souris (*Mus musculus*) et les protéines de *A. phagocytophilum* (19223 entrées avril 2016 extraites de UniProtKB/TrEMBL). Les recherches ont été effectuées à l'aide de deux algorithmes de

recherche : Mascot (version 2.5.1) et OMSSA (version 2.1.9). La validation des protéines a été réalisée à l'aide du logiciel Proline version 1.3. Seuls les spectres respectant les pré-filtres suivants ont été gardés : Score d'ion Mascot > 25 et $-\log(e\text{-value})$ OMSSA > 9, longueur minimale du peptide de 7 résidus, rang du spectre = 1. Les protéines identifiées ont ensuite été validées en utilisant un FDR protéique inférieur ou égal à 1 %.

Ces données ont été utilisées pour le Chapitre III.III-2.2.2 Algorithmes de recherche dans les banques de données (Figure III-8).

22. Chapitre III.IV-3.1.2 *Babesia microti* et le paragraphe Chapitre III.IV-3.2 Peaux de souris infectées via des tiques collectées dans la nature

Les protocoles utilisés sont décrits dans les points 3 et 4.

Chapitre IV. Analyse protéomique de *Pseudomonas aeruginosa*

23. Chapitre IV.II- Développements méthodologiques

Echantillon témoin : les cellules épithéliales pulmonaires humaines A549 (ATCC® CCL-185™) ont été cultivées en milieu RPMI 1640 (Gibco) enrichi par 10% v/v de FBS (Gibco) à 37°C et 5% CO₂.

Cellules infectées : le milieu de culture contenant les bactéries *P. aeruginosa* 5x10⁵ CFU (multiplicité d'infection de 50) a été ajouté au milieu de culture des cellules A549. L'infection a été réalisée pendant 3 h à 37°C et 5 % CO₂. A la fin de l'infection, les plaques ont été lavées une fois avec du tampon PBS, puis les cellules A549 et bactéries adhérant au fond de la plaque ont été collectées à l'aide d'un grattoir cellulaire. Les cellules A549 et les bactéries ont été centrifugées, le surnageant a été collecté et le culot sec a été remis en suspension dans un volume de tampon PBS et deux volumes de réactif antibactérien ARNprotect (Qiagen) avant centrifugation et stockage à -80°C.

Les cellules A549 infectées ou non ont été ensuite diluées dans une solution de Laemmli diluée au demi. Pour la digestion en solution (DS), les cellules A549 infectées ont été diluées dans une solution de 8 M urée dans 100 mM NH₄HCO₃.

Une étape de lyse par sonication à l'aiguille dans un bac à glace a été réalisée pendant 1 min (en alternant 5 s on/off). Le dosage des protéines a été réalisé comme dans le point 2. L'étape de lyse et dosage a été appliqué pour les trois protocoles.

Les protocoles de préparation du gel pour le préfractionnement SDS-PAGE et du SG ont été les mêmes que dans le point 3. Dans le cas du SG, les protéines ont été migrées dans la partie supérieure du gel (gel de concentration à 4% d'acrylamide) avec un voltage constant de 50V. La bande migrée a été excisée, puis découpée en trois morceaux équivalents. Pour le TG, la préparation a été réalisée comme cité précédemment dans le point 10. Au total 12,5 µg de protéines ont été préparées dans le TG.

Prédigestion et digestion : l'étape de préfractionnement a été réalisée avec le robot Massprep, comme décrit dans le point 3. Pour le SG, le TG et la DS cette étape a été réalisée manuellement en appliquant les protocoles suivants :

Pour le SG et le TG, est réalisé comme cité dans les points 3. Le protocole de prédigestion cité précédemment a été utilisé pour le TG en utilisant 4 fois les volumes de solution indiqué à chaque étape. La digestion enzymatique a été réalisée avec un ratio enzyme:substrat (1 :25) pendant 14h à 37°C.

Pour la DS, la réduction a été réalisée par l'ajout de DTT jusqu'à une concentration finale de 12 mM pendant 1h à 37 °C. L'alkylation des cystéines a été réalisée par ajout d'IAA jusqu'à une concentration finale de 40 mM. La solution a ensuite été diluée à 1M d'urée. La digestion enzymatique a été réalisée avec un ratio enzyme:substrat (1:25) pendant 14h à 37°C. Après la digestion, du TFA (10%) a été ajouté de manière à obtenir une concentration finale de 0,5 % en TFA.

L'ensemble des échantillons DS a été dessalé par SPE avec une cartouche Sep-Pak C18 (Waters) selon le protocole suivant : Conditionnement (2 x 1 mL MeOH, 3 x 1 mL ACN et 3 x 1 mL 0,1 % FA) ; Chargement de l'échantillon ; Lavage (2 x 1 mL 0,1 % FA) ; Elution (600 µL 60 % ACN 0,1 % FA).

L'ensemble des échantillons des trois différents protocoles a été évaporé à sec sous vide (SpeedVac), et les échantillons ont été repris avec une solution (H₂O/ACN/FA (98/2/0,1) contenant les peptides iRT).

Analyse nanoLC-MS/MS : les analyses ont été réalisées sur un système nanoAcquity Ultra-performance-LC system (UPLC, Waters, Milford, MA, USA) couplé à un Q-Exactive+ (Thermo

Scientific, Waltham, MA, USA), comme décrit précédemment dans le point 4, avec une modification de gradient, en utilisant un gradient de solvant B de 1 à 35 % sur 120 min, puis 35 à 90% de solvant B sur 1 min.

La tension du capillaire était fixée à 1,8 kVolts et la température de la source était fixée à 250 °C. L'analyse des ions précurseurs a été effectuée en mode positif sur une gamme de masse allant de 300 à 1 800 m/z avec une résolution de 70 000 @ 200 m/z, un AGC fixé à $3 \cdot 10^6$ et un temps d'accumulation maximum de 50 ms. Les 10 précurseurs les plus intenses avec une intensité supérieure à $2 \cdot 10^5$ sur le spectre MS ont été sélectionnés afin d'être isolés et fragmentés par HCD. Les ions monochargés, et dont la charge n'a pas été assignée, ont été exclus de cette sélection. L'acquisition des spectres MS/MS a été effectuée sur une gamme de masse allant de 200 à 2 000 m/z avec une résolution de 17 500 @ 200 m/z, un AGC fixée à $1 \cdot 10^5$ et un temps d'accumulation maximum de 100 ms. L'exclusion dynamique des précurseurs a été fixée à 60 s. L'ensemble du couplage a été piloté par le logiciel Xcalibur 3.1.66.10 (Thermo Scientific).

Interprétation des données : les données de masse (.raw) collectées au cours des analyses ont été converties en listes de masses (.mgf) par le logiciel MSConvert (Proteowizard, version 3.0.6090), et interprétées par les algorithmes Mascot (Matrix Science, version 2.6.2). Les recherches ont été effectuées dans une banque générée manuellement comprenant les séquences des protéines humaines (20194, mars 2017) et *P. aeruginosa* PAO1 (5677 entrées, mars 2017) extraites de UniprotKB/Swissprot et UniprotKB respectivement. Les contaminants connus, tels que les kératines humaines, ont été ajoutés. L'ensemble des séquences protéiques a été rassemblé avec les copies inverses de chaque séquence (*decoy*). La trypsine a été sélectionnée comme enzyme, la carbamidométhylation des résidus cystéines (+57 Da) a été définie comme modification fixe et l'oxydation des résidus méthionines (+16 Da) a été définie comme modification variable. Pour les spectres MS/MS, les tolérances des ions parents et des fragments ont été fixées à 5 ppm et 0,07 Da respectivement avec un maximum d'une coupure manquée. Les résultats d'identification des algorithmes Mascot ont été chargés indépendamment dans le logiciel Proline (Proline Studio Release 2.0). Pour la validation des protéines, seuls les spectres respectant les pré-filtres suivants ont été gardés : rang égal à 1, un FDR de 1% à niveau peptidique (adjusted e-value) et protéique (modified Mudpit)

24. Chapitre IV.III- Réalisation des études protéomiques différentielles

Pour les deux études différentielles, le protocole de préparation d'échantillon utilisé est le TG. Un total de 12,5 µg de protéines ont été préparées dans le TG. L'étape de préparation d'échantillon, d'analyses par LC-MS/MS et d'interprétation de données ont été décrites précédemment.

Les différents composés (exosidérophores, sidéromicines et chélateurs) ont été ajoutés aux cellules A549 en même temps que *P. aeruginosa*, comme décrit précédemment (point 23). L'étude différentielle a été réalisée avec 5 réplicats biologiques pour chaque condition. Les différentes conditions sont citées dans les paragraphes III-1.2 et III-2.2

Pour l'étape de quantification, deux traitements différents ont été utilisés :

Traitement 1 : Les données brutes ont été traitées avec le logiciel Maxquant (version 1.6.0.16). La recherche a été effectuée à l'aide du moteur de recherche Andromeda en utilisant la trypsine comme enzyme. La banque utilisée était la même que précédemment décrite (point 1). La longueur minimale requise pour un peptide était de 7 acides aminés, et une seule coupure manquée a été autorisée. L'oxydation des méthionines a été utilisée comme modification variable. La carbamidométhylation des cystéines a été paramétrée comme modification fixe. L'option « *Match between runs* » a été utilisée, et un FDR inférieur ou égal à 1% a été utilisé à la fois au niveau peptidique et protéique en utilisant la stratégie *target-decoy*.

Le traitement statistique a été effectué avec le logiciel Perseus (version 1.5.6.0) en utilisant une distribution normale pour l'imputation des valeurs manquantes. Le test T de Welch a été utilisé pour identifier les protéines différentiellement exprimées. Cette analyse a été appliquée aux cinq réplicats biologiques. Les protéines ont été considérées comme exprimées de manière différentielle avec une p-value inférieure à 0,05.

Traitement 2 : Pour la quantification Proline (version 2.0), les fichiers ".raw" ont été convertis en fichiers ".mzDB" avec le logiciel MS Angel (version 1.6.2). La quantification XIC a été effectuée en utilisant 5 ppm comme tolérance m/z pour l'extraction de l'abondance des peptides. Un alignement des spectres MS, ainsi qu'un alignement de temps de rétention ont été réalisés avec la méthode Loess. L'option « *Cross assignments* » des peptides a été utilisée en utilisant 5 ppm comme tolérance m/z et 60 s comme tolérance de temps de rétention. Seules les protéines identifiées avec au moins un peptide unique ont été prises en

compte, et seuls des peptides spécifiques ont été conservés. Les contaminants ont été exclus de la liste de protéines avant l'analyse statistique.

Le traitement statistique a été effectué avec le logiciel ProStaR (version 1.16.6). Les protéines identifiées dans au moins 4 réplicats ont été prises en compte et la normalisation a été effectuée en utilisant le mode « *median quantile centering* » pour toutes les conditions. L'imputation a été effectuée au niveau des protéines avec l'algorithme SLSA pour les valeurs partielles observées par condition (POV) et avec l'algorithme DetQuantile (2,5 %, facteur 1) pour les valeurs manquantes dans toutes les conditions (MEC). Le test T de Welch a été utilisé pour identifier les protéines différentiellement exprimées. Pour chaque comparaison, une p-value inférieure à 0,001 permettait de considérer une protéine comme étant exprimée de manière différentielle.

Bibliographie

- Abbatiello, S. E., Mani, D. R., Keshishian, H., & Carr, S. A. (2010). Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin Chem*, 56(2), 291-305.
- Abdallah, M., Benoliel, C., Drider, D., Dhulster, P., & Chihib, N. E. (2014). Biofilm formation and persistence on abiotic surfaces in the context of food and medical environments. *Arch Microbiol*, 196(7), 453-472.
- Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., . . . Carr, S. A. (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol*, 27(7), 633-641.
- Adeolu, M., & Gupta, R. S. (2014). A phylogenomic and molecular marker based proposal for the division of the genus *Borrelia* into two genera: the emended genus *Borrelia* containing only the members of the relapsing fever *Borrelia*, and the genus *Borrelia* gen. nov. containing the members of the Lyme disease *Borrelia* (*Borrelia burgdorferi* sensu lato complex). *Antonie Van Leeuwenhoek*, 105(6), 1049-1072.
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., . . . Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*, 46(W1), W537-W544.
- Aguero-Rosenfeld, M. E., Wang, G., Schwartz, I., & Wormser, G. P. (2005). Diagnosis of lyme borreliosis. *Clin Microbiol Rev*, 18(3), 484-509.
- Ankney, J. A., Adil Muneer, A., & Chen, X. (2018). Relative and Absolute Quantitation in Mass Spectrometry-Based Proteomics. *Annual Rev. Anal. Chem*, 11, 49-77.
- Apweiler, R., Bairoch, A., & Wu, C. H. (2004). Protein sequence databases. *Curr Opin Chem Biol*, 8(1), 76-80.
- Aubry, C., Socolovschi, C., Raoult, D., & Parola, P. (2016). Bacterial agents in 248 ticks removed from people from 2002 to 2013. *Ticks Tick Borne Dis*, 7(3), 475-481.
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem*, 389(4), 1017-1031.
- Barthold, S. W., deSouza, M. S., Janotka, J. L., Smith, A. L., & Persing, D. H. (1993). Chronic Lyme Borreliosis in the Laboratory Mouse. *Am J Pathol*, 143, 959-972.
- Benson, D., Boguski, M., Lipman, D., & Ostell, J. (1990). The National Center for Biotechnology Information *Genomics*, 6(2), 389-391.
- Berghoff, W. (2012). Chronic Lyme Disease and Co-infections: Differential Diagnosis. *Neurol J*, 6, 158-178.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, 28(1), 235-242.
- Bernard, Q., Grillon, A., Lenormand, C., Ehret-Sabatier, L., & Boulanger, N. (2020). Skin Interface, a Key Player for *Borrelia* Multiplication and Persistence in Lyme Borreliosis. *Trends Parasitol*, 36(3), 304-314.
- Bernard, Q., Jaulhac, B., & Boulanger, N. (2014). Smuggling across the border: how arthropod-borne pathogens evade and exploit the host defense system of the skin. *J Invest Dermatol*, 134(5), 1211-1219.
- Betancourt, L. H., Sanchez, A., Pla, I., Kuras, M., Zhou, Q., Andersson, R., & Marko-Varga, G. (2018). Quantitative Assessment of Urea In-Solution Lys-C/Trypsin Digestions Reveals Superior Performance at Room Temperature over Traditional Proteolysis at 37 degrees C. *J Proteome Res*, 17(7), 2556-2561.

- Beynon, R. J., Doherty, M. K., Pratt, J. M., & Gaskell, S. J. (2005). Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods*, 2(8), 587-589.
- Biemann, K. (1990). Nomenclature for Peptide Fragment Ions (Positive Ions) *Methods Enzymol*, 193, 886-887.
- Blaschitz, M., Narodoslavsky-Gfoller, M., Kanzler, M., Stanek, G., & Walochnik, J. (2008). Babesia species occurring in Austrian Ixodes ricinus ticks. *Appl Environ Microbiol*, 74(15), 4841-4846.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., . . . Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 31(1), 365-370.
- Boeuf, A. (2013). Développement d'approches protéomiques pour l'étude des interactions tique / Borrelia / peau. *Thèse de Doctorat. Université de Strasbourg*.
- Bons, J. (2019). Développement de méthodes analytiques par spectrométrie de masse pour une meilleure quantification et une caractérisation fine des protéomes. *Thèse de Doctorat. Université de Strasbourg*.
- Borchers, A. T., Keen, C. L., Huntley, A. C., & Gershwin, M. E. (2015). Lyme disease: a rigorous review of diagnostic criteria and treatment. *J Autoimmun*, 57, 82-115.
- Borras, E., & Sabido, E. (2017). What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics*, 17(17-18).
- Boukhalfa, H., & Crumbliss, A. L. (2002). Chemical aspects of siderophore mediated iron transport. *BioMetals*, 15, 325-339.
- Boulanger, N., & Lipsker, D. (2015). Protection contre les piqûres de tiques. *Ann Dermatol Vener*, 142, 245-251.
- Bourmaud, A., Gallien, S., & Domon, B. (2016). Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics*, 16(15-16), 2146-2159.
- Bouyssie, D., Hesse, A. M., Mouton-Barbosa, E., Rompais, M., Macron, C., Carapito, C., . . . Bruley, C. (2020). Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics*, 36(10), 3148-3155.
- Brackley, K. I., & Grantham, J. (2009). Activities of the chaperonin containing TCP-1 (CCT): implications for cell cycle progression and cytoskeletal organisation. *Cell Stress Chaperones*, 14(1), 23-31.
- Brownridge, P., Holman, S. W., Gaskell, S. J., Grant, C. M., Harman, V. M., Hubbard, S. J., . . . Beynon, R. J. (2011). Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics*, 11, 2957-2970.
- Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., . . . Garin, J. (2007). Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics*, 6(12), 2139-2149.
- Buczek, A. M., Buczek, W., Buczek, A., & Bartosik, K. (2020). The Potential Role of Migratory Birds in the Rapid Spread of Ticks and Tick-Borne Pathogens in the Changing Climatic and Environmental Conditions in Europe. *Int J Environ Res Public Health*, 17(6).
- Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., . . . Righetti, P. G. (2004). Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis*, 25(9), 1327-1333.
- Cappadona, S., Baker, P. R., Cutillas, P. R., Heck, A. J., & van Breukelen, B. (2012). Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids*, 43(3), 1087-1108.
- Carapito, C., Burel, A., Guterl, P., Walter, A., Varrier, F., Bertile, F., & Van Dorsselaer, A. (2014). MSDA, a proteomics software suite for in-depth Mass Spectrometry Data Analysis using grid computing. *Proteomics*, 14(9), 1014-1019.

- Carcy, B., Précigout, E., Valentin, A., Gorenflot, A., Reese, R. T., & Schrével, J. (1991). Heat shock response of *Babesia divergens* and identification of the hsp70 as an immunodominant early antigen during ox, gerbil and human babesiosis. *Biol Cell*, 72, 93-102.
- Channaveerappa, D., Ngounou Wetie, A. G., & Darie, C. C. (2019). Bottlenecks in Proteomics: An Update. *Adv Exp Med Biol*, 1140, 753-769.
- Chatterjee, S., Stupp, G. S., Park, S. K., Ducom, J. C., Yates, J. R., 3rd, Su, A. I., & Wolan, D. W. (2016). A comprehensive and scalable database search system for metaproteomics. *BMC Genomics*, 17(1), 642.
- Chelius, D., & Bondarenko, P. V. (2002). Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.*, 1, 317 - 323.
- Chen, C., Huang, H., & Wu, C. H. (2017). Protein Bioinformatics Databases and Resources. *Methods Mol Biol*, 1558, 3-39.
- Chen, E. I., Cociorva, D., Norris, J. L., & Yates, J. R., 3rd. (2007). Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res*, 6(7), 2529-2538.
- Chevallet, M., Diemer, H., Van Dorssealer, A., Villiers, C., & Rabilloud, T. (2007). Toward a better analysis of secreted proteins: the example of the myeloid cells secretome. *Proteomics*, 7(11), 1757-1770.
- Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., & Gygi, S. P. (2015). A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol*, 33(7), 743-749.
- Comstedt, P., Schuler, W., Meinke, A., & Lundberg, U. (2017). The novel Lyme borreliosis vaccine VLA15 shows broad protection against *Borrelia* species expressing six different OspA serotypes. *PLoS One*, 12(9), 1-13.
- Cornelis, P., & Dingemans, J. (2013). *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Front Cell Infect Microbiol*, 3, 75.
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 26(12), 1367-1372. doi:10.1038/nbt.1511
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*, 10(4), 1794-1805.
- Craig, R., & Beavis, R. C. (2004a). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9), 1466-1467.
- Craig, R., Cortens, J. P., & Beavis, R. C. (2004b). Open Source System for Analyzing, Validating, and Storing Protein Identification Data *J Proteome Res*, 3(6), 1234-1242.
- Cretu, D., Prassas, I., Saraon, P., Batruch, I., Gandhi, R., Diamandis, E. P., & Chandran, V. (2014). Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry *Clin Proteomics*, 11, 27.
- Cunrath, O., Geoffroy, V. A., & Schalk, I. J. (2016). Metallome of *Pseudomonas aeruginosa*: a role for siderophores *Environ Microbiol.*, 18, 3258-3267.
- Cutler, S. J., Rudenko, N., Golovchenko, M., Cramaro, W. J., Kirpach, J., Savic, S., . . . Amaro, A. (2017). Diagnosing Borreliosis. *Vector Borne Zoonotic Dis*, 17(1), 2-11.
- Dagley, L. F., Infusini, G., Larsen, R. H., Sandow, J. J., & Webb, A. I. (2019). Universal Solid-Phase Protein Preparation (USP(3)) for Bottom-up and Top-down Proteomics. *J Proteome Res*, 18(7), 2915-2924.
- Dantas-Torres, F., Fernandes Martins, T., Munoz-Leal, S., Onofrio, V. C., & Barros-Battesti, D. M. (2019). Ticks (Ixodida: Argasidae, Ixodidae) of Brazil: Updated species checklist and taxonomic keys. *Ticks Tick Borne Dis*, 10(6), 101252.
- Dantas-Torres, F., & Otranto, D. (2016). Best Practices for Preventing Vector-Borne Diseases in Dogs and Humans. *Trends Parasitol*, 32(1), 43-55.

- de la Fuente, J., Antunes, S., Bonnet, S., Cabezas-Cruz, A., Domingos, A. G., Estrada-Pena, A., . . . Rego, R. O. M. (2017a). Tick-Pathogen Interactions and Vector Competence: Identification of Molecular Drivers for Tick-Borne Diseases. *Front Cell Infect Microbiol*, 7, 114.
- de la Fuente, J., Contreras, M., Estrada-Pena, A., & Cabezas-Cruz, A. (2017b). Targeting a global health problem: Vaccine design and challenges for the control of tick-borne diseases. *Vaccine*, 35(38), 5089-5094.
- de Silva, A. M., Telford, S. R., Brunet, L. R., Barthold, S. W., & Fikrig, E. (1996). *Borrelia burgdorferi* OspA Is an Arthropod-specific Transmission-blocking Lyme Disease Vaccine. *J. Exp. Med.*, 183, 271-275.
- Dehhaghi, M., Kazemi Shariat Panahi, H., Holmes, E. C., Hudson, B. J., Schloeffel, R., & Guillemin, G. J. (2019). Human Tick-Borne Diseases in Australia. *Front Cell Infect Microbiol*, 9, 3.
- Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., . . . Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Res*, 34(Database issue), D655-658.
- Deutsch, E. W., Lam, H., & Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep*, 9(5), 429-434.
- Dhabaria, A., Cifani, P., Reed, C., Steen, H., & Kentsis, A. (2015). A High-Efficiency Cellular Extraction System for Biological Proteomics. *J Proteome Res*, 14(8), 3403-3408.
- Di Venere, M., Fumagalli, M., Cafiso, A., De Marco, L., Epis, S., Plantard, O., . . . Sasser, D. (2015). *Ixodes ricinus* and Its Endosymbiont *Mitochondria*: A Comparative Proteomic Analysis of Salivary Glands and Ovaries. *PLoS One*, 10(9), 1-16.
- Diuk-Wasser, M. A., Vannier, E., & Krause, P. J. (2016). Coinfection by *Ixodes* Tick-Borne Pathogens: Ecological, Epidemiological, and Clinical Consequences. *Trends Parasitol*, 32(1), 30-42.
- Domanski, D., Percy, A. J., Yang, J., Chambers, A. G., Hill, J. S., Freue, G. V., & Borchers, C. H. (2012). MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics*, 12(8), 1222-1243.
- Domon, B., & Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. *Science*, 312(5771), 212-217.
- Domon, B., & Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol*, 28(7), 710-721.
- Dongre, A. R., Somogyi, A., & Wysocki, V. H. (1996). Surface-induced Dissociation : An Effective Tool to Probe Structure, Energetics and Fragmentation Mechanisms of Protonated Peptides. *J Mass Spectrom*, 31, 339-350.
- Eder, M., Cortes, F., Teixeira de Siqueira Filha, N., Araujo de Franca, G. V., Degroote, S., Braga, C., . . . Turchi Martelli, C. M. (2018). Scoping review on vector-borne diseases in urban areas: transmission dynamics, vectorial capacity and co-infection. *Infect Dis Poverty*, 7(1), 1-24.
- Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., . . . MacCoss, M. J. (2013). Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods*, 10(8), 744-746.
- Eldin, C., Raffetin, A., Bouiller, K., Hansmann, Y., Roblot, F., Raoult, D., & Parola, P. (2019). Review of European and American guidelines for the diagnosis of Lyme borreliosis. *Med Mal Infect*, 49(2), 121-132.
- Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, 4(3), 207-214.
- Embers, M. E., Hasenkampf, N. R., Jacobs, M. B., Tardo, A. C., Doyle-Meyers, L. A., Philipp, M. T., & Hodzic, E. (2017). Variable manifestations, diverse seroreactivity and post-treatment persistence in non-human primates exposed to *Borrelia burgdorferi* by tick feeding. *PLoS One*, 12(12), e0189071.

- Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An Approach to Correlate Tandem Mass Spectral Data of Peptides With Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom*, 5(11), 976-989.
- Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., . . . Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics*, 12, 1111-1121.
- Feist, P., & Hummon, A. B. (2015). Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci*, 16(2), 3537-3563.
- Figoni, J., Chirouze, C., Hansmann, Y., Lemogne, C., Hentgen, V., Saunier, A., . . . endorsed by scientific, s. (2019). Lyme borreliosis and other tick-borne diseases. Guidelines from the French Scientific Societies (I): prevention, epidemiology, diagnosis. *Med Mal Infect*, 49(5), 318-334.
- Fischer, R., & Kessler, B. M. (2015). Gel-aided sample preparation (GASP)--a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics*, 15(7), 1224-1229.
- Folkesson, A., Jelsbak, L., Yang, L., Johansen, H. K., Ciofu, O., Høiby, N., & Molin, S. (2012). Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nat Rev Microbiol*, 10, 841-851.
- Foulongne, V., Michaux-Charachon, S., O'Callaghan, D., & Ramuz, M. (2002). Systèmes de sécrétion des protéines de type IV et virulence bactérienne. *médecine/sciences*, 18(4), 439-447.
- Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F., Mechtler, K., Heck, A. J., & Mohammed, S. (2013). Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (ETHcD). *J Proteome Res*, 12(3), 1520-1525.
- Gallien, S., Bourmaud, A., Kima, S. Y., & Domon, B. (2014). Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics*, 100, 147-159.
- Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., & Domon, B. (2012). Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics*, 11(12), 1709-1723.
- Gallien, S., Duriezan, E., & Domon, B. (2011). Selected reaction monitoring applied to proteomic. *J. Mass. Spectrom*, 46, 298-312.
- Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., . . . Bryant, S. H. (2004). Open Mass Spectrometry Search Algorithm. *J Proteome Res*, 3, 958-964.
- Geiger, T., Cox, J., & Mann, M. (2010a). Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. *Mol Cell Proteomics*, 9, 2252-2261.
- Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., & Mann, M. (2010b). Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods*, 7(5), 383-385.
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., & Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*, 100(12), 6940-6945.
- Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., . . . Langridge, J. I. (2009). The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics*, 9(6), 1683-1695.
- Gillet, L. C., Leitner, A., & Aebersold, R. (2016). Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem*, 9(1), 449-472.
- Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., . . . Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics*, 11(6), 1-17.

- Gilmore Jr, R. D., & Mbow, M. L. (1998). A monoclonal antibody generated by antigen inoculation via tick bite is reactive to the *Borrelia burgdorferi* Rev protein, a member of the 2.9 gene family locus *Infect Immun*, 66, 980-986.
- Gonnelli, G., Stock, M., Verwaeren, J., Maddelein, D., De Baets, B., Martens, L., & Degroeve, S. (2015). A decoy-free approach to the identification of peptides. *J Proteome Res*, 14(4), 1792-1998.
- Gorska, A., Sloderbach, A., & Marszall, M. P. (2014). Siderophore-drug complexes: potential medicinal applications of the 'Trojan horse' strategy. *Trends Pharmacol Sci*, 35(9), 442-449.
- Gray, J. S., Kahl, O., Lane, R. S., Levin, M. L., & Tsao, J. I. (2016). Diapause in ticks of the medically important *Ixodes ricinus* species complex. *Ticks Tick Borne Dis*, 7(5), 992-1003.
- Grillon, A., Westermann, B., Cantero, P., Jaulhac, B., Voordouw, M. J., Kapps, D., . . . Boulanger, N. (2017). Identification of *Borrelia* protein candidates in mouse skin for potential diagnosis of disseminated Lyme borreliosis. *Sci Rep*, 7(1), 16719.
- Grimm, D., Tilly, K., Byram, R., Stewart, P. E., Krum, J. G., Bueschel, D. M., . . . Rosa, P. A. (2004). Outer-surface protein C of the Lyme disease spirochete: a protein induced in ticks for infection of mammals. *Proc Natl Acad Sci U S A*, 101(9), 3142-3147.
- Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D. L., Dianes, J. A., Del-Toro, N., . . . Vizcaino, J. A. (2016). Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods*, 13(8), 651-656.
- Grochowska, A., Milewski, R., Pancewicz, S., Dunaj, J., Czupryna, P., Milewska, A. J., . . . Moniuszko-Malinowska, A. (2020). Comparison of tick-borne pathogen prevalence in *Ixodes ricinus* ticks collected in urban areas of Europe. *Sci Rep*, 10(1), 6975.
- Guadagno, N. A., & Progida, C. (2019). Rab GTPases: Switching to Human Diseases. *Cells*, 8(8).
- Guo, B. P., Brown, E. L., Dorward, D. W., Rosenberg, L. C., & Höök, M. (1998). Decorin-binding adhesins from *Borrelia burgdorferi*. *Mol Microbiol*, 30, 711-723.
- Gupta, R. S. (2019). Distinction between *Borrelia* and *Borrelia* is more robustly supported by molecular and phenotypic characteristics than all other neighbouring prokaryotic genera: Response to Margos' et al. "The genus *Borrelia* reloaded" (PLoS ONE 13(12): e0208432). *PLoS One*, 14(8), 1-22.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags *Nat Biotechnol*, 17(10), 994-999.
- Hajdusek, O., Sima, R., Ayllon, N., Jalovecka, M., Perner, J., de la Fuente, J., & Kopacek, P. (2013). Interaction of the tick immune system with transmitted pathogens. *Front Cell Infect Microbiol*, 3, 26.
- Halperin, J. J. (2019). Lyme neuroborreliosis. *Curr Opin Infect Dis*, 32(3), 259-264.
- Hanke, S., Besir, H., Oesterheld, D., & Mann, M. (2008). Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. *J. Proteome Res.*, 7, 1118-1130.
- Hansmann, Y. (2009). Treatment and Prevention of Lyme Disease *Curr Probl Dermatol.*, 37, 111-129
- Herrmann, C., & Gern, L. (2015). Search for blood or water is influenced by *Borrelia burgdorferi* in *Ixodes ricinus*. *Parasit Vectors*, 8, 6.
- Hober, A., Edfors, F., Ryaboshapkina, M., Malmqvist, J., Rosengren, L., Percy, A. J., . . . Miliotis, T. (2019). Absolute Quantification of Apolipoproteins Following Treatment with Omega-3 Carboxylic Acids and Fenofibrate Using a High Precision Stable Isotope-labeled Recombinant Protein Fragments Based SRM Assay. *Mol Cell Proteomics*, 18(12), 2433-2446.

- Hoffman, M. A., Fang, B., Haura, E. B., Rix, U., & Koomen, J. M. (2018). Comparison of Quantitative Mass Spectrometry Platforms for Monitoring Kinase ATP Probe Uptake in Lung Cancer. *J Proteome Res*, 17(1), 63-75.
- Holstein-Sherwood, C. A., Gafken, P. R., & Martin, D. B. (2011). Collision Energy Optimization of b- and y-Ions for Multiple Reaction Monitoring Mass Spectrometry. *J Proteome Res*, 10, 231-240.
- Hong, S. H., Kim, S. Y., Song, B. G., Rho, J. R., Cho, C. R., Kim, C. N., . . . Lee, S. E. (2019). Detection and characterization of an emerging type of Babesia sp. similar to Babesia motasi for the first case of human babesiosis and ticks in Korea. *Emerg Microbes Infect*, 8(1), 869-878.
- Hubálek, Z. (2009). Epidemiology of Lyme Borreliosis. *Curr Probl Dermatol.*, 37, 31-50.
- Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., & Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol*, 10, 1-14.
- Huillet, C., Adrait, A., Lebert, D., Picard, G., Trauchessec, M., Louwagie, M., . . . Brun, V. (2012). Accurate quantification of cardiovascular biomarkers in serum using Protein Standard Absolute Quantification (PSAQ) and selected reaction monitoring. *Mol Cell Proteomics*, 11(2), M111 008235.
- Husson, G., Delangle, A., O'Hara, J., Cianferani, S., Gervais, A., Van Dorsselaer, A., . . . Carapito, C. (2018). Dual Data-Independent Acquisition Approach Combining Global HCP Profiling and Absolute Quantification of Key Impurities during Bioprocess Development. *Anal Chem*, 90(2), 1241-1247.
- Huttenhain, R., Soste, M., Selevsek, N., Rost, H., Sethi, A., Carapito, C., . . . Aebersold, R. (2012). Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci Transl Med*, 4(142), 142ra194.
- Izac, J. R., & Marconi, R. T. (2019). Diversity of the Lyme Disease Spirochetes and its Influence on Immune Responses to Infection and Vaccination. *Vet Clin North Am Small Anim Pract*, 49(4), 671-686.
- Jahfari, S., Hofhuis, A., Fonville, M., van der Giessen, J., van Pelt, W., & Sprong, H. (2016). Molecular Detection of Tick-Borne Pathogens in Humans with Tick Bites and Erythema Migrans, in the Netherlands. *PLoS Negl Trop Dis*, 10(10), 1-15.
- Jaulhac, B., Saunier, A., Caumes, E., Bouillier, K., Gehanno, J. F., Rabaud, C., . . . endorsed by scientific, s. (2019). Lyme borreliosis and other tick-borne diseases. Guidelines from the French scientific societies (II). Biological diagnosis, treatment, persistent symptoms after documented or suspected Lyme borreliosis. *Med Mal Infect*, 49(5), 335-346.
- Jensen, P. M., Christoffersen, C. S., Moutailler, S., Michelet, L., Klitgaard, K., & Bodker, R. (2017). Transmission differentials for multiple pathogens as inferred from their prevalence in larva, nymph and adult of Ixodes ricinus (Acari: Ixodidae). *Exp Appl Acarol*, 71(2), 171-182.
- Jia, Y., Wu, T., Jelinek, C. A., Bielekova, B., Chang, L., Newsome, S., . . . Cotter, R. J. (2012). Development of protein biomarkers in cerebrospinal fluid for secondary progressive multiple sclerosis using selected reaction monitoring mass spectrometry (SRM-MS) *Clin Proteomics*, 9, 9.
- Johnson, K. L., & Muddiman, D. C. (2004). A method for calculating 16O/18O peptide ion ratios for the relative quantification of proteomes. *J Am Soc Mass Spectrom*, 15(4), 437-445.
- Johnson, S. M., Saint John, B. E., & Dine, A. P. (2008). Local anesthetics as antimicrobial agents: a review. *Surg Infect (Larchmt)*, 9(2), 205-213.
- Jones, P., Cote, R. G., Martens, L., Quinn, A. F., Taylor, C. F., Derache, W., . . . Apweiler, R. (2006). PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res*, 34(Database issue), D659-663.

- Kazimirova, M., Thangamani, S., Bartikova, P., Hermance, M., Holikova, V., Stibraniova, I., & Nuttall, P. A. (2017). Tick-Borne Viruses and Biological Processes at the Tick-Host-Virus Interface. *Front Cell Infect Microbiol*, 7, 339.
- Kenedy, M. R., Lenhart, T. R., & Akins, D. R. (2012). The role of *Borrelia burgdorferi* outer surface proteins. *FEMS Immunol Med Microbiol*, 66(1), 1-19.
- Kern, A., Collin, E., Barthel, C., Michel, C., Jaulhac, B., & Boulanger, N. (2011). Tick saliva represses innate immunity and cutaneous inflammation in a murine model of Lyme disease. *Vector Borne Zoonotic Dis*, 11(10), 1343-1350.
- Kern, A., Schnell, G., Bernard, Q., Boeuf, A., Jaulhac, B., Collin, E., . . . Boulanger, N. (2015). Heterogeneity of *Borrelia burgdorferi* Sensu Stricto Population and Its Involvement in *Borrelia* Pathogenicity: Study on Murine Model with Specific Emphasis on the Skin Interface. *PLoS One*, 10(7), e0133195.
- Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., . . . Domon, B. (2011). Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomic. *Mol Cell Proteomics*, 10, 1-11.
- Klitgaard, K., Kjaer, L. J., Isbrand, A., Hansen, M. F., & Bodker, R. (2019). Multiple infections in questing nymphs and adult female *Ixodes ricinus* ticks collected in a recreational forest in Denmark. *Ticks Tick Borne Dis*, 10(5), 1060-1065.
- Koedel, U., Fingerle, V., & Pfister, H. W. (2015). Lyme neuroborreliosis-epidemiology, diagnosis and management. *Nat Rev Neurol*, 11(8), 446-456.
- Kuhn, S., & Mannherz, H. G. (2017). Actin: Structure, Function, Dynamics, and Interactions with Bacterial Toxins. *Curr Top Microbiol Immunol*, 399, 1-34.
- Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods*, 11(3), 319-324.
- Kurtenbach, K., Hanincova, K., Tsao, J. I., Margos, G., Fish, D., & Ogden, N. H. (2006). Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol*, 4(9), 660-669.
- Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227, 680-685.
- LaFleur, R. L., Dant, J. C., Wasmoen, T. L., Callister, S. M., Jobe, D. A., Lovrich, S. D., . . . Schell, R. F. (2009). Bacterin That Induces Anti-OspA and Anti-OspC *Borrelia burgdorferi* Antibodies Provides a High Level of Protection against Canine Lyme Disease. *Clin. Vaccine. Immunol.*, 16 (2) 253-259.
- Lam, H., & Aebersold, R. (2011). Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods*, 54(4), 424-431.
- Lange, V., Malmstrom, J. A., Didion, J., King, N. L., Johansson, P. B., Schafer, J., . . . Aebersold, R. (2008a). Targeted Quantitative Analysis of *Streptococcus pyogenes* Virulence Factors by Multiple Reaction Monitoring. 7, 1489–1500.
- Lange, V., Picotti, P., Domon, B., & Aebersold, R. (2008b). Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol*, 4, 222.
- Lau, K. W., Hart, S. R., Lynch, J. A., Wong, S. C., Hubbard, S. J., & Gaskell, S. J. (2009). Observations on the detection of b- and y-type ions in the collisionally activated decomposition spectra of protonated peptides. *Rapid Commun. Mass Spectrom.*, 23, 1508–1514.
- le Maire, M., Champeil, P., & Möller, J. V. (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta*, 1508(1-2), 86-111.
- Lejal, E., Marsot, M., Chalvet-Monfray, K., Cosson, J. F., Moutailler, S., Vayssier-Taussat, M., & Pollet, T. (2019). A three-years assessment of *Ixodes ricinus*-borne pathogens in a French peri-urban forest. *Parasit Vectors*, 12(1), 551.

- Levy, S. A. (2002). Use of a C6 ELISA Test to Evaluate the Efficacy of a Whole-Cell Bacterin for the Prevention of Naturally Transmitted Canine *Borrelia burgdorferi* Infection. *Vet. Ther.*, 3(4), 420-424.
- Liang, F. T., Nelson, F. K., & Fikrig, E. (2002). Molecular adaptation of *Borrelia burgdorferi* in the murine host. *J Exp Med*, 196(2), 275-280.
- LIMA, C. M. R., ZEIDNER, N. S., BEARD, C. B., SOARES, C. A. G., DOLAN, M. C., DIETRICH, G., & PIESMAN, J. (2005). Differential Infectivity of the Lyme Disease Spirochete *Borrelia burgdorferi* Derived from *Ixodes scapularis* Salivary Glands and Midgut. *J. Med. Entomol.*, 42(3), 506-510.
- Link, A., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., . . . Yates, J. R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7), 676-682.
- Liu, H., Sadygov, R. G., & Yates, J. R. (2004). A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.*, 76, 4193-4201.
- Lu, X., & Zhu, H. (2005). Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Mol Cell Proteomics*, 4, 1948-1958.
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*, 14(8), e8126.
- Lundgren, D. H., Hwang, S.-I., Wu, L., & Han, D. K. (2010). Role of Spectral Counting in Quantitative Proteomics. *Expert Rev Proteomics*, 7(1), 39-53.
- Lyczaka, J. B., Cannon, C. L., & Pier, G. B. (2000). Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.*, 2, 1051-1060.
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., . . . MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26(7), 966-968.
- Madison-Antenucci, S., Kramer, L. D., Gebhardt, L. L., & Kauffman, E. (2020). Emerging Tick-Borne Diseases. *Clin. microbiol. Rev.*, 33, 1-34.
- Madsen, J. A., Boutz, D. R., & Brodbelt, J. S. (2010). Ultrafast ultraviolet photodissociation at 193 nm and its applicability to proteomic workflows. *J Proteome Res*, 9(8), 4205-4214.
- Magni, R., Luchini, A., Liotta, L., & Molestina, R. E. (2019). Analysis of the *Babesia microti* proteome in infected red blood cells by a combination of nanotechnology and mass spectrometry. *Int J Parasitol*, 49(2), 139-144.
- Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., . . . Aebersold, R. (2007). Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol*, 25, 125-131.
- Matthiesen, R., & Jensen, O. N. (2008). Analysis of Mass Spectrometry Data in Proteomics. *Methods Mol Biol*, 453, 105-122.
- McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L., Jedrychowski, M. P., Rogers, J. C., . . . Gygi, S. P. (2012). Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem*, 84(17), 7469-7478.
- McCoy, K. D., & Boulanger, N. (2017). Tiques et maladies à tiques. Biologie, écologie évolutive et épidémiologie. *IRD Éditions*, 344p.
- McHugh, L., & Arthur, J. W. (2008). Computational Methods for Protein Identification From Mass Spectrometry Data. *PLoS Comput Biol*, 4(2), 1-12.
- Mehaffy, C., Dobos, K. M., Nahid, P., & Kruh-Garcia, N. A. (2017). Second generation multiple reaction monitoring assays for enhanced detection of ultra-low abundance *Mycobacterium tuberculosis* peptides in human serum. *Clin Proteomics*, 14, 21.
- Meng, F., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M., & Kelleher, N. L. (2002). Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal Chem*, 74, 2923-2929.

- Michalski, A., Cox, J., & Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res*, 10(4), 1785-1793.
- Mislin, G. L., & Schalk, I. J. (2014). Siderophore-dependent iron uptake systems as gates for antibiotic Trojan horse strategies against *Pseudomonas aeruginosa*. *Metallomics*, 6(3), 408-420.
- Miyazaki, S., Sugawara, H., Gojobori, T., & Tateno, Y. (2003). DNA Data Bank of Japan (DDBJ) in XML. *Nucleic Acids Res*, 31(1), 13-16.
- Moniuszko, A., Dunaj, J., Swiecicka, I., Zambrowski, G., Chmielewska-Badora, J., Zukiewicz-Sobczak, W., . . . Pancewicz, S. (2014). Co-infections with *Borrelia* species, *Anaplasma phagocytophilum* and *Babesia* spp. in patients with tick-borne encephalitis. *Eur J Clin Microbiol Infect Dis*, 33(10), 1835-1841.
- Montero, H., Perez-Gil, G., & Sampieri, C. L. (2019). Eukaryotic initiation factor 4A (eIF4A) during viral infections. *Virus Genes*, 55(3), 267-273.
- Moradian, A., Kalli, A., Sweredoski, M. J., & Hess, S. (2014). The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics*, 14(4-5), 489-497.
- Morita, Y., Tomida, J., & Kawamura, Y. (2014). Responses of *Pseudomonas aeruginosa* to antimicrobials. *Frontiers in Microbiology*, 4, 1-8.
- Moutailler, S., Valiente Moro, C., Vaumourin, E., Michelet, L., Tran, F. H., Devillers, E., . . . Vayssier-Taussat, M. (2016). Co-infection of Ticks: The Rule Rather Than the Exception. *PLoS Negl Trop Dis*, 10(3), 1-17.
- Muller, L., Fornecker, L., Cianferani, S., & Carapito, C. (2019). Tube-Gel: A Fast and Effective Sample Preparation Method for High-Throughput Quantitative Proteomics. *Methods Mol Biol*, 1959, 123-127.
- Muller, L., Fornecker, L., Van Dorsselaer, A., Cianferani, S., & Carapito, C. (2016). Benchmarking sample preparation/digestion protocols reveals tube-gel being a fast and repeatable method for quantitative proteomics. *Proteomics*, 16(23), 2953-2961.
- Nagaraj, N., Lu, A., Mann, M., & Wiśniewski, J. R. (2008). Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs. *J Proteome Res*, 7, 5028-5032.
- Navarro, P., & Vázquez, J. (2009). A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases. *J. Proteome Res.*, 8, 1792-1796.
- Ndiaye, M. M., Ta, H. P., Chiappetta, G., & Vinh, J. (2020). On-Chip Sample Preparation Using a ChipFilter Coupled to NanoLC-MS/MS for Bottom-Up Proteomics. *J Proteome Res*, 19(7), 2654-2663.
- Nebbak, A., Dahmana, H., Almeras, L., Raoult, D., Boulanger, N., Jaulhac, B., . . . Parola, P. (2019). Co-infection of bacteria and protozoan parasites in *Ixodes ricinus* nymphs collected in the Alsace region, France. *Ticks Tick Borne Dis*, 10(6), 101241.
- Nesvizhskii, A. (2007). Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. *Methods Mol Biol.*, 367, 87-119.
- Nesvizhskii, A. I., & Aebersold, R. (2005). Interpretation of Shotgun Proteomic Data. *Mol Cell Proteomics*, 4(10), 1419-1440.
- Nesvizhskii, A. I., Vitek, O., & Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods*, 4(10), 787-797.
- Norris, S. J. (2014). vls Antigenic Variation Systems of Lyme Disease *Borrelia*: Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity. *Microbiol Spectr*, 2(6).
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745.

- Obritsch, M. D., Fish, D. N., MacLaren, R., & Jung, R. (2012). Nosocomial Infections Due to Multidrug-Resistant *Pseudomonas aeruginosa*: Epidemiology and Treatment Option. *PHARMACOTHERAPY*, 25(10), 1353–1364.
- Old, W. M., Meyer-Arendt, K. M., Aveline-Wolf, L. A., Pierce, K. G., Mendoza, A., Sevinsky, J. R., . . . Ah, N. G. (2005). Comparison of Label-Free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol Cell Proteomics*, 4 (10), 1487-1502.
- Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., & Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*, 4(9), 709-712.
- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics*, 1(5), 376-386.
- Ong, S. E., & Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol*, 1(5), 252-262.
- Orwick-Rydmark, M., Arnold, T., & Linke, D. (2016). 8The Use of Detergents to Purify Membrane Protein. *Curr Protoc Protein Sci*, 84, 1-35.
- Parkins, M. D., Somayaji, R., & Waters, V. J. (2018). Epidemiology, Biology, and Impact of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. *Clin Microbiol Rev*, 31(4).
- Parola, P., & Raoult, D. (2001). Tick-borne bacterial diseases emerging in Europe. *Clin Microbiol Infect*, 7, 80-83.
- Perez-Eid, C., & Gilot, B. (1998). Les tiques : cycles, habitats, hôtes, rôle pathogène, lutte. *Médecine et Maladies Infectieuses*, 28(4), 335-343.
- Perez-Riverol, Y., Alpi, E., Wang, R., Hermjakob, H., & Vizcaino, J. A. (2015). Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics*, 15(5-6), 930-949.
- Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, 3551-3567.
- Perraud, Q., Moynié, L., Gasser, V., Munier, M., Godet, J., Hoegy, F., . . . Schalk, I. J. (2018). A Key Role for the Periplasmic PfeE Esterase in Iron Acquisition via the Siderophore Enterobactin in *Pseudomonas aeruginosa*. *ACS Chemical Biology*, 13(9), 2603-2614.
- Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol Cell Proteomics*, 11(11), 1475–1488.
- Picard, G., Lebert, D., Louwagie, M., Adrait, A., Huillet, C., Vandenesch, F., . . . Brun, V. (2012). PSAQ™ standards for accurate MS-based quantification of proteins: from the concept to biomedical applications. *Journal of Mass Spectrometry*, 47(10), 1353-1363.
- Picotti, P., & Aebersol, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future direction. *Nat Methods*, 9 (6), 555-566.
- Picotti, P., Aebersold, R., & Domon, B. (2007). The Implications of Proteolytic Background for Shotgun Proteomics. *Mol Cell Proteomics*, 6, 1589–1598.
- Popara, M., Villar, M., & de la Fuente, J. (2015). Proteomics Characterization of Tick-Host-Pathogen Interactions. *Methods Mol Biol.*, 1247, 513-527.
- Portillo, A., Santibanez, P., Palomar, A. M., Santibanez, S., & Oteo, J. A. (2018). 'Candidatus *Neoehrlichia mikurensis*' in Europe. *New Microbes New Infect*, 22, 30-36.
- Radolf, J. D., Caimano, M. J., Stevenson, B., & Hu, L. T. (2012). Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nat Rev Microbiol*, 10(2), 87-99.
- Raffetin, A., Saunier, A., Bouillier, K., Caraux-Paz, P., Eldin, C., Gallien, S., . . . Grillon, A. (2020). Unconventional diagnostic tests for Lyme borreliosis: a systematic review. *Clin Microbiol Infect*, 26(1), 51-59.

- Raileanu, C., Moutailler, S., Pavel, I., Porea, D., Mihalca, A. D., Savuta, G., & Vayssier-Taussat, M. (2017). *Borrelia* Diversity and Co-infection with Other Tick Borne Pathogens in Ticks. *Front Cell Infect Microbiol*, 7, 36.
- Ramamoorthi, N., Narasimhan, S., Pal, U., Bao, F., Yang, X. F., Fish, D., . . . Fikrig, E. (2005). The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature*, 436(7050), 573-577.
- Rar, V., & Golovljova, I. (2011). *Anaplasma*, *Ehrlichia*, and "Candidatus *Neoehrlichia*" bacteria: pathogenicity, biodiversity, and molecular genetic characteristics, a review. *Infect Genet Evol*, 11(8), 1842-1861.
- Razavi, B. M., & Fazly Bazzaz, B. S. (2019). A review and new insights to antimicrobial action of local anesthetics. *Eur J Clin Microbiol Infect Dis*, 38(6), 991-1002.
- Rebuffat, S. (2012). Microcins in action: amazing defence strategies of Enterobacteria. *Biochem. Soc. Trans.*, 40, 1456-1462.
- Rice Conlon, J. A., Mather, T. N., Tanner, P., Gallo, G., & Jacobson, R. H. (2000). Efficacy of a Nonadjuvanted, Outer Surface Protein A, Recombinant Vaccine in Dogs After Challenge by Ticks Naturally Infected with *Borrelia burgdorferi*. *Vet. Ther.*, 1(2), 96-107.
- Rifflet, A., Filali, S., Chenau, J., Simon, S., Fenaille, F., Junot, C., . . . Becher, F. (2019). Quantification of low abundance *Yersinia pestis* markers in dried blood spots by immuno-capture and quantitative high-resolution targeted mass spectrometry. *Eur J Mass Spectrom (Chichester)*, 25(3), 268-277.
- Rikihisa, Y. (2010). *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. *Nat Rev Microbiol*, 8(5), 328-339.
- Rikihisa, Y. (2011). Mechanisms of obligatory intracellular infection with *Anaplasma phagocytophilum*. *Clin Microbiol Rev*, 24(3), 469-489.
- Rizzoli, A., Silaghi, C., Obiegala, A., Rudolf, I., Hubalek, Z., Foldvari, G., . . . Kazimirova, M. (2014). *Ixodes ricinus* and Its Transmitted Pathogens in Urban and Peri-Urban Areas in Europe: New Hazards and Relevance for Public Health. *Front Public Health*, 2, 251.
- Robles, A., Fong, J., & Cervantes, J. (2018). [Borrelia Infection in Latin America]. *Rev Invest Clin*, 70(4), 158-163.
- Roepstorff, P., & Fohlman, J. (1984). Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectro*, 11(11), 601.
- Ronsein, G. E., Pamir, N., von Haller, P. D., Kim, D. S., Oda, M. N., Jarvik, G. P., . . . 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.
- Rose, C. M., Merrill, A. E., Bailey, D. J., Hebert, A. S., Westphall, M. S., & Coon, J. J. (2013). Neutron encoded labeling for peptide identification. *Anal Chem*, 85(10), 5129-5137.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., . . . Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 3(12), 1154-1169.
- Rupprecht, T. A., Manz, K. M., Fingerle, V., Lechner, C., Klein, M., Pfirrmann, M., & Koedel, U. (2018). Diagnostic value of cerebrospinal fluid CXCL13 for acute Lyme neuroborreliosis. A systematic review and meta-analysis. *Clin Microbiol Infect*, 24(12), 1234-1240.
- Savitski, M. M., Wilhelm, M., Hahne, H., Kuster, B., & Bantscheff, M. (2015). A Scalable Approach for Protein FalseDiscovery Rate Estimation in Large ProteomicData Sets. *Mol Cell Proteomics*, 14(9), 2394-2404.
- Savoldi, A., Carrara, E., Gladstone, B. P., Azzini, A. M., Gopel, S., & Tacconelli, E. (2019). Gross national income and antibiotic resistance in invasive isolates: analysis of the top-ranked antibiotic-resistant bacteria on the 2017 WHO priority list. *J Antimicrob Chemother*, 74(12), 3619-3625.

- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., . . . Ostell, J. (2020a). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 48, D9-D16.
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2020b). GenBank. *Nucleic Acids Res*, 48(D1), D84-D86.
- Schiffmann, C., Hansen, R., Baumann, S., Kublik, A., Nielsen, P. H., Adrian, L., . . . 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.
- Schilling, B., MacLean, B., Held, J. M., Sahu, A. K., Rardin, M. J., Sorensen, D. J., . . . Gibson, B. W. (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(20), 10222-10229.
- Schnell, G. (2014). Développement d'approches protéomiques pour l'étude de la borréliose de Lyme. *Thèse de Doctorat. Université de Strasbourg*.
- Schnell, G., Boeuf, A., Jaulhac, B., Boulanger, N., Collin, C., Barthel, C., . . . Ehret-Sabatier, L. (2015a). Proteomic analysis of three *Borrelia burgdorferi* sensu lato native species and disseminating clones: Relevance for Lyme vaccine design. *Proteomics*, 15, 1280–1290.
- Schnell, G., Boeuf, A., Westermann, B., Jaulhac, B., Lipsker, D., Carapito, C., . . . Ehret-Sabatier, L. (2015b). Discovery and Targeted Proteomics on Cutaneous Biopsies Infected by *Borrelia* to Investigate Lyme Disease. *Mol Cell Proteomics*, 14, 1254–1264.
- Schnittger, L., Rodriguez, A. E., Florin-Christensen, M., & Morrison, D. A. (2012). Babesia: a world emerging. *Infect Genet Evol*, 12(8), 1788-1809.
- Schuler, G. D., Epstein, J. A., Ohkawa, H., & Kans, J. A. (1996). Entrez: Molecular Biology Database and Retrieval System *Methods Enzymol.*, 266, 141-162.
- Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Kuster, B., & Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat Commun*, 11(1), 1548.
- Severo, M. S., Stephens, K. D., Kotsyfakis, M., & Pedra, J. H. (2012). Anaplasma phagocytophilum: deceptively simple or simply deceptive? *Future Microbiol*, 7(6), 719-731.
- Sherwood, C. A., Eastham, A., Lee, L. W., Risler, J., Mirzaei, H., Falkner, J. A., & Martin, D. B. (2009). Rapid optimization of MRM-MS instrument parameters by subtle alteration of precursor and product m/z targets. *J Proteome Res*, 8(7), 3746-3751.
- Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., & Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, 1(6), 2856-2860.
- Shi, T., Quek, S.-I., Gao, Y., Nicora, C. D., Nie, S., Fillmore, T. L., . . . Qian, W.-J. (2017). Multiplexed targeted mass spectrometry assays for prostate cancer-associated urinary proteins *Oncotarget*, 8, 101887-101898.
- Shiio, Y., & Aebersold, R. (2006). Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat Protoc*, 1(1), 139-145.
- Shin, J., Kwon, Y., Lee, S., Na, S., Hong, E. Y., Ju, S., . . . Lee, C. (2019). Common Repository of FBS Proteins (cRFP) To Be Added to a Search Database for Mass Spectrometric Analysis of Cell Secretome. *J Proteome Res*, 18(10), 3800-3806.
- Sidoli, S., & Garcia, B. A. (2017). Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev Proteomics*, 14(7), 617-626.
- Singh, S., Springer, M., Steen, J., Kirschner, M. W., & Steen, H. (2009). FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J Proteome Res*, 8(5), 2201-2210.

- Sleno, L., & Volmer, D. A. (2004). Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.*, 39, 1091-1112.
- Smith, S. M. (2017). Strategies for the Purification of Membrane Proteins. *Methods Mol Biol*, 1485, 389-400.
- Sohn, A., Kim, H., Yeo, I., Kim, Y., Son, M., Yu, S. J., . . . Kim, Y. (2018). Fully validated SRM-MS-based method for absolute quantification of PIVKA-II in human serum: Clinical applications for patients with HCC. *J Pharm Biomed Anal*, 156, 142-146.
- Speers, A. E., & Wu, C. C. (2007). Proteomics of Integral Membrane Proteins Theory and Application. *Chem. Rev.*, 107, 3687-3714.
- Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., & Domon, B. (2007). High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosite. *Mol Cell Proteomics*, 6, 1809-1817.
- Stanek, G., Fingerle, V., Hunfeld, K. P., Jaulhac, B., Kaiser, R., Krause, A., . . . Gray, J. (2011a). Lyme borreliosis: clinical case definitions for diagnosis and management in Europe. *Clin Microbiol Infect*, 17(1), 69-79.
- Stanek, G., & Reiter, M. (2011b). The expanding Lyme Borrelia complex--clinical significance of genomic species? *Clin Microbiol Infect*, 17(4), 487-493.
- Stanek, G., Wormser, G. P., Gray, J., & Strle, F. (2012). Lyme borreliosis. *The Lancet*, 379(9814), 461-473.
- Steere, A. C., Strle, F., Wormser, G. P., Hu, L. T., Branda, J. A., Hovius, J. W., . . . Mead, P. S. (2016). Lyme borreliosis. *Nat Rev Dis Primers*, 2, 16090.
- Stoesser, G., Baker, W., van den Broek, A., Garcia-Pastor, M., Kanz, C., Kulikova, T., . . . Vaughan, R. (2003). The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res*, 31(1), 17-22.
- Straubinger, R. K., Summers, B. A., Chang, Y. F., & Appel, M. J. (1997). Persistence of Borrelia burgdorferi in Experimentally Infected Dogs After Antibiotic Treatment. *J Clin Microbiol.*, 35(1), 111-116.
- Svitalkova, Z. H., Harustiakova, D., Mahrikova, L., Mojsova, M., Berthova, L., Slovak, M., . . . Kazimirova, M. (2016). Candidatus Neorhlichia mikurensis in ticks and rodents from urban and natural habitats of South-Western Slovakia. *Parasit Vectors*, 9, 2.
- Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., & Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*, 101(26), 9528-9533.
- Taoka, M., Fujii, M., Tsuchiya, M., Uekita, T., & Ichimura, T. (2017). A Sensitive Microbead-Based Organic Media-Assisted Method for Proteomics Sample Preparation from Dilute and Denaturing Solutions. *ACS Appl Mater Interfaces*, 9(49), 42661-42667.
- Thepparit, C., Bourchookarn, A., Petchampai, N., Barker, S. A., & Macaluso, K. R. (2010). Interaction of Rickettsia felis with histone H2B facilitates the infection of a tick cell line. *Microbiology*, 156(Pt 9), 2855-2863.
- Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., . . . Hamon, C. (2003). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal Chem*, 75(8), 1895-1904.
- Toby, T. K., Fornelli, L., & Kelleher, N. L. (2016). Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem*, 9(1), 499-519.
- Toledo, A., Olmeda, A. S., Escudero, R., Jado, I., Valcárcel, F., Casado-Nistal, M. A., . . . Anda, P. (2009). Tick-Borne Zoonotic Bacteria in Ticks Collected from Central Spain. *Am. J. Trop. Med. Hyg*, 81 (1), 67-74.
- Tsiatsiani, L., & Heck, A. J. (2015). Proteomics beyond trypsin. *FEBS J*, 282(14), 2612-2626.
- Tubaon, R. M., Haddad, P. R., & Quirino, J. P. (2017). Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics*, 17(20), 1-27.

- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., . . . Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*, 13(9), 731-740.
- UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 47(D1), D506-D515.
- Unwin, R. D., Griffiths, J. R., Leverentz, M. K., Grallert, A., Hagan, I. M., & Whetton, A. D. (2005). Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics*, 4(8), 1134-1144.
- Vandermarliere, E., Mueller, M., & Martens, L. (2013). Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev*, 32(6), 453-465.
- Vaulont, S., & Isabelle Schalk, I. (2015). Roles of bacterial and mammalian siderophores in host-pathogen interactions. *Med Sci (Paris)*, 31, 756-763.
- Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., & Yate, J. R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods*, 1 (1), 39-45.
- Verheggen, K., Raeder, H., Berven, F. S., Martens, L., Barsnes, H., & Vaudel, M. (2020). Anatomy and evolution of database search engines-a central component of mass spectrometry based proteomic workflows. *Mass Spectrom Rev*, 39(3), 292-306.
- Villar, M., Ayllon, N., Kocan, K. M., Bonzon-Kulichenko, E., Alberdi, P., Blouin, E. F., . . . de la Fuente, J. (2015). Identification and Characterization of *Anaplasma phagocytophilum* Proteins Involved in Infection of the Tick Vector, *Ixodes scapularis*. *PLoS One*, 10(9), 1-26.
- Wee, S., Alli-Shaika, A., Kek, R., Swa, H. L. F., Tien, W.-P., Lim, V. W., . . . Gunaratnea, J. (2019). Multiplex targeted mass spectrometry assay for one-shot flavivirus diagnosis. *Proc. Natl. Acad. Sci*, 116, 6754-6759.
- Westermann, B. (2016). Développement de méthodologies protéomiques pour l'étude des maladies infectieuses. *Thèse de Doctorat. Université de Strasbourg*.
- Wieczorek, S., Combes, F., Lazar, C., Giai Gianetto, Q., Gatto, L., Dorffer, A., . . . Burger, T. (2017). DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics*, 33(1), 135-136.
- Williamson, J. C., Scheipers, P., Schwammle, V., Zibert, J. R., Beck, H. C., & Jensen, O. N. (2013). A proteomics approach to the identification of biomarkers for psoriasis utilising keratome biopsy. *J Proteomics*, 94, 176-185.
- Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat Methods*, 6(5), 359-362.
- WO2015022470A2, B. N. (2015). Brevet N° WO2015022470A2. 1-47.
- Wolters, D. A., Washburn, M. P., & Yates, J. R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*, 73, 5863-5890.
- Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., & Yates, J. R. (2004). Metabolic Labeling of Mammalian Organisms With Stable Isotopes for Quantitative Proteomic Analysis. *Anal Chem*, 76(17), 4951-4959.
- Wu, C. H., Yeh, L. S., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., . . . Barker, W. C. (2003). The Protein Information Resource. *Nucleic Acids Res*, 31(1), 345-347.
- Wysocki, V. H., Tsaprailis, G., Smith, L. L., & Brechi, L. A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.*, 35, 1399-1406.
- Xiang, F., Ye, H., Chen, R., Fu, Q., & Li, L. (2010). N,N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Anal Chem*, 82(7), 2817-2825.
- Yao, X., Freas, A., Ramirez, J., Demirev, P. A., & Fenselau, C. (2001). Proteolytic ¹⁸O Labeling for Comparative Proteomics: Model Studies With Two Serotypes of Adenovirus. *Anal Chem*, 73(13), 2836-2842.

- Yocum, A. K., Khan, A. P., Zhao, R., & Chinnaiyan, A. M. (2010). Development of selected reaction monitoring-MS methodology to measure peptide biomarkers in prostate cancer. *Proteomics*, 10(19), 3506-3514.
- Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., & Mann, M. (2012). A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol Cell Proteomics*, 11(3), O111 009613.
- Zhang, J., Wang, Y., & Li, S. (2010). Deuterium Isobaric Amine-Reactive Tags for Quantitative Proteomics. *Anal Chem*, 82(18), 7588-7595.
- Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R., 3rd. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*, 113(4), 2343-2394.
- Zhang, Y., Wen, Z., Washburn, M. P., & Florens, L. (2009). Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal Chem*, 81, 6317–6326.
- Zhang, Y., Wen, Z., Washburn, M. P., & Florens, L. (2010). Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Anal. Chem.*, 82, 2272–2281.
- Zhang, Y., Wen, Z., Washburn, M. P., & Florens, L. (2019). Evaluating Chromatographic Approaches for the Quantitative Analysis of a Human Proteome on Orbitrap-Based Mass Spectrometry Systems. *J Proteome Res*, 18(4), 1857-1869.
- Zhang, Z., Wu, S., Stenoien, D. L., & Pasa-Tolic, L. (2014). High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)*, 7, 427-454.
- Zhou, Y., Qin, S., Sun, M., Tang, L., Yan, X., Kim, T. K., . . . Hood, L. (2020). Measurement of Organ-Specific and Acute-Phase Blood Protein Levels in Early Lyme Disease. *J Proteome Res*, 19(1), 346-359.
- Zougman, A., Selby, P. J., & Banks, R. E. (2014). Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics*, 14(9), 1006-1000.
- Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., . . . McLafferty, F. W. (2000). Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cation. *Anal. Chem*, 72, 563-573.

Annexes

Annexe 1 : Liste des protéines et peptides utilisés dans les méthodes de quantification ciblée.

Méthode 1 (M1) utilisée pour l'approche vaccinale contre la borréliose de Lyme

Protéine	Peptide	AQUA	Tag	Fournisseur
BB0758	SFDLIHK	Oui	Oui	JPT Technologies
	NVLDLINNR	Non	Non	JPT Technologies
	VLSVMQDSCDLK	Non	Non	JPT Technologies
	FENNHYGIVIYNLSK	Non	Non	JPT Technologies
	SIANLQISEQILYLK	Non	Non	JPT Technologies
	VFSNFFAR	Non	Non	JPT Technologies
DbpA	DITDEIDAIK	Non	Non	JPT Technologies
	GVNFDAFK	Non	Non	JPT Technologies
	TGSGVSENPFILEAK	Oui	Oui	JPT Technologies
	LGIQEMTK	Non	Non	JPT Technologies
	ENSTFTDEK	Non	Non	JPT Technologies
Flagelline	MIINHNTSAINASR	Non	Non	JPT Technologies
	NNGINAANLSK	Non	Non	JPT Technologies
	ASDDAAGMGVSGK	Non	Non	JPT Technologies
	AINFIQTTEGNLNEVEK	Oui	Non	Thermo Fisher Scientific
	ELAVQSGNGTYSADR	Non	Non	JPT Technologies
	GSIQIEIEQLTDEINR	Oui	Oui	JPT Technologies
	IADQAQYNQMHLNSK	Non	Non	JPT Technologies
	TAEELGMQPAK	Non	Non	JPT Technologies
	INTPASLSGSQASWTLR	Oui	Non	Thermo Fisher Scientific
	ANLGAFQNR	Oui	Non	Thermo Fisher Scientific
	DSTEYAIENLK	Non	Non	JPT Technologies
	NNAINAANLSK	Non	Non	JPT Technologies
	ELAVQSGNGTYSDDSR	Non	Non	JPT Technologies
	NSTEYAIENLK	Non	Non	JPT Technologies
Lipoprotéine gi 365823350	EISLTPEEAELSLK	Oui	Oui	JPT Technologies
	EISLTPEEAELK	Non	Non	JPT Technologies
	DAMSVNGR	Non	Non	JPT Technologies
	EFISSFDNISSIVSK	Non	Non	JPT Technologies
	RPTEQQSLGFK	Non	Non	JPT Technologies
	GEALSLFFQK	Oui	Oui	JPT Technologies
	VVDAFGADPYK	Non	Non	JPT Technologies
	DNDESQKPVK	Non	Non	JPT Technologies
OspC	VLTESESNNELK	Non	Non	JPT Technologies
	AILITDAK	Non	Non	JPT Technologies
	AVENLAK	Non	Non	JPT Technologies
	ITDSNAFVLAVK	Oui	Non	Thermo Fisher Scientific
	GNPLTEISK	Oui	Non	Thermo Fisher Scientific
	ITDSNAVLLAVK	Non	Non	JPT Technologies
	EVEALLSSIDEIAAK	Non	Non	JPT Technologies
	CSETFTNK	Non	Non	JPT Technologies
	LFESVEVLK	Non	Non	JPT Technologies
	EMLANSVK	Non	Non	JPT Technologies
	ELTSPVVAESPK	Non	Non	JPT Technologies
	AAQVALTNSVK	Non	Non	JPT Technologies
	EVETLLASIDEVAK	Non	Non	JPT Technologies
	AAQETLNNSVK	Non	Non	JPT Technologies
PGP	FVIDTLK	Oui	Oui	JPT Technologies
	LLSLEHDNPNLQEK	Non	Non	JPT Technologies
	NLSFQTKPYENIKPLLETMNK	Non	Non	JPT Technologies
	LNIPIGILSNK	Non	Non	JPT Technologies
	NHEELINLVK	Non	Non	JPT Technologies

Méthode 2 (M2) utilisée pour l'approche vaccinale contre de la borréliose de Lyme

Protéine	Peptide	AQUA	Tag	Fournisseur
Borrelia REV	QIDSLSTGVSTLNNK	Non	Non	JPT Technologies
	DVGNAELK	Non	Non	JPT Technologies
	IEGTTDADNNTAK	Non	Non	JPT Technologies
	EAVEQIDEIHK	Non	Non	JPT Technologies
lipoprotéine P27	IISVASIQR	Non	Non	JPT Technologies
	ASNQNPISIK	Non	Non	JPT Technologies
	EHEFQQAEDK	Non	Non	JPT Technologies
	VESLSILINK	Non	Non	JPT Technologies
	DYVLLGLK	Non	Non	JPT Technologies
	GVGEQYANPIYR	Non	Non	JPT Technologies
	DYHSIDYNK	Non	Non	JPT Technologies
	ENSIELDLLPHNQIFK	Non	Non	JPT Technologies
	ILEILTDK	Oui	Oui	JPT Technologies
Lipoprotéine gj 365823350	DAMSVNGR	Non	Non	JPT Technologies
	SLTEIDSGNGIPLVSDVVK	Oui	Oui	JPT Technologies
	EISLTPEEA EK	Non	Non	JPT Technologies
	EFFDWLSK	Non	Non	JPT Technologies
	EFISSFDNISSIVSK	Non	Non	JPT Technologies
	RPTEQQSLGFK	Non	Non	JPT Technologies
	GEALSLFFQK	Oui	Oui	JPT Technologies
	VVDAFGADPYK	Non	Non	JPT Technologies
	DNDESVQKPVK	Non	Non	JPT Technologies
OppB	VLTESESNNELK	Non	Non	JPT Technologies
	GDLGPSLK	Non	Non	JPT Technologies
	DLTVNQYIK	Oui	Oui	JPT Technologies
	WGLLYTSGWITER	Non	Non	JPT Technologies

Méthode 4 (M4) utilisée dans l'approche diagnostic de la phase précoce de la borréliose de Lyme

Protéine	Peptide	AQUA	Tag	Fournisseur
DbpA	DITDEIDAIK	Non	Non	Thermo Fisher Scientific
	GVNFDAFK	Non	Non	Thermo Fisher Scientific
	TTANGIIEIVK	Non	Non	Thermo Fisher Scientific
	TVTDAAEQHPTTTAEGILEIAK	Non	Non	Thermo Fisher Scientific
	VSENSFILEAK	Non	Non	Thermo Fisher Scientific
Enolase	IQLVGDDLFTNTSFLK	Non	Non	Thermo Fisher Scientific
	SGETEDTTIADLVVALGTGQIK	Non	Non	Thermo Fisher Scientific
Flagelline	AINFIQTTEGNLNEVEK	Oui	Non	Thermo Fisher Scientific
	ANLGAFQNR	Oui	Non	Thermo Fisher Scientific
	ASDDAAGMGVSGK	Non	Non	Thermo Fisher Scientific
	ELAVQSGNGTYSDDSR	Non	Non	Thermo Fisher Scientific
	GSIQIEIEQLTDEINR	Non	Non	Thermo Fisher Scientific
	IADQAQYNQMHLSNK	Non	Non	Thermo Fisher Scientific
	INTPASLSGSQASWTLR	Oui	Non	Thermo Fisher Scientific
	MIINHNTSAINASR	Non	Non	Thermo Fisher Scientific
	NNAINAANLSK	Non	Non	Thermo Fisher Scientific
	NNGINAANLSK	Non	Non	Thermo Fisher Scientific
	NSTEYAIENLK	Non	Non	Thermo Fisher Scientific
	TAEELGMQPAK	Non	Non	Thermo Fisher Scientific
GADPH	GGYLDHVNHAGAK	Non	Non	Thermo Fisher Scientific
	GIDIVAINDLTDPK	Non	Non	Thermo Fisher Scientific
Lipoprotéine gi 365823350	EFFDWLSK	Non	Non	Thermo Fisher Scientific
	GEALSLFFQK	Non	Non	Thermo Fisher Scientific
	SLTEIDSGNGIPLVSDVVK	Non	Non	Thermo Fisher Scientific
	VLTESESNNELK	Non	Non	Thermo Fisher Scientific
OspC	AAQAALNSVK	Non	Non	Thermo Fisher Scientific
	AAQVALNSVK	Non	Non	Thermo Fisher Scientific
	AILITDAAK	Non	Non	Thermo Fisher Scientific
	AQLGVQNGGATNDHAK	Non	Non	Thermo Fisher Scientific
	AVENLAK	Non	Non	Thermo Fisher Scientific
	AVENLSK	Non	Non	Thermo Fisher Scientific
	DGNASANSADSVK	Non	Non	Thermo Fisher Scientific
	DLFESVEGLLK	Non	Non	Thermo Fisher Scientific
	DSHAELGIQSVQDDNAK	Non	Non	Thermo Fisher Scientific
	EALSNSVK	Non	Non	Thermo Fisher Scientific
	ELTNPVVAETPK	Non	Non	Thermo Fisher Scientific
	ELTSPVVAEAPK	Non	Non	Thermo Fisher Scientific
	ELTSPVVAESPK	Non	Non	Thermo Fisher Scientific
	ELTSPVVVESPK	Non	Non	Thermo Fisher Scientific
	EVEALLSSIDEIAAK	Non	Non	Thermo Fisher Scientific
	EVEALLSSIDELSK	Non	Non	Thermo Fisher Scientific
	EVETLLASIDELATK	Non	Non	Thermo Fisher Scientific
	EVETLVLSIDELAK	Non	Non	Thermo Fisher Scientific
	GGDSASTNPADESAK	Non	Non	Thermo Fisher Scientific
	GPNLTEISK	Oui	Non	Thermo Fisher Scientific
	GPNLTVISK	Non	Non	Thermo Fisher Scientific
	GVEELEK	Non	Non	Thermo Fisher Scientific
	IGNNGLEANQSK	Non	Non	Thermo Fisher Scientific
	ITDSNAFLAVK	Non	Non	Thermo Fisher Scientific
	ITDSNAFVLAVK	Oui	Non	Thermo Fisher Scientific
	ITDSNAVLLAVK	Non	Non	Thermo Fisher Scientific
	LFESVEVLSK	Non	Non	Thermo Fisher Scientific
	LSGSLESLSK	Non	Non	Thermo Fisher Scientific
	LSVLNSEEK	Non	Non	Thermo Fisher Scientific
	NESLIAGAYEISK	Non	Non	Thermo Fisher Scientific
	NTSLLAGAYSISTLITK	Non	Non	Thermo Fisher Scientific

Méthode 6 (M6) utilisée dans l'approche diagnostic de la phase tardive disséminée de la borréliose de Lyme

Protéine	Peptide	AQUA	Tag	Fournisseur
BAPKO0593	LPLALNLAVSR	Oui	Oui	JPT Technologies
BAPKO4515	ALINLNFNQLLEK	Oui	Oui	JPT Technologies
DbpA	TTANGIIEIVK	Oui	Oui	JPT Technologies
	DITDEIDAIK	Non	Non	Thermo Fisher Scientific
	GVNFDAFK	Non	Non	Thermo Fisher Scientific
	VSENSFILEAK	Non	Non	Thermo Fisher Scientific
	TVTDAAEQHPTTTAEGILEIAK	Non	Non	Thermo Fisher Scientific
Enolase	SGETEDTTIADLVVALGTGQIK	Oui	Oui	JPT Technologies
Flagelline	MIINHNTSAINASR	Non	Non	Thermo Fisher Scientific
	NNAINAANLSK	Non	Non	Thermo Fisher Scientific
	ASDDAAGMGVSGK	Non	Non	Thermo Fisher Scientific
	AINFIQTTEGNLNEVEK	Oui	Non	Thermo Fisher Scientific
	ELAVQSGNGTYSDDSR	Non	Non	Thermo Fisher Scientific
	GSIQIEIEQLTDEINR	Non	Non	Thermo Fisher Scientific
	IADQAQYNQMHLMSNK	Non	Non	Thermo Fisher Scientific
	TAEELGMQPAK	Non	Non	Thermo Fisher Scientific
	INTPASLSGSQASWTLR	Oui	Non	Thermo Fisher Scientific
	ANLGAFQNR	Oui	Non	Thermo Fisher Scientific
	NSTEYAIENLK	Non	Non	Thermo Fisher Scientific
	NNGINAANLSK	Non	Non	Thermo Fisher Scientific
GAPDH	GIDIVAINDLTDPK	Oui	Oui	JPT Technologies
	GGYLDHVNHAGAK	Non	Non	Thermo Fisher Scientific
	VPVPTGSIVDLTVQLK	Non	Non	Thermo Fisher Scientific
Histidine kinase	FAILDFLYSK	Oui	Oui	JPT Technologies
BB081	FNISGIIETGLK	Oui	Oui	JPT Technologies
BBP42	LTNSSINFDK	Oui	Oui	JPT Technologies
HSP90	YSNHINYPIYIK	Oui	Oui	JPT Technologies
Lipoprotéine gi 365823350	GGALSLFFQK	Non	Non	JPT Technologies
	VFTESDSNNELK	Non	Non	JPT Technologies
	SLTEIDSGNGIPLVSDVVK	Non	Non	Thermo Fisher Scientific
	GEALSLFFQK	Non	Non	Thermo Fisher Scientific
	VLTESESNNELK	Non	Non	Thermo Fisher Scientific
RNA Polymerase	LNPNTLEFK	Oui	Oui	JPT Technologies
VLsE	AAEEAIVGATGDGK	Oui	Oui	JPT Technologies
	IGESADNGAAADADSVK	Non	Non	JPT Technologies
	AAEAVSSVSQEILK	Non	Non	Thermo Fisher Scientific
	GIVDAAGTAAGK	Oui	Oui	JPT Technologies
	GIVDAAGTAAGKK	Non	Non	Thermo Fisher Scientific
	AAAAVSSVSQEILK	Non	Non	Thermo Fisher Scientific
	AVSSVSQEILK	Non	Non	JPT Technologies
	VAAALVLR	Non	Non	JPT Technologies
	IVAAIVLR	Non	Non	Thermo Fisher Scientific
	AVTAVSQEILK	Non	Non	JPT Technologies
	IGDSANQGAVADADSVK	Non	Non	Thermo Fisher Scientific
	GIVDAAGK	Non	Non	Thermo Fisher Scientific
	AAGAVSAVSQEILK	Non	Non	Thermo Fisher Scientific
	AGAVSAVSQEILK	Non	Non	JPT Technologies
	AVSAVSQEILK	Non	Non	Thermo Fisher Scientific
	IAAAIVLR	Non	Non	Thermo Fisher Scientific
	AAAAVTAVSQEILK	Oui	Oui	JPT Technologies
	IGDVGADNK	Non	Non	JPT Technologies
	GIVAAAGK	Non	Non	Thermo Fisher Scientific
	AAGAVTAVSQEILK	Non	Non	Thermo Fisher Scientific

Annexe 2 : Liste des protéines et peptides détectés dans les analyses LC-SRM et LC-PRM pour les patients de la cohorte (phase précoce de la borréliose de Lyme)

*Lipo = Lipoprotéine gj|365823350

sans culture	SRM	PRM	SRM et PRM	Non proteotypique																																					
Protéines	OspC																			Flagellin										DbpA			GAPDH		Enolase	Lipo*					
	ITDSNAFLAVK	ALITDAAK	LSGSLESLSK	ELTNPVAETPK	AVENLSK	AAQAALTNSVK	EVETLLASIDELATK	ITDSNAFLAVK	LSVLNSEELK	LFESVEVLSK	IGNNGLEANQSK	GPNLTEISK	AAQVALTNSVK	DLFESVEGLLK	ELTSPVVAESPK	EVEALLSIDELSK	EVETLVLSIDELAK	GGDSASTNPADSAK	GPNLTVISK	NESLIAGAYEISK	AINFIQTTEGNLNEVEK	NNGINAANLSK	TAEELGMOPAK	GSIQIEIQLTDEINR	ANLGAFQNR	ASDDAAGMGVSGK	ELAVQSGNGTYSDSR	INTPASLGSQASWTLR	MIINHNTSAINASR	NNAINAANLSK	IADQAQYNQMHMLSNK	NSTEYAIENLK	TTANGIEVK	DITDEIAIK	GVNFDAFK	TVTDAAEQHPHTTAEGLIEIAK	GGYLDHVNHAGAK	GIDIVANDLTPK	SGETEDTTIADLWALGTGQIK	IQLVGDDLEVTNTSFLK	EFFDWLSK
003-044	X											X		X	X			X			X		X	X	X	X	X	X	X	X	X					X		X	X	X	
003-027	X											X	X	X	X		X	X			X		X		X	X	X	X	X	X	X					X					
003-043	X											X						X			X				X		X	X	X	X	X					X					
002-003	X											X						X			X	X	X	X	X	X	X	X	X	X	X					X	X				
003-028	X											X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X					X					
003-031				X		X		X	X			X		X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X										
001-020	X											X								X	X	X	X	X	X	X	X	X	X	X	X										
001-023	X								X			X								X	X	X	X	X	X	X	X	X	X	X	X										
003-001				X						X		X									X												X	X							
003-013					X															X													X	X							
003-024			X																	X			X										X	X							
002-001				X																X			X											X	X						
002-002	X	X																					X																		
003-009										X														X																	
003-014				X				X												X																					
003-015								X					X																												
003-017	X					X						X																													
003-022						X					X																														
003-032																																		X	X	X	X				
003-036							X																																		
003-037							X																																		
003-040																																									
005-001																																									
005-004																																									
001-005																					X																				
001-007																						X																			
001-010																																									
001-012			X										X																												
001-019		X																																							
003-002						X																																			
003-011							X																																		
003-012																																									
003-019																																									
003-025									X																																
003-026																																									

[illegible]

Annexe 3 : Tableau récapitulatif des banques de données utilisées dans le chapitre III

Banque de données	Application
A	Souris infectées par inoculation avec <i>A. phagocytophilum</i> (2016)
B	Tiques
C	Souris infectées via tiques
D	Souris infectées par inoculation avec <i>A. phagocytophilum</i> (2018)
E	Souris infectées par inoculation avec <i>B. microti</i>

Développement d'approches protéomiques pour l'identification, la détection spécifique et la quantification de pathogènes

Résumé

L'objectif était de développer de nouvelles approches protéomiques pour l'étude, la prévention ou le diagnostic de maladies infectieuses : borréliose de Lyme (LB), maladies vectorielles à tiques, infections par *Pseudomonas aeruginosa*. Pour la LB, nous avons identifié des candidats-vaccin, qui ont pu ensuite être testés dans un essai vaccinal chez le chien. Nous avons validé sur une cohorte humaine la protéomique ciblée comme méthode diagnostique de la phase précoce de la LB. L'extension au diagnostic de la phase disséminée est en cours. L'approche de protéomique non ciblée a été étendue aux autres maladies vectorielles à tiques, et a permis d'identifier des marqueurs d'infection active. Enfin, des approches différentielles quantitatives ont permis de comprendre l'adaptation de *P. aeruginosa* à son environnement, et comment les voies d'acquisition du fer peuvent être exploitées pour développer de nouveaux antibiotiques.

Mots-clefs : Protéomique non ciblée - Protéomique ciblée – Pathogènes

Summary

The aim was to develop new proteomic approaches for the study, the prevention or the diagnosis of infectious diseases: Lyme borreliosis (LB), tick-borne diseases (TBD), *Pseudomonas aeruginosa* infections. In case of LB, we identified vaccine-candidates which have been further evaluated in a vaccine assay in dogs. In a human cohort, we validated targeted proteomics for the diagnosis of the early phase of LB. Application for the diagnosis of the disseminated phase is in progress. For the extension to other TBD, discovery non targeted approaches allowed the identification of active infection markers. Finally, label free approaches highlighted the phenotypical adaptation of *P. aeruginosa* to its environment, and how iron acquisition pathways can be exploited to develop new antibiotics.

Keywords: Non targeted proteomics - Targeted proteomics – Pathogens